Intraspecific comparison and annotation of two complete mitochondrial genome sequences from the plant pathogenic fungus *Mycosphaerella graminicola*

Stefano F.F. Torriani\(^a\), Stephen B. Goodwin\(^b\), Gert H.J. Kema\(^c\), Jasmyn L. Pangilinan\(^d\), Bruce A. McDonald\(^a,\)\(^*\)

\(^a\) Plant Pathology Group, Institute of Integrative Biology, ETH Zürich/LFW, Universitätstrasse 2, CH-8092 Zürich, Switzerland
\(^b\) USDA-ARS, Crop Production and Pest Control Research Unit, Department of Botany and Plant Pathology, 915 West State Street, Purdue University, West Lafayette, IN 47907-2054, USA
\(^c\) Plant Research International B.V., P.O. Box 16, 6700 AA Wageningen, The Netherlands
\(^d\) Department of Energy Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

Received 13 August 2007; accepted 10 December 2007
Available online 23 December 2007

**Abstract**

The mitochondrial genomes of two isolates of the wheat pathogen *Mycosphaerella graminicola* were sequenced completely and compared to identify polymorphic regions. This organism is of interest because it is phylogenetically distant from other fungi with sequenced mitochondrial genomes and it has shown discordant patterns of nuclear and mitochondrial diversity. The mitochondrial genome of *M. graminicola* is a circular molecule of approximately 43,960 bp containing the typical genes coding for 14 proteins related to oxidative phosphorylation, one RNA polymerase, two rRNA genes and a set of 27 tRNAs. The mitochondrial DNA of *M. graminicola* lacks the gene encoding the putative ribosomal protein (*rps5*-like), commonly found in fungal mitochondrial genomes. Most of the tRNA genes were clustered with a gene order conserved with many other ascomycetes. A sample of 35 additional strains representing the known global mt diversity was partially sequenced to measure overall mitochondrial variability within the species. Little variation was found, confirming previous RFLP-based findings of low mitochondrial diversity. The mitochondrial sequence of *M. graminicola* is the first reported from the family Mycosphaerellaceae or the order Capnodiales. The sequence also provides a tool to better understand the development of fungicide resistance and the conflicting pattern of high nuclear and low mitochondrial diversity in global populations of this fungus.

© 2007 Elsevier Inc. All rights reserved.

**Keywords:** Comparative genomics; Genome organization; Microsatellites; Mitochondrial genome (mtDNA); *Septoria tritici*

1. **Introduction**

*Mycosphaerella graminicola* (anamorph *Septoria tritici*) is the causal agent of Septoria tritici blotch of wheat and other poaceous hosts, and occurs worldwide across a wide range of climates (Eyal, 1999). The life cycle of *M. graminicola* includes both sexual and asexual stages. The sexual stage permits genetic recombination and produces airborne ascospores with the potential to be dispersed over several kilometers (Sanderson, 1972), whereas the asexual phase (*S. tritici*) produces pycnidiospores disseminated over limited distances from plant to plant via rain splash (Bannon and Cooke, 1998).

In most fungi studied to date there is concordance between genetic variation in the mitochondrial (mt) and nuclear genomes (Sommerhalder et al., 2007; Zhan et al., 2004), with some fungi having high levels of mt and nuclear diversity (Kudla et al., 2002; Liu et al., 1996) and others having low mt and nuclear genetic variability (Kurdyla
et al., 1995; Xia et al., 2000). However, the pattern of variability in *M. graminicola* is different; a comparison of RFLP markers in nuclear and mt genomes showed a pattern of high nuclear and low mt diversity in populations around the world (Zhan et al., 2003). Over 1300 nuclear RFLP genotypes were found among 1673 isolates, with an average of 18 alleles per nuclear RFLP locus. In contrast, only seven mtDNA haplotypes were found globally, with the two most common representing 93% of the world population. The high nuclear diversity is thought to be the consequence of high gene flow (Boeger et al., 1993), coupled with large effective population sizes (Zhan and McDonald, 2004) and recurring sexual reproduction (Chen and McDonald, 1996; Kema et al., 1996; Hunter et al., 1999). Zhan et al. (2004) suggested a selective sweep to explain the low diversity found in the mtDNA. Selective sweeps may be common in mt genomes because all of the genes are linked in one molecule so that selection on one gene can affect the frequency of all genes through hitchhiking.

Mt genomes have proven to be highly useful for research in evolutionary biology and systematics because of their uniparental inheritance, the near absence of genetic recombination, and uniform genetic backgrounds (Chen and Hebert, 1999). The evolution of mtDNAs has been characterized by extensive loss and translocation of genes to the nucleus (Adams et al., 2000) since their origin by endosymbiosis of a bacterial ancestor (John and Whatley, 1975). The result of this process is that most mt proteins are encoded by nuclear genes whose products are imported into the mitochondrion by translocase complexes, leaving relatively few mt proteins that are synthesized directly within the organelle (Brennicke et al., 1993; Hartl et al., 1989).

Mt genomes are characterized by high A + T content, lack of methylation, conservation in gene function, and high copy number (Campbell et al., 1999), and they can evolve at their own rate relative to the nuclear genomes of the organisms in which they occur (Ballard and Whitlock, 2004). The size and topology of the mt genome, the number and nature of the proteins it encodes, and even the genetic code itself can vary greatly among species (Gray et al., 1999). Fungal mtDNAs are generally an order of magnitude smaller than those of plants but larger than animal mtDNAs (Burger et al., 2003) and usually contain 14 genes encoding hydrophobic subunits of respiratory chain complexes, as well as genes for the large (*rnl*) and small (*rns*) ribosomal subunits and a set of tRNAs (Gray et al., 1999). The coding percent ranges between 40% and 60% in the Pezizomycotina. Among fungi, mt genomes vary widely in size, from approximately 18 to 109 kb (NCBI database). The variability of mt genome size among species is strongly influenced by differences in length and organization of intergenic regions, as well as by differences in intron content (from 0 to 30) and size (ranging from 0.15 to 4 kb). Burger et al. (2003) showed that there is no correlation between mtDNA size and gene content.

The taxonomic placement of *Mycosphaerella* within the class Dothideomycetes until recently was uncertain, and it usually was placed near *Dothidea* in the Dothideales (Kirk et al., 2001; Goodwin et al., 2004). However, recent analyses of a multigene phylogeny showed that *Mycosphaerella* belongs in the Capnodiales, a sister group to the Dothideales and Myriangiales (Schoch et al., 2006). Though the mt genome of *Stagonospora nodorum*, another member of the Dothideomycetes, was recently published (Hane et al., 2007), no mt genomes have been published from *Mycosphaerella* or any species in the Capnodiales, Dothideales, or Myriangiales.

The goals of this research were to obtain and annotate the first complete mitochondrial genome sequence from the *Mycosphaerella* branch of the fungal evolutionary tree, and to test a previous hypothesis of low mitochondrial diversity within global populations of *M. graminicola*. Complete sequences of the mtDNA genomes from two isolates of *M. graminicola* (one from North America and one from Europe) plus sequences at three mitochondrial loci for 35 additional isolates representing most of the known global diversity were compared, first to quantify the overall mtDNA sequence diversity in *M. graminicola* and, second, to compare it with earlier findings of low diversity based on RFLP analysis. An interspecific analysis of the tRNA genes flanking *rnl* of species in the Pezizomycotina revealed a consensus in tRNA gene content and order.

2. Materials and methods

2.1. Fungal strains, DNA extraction, and library construction

Strain IPO323 was isolated from a naturally infected leaf of the soft white wheat cultivar Arminda collected in Brabant, the Netherlands during 1981 (Kema and Van Silfhout, 1997). Fungal mycelia were produced on liquid shake cultures, harvested, stored and prepared for DNA extraction as described in Kema et al. (2002). Fungal spores and mycelia were ground with a Hybaid Ribolyser (model FP120HY-230) for 10 s at 2500 rpm with two tungsten carbide beads, and total genomic DNA was extracted using the Promega Wizard Magnetic DNA Purification System for Food as described by the manufacturer except with only 50 mg of lyophilised fungal material and 500 μl of lysis buffer. Plasmid libraries with insert sizes of 3 and 8 kb were created at the U.S. Department of Energy’s Joint Genome Institute (JGI) and sequenced to 4× genomic sequence coverage (~150,000 clones each).

Strain STBB1 was isolated from a wheat field 5 km southwest of College Station, Texas, USA, during 1989. The entire mt genome of this isolate was purified from total DNA by cesium chloride (CsCl) ultracentrifugation as described by Garber and Yoder (1983), with a CsCl density of 1.6 g/ml. A library was constructed by digesting the purified mtDNA to completion using the restriction enzyme HinfI, followed by gel electrophoresis and elution.
enzyme HindIII, ligating the fragments into the plasmid vector pUC18 and cloning in Escherichia coli strain DH5α.

2.2. DNA sequencing and assembly

Shotgun sequencing of the nuclear and mt genomes of isolate IPO323 was through the Community Sequencing Program of the JGI (www.jgi.doe.gov/CSP/) by analysis of libraries with insert sizes averaging 3, 8 and 40 kb. The mt genome was assembled from ~7680 sequencing reads from 10 plates of the 3-kb library using phrap (http://www.phrap.org/) with its standard parameters. This corresponds to roughly 5–6 Mb of sequence. Approximately 5.5% of the reads (~260 kb) represented mtDNA so the initial sequence was assembled at a depth of about 6×. The average depth of coverage for the entire project was 8.9× and was released publicly (http://genome.jgi-psf.org/Mycgr1/Mycgr1.home.html) during November 2006.

The mtDNA library obtained from isolate STBB1 was sequenced using the BigDye™ Terminator v3.0 Cycle Sequencing kit and the primer walking strategy. The sequencing reactions were in a total volume of 10 μl using 20–40 ng of plasmid DNA, 10 pmol of primers and 2 μl of BigDye reaction mix, previously diluted 1:4. The cycling profile was 10 s denaturation at 95°C, 5 s annealing at 50°C and 4 min extension at 60°C for 100 cycles. The sequencing reactions were purified through Sephadex G-50 DNA Grade F (Amersham Biosciences, Switzerland) before being loaded into an ABI 3100 automated sequencer (Applied Biosystems). The sequences were aligned and analyzed with the Sequencher version 4.2 software package (Gene Codes Corporation, Ann Arbor, MI) using the genetic code of Pezizomycotina that diverged from the standard nuclear code for the codon TGA, which was read as Trp and not as Stop.

Sequencing of the isolate STBB1 mt library generated approximately 75% of the entire mtDNA genome. Gaps in the STBB1 sequence were filled by aligning the sequenced HindIII fragments to the complete mtDNA sequence of isolate IPO323 and designing pairs of primers to amplify the missing regions in STBB1. The amplicons were sequenced as described above to obtain the entire mt genome of isolate STBB1.

2.3. Sequence annotation

The mtDNA sequence of M. graminicola was screened for similarity with those from other organisms in the NCBI database using the BlastN tool. Sequences showing matches with protein-coding genes of other organisms were subsequently compared using the BlastX tool (Altschul et al., 1990). The mt sequences of strains IPO323 and STBB1 were aligned using the Sequencher program and screened manually for polymorphisms including transitions, transversions, insertions and deletions (indels). The genes coding for ribosomal RNAs were determined by comparison with sequences from other fungi. The tRNAs were defined by tRNAscan-SE v.1.21 (Lowe and Eddy, 1997) and by comparison with the NCBI database. Expression of mt genes was tested by blast searches against databases of EST sequences (Goodwin et al., 2007; Kema et al., 2003; Soanes and Talbot, 2006). Repetitive elements, including minisatellites, simple-sequence repeats (SSRs) and mononucleotide repeats were identified using the online program Perfect Microsatellite Repeat Finder (http://sgdp.iop.kcl.ac.uk/nikammar/repeatfinder.html).

2.4. Intraspecific comparison and haplotype network

Three mtDNA loci, named Mgl, Mg2 and Mg3, were used to assess the overall mt diversity within the species. These loci were located in different regions of the mtDNA and were chosen because they displayed different degrees of polymorphism in the comparison between STBB1 and IPO323. Mg1 was located within orf1 and had no polymorphism. Mg2 included a portion of orf4 and had several polymorphisms including single-nucleotide polymorphisms (SNPs), indels and homopolymers of various lengths. Mg3 included the region with tRNA-Gly, tRNA-Asp, tRNA-Ser, and tRNA-Trp and had two polymorphic microsatellites and one homopolymer. Thirty-five isolates (Table 1) belonging to four RFLP haplotypes (Zhan et al., 2003) and originating from five continents were amplified using primers: MglF (5′-CCG GTC CCT CTA ATAT GTG G-3′) and MglR (5′-TAA GCC ATT ACT TCT CAG G-3′); Mg2F (5′-GGT TCC AAT GGG TTT AAT GCT A-3′) and Mg2R (5′- TGG GTG TAG CTA GAA ACC CTT C-3′); Mg3F (5′-AAG CTA CGC GTA TGG CTA ACA C-3′) and Mg3R (5′-AGG TAA GAC GCA CGC ATT TC-3′). Each PCR reaction contained 5–10 ng of DNA in a 20-μl reaction volume containing 10 pmol of each primer, 100 μM of each nucleotide, 2 μl of 10× PCR buffer (1× PCR buffer: 10 mM KCl, 10 mM (NH4)2SO4, 20 mM Tris–HCl, 2 mM MgSO4, 0.1% Triton X-100 [pH 8.8]) and 1 U of Taq DNA Polymerase (New England Biolabs). The PCR amplifications were carried out under the following conditions: initial denaturation at 96°C for 2 min, followed by 35 cycles of 96°C for 1 min, 56°C for 1 min, and 72°C for 1 min, with a final extension at 72°C for 5 min. The PCR products were sequenced using the primers MglF, Mg2F, and Mg3R, generating a total of 1339 bp (338 bp for Mgl, 333 bp for Mg2 and 668 bp for Mg3). Sequencing reactions were performed as described previously for STBB1. The program SNAP Workbench (Price and Carbone, 2005) was used to collapse the sequences into haplotypes and DnaSP (Rozas et al., 2003) was used to test for recombination within and among the tested loci. The software package TCS version 1.21 (Clement et al., 2000) was used to infer intraspecific evolution of the M. graminicola mtDNA. This program applies a statistical parsimony method to infer unrooted cladograms based on Templeton’s 95% parsimony connection limit (Templeton et al., 1992).
3. Results

3.1. Gene content and genome organization

The mt genome of *M. graminicola* is a circular molecule of approximately 43,960 bp containing 15 protein-coding genes, the large (*rnl*) and small (*rns*) ribosomal subunits, 27 tRNAs and eight putative open reading frames (*orf1–8*) of unknown function (Fig. 1 and Table 2). The protein-coding genes included three ATP synthase subunits (*atp6*, *atp8*, and *atp9*), the three cytochrome oxidase subunits I, II, and III (*cox1–3*), cytochrome *b* (*cytb*), seven nicotinamide adenine dinucleotide ubiquinone oxidoreductase subunits (*nad1–6*, *nad4L*) and a DNA-directed RNA polymerase (*RNA-Pol*). These genes were transcribed in two contiguous segments of opposite direction (Fig. 1).

A putative ribosomal protein (rps5-like) commonly found within *rnl* of ascomycetes was missing (Fig. 1, Table 2). To test whether this gene could have been transferred to the nuclear genome, *blastp* and *tblastn* searches were performed on the 8.9/C2 draft genomic sequence of *M. graminicola*. The *blastp* searches identified no matching proteins among the list of annotated genes. However, the *tblastn* searches identified matches at better than e/C05 on scaffold 5 to rps5-like proteins from *Phaeosphaeria nodorum* (e/C09) and *Penicillium marneffei* (e/C07), but not to those from the Sordariomycetes *Hypocrea jecorina* or *Verticillium dahliae*. Therefore, this gene most likely occurs in the nuclear rather than the mitochondrial genome of *M. graminicola*.

The eight putative *orfs* of unknown function are predicted to produce proteins containing from 126 to 481 amino acids. The 30 terminus of *orf4* overlapped with *orf3* for 52 nucleotides.

### Table 1

*Mycosphaerella graminicola* isolates included in the analysis of mtDNA variation

<table>
<thead>
<tr>
<th>Isolate</th>
<th>Haplotype</th>
<th>Host</th>
<th>Year</th>
<th>Location</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>AU49</td>
<td>1</td>
<td>Bread wheat</td>
<td>1993</td>
<td>Australia</td>
<td>B. Ballantyne</td>
</tr>
<tr>
<td>AU54</td>
<td>1</td>
<td>Bread wheat</td>
<td>1993</td>
<td>Australia</td>
<td>B. Ballantyne</td>
</tr>
<tr>
<td>AU58</td>
<td>1</td>
<td>Bread wheat</td>
<td>1993</td>
<td>Australia</td>
<td>B. Ballantyne</td>
</tr>
<tr>
<td>CH9B12A</td>
<td>1</td>
<td>Bread wheat</td>
<td>1993</td>
<td>Switzerland</td>
<td>B.A. McDonald</td>
</tr>
<tr>
<td>AU59</td>
<td>1</td>
<td>Bread wheat</td>
<td>1993</td>
<td>Australia</td>
<td>B. Ballantyne</td>
</tr>
<tr>
<td>OR402</td>
<td>1</td>
<td>Bread wheat</td>
<td>1993</td>
<td>Oregon</td>
<td>J. Boeger, B.A. McDonald, M. Schmitt</td>
</tr>
<tr>
<td>CH9B12C</td>
<td>1</td>
<td>Bread wheat</td>
<td>1993</td>
<td>Switzerland</td>
<td>B.A. McDonald</td>
</tr>
<tr>
<td>IN11</td>
<td>1</td>
<td>Bread wheat</td>
<td>1993</td>
<td>Indiana</td>
<td>G. Shaner</td>
</tr>
<tr>
<td>IN12</td>
<td>1</td>
<td>Bread wheat</td>
<td>1993</td>
<td>Indiana</td>
<td>G. Shaner</td>
</tr>
<tr>
<td>IN13</td>
<td>1</td>
<td>Bread wheat</td>
<td>1993</td>
<td>Indiana</td>
<td>G. Shaner</td>
</tr>
<tr>
<td>OR389</td>
<td>1</td>
<td>Bread wheat</td>
<td>1993</td>
<td>Oregon</td>
<td>J. Boeger, B.A. McDonald, M. Schmitt</td>
</tr>
<tr>
<td>OR409</td>
<td>1</td>
<td>Bread wheat</td>
<td>1993</td>
<td>Switzerland</td>
<td>B.A. McDonald</td>
</tr>
<tr>
<td>CH9C1A</td>
<td>1</td>
<td>Bread wheat</td>
<td>1993</td>
<td>Indiana</td>
<td>G. Shaner</td>
</tr>
<tr>
<td>IN9</td>
<td>1</td>
<td>Bread wheat</td>
<td>1993</td>
<td>Indiana</td>
<td>G. Shaner</td>
</tr>
<tr>
<td>CH9B5A</td>
<td>2</td>
<td>Bread wheat</td>
<td>1993</td>
<td>Switzerland</td>
<td>B.A. McDonald</td>
</tr>
<tr>
<td>CH9B9A</td>
<td>2</td>
<td>Bread wheat</td>
<td>1993</td>
<td>Switzerland</td>
<td>B.A. McDonald</td>
</tr>
<tr>
<td>GEA2a.2</td>
<td>2</td>
<td>Bread wheat</td>
<td>1992</td>
<td>Germany</td>
<td>R. Huang, G Koch</td>
</tr>
<tr>
<td>IPO323</td>
<td>2</td>
<td>Bread wheat</td>
<td>1981</td>
<td>Netherlands</td>
<td>G.H.J. Kema</td>
</tr>
<tr>
<td>CH9B7B</td>
<td>2</td>
<td>Bread wheat</td>
<td>1999</td>
<td>Switzerland</td>
<td>B.A. McDonald</td>
</tr>
<tr>
<td>GEE2b.2</td>
<td>2</td>
<td>Bread wheat</td>
<td>1992</td>
<td>Germany</td>
<td>R. Huang, G Koch</td>
</tr>
<tr>
<td>GEE3a.2</td>
<td>2</td>
<td>Bread wheat</td>
<td>1992</td>
<td>Germany</td>
<td>R. Huang, G Koch</td>
</tr>
<tr>
<td>U15</td>
<td>2</td>
<td>Bread wheat</td>
<td>1981</td>
<td>Netherlands</td>
<td>G.H.J. Kema</td>
</tr>
<tr>
<td>GEE1a.2</td>
<td>2</td>
<td>Bread wheat</td>
<td>1992</td>
<td>Germany</td>
<td>R. Huang, G Koch</td>
</tr>
<tr>
<td>OR428</td>
<td>2</td>
<td>Bread wheat</td>
<td>1990</td>
<td>Oregon</td>
<td>J. Boeger, B.A. McDonald, M. Schmitt</td>
</tr>
<tr>
<td>STIBI1</td>
<td>2</td>
<td>Bread wheat</td>
<td>1989</td>
<td>Texas</td>
<td>B.A. McDonald</td>
</tr>
<tr>
<td>AU57</td>
<td>3</td>
<td>Bread wheat</td>
<td>1993</td>
<td>Australia</td>
<td>B. Ballantyne</td>
</tr>
<tr>
<td>AU70</td>
<td>3</td>
<td>Bread wheat</td>
<td>1993</td>
<td>Australia</td>
<td>B. Ballantyne</td>
</tr>
<tr>
<td>AU72</td>
<td>3</td>
<td>Bread wheat</td>
<td>1993</td>
<td>Australia</td>
<td>B. Ballantyne</td>
</tr>
<tr>
<td>MX156</td>
<td>3</td>
<td>Bread wheat</td>
<td>1993</td>
<td>Mexico</td>
<td>L. Gilchrist</td>
</tr>
<tr>
<td>MX160</td>
<td>3</td>
<td>Bread wheat</td>
<td>1993</td>
<td>Mexico</td>
<td>L. Gilchrist</td>
</tr>
<tr>
<td>MX163</td>
<td>3</td>
<td>Bread wheat</td>
<td>1993</td>
<td>Mexico</td>
<td>L. Gilchrist</td>
</tr>
<tr>
<td>MX167</td>
<td>3</td>
<td>Bread wheat</td>
<td>1993</td>
<td>Mexico</td>
<td>L. Gilchrist</td>
</tr>
<tr>
<td>MX169</td>
<td>3</td>
<td>Bread wheat</td>
<td>1993</td>
<td>Mexico</td>
<td>L. Gilchrist</td>
</tr>
<tr>
<td>U2</td>
<td>4</td>
<td>Durum</td>
<td>1991</td>
<td>Syria</td>
<td>G.H.J. Kema</td>
</tr>
<tr>
<td>U7</td>
<td>4</td>
<td>Durum</td>
<td>1991</td>
<td>Tunisia</td>
<td>G.H.J. Kema</td>
</tr>
<tr>
<td>U6</td>
<td>4</td>
<td>Durum</td>
<td>1991</td>
<td>Tunisia</td>
<td>G.H.J. Kema</td>
</tr>
</tbody>
</table>

a RFLP haplotypes are groups of isolates having identical RFLP patterns following Zhan et al. (2003).
b Sequence haplotypes are groups of isolates having identical concatenated sequences for mitochondrial loci *Mg1*, *Mg2*, and *Mg3*.
c SNP haplotypes are groups of isolates having the identical concatenated sequence for *Mg1*, *Mg2*, and *Mg3* after removing all indels.
proteins remains to be determined, but the TMHMM2 method (Krogh et al., 2001) predicted orf2 to encode a non-membrane protein, whereas the other orfs were predicted to encode proteins having from one (orf5 and orf8) to ten (orf7) transmembrane domains. Expressed sequence tag (EST) databases (Goodwin et al., 2007; Kema et al.,

![Circular map of the mitochondrial genome of Mycosphaerella graminicola.](image)

**Table 2**

A comparison of the principal features of some completely sequenced fungal mt genomes

<table>
<thead>
<tr>
<th>Species</th>
<th>Size (kb)</th>
<th>A + T content (%)</th>
<th>Coding genes&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Orfs</th>
<th>Percent coding&lt;sup&gt;c&lt;/sup&gt; (%)</th>
<th>RNAs&lt;sup&gt;d&lt;/sup&gt;</th>
<th>Accession Nos.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aspergillus niger</td>
<td>31.1</td>
<td>74</td>
<td>14</td>
<td>2</td>
<td>47&lt;sup&gt;*&lt;/sup&gt;</td>
<td>27</td>
<td>DQ207726</td>
</tr>
<tr>
<td>Aspergillus tubingensis</td>
<td>33.6</td>
<td>74</td>
<td>14</td>
<td>2</td>
<td>43&lt;sup&gt;*&lt;/sup&gt;</td>
<td>27</td>
<td>DQ217399</td>
</tr>
<tr>
<td>Penicillium marneffei</td>
<td>35.5</td>
<td>76</td>
<td>15</td>
<td>5</td>
<td>67&lt;sup&gt;*&lt;/sup&gt;</td>
<td>27</td>
<td>AY347307</td>
</tr>
<tr>
<td>Epidermophyton floccosum</td>
<td>30.9</td>
<td>77</td>
<td>15</td>
<td>5</td>
<td>67&lt;sup&gt;*&lt;/sup&gt;</td>
<td>27</td>
<td>AY916130</td>
</tr>
<tr>
<td>Mycosphaerella graminicola</td>
<td>43.9</td>
<td>68</td>
<td>15</td>
<td>8</td>
<td>52&lt;sup&gt;*&lt;/sup&gt;</td>
<td>27</td>
<td>EU090238</td>
</tr>
<tr>
<td>Lecanicillium muscarium</td>
<td>24.5</td>
<td>73</td>
<td>15</td>
<td>0</td>
<td>58</td>
<td>27</td>
<td>AF087277</td>
</tr>
<tr>
<td>Verticillium dahliae</td>
<td>27.2</td>
<td>73</td>
<td>15</td>
<td>0</td>
<td>53</td>
<td>27</td>
<td>DQ351941</td>
</tr>
<tr>
<td>Fusarium oxysporum</td>
<td>34.5</td>
<td>69</td>
<td>15</td>
<td>1</td>
<td>44</td>
<td>27</td>
<td>AY945289</td>
</tr>
<tr>
<td>Metarhizium anisopliae</td>
<td>24.7</td>
<td>72</td>
<td>15</td>
<td>0</td>
<td>59</td>
<td>26</td>
<td>AY884128</td>
</tr>
</tbody>
</table>

<sup>a</sup> All fungi in this list have mt genomes with circular topologies.
<sup>b</sup> If present the fifteenth gene is rps5 except for M. graminicola that has a RNA-pol.
<sup>c</sup> Asterisks mark the genomes in which orfs were considered as coding genes in the calculation of percent coding.
<sup>d</sup> All fungi in this list have two genes encoding for the large and small ribosomal subunit.
2003; Soanes and Talbot, 2006) provided evidence for the transcription of orf5, orf6, and orf8.

Putative protein-coding genes covered 51.8% of the genome (including 15.9% composed of putative orfs), while 4.5% and 11.5% corresponded to tRNA genes and both rnl and rns, respectively. These values were similar to those reported for other ascomycetes (Table 2). Overall, the M. graminicola mtDNA was 34.5% A, 33.5% T, 16.3% G, and 15.7% C. MtDNA AT-content was 68% with coding and non-coding parts of the genome having, on average, the same AT-percentage.

3.2. Codon usage and tRNA genes

As expected given the transmembrane location of most mt proteins, the three most frequent codons were TTA (377 counts), ATA (371 counts) and TTT (270 counts) encoding Leu, Ile, and Phe, respectively. These amino acids have hydrophobic side chains commonly found in transmembrane helices. These three codons accounted for 19.3% of all codons in the mt genome. One codon was not used at all (CGA, Arg) and eight codons (CGC, TGG, TGC, CGG, CTC, GCC, GTC, CCG) were under-represented, being used from one to ten times each. All 15 protein-coding genes started with the canonical translation initiation codon ATG. The preferred stop codon was TAG. Codon usage of the orfs was similar to that of the protein-coding loci.

The 27 tRNAs encoded by the mt genome of M. graminicola could carry all 20 amino acids (Fig. 1). Two tRNA isoacceptors were identified for serine and leucine, three for arginine and four for methionine. Among the 27 tRNAs, only tRNA-Val occurred singly. The remaining 26 tRNA genes were grouped into five clusters, composed of 12, 5, 4, 3, and 2 tRNA genes (Fig. 1). As in other filamentous fungi, several tRNA genes flanked the rnl gene (Table 3, Tambor et al., 2006). In M. graminicola, these tRNA genes had an order similar to that of Eurotiomycetes and generally followed a conserved pattern found in other fungi (Table 3, Ghikas et al., 2006). Surprisingly, M. graminicola did not possess the TEM-tRNA genes at the beginning of the 3′ tRNA gene consensus, in contrast to both Eurotiomycetes and Sordariomycetes, suggesting an independent rearrangement in this species. The secondary structures of tRNA-Phe and tRNA-Thr diverged from the expected cloverleaf form as they contained nine instead of the canonical seven nucleotides in the anticodon loop.

3.3. Repetitive elements and comparative genomics

One 27-mer minisatellite repeated three times (located between the nad4 and nad4L genes), 186 SSRs and 51 mononucleotide repeats larger than seven nucleotides (mainly located in non-coding regions), were found in the mt genome of M. graminicola.

The total nucleotide diversity between IPO323 and STBB1 was 0.16%. The two M. graminicola isolates differed by only 23 base substitutions, including fourteen transitions and nine transitions. These changes represented 0.05% of the entire mt genome. Twenty-two additional mutations were found between IPO323 and STBB1: 18 were mononucleotide repeats of different lengths (9 poly-A and 9 poly-T), two were tetra- (AAAT) or penta-nucleotide microsatellite repeats (ATTTA), one was a frameshift mutation and the last was a 17-base deletion (Fig. 2). The nucleotide diversity among the global sample of 35 isolates was 35% greater than that between only IPO323 and STBB1 for the same three mitochondrial loci (Mg1, Mg2, and Mg3). Mg2 was the most variable locus, having 3 SNPs, a 17-bp indel, a polymorphic microsatellite with 2 alleles, and a mononucleotide repeat with 3 alleles. Mg1 had the fewest mutations, with two SNPs that were exclusive to isolates collected from durum wheat (Triticum turgidum). Mg3 had a mononucleotide repeat with 2 alleles and 2 microsatellites, respectively, with 2 and 4 alleles. All microsatellite alleles were due to differences in the number of repeats. The concatenated sequences of Mg1, Mg2, and
Mg3 from all 37 isolates identified 14 haplotypes. If all mutations other than SNPs were excluded from the analysis, only three haplotypes were found (Table 1). If the increase of 35% in nucleotide diversity detected for the Mg loci is extrapolated to the total genome, it results in a value of 0.22% for mitochondrial nucleotide diversity in a global sample of 37 isolates representing most of the known mt variants.

A haplotype network was inferred from all three Mg loci using the concatenated alignments (Fig. 3). The haplotype network did not show a clear pattern of geographical association, although isolates from North America were in the top half and all of those at the bottom were from Europe. Some frequent haplotypes such as H5, H6, and H7 included isolates of mixed origin, while others (H1, H11, and H13) were geographically limited. The *M. graminicola* haplotypes originating from durum wheat (*T. turgidum* ssp. *durum*) were distinguished from those originating from bread wheat (*T. aestivum*) by three SNPs. The two sequenced haplotypes (H5 and H9) represented different parts of the network. No evidence for recombination was found in the mtDNA of *M. graminicola* using the DnaSP program.

4. Discussion

The mt genomes of two strains of the plant pathogenic fungus *M. graminicola* originating from different continents (Europe and North America) were sequenced completely, annotated and compared to identify polymorphisms. Both isolates had mt genomes belonging to RFLP haplotype 2 (Zhan et al., 2003; Table 1). The mtDNA of *M. graminicola* was circular and A + T biased like those of most other fungi (Table 2).

These two *M. graminicola* sequences represent the first complete mt genomes of any species in the genus *Mycosphaerella* or from the branch of the fungal evolutionary tree that includes the Capnodiales, Dothideales, or Myriangiales (Schoch et al. 2006). *Mycosphaerella* and its related asexual genera (e.g., *Cercospora*, *Septoria*) comprise one of the largest and most economically important groups of pathogenic fungi (Goodwin et al., 2001) with several thou-
sand species infecting virtually every major family of plants (Corlett, 1991). Species of Mycosphaerella are not closely related to model fungi or those with completely sequenced mt genomes, so represent a previously unsampled branch of the fungal evolutionary tree.

The mtDNA of M. graminicola contains genes for 14 inner mt membrane proteins involved in electron transport and coupled oxidative phosphorylation, as well as rnl, rns, and RNA-Pol genes (Fig. 1). Except for presence of the RNA-Pol gene and absence of a gene encoding a putative ribosomal protein (rps5-like), this is the standard set of mtDNA-encoded genes found in other fungi. The rps5-like gene is found commonly in mt genomes of different fungal species and it was postulated that mtDNA-encoded rps5 was present in the common ancestor of fungal and animal mtDNAs ( Bullerwell et al., 2000 ). As M. graminicola is one of the few ascomycetes known to be lacking rps5, the absence of this gene could indicate an independent loss in this species. A possible homolog of this gene was identified on scaffold 5 of the 8.9 × draft genomic sequence of M. graminicola, so it may have been transferred to the nuclear genome rather than having been lost.

Genes in the M. graminicola mtDNA had no introns, a finding that contrasts with other fungal mtDNAs that possess large introns containing intron-encoded proteins, as found in Podospora anserina ( Cummings et al., 1990 ) and Penicillium marneffei ( Woo et al., 2003 ). Eight orfs, with no obvious homology to any other sequenced genes present in the GenBank database, were found in the mt genome of M. graminicola. The functions of these putative genes remain unclear, although some of them may represent highly diverged versions of known mtDNA-encoded genes, no longer recognizable by identity searches ( Gray et al., 1998 ). EST databases provided evidence for transcription of orf5, orf6, and orf8, indicating that they may be expressed. Interestingly, these three orfs were the only ones of the eight that were located adjacent to tRNA genes, so possibly they may be transcribed along with the tRNAs but not translated.

All tRNA secondary structures had the expected clover-leaf form, but particularly interesting were tRNA-Thr ( UGU as anticodon) and tRNA-Phe ( GAA as anticodon) because they had nine nucleotides in the anticodon loop instead of the canonical seven. This rare tRNA structure was described previously in Metarhizium anisopliae for tRNA-Thr and tRNA-Glu ( Ghikas et al., 2006 ), and in Verticillium dahliae for tRNA-Thr, tRNA-Glu, tRNA-Arg, and tRNA-Ser ( Pantou et al., 2006 ).

Nucleotide genomes, including that of M. graminicola ( Goodwin et al., 2007 ), possess SSRs that are known to be highly variable in terms of motif repeat number and distribution ( Katti et al., 2001 ; Toth et al., 2000 ). This study presents a similar picture for the mt genome of M. graminicola. SSRs and mononucleotide repeats may play a significant role in the regulation and evolution of the entire molecule. In nuclear genomes it was demonstrated that these highly variable tracts, if placed in promoter regions, could influence transcriptional activity ( Kashi et al., 1997 ) and could play an important role in creating and maintaining quantitative genetic variation ( Kashi et al., 1997 ; Tautz et al., 1986 ). In the mtDNA of M. graminicola, mononucleotide repeats became less common in coding regions as their length increased. Because most long mononucleotide repeats are located 5'-upstream of ATG start codons ( Fig. 2 ), we hypothesize that they might play a role in regulating transcription. These tracts could be protein binding signals and, more precisely, upstream promoter elements, as demonstrated previously in nuclear genomes ( Kashi et al., 1997 ).

The intraspecific mt diversity was first assessed by comparing the total genome sequences of two isolates ( STBB1 and IPO323 ), giving a nucleotide diversity of 0.16%. In order to assess species-wide variation, another 35 isolates were chosen, originating from five continents and belonging to four of the seven known RFLP haplotypes ( Table 1 ). Using these additional isolates, the total mtDNA variation in M. graminicola was estimated to range from 0.16% to 0.22%, falling within the lower range of published intraspecific nucleotide diversities. The nucleotide diversity would decrease to 0.12% if the 17-bp indel was excluded. This 17-bp indel appears to be a recent mutation that occurred during the 1970s ( Torriani SFF, unpublished), suggesting that the M. graminicola mtDNA may be increasing in diversity following the hypothesized selective sweep ( Zhan et al., 2003 ). Other examples of low intraspecific mtDNA nucleotide diversity based on complete mtDNA sequences were 0.2% for the olive fly Bactrocera oleae ( Nardi et al., 2003 ) and 0.36% for Drosophila simulans ( Ballard, 2000 ).

These results support earlier findings of low mt diversity in M. graminicola obtained by RFLP analysis ( Zhan et al., 2003 ). While the haplotypic diversity based on sequences was higher than that found using RFLPs, the total nucleotide diversity remains the lowest reported to date in fungi. The greater number of haplotypes found through sequencing reflects the higher resolution of this method, especially the ability to resolve small indels that are missed by RFLP analysis. In fact, if indels were removed from the sequence analysis and only SNPs were considered, only three mt haplotypes were found, but they did not always correspond with the RFLP data ( Table 1 ). For example, isolates with RFLP haplotypes 1 and 2 were the most polymorphic and could have SNP haplotypes 1 or 3. Isolates with RFLP haplotype 3 always had SNP haplotype 1. It was interesting that isolates of M. graminicola adapted to durum wheat had unique RFLP and SNP haplotypes 4 and 2, respectively ( Table 1 ). The nonrandom association between mitochondrial RFLP haplotypes and host species, presumably caused by natural selection operating on the mt genome, was noted previously in M. graminicola ( Zhan et al., 2004 ) and other fungi ( Demanche et al., 2001 ; Gomes et al., 2000 ). The intraspecific haplotype network ( Fig. 3 ) that included all mutational events also distinguished between haplotypes originating from bread wheat and durum wheat.
The contrasting genetic diversity among mt and nuclear genomes in *M. graminicola* (Zhan et al., 2003, 2004) raises intriguing questions about the mechanisms leading to this phenomenon. At least two hypotheses can be proposed to account for the observed low levels of mt variation, including a lower mutation rate in the mt genome or a selective sweep. A comparison among the three yeast species *Saccharomyces cerevisiae*, *Klyveromyces lactis*, and *Candida glabrata* showed that the frequency of nucleotide changes is higher in nuclear than in mt genomes (Clark-Walker, 1991), which is the opposite of mammals where nuclear genes evolve slower than mt genes (Saccone et al., 2000).

On the other hand, the low level of polymorphism in the *M. graminicola* mtDNAs may have been generated through fixation of an advantageous mt mutation during a selective sweep. The selection of a favored mt haplotype leading to fixation of an advantageous mt mutation during a selective sweep. The selection of a favored mt haplotype leading to the evolution of resistance to strobilurin fungicides in *M. graminicola* (Torriani and McDonald, unpublished).

Acknowledgments

We gratefully acknowledge Marcello Zala for technical support and Charles Crane for bioinformatics analyses. DNA sequencing of isolate IPO323 was performed at the Department of Energy’s Joint Genome Institute through the Community Sequencing Program (www.jgi.doe.gov/ csp/). All sequencing data are public and may be accessed through http://www.jgi.doe.gov/Mgraminicola. This project was supported by the Swiss National Science Foundation (Grant 3100A0-104145) and by USDA CRIS Project 3602-22000-013-00D.

References


Genome structure and gene content in protist mitochondrial DNAs. Nucleic Acids Res. 26, 865–887.


