Interspecific transfer of mitochondrial genes in fungi and creation of a homologous hybrid gene

BRUNO PAQUIN†, MARIE-JOSÉE LAFOREST, AND B. FRANZ LANG‡

Département de Biochimie, Université de Montréal, CP 6128, succ. Centre-Ville, Montréal, H3C 3J7, PQ, Canada

Communicated by Russell F. Doolittle, August 8, 1994 (received for review March 30, 1994)

ABSTRACT In eukaryotes, horizontal gene transfer is a rare event. Here we show that the mitochondrial genome of a lower fungus, Allomyces macrognynus, has an extra DNA segment not present in a close relative, Allomyces arbusculus. This insert consists of the C-terminus of a foreign gene encoding a subunit of the ATP synthetase complex (atp6) plus an open reading frame encoding an endonuclease. The inserted atp6 portion is fused in phase to the resident gene, resulting in expression of a hybrid atp6 gene and the displacement of the original C-terminal atp6 region. We present evidence that this insertion may have been acquired by interspecific transfer and we discuss the possible role of the endonuclease in this process.

DNA exchange across taxonomic boundaries (1) has played an important role in genome evolution, especially in the prokaryotic kingdom. Many bacterial genomes are mosaic as a consequence of lateral transfer of DNA mediated by conjugation, phage infection, or direct plasmid transformation (2). Stable inheritance of the transferred DNA results either after integration through homologous or heterologous recombination or integrase- or transposase-mediated recombination or simply by self-replication of the imported DNA. Although of eubacterial origin, mitochondria are not capable of such efficient DNA transfer, because they are excluded within the eukaryotic cell. Exchange of nuclear genes between different eukaryotic taxa is very limited and seems to occur preferentially through transposable elements (3). Similar constraints are likely to apply to mitochondrial gene flow. In support of this, only transfer of mitochondrial plasmids between Neurospora isolates has been clearly demonstrated (4, 5), and just a few cases of interspecific transfer of mitochondrial group I and II introns have been postulated (6–8). No cases of mitochondrial gene transfers have been described. Here, we present evidence for a recent interspecific transfer of mitochondrial sequences involving formation of a hybrid, expressed gene. The transferred sequence includes part of a structural gene required for mitochondrial function.

MATERIALS AND METHODS

Strains and Sequence Analysis. The strains used in this analysis were Allomyces macrognynus Burma 3-35 (35°C) his1 (ATCC 46923) and Allomyces arbusculus (ATCC 10983). Cloning, sequencing, and sequence analysis will be described elsewhere (B.P., L. Forget, and B.F.L., unpublished data).

Northern Analysis. RNA was extracted according to Deeley et al. (9) except that a 6 M guanidine hydrochloride solution was employed for lysis of cellular components. One microgram of total RNA was separated in a 1% agarose gel containing formaldehyde (10). Capillary transfer and hybridization on Hybond-N (Amersham) were performed following the manufacturer’s instructions. The probes were labeled using a modified protocol of an oligolabeling kit (Pharmacia) and [α-32P]dATP. The atp6 probe consisted of a Dra I–Taq I restriction fragment covering the 5' part of the gene. The ORF360 probe was made with a PCR-amplified fragment of the entire open reading frame (ORF).

cDNA Amplifications. One microgram of total RNA was incubated with FPLC pure DNase I (Pharmacia) at 37°C for 15 min. The RNA was then extracted with phenol/chloroform and precipitated with ethanol. First strand synthesis was carried out with avian myeloblastosis virus reverse transcriptase (Boehringer Mannheim, as detailed by the manufacturer) and either primer 2 (5'-ATCTTAGATCAAGTTCAAACGC-3') for the amplification of the precursor cDNA or primer 4 (5'-TGATACCCATATTCC-3') for generation of the cDNA sequence of the 3' region of atp6 mRNA. PCR amplification was carried out by adding the appropriate primer [primer 1 (5'-ATCTTAGATCAAGTTCAAACGC-3') for the precursor RNA or primer 3 (5'-CGTATCGCCAGTG-3') for the cDNA of 3' atp6] with Vent polymerase (New England Biolabs). Hybridization sites for primers are shown in Fig. 1. PCR fragments were then gel purified, cloned, and sequenced.

Protein Sequence Alignment. The optimum alignment was computed using the multiple alignment software packages MULTALIN (11) and CLUSTAL (12).

Endonuclease Assay. Endonuclease reactions were performed as described (13). The substrate DNA was obtained by PCR amplification using primers 1 and 4, on A. arbusculus mtDNA. The expression vector used was pBlueScript (Stratagene), which was modified by in vitro mutagenesis such that the third codon (AUG) of the lacZ gene overlaps an Nco I restriction site. The ORF360 sequence was then inserted with its initiation codon at this Nco I site of the vector.

RESULTS AND DISCUSSION

When sequencing the complete mtDNA of A. macrognynus (ref. 15 and unpublished data), we observed that the region of the gene coding for the subunit 6 of the ATP synthetase complex (atp6) had a peculiar organization (Fig. 1). In addition to an ORF encoding a complete, bona fide atp6 protein, a second copy of the C-terminal portion (atp6*-C) was found about 1500 bp downstream. Sequence similarity between atp6*-C and the corresponding part of the complete atp6 (3' atp6) extends over a stretch of 64 amino acids, with 92% identical positions at the protein level and 80% at the nucleotide level (Fig. 2). The high identity at the protein level reflects the fact that most of the nucleotide changes are silent, either occurring at the third position of a codon or specifying the same amino acid by a member of a different codon family. Note that four of five amino acid changes between 3' atp6 and atp6*-C are conservative and that the sequence similarity at

Abbreviation: ORF, open reading frame.
†Present address: Department of Biological Sciences, University at Albany, State University of New York, Albany, NY 12222.
‡To whom reprint requests should be addressed.
the nucleotide level stops abruptly after the respective termination codons. The sequence between 3' atp6 and atp6*-C potentially codes for a polypeptide of 360 residues and is referred to as ORF360 (previously called ORF414 (15)). ORF360 has a codon bias indistinguishable from other protein-coding genes in this genome (unpublished results).

To obtain clues on the origin of this unusual gene organization, a comparative analysis of the atp6 regions from A. macrogyrus and another species of the same genus, A. arbusculus, was performed. We carried out PCR amplification of the corresponding atp6 regions and obtained a single 222-bp fragment in A. arbusculus as opposed to the fragments of 222 bp and 1.7 kb found in A. macrogyrus (Fig. 3), suggesting that ORF360 is absent in the mtDNA of A. arbusculus. This was confirmed by sequencing the atp6 region in A. arbusculus, which revealed a single, continuous atp6 gene without any duplication of its 3' region. Comparison of the atp6 sequences of the two species shows >99% nucleotide identity between the 5' portions of the gene, but only 80% for the C terminus. Surprisingly, however, atp6*-C of A. macrogyrus and the 3' end of the A. arbusculus atp6 gene are identical and even the downstream adjacent intergenic sequence contains few nucleotide changes (Fig. 2).

Therefore, we hypothesize that (i) the gene organization in A. arbusculus is the original one and (ii) atp6*-C in A. macrogyrus has recently been separated from the 5' atp6 coding region by an insertion, composed of 3' atp6 and ORF360 (Fig. 1). The insertion site is located within a block of 9 nt that starts exactly at the position corresponding to the beginning of atp6*-C (Fig. 2). The insert is fixed rather than polymor-
Table 1. Nucleotide and amino acid changes between atp6*C and 3' atp6 and corresponding coding regions of A. nidulans and N. crassa

<table>
<thead>
<tr>
<th>Changes</th>
<th>A. nidulans</th>
<th>atp6*C/3' atp6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nucleotide</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transitions</td>
<td>17</td>
<td>19</td>
</tr>
<tr>
<td>Transversions</td>
<td>12</td>
<td>21</td>
</tr>
<tr>
<td>Total</td>
<td>29</td>
<td>40</td>
</tr>
<tr>
<td>Amino acid</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conservative</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Nonconservative</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>6</td>
<td>5</td>
</tr>
</tbody>
</table>

As an alternative explanation we propose that site-specific integration created a hybrid atp6 gene consisting of its original 5' portion fused to a foreign 3' atp6 sequence. Such an integration restores a full-length atp6 gene, mitochondrial function would not be impaired. If this model is correct, it is probable that the inserted DNA segment containing the C-terminal part of atp6 plus ORF360 originated from a horizontal gene transfer event. Coexistence of two copies of the C-terminal sequence and relatively low sequence identity between the copies are properties indicating interspecific gene transfer (17). Any type of intragenomic recombination event leading to the described gene organization can be clearly ruled out because only one copy of ORF360 is found within the complete sequenced mtDNA of A. macrognynus (B.P. and B.F.L., unpublished data).

The gene transfer postulated here apparently occurred recently, because atp6*C is identical to that found in A. arbusculus. If the integration was an ancient event, some differences between the two sequences would be expected, at least in third positions of codons. Based on its high (A+T) content and a codon bias characteristic of mitochondrial protein-coding genes, the insert (3' atp6 plus ORF360) is almost certainly of mitochondrial origin. From the degree of sequence divergence between C. and the accession 3' atp6, we can extrapolate that the foreign sequence is as distant from A. macrognynus as Aspergillus nidulans (Ascomycetes, Plectomycetes) is from Neurospora crassa (Ascomycetes, Pyrenomycetes) (see Table 1). It is conceivable that the acquired sequence originated from fungi parasitizing Allomyces sp. (ref. 18, pp. 178–180 and 655–657), and it would be interesting to survey mtDNAs of known parasitic fungi for the occurrence of ORF360-related sequences.

The organization of the atp6 region in A. macrognynus is reminiscent of the integration pattern of the cryptic prophage e14 (19) and other lambda phages (20). The DNAs of these phages can specifically integrate into the E. coli chromosome, replacing the 3' end of a resident gene by a highly similar sequence and displacing the original 3' end. The gene portion carried by the phage and allowing targeted insertion, originates from the host chromosome (20). This site-specific recombination system of the lambda phages relies on integrases (21) that share little sequence similarity among each other, except four conserved amino acid positions (22–23), a motif not present in ORF360. However, a search of ORF360 in the protein data bases did reveal some similarity to endonucleases of the GIY...YIG type (24) (Fig. 5). Further experiments confirm that ORF360 indeed codes for an endonuclease that cuts the atp6 sequence of A. arbusculus at a single site (within the limits of detection of an agarose gel separation), corresponding to the insertion site in the A. macrognynus atp6 gene (Fig. 6). The hybrid atp6 sequence of A. macrognynus was not cut by the endonuclease to any detectable level (results not shown).

We propose that the insert in the atp6 gene of A. macrognynus represents a mobile element that encodes an endonuclease and part of a foreign structural gene. The rationale of this proposal relies on previous findings showing that "intron homing" is mediated by an endonuclease of a highly specific endonuclease, independent of the conserved RNA structure of the intron (30), and that a gene encoding an endonuclease by itself can be mobile (32). The role of the endonuclease would be to cut an insertless gene at a specific target site, initiating the integration of the mobile element through gene conversion—a mechanism proposed for group I introns called homing (for review, see ref. 31). The site specificity of the cleavage would ensure an in-phase fusion of the structural gene portion that is carried by the mobile element. As
opposed to introns, which split genes and restore their integrity by RNA splicing, these insertion events result in the expression of complete hybrid genes. Interspecific transfer of the element to more distant species would have to meet several additional criteria: (i) the hybrid genes created in this way have to be functional for the cell, (ii) the endonuclease must be able to recognize a specific cut site in a sufficiently conserved gene sequence of the host, and (iii) the recognition site needs to be sufficiently changed by the insertion in order to be protected against a subsequent deleterious endonuclease cleavage.

Like mobile introns that also spread through horizontal transfer (6–8, 33), one might predict the presence of this mobile element in many different genes from a wide variety of species and different genomes. Other examples of such elements may have been overlooked because an integration event would not necessarily be apparent in cDNA or gene sequence analysis, particularly when the displaced C-terminal sequence of the original gene or the ORF of the insertion element is subsequently deleted.

ORFs encoding endonucleases similar to those found in group I introns (31) have also been identified in an archael intron (34), as free-standing sequences (35–38), as part of “protein introns” (39), and now in the element described here. The variety of forms, functions, and genome locations occupied by these endonucleases further supports the view that they were not originally integral parts of ancestral introns but have, instead, evolved as separate mobile entities and are only associated secondarily with group I introns (30).

We thank G. Burger, T. Littlesjohn, I. Debie, C. J. O’Kelly, B. Golding, and M. W. Gray for discussions and for helpful comments on the manuscript. This work was supported by the Medical Research Council (Canada) and Fonds pour la Formation de Chercheurs et l’Aide à la Recherche (Canada), a fellowship (B.F.L.) from the Canadian Institute for Advanced Research, and scholarships (B.P.) from the Natural Sciences and Engineering Research Council (Canada) and Fonds pour la Formation de Chercheurs et l’Aide à la Recherche (Canada).