Title	Complete Structure of Nuclear rDNA of the Obligate Plant Parasite Plasmodiophora brassicae : Intraspecific Polymorphisms in the Exon and Group I Intron of the Large Subunit rDNA
Author(s)	Niwa, Rieko; Kawahara, Ai; Murakami, Hiroharu; Tanaka, Shuhei; Ezawa, Tatsuhiro
Citation	Protist, 162(3), 423-434 https://doi.org/10.1016/j.protis.2011.02.005
Issue Date	2011-07
Doc URL	http://hdl.handle.net/2115/46765
Type	article (author version)
File Information	Pro162-3_423-434.pdf

[Instructions for use](#)

Title: Complete structure of nuclear rDNA of the obligate plant parasite
Plasmodiophora brassicae: intraspecific polymorphisms in the exon and group I
intron of large subunit rDNA

Names of authors: Rieko Niwa^{1, a, b}, Ai Kawahara¹, Hiroharu Murakami², Shuhei
Tanaka³ and Tatsuhiko Ezawa¹

Affiliations:

¹Graduate School of Agriculture, Hokkaido University, Sapporo, Hokkaido
060-8589, Japan

²National Agricultural Research Center for Western Region (WeNARC), National
Agriculture and Food Research Organization (NARO), Ayabe, Kyoto, 623-0035,
Japan

³Faculty of Agriculture, Yamaguchi University, Yoshida, Yamaguchi, 753-8515,
Japan

^aCorresponding author:

Rieko Niwa

Tel: +81-29-838-8257

Fax: +81-29-838-8351

E-mail: rniwa@affrc.go.jp

^bPresent address:

National Institute for Agro-Environmental Sciences (NIAES), 3-1-3 Kannondai,
Tsukuba, Ibaraki, 305-8604, Japan

Abstract

Plasmodiophora brassicae is a soil-borne obligate intracellular parasite in the phylum Cercozoa of the Rhizaria that causes clubroot disease of crucifer crops. To control the disease, understanding the distribution and infection routes of the pathogen is essential, and thus development of reliable molecular markers to discriminate geographic populations is required. In this study, the nuclear ribosomal RNA gene (rDNA) repeat unit of *P. brassicae* was determined, with particular emphasis on the structure of large subunit (LSU) rDNA, in which polymorphic regions were expected to be present. The complete rDNA complex was 9513 bp long, which included the small subunit, 5.8S and LSU rDNAs as well as the internal transcribed spacer and intergenic spacer regions. Among eight field populations collected from throughout Honshu Island, Japan, a 1.1 kbp region of the LSU rDNA, including the divergent 8 domain, exhibited intraspecific polymorphisms that reflected geographic isolation of the populations. Two new group I introns were found in this region in six out of the eight populations, and the sequences also reflected their geographic isolation. The polymorphic region found in this study may have potential for the development of molecular markers for discrimination of field populations/isolates of this organism.

Keywords

Cercozoa, divergent domain in LSU rDNA, group I intron, intraspecific polymorphism, nuclear rDNA, *Plasmodiophora brassicae*

Introduction

Plasmodiophora brassicae is a soil-borne plant pathogen that causes clubroot disease of crucifers. Although it was traditionally classified as a fungus, recent molecular phylogenetic analyses of the small subunit (SSU) ribosomal RNA (Castlebury and Domier 1998; Cavalier-Smith and Chao 1997), actin and polyubiquitin genes (Archibald and Keeling 2004) suggest that it belongs to the protist phylum Cercozoa in the kingdom Rhizaria (Archibald and Keeling 2004; Cavalier-Smith and Chao 2003). Within the Cercozoa, the order Plasmodiophorida has attracted particular attention because it includes not only *P. brassicae* but also several plant pathogens of worldwide importance, such as *Spongospora subterranea*, which causes potato powdery scab disease, and *Polymyxa* spp., which transmit various plant viruses (Braselton 1995; Bulman et al. 2001). Information about the genomic structure of plasmodiophorids, however, is limited due to their nature as obligate intracellular parasites.

Pathotypes of *P. brassicae* have been characterized from sets of different hosts (Buczaki et al. 1975; Hatakeyama et al. 2004; Kuginuki et al. 1999; Somé et al. 1996; Williams 1966), but this process is time-consuming, labor-intensive, and subject to varying environmental conditions. Identification of the geographic populations/isolates of *P. brassicae* is also important to understand the distribution and infection routes. Therefore, development of reliable molecular markers that enable rapid identification of the pathotypes and geographic populations is required. For this purpose, molecular characterization of the genome is essential. Recently, Bulman et al. (2007) compared the nucleotide sequences of 24 genes in the *P. brassicae* genome with those of the corresponding cDNA and found that the genome was rich in spliceosomal introns. Random amplified polymorphic DNA (RAPD) analysis on *P. brassicae* populations that differed in pathogenicity revealed that the genome was highly polymorphic among the populations, although correlations between the pathogenicity and RAPD patterns were not clear (Manzanares-Dauleux et al. 2001; Osaki et al. 2008). In *Polymyxa graminis*, on the other hand, sequence variation in the two transcribed spacer (ITS1 and ITS2) regions of the rDNA were observed among the geographical populations and also among those with different host ranges (Legrève et al. 2002). In contrast, there was no sequence variation in the ITS sequences of *Spongospora subterranea* among Australasian and European populations, except for one from Scotland (Bulman and Marshall 1998). Among *P. brassicae* populations, however, little information about sequence variation among geographic populations of particular genes is available so far.

In eukaryotes, large variation in length and sequence have been reported in the

large subunit (LSU) rDNA. Twelve divergent domains that showed sequence variability in the LSU rDNA were first identified in mouse (Hassouna et al. 1984). These domains have been employed for identification and/or detailed characterization of species/isolates in a range of organisms (e.g., Chenuil et al. 2008; Lachance et al. 2000). The intergenic spacer (IGS) region of rDNA repeats in eukaryotes also shows extensive sequence variation and has been used widely as a molecular marker in the population biology of pathogenic fungi (e.g., Chang et al. 2008; Jackson et al. 1999; Pantou et al. 2003; Pramateftaki et al. 2000). These observations led us to expect that several polymorphic regions would be present in the unexplored region of rDNA of *P. brassicae*, although no sequence information is currently available. In the present study, the complete structure of the rDNA repeat unit of *P. brassicae* was determined, with particular emphasis on polymorphisms among field populations of different geographic origin.

Results

Structure of the ribosomal RNA gene repeat unit

Clubroot galls formed on *B. rapa* var. *pekinensis* grown in an experimental field of Nagoya University were harvested and designated as population NGY. To determine the complete sequence of the rDNA repeat unit of this population, spores were purified from several galls, and then genomic DNA was extracted. Seven DNA fragments were amplified with the primer pairs Pbr1/Pbr2r, Pb121/Pbr4r, NS7/ITS4, Pbr4/NDL22, NDL22f/IGS1r, V282/Pbr1r, and IGSa-10f/Pb121r (Supplemental Table 1 and Fig. 1a), cloned, sequenced from both ends of each PCR product, and assembled. The total length of the rDNA repeat unit was 9513 bp. The lengths of the SSU rDNA, internal transcribed spacer 1 (ITS1)-5.8S rDNA-ITS2 region, LSU rDNA, and IGS region were 3105, 465, 3611, and 2332 bp, respectively, based on sequence alignment and comparisons to those of other eukaryotes. Three introns in the SSU rDNA that were described in Castlebury and Domier (1998) were named Pbr.S516, Pbr.S943, and Pbr.S1506, with reference to the corresponding positions in *Escherichia coli* rDNA. To examine the presence of introns in the LSU rDNA, four PCR fragments that covered 3145 bp of the LSU rRNA were amplified with the primer pairs LRP4/NDL22, NDL22f/28s1r, 28s1/28s3r, and 28s3/P282r using cDNA that was reverse-transcribed from RNA as template and sequenced (Supplemental Table 1 and Fig. 1b). Comparative analysis between sequences of the rRNA and rDNA revealed that no intron was present

in the LSU rDNA of this population. The IGS region located between 8168–8429 nt contained five complete copies of a 14 bp palindrome (Fig. 2).

Polymorphisms among geographical populations

Because a preliminary database search revealed that little sequence variation was present in the 3.6 kbp genomic region consisting of SSU rDNA and ITS, including the three introns in the SSU rDNA, among the geographical populations across continents (those from Japan, Korea, Australia, and the UK; GenBank accession numbers AB094977, AF353997, AF231027, and Y12831, respectively; data not shown), we focused on the LSU rDNA and IGS regions thereafter and tried to obtain sequence information on these regions from several geographic populations from Japan. Although the amplicon for the IGS region was not obtained from any of the populations with the primer combinations used to amplify this region of the NGY, several parts of the LSU rDNA were successfully amplified and aligned. Among these, a PCR product amplified with the 28s1/28s3r primer set were found to be approximately 300 bp longer in some populations than in NGY and showed extensive sequence polymorphisms. Therefore, this region of the eight field populations collected from throughout Honshu Island (Fig. 3) was amplified and cloned. Sequence analysis on three to four clones raised from each of the populations revealed that six (Akita, Fukushima, Shiraoka, Matsuzaka, Hagi, and Hiroshima) out of the eight populations possessed two different introns at positions after 1440 or 1472 nt of the LSU rDNA of NGY, which were named Pbr.L1064 and Pbr.L1094, respectively, with reference to the corresponding position of *E. coli* rDNA (Fig. 1c). The Pbr.L1094 introns were found in all six populations, whereas the Pbr.L1064 intron was found only in two out of the three clones of the Matsuzaka population. Neither of the introns was found in the Yokosuka and NGY populations. Further alignment analysis on the 1.1 kbp exon sequences of this region revealed that extensive sequence polymorphisms were concentrated in the divergent 8 (D8) domain (Hassouna et al. 1984), which starts from 2135 to 2326 nt of the LSU rDNA of the NGY (Fig. 1c): nucleotide substitutions among the eight populations occurred in 65 positions within the D8 domain that is 192 bp in length, while that in the remaining 965 bp 5'-upstream of the D8 domain was observed only in 69 positions (Fig. 4a). Secondary structural analysis on the domain indicated that the region formed a typical Y-shaped structure of the D8 domain, with D8a and D8b helices (Michot and Bachellerie 1987) (Supplemental Figure 1), and that most of the sequence polymorphisms occurred in the D8b helix (Fig. 4b).

Based on the fact that the D8 domain harbors intraspecific polymorphisms, three to four clones of the 1.1 kbp exon sequences of LSU rDNA, including the domain (Fig. 1c) from each of the eight field populations, were subjected to phylogenetic analysis (Fig. 5). All of the geographic populations of *P. brassicae* formed a robust cluster, showing close relationships with members of the phylum Cercozoa, including *Cercomonas* sp., *Paracercomonas marina*, *Cercozoa* sp., *Protaspis grandis*, and *Bigelowiella natans*. Even though this analysis was performed on the exon sequence, the eight *P. brassicae* populations were differentiated into two subclusters according to the presence/absence of the introns in the LSU rDNA, with high bootstrap values: the NGY and Yokosuka populations that harbor no intron, and those harboring either the Pbr.L1064 or Pbr.L1094. Generally, geographic isolation of the populations was well reflected in the tree topology, except for the Matsuzaka population, in which the three clones were split into two distinct clusters. In this population, one clone that harbored the Pbr.L1094 (a common intron present in six out of the eight populations) formed a cluster with the Shiraoka, Hagi, and Hiroshima populations, and the other two that harbored the Pbr.L1064 (a rare intron found only in this population) formed an independent cluster related to the Akita and Fukushima populations. It is also noteworthy that sequence polymorphisms were observed not only among the geographic populations but also within the individual populations.

Characterization and polymorphisms of the introns in large subunit rDNA

The Pbr.L1064 and Pbr.L1094 in the LSU rDNA were characterized as group I introns according to primary and secondary structural analyses, as follows: i) the occurrence of group I introns in the same positions is known in other organisms (Jackson et al. 2002); ii) the last exon and intron bases were uracil and guanine, respectively (Cech 1988); and iii) all of the paired elements P1–P10 (Burke et al. 1987; Cech et al. 1994; Michel and Westhof 1990) were found through the secondary structure predictions (Fig. 6a, b). Pbr.L1064 and Pbr.L1094 were further classified into subgroup IC1, which possesses a bulged nucleotide A in the P7 helix and the P2.1 and P5abc helices (Michel and Westhof 1990).

The LSU rDNA intron sequences of three clones from each of the six populations were subjected to phylogenetic analysis (Fig. 7). The type of introns and geographic isolation of the populations were also reflected in the tree topology to the same extent as in the phylogeny of the exon sequences of the partial LSU rDNA, i.e., the Pbr.L1064 found only in two clones of the Matsuzaka population formed an

independent cluster, whereas the Pbr.L1094 formed two small clusters of the Akita and Fukushima populations, and one large cluster of the other populations, including a clone of the Pbr.L1094 from the Matsuzaka population.

Discussion

Structure of nuclear rDNA repeat unit of Plasmodiophora brassicae

The complete nucleotide sequence of the *P. brassicae* nuclear rDNA repeat unit that consisted of the three coding regions SSU, 5.8S, and LSU rDNA, interrupted by the three non-coding regions ITS1, ITS2, and IGS, was determined in one of the field populations in the present study. In the Rhizaria, sequence information on the complete rDNA repeat unit including the LSU rDNA and IGS region is quite limited, although that on the SSU rDNA and ITS regions of plasmodiophorids is abundant in the database. Moreira et al. (2007) demonstrated that not only SSU rDNA sequences but also LSU rDNA sequences are effective taxonomic markers for eukaryote phylogeny, especially for unculturable organisms to which application of multi-gene analysis is difficult. Our phylogenetic analysis based on the exon sequences of LSU rDNA confirmed that *P. brassicae* belongs to the Cercozoa in the Rhizaria, as proposed by Archibald and Keeling (2004) and Cavalier-Smith and Chao (2003), in which the sequences of actin and polyubiquitin and that of the SSU rDNA were used as taxonomic markers.

The complete sequence of the rDNA repeat unit obtained in the present study is actually a 'conceptual assemblage' of seven PCR fragments that may originate from different genotypes/rDNA copies within the population NGY. However, we consider that the nucleotide sequence, at least all except for the IGS region, would provide sufficiently realistic information about a representative sequence of one particular genotype/rDNA copy of the population for the following reasons: i) the SSU rDNA, ITS, and 5' end of LSU rDNA regions that were covered by four PCR fragments (Fig. 1) are highly conserved within the population, and ii) the rest of the LSU rDNA, in which the newly found introns and D8 domain (the polymorphic regions) are present, were covered by a 2.8 kbp single PCR fragment. The IGS region, however, was covered by two fragments in which the polymorphic regions of the LSU rDNA was uncovered, implying that there is a possibility that the sequences of the LSU rDNA and IGS were from different genotypes/rDNA copies, if the IGS harbors polymorphisms within the population. Establishing single spore isolates from the population and subsequent analysis on the regions is required.

Two new self-splicing group I introns were found in the LSU rDNA region of *P. brassicae*. Insertions of several group I introns into an rDNA repeat unit have been reported in fungi (Gargas et al. 1995; Nikoh and Fukatsu 2001) and algae (Turmel et al. 1993). The fact that *P. brassicae* possesses five types of group I intron in the rDNA repeat unit, three in the SSU, and two in LSU rDNAs supports the idea that its genome is rich in introns (Bulman et al. 2007). Group I introns have been found in a diverse range of higher organisms, including fungi, protists, and algae, where they occur in the nuclear, mitochondrial, and chloroplast genomes (Cech 1988). An important insight into group I intron evolution is that the introns inserted at the same site are phylogenetically closely related (monophyletic) even though they reside in distantly related organisms (Nikoh and Fukatsu 2001). In fact, the newly found introns Pbr.L1094 and Pbr.L1064 were phylogenetically distinct, suggesting that the acquisition processes of the two introns might be evolutionarily distinct even though their insertion sites are quite close.

Intraspecific polymorphisms among the geographic populations

The D8 domain in eukaryotic LSU rDNA is one of the most rapidly evolving divergent domains that show extensive variability in length and sequence among species (Hassouna et al. 1984). Intraspecific polymorphisms in the domain, however, have been reported only in the ascomycete fungus *Clavispora opuntiae* (Lachance et al. 2000) and the echinoderm *Echinocardium cordatum* (Chenuil et al. 2008). The present study demonstrated that the protist *P. brassicae* also harbors intraspecific polymorphisms in this domain. Chenuil et al. (2008) estimated the evolutionary rate of echinoderms based on the nucleotide substitutions in the D8 domain and suggested that the D8a helix evolved more rapidly than other parts of the domain, because most of the nucleotide substitutions were concentrated within the D8a helix. In contrast, the substitutions among the *P. brassicae* populations occurred mainly in the D8b helix, suggesting that the helix is likely to evolve more rapidly in this organism.

The newly found introns in the LSU rDNA also showed intraspecific polymorphisms, reflecting geographic isolation, in contrast to those in the SSU rDNA, which are highly conserved within the species. Furthermore, it is of particular interest that the phylogenetic relationships among the populations revealed by the intron sequences were quite consistent with those revealed by the exon sequences that include the D8 domain, even though the introns were located more than 600 bp upstream of the domain. In addition, the presence/absence of introns, as well as the types of intron (in the case of the Matsuzaka population) were also reflected in the sequences of the

domain. These results strongly suggest that the introns and D8 domain evolved in parallel within the individual populations and that acquisition/deletion of the introns rarely occurred during evolution.

Variation in the intraspecific sequences has been found in the IGS region of phytopathogenic (Pramateftaki et al. 2000) and entomopathogenic (Pantou et al. 2003) fungi, likely due to multiple insertions of perfect and/or imperfect copies and deletions. The IGS region of *P. brassicae*, however, was successfully cloned only in the NGY, suggesting that the region in other populations may also contain several palindrome sequences and AT-rich regions as found in the NGY, and that these primary structures might interfere with PCR amplification. Although it has been generally accepted that the IGS region has strong potential as a molecular marker for the population biology of eukaryotic microorganisms, its practical application to *P. brassicae* seems difficult.

The D8 domain and introns in the LSU rDNA showed sequence variation not only among the geographic populations but also within individual populations. These sequences were obtained from genomic DNA prepared from multiple galls collected from each of the populations, and under these conditions, two possible interpretations for the observed variation are considered. One is that the field populations of *P. brassicae* were actually heterogeneous, probably even in single galls, i.e., several genotypes coexist in the field (e.g., Manzanares-Dauleux et al. 2001; Somé et al. 1996; Xue et al. 2008). Another interpretation is that the individual populations consisted of homogeneous genotypes but several different rDNA copies coexist in their genomes, as found in Foraminifera, which also belong to the Rhizaria (Holzmann et al. 1996). Sequence analysis on the region of single spore isolates of *P. brassicae* will answer this question.

Conclusion

The present study provides new information about intraspecific sequence polymorphisms in the rDNA repeat unit of *P. brassicae*. The sequences of the D8 domain and newly found group I introns in the LSU rDNA showed extensive sequence variation, reflecting geographic isolation. The IGS region was also expected to harbor polymorphisms, although the sequence could not be compared among the geographic populations in the present study. We conclude, however, that using an exon sequence as a molecular marker for discrimination of field populations/isolates is more reliable and practical than using a long non-coding region such as IGS, because a short polymorphic region is generally located between conserved regions, for which reliable PCR primers

could easily be designed (e.g., the D1 and D2 regions of LSU rDNA that can be amplified by universal primers have been extensively applied for the identification and community analysis of glomeromycotan fungi, which are obligate plant symbionts; An et al. 2008; Li et al. 2009; Pivato et al. 2007). Furthermore, the polymorphic regions found in the present study may be applicable to the identification of geographic population/isolates of other plasmodiophorid organisms. Further study to obtain sequence information on this region from a range of organisms is required.

Methods

Collection of field populations and subculture conditions

The clubroot galls of *P. brassicae* on Chinese cabbage (*Brassica rapa* var. *pekinensis*) or cabbage (*B. oleracea* var. *capitata*) were collected from an experimental field of Nagoya University (NGY) (Aichi prefecture, N35°, E137°) in 2003, from arable land in Akita City (Akita prefecture, N39°, E140°) in 2009, Fukushima City (Fukushima prefecture, N37°, E140°) in 2009, Shiraoka Town (Saitama prefecture, N35°, E139°) in 2009, Yokosuka City (Kanagawa prefecture, N35°, E139°) in 2009, Matsuzaka City (Mie prefecture, N34°, E136°) in 2009, Hagi City (Yamaguchi prefecture, N34°, E131°) in 2000, and Hiroshima City (Hiroshima prefecture, N34°, E132°) in 1993, and stored at –80°C (Fig. 3 and Supplemental Table 2). If nucleic acids extracted from the stored galls were degraded, fresh material was obtained as described below. Resting spores, prepared by homogenizing the frozen galls in water, were mixed with soil, and then *B. rapa* var. *perviridis* cv. Komatsuna was grown on the soil in 2 L square plastic pots in a multi-compartment greenhouse at 26°C for 40 d. To avoid cross-contamination, pathogens from different geographic origins were not cultured in the same compartment of the greenhouse. The galls formed on the roots were harvested and subjected to immediate DNA/RNA extraction.

DNA extraction

The frozen galls were thawed in water, washed thoroughly under running tap water, surface-sterilized with 70% (v/v) ethanol for 30 s, homogenized in sterilized water, and filtered through eight layers of cheesecloth. The filtrate was centrifuged at 1450×g for 15 min. The pellet was washed twice with sterilized water, resuspended in a small amount of sterilized water, layered on 10 mL Percoll (GE Healthcare Bio-Sciences,

Tokyo, Japan) in a 50-mL plastic tube, and centrifuged at 1010×g for 10 min. The supernatant was transferred to a new tube, mixed with 40 mL sterilized water, and centrifuged at 1450×g for 15 min. The resultant pellet was washed three times with sterilized water and resuspended in a small amount of sterilized water. The suspension was transferred to a 2 mL tube with an O-ring sealed cap (Yasui Kikai, Osaka, Japan) and centrifuged at 13200×g for 1 min. Then the pellet (resting spores) was frozen at −80°C for 1 h, freeze-dried overnight, and ground in a Multi-Beads Shocker (Yasui Kikai, Osaka, Japan) with zirconium beads (0.5 mm in diameter) at 2500 rpm for 6 × 30 s at 30 s intervals. DNA was extracted from the resting spores using the DNeasy Plant Mini Kit (Qiagen, Tokyo, Japan) according to the manufacturer's instructions and stored at −30°C.

PCR amplification, cloning, and sequencing

PCR amplification was performed in a 25 µL reaction mixture with the Expand High-Fidelity PLUS PCR System (Roche Diagnostics, Tokyo, Japan) containing 0.5 µM primers and 1 µL or 10 µL of the template DNA solution in the PTC-225 DNA Engine Tetrad thermal cycler (GMI, Minnesota, USA). An initial denaturation step of 94°C for 2 min was followed by 30 cycles of denaturation, annealing, and extension as follows: 94°C for 15 s, annealing for 60 s (the temperatures for each primer are listed in Supplemental Table 1), and 72°C for 60 s in the first 10 cycles. The same parameters were used for the next 11–30 cycles, except that the 72°C extension steps were lengthened to 10 s cycle^{−1}. After 30 cycles, the samples were incubated for an additional 10 min at 72°C. The PCR products were analyzed on a 1.2% agarose gel in 0.5 × TBE buffer.

The PCR primers used to amplify different regions of the gene complex are shown in Supplemental Table 1, and their relative positions in the rDNA transcription unit are indicated in Fig. 1a and Fig. 1c. All primers used to amplify the rDNA transcription unit of the population NGY were designed based on the sequence information of *P. brassicae* SSU rDNA and those of LSU rDNA of other eukaryotes in the database. The primers used to amplify the LSU rDNA regions of the other populations were designed based on the sequence of the NGY determined in this study.

The PCR products were purified by the MonoFas DNA Purification Kit (GL Science, Tokyo, Japan) and cloned into the pT7Blue T-vector (Novagen/Merck, Tokyo, Japan) according to the manufacturer's instructions. The nucleotide sequences of three to four randomly chosen clones were determined via the dideoxy-sequencing method

using the BigDye Terminator v3.0 or v3.1 Cycle Sequencing Kit with the ABI PRISM 3100 Genetic Analyzer (Applied Biosystems, Tokyo, Japan). The GenBank accession numbers of the sequenced clones are listed in Supplemental Table 2.

LSU rRNA sequencing

Approximately 100 mg fresh gall was frozen with liquid nitrogen in a 2 mL tube with an O-ring sealed cap and ground in a Multi-Beads Shocker with a metal cone at 2500 rpm for 6 × 30 s at 30 s intervals, during which the sample was cooled in liquid nitrogen. Total RNA was extracted using the RNeasy Plant Mini Kit (Qiagen) according to the manufacturer's instructions. First-strand cDNA was synthesized using AMV reverse transcriptase (Roche Diagnostics, Tokyo, Japan) with the P282r or random primers according to the manufacturer's instructions. The PCR primers used in this experiment are shown in Supplemental Table 1, and their relative positions are indicated in Fig. 1b. Detailed conditions for PCR amplification, cloning, and sequencing are described in the section 'PCR amplification, cloning and sequencing.'

Sequence similarity and phylogenetic analyses

Sequence similarity of the LSU rDNA among the different field populations was analyzed using the AlignX program of the Vector NTI Suite 9.0 (Invitrogen, California, USA) to obtain an overview of the distribution of variable regions in the gene. For phylogenetic analysis on the LSU rDNA, the 1.1 kbp exon sequences including the D8 domain (Fig. 1c) and those of the corresponding region from other eukaryotes obtained from the GenBank were employed. For analysis of the LSU rDNA introns, approximately 150 nt sequences that form the conserved secondary structures of the group I intron (from P3 to P7 elements, see Fig. 6) were employed. These datasets were aligned using the ClustalX program (Thompson et al. 1997), and maximum likelihood analysis with the GTR + SS and GTR + γ models was applied for phylogenies of the LSU rDNA and introns, respectively, using the PAUP 4.0b software (Swofford 2002). The reliability of the tree topologies were evaluated both from the bootstrap values calculated from the maximum likelihood analysis with 100 replications (Felsenstein 1985) and from posterior probabilities calculated with Bayesian analysis using the MrBayes 3.1.3 software (Ronquist and Huelsenbeck 2003).

Intron secondary structure prediction

The secondary structures were predicted using the models proposed in Cech (1988) and Michel and Westhof (1990), and the structural elements were defined based on the nomenclature of Johansen and Haugen (2001) and from the standard secondary structure representation for group I introns (Cech et al. 1994). The individual elements (P1–P9 stem loops) were identified by comparison with available group I intron sequences from the Comparative RNA website at <http://www.rna.cccb.utexas.edu/> (Cannone et al. 2002) and drawn using the *mfold* web server at <http://mfold.bioinfo.rpi.edu/> (Zuker 2003). The final structures were drawn using the XRNA software provided by the RNA Center, University of California, Santa Cruz, at <http://rna.ucsc.edu/rnacenter/xrna/xrna.html>.

Acknowledgments

We are grateful to Y. Tahara, M. Maesaka, and S. Mizuno at Nagoya University for providing the clubroot inocula. This study was supported by a Grant-in-Aid for Scientific Research (17658031) from the Japan Society for the Promotion of Science (TE).

References

- An G-H, Miyakawa S, Kawahara A, Osaki M, Ezawa T** (2008) Community structure of arbuscular mycorrhizal fungi associated with pioneer grass species *Miscanthus sinensis* in acid sulfate soils: habitat segregation along pH gradients. *Soil Sci Plant Nutr* **54**: 517-528
- Archibald JM, Keeling PJ** (2004) Actin and ubiquitin protein sequences support a cercozoan/foraminiferan ancestry for the plasmodiophorid plant pathogens. *J Eukaryot Microbiol* **51**: 113-118
- Braselton JP** (1995) Current status of the plasmodiophorids. *Crit Rev Microbiol* **21**: 263 - 275
- Buczacki S, Toxopeus H, Mattusch P, Johnston T, Dixon GR, Hobolth LA** (1975) Study of physiologic specialization in *Plasmodiophora brassicae*: proposals for attempted rationalization through an international approach. *Trans Br Mycol Soc* **65**: 295-303
- Bulman SR, and Marshall JW** (1998) Detection of *Spongospora subterranea* in potato tuber lesions using the polymerase chain reaction (PCR). *Plant Pathol* **47**:

- 443 759-766
- 444 **Bulman SR, Kühn SF, Marshall JW, Schnepf E** (2001) A phylogenetic analysis of
445 the SSU rRNA from members of the Plasmodiophorida and Phagomyxida.
446 Protist **152**: 43-51
- 447 **Bulman SR, Ridgway HJ, Eady C, Conner AJ** (2007) Intron-rich gene structure in
448 the intracellular plant parasite *Plasmodiophora brassicae*. Protist **158**: 423-433
- 449 **Burke JM, Belfort M, Cech TR, Davies RW, Schweyen RJ, Shub DA, Szostak JW,**
450 **Tabak HF** (1987) Structural conventions for group I introns. Nucl Acids Res
451 **15**: 7217-7221
- 452 **Cannone JJ, Subramanian S, Schnare MN, Collett JR, D'Souza LM, Du YS, Feng**
453 **B, Lin N, Madabusi LV, Müller KM et al** (2002) The Comparative RNA Web
454 (CRW) Site: an online database of comparative sequence and structure
455 information for ribosomal, intron, and other RNAs. BMC Bioinformatics **3**: 2,
456 Erratum in: BMC Bioinformatics, 2002, 3:15
- 457 **Castlebury LA, Domier LL** (1998) Small subunit ribosomal RNA gene phylogeny of
458 *Plasmodiophora brassicae*. Mycologia **90**: 102-107
- 459 **Cavalier-Smith T, Chao EE** (1997) Sarcomonad ribosomal RNA sequences, rhizopod
460 phylogeny, and the origin of euglyphid amoebae. Arch Protist **147**: 227-236
- 461 **Cavalier-Smith T, Chao EE** (2003) Phylogeny and classification of phylum Cercozoa
462 (Protozoa). Protist **154**: 341-358
- 463 **Cech TR** (1988) Conserved sequences and structures of group I introns: building an
464 active site for RNA catalysis - a review. Gene **73**: 259-271
- 465 **Cech TR, Damberger SH, Gutell RR** (1994) Representation of the secondary and
466 tertiary structure of group I introns. Nat Struct Biol **1**: 273-280
- 467 **Chang JC, Hsu MML, Barton RC, Jackson CJ** (2008) High-frequency intragenomic
468 heterogeneity of the ribosomal DNA intergenic spacer region in *Trichophyton*
469 *violaceum*. Eukaryot Cell **7**: 721-726
- 470 **Chenuil A, Egea E, Rocher C, Touzet H, Féral JP** (2008) Does hybridization increase
471 evolutionary rate? Data from the 28S-rDNA D8 domain in echinoderms. J Mol
472 Evol **67**: 539-550
- 473 **Felsenstein J.** (1985) Confidence limits on phylogenies: an approach using the
474 bootstrap. Evolution **39**: 783-791
- 475 **Gargas A, DePriest PT, Taylor JW** (1995) Positions of multiple insertions in SSU
476 rDNA of lichen-forming fungi. Mol Biol Evol **12**: 208-218
- 477 **Hassouna N, Michot B, Bachellerie JP** (1984) The complete nucleotide-sequence of
478 mouse 28S ribosomal-RNA gene. Implications for the process of size increase of

the large subunit ribosomal RNA in higher eukaryotes. Nucl Acids Res **12**:
3563-3583

Hatakeyama K, Fujimura M, Ishida M, Suzuki T (2004) New classification method
for *Plasmodiophora brassicae* field isolates in Japan based on resistance of F₁
cultivars of chinese cabbage (*Brassica rapa* L.) to clubroot. Breed Sci **54**:
197-201

Holzmann M, Piller W, Pawlowski J (1996) Sequence variations in the large-subunit
ribosomal RNA gene of *Ammonia* (Foraminifera, Protozoa) and their
evolutionary implications. J Mol Evol **43**: 145-151

Jackson CJ, Barton RC, Evans EGV (1999) Species identification and strain
differentiation of dermatophyte fungi by analysis of ribosomal DNA intergenic
spacer regions. J Clin Microbiol **37**: 931-936

Jackson SA, Cannone JJ, Lee JC, Gutell RR, Woodson SA (2002) Distribution of
rRNA introns in the three-dimensional structure of the ribosome. J Mol Biol
323: 35-52

Johansen S, Haugen P (2001) A new nomenclature of group I introns in ribosomal
DNA. RNA **7**: 935-936

Kuginuki Y, Yoshikawa H, Hirai M (1999) Variation in virulence of *Plasmodiophora*
brassicae in Japan tested with clubroot-resistant cultivars of chinese cabbage
(*Brassica rapa* L. ssp. *pekinensis*). Eur J Plant Pathol **105**: 327-332

Lachance MA, Starmer WT, Bowles JM, Phaff HJ, Rosa CA (2000) Ribosomal
DNA, species structure, and biogeography of the cactophilic yeast *Clavispora*
opuntiae. Can J Microbiol **46**: 195-210

Legrève A, Delfosse P and Maraite H (2002) Phylogenetic analysis of *Polymyxa*
species based on nuclear 5.8S and internal transcribed spacers ribosomal DNA
sequences. Mycol Res **106**: 138-147

Li LF, Li T, Zhang Y, Zhao ZW (2010) Molecular diversity of arbuscular mycorrhizal
fungi and their distribution patterns related to host-plants and habitats in a hot
and arid ecosystem, southwest China. FEMS Microbiol Ecol **71**: 418-427

Manzanares-Dauleux MJ, Divaret I, Baron F, Thomas G (2001) Assessment of
biological and molecular variability between and within field isolates of
Plasmodiophora brassicae. Plant Pathol **50**: 165-173

Michel F, Westhof E (1990) Modeling of the 3-dimensional architecture of group I
catalytic introns based on comparative sequence-analysis. J Mol Biol **216**:
585-610

Michot B, Bachellerie JP (1987) Comparisons of large subunit ribosomal RNAs reveal

- some eukaryote-specific elements of secondary structure. *Biochimie* **69**: 11-23
- Moreira D, von der Heyden S, Bass D, López-García P, Chao E, Cavalier-Smith T** (2007) Global eukaryote phylogeny: combined small- and large-subunit ribosomal DNA trees support monophyly of Rhizaria, Retaria and Excavata. *Mol Phylogenet Evol* **44**: 255-266
- Nikoh N, Fukatsu T** (2001) Evolutionary dynamics of multiple group I introns in nuclear ribosomal RNA genes of endoparasitic fungi of the genus *Cordyceps*. *Mol Biol Evol* **18**: 1631-1642
- Osaki K, Fujiyama S, Nakayama A, Shimizu Y, Ito S, Tanaka S** (2008) Relation between pathogenicity and genetic variation within *Plasmodiophora brassicae*. *J General Plant Pathol* **74**: 281-288
- Pantou MP, Mavridou A, Typas MA** (2003) IGS sequence variation, group I introns and the complete nuclear ribosomal DNA of the entomopathogenic fungus *Metarhizium*: excellent tools for isolate detection and phylogenetic analysis. *Fungal Genet Biol* **38**: 159-174
- Pivato B, Mazurier S, Lemanceau P, Siblot S, Berta G, Mougél C, van Tuinen D** (2007) *Medicago* species affect the community composition of arbuscular mycorrhizal fungi associated with roots. *New Phytol* **176**: 197-210
- Pramateftaki PV, Antoniou PP, Typas MA** (2000) The complete DNA sequence of the nuclear ribosomal RNA gene complex of *Verticillium dahliae*: intraspecific heterogeneity within the intergenic spacer region. *Fungal Genet Biol* **29**: 19-27
- Ronquist F, Huelsenbeck JP** (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**: 1572-1574
- Somé A, Manzanares MJ, Laurens F, Baron F, Thomas G, Rouxel F** (1996) Variation for virulence on *Brassica napus* L. amongst *Plasmodiophora brassicae* collections from France and derived single-spore isolates. *Plant Pathol* **45**: 432-439
- Swofford DL** (2002) PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version 4. Sinauer Associates, Sunderland, Massachusetts
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG** (1997) The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucl Acids Res* **25**: 4876-4882
- Turmel M, Gutell RR, Mercier JP, Otis C, Lemieux C** (1993) Analysis of the chloroplast large subunit ribosomal RNA gene from 17 Chlamydomonas taxa. Three internal transcribed spacers and 12 group I intron insertion sites. *J Mol Biol* **232**: 446-467

- van Tuinen D, Jacquot E, Zhao B, Gollotte A, Gianinazzi-Pearson V** (1998)
Characterization of root colonization profiles by a microcosm community of
arbuscular mycorrhizal fungi using 25S rDNA-targeted nested PCR. *Mol Ecol* **7**:
879-887
- White TJ, Bruns T, Lee S, Taylor J** (1990) Amplification and direct sequencing of
fungal ribosomal RNA genes for phylogenetics. In: *PCR Protocols, a Guide to
Methods and Applications*. Academic Press, San Diego, pp 315-322
- Williams PH** (1966) A system for determination of races of *Plasmodiophora brassicae*
that infect cabbage and rutabaga. *Phytopathol* **56**: 624-626
- Xue S, Cao T, Howard RJ, Hwang SF, Strelkov SE** (2008) Isolation and variation of
single-spore isolates of *Plasmodiophora brassicae* from Canada. *Plant Disease*
92: 456-462
- Zuker M** (2003) Mfold web server for nucleic acid folding and hybridization prediction.
Nucl Acids Res **31**: 3406-3415

567

582

Figure 2. Schematic representation of the 14 bp palindrome sequences (box) found in the IGS region of *Plasmodiophora brassicae* population NGY. Arrows indicate the centers of palindromes interrupted by the repetitive sequences. Guanine in parentheses represents both the insertion and deletion clones observed. Numbers in bold at the 5' and 3' ends indicate nucleotide positions within the IGS region. Numbers in brackets represent the number of nucleotides between palindromes.

Figure 3. Sampling sites of *Plasmodiophora brassicae* field populations on Honshu Island, Japan.

Figure 4. Sequence polymorphisms in the LSU rDNA among the *Plasmodiophora brassicae* field populations. a, Distribution of sequence polymorphisms in the 1.1 kbp exon sequences of the LSU rDNA. Representative single clones from each of the eight populations were aligned, and the levels of similarity were scored in a 0–1 range at each alignment position using the AlignX program of Vector NTI. Arrows indicate the insertion sites of the Pbr.L1064 and Pbr.L1094 introns. b, Alignment of the D8 domain of the representative clones from each of the field populations using the AlignX program. Identical nucleotides and deletions with reference to the NGY-1 sequence are indicated by dots and dashes, respectively. The yellow, pink, and blue boxes represent nucleotide sequences that correspond to the basal, D8a, and D8b helices, respectively.

Figure 5 Maximum likelihood phylogenetic tree of the 1.1 kbp exon of LSU rDNA (corresponding to 1218–2322 nt of the NGY sequence) of *Plasmodiophora brassicae* field populations NGY, Akita, Fukushima (Fuku), Shiraoka (Shira), Yokosuka (Yoko), Matsuzaka (Matsu), Hagi, and Hiroshima (Hiro) with those of selected fungi and glaucophyta in eukaryotes. The sequences from three or four randomly chosen clones from each population were sequenced and subjected to analysis. Numbers at the branches are maximum likelihood bootstrap values (left) and posterior probabilities (right), and only those higher than 50% and 0.50, respectively, are shown. The GenBank accession numbers of the sequences are shown in parentheses. *Acanthobothrium parviuncinatum* and *Clistobothrium montaukensis*, classified as Metazoa, were employed as the outgroup.

(a) Pbr.L1064

(b) Pbr.L1094

Figure 6. Predicted RNA secondary structures of group I introns found in the LSU rDNA of *Plasmodiophora brassicae*. The structures of Pbr.L1064 (a) and Pbr.L1094 (b) introns were drawn based on sequences of the Matsuzaka (Matsu-2 clone) and Hagi (Hagi-2 clone) populations, respectively. Uppercase and lowercase letters indicate intron and exon sequences, respectively. The conserved paired elements of P1–P10 and joining regions of J3/J4 are indicated.

Figure 7. Maximum likelihood phylogenetic tree of the Pbr.L1064 and Pbr.L1094 introns in the LSU rDNA of the *Plasmodiophora brassicae* field populations Akita, Fukushima (Fuku), Shiraoka (Shira), Matsuzaka (Matsu), Hagi, and Hiroshima (Hiro). Numbers at the branches are maximum likelihood bootstrap values (left) and posterior probabilities (right), and only those higher than 50% and 0.50, respectively, are shown. The Pbr.S516 intron (subgroup IE) in the SSU rDNA was employed as the outgroup.

652

Supplemental table 1

PCR primers and annealing temperatures used for sequencing the ribosomal RNA gene repeat unit of *Plasmodiophora brassicae*.

Primer	Sequence (5'-3')	Annealing (°C) ^a	Targets (source)
Pbr1	GGTTGATCCTGCCAGTAGTC	54	SSU rDNA (This study)
Pbr1r	GACTACTGGCAGGATCAACC	54	SSU rDNA (This study)
Pb121	GGATACAAAAACCAAACCTGGC	52	SSU rDNA (This study)
Pb121r	GCCAGGTTTGGTTTTTGTATCC	53	SSU rDNA (This study)
Pbr2r	CTCTATGCCCGAATCGCTTC	53	SSU rDNA (This study)
NS7	GAGGCAATAACAGGTCTGTGATGC	61	SSU rDNA (White et al. 1990)
Pbr4	GTGTCGCTTAAGATATAGTC	48	SSU rDNA (This study)
Pbr4r	GACTATATCTTAAGCGACAC	48	SSU rDNA (This study)
ITS4	TCCTCCGCTTATTGATATGC	50	LSU rDNA (White et al. 1990)
LRP4	CGTTGGAAGGCGGCATCA	54	LSU rDNA (This study)
NDL22f	CGTCTTGAAACACGGACCA	52	LSU rDNA (This study)
NDL22	TGGTCCGTGTTTCAAGACG	52	LSU rDNA (van Tuinen et al. 1998)
28s1	CCGGAGCTGAAACAGCTT	52	LSU rDNA (This study)
28s3r	AATTAAACAGTCGGATTCCCC	50	LSU rDNA (This study)
V282	GTGGGATAACGGCTGAACG	54	LSU rDNA (This study)
P282r	CGTTCAGCCGTAATCCTAC	52	LSU rDNA (This study)
IGS1r	GTTGATGTGTTGTAAGGGGTATG	53	IGS (This study)
IGSa-10f	GCATCACCTTAGCATTCGTTC	52	IGS (This study)

^aAnnealing temperature in PCR.

654

654

Supplemental table 2

655

GenBank accession numbers of the sequenced clones.

656

Isolates (origin)/clone names	Sequence encoding	Accession no.
NGY (Aichi pref., N35° E137°)		
Assemblage of 7 PCR fragments	Complete rRNA gene repeat unit	AB526843
NGY-1	Partial LSU rRNA gene	AB588900
NGY-2	Partial LSU rRNA gene	AB588901
NGY-3	Partial LSU rRNA gene	AB588902
NGY-4	Partial LSU rRNA gene	AB588903
Akita (Akita pref., N39° E140°)		
Akita-5	Partial LSU rRNA gene	AB588904
Akita-7	Partial LSU rRNA gene	AB588905
Akita-8	Partial LSU rRNA gene	AB588906
Fukushima (Fukushima pref., N37° E140°)		
Fuku-4	Partial LSU rRNA gene	AB588907
Fuku-7	Partial LSU rRNA gene	AB588908
Fuku-8	Partial LSU rRNA gene	AB588909
Shiraoka (Saitama pref., N35° E139°)		
Shira-7	Partial LSU rRNA gene	AB588910
Shira-9	Partial LSU rRNA gene	AB588911
Shira-10	Partial LSU rRNA gene	AB588912
Yokosuka (Kanagawa pref., N35° E139°)		
Yoko-1	Partial LSU rRNA gene	AB588913
Yoko-2	Partial LSU rRNA gene	AB588914
Yoko-10	Partial LSU rRNA gene	AB588915
Matsuzaka (Mie pref., N34° E136°)		
Matsu-2	Partial LSU rRNA gene	AB588916
Matsu-3	Partial LSU rRNA gene	AB588917
Matsu-9	Partial LSU rRNA gene	AB588918
Hagi (Yamaguchi pref., N34° E131°)		
Hagi-2	Partial LSU rRNA gene	AB526844
Hagi-7	Partial LSU rRNA gene	AB526845
Hagi-8	Partial LSU rRNA gene	AB526846
Hiroshima (Hiroshima pref., N34° E132°)		
Hiro-49	Partial LSU rRNA gene	AB526847
Hiro-50	Partial LSU rRNA gene	AB526848
Hiro-51	Partial LSU rRNA gene	AB526849

Supplemental Figure 1. Predicted RNA secondary structure of the D8 domain of *Plasmodiophora brassicae*. The basal D8a and D8b helices are indicated. The structure was drawn based on the sequence of population NGY (NGY-1 clone).