



## Sequence variability of the *MspI* satellite DNA family of the pinewood nematode *Bursaphelenchus xylophilus* at different geographic scales



Paulo Vieira<sup>a,\*</sup>, Chantal Castagnone<sup>b</sup>, Sophie Mallez<sup>b</sup>, Margarida Espada<sup>a</sup>, Alfonso Navas<sup>c</sup>, Manuel Mota<sup>a,d</sup>, Philippe Castagnone-Sereno<sup>b</sup>

<sup>a</sup> NemaLab/ICAAM, Instituto de Ciências Agrárias e Ambientais Mediterrânicas, Universidade de Évora, Núcleo da Mitra, Ap. 94, 7002-554 Évora, Portugal

<sup>b</sup> INRA, UNS, CNRS, UMR1355 Institut Sophia Agrobiotech, 400 route des Chappes, F-06603 Sophia Antipolis, France

<sup>c</sup> Dpto Biodiversidad y Biología Evolutiva, Museo Nacional de Ciencias Naturales, Consejo Superior de Investigaciones Científicas, Calle José Gutiérrez Abascal, 2 Madrid 28006, Spain

<sup>d</sup> INIAV/Unidade Estratégica de Investigação e Serviços de Sistemas Agrários e Florestais e Sanidade Vegetal Av. da República, Quinta do Marquês, 2784-159 Oeiras, Portugal

### ARTICLE INFO

#### Article history:

Received 3 May 2013

Revised 16 September 2013

Accepted 18 September 2013

Available online 27 September 2013

#### Keywords:

Pinewood nematode  
Aphelenchoididae  
Satellite DNA library  
Sequence variability  
Homogenization

### ABSTRACT

Tandemly repeated sequences known as satellite DNA (satDNA) generally exhibit complex evolutionary patterns of concerted evolution in which mutations are homogenized and fixed in a stochastic process of molecular drive. Here, the nucleotide variability of the *MspI* satDNA family of the pinewood nematode *Bursaphelenchus xylophilus* is analyzed in order to understand the evolutionary dynamics of satDNA at the intraspecific level. A total of 425 *MspI* monomer units, either PCR-amplified from isolates of local (Peninsula of Setúbal, Portugal) or worldwide origin, or retrieved from the *B. xylophilus* genome sequence, were characterized and compared. Whatever their origin, sliding window analysis of sequence variability patterns among monomers revealed low, moderate and highly variant domains, indicating that variable levels of evolutionary constraint may act upon the entire monomers. The phylogenetic inference based on the different sets of *MspI* satDNA family for this species shows a broad polymorphism of the individual monomers, which were distributed into four main clusters. However, such clustering appeared independent from the geographic origin of the nematodes, and could not discriminate isolates or groups of geographically close isolates. Rather, the formation of different phylogenetic groups within this satDNA family suggests an *a priori* embodying of a set of diverging repeats from a common ancestor satDNA library, which have been differently amplified along the evolutionary pathway of this species. The present work improves knowledge on the evolutionary dynamics of satDNA at the intraspecific level, and provides new information on satDNA sequence variability among natural populations sampled at a local geographic scale.

© 2013 Elsevier Inc. All rights reserved.

### 1. Introduction

Most eukaryotic genomes possess large amounts of non-coding DNA, arranged in highly repeated tandem units, and commonly referred to as satellite DNA (satDNA). These highly repetitive sequences make up as much as half or more of the genomic content and are organized into arrays of tens to thousands of copies primarily clustered in heterochromatin at centromeres and other heterochromatin regions of chromosomes (Charlesworth et al., 1994). These sequences are generally A–T rich and show variability affecting monomer size, nucleotide sequence and copy number. Contrary to a former established idea, the chromosomal regions consisting of satDNA are now known to play important but incompletely understood roles in some intragenomic mechanisms, such as centromeric condensation, sister chromatid pairing,

karyotypic evolution and chromosome arrangement (Plohl et al., 2008; Malik and Henikoff, 2009) and also seem to play some role in gene regulation (Ugarković, 2005; Tomilin, 2008).

SatDNA sequences evolve according to the principles of concerted evolution driven by different mechanisms of nonreciprocal DNA transfer, such as unequal crossing-over between repeating units, gene conversion, rolling circle replication followed by reinsertion, replicative transposition, and slippage replication (Dover, 1982, 2002). Despite rapid changes in copy numbers, satDNAs are inherited in a linear fashion through generations, and their genomic composition is expected to follow species evolutionary history. Sequence rearrangements such as duplications, insertions, deletions, and inversions within monomers may play a dominant role in the evolution of some satellite sequences (Mravinac and Plohl, 2007), and rearranged monomer variants often follow the principles of the “library model”, which is based on the differential amplification/contraction of satDNA subfamilies through species/lineages diversification (Salser et al., 1976; Meštrović et al.,

\* Corresponding author.

E-mail address: [pvieira@uevora.pt](mailto:pvieira@uevora.pt) (P. Vieira).

1998). In this model, related species share a “library” of conserved satDNA sequence variants, some of which could be amplified into a major satellite due to the evolutionary dynamics of satDNAs. Quantitative changes in satDNAs, induced by occasional amplification of monomeric variants from the library, could possibly occur in the course of the speciation process, thus forming a species-specific profile of satDNAs. Mechanisms of sequence homogenization, such as unequal crossing-over and gene conversion, act more efficiently on contiguous monomers within a satellite array than among those located distantly, ultimately leading to the grouping of monomer sequence variants and formation of satellite sub-families (Dover, 1982; Hall et al., 2005).

The pattern of overall variability of particular satDNAs or satellite subfamilies could reveal evolutionary processes affecting these sequences. The rates at which homogenization takes place in different organisms show considerable variation as they are likely dependent on many factors at the molecular and species levels (Plohl et al., 2008; Mravinac and Plohl, 2010). The rapid changes in copy number and nucleotide sequence of satDNAs may result in the evolution of genus and species-specific repeats (Pons and Gillepsie, 2004; Mravinac and Plohl, 2010), as these repetitive units are usually highly homogenized within a species due to concerted evolution. Although satDNA has been occasionally considered as phylogenetically informative of different under-species hierarchical ranks, such as ecotype-specific variants (Hall et al., 2003), populations (Felicello et al., 2011) or phylogeographic clades (Robles et al., 2004), detailed studies on sequence variation at the intraspecific level are still very scarce.

The pinewood nematode (PWN), *Bursaphelenchus xylophilus*, is a naturally endemic species from North America, occurring throughout most of the territories of Canada and United States (Rutherford et al., 1990), and also occasionally found in Mexico (Dwinell, 1993). During the last century, this species has been scattered by man to several non-native regions of the world, such as Asia and Europe, associated with trade and global flow of wood products (Evans et al., 1996; Webster, 2004). The current known worldwide distribution of the PWN in non-native areas is broad, including several regions of Asia (China, Japan, Korea and Taiwan) and some localized areas in Europe (Portugal, Madeira island and Spain) (Pereira et al., 2013).

The species-specific *MspI* satDNA of *B. xylophilus* makes up to 30% of the nematode genome, consisting of monomer units  $\approx 160$  bp-long and slightly A–T rich (Tarès et al., 1993). Although this satDNA family has been recurrently used for species diagnosis (François et al., 2007; Cardoso et al., 2012), the available information regarding the structure and variation of this satDNA family is still scantily studied (Tarès et al., 1993; Castagnone-Sereno et al., 2008). Despite the recent release of the genome sequence of *B. xylophilus* (Kikuchi et al., 2011), no output was provided for such highly repetitive regions.

In order to obtain a more comprehensive insight into the structure, organization and variability of *MspI* satDNA of *B. xylophilus* at the intraspecific level, we compared a set of different isolates originating from the initial point of entry in Portugal (Peninsula of Setúbal) (Mota et al., 1999), and a set of single monomer sequences that we have previously generated (Castagnone-Sereno et al., 2008) from isolates maintained in collection either originating from native areas, or non-native regions where this species has been introduced. In addition, a set of monomers extracted from the genome sequence database was also combined into this analysis. Based on such a large sample of monomers, we examined here in detail the intraspecific variability of the *B. xylophilus* *MspI* satDNA family and investigated the potential mechanisms underlying the different variability levels that occur within the well separated sources of monomer variants identified.

## 2. Material and methods

### 2.1. Nematode sampling from the Peninsula of Setúbal

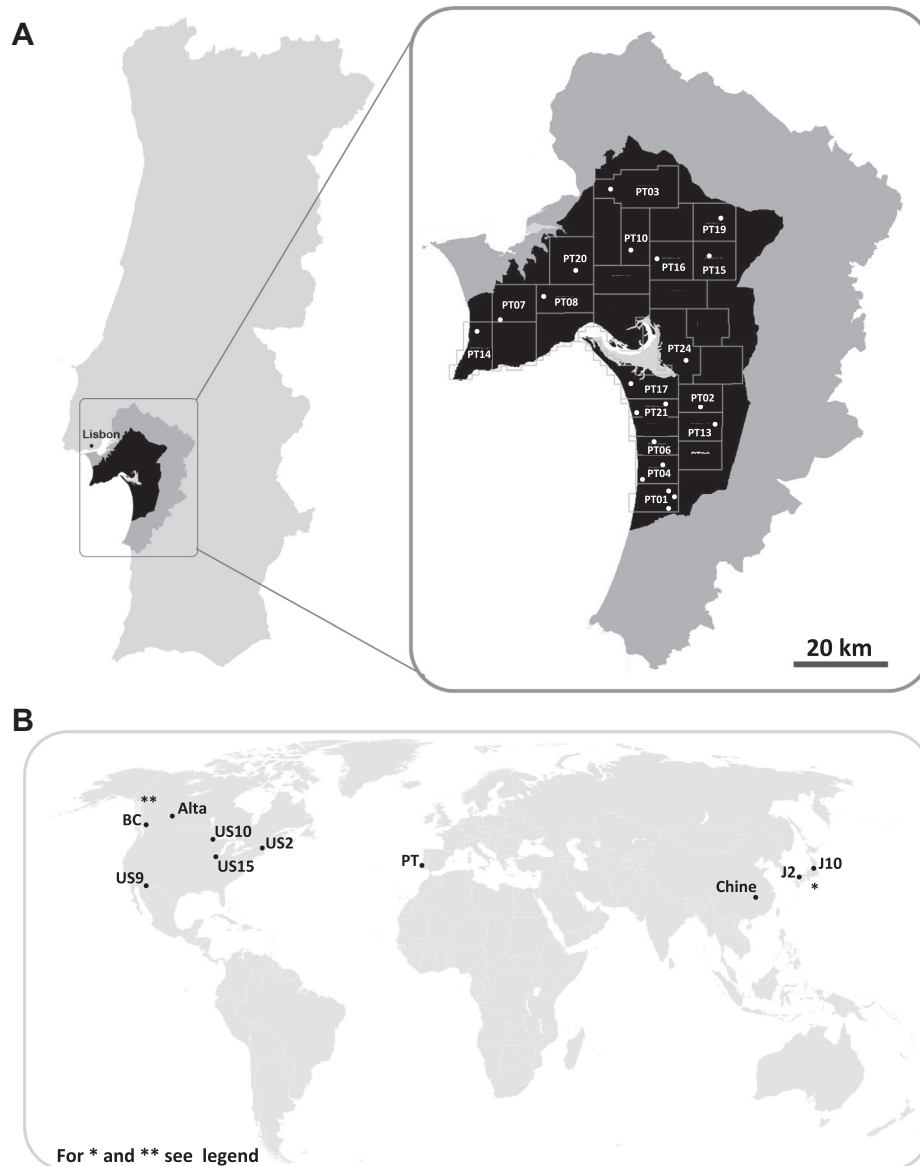
A total of 21 isolates of *B. xylophilus* were obtained from the Peninsula of Setúbal, Portugal. Each isolate was collected from a wood sample (40–80 g each) extracted from individual trees distributed among 17 different blocks, according to the design established during the national PROLUNP program for monitoring the PWN in Portugal (Fig. 1A). From each wood sample, several hundred nematodes (without separation according to sex or developmental stage) were collected, washed several times in distilled water and stored at  $-20^{\circ}\text{C}$  until subsequent analyses.

### 2.2. Isolation of *MspI* satellite DNA monomers

The amplification of satDNA was performed based on a PCR procedure avoiding extraction of the genomic DNA (Castagnone et al., 2005). For each isolate, three nematodes were transferred to a dry thin walled PCR tube, covered with 5  $\mu\text{l}$  lysis buffer (50 mM KCl, 10 mM Tris pH 8.2, 2.5 mM  $\text{MgCl}_2$ , 60  $\text{mg ml}^{-1}$  proteinase K, 0.45% NP40, 0.45% Tween 20, 0.01% gelatin), and overlaid with mineral oil. Tubes were put at  $-80^{\circ}\text{C}$  for 45 min, and immediately transferred to  $60^{\circ}\text{C}$  for 60 min and then  $95^{\circ}\text{C}$  for 15 min in the thermal cycler. The satDNA sequences were then amplified with specific primers J10-1 5'-GGTGTCTAGTATAATATCAGAG-3' and J10-2Rc 5'-GTGAATTAGTGACGACGGAGTG-3' (Castagnone et al., 2005), designed according to the sequence derived from the *MspI* satDNA family previously characterized for *B. xylophilus* (Tarès et al., 1993). PCR was carried in 25  $\mu\text{l}$  reaction mixtures in the same tubes in which were added 2.5 mM dNTP, 100 ng of each of the primers and 0.5 Unit of *Taq* DNA polymerase (Q-Biogene), using a TRIO-Thermoblock thermal cycler (Biometra). After denaturation at  $94^{\circ}\text{C}$  for 5 min, cycling was performed for 25 cycles of 30 s at  $94^{\circ}\text{C}$ , 1 min at  $64^{\circ}\text{C}$  and 1 min at  $72^{\circ}\text{C}$ , with a post-cycling extension at  $72^{\circ}\text{C}$  for 5 min. The resulting fragments were separated on 1.3% agarose gel and stained with ethidium bromide. Bands corresponding to monomers were recovered from the agarose gel, using a gel extraction column (MinElute Gel Extraction Kit, Qiagen Inc.). SatDNA fragments recovered from the gel were ligated into the plasmid vector pGEM-T using protocols provided by the manufacturer (pGEM-T Vector System, Promega). The ligation was used to transform competent *Escherichia coli* DH5 $\alpha$  cells, which were spread on L-agar plates with ampicillin, and grown overnight at  $37^{\circ}\text{C}$ . The positive recombinant clones were identified as white colonies, and 12 random recombinants were amplified by PCR using SP6 and T7 universal primers ( $94^{\circ}\text{C}$  for 2 min, followed by 35 cycles of:  $94^{\circ}\text{C}$  for 30 s,  $55^{\circ}\text{C}$  for 30 s and  $72^{\circ}\text{C}$  for 1 min; followed by a 7 min final extension at  $72^{\circ}\text{C}$ ) and insert size was estimated on agarose gels. The molecular procedures adapted herein were based on Tomalova et al. (2012), which showed a high reproducibility of these molecular protocols. Clones with inserts of approximately 150 bp-length were used to inoculate 3 mL LB, grown overnight at  $37^{\circ}\text{C}$ , followed by plasmid DNA extraction (QIAprep Spin Miniprep Kit, QIAGEN, Inc.). The successful clones were sequenced in one direction by Genome Express (Meylan, France).

### 2.3. Additional monomeric sequences included in the analyses

In addition to the monomers obtained as described above (207 sequences), two additional sets of monomer sequences were included in our analyses: 117 monomer sequences obtained from a collection of different isolates originating from a worldwide geographic distribution (Castagnone-Sereno et al., 2008; Fig. 1B), and



**Fig. 1.** Geographic distribution of the *B. xylophilus* isolates used in this study. (A) Local distribution of isolates collected within the initial entry area at the Peninsula of Setúbal (Portugal). The black area marks the area delimited by the blocks designed during the PLOLUNP program for the PWN survey in 2005. White dots correspond to the location of the pine trees sampled within the respective block. (B) Worldwide distribution of isolates. The black dots represent the approximate origin of the isolates. \* and \*\* = Non-specified origin of the isolates “Japan” and “601” and “667” from Japan and Canada, respectively.

monomeric sequences extracted from the *B. xylophilus* genome sequence database (<http://www.genedb.org/Homepage/Bxylophilus>), which DNA was extracted from a mixed-stage nematode population originally isolated in Iwakari prefecture (Japan; Kikuchi et al., 2011). In the last case, a random collection of 101 repeats was obtained by performing a BLAST search of satDNA repeats on the whole genome database of *B. xylophilus* (Kikuchi et al., 2011), using as query the consensus sequence previously reported for this species (Tarès et al., 1993).

#### 2.4. Monomer sequence comparison and phylogenetic analysis

Basic sequence statistics were estimated using the DnaSP ver. 5.1 software (Librado and Rojas, 2009). Conserved and variable domains were determined by examining nucleotide occurrence values over a sliding window of 10 bp. Variable sequence blocks were considered as significant when nucleotide diversity (Pi) exceeded two times the standard deviation (s.d.) of the average value.

Sequences alignments were carried on using the default parameters of the CLUSTAL X program available on the software Geneious v6 (Drummond et al., 2010). Phylogenetic analyses were constructed using two approaches, i.e., Neighbor-Joining and Maximum-Likelihood with F84 distance parameter with the program PHYML available on Geneious v6. One thousand bootstrap replicates were performed to assess internal support for nodes. The phylogenetic trees were then build using FigTree v1.3.1 software (<http://tree.bio.ed.ac.uk/software/figtree>).

### 3. Results

#### 3.1. Characterization of *MspI* satellite monomers from *B. xylophilus* isolates

Although satDNAs have been characterized in other nematode species, such as plant-parasitic species of the genus *Meloidogyne* (Piotte et al., 1994; Castagnone-Sereno et al., 2000; Meštrović

et al., 2006) or entomopathogenic nematodes of the genus *Heterorhabditis* (Abadon et al., 1998), our BLAST search did not give any significant similarity with other sequences deposited in the GenBank database, neither with particular subdomains shared with other recognized satDNA sequences.

As expected, agarose gel electrophoresis of the DNA fragments amplified using specific primers previously designed from the *MspI* satDNA family of *B. xylophilus* produced a typical and well-defined ladder pattern of ~150-bp monomers, followed by bands of the corresponding dimers (~300 bp), trimers (~450 bp), and so on fragments (Supplementary Fig. S1).

The nucleotide monomer sequences of the *MspI* satDNA family of *B. xylophilus* were determined for 21 isolates distributed along the Peninsula of Setúbal (Fig. 1A and Table 1). For each isolate, a minimum of five to twelve monomers were successfully sequenced, performing a total of 207 satDNA sequences generated for this species. The size of the satellite monomers was established by using the set of primer regions as the arbitrary starting and ending of the repeats. The amplified *MspI* monomer sequences showed a high conserved unit size of 146 bp among 200 clones. The very few exceptions occurred for some monomers units displaying 145 bp (PT1a.10, PT17.9, PT24.7) or 148 bp (PT13.3, PT13.6, PT13.7, PT13.8; Supplementary Fig. S2), respectively. Their defining features, in addition to their very conserved unit length, include an average nucleotide composition of  $T = 31.75\%$ ,  $C = 21.04\%$ ,  $A = 25.24\%$  and  $G = 21.94\%$  (Table 1).

In parallel, we conducted the same type of analyses on the 117 monomers previously obtained using the same PCR-amplified methodology described above (Castagnone-Sereno et al., 2008), from twelve isolates with a worldwide geographic distribution (Fig. 1B), and on 101 monomers randomly obtained from different contigs of the *B. xylophilus* genome database. Basically, the data obtained for these two sets of monomers exhibited the same patterns as for the isolates collected from the Peninsula of Setúbal in terms of sequence length (146 bp) and A+T content (56–58%) (Supplementary Table S1).

### 3.2. Variability distribution across the *MspI* monomers

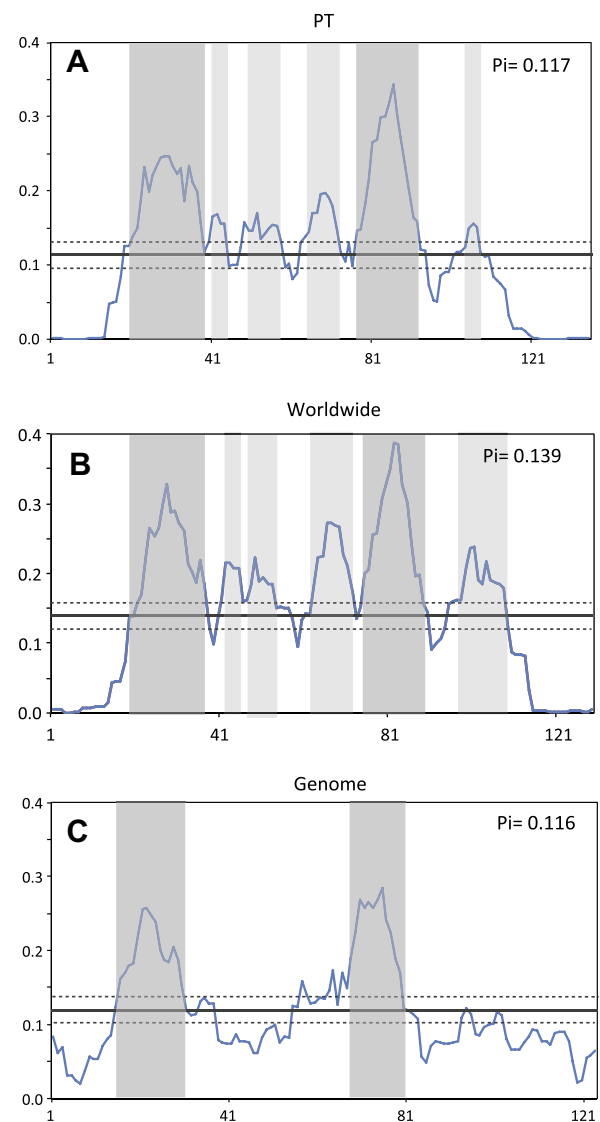
The alignments of the different individual sets of monomers exhibited internal sequence variability (Supplementary

**Table 1**  
Peninsula of Setúbal isolates list containing the number of complete monomers obtained, including their monomer size and A–T rich content.

Block	Isolate code	N	Size of monomers in bp	A–T (%)
<i>Peninsula of Setúbal</i>				
PT1	PT1	9	146	57
	PT1a	11	146 (10) and 145 (1)	58
	PT1b	12	146	57
PT2	PT2	5	146	56
PT3	PT3	7	146	57
PT4	PT4	11	146	58
	PT4a	12	146	57
PT6	PT6	10	146	58
PT7	PT7	11	146	57
PT8	PT8	12	146	56
PT10	PT10	6	146	56
PT13	PT13	10	146 (6) and 148 (4)	57
PT14	PT14	10	146	58
PT15	PT15	8	146	56
PT16	PT16	10	146	58
PT17	PT17	12	146 (11) and 145 (1)	57
PT19	PT19	7	146	57
PT20	PT20	11	146	56
PT21	PT21	11	146	57
PT24	PT21a	11	146	57
	PT24	11	146 (10) and 145 (1)	57
Total		207		

Figs. S3–S5). Pairwise comparisons of the monomers revealed sequence similarity of 88.1% for the Peninsula of Setúbal isolates, 85.3% and 89.1% for the genome and worldwide isolates, respectively.

The distribution of nucleotide diversity was measured by sliding window analysis among the sequences from the three data sets. This analysis revealed an alternation of highly to moderately variable and conserved domains within the individual monomeric sequences of the Peninsula of Setúbal isolates (Fig. 2A). The two most highly variable domains within this satDNA accumulated in the sequence inner-central regions between the base positions 22–40 and 78–97 bp (Fig. 2A, dark gray). The sliding window obtained for sequences from the isolates with a worldwide geographic distribution exhibited the same profile with alternate conserved and variable domains overlapping the same regions as



**Fig. 2.** Nucleotide variation across the *MspI* satellite DNA monomers from *B. xylophilus*, represented by sliding window analyses of the alignment of the different sets of sequences. (A) Monomers cloned from isolates collected at the Peninsula of Setúbal. (B) Monomers cloned from the isolates collected worldwide. (C) Monomers extracted from the genome sequence database. The graphs show the value of nucleotide diversity (Pi) in a sliding window size of 10 with step size 2. Each value is depicted at its mid-point. The solid line indicate the average Pi, and the dashed lines represent two-fold standard deviation (SD) from the average, respectively.



for the isolates collected from the Peninsula of Setúbal (Fig. 2B). Overall, the nucleotide diversity obtained for the worldwide monomers was relatively higher in comparison with that of the Peninsula of Setúbal ( $P_i = 0.139$  versus  $P_i = 0.117$ ).

To study the potential bias in the estimation of variability due to primer specificity in both PCR-amplified monomer sets, we assessed monitoring the nucleotide conservation across these sequences by alignment of the 101 satDNA monomers obtained from the genome database (Supplementary Fig. S5), and applying the same criteria as described above. The sliding window of the genome monomer dataset displayed an overall reduced variability when compared to the counterparts obtained from the PCR-amplified repeats of the isolates with a worldwide distribution (Fig. 2C), but showing similar nucleotide diversity ( $P_i = 0.116$ ) to the Peninsula of Setúbal isolates. The limited variability found for set of monomers extracted from the genome, seems to be related to the low variability presented by the single population used for the genome sequencing (isolated in the Ibakari prefecture, in Japan). It is worth noting the highly conserved overlapping short segments within the three datasets, possibly reflecting some molecular constraint acting upon these specific regions.

We next used our entire data sets of satDNA monomers to explore the potential evolutionary relationships of the *MspI* monomers within this species. Initially, the three independent sets of monomers were independently analyzed making a total of 425 monomers (207 from the Peninsula of Setúbal and 117 from worldwide isolates obtained from PCR amplification, and 101 obtained directly by the genome database). The alignments of the three original sets of *MspI* monomers were initially used independently to infer phylogenetic relationships using neighbor-joining and maximum likelihood algorithms. The resulting phylogenetic trees showed very similar topologies in relation to the internal nodes (Fig. 3A–C), supported by high bootstrap values, which suggested that this family of satDNA can be divided into different clusters of monomer variants. In each set, the majority of the total number of randomly selected monomers was allocated to the same large cluster D (gathering approximately 50% of the monomers distributed among the worldwide isolates, to more than 79% in both monomers obtained from the Peninsula of Setúbal isolates or the genome), while the remaining monomers were distributed within the other clusters, and represented with a lower number of sequences (clusters A, B and C). Only a few number of monomers were positioned out of the four main clusters identified (black branches on the phylogenetic trees). Within each individual cluster we could still observe some levels of sequence heterogeneity (Fig. 3D), but in all cases the highest intra-cluster variability was detected among the monomer sequences that composed the major cluster D. In each single cluster, we could observe an intermingling and random distribution of monomers derived from different isolates, and generally no correlation could be drawn between a specific cluster and a group or groups of geographic isolates.

We further computed global NJ and ML phylogenetic trees based on the alignment of the three types of sequences, to examine the phylogenetic relationships among all the *MspI* monomers (Fig. 4). Overall, the NJ and ML trees shared the same general topology, and monomeric sequences from either the genome, Peninsula of Setúbal or worldwide origin were not clustered in specific clades. Notably, the monomers belonging to clusters A and B were well separated in each set-specific origin, at the root of the dendrogram, without clearing intermixing (the only exceptions were for monomer PT8.9 that grouped together with cluster A and monomer PT21a.2 that grouped together with cluster B of the worldwide set). The remaining monomers, belonging to both clusters C and D,

fall deeply intermingled within the same group, independently of their geographic origin or set.

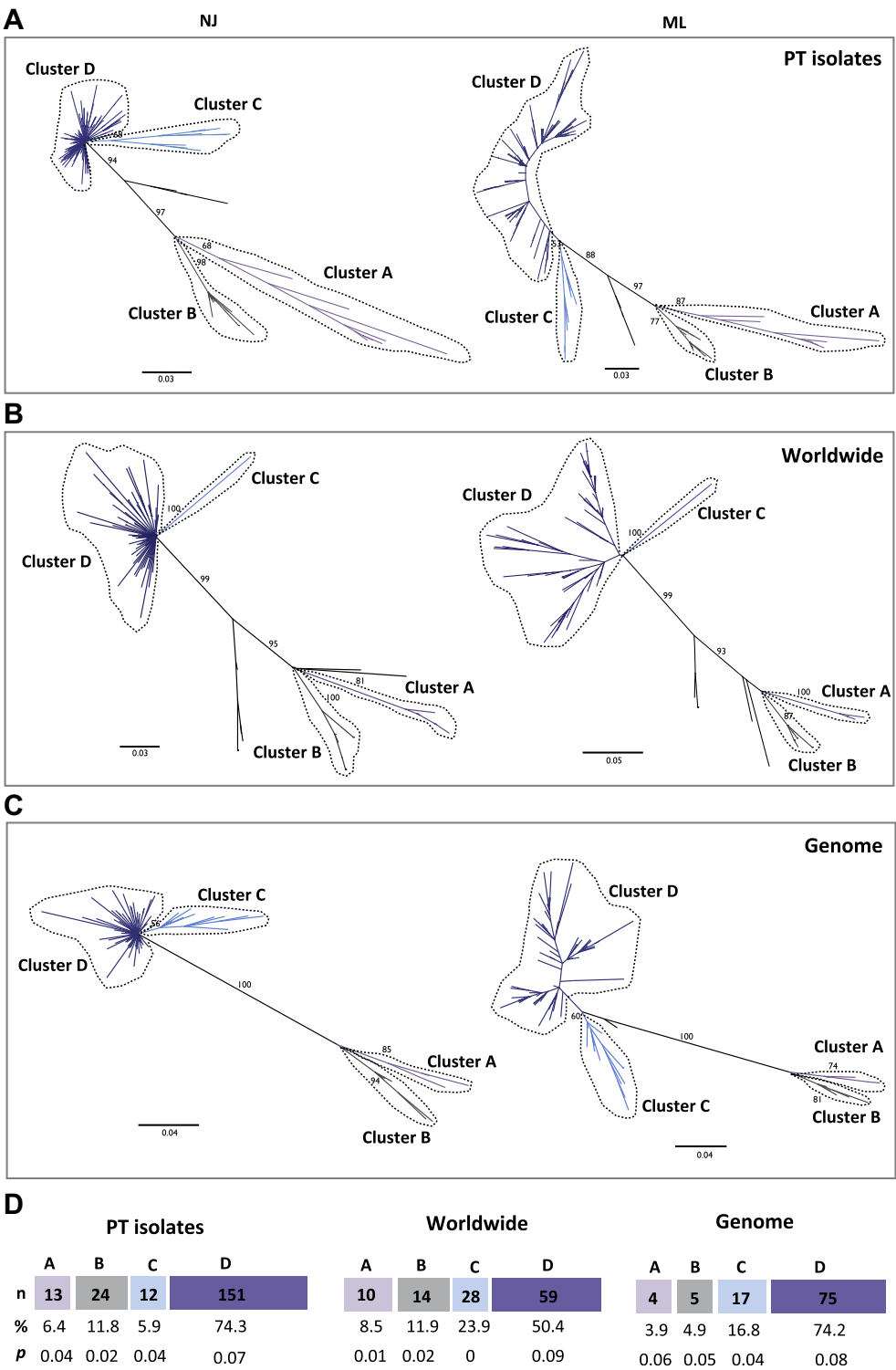
### 3.3. Comparative analysis of *B. xylophilus* isolates from a local origin and worldwide distribution

To analyze the extent of sequence variation among the different geographic isolates from the Peninsula of Setúbal, we first calculated the genetic distances within and between the repeats of each geographic isolates. Genetic distances within monomers of the same isolate ranged from 0.165 to 0.047 ( $p$ -mean distances), showing a similar magnitude of sequence diversity obtained against different geographic isolates collected within the Peninsula of Setúbal (0.172–0.051) (Table 2). For a more detailed analysis on the intra-variability of monomers from isolates distributed in adjacent trees within the same area, two to three isolates were collected within the same block for three blocks (i.e., PT1, PT4 and PT21). The mean distance obtained for isolates within the same block showed the same range of variation exhibited for the total number of isolates (Table 2). The  $p$ -mean distances were then calculated within isolates distributed worldwide (Supplementary Table S2), showing a similar trend as the isolates of the Peninsula of Setúbal (0.162–0.067). Interestingly, when  $p$ -mean distances are taken into account among all the isolates (Supplementary Table S2), a high level of sequence variability emerges between isolates (up to 0.25), with the highest differences found between isolates collected from North America (e.g., US9, US10) and the Peninsula of Setúbal isolates.

## 4. Discussion

The sequence classification and structural arrangement of satDNA have been mainly studied in nematode species belonging to root-knot nematodes (family Meloidogynidae) or entomopathogenic nematodes (family Heterorhabditidae) (Piotte et al., 1994; Abadon et al., 1998; Castagnone-Sereno et al., 2000; Meštrović et al., 2006). In order to achieve a global portrait of the *MspI* satDNA family in *B. xylophilus* (family Aphelenchoididae), disregarding their origin (geographical distribution or biological conditions, i.e. field versus lab collection), we conducted combined analyses on three different sets of both PCR-amplified and genome-retrieved monomers. No specific clustering according to the monomer origin was detected in the sequence analyses performed in this study, which indicates that the experimental origin of the monomers did not generate any apparent bias in term of sequence variability, thus reinforcing earlier results (Castagnone-Sereno et al., 2008).

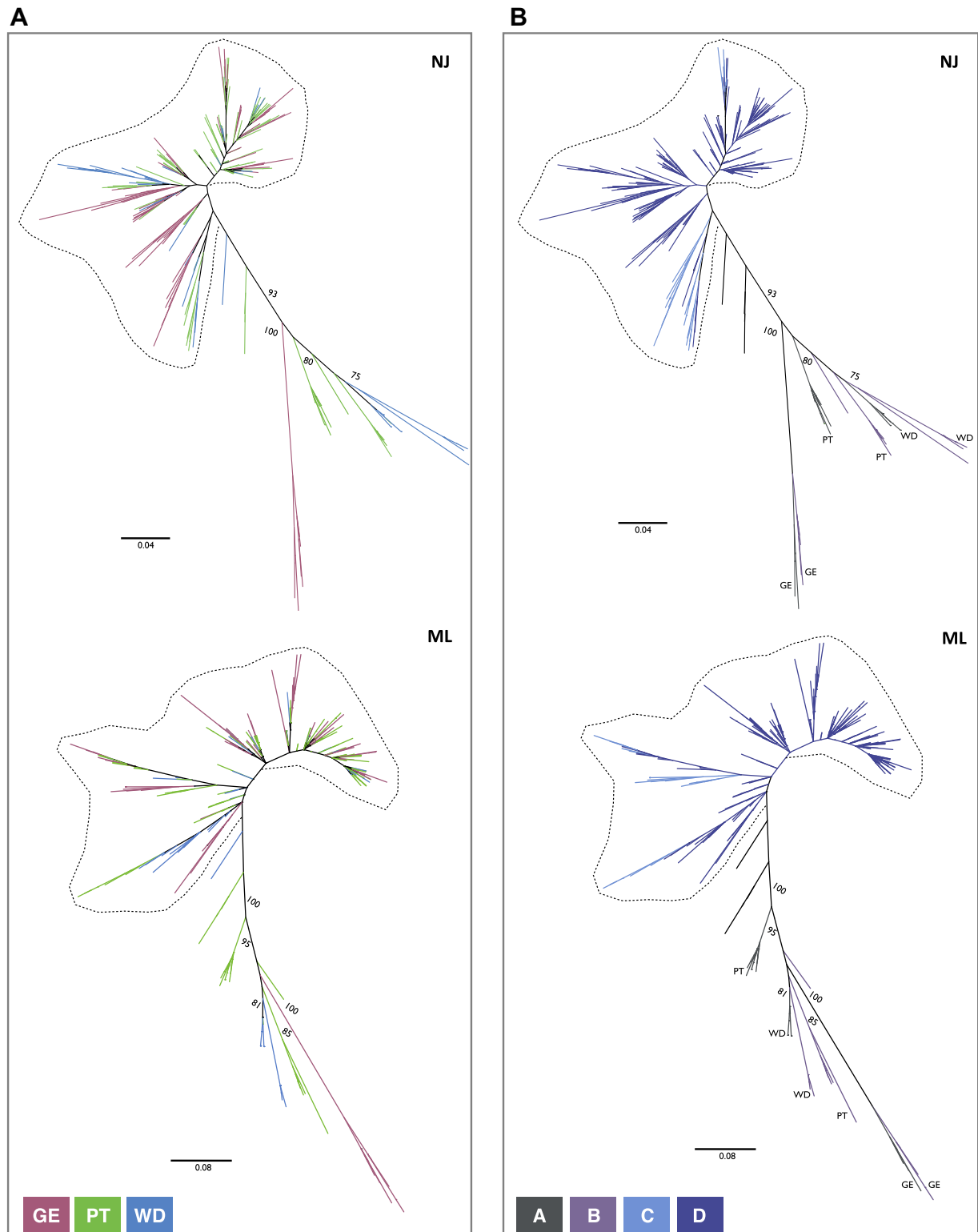
Basic sequence analyses revealed a strict conservation of the sequence length in each set of monomers analyzed. By evaluating a higher number of monomer units, the average sequence variability showed a more than twofold increase ( $\approx 10\%$ ), when compared to the 3.9% average variability exhibited among the thirteen monomers originally cloned for this species (Tarès et al., 1993). Intra-monomer analyses of *MspI* satDNA disclose several domains showing high variability, while other domains presented a strong conservation among the full set of monomers. Nevertheless, the divergence within the satDNA monomers seems to occur in a gradual manner mostly due to the accumulation of nucleotide substitutions, while deletions and insertions events are less common for this family. The preservation of both conserved and variable domains across satellite sequences are characteristic of this kind of repetitive DNA. Such distribution of variable domains has been found in other nematode species (e.g., Meštrović et al., 2006), and is well characterized in satDNAs from model organisms such as the centromeric satellite of *Arabidopsis thaliana* (Hall et al.,



**Fig. 3.** Neighbor-joining (NJ) and Maximum Likelihood (ML) phylogenetic trees of *MspI* satDNA monomers from *B. xylophilus*. Trees built with monomeric repeats from (A) isolates of the Peninsula of Setúbal (207 monomers); (B) isolates with a worldwide distribution (117 monomers); (C) the genome sequence database (101 monomers). One thousand bootstrap replications were used to generate the phylogenetic trees. The *MspI* satDNA monomers are grouped into four main clusters defined by different colors. (D) Distribution of the number of monomers within the main clusters identified. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

2003) and human  $\alpha$ -satDNA (Choo et al., 1991). The variability patterns observed here for *B. xylophilus* monomeric sequences follow the hypothesis that some repeat regions are under selective pressure to maintain a particular DNA sequence, whereas other regions evolve without any constraint. The growing knowledge on the

effective functional roles of the highly conserved domains underpins the idea that this non-random pattern of variability could enhance binding with specific centromere proteins (Ugarković, 2008), or even the control of gene expression (Ugarković, 2005; Tomilin, 2008).



**Fig. 4.** Phylogenetic trees of combined satellite monomers obtained from the three sets of monomers. Neighbor-joining (NJ) and maximum likelihood (ML) trees containing a total of 425 sequences were generated for *B. xylophilus*. One thousand bootstrap replications were used to generate each phylogenetic tree. Only significant bootstrap values are shown close to the main nodes of the dendrogram. In (A) monomers were represented by colors related to their set of origin (GE: genome; PT: Peninsula of Setúbal; WD: worldwide), while in (B) monomers are represented by colors representing the main clusters obtained in Fig. 3 (Clusters A–D, respectively). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

The phylogenetic inference based on the different sets of *MspI* satellite DNA family for this species shows a broad polymorphism of the individual monomers, when compared for example to other

species of nematodes (e.g. Meštrović et al., 2005, 2006). In both analyses obtained either from the PCR-amplified monomers or from monomers retrieved from the genome database, the

Table 2

Pairwise genetic distances (*p*-distances) among different *B. xylophilus* isolates collected from the Peninsula of Setúbal (Maximum Composite Likelihood).

	PT1	PT1a	PT1b	PT2	PT3	PT4	PT4a	PT6	PT7	PT8	PT10	PT13	PT14	PT15	PT16	PT17	PT19	PT20	PT21	PT21a	PT24
PT1	<b>0.111</b>																				
PT1a	0.103	<b>0.091</b>																			
PT1b	0.089	0.076	<b>0.064</b>																		
PT2	0.116	0.112	0.100	<b>0.143</b>																	
PT3	0.131	0.125	0.117	0.141	<b>0.148</b>																
PT4	0.089	0.069	0.058	0.101	0.118	<b>0.047</b>															
PT4a	0.116	0.107	0.100	0.122	0.134	0.096	<b>0.127</b>														
PT6	0.104	0.088	0.075	0.114	0.122	0.071	0.110	<b>0.094</b>													
PT7	0.104	0.083	0.075	0.113	0.127	0.061	0.106	0.086	<b>0.081</b>												
PT8	0.140	0.144	0.137	0.147	0.152	0.142	0.146	0.145	0.149	<b>0.165</b>											
PT10	0.123	0.121	0.113	0.126	0.146	0.110	0.126	0.124	0.120	0.148	<b>0.148</b>										
PT13	0.095	0.084	0.070	0.108	0.121	0.068	0.106	0.085	0.085	0.140	0.117	<b>0.081</b>									
PT14	0.086	0.073	0.057	0.101	0.113	0.051	0.098	0.072	0.069	0.135	0.112	0.069	<b>0.053</b>								
PT15	0.126	0.126	0.117	0.129	0.148	0.119	0.129	0.129	0.128	0.148	0.131	0.122	0.118	<b>0.146</b>							
PT16	0.093	0.080	0.065	0.107	0.122	0.059	0.103	0.079	0.076	0.144	0.117	0.075	0.060	0.124	<b>0.070</b>						
PT17	0.118	0.115	0.104	0.127	0.140	0.103	0.126	0.116	0.116	0.148	0.132	0.112	0.100	0.136	0.109	<b>0.137</b>					
PT19	0.125	0.127	0.117	0.135	0.142	0.119	0.133	0.127	0.129	0.149	0.137	0.118	0.117	0.139	0.123	0.137	<b>0.156</b>				
PT20	0.122	0.125	0.113	0.127	0.145	0.117	0.130	0.125	0.128	0.143	0.129	0.119	0.113	0.129	0.121	0.131	0.135	<b>0.135</b>			
PT21	0.137	0.131	0.128	0.139	0.144	0.126	0.134	0.134	0.132	0.172	0.139	0.131	0.126	0.140	0.133	0.143	0.145	0.140	<b>0.153</b>		
PT21a	0.104	0.095	0.083	0.114	0.125	0.081	0.111	0.096	0.094	0.138	0.119	0.092	0.080	0.124	0.088	0.114	0.124	0.119	0.130	<b>0.101</b>	
PT24	0.098	0.083	0.071	0.108	0.122	0.066	0.104	0.084	0.082	0.141	0.117	0.080	0.068	0.121	0.074	0.110	0.124	0.119	0.130	0.090	<b>0.082</b>

The average within-isolate genetic distance is shown in bold values.

individual monomers intermingled forming four main clusters whatever their origin. In addition, the general allocation of monomers from a single isolate among different clusters provided evidence that most of the different geographic isolates embodies the whole population of repeats that compose the *MspI* family. The patterns of variability within satDNA sequences have classically been explained as a result of the combination of random mutation and horizontally intragenomic spreading of new mutants (Dover, 2002). Various mechanisms, including unequal recombination, gene conversion events, and replication slippage, have been proposed to explain how individual satellite blocks can evolve rapidly (Plohl et al., 2008). Taking into account the formation of these different phylogenetic groups, the topology of each tree may be interpreted according to the hypothesis of *B. xylophilus* embodying *a priori* a set of diverging repeats from a common ancestor satDNA library, which have been differently amplified along the evolutionary pathway of this species. The dispersion of these different sequence variants among the nematode genome reinforces the idea that the “library model” may be also extended to single subsets of diverged monomers of the same satellite family within the same species (Mravinac et al., 2002; Bruvo et al., 2003; Cesari et al., 2003; Pons et al., 2004; Lee et al., 2011). The occurrence of different satDNA repeats in the form of a library coexisting in the same species might be considered as independent evolutionary units (Ugarković and Plohl, 2002). Hence, the dynamics observed for both high and low-copy satellite clusters suggests a continuous nucleotide change along their evolutionary clock. However, the higher number of monomers falling in cluster D could correspond to the major satellite group in this species, where the pronounced intermingle of these sequences could suggest a more homogenizing process of this variant across the genome of *B. xylophilus*. While clusters with a lower number of monomers (such as clusters A and B) can be predicted to evolve or homogenise at different rates, resulting in less conserved monomers in comparison to the most abundant variants. Over time, the abundance of particular variants could stochastically change through both expansion and shrinkage, resulting in replacement of the most abundant variant with a different variant (Ugarković and Plohl, 2002).

Based on the structural and evolutionary studies of diverse satDNAs, it is suggested that satellite sequences within a library are highly adapted and relatively conserved in the sense that they can tolerate limited sequence change (Dover, 2002; Ugarković and

Plohl, 2002). In addition to selection, the rate of sequence change depends on mutation rate, rates of recombination processes that spread mutations horizontally through the repetitive family, as well as on the rate of fixation within a population (Ugarković and Plohl, 2002). The rapid and complex molecular coevolution of satellite repeats could conceivably create isolate-specific profiles for well separated distant regions. Here, our goal was to study the sequence variability of this satDNA family within different geographic scales of *B. xylophilus*, and their potential use for discrimination of their area of distribution. Our analyses suggest that no specific nucleotide substitutions or indels were identified in *MspI* satDNA monomers, which could discriminate single populations or groups of geographically close populations at the scale of the Peninsula of Setúbal. Such a lack of intraspecific molecular structure in satDNA sequences has been observed in several *Drosophila* species (Kuhn and Sene, 2004; Franco et al., 2006), and was suggested to arise due to the combination of both past and current gene flow events among populations and independent homogenization mechanisms favoring monomers shared among the populations (Kuhn and Sene, 2004; Franco et al., 2010). Considering the very recent invasion of Portugal by *B. xylophilus* (Mota et al., 1999), and the limited dispersion capacity of the nematode, it is very unlikely that a significant amount of gene flow has occurred within the investigated area. However, the natural dispersion of the PWN from tree to tree is totally dependent on its insect vector, *Monochamus* spp., and linked to the insect flight (Linit, 1988). As we have been working with isolates from a relative limited area, this suggests that intermingle isolates could result from the insect dispersion way among this area, resulting in uninterrupted crossing among the established genetic background within the Peninsula of Setúbal. Nevertheless, the exchange of infected material goods within these areas by human activities cannot be excluded, as they represent one of the most effective ways of nematode dispersal into new areas (Robinet et al., 2009).

Few experimental data have been reported on the evolutionary dynamics of satDNA at the intraspecific level. From that point of view, the present study constitutes an original approach, and provides new information on satDNA sequence variability among natural populations of an invasive species that has been sampled in a newly infested area. However, the lack of correlation observed herein between satDNA nucleotide diversity and geographic distribution of the nematode precludes the use of this kind of repetitive



sequence as a relevant marker for population genetics studies in *B. xylophilus*, which are of crucial importance to decipher the invasion routes of this pest and anticipate further expansion of the PWD in Europe. As an alternative approach, investigation on the geographic diversity of *B. xylophilus*, including a higher number of isolates distributed worldwide among the native and invaded regions, is now in progress using more powerful molecular markers, i.e., microsatellites (Mallez et al., 2013). Definitely, population genetics analyses should provide a better resolution on the genetic variability within this species. Hence, a better and more accurate idea of the local and worldwide genetic distribution of this important pathogenic species will be provided in near future.

## Acknowledgments

This work was partially supported by the COST Action 872 NEMAGENICS ('Exploiting genomics to understand plant-nematode interactions') and the EC 7th Framework project REPHRAME (KBBE.2010.1.4-09, 'Analysis of the potential of the pine wood nematode (*Bursaphelenchus xylophilus*) to spread, survive and cause pine wilt in European coniferous forests in support of EU plant health policy').

## Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.ympev.2013.09.017>.

## References

- Abadon, M., Grenier, M., Laumond, C., Abad, P., 1998. A species-specific satellite DNA from the entomopathogenic nematode *Heterorhabditis indicus*. *Genome* 41, 148–153.
- Bruvo, B., Pons, J., Ugarković, D., Juan, C., Petitpierre, E., Pohl, M., 2003. Evolution of low-copy number and major satellite DNA sequences coexisting in two *Pimelia* species-groups (Coleoptera). *Gene* 312, 85–94.
- Cardoso, J.M.S., Fonseca, L., Abrantes, I., 2012. Direct molecular detection of the pinewood nematode (*Bursaphelenchus xylophilus*). *Eur. J. Plant. Pathol.* 133, 419–425.
- Castagnone, C., Abad, P., Castagnone-Sereno, P., 2005. Satellite DNA-based species-specific identification of single individuals of the pinewood nematode *Bursaphelenchus xylophilus* (Nematoda: Aphelenchoididae). *Eur. J. Plant. Pathol.* 112, 191–193.
- Castagnone-Sereno, P., Leroy, F., Abad, P., 2000. Cloning and characterization of an extremely conserved satellite DNA family from the root-knot nematode *Meloidogyne arenaria*. *Genome* 43, 346–353.
- Castagnone-Sereno, P., Castagnone, C., François, C., Abad, P., 2008. Satellite DNA as a Versatile Genetic Marker for *Bursaphelenchus xylophilus*. In: Mota M. Vieira (Ed.), *Pine Wilt Disease: a worldwide threat to forest ecosystems*. Springer Science + Business Media BV, p. 187–195.
- Cesari, M., Luchetti, A., Passamonti, M., Scali, V., Mantovani, B., 2003. Polymerase chain reaction amplification of the *Bag320* satellite family reveals the ancestral library and past gene conversion events in *Bacillus rossius* (Insecta Phasmatodea). *Gene* 312, 289–295.
- Charlesworth, B., Sniegowski, P., Stephan, W., 1994. The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature* 371, 215–220.
- Choo, K.H., Vissel, B., Nagy, A., Earle, E., Kalitsis, P., 1991. A survey of the genomic distribution of alpha-satellite DNA on all human chromosomes. *Nucleic Acids Res.* 19, 1179–1182.
- Dover, G., 1982. Molecular drive: a cohesive mode of species evolution. *Nature* 299, 111–117.
- Dover, G., 2002. Molecular drive. *Trends Genet.* 18, 587–589.
- Drummond, A., Ashton, B., Cheung, M., Heled, J., Kearse, M., Markowitz, S., Mir, R., Stones-Havas, S., Sturrock, S., Thierer, T., Wilson, A., 2010. Geneious v6. Geneious website. <<http://www.geneious.com/>>.
- Dwinell, L.D., 1993. First report of pinewood nematode (*Bursaphelenchus xylophilus*) in Mexico. *Plant Dis.* 77, 846.
- Evans, H.F., McNamara, D.G., Braasch, H., Chadouef, J., Magnusson, C., 1996. Pest risk analysis (PRA) for the territories of the European Union (as PRA area) on *Bursaphelenchus xylophilus* and its vectors in the genus *Monochamus*. *EPPO Bulletin* 26, 199–249.
- Feliciello, I., Chinali, G., Ugarković, D., 2011. Structure and population dynamics of the major satellite DNA in the red flour beetle *Tribolium castaneum*. *Genetica* 139, 999–1008.
- Franco, F.F., Kuhn, G.C.S., Sene, F.M., Manfrin, M.H., 2006. Conservation of pBuM-2 satellite DNA sequences among geographically isolated *Drosophila gouveai* populations from Brazil. *Genetica* 128, 287–295.
- Franco, F.F., Sene, F.M., Manfrin, M.H., 2010. Low satellite DNA variability in natural populations of *Drosophila antonietae* involved in different evolutionary events. *J. Hered.* 101, 650–656.
- François, C., Castagnone, C., Boonham, N., Tomlinson, J., Lawson, R., Hockland, S., Quill, J., Vieira, P., Mota, M., Castagnone-Sereno, P., 2007. Satellite DNA as a target for TaqMan real-time PCR detection of the pinewood nematode, *Bursaphelenchus xylophilus*. *Mol. Plant. Pathol.* 8, 803–809.
- Hall, S.E., Kettler, G., Preuss, D., 2003. Centromere satellites from *Arabidopsis* populations: maintenance of conserved and variable domains. *Genome Res.* 13, 195–205.
- Hall, S.E., Luo, S., Hall, A., Preuss, D., 2005. Differential rates of local and global homogenization in centromere satellites from *Arabidopsis* relatives. *Genetics* 170, 1913–1927.
- Kikuchi, T., Cotton, J.A., Dalzell, J.J., Hasegawa, K., Kanzaki, N., McVeigh, P., Takanashi, T., Tsai, I.S., Assefa, S.A., Cock, P.J.A., Dan Otto, T., Hunt, M., Reid, A.J., Sanchez-Flores, A., Tsuchihara, K., Yokoi, T., Larsson, M.C., Miwa, J., Maule, A.G., Sashisi, N., Jones, J.T., Berriman, M., 2011. Genomic insights into the origin of parasitism in the emerging plant pathogen *Bursaphelenchus xylophilus*. *PLoS Pathogens* 7, e1002219.
- Kuhn, G.C.S., Sene, F.M., 2004. Characterisation and interpopulation variability of a complex *HpaI* satellite DNA of *Drosophila seriema* (repleta group). *Genetica* 121, 241–249.
- Lee, H.-R., Hayden, K.E., Willard, F., 2011. Organization and molecular evolution of a CENP-A-associated satellite DNA families in a basal primate genome. *Genome Biol. Evol.* 3, 1136–1149.
- Librado, P., Rojas, J., 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25, 1451–1452.
- Limit, M., 1988. Nematode-vector relationships in the pine wilt disease system. *J. Nematol.* 20, 227–235.
- Malik, H.S., Henikoff, S., 2009. Major evolutionary transitions in centromere complexity. *Cell* 138, 1067–1082.
- Mallez, S., Castagnone, C., Espada, M., Vieira, P., Eisenback, J.D., Mota, M., Guillemaud, T., Castagnone-Sereno, P., 2013. First insights into the genetic diversity of the pinewood nematode in its native area using new polymorphic microsatellite loci. *PLoS One* 8, e59165.
- Meštrović, N., Pohl, M., Mravinac, B., Ugarković, D., 1998. Evolution of satellite DNA from the genus *Palorus* – experimental evidence for the “library” hypothesis. *Mol. Biol. Evol.* 15, 1062–1068.
- Meštrović, N., Randig, O., Abad, P., Pohl, M., Castagnone-Sereno, P., 2005. Conserved and variable domains in satellite DNAs of mitotic parthenogenetic root-knot nematode species. *Gene* 362, 44–50.
- Meštrović, N., Castagnone-Sereno, P., Pohl, M., 2006. High conservation of the differentially amplified MPA2 satellite DNA family in parthenogenetic root-knot nematodes. *Gene* 376, 260–267.
- Mota, M., Braasch, H., Bravo, M.A., Penas, A.C., Burgermeister, W., Metge, K., Sousa, E., 1999. First report of *Bursaphelenchus xylophilus* in Portugal and in Europe. *Nematology* 1, 727–734.
- Mravinac, B., Pohl, M., 2007. Satellite DNA junctions identify the potential origin of new repetitive elements in the beetle *Tribolium madens*. *Gene* 394, 45–52.
- Mravinac, B., Pohl, M., 2010. Parallelism in evolution of highly repetitive DNAs in sibling species. *Mol. Biol. Evol.* 27, 1857–1867.
- Mravinac, B., Pohl, M., Meštrović, N., Ugarković, D., 2002. Sequence of PRAT satellite DNA “frozen” in some coleopteran species. *J. Mol. Evol.* 54, 774–783.
- Pereira, F., Moreira, C., Fonseca, L., van Asch, B., Mota, M., Abrantes, I., Amorim, A., 2013. New insights into the phylogeny and worldwide dispersion of two closely related nematode species, *Bursaphelenchus xylophilus* and *Bursaphelenchus mucronatus*. *PLoS One* 8, e56288.
- Piotte, C., Castagnone-Sereno, P., Bongiovanni, M., Dalmasso, A., Abad, P., 1994. Cloning and characterization of two satellite DNAs in the low-C-value genome of the nematode *Meloidogyne* spp.. *Gene* 138, 175–180.
- Pohl, M., Luchetti, A., Meštrović, N., Mantovani, B., 2008. Satellite DNAs between selfishness and functionality: structure, genomics and evolution patterns of tandem repeats in centromeric (hetero)chromatin. *Gene* 409, 72–82.
- Pons, J., Gillepsie, R.G., 2004. Evolution of satellite DNAs in a radiation of endemic Hawaiian spiders: Does concerted evolution of highly repetitive sequences reflect evolutionary history? *J. Mol. Evol.* 59, 632–641.
- Pons, J., Bruvo, B., Petitpierre, E., Pohl, M., Ugarković, D., Juan, C., 2004. Complex structural features of satellite DNA sequences in the genus *Pimelia* (Coleoptera: Tenebrionidae): random differential amplification from a common “satellite DNA library”. *Heredity* 92, 418–427.
- Robinet, C., Roques, A., Pan, H., Fang, G., Ye, J., Zhang, Y., Sun, J., 2009. Role of human-mediated dispersal in the spread of the pinewood nematode in China. *PLoS One* 4, e4646.
- Robles, F., de la Herrán, R., Ludwig, A., Ruiz Rejón, C., Ruiz Rejón, M., Garrido-Ramos, M.A., 2004. Evolution of ancient satellite DNAs in sturgeon genomes. *Gene* 18, 133–142.
- Rutherford, T.A., Mamiya, Y., Webster, J.M., 1990. Nematode-induced pine wilt disease: factors influencing its occurrence and distribution. *Forest Sci.* 36, 145–155.
- Salser, W., Owen, S., Browne, D., Eladli, F., Fedoroff, N., Fry, K., Heindell, H., Paddock, G., Poon, R., Wallace, B., Whitcome, P., 1976. Investigation of the organization of mammalian chromosomes at DNA sequence level. *Fed. Proc.* 35, 23–35.

- Tarès, S., Lemontey, J.M., Guiran, G., Abad, P., 1993. Cloning and characterization of a highly conserved satellite DNA sequence specific for the phytoparasitic nematode *Bursaphelenchus xylophilus*. *Gene* 129, 269–273.
- Tomalova, I., Iachia, C., Mulet, K., Castagnone-Sereno, P., 2012. The *map-1* gene family in root-knot nematodes, *Meloidogyne* spp.: a set of taxonomically restricted genes specific to clonal species. *Plos One* 7, e38656.
- Tomilin, N.V., 2008. Regulation of mammalian gene expression by retroelements and non-coding tandem repeats. *BioEssays* 30, 338–348.
- Ugarković, D., 2005. Functional elements residing with satellite DNAs. *EMBO Rep.* 6, 1035–1039.
- Ugarković, D., 2008. Satellite DNA libraries and centromere evolution. *Open Evol. J.* 2, 1–6.
- Ugarković, D., Plohl, M., 2002. Variation in satellite DNA profiles – cause and effects. *EMBO J.* 21, 5955–5959.
- Webster, J., 2004. The pine wood nematode: implications of factors past and present for pine wilt disease. In: Mota M, Vieira (Ed.), The pinewood nematode, *Bursaphelenchus xylophilus*. Nematology Monographs and Perspectives, vol. 1. Brill Academic Publishers. p. 55–64.