Mycoscience (2012) 53:471–475
DOI 10.1007/s10267-012-0183-2

NOTE

The mtDNA *rps3* locus has been invaded by a group I intron in some species of *Grosmannia*

Shelly M. Rudski · Georg Hausner

Received: 30 November 2011 / Accepted: 18 January 2012 / Published online: 9 February 2012

Abstract The mitochondrial *rps3* gene in some filamentous ascomycetes fungi is encoded within an *rnl* group I intron. In *Grosmannia piceiperda* the N-terminal segment of the intron-encoded *rps3* gene has been invaded by an IC2-type group I intron. This intron disrupts the recipient *rps3* and fragments this gene into two open reading frames (ORFs). The IC2 group I intron encodes a putative double-motif LAGLIDADG ORF, which is fused in-frame to the upstream *rps3* exon sequence. The presence of the LAGLIDADG amino acid motif is indicative of an enzyme that has endonuclease and/or maturase activity and thus the intron encoded protein could be involved in promoting splicing and mobility. Reverse transcriptase polymerase chain reaction (RT-PCR) confirmed that this intron is spliced in vivo and as a result this could allow for the expression of a functional ribosomal Rps3 protein.

Keywords Blue-stain fungi · Homing endonuclease · Ribosomal protein · RT-PCR

The mitochondrial (mt) genomes of fungi contain numerous types of mobile genetic elements, such as group I and group II introns and homing endonuclease genes (HEGs) (Hausner 2012). Group I and group II introns are self-splicing elements and therefore they are spliced from transcripts, thus minimizing their impact to the genes that host them. Many group I introns contain open reading frames (ORFs) that either promote their mobility or assist in their splicing activity such as homing endonucleases (HEase) or maturases, respectively (Belfort et al. 2002). HEases are enzymes that can initiate intron mobility by introducing double-stranded breaks (DSBs) at specific target sites, which activates the cellular DSB-repair pathway (Belfort et al. 2002). Maturase proteins assist the intron with folding into a splicing competent ribozyme (Caprara and Waring 2005). There are two families of HEGs commonly encountered among fungal mtDNAs; named according to conserved amino acid motifs, the LAGLIDADG and GIY-YIG proteins (Stoddard 2006). HEGs have also been shown to be mobile elements that can move independently from their ribozyme counterparts (Mota and Collins 1988). Mobile introns and HEGs are of interest as they can generate mtDNA polymorphisms and they have applications in biotechnology (Takeuchi et al. 2011; Hausner 2012).

Previously we characterized the mt DNA *rps3* locus within species of *Ophiostoma* Syd. & P. Syd. and *Grosmannia* Goid. emend. Z.W. de Beer, Zipfel & M.J. Wingf. (Sethuraman et al. 2009a). Many species of *Ophiostoma* and *Grosmannia* are blue-stain fungi that can discolor timber and thus reduce its economic value, and so far very little information is available on their mt genomes and their intron complements. Rps3 is a ribosomal protein that is encoded by an ORF that is embedded within a group I A1 intron (mL2449) inserted within the U11 region of the large ribosomal subunit (*rnl*) gene (Burke and RajBhandary 1982; reviewed in Sethuraman et al. 2009a, b). Among species of *Ophiostoma* and *Grosmannia* we previously showed that the *rps3* locus can be invaded by three different HEGs (Sethuraman et al. 2009a).

Homing endonuclease genes were noted to insert within two different sites located within the C-terminal coding segment of the *rps3* gene in a manner where the HE coding sequences are fused to the upstream *rps3* coding region forming one complex ORF that encodes two different

S. M. Rudski · G. Hausner (✉)
Department of Microbiology, University of Manitoba, Winnipeg, MB R3T 2N2, Canada
e-mail: hausnerg@cc.umanitoba.ca

peptides (Sethuraman et al. 2009a). Most likely these HEases are initially expressed as fusion proteins in combination with their host Rps3 protein. Protease activity would resolve the fusion protein and release the endonuclease protein from the Rps3 protein. This arrangement would allow the HEGs to insert in a manner that minimizes toxicity to the host genome.

A third type of HEG was discovered in *Grosmannia piceiperda* (Rumbold) Goid. and *Grosmannia laricis* (K. van der Westhuizen, Yamaoka & M.J. Wingf.) Zipfel, Z.W. de Beer & M.J. Wingf., where the HEG is inserted into the N-terminal coding region of the *rps3* gene; here the insertion generates a bi-ORFic complex (two putative ORFs within one intron; Sellem and Belcour 1997), as the HEG brings along its own stop codon (Sethuraman et al. 2009a). The introduction of this HEG stop codon truncates the *rps3* coding region, rendering it a potential pseudogene. However, one could also speculate that a functional Rps3 protein can still be assembled by expressing the two ORFs separately using an alternative start codon, followed by interaction of the N- and C-terminal peptide segments, or an alternative splicing event could generate a continuous *rps3* transcript. The objective of the work is to re-examine the N-terminal HEG insertion discovered by Sethuraman et al. (2009a) and gain a better understanding of plausible models that could explain the expression of the *rps3* locus in *G. piceiperda*.

The *rnl*-U11 region of *G. piceiperda* [strains WIN(M) 975 (=UAMH 9784, University of Alberta Microfungus Collection and Herbarium) and 979)] was amplified with primers IP1 (GGAAAAGCTACGCT AGGG) and IP2 (CTTGCGCAAATTAGCC; Bell et al. 1996) under conditions previously described in Sethuraman et al. (2009a). Polymerase chain reaction (PCR) products were cloned into *Escherichia coli* (DH5α) using the TOPO TA Cloning® kit (Invitrogen Life Technologies, Burlington, ON, Canada).

DNA sequencing templates were prepared by purifying PCR products with the Wizard® SV Gel and PCR Clean-Up system (Promega, Madison, WI, USA). Recombinant plasmids from positive clones were purified with the Wizard® Plus Minipreps DNA purification system (Promega). The double-stranded DNA fragments or plasmids were sequenced using the BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, CA, USA) following the manufacturer's instructions. The sequencing products were denatured and resolved on a 3130 genetic analyzer (Applied Biosystems).

Sequences were obtained with primers IP1 and IP2 and additional primers were designed as needed. The sequences were compiled and assembled manually into contigs using the GeneDoc program v2.5.010 (Nicholas et al. 1997) and nucleotide sequence alignments were done with the Clustal-X program (Thompson et al. 1997). The ORF finder program (http://www.ncbi.nlm.nih.gov/gorf/gorf.html; using the Genetic code setting: 4 for mold mtDNAs) was used to search for potential ORFs within the *rnl*-U11 region. The online program RNAweasel (http://megasun.bch.umontreal.ca/RNAweasel/; Lang et al. 2007) was used to analyze the *rnl*-U11 sequence for the presence of introns. Based on the RNAweasel output and previously published structures (Michel and Westhof 1990; Cech et al. 1994) the intron secondary structure was generated with the mfold program (online version 3.2; http://www.bioinfo.rpi.edu/~zukerm/rna/; Zuker 2003).

RNA was isolated from strain *G. piceiperda* WIN(975) using the RNeasy kit for total RNA isolation (Qiagen Inc., Valencia, CA, USA) with some modifications. Initially the mycelium was ground in liquid nitrogen. However, once the cell walls were broken the RNA was extracted and purified following the yeast protocol of the RNeasy kit. The RNA was treated with the TURBO™ DNase kit (Ambion, Austin, TX, USA) following the manufacturer's recommendations, and 1 μg of RNA was used as template for reverse transcriptase-PCR (RT-PCR) using the ThermoScript RT-PCR system (Invitrogen) according to the manufacturer's recommendations. First-strand synthesis was carried out with primer A-rps3aR2 (AATGCTCTAT CTGCTCTATAACG) at a final concentration of 100 pM and subsequent PCR amplification was carried out with primers A-rps3aF2 (ATATTAAAATTAACTTTGTAG GTG) and A-rps3aR2, both at 100 pM concentrations.

The *rnl*-U11 region for a second strain of *G. piceiperda* WIN(M) 975 was sequenced and this strain yielded the same results as were previously obtained for WIN(M) 979 (Genbank accession: FJ717847; Sethuraman et al. 2009a). Essentially the *rnl*-U11 region contains the mL2449 intron, which is 2.914 kb in length, and when compared to the closely related species *Grosmannia aurea* (R.C. Rob.-Jeffr. & R.W. Davidson) Zipfel, Z.W. de Beer & M.J. Wingf. (strain CBS 438.69; Genbank accession: FJ717837), which has a 1.6 kb version of the mL2449 intron but without any HEG insertions, we could determine the HEG insertion point within the *rps3* gene of *G. piceiperda* (Fig. 1a). The insertion of this double-motif LAGLIDADG HEG within the *rps3* N-terminus generated two ORFs; the first ORF is 1.446 kb and encodes a putative 482-amino acid fusion protein consisting of the first 189 bp of the recipient *rps3* ORF (the N-terminal 63 amino acids) followed by 1.257 kb (419 amino acids) that corresponds to a double-motif LAGLIDADG homing endonuclease. The second ORF is 1.041 kb long, and the first 39 nucleotides from the second ORF were found to have no homology to any sequences in the NCBI data base, but thereafter the sequence was found to be identical to the corresponding *rps3* sequence of *G. aurea*. The two ORFs are separated by a 79 bp spacer region (Sethuraman et al. 2009a).

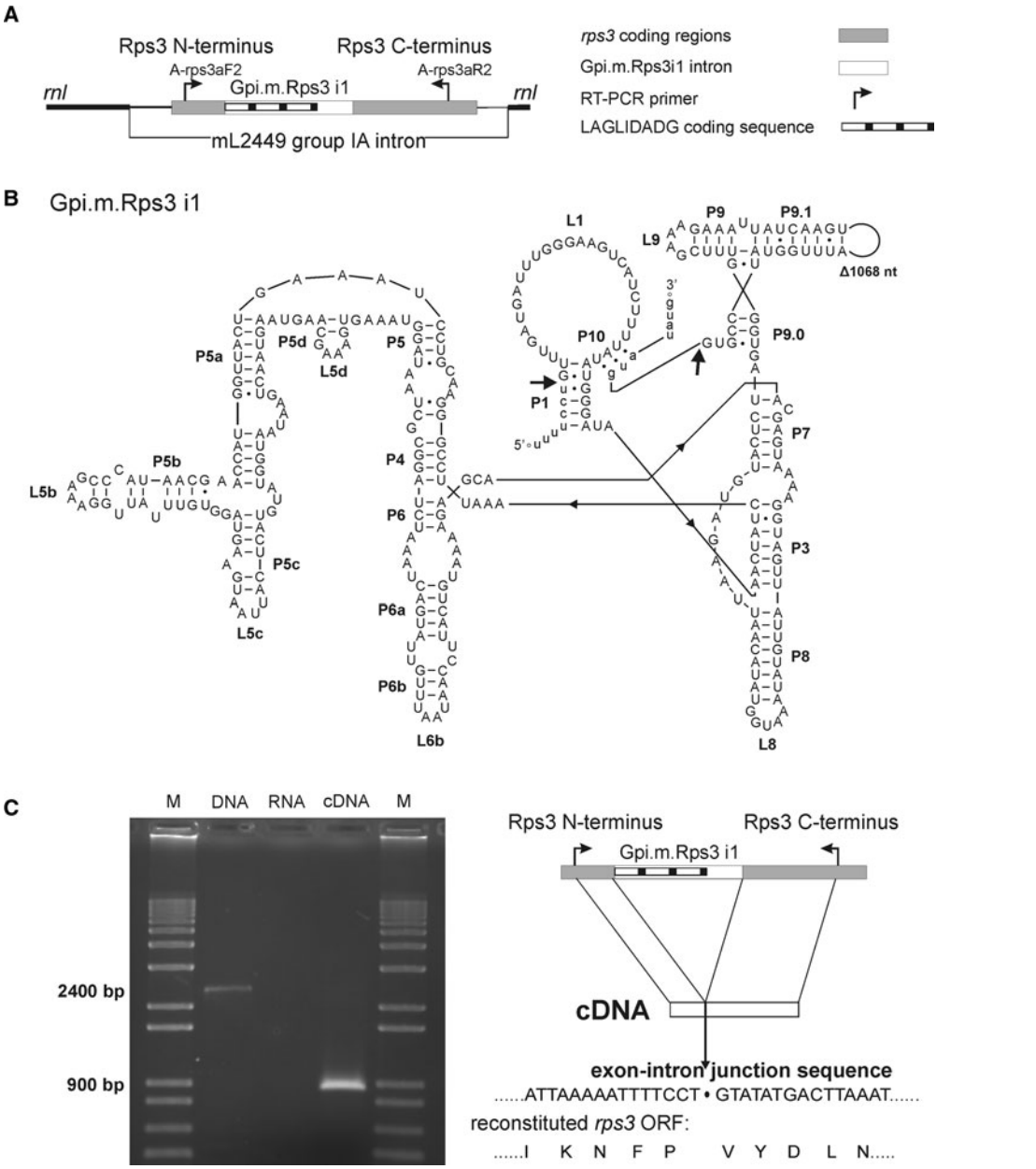

Fig. 1 **a** Schematic diagram of the *rps3* locus in *Grosmannia piceiperda*; the *rps3* open reading frame (*ORF*) is embedded within an *rnl* group I A1 intron at *rnl* position 2449 according to Johansen and Haugen's (2001) system of designating group I intron insertions. The *rps3* coding region appears to be interrupted by the insertion of a sequence element that encodes a double-motif LADLIDADG peptide, thus splitting the *rsp3* coding region into two distinct ORFs (the N- and C-termini "coding segments"). The relative positions for reverse transcriptase polymerase chain reaction (*RT-PCR*) primers A-rps3aF2 and A-rps3aR2 are indicated by *arrows*. **b** Secondary structure prediction for the *rps3* group I C2 intron sequence (Gpi.m.Rps3i1). Typically group I introns can be folded into 10 paired (helical) regions (*P*) along with sequence segments (*L*) that connect these helical regions (Cech et al. 1994). Exon sequences are in *lower case* and intron sequences are in *upper case*. *Arrows* indicate the 5′ and 3′ terminal ends of the group I intron. Labeling of the structural components is based on the computer program RNAweasel and conventions proposed by Michel and Westhof (1990) and Cech et al. (1994). **c** RT-PCR analysis for the *rps3* transcript to demonstrate in vivo splicing of the Gpi.m.Rps3i1 intron and to define the exon–intron junction. *Lanes marked M* contain the 1 kb plus DNA ladder (Invitrogen), *lane marked DNA* represents a standard PCR reaction using primers A-rps3aF2 and A-rps3aR2 using whole cell DNA as a template, while the *lane denoted cDNA* contains amplicons derived from the RT-generated cDNA template. The negative control is shown in the *lane marked RNA*, where a standard PCR reaction was carried out with primers A-rps3aF2 and A-rps3aR2 using whole cell RNA as a template. The latter yielded no bands, confirming that the RNA sample was DNA-free. The schematic *on the right* shows that, based on the identified exon–intron junction, the intron is located in such a way (phase 0) that its removal joins the N-terminal and C-terminal coding regions for the Rps3 protein; thus, the intron and its associated homing endonuclease (HE) sequence can minimize their impact on the expression of the *rps3* gene

It was originally assumed that HEGs that can move independently or without ribozyme components had invaded the *rps3* loci of *Ophiostoma* and *Grosmannia* species and this appears to be the case for the C-terminal inserted HEGs (Sethuraman et al. 2009a). However, a more detailed reexamination of the sequences representing the *rps3* loci with the different types of HEG insertions with the RNA-weasel program (Lang et al. 2007) revealed that the N-terminal inserted HEG sequence was associated with an IC2 group I intron sequence (Fig. 1b). This intron is named Gpi.m.Rps3i1 as it is a group I intron and the first one noted to be inserted within the mt *rps3* gene of *G. piceiperda*. This intron is very compact, with the ORF sequence overlapping with most of the intron core sequences (Fig. 1a, b). This configuration would explain how the N-terminal HEG inserted into the *rps3* coding region; i.e., by being part of a composite intron/HEG mobile element. Also a potentially self-splicing intron would not be toxic as it permits the expression of the Rps3 protein and the maturation of the *rnl* transcript. Reverse transcriptase PCR was employed to examine whether the putative intron was removed from the *rps3* transcript and to characterize the exon–intron junctions.

The *rps3* locus was amplified by PCR and the resulting amplicon was 2.4 kb; however, using the same primers as those used for the RT-PCR assay showed that the cDNA amplicon representing the same region was an estimated 900 bp (Fig. 1c). The data show that about 1500 bases are removed during the maturation of the *rps3* transcript. Sequence characterization of the cDNA and alignment with the original *rnl*-U11 sequences for *G. piceiperda*, and *G. larics* (Genbank accession: FJ717837) and the HEG minus version found in *G. aurea* showed that the exon/intron junction corresponds to the predicted IC2 intron insertion site (Fig. 1b, c). Originally the first 39 nucleotides of the putative second ORF were an enigma as they did not match any *rps3* sequences within the NCBI databases (Sethuraman et al. 2009a); however, the cDNA analysis shows that these 39 nucleotides are part of the intron as they are spliced out.

The RT-PCR reaction demonstrated that the putative IC2 intron is removed from the *rps3* transcript and that the intron–exon junctions allow for the original *rps3* N and C terminal coding components to be spliced together in frame forming a continuous ORF (Fig. 1c). The IC2 intron is in phase 0, which means the intron inserted in a manner that does not disrupt a codon in the "host" *rps3* reading frame.

This new finding explains the possible origin of the *rps3* N-terminal HEGs within certain species of *Grosmannia*, but it adds complexity to the mL2449 intron (Fig. 1a). First, we have an *rnl* group I intron encoding a potentially essential gene that is required for ribosome assembly (reviewed in Sethuraman et al. 2009a); second, a group I intron has invaded an intron encoded ORF. In order for the *rnl* RNA to mature the mL2449 intron has to be removed (as shown in Sethuraman et al. 2009a), this would generate the mRNA for the *rps3* ORF (see Fig. 2). However, the IC2 intron must be spliced from the *rps3* precursor transcript.

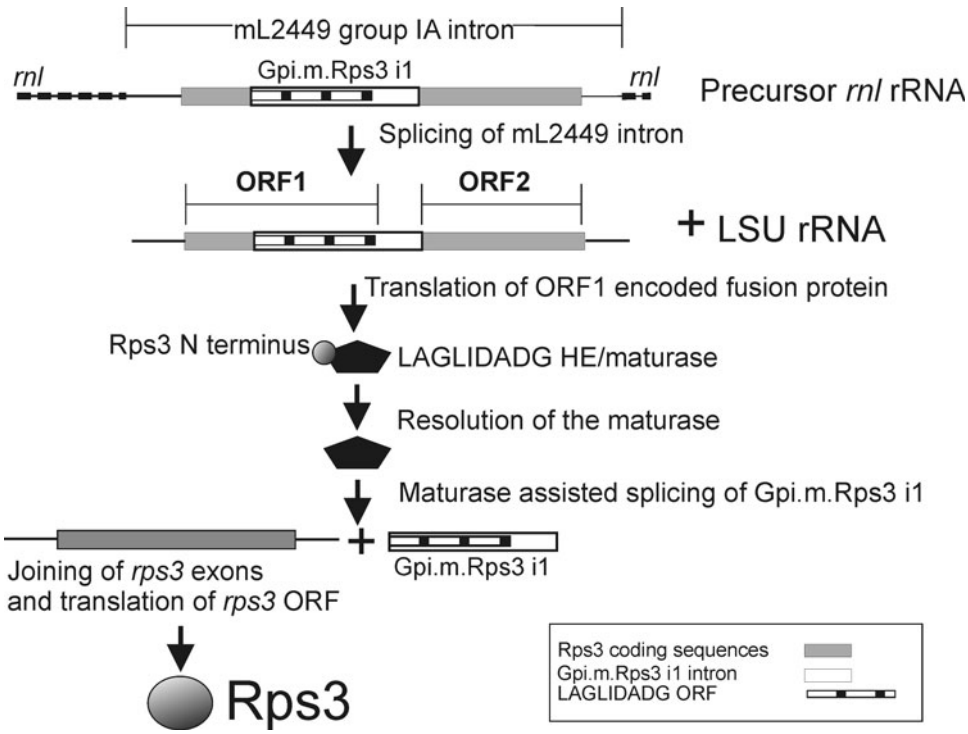

Fig. 2 A potential model showing the expression of the Rps3 protein from an intron encoded open reading frame (*ORF*) embedded within an *rnl* group IA intron (mL2449) where the *rps3* gene has been fragmented by the insertion of a group I intron (Gpi.m.Rps3i1) encoding a LAGLIDADG type ORF. See text for more detail. *LSU* large subunit

As the intron encoded protein (IEP) is fused, in frame, to the upstream *rps3* exon one must assume that the IEP is translated as a fusion protein, thus benefiting from the *rps3* ORF's start codon and regulatory regions that permit translation. The Rps3 N-terminus is either proteolytically removed from the IEP or it remains attached and does not alter the function of the IEP. The IEP could have maturase activity that is needed to facilitate the efficient removal of the IC2 intron from the *rps3* precursor transcript, thus generating the *rps3* mRNA that can be translated to yield a functional Rps3. This scenario of stepwise splicing has been evoked to explain the expression of mtDNA genes interrupted by several introns that encode putative proteins whose reading frames are fused to the upstream exon regions (Grivell 1995). The insertion of the IC2 intron within the N terminus of the *rps3* ORF probably requires a cascade of splicing events (Fig. 2) that allow for the expression of the IEP and for the Rps3 protein.

Acknowledgments This work is supported by a Discovery Grant from the Natural Sciences and Engineering Research Council of Canada (NSERC) to G.H.

References

Belfort M, Derbyshire V, Cousineau B, Lambowitz A (2002) Mobile introns: pathways and proteins. In: Craig N, Craigie R, Gellert M, Lambowitz A (eds) Mobile DNA II. ASM Press, Washington, DC, pp 761–783

Bell JA, Monteiro-Vitorello CB, Hausner G, Fulbright DW, Bertrand H (1996) Physical and genetic map of the mitochondrial genome of *Cryphonectria parasitica* Ep155. Curr Genet 30:34–43

Burke JM, RajBhandary UL (1982) Intron within the large rRNA gene of *N. crassa* mitochondria: a long open reading frame and a consensus sequence possibly important in splicing. Cell 31:509–520

Caprara MG, Waring RB (2005) Group I introns and their maturases: uninvited, but welcome guests. Nucleic Acids Mol Biol 16:103–119

Cech TR, Damberger SH, Gutell ER (1994) Representation of the secondary and tertiary structure of group I introns. Nat Struct Biol 1:273–280

Grivell LA (1995) Nucleo-mitochondrial interactions in mitochondrial gene expression. Crit Rev Biochem Mol Biol 30:121–164

Hausner G (2012) Introns, mobile elements and plasmids. In: Bullerwell CE (ed) Organelle genetics: evolution of organelle genomes and gene expression. Springer, Berlin, pp 329–358

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

Lang BF, Laforest MJ, Burger G (2007) Mitochondrial introns: a critical view. Trends Genet 23:119–125

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

Mota EM, Collins RA (1988) Independent evolution of structural and coding regions in a *Neurospora* mitochondrial intron. Nature 332:654–656

Nicholas KB, Nicholas HB Jr, Deerfield DW (1997) GeneDoc: analysis and visualization of genetic variation. EMB News 4:14

Sellem CH, Belcour L (1997) Intron open reading frames as mobile elements and evolution of a group I intron. Mol Biol Evol 14:518–526

Sethuraman J, Majer A, Friedrich N, Edgell D, Hausner G (2009a) Genes-within-genes: multiple LAGLIDADG homing endonucleases target the ribosomal protein S3 gene encoded within an *rnl* group I intron of *Ophiostoma* and related taxa. Mol Biol Evol 26:2299–2315

Sethuraman J, Majer A, Iranpour M, Hausner G (2009b) Molecular evolution of the mtDNA encoded *rps3* gene among filamentous ascomycetes fungi with an emphasis on the ophiostomatoid fungi. J Mol Evol 69:372–385

Stoddard BL (2006) Homing endonuclease structure and function. Q Rev Biophys 1:49–95

Takeuchi R, Lambert AR, Mak AN, Jacoby K, Dickson RJ, Gloor GB, Scharenberg AM, Edgell DR, Stoddard BL (2011) Tapping natural reservoirs of homing endonucleases for targeted gene modification. Proc Natl Acad Sci USA 108:13077–13082

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 25:4876–4882

Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res 31:3406–3415