Open Access

Long mRNAs coding for yeast mitochondrial proteins of prokaryotic origin preferentially localize to the vicinity of mitochondria

Julien Sylvestre, Stéphane Vialette, Marisol Corral Debrinski and Claude Jacq

Address: Laboratoire de Génétique Moléculaire, Ecole Normale Supérieure, 46 rue d'Ulm 75230 Paris cedex 05, France.

Correspondence: Claude Jacq. E-mail: Claude.Jacq@biologie.ens.fr.

Abstract

Background: Subcellular messenger RNA localization is important in most eukaryotic cells, even in unicellular organisms like yeast for which this process has been underestimated. Microarrays are rarely used to study subcellular mRNA localization at whole-genome level, but can be adapted to that purpose. This work focuses on studying the repartition of yeast nuclear transcripts encoding mitochondrial proteins between free cytosolic polysomes and polysomes bound to the mitochondrial outer membrane.

Results: Combining biochemical fractionations with oligonucleotide array analyses permits clustering of genes on the basis of the subcellular sites of their mRNA translation. A large fraction of yeast nuclear transcripts known to encode mitochondrial proteins is found in mitochondrial outer-membrane-bound fractions. These results confirm and extend a previous analysis conducted with partial genomic microarrays. Interesting statistical relations among mRNA localization, gene origin and mRNA lengths were found: longer and older mRNAs are more prone to be localized to the vicinity of mitochondria. These observations are included in a refined model of mitochondrial protein import.

Conclusions: Mitochondrial biogenesis requires concerted expression of the many genes whose products make up the organelle. In the absence of any clear transcriptional program, coordinated mRNA localization could be an important element of the time-course of organelle construction. We have built a 'MitoChip' localization database from our results which allows us to identify interesting genes whose mRNA localization might be essential for mitochondrial biogenesis in most eukaryotic cells. Moreover, many components of the experimental and data-analysis strategy implemented here are of general relevance in global transcription studies.

Background

Subcellular messenger RNA localization appears to be of great importance in a wide variety of biological contexts [1–4].

In this study, we focus on the subcellular localization of yeast nuclear mRNAs encoding mitochondrial proteins. We wish to show that the genome-wide approach used in this work is

both useful and adaptable to addressing global mRNA locali-
zation studies. In most eukaryotic cells, mitochondrial bio-
genesis relies on the expression of hundreds of nuclear genes
whose protein products have to be properly addressed to
mitochondrial compartments in synchrony with the biogen-
esis program. Seven of the eight major products encoded by
the yeast *Saccharomyces cerevisiae* mitochondrial DNA are
hydrophobic subunits of respiratory complexes in the mito-
chondrial inner membrane. The translation of the subunits
occurs on ribosomes bound to the mitochondrial inner mem-
brane, which implies membrane-bound nuclear-encoded
translation activators. Localization of the translation process
in the same cellular compartment ensures synthesis of mito-
chondrially encoded proteins near sites of their assembly into
multimeric respiratory complexes [5].

All other mitochondrial proteins are encoded by the nuclear
genome. The way these proteins are addressed to mitochon-
dria has been studied extensively for the past three decades.
Best known are the mitochondrial targeting sequence (mts),
which is an amino-terminal presequence, and its interactions
with the mitochondrial receptor-translocator TOM-TIM
(transporter of the outer membrane - transporter of the inner
membrane) [6,7]. The steps leading to the translated precur-
sor in its mature form, localized in the mitochondria, have
been fully analyzed by *in vitro* experiments [8,9]. Neverthe-
less, various observations suggest that this general post-
translational targeting mechanism could be preceded, *in
vivo*, by many processes. In particular, mRNA localization
could guide protein translation to the vicinity of mitochon-
dria. Thirty years ago, ribosome-binding sites were observed
on yeast mitochondria [10]. Additionally, the time lag
between the completion of translation and the beginning of
the import process is very short, which suggests that the two
processes might be tightly connected [11].

More recently, several studies in human and in yeast cells
presented compelling evidence that some mRNAs are local-
ized close to mitochondria *in vivo* [12,13]. In previous studies
we used microarray analyses to explore the mRNA targeting
process [14]. These analyses, which concerned only half of the
mitochondrial 'mRNA localizome', showed that a proportion
of nuclear-encoded mitochondrial transcripts are addressed
to the vicinity of the mitochondrion. They are translated on
polysomes associated with the outer mitochondrial mem-
brane, which, in a way, is reminiscent of the case of the seven
mRNAs encoded by the mitochondrial genome and translated
on the mitochondrial inner membrane. In the study reported
here we carried out a more complete analysis based on an
improved purification protocol; this explored all the yeast
genes. These experiments provided results for 95% of nuclear
genes known to encode a mitochondrial product. This more
complete dataset revealed interesting statistical relations
involving mRNA localization. These relations and the gener-
ality of our micropurification/microarray (MPMA) approach
are discussed.

Results and discussion
Subcellular fractionation
We adapted an improved micropurification procedure to
obtain enriched mitochondrial fractions. The purity of these
fractions in terms of presence or absence of mitochondria was
routinely assessed by northern blots or by reverse tran-
scriptase (RT)-PCR analysis using genes for mitochondrial,
cytosolic and plasma membrane proteins as molecular probes
(see Additional data files). However, the mitochondrial frac-
tions thus obtained always included mRNAs from other intra-
cellular membranes (data not shown). This is not surprising,
as mitochondria and endoplasmic reticulum are known to be
highly interconnected in *S. cerevisiae* [15]. This was taken
into account in the following analyses. After growth in galac-
tose and RNA extraction, the amplified samples were treated
according to the manufacturer's instructions. To assess
mRNA quality, a quality-control was included in the protocol
before amplification. Figure 1 summarizes the biological
question and the experimental strategy.

Microarray data analysis strategy
A major difficulty in studying mRNA localization with micro-
arrays is microarray normalization. All current methods of
normalizing microarray data were developed to allow
between-chip comparison in transcriptome expression exper-
iments [16,17]. The hypotheses they rely on - the constant
expression levels of most genes across experiments, for
instance - do not hold true when different subcellular mRNA
populations are compared. In a previous experiment, Cy5/
Cy3 ratios were ranked into percentiles and averaged to com-
pare different experiments [14]. Diehn *et al.* [18] adapted a
moving-average analysis to study membrane-bound polys-
omes. In both cases, the method of combining results from
different experiments was not straightforward and the ques-
tion of the reproducibility was not readily addressed.

In this study, a more adequate experimental design and a
more complete dataset allowed us to design a more direct
strategy, which is statistically more satisfactory. Three mRNA
populations were isolated and hybridized to distinct chips:
MP (mitochondrial and membrane bound polysomes); FP
(free cytosolic polysomes); and TOT (total polysomes, that is,
the whole transcriptome). This experiment was repeated
twice, giving six results for each probe set. Pairwise quantile
normalization was then applied (MP_1 and MP_2; FP_1 and FP_2;
TOT_1 and TOT_2). The expression levels were then calculated
(see Materials and methods for details). Spike mRNAs
present on Affymetrix gene chips were used to check whether
an additional normalization step was necessary to allow com-
parisons among the three normalized couples. As these exog-
enous control mRNAs, which were introduced before RNA
amplification and chip hybridization, gave almost identical
signals for the six chips, no further correction was applied
[19]. Of the probe sets 3% had at least one expression value
under an arbitrary threshold. They were eliminated from sub-
sequent analysis, to filter against uncertain extremely low
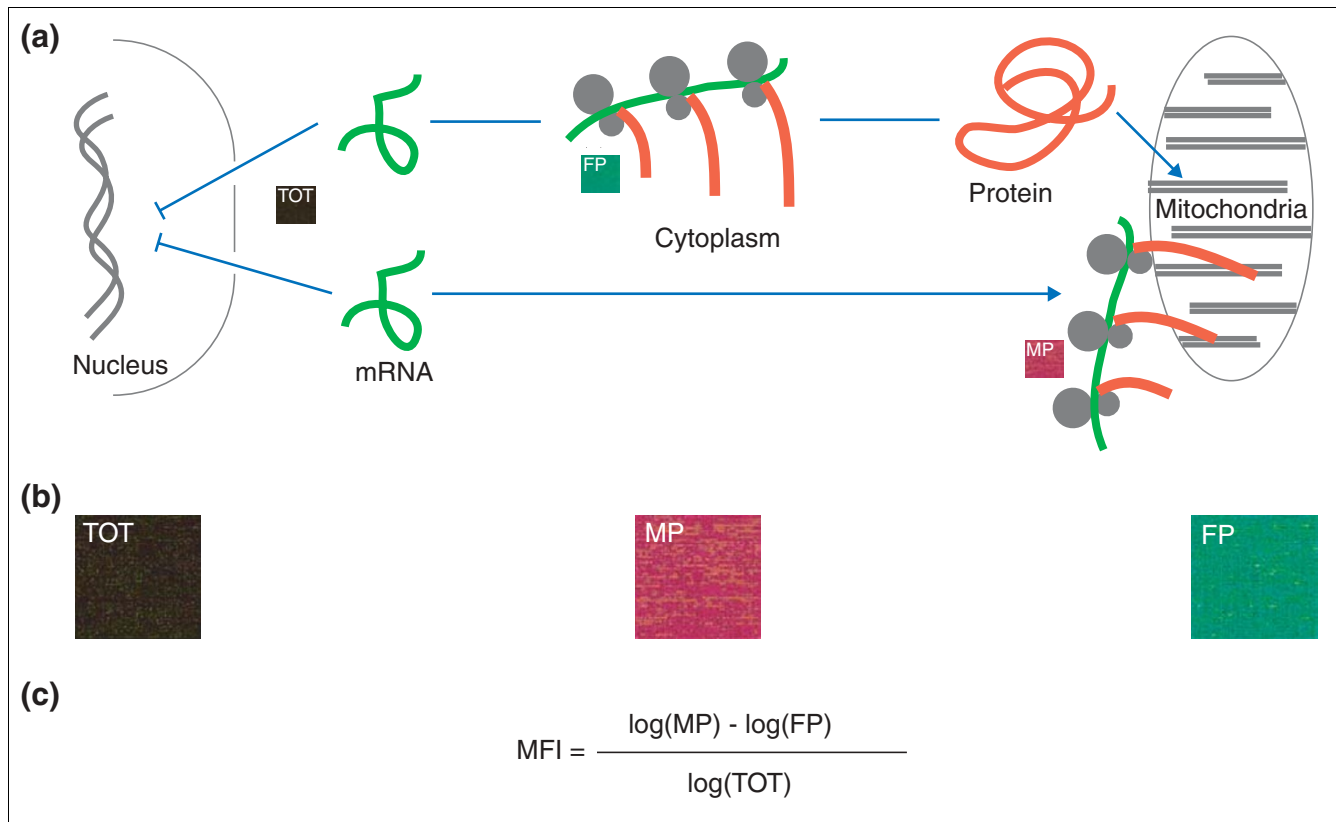
Figure 1
A microarray analysis to study the subcellular localization of mRNAs encoding mitochondrial proteins. **(a)** After cycloheximide blocking of translation, polysomes associated with mitochondria (MP) and free cytoplasmic polysomes (FP) were purified and the corresponding mRNA species were purified and treated for analysis **(b)** using high-density oligonucleotide microarrays. The total mRNA population from spheroplasts (TOT) was extracted in parallel with the two other fractions and used as a reference. **(c)** After normalization and signal treatment, a synthetic localization value called MFI (membrane versus free index) was calculated as indicated for each mRNA. Ninety-seven percent of the probe sets, corresponding to 96% of the genes and 94% of the mitochondrial genes, have MFI values assigned to them.

intensities. A localization parameter called MFI (for membrane versus free index) was defined for each triplet of chips (MP_1, FP_1, TOT_1 and MP_2, FP_2, TOT_2) and each transcript by the formula:

$$MFI_i = (\log(MP_i) - \log(FP_i))/\log(TOT_i) \text{ where } (i = 1 \text{ or } 2).$$

The MFI_1 and MFI_2 values were very close, indicating good reproducibility of the whole process. Hence, the MFI_1 and MFI_2 values were averaged. The methodology described above, in particular the normalization procedure and the utilization of the total mRNA population (TOT), can be easily generalized to efficiently analyze larger-scale experiments involving more subcellular mRNA populations and more chips.

Validation
The MFI values were compared to the 'Mitochondrial Localization of RNA' (MLR) values previously obtained from a different experimental approach. This comparison was limited to the available MLR data as the previous analysis covered only about half the genome. Despite quite different experimental designs, the comparison resulted in good correlation values, thereby validating our new approach. More precisely, previous localization values were calculated using data from one to six independent experiments. Figure 2 is a plot of the rank correlation coefficient between new values (MFI) and old values (MLR) as a function of the number of experiments included in the former. The graph includes 95% confidence intervals calculated with non-parametric bootstrap [20]. The correlation increases quite sharply from one to four experiments, after which the curve is almost flat, with a correlation coefficient between ranks of above 0.6 from four to six experiments. This clearly demonstrates that MFIs more accurately quantify the localization than the previous MLRs.

Many mitochondrial proteins have relatively high MFIs
The localization results for 97% of the probe sets, corresponding to 96% of the genes (95% of the genes encoding mitochondrial products) are available online at [20]. Individual results or lists of results can be retrieved by ORF name, gene name or Affymetrix probe set identifier (affxid). To facilitate data
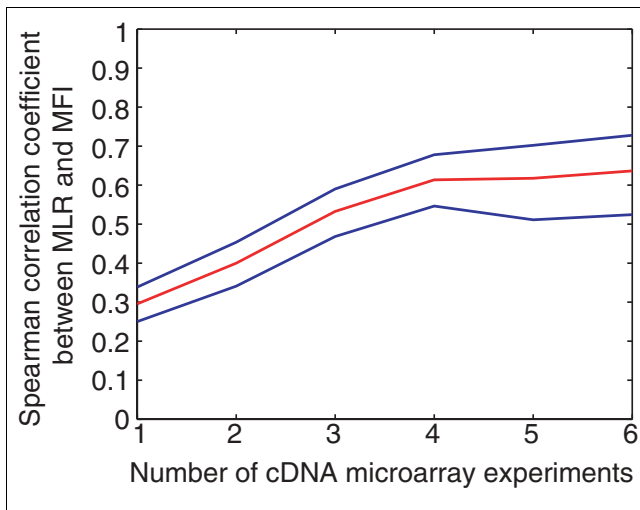
Figure 2
Comparison of the high-density oligonucleotide analyses (MFI) with the previous cDNA microarray analyses (MLR). A Spearman rank correlation coefficient between the MFI values obtained in this work and the previous MLR values from previous experiments was established as a function of the number of MLR experiments. The MFI experiments, which are more accurate, provide results for more than 95% of the yeast genes. Number of genes for which the MLR was calculated: one experiment, 1,268; two experiments, 632; three experiments, 412; four experiments, 247; five experiments, 110; six experiments, 94.

interpretation, MFIs were grouped in 10 evenly divided gene classes. For instance, a mitochondrial gene belonging to MFI class 10 has its MFI among the genome's 10% highest, and hence its mRNA is much enriched in mitochondrial outer-membrane-bound polysomes. Figure 3 presents the genome-wide distribution of MFI values and highlights two lists of mitochondrial genes with extreme MFI values. More precisely, the mitochondrial genes listed on the right of the plot have MFI values among whole-genome highest 5% (the upper half of MFI class 10) and correspond to mitochondria-localized transcripts. In contrast, the list of mitochondrial genes on the left of the plot have MFI values among whole-genome lowest 5% (the lower half of MFI class 1) and correspond to transcripts that are translated on free cytosolic polysomes. The subcellular localization of many yeast proteins is reasonably well known. We used the public database MIPS CYGD [21] as a reference. We then plotted the empirical cumulative distribution functions of the MFIs for mRNAs belonging to three separate known subcellular protein localization categories (Figure 4). As expected, genes encoding cytosolic proteins have low MFIs, whereas genes encoding plasma membrane and endoplasmic reticulum proteins have higher MFIs. If mitochondrial proteins were all translated from free cytosolic polysomes, one would expect the corresponding curve to be quite similar to that of cytosolic proteins. In contrast, the curve for mRNAs encoding mitochondrial proteins is between the two reference curves for mRNAs translated on cytosolic and on membrane-bound polysome fractions, respectively. This is the most provocative aspect of these data

because it shows that a large proportion of the mRNAs coding for mitochondrial proteins is actually localized to the vicinity of mitochondria. Thus, what had been previously observed by various experimental approaches is confirmed and quantified at a genome-wide level.

A critical point in such an approach is the procedure used to purify the mitochondria. The basic protocol used for isolation of mitochondria-associated polysomes is very similar to the previous procedure published by Butow's group [10]. However, we previously observed [14] that the linear sucrose gradient step was not necessary to obtain reliable results. Accordingly our results are consistent with those previously obtained by Butow's group on a limited set of genes. However it should be emphasized that this purification protocol does not lead to pure mitochondrial fractions. Considerable contamination by diverse membrane fractions is generally considered unavoidable; nevertheless these contaminants do not affect the general conclusion drawn from the results of Figure 3: the nuclear genes known to code for mitochondrially localized products are represented in the two extreme values of the MFI curve.

Common features of mRNAs that are preferentially translocated to the vicinity of mitochondria
Our extensive global study shows that more than 88 genes coding for mitochondrial products exhibit an MFI value over 0.15. This means that the corresponding mRNA are, in normal growth conditions in galactose, more prone to be in found in polysomes associated with mitochondria. We addressed the specific features shared by this important mRNA class. It should be kept in mind that for the six cases we examined, the 3' untranslated region (3' UTR) sequence of these mRNAs contains the mitochondrial addressing information that can target, *in vivo*, a tagged version of this RNA fragment [14]. Significantly, this *in vivo* targeting process did not require any translational activity as it was observed with non-translatable RNAs produced by RNA polymerase III. If this observation might be extended to the whole class of mRNAs that are translocated to the vicinity of mitochondria, it does not explain the *raison d'être* of this targeting process. To address this question we systematically analyzed the influence of several other parameters on mRNA localization (codon bias, codon adaptation index, amino-acid content, presence and length of the mitochondrial presequence as predicted by mitoprot or targetP, intra-mitochondrial protein localization, 3' UTR M-FOLD structure, and so on). Two characteristics - the phylogenetic origin and the length of the genes - turned out to correlate with the mRNA localization properties.

By making batches of alignments, Karlberg *et al.* [22] classified the yeast mitochondrial proteome on the basis of each gene's evolutionary origin. The main categories they considered were: ABE (yeast genes having at least one homolog in Archaea, one homolog in Bacteria and one homolog in the
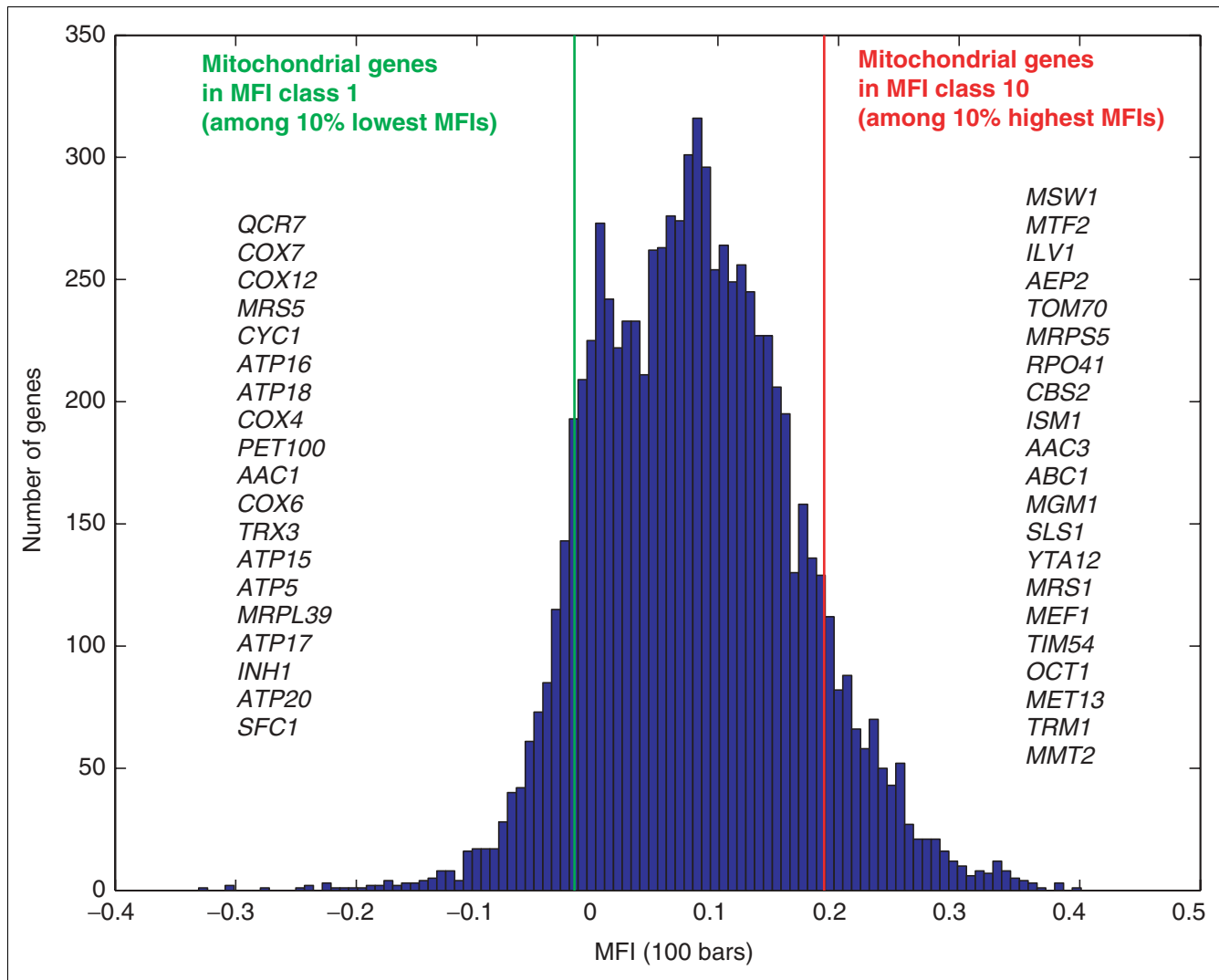
Figure 3
Whole-genome MFI distribution. Individual results for most of the yeast genes are available at [29]. As examples, genes which are known to code for mitochondrially localized products and belonging to the two extreme 5% MFI values are listed; thus the mRNAs of genes named on the right are translated to the vicinity of mitochondria.

eukaryotes); BE (at least one homolog in both a bacterium and a eukaryote); E (at least one homolog in a eukaryote); and SC (genes specific to *S. cerevisiae*). They concluded that the yeast mitochondrial proteome has a dual origin, around half of it having originated from the endosymbiont genome, while the other part evolved from the nuclear genome.

Our previous study had shown that there was indeed a link between the eukaryotic/prokaryotic origin of a mitochondrial gene and the localization of the corresponding transcript. We found a statistically significant relation when we applied a one-way analysis of variance (ANOVA) to explain 358 MFI values by four categorical evolutionary origin values. To further characterize this relation, a multiple comparison among the four groups was done. Figure 5 highlights the results of

that more detailed analysis, with 99% confidence intervals for each category's mean MFI. This reveals significant and ordered differences, with (ABE, BE) having more mitochondrial-membrane-localized translation than (E, SC). Although the results are not amenable to experimental test, this statistical observation might help in better understanding the cooperation between the nuclear and the organellar genomes. Interestingly, like the few remaining mitochondrially encoded genes translated near the inner membrane, the nuclear genes coming from the mitochondrial genome tend to be translated in the vicinity of the outer mitochondrial membrane.

Concerning the length of the proteins, which in yeast is closely connected to the length of the mRNA, we found a
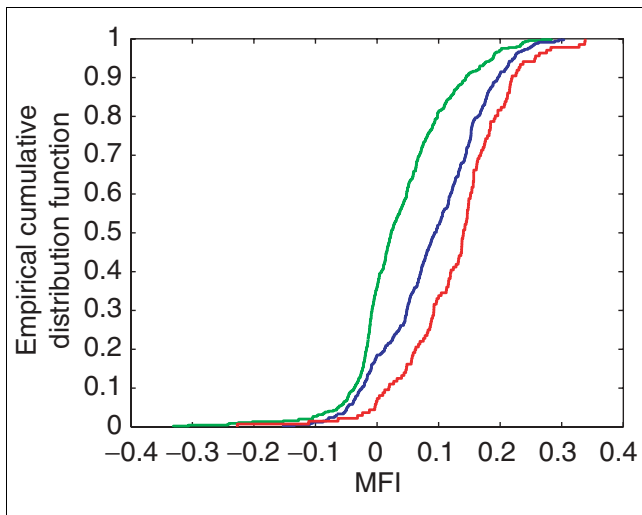
Figure 4
Cumulative distribution of MFIs and protein localization. Empirical cumulative distribution functions of MFIs for mRNAs encoding proteins localized in three kinds of subcellular compartments. The left (green) and right (red) curves correspond to reference populations of 388 cytosolic and 245 plasma membrane/endoplasmic reticulum localized proteins, respectively (corresponding mRNAs are known to be translated on free cytosolic and membrane-bound polysomes). The 358 mitochondrial proteins are in the curve in the middle (blue), their mRNA localization being between that of the two reference populations.

surprising statistical relation between mRNA length and mRNA localization; longer mRNAs tend to be localized to the vicinity of mitochondria (correlation coefficient: 0.67; see Figure 6).

George *et al.* [23] recently proposed a molecular mechanism and a model for the coupling of translation and translocation of mitochondrial mRNAs *in vivo*. Their model is based on the experimental observation that the nascent-polypeptide-associated complex (NAC) promotes interaction of the ribosome with the mitochondrial surface *in vivo*. Lithgow's two-step model proposes that in an early phase most mitochondrial mRNAs begin to be translated in the cytosol, whereas in the late phase, when a ribosome eventually encounters the surface of a mitochondrion, the mRNA molecule remains 'stuck' to the mitochondrion until its translation is completed. On average, longer mRNAs have more ribosome-NAC complexes, which results in a higher affinity for mitochondrial outer membrane. Interestingly, in a quite different context, flock-house virus mRNA polymerase is translated on *Drosophila* mitochondria while its amino-terminal sequence is inserted into the mitochondrial membrane [24]. That example shows that the existence of a mitochondrial targeting sequence (mts) and a mitochondria-localized mRNA is not contradictory. In addition to the coordinated action of NAC and mts, *in vivo* RNA targeting analyses have shown that the 3' UTR contains sufficient information to address the corresponding RNA to the vicinity of mitochondria [14].

This means that several mechanisms can control the mRNA targeting process. Some of these are essential; for instance, we recently obtained direct evidence for the genes *ATP2* and *ATM1* that the 3' UTR sequence is necessary for functional mitochondria. This suggests that, in the case of these two genes, the mRNA targeting process linked to the 3' UTR signal is required for mitochondrial biogenesis [25]. In addition, the initiation of translation of many mRNAs is, at least in part, controlled by their 3' UTR structure. Thus, early translation *cis* (mts, 3' UTR) and *trans* (NAC) factors act synergistically to control the targeting of many mRNAs coding for mitochondrial products. It has recently been demonstrated that ribosomes can 'sense' features of nascent polypeptide chains and modulate translation accordingly. Our results suggest that the mts and the 3' UTR could be such features that control mRNA translation in connection with the mitochondrial import process [26].

What could be the role of the mRNA targeting process in the biogenesis of mitochondria? Our favourite hypothesis is to imagine that it might be part of the control of the time course of production of mitochondrial proteins. One can reasonably propose that assembly of large mitochondrial complexes requires a temporally controlled import of the different subunits. Our observation would imply that the largest subunits are more likely to be the basic elements of the architecture of
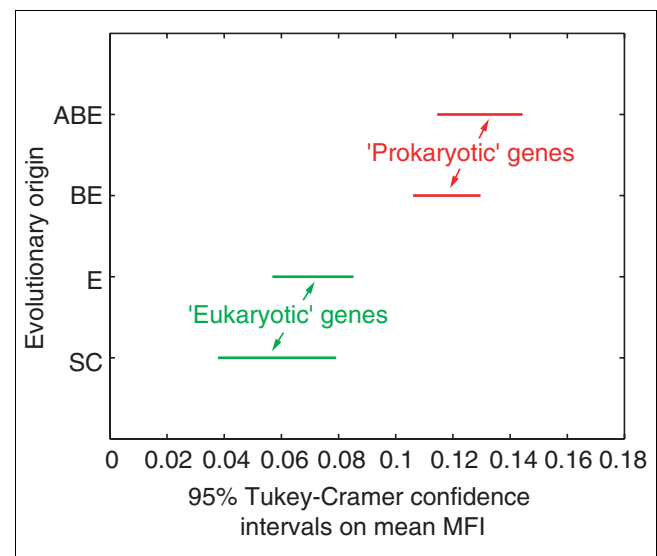

Figure 5
Relationship between phylogenetic categories of genes coding for mitochondrial proteins and their MFI values. Mitochondrial proteins were previously grouped into four classes corresponding to: SC, 56 genes specific to *S. cerevisiae*; E, 115 genes with homologs in eucaryotes exclusively; BE, 80 genes with homologs in bacteria and eucaryotes; ABE, 100 genes with homologs in archaea, bacteria and eucaryotes. Genes of prokaryotic origin are in red, and those of eukaryotic origin in green. Bars are Tukey-Cramer 99% confidence intervals. This graph shows a very significant effect of the prokaryotic/eukaryotic origin of genes on the MFI of their mRNAs.
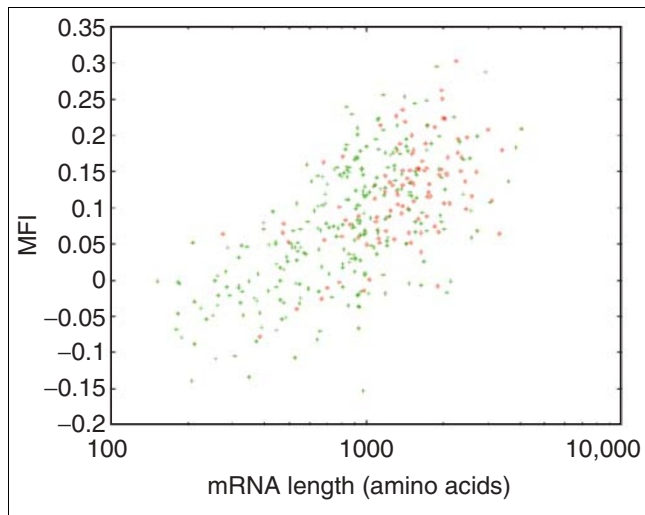
Figure 6
Mitochondrial mRNA localization and mRNA length. The MFI values of 358 genes known to have mitochondrially localized products are plotted against the corresponding mRNA length (semi-log scale). The Pearson correlation coefficient between MFIs and log mRNA length is 0.67. The red dots correspond to genes of prokaryotic origin (ABE in Figure 5). The higher number of nascent polypeptide-chain-associated complexes (NAC) found on the longer mRNAs could attract the corresponding polysomes to the vicinity of mitochondria (see text for details).

the mitochondrial complexes. This assumption should be amenable to experimental assessment. This mitochondrial mRNA targeting process is not specific to *S. cerevisiae*. Several studies in various human cells have shown that some specific mRNAs can be found next to mitochondria [27,28], which strongly suggests that our observation can be applied to many organisms.

Conclusions
This study shows that a combination of subcellular fractionation, high-density oligonucleotide arrays and an appropriate data-analysis strategy can reveal subcellular mRNA localization at a genome-wide level. Such approaches, which rely on biochemical purification, require rigorous independent controls, as it can be problematic to distinguish between mRNAs associated with mitochondria and with other cellular membranes. In the case of yeast mitochondrial fractions, which always co-purified with endoplasmic reticulum membranes, we checked, by *in vivo* experiments, that several mRNAs coding for mitochondrially localized proteins are actually addressed to the vicinity of mitochondria [12]. The global analysis presented here, which is based on oligonucleotide arrays, describes two subcellular localizations for mRNAs from which mitochondrially localized proteins are translated. Such a dual localization is clearly independent of the biochemical contaminants of mitochondrial fractions. This confirms and extends our previous study [14] showing that mRNAs of prokaryotic origin are preferentially localized to

the vicinity of mitochondria. Moreover, this new analysis shows an unexpected relationship between mRNA length and localization. It looks as if the primitive mRNA mitochondrial targeting process has been predominantly conserved for longer mRNAs, a feature that might be connected with the mitochondrial biogenesis program. The 'MitoChip' database we compiled, and which is available online ([29] and Additional data files), should help in designing new experiments, on a gene-by-gene basis, to test and refine this model. We are currently generalizing this kind of mRNA localization analysis to other types of cells and subcellular fractions.

Materials and methods
Cell culture and fractionation
Saccharomyces cerevisiae strain CW04 (isonuclear to W303-1B: *MATa, leu2-3, ura3-1, trp1-1, ade1-2, his3-11, can1-100*) was used. Yeast was grown in a galactose-rich medium (1% bacto-yeast extract, 1% bactopeptone, 0.1% KH_2PO_4, 0.12% $(NH_4)_2SO_4$, 2% galactose).

Mitochondria-bound polysomes and free cytoplasmic polysomes were purified as described by Ades and Butow [30], with few modifications. To convert cells to spheroplasts, they were treated for 60 min with Quantazyme (1,000 U per g per wet weight cells) (Quantum Appligene, Illkirch, France) at 30°C with shaking. Spheroplasts were allowed to regenerate for 3 h with gentle shaking at 28°C in 1 M sorbitol, 1% bactopeptone, 1% yeast extract, 0.1% KH_2PO_4, 0.12% $(NH4)_2SO_4$ and 2% galacotose. Regeneration was stopped by adding cycloheximide to a concentration of 300 μ/ml and 20 min later the culture was poured over frozen 1 M sorbitol, 200 μ/ml cycloheximide. The spheroplasts were harvested and washed with 1 M sorbitol, 200 μ/ml cycloheximide. One-twentieth of the spheroplast volume was frozen at -80°C. The rest of the spheroplasts were homogenized in 6 ml of the following breaking buffer: 0.6 M mannitol, 30 mM Tris-HCl pH 7.4, 100 mM KCl, 5 mM magnesium acetate, 5 mM 2-mercaptoethanol, 200 μ/ml cycloheximide, 2,000 U/g yeast protease inhibitor cocktail (Sigma) 10,000 units RNAsin ribonuclease inhibitor (Promega, Madison, WI) and 500 μ/ml heparin. Cells were disrupted with 15 strokes using a glass Teflon pestle. The homogenate was centrifuged at 4,500 rpm for 10 min; the pellet was once again disrupted and centrifuged. Both homogenates were combined and centrifuged twice to eliminate unbroken cells and nuclei. The resulting supernatant was further centrifuged at 11,000 rpm for 30 min. The pellet of crude mitochondria associated with polysomes was washed twice in mannitol-polysome buffer before storage at -80°C. Free cytoplasmic polysomes were purified from the post-mitochondrial supernatant by sedimentation through a step gradient of 2 M and 0.5 M sucrose, each of these solutions prepared in 30 mM Tris-HCl pH 7.4, 100 mM KCl, 5 mM magnesium acetate, 5 mM 2-mercaptoethanol, 200 μ/ml cycloheximide, 600 U/g yeast protease inhibitor cocktail (Sigma) 2,500 units RNAsin ribonuclease inhibitor

(Promega) and 500 μ/ml heparin. The gradients were centrifuged at 40,000 rpm for 20 h at 4°C using the TST41 Ti rotor. The pellet, containing free cytoplasmic polysomes, was used for RNA extraction.

RNA was extracted from spheroplasts, mitochondria-bound polysomes and free cytoplasmic polysomes using the RNeasy Protect MIDI kit (Qiagen, Valencia, CA).

Oligonucleotide array hybridizations and data analysis
Detailed information on this aspect of the experiments can be found in Additional data files and at [29] All the raw-values and pictures (.cel and .dat files) are available for download. For each of the six arrays, mRNA quality was controlled (Agilent Bioanalyzer; Agilent, Palo Alto, CA) before hybridization. Data analysis was carried out using the Affy Package (L. Gautier and R. Izarry, unpublished), which is part of Bioconductor R bundle. Classical normalization procedures, built on the assumption that most genes expression values do not vary across arrays, could not be applied for all six arrays. Therefore, arrays were normalized on a pairwise basis using the Bolstad method: MP_1 with MP_2; FP_1 with FP_2 and TOT_1 with TOT_2. The Bolstad method averages independently perfect match and mismatch signal intensity values among chips at each quantile level, which reshapes the entire intensity distributions to make them comparable. Quantile normalization tends to align intensity boxplots and flatten MvA plots. We then checked whether spike mRNAs gave similar results for the six chips; as this was almost the case, no further correction was applied. Expression values were then computed using Affymetrix's playerout algorithm. Genes (3% of total) that had at least one of the six intensity values under an arbitrary threshold were discarded from further analysis. Interestingly, when conducted with MAS5.0, the analysis gave the same kinds of results, but with larger variability.

Link between localization and evolution categories
We used standard one-way analysis of variance. Multiple comparison analysis was then applied using the Tukey-Kramer test with $\alpha = 1\%$. This test is known to be somewhat conservative for unequally sized subpopulations. The Bonferroni test gave the same results (ABE, BE) > (E, SC).

Additional data files
Additional data files available with the online version of this paper and from [29] comprise supplementary figures showing: the effect of quantile normalization on MvA plots and boxplots of log-intensities (Additional data file 1); log-intensities for spike mRNAs (Additional data file 2); MFI for replicate 1 (chips MP1, FP1, TOT1) versus MFI for replicate 2 (chips MP2, FP2, TOT2) (Additional data file 3); ANOVA results (Additional data files 4 and 5); comparison of the length of genes coding for mitochondrial proteins and cytoplasmic proteins in yeast (Additional data file 6); analysis of the quality of RNA purified from the two polysomal preparations examined (Additional data file 7); figure legends and titles (Additional data file 8), and a full dump of the online database (Additional data files 9 and 10). Each row of the database is composed of the following columns: Gene Name, ORF, Affymetrix Id, MFI Value and MIPS Localization.

References
1. Palacios IM, Johnston DS: **Getting the message across: the intracellular localization of mRNAs in higher eukaryotes.** *Annu Rev Cell Dev Biol* 2001, **17:**569-614.
2. Chartrand P, Singer RH, Long RM: **RNP localization and transport in yeast.** *Annu Rev Cell Dev Biol* 2001, **17:**297-310.
3. Okita TW, Choi SB: **mRNA localization in plants: targeting to the cell's cortical region and beyond.** *Curr Opin Plant Biol* 2002, **5:**553.
4. Steward O: **mRNA at synapses, synaptic plasticity, and memory consolidation.** *Neuron* 2002, **36:**338-340.
5. Sanchirico ME, Fox TD, Mason TL: **Accumulation of mitochondrially synthesized *Saccharomyces cerevisiae* Cox2p and Cox3p depends on targeting information in untranslated portions of their mRNAs.** *EMBO J* 1998, **17:**5796-5804.
6. Schatz G: **Protein transport: The doors to organelles.** *Nature* 1998, **395:**439-440.
7. Roise D, Schatz G: **Mitochondrial presequences.** *J Biol Chem* 1988, **263:**4509-4511.
8. Neupert W, Brunner M: **The protein import motor of mitochondria.** *Nat Rev Mol Cell Biol* 2002, **3:**555-565.
9. Pfanner N: **Protein sorting: recognizing mitochondrial presequences.** *Curr Biol* 2000, **10:**R412-R415.
10. Kellems RE, Butow RA: **Cytoplasmic-type 80 S ribosomes associated with yeast mitochondria. I. Evidence for ribosome binding sites on yeast mitochondria.** *J Biol Chem* 1972, **247:**8043-8050.
11. Verner K: **Co-translational protein import into mitochondria: an alternative view.** *Trends Biochem Sci* 1993, **18:**366-371.
12. Corral-Debrinski M, Blugeon C, Jacq C: **In yeast, the 3' untranslated region or the presequence of ATM1 is required for the exclusive localization of its mRNA to the vicinity of mitochondria.** *Mol Cell Biol* 2000, **20:**7881-7892.
13. Ricart J, Izquierdo JM, Di Liegro CM, Cuezva JM: **Assembly of the ribonucleoprotein complex containing the mRNA of the beta-subunit of the mitochondrial H+-ATP synthase requires the participation of two distal *cis*-acting elements and a complex set of cellular *trans*-acting proteins.** *Biochem J* 2002, **365:**417-428.
14. Marc P, Margeot A, Devaux F, Blugeon C, Corral-Debrinski M, Jacq C: **Genome-wide analysis of mRNAs targeted to yeast mitochondria.** *EMBO Rep* 2002, **3:**159-164.
15. Achleitner G, Gaigg B, Krasser A, Kainersdorfer E, Kohlwein SD, Perktold A, Zellnig G, Daum G: **Association between the endoplasmic reticulum and mitochondria of yeast facilitates interorganelle transport of phospholipids through membrane contact.** *Eur J Biochem* 1999, **264:**545-553.
16. Hoffmann R, Seidl T, Dugas M: **Profound effect of normalization on detection of differentially expressed genes in oligonucleotide microarray data analysis.** *Genome Biol* 2002, **3:**research0033.1-0033.11.
17. Schadt EE, Li C, Ellis B, Wong WH: **Feature extraction and normalization algorithms for high-density oligonucleotide gene expression array data.** *J Cell Biochem Suppl* 2001, **Suppl 37:**120-125.
18. Diehn M, Eisen MB, Botstein D, Brown PO: **Large-scale identification of secreted and membrane-associated gene products using DNA microarrays.** *Nat Genet* 2000, **25:**58-62.
19. Hill AA, Brown EL, Whitley MZ, Tucker-Kellogg G, Hunter CP, Slonim DK: **Evaluation of normalization procedures for**

oligonucleotide array data based on spiked cRNA controls. *Genome Biol* 2001, **2:**research0055.1-0055.13.

20.　Tibshirani RJ: *An Introduction to the Bootstrap.* London, UK: Chapman & Hall; 1993.

21.　Mewes HW, Frishman D, Guldener U, Mannhaupt G, Mayer K, Mokrejs M, Morgenstern B, Munsterkotter M, Rudd S, Weil B: **MIPS: a database for genomes and protein sequences.** *Nucleic Acids Res* 2002, **30:**31-34.

22.　Karlberg O, Canback B, Kurland CG, Andersson SG: **The dual origin of the yeast mitochondrial proteome.** *Yeast* 2000, **17:**170-187.

23.　George R, Walsh P, Beddoe T, Lithgow T: **The nascent polypeptide-associated complex (NAC) promotes interaction of ribosomes with the mitochondrial surface *in vivo*.** *FEBS Lett* 2002, **516:**213-216.

24.　Miller DJ, Ahlquist P: **Flock house virus RNA polymerase is a transmembrane protein with amino-terminal sequences sufficient for mitochondrial localization and membrane insertion.** *J Virol* 2002, **76:**9856-9867.

25.　Margeot A, Blugeon C, Sylvestre J, Vialette S, Jacq C, Corral-Debrinski M: **In *Saccharomyces cerevisae*, ATP2 mRNA sorting to the vicinity of mitochondria is essential for respiratory function.** *EMBO J* 2002, **21:**6893-6904.

26.　Gong F, Yanofsky C: **Instruction of translating ribosome by nascent peptide.** *Science* 2002, **297:**1864-1867.

27.　Rings EH, Buller HA, de Boer PA, Grand RJ, Montgomery RK, Lamers WH, Charles R, Moorman AF: **Messenger RNA sorting in enterocytes. Co-localization with encoded proteins.** *FEBS Lett* 1992, **300:**183-187.

28.　Heinrich PC, Schmelzer E, Northemann W, Kaiser C, Witt I: **Rat liver cytochrome c oxidase subunits IV and V: cell-free synthesis as larger molecular weight precursors, mRNA sizes and sites of synthesis.** *Prog Clin Biol Res* 1982, **102:**149-159.

29.　**LGM MitoChip Project** [http://www.biologie.ens.fr/lgmgml/Mito]

30.　Ades IZ, Butow RA: **The products of mitochondria-bound cytoplasmic polysomes in yeast.** *J Biol Chem* 1980, **255:**9918-9924.

comment

reviews

reports

deposited research

refereed research

interactions

information