Infection, Genetics and Evolution 31 (2015) 221–230

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

In-vineyard population structure of '*Candidatus* Phytoplasma solani' using multilocus sequence typing analysis

Sergio Murolo, Gianfranco Romanazzi *

Department of Agricultural, Food and Environmental Sciences, Marche Polytechnic University, Ancona, Italy

ARTICLE INFO

Article history:
Received 6 September 2014
Received in revised form 28 January 2015
Accepted 29 January 2015
Available online 4 February 2015

Keywords:
Genetic variability
Molecular typing
Selective pressure
Stolbur
Surface membrane protein

ABSTRACT

'*Candidatus* Phytoplasma solani' is a phytoplasma of the stolbur group (16SrXII subgroup A) that is associated with 'Bois noir' and causes heavy damage to the quality and quantity of grapevine yields in several European countries, and particularly in the Mediterranean area. Analysis of '*Ca.* P. solani' genetic diversity was carried out for strains infecting a cv. 'Chardonnay' vineyard, through multilocus sequence typing analysis for the *vmp*1, *stamp* and *secY* genes. Several types per gene were detected: seven out of 20 types for *vmp*1, six out of 17 for *stamp*, and four out of 16 for *secY*. High correlations were seen among the *vmp*1, *stamp* and *secY* typing with the *tuf* typing. However, no correlations were seen among the *tuf* and *vmp*1 types and the Bois noir severity in the surveyed grapevines. Grouping the '*Ca.* P. solani' sequences on the basis of their origins (i.e., study vineyard, Italian regions, Euro-Mediterranean countries), dN/dS ratio analysis revealed overall positive selection for *stamp* (3.99, $P = 0.019$) and *vmp*1 (2.28, $P = 0.001$). For *secY*, the dN/dS ratio was 1.02 ($P = 0.841$), showing neutral selection across this gene. Using analysis of the nucleotide sequencing by a Bayesian approach, we determined the population structure of '*Ca.* P. solani', which appears to be structured in 3, 5 and 6 subpopulations, according to the *secY*, *stamp* and *vmp*1 genes, respectively. The high genetic diversity of '*Ca.* P. solani' from a single vineyard reflects the population structure across wider geographical scales. This information is useful to trace inoculum source and movement of pathogen strains at the local level and over long distances.

1. Introduction

Bois noir (BN) is a grapevine disease that is associated to '*Candidatus* Phytoplasma solani' ('*Ca.* P. solani'; 16SrXII-A subgroup) (Quaglino et al., 2013), and it is most common and widespread in Euro-Mediterranean regions (Maixner, 2011; Zahavi et al., 2013). In recent years, rapid spreading of BN has been seen frequently, and this can be a limiting factor in several grapevine-growing regions (Maixner, 2011; Zahavi et al., 2013). BN spreading is influenced by the biology of the main vector, *Hyalesthes obsoletus*, together with other potential vectors (Imo et al., 2013; Landi et al., 2013; Cvrković et al., 2014), and their wide range of host plants (Kessler et al., 2011; Johannesen et al., 2012; Riolo et al., 2012; Cvrković et al., 2014). Disease management for BN is very difficult, as the phytoplasma and insect vectors are both not host specific, and thus their control is mainly based on agronomic approaches. Innovative perspectives to mitigate disease spread and severity have been based on the induction of host defences (Romanazzi et al., 2013). In cv. 'Chardonnay', which is particularly sensitive to BN infection, the drying up of grape bunches can result in production losses of about 50%, with lower sugar content in the grapes of symptomatic plants (Endeshaw et al., 2012).

'*Ca.* P. solani' isolates are characterised by different degrees of genetic variability according to the genes involved (Quaglino et al., 2009, 2013; Foissac et al., 2013). The most variable genes are those that code for surface membrane proteins, which have been sequenced for several phytoplasma, and have been classified into three types: immunodominant membrane protein (Imp); immunodominant membrane protein A (ImpA); and antigenic membrane protein (Amp) (Kakizawa et al., 2006a). Several genes that code for surface membrane proteins in '*Ca.* P. solani' have been sequenced and characterised (Cimerman et al., 2009; Fabre et al., 2011). In particular, in samples from the Euro-Mediterranean basin (e.g., vectors, different hosts), 23 restriction fragment-length polymorphism (RFLP) genotypes have been recorded in the gene coding for variable membrane protein-1 (*vmp*1), 14 in the gene coding for antigenic membrane protein (*stamp*), and 39 in the *secY* gene (Fabre et al., 2011; Foissac et al., 2013).

The study of phytoplasma population genetics can be based on genotyping by multilocus sequence typing analysis (Arnaud et al., 2007; Danet et al., 2011), and on the estimation of the dN/dS ratio,

* Corresponding author. Tel.: +39 071 2204336; fax: +39 071 2204856.
E-mail address: g.romanazzi@univpm.it (G. Romanazzi).

http://dx.doi.org/10.1016/j.meegid.2015.01.028
1567-1348/© 2015 Elsevier B.V. All rights reserved.

which is the ratio between the non-synonymous (dN) and the synonymous (dS) substitution rates in an alignment of amino-acid-coding sequences (Nielsen, 2005). Synonymous substitutions are usually regarded as having much smaller effects on fitness than non-synonymous substitutions. This ratio has been used to identify the types of selection in genes; i.e., dN/dS >1, =1 and <1, for positive, neutral and negative selection, respectively (Nielsen, 2005). This approach can reveal useful information on the molecular ecology, to trace the route of spread of phytoplasma strains, so as to better understand the epidemiology of phytoplasma diseases. Moreover, it can be helpful to hypothesise natural events that can provide pressure of selection, particularly for those microrganisms for which phenotypic differentiation is complicated by the difficulty of cultivation.

The aims of this study were: (i) to determine the genetic variability and strain composition of '*Ca.* P. solani' in the investigated vineyard, using multilocus sequence typing analysis of the *vmp*1, *stamp* and *secY* genes; (ii) to estimate the selection (i.e., positive, negative, neutral) on the same genes moving from the small scale (vineyard), through the national scale (Italian regions), to the international scale (Euro-Mediterranean countries); and (iii) to determine the population structure of '*Ca.* P. solani', considering the sequencing related to the grapevine and the main vectors in the vineyard.

2. Materials and methods

2.1. Sample sources

The present study included part of a DNA collection (329 samples) that was obtained from a total of 357 infected grapevines that were sampled in 2011. This collection was carried out in a commercial vineyard in central-eastern Italy (42° 59′ 00″ N, 13° 36′ 00″ E), with the infection previously assessed by molecular tools to belong to '*Ca.* P. solani' (Murolo et al., 2014). Based on the information of BN severity of each infected vine recorded according to an empirical scale, as proposed in a previous study (Murolo et al., 2014), the disease severity that occurred in the vineyard and that was induced by different strains of '*Ca.* Phytoplasma solani' was calculated according to the formula $S = \sum(c \times f)/n$, where c is the severity class, f is the frequency of the class, and n is the number of symptomatic plants.

2.2. PCR and molecular characterisation of '*Ca.* P. solani' for the *tuf* gene using RFLP

The 329 DNA samples were amplified in nested PCR with tufAYf/r (Schneider et al., 1997), then molecular characterisation of the *tuf* gene was carried out by PCR-RFLP, according to the protocol described by Langer and Maixner (2004).

Moreover, the same samples were amplified with additional primer pairs designed on the *stamp* and *secY* regions of the '*Ca.* P. solani' genome, as available in the literature (Fialova et al., 2009; Fabre et al., 2011).

2.3. Multilocus sequence typing analysis

On the basis of the RFLP characterisation of the *tuf* gene (the present study) and the *vmp*1 gene (Murolo et al., 2014), representative samples amplified with specific primer pairs for the *vmp*1, *stamp* and *secY* genes were purified using Wizard SV gels and PCR Clean-Up kits (Promega Corporation, Madison, WI, USA), and quantified using a BioPhotometer (Eppendorf, Hamburg, Germany). These samples were sequenced using an ABI 3730XL DNA sequencer (Applied Biosystems, Carlsbad, CA, USA) at Beckman Coulter Genomics (Essex, UK), edited using the Chromas version 2.33 software, and assembled using the GAP4 Staden Package (http://www.staden.sourceforge.net), to obtain a consensus sequence. The Bioedit software, v. 7.0.0 (http://www.mbio.ncsu.edu/Bioedit/bioedit.html) was used to cut off ∼20–30 bp of the terminal end sequence. All of the consensus sequences for each gene were used as the query sequences in the BLAST searches (http://blast.ncbi.nlm.nih.gov/Blast.cgi), to determine the nucleotide identities for the '*Ca.* P. solani' strains available in GeneBank NCBI (http://www.ncbi.nlm.nih.gov).

Fifteen *vmp*1, 17 *stamp*, and 13 *secY* nucleotide sequences from the samples of the study vineyard (Table 1) were aligned with the nucleotide sequences available in Genebank, with the selection of '*Ca.* P. solani' strains from the same host (grapevine) and from vectors and potential vectors only from the vineyard (*H. obsoletus*, *Reptalus panzeri, R. quinquecostatus*) (Table 1). The multiple alignment was carried out using the Clustal X v. 1.8 programme (Thompson et al., 1997).

2.4. Phylogenetic analysis

Distance-based trees were produced based on distance matrices generated by the Jukes–Cantor distance correction method (Jukes and Cantor, 1969), followed by tree construction using the Mega v. 5.1 software (Tamura et al., 2011). The percentage of replicate trees in which the associated strains clustered together in the bootstrap test (1000 replicates) was estimated, as shown next to the tree branches. The trees were drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic trees. All of the positions containing gaps and missing data were eliminated from the dataset. The nucleotide sequences were analysed with DnaSP version 5 (Librado and Rozas, 2009), to distinguish between the different types for each gene, to assess the '*Ca.* P. solani' strain composition in the context of the study vineyard, and that of the national and international phytoplasma infections of grapevine.

2.5. Genetic analysis and selective pressure of '*Ca.* P. solani' from the study vineyard to the Euro-Mediterranean countries

DnaSP version 5 was used to estimate: number of polymorphic sites (S) (singletons and parsimony-informative sites), total number of mutations (η), number of molecular types (Hn), molecular type diversity (Hd), average number of nucleotide differences between sequences from the same population (k), and average pairwise nucleotide diversity (Pi) (Tajima, 1983).

The ratio between the proportion of non-synonymous and synonymous substitutions, known as the dN/dS ratio, was determined for the nucleotide sequences at the study vineyard level, with respect to the national (Italian regions) and international (other Euro-Mediterranean countries) contexts. Additionally, the dS and dN variances were studied, as Var(dS) and Var(dN), respectively. The dN/dS ratios and the null hypothesis of no selection (H0: dN = dS) *versus* the positive selection hypothesis (H1: dN > dS) were calculated using the Nei–Gojobori method in a codon-based Z-selection test (Nei and Gojobori, 1986). The analysis was carried out in MEGA5, and the variance of the differences was computed using the bootstrap method (1000 replicates) (Tamura et al., 2011). The synonymous and non-synonymous nucleotide substitution rates for the nucleotide sequences of the target genes (*vmp*1, *stamp*) with the housekeeping *secY* gene (Fabre et al., 2011) were calculated. Positive selection happens when dN/dS ratio >1.0 and p-value for the Z-test < 0.05, on the other hand a ratio <1.0 suggests purifying selection process (Nei and Kumar, 2000).