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 Takeda[1,2,*], Shohei Nakamura[1] and Satoshi Matsushita[1]

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

Abstract: The *ATP1* gene (YBL099w) encoding the F_1F_0-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 *ATP1*s 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* repetition 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 *ATP1*s 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 (F_1F_0-ATPase complex = F_1F_0) 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 F_1-ATPase (F_1) and the transmembrane sector, or proton channel (F_0) [3, 4]. Both F_1 and F_0 are necessary for ATP synthase activity, whereas F_1 alone retains the ability to hydrolyze ATP [5].

F_1 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 F_1 resides on the α-β-subunit dimer [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 F_1F_0 subunit genes were not coincident with the subunit-stoichiometry of F_1 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 F_1F_0 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 F_1-α subunit ($F_1\alpha$) 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 *ATP1*s and other F_1F_0 subunit genes was discussed.

*Address correspondence to this author at the Department of Applied Life Science, Sojo University, 4-22-1 Ikeda, Kumamoto 860-0082, Japan; Tel/Fax: +81-235-25-9130. E-mail: mtakeda@tsuruoka-nct.ac.jp

1875-693X/10 **2010 Bentham Open**

MATERIALS AND METHODS

Yeast Strains

Yeast strains used in this study were Saccharomyces cerevisiae DC5 (MATa, leu2-3, leu2-112, his 3, can1-11), LL20 (MATα, leu2, his3), W301-1A (MATa, leu2-3, leu2-112, his3-11 his3-15, trp1-1, ura3-1, ade2-1, can1-100), W303-1B (MATα, leu2-3, leu2-112, his3-11 his3-15, trp1-1, ura3-1, ade2-1, can1-100), YPH499 (MATa, ade2, his3, leu2, trp1, ura3, lys2), S288C (MATα, SUC2, mal, mel, gal2, CUP1), SKY2A11 (MATa, leu2-3, leu2-112, his 3, can1-11, atp1b::LEU2), SKY4A11 (MATa, leu2-3, leu2-112, his 3, can1-11,atp1a::HIS3, 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$^-$), Δ(mcrCB-hsdSMR-mrr)171, endA1, supE44, thi-1, gyrA96, relA1, lac, recBrecJ, sbcC, umuC,::Tn5 (Kan^r), uvrC[F' proAB, lac1^{q}Z_M15, Tn10 (Tet^r)], DH10B (F, mcrA, Δ(mrr-hsdRMS-mrcBC), Φ80dlacZΔM15, ΔlacX74, deoR, recA1, endA1, araD139, Δ(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% $(NH_4)_2SO_4$, 0.2% KH_2PO_4, 0.1% $MgSO_4$, 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^{TM} 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

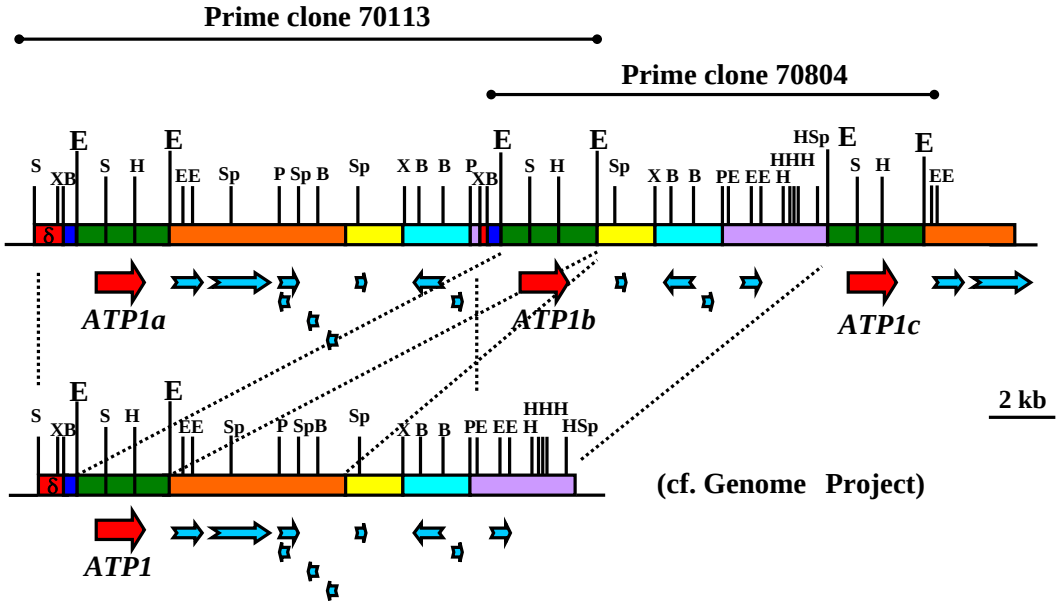

Fig. (1). Gene organization of three copies of the *ATP1* gene on the left arm of chromosome II in *S. cerevisiae*. Template DNAs for sequencing were isolated from *S. cerevisiae* strains DC5, W301-1A, S288C, SKY4A11 and TKY4011, and the prime clones 70113 and 70804 were purchased from ATCC [8, 9, 17]. The digested DNA fragments were purified and manipulated according to previously published procedures [8, 9].

The prime clone 70113 was used for the isolation of *ATP1a* and *ATP1b* and their neighboring fragments, and prime clone 70804 was used for *ATP1b* and *ATP1c* and their neighboring fragments. The identification of *ATP1* and the neighboring DNA fragments was performed by using Southern hybridization with the appropriate probes [8, 9]. To confirm the sequencing data from the prime clones, we sequenced the genomic DNAs isolated from the wild-type yeast strains as described previously [8, 9].

PCR and nucleotide sequencing were performed as described in Materials and Methods.

(a), Arrows indicate open reading frames (ORFs) and the direction of transcription. The color of the region in the figure indicates a region maintained the same nucleotide sequence. *ATP1* neighboring ORFs were reported by the Yeast Genome Project (designated as G.P.).

B, *Bam*HI; E, *Eco*RI; H, *Hin*dIII; P, *Pst*I; S, *Sal*I; Sp, *Sph*I; X, *Xho*I.

Nucleotide numbers (nt) of the repeated sequences of the three copies of *ATP1*, *ATP1a*, *ATP1b* and *ATP1c* (red, purple, green, orange, yellow, blue, violet) in Fig.1a are as follows:

ATP1a: red (927 nt)-purple (663 nt)-green (3,278 nt containing *ATP1a*)-orange (5,730 nt)-yellow (1,532 nt)-blue (2,144 nt)-violet (105 nt).

ATP1b: red (113 nt)-purple (663 nt)-green (3,278 nt containing *ATP1b*)-orange (14 nt)-yellow (1,532 nt)-blue (2,144 nt)-violet (4,333 nt).

ATP1c: green (3,278 nt containing *ATP1c*)-orange (5,730 nt)-below, the same as reported by the G.P.

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 dideoxy 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 (w/v) 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 ^{o}C) with a 2.8 to 3.4 s linear gradient, as described previously [8].

Miscellaneous

Southern hybridization of DIG-labeled *ATP1* and other *ATP1*-neighboring probes used in the experiments were prepared as described previously [8].

RESULTS AND DISCUSSION

Nucleotide Sequence

The *ATP1* gene (YBL099w), encoding the $F_1\alpha$, 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 *Eco*R1 fragment) [8] and (3) long-PCR analysis using primers located just outside the

2a

Orange region →

*Eco*RI

```
                1       10        20        30        40        50        60
ATP1 (Genome)   GAATTCAAGC CAAACTATGG CGGAAATTTT GCAATAGCTC CTGACCATTT GCACATTTGG
ATP1a-b         GAATTCAAGC CAAACTATGG CGGAAATTTT GCAATAGCTC CTGACCATTT GCACATTTGG
ATP1b-c         GAATTCAAGC CAAAC***** ********** ********** ********** **********

                       70        80        90       100       110       120
ATP1 (Genome)   CCTCGTCATA AATTCATGTT AATTGCGCTC GCCAACAGTG ACGGCTCGTT CACTTCAACC
ATP1a-b         CCTCGTCATA AATTCATGTT AATTGCGCTC GCCAACAGTG ACGGCTCGTT CACTTCAACC
ATP1b-c         ********** ********** ********** ********** ********** **********

                                                                  EcoRI
                      130       140       150       160       170       180
ATP1 (Genome)   TTTTTCGGTT CTAAAGATCA AATATCAGAT CTGATAACTT CCAAGTCACG TGTGAGGGAA
ATP1a-b         TTTTTCGGTT CTAAAGATCA AATATCAGAT CTGATAACTT CCAAGTCACG TGTGAGGGAA
ATP1b-c         ********** ********** ********** ********** ********** **********

                      190       200       210       220       230       240
ATP1 (Genome)   TTCTTAATCG AGAACTTTCC CGATATTATT AATATTATGG ATTTGGACGA TGCTGTCAAA
ATP1a-b         TTCTTAATCG AGAACTTTCC CGATATTATT AATATTATGG ATTTGGACGA TGCTGTCAAA
ATP1b-c         ********** ********** ********** ********** ********** **********

                      250       260       270       280       290       300
ATP1 (Genome)   AGGTTTATCA CTTATCCAAA GGAAAGTCTT GTCTGTGTAA ACTGTAAGCC ATACGATGTA
ATP1a-b         AGGTTTATCA CTTATCCAAA GGAAAGTCTT GTCTGTGTAA ACTGTAAGCC ATACGATGTA
ATP1b-c         ********** ********** ********** ********** ********** **********


ATP1 (Genome)   CCA   (5,430 nt, EcoRI, XbaI, HindIII, SphI etc)
ATP1a-b         CCA   (5,430 nt, EcoRI, XbaI, HindIII, SphI etc))
ATP1b-c         ***   (no nucleotides)
```

Yellow region →

```
                     5743      5753      5763      5773      5783      5793
ATP1 (Genome)   AGAGTGTTTG ATTCCAGCAG AAGGTAATAC GCACCTTTCT CATCTATTTG CAGAATCGTT
ATP1b-c         AGAGTGTTTG ATTCCAGCAG AAGGTAATAC GCACCTTTCT CATCTATTTG CAGAATCGTT
ATP1b-c         AGAGTGTTTG ATTCCAGCAG AAGGTAATAC GCACCTTTCT CATCTATTTG CAGAATCGTT
```

Blue region →

```
                                      XhoI      7273      7283      7293
ATP1 (Genome)   ---------(1,480 nt)-------------TCTCGAGCTT TATATACTCT GTTGATTGTT
ATP1a-b         ---------(1,480 nt)-------------TCTCGAGCTT TATATACTCT GTTGATTGTT
ATP1b-c         ---------(1,480 nt)-------------TCTCGAGCTT TATATACTCT GTTGATTGTT
```

Violet region →

```
                                                PstI
                                              9393      9403
ATP1 (Genome)   ----------(2,100 nt)------------TGAATAAGAA CAACCCTGCA G
ATP1a-b         ----------(2,100 nt)------------TGAATAAGAA CAACCCTGCA G
ATP1b-c         ----------(2,100 nt)------------TGAATAAGAA CAACCCTGCA G
```

Fig. (2). contd….

2b

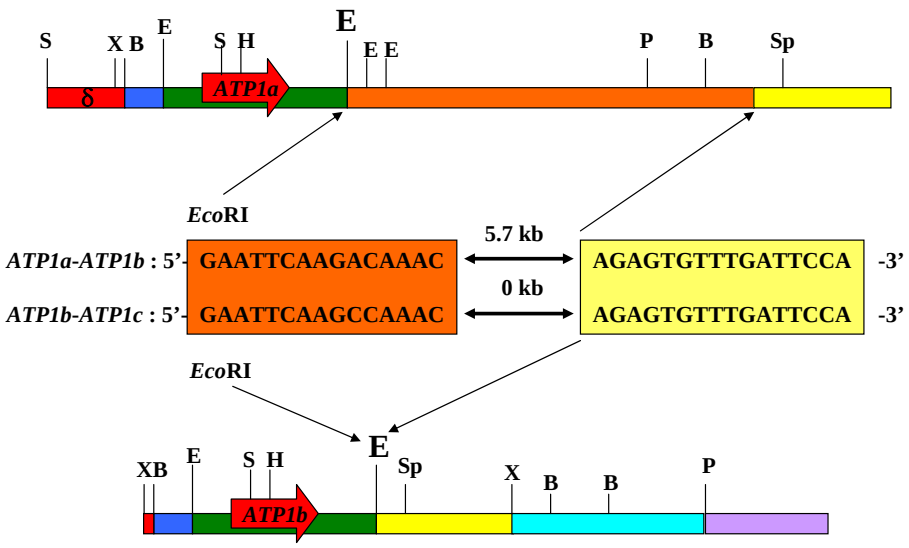

Fig. (2). Nucleotide sequence of three copies of *ATP1*s and the neighboring (the orange-yellow) region. Isolation, identification, PCR experiments and sequencing of DNA fragments are as described in the legend to Fig. 1. Colors are the same as the appropriate region of the *ATP1* neighboring fragment 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: YBLs098w, 097w, 096c, 095w, 094c and 093c), the *Eco*RI-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 *Xho*I-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

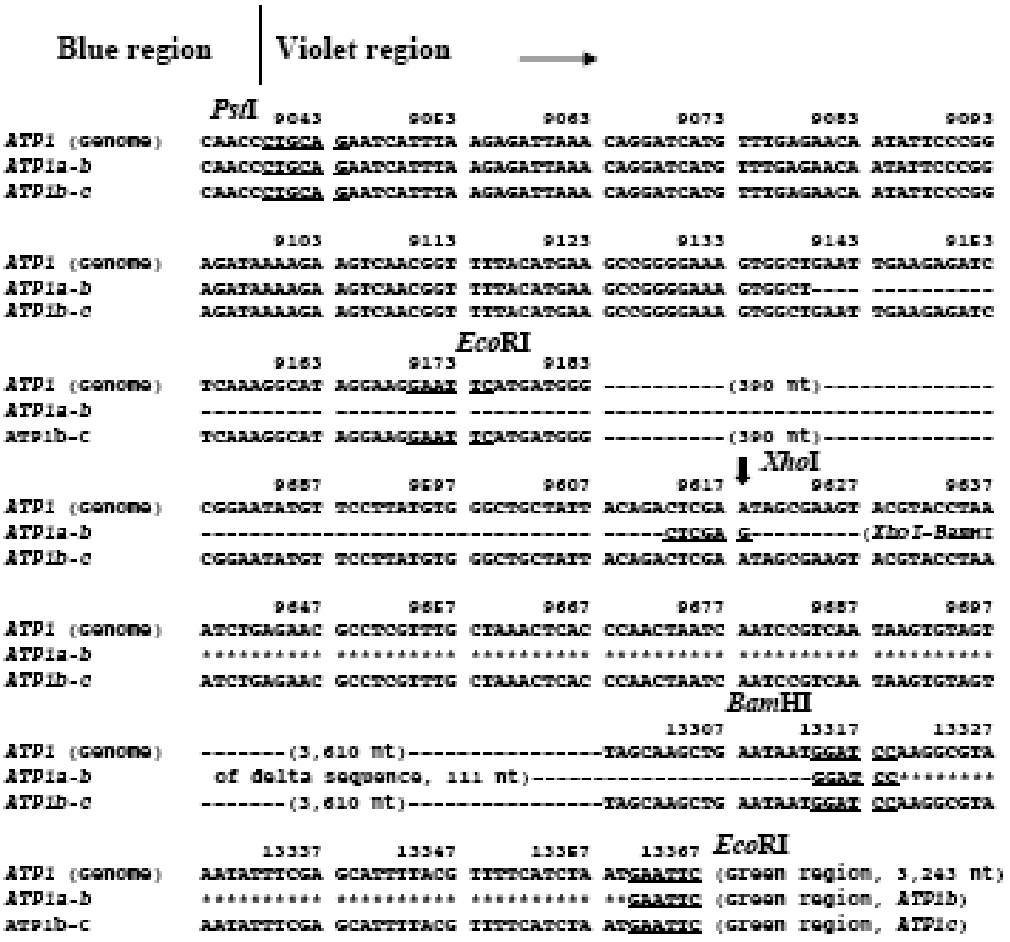

Fig. (3). contd…..

3b

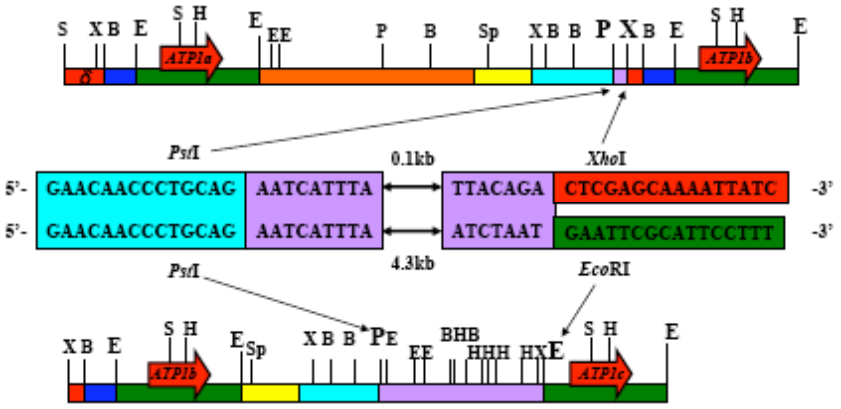

Fig. (3). Nucleotide sequence of three copies of *ATP1*s and the neighboring DNA. Isolation, identification, PCR experiments and sequencing of DNA fragments were as described in the legend to Fig. (**1**). Colors are the same as the appropriate regions of the *ATP1* neighboring fragment on chromosome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

4a

Red region

*Sal*I →

```
                1       10        20        30        40        50        60
ATP1 (Genome)   GTGGACAAAG GCTTCGTTAA CATGCACTTA AACTACAGAA GCAGAAATAC ACGCAGTCAG
ATP1a           GTCGACAAAG GCTTCGTTAA CATGCACTTC AACTACAGAA GCAGAAATAC ACGCAGTCAG
ATP1b           ********** ********** ********** ********** ********** **********
ATP1c           ********** ********** ********** ********** ********** **********

                       70
ATP1 (Genome)   TGAAGCTATA--------------------(650 nt)--------------------------
ATP1a           TGAAGCTATA--------------------(650 nt)--------------------------
ATP1b           **********************(no nucleotides)**********************
ATP1c           **********************(no nucleotides)**********************

                      730       740       750       760       770       780
ATP1 (Genome)   ATAATGAACG ATAACACACA CTATGAAAGA AGAATAATAA TAATAACACT GTATAGAAAT
ATP1a           ATAATGAACG ATAACACACA CTATGAAAGA AGAATAATAA TAATAACACT GTATAGAAAT
ATP1b           ********** ********** ********** ********** ********** **********
ATP1c           ********** ********** ********** ********** ********** **********
```

*Xho*I → **(Red region)**

```
                      790       800       810       820       830       840
ATP1 (Genome)   AGCGGCTCCC TCTTGTTTAT TCTCACATCC TCGAGCAAAA CTTCTAGCAA ATCCTGTGTA
ATP1a           AGCGGCTCCC TCTTGTTTAT TCTCACATCC TCGAGCAAAA CTTCTAGCAA ATCCTGTGTA
ATP1b           ********** ********** *********C TCGAGCAAAA CTTCTAGCAA ATCCTGTGTA
ATP1c           ********** ********** ********** ********** ********** **********

                      850       860       870       880       890       900
ATP1 (Genome)   TTTAATATTA TGGCCTCTAT CAGCAATGGA CTCCCAATAA TTATCCAATT ACTCACCAAT
ATP1a           TTTAATATTA TGGCCTCTAT CAGCAATGGA CTCCCAATAA TTATCCAATT ACTCACCAAT
ATP1b           TTTAATATTA TGGCCTCTAT CAGCAATGGA CTCCCAATAA TTATCCAATT ACTCACCAAT
ATP1c           ********** ********** ********** ********** ********** **********
```

*Bam*HI → **Purple region**

```
                      910       920       930       940       950       960
ATP1 (Genome)   TTTTCAATAT TAGTGTAGAT AGGAAAGGAT CCTCGATGAA ATCGTTATGG TTAGTGTCTC
ATP1a           TTTTCAATAT TAGTGTAGAT AGGAAAGGAT CCTCGATGAA ATCGTTATGG TTAGTGTCTC
ATP1b           TTTTCAATAT TAGTGTAGAT AGGAAAGGAT CCTCGATGAA ATCGTTATGG TTAGTGTCTC
ATP1c           ********** ********** ********** ********** ********** **********
```

Green region

*Eco*RI →

```
                      970                                      1586      1596
ATP1 (Genome)   TGTTGATAAT -----------(606 nt)--------------TGTTCGATCC ACAGAATTCG
ATP1a           TGTTGATAAT -----------(606 nt)--------------TGTTCGATCC ACAGAATTCG
ATP1b           TGTTGATAAT -----------(606 nt)--------------TGTTCGATCC ACAGAATTCG
ATP1c           ********** **********(no nucleotides)*************** ***GAATTCG
```

Orange region

*Eco*RI →

```
                    1606      1616                                4859
ATP1 (Genome)   CATTCCTTTT TGCTAGCATT ---(3,243 nt, ATP1)---TACTGAGGAA TTC
ATP1a           CATTCCTTTT TGCTAGCATT ---(3,243 nt, ATP1a)--TACTGAGGAA TTC
ATP1b           CATTCCTTTT TGCTAGCATT ---(3,243 nt, ATP1b)--TACTGAGGAA TTC
ATP1c           CATTCCTTTT TGCTAGCATT ---(3,243 nt, ATP1c)--TACTGAGGAA TTC
```

Fig. (4). contd…..

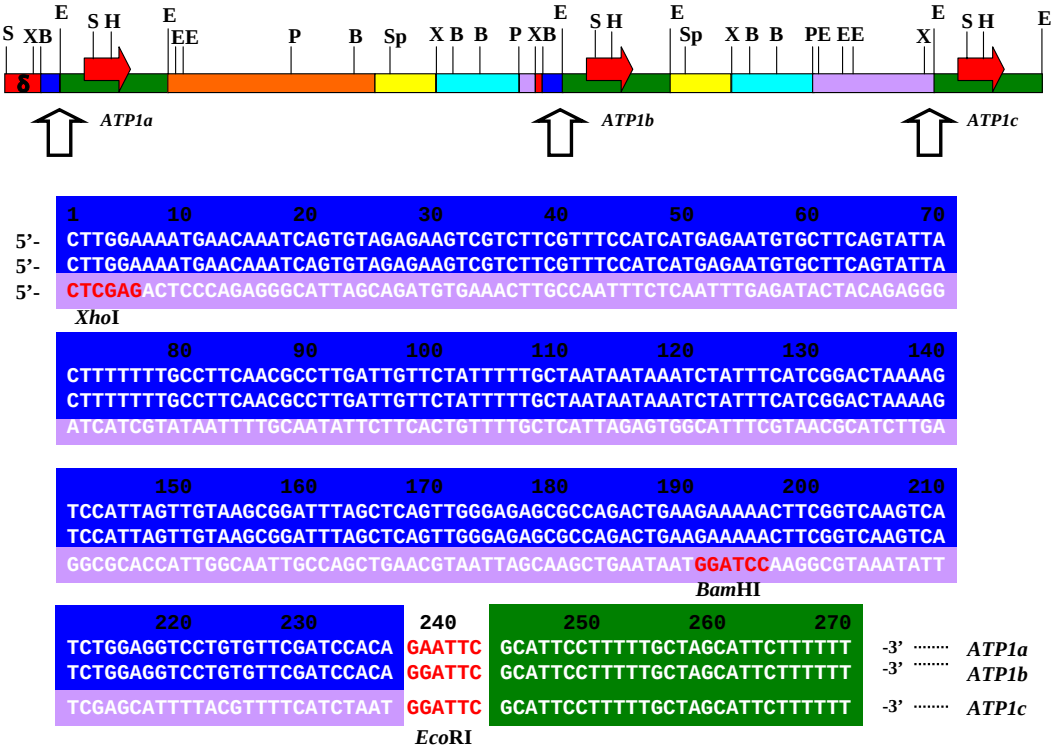

Fig. (4). The junction sequences of the red-, purple-, green- and orange regions. (a), the nucleotide sequence of the above junction. G.P.: Genome Project data. (b), the map of the above region revealed by the nucleotide sequence.

Both the 0.6 kb *Bam*HI-*Eco*RI fragment (purple-region) and the delta sequence (red-region) upstream of *ATP1c* were completely missing. The 0.1 kb *Xho*I-*Bam*HI fragment of the delta sequence was conserved upstream of *ATP1b*.

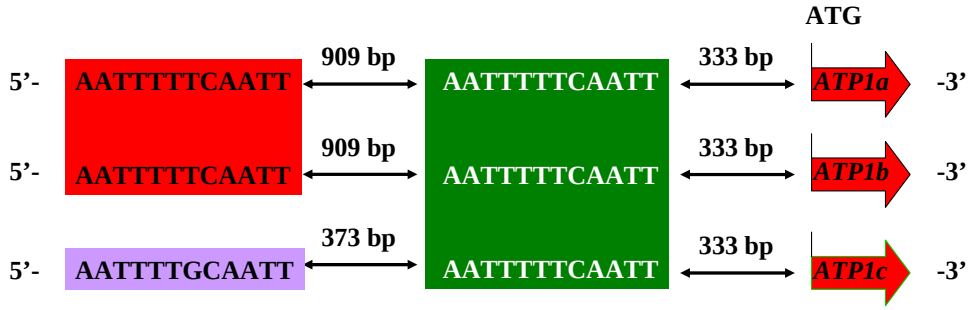

Fig. (5). Upstream consensus sequence of three *ATP1* genes. Isolation, identification, PCR experiments and sequencing of DNA fragments were as described in the legend to Fig. (**1**). Colors are the same as the appropriate region of the *ATP1* neighboring fragment on chromosome II (Fig. **1**).

start and stop codons [9]. In addition, these three *ATP1*s, *ATP1a*, *ATP1b* and *ATP1c* apparently showed no difference in ATPase activities although the distances between *ATP1a* and *ATP1b*, and *ATP1b* and *ATP1c* were different from each other as determined by long-PCR analysis [8, 9].

The nucleotide sequence of the repetitive region (ca. 35 kb) containing the three *ATP1*s, *ATP1a*, *ATP1b* and *ATP1c* was determined using the prime clones 70113 and 70804 from ATCC [20], and genomic DNAs isolated from yeast strains DC5, W303-1A [8-11] and the gene copy-specific *ATP1*-disruptaption strains, SKY2A11, SKY4A11 and TKY4011 [8]. The results are shown in Fig. (**1**). The arrows were indicated the ORFs and the direction of transcription. Surprisingly, three identical *ATP1*s were repeated accompanied by the neighboring ORFs reported by the Genome Project [12]. That is, the region downstream of *ATP1b* had a 5.7-kb (orange region) deletion compared with those of *ATP1a*, but the following yellow and blue regions were conserved (Figs. **1** and **2**). However, most of the violet regions were deleted downstream of *ATP1a* (the region between *ATP1a* and *ATP1b*, Figs. **1** and **3**).

The delta sequence (red region, a transposable element) present upstream of *ATP1a* was conserved. In contrast, that of *ATP1b* had most of this sequence deleted except that the

*Xho*I-*Bam*HI 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 *Bam*HI-*Eco*R1 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 *ATP1*s and the 3.4 kb *Eco*RI 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 30 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 *ATP1*s on chromosome II.

DNA Sequence Present Upstream of the Repetitive Genes

The *Xho*I-*Bam*HI DNA sequence, in particular 5'-AATTTTTCAATT-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 the *ATP1b*-start codons. In addition, the 5'-AATTTTTCAATT-3' sequence was also completely conserved 332 bases upstream of the three *ATP1*s within the 3.4 kb *Eco*RI fragment (Fig. **4**). The homologous DNA sequence 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 F_1F_0 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'-AATTTTTCAAT-3', located upstream of the three *ATP1*s 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 F_1F_0 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].

F_1F_0-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 repletion on chromosome and the physiological meaning of these multiple copies of the F_1F_0 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 F_1F_0-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

[1] Cox GB, Devenish RJ, Gibson F, Howitt SM, Nagley P. The structure and assembly of ATP synthase. In Ernster L, Ed. Molecular Mechanism in Bioenergetics, Amsterdam. Elsevier 1992; pp. 283-315.

[2] Gresser M, Meyers JA, Boyer PD. Catalytic site cooperativity of beef heart mitochondrial F_1 adenosine triphosphatase. J Biol Chem 1982; 257: 12030-8.

[3] Boyer, PD. The ATP synthase- A splendid molecular machine. Annu Rev Biochem 1997; 66: 717-749.

[4] Futai M, Omote H, Sambongi Y, Wada Y. Synthase (H^+-ATPase) coupling between catalysis, mechanical work, and proton translocation. Biochim Biophys Acta 2000; 1458: 276-288.

[5] Noji H, Yasuda R, Yoshida M, Kinoshita K Jr. Direct observation of the rotation of F_1-ATPase. Nature 1997; 386: 299-302.

[6] Harada M, Ito Y. Sato M, Ohta S, Kagawa Y. Small-angle X-ray scatting studies of Mg.AT (D) P-induced hexamer to dimmer dissociation in the reconstitute alpha 3 beta 3 complex of ATP synthase from thermophilic bacterium PS3. J Biol Chem 1991; 25:11455-60.

[7] Douglas MG, Koh Y, Docker ME, Schatz G. Aureovertin binds to the β subunit of yeast mitochondrial ATPase. J Biol Chem 1977; 252: 8333-8335.

[8] Takeda M, Okushiba T, Satoh T, Kuniyoshi S, Morishita C, Ichimura Y. Three *ATP1* genes are present on chromosome II in *Saccharomyces cerevisiae*. J Biochem 1995; 118: 607-613.

[9] Takeda M, Satoh H, Ohnishi K, Satoh T, Mabuchi T. The three copies of *ATP1* gene are arranged in tandem on chromosome II of the yeast *Saccharomyces cerevisiae* S288C. Yeast 1999; 15: 873-878.

[10] Takeda M, Katayama H, Satoh T, Mabuchi T. The three copies of *ATP2* gene are arranged in tandem on chromosome X of yeast *Saccharomyces cerevisiae*. Curr Genet 2005; 47: 265-272.

[11] Ohnishi K, Ishibashi S, Kunihiro M, Satoh T, Matsubara K, Oku S, Ono B, Mabuchi T, Takeda M. Studies on the *ATP3* gene of *Saccharomyces cerevisiae*: presence of two closely linked copies, *ATP3a* and *ATP3b*, on the right arm of chromosome II. Yeast 2003; 20: 943-954.

[12] Feldmann H, Aigle M, Aljinovie G, *et al*. Complete DNA sequence of yeast chromosome II. EMBO J 1994; 13: 5795-5809.

[13] Mewes HW, Albermann K, Bahr M, *et al*. Overview of the yeast genome. Nature 1997; 387: 7-65.

[14] Galibert F, Alexandraki D, Baur A, *et al*. Complete nucleotide sequence of *Saccharomyces cerevisiae* chromosome X. EMBO J 1996; 15: 2031-2049.

[15] Takeda M, Nakahara M. Structural Features of the Nucleotide Sequences of Genomes. J Comput Aided Chem 2009; 10: 38-52.

[16] Nakahara M, Takeda M. Characterization of the sequence spectrum of DNA based on the appearance frequency of the nucleotide sequences of the genome - A new method for analysis of genome structure-. J Biomed Sci Eng 2010(a); 3: 340-350.

[17] Nakahara M, Takeda M. Identification of the Interactive Region by the Homology of the Sequence Spectrum. J Biomed Sci Eng 2010(b); 3: 868-883.

[18] Takeda M, Chen W-J, Saltzgaber J, Douglas MG. Nuclear genes encoding the yeast mitochondrial ATPase complex-analysis of *ATP1* coding the F_1-ATPase α-subunit and its assembly. J Biol Chem 1986; 261: 15126-15133.

[19] Takeda M, Okushiba T, Hayashida T, Gunge N. *ATP1* and *ATP2*, the F_1F_0-ATPase α and β subunit genes of *Saccharomyces cerevisiae*, are respectively located on chromosome II and X. Yeast 1994; 10: 1531-4.

[20] Olson MV, Dutchik JE, Graham MG, *et al*. Random-clone strategy for genomic restriction mapping in yeast. Proc Natl Acad Sci USA 1986; 83: 7826-7830.

[21] Shapiro JA. Transposable element as the key to a 21st century view of evolution. Genetica 1999; 107: 171-9.

[22] Llorente B, Durrens P, Malpertuy A, *et al*. Genome exploration of the Hemiascomycetous yeast: 20. Evolution of gene redundancy compared to *Saccharomyces cerevisiae*. FEBS Lett 2000; 487: 122-133.

[23] Piskur J, Langkjar RB. Yeast genome sequencing: the power of comparative genomics. Mol Microbiol 2004; 53: 381-389.

[24] Dujon B, Sherman D, Fisher G, *et al*. Genome evolution in yeasts. Nature 2004; 430: 35-44.

[25] The *C. elegans* Sequencing Consortium. Genome sequence of the Nematode *C.elegance*: A platform for investigating biology. Science 1998; 282: 2012-2018.

[26] International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. Nature 2001; 409: 860-921.

[27] Town CD, Cheung F, Maiti R, *et al*. Comparative genomics of *Brassica oleracea* and *Arabidopsis thaliana* reveal gene loss, fragmentation, and dispersal after polyploidy. Plant Cell 2006; 18: 1348-1359.

[28] NCBI genome data base. 2010. Available from http://www.ncbi.nlm.nih.gov/sites/entrez?db=genome.

[29] Velculescu VE, Zhang L, Zhou W, *et al*. Characterization of the yeast transcriptome. Cell 1997; 88: 243-51.

[30] Brown CE, Lechner T, Howe L, Workman JL. The many HATs of transcription coactivators. Trends Biol Sci (TIBS*)* 2000; 25: 15-19.

[31] Wan XF, VerBerkmoes NC, McCue LA, *et al*. Transcriptomic and proteomic characterization of the fur modulon in the metal-reducing bacterium Shewanella oneidensis. J Bacteriol 2004; 186: 8385-8400.

[32] Olson LE, Richtsmeiser JT, Leszl J, Reeves RH. A chromosome 21 critical region does not cause specific down syndrome phenotypes. Science 2004; 306: 687-690.

[33] Sultan M, Piccini I, Balzereit D, *et al*. Gene expression variation in Down's syndrome mice allows prioritization of candidate genes. Genome Biology 2007; 8: R91.

[34] Couzin J. Small RNAs make big splash. Science 2002; 298: 2296-2297.

[35] Kawasaki H, Wadhwa R, Taira K. World of small RNAs: from ribozymes to siRNA and miRNA. Differentiation 2004; 72: 58-64.

[36] Ketting RF, Plasterk RH. What's new about RNAi? Meeting on siRNAs and miRNAs. EMBO Rep 2004; 5: 762-765.

[37] Webb CF, Das C, Eneff K, Tucker PW. Identification of a matrix-associated region 5' of an immunoglobulin heavy chain variable region gene. Mol Cell Biol 1992; 11: 5206-5211.

[38] Morton BR. Neighboring base composition and transversion/transition bias in a comparison of rice and maise chloroplast noncoding regions. Proc Natl Acad Sci USA 1995; 92: 9791-9721.

[39] West AG, Gaszner M, Felsenfeld G. Insulators: many functions, many mechanisms. Genes Dev 2002; 16: 271-288.

[40] Levine M, Tjian R. Transcription regulation and animal diversity. Nature 2003; 424: 147-151.

[41] Lai EC, Roegirs F, Qin X, Jan YN, Rubin GM. The ubiquitin ligase *Dorosophila* Mind bomb promoters Notch signaling by regulating the localization and activity of Serrate and Delta. Development 2005; 132: 2319-2332.

[42] Mattick JS. RNA regulation: a new genetics? Nat Rev Genet 2004; 5: 316-323.

[43] Martens JS, Wu P-YJ, Winston F. Regulation of an intergenic transcript controls adjacent gene transcription in *Saccharomyces cerevisiae*. Genes Dev 2005; 19: 2695-2704.

[44] Taft RJ, Pheasant M, Mattick JS. The relationship between non-protein-coding DNA and eukaryotic complexity. Bioessays 2007; 29: 288-299.

Received: October 08, 2010 Revised: November 25, 2010 Accepted: November 26, 2010