Protist, Vol. 158, 423—433, October 2007
http://www.elsevier.de/protis
Published online date 11 July 2007

Protist

ORIGINAL PAPER

Intron-rich Gene Structure in the Intracellular Plant Parasite *Plasmodiophora brassicae*

Simon Bulman[a,b,1], Hayley J. Ridgway[b], Colin Eady[a,b], and Anthony J. Conner[a,b]

[a]National Centre for Advanced Bio-Protection Technologies, Lincoln University, Canterbury, New Zealand
[b]New Zealand Institute for Crop & Food Research, Christchurch, New Zealand

Submitted December 22, 2006; Accepted April 28, 2007
Monitoring Editor: Keith Gull

***Plasmodiophora brassicae*, a pathogen of Brassicaceae plants, is grouped within the eukaryotic supergroup, the Rhizaria. Although a large diversity of protists is found in the Rhizaria, genomes of organisms within the group have barely been examined. In this study, we identified DNA sequences spanning or flanking 24 *P. brassicae* genes, eventually sequencing close to 44 kb of genomic DNA. Evidence from this preliminary genome survey suggested that splicing is an important feature of *P. brassicae* gene expression; the *P. brassicae* genes were rich in spliceosomal introns and two mini-exons of less than 20 bp were identified. Consensus splice sites and branch-point sequences in *P. brassicae* introns were similar to those found in other eukaryotes. Examination of the promoter and transcription start sites of genes indicated that *P. brassicae* transcription is likely to begin from initiator elements rather than TATA-box containing promoters. Where neighbouring genes were confirmed, intergenic distances were short, ranging from 44 to 470 bp, but a number of larger DNA fragments containing no obvious genes were also sequenced.**

Key words: genome structure; introns; *Plasmodiophora brassicae*; Rhizaria.

1. Introduction

Molecular phylogenetic studies are continually increasing knowledge of eukaryotic evolution and the breadth of microbial diversity. Eukaryotes are currently thought to coalesce into five or six supergroups, one of which, the Rhizaria, contains mostly amoeboid protists. Prominent protistan groups in the Rhizaria include the radiolarians, Foraminifera and Cercozoa (Keeling et al. 2005). Despite the Rhizaria being circumscribed largely on the basis of recent molecular evidence, very little is known about the genomes of these organisms. cDNA sequences have only been identified in large numbers from the amoeboflagellate alga *Bigelowiella natans* (Archibald et al. 2003; Rogers et al. 2004) and the foraminiferan *Reticulomyxa filosa* (Burki et al. 2006). The first genome sequencing project on a rhizarian (*B. natans*) has just begun (http://www.jgi.doe.gov/sequencing/cspseqplans2007.html).

The Plasmodiophorida is a cercozoan order (Archibald and Keeling 2004; Bass et al. 2005; Cavalier-Smith and Chao 1997) of special interest because it contains several plant-pathogens of worldwide importance (Braselton 1995, 2001). These include *Spongospora subterranea*, the cause of potato powdery scab disease, and *Polymyxa* species that serve as vectors of many plant viruses (Rochon et al. 2004).

[1]Corresponding author;
fax +64 3 3252 074
e-mail bulmans@crop.cri.nz (S. Bulman).

doi:10.1016/j.protis.2007.04.005

However, because plasmodiophorids are obligate intracellular parasites, gene sequences from these organisms are only now being identified. The most studied plasmodiophorid is *Plasmodiophora brassicae*, which causes club root disease in Brassicaceae including *Arabidopsis thaliana* (Ludwig-Müller 1999). Reports of genomic structures in *P. brassicae* are largely confined to a few examples of introns in genes. An early study showed the presence of self-splicing type I introns in the *P. brassicae* ribosomal small subunit gene (Castlebury and Domier 1998), and a single spliceosomal intron sequence was identified in a DNA fragment from the *P. brassicae* trehalose-6-phosphate synthase (TPS) gene (Brodmann et al. 2002). More recently, cDNA and genomic DNA sequences of a *P. brassicae* serine-threonine kinase gene (*PbSTKL1*) were published (Ando et al. 2006). The presence of multiple introns in *PbSTKL1* was reported, but no information on the structure of these introns was presented.

We recently identified a collection of cDNA sequences from *P. brassicae* (Bulman et al. 2006). With an initial aim to gain information about the structure of *P. brassicae* promoters, we set out to isolate sequences adjacent to a selection of these cDNAs. During this process, the sequences of many introns were also discovered. Here we present the results of this small-scale genome survey from *P. brassicae*.

Results

Genomic DNA from Actin and TPS Sequences

Genomic DNA sequence stretching 2.2 kb in the 5′ direction and 350 bp in the 3′ direction from the GenBank *P. brassicae* actin sequences was obtained (see Fig. 1 for a diagram of all genomic sequences obtained). This sequence contained the entire actin coding region; with PCR primers on either side of this region, we re-amplified and directly sequenced this gene (*PbACTI*).

During sequencing of *P. brassicae* actin cDNAs, a clone was found with a 5′ UTR sequence that did not match that of *PbACTI*. DNA-walking with primers based on this cDNA yielded the sequence of a second *P. brassicae* actin gene (*PbACTII*). *PbACTI* and *PbACTII* differed by 44 bp across the 1128 bp coding sequences. Neither gene sequence exactly matched any of the three actin *P. brassicae* GenBank sequences (Archibald and Keeling 2004). Both *PbACTI* and *PbACTII* had greatest similarity to the actin 3 sequence (AY452181), matching 711 and 715 of 733 bp respectively.

With PCR primers designed from the short GenBank *P. brassicae* TPS sequence (*PbTPS*; Brodmann et al. 2002) a complete 2640 bp *PbTPS* cDNA was amplified by RACE. By DNA walking, genomic DNA sequence spanning from 870 bp upstream of the *PbTPS* transcription start site (TSS) to 80 bp from the 3′ end of the *PbTPS* cDNA was obtained (Fig. 1).

New Genes were Identified in Flanking DNA

DNA sequences that flanked a further 12 genes from Bulman et al. (2006) were obtained (Table 1A). The new genomic DNA sequences obtained in this study are deposited under EMBL accessions AM411663—AM411680. Using RACE and RT-PCR, the sequences of seven new cDNAs in these genomic DNA sequences were identified (Table 1B).

In total, about 44 kb of *P. brassicae* genomic DNA was sequenced (Fig. 1). The longest DNA fragment of 6.4 kb contained four closely spaced genes, separated by 44 bp, 215 bp and 122 bp (Fig. 1). On another genomic DNA fragment, 97 bp separated the 3′ end of the *PbGTPX* cDNA from the TSS of the *PbPDA* cDNA. The maximum distance between confirmed cDNAs was 474 bp (*PbURM-PbACTI*). Several *P. brassicae* genomic DNA fragments of up to 2.2 kb in length contained no obvious genes by BLASTX similarity (Fig. 1).

Genomic DNA sequence was obtained containing the *PbCHAP* gene, a likely cysteine, histidine-dependent amino-peptidase (CHAP). The CHAP domain is predominantly found in bacterial genes, but has also been characterised in trypanosome genes (Bateman and Rawlings 2003; Rigden et al. 2003). A ribosomal L7A gene was found upstream of *PbCHAP* (Fig. 1). Phylogenetic comparisons of the *P. brassicae* L7A amino acid sequence with genes from a broad variety of organisms (Russell et al. 2005) showed *PbL7A* to group most closely with a sequence from the chlorarachniophyte *B. natans*, albeit with low bootstrap support throughout the tree (Supplementary Fig. 1).

Introns were Located in *P. brassicae* Genes

Genomic DNA that spanned the coding region of several genes was sequenced. Alignment of the genomic sequences with the corresponding cDNAs revealed the presence of 73 spliceosomal introns in 14 of the *P. brassicae* genes (see Fig. 1). The greatest number of introns (20) was found in the *PbPSA* gene.