

Statistical modeling of long-term grapevine response to ‘*Candidatus Phytoplasma solani*’ infection in the field

Ana Rotter · Petra Nikolić · Neža Turnšek · Polona Kogovšek ·
Andrej Blejec · Kristina Gruden · Marina Dermastia

Accepted: 28 July 2017 / Published online: 16 August 2017
© The Author(s) 2017. This article is an open access publication

Abstract Bois noir (BN) is the most widespread European grapevine yellows disease caused by ‘*Candidatus Phytoplasma solani*’. Although our knowledge of the mechanisms of interactions of this pathogenic bacteria with host is largely unknown, the plant-pathogen system of BN is commonly used as a model system for studying grapevine yellows diseases. We applied here a conceptual model of general plant pathology – a disease triangle for describing interactions among the host plant, the pathogen and the environment. We generated a proof-of-concept statistical model for disease triangle using original experimental data and different statistical and data mining approaches for a selected system of ‘*Ca. P. solani*’ infection of cv. ‘Chardonnay’ grapevine plants. We monitored individual plants from a single vineyard over a period of six years. Phytoplasma content, the expression of 21 selected grapevine genes and environmental conditions were recorded and related to disease severity. Our model predicts that in described conditions BN is a function of the expression of grapevine gene *VvDMR6*, summer rainfall and abundance of ‘*Ca. P. solani*’. The greatest impact among elements of the disease triangle is attributed to the pathogen, and is independent of the pathogen titer. We showed that this

first de facto representation of the disease triangle is useful for showing disease dynamics over several years and could be applied to other plant-pathogen systems. The overall results of this study will contribute to understanding of ‘*Ca. P. solani*’ biology and its interactions with grapevine host.

Keywords Bois noir · Disease triangle · Environment · Grapevine · Phytoplasma · Statistical model

Introduction

Research on phytoplasma-incited plant diseases has received increasing attention in recent years (Strauss 2009). Phytoplasmas are cell-wall-less plant pathogenic bacteria belonging to the class Mollicutes (reviewed in Bertaccini and Duduk 2009; Marcone 2014; Zhao et al. 2015). They are transmitted from plant to plant by sap-feeding insect vectors, and they propagate within the cytoplasm of both insects and plants. In plants they exclusively inhabit nutrient-rich phloem tissue. They are the most poorly characterized plant pathogens, primarily because their cultivation in artificial media has just been started (Contaldo et al. 2016), and because efforts of gene delivery and mutagenesis have been so far unsuccessful. Phytoplasmas have a broad range of plant hosts, and they affect several hundred economically important crops and cause substantial losses in yield as well as economic losses (Bertaccini and Duduk 2009). Because of the difficulties associated with

Electronic supplementary material The online version of this article (doi:10.1007/s10658-017-1310-x) contains supplementary material, which is available to authorized users.

A. Rotter · P. Nikolić · N. Turnšek · P. Kogovšek ·
A. Blejec · K. Gruden · M. Dermastia (✉)
Department of Systems Biology and Biotechnology, National
Institute of Biology, Večna pot 111, SI-1000 Ljubljana, Slovenia
e-mail: marina.dermastia@nib.si

phytoplasma research, early detection is difficult, as is development of efficient strategies for disease control.

Phytoplasmas from several molecularly distinct taxonomic groups cause yellows diseases on grapevine (Constable and Bertaccini 2017). In Europe, ‘*Candidatus Phytoplasma solani*’ from the stolbur group 16 SrXII-A (Quaglino et al. 2013), which causes bois noir (BN) disease, is the most widespread and may lead to up to 50% loss of grape clusters (Starý et al. 2013). BN is usually endemic in the Euro-Mediterranean area, and its spread occurs via a complicated disease cycle which includes insect vectors and multiple herbaceous plants as phytoplasma reservoirs (Belli et al. 2010; Cvrković et al. 2014).

Based on the assumption that similar symptoms of grapevine yellows diseases, although caused by taxonomically unrelated phytoplasmas (Belli et al. 2010), are the result of similar responses of grapevine to the infection, the plant-pathogen systems of BN is commonly used as an experimental model for grapevine yellows diseases. Considering very low titers of phytoplasmas in host cells and their uneven distribution in different parts of the plant (Prezelj et al. 2013), the amount of damage caused to hosts by these pathogens is much greater than would be expected from the mere removal of nutrients by them and likely involves plant responses to the infection. Recent studies on BN have shown several transcriptional and metabolic changes in the host plant. Specifically, the genes involved in primary and secondary metabolism are changed in response to infection; several photosynthetic genes are largely down-regulated in infected plant tissues, whereas the genes involved in carbohydrate metabolism, genes associated with the flavonoid synthesis pathway, some genes encoding pathogenesis-related (PR) proteins and genes associated with reactive oxygen species production are significantly induced (Hren et al. 2009a, b; Albertazzi et al. 2009; Landi and Romanazzi 2011; Santi et al. 2013; Covington Dunn et al. 2016). Several metabolites associated with activities of the enzyme products of these genes have also been demonstrated as increased in infected plants (Prezelj et al. 2016b). These results have been largely obtained by novel high-throughput molecular biology techniques which, as has been suggested by Ouzounis (2012), require new interpretive approaches with the widening range of bioinformatics analysis.

Combining all these imperfect and very diverse data in a suitable model to describe and explain the pathogenicity is a methodological challenge, because several factors may introduce uncertainties at different stages

of the model construction (Taylor et al. 2016). Recently, a Bayesian inference with generalized linear mixed effect models has suggested that vector presence is not the only and not the most important factor for BN prevalence, but environment and grapevine cultivar also contribute to the disease development (Panassiti et al. 2015). In addition, mathematical modeling of another phytoplasma grapevine yellows disease, flavescence dorée has been proposed, in which acquisition of the disease, latency and expression of symptoms, recovery rate, removal and replacement of infected plants, and insecticidal treatment have been taken into account (Lessio et al. 2015). Flavescence dorée has been also addressed in a space-time point pattern analysis, which highlighted statistics of the disease progression and regression over time (Maggi et al. 2017).

The present study uses statistical and data mining approaches to interpret host transcriptomic response in combination with pathogen abundance and environmental conditions acquired during six-year-long tracking of BN in a commercial vineyard of cv. ‘Chardonnay’ infected with ‘*Ca. P. solani*’. The experimental results were used to construct a disease triangle; a concept often mentioned (Agrios 2005; Scholthof 2007) but never before realized using experimental data. Our work here may represent the basis for improvement of the knowledge on ‘*Ca. P. solani*’–grapevine interaction, as well as the expansion of the general disease triangle idea for interpreting and even predicting disease dynamics. The developed approach is applicable also for other studies of other diseases.

Materials and methods

The experimental overview of this study is shown in the Supplementary Fig. S1.

Plant material and RNA extraction

The study was carried out in a production vineyard of grapevine (*Vitis vinifera* L.) cv. ‘Chardonnay’ in the southwestern part of Slovenia (45°58' N, 13°32' E) in the years 2004 and 2005, and from 2007 to 2010. Grapevine cv. ‘Chardonnay’ has been confirmed as highly susceptible to phytoplasma ‘*Ca. P. solani*’ (Panassiti et al. 2015). The vineyard was regularly treated with fungicides against *Plasmopara viticola*, *Erysiphe necator* and *Phomopsis viticola*, other pesticides were not

applied. There were sporadic occurrences of the potential ‘*Ca. P. solani*’ reservoir *Convolvulus arvensis* in the vineyard. At the beginning of the experiment in 2004, 15 uninfected and 15 plants infected with ‘*Ca. P. solani*’ were chosen (Table 1). In 2004 all chosen infected plants

had highly pronounced phytoplasma symptoms of level 4. Because of uneven distribution of phytoplasmas within the host plant and their low titers in plant cells, which are detectable with certainty merely in the symptomatic tissues (Prezelj et al. 2013), only symptomatic parts of plant

Table 1 Disease severity of samples plants in different years

Plant	Sampled year					
	2004	2005	2007	2008	2009	2010
1	0	0	0	0	0	0
2	4	4	1	1	0	0
3	0	0	0	0	0	0
4	0	0	0	1	0	0
5	0	0	0	0	0	0
6	0	0	0	0	0	0
7	0	0	1	0	0	0
8	0	0	0	0	0	0
9	0	0	0	0	0	0
10	0	0	0	0	0	0
11	0	0	0	0	0	0
12	0	0	0	2	0	0
13	0	0	0	0	0	3
14	0	0	0	0	0	0
15	4	4	0	2	0	0
16	0	0	0	0	1	0
17	4	2	0	1	1	0
18	4	3	3			
19	4	4	4			
20	4	1	0	0	0	3
21	4	4				
22	4	4	3	4	4	4
23	4	3	3	4	3	4
24	4	3	4	4	2	3
25	0	0	0	0	1	0
26	4	2	0	3	3	0
27	4	3	2	2	2	0
28	4	3				
29	4	2	3	4	1	0
30	4	3	1	3	0	0

The categorization was performed based on pathogen ‘*Ca. P. solani*’ abundance and categorization of symptoms is presented for the studied growing seasons. Color code: green, pathogen not detected by qPCR; yellow, low phytoplasma abundance; red, high pathogen abundance. Level of symptoms are scaled from 0 to 4; 0, asymptomatic; 4, highly pronounced phytoplasma symptoms; grey, year when plant died or was cut down due to infection

were sampled as potentially infected. The same plants were sampled throughout the experiment duration although their sanitary status changed over the years. Grapevine leaf samples were collected at the véraison stage of berry development when pronounced symptoms of BN were observed on the plants (in August). In 2008 and 2010 the plants were additionally sampled in early June before the symptoms developed. In 2008 and 2010 all sampled plants were tested for presence of common grapevine viruses (Supplementary Table S1) that might contribute to symptom development (Table 1). None of the detected viruses could be attributed to any disease severity group (see below) ($p > 0.05$, Fisher's exact test), (Supplementary Fig. S2).

Each sample consisted of the phloem-enriched midrib leaf tissue of the three youngest fully developed leaves from the same shoot. Leaf tissue pieces with 1–2 mm of lamina on each side of the midrib were cut in the field and immediately stored in liquid nitrogen. All samples were collected from 1 to 2 m above ground between 10.00 and 12.00 on the same day. Total RNA was extracted from samples using an RNeasy Plant Mini Kit (Qiagen) and treated with DNaseI (Invitrogen) following the manufacturer's recommendations. The resulting RNA was quantified spectrophotometrically (Nanodrop, Nanodrop Technologies) and quality checked using Bioanalyser (Agilent).

Categorization of plants into disease severity groups

Initial leaf sample categorization was done according to the expressed symptoms in the field (Table 1). Class 0 denotes an asymptomatic plant, which had the same appearance as the uninfected reference; class 1 showed only slight yellowing; class 2 showed slight yellowing; class 3 showed prominent yellowing; and samples in class 4 had typical phytoplasma related symptoms: yellowing, downward curling leaf laminae, and brittle leaf laminae (Hren et al. 2009a). Prior to data analysis, new categorization into what we call the disease severity groups was performed based on symptom severity classes and pathogen abundance. The basis for grouping samples from classes 0 and 1 into disease severity group 1 was absence of symptoms or extremely mild and untypical phytoplasma symptoms and the absence of '*Ca. P. solani*' (Table 1). The samples in disease severity group 2 had symptoms from classes 2 and 3, e.g. pronounced yellowing, although '*Ca. P. solani*' was not always detected in these samples (Table 1). The disease

severity group 3 comprised samples with highly developed phytoplasma symptoms from class 4 in which '*Ca. P. solani*' was always detected (Table 1).

Pathogen abundance

Abundance of '*Ca. P. solani*' was gauged by relative expression of the *Sto111* phytoplasma genomic fragment, which encodes acyl-sn-glycerol-3-phosphate acyltransferase involved in primary metabolism phospholipid biosynthesis, using qPCR developed in Hren et al. (2007). Pathogen abundance was assessed (Table 1) in the same samples as were collected for evaluation of host plant gene expression to ensure that phytoplasma was alive and active at the time of sampling. The expression of *Sto111*cDNA was normalized to expression of *COX* and *18S rDNA* reference genes and relatively quantified as previously described in detail (Hren et al. 2009b). For the relative quantification, plant 24 was used as a calibration sample. The exception was the year 2009, when we were not able to detect '*Ca. P. solani*' in sample collected from this plant; in this case plant 22 was used instead. The relative abundance of *Sto111*cDNA is given as relative copy number of *Sto111* in individual samples. Based on negative or positive result of this testing, plant's sanitary status was assigned as uninfected or infected.

Host gene expression analysis using RT-qPCR

The RT-qPCR analysis of 21 selected genes was performed exactly as previously described (Hren et al. 2009b). Among the genes that showed statistically significant differential expression between '*Ca. P. solani*' infected and uninfected samples in the previous microarray study (Hren et al. 2009a), the ones being part of metabolic pathways presumably associated with the infection were selected (Supplementary Table S2). Based on new data about the association of phytoplasmal diseases with a reactive oxygen species production (Musetti et al. 2007), some genes involved in this process were added to analysis in 2007 (Supplementary Table S2). In year 2010 only the expressions of genes that had been revealed by decision trees (see below) were analyzed to validate the results of statistical modeling and decision trees (Table S2).

Universal Master Mix (TaqMan or SybrGreen, Life Technologies) together with primers and probes of appropriate concentration was used for analyzing 2 μ l of

cDNA (Supplementary Table S2). 10-fold and 100-fold diluted cDNA was tested in two parallel reactions in order to control for the inhibition of amplification. All RT-qPCRs were carried out in 384-well reaction plates on the Lightcycler LC480 (Roche) using real time data collection. For amplification, standard cycling conditions were used with added melting curve stage for SybrGreen reactions. Dissociation curve analysis was done and primer dimer formation examined. Data was first analyzed using the Lightcycler 480 software (Roche). The standard curve method was used for quantification of relative gene expression (Hren et al. 2009b), where the efficiency of amplification of each amplicon was extrapolated from the slope of the standard curve. The results of the gene expression analysis were calculated as numbers of copies relative to the geometrical mean of *COX* and *18S* (Eukaryotic 18S rRNA TaqMan endogenous control, Life Technologies) copies. Samples with low expression of the target gene, falling out of the dynamic range of the assay, and samples with failed control of inhibition (standard deviation of relative copy number calculated from different dilutions of the sample higher than 0.5) were excluded from further calculations (Supplementary Table S3).

All details on qPCR reactions and normalization are given in agreement with MIQE précis (Bustin et al. 2010) recommendations (Supplementary Table S2).

For the disease triangle construction, data on gene expression were further standardized (Supplementary Table S4). Thirteen plants with the assigned uninfected sanitary status for the entire experiment duration were used as control plants. For each growing season j , the average (\bar{y}_{c_j}) and the standard deviation (sd_{c_j}) of these control genes were calculated. Then the standardized value for each measured copy number (y_{s_j}) was calculated by

$$y_{s_j} = \frac{y_{i_j} - \bar{y}_{c_j}}{sd_{c_j}}$$

where, j denotes the year tested and (y_{i_j}) denotes copy numbers as calculated from individual plant samples. Since each plant was tested at least twice, tested shoots from the same plant were averaged.

Environmental data

Several environmental factors that might influence a disease progression (Contreras-Medina et al. 2009) and

for which public data are available (the portal of the Statistical Office of the Republic of Slovenia for the weather station Bilje, nearest to the selected vineyard – 45° 53' N, 13° 38' E) were followed during the experimental years in the studied vineyard. Daily data for temperature, rainfall and relative humidity were averaged to the average winter temperature, annual rainfall, average temperature in the sampling month, summer months rainfall, average rainfall in the growing season prior to sampling (May to October), and annual relative humidity (Supplementary Table S5). Additionally data for number of rainy days per growing season, number of days with snow cover, number of days with thunderstorms, number of days with hail or sleet, annual snow cover in centimeters, number of cloudy days, number of sunny days, annual sun duration, prevalent wind direction and frequency of wind were also considered (Supplementary Table S5).

Construction of decision trees

Decision trees were calculated with Weka software (Hall et al. 2009), using the J48 classifier (Quinlan 1993). The response variable in the constructed decision trees was plant sanitary status (i. e. uninfected/infected based on the relative expression of the *Sto111* phytoplasma genomic fragment, see above). Data used for the construction of all decision trees were standardized relative gene expression values. Two types of decision trees were constructed. Data used for building the first decision tree were gene expression from plants sampled in late summer of 2004, 2005, 2007–2009 growing seasons. This decision tree was constructed using data from plants sampled in late summer in order to predict their late summer sanitary status. Data used for building the second decision tree were gene expression in plants sampled in early summer of 2008. The second decision tree predicted the sanitary status of plants in late summer based on the expression data from early summer sampling. The performance of both decision trees was validated by gene expression data from 2010.

Selection of disease triangle elements

Except where stated otherwise all data was analyzed in R language (R Core Team 2014) using the libraries rms (Harrell 2013) and compositions (van den Boogaart et al. 2012).

For construction of the disease triangle, a single and the most influential covariate per disease triangle element (i.e. pathogen, host and environment) was selected. Various steps and statistical methods used to identify the covariate that could be considered as a good representative for each triangle element are described in the following paragraphs and partially in the Results section.

Pathogen abundance was selected as a pathogen element in disease triangle.

In order to find the most suitable candidate to represent the host element in the disease triangle, we checked the difference in gene expression due to the sanitary status of the plants using Wilcoxon test with Bonferroni correction in each growing season. Genes that had Bonferroni corrected p -values below 0.05 in at least half of the tested growing seasons were considered as potential host element candidates. Their ability to discriminate between uninfected and infected plant sanitary statuses was assessed with discriminant analysis. The gene with the highest linear discriminant coefficient was considered to represent host element in disease triangle. The selection was additionally confirmed by logistic regression function, calculated for each of candidate genes to test the significance of their influence on disease severity.

The environmental indicators (Supplementary Table S5) were separately added into logistic regression models, and their influence on disease severity was assessed.

Several modeling approaches considered the longitudinal / repeated measures model for evaluation of combined pathogen, host and environment data. (i) In generalized estimating equations (gee), using a geepack library (Højsgaard et al. 2006), data were ordered by plant to take into account the repeated component, and by time-year of sampling to taking into account the longitudinal component. (ii) A MASS library (Venables and Ripley 2002) was used for mixed effects models on the same ordered data, where the random effects source was defined as the plant sampled within a year. (iii) When plant symptoms was the ordered factor and the plant sampled each year was the source of random variation, a cumulative link mixed model in package ordinal (Christensen 2015) was applied. (iv) Finally, the package rms (Harrell 2015) was used with logistic regression and robust covariance matrix estimates to correct for the repeated measures (i.e. the same plants sampled in different years). For calculation of

coefficients used for the disease triangle construction, each model for each combination of host and environment variables was tested using the whole dataset. From the model in which all combined variables had a significant influence ($p < 0.05$) on disease severity, the representative for an element in the disease triangle was selected.

Predictions based on the model were calculated and plotted using a nomogram, in which the complex model calculations are replaced by simpler readings of nomogram scales (Banks 2006; Ambaum 2007; Weisstein 2016). In the nomogram presented in this study, a disease is expressed as a function of dependent variables, in our case triangle elements. The influences of each variable are added and the probability of the sample being infected is then read from the nomogram. By aligning the measurement of each of the variables to the “Point” ruler, the actual measurements can be converted to points. By summing “Points” of each individual variable a “Total points” ruler can be aligned to the “Probability” ruler. The comparison of performance of nomogram used in this study with the probability of infection calculated from the model gave similar results. This suggests that the presented nomogram can be a practical and fast tool in the diagnostics of phytoplasmas.

The nomogram was used to enumerate the influence of each of the triangle elements. Specifically, the “Points” ruler was recorded for each element and these “Points” were then converted separately for each year to the compositions (van den Boogaart et al. 2012) – the proportion (R) of each element for individual plants in individual year, using the equation:

$$R_{P,H,E} = \frac{\text{points}_{P,H,E}}{\sum \text{points}_{P,H,E}}$$

where, P denotes pathogen, H denotes host and E denotes environmental triangle element.

The proportions (R) were used for visualizing the influence of each factor with a ternary plot. For presentation of temporal changes in influences over the studied period, ternary plots were combined into a sequence of animated plots. Animated plots with smooth transitions between states can enhance the perception of changes in time. In addition to the values of factors, plant sanitary status and its assigning to the disease severity group were coded with color and symbol size. The R package for animated graphics (animatoR) developed by our

team (Blejec 2011, 2016) was used for construction of animated plots.

Results

The experimental workflow (Supplementary Fig. S1) developed for study of BN disease included: (i) collection of experimental data describing host response, environment and pathogen abundance, (ii) identification of a proper measure of disease severity that is valid in both statistical (i.e. used in the appropriate model) and biological (i.e. easy to measure) senses for the triangle elements; (iii) construction of decision trees for predicting the sanitary status of plants, and statistical modeling for disease triangle construction; and (iv) dynamic visualizations of the disease triangle to show development and fluctuations of BN over several years. To compute the BN disease triangle model and keep it as simple as possible by not overfitting, different types of experimental data were generated: pathogen abundance estimated from collected samples, relative expression of several host plant genes that have been shown to be associated with development of grapevine yellows symptoms, symptom evaluation, and environmental factors collected from the nearby weather station. The computed model of disease triangle was validated by data acquired in a year that was excluded from the statistical analysis.

Abundance of ‘*Ca. P. solani*’ and disease severity

According to the relative copy number of the ‘*Ca. P. solani*’ transcript of the *Sto111* genomic fragment, samples were discretized into three groups as uninfected, slightly infected and severely infected (Table 1). Because a bimodal distribution of the *Sto111* transcript’s relative copy number was observed in positive samples in almost every growing season, the limit between slightly infected (henceforth: low pathogen abundance) and severely infected (henceforth: high pathogen abundance) samples was set for each growing season (Supplementary Fig. S3).

In the first experimental growing season 15 symptomatic plants with confirmed infection with ‘*Ca. P. solani*’ and 15 asymptomatic plants in which such infection was not detected were chosen. However, their sanitary status changed over the years. At the end of experiment in only one plant the pathogen was detected in all growing seasons, five plants were tested negative

to the pathogen presence in the second growing season, and two in the third one. In the third and fourth growing season some infected plants died. Additional five plants were re-tested positive to the pathogen usually one or two growing seasons after the negative test in an interim. Only one initially uninfected plant became infected during the experiment (Table 1).

Identification of genes that best predict the sanitary status of plants

In order to identify plant genes that may be important for determining uninfected or infected plants, we followed the expression of 21 genes chosen from various pathways in the primary and secondary metabolism (Fig. 1, Supplementary Table S3, S4). In the first step we aimed at identification of genes whose expression profile can predict the sanitary status of the plant, i.e. infected or uninfected with ‘*Ca. P. solani*’. We constructed two decision trees (Fig. 2).

The first decision tree (Fig. 2a) was constructed using the results of the standardized relative gene expression data from plants sampled late in five growing seasons (Supplementary Table S5) to predict their sanitary status. It exposed three genes – *VvAGPL*, *VvDMR6* and *VvHP*. According to the analysis the sample was infected if the standardized amount of *VvDMR6* transcript was higher than 2.88, the standardized amount of *VvHP* transcript lower than –1.13 and that of *VvAGPL* higher than –0.25. Similarly, the sample was infected if the standardized amount of *VvDMR6* transcript was higher than 2.88 and at the same time that of *VvAGPL* higher than 1.34.

The second decision tree (Fig. 2b) was built on the data set from samples collected in early summer when the symptoms had not been developed yet, to predict plant sanitary status late in the growing season. This decision tree uncovered two genes – *VvAGPL*, which was also part of the first decision tree, and *VvGLC2*. Based on their standardized relative gene expressions the sample was infected with ‘*Ca. P. solani*’ if the standardized amount of *VvGLC2* transcripts was higher than 0.76 while at the same time the standardized amount of *VvAGPL* was higher than –0.22.

Performance (Table 2) of the first decision tree was worse in terms of accuracy and sensitivity compared to the second one. This might indicate that other factors besides selected gene expression participate in the development of BN late in the growing season. The overall

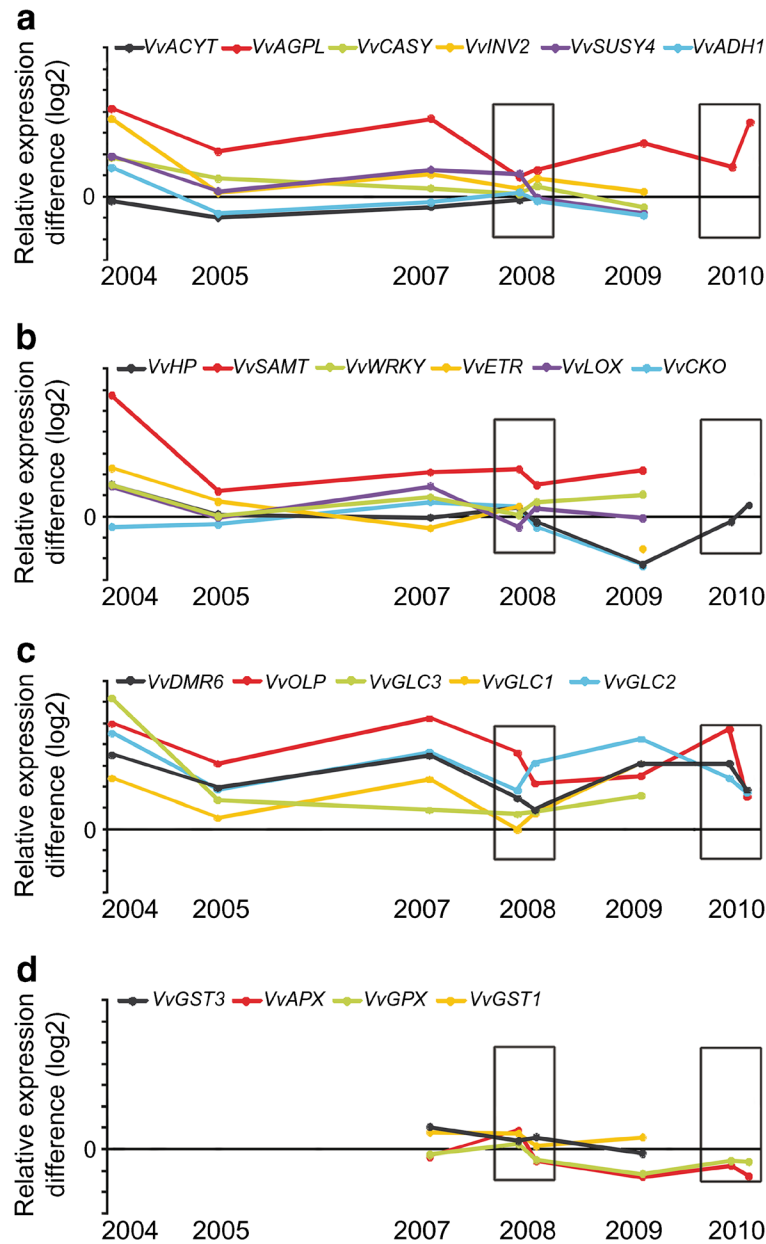


Fig. 1 The difference in expression profile of infected and uninfected grapevine plants between relative gene expression in infected and uninfected samples in several growing seasons. The genes were chosen from various metabolic pathways: expression of genes associated with primary metabolism (a), signaling and hormone metabolism (b), pathogenesis related proteins and protein with unknown function (c) and ascorbate–glutathione metabolism (d). The y-axis shows \log_2 ratio of the average relative gene expressions in infected plants relative to uninfected plants. Frames indicate years 2008 and 2010 when plants were sampled both early and late in the growing season. The following genes encode proteins as follows: *VvACYT*, apocytochrome f precursor; *VvINV2*,

vacuolar acid invertase 2; *VvAGPL*, large subunit of ADP-glucose pyrophosphorylase; *VvSUSY4*, sucrose synthase; *VvCASY*, callose synthase; *VvADH1*, alcohol dehydrogenase 1; *VvLOX*, lipoxygenase; *VvETR*, ethylene receptor; *VvHHP*, histidine-containing phosphotransfer protein; *VvCKO*, cytokinin oxidase; *VvSAMT*, S-adenosyl-L-methionine: salicylic acid carboxyl methyltransferase; *VvWRKY*, transcription factor WRKY54; *VvDMR6*, 2-oxoglutarate and Fe(II)-dependent oxygenase; *VvOLP*, osmotin; *VvGLC1*, *VvGLC2* and *VvGLC3*, β -1,3-glucanases 1, 2 and 3, respectively; *VvAPX*, ascorbate peroxidase; *VvGPX*, glutathione peroxidase; *VvGst1* and *VvGST3*, glutathione S-transferase 1 and 3, respectively

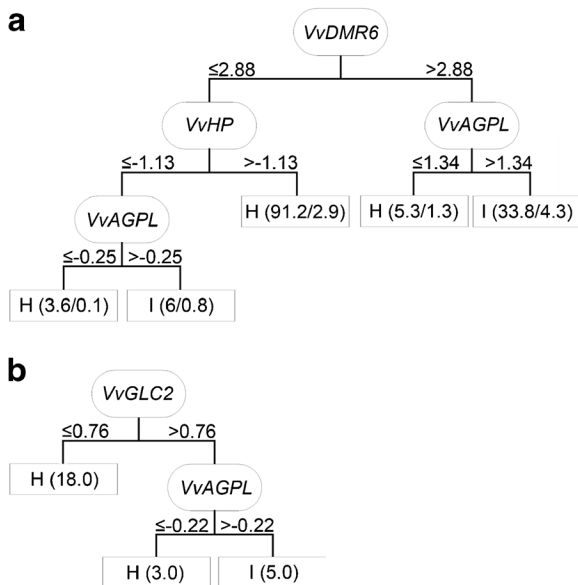


Fig. 2 Decision trees for plant sanitary status prediction. Decision trees to classify plants as uninfected or infected with ‘*Ca. P. solani*’ based on the standardized relative gene expression data: **a** prediction of the plant sanitary status in late summer based on testing of the late summer plants; **b** prediction of the sanitary status in late summer based on testing of plants in early summer. Gene abbreviations are denoted in ovals. The numbers above each branch of the decision tree denote standardized relative gene expression cut-off for the decision. Each result (H-uninfected, I-infected) is accompanied by a number in parentheses representing the number of correctly/incorrectly classified samples after 10-fold cross validation. The first number in parentheses represents the number of correctly classified samples, while the second number represents the number of misclassified instances. Due to missing values for genes that did not pass the qPCR quality control some of the numbers are not integers

Table 2 Performance of constructed decision trees

Performance measure	First decision tree late-late	Second decision tree early-late
True positives	0.29	0.6
True negatives	1	1
False positives	0	0
False negatives	0.71	0.4
Accuracy*	0.64	0.86
Precision**	1	1
Sensitivity***	0.29	0.6

*ratio of correctly classified instances

**ratio of positive instances, correctly classified as such (also known as positive predictive value)

***power of test to identify positive results (i.e. infected plants)

accuracy for the second decision trees was high (86%) (Table 2), meaning that the results of these decision trees were trustworthy due to a high proportion of true positives and true negatives. In the second decision tree, when a sample is identified as infected it is indeed infected. However, we must take into consideration that some true positives (de facto infected plants) in late summer could not be predicted as such from the early summer samples. Similarly as for the first decision tree, these results suggest an involvement of additional factors in the development of BN.

BN disease triangle construction

For the BN disease triangle construction, only one covariate was considered for each triangle element. The rationale for this decision was in retaining the computed disease triangle as simple as possible, but with keeping in mind that in its new versions the number of covariates may be increased.

In the first step we used standardized gene expression for all examined genes in each growing season and compared their expression from infected and uninfected plants using Wilcoxon test with Bonferroni correction. Using this approach, we identified the genes *VvGLC2*, *VvAGPL*, *VvOLP*, *VvDMR6* and *VvAPX* that were differentially expressed between infected and uninfected plants in all years (Supplementary Fig. S4). In a discriminant analysis based only on expression of these genes, the prior probabilities of group uninfected and group infected were 0.7 and 0.3, respectively. The analysis showed that the highest coefficient of linear discriminant was attributed to the *VvDMR6* gene and thus *VvDMR6* had the highest power to discriminate between infected and uninfected samples.

Additional logistic regression models having the expression of *VvDMR6*, *VvAGPL*, *VvOLP* and *VvGLC2* genes with the highest discriminatory power between uninfected and infected categories as explanatory variables, showed that only the expression of *VvDMR6* made a significant contribution to the progression of samples from lower to higher disease severity group ($\beta = 0.5, p = 0.0002$). Other tested genes had a marginal or insignificant impact.

Logistic regression analysis of environmental factors revealed that winter temperature, rainfall in the previous growing season and summer rainfall were the only environmental factors which had, in combination with the expression of the most influential host gene

VvDMR6 and pathogen abundance, a significant albeit small influence on disease severity. Of these the summer rainfall was chosen for further analysis as the most significant one ($\beta = 0.03$, $p = 0.04$) for disease severity grouping, while not affecting the significance of other elements in the statistical model (Table 3).

In the final step of statistical modeling we combined the influence of pathogen, host and environment elements on the disease severity in the particular growing seasons and over several years. The model predicts that disease is a function of the expression of host gene *VvDMR6* ($\beta = 0.2$, $p = 0.04$), summer rainfall ($\beta = 0.03$, $p = 0.04$) and abundance of pathogen ($\beta = 12.2$, $p < 0.0001$ even for low '*Ca. P. solani*' abundance) (Table 3). The coefficients of all three variables were positively connected with disease severity (Table 3).

To visualize the impact of each variable of the model, a nomogram was applied (see example in Fig. 3, Supplementary Table S6). The nomogram shows that the greatest impact among elements that comprise the BN disease triangle is attributed to the pathogen element. Even low abundance of the pathogen counts for more than 90 points, which corresponds to 100% probability of disease appearance.

Finally, the influence of each element was depicted with a ternary plot (Figs. 4 and 5, Supplementary Fig. S5). A special feature of the BN disease triangle is bubbles that represent single analyzed plants and are located inside the triangle. In the first step of building the BN disease triangle, data were summarized to visualize year-to-year differences within the same plot (Fig. 4). In the next step a dynamic visualization was included in the disease triangle (Fig. 5, Supplementary Fig. S5). Here each individual plant was plotted throughout all growing seasons (Fig. 5). While the size of bubble denotes the disease severity, their colors can be used for monitoring changes in plant sanitary status. The bubbles

Table 3 Univariate logistic regression models for testing the significance of examined elements to predict the disease severity

	Coefficient β	Standard error	Wald Z statistic	p -value
<i>VvDMR6</i>	0.2	0.1	2.1	0.04
Summer rainfall	0.03	0.02	2.1	0.04
$y = \text{low}$	12.2	0.7	18.5	<0.0001
$y = \text{high}$	13.2	1.0	13.0	<0.0001

VvDMR6 expression, pathogen abundance (y), summer rainfall

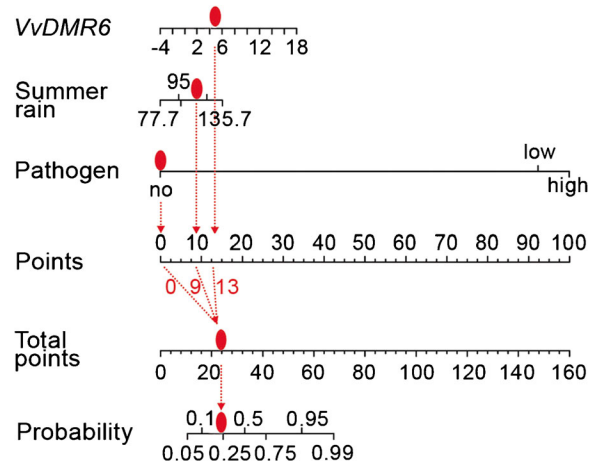


Fig. 3 Nomogram for predicting the influence of each disease triangle element on the plant sanitary status. Points (i.e. influence) for each variable may be read from the “Points” scale and their sum applied to the “Total points” line underneath. The Total points scale is aligned with the Probability of higher symptoms (i.e. probability of infection) on the bottom line of the nomogram. Red dots represent an example calculation: *VvDMR6* relative standardized expression was 5 (corresponding to 13 points), summer rainfall was 100 mm (corresponding to 9 points), and there was no detected phytoplasma infection (corresponding to 0 points). Total points are then calculated by summing the points from each variable, in this case 22 points. Twenty-two total points corresponds to a probability of infection of approximately 0.2

representing uninfected plants are always at the side opposite to the pathogen corner because they have no trace of the pathogen. The bubbles that represent infected plants are always in the pathogen corner because of the very high impact of the pathogen component. The plants that were tested positive to the pathogen in one year and negative in the subsequent one are depicted with blue bubbles. They show mixed behavior; some of them behave like uninfected and some like infected. Bubbles representing plants in which testing to pathogen was only temporary negative are usually positioned nearer the pathogen corner. However, both plants that became negative to the pathogen presence and those that were again tested positive in subsequent growing seasons tend to appear among the uninfected plants in triangles in growing seasons in which the impact of the environmental element was high (e.g. in 2005).

Discussion

Here we analyzed a long-term grapevine response to '*Ca. P. solani*' infection in the vineyard in combination

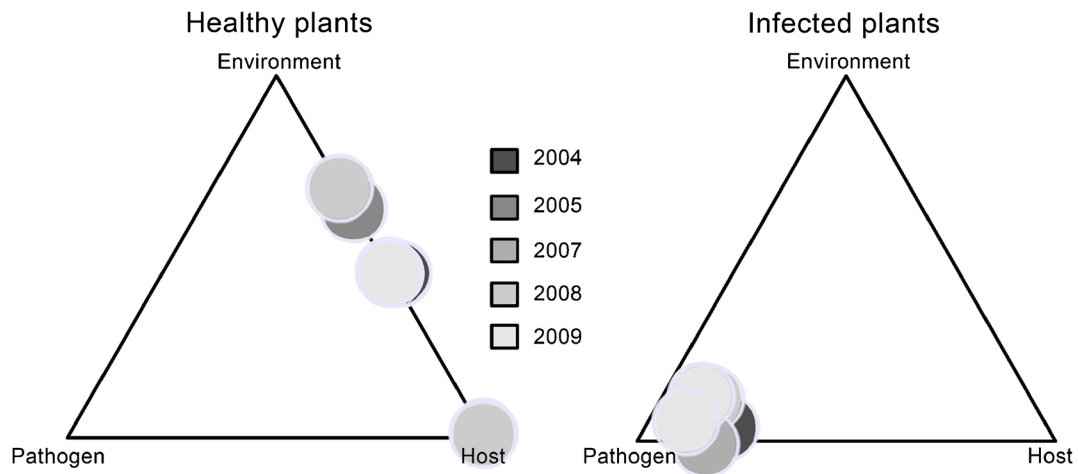


Fig. 4 BN disease triangle summarized over the sampling years. The left triangle depicts the average aggregation of uninfected plants over the years 2004–2009 while the right one denotes the average effects of the disease triangle elements on the infected

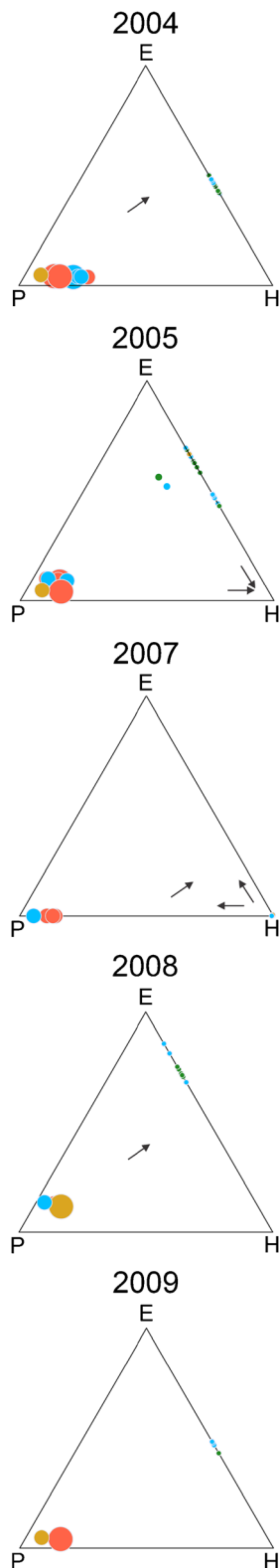
plants. Each bubble represents the average measured values in a particular year. The bubble position indicates influence of a particular variable within a certain year. The size of the bubble indicates the disease severity group

with environmental factors with different statistical tools, and the results were finally used for the construction of a model BN disease triangle. With this approach we gained some novel insights into biology of the studied disease as well as confirmed some information that had been previously published on BN and other grapevine yellows diseases. The later demonstrates the accuracy of the model and increases a confidence in new information.

The set of genes included in this study was based on our previous microarray dataset (Hren et al. 2009a) and some additional literature search (e.g. Musetti et al. 2007). Here we evaluated their expression over multiple years in uninfected and infected plants and applied a decision tree approach to identify the genes that best predict the disease status of a given plant. We have previously shown that an accurate classification of plants as uninfected or infected with ‘*Ca. P. solani*’ is also possible with other tools, for example with a support vector machines algorithm (Hren et al. 2009b). Although the results of such classification have been confirmed to be very accurate, they are without real applied value because the analysis is very laborious, requiring RT-qPCR analysis of 17 genes. On the other hand, genes selected by tools in this study have potential use in phytoplasma diagnostics. The results here indicated that specific combinations of the expression of three or two genes from the first and second decision tree, respectively, may be used as probable markers of ‘*Ca. P. solani*’ infection.

Several outcomes of this study are noteworthy. Both the standard identification of differentially expressed genes as well as decision tree construction revealed the gene *VvDMR6* from the 2OG-Fe(II) oxygenase gene superfamily, which showed the highest similarity (69%) to the sequence of the gene *DMR6* from *Arabidopsis* (At5g24530) (Van Damme et al. 2008). Although its biological role is not known at the moment, it has been shown that *Arabidopsis* plants lacking a functional *DMR6* gene have reduced susceptibility to downy mildew (Van Damme et al. 2008). In addition, it has been suggested that it acts as a suppressor of plant immunity (Zeilmaker et al. 2015). In a recent study using the CRISPR/Cas-9 system, the generated tomato plants with small deletions in the *DMR6* gene do not have significantly affected growth and development, but show disease resistance against different pathogens (de Toledo Thomazella et al. 2016). Based on the results of the model presented here, we have already tested the involvement of *VvDMR6* in pathogenicity of flavescence dorée phytoplasma that causes the most devastating grapevine phytoplasma disease. *VvDMR6* has shown similar pattern and has been suggested as a potential early marker in diagnosis of grapevine yellows (Prezelj et al. 2016a).

The second gene identified in both analysis was *VvAGPL*, which encodes a large subunit of ADP-glucose pyrophosphorylase (AGPase) involved in starch biosynthesis. *VvAGPL* was also exposed in the second decision tree, which predicts sanitary status of plants in



◀ **Fig. 5** Disease triangles calculated for each growing season. Each bubble represents an individual plant. The position of a bubble within each triangle denotes the importance of each of the elements: pathogen (P), host plant (H) and environment (E). The plant sanitary status is represented by bubble color: green, plants that remained uninfected during the whole period; red, plants infected during the whole period; blue, plants tested positive to the pathogen in one season and negative in the subsequent; ochre, plants in which testing to pathogen was only temporary negative. The size of the bubbles denotes the disease severity group. The arrows indicate transitions between the sanitary statuses of individual plants from one growing season to another. These results can be visualized dynamically throughout the years using anima-toR (Supplementary Fig. S5)

late summer based on results from early growing season testing. It is of note that the increased expression of *VvAGPL*, the activity of AGPase, as well as starch concentration have been reported in grapevines infected with phytoplasma *flavescence dorée* (Prezelj et al. 2016a) indicating that this pattern of expression is typical for GY and not for specific grapevine–phytoplasma interaction. The third gene in this group is *VvGLC2*, encoding an isoform of β -1,3-glucanases, a member of the class 2 pathogenesis related proteins (PR-2). Increased levels of *VvGLC2*, *VvGLC1* and *VvGLC3* have been shown before to be involved in development of BN (Hren et al. 2009a; Landi and Romanazzi 2011; Dermastia et al. 2015). In addition, increased amounts of β -1,3-glucanase protein product have been demonstrated in grapevines infected with phytoplasma *flavescence dorée* (Margaria et al. 2013).

An important gene associated with BN and here revealed by the Wilcoxon test was *VvOLP*, encoding osmotin-like protein from the class 5 of pathogenesis related proteins (PR-5). Increased transcript of *VvOLP* has also been reported in grapevine upon infection with phytoplasma *flavescence dorée* (Prezelj et al. 2016a, b), fungal pathogens (Spagnolo et al. 2012), as well as in *Arabidopsis* upon infection with beet cyst (*Heterodera schachtii*) nematode (Hamamouch et al. 2011). Osmotins have strong antifungal/antibacterial properties through their action on membrane permeability, disruption of lipid bilayers and delayed growth of microbes (Anžlovar and Dermastia 2003). During stress conditions, osmotins help in accumulation of the osmolyte proline, which quenches reactive oxygen species and free radicals (Anil Kumar et al. 2015) and is highly increased in grapevines infected with ‘*Ca. P. solani*’ (Prezelj et al. 2016b). The amount of *VvOLP* transcript may indicate an intensive response by phytoplasma-

susceptible grapevine cv. ‘Chardonnay’ to the appearance of ‘*Ca. P. solani*’ in leaves early in the growing season and a constant state of priming in infected plants throughout of the growing season. These suggestions are supported by strongly increased amount of *VvOLP* transcript in infected grapevine plants in comparison with uninfected ones, together with a relative drop in its expression later in the growing season.

PR-2 and PR-5 genes are commonly used as molecular markers for the salicylic acid (SA)-dependent systemic acquired resistance (SAR)-signaling and their expression is coordinately regulated by SA (Frias et al. 2013). In addition, it has been suggested that ‘*Ca. P. solani*’ induces SA-dependent SAR in leaves of infected tomatoes and grapevine (Ahmad et al. 2015; Dermastia et al. 2015). The results of this study associated with PR-2 and PR-5 genes and additionally with increased gene expression of *VvSAMT*, which encodes S-adenosyl-L-methionine: salicylic acid carboxyl methyltransferase in infected plants, as well as with 26-fold increase in salicylic acid-glucopyranoside (Prezelj et al. 2016b), support this idea.

It is generally assumed that the environment is a driving force in disease (Hardwick 2006). Accordingly, several models formulate pathogen variables as a function of driving variables including temperature, rainfall, moisture, wind direction, and radiation (reviewed in

Contreras-Medina et al. 2009). However, according to our results in our experimental system the environment has some impact on uninfected and newly infected plants, but has only a slight effect on already infected plants. These findings might reflect the obligate intracellular nature of phytoplasmas, which do not spread via spores or other external means that are more influenced by environmental variables. The results corroborate a recent model of BN (Panassiti et al. 2015), which predicts that altitude has the highest impact on the disease prevalence but most of other environmental parameters only slightly affect it. In addition, the model of Panassiti et al. (2015) shows that BN is negatively correlated with the minimal temperature of coldest period, which is opposite to our findings. The discrepancy is likely related to different climate conditions in southwestern Germany in comparison with our vineyard location in a mild sub-Mediterranean climate, where the winter temperature over the experimental years never dropped below 3 °C but was in average higher than 3 °C.

The effects of the host plant factor seem to be minor compared to the very strong pathogen impact. Nevertheless, there were some differences in the locations of the bubbles on the pathogen-to-host side of the triangles that indicated influence by host metabolism. It is of note that plants in which pathogen was not detected in every growing season show mixed behavior (i.e. have

Table 4 Differences in disease triangle construction between the original disease triangle idea (Francl 2001; Scholthof 2007) and our solution focusing on response of individual plants

	Original triangle	Constructed triangle
Element		
Host	Inherent susceptibility Seasonal development	Gene expression of selected genes
Pathogen	Virulence Prevalence	Pathogen abundance
Environment	Severity Duration of infection period	Most influential environmental factor
Element features		
Disease severity	Determined by the apex of the triangle	Determined by the size of the bubble plot
Number of components within an element	Maximum of 2	≥1, depending on the applied statistical model
Influence of each element	Determined by angles and height	Determined by the bubble distance from corners
Dynamic visualization	No	Yes
Visualization of the status of individual plants	No	Yes
Triangle appearance		
Side lengths	The influence of each element is deducted from the length of the triangle side	The triangle size and side is not important

changed sanitary status) in the disease triangle. This is in agreement with reports investigating changes in the gene expression differences among grapevines infected with ‘*Ca. P. solani*’ (Dermastia et al. 2015) and apple trees infected with ‘*Ca. P. mali*’ (Musetti et al. 2013), and the proteome of grapevine infected with the phytoplasma *flavescence dorée* (Margaria et al. 2013) in which the amounts of specific proteins/transcripts overlapped among plants with different sanitary status. Our observation here might reflect still very unclear phenomenon of recovery, described as a spontaneous disappearance of symptoms from the crown of the affected plants (Bertaccini and Duduk 2009). Recovery may represent either a latent infection with temporal remission of symptoms or, as has been proposed, if there is no symptoms for three consecutive years, the plant should be considered to have recovered from phytoplasma (Maixner 2006). In our case with very thorough following of symptoms and phytoplasma presence we can conclude that plants without confirming pathogen and categorized in a disease severity group 1 for at least three consecutive years do not progress to the disease severity group 2 or 3. However, they may be re-infected after this recovered period.

In building the BN disease triangle we considered and redefined the original triangle concept of McNew (1960) in order to make it useful for describing disease on the level of individual plant rather than taking an epidemiological view of disease (Table 4). With a combination of biological knowledge and statistical tools we identified a minimum number of components required to construct the BN disease triangle. In contrast to the original disease triangle, in which a maximum of two components within each of the triangle elements was used, in our model conception a single component was chosen to preserve the simplicity of the model. However, extra variables can be added, but care must be taken to avoid model overfitting. While in the original disease triangle only its shape and height determine disease severity, a new feature, bubbles, representing status of individual plants, was added to the new BN disease triangle. Therefore, the shape of the triangle is always the same, but the size, color and positioning of the bubbles represent the disease dynamics. The disease triangle redefinition circumvents disadvantages of the original model, for example, situations in which the triangle cannot be drawn because one of the sides is too short or if the total angle sum is not 180°.

We have shown that with a disease triangle it is possible to create a dynamic representation of the combined influence of pathogen, host and environment on disease severity. In this first version of the model we did not consider some important contributors to BN, including insect vectors and wild reservoirs of ‘*Ca. P. solani*’ or even bacterial endophytic communities in grapevine (Campisano et al. 2014). Therefore, in the future more complex models with more variables for each element may be constructed using similar approaches. With properly chosen variables, this new concept for building a disease triangle may be expanded to other grapevine varieties or to other economically important grapevine phytoplasmas.

Acknowledgements The authors thank Mr. Boris Jakončič, who gave us permission to conduct the study on his land. They also thank Marko Chersicola and Dr. Nina Prezelj for their help with data collection and Dr. Elizabeth Covington for critical reading of the manuscript. This work was financed by the Slovenian Research Agency grant P4-0165 and young researcher grant No. 1000-08-310001 awarded to PN. The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Agrios, G. N. (2005). *Plant Pathology*, (p. 79). New York: Elsevier.
- Ahmad, J. N., Renaudin, J., & Eveillard, S. (2015). Molecular study of the effect of exogenous phytohormones application in “stolbur” phytoplasma infected tomatoes on disease development. *Phytopathogenic Mollicutes*, 5(1-Suppl), S121–S122.
- Albertazzi, G., Milc, J., Caffagni, A., Francia, E., Roncaglia, E., Ferrari, F., et al. (2009). Gene expression in grapevine cultivars in response to Bois Noir phytoplasma infection. *Plant Science*, 176, 792–804.
- Ambaum, M. H. P. (2007). A nomogram for the atmosphere. *Weather*, 62(12), 344–345
- Anil Kumar, S., Hima Kumari, P., Shravan Kumar, G., Mohanalatha, C., & Kavi Kishor, P. B. (2015). Osmotin: a

- plant sentinel and a possible agonist of mammalian adiponectin. *Frontiers in Plant Science*, 6, 63.
- Anžlovar, S., & Dermastia, M. (2003). The comparative analysis of osmotins and osmotin-like PR-5 proteins. *Plant Biology*, 5, 116–124.
- Banks, J. (2006). Nomograms. In: Encyclopedia of statistical sciences. Wiley Online Library. doi:10.1002/0471667196.ess1795.pub2
- Belli, G., Bianco, P. A., & Conti, M. (2010). Grapevine yellows in Italy: Past, present and future. *Journal of Plant Pathology*, 92(2), 303–326.
- Bertaccini, A., & Duduk, B. (2009). Phytoplasma and phytoplasma diseases: a review of recent research. *Phytopathologia Mediterranea*, 48, 355–378.
- Blejec, A. (2011). AnimatoR: dynamic graphics in R. Page 39 in The R User Conference, user! Book of Contributed Abstracts, University of Warwick, Coventry, UK. http://web.warwick.ac.uk/statsdept/useR-2011/abstract_booklet.pdf. Accessed 25 October 2016.
- Blejec, A. (2016). AnimatoR: R package for Animated Graphics. Zenodo. <https://github.com/ablejec/animatoR/>. doi:10.5281/zenodo.60228.
- Bustín, S. A., Beaulieu, J.-F., Huggett, J., Jaggi, R., Kibenge, F. S. B., Olsvik, P. A., et al. (2010). MIQE précis: Practical implementation of minimum standard guidelines for fluorescence-based quantitative real-time PCR experiments. *BMC Molecular Biology*, 11(1), 74.
- Campisano, A., Antonielli, L., Pancher, M., Yousaf, S., Pindo, M., & Pertot, I. (2014). Bacterial endophytic communities in the grapevine depend on pest management. *PloS One*, 9(11), e112763.
- Christensen, R. H. B. (2015). Ordinal - Regression models for ordinal data. R package version 2015.6–28. <https://cran.r-project.org/web/packages/ordinal/ordinal.pdf>.
- Constable, F., & Bertaccini, A. (2017). Worldwide distribution and identification of grapevine yellows diseases. In M. Dermastia, A. Bertaccini, F. Constable, & N. Mehle (Eds.), *Grapevine yellows diseases and their phytoplasma agents: Biology and detection* (pp. 17–46). Cham: Springer.
- Contaldo, N., Satta, E., Zambon, Y., Paltrinieri, S., & Bertaccini, A. (2016). Development and evaluation of different complex media for phytoplasma isolation and growth. *Journal of Microbiological Methods*, 127, 105–110.
- Contreras-Medina, L., Torres-Pacheco, I., Guevara-González, R. G., Romero-Troncoso, R. J., Terol-Villalobos, I. R., & Osornio-Rios, R. A. (2009). Mathematical modeling tendencies in plant pathology. *African Journal of Biotechnology*, 8(25), 7399–7408.
- Covington Dunn, E., Roitsch, T., & Dermastia, M. (2016). Determination of the activity signature of key carbohydrate metabolism enzymes in phenolic-rich grapevine tissues. *Acta Chimica Slovenica*, 63, 757–762.
- Cvrković, T., Jović, J., Mitrović, M., Krstić, O., & Toševski, I. (2014). Experimental and molecular evidence of *Reptalus panzeri* as a natural vector of bois noir. *Plant Pathology*, 63(1), 42–53.
- de Toledo Thomazella, D. P., Brail, Q., Douglas, D., & Staskawicz, B. (2016). CRISPR-Cas9 mediated mutagenesis of a DMR6 ortholog in tomato confers broad-spectrum disease resistance. *bioRxiv*. doi:10.1101/064824.
- Dermastia, M., Nikolić, P., Chersicola, M., & Gruden, K. (2015). Transcriptional profiling in infected and recovered grapevine plant responses to ‘*Candidatus Phytoplasma solani*’. *Phytopathogenic Mollicutes*, 5(1-Suppl), 123–124.
- Francl, L. J. (2001). The disease triangle: a plant pathological paradigm revisited. The Plant Health Instructor. Retrieved from <http://www.apsnet.org/edcenter/instcomm/TeachingArticles/Pages/DiseaseTriangle.aspx>. doi:10.1094/PHI-T-2001-0517-01.
- Frías, M., Brito, N., & González, C. (2013). The *Botrytis cinerea* cerato-platanin BcSpl1 is a potent inducer of systemic acquired resistance (SAR) in tobacco and generates a wave of salicylic acid expanding from the site of application. *Molecular Plant Pathology*, 14(2), 191–196.
- Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software: an update. *ACM SIGKDD Explorations Newsletter*, 11, 10–18.
- Hamamouch, N., Li, C., Seo, P. J., Park, C.-M., & Davis, E. L. (2011). Expression of Arabidopsis pathogenesis-related genes during nematode infection. *Molecular Plant Pathology*, 12, 355–364.
- Hardwick, N. V. (2006). Disease forecasting. In B. M. Cooke, D. Gareth Jones, & B. Kaye (Eds.), *The epidemiology of plant disease* (pp. 239–267). Dordrecht: Springer.
- Harrell Jr, F. E. (2015). Regression modeling strategies. Springer. <https://link.springer.com/book/10.1007%2F978-3-319-19425-7>.
- Højsgaard, S., Halekoh, U., & Yan, J. (2006). The R package geepack for generalized estimating equations. *Journal of Statistical Software*, 15(2), 1–11. doi:10.18637/jss.v015.i02.
- Hren, M., Boben, J., Rotter, A., Kralj, P., Gruden, K., & Ravnikar, M. (2007). Real-time PCR detection systems for Flavescence dorée and Bois noir phytoplasmas in grapevine: comparison with conventional PCR detection and application in diagnostics. *Plant Pathology*, 56, 785–796.
- Hren, M., Nikolić, P., Rotter, A., Blejec, A., Terrier, N., Ravnikar, M., Dermastia, M., & Gruden, K. (2009a). “Bois noir” phytoplasma induces significant reprogramming of the leaf transcriptome in the field grown grapevine. *BMC Genomics*, 10(1), 460.
- Hren, M., Ravnikar, M., Brzin, J., Ermacor, P., Carraro, L., Bianco, P. A., Casati, P., Borgo, M., Angelini, E., Rotter, A., & Gruden, K. (2009b). Induced expression of sucrose synthase and alcohol dehydrogenase I genes in phytoplasma-infected grapevine plants grown in the field. *Plant Pathology*, 58(1), 170–180.
- Landi, L., & Romanazzi, G. (2011). Seasonal variation of defense-related gene expression in leaves from Bois noir affected and recovered grapevines. *Journal of Agricultural and Food Chemistry*, 59, 6628–6637.
- Lessio, F., Portaluri, A., Paparella, F., & Alma, A. (2015). A mathematical model of flavescence dorée epidemiology. *Ecological Modelling*, 312, 41–53.
- Maggi, F., Bosco, D., Galetto, L., Palmano, S., & Marzachi, C. (2017). Space-time point pattern analysis of flavescence dorée epidemic in a grapevine field: disease progression and recovery. *Frontiers in Plant Science*, 7, 1987. doi:10.3389/fpls.2016.01987.
- Maixner, M. (2006). Temporal behaviour of grapevines infected by type II of Vergilbungskrankheit (Bois noir). <http://www>.

- jki.bund.de/downloadFatPdf.php?file=2006_0071.pdf. Accessed 25 October 2016.
- Marcone, C. (2014). Molecular biology and pathogenicity of phytoplasmas. *Annals of Applied Biology*, 165(2), 199–221.
- Margaria, P., Abbà, S., & Palmano, S. (2013). Novel aspects of grapevine response to phytoplasma infection investigated by a proteomic and phospho-proteomic approach with data integration into functional networks. *BMC Genomics*, 14, 38.
- McNew, G. L. (1960). The nature, origin and evolution of parasitism. In A. E. Horsfall & J. G. Dimond (Eds.), *Plant Pathology: An advanced treatise* (pp. 19–69). New York: Academic Press.
- Musetti, R., Marabottini, R., Badiani, M., Martini, M., Sanita di Toppi, L., Borselli, S., Borgo, M., & Osler, R. (2007). On the role of H₂O₂ in the recovery of grapevine (*Vitis vinifera* cv. Prosecco) from Flavescence dorée disease. *Functional Plant Biology*, 34, 750–758.
- Musetti, R., Farhan, K., De Marco, F., Polizzotto, R., Paolacci, A., Ciaffi, M., Ermacor, P., Grisan, S., Santi, S., & Osler, R. (2013). Differentially-regulated defence genes in *Malus domestica* during phytoplasma infection and recovery. *European Journal of Plant Pathology*, 136, 13–19.
- Ouzounis, C. A. (2012). Rise and demise of bioinformatics? Promise and Progress. *PLoS Computational Biology*, 8, e1002487.
- Panassiti, B., Hartig, F., Breuer, M., & Biedermann, R. (2015). Bayesian inference of environmental and biotic factors determining the occurrence of the grapevine disease ‘bois noir’. *Ecosphere*, 6(8), 143.
- Prezelj, N., Nikolić, P., Gruden, K., Ravnikar, M., & Dermastia, M. (2013). Spatiotemporal distribution of flavescence dorée phytoplasma in grapevine. *Plant Pathology*, 62(4), 760–766.
- Prezelj, N., Covington, E., Roitsch, T., Gruden, K., Fagner, L., Weckwerth, W., Chersicola, M., Vodopivec, M., & Dermastia, M. (2016a). Metabolic consequences of infection of grapevine (*Vitis vinifera* L.) cv. “Modra frankinja” with Flavescence Dorée phytoplasma. *Frontiers in Plant Science*, 7, 711.
- Prezelj, N., Fagner, L., Weckwerth, W., & Dermastia, M. (2016b). Metabolome of grapevine leaf vein-enriched tissue infected with *Candidatus* Phytoplasma solani. *Mitteilungen Klosterneuburg Rebe und Wein, Obstbau und Früchteverwertung*, 66(S1), 74–78.
- Quaglino, F., Zhao, Y., Casati, P., Bulgari, D., Bianco, P. A., Wei, W., & Davis, R. E. (2013). “*Candidatus* Phytoplasma solani”, a novel taxon associated with stolbur- and bois noir-related diseases of plants. *International Journal of Systematic and Evolutionary Microbiology*, 63, 2879–2894.
- Quinlan, J. R. (1993). *C4.5: Programs for machine learning*. San Francisco: Morgan Kaufmann Publishers Inc..
- R Core Team. (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>. Accessed 11 Aug 2017.
- Santi, S., De Marco, F., Polizzotto, R., Grisan, S., & Musetti, R. (2013). Recovery from stolbur disease in grapevine involves changes in sugar transport and metabolism. *Frontiers in Plant Science*, 4, 171.
- Scholthof, K.-B. G. (2007). The disease triangle: pathogens, the environment and society. *Nature Reviews. Microbiology*, 5, 152–156.
- Spagnolo, A., Magnin-Robert, M., Alayi, T. D., Cilindre, C., Mercier, L., Schaeffer-Reiss, C., Van Dorsselaer, A., Clément, C., & Fontaine, F. (2012). Physiological changes in green stems of *Vitis vinifera* L. cv. Chardonnay in response to esca proper and apoplexy revealed by proteomic and transcriptomic analyses. *Journal of Proteome Research*, 11, 461–475.
- Starý, M., Válová, P., Šafářová, D., Lauterer, P., Ackermann, P., & Navrátil, M. (2013). Survey and molecular detection of Bois noir in vineyards of the Czech Republic— Short communication. *Horticultural Science (Prague)*, 40, 83–87.
- Strauss, E. (2009). Phytoplasma Research Begins to Bloom. *Science*, 325, 388–390.
- Taylor, R. A., Mordecai, E. A., Gilligan, C. A., Rohr, J. R., & Johnson, L. R. (2016). Mathematical models are a powerful method to understand and control the spread of Huanglongbing. *Peer J*, 4, e2642. doi:10.7717/peerj.2642.
- Van Damme, M., Huibers, R. P., Elberse, J., & Van den Ackerveken, G. (2008). Arabidopsis DMR6 encodes a putative 2OG-Fe(II) oxygenase that is defense-associated but required for susceptibility to downy mildew. *The Plant Journal*, 54(5), 785–793.
- van den Boogaart, G. K., Tolosana, R., & Bren, M. (2012). Compositions: compositional data analysis. <http://cran.r-project.org/package=compositions>. Accessed 25 October 2016.
- Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S* (Fourth ed.). New York: Springer.
- Weisstein, E. W. (2016). Nomogram. Retrieved from MathWorld— A Wolfram Web Resource. <http://mathworld.wolfram.com/Nomogram.html>. Accessed 25 October 2016.
- Zeilmaker, T., Ludwig, N. R., Elberse, J., Seidl, M. F., Berke, L., Van Doorn, A., Schuurink, R. C., Snel, B., & Van den Ackerveken, G. (2015). DOWNY MILDEW RESISTANT 6 and DMR6-LIKE OXYGENASE 1 are partially redundant but distinct suppressors of immunity in Arabidopsis. *The Plant Journal*, 81(2), 210–222.
- Zhao, Y., Davis, R. E., Wei, W., & Lee, I.-M. (2015). Should “*Candidatus* Phytoplasma” be retained within the order Achaeoplasmatales? *International Journal of Systematic and Evolutionary Microbiology*, 65, 1075–1082.