The Unusual *var1* Gene of Yeast Mitochondrial DNA

Ronald A. Butow, Philip S. Perlman, Lawrence I. Grossman

Summary. The *var1* gene specifies the only mitochondrial ribosomal protein known to be encoded by yeast mitochondrial DNA. The gene is unusual in that its base composition is nearly 90 percent adenine plus thymine. It and its expression product show a strain-dependent variation in size of up to 7 percent; this variation does not detectably interfere with function. Furthermore, *var1* is an expandable gene that participates in a novel recombinational event resembling gene conversion whereby shorter alleles are preferentially converted to longer ones. The remarkable features of *var1* indicate that it may have evolved by a mechanism analogous to exon shuffling, although no introns are actually present.

Mitochondria synthesize a dozen or so polypeptides required for mitochondrial function. Synthesis of these polypeptides takes place on mitochondrial ribosomes, assembled from the ribosomal RNA's (rRNA's) that are encoded by mitochondrial DNA (mtDNA) and from the ribosomal proteins encoded largely or exclusively by the nuclear genome. This nucleo-cytoplasmic division of labor necessary for assembling mitochondrial ribosomes appears to be the rule also for all other identified protein complexes containing mtDNA products; such complexes always contain subunits encoded by the nucleus. In that sense, mitochondria can be considered to exist in symbiosis with the nucleus.

The known mitochondrial gene products vary little among eukaryotes as diverse as fungi and vertebrates. They always include subunits I, II, and III of cytochrome c oxidase, subunit 6 of the adenosinetriphosphatase (ATPase) complex, cytochrome b of the bc_1 complex, the small and large rRNA's, and at least 22 transfer RNA's (*1–3*). Recently, ATPase subunit 8 was identified on yeast mtDNA and, by DNA sequence homology, on animal mtDNA (*4*). Other genes are mitochondrial only in some cases. For example, subunit 9 of the ATPase complex is encoded by mtDNA in the ascomycete yeast *Saccharomyces* (*5*), but is a nuclear gene product in the ascomycetes *Neurospora* (*6*) and *Aspergillus* (*7*), and probably in vertebrates as well (*8–10*).

In both yeast (*11*) and *Neurospora* (*12*) and possibly in other lower eukaryotes (*13, 14*), one mitochondrial small subunit ribosomal protein (*var1* in yeast and *S5* in *Neurospora*) is encoded by mtDNA, while all other ribosomal proteins of both subunits are encoded in the nucleus (*15*). This adds another dimension to the division of labor for mitochondrial protein synthesis noted above. On the basis of studies with mutants and with specific inhibitors of mitochondrial protein synthesis, both var1 and *S5* appear to be required for the assembly of mature mitochondrial small ribosomal subunits (*16, 17*). In contrast to these fungi, there is no evidence that any of the proteins of either mitochondrial ribosomal subunit is encoded by mtDNA in vertebrates (*18*).

Why yeast or *Neurospora* mtDNA harbors a gene for just one of the 70 or so mitochondrial ribosomal proteins is unknown, as are the answers to several even more fundamental questions. For instance, why should mitochondrial genomes encode any of the particular proteins that they do? Why do mitochondrial gene products function in complexes with nuclear-encoded polypeptides? And why, in fact, do cells expend more than a hundred nuclear genes just to allow the mitochondrial genetic system to translate about a dozen polypeptides in situ? The unusual properties of the var1 ribosomal protein and its structural gene have led us to consider some of these questions, and to reevaluate some generally held views on the structure, function, and evolution of elements of the yeast mitochondrial genome, and of mitochondrial genes as a whole.

Var1

The gene encoding var1 protein in yeast has turned out to be one of the most unusual genes seen thus far in any system. On the basis of electrophoretic mobility, more than a dozen different strain-dependent variants of var1 protein ranging in apparent size between 40,000 and 44,000 daltons (*19*) have been detected in unselected yeast stocks by sodium dodecyl sulfate–polyacrylamide gel electrophoresis (*20–22*). However, in any given yeast strain only one form of var1, whose size is stable throughout vegetative growth, has been observed. Furthermore, we have not seen any phenotypic differences among respiration-competent strains containing different forms of var1 that can be attributed unambiguously to the size polymorphism. Thus, although these var1 forms differ by as many as 26 amino acid residues in various combinations, mitochondrial function, including protein synthesis, is not affected in any way we have been able to measure.

Two lines of evidence indicate that var1 protein is essential for the biogenesis of functional mitochondria. (i) The assembly of mature small mitochondrial ribosomal subunits is completely dependent on mitochondrial protein synthesis, presumably because of a limitation in the amount of var1. If mitochondrial protein synthesis is selectively inhibited, small ribosomal subunits are not assembled (*17*). However, assembly of large subunits, all of whose proteins are encoded in the nucleus (*23*), is unaffected. (ii) Genetic studies have uncovered three mutants falling within a single complementation group that map to the *var1* region (*24, 25*). In addition to being respiratory-deficient phenotypes, these mutants have very low levels of mitochondrial protein synthesis. One maps within the *var1* reading frame, and the other two map within the *var1* transcription unit within previously presumed "spacer" DNA, one 5′ and the other 3′ to the coding region.

Those mutants were the first indications that functionally important sequences affecting the expression of a structural gene lie in the AT-rich (A, adenosine; T, thymine) spaces between genes. In the mutant mapping 5′ to the coding region, there appears to be a

Ronald A. Butow is a professor in the Department of Biochemistry, University of Texas Health Science Center, Dallas 75235. Philip S. Perlman is a professor in the Department of Genetics, Ohio State University, Columbus 43210. Lawrence I. Grossman is an assistant professor in the Department of Cellular and Molecular Biology, Division of Biological Sciences, University of Michigan, Ann Arbor 48109. Send requests for reprints to L. I. Grossman at the above address.

block in the processing of precursor to functional *var1* messenger RNA (mRNA) (*26*). Interestingly, the unprocessed *var1* transcript in this mutant contains the entire reading frame in addition to 5′ and 3′ extensions; it differs from the mRNA only in that it has a longer 5′ noncoding segment, which apparently interferes with its functioning as an mRNA. This result, together with recent studies of transcripts of several other mitochondrial genes (*27, 28*), indicates that the long AT-rich leader sequences are functionally important not only in transcript processing but also in the formation of productive translation initiation complexes.

Genetic Analysis of *var1*

We have taken advantage of the stable var1 polymorphisms in genetic studies, first to show that the molecular size differences observed in the protein are specified by a locus on mtDNA we have termed *var1* (*20*), and second to map this locus to a region of the mitochondrial genome between the large (21*S*) rRNA gene and the ATPase subunit 9 gene (*29*). The gene *var1* is the first structural gene to be mapped on yeast mtDNA by direct analysis of its protein product.

A systematic analysis of the behavior of *var1* in crosses has provided a satisfying genetic framework to explain its remarkable size polymorphism. By applying conventional mitochondrial transmission genetics (*30–32*) and a complementation assay that we developed and termed zygotic gene rescue (*33, 34*), we identified two types of genetic elements associated with *var1* that account for all observed forms of var1 protein. The genotypic designations of the *var1* alleles are based on the combinatorial arrangement of these specific elements, which we have designated *a, b,* and variant forms of *b,* termed b_p (Table 1).

The key observation that led to this genetic formulation was the finding that, when two yeast strains are crossed, one containing, for example, the smallest detectable var1 (var1[40.0]), and the other the largest (var1[44.0]), two major classes of nonparental or recombinant forms appear in the diploid progeny that are intermediate in size between the parental forms (Fig. 1) (*30*). One class appears at a frequency of 20 to 30 percent of the population, and the other at about 2 percent. However, in crosses between strains with identical *var1* alleles, only the parental form of the protein is observed among the diploid progeny. From these and other results, we have been able to establish the genotypes indicated in Table 1 and unambiguously assign them to the *var1* forms listed there.

Table 1. The genotypes of forms of var1 protein.

Var1 allele* (kD)	Genotype
40.0	a^-b^-
41.8	a^+b^-
42.0	$a^-b_p^+$
42.2	a^-b^+
43.8	$a^+b_p^+$
44.0	a^+b^+

*Designated by the apparent molecular size (kilodaltons) of the var1 protein as estimated by sodium dodecyl sulfate–polyacrylamide gel electrophoresis.

These genotypes reflect all possible combinations of *a, b,* and b_p. Thus, the *a* element is either present (a^+) or absent (a^-) from *var1* while the *b* element, although usually recombining as a unit, can be subdivided further into a class denoted b_p (for *partial b* elements) (*32*). These latter events occur at low frequency (<1 percent) and are detected as *var1* forms *smaller* than the form specified by the a^-b^+ allele but *larger* than that of the a^-b^- allele. Support for subclasses of *b* comes primarily from the fact that crosses between different b_p^+ strains yield some progeny that are indistinguishable, both in size of var1 and recombination behavior, from strains containing the original b^+ element. The fact that the *b* element usually recombines as a unit is all the more surprising given the observation (to be discussed below) that it actually consists of two DNA insertions separated by about 350 base pairs (bp).

As presented schematically in Fig. 1, these elements can be considered as DNA sequences capable of transfer in crosses from one *var1* allele to another. By comparing the input of parental *var1* forms to the output of all forms issued from a cross, we have concluded that recombination, particularly that of the *a* element, has the characteristics of an asymmetric gene conversion event whereby "short" alleles are preferentially converted to "long" ones (*30*). Such asymmetry in recombination is known to occur at only one other location on the yeast mitochondrial genome and involves alleles of the 21*S* rRNA gene (*1*). Like *var1*, short rRNA alleles (that lack an intervening sequence) are preferentially and efficiently converted to long ones (containing the intervening sequence) (*35*).

Structure of the *var1* Gene

In Figure 2, we show a schematic representation of the *var1* gene derived from the complete nucleotide sequence of the shortest allele (*var1[40.0]*) (*36*) and its flanking sequences (*37*), and from comparisons of this sequence with alleles containing various combinations of the *a, b,* and b_p elements (*38*). There is an uninterrupted reading frame of 1188 bp, beginning with an ATG and ending with an ochre terminator, that encodes a protein with a molecular size of 46,786 daltons.

Our DNA sequence of *var1* has confirmed that this gene is not typical of other yeast mitochondrial structural genes. Yeast mtDNA has the unusual base composition of 82 percent A plus T (*39*) and, from extensive studies (*1*), appears organized as genes of modest GC (G, guanine; C, cytosine) content (18 to 25 percent) embedded within so-called "spacer" DNA. The latter comprises about one-half of the 75-kb mitochondrial genome and consists of long stretches of AT-rich DNA (95 percent), interspersed with about 100 short GC-rich clusters, each 20 to 75 bp long (*40–42*). Previous DNA sequence analysis of the *var1* region (*43*) had indicated that it is

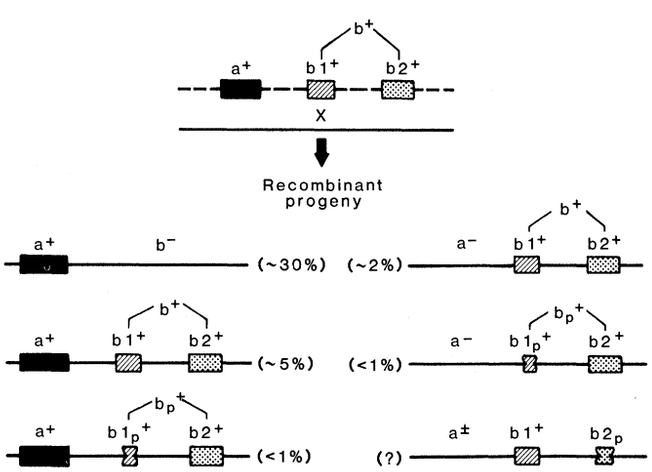

Fig. 1. Elements specifying forms of var1 polypeptide. Recombinant progeny issuing from an $a^-b^- \times a^+b^+$ cross and their frequencies are indicated. The *b* element has been subdivided by DNA sequencing into two elements, *b1* and *b2*. The level of a^+b^+ recombinants is estimated.

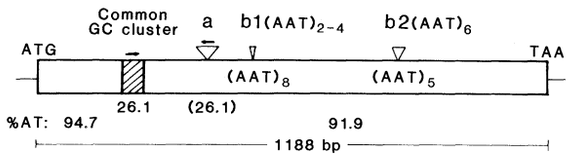

Fig. 2. Organization of the *var1* gene. The schematic representation of the 1188 bp *var1* reading frame shows the common GC cluster (cross hatches) and the positions of the three genetic elements that can expand the gene. The arrows above the cluster and *a* element shows their relative orientations. For *b1* and *b2*, the number of asparagine (AAT) residues in each element is shown. The base composition of several regions is shown below the gene, with that for *a* in parenthesis.

very AT-rich, with interspersed GC-clusters, thereby more closely resembling "nongenic" mtDNA than a typical gene sequence.

In fact, the *var1*[*40.0*] allele is 89.6 percent A+T and is among the most AT-rich protein genes that we know of anywhere in nature (*36*). Not surprisingly, the extraordinarily high AT-content of the *var1* gene is reflected in a high proportion (83.4 percent on a molar basis) of amino acids encoded by AT-rich codons. Most striking is the extraordinarily high number of asparagines, which comprise nearly one-third of the total amino acid residues in the expressed protein! Of the total GC residues present within the reading frame, nearly 40 percent are concentrated within a 46-bp palindromic GC cluster located 186 bp downstream from the ATG initiator; we refer to it as the "common GC cluster" (Fig. 2) because all *var1* alleles contain it. This same GC-rich sequence is found in at least two other locations on the yeast mitochondrial genome (see below). With the exception of *var1*, GC-rich clusters as a class are not found within mitochondrial protein genes, although they have been detected within the 21*S* (*35*) and 15*S* (*44, 45*) mitochondrial rRNA genes. In those cases, a 66- and 40-bp GC cluster, respectively, are found within the gene in some, but not all, yeast strains; when present, however, these GC clusters are transcribed and appear in the mature rRNA's (*44–46*).

The identification and nucleotide sequence of the *var1* gene has led to two important conclusions concerning the yeast mitochondrial genetic code (*36*). First, it allowed identifying AUA (U, uracil) as a methionine codon, a deviation from the "universal" genetic code also seen previously in vertebrate (*47*) and, subsequently, in *Drosophila* (*48*) mtDNA's. Second, it confirmed the assignment of CUN (N, unassigned base) as threonine in yeast mitochondria (*49*), a unique assignment thus far. These variations of the genetic code used by yeast mitochondria are especially provocative because the mitochondria from another ascomycete, *Neurospora crassa*, use the universal codon assignments of AUA as isoleucine and CUN as leucine (*50, 51*).

A comparison of the codon usage in the *var1* gene with other yeast mitochondrial genes reveals a bias toward those codons used preferentially in unidentified genes (*36*). Excluding *var1*, there are 7 identified and up to 14 unidentified reading frames. For instance, UUU is used preferentially (about 7:1) over UUC for phenylalanine in unidentified genes and exclusively in *var1*, although these codons are used about equally in other yeast mitochondrial genes. We have previously interpreted this division in codon usage between unidentified genes and the identified ones, which are all involved in oxidative phosphorylation, as suggesting that these two groups of genes represent evolutionarily independent events in the colonization of fungi by endosymbionts (*36*). Since the group of unidentified genes shows differences in different organisms, and the split genetic code differs between yeast and *Neurospora*, we suggest that mitochondrial genomes may have evolved along parallel but not identical lines even in related organisms.

Molecular Basis of *var1* Polymorphism: An Expandable Gene

Comparisons of the coding sequence of different *var1* alleles and analyses of their products revealed a number of unexpected and novel molecular features underlying *var1* size polymorphism. As suggested by the genetic studies, the physical correlates to the *a*, *b*, and b_p elements are, in fact, in-frame DNA insertions that provide additional coding information within *var1* and thus "expand" the gene (*38*). The types of insertions associated with the *a* and *b* elements, however, are quite different.

The *b* (and b_p) elements correlate with in-frame insertions of asparagine codons (AAT) within strings of AAT repeats. Thus far, we have detected two such acceptor repeats consisting of five and eight AAT's into which additional AAT codons are inserted. These acceptor sequences are indicated in Fig. 2 as b_2 and b_1, respectively. In all strains genetically defined as b^+, an insertion into b_2 of six AAT codons is observed, expanding that string from 5 to 11. The insertion into b_1, however, is not identical in all b^+ strains. For example, when we compare the sequence of an a^-b^- *var1* allele with that of a strain whose *var1* is genetically defined as a^-b^+, an insertion into b_1 is observed of four AAT codons, expanding that run from 8 to 12. When strains whose *var1* is b^+_p are examined, only two AAT codons are inserted into b_1. This finding that b_p is physically a subset of the more common insertion into b_1 rationalizes the genetic result that some progeny of crosses between b_p strains are b^+.

Our current picture of the anatomy of *b*-containing alleles is that they contain an 18-bp b_2 insert, and an insertion into b_1, which is most commonly 12 bp (in b^+ strains) but may be as few as 6 bp (in b^+_p strains). Although b_1 and b_2 are separated by about 350 bp, the conversion of b^- to b^+ in a cross nevertheless involves the insertion of all ten "extra" b^+ AAT codons into the b^- allele. From quantitative estimates of the recombinant progeny issuing from $b^+ \times b^-$ crosses, in about one out of five pairings between these alleles not all of the b^+-associated asparagine codons are inserted into b^-. The products of these minority conversion events are those which specify a *var1* protein shorter than the b^+ product that we have designated b_p^+. This class may actually contain a number of alleles, reflecting in both number and location all possible combinations of these additional asparagine codons within the *var1* gene. Indeed, examination of even rarer recombinants with protein sizes between that of *b* and b_p may reveal just such combinations.

Similar comparisons between a^+ and a^- alleles show that the a^+ insert correlates with the presence of an additional GC cluster (Fig. 3) located within the gene 146 bp downstream from the common GC cluster (*38*). Remarkably, this second GC cluster is identical in sequence to the 46-bp common GC cluster, but is in the opposite orientation. It is flanked by two additional A residues compared to a^- strains, so that the total addition of 48 bp remains in-frame for coding.

An interesting feature of these two GC clusters is that their flanking sequences are identical (Fig. 3); these are pure AT sequences, 11 bp on one side and 9 on the other, that occur only within the *var1* gene. These same pure AT sequences are also found near the *oli1* gene flanking a GC cluster identical in sequence to those in *var1* (*43*); in that case, however,

the flanking sequences are inverted relative to their orientation in *var1* (Fig. 3). These correlations suggest that the GC clusters are not inserted randomly, but rather are directed toward some specific acceptor sequence. A possible mechanism for such insertions is suggested by the fact that the putative acceptor sequence in a^- alleles is shorter by 2 bp. Thus, insertion of the GC cluster from a^+ to a^- could occur by a transposition-like process, with a staggered 2-bp cleavage and duplication within the 20-bp acceptor sequence generating the 22-bp flanking sequences. From these considerations, and the well-known variability in the location of GC clusters among different wild-type yeast mitochondrial genomes (*35, 38, 45, 52, 54*), these GC clusters may be analogous to insertion elements in other genomes.

Transposition is an attractive hypothesis to explain the original formation of an a^+ allele. However, it precludes, without special assumptions, the transposition of the common GC cluster to the *a* acceptor site, because experimentally a^+ progeny are never recovered from crosses of the form $a^- \times a^-$. A simple interpretation of the recombination behavior of the *a* element is that the a^- allele acquires the *a* element by asymmetric gene conversion; that is, a mismatch correction to a^+ of heteroduplexes occurs between a^+ and a^- strands.

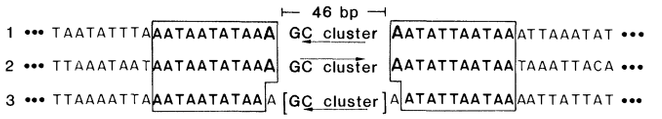

Fig. 3. Location of the *var1* GC cluster on the yeast mitochondrial genome. The *var1* GC cluster is shown embedded in identical flanking sequences (boxed) at all known locations on the yeast mitochondrial genome. (Line 1) 3′ to the *oli1* (ATPase subunit 9) gene; (line 2) in the common site within *var1*; (line 3) at the optional site in *var1* to yield an a^+ genotype. The two A residues shown flanking this GC cluster are not present at that location in a^- strains. Arrows show the relative orientation of the GC cluster at each location.

Evolution of the *var1* Gene

The unusual organization of the *var1* gene, whereby it differs from noncoding AT-rich "spacer" sequences only in representing an open reading frame, raises the questions of the origins both of the *var1* gene and of the spacer sequences.

The most common supposition about mitochondrial origin is that mitochondria originated as endosymbionts, which conferred an advantage to emerging eukaryotic cells, such as the ability to carry out oxidative metabolism (*55*). Of the genes which entered with the proposed endosymbiont, most probably duplicated existing functions and were lost, some were transferred to the nucleus, and some remained in mitochondria to constitute their present-day gene complement.

Several lines of evidence indicate that genes can and do move between mitochondria and nucleus. The most-studied case is subunit 9 of the ATPase complex. This gene has been found on mtDNA only in yeast, and has been shown explicitly to be a nuclear gene in *N. crassa* (*5*). In addition to ATPase 9, other mtDNA sequences have been found in the nucleus in a variety of organisms, although most as fragmented and rearranged versions of their mitochondrial counterparts (*56–60*).

If mitochondrial genes can move to the nucleus (or *vice versa*), and have done so in the past, why was transfer not quantitative? Past speculation has centered on the suggestion that mitochondrial translation products are so hydrophobic that they are best made in situ (*61*). Other hydrophobic proteins are successfully transported in the cell [for example, (*62*)], however, and some mitochondrially coded proteins are as hydrophilic as typical soluble proteins (*63*). An alternative suggestion, supported by the different locations of ATPase 9 in different ascomycetes, is that mitochondria contain genes not for a profound functional reason, but because some genes were trapped there when the transfer of mitochondrial genes to the nucleus was interrupted by an unknown event in the evolutionary past. For instance, as suggested by Fox (*64*), the evolution of the mitochondrial genetic code, in which the nuclear opal terminator, UGA, is translated as tryptophan in mitochondria, would have served as an effective barrier to the further transfer of usable genes between the two compartments. It is interesting that ATPase 9 is the only present-day mitochondrial protein that lacks tryptophan, and thus could have transferred at any time after the divergence of yeast and *Neurospora* without requiring mutational changes in UGA tryptophan codons.

If functional nuclear genes eventually predominate over mitochondrial copies, as appears to have been the general case, then such mitochondrial genes, once removed from selective pressure, are free to change. One can picture three pathways these mitochondrial genes might follow: (i) mutation to noncoding sequences, (ii) elimination by deletion, and (iii) recombination or mutation (or both) to form coding sequences with new functions.

Mutation to noncoding sequences in mitochondria would be rapid, given the apparent lack of at least a number of

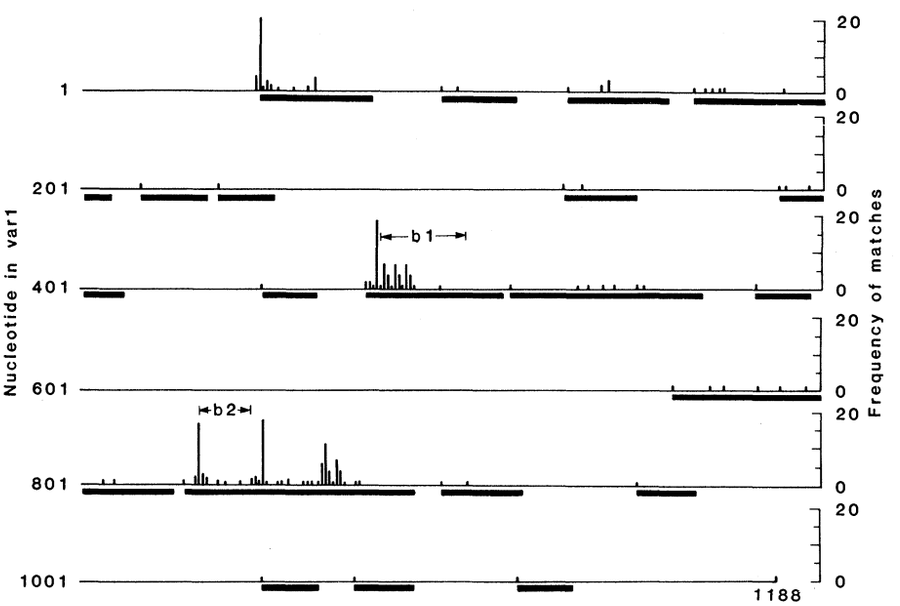

Fig. 4. Homology of *var1* to sequences in the remainder of the yeast mitochondrial genome. The reading frame is indicated by the thin horizontal lines, each representing 200 bp. Heavy horizontal lines show regions that match at least 16 bp exactly one or more times in the remainder of the genome. Vertical bars show the frequency with which those positions in *var1* start matches of at least 16 bp with the remaining genome. The positions of *b1* and *b2* are also shown.

repair mechanisms (*65, 66*). If mitochondria also lack a dUMP (deoxyuradylate) excision system to remove deoxyuridine (resulting from spontaneous deamination of deoxycytidine), as seems likely, then noncoding sequences would become AT-rich, as commonly found for fungal mitochondrial genomes. Elimination by deletion of noncoding sequences, if it occurs in an infrequent, punctuated manner, would cause mitochondrial genomes to vary widely in size, and approach the minimum size necessary to encode the remaining mitochondrial functions, a prediction which is also fulfilled.

Recombination to form new coding sequences could take place between AT-rich spacer DNA segments that bracket (short) unexpressed open reading frames. A functional protein might then be generated by an extended open reading frame, including the AT-rich recombination joints. Examination of *var1* shows an organization that suggests such a recombinational shuffling model analogous to exon shuffling (*67*).

The *var1* reading frame was matched by computer (*68*) with all other sequenced regions of the yeast mitochondrial genome to find regions of homology of length >15. These regions are shown in two ways in Fig. 4. The most prevalent matches by far correspond to the repeat $[AAT]_n$.

Thus, regions that appear commonly are interspersed with regions that occur less commonly, and with long regions that do not occur at all elsewhere on the known genome. These findings suggest that the *var1* gene may have originated by recombination between highly repeated AT-rich segments, such as the [AAT] repeats, juxtaposing domains that may have existed earlier as individual genes or coding units. Of importance is that the recombination "joints" were retained and remained open, thereby preserving a full reading frame.

Two of the three regions in *var1* that contain the highest frequency of matches with other regions are sites of *var1* polymorphism and are very asparagine-rich. The high frequency with which these polymorphic forms appear in genetic crosses under nonselective conditions, as well as the lack of any obvious phenotypic differences among them, suggests that var1 protein may be functionally indifferent to these structural changes. While it does not follow that these domains are dispensable, the observations are consistent with their not having been selected for encoding a functional domain, but rather for a secondary and more flexible role, such as linking functional domains. Indeed, secondary structure predictions of var1 indicate that potential domains of the protein are separated by asparagine-rich β-turn regions (*69*), encoded by AAT strings.

Some support for such a model comes from recent correlative studies in which the splice junction region of related proteins was examined. For example, a comparison of eubacterial and eukaryotic serine proteases and dihydrofolate reductases shows that many of the splice junctions in the eukaryotic proteins and the corresponding regions of the eubacterial proteins coincide with regions of polypeptide length variability (*70*). A plausible interpretation of these observations is that, in the evolution of these proteins, the insertion or removal of intervening DNA from these regions may have been relatively imprecise, so that some exon sequences are removed, or some intron sequences become protein coding information. In effect, such nonessential sequences can be dispensed with at either the genomic level by deletion or the RNA level by splicing. When present, however, these optional sequences may fulfill some role—for example, the maturase function encoded by the optional intron 2 of the *cob* gene of yeast mitochondria (*71*). Alternatively, such sequences may be informational and expressed at present but could evolve in the future to become dispensable and eliminated entirely.

Stability of Mitochondrial Genomes

Since the initial discovery of *var1* size polymorphism, many additional examples of mitochondrial genomic variation have been described in comparable detail. These are optional introns, "mini-inserts" in rRNA genes, and optional A+T- or G+C-rich clusters in intergenic spaces [reviewed in (*1*)]. Thus, the yeast mitochondrial genome consists of a large set of conserved sequences, interrupted by an assortment of optional sequences. It is now clear that the genome of a given strain is very stable. When genomes of different strains are compared, their maps of coding sequences are perfectly conserved and each optional sequence is present or absent at a particular position. From detailed comparative studies, it appears that the optional sequences could be traced to different species of the genus (*72, 73*). The wide range of functionally acceptable combinations probably arose in laboratory strains by unselected mitochondrial recombination events during the decades of genetic manipulation in which current laboratory strains were derived. The spontaneous loss of an optional sequence is an exceedingly rare event. Indeed, there are only a few reports of the loss of a portion of the genome without creating a mutant phenotype—for example, the excision of the last two introns from the cytochrome *b* (*cob*) gene, indicating these introns are optional, and are capable of a faithful excision process (*74, 75*).

GC Clusters

The *var1* common GC cluster and the *a* insert are unusual in representing GC clusters within a protein coding gene (*36, 38*). Nevertheless, the presence of short GC-rich sequences, usually containing Hae III and Hpa II sites, is known to be a prominent part of the yeast mitochondrial genome, with more than 100 scattered copies located primarily in intergenic regions (*41*). These clusters have been suggested to function in control of initiation or termination of transcription (*41, 76, 77*), in RNA processing (*43*), or to be sites for recombinational excision of petite mutants from the wild-type genome (*78, 79*).

In the course of examining the *var1* region in a number of alleles, we discovered a total of five GC clusters (*36–38*), allowing us to examine some of the models for the function of these unusual sequences. Two of these clusters are invariant: the common GC cluster in the reading frame and a cluster 3′ to the reading frame that is homologous to a sequence in 15*S* rRNA. Three others, however, are optional; the *a* insert in the reading frame, and two clusters we have previously called *x* and *z*, that are 5′ and 3′, respectively, to the reading frame. Neither *z* nor probably *x* are allelic determinants. Our detection of optional GC clusters in the *var1* region complements the recent detection of other optional GC clusters on the yeast mitochondrial genome. These are GC clusters near *oli1* (compare *38* with *42*), within the 15*S* and 21*S* rRNA genes (*35, 45*), and within the reading frame of *cob* intron 5 (*75*).

Because the presence of specific GC clusters at specific positions appears to be strain dependent, whereas the location of coding sequences is not, it is increasingly difficult to sustain the idea that GC clusters play a significant role in the control of gene expression, since a necessary control element cannot be optionally present. In the case of RNA processing, furthermore, we have recently determined the origin of *var1* mRNA and found that it derives from a large polycistronic precursor that includes sequences from the upstream *oli1*

and *tRNA*^{Ser} genes. We identified at least six RNA processing sites in this precursor RNA and *all* are AT-rich (*80*).

One is left, then, with the question of what GC clusters do. With the exception of two of them in *var1* that encode amino acids, we don't know. However, it is difficult to escape the comparison between GC-rich *Hae* III and *Hpa* II clusters in *Saccharomyces* mitochondria (and possibly analogous *Pst* I clusters in *Neurospora*) (*81, 82*) with nuclear *Alu* I or *Kpn* I families in primates (*83, 84*), which may be species-specific promotors of genomic rearrangement. We have previously suggested that "jumping GC clusters" may be a normal feature of the yeast mitochondrial genome (*85*). If recombination takes place preferentially at such sites, either because it is site-specific or because such heteroduplexes would be particularly stable, then the movement of such clusters could be an important mechanism for generating the diversity of genome size and organization seen in fungal mitochondria.

References and Notes

1. B. Dujon, in *The Molecular Biology of the Yeast Saccharomyces: Life Cycle and Inheritance*, J. Strathern, E. Jones, J. Broach, Eds. (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1981), pp. 505–635.
2. D. C. Wallace, *Microbiol. Rev.* **46**, 208 (1982).
3. D. Clayton, *Annu. Rev. Biochem* **53**, 573 (1984).
4. I. G. Macreadie, C. E. Novitski, R. J. Maxwell, U. John, B.-G. Ooi, G. L. McMullen, H. B. Lukins, A. W. Linnane, P. Nagley, *Nucleic Acids Res.* **11**, 4435 (1983).
5. G. Macino and A. Tzagoloff, *J. Biol. Chem.* **254**, 4617 (1979); L. A. M. Hensgens *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **76**, 1663 (1979).
6. W. Sebald, J. Hoppe, E. Wachter, in *Function and Molecular Aspects of Biomembrane Transport*, E. Quagliariello, Ed. (North-Holland, Amsterdam, 1979), p. 63.
7. G. Turner, G. Imam, H. Kuntzel, *Eur. J. Biochem.* **97**, 565 (1979).
8. ATPase 9 has subsequently been shown to be present on *Neurospora crassa* mtDNA, although apparently as an inactive pseudogene, since the protein actually found in *N. crassa* mitochondria does not have the amino acid sequence predicted by the mitochondrial gene [(*5*) with (*9*) and (*10*)].
9. P. Van den Boogaart, J. Samallo, E. Agsteribbe, *Nature (London)* **298**, 187 (1982).
10. H. de Vries, J. C. de Jonge, A. Arnberg, A. A. C. M. Peijnenburg, E. Agsteribbe, in *Mitochondria 1983: Nucleo-Mitochondrial Interactions*, R. J. Schweyen, K. Wolf, F. Kaudewitz, Eds. (de Gruyter, Berlin, 1983), pp. 343–356.
11. P. Terpstra, E. Zanders, R. A. Butow, *J. Biol. Chem.* **254**, 12653 (1979).
12. A. M. Lambowitz, R. J. LaPolla, R. A. Collins, *J. Cell Biol.* **82**, 17 (1979).
13. R. Netzker, H. G. Kochel, N. Basak, H. Kuntzel, *Nucleic Acid Res.* **10**, 4783 (1982).
14. W. M. Ainley, I. Macreadie, R. A. Butow, *J. Mol. Biol.*, in press.
15. A. M. Lambowitz, N.-H. Chua, D. J. L. Luck, *ibid.* **107**, 223 (1976).
16. R. J. LaPolla and A. M. Lambowitz, *ibid.* **116**, 189 (1977).
17. P. Terpstra and R. A. Butow, *J. Biol. Chem.* **254**, 12662 (1979).
18. G. L. Schieber and T. W. O'Brien, in *Mitochondria 1983: Nucleo-Mitochondrial Interactions*, R. J. Schweyen, K. Wolf, F. Kaudewitz, Eds. (de Gruyter, Berlin, 1983), pp. 469–480.
19. The actual molecular weights of these var1 forms determined by DNA sequence analysis of the *var1* gene predict polypeptides somewhat larger than the sodium dodecyl sulfate–polyacrylamide gel electrophoresis estimates. Since we initially designated different var1 forms and alleles according to their apparent molecular weights (for example, *var1*[*40.0*], and *var1*[*44.0*]), we continue to do so here.
20. M. G. Douglas and R. A. Butow, *Proc. Natl. Acad. Sci. U.S.A.* **73**, 1083 (1976).
21. R. A. Butow, R. D. Vincent, R. L. Strausberg, E. Zanders, P. S. Perlman, in *Mitochondria 1977: Genetics and Biogenesis*, W. Bandlow, R. J. Schweyen, K. Wolf and F. Kaudewitz, Eds. (de Gruyter, Berlin, 1977), pp. 317–335.
22. R. A. Butow, R. L. Strausberg, R. D. Vincent, L. D. Paulson, P. S. Perlman, in *International Symposium on Pure and Applied Biochemistry of Yeasts*, M. Bacilla and B. Horecker, Eds. (Academic Press, New York, 1978), pp. 403–411.
23. G. Faye, *FEBS Lett.* **69**, 167 (1976); K. K. Maheshwari and S. Marzuki, *Biochem. Biophys. Acta* **781**, 153 (1984).
24. H. P. Zassenhaus and P. S. Perlman, *Curr. Genet.* **6**, 179 (1982).
25. M. Murphy, K.-B. Choo, I. MacReadie, S. Marzuki, H. B. Lukins, P. Nagley, A. W. Linnane, *Arch. Biochem. Biophys.* **203**, 260 (1980).
26. H. P. Zassenhaus and P. S. Perlman, in preparation.
27. C. L. Dieckmann, T. J. Koerner, A. Tzagoloff, *J. Biol. Chem.* **259**, 4722 (1984).
28. P. P. Muller, M. K. Reif, S. Zonghou, C. Sengstag, T. L. Mason, T. D. Fox, *J. Mol. Biol.*, **175**, 431 (1984).
29. P. S. Perlman, M. G. Douglas, R. L. Strausberg and R. A. Butow, *J. Mol. Biol.* **115**, 675 (1977).
30. R. L. Strausberg, R. D. Vincent, P. S. Perlman, R. A. Butow, *Nature (London)* **276**, 577 (1978).
31. R. D. Vincent, P. S. Perlman, R. L. Strausberg, R. A. Butow, *Curr. Genet.* **2**, 27 (1980).
32. R. L. Strausberg and R. A. Butow, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 494 (1981).
33. R. L. Strausberg and R. A. Butow, *ibid.* **74**, 2715 (1977).
34. I. C. Lopez, F. Farrelly, R. A. Butow, *J. Biol. Chem.* **256**, 6496 (1981).
35. B. Dujon, *Cell* **20**, 185 (1980).
36. M. E. S. Hudspeth, W. M. Ainley, D. S. Shumard, R. A. Butow, L. I. Grossman, *ibid.* **30**, 617 (1982).
37. H. P. Zassenhaus, F. Farrelly, M. E. S. Hudspeth, L. I. Grossman, R. A. Butow, *Mol. Cell. Biol.* **3**, 1615 (1983).
38. M. E. S. Hudspeth, R. D. Vincent, P. S. Perlman, D. S. Shumard, L. O. Treisman, L. I. Grossman, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 3148 (1984).
39. G. Bernardi, M. Faures, G. Piperno, P. P. Slonimski, *J. Mol. Biol.* **48**, 23 (1970).
40. A. Prunell and G. Bernardi, *ibid.* **86**, 825 (1974).
41. ——— *ibid.* **110**, 53 (1977).
42. G. Bernardi, *Trends Biochem. Sci.* **4**, 197 (1979).
43. A. Tzagoloff, M. Nobrega, A. Akai, G. Macino, *Curr. Genet.* **2**, 149 (1980).
44. F. Sor and H. Fukuhara, *Nucleic Acids Res.* **10**, 1625 (1982).
45. F. Sor and H. Fukuhara, in *Mitochondrial Genes*, P. Slonimski, P. Borst and G. Attardi, Eds. (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1982), pp. 255–262.
46. F. Sor and H. Fukuhara, *Nucleic Acids Res.* **11**, 339 (1983).
47. B. G. Barrell, A. T. Bankier, J. Drouin, *Nature (London)* **282**, 189 (1979).
48. M. H. L. de Brujin, *ibid.* **304**, 234 (1983).
49. M. Li and A. Tzagoloff, *Cell* **18**, 47 (1979).
50. K. S. Browning and U. L. RajBhandary, *J. Biol. Chem.* **257**, 5253 (1982).
51. J. Heckman, J. Sarnoff, B. Alzner-DeWeerd, S. Yin, U. L. RajBhandary, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 3159 (1980).
52. A. Prunell, H. Kopecka, F. Strauss, G. Bernardi, *J. Mol. Biol.* **110**, 17 (1977).
53. G. Fonty, R. Goursot, D. Wilkie, G. Bernardi, *ibid.* **119**, 213 (1978).
54. H. Blanc and B. Dujon, in *Mitochondrial Genes*, P. Slonimski, P. Borst, G. Attardi, Eds. (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1982), pp. 279–294.
55. L. Margulis, *Origin of Eukaryotic Cells* (Yale Univ. Press, New Haven, Conn., 1970).
56. F. Farrelly and R. A. Butow, *Nature (London)* **301**, 196 (1983).
57. G. Gellissen, J. Y. Bradfield, B. N. White, G. R. Wyatt, *ibid.*, p. 631.
58. R. M. Wright and D. J. Cummings, *ibid.* **302**, 86 (1983).
59. H. I. Hadler, B. Dimitrijevic and R. Mahalingam, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 6495 (1983).
60. H. T. Jacobs *et al.*, *J. Mol. Biol.* **165**, 509 (1983).
61. P. Borst, in *Autonomy and Biogenesis of Mitochondria and Chloroplasts*, N. K. Boardman, A. W. Linnane, R. M. Smillie, Eds. (North-Holland, Amsterdam, 1979), pp. 260–266.
62. B. Schmidt, B. Hennig, R. Zimmermann, W. Neupert, *J. Cell Biol.* **96**, 248 (1983).
63. R. Capaldi and G. Vanderkooi, *Proc. Natl. Acad. Sci. U.S.A.* **69**, 930 (1972).
64. T. D. Fox, *ibid.* **76**, 6534 (1979).
65. D. A. Clayton, J. N. Doda, E. C. Friedberg, *ibid.* **71**, 2777 (1974).
66. R. A. Lansman and D. A. Clayton, *J. Mol. Biol.* **99**, 761 (1975).
67. W. Gilbert, *Nature (London)* **271**, 501 (1978); J. E. Darnell, *Science* **202**, 1257 (1978).
68. All yeast mitochondrial DNA sequences available in version 1.1 of the EMBL nucleotide sequence data library were compared to the *var1* reading frame sequence (*36*). Matching was carried out on an IBM PC according to a routine that allowed specifying the length of a minimum match size. A match of minimum length, once found, was expanded to the largest possible match and the positions and length of the latter reported.
69. W. M. Ainsley, P. Hensley, R. A. Butow, *J. Biol. Chem.* **259**, 8422 (1984).
70. C. S. Craik, W. J. Rutter, R. Fletterick, *Science* **220**, 1125 (1983).
71. J. Lazowska, C. Jacq, P. P. Slonimski, *Cell* **22**, 333 (1980).
72. J. P. M. Sanders, C. Heyting, M. P. Vorbeet, F. C. P. W. Meijlink, P. Borst, *Mol. Gen. Genet.* **157**, 239 (1977).
73. R. Jarrell and P. S. Perlman, unpublished observations.
74. C. Jacq *et al.*, in *Mitochondrial Genes*, P. Slonimski, P. Borst, G. Attardi, Eds. (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1982), pp. 155–183.
75. A. Gargouri, J. Lazowska, P. P. Slonimski, in *Mitochondria 1983: Nucleo-Mitochondrial Interactions*, R. J. Schweyen, K. Wolf, F. Kaudewitz, Eds. (de Gruyter, Berlin, 1983), pp. 259–268.
76. G. Bernardi, in *Mitochondrial Genes*, P. Slonimski, P. Borst, G. Attardi, Eds. (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1982), pp. 269–278.
77. G. Baldacci and G. Bernardi, in *Mitochondria 1983: Nucleo-Mitochondrial Interactions*, R. J. Schweyen, K. Wolf, F. Kaudewitz, Eds. (de Gruyter, Berlin, 1983), pp. 57–64.
78. G. Faugeron-Fonty, F. Culard, G. Baldacci, R. Goursot, A. Prunell, G. Bernardi, *J. Mol. Biol.* **134**, 493 (1979).
79. C. Gaillard, F. Strauss, G. Bernardi, *Nature (London)* **283**, 218 (1980).
80. H. P. Zassenhaus, N. C. Martin and R. A. Butow, *J. Biol. Chem.* **259**, 6019 (1984).
81. S. Yin, J. Heckman, U. L. RajBhandary, *Cell* **26**, 325 (1981).
82. J. M. Burke and U. L. RajBhandary, *ibid.* **31**, 509 (1983).
83. P. Jagadeeswaran, B. G. Forget, S. M. Weissman, *ibid.* **26**, 141 (1981).
84. M. Singer, *Int. Rev. Cytol.* **76**, 67 (1982).
85. R. A. Butow, W. M. Ainley, H. P. Zassenhaus, M. E. Hudspeth, L. I. Grossman, in *Manipulation and Expression of Genes in Eukaryotes*, P. Nagley, A. W. Linnane, A. W. Peacock, J. A. Pateman, Eds. (Academic Press, Sydney, 1983), pp. 269–277.

86. We acknowledge with pleasure the members of our three laboratories who participated in this work and thank Robert B. Helling for comments on the manuscript and the National Institutes of Health (GM-26546 and GM-22525) and the Robert A. Welch Foundation for continued support.