Dicistronic tRNA–5S rRNA genes in *Yarrowia lipolytica*: an alternative TFIIIA-independent way for expression of 5S rRNA genes

Joël Acker[1], Christophe Ozanne[2], Rym Kachouri-Lafond[3], Claude Gaillardin[2], Cécile Neuvéglise[2] and Christian Marck[1,*]

[1]Saclay Biology and Technologies Institute (iBiTec-S), CEA, 91191 Gif-sur-Yvette CEDEX, [2]INRA UMR1238, CNRS 2585, AgroParisTech, Microbiology and Molecular Genetics, BP 01, 78850 Thiverval-Grignon and [3]Architecture et Réactivité de l'ARN, Strasbourg Louis Pasteur University , CNRS, IBMC, 15 rue René Descartes, 67084 Strasbourg, France

ABSTRACT

In eukaryotes, genes transcribed by RNA polymerase III (Pol III) carry their own internal promoters and as such, are transcribed as individual units. Indeed, a very few cases of dicistronic Pol III genes are yet known. In contrast to other hemiascomycetes, 5S rRNA genes of *Yarrowia lipolytica* are not embedded into the tandemly repeated rDNA units, but appear scattered throughout the genome. We report here an unprecedented genomic organization: 48 over the 108 copies of the 5S rRNA genes are located 3′ of tRNA genes. We show that these peculiar tRNA–5S rRNA dicistronic genes are expressed *in vitro* and *in vivo* as Pol III transcriptional fusions without the need of the 5S rRNA gene-specific factor TFIIIA, the deletion of which displays a viable phenotype. We also report the existence of a novel putative non-coding Pol III RNA of unknown function about 70 nucleotide-long (RUF70), the 13 genes of which are devoid of internal Pol III promoters and located 3′ of the 13 copies of the tDNA-Trp (CCA). All genes embedded in the various dicistronic genes, fused 5S rRNA genes, RUF70 genes and their leader tRNA genes appear to be efficiently transcribed and their products correctly processed *in vivo*.

INTRODUCTION

Co-transcription of two or more genes under the control of a single promoter is very scarce in the eukaryotic world (1). Very exceptionally several genes coding for proteins are cotranscribed as a single primary mRNA. For example, all genes from chromosome 3 of *Leishmania major* (with the exception of a single gene and one tRNA genes) are transcribed as two convergent polycistronic clusters (2). More numerous cases of polycistronic non-coding RNA (ncRNA) genes are known in plants. Besides the special case of snoRNAs encoded in introns of protein coding genes, non-intronic snoRNAs gene clusters are transcribed as polycistronic pre-snoRNAs and later processed. In *Arabidopsis thaliana*, 43 clusters representing 71 snoRNA genes were identified (3), while in *Oryza sativa*, 270 snoRNA genes are involved in 70 clusters (4). A single case of dicistronic snoRNAs is known in yeast *Saccharomyces cerevisiae* where snR190 and U14 are separated by only 67 nt and the primary transcript processed by the RNA endonuclease Rnt1 (5). In higher eukaryotes, many microRNAs are encoded in polycistronic transcripts (6,7). In *S. cerevisiae* also, seven box C/D snoRNAs are expressed as a 1.4 kb polycistronic RNA processed by Rnt1 and Rat1 nucleases (8). All the above examples concern genes transcribed by RNA polymerase II (Pol II). Pol I task is devoted to the transcription of a single polycistronic gene coding for the 35S rRNA, which is the precursor of the ribosomal 18S, 5.8S and 25S rRNAs.

In contrast with Pol I and Pol II genes, Pol III genes feature internal promoters (known as the A and B boxes in the cases of tRNA genes) and display at their 3′-end a transcription termination signal made of poly-T in the RNA-like strand (9,10). Therefore, clustering of Pol III genes appears at first glance unnecessary as well as unlikely. Nevertheless two types of such clustering are yet known, homologous and heterologous. First, in most hemiascomycetous yeast, tandemly repeated tRNA genes (tDNA) are quite common. For example, *Debaryomyces hansenii* displays 17 tandem tDNAs (11). In such Pol III tandems, the termination signal is located 3′ of the second gene and the primary transcript is probably processed into two functional tRNAs by the usual tRNA maturation

machinery, as was demonstrated for tandem tRNA genes from *Schizosaccharomyces pombe* (12). The distance between the two genes is often very short, down to 3 nt; and the same types of tDNA pairs are encountered in neighbour genomes (13). Only two types of dicistronic heterologous Pol III transcripts are known yet: tRNA–snoRNA and tRNA–miRNA genes. In *A. thaliana*, 12 copies of the C/D snoRNA sno43 each located 3′ of copies of tDNA-Gly (GCC) were reported by Kruszka and coworkers (14). The same team also identified in rice two copies of a similar fusion between a novel C/D snoRNA and tDNA-eMet. tRNA–miRNA genes were discovered both in the C19MC cluster of the human genome (15) and in the mouse gammaherpes virus 68 (16). The present work reports the identification of two novel types of dicistronic heterologous Pol III transcripts in the genome of *Yarrowia lipolytica*. As a matter of fact, this hemiascomycete yeast contains multiple copies of dicistronic genes featuring in all cases a leading tRNA gene followed either by a 5S rRNA gene or a yet uncharacterized ncRNA gene.

In most hemiascomycetous yeasts, 5S rRNA genes are embedded in the highly repeated rDNA unit (~10 kb, present in about 100 copies). In each active rDNA unit, Pol III and Pol I divergently transcribe 5S and 35S precursor RNA genes, respectively. Transcription of all Pol III genes (tRNA, 5S rRNA, RPR1 RNA, SCR1 RNA genes and a few others) first requires the binding onto the gene of the assembly transcription factor TFIIIC and, in a second step, the recruitment of TFIIIB which, in turn, recruits RNA polymerase III for multiple rounds of transcription (10). For all Pol III genes (except 5S rRNA genes), the primary recognition by TFIIIC relies on two short internal sequences known as the A- and B-boxes [TRGYnnAnnnG and GWTCRAnnC, respectively, see (13)]. Transcription of 5S rRNA genes (in which no B-box is present) is even more complex as it requires the binding of the additional factor TFIIIA, prior to the assembly of TFIIIC (17,18).

TFIIIA harbours nine zinc fingers forming three functional subdomains (19). In *Xenopus*, fingers 1–3 and 7–9 bind to the A and C promoter elements of the 5S rRNA gene (20) while the central fingers 4–6 interact with the intermediate promoter element (21,22). Besides its role in transcription, TFIIIA also interacts as a chaperone with the 5S rRNA transcript itself for nucleocytoplasmic transport and RNA storage (23–25). Different fingers form the minimal interaction domains with either type of nucleic acids: fingers 1–3 are necessary and sufficient for DNA binding (26) while fingers 4–6 form the minimal RNA binding domain (27–29). Although TFIIIA is an essential protein both in *Xenopus* and yeast, its only essential function, in *S. cerevisiae*, is the transcription of 5S rRNA genes (30). As a matter of fact, cells devoid of TFIIIA may survive if a single copy of the 5S rRNA gene is expressed from a RPR1 Pol III promoter, showing that the 5S rRNA transport and storage function of TFIIIA is dispensable in yeast. Consistent with this result is the survival of several mutants affecting the central fingers 4–6 shown to form the minimal 5S rRNA binding domain in *Xenopus* (27–29): disruption of yeast TFIIIA fingers 4, 5, 6, 4 + 5 or 4 + 6 yielded viable phenotypes (31).

In *Y. lipolytica*, the genomic organization of 5S rRNA genes is known to depart from that in other hemiascomycetes (32,33). As a matter of fact, none of the 5S rRNA genes are part of the rDNA unit but appear to lie dispersed in the whole genome (11). In this work, we noticed that nearly half (48 over 108) of the 5S rRNA gene copies of *Y. lipolytica* are closely located 3′ to tDNAs, suggesting that these peculiar tDNA-fused 5S rRNA genes might be expressed, without the need of TFIIIA, as Pol III transcriptional fusions, similarly to tandem tDNAs already known in hemiascomycetes (13). Again in *Y. lipolytica*, we also identified a novel 70 nt ncRNA present in 13 copies, each of them located 3′ to the 13 copies of the tRNA-Trp gene.

We first checked, by RT–PCR that all types of fusions are transcribed *in vivo* as a single RNA as well as *in vitro* by whole cell extracts (WCE) from either *Y. lipolytica* or *S. cerevisiae*. Through RNA analysis and northern experiments we then checked that both tRNA and 5S rRNA genes appear to be correctly matured. Finally, the deletion of the TFIIIA ortholog of *Y. lipolytica* yielded a viable phenotype suggesting that the transcription of only half of the 5S rRNA genes (those 3′-fused to tDNAs) can sustain viability. This very peculiar organization of 5S rRNA genes was searched for in other ascomycete genomes but, up to now, remains unique.

MATERIALS AND METHODS

Identification of 5S rRNA and Pol III composite genes

5S rRNA genes (5S rDNA) were searched with the Génolevures Blast facility (http://cbi.labri.fr/Genolevures/blast.php#) using the *S. cerevisiae* gene as seed. Of the 5S DNA, 117 copies were found of which eight are relics [deleted at either or both end(s)]. Out of the 109 remaining copies, 60 are regular isolated copies and 49 are located immediately downstream of five different types of tDNA. In one of this dicistronic tDNA–5S DNA, the 5S DNA is interrupted by a Ylt1 transposable element. Close examination of Pol III terminators (T-track) revealed a putative ncRNA [named RNA of unknown function about 70 nt-long (RUF70)] present at the 3′-end of all tDNA-Trp (CCA) genes. Detailed alignments of all 5S rRNA, tDNA–5S rDNA and tDNA-Trp (CCA)-RUF 70 genes are given in Supplementary Data 1, 2 and 3.

RT–PCR

Total RNA was extracted using the RNeasy Midi Kit (Qiagen, Courtaboeuf, France) and DNA contamination was removed with the Turbo DNA-free kit (Applied Biosystems/Ambion, Austin, Texas, USA). RT–PCR were performed with Ready-To-Go™ RT-PCR Beads (GE Healthcare Life Sciences, Orsay, France) and PCR control with PuReTaq Ready-To-Go™ PCR Beads (GE Healthcare Life Sciences). Primers used are listed in Supplementary Data 4.

Cloning of tRNA–5S rRNA and tRNA-Trp (CCA)–RUF70 genes

Genomic DNA from strain E150 (CLIB122, *MatB lys11–23 ura3–302 leu2–270 xpr2–322*) was prepared as described (34). PCR amplifications were carried out with the pfu DNA polymerase purified in the lab. Primers used are listed as Supplementary Data 4. PCR products were then introduced into an *Eco*RV-digested pBluescript SK and competent cells of strain DH10B were transformed by electroporation (1.7 kV/mm, 25 µF, 100 Ω). Transformants were selected on LB media with ampicillin (100 µg/ml) and Xgal (40 µg/ml), verified by both PCR amplification with primers used for DNA amplification and plasmid sequencing.

In vitro transcriptions

Transcription in yeast (*Y. lipolytica* or *S. cerevisiae*) WCE [50 µg of S100 (35)] was carried out with the same buffer conditions as for the reconstituted system (see below). 50 µg of S100 extract were incubated with 60 ng of the indicated templates, 1 h at 25°C. After the addition of NTP mix [0.6 mM (ATP, GTP, CTP), 0.03 mM UTP and 1 µl of ^{32}P UTP at 10 µCi/µl], a further 20 min incubation was allowed before stopping the reaction. The transcripts were analysed by electrophoresis on 6% acrylamide–urea gel.

Standard *in vitro* transcriptions were performed as previously described (36,37) using affinity purified whole recombinant rTFIIIC, partially purified B″ fraction (38,39), pure rBdp1, rTBP, rBrf1, highly purified RNA polymerase III and the respective plasmid DNA template [see (40) for details].

Deletion of *ylTFIIIA*

The complete deletion of *Y. lipolytica TFIIIA* gene (*ylTFIIIA, YALI0F05104g*) was performed as follows (41). Promoter (P = 990 bp) and terminator (T = 588 bp) regions of haploid strain PO1d (CLIB139, *MatA ura3–302 leu2–270 xpr2–322*) were amplified by PCR with primers Tf3A-P1 (caccagaatcttcatcatcagcctactg), Tf3A-P2 (**CATTACCCTGTTATCCCTAC**aaacatgctcttagagctgactctcgag) including an I-*Sce*I restriction site in bold, Tf3A-T1 (**CTAGGGATAACAGGGTAATC**tatggagtactcgtccaatgaa gaggtc) also with an I-*Sce*I restriction site and Tf3A-T2 (gtagacagttcatcagcgcataacgtac). A ~1.9 kb I-*Sce*I_LEU2_I-*Sce*I fragment [ML cassette (41)] was introduced into the PT cassette at the I-*Sce*I restriction site. Cells of PO1d were transformed by the lithium acetate method (42) with about 400 ng of purified disruption cassette. Transformants were selected on YNB medium with NH₄Cl (5 g/l), glucose (10 g/l), sodium potassium phosphate buffer, pH 6.8 (50 mM), agar (2%) and uracyl (100 µg/ml). Verification of the correct *ylTFIIIA* deletion was realized by PCR with primers external to the disruption cassette, i.e. upstream of P (V1 gttgcatctaagtgctatc-cacg) and downstream of T (V2 atgcagcatgcagccaggtattg). As the PCR products of the wild copy and of the disrupted copy are the same size, PCR amplification was followed by an *Eco*RI restriction that generates three restriction fragments for the wild copy and four for the disrupted one. Three independent clones were retained (Δ*ylTFIIIA* C6, C7 and C8).

Northern blot analyses

For northern blot experiments, total RNA extractions were performed using the hot phenol method as described (43) and not using RNeasy Midi Kit to obtain large amount of small RNA products. Total RNA (20 µg) was analysed by northern blot using the following oligonucleotides according to the previously described procedure (44,45)

Trp: TGGAGTCGAAAGCTCTACCATTG

RUF 70: The various oligonucleotides were mixed during the hybridization step.

WX01:GCTATGGGACTTAAACCCACAAT;
WX06:GCTAGGGGAATTAAACCCCATA;
WX09:GCGATGGGACTCTAACCCATATT;
WX10:GCTAGGGGAGTCAAACCCCTTAT;
WX11:GCTACCGGAATCAAACCGGCTTA;
WX12:GCTACCGGAATTTAACCGGCTAT;
WX13:AACAGAGGAATTAAACCTCCTTG.

RESULTS

Variety of 5S rRNA genes organization throughout eukaryotic genomes

In most hemiascomycetes 5S rRNA genes are embedded in the rDNA unit (Table 1). In *S. cerevisiae*, the 35S rRNA which is the precursor of 18S, 5.8S and 25S ribosomal RNAs is synthesized by RNA polymerase I machinery while a single copy of 5S rRNA gene is synthesized by RNA polymerase III with the help of the TFIIIA, -C and -B transcription factors. The tandemly repeated rDNA units form a single cluster internal to one chromosome (as in *S. cerevisiae*) or multiple clusters often located in subtelomeric regions (as in *Candida glabrata*). *Yarrowia lipolytica* is presently the only fully sequenced hemiascomycete in which 5S rRNA genes are not included in the rDNA units, similarly to *S. pombe* and most other eukaryotes.

In *Y. lipolytica* half of the 5S rRNA gene copies are located 3′ of tRNA genes

In *Y. lipolytica*, the 108 full length 5S rRNA genes are dispersed throughout the whole genome. The genome also contains eight 5S rRNA gene relics (with various 5′ or 3′ or 5′ and 3′ deletions, see alignment in Supplementary Data 1). Very remarkably, 48 of the 5S rRNA genes are located immediately 3′ of tDNAs. Some examples of dicistronic Pol III genes are already known: dual tDNAs in *S. pombe* (12,53) or hemiascomycetes (13) and tRNA–snoRNA genes (14); but tRNA–5S rRNA composite genes are an unprecedented case. Figure 1 summarizes the different elements of the Pol III transcription machinery requested for transcription of tRNA or 5S rRNA genes. The *in vivo* transcription of a duplex tRNA gene is known to rely solely on the promoter elements of the upstream gene (12). In the present case of a dicistronic tRNA–5S rRNA gene, we hypothesized

Table 1. Variability of the genomic organization of 5S rRNA and other rRNA genes in yeasts

	Number of rDNA cluster	Internal or Subtelomeric	rDNA units number	rDNA unit length (bp)	5S rRNA genes status included (in rDNA unit) or gene number if external	References
Saccharomyces cerevisiae	Unique	I	~100	9137	Included	SGD
Candida glabrata	2	S	?	–	Included two copies/repeat	(11)
Zygosaccharomyces rouxii	?	I	?	9868	Included	(46)
Lachancea kluyveri	Unique	I	?	8656	Included	(47)
Lachancea thermotolerans	Unique	I	~200	8506	Included	(48) (46)
Kluyveromyces lactis	Unique	I	?	8711	Included	(49) (46)
Eremothecium gossypii	Unique	I	~50	8187	Included	(50)
Debaryomyces hansenii	3	I	?	7705	Included two copies/repeat	(51)
Yarrowia lipolytica	6	S	~100	–	109	(32) (11)
Schizosaccharomyces pombe	2	S	~100	–	30	(32, 52)

The rDNA unit comprises 18S, 5.8S and 25S ribosomal RNA genes, which are transcribed as a single 35S precursor rRNA by RNA polymerase I. The 5S rRNA genes are transcribed by RNA polymerase III. In all yeasts but *Y. lipolytica* and *S. pombe*, one copy of 5S gene is embedded in the rDNA (two copies for *C. glabrata* and *D. hansenii*). *S. cerevisiae* data were obtained from the Saccharomyces Genome Database (SGD, http://www.yeastgenome.org/).
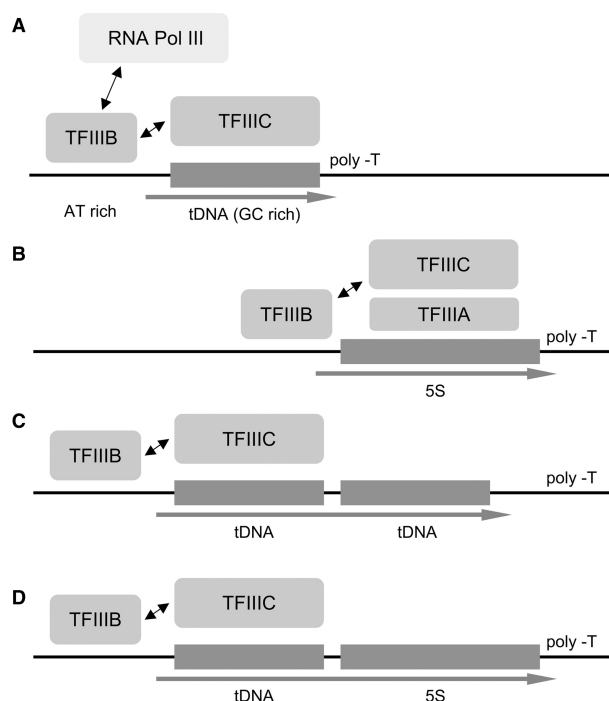
Figure 1. Overview of the transcription mechanisms of tRNA, 5S rRNA and tRNA–5S rRNA dicistronic genes. (**A**) In *S. cerevisiae*, transcription of a tRNA gene by Pol III requires the assembly of transcription factor TFIIIC onto the tDNA (GC-rich) followed by that of TFIIIB (onto a AT-rich region) that finally recruits Pol III for multiple cycles of transcription. (**B**) A regular 5S rRNA gene is first recognized by its specific factor TFIIIA; then the 5S rRNA gene–TFIIIA complex is bound by TFIIIC and the next steps of transcription are identical to that of tRNA genes. In both cases, transcription stops and efficiently recycles when Pol III reaches the terminal T-track in the RNA-like strand. (**C**) Transcription of a dicistronic tRNA–tRNA gene by a unique TFIIIC molecule binding on the upstream gene. Assembly of TFIIIB onto the upstream gene (GC-rich) triggered by TFIIIC bound on the downstream gene to transcribe this only gene is penalized *in vivo*. (**D**) Hypothetical transcription of a dicistronic tRNA–5S rRNA gene may proceed similarly through the recognition of the promoter elements of the upstream gene by TFIIIC. In this case a single primary RNA (arrow) is produced and later matured into two functional products (tRNA and 5S rRNA) without the need of TFIIIA (see text for details).

Table 2. Genomic organization of the 47 tRNA–5S RNA dicistronic genes and of a single tRNA–tRNA–5S RNA tricistronic gene in *Y. lipolytica*

tRNA (type)	Number of tRNA gene copies	Number of 5S rRNA gene-fused copies (full length)	Intergenic distance (nt)
Gly (GCC)	30	28	4 or 5
Asp (GTC)	28	8	4 or 5
Glu (CTC)	27	9	4
Gln (CTG)	13	1	4
Thr (AGT)	23	1	7
Lys (CTT)-Glu (CTC)	3	1	17 and 4

that, for steric reasons, TFIIIA may not bind the downstream 5S rRNA gene and therefore that this composite gene could be expressed without the need of TFIIIA.

The type of the tDNAs to which 5S rRNA genes are linked appears to be non-random (Table 2). Of the 30 tDNA-Gly (GCC), 29 are followed by 5S rRNA genes but none of the 11 tDNA-Gly (TCC). In one of these 29 dual genes, the 5S rRNA gene is interrupted by an Ylt1 element. Of the 28 tDNA-Asp (GTC) and 27 tDNA-Glu, 8 and 9 are followed by 5S rRNA genes, respectively. Two single tDNA-Thr (AGT) and tDNA-Gln (CTG) are also followed by 5S rRNA genes. The distance separating nt 73 of the tDNAs and nt 1 of the mature 5S rRNA is most of time 4 or 5 nt and only once, 7 nt. We also observed the presence of a single tricistronic gene tRNA-Lys (CTT)-tRNA-Glu (CTC)-5S rRNA gene with intergenic distances of 17 and 4 nt. From the examination of their sequence, all 3′ fused 5S rRNA genes appear functional (with the exception of a single one 3′ deleted by the insertion of an Ylt1 element). The sequences of the 5S rRNA fused genes do not depart from that of isolated genes. Coordinates of these 47 dicistronic and of the single tricistronic genes are given in Supplementary Data 2.

A novel putative ncRNA, 'RUF70', is located 3′ of the 13 tRNA-Trp genes of *Y. lipolytica*

Following the discovery of tRNA genes located immediately upstream of about half of the 5S rRNA genes, we thoroughly checked for the presence of a Pol III transcriptional terminator (T-track) in the 50 nt at the 3′-end of all other tRNA genes. With the exception of a few tandem tRNA genes already reported (11), a feasible terminator was always identified except in all 13 copies of the essential tRNA-Trp (CAA) genes in which the Pol III terminator was always found about 70 nt downstream of the tDNA 3′-end. According to previously used nomenclature (54), we named this putative ncRNA 'RUF70'. As shown in Figure 2A, a common tertiary structure is possible for all 13 copies of *Y. lipolytica* RUF70. Noteworthy, base changes leading to compensatory base pairing (shown with stars in Figure 2A) support conservation of secondary structure and argue positively for function. An alternative structure is shown in Figure 2B and the corresponding 2D structures in Figure 2C and D. Detailed coordinates of the 13 tDNA-Trp (CCA)–RUF70 dicistronic genes are given in Supplementary Data 3.

Characterization of the tRNA–5S rRNA and tRNA-Trp– RUF70 genes dicistronic transcripts

Performing RT–PCR, we demonstrated that tRNA–5S rRNA and tRNA-Trp–RUF70 genes are transcribed into a single dicistronic pre-RNA. Oligonucleotides were designed to amplify, after reverse transcription, a PCR double stranded DNA extending from nt 1 of the mature tRNA to nt 118 of the 5S rRNA (or nt 68 in the case of RUF70). Figure 3A shows that, in all cases, PCR bands of the expected length were obtained.

One copy of each type of fusions was then selected and cloned into an *Eco*RV-digested pBluescriptSK vector for further transcriptional studies. Using WCE from both *S. cerevisiae* and *Y. lipolytica* strains, we observed one major RNA species for the various DNA templates that, according to their size, may represent the primary transcripts (Figure 3B). As previously described (12,53,55), our data confirmed that the tRNA and the 5S rRNA or the tRNA-Trp and RUF70 genes are transcribed coordinately, starting from the promoter of the leader tRNA gene. The same results are obtained using an *in vitro* reconstituted Pol III transcription system. In fact, in this case some minor bands with slower mobility can be detected, suggesting that some processing events may have occurred (Fig. 3C), as previously described (40). Yeast *S. cerevisiae* cell free extracts were previously shown to be able to process precursor tRNA but also tRNA–tRNA fusion (55,56). However, under our experimental conditions, none of our tRNA fusions were completely cleaved (Figure 3C), suggesting that these fusions are not correctly processed in our *in vitro* experiments.

Characterization of the matured RNA products

To assess the correct maturation of the tRNA-Trp and RUF70 molecules, we perform northern experiments with RNA extracted from *Y. lipolytica* and *S. cerevisiae*

growing in exponential phase (Figure 3D). Only one strong signal with the same size was observed with tRNA-Trp specific oligonucleotides in total RNA prepared from both species (Figure 3D, top pannel). As a matter of fact, tDNA-Trp from both species are nearly identical [only 3 nt are different]. On the other hand, oligonucleotides targeted to RUF70 RNAs revealed a signal only in RNA extracted from *Y. lipolytica* as expected (bottom panel). It corresponds to a band that migrates slightly more rapidly that the tRNA-Trp (∼70 versus 78 nt, data not shown). We hardly detected the dicistronic transcript tRNA-Trp–RUF70 (Figure 3A), however, its presence *in vivo* as a mature ∼70 base product is unambiguously evidenced by northern experiment (Figure 3D). We, therefore, favour the hypothesis of Pol III dicistronic transcripts for the expression of the 13 tRNA-Trp–RUF70 genes.

Although two types of 5S rRNA are present in *Y. lipolytica* (expressed from the isolated genes and potentially from the tDNA-fused genes), only a single species of 5S rRNA is revealed when total RNA was analysed in acrylamide–urea gel (Figure 4A, lanes 2 and 3). Altogether these data strongly suggest that all kinds of discistronic genes are both expressed as a primary RNA product that is next correctly processed into tRNA, 5S rRNA or RUF70 functional components.

Inactivation of *ylTFIIIA* yields a viable phenotype in *Y. lipolytica* and suggests that *ylTFIIIA* is possibly implicated in 5S rRNA maturation

Since 5S rRNA expressed by the transcription of dicistronic tRNA–5S rRNA genes appear to be correctly matured, we wondered whether TFIIIA was really, in *Y. lipolytica*, an essential protein. In *S. cerevisiae*, TFIIIA is encoded by *YPR186c* (Tfc2, also name Pzf1); this protein is essential for the transcription of 5S rRNA genes and the deletion of its gene is lethal for the cells. (57). Using BlastP, we unambiguously identified orthologs in *C. glabrata*, *Kluyveromyces lactis*, *D. hansenii* and *Y. lipolytica* (with E values e-118, e-118, e-76 and e-29, respectively). Sequences and alignment of TFIIIA orthologs in hemiascomycetes are given in Supplementary Data 5. We performed the deletion of the identified ortholog (*YALI0F05104g* gene, see Materials and Methods) in the *Y. lipolytica* haploid strain PO1d. Among the numerous resulting clones, 10 independent clones were analysed by PCR and *Eco*RI restriction to verify the deletion. Three of them ($\Delta ylTFIIIA$ C6, C7, C8, circled in Figure 4B) showed only the expected restriction pattern of the disrupted cassette and not the pattern of the wild copy. This implies that *ylTFIIIA* gene was successfully deleted in these transformants, with no remaining wild gene copy. These transformants displayed viable phenotype with normal growth at 18°C and 28°C and only a slight morphological phenotype: colonies are smooth instead of rough (Figure 4C). We then examined the total RNA extracted from both WT and $\Delta ylTFIIIA$ mutant strains (Figure 4A). In both types of cells, we observed a single band for matured 5S rRNA; however in the mutant strains, the 5S rRNA appear to be slightly longer
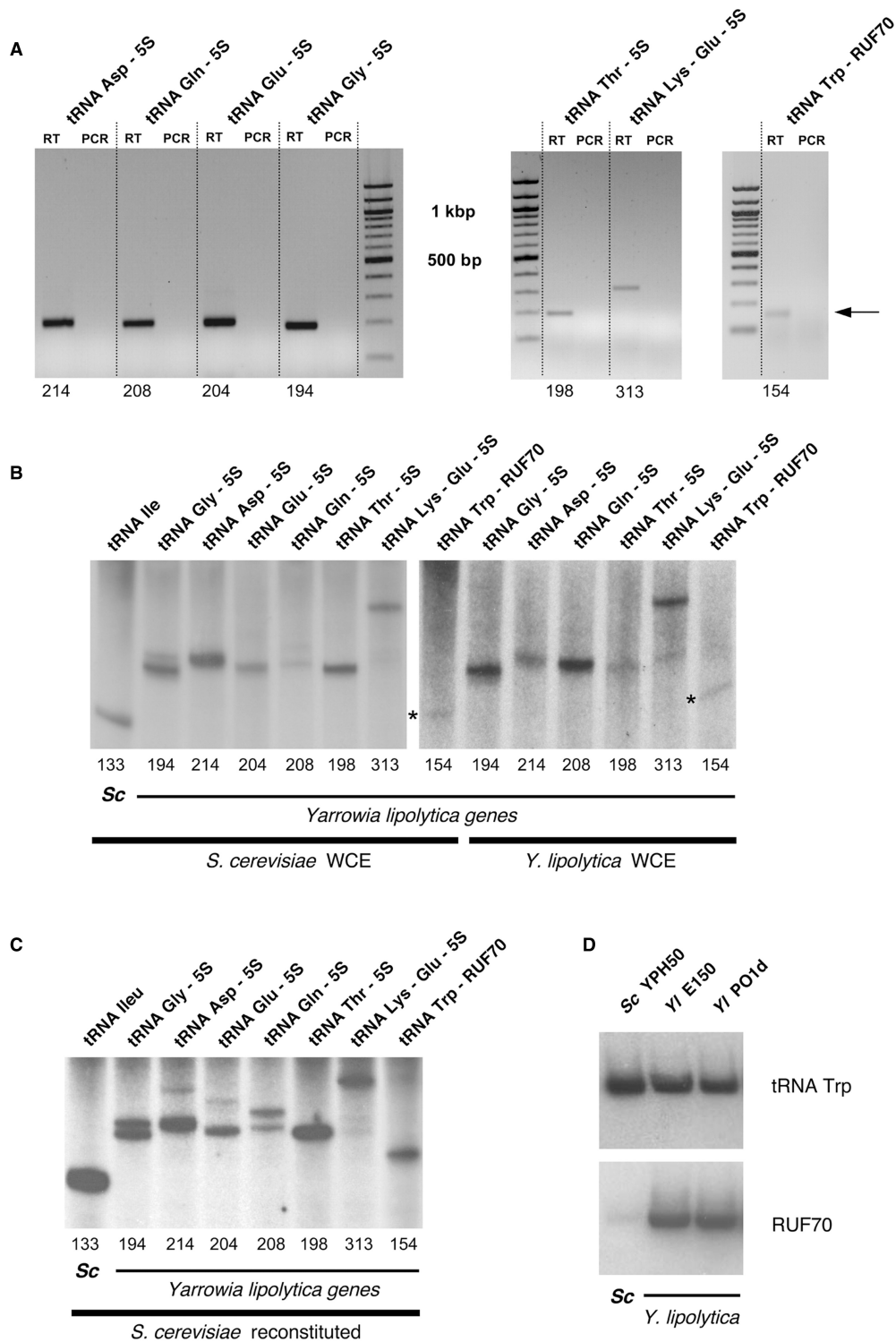
A

```
tDNA-Trp #73  <---------------------------------  RUF70 gene  --------------------------------->
         |
         |    *            * *           .(.......       * *        * *              *
     #   | .((((.....)))))((((((.(.......((((........(((((.......)))))).))))...)))).))..............
     1 TCCATGTCA AGTTCTCT GAACTGACAGTGG     AGGAAACAT TGTGGGTTTAAGTCCCATAGCTTC  AGCTGATCGGTGTA  TTTTTTG 68
     2 TCCATGTCA AGTTCTAATAGAACTGACAGTGG     AGGAAATTA TGGGGGTTTAAATTCCCCTAGCTTCT AGCTGATCAGTGATA TTTTTTTG 72
     6 TCCATGTCA AGTTCTAACAGAACTGACAGTGG     AGGAAACATA TGGGGGTTTAATTCCCCTAGCTTCT AGCTGATCAGTGA  TTTTTTTG 70
     3 TCCATGTCA TGTTCTACGAGAACTGACAGTGG     AGGAACATA TGTGGGTTTGATTCCCATAGCTTCT AGCTGATCAGTGA  TTTTTTTG 70
     7 TCCATGTCA AGTTCGCAA GAACTGGCAGTGG     AGGAACATA TGTGGGTTTAATTCCCATAGCTTCT AGCTGATCAGTG   TTTTTTTG 68
     9 TCCATGTCA AGTTCTCAC GAATTGACAGTGG     AGGAAACAA TATGGGTTAGAGTCCCATCGCTTCT AGCTGATCAGTG   TTTTTTTTG 68
    12 TCCATGTCA AATTCTAAG GAATTGACAGTGG     AGGAAACAATAGCCGGTTAAATTCCGGTAGCTTCTC AGCTGATCAACATC TTTTTTT 68

     4 TCCATGTCA AATTCTCAA GGATTGACGATGGCTTAACAAGGAATATA TGGGAGTTCAATTCTCCTAGCTTCTGA AGCTGATCAACACA TTTTTTG 79
    10 TCCATGTCA AATTCCTATGGGATTGACGAATGGCTTAAACTGGAACAAATAAGGGGTTTGACTCCCCTAGCTTCTGA AGCTAGTCAACAC TTTTTTTTTG 80
    13 TCCATGTCA TTGTTCTTCG GGACTGACGATGGCCTCAATAAGGAAATCAAGGAGGTTTAATTCCTCT AGCTTAGTCAACAC TTTTTTTG 70

     5 TCCATGTCA TGTT   AAT GACAGTGG     AGGAAAAAA GGCCGGTTTGATTCCGGTAGCTTCTCTAAGCTGATCAACACTA TTTTTTC 67
    11 TCCATGTCA TGTA   AAT GACAGTGG     AGGAAAATA AGCCGGTTTGATTCCGGTAGCTTCTCTAAGCTGATCAACACA TTTTTTA 66
     8 TCCATGTCA TGTT   AAT GACAGTGG     AGGAACATT AGCCGGTTTGATTCCGGTGGCTTCTCAAAGCTGATCAACAAC  TTTTTG 66
```

B

```
   (((((((((  (((((.....)))))))))) )))      )((.......((((((.......)))))).)) .....
 1 TCCATGTCA AGTTCT CT GAACTGACAGTGG     AGGAAACAT TGTGGGTTTAAGTCCCATAGCTTC  AGCTGATCGGTGTA  TTTTTTG
```

C

D

Figure 2. Structural alignment of the 13 RUF70 genes located beyond the 13 copies of tDNA-Trp (CCA). (**A**) For sake of clarity, the DNA sequence, instead of RNA sequence, is used. Nucleotides 67–73 of the tDNA-Trp (CCA) are shown at left followed by the sequence of the RUF70 genes (sorted according to similarity). Coloured background in stems indicate correct base pairing (including GT pairs, GU in RNA). Stars above base pairing line (shown as opening and closing brackets) denote compensatory base pairing (e.g. TA or GC or CG pairs exchange) in stems. The length indicated at right is computed from the nucleotide following nt 73 of tDNA-Trp to the transcription termination signal (T-track, not included). The oligonucleotides used in northern experiments (see Figure 3D) hybridize with the underlined sequences. Detailed coordinates are given in Supplementary Data 3. (**B**) Alignment showing an alternative structure expanding the first stem at the expense of the tRNA acceptor stem. (**C**) 2D RNA structure corresponding to alignment shown in (A) for the first sequence. (**D**) Same for the alignment shown in (B).

Figure 3. Expression detected by RT–PCR, *in vitro* trancriptions and northern analyses. (**A**) Expression of the five different tRNA–5S rRNA, one tRNA–tRNA–5S rRNA and tRNA-Trp–RUF70 composite genes detected by RT–PCR. 'RT' refers to regular RT–PCR experiment, the reverse transcriptase was omitted in the lanes labelled 'PCR'. The predicted length of the RT–PCR products is indicated at bottom. Only a faint band was obtained when testing the expression of the tRNA-Trp–RUF70 gene (arrow). (**B**) *In vitro* transcription of cloned copies of the same composite genes with WCE from either *S. cerevisiae* or *Y. lipolytica* as indicated. A longer exposure was used for the rightmost seven lanes. A tRNA-Ile (TAT) gene from *S. cerevisiae* (labelled '*Sc*') is used as a control in the left lane. Expected lengths are indicated at bottom. Asterisks denote the faint bands obtained in the transcription experiments of tRNA-Trp–RUF70 gene. (**C**) *In vitro* transcription of the same genes using a fully recombinant TFIIIC reconstituted system from *S. cerevisiae*. (**D**) Northern blot analysis of the mature transcription products of the tRNA-Trp (CCA)–RUF70 composite genes. Total RNAs were extracted from the indicated strains growing in exponential phase. RUF70 genes were probed with a mixture of oligos targeted at the main stem-loop (sequences underlined in Figure 2A). The positive response of *S. cerevisiae* RNAs is due to the nearly perfect sequence conservation of the tRNA-Trp (CCA) genes between *S. cerevisiae* and *Y. lipolytica*.
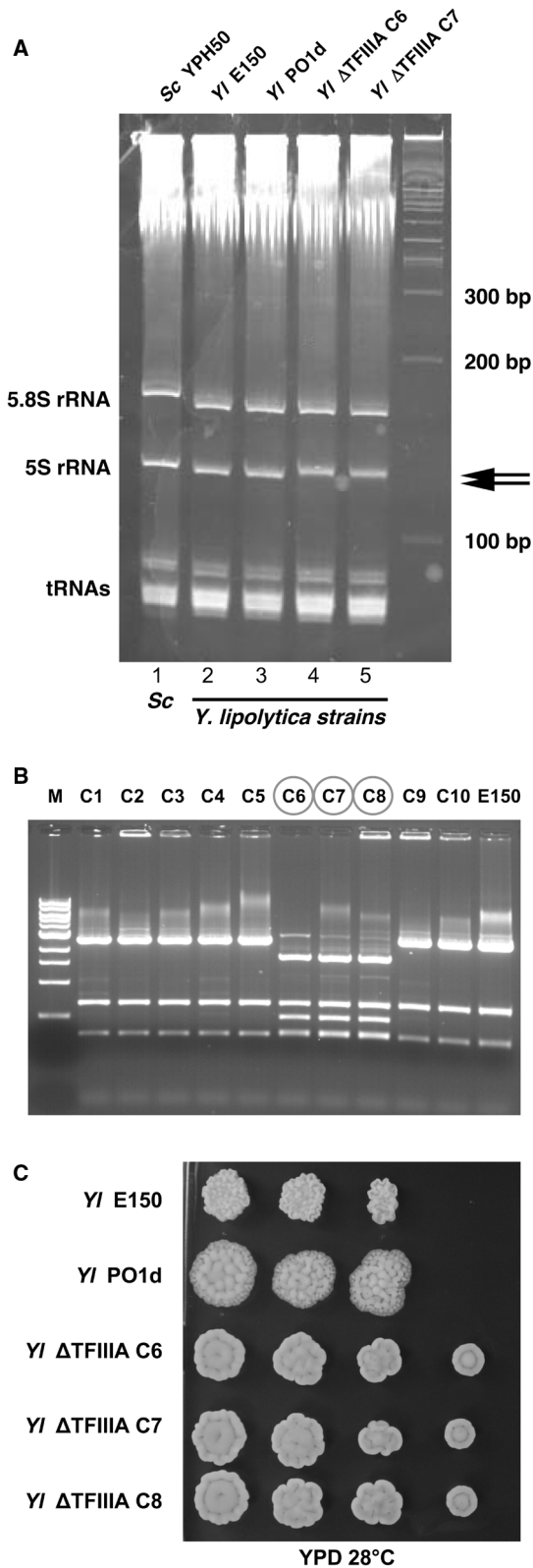
Figure 4. Phenotypic and 5S rRNA analyses of *Y. lipolytica* Δ*ylTFIIIA* mutants. (**A**) Total RNA content of wild-type *S. cerevisiae* strain (YPH500), wild-type *Y. lipolytica* strains (E150 and PO1d) and two Δ*ylTFIIIA Y. lipolytica* deleted strains (Δ*ylTFIIIA* C6 and Δ*ylTFIIIA* C7, similar results were obtained for Δ*ylTFIIIA* C8, not shown). Cells were grown in exponential phase, total RNA was

suggesting that *ylTFIIIA* could be, in some way, implicated in the maturation of 5S rRNA [compare lanes 4 or 5 (wild-type) with lanes 2 or 3 (mutant) in Figure 4A]. This difference in length appears to be comparable to the difference between 5S rRNA from *S. cerevisiae* (lane 1) and *Y. lipolytica* (lane 2) which is only 1 nt.

Evolution of TFIIIA factor throughout eukaryotes: specificity of the *Y. lipolytica* ortholog

The first TFIIIA sequence was obtained from a *Xenopus laevis* cDNA (58). Proteolytic digestion at 3-K interval and the presence of 9–11 atoms of zinc per 7S particle (5S rRNA + TFIIIA) together with the repetitive presence of numerous Cys and His residues suggested the 9 zinc-finger model (19). Further sequencing of the *X. laevis* TFIIIA gene itself revealed that exon–intron boundaries also correspond closely to the repetitive unit, supporting the idea that TFIIIA evolved by successive duplication of a primordial 30-residue unit (59). The second TFIIIA sequence, that of *S. cerevisiae* was shown to diverge extensively from the *Xenopus* sequence, especially fingers 8 and 9 were separated by an essential 81-residue intervening sequence (60,61). The gene is unique in *S. cerevisiae* and TFIIIA is essential. Next came the sequence of *S. pombe* (52) and a careful comparison of the three sequences revealed that the *S. pombe* ortholog has in fact 10 zinc-finger, with its finger 9 corresponding to finger 9 of the *Xenopus* sequence and finger 10 corresponding to finger 9 of the yeast sequence (62); this is schematized in Figure 5. Curiously, the *S. pombe* ortholog terminates immediately after finger 10. Examination of several hemiascomycete TFIIIA sequences shows that they are all similar to the *S. cerevisiae* model: the Cys and His residues of the last but one finger (numbered 8 in *X. laevis*) were lost (see multiple alignment in Supplementary Data 5). Interestingly, the *Y. lipolytica* ortholog (*ylTFIIIA*) is the longest among TFIIIA proteins (655 residues) and it displays two specific extensions with respect to the sequences of other hemiascomycetes. A first 37-residue intervening sequence is located between fingers 9 and the last homology block common to all hemiascomycetes, a 10th zinc finger is present in this extension; it has no equivalent in other TFIIIA's. A second long N-terminal extension extends the sequence by 133 residues with respect to *S. cerevisiae* ortholog (Figure 5).

extracted and analysed in 6% acrylamide–urea gel and stained with ethidium bromide. The migration of tRNA, 5.8S and the two distinct 5S rRNA species (arrows at right) are indicated. (**B**) Restriction pattern of the *ylTFIIIA* amplified locus of *Y. lipolytica* transformants. Only three out 10 transformants lost accurately the *ylTFIIIA* gene (C6, C7 and C8). Strain E150 was used as a control for the wild restriction pattern. M, molecular marker, 1 kb DNA ladder (New England Biolabs). (**C**) Effects of the disruption of the *ylTFIIIA* gene (*YALI0F05104g*) on growth ability. Serial dilutions (from 3×10^3 to 3 cells) of overnight cultures of the wild-type strains (E150 and PO1d) and of three independent (Δ*ylTFIIIA* C6, C7 and C8) mutant strains were inoculated on YPD plates (or YNB supplemented, not shown) and incubated at 28°C (or 18°C, not shown) for 3 days.
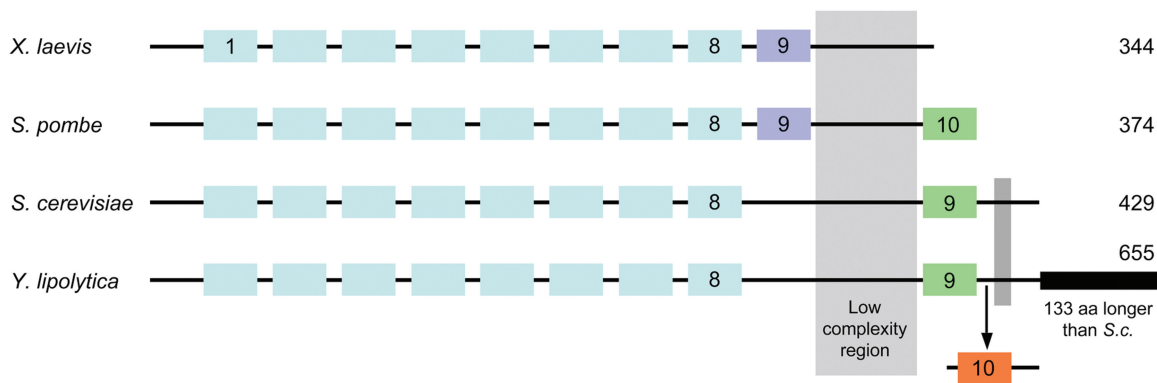
Figure 5. Evolution of TFIIIA sequence throughout eukaryotes. The nine filled rectangles in the top sequence (*X. laevis*) symbolize the nine zinc fingers. The last zinc finger in *S. cerevisiae* TFIIIA is also traditionally numbered nine because this sequence was obtained prior to that of *S. pombe* that exhibits 10 zinc fingers. Multiple sequences alignment (not shown) confirm that finger 9 of *S. pombe* sequence corresponds to finger 9 of *X. laevis* sequence and that fingers 9 of *S. cerevisiae* and *Y. lipolytica* correspond to finger 10 of *S. pombe*. In TFIIIA from *Y. lipolytica*, an extra 10th zinc is present between finger 9 and the last block of homology common to all hemiascomytes (small dark vertical rectangle). See also alignment in Supplementary Data 5.

DISCUSSION

In this work we report the identification and first characterization of two novel types of dicistronic Pol III genes in *Y. lipolytica*. In both types, similarly to other cases already known, the leader gene encodes a tRNA and the second gene encodes here either a 5S rRNA (48 cases) or a novel RNA (13 cases), RUF70, of unknown function. These two types enlarge the present collection of dicistronic Pol III genes: tandem tRNA genes in *S. pombe* and hemiascomycetes (11–13); tRNA–snoRNA genes in *A. thaliana* and *O. sativa* (14); tRNA–miRNA genes in human (15) and mouse gammaherpes virus 68 (16). Nearly half of the 5S rRNA gene copies present in *Y. lipolytica* are implicated in Pol III fusions (48 over 108 full length genes), splitting the 5S rRNA genes into two families: the isolated genes (60 copies) and the fused genes (48 copies). The existence of such dicistronic genes linking a tRNA gene and a 5S rRNA gene is an unprecedented and still unique case. Interestingly, *ylTFIIIA* becomes no more essential in normal growth conditions and its absence revealed a possible role of *ylTFIIIA* in 5S rRNA processing.

The various composite genes appear to be active in normal growth conditions. Pol III promoter elements, defined as containing the sequences TRGYnnAnnnG and GWTCRAnnC as the A- and B-block, respectively (13), are present and functional as evidenced by *in vitro* experiments performed with a reconstituted Pol III transcription system. Furthermore, these Pol III promoter elements are correctly cross recognized through species (Figure 3B and C) since transcription can be detected with the Pol III transcription machinery derived from *S. cerevisiae*. Transcription of these genes occurs *in vivo* and starts with the synthesis of a primary RNA product expressing both genes, as demonstrated by RT–PCR (Figure 3A). Further experiments demonstrate that these RNA are processed *in vivo*. All copies of the tDNA-Gly(GCC) but two (28 out of 30) are followed by a full length 5S rRNA gene, RUF70 is only present in dicistronic genes in fusion with tDNA-Trp. However, neither uncleaved tRNA-Gly (GCC)-5S rRNA, nor tRNA-Trp–RUF70

primary product could be detected by northern blot analysis contrary to the matured tRNA-Gly, tRNA-Trp, 5S rRNA and RUF70 (Figure 3D and data not shown). As shown experimentally by Kruszka and coworkers, the accurate processing of a snoRNA adjacent to tRNA-Gly (GCC) in *A. thaliana* implicates the regular tRNA 3′-processing activity performed by tRNase Z (14). In *S. cerevisiae*, the 3′-tRNase activity is supported by Trz1 (encoded by *YKR079c*) (63,64) and this protein appears to be well conserved among hemiascomycetes (E-value e-32 for the ortholog of *Y. lipolytica*, YALI0A13057g, with respect to that of *S. cerevisiae*). Yeast cell free extracts were generally found to allow the synthesis of processed transcripts, including internal cleavage when dicistronic genes are used (55,56), contrary to what we observed using both *S. cerevisiae* and *Y. lipolytica* extracts (Figure 3B), suggesting that the reaction conditions are not optimal for the cleavage. For instance, metal ion were found to be required for *in vitro* pre-tRNA processing reactions catalyzed by TRZ1, the tRNase Z from *A. thaliana* (65). tRNAse Z also appears to request *in vitro* finely tuned pH and salt concentration values from one species to another (66) and we concluded that the cleavage by tRNase Z is not observed in our transcription experiments due to unsuitable pH and salt conditions.

Tandem tRNA gene transcription was first investigated two decades ago with a tDNA-Ser–tDNA-Met from *S. pombe*. Both genes were shown to be transcribed as a dimeric tRNA precursor followed by processing of the two tRNAs (12). *In vivo* transcription was drastically reduced by mutations in either the A-box (A at position 19) or the B-Box (A at 53) of the leader tRNA-Ser gene. This provided a direct evidence that the trailer tRNA-Met gene is derived through processing of the dimeric precursor and not initiated independently from the internal tRNA-Met gene promoters (12). The novel case of tRNA–5S rRNA genes in *Y. lipolytica* raises the question of a possible conflict between the Pol III transcription factors requested for both genes. We reasoned similarly in the case of the composite tRNA–5S rRNA

genes of *Y. lipolytica* and assumed that *ylTFIIIA* might be dispensable in this special case. We also considered a possible steric conflict between *ylTFIIIA* bound on the 5S rRNA gene and TFIIIC bound on the leader tRNA gene. Furthermore, if *ylTFIIIA* and TFIIIC bind on the 5S rRNA gene, TFIIIB would have to bind DNA inside the leader tRNA gene, the high GC content of which is known to prevent efficient TFIIIB binding (67). For these reasons, we attempted the deletion of the *Y. lipolytica* ortholog of *S. cerevisiae* TFIIIA.

Deletion of *ylTFIIIA* turned out to be viable and leads only to a slight morphological phenotype. More important, 5S rRNA appears slightly longer (a few nucleotides) in the deletant strains and we noticed that a single 5S rRNA species, homogenous in length is observed in both the wild-type strains and in the *ylTFIIIA*-deleted strains. This may appear illogical, at first glance, if one considers that, in wild-type strains, two types of 5S rRNAs coexist: those expressed from the isolated 5S rRNA gene and those resulting from the maturation of the dicistronic primary transcript of the tRNA–5S rRNA genes (both types being present in nearly equal gene amounts). In the *ylTFIIIA*-deleted strains, in which the transcription of isolated genes is no more possible, only the latter type of 5S rRNA is present. The defect in 5S RNA processing we observed in our Δ*ylTFIIIA* strains (a few nucleotides longer) suggests that *ylTFIIIA* itself could participate, in some way, in this processing possibly through its role of 5S RNA-chaperone (23–25) or in functional interaction with the YL3 protein (68). However, it also indicates that the final processing event that occurs in the presence of *ylTFIIIA* is not absolutely required for 5S rRNA to be functional, at least under normal growth conditions. As a matter of fact, 3′-extended 5S RNA was also observed in a strain deleted for the 3′ exoribonuclease Rex1p exhibiting no growth defect (69). With respect to TFIIIA proteins of other hemiascomycetes, *ylTFIIIA* harbours a 10th zinc finger and a long C-terminal extension. These specific features might be also implicated in novel roles for TFIIIA in *Y. lipolytica*.

Our results nicely parallel those obtained earlier on the artificial Pol III fusion RPR1 promoter-5S rRNA studied by Camier and coworkers (30). A chimeric gene was constructed in which the first 84 nt of the RPR1 gene containing a tRNA-like promoter sequence was fused to a 5S rRNA gene. This gene was transcribed in an *in vitro* Pol III transcription assay and a single RNA transcript was observed at the expected length for initiation at the RPR1 promoter, however no cleavage of the two products took place. No transcript initiating at the 5S promoter was observed, even after addition of TFIIIA, suggesting that the RPR1 promoter is functional and dominant over the downstream 5S promoter. *In vivo*, cells devoid of TFIIIA could survive if they carry such a construction on a multicopy plasmid. Similarly with our present observations, 5S rRNA expressed from RPR1-5S rRNA hybrid genes was slightly longer than that produced from endogenous 5S rRNA gene copies.

A number of tandem tDNAs are already known in hemiascomycetes (11). We first checked that the Pol III gene fusions we observed in the sequenced strain of *Y. lipolytica*

(strain E150) were also present in the strain investigated during earlier studies [strain PO1d, (70)]. We then checked the set of nine hemiascomycetes + *S. pombe* previously studied (13) for the presence of a Pol III terminator 3′ to all tDNAs (within 40 nt 3′). No new cases of Pol III fusions other than tandem tDNA genes already reported were discovered. We then extended our search to other ascomycetes: *Pichia stipitis*, *Neurospora crassa*, *Magnaporthe grisea*, *Coprinus cinereus*, *Aspergillus nidulans*, *Fusarium graminearum*, *Ustilago maydis* and *Podospora anserina*. No other type of Pol III fusions than tandem tDNAs was discovered (with the exception of a triple tDNA-Lys (CTT) in *Pichia stipitis*). We also further investigated two recently published plant genomes: *Populus trichocarpa* and *Vitis vinifera*. In *P. trichocarpa*, we disclosed 12 copies of the tDNA-Gly (GCC)–snoRNA (Figure 6) identical to those of *A. thaliana* (14) and only two copies in *V. vinifera*. We finally noticed that, curiously, the very same type of tDNA gene, namely tDNA-Gly (GCC) is implicated in Pol III gene fusions in several organisms: tRNA-Gly (GCC)–snoRNA genes in *A. thaliana* (14) and also in *P. trichocarpa* (this work), tDNA-Gly (GCC)-5S rRNA genes in *Y. lipolytica* (this work). We do not know whether this might indicate for tDNA-Gly (GCC) a propensity to generate fusions more easily than other tRNAs. Remarkably, the other tDNA-Gly, with anticodon (TGC), is never implicated in fusions.

Remarkably, the *Y. lipolytica* genome contains another type of Pol III fusions: each of the 13 copies of the tRNA-Trp (CCA) gene is followed by an extension of about 70 nt. This extension is cotranscribed and appears to be stable enough as northern blot can detect it. We noticed that between the 13 copies, the sequence is only loosely conserved, suggesting some degeneration. We first thought about a RNA implicated in the maturation of the tRNA-Trp or in the splicing of its intron. However, this tRNA is highly conserved throughout hemiascomycetes (only 3 nt are changed with respect to the *S. cerevisiae* tRNA-Trp) and the intron is not present in some species such as *D. hansenii*. We noticed some structural similarity (stems/loops) with the ncRNA CeN76 identified in the non-coding transcriptome of *Caenorhabditis elegans* [Figure 2C in (71)]. In the model we propose (Figure 2A and C), a stem is immediately adjacent to the 3′ end of the tRNA, similarly with the models proposed for CeN76 and the snoRNA following the tRNA-Gly (GCC) in *A. thaliana* [Figure 6 in (14)]. Attempt to identify a similar ncRNA in other ascomycetes and *S. pombe* remained unsuccessful. The role of RUF70 in *Y. lipolytica* remains unknown as are unknown the role of many ncRNAs recently discovered in higher eukaryotes (54,71).

In each yeast species, all tRNA genes encoding a given tRNA type (same anticodon) remain identical or nearly so. This intraspecific sequence homogeneity, combined with the rapid variation of the number of paralogous gene copies, is puzzling. It contrast with the rapid sequence variation between paralogous copies of protein coding genes, suggesting the existence of a mechanism for a quick loss and proliferation for tRNA genes in genome. The existence of a proliferation mechanism is further suggested by the frequent occurrence, in yeast genomes, of a variety of Pol
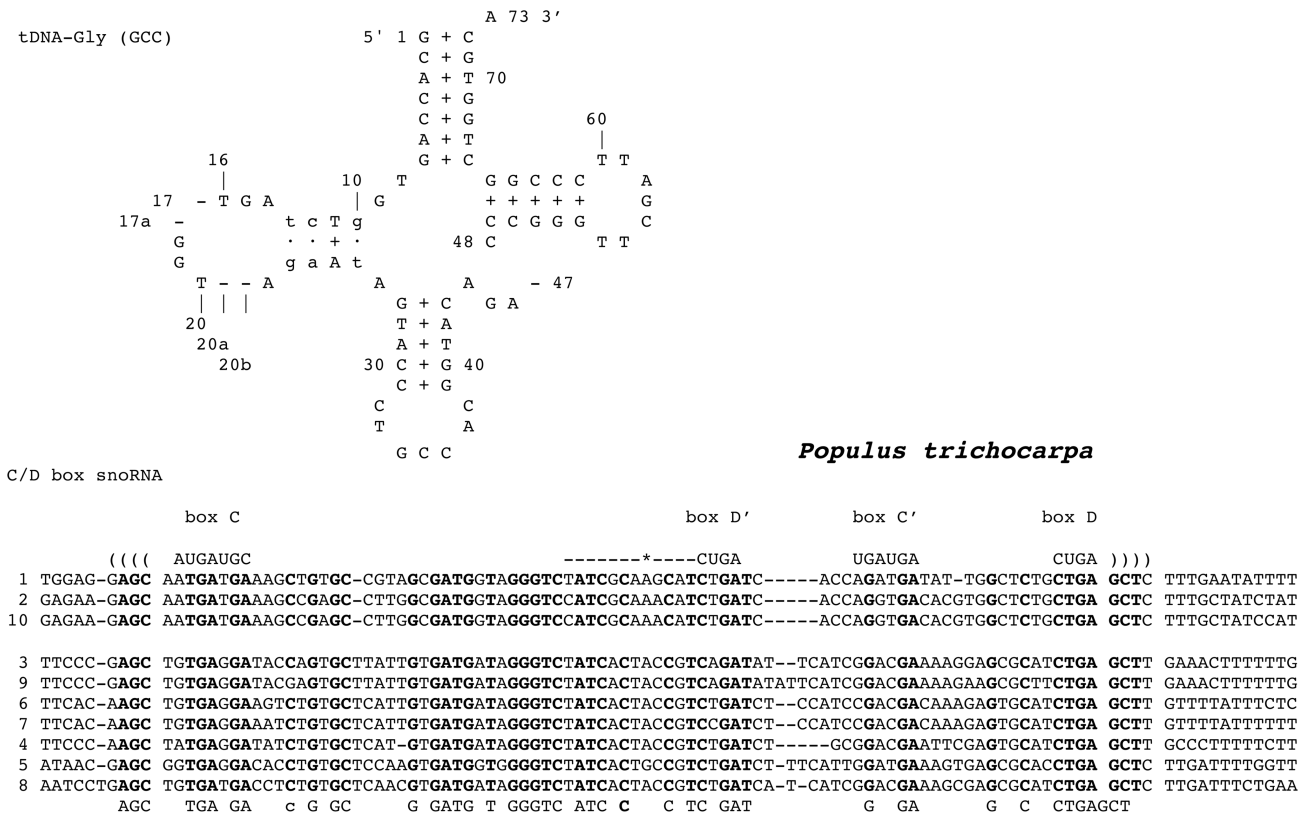
```
                                        A 73 3'
tDNA-Gly (GCC)              5' 1 G + C
                                  C + G
                                  A + T 70
                                  C + G
                                  C + G            60
                                  A + T             |
                                  G + C            T T
              16             10   T      G G C C C      A
               |          - T G A  | G      + + + + +       G
          17  - T G A            t c T g    C C G G G       C
          17a  -                 · · + ·         48 C    T T
              G                  g a A t
              G             A          A        A    - 47
          T - - A         G + C       G A
           | | |          T + A
          20               A + T
          20a          30 C + G 40
          20b              C + G
                          C        C
                          T        A
                           G C C
```
<p style="text-align:right">***Populus trichocarpa***</p>

C/D box snoRNA

```
        box C                                    box D'      box C'       box D
      ((((  AUGAUGC                -------*----CUGA   UGAUGA      CUGA ))))
 1 TGGAG-GAGC AATGATGAAAGCTGTGC-CGTAGCGATGGTAGGGTCTATCGCAAGCATCTGATC-----ACCAGATGATAT-TGGCTCTGCTGA GCTC TTTGAATATTTT
 2 GAGAA-GAGC AATGATGAAAGCCGAGC-CTTGGCGATGGTAGGGTCCATCGCAAACATCTGATC-----ACCAGGTGACACGTGGCTCTGCTGA GCTC TTTGCTATCTAT
10 GAGAA-GAGC AATGATGAAAGCCGAGC-CTTGGCGATGGTAGGGTCCATCGCAAACATCTGATC-----ACCAGGTGACACGTGGCTCTGCTGA GCTC TTTGCTATCCAT

 3 TTCCC-GAGC TGTGAGGATACCAGTGCTTATTGTGATGATAGGGTCTATCACTACCGTCAGATAT--TCATCGGACGAAAAGGAGCGCATCTGA GCTT GAAACTTTTTTG
 9 TTCCC-GAGC TGTGAGGATACGAGTGCTTATTGTGATGATAGGGTCTATCACTACCGTCAGATATATTCATCGGACGAAAAGAAGCGCTTCTGA GCTT GAAACTTTTTTG
 6 TTCAC-AAGC TGTGAGGAAGTCTGTGCTCATTGTGATGATAGGGTCTATCACTACCGTCTGATCT--CCATCCGACGACAAAGAGTGCATCTGA GCTT GTTTTATTTCTC
 7 TTCAC-AAGC TGTGAGGAAATCTGTGCTCATTGTGATGATAGGGTCTATCACTACCGTCCGATCT--CCATCCGACGACAAAGAGTGCATCTGA GCTT GTTTTATTTTTT
 4 TTCCC-AAGC TATGAGGATATCTGTGCTCAT-GTGATAGAGGGTCTATCACTACCGTCTGATCT-----GCGGACGAATTCGAGTGCATCTGA GCTT GCCCTTTTTCTT
 5 ATAAC-GAGC GGTGAGGACACCTGTGCTCCAAGTGATGGTGGGGTCTATCACTGCCGTCTGATCT-TTCATTGGATGAAAGTGAGCGCACCTGA GCTC TTGATTTTGGTT
 8 AATCCTGAGC TGTGATGACCTCTGTGCTCAACGTGATGATAGGGTCTATCACTACCGTCTGATCA-T-CATCGGACGAAAGCGAGCGCATCTGA GCTC TTGATTTCTGAA
       AGC    TGA GA   c G GC      G GATG T GGGTC ATC  C    C TC GAT       G  GA      G C  CTGAGCT
```

Figure 6. Structural alignment of the 10 C/D box snoRNA genes located beyond the tDNA-Gly (GCC) genes of *P. trichocarpa*. Ten dicistronic tRNA-Gly (GCC)–snoRNA genes were identified in the genome of *P. trichocarpa*. The 10 copies of tDNA are identical and are shown under the usual cloverleaf structure (top). The 10 C/D snoRNA gene sequences shown (bottom) immediately follows nt 73 of the tDNAs. Dashes indicate the antisense element and the asterisk indicates the nucleotide targeting the methylation. Most conserved residues (in bold) are displayed in the last line.

III gene fusion involving a leader tRNA gene. Such fusions might appear unnecessary since every Pol III gene carries its own internal promoter elements. This apparently unlikely gene fusion event is now known to have occurred many more times in hemiascomycetes where a variety of pairs of tRNA genes separated by a very short distance are observed often in multiple copies in genomes (11). Multiple paralogous copies of dicistronic tRNA–snoRNA are also present in plants (14). In *Y. lipolytica*, fusions event including a 5S rRNA gene have occurred at least six times (as six different tDNA types are implicated), some of the fusions being amplified up to 29 times in the case of the tRNA-Gly (GCC)-5S rRNA gene. Similarly, the tRNA-Trp–RUF70 gene was amplified into 13 dispersed copies. We still do not know the possible mechanisms of Pol III gene fusion and proliferation in hemiascomycetes.

A strikingly similar case of 5′ maturation by tRNase Z was very recently reported for the 5S rRNA gene in the euryarchaeota *H. volcanii* (72). The 5S gene is expressed from the ribosomal operon that also includes the 16S and 23S rRNA genes and a distal tRNA gene. The 5S rRNA gene is not preceded by a tRNA gene but the sequences immediately upstream fold into a tRNA-like structure and therefore the 5′ splicing of the 5S rRNA (from the primary transcript) can be performed by tRNase Z. Beside that of *H. volcanii*, tRNase Z from *A. thaliana* or *S. cerevisiae* were also able to perform this maturation *in vitro*. With this information in mind, we examined the sequences upstream the mature snR52, RPR1 and Zod1 RNA genes of *S. cerevisiae* (and related species) which contains regular A- and B-promoter elements (13,73). However, we could not arrange any of these sequences in a way similar to the tRNA-like structure discovered in *H. volcanii*. The quite similar case of tRNA-helped 5S rRNA 5′ processing we report here in *Y. lipolytica* is therefore the only one yet known in eukaryotes.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

Conflict of interest statement. None declared.

REFERENCES

1. Blumenthal,T. (2004) Operons in eukaryotes. *Brief. Funct. Genomic. Proteomic.*, **3**, 199–211.
2. Worthey,E.A., Martinez-Calvillo,S., Schnaufer,A., Aggarwal,G., Cawthra,J., Fazelinia,G., Fong,C., Fu,G., Hassebrock,M., Hixson,G. *et al.* (2003) *Leishmania major* chromosome 3 contains two long convergent polycistronic gene clusters separated by a tRNA gene. *Nucleic Acids Res.*, **31**, 4201–4210.
3. Brown,J.W., Clark,G.P., Leader,D.J., Simpson,C.G. and Lowe,T. (2001) Multiple snoRNA gene clusters from Arabidopsis. *RNA*, **7**, 1817–1832.
4. Chen,C.L., Liang,D., Zhou,H., Zhuo,M., Chen,Y.Q. and Qu,L.H. (2003) The high diversity of snoRNAs in plants: identification and comparative study of 120 snoRNA genes from *Oryza sativa*. *Nucleic Acids Res.*, **31**, 2601–2613.
5. Chanfreau,G., Rotondo,G., Legrain,P. and Jacquier,A. (1998) Processing of a dicistronic small nucleolar RNA precursor by the RNA endonuclease Rnt1. *EMBO J.*, **17**, 3726–3737.
6. Tanzer,A. and Stadler,P.F. (2004) Molecular evolution of a microRNA cluster. *J. Mol. Biol.*, **339**, 327–335.
7. Thatcher,E.J., Bond,J., Paydar,I. and Patton,J.G. (2008) Genomic organization of zebrafish microRNAs. *BMC Genomics*, **9**, 253.
8. Qu,L.H., Henras,A., Lu,Y.J., Zhou,H., Zhou,W.X., Zhu,Y.Q., Zhao,J., Henry,Y., Caizergues-Ferrer,M. and Bachellerie,J.P. (1999) Seven novel methylation guide small nucleolar RNAs are processed from a common polycistronic transcript by Rat1p and RNase III in yeast. *Mol. Cell. Biol.*, **19**, 1144–1158.
9. Geiduschek,E.P. and Kassavetis,G.A. (2001) The RNA polymerase III transcription apparatus. *J. Mol. Biol.*, **310**, 1–26.
10. Dieci,G., Fiorino,G., Castelnuovo,M., Teichmann,M. and Pagano,A. (2007) The expanding RNA polymerase III transcriptome. *Trends Genet.*, **23**, 614–622.
11. Dujon,B., Sherman,D., Fischer,G., Durrens,P., Casaregola,S., Lafontaine,I., De Montigny,J., Marck,C., Neuvéglise,C., Talla,E. *et al.* (2004) Genome evolution in yeasts. *Nature*, **430**, 35–44.
12. Willis,I., Hottinger,H., Pearson,D., Chisholm,V., Leupold,U. and Söll,D. (1984) Mutations affecting excision of the intron from eukaryotic dimeric tRNA precursor. *EMBO J.*, **3**, 1573–1580.
13. Marck,C., Kachouri-Lafond,R., Lafontaine,I., Westhof,E., Dujon,B. and Grosjean,H. (2006) The RNA polymerase III-dependent family of genes in hemiascomycetes: comparative RNomics, decoding strategies, transcription and evolutionary implications. *Nucleic Acids Res.*, **34**, 1816–1835.
14. Kruszka,K., Barneche,F., Guyot,R., Ailhas,J., Meneau,I., Schiffer,S., Marchfelder,A. and Echeverria,M. (2003) Plant dicistronic tRNA-snoRNA genes: a new mode of expression of the small nucleolar RNAs processed by RNase Z. *EMBO J.*, **22**, 621–632.
15. Borchert,G.M., Lanier,W. and Davidson,B.L. (2006) RNA polymerase III transcribes human microRNAs. *Nat. Struct. Mol. Biol.*, **13**, 1097–1101.
16. Pfeffer,S., Sewer,A., Lagos-Quintana,M., Sheridan,R., Sander,C., Grasser,F.A., van Dyk,L.F., Ho,C.K., Shuman,S., Chien,M. *et al.* (2005) Identification of microRNAs of the herpesvirus family. *Nat. Methods*, **2**, 269–276.
17. Wang,C.K. and Weil,P.A. (1989) Purification and characterization of *Saccharomyces cerevisiae* transcription factor IIIA. *J. Biol. Chem.*, **264**, 1092–1099.
18. Gottesfeld,J. and Bloomer,L.S. (1982) Assembly of transcriptionally active 5S RNA gene chromatin in vitro. *Cell*, **28**, 781–791.
19. Miller,J., McLachlan,A.D. and Klug,A. (1985) Repetitive zinc-binding domains in the protein transcription factor IIIA from *Xenopus* oocytes. *EMBO J.*, **4**, 1609–1614.
20. Pieler,T., Hamm,J. and Roeder,R.G. (1987) The 5S gene internal control region is composed of three distinct sequence elements, organized as two functional domains with variable spacing. *Cell*, **48**, 91–100.
21. Clemens,K.R., Liao,X.L., Wolf,V., Wright,P.E. and Gottesfeld,J.M. (1992) Definition of the binding sites of individual zinc fingers in the transcription factor IIIA-5S RNA gene complex. *Proc. Natl Acad. Sci. USA*, **89**, 10822–10826.
22. Hayes,J.J. and Tullius,T.D. (1992) Structure of the TFIIIA-5S DNA complex. *J. Mol. Biol.*, **227**, 407–417.
23. Pelham,H.R.B. and Brown,D.D. (1980) A specific transcription factor that can bind either the 5S RNA gene or 5S RNA. *Proc. Natl Acad. Sci. USA*, **77**, 4170–4174.
24. Guddat,U., Bakken,A.H. and Pieler,T. (1990) Protein-mediated nuclear export of RNA: 5S rRNA containing small RNPs in xenopus oocytes. *Cell*, **60**, 619–628.
25. Honda,B.M. and Roeder,R.G. (1980) Association of a 5S gene transcription factor with 5S RNA and altered levels of the factor during cell differentiation. *Cell*, **22**, 119–126.
26. Liao,X.L., Clemens,K.R., Tennan,L., Wright,P.E. and Gottesfeld,J.M. (1992) Specific interaction of the first three zinc fingers of TFIIIA with the internal control region of the *Xenopus* 5S RNA gene. *J. Mol. Biol.*, **223**, 857–871.
27. Theunissen,O., Rudt,F., Guddat,U., Mentzel,H. and Pieler,T. (1992) RNA and DNA binding zinc fingers in *Xenopus* TFIIIA. *Cell*, **71**, 679–690.
28. Clemens,K.R., Wolf,V., McBryant,S.J., Zhang,P., Liao,X., Wright,P.E. and Gottesfeld,J.M. (1993) Molecular basis for specific recognition of both RNA and DNA by a zinc finger protein. *Science*, **260**, 530–533.
29. Lee,B.M., Xu,J., Clarkson,B.K., Martinez-Yamout,M.A., Dyson,H.J., Case,D.A., Gottesfeld,J.M. and Wright,P.E. (2006) Induced fit and "lock and key" recognition of 5S RNA by zinc fingers of transcription factor IIIA. *J. Mol. Biol.*, **357**, 275–291.
30. Camier,S., Dechampesme,A.M. and Sentenac,A. (1995) The only essential function of TFIIIA in yeast is the transcription of 5S rRNA genes. *Proc. Natl Acad. Sci. USA*, **92**, 9338–9342.
31. Rothfels,K., Rowland,O. and Segall,J. (2007) Zinc fingers 1 and 7 of yeast TFIIIA are essential for assembly of a functional transcription complex on the 5S RNA gene. *Nucleic Acids Res.*, **35**, 4869–4881.
32. van Heerikhuizen,H., Ykema,A., Klootwijk,J., Gaillardin,C., Ballas,C. and Fournier,P. (1985) Heterogeneity in the ribosomal RNA genes of the yeast *Yarrowia lipolytica*; cloning and analysis of two size classes of repeats. *Gene*, **39**, 213–222.
33. Clare,J.J., Davidow,L.S., Gardner,D.C. and Oliver,S.G. (1986) Cloning and characterisation of the ribosomal RNA genes of the dimorphic yeast, *Yarrowia lipolytica*. *Curr. Genet.*, **10**, 449–452.
34. Querol,A., Barrio,E., Huerta,T. and Ramon,D. (1992) Molecular monitoring of wine fermentations conducted by active dry yeast strains. *Appl. Environ. Microbiol.*, **58**, 2948–2953.
35. Huet,J., Manaud,N., Dieci,G., Peyroche,G., Conesa,C., Lefebvre,O., Ruet,A., Riva,M. and Sentenac,A. (1996) RNA polymerase III and class III transcription factors from *Saccharomyces cerevisiae*. *Methods Enzymol.*, **273**, 249–267.
36. Deprez,E., Arrebola,R., Conesa,C. and Sentenac,A. (1999) A subunit of yeast TFIIIC participates in the recruitment of TATA-binding protein. *Mol. Cell. Biol.*, **19**, 8042–8051.
37. Huet,J., Conesa,C., Carles,C. and Sentenac,A. (1997) A cryptic DNA binding domain at the COOH terminus of TFIIIB70 affects formation, stability and function of preinitiation complexes. *J. Biol. Chem.*, **272**, 18341–18349.
38. Kassavetis,G.A., Joazeiro,C.A., Pisano,M., Geiduschek,E.P., Colbert,T., Hahn,S. and Blanco,J.A. (1992) The role of the TATA-binding protein in the assembly and function of the multisubunit yeast RNA polymerase III transcription factor, TFIIIB. *Cell*, **71**, 1055–1064.
39. Ferrari,R., Rivetti,C., Acker,J. and Dieci,G. (2004) Distinct roles of transcription factors TFIIIB and TFIIIC in RNA polymerase III transcription reinitiation. *Proc. Natl Acad. Sci. USA*, **3**, 13442–13447.
40. Ducrot,C., Lefebvre,O., Landrieux,E., Guirouilh-Barbat,J., Sentenac,A. and Acker,J. (2006) Reconstitution of the yeast RNA polymerase III transcription system with all recombinant factors. *J. Biol. Chem.*, **281**, 11685–11692.
41. Fickers,P., Le Dall,M.T., Gaillardin,C., Thonart,P. and Nicaud,J.M. (2003) New disruption cassettes for rapid gene disruption and marker rescue in the yeast *Yarrowia lipolytica*. *J. Microbiol. Methods*, **55**, 727–737.
42. Le Dall,M.T., Nicaud,J.M. and Gaillardin,C. (1994) Multiple-copy integration in the yeast *Yarrowia lipolytica*. *Curr. Genet.*, **26**, 38–44.

43. Oficjalska-Pham,D., Harismendy,O., Smagowicz,W.J., Gonzalez de Peredo,A., Boguta,M., Sentenac,A. and Lefebvre,O. (2006) General repression of RNA polymerase III transcription is triggered by protein phosphatase type 2A-mediated dephosphorylation of Maf1. *Mol. Cell.*, **22**, 623–632.

44. Sethy-Coraci,I., Moir,R.D., Lopez-de-Leon,A. and Willis,I.M. (1998) A differential response of wild type and mutant promoters to TFIIIB70 overexpression *in vivo* and *in vitro*. *Nucleic Acids Res.*, **26**, 2344–2352.

45. Li,Y., Moir,R.D., Sethy-Coraci,I.K., Warner,J.R., Willis,I.M., Huang,Y., McGillicuddy,E., Weindel,M., Dong,S. and Maraia,R.J. (2000) Repression of ribosome and tRNA synthesis in secretion-defective cells is signaled by a novel branch of the cell integrity pathway. *Mol. Cell. Biol.*, **20**, 3843–3851.

46. The Génolevures-3 Consortium. (2008) Comparative genomics of five hemiascomycetous species identifies the basic yeast genome repertoire. *Genome Research*, in preparation.

47. Neuvéglise,C., Bon,E., Lépingle,A., Wincker,P., Artiguenave,F., Gaillardin,C. and Casaregola,S. (2000) Genomic exploration of the hemiascomycetous yeasts: 9. *Saccharomyces kluyveri*. *FEBS Lett.*, **487**, 56–60.

48. Malpertuy,A., Llorente,B., Blandin,G., Artiguenave,F., Wincker,P. and Dujon,B. (2000) Genomic exploration of the hemiascomycetous yeasts: 10. *Kluyveromyces thermotolerans*. *FEBS Lett.*, **487**, 61–65.

49. Bolotin-Fukuhara,M., Toffano-Nioche,C., Artiguenave,F., Duchateau-Nguyen,G., Lemaire,M., Marmeisse,R., Montrocher,R., Robert,C., Termier,M., Wincker,P. *et al.* (2000) Genomic exploration of the hemiascomycetous yeasts: 11. *Kluyveromyces lactis*. *FEBS Lett.*, **487**, 66–70.

50. Wendland,J., Pohlmann,R., Dietrich,F., Steiner,S., Mohr,C. and Philippsen,P. (1999) Compact organization of rRNA genes in the filamentous fungus *Ashbya gossypii*. *Curr. Genet.*, **35**, 618–625.

51. Lépingle,A., Casaregola,S., Neuvéglise,C., Bon,E., Nguyen,H., Artiguenave,F., Wincker,P. and Gaillardin,C. (2000) Genomic exploration of the hemiascomycetous yeasts: 14. *Debaryomyces hansenii* var. *hansenii*. *FEBS Lett.*, **487**, 82–86.

52. Wood,V., Gwilliam,R., Rajandream,M.A., Lyne,M., Lyne,R., Stewart,A., Sgouros,J., Peat,N., Hayles,J., Baker,S. *et al.* (2002) The genome sequence of *Schizosaccharomyces pombe*. *Nature*, **415**, 871–880.

53. Reyes,V.M., Newman,A. and Abelson,J. (1986) Mutational analysis of the coordinate expression of the yeast tRNAArg-tRNAAsp gene tandem. *Mol. Cell. Biol.*, **6**, 2436–2442.

54. Chakrabarti,K., Pearson,M., Grate,L., Sterne-Weiler,T., Deans,J., Donohue,J.P. and Ares,M. Jr. (2007) Structural RNAs of known and unknown function identified in malaria parasites by comparative genomics and RNA analysis. *RNA*, **13**, 1923–1939.

55. Engelke,D.R., Gegenheimer,P. and Abelson,J. (1985) Nucleolytic processing of a tRNAArg-tRNAAsp dimeric precursor by a homologous component from *Saccharomyces cerevisiae*. *J. Biol. Chem.*, **260**, 1271–1279.

56. Klekamp,M.S. and Weil,P.A. (1982) Specific transcription of homologous class III genes in yeast-soluble cell-free extract. *J. Biol. Chem.*, **257**, 8432–8441.

57. Giaever,G., Chu,A.M., Ni,L., Connelly,C., Riles,L., Veronneau,S., Dow,S., Lucau-Danila,A., Anderson,K., Andre,B. *et al.* (2002) Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature*, **418**, 387–391.

58. Ginsberg,A.M., King,B.O. and Roeder,R.G. (1984) *Xenopus* 5S gene transcription factor, TFIIIA: characterization of a cDNA clone and measurement of RNA levels throughout development. *Cell*, **39**, 479–489.

59. Tso,J.Y., Van Den Berg,D.J. and Korn,L.J. (1986) Structure of the gene for Xenopus transcription factor TFIIIA. *Nucleic Acids Res.*, **14**, 2187–2200.

60. Archambault,J., Milne,C.A., Schappert,K.T., Baum,B., Friesen,J.D. and Segall,J. (1992) The deduced sequence of the transcription factor TFIIIA from *Saccharomyces cerevisiae* reveals extensive divergence from *Xenopus* TFIIIA. *J. Biol. Chem.*, **267**, 3282–3288.

61. Milne,C.A. and Segall,J. (1993) Mapping regions of yeast transcription factor IIIA required for DNA binding, interaction with transcription factor IIIC and transcription activity. *J. Biol. Chem.*, **268**, 11364–11371.

62. Schulman,D.B. and Setzer,D.R. (2002) Identification and characterization of transcription factor IIIA from *Schizosaccharomyces pombe*. *Nucleic Acids Res.*, **30**, 2772–2781.

63. Takaku,H., Minagawa,A., Takagi,M. and Nashimoto,M. (2003) A candidate prostate cancer susceptibility gene encodes tRNA 3′ processing endoribonuclease. *Nucleic Acids Res.*, **31**, 2272–2278.

64. Chen,Y., Beck,A., Davenport,C., Chen,Y., Shattuck,D. and Tavtigian,S.V. (2005) Characterization of TRZ1, a yeast homolog of the human candidate prostate cancer susceptibility gene ELAC2 encoding tRNase Z. *BMC Mol. Biol.*, **6**, 12.

65. Späth,B., Settele,F., Schilling,O., D'Angelo,I., Vogel,A., Feldmann,I., Meyer-Klaucke,W. and Marchfelder,A. (2007) Metal requirements and phosphodiesterase activity of tRNase Z enzymes. *Biochemistry*, **46**, 14742–14750.

66. Späth,B., Schubert,S., Lieberoth,A., Settele,F., Schutz,S., Fischer,S. and Marchfelder,A. (2008) Two archaeal tRNase Z enzymes: similar but different. *Arch Microbiol*, ePub.

67. Joazeiro,C.A., Kassavetis,G.A. and Geiduschek,E.P. (1996) Alternative outcomes in assembly of promoter complexes: the roles of TBP and a flexible linker in placing TFIIIB on tRNA genes. *Genes Dev.*, **10**, 725–739.

68. Lee,Y. and Nazar,R.N. (2003) Terminal structure mediates 5 S rRNA stability and integration during ribosome biogenesis. *J. Biol. Chem.*, **278**, 6635–6641.

69. van Hoof,A., Lennertz,P. and Parker,R. (2000) Three conserved members of the RNase D family have unique and overlapping functions in the processing of 5S, 5.8S, U4, U5, RNase MRP and RNase P RNAs in yeast. *EMBO J.*, **19**, 1357–1365.

70. Souciet,J., Aigle,M., Artiguenave,F., Blandin,G., Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S., de Montigny,J., Dujon,B. *et al.* (2000) Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies. *FEBS Lett.*, **487**, 3–12.

71. Deng,W., Zhu,X., Skogerbo,G., Zhao,Y., Fu,Z., Wang,Y., He,H., Cai,L., Sun,H., Liu,C. *et al.* (2006) Organization of the *Caenorhabditis elegans* small non-coding transcriptome: genomic features, biogenesis and expression. *Genome Res.*, **16**, 20–29.

72. Hölzle,A., Fischer,S., Heyer,R., Schütz,S., Zacharias,M., Walther,P., Allers,T. and Marchfelder,A. (2008) Maturation of the 5S rRNA 5′ end is catalyzed in vitro by the endonuclease tRNase Z in the archaeon *H. volcanii*. *RNA*, **14**, 928–937.

73. Guffanti,E., Percudani,R., Harismendy,O., Soutourina,J., Werner,M., Iacovella,M.G., Negri,R. and Dieci,G. (2006) Nucleosome depletion activates poised RNA polymerase III at unconventional transcription sites in *Saccharomyces cerevisiae*. *J. Biol. Chem.*, **281**, 29155–29164.