Genome Organization of the Three Identical ATP1 Genes on the Left Arm of Chromosome II of Saccharomyces cerevisiae: Sequence Analysis of the 35-kb Region Containing Three ATP1 Genes

Masaharu Takeda1,2,*, Shohei Nakamura1 and Satoshi Matsushita1

1Department of Applied Life Science, Sojo University, 4-22-1 Ikeda, Kumamoto 860-0082, Japan; 2Department of Materials and Biological Engineering, Tsuruoka National College of Technology, Tsuruoka, Yamagata 997-8511, Japan

Abstract: The ATP1 gene (YBL099w) encoding the F1F0-ATPase complex α subunit of Saccharomyces cerevisiae is present on the left arm close to the telomere of chromosome II, and only one copy was reported by the Genome Project. Recently, we reported that three ATP1s designated ATP1a, ATP1b and ATP1c are located on chromosome II, with different distances between ATP1a and ATP1b, and ATP1b and ATP1c as identified previously. To elucidate the ATP1 replication and their junction sites, we report here the complete nucleotide sequence of this region (approximately 35 kb) on the left arm of chromosome II in Saccharomyces cerevisiae using the prime clones 70113 and 70804 from ATCC, and genomic DNAs from various yeast strains S288C, DC5, W303-1A and those of the gene copy-specific ATP1-disruptants. The nucleotide sequences of the three ATP1s were identical, and they were repeated along with differing amounts of neighboring DNA sequences.

The nucleotide sequence of the ATP1 repeat region has been deposited to DDBJ (AB304259, 20070511181828.08607).

Keyword: ATP1, Nucleotide sequence, Repetitive genes, Junction, Saccharomyces cerevisiae.

INTRODUCTION

Mitochondrial ATP synthase (F1F0-ATPase complex = F1:F0) functions as a key enzyme for ATP production in eukaryotic cells [1]. The enzyme is controlled in response to the energy demands of cells [2]. The enzyme complex is composed of the F1-ATPase (F1) and the transmembrane sector, or proton channel (F0) [3, 4]. Both F1 and F0 are necessary for ATP synthase activity, whereas F1 alone retains the ability to hydrolyze ATP [5].

F1 consists of five different subunits, α, β, γ, δ and ε in a stoichiometry of 3:3:1:1:1 in all aerobic cells. In the yeast S. cerevisiae, the subunits are encoded by the nuclear genes ATP1, ATP2, ATP3, ATP16 and ATP15, respectively. The minimum unit for F1 resides on the Fε-subunit [6]. The catalytic center is considered to be the β-subunit [7]. Recently, we revealed that multiple copies of the ATP1, ATP2, and ATP3 genes are arranged in tandem on each chromosome on which these genes are located [8-11], which were apparently different copy number from the sequence reported by the Genome Project [12-14]. The copy numbers of these F1:F0 subunit genes were not coincident with the subunit-stoichiometry of F1 mentioned above.

Recently, we have revealed that the four bases should be arranged in a sophisticated fashion in the genome, and DNA sequences were deeply affected by the adjoining sequences. The non-coding sequences might play some important roles to express each gene (the coding sequences) in genome. That is, not only the coding region, but also the non-coding region might be necessary to transmit and to transform the biological information precisely, rapidly, and stably [15, 16]. Therefore, in the case of the discussion for the gene(s) in living cells, the entire structure of the genomic DNA including both coding- and non-coding regions should be targeted, and many biological phenomena might be deeply affected on the genomic DNAs [15-17].

From these results, these genes were repeated twice or three times accompanied by neighboring ORFs and other DNA sequences on each chromosome as reported by the Genome Project. Therefore, more repetitive genes such as the F1:F0 subunit genes might be present on various chromosomes than previously thought. Gene repetition on each chromosome might make sense biologically and evolutionarily in addition to providing gene-backups. The DNA sequences to engender the gene repetition might be present on genomic DNA.

The ATP1 gene encoding the F1-α subunit (F1:α) of S. cerevisiae [18, 19] is present on the left arm close to the telomere of chromosome II. We report here the complete nucleotide sequence of the ATP1-repeated region (approximately 35 kb) on the left arm of chromosome II, and reveal that three ATP1 genes are arranged in tandem on chromosome II accompanied by the neighboring ORFs as reported previously [8, 9]. In this manuscript, the precise distances between the three ATP1 genes were determined by the complete nucleotide sequence analysis of the repeated region and junction sequences, and a common nucleotide sequence that was observed upstream of the three ATP1s and other F1:F0 subunit genes was discussed.
**MATERIALS AND METHODS**

**Yeast Strains**

Yeast strains used in this study were *Saccharomyces cerevisiae* DC5 (MATa, leu2-3, leu2-112, his3, can1-11), LL20 (MATa, leu2, his3), W301-1A (MATa, leu2-3, leu2-112, his3-11 his3-15, trp1-1, ura3-1, ade2-1, can1-100), W303-1B (MATa, leu2-3, leu2-112, his3-11 his3-15, trp1-1, ura3-1, ade2-1, can1-100, atp1a::HIS3), SKY4A11 (MATa, leu2-3, leu2-112, his3-11 his3-15, trp1-1, ura3-1, ade2-1, can1-100, atp1c::HIS3) and TKY4011 (MATa, leu2-3, leu2-112, his3-11 his3-15, trp1-1, ura3-1, ade2-1, can1-100, atp1c::HIS3).

**E. Coli Strains**

Sure (el4 - (McrA -), (mcrCB-hsdSMR-mrr)171, endA1, supE44, thi-1, gyrA96, relA1, lac, recBrecJ, sbcC, umuC::Tn5 (kan'), uvrC[F+ proAB, lacI95Z_M15, Tn10 (Tet r)]), DH10B (F, mcrA, (mrr-hsdRMS-mrcBC), 80dlacZ_M15, lacX74, deoR, recA1, endA1, araD139, (ara, leu)7697, galU galK λ, rpsL, nupG).

**Media**

*E. coli* carrying plasmid was grown in LB (0.5% yeast extract, 1% bacto-tryptone, 1% NaCl) containing 50 μg of ampicillin per ml. Yeast strains were grown on YPD (1% yeast extract, 2% bacto-peptone, 2% glucose), YPG (1% yeast extract, 2% bacto-peptone, 3% glycerol), YPDM (1% yeast extract, 0.5% bacto-peptone, 0.1% (NH4)2SO4, 0.2% KH2PO4, 0.1% MgSO4, 0.8% glucose), or SD (0.67% yeast nitrogen base without amino acids, 2% glucose, and appropriate nutrients) in respective experiments. Solid medium contained 2% agar.

**Polymerase Chain Reaction (PCR)**

PCR was performed according to the procedure provided with the Takara Pyrobest™ polymerase PCR kit (Takara Shuzo Co., Ltd, Osaka, Japan). The primer pairs used in these experiments were designed according to the sequence data and purchased from Hokkaido System Science, Co. These primers are located in the ATP1-coding, 5'-, 3'- non-coding region and the neighboring ATP1 gene. DNA was amplified in PCR processors (Astec, Program TEMP Control...
System PC-700, Fukuoka, Japan) by using 30 cycles. Yeast genomic DNA was purified from each strain using a previously reported method [9]. PCR products from each template were cloned into the vector pBluescript (Stratagene, La Jolla, CA) for sequencing according to the procedure for the Takara Blunting Kination Ligation kit (Takara Shuzo Co., Ltd, Osaka, Japan).

DNA Sequencing

Nucleotide sequencing was performed by the dye-terminator chain termination method with ABI models 373 and 310, and LI-COR model 4200L-2 sequencers.

Pulse-Field Gel Electrophoresis

The amplified DNAs were separated on 1% agarose gels on an alternating CHEF gel apparatus (Bio-Rad, CA). Electrophoresis was carried out for 16 h in 0.5 x TBE buffer at 200 V (14°C) with a 2.8 to 3.4 s linear gradient, as described previously [8].

RESULTS AND DISCUSSION

Nucleotide Sequence

The ATP1 gene (YBL099w), encoding the F$_{1}$α, was mapped on the left arm close to the telomere of chromosome II in S. cerevisiae [19]. The yeast Genome Project reported that ATP1 was a single-copy gene mapping approximately 35 kb from the left telomere of chromosome II [12, 13]. We reported that three ATP1 genes were arranged in tandem on the left arm of chromosome II of S. cerevisiae (even in strain S288C) based on (1) chromosome II fragmentation at the site of the ATP1 gene, (2) Southern hybridization of the prime clones with the ATP1 probe (3.4 EcoRI fragment) [8] and (3) long-PCR analysis using primers located just outside the

Miscellaneous

Southern hybridization of DIG-labeled ATP1 and other ATP1-neighboring probes used in the experiments were prepared as described previously [8].
Fig. (2). Nucleotide sequence of three copies of $ATP1s$ and the neighboring (the orange-yellow) region. Isolation, identification, PCR experiments and sequencing of DNA fragments are as described in the legend to Fig. 1. Colors are the same as the appropriate region of the $ATP1$ neighboring fragment on chromosome II. Restriction sites are the same as in the legend to Fig. (1).

(a), the nucleotide sequence of the orange region (contains ORF: YBL098w, 097w, 096c, 095w, 094c and 093c), the EcoRI-site located 1,067 bases downstream of the $ATP1$-stop codon to the 5'-TGGAGA (5'-GAATTC to 5'-TGGAGA, 5,733nt), the yellow region (contains ORF: YBL092w), 5'-AGAGTG located 6,800 bases downstream of the $ATP1$-stop codon to the XhoI-site (1,532 nt), the blue region (contains ORF: YBL091c and 090w), and the violet region (contains ORF: YBL089w). G.P.: Genome Project data.

(b), the map of the above region revealed by the nucleotide sequence. The 5.7 kb orange-yellow region upstream of $ATP1b$ was lacking upstream of $ATP1c$. The following 1.53 kb yellow-region and the 2.15 kb blue-region were conserved upstream of both $ATP1b$ and $ATP1c$. 

3a
Fig. (3). Nucleotide sequence of three copies of ATP1s and the neighboring DNA. Isolation, identification, PCR experiments and sequencing of DNA fragments were as described in the legend to Fig. (1). Colors are the same as the appropriate regions of the ATP1 neighboring fragment on chromosome II. Restriction sites are the same as in the legend to Fig. (1).

(a), the nucleotide sequence of the blue and the violet regions. G.P.: Genome Project data.

(b), the map of the above region revealed by the nucleotide sequence. The violet-region upstream of ATP1b was composed of approximately 0.1 kb. On the other hand, that of ATP1c was composed of approximately 4.3 kb.

Red region

SalI →

4a

ATP1 (Genome)  1      10      20      30      40      50      60
ATP1a          GTGGAC         AAAG GCTTCGTTAA CATGCACTTA AACTACAGAA GCAGAAATAC ACGCAGTCAG
ATP1b          **********         **********         **********         **********         **********         **********
ATP1c          **********         **********         **********         **********         **********         **********

70

ATP1 (Genome)  700     800     900
ATP1a          TCGAGCTATA         --------------------(650 nt)---------------------------
ATP1b          *************************(no nucleotides)************************
ATP1c          *************************(no nucleotides)************************

730 740 750 760 770 780

ATP1 (Genome)  730     740     750     760     770     780
ATP1a          ATATGAGAAG         ATACACACCA         CTATGAAAGA         AGAATAATAA         TAATAACACT         GTATAGAAAT
ATP1b          **********         **********         **********         **********         **********         **********
ATP1c          **********         **********         **********         **********         **********         **********

(Blue region)

XhoI →

Red region

XhoI →

XhoI →

SalI →

4a

ATP1 (Genome)  970     1586     1596
ATP1a          TTTTAATATAT         970     980     990     1000
ATP1b          TTTTAATATAT         1010    1020    1030    1040
ATP1c          TTTTAATATAT         1050    1060    1070    1080

Green region

EcoRI →

EcoRI →

EcoRI →

ATP1 (Genome)  1606    1616    4859
ATP1a          1606    1616    4859
ATP1b          1626    1636    4859
ATP1c          1646    1656    4859

Orange region
Fig. (4). The junction sequences of the red-, purple-, green- and orange regions. (a), the nucleotide sequence of the above junction. G.P.: Genome Project data. (b), the map of the above region revealed by the nucleotide sequence. Both the 0.6 kb BamHI-EcoRI fragment (purple-region) and the delta sequence (red-region) upstream of \(ATP1c\) were completely missing. The 0.1 kb XhoI-BamHI fragment of the delta sequence was conserved upstream of \(ATP1b\).

Fig. (5). Upstream consensus sequence of three \(ATP1\) genes. Isolation, identification, PCR experiments and sequencing of DNA fragments were as described in the legend to Fig. (1). Colors are the same as the appropriate region of the \(ATP1\) neighboring fragment on chromosome II (Fig. 1).
XhoI-BamHI fragment (0.1 kb) of the delta sequence was conserved, but that of ATP1c was completely deleted (Figs. 1, 3 and 4). The 0.6 kb BamHI-EcoRI DNA sequence was conserved in both the 5’-upstream regions of ATP1a and ATP1b, but completely deleted upstream of ATP1c (Figs. 3 and 4).

The nucleotide sequences of the three ATP1s and the 3.4 kb EcoRI fragment (green region), and those of ATP1a, ATP1b and ATP1c were completely identical (data not shown).

Thus, three copies of the ATP1 gene are arranged in tandem accompanied by the neighboring ORFs and DNAs of ATP1 in a region of at least 50 kb on chromosome II reported by the yeast genome project. The differences [9] in the distances between ATP1a and ATP1b, and ATP1b and ATP1c could be explained to reveal the genome organization of the three ATP1s on chromosome II.

DNA Sequence Present Upstream of the Repetitive Genes

The XhoI-BamHI DNA sequence, in particular 5’-AATTTTTCATAGCT-3’, which was part of the delta sequence (red region, a transposable element) [1, 21], was located 1,242 bases upstream of the ATP1a and ATP1b start codons. Although the delta sequence was completely deleted upstream of ATP1c, a homologous 12 nucleotide sequence, 5’-AATTTTGCAATT-3’, was located 706 bases from the upstream of the ATP1c start codon, which was as close as to the ATP1a- and ATP1b-start codons. In addition, the 5’-AATTTTTCAATG-3’ was also found in the 1,294 bases upstream of the repetitive ATP2-region on chromosome X (data not shown). Moreover, other nuclear-encoded F1F0 subunit genes, ATP3, ATP16, ATP15, ATP4, ATP5 or ATP7, each have a similar sequence upstream of the start codon, although the distance from the start codon varied (manuscript in preparation). Based on these results, it may be possible that the DNA sequence 5’-AATTTTTTCATAGCT-3’, located upstream of the three ATP1s could be driving gene repetition on the chromosome.

The mechanism of gene repetition that results in multiple gene copies of genes like that encoding the F1F0 subunit is poorly understood, but the biological meaning of gene multiplicity and the relationship of gene duplication in the S. cerevisiae genome is an exciting subject that is currently under investigation [22-24].

F1F0-ATPase is one of the essential enzymes in eukaryotic cells, so the genes encoding subunits of the complex must be maintained in the event that some of the genes become mutated. Also, expression of the genes might be controlled by sophisticated regulatory mechanisms using multiple gene copies. The mechanism of the gene repetition on chromosome and the physiological meaning of these multiple copies of the F1F0 subunit genes including ATP1 and other adjacent regions containing repeated units are still unknown in living cells. However, we need continue to consider them to understand why living cells harbor gene repetitions on chromosomes [24].

In recent genome projects, the shotgun methods had been used to rapidly sequence large genomes and the data is organized into contigs by computer analysis. With this method it might be easy to overlook or ignore repetitive sequences such those identified in the ATP1 region (more than 35 kb). Typical repetitive genes were observed for ribosomal RNA on the right arm of chromosome XII of S. cerevisiae. This region was very large (1-2 Mb), but the exact size (base number) was not known. However, the ATP1-repetition region was an appropriate size for a detailed analysis of the DNA structure. Similar gene-repetitions had been observed for other ATP genes in S. cerevisiae. ATP provides cellular energy and participates in all biological phenomena in living cells. To perform efficiently, each ATP gene might be organized on the chromosome in a sophisticated fashion. The gene organization might be regulating not only ORFs, but also other DNA sequences including the non-coding regions on chromosomes. Thus, we might modify our view of the S. cerevisiae genome based on the results reported in this manuscript.

In recent years, genome sequencing projects for many other species both the eukaryotes and the prokaryotes had been completed [25-28]. Presently, most of research projects were focused on the analysis of ORFs, and the functions of the encoded proteins using methods such as proteome and transcriptome analyses [29-31]. However, from this viewpoint, it might be impossible to resolve the structure of the entire genome as a molecule in the biological system, and it still remained totally obscure whether or not there was significant structure in the genome [32, 33].

Much of the genome, especially those of the eukaryotic cells were occupied by the non-coding regions such as RNAs, promoters, introns, SINE, LINE, MAR and poly(A) associated signals, etc., in addition to genes that were translated into proteins. In the eukaryotes, multiple complicated regulatory sequences were needed to express genes. These elements, which were located in the non-coding regions, were all needed to express a gene. For instance, miniRNA consisting of 22 nucleotides, was one type of non-coding region in genomes that may participate in regulating gene expression [34-36].

Certain non-coding elements, including the transposable elements might trigger gene rearrangements in the chromosome. The delta sequence, which was a transposable element, was present in the region upstream of the ATP1 gene. It was possible that the delta sequence might be able to promote gene repetition, but all ATP genes did not have the

*Foot note; other nuclear-coded F1F0-ATPase subunit genes of S. cerevisiae, ATP16 and ATP15, were also observed the gene-repetition accompanying the neighboring ORFs reported by the Genome Project (manuscripts in preparation).
sequence in their upstream regions. A detailed study of the mechanism of ATP gene repetition is in progress.

The non-translated DNA regions vary in different species, and these sequences were important roles for perform the gene expression and the regulation of biological phenomena, and these detailed study are under the progress [37-44]. It might be necessary to analyze the detailed study of the 35-kb region of three ATP1-repeated, especially the non-coding sequences.

ACKNOWLEDGEMENT

The authors wish to thank to Dr. Tadashi Mabuchi at the University of Yamanashi for his continuous encouragement on the research and discussions about this manuscript.

REFERENCES

[19] Takeda M, Okushiba T, Hayashida T, Gunge N, ATP1 and ATP2, the F1,F0-ATPase alpha and beta subunit genes of Saccharomyces cerevisiae, are respectively located on chromosome II and X. Yeast 1994; 10: 1531-4.
Genome Organization of the Three Identical ATP1 Genes on the Left Arm


© Takeda et al.; Licensee Bentham Open.
This is an open access article licensed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/3.0/) which permits unrestricted, non-commercial use, distribution and reproduction in any medium, provided the work is properly cited.