Duplication of a Truncated Paralog of the Family B DNA Polymerase Gene *Aa-polB* in the Agrocybe aegerita Mitochondrial Genome

GERARD BARROSO, FREDERIC BOIS, AND JACQUES LABARE`RE*  
Laboratoire de Génétique Moléculaire et d’Amélioration des Champignons Cultivés, Université Victor Segalen Bordeaux 2—INRA, I.B.V.M., CRA de Bordeaux, 33883 Villenave d’Ornon Cédex, France

Received 24 July 2000/Accepted 29 January 2001

The *Agrocybe aegerita* mitochondrial genome contains a truncated family B DNA polymerase gene (*Aa-polB P1*) whose nucleotide sequence is 86% identical to the previously described and potentially functional *Aa-polB* gene. A tRNA\(^{\text{Met}}\) gene occurs at the 3’ end of the *Aa-polB P1* gene. The *Aa-polB P1* gene could result from reverse transcription of an *Aa-polB* mRNA primed by a tRNA\(^{\text{Met}}\) followed by the integration of the cDNA after recombination at the mitochondrial tRNA locus. Two naturally occurring alleles of *Aa-polB P1* carry one or two copies of the disrupted sequence. In strains with two copies of *Aa-polB P1*, these copies are inverted relative to one another and separated by a short sequence carrying the tRNA\(^{\text{Met}}\) gene. Both *A. aegerita* mitochondrial family B DNA polymerases were found to be related to other family B DNA polymerases (36 to 53% amino acid similarity), including the three enzymes of the archaebacterium *Sulfolobus solfataricus*. If mitochondria originated from a fusion between a *Clostridium*-like eubacterium and a *Sulfolobus*-like archaebacterium, then the *A. aegerita* family B DNA polymerase genes could be remnants of the archaebacterial genes.

The genes in the fungal mitochondrial genome generally belong to a small set of highly conserved genes that probably originated from a prokaryotic ancestral endosymbiont (for a review, see reference 12). They encode components of complexes I to V of the electron transport chain, as well as the RNA portion of the translation system, mitochondrial rRNAs, and mitochondrial tRNAs. Most other genes encoding mitochondrial proteins are located on the nuclear chromosomes. In particular, the highly conserved \(\gamma\) DNA polymerases (17) responsible for the replication of the mitochondrial DNA (mtDNA) are nucleus encoded.

*Agrocybe aegerita* is a cultivated basidiomycete whose mitochondrial genome has been cloned and mapped (21) and from which three mitochondrial genes have been sequenced: *cox1* (10) and the small-subunit (SSU) and large-subunit rRNA genes (rDNAs) (9, 11). A potentially functional family B DNA polymerase gene named *Aa-polB* (17) re-arranges and were previously described (3). This gene putatively encodes a 571-amino-acid (aa) protein possessing all the conserved domains and activities (4). Other sequences similar to *Aa-polB* in the mitochondrial genome about 20 kb from the *cox1* gene. The *Aa-polB* region, named H4, has two alleles, H4-1 and H4-2, that differ in length and that are present in 28 and 8 strains, respectively, of 36 *A. aegerita* field strains (3).

Our objectives in this study were to determine the relationship between *Aa-polB* and a putative copy of this gene some 20 kb distant and to determine if the distal copy of this sequence affects allelic variability. The duplicated copies of this family B DNA polymerase gene may have arisen by reverse transcription of mRNAs primed by a mitochondrial tRNA.

MATERIALS AND METHODS

Strains, media, and culture conditions. We sequenced the *HindIII* restriction fragments H4 and H4a, corresponding respectively to the allelic forms H4-2 and H4-1. The 4.2-kb H4 fragment from the H4-2 allele was previously isolated from an mtDNA library from *A. aegerita* strain WT-3 (\(\sim SM47\)) (21); the 3.5-kb H4a fragment from the H4-1 allele was isolated from a library of mtDNA from strain WT-11 (\(\sim SM751002\)). Both strains are preserved on CYM medium (23) in the International Culture Collection of the Laboratory of Molecular Genetics and Breeding of Cultivated Mushrooms (collection number MACC WDCM 786) and were previously described (3). *Escherichia coli* JM83 (29) was used for cloning and propagation of plasmids in *Luria-Bertani* medium (20).

DNA manipulations. *HindIII* or *HinII* mitochondrial fragments were cloned into the *HindIII* or *SmaI* sites of pUC18, respectively, by using conventional cloning procedures (8, 20). The H4a restriction fragment was isolated by colony hybridization, using H4 as a probe. Probes were digested with the appropriate restriction endonucleases, separated by electrophoresis in a 0.8% (wt/vol) NuSieve GTG agarose gel (FMC Bioproducts, Rockland, Maine), recovered by using a GeneClean kit (Bio 101 Inc., Vista, Calif.), and labeled with [\(\alpha-^{32}\text{P}\)]dCTP (Amersham Pharmacia Biotech, Little Chalfont, Buckinghamshire, United Kingdom) by using a Promega (Madison, Wis.) Random Primer DNA labeling kit. *A. aegerita* DNA was extracted from vegetative dikaryotic mycelium by using the Erase-a-Base system according to the manufacturer’s recommendations. Recombinant plasmids were purified from the *E. coli* JM83 clones by a conventional miniprep method (20). Both strands were sequenced in reactions using the M13-40 primer, the M13 reverse primer, or specific 18-mer oligonucleotides (Eurogentec, Seraing, Belgium). The sequencing reactions were performed by the dye-deoxy chain termination method (26) with a Sequenase II kit (United States Biochemical Corp., Cleveland, Ohio). Labeled DNA sequencing and sequence analyses. Mitochondrial sequences were subcloned in both orientations in pUC18, then processed to generate nested deletions by using the Erase-a-Base system according to the manufacturer’s (Promega) recommendations. Recombinant plasmids were purified from the *E. coli* JM83 clones by a conventional miniprep method (20). Both strands were sequenced in reactions using the M13-40 primer, the M13 reverse primer, or specific 18-mer oligonucleotides (Eurogentec, Seraing, Belgium). The sequencing reactions were performed by the dye-deoxy chain termination method (26) with a Sequenase II kit (United States Biochemical Corp., Cleveland, Ohio). Labeled sequences were cloned in both orientations in pUC18, then processed to generate nested deletions by using the Erase-a-Base system according to the manufacturer’s (Promega) recommendations. Recombinant plasmids were purified from the *E. coli* JM83 clones by a conventional miniprep method (20). Both strands were sequenced in reactions using the M13-40 primer, the M13 reverse primer, or specific 18-mer oligonucleotides (Eurogentec, Seraing, Belgium). The sequencing reactions were performed by the dye-deoxy chain termination method (26) with a Sequenase II kit (United States Biochemical Corp., Cleveland, Ohio). Labeled
**RESULTS**

Hybridization of the *Aa-polB* gene with a sequence located in a polymorphic region of the *A. aegerita* mtDNA. The *Aa-polB* gene is carried on a 4.2-kb *HaeIII* and a 7.2-kb *HindIII* restriction fragment (4). Four fragments were identified in Southern hybridizations of the *HindIII*- or *HaeIII*-digested mtDNA from WT-3, using the 4.2-kb *HaeIII* fragment or the H4 fragment (21) as a probe: 7.2- and 4.2-kb (H4) *HindIII* fragments and 4.2- and 11-kb *HaeIII* fragments. If mtDNA from WT-11 was probed in a similar manner, then the DNA fragments typical of the H4-2 allele (4.2-kb *HindIII* and 11-kb *HaeIII*) were not seen and those typical of the H4-1 allele (3.5-kb *HindIII* and 5.3-kb *HaeIII*) were evident instead. From these results, we concluded that a sequence hybridizing with the *Aa-polB* gene was present in two locations in both the WT-3 and WT-11 mtDNAs.

We sequenced the 4.2-kb *HindIII* fragment from WT-3, termed H4 (GenBank accession no. AF269234), and the 3.5-kb *HindIII* fragment from WT-11, termed H4a (GenBank accession no. AF269233). H4a was 3,401 nucleotides (nt) in length and was 77% A+T. H4 was 4,225 nt in length and 78% A+T, and it carried two long inverted repeats of 1,954 nt (R), each beginning at a *HindIII* restriction site, separated by a nonrepeated central sequence of 317 nt (Fig. 1). H4 and H4a have 2,282 nt of identical sequence which contains (i) a complete sequence of the inverted repeat R (1,954 nt), (ii) the central nonrepeated sequence (317 nt), and (iii) the last 11 nucleotides of the second copy of the inverted repeat. H4 has a complete second copy of the inverted repeat, while H4a has a 1,119-nt sequence not found on H4. We used the 0.6-kb *HaeIII* fragment from H4a to probe total DNA of strains WT-3 and WT-11. This fragment hybridized as expected with the 3.5-kb *HindIII* and 0.6-kb *HaeIII* fragments from H4a but did not hybridize with any fragments from WT-11.

**Sequence analysis.** We identified open reading frames (ORFs) in the H4 and H4a sequences by following coding rules for *Neurospora crassa* mtDNA (9). In H4a, we found a single large ORF of 1,857 nt, beginning at the *HindIII* site. In H4, this ORF was found in both inverted copies; no additional ORF was present in the 317-nt central nonrepeated sequence. The ORF was interrupted at its 5′ end by the *HindIII* site. We sequenced the region surrounding the H4a fragment on the overlapping 5.3-kb *HaeIII* fragment (Fig. 1) and found a TAA stop codon in the same reading frame immediately before the *HindIII* site. Thus, the ORF is entirely contained in the H4a sequence. The putatively encoded protein had a size of 500 aa from the first ATG (nt 355 to 357) codon or of 617 aa from the termination codon just before the *HindIII* site.

The ORF is 86% (nt) and 96% (aa) identical to the previously described *Aa-polB* gene and gene product (4), respectively, from the *A. aegerita* mitochondrial genome (GenBank accession no. AF061244). Lower percentages of amino acid similarity were obtained with other family B DNA polymerases, encoded by linear protein-primed replicating plasmids (24, 27), such as the bacteriophage Φ 29 (GenBank accession no. X53370; 42% aa identity), or fungal mitochondrial plasmids such as the plasmid pEM of *Agaricus bisporus* (GenBank accession no. P30322; 40% aa similarity), the plasmid pHc2 of *Hebeloma circinans* (GenBank accession no. Y11504; 39% aa similarity), or the kalilo plasmid of *Neurospora intermedia* (SwissProt accession no. P33538; 37% aa similarity).

The 1,119 nt specific to the H4a fragment has no significant sequence identity at the nucleotide level to sequences in the GenBank and EMBL databases. The largest ORF it contained that began with an ATG codon was 210 nt.

The central nonrepeated sequence from H4 (Fig. 1) has 72% identity with two *Saccharomyces cerevisiae* mitochondrial tRNA genes, tRNA^{Met} and tRNA^{Pro}. Based on primary sequence and on secondary structure, both H4 and H4a carry the same mitochondrial tRNA^{Met}, located between nt 2071 (5′ end) and nt 2141 (3′ end), i.e., 213 nt following the first stop codon of the ORF with sequence similarity to the family B DNA polymerase genes.

**Comparison of the large ORF of the polymorphic region with *Aa-polB* and other related family B DNA polymerase genes.** In the 2,282-nt region from the *HindIII* site (nt 1) to the end of the tRNA^{Met} (nt 2071) in both H4 and H4a, the ORF region (nt 1 to 1955) can be aligned with a part (nt 644 to 2602) of the *A. aegerita* mitochondrial sequence (GenBank accession no. AF061244) containing the *Aa-polB* gene preceded by the intergenic region between the 5′ end of the SSU rDNA and *Aa-polB* (86% sequence identity). The remaining 116 nt of H4 or H4a, from nt 1956 to the 5′ end of the tRNA^{Met} (nt 2071), had no significant sequence identity with the *Aa-polB* gene sequence, suggesting that the 3′ end of this gene was missing from the H4 or H4a sequence. If the two protein sequences are
We compared the amino acid sequence similarities of Aa-POLB, Aa-POLB P1, the bacteriophage Φ 29 replicase, the DNA polymerase of the mitochondrial linear plasmid pEM from *Agaricus bitorquis*, and three family B DNA polymerases of *Sulfolobus solfataricus* in the portion of each protein between the highly conserved Pol I and Pol III domains (Table 1). Aa-POLB and Aa-POLB P1 were found to be distinctly related (36 to 53% aa similarity) to the other family B DNA polymerases in the GenBank database. The *A. aegerita* proteins are about as closely related to the three *S. solfataricus* DNA polymerases (36 to 44%) as these proteins are to each other (44 to 50% similarity).

**DISCUSSION**

The *A. aegerita* mitochondrial genome contains the previously described *Aa-polB* gene, which encodes a putatively functional family B DNA polymerase (4), and a truncated, probably nonfunctional paralog, *Aa-polB P1*. Moreover, two widely distributed mitochondrial alleles carry one or two copies of the truncated *Aa-polB P1* sequence. These DNA polymerase genes may have arisen by duplications occurring in two steps (Fig. 3). A first duplication of *Aa-polB*, occurring before *A. aegerita* speciation, led to the divergent (14%) paralog *Aa-polB P1* present in all strains (Fig. 3A and B); then a duplication of *Aa-polB P1*, occurring after *A. aegerita* speciation, led to the copies accounting for the allelic variability of the H4 region (Fig. 3C).

The putative origin of both duplications could be due to an illegitimate recombination between two mitochondrial genomes or, more probably, to the integration of a cDNA at a tRNA mitochondrial locus (Fig. 3). The first hypothesis is supported only by the fact that recombinant mtDNA molecules in *A. aegerita* heteroplasmons have been described (2). Because the *Aa-polB P1* gene is followed by a tRNA<sup>Met</sup>, it seems more probable that an mRNA of the *Aa-polB* gene captured a tRNA<sup>Met</sup> whose 3'-OH end was used as a primer for a mitochondrial reverse transcriptase activity (Fig. 3A) (see, e.g., references 5 and 18). The resulting cDNA (*Aa-polB-tRNA<sup>Met</sup>*) was integrated into the mtDNA by recombination at the tRNA<sup>Met</sup> locus (Fig. 3A and B). We have no information about the other recombination site on the mtDNA; this region of

---

**TABLE 1. Comparison of the *A. aegerita* family B DNA polymerases Aa-POLB and Aa-POLB P1 with five related family B DNA polymerases**

<table>
<thead>
<tr>
<th>Microorganism</th>
<th>Family B DNA polymerase (GenBank accession no.)</th>
<th>Size (aa)</th>
<th>Size (aa) and location of the Pol I to Pol III sequences</th>
<th>% aa similarity&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Aa-POLB</td>
<td>571</td>
<td>271 (IRKS–SKEL)</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>(AF061244)</td>
<td></td>
<td></td>
<td>96</td>
</tr>
<tr>
<td></td>
<td>Aa-POLB P1</td>
<td>500</td>
<td>272 (IRIG–YFYD)</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>(AF268233)</td>
<td></td>
<td></td>
<td>41</td>
</tr>
<tr>
<td></td>
<td>S. solfataricus strain P2</td>
<td>882</td>
<td>327 (RTSA–VNVL)</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>Ss-B1 (U92875)</td>
<td>882</td>
<td></td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>Ss-B2 (X71597)</td>
<td>626</td>
<td>298 (IMMV–RVYL)</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>Ss-B3 (Y08257)</td>
<td>764</td>
<td>242 (IRFI–SKKK)</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>Bacteriophage Φ 29</td>
<td>882</td>
<td></td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>Replicase (X53370)</td>
<td>575</td>
<td>245 (VRYA–TGTE)</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>Replicase (P30322)</td>
<td>773</td>
<td>286 (IRSS–RKPL)</td>
<td>100</td>
</tr>
</tbody>
</table>

<sup>a</sup> The complete sequences of *Aa-polB* as well as those of *Aa-polB P1* in strains WT-3 and WT-11 were identical (see text and reference 4).

<sup>b</sup> Based on the amino acid sequences located between the highly conserved Pol I and Pol III domains.
recombination could be nonhomologous (illegitimate recombination) or, more probably, consist of a short homologous sequence present both on the cDNA and on the mtDNA near the tRNA<sup>Met</sup> locus. The high percentage of A+T in the recombining molecules favors the presence of such a homologous small sequence.

The second duplication event leading to the H4-2 allele appears to require a complex mechanism involving the duplication of Aa-polB P1 and the deletion of a large sequence of size greater than 0.6 kb (Fig. 3C). Since this duplication could have affected the regions preceding the sequenced fragment, our results do not allow the determination of the size of the deleted sequence or the size of the duplicated one.

The mechanism(s) leading to the formation of the Aa-polB P1 truncated gene is unknown; in particular, we failed to find in the short sequence separating the truncation site and the tRNA<sup>Met</sup> any sequence or secondary structure reminiscent of a group I or II mitochondrial intron.

The γ DNA polymerases represent a highly conserved family of nucleus-encoded DNA polymerases responsible for the replication of circular mitochondrial genomes of all eukaryotic organisms. However, recent reports on plant or fungal mitochondria have described the presence of additional DNA polymerase activities. For example, a nucleus-encoded β DNA polymerase activity in yeast mitochondria has been recently reported (19), and family B DNA polymerase genes in the chrysophyte alga <i>Ochromonas danica</i> (6) and in the plant <i>Beta vulgaris</i> (GenBank accession no. Z34298) have been described. The products of these family B DNA polymerase genes could be involved in the replication of linear genomes, and it has been recently reported that most of the plant and fungal genomes are present as linear multimeric molecules (16).

Comparison of Aa-POLB and Aa-POLB P1 sequences with other family B DNA polymerases shows that they are distantly related to the polymerases of <i>Sulfolobus solfataricus</i> (7). If mitochondria originated from a fusion between a <i>Clostridium</i>-like eubacterium and a <i>Sulfolobus</i>-like archaeabacterium (15), the A. aegerita family B DNA polymerase genes could be remnants of the archaeabacterial genes.

The presence of Aa-polB and its paralog Aa-polB P1 in the mtDNA of the basidiomycete <i>A. aegerita</i> suggests that a family B DNA polymerase activity could exist in a fungal mitochondrion; the lack of such a gene in all the ascomycete mitochondrial genomes sequenced to date leads us to hypothesize that in these fungi this gene was eliminated after transfer of a duplicated copy to the nucleus. The search for sequences homologous to Aa-polB in other basidiomycetes and the determination of their genomic location(s) (nucleus and/or mitochondria) will allow us to assess this hypothesis and to reconstruct the evolutionary histories of such mitochondrial family B DNA polymerase genes.

ACKNOWLEDGMENTS

This work was supported by grants from the European Community (Fond Européen de Développement Régional), the Conseil Scientifique de l’Université Victor Segalen Bordeaux 2, the Conseil Régional d’Aquitaine, Monsieur le Préfet de la région Aquitaine, Préfet de la Gironde (Fonds National d’Aménagement et de Développement du Territoire), the Conseil Régional d’Aquitaine, and the Institut National de la Recherche Agronomique.

REFERENCES