

## RESEARCH ARTICLE

# Protein markers of *Bursaphelenchus xylophilus* Steiner & Buhner, 1934 (Nickle, 1970) populations using quantitative proteomics and character compatibility

Sergio Ciordia<sup>1</sup>, Lee Robertson<sup>2</sup>, Susana C. Arcos<sup>2</sup>, María Rosa González<sup>2</sup>, María del Carmen Mena<sup>1</sup>, Paula Zamora<sup>3</sup>, Paulo Vieira<sup>4</sup>, Isabel Abrantes<sup>5</sup>, Manuel Mota<sup>4</sup>, Philippe Castagnone-Sereno<sup>6</sup> and Alfonso Navas<sup>2</sup>

<sup>1</sup> Unidad de Proteómica Centro Nacional de Biotecnología, CSIC, Madrid, Spain

<sup>2</sup> Department of Biodiversity and Evolutionary Biology, Museo Nacional de Ciencias Naturales, CSIC, Madrid, Spain

<sup>3</sup> Centro de Sanidad Forestal de Calabazanos, Consejería de Fomento y Medio Ambiente, Dirección General del Medio Natural, Junta de Castilla y León, Polígono Industrial de Villamuriel de Cerrato, Palencia, España

<sup>4</sup> NemaLab/ ICAAM (Instituto de Ciências Agrárias e Ambientais Mediterrânicas), Universidade de Évora, Herdade da Mitra, Évora, Portugal

<sup>5</sup> Centre for Functional Ecology, Department of Life Sciences, University of Coimbra, Calçada Martim de Freitas, Coimbra, Portugal

<sup>6</sup> INRA UMR1355/Université de Nice-SophiaAntipolis/CNRS UMR7254, Institut Sophia Agrobiotech, Sophia Antipolis, France

The Pine Wood Nematode (PWN) *Bursaphelenchus xylophilus* is a severe forest pathogen in countries where it has been introduced and is considered a worldwide quarantine organism. In this study, protein markers for differentiating populations of this nematode were identified by studying differences among four selected Iberian and one American population. These populations were compared by quantitative proteomics (iTRAQ). From a total of 2860 proteins identified using the public database from the *B. xylophilus* genome project, 216 were unambiguous and significantly differentially regulated in the studied populations. Comparisons of their pairwise ratio were statistically treated and supported in order to convert them into discrete character states, suggesting that 141 proteins were not informative as population specific markers. Application of the Character Compatibility methodology on the remaining 75 proteins (belonging to families with different biological functions) excludes 27 which are incompatible among them. Considering only the compatible proteins, the method selects a subset of 30 specific unique protein markers which allowed the compared classification of the Iberian isolates. This approach makes it easier search for diagnostic tools and phylogenetic inference within species and populations of a pathogen exhibiting a high level of genetic diversity.

Received: March 17, 2015  
Revised: November 6, 2015  
Accepted: December 29, 2015

## Keywords:

Animal proteomics / *Bursaphelenchus xylophilus* / Character compatibility / iTRAQ / Proteins population biomarkers / Quantitative proteomics



Additional supporting information may be found in the online version of this article at the publisher's web-site

**Correspondence:** Dr. Alfonso Navas, Department of Biodiversity and Evolutionary Biology, Museo Nacional de Ciencias Naturales, CSIC, calle José Gutiérrez Abascal 2, Madrid 28006, Spain  
**Fax:** +34-91-5645078  
**E-mail:** anavas@mncn.csic.es

**Abbreviation:** PWN, pinewood nematode

## 1 Introduction

The pinewood nematode (PWN) *Bursaphelenchus xylophilus* Steiner & Buhner, 1934 [1], native to North America (Canada and USA), constitutes a severe pathogen and pest in countries where it has been introduced [2]. The nematode was

## Significance of the study

The main focus of this paper is leading to characterize unique proteins markers for populations of the pinewood nematode *Bursaphelenchus xylophilus*, a severe forest pathogen in countries where it has been introduced and considered a worldwide quarantine organism. We have used contrasted statistics and the Character Compatibility method (Perfect Phylogeny) applied to quantitative proteomics (iTRAQ) as the most robust criteria of selection. The property required for this is to have binary characters, having only two possible states (mathematical factorization is used in case of ambiguity) in such a way that the set of all members which

have a state (0 or 1) for any character connect subtrees in a hierarchical manner. We demonstrate how quantitative proteomics, as a very precise method in order of detect significant differences in protein regulation level, can be used in combination with Character Compatibility. The potential for further studies including those on functional physiological approaches such as changes in virulence of pathogens or transmitting insects is a way of application. However this approach is also efficient as a general selection method of biomarkers, especially when hundreds or thousands of proteins are considered as potential targets.

introduced to Japan at the beginning of the XXth Century. In 1905 it was suggested to be linked to the pine wilt disease (PWD) [3] and confirmed to be the causal agent in 1970 [4]. *Bursaphelenchus xylophilus* was further spread to Manchuria (China) in the 1930s, as well as to other Asian countries including Korea and Taiwan [5]. The nematode is listed as a quarantine organism by the European and Mediterranean Plant Protection Organization and phytosanitary regulations have been implemented to prevent its introduction and spread within Europe (EU directive 2000/29/EC).

*Bursaphelenchus xylophilus* was first recorded in Europe (Portugal) in 1999 associated with *Pinus pinaster* [6] and the insect vector *Monochamus galloprovincialis* [7]. In 2010, it was reported from Madeira island [8]. In 2008, 2010, 2012 and 2013 *B. xylophilus* was detected in wood samples from four localities in Spain close to the Portuguese border in *P. pinaster* (Extremadura and Galicia) [9] and in *P. pinaster* and *P. radiata* (Salamanca) (official report to the European Commission from the Spanish authorities). Due to the pine wood trade from Portugal to Spain and the rest of Europe, infected pine material (logs, wood pallets and others pine derivatives such as chips and planks) near the Spanish border have been intercepted.

The most important *Pinus* species (*P. sylvestris*) in Europe are highly susceptible to *B. xylophilus* [10] and its life cycle is clearly associated to summer isotherms (above 20°C) [11]. For this reason, PWN is a serious threat for European forest ecosystems whose importance will increase in the foreseeable future (taking into account climatic change scenarios) in spite of the measures adopted for its control within the European Union [2] in addition to Portugal and Spain.

The selection processes to which this pathogen is subjected to dispersal pathways are keys for its evolution, which can be seen within the context of a metapopulation structure [12,13]. In this respect, the biological traits of the pathogen, including pathogenicity, aggressiveness or virulence, can be exploited to decipher its evolutionary history or phylogeny. *B. xylophilus* has a wide geographical range and can be considered as one ubiquitous species occupying very diverse pine forest

ecosystems throughout the world. Although the taxonomic resolution at the species level is very stable, a wide range of *B. xylophilus* intraspecific differences have been demonstrated [14]. The assessment of differences among populations through various DNA molecular marker approaches has been the main aspect of prospective studies on *B. xylophilus* [15,16]; however few results have been published on proteome differences regarding this important plant quarantine pathogen [17] although its genome sequence has recently been made available [18]. This is all the more surprising as new methodological developments in proteomics could be successfully applied to research into the genetic, pathogenic and evolutionary characteristics of parasites [13,19–21]. In particular, wide analyses and comparison of protein regulation profiles support the statement that the proteome could be considered the “molecular phenotype” of genome expression [22,23]. Indeed, the regulation of proteins can be treated as quantitative inheritance characters [24] and its codification in discrete states would facilitate the finding of unambiguous proteins markers with taxonomic, evolutionary and phylogenetic value. Here, we applied such an approach to the PWN in order to discriminate and select a unique set of candidate population-specific marker proteins within this species.

## 2 Materials and methods

### 2.1 Population background of *Bursaphelenchus xylophilus* in the Iberian Peninsula

Nematodes were unambiguously identified by microscopy, based on morphological and morphometrical features following the diagnostic protocols for regulated pests regarding the PWN (EPPO Standards PM7) [5]. When the number of individuals was abundant, their identity was confirmed by ITS-RFLP [25] and when the number of individuals was small, their identification was based on species-specific satellite DNA (satDNA) [26] that have been shown to be excellent taxonomic markers for nematodes of agro-forestry interest

[27], which accomplish the criteria of high and versatile variability in the DNA sequences [15, 16]. To this end, the systematization of representative satDNA sequences of *B. xylophilus* isolates from the Iberian Peninsula has been carried out in order to define genetic or phylogenetics groups within this pest (reported to EU Commission, REPHRAME project, data not shown). For this experiment two *B. xylophilus* populations from Portugal (Bx1406PT, Bx2709PT) and two from Spain (Bx7333SA, BxPO) have been selected, being representative members of the different satDNA groups reported to EU Commission (REPHRAME project). These populations were compared with an American population (US10) as original reference of worldwide dispersion. Two biological replicates were performed for each population and three technical replicates for each biological replicate.

## 2.2 Extraction of nematodes proteins

Samples of a mixture of male and female specimens for each isolate were crushed by hand with a pestle in a 1.5 mL Eppendorf tube on ice to prevent protein degradation. The crushed extract was suspended immediately in 200  $\mu$ L of lysis buffer (7 M urea, 2 M thiourea, 2% (v/v) Triton-100, IPG buffer pH 3–11 2% (v/v), 40 mM DTT with protease and phosphatase inhibitors). Total proteins were precipitated using methanol/chloroform [28]. Protein pellets were resuspended and denatured in 7 M Urea/2 M Thiourea/100 mM TEAB, pH 7.5. Protein concentration was estimated using the RC DC Protein Assay kit (BIO-RAD).

## 2.3 Protein digestion and tagging with iTRAQ-4-plex<sup>®</sup> reagent

For digestion, 40  $\mu$ g of protein from each condition/population was reduced with 2  $\mu$ L of 50 mM Tris(2-carboxyethyl) phosphine (TCEP, AB SCIEX), pH 8.0, at 37°C for 60 min and followed by 2  $\mu$ L of 200 mM cysteine-blocking reagent (methyl methanethiosulfonate (MMTS, Pierce) for 10 min at room temperature. Samples were diluted up to 1 M Urea concentration with 25 mM TEAB. Digestions were initiated by adding sequence grade-modified trypsin (Sigma-Aldrich) to each sample in a ratio of 1:20 (w/w), which were then incubated at 37°C overnight on a shaker. Sample digestions were evaporated to dryness.

Digested samples were labeled at room temperature for 2 h with iTRAQ Reagent Multi-plex kit (AB SCIEX) according to the manufacturer's instructions. The iTRAQ labeling was performed separately in two 4-plex designs using tags 114, 115, 116 and 117 (Labeling Scheme, Table 1S). In the first labeling (iTRAQ1), tags were used for *B. xylophilus* US10-1, Bx2709PT-1 and US10-2, respectively. In parallel, in the second labeling (iTRAQ2), tags were used for *B. xylophilus* US10-3, Bx7333SA-1, BxPO-1 and US10-4. The third labeling (iTRAQ3), tags were used for US10-5, Bx2709PT-2, BxPO-2

and US10-6. The fourth labeling (iTRAQ4) tags were used for US10-7, Bx1406PT-1, Bx7333SA-2 and US10-8. Finally the last labeling (iTRAQ5) the tags were for labeling USA10-9 and Bx1406PT. After labeling, the samples were pooled, dried and desalted using a SEP-PAK C18 Cartridge (Waters). Finally, the cleaned tryptic peptides were evaporated to dryness and stored at –20°C until analysis.

## 2.4 Liquid chromatography and mass spectrometry analysis

A 2.5  $\mu$ g aliquot of the resulting mixture was subjected to 2D-nano LC ESI-MS/MS analysis using a nano liquid chromatography system (Eksigent Technologies nanoLC Ultra<sup>™</sup> 1D plus, AB SCIEX) coupled to high-speed TripleTOF<sup>®</sup> 5600 mass spectrometer (AB SCIEX) with a duo spray ionization source. The analytical column used was a silica-based reversed phase column C18 ChromXP 75  $\mu$ m  $\times$  15 cm, 3  $\mu$ m particle size and 120 Å pore size (Eksigent Technologies, AB SCIEX). The trap column was also a C18 ChromXP model, in line with the analytical column. The nano-pump provided a flow rate of 300 nL/min and was operated under gradient elution conditions using 0.1% formic acid in water as mobile phase A, and 0.1% formic acid in acetonitrile as mobile phase B. Peptides with iTRAQ labels were separated using a 200 nL/min gradient ranging from 2 to 90% mobile phase B (mobile phase A: 2% acetonitrile, 0.1% formic acid; mobile phase B: 100% acetonitrile, 0.1% formic acid). The injection volume was 5  $\mu$ L.

Data were acquired using a TripleTOF 5600 System (AB SCIEX) with the following settings: ion spray voltage floating (ISVF) 2800 V, curtain gas (CUR) 20 L/h, interface heater temperature (IHT) 150°C, ion source gas 1 (GS1) 20 L/h, declustering potential (DP) 85 V. All data were collected in information-dependent acquisition (IDA) mode using Analyst TF 1.5 software (AB SCIEX). A 0.25 s MS survey scan was performed in the mass range 350–1250 Da, followed by 30 MS/MS scans of 150 ms in the mass range 100–1800 (total cycle time: 4.04 s). Switching criteria were set to ions with a mass/charge ratio ( $m/z$ ) of >350 but <1250, with a charge state of 2–5 and an abundance threshold of more than 70 counts (cps). Former target ions were excluded for 20 s. The IDA rolling collision energy (CE) parameters script option was used for automatically controlling the CE. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium [29] via the PRIDE partner repository with the dataset identifier PXD003129 and 10.6019/PXD003129.

## 2.5 Data analysis

MS and MS/MS data obtained for pooled samples were processed using Analyst<sup>®</sup> TF 1.5.1 software (AB SCIEX). Raw data file conversion tools generated mgf files which were



Table 1. Continued

61.	1X1111X111XXX1XXX11XX1111111111XX111111111111111X1X1XXX1XX1XXX1111X1X1XX1111X1XX
62.	111
63.	111X11X11XX1XXXX1X1X111111X111XX111X1111X1XXX111XXX1XXX1XX11X111XXX1111X1X1
64.	111
65.	1X1X11X11X1X1XX1XX1XX11111X111X1111X1111X1XX11X1XX1111XXXX11X1X11X1X11111XX
66.	1X1111X111XXX1XXX11XX1111111111XX111111111111111X1X1XXX1XX1XXX1111X1X1XX1111X1XX
67.	1X1X11X11X1X1XX1XX1XX11111X111X1111X1111X1XX11X1XX1111XXXX11X1X11X1X11111XX
68.	111X11X11XX1XXXX1X1X111111X111XX111X1111X1XXX111XXX1XXX1XX11X111XXX1111X1X1
69.	111
70.	111
71.	111111X111XXX1X1X11XX1111111111XX111111111111111X1XX11XX1XXX1111111111111111XX
72.	1X1X11X11X1X1XX1XX1XX11111X111X1111X1111X1XX11X1XX1111XXXX11X1X11X1X11111XX
73.	111111X111XXX1X1X11XX1111111111XX111XX
74.	1X11111XXX1X1X1XX1X1X1111X1111XX1111111111X1X1XXX1XX1XXX111X1X1XXXX111XXX11
75.	111X111XXXXXXXXXXXXX1X1111XX1X11XX11XX111X1XXX1X1XXX111XXXX11X111XXX111XXX11

This matrix is obtained after factorization in 133 characters (1 if compatible; X if not). Dark shaded lines represent the 27 characters which are not compatible with the rest.

compared against the *Bursaphelenicus xylophilus* database (which contains 36 148 protein-coding genes and their corresponding reversed entries), using the MASCOT Server v. 2.5.0 (Matrix Science). The search parameters were: enzyme, trypsin; allowed missed cleavages, 1; fixed modifications, iTRAQ4plex (N-term and K) and beta-methylthiolation of cysteine; variable modifications, oxidation of methionine. Peptide mass tolerance was set at  $\pm 25$  ppm for precursors and 0.05 Da for fragment masses. The confidence interval for protein identification was set to  $\geq 95\%$  ( $p < 0.05$ ); only peptides with an individual ion score above the 1% false discovery rate (FDR) threshold were considered correctly identified. Only proteins with at least two quantified peptides were considered in quantitation statistical model. A 5% quantitation FDR threshold was used to consider proteins differentially regulated.

## 2.6 Coding protein regulation values and Strict Character Compatibility Analysis

In order to select the proteins which can be considered as unambiguous markers, quantification of proteins expression has been coded to obtain significant discrete states (0, 1). Codification (0, 1) was based on level of significance of ratio ( $\text{Log}_2$  ratios), and according to the criterion of significance for continuous character according to [30–32]. For definition of marker importance, differential proteins were subject to strict character compatibility (used for what is also called "Perfect Phylogeny"[33–35]; it requires that every change be done only once. Two binary characters are incompatible if and only if all four possible combinations (0–0, 0–1, 1–0, 1–1) are present in the dataset. It has been demonstrated that a group of two-state characters which are pairwise compatible would be jointly compatible [36, 37]. The analysis of character compatibility was performed by CLIQUE software. However because of some proteins (characters) states were not unambiguously assigned in the two biological replicates [they were

considered as unknown states (?) in Supporting Information Table S2], a mathematical factorization by FACTOR software which convert multiple states in binary states (input instruction in Supporting Information Table S3a) was done before application of CLIQUE (input instruction in Supporting Information Table S3b). FACTOR and CLIQUE software are from Phylip packages programs (version 3.695 [38]. For representing the phylogenetic relationships of the populations we have used the tree editor of PAUP [39].

## 3 Results and discussion

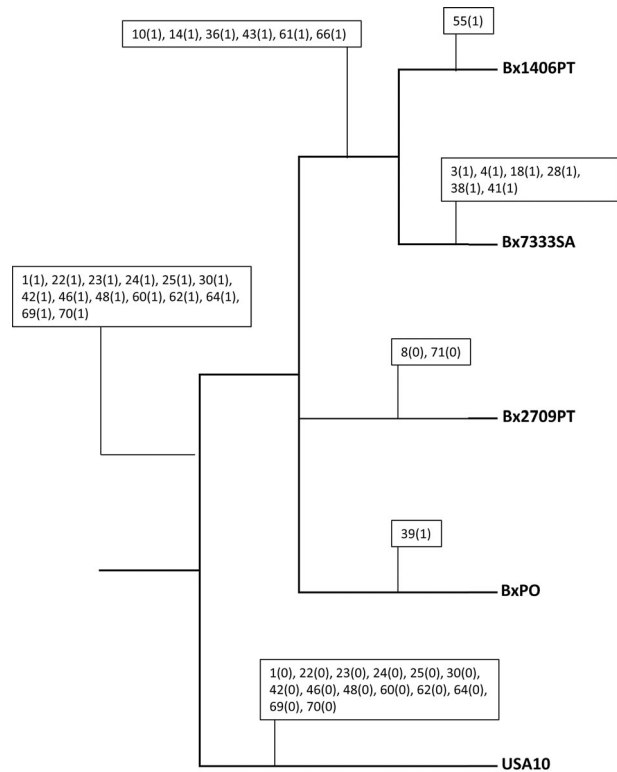
In total, 2860 proteins were unambiguously identified by the iTRAQ experiments (Supporting Information Table S1). Criteria for identification of differential regulation among them, were at least two peptides showing a 95% level of significance ( $p < 0.05$ ) of differentially regulated proteins for 1% FDR at quantitation level, measured by  $\text{Log}_2$  ratios of the relative protein/peptide abundances of all populations compared with the reference (USA10). Two hundred and sixteen proteins accomplished this criterion (Supporting Information Table S4). These proteins were assigned according to their accession number in the sequenced genome of *B. xylophilus* [17]; their description as metabolic and functional proteins and homology with other nematodes are available in the Supporting Information (Table S1). Because all isolates were compared with USA10, the reference ratio of  $\text{Log}_2$  ratios approximates 1 (i.e., ratio of comparison of USA10 with itself) defining in this way the tendency of down or upregulation of proteins. The overall level of significance were obtained considering the five  $\text{Log}_2$  ratios (Bx1406PT/USA10, Bx2709PT/USA10, Bx7333SA/USA10, BxPO/USA10, USA10/USA10) and their relations with the values of average ( $\bar{X}$ ) and standard deviation ( $\delta x$ ) ratios (gap-coding criterion of Archie [30]) (Supporting Information Table S2). According to that, the level of significance would be established by  $2\delta x$  if  $\delta x$  is  $\leq \bar{X}$  or  $\delta x$  if  $\delta x \geq \bar{X}$ . This allows the conversion of the significantly regulated

proteins as continuous characters in discrete or binary states (1, overregulated protein; 0, downregulated protein) [40, 41]. Binary conversion allowed to recognize that 21 proteins are not informative (state 1 in all the populations) and 58 have ambiguity at least for one population (it can be 0 or 1), referred as (?) in Supporting Information Table S2. In total, 75 proteins are clearly informatives. Mathematical factorization converts the unambiguous character (?) in two new possible states for using character compatibility approach (resolved with 133 characters, 58 factorized; Supporting Information Table S3b).

### 3.1 Character compatibility

The output of CLIQUE software has the option to get the compatibility of the original characters (proteins) although factorization was used. Consequently, among these 75 significant proteins (i.e., original characters), 27 were not compatible with the rest (Table 1) indicating that they can not be considered as markers of the studied *B. xylophilus* populations. From the seven assumptions of compatibility character methods which are summarized in the CLIQUE software from PHYLIP Program<sup>®</sup> [38, 42], we want to point out that each character and lineage is considered as evolutionary independent and the ancestral state is not known. The largest group of proteins (48 proteins) was mutually compatible and defined the branching pattern of the populations (Fig. 1). Twenty eight proteins were upregulated in the *B. xylophilus* Iberian populations compared with the American USA10 and only one was downregulated. In Fig. 1, proteins expressing relationships are represented. Iberian populations constitute a differentiated group which was marked by the largest “clique” of upregulated proteins (14 proteins). On contrary, what can be considered as the “primitive” condition for those proteins, are downregulated in the American population USA10. The population Spanish BxPO is clearly characterized by one up-regulated protein while the population Bx2709PT is characterized by two downregulated proteins which would constitutes a regression to the states of the American USA10. Two populations (Portuguese 1406PT and Spanish Bx7333SA) are grouped by other big “clique” formed by six up-regulated proteins. The Spanish BxPO and Portuguese Bx2709PT do not form a group with specific markers in common.

The evolutionary history of *B. xylophilus* has to be reconstructed without the support of geographical evidence due to the human impact on its worldwide dispersion [43]. Available molecular markers such as homologous DNA probes [44], ITS-RFLP [45, 46] and especially polymorphic microsatellite loci [47, 48] have been used to demonstrate the variability of different geographical isolates of *B. xylophilus*. Such variability is suitable to detect the occurrence of genetically differentiated populations, and can eventually be used for establishing a probable origin of propagation, especially when phytosanitary programs have been implemented in a country [8]. The putative dispersal history of *B. xylophilus* is linked to human



**Figure 1.** Phylogenetic tree obtained by Character Compatibility Analysis. There are 30 proteins which are unique markers for the populations or groups. Numbers in rectangles are the non-factorized character used for Compatibility Analysis which correspond to proteins accession number of Supporting Information Table S2.

influence while genetic analyses supports the hypothesis that Asian populations originated from a single introduction of the pest from North America [49, 50]. This is coincidental with our results when tree is rooted because this option located the American population (USA 10) as the outgroup of the Iberian ones. The unrooted option did not alter the results regarding the location of protein markers (data not shown). However, a recent population genetics study highlighted the need for more sophisticated methods of invasion route inference and broader sampling to conclusively determine the origin of the Iberian PWN outbreaks [48]. The results of this study support the hypothesis that there are at least three genetics groups of *B. xylophilus* populations in Iberia Peninsula within its specific taxonomic proteins markers. Bx2709PT and BxPO populations are independent one to other and very different that populations 1406PT-Bx7333SA. The accession number of markers proteins is in the Supporting Information Table S2.

In spite of the importance of this nematode as a pine forest pathogen (extensive to other conifers), few attempts have been made to reconstruct its phylogeny at the species level, taking into account that some authors consider *B. xylophilus* as a supraspecies or a species complex [51, 52]. The

proteomics approach presented here would help characterizing the level of genetic diversity in order to obtain groups of populations/isolates to be tested in assays for host resistance, considering that populations/isolates which share common ancestor or origin, also share pathogenic or biological features. Such a differential interaction, comparing the aggressiveness or virulence of different isolates or populations, has been demonstrated for *B. xylophilus* inoculated onto various *P. thunbergii* genotypes [14].

Character compatibility method was proposed as extremely suitable for data such as those obtained by proteomics [19]. These data are clear when a matrix of presence and absence can be constructed (e.g., presence or absence of protein spots in a gel) [53]. Gupta and Sneath [54] used character compatibility in conserved protein sequences; these authors used those sites in the alignments where only two amino acids were present. They selected them and such sites were converted into a binary file of "0, 1" characters, resolving in this way/by this method the phylogenetic topology of 23 proteobacterial species. However, when data are expressed as continuous values with reliable confidence as are those obtained by iTRAQ analysis, they can be converted into discrete (binary) states and used with proper methodology to potential the quantitative proteomics. The basic concepts and mathematical foundations of character compatibility have been used since 1969 [37, 55–60]. This method has primarily been used with morphological characters even when they are multistate [61] and its applicability to molecular sequence data has not been explored [42] until Gupta and Sneath [54].

#### 4 Concluding remarks

The reliability of compatibility analysis has scarcely been investigated in the field of proteomics. The main limitation of character compatibility algorithms is that it works mainly with binary character states [37, 42]. Nevertheless, due to the extraordinary confident limits imposed by iTRAQ methodology, the statistical conversion of continuous characters ( $\log_2$  ratios of the relative protein/peptide abundances) by contrasted statistical approach [30] in binary states, allowed the identification of unambiguous proteins markers which are specific of populations. The concepts applied in this study were developed by several investigators [37, 57–60] to prove the Pairwise Compatibility Theorem which states that if provided all characters, these are binary (i.e., have only two states, such as 0 or 1, or present or absent, or can be recoded into this form), if all pairs are compatible, the entire clique is compatible with a single tree, it means the "perfect phylogeny". In other words, a group of two-state characters which were pairwise compatible would be jointly compatible [36, 37]. Based on comparative proteomics analyses we propose a method for evaluating reliable precise data obtained by quantitative proteomics in order obtain unique protein markers. However, the proposed methodology (Character Compatibility approach after significant statistical evaluation) is very reliable and can be used

not only for taxonomic or phylogenetic studies, but also in other studies in which decisions have to be taken based on robust criteria. So, the search for disease biomarkers, diagnosis, epidemiology, population genetics and tissue comparison is strengthened and easier affordable.

*This work has been funded by the CSIC (PI:201130E033), and the EU project REPHRAME FP7-KBBE-2010-4. We are grateful to Adela Abelleira (Estación Fitopatológica de Areiro, Pontevedra) for collaboration in supplying the nematode BxPO population. The proteomic analysis was performed in the proteomics facility of National Center for Biotechnology (CNB-CSIC) that belongs to ProteoRed, PRB2-ISCIII, supported by grant PT13/0001.*

*The authors have declared no conflict of interest.*

#### 5 References

- [1] Nickle, W. R. A taxonomic review of the genera of the Aphelenchoidea (Fuchs, 1937) Thorne, 1949 (Nematoda: Tylenchida). *J. Nematol.* 1970, 2, 375–392.
- [2] Mota, M., Vieira, P. (Eds). *Pine Wilt Disease: A Worldwide Threat to Forest Ecosystems*. Springer, Dordrecht 2008.
- [3] Mamiya, Y. Pathology of the pine wilt disease caused by *Bursaphelenchus xylophilus*. *Ann. Rev. Phytopathol.* 1983, 21, 201–220.
- [4] Kiyohara, T., Tokushige, Y. Inoculation experiments of a nematode *Bursaphelenchus* sp. onto pine trees. *Jpn. J. Forestry Soc.* 1971, 53, 210–218.
- [5] EPPO: PM 7/4 (3) *Bursaphelenchus xylophilus*. *EPPO Bull.* 2013, 43, 105–118.
- [6] Mota, M. M., Braasch, H., Bravo, M. A., Penas, A. C. et al., First report of *Bursaphelenchus xylophilus* in Portugal and in Europe. *Nematology* 1999, 1, 727–734.
- [7] Sousa, E., Bravo, M., Pires, J., Naves, P. et al., *Bursaphelenchus xylophilus* (Nematoda; Aphelenchoididae) associated with *Monoctonus galloprovincialis* (Coleoptera; Cerambycidae) in Portugal. *Nematology* 2001, 3, 89–91.
- [8] Fonseca, L., Cardoso, J. M. S., Lopes, A., Pestana, M. et al., The pinewood nematode, *Bursaphelenchus xylophilus*, in Madeira Island. *Helminthologia* 2012, 49, 96–103.
- [9] Robertson, L., Arcos, S. C., Escuer, M., Merino, R. S. et al., Incidence of the pinewood nematode *Bursaphelenchus xylophilus* Steiner & Buhner, 1934 (Nickle, 1970) in Spain. *Nematology* 2011, 13, 755–757.
- [10] Braasch, H. Host and pathogenicity tests with pine wood nematode (*Bursaphelenchus xylophilus*) from North America under Central European weather conditions. *Nachrichtenblatt Deutschen Pflanzenschutzdienstes* 1997, 49, 209–214.
- [11] Melakeberhan, H., Rutherford, T. A., Webster, J. M. Influence of temperature on reproduction of *Bursaphelenchus xylophilus* and *Pinus sylvestris* mortality. *Nematologica* 1992, 38, 80–87.
- [12] Hanski, I. In the midst of ecology, conservation, and competing interests in the society. *Ann. Zool. Fenn.* 2002, 39, 183–186.

- [13] Biron, D. G., Loxdale, H. D., Ponton, F., Moura, H. et al., Population proteomics: an emerging discipline to study metapopulation ecology. *Proteomics* 2006, *6*, 1712–1715.
- [14] Akiba, M., Ishihara, M., Sahashi, N., Nakamura, K. et al., Virulence of *Bursaphelenchus xylophilus* isolated from naturally infested pine forests to five resistant families of *Pinus thunbergii*. *Plant Dis.* 2012, *96*, 249–252.
- [15] François, C., Castagnone, C., Boonham, N., Tomlinson, J. et al., Satellite DNA as a target for TaqMan real-time PCR detection of the pinewood nematode, *Bursaphelenchus xylophilus*. *Mol. Plant Pathol.* 2007, *8*, 803–809.
- [16] Vieira, P., Castagnone, C., Mallez, S., Espada, M. et al., Sequence variability of the MspI satellite DNA family of the pinewood nematode *Bursaphelenchus xylophilus* at different geographic scales. *Mol. Phylog. Evol.* 2014, *70*, 120–129.
- [17] Shinya, R., Morisaka, H., Kikuchi, T., Takeuchi, Y. et al. Secretome Analysis of the Pine Wood Nematode *Bursaphelenchus xylophilus* reveals the tangled roots of parasitism and its potential for molecular mimicry. *PLoS ONE* 2013, *8*, e67377. doi:10.1371/journal.pone.0067377.
- [18] Kikuchi, T., Cotton, J. A., Dalzell, J. J., Hasegawa, K. et al., Genomic insights into the origin of parasitism in the emerging plant pathogen *Bursaphelenchus xylophilus*. *PLoS Pathog* 2011, *7*, e1002219.
- [19] Navas, A., Albar, J. P. Application of proteomics in phylogenetic and evolutionary studies. *Proteomics* 2004, *4*, 299–302.
- [20] Telleria, J., Biron, D. G., Brizard, J. P., Demetree, E. et al., Phylogenetic character mapping of proteomic diversity shows high correlation with subspecific phylogenetic diversity in *Trypanosoma cruzi*. *Proc. Natl. Acad. Sci. USA*, 2010, *107*, 20411–20416.
- [21] Holzmüller, P., Grébaud, P., Brizard, J. P., Berthier, D. et al., “Pathogeno-proteomics”: toward a new approach of host-vector-pathogen interactions. *Ann. NY Acad. Sci.* 2008, *1149*, 66–70.
- [22] Feder, M. E., Mitchell-Olds, T. Evolutionary and ecological functional genomics. *Nat. Rev. Genet.* 2003, *4*, 651–657.
- [23] Kellner, R. Proteomics. Concept perspectives. *Fresenius J. Anal. Chem.* 2000, *366*, 517–524.
- [24] Di, G., You, W., Yu, J., Wang, D. et al., Genetic changes in muscle protein following hybridization between *Haliothis diversicolor* Reeve Japan and Taiwan populations, revealed using a proteomic approach. *Proteomics* 2013, *13*, 845–859.
- [25] Burgermeister, W., Metge, K., Braasch, H., Buchbach, E. ITS-RFLP patterns for differentiation of 26 *Bursaphelenchus* species (Nematoda: Parasitaphelenchidae) and observations on their distribution. *Russ. J. Nematol.* 2005, *13*, 29–42.
- [26] Castagnone, C., Abad, P., Castagnone-Sereno, P. Satellite DNA-based species-specific identification of single individuals of the pinewood nematode *Bursaphelenchus xylophilus* (Nematoda: Aphelenchoididae). *Eur. J. Plant Pathol.* 2005, *112*, 191–193.
- [27] Grenier, E., Castagnone-Sereno, P., Abad, P. Satellite DNA sequences as taxonomic markers in nematodes of agronomic interest. *Parasitology Today* 1997, *13*, 398–401.
- [28] Wessel, D., Flugge, U. I. A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. *Anal. Biochem.* 1984, *138*, 141–143.
- [29] Vizcaino, J. A., Deutsch, E. W., Wang, R., Csordas, A. et al., ProteomeXchange provides globally co-ordinated proteomics data submission and dissemination. *Nature Biotechnol.* 2014, *30*, 223–226. PubMed PMID:24727771.
- [30] Archie, J. W. Methods for coding variable morphological features for numerical taxonomic analysis. *Syst. Zool.* 1985, *34*, 326–345.
- [31] Cranstron, P. S., Huphries, C. J. Cladistic and computers: achironomid conundrum. *Cladistics* 1988, *4*, 72–92.
- [32] Thiele, K., Ladiges, P. Y. A cladistics analysis of Angophora Cav. (Myrtaceae). *Cladistics* 1988, *4*, 23–42.
- [33] Fernandez-Baca. The Perfect Phylogeny Problem, in: Cheng, X., Du, D. Z. (Eds.), *Combinatorial Optimization: Stainer Trees in Industry*, 2001, Kluwer Academic Publishers, Dordrecht, pp. 203–234.
- [34] Pe’er, I., Pupko, T., Shamir, R., Sharan, R. Incomplete directed perfect phylogeny. *SIAM J. Comput.* 2004, *33*, 590–607.
- [35] Bonizzoni, P., Braghin, C., Dondi, R., Trucco, G. The binary perfect phylogeny with persistent characters. *Theoret. Comp. Sci.* 2012, *454*, 51–63.
- [36] Estabrook, G. F., Johnson, Jr, C. S., McMorris, F. R. An idealized concept of the true cladistic character. *Math. Biosci.* 1975, *23*, 263–272.
- [37] Estabrook, G. F., Johnson, Jr, C. S., McMorris, F. R. A mathematical foundation for the analysis of cladistic character compatibility. *Math. Biosci.* 1976, *29*, 181–187.
- [38] Felsenstein, J. PHYLIP (Version 3.695) 2014. <http://evolution.genetics.washington.edu/phylip.html>.
- [39] Swofford, D. *PAUP. Phylogenetic Analysis Using Parsimony (and Other Methods)*. 2002. Sinauer Associates, Inc., Publishers. Sunderland, Massachusetts. <http://paup.csit.fsu.edu/index.html>
- [40] Goldman, N. Methods for discrete coding of morphological characters for numerical analysis. *Cladistics* 1988, *4*, 59–71.
- [41] García-Cruz, J., Sosa, V. Coding quantitative character data for phylogenetic analysis: a comparison of five methods. *Syst. Bot.* 2006, *31*, 302–309.
- [42] Felsenstein, J., *Inferring Phylogenies*. Sinauer Associates, Sunderland, MA 2004.
- [43] Robinet, C., Roques, A., Pan, H. Y., Fang, G. et al. Role of human-mediated dispersal in the spread of the pinewood nematode in China. *PLoS ONE* 2009, *4*, e4646.
- [44] Tàres, S., Abad, P., Bruguier, N., de Guiran, G. Identification and evidence for relationships among geographical isolates of *Bursaphelenchus* spp. (pinewood nematode) using homologous DNA probes. *Heredity* 1992, *68*, 157–164.
- [45] Iwahori, H., Kanzak, N., Futai, K., in: Futai, K., Togashi, K., Ikeda, T. (Eds.), *Sustainability of Pine Forests in Relation to Pine Wilt and Decline. Proc. Int. Symp.* Nakanishi Printing, Tokyo 1998, pp. 35–38.
- [46] Lu, W., Chen, X. X., Zheng, J. W., Wu, H. et al., A comparative study on ITS region of *Bursaphelenchus xylophilus* and



- B. mucronatus* by PCR-RFLP. *J. Agri. Biotech.* 2001, 9, 387–390.
- [47] Mallez, S., Castagnone, C., Espada, M., Vieira, P. et al., First insights into the genetic diversity of the pinewood nematode in its native area using new polymorphic microsatellite loci. *Plos One* 2013, 8, e59165.
- [48] Mallez, S., Castagnone, C., Espada, M., Vieira, P. et al. World-wide invasion routes of the pinewood nematode: what can we infer from population genetics analyses? *Biol. Invasions* 2015, 17, 1199–1213.
- [49] Zhang, K., Liu, H., Sun, J., Liu, J. et al. Molecular phylogeny of geographical isolates of *Bursaphelenchus xylophilus*: Implications on the origin and spread of this species in China and worldwide. *J. Nematol.* 2008, 40, 127–137.
- [50] Cheng, X. Y., Cheng, F. X., Xu, R. M., Xie, B. Y. Genetic variation in the invasive process of *Bursaphelenchus xylophilus* (Aphelenchida: Aphelenchoididae) and its possible spread routes in China. *Heredity* 2008, 100, 356–365.
- [51] De Guiran, G., Bruguier, N. Hybridization and phylogeny of the pine wood nematode (*Bursaphelenchus* spp.). *Nematologica* 1989, 35, 321–330.
- [52] Bolla, R. I., Boschert, M. Pinewood nematode species complex-interbreeding potential and chromosome-number. *J. Nematol.* 1993, 25, 227–238.
- [53] Navas, A., López, J. A., Espárrago, G., Camafeita, E. et al., Protein variability in *Meloidogyne* spp. (Nematoda: Meloidogynidae) revealed by two-dimensional gel electrophoresis and mass spectrometry. *J. Proteome Res.* 2002, 1, 421–427.
- [54] Gupta, R. S., Sneath, P. H. Application of the character compatibility approach to generalized molecular sequence data: branching order of the proteobacterial subdivisions. *J. Mol. Evol.* 2007, 64, 90–100.
- [55] Le Quesne, W. J. A method of selection of characters in numerical taxonomy. *Syst. Zool.* 1969, 18, 201–205.
- [56] Le Quesne, W. J. The uniquely evolved character concept and its cladistics application. *Syst. Zool.* 1974, 23, 513–517.
- [57] Estabrook, G. F., McMorris, F. R. When is one estimate of evolutionary relationships a refinement of another? *J. Math. Biol.* 1980, 10, 367–373.
- [58] Sneath, P. H. A., Sackin, M. J., Ambler, R. P. Detecting evolutionary incompatibilities from protein sequences. *Syst. Zool.* 1975, 24, 311–332.
- [59] Wilson, E. O. A consistency test for phylogenies based on contemporaneous species. *Syst. Zool.* 1965, 14, 214–220.
- [60] Meacham, C. A., Estabrook, G. F. Compatibility methods in systematics. *Ann. Rev. Ecol. Syst.* 1985, 16, 431–446.
- [61] Camacho, A. I., Bello, E., Estabrook, G. F. A statistical approach to the evaluation of characters to estimate evolutionary relationships among the species of the aquatic subterranean genus, *Iberobuthynella* (Crustacea, Syncarida). *Biol. J. Linn. Soc.* 1997, 60, 221–241.