
This copy is for your personal, non-commercial use only.

If you wish to distribute this article to others, you can order high-quality copies for your colleagues, clients, or customers by [clicking here](#).

Permission to republish or repurpose articles or portions of articles can be obtained by following the guidelines [here](#).

The following resources related to this article are available online at www.sciencemag.org (this information is current as of June 24, 2011):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:

<http://www.sciencemag.org/content/302/5649/1399.full.html>

Supporting Online Material can be found at:

<http://www.sciencemag.org/content/suppl/2003/11/20/302.5649.1399.DC1.html>

This article has been **cited by** 94 article(s) on the ISI Web of Science

This article has been **cited by** 53 articles hosted by HighWire Press; see:

<http://www.sciencemag.org/content/302/5649/1399.full.html#related-urls>

This article appears in the following **subject collections**:

Molecular Biology

http://www.sciencemag.org/cgi/collection/molec_biol

Nucleolar Clustering of Dispersed tRNA Genes

Martin Thompson, Rebecca A. Haeusler,
Paul D. Good, David R. Engelke*

Early transfer RNA (tRNA) processing events in *Saccharomyces cerevisiae* are coordinated in the nucleolus, the site normally associated with ribosome biosynthesis. To test whether spatial organization of the tRNA pathway begins with nucleolar clustering of the genes, we have probed the subnuclear location of five different tRNA gene families. The results show that tRNA genes, though dispersed in the linear genome, colocalize with 5S ribosomal DNA and U14 small nucleolar RNA at the nucleolus. Nucleolar localization requires tRNA gene transcription-complex formation, because inactivation of the promoter at a single locus removes its nucleolar association. This organization of tRNA genes must profoundly affect the spatial packaging of the genome and raises the question of whether gene types might be coordinated in three dimensions to regulate transcription.

Little is known about how specific nuclear genes are arranged in three dimensions, although the position of some chromosome regions are dynamic and responsive to the general transcription state (1). Examples of documented DNA positions are the telomeric regions, preferentially found near the nuclear periphery in yeast, and the tandemly repeated ribosomal RNA (rRNA) genes, which create distinctive nucleolar structures where ribosome assembly is coordinated.

Components of the early tRNA processing pathway in *Saccharomyces cerevisiae* and other eukaryotes are often at the nucleolus (2–5), and additional observations suggest that transcription of the tRNA genes might also be nucleolar. Transcription by RNA polymerase II (Pol II) is suppressed in the vicinity of tRNA genes (6, 7), and promoters for Pol II are underrepresented within 500 base pairs upstream of tRNA genes, with the exception of Ty retrotransposons (8, 9). This silencing near tRNA genes is relieved by a mutation in the rRNA processing enzyme, *CBF5* (10). These observations suggest that tRNA and rRNA biogenesis might be coordinated at the nucleolus and that spatial coordination might begin by localizing the tRNA genes at the nucleolus (10). Such a clustering of tRNA genes would create a region that was highly concentrated for Pol III as well as Pol I transcription. In yeast, the Pol III-transcribed 5S rRNA genes are nucleolar by definition, because they adjoin the large rRNA gene repeats (11). Nucleolar localization of 5S rRNA genes has also been found in higher eukaryotes (12–16), in which the 5S gene clusters are not attached to the large rRNA genes in the linear DNA.

We investigated the location of five different families of tRNA genes: tRNA^{Leu}(CAA) and tRNA^{Lys}(CUU), which contain introns, and tRNA^{Gly}(GCC), tRNA^{Gln}(UUG), and tRNA^{Glu}(UUC), which do not contain introns. Each family contains 9 to 16 genes at

scattered positions throughout the yeast genome (fig. S1). Fluorescent oligonucleotide probes specific for the antisense strand of the genes were hybridized to fixed, permeabilized yeast, and the signal was compared to that obtained with probes to either a small nucleolar RNA (U14 snoRNA) or a known nucleolar DNA segment from chromosome XII, the 5S rRNA gene (17).

Most of the signal from all five families of tRNA genes overlaps the 5S ribosomal DNA (rDNA) at the nucleolus (Fig. 1). The tRNA genes tend to be on the periphery of the 5S rDNA signal, but it is not clear whether they are within or at the edge of the nucleolus (5S rDNA and U14 snoRNA positions). tRNA gene probes were specific for DNA, because the signal was sensitive to deoxyribonuclease but not ribonuclease or proteinase (fig. S2). We are probably able to detect only clustered tRNA genes; individual genes would likely fall below the detection limit. Consistent with this, some pre-tRNAs are nucleoplasmic, although most of the pre-tRNA signal is nucleolar (2, 8, 18, 19). The nucleolar signal is specific to tRNA genes, because telomeric probes give a signal in the nucleoplasm and nuclear periphery, as expected (20).

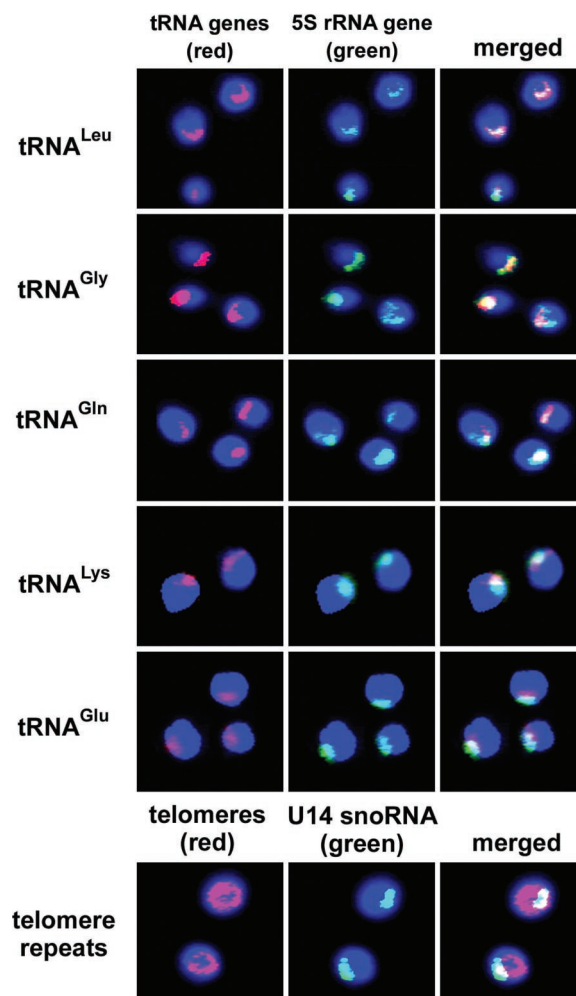


Fig. 1. Nucleolar localization of tRNA genes. Fluorescent oligonucleotide probes were annealed to individual tRNA gene families (Leu, Gly, Gln, Lys, and Glu) or to the telomeric repeats. Each family of tRNA genes has 9 to 16 members that are dispersed in the linear genome map (28) (fig. S1). The Cy3-labeled probes to the tRNA genes and telomeres are red, and the Oregon Green 488 probe to the 5S rRNA gene is green. 4',6'-diamidino-2-phenylindole (DAPI) staining of the nucleoplasm is shown as blue. Most of the tRNA gene signal overlaps the nucleolar 5S rRNA genes. In contrast, the telomeric repeat probe stains the nuclear periphery and nucleoplasm. Cells shown are representative of multiple experiments.

Department of Biological Chemistry, University of Michigan, Ann Arbor, MI 48109–0606, USA.

*To whom correspondence should be addressed. E-mail: engelke@umich.edu

REPORTS

To test whether tRNA gene position is dependent on Pol III transcription, we compared the position of a single tRNA gene locus, *SUP53* on chromosome III (Fig. 2), to the position of the same locus after the *SUP53* promoter had been inactivated by point mutations that precluded stable association of all Pol III transcription-complex components (21). When the tRNA gene is active, the *SUP53* signal overlaps the nucleolar signal ~50% of the time, which is substantially more frequent than the signal overlap that occurs when the tRNA gene is inactive (52% of 440 cells compared with 13% of 715 cells). Similar distributions of the locus were found when the same active and inactive tRNA genes were placed with *URA3* on a low-copy plasmid (7) (fig. S3). These results suggest that a given tRNA gene might not necessarily be found continuously at the nucleolus but that the tendency for nucleolar association is dependent on some aspect of transcription-complex formation.

To test whether tRNA gene clustering is sensitive to nucleolar perturbation, we probed the tRNA^{Leu} family in cells with a nonlethal

deletion of the Pol I subunit, $\Delta rpa49$ (Fig. 3), which compromises Pol I transcription and results in a less compact nucleolus than normal (22) (Fig. 3). In keeping with results from previous nucleolar disruption (10), the pre-tRNA transcript signal in $\Delta rpa49$ disperses into the nucleoplasm. The tRNA gene signal disappears, because the dispersed tRNA genes no longer individually produce enough of a signal to be seen above the background fluorescence. We conclude that tRNA gene localization is, to some extent, dependent on normal rRNA biosynthesis in the nucleolus.

Clustering tRNA genes near 5S rRNA genes in the nucleolus forms a nuclear subregion specializing in Pol III transcription. Transcription complexes on 5S rRNA genes and tRNA genes both contain Pol III and two multisubunit transcription factors, TFIIB and TFIIC, although the 5S rRNA genes require one extra recognition factor, TFIIA (23). Spatial coordination of the Pol III-transcribed genes would allow concentration of these factors, as well as initiating nucleolar organization of the early tRNA processing pathway (1, 9).

Localization of most tRNA genes to the yeast nucleolus has drastic implications for three-dimensional genome organization, because the genes are scattered throughout the linear genome map (fig. S1). It is not clear whether these results reflect the organization of tRNA genes in metazoans. There is currently little information on the position of actively transcribed tRNA-class promoters in metazoans, including both tRNA genes and short interspersed nuclear elements (24). Further, morphology of nucleoli in multicellular eukaryotes is distinct from the single nucleolus in yeast.

The clustering of tRNA genes in subnuclear regions specific for Pol III expression (or Pol I and Pol III expression) is consistent with Pol II transcription being relatively poor near actively transcribed tRNA genes (7, 10) and with Pol II promoters being under-represented near tRNA genes (9). A case of transcription-unit segregation, the rRNA genes in nucleoli, results in the exclusion of Pol II from nucleoli and the silencing of Pol II transcription units when they are inserted between rRNA gene repeats (25). If Pol III transcription units are also localizing in or near the nucleolus, Pol II transcription might be suppressed by either simple exclusion of transcription components or a more active mechanism. This clustering of tRNA genes in specialized subnuclear locations might also partially explain why propagating a chromatin-modeling signal beyond a tRNA gene is difficult, as observed in cases where tRNA genes serve as boundary elements for chromatin domains (26, 27).

Fig. 2. tRNA gene localization depends on Pol III complex formation. The *SUP53* tRNA gene locus on chromosome III was probed by replacing the neighboring *LEU2* coding region with *URA3* coding sequences and simultaneously probing fixed cells with 14 fluorescent oligonucleotides (red) complementary to the non-RNA strand of *URA3*. Nucleoli were identified with a fluorescent oligonucleotide to U14 snoRNA (green), and the nucleoplasm is stained with DAPI (blue). The endogenous *URA3* coding sequences are deleted from chromosome V in these strains. As a negative control, a strain in which the *URA3* sequences were deleted by replacement with a kanamycin resistance gene (Kan) showed no signal with the *URA3* probes (lowest panel). *SUP53* frequently localized in or showed some overlap with the nucleolus when transcriptionally active (52% out of 440 cells imaged), but did so less often when the promoter was mutated to be inactive (13% out of 715 cells imaged).

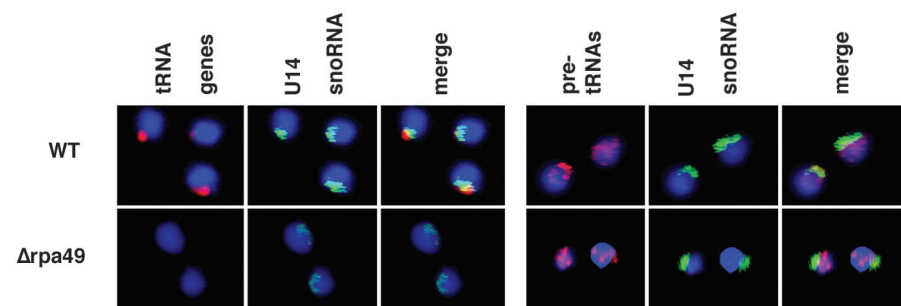
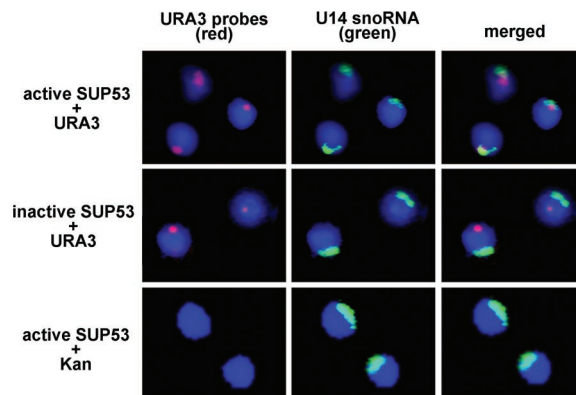


Fig. 3. Pol I affects tRNA gene clusters. The positions (red) of the tRNA^{Leu} gene family and its pre-tRNA transcripts were probed (17) in a strain with a deleted, nonessential subunit of Pol I ($\Delta rpa49$) and compared with the parental wild-type strain (WT). Nucleoli are marked by a probe to the U14 snoRNA (green) and nucleoplasm is stained with DAPI (blue) in all panels.

References and Notes

1. S. M. Gasser, *Science* **296**, 1412 (2002).
2. E. Bertrand, F. Houser-Scott, A. Kendall, R. H. Singer, D. R. Engelke, *Genes Dev.* **12**, 2463 (1998).
3. B. Lee, A. G. Matera, D. C. Ward, J. Craft, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 11471 (1996).
4. M. R. Jacobson, L.-G. Cao, K. Taneja, R. H. Singer, Y.-L. Wang, *J. Cell Sci.* **110**, 829 (1997).
5. N. Jarrous, J. S. Wolenski, D. Wesolowski, C. Lee, S. Altman, *J. Cell Biol.* **146**, 559 (1999).
6. P. T. Kinsey, S. B. Sandmeyer, *Nucleic Acids Res.* **19**, 1317 (1991).
7. M. W. Hull, J. Erickson, M. Johnston, D. R. Engelke, *Mol. Cell Biol.* **14**, 3244 (1994).
8. J. D. Boeke, S. B. Sandmeyer, in *The Molecular and Cellular Biology of the Yeast Saccharomyces cerevisiae*, J. Broach, E. Jones, J. Pringle, Eds. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1991), pp. 193–261.
9. E. C. Bolton, J. D. Boeke, *Genome Res.* **13**, 254 (2003).
10. A. Kendall *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 13108 (2001).
11. Y. Lee, A. M. Erkin, V. I. Van Ryke, R. N. Nazar, *Nucleic Acids Res.* **23**, 634 (1995).
12. R. L. Metzberg, J. N. Stevens, E. U. Selker, E. Morzycka-Wroblewska, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 2067 (1985).
13. S. Narayanswami, B. A. Hamkalo, *Cytometry* **11**, 144 (1990).
14. F. S. Kaplan *et al.*, *Genomics* **15**, 123 (1993).
15. M. I. Highett, A. F. Beven, P. J. Shaw, *J. Cell Sci.* **105**, 1151 (1993).
16. M. B. Montijn, A. B. Noutsmuller, R. tenHooft, J. L. Oud, N. Nanninga, *Chromosome Res.* **7**, 387 (1999).

17. Materials and methods are available as supporting material on Science Online and are derived from protocols at http://singerlab.aecom.yu.edu/protocols/insitu_yeast.htm.
18. J. P. O'Connor, C. L. Peebles, *Mol. Cell. Biol.* **11**, 425 (1991).
19. S. Sarkar, A. K. Hopper, *Mol. Biol. Cell* **11**, 3041 (1998).
20. F. Hediger, F. R. Neumann, G. Van Houwe, K. Dubrana, S. M. Gasser, *Curr. Biol.* **12**, 2076 (2002).
21. J. M. Huibregtse, D. R. Engelke, *Mol. Cell. Biol.* **4**, 3244 (1989).
22. P. Liljelund, S. Sariotte, J. M. Buhler, A. Sentenac, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 9302 (1992).
23. M. R. Paule, R. J. White, *Nucleic Acids Res.* **28**, 1283 (2000).
24. P. L. Deininger, M. A. Batzer, *Genome Res.* **12**, 1455 (2002).
25. J. S. Smith, J. D. Boeke, *Genes Dev.* **11**, 241 (1997).
26. D. Donze, C. R. Adams, J. Rine, R. T. Kamakaka, *Genes Dev.* **13**, 698 (1999).
27. D. Donze, R. T. Kamakaka, *EMBO J.* **20**, 520 (2001).
28. The *Saccharomyces* Genome Database is available at <http://genome-www.stanford.edu/Saccharomyces/>.
29. We thank A. Hopper and members of the Engelke lab for experimental suggestions, and D. Thiele, E. Phisicky, and J. Abelson for helpful comments on the manuscript. Supported by research grant GM63142 from NIH.

Supporting Online Material
www.sciencemag.org/cgi/content/full/302/5649/1399/DC1
 Materials and Methods
 Figs. S1 to S3
 References

30 July 2003; accepted 10 October 2003

The Origins of Genome Complexity

Michael Lynch^{1*} and John S. Conery²

Complete genomic sequences from diverse phylogenetic lineages reveal notable increases in genome complexity from prokaryotes to multicellular eukaryotes. The changes include gradual increases in gene number, resulting from the retention of duplicate genes, and more abrupt increases in the abundance of spliceosomal introns and mobile genetic elements. We argue that many of these modifications emerged passively in response to the long-term population-size reductions that accompanied increases in organism size. According to this model, much of the restructuring of eukaryotic genomes was initiated by nonadaptive processes, and this in turn provided novel substrates for the secondary evolution of phenotypic complexity by natural selection. The enormous long-term effective population sizes of prokaryotes may impose a substantial barrier to the evolution of complex genomes and morphologies.

The ~100 fully sequenced eubacterial and archaeal genomes contain between 350 and 6000 genes, packed into 0.6 to 7.6 megabases (Mb) (*1*). Whereas some unicellular eukaryotes have genomes well within the range of these prokaryotes (such as 2000 genes in 2.9 Mb for the parasitic microsporidian *Encephalitozoon cuniculi*), all well-characterized genomes of multicellular animals and plants contain more than 13,000 genes in at least 100 Mb. The amount of DNA associated with just 30 human genes is equivalent to the entire genome size of an average prokaryote. Accompanying the increase in gene number in multicellular species is an expansion in the size and number of intragenic spacers (introns) and a dramatic proliferation of mobile genetic elements.

It remains unclear whether the expansions of genome size and complexity during eukaryotic evolution were essential for adaptive phenotypic diversification. After all, there are many ways to generate multiple functions from individual genes, such as tissue-specific gene regulation, alternative splicing, and RNA editing. In addition, the millions of

mobile elements in the human genome and the massive increase in the average intron size in some multicellular eukaryotes have no obvious advantages. Finally, given that some prokaryotes are capable of cell differentiation, have linear chromosomes, and in rare cases have nuclear membranes, it is unclear whether the relatively simple genomes of microbes are merely reflections of unusual physiological constraints. Any general theory of genomic architecture evolution must account for the peculiar molecular attributes of various genetic elements, in addition to being compatible with the principles of population genetics. We argue here that the transitions from prokaryotes to unicellular eukaryotes to multicellular eukaryotes are associated with orders-of-magnitude reductions in population size; by magnifying the power of random genetic drift, reduced population size provides a permissive environment for the proliferation of various genomic features that would otherwise be eliminated by purifying selection.

Direct counts from multicellular and unicellular eukaryotes consistently show an inverse relationship between population density per unit of area and average individual body mass within a species (*2–5*). Such scaling need not reflect the pattern for total population size, given that it does not account for total species ranges. Moreover, the total abundance of a species need not reflect the

more evolutionarily relevant genetic effective population size (N_e), which determines the degree to which gene frequencies are faithfully transmitted across generations. For example, a large population can behave genetically like a small one if a minor fraction of individuals contribute to the reproductive pool or if beneficial chromosomal segments periodically sweep through the population. Insight into long-term effective population sizes can be acquired from the nucleotide variation at silent sites in protein-coding genes (i.e., sites at which a nucleotide substitution leaves the encoded amino acid unchanged). The rate of introduction of new variation per site in two randomly compared alleles is $2u$ (twice the mutation rate per nucleotide), whereas the expected rate of loss of variation from neutral sites is $1/(2N_e)$ in a randomly mating diploid population. At equilibrium, the average number of nucleotide substitutions at neutral sites is $4N_e u$, with slight modifications required for other modes of inheritance (*1*). Thus, levels of silent-site variation among random alleles within a species provide an estimate of the composite parameter $N_e u$.

In a broad phylogenetic sense, there is an inverse relationship between organism size and $N_e u$. Proceeding from top to bottom of Fig. 1A, with two exceptions (*Streptococcus pyogenes* and *Pseudomonas aeruginosa*), all surveyed prokaryotes have $N_e u > 0.025$, whereas, with the exception of the malarial parasite *Plasmodium falciparum* and the ciliate *Tetrahymena thermophila*, the physically larger unicellular eukaryotes have $0.0035 < N_e u < 0.025$. For the still larger vascular plants and invertebrates, $0.00077 < N_e u < 0.0037$, whereas for vertebrates, $0.00027 < N_e u < 0.0010$. N_e can be disentangled from u by noting that the mutation rate per base per cell division ranges from 5×10^{-11} to 5×10^{-10} , with an average value of $\sim 2.3 \times 10^{-10}$ (*6*). This implies that N_e is generally greater than 10^8 for prokaryotes and often in the range of 10^7 to 10^8 for unicellular eukaryotes. The number of germline cell divisions per generation is ~ 10 in nematodes and ~ 25 in flies (*6*), implying that N_e is in the range of $\sim 10^5$ to 10^6 for invertebrates; the number of germline cell divi-

¹Department of Biology, Indiana University, Bloomington, IN 47405, USA. ²Department of Computer and Information Science, University of Oregon, Eugene, OR 97403, USA.

*To whom correspondence should be addressed. E-mail: mlynch@bio.indiana.edu