CORE
Metadata, citation and similar papers at core.ac.uk
Provided by Yamagata University Academic Repository

山形大学紀要（工学）第25巻　第１号　平成10年１月
Bull.Yamagata Univ.（Eng.）Vol.25 No.1 Jan.1998

Detection of bacterial tRNA and rRNA gene clusters hidden in long DNA sequences: Prediction of a putative tRNA (Glu) (TTC) -tRNA (Asp) (GTC) -tRNA (Phe) (GAA) gene cluster of the Gram-positive bacterium *Mycobacterium leprae*

KUDO Yoshihiro, KANAYA Shigehiko, and SATO Yuji
Department of Electric and Information Engineering, Faculty of Engineering,
Yamagata University, Yonezawa, 992 Japan.

(Received September 1, 1997)

Abstract

From the genomic DNA sequences of Archebacteria, Gram-negative and Gram-positive bacteria, Mycoplasmas, and other bacteria accumulated by the GenBank /EMBL/DDBJ (Release 19), records of only the tRNA and rRNA gene clusters were extracted, and arranged in a KWIC index-like list of gene clusters. In the genomic 40123-base DNA fragment (Locus ML15182) of the Gram-positive bacterium *Mycobacterium lepre*, we found two putative genes, tRNA $^{\mathrm{Glu}}$(TTC) (37225 to 37296, 72bp) and tRNA $^{\mathrm{Asp}}$(GTC) (37328 to 37400, 73 bp). The gene cluster list currently has only two entries whose headwords begin with [ED] (i.e., the tRNA $^{\mathrm{Glu}}$(TTC) -tRNA $^{\mathrm{Asp}}$(GTC) popular gene cluster) : (1) the 5S rRNA -tRNA $^{\mathrm{Gly}}$(GCC) -tRNA $^{\mathrm{Arg}}$(ACG) -tRNA $^{\mathrm{Val}}$(TAC) - [ED], and (2) the tRNA $^{\mathrm{Lys}}$(TTT) - [ED] -tRNA $^{\mathrm{Phe}}$(GAA) of the Gram-positive bacteria *Bacillus subtilis* and *Lactobacillus delbruecki*, respectively. Therefore we tried to find corresponding tRNA $^{\mathrm{Val}}$(TAC), tRNA $^{\mathrm{Lys}}$(TTT), and/or tRNA $^{\mathrm{Phe}}$(GAA) genes also in *M. lepre*, and found a tRNA $^{\mathrm{Phe}}$(GAA) (37432 to 37509, 73bp). The resulted new putative tRNA$^{\mathrm{Glu}}$(TTC) -tRNA $^{\mathrm{Asp}}$(GTC) -tRNA $^{\mathrm{Phe}}$(GAA) gene cluster overlaps the known tRNA $^{\mathrm{Asp}}$(ATC) -tRNA $^{\mathrm{Phe}}$(GGC). Also one more putative tRNA $^{\mathrm{Lys}}$(TTT) (37101 to 37029, complement, 73 bp) came into sight, which overlaps the known tRNA $^{\mathrm{Lys}}$(TTT) (37095 to 37023, complement, 73 bp). They are separated from the known coding regions.

1. Notation

Since the present paper describes the tRNA and rRNA gene clusters in terms of bacteria, their categories, and their tRNA and rRNA genes, some important notations for representing them are defined here in advance.

[RNA gene] The present paper represents either a tRNA gene with its relevant amino acid 1-letter code with codon in lower letters and RNA-mode, or an amino acid 3-letter code with ANTICODON in CAPITAL letters and DNA-mode, e.g. Maug, MetCAT, a tRNA $^{\mathrm{Met}}$(aug) gene and a tRNA $^{\mathrm{Met}}$(CAT) gene are perfectly equivalent. A special kind of methionine, so-called fM, is given not M but X if and only if it is identified. Ribosomal RNAs are represented with 16S, 23S, 5S, 7S and so on.

［Category code］ Each bacterium is classified into the five categories as follows: Archaebacteria （Code A）, Gram-negative （N） and Gram-positive （P） bacteria, Mycoplasmas （M）, and a group of members of other categories and bacteria whose categories are unknown （O）.

［Bacterial species］ Each bacterium is represented with a category code, a hyphen mark, and 5-letter abbreviation of its name: e.g. N-escco for Gram-negative bacterium *Escherichia coli*, and P-bacsu for the Gram-positive one *Bacillus subtilis*.

［A gene cluster］ A gene cluster is defined as a string of two or more tRNA and rRNA genes whose spacer sequences are, in principle, shorter than 100 bp. Some of gene clusters which appeared more than one time are treated as popular gene clusters, and are given short names: e. g., ［GRV］ for GggcRcguVgua.

2. Introduction

It is known that the archaebacteria has several consensus gene clusters, e.g., "16S-23S.......5S" in some species of the crenarchaeota, and "16S-Ala-23S-5S" in many species of the eutyarchaeota [1]. Phylogenetical relationship among *B. Subtilis* （P-bacsu） and some *Mycoplasmas* species was discussed using several consensus gene clusters such as "AgcaIaugI (or M)augSuca" [2]. In fact, *Mycoplasmas* bacteria are sometimes treated as Gram-positive ones. We collected as many records of bacterial tRNA and rRNA gene clusters as possible, and found, in addition that there were many consensus gene clusters, that some of them could be often related to the same respective categories. For example, so far the Glu-Cys-Val and the 16S-Glu-23S gene clusters belong to only Gram-positive （*Micrococcus luteus and Streptmyces lividanlls*） and Gram-negative bacteria （*E. coli and Plesiomonas shigellodes*）, respecstively. Moreover, we confirmed not only that the above mentioned AgcaIaugI (or M) augSuca gene cluster has been found in Gram-positive bacteria and Mycoplasmas [2], but also that it had not been found in Gram-negative bacteria and Archarbacteria as yet [3]. The present paper mainly reports that we deductively found a new putative EgaaDgacFuuc gene cluster of the Gram-positive bacterium *Mycobacterium leprae* （P-mycle）.

3. Data source and the basic tool for analysis.

All data were extracted from the internatioanl database of DNA sequences accumulated by the GenBank /EMBL/DDBJ （Release 19） with annotation by original authors. We collected as many records of tRNA （and rRNA） gene clusters as possible, and edited a KWIC index-like gene cluster list. Each entry is added a unique code indicating its relevant bacterium such as N-escco for *Escherichia coli*. Table 1 lists the bacteria concerned and their name codes. Figure 1 exemplifies gene clusters of some of the bacteria. Table 2 is the current list of the popular gene clusters. Figure 2 denotes all entries of the current gene cluster list, where the mark '＃' is the indicator of each headword.

Table 1. List of relevant bacteria and their codes

A Archaebacteria; N Gram-Negative bacteria; P Gram-positive bacteria; M Mycoplasmas; O Others

Acholeplasma laidlawii	(M-achla)
Actinobacillus actinomycetemcomitans	(N-actac)
Actinoplanes utahensis	(P-actut)
Aeromonas hydrophila	(N-aerhy)
Aeromonas salmonicida	(N-aersa)
Amycolatopsis orientalis	(Pamyor)
Anacystis nidulans	(O-anani)
Arthrospira sp.	(O-artsp)
Azospirillum lipoferum	(N-azoli)
Bacillus sp.	(P-bacsp)
Bacillus subtilis	(P-bacsu)
Bartonella bacilliformis	(N-barba)
Bifidobacterium breve	(P-bifbr)
Bifidobacterium infantis	(P-bifin)
Borrelia burgdorferi	(N-borbu)
Bradyrhizobium japonicum	(O-braja)
Buchnera aphidicola	(N-bucap)
Campylobacter jejuni	(N-capje)
Caulobacter crescentus	(N-caucr)
Chlamydia trachomatis	(N-chltr)
Clostridium acetobutylicum	(P-cloac)
Clostridium perfringens	(P-clope)
Clostridium tyrobutyricum	(P-cloty)
Coxiella burnetii	(N-coxbu)
Desulfurococcus faecalis	(A-desfa)
Enterococcus faecalis	(P-entfa)
Enterococcus hirae	(P-enthi)
Escherichia coli	(N-escco)
Frankia sp.	(P-frasp)
Haemophilus influenzae	(N-halin)
Haloarcula marismortui	(O-halma)
Halobacterium cutirubrum	(A-halcu)
Halobacterium halobium	(A-halha)
Halobacterium marismortui	(A-halma)
Halococcus morrhuae	(A-halmo)
Haloferax volcanii	(A-halvo)
Lactobacillus brevis	(P-lacbr)
Lactobacillus casei	(P-lacca)
Lactobacillus delbrueckii	(P-lacde)
Lactobacillus hetero	(P-Lachet)
Lactobacillus japonics	(P-Lacja)
Lactococcus lactis	(P-lacla)
Legionella anisa	(N-Legan)
Legionella brunensis	(N-legbr)
Legionella cherii	(N-legch)
Legionella cincinatiensis	(N-legci)
Legionella moravica	(N-legmo)
Legionella oakridgenesis	(N-legoa)
Methanobacterium thermoautotrophicum	(A-metth)
Methanococcus vannielii	(A-metva)
Methanosaeta concilii	(A-metco)
Methanosarcina frisia	(A-metfr)
Methanothermus fervidus	(A-metfe)
Micrococcus luteus	(P-miclu)
Mollicutes sp.	(M-molsp)
Mycobacterium africanum	(P-mycaf)
Mycobacterium avium	(P-mycav)
Mycobacterium bovis	(P-mycbo)
Mycobacterium intracellulare	(P-mycin)
Mycobacterium leprae	(P-mycle)
Mycobacterium paratuberculosis	(P-mycpa)
Mycobacterium phlei	(P-mycph)
Mycobacterium smegmatis	(P-mycsm)
Mycobacterium tuberculosis	(P-myctu)
Mycoplasma capricolum	(M-mycca)
Mycoplasma flocculare	(M-mycfl)
Mycoplasma gallisepticum	(M-mycga)
Mycoplasma mycoides	(M-mycmy)
Mycoplasma pneumoniae	(M-mycpn)
Mycoplasma sp.	(M-mycsp)
Mycoplasma sp. strain PG50	(M-mycsp)
Mycoplasma -like organism	(M-mylor)
Mycoplasma -like sp.	(M-mylsp)
Natronobacterium magadii	(A-natma)
Neisseria meningitidis	(N-neime)
Photobacterium leiognathi	(N-phole)
Photobacterium phosphoreum	(N-phoph)
Plesimonas shigelloides	(N-plesh)
Proteus vulgaris	(N-provu)
Pseudomonas aeruginosa	(N-pseae)
Pseudomonas cepacia	(N-psece)
Pseudomonas fluorescens	(N-psefl)
Pseudomonas gladioli	(N-psegl)
Pseudomonas mallei	(N-psema)
Pseudomonas mendocina	(N-pseme)
Pseudomonas pickettii	(N-psepi)
Pseudomonas pseudomallei	(N-pseps)
Pyrococcus furiosus	(A-pyrfu)
Pyrodictium brockii	(A-pyrbr)
Rhodobacter sphaeroides	(N-rhosp)
Rhodothermus marinus	(O-rhoma)
Salmonella typhimurium	(N-salty)
Spiroplasma citri	(M-spici)
Spiroplasma melliferum	(M-spime)
Sprirulina sp.	(O-sprsp)
Staphylococcus aureus	(P-staau)
Staphylococcus hominis	(P-staho)
Staphylococcus warneri	(P-stawa)
Stigmatella aurantiaca	(N-stiau)
Streptococcus agalactiae	(P-strag)
Streptococcus pneumoniae	(P-strpn)
Streptococcus pyogenes	(P-strpy)
Streptococcus salivarius	(P-strsa)
Streptomyces ambofaciens	(P-stram)
Streptomyces coelicolor	(P-strco)
Streptomyces griseus	(P-strgr)
Streptomyces lividans	(P-strli)
Streptomyces rimosus	(P-strri)
Sulfolobus acidocaldarius	(A-sulac)
Sulfolobus solfataricus	(A-sulso)
Synechococcus sp. (PCC 6301)	(O-synsp)
Thermococcus celer	(A-thece)
Thermofilum pendens	(A-thepe)
Thermotoga maritima	(N-thema)
Thermus aquaticus	(N-theaq)
Thermus aquaticus thermophilus	(N-theth)
Thiobacillus ferrooxidans	(N-thife)
Trichdesmium sp.	(O-trisp)
Vibrio harveyi	(N-vibha)
Xanthomonas maltophilia	(N-Xanma)

Pyrodictium brockii (A-pyrbr) 49#
16S-23S- (A-pyrbr) L36461

Rhodobacter sphaeroides (N-rhosp) 227#
16S-IaucAgca23S-5S--Xaug (N-rhosp) X53853
16S-IaucAgca23S-5S--Xaug (N-rhosp) X53854
16S-IaucAgca23S-5S--Xaug (N-rhosp) X53855

Rhodothermus marinus (O-rhoma)
16S-Iauc (O-rhoma) X77140
16S-IaucAgca (O-rhoma) X80994
Salmonella typhimurium (N-salty)
S---RcguRcguRcguRcgu (N-salty) X64175 X70226
RcggHcacLcugPcca (N-salty) X00066 M10156
RcggHcacLcugPcca (N-salty) M95047

Spiroplasma citri (M-spici) 111
WugaWuggSucg (M-spici) M91385

Spiroplasma melliferum (M-spime) 111#
CugcRcguPccaAgcaMaugJaugSucaXaugDgacFuuc (M-spime) X03715

Sprirulina sp. (O-sprsp)

Staphylococcus aureus (P-staau)
X---DgacF--- (P-staau) X66088 S42081
16S-Iauc23S- (P-staau) U11773
16S-23S- (P-staau) U11774
16S-23S- (P-staau)U11775
16S-Iauc23S- (P-staau) U11776
16S-Iauc23S- (P-staau) U11777
16S-23S- (P-staau) U11778
16S-Iauc23S- (P-staau) U11779
16S-Iauc23S- (P-staau) U11780
16S-IaucAgca23S- (P-staau) U11781
16S-Iauc23S- (P-staau) U11782
16S-Iauc23S- (P-staau) U11783
16S-IaucAgca23S- (P-staau) U11784
16S-23S- (P-staau)U11785
16S-IaucAgca23S- (P-staau) U11786
16S-Iauc23S- (P-staau) U11787
16S-23S-(P-staau) U11788
16S-Iauc23S-(P-staau) U11789
5S--VguaTacaKkaaGggcLuuaRcguPccaAgca16S-Iauc23S-5S--16S-23S--
(P-staau) L36472
5S--VguaTacaKaaaLcuaGggcLuuaRcguPccaAgcaMaugMaugSucaDgacSucaMaug
DgacFuucTacaYuacWuggHcacQcaaCugcGggaGggaGggaLuug(P-staau) L11530

Figure 1. Bacteria and their tRNA (and rRNA) gene clusters (part)

Table 2. The current popular tRNA and rRNA gene clusters

Code	The consensus gene	Remarks	Code	The consensus gene	Remarks
[YQ]	YuacQcaa		[IA]	IaucAgca	
[WW]	WugaWugg		[HQ]	HcacQcaa	
[VcVc]	VgucVguc		[GgT]	GgggTacc	
[VVKa]	VguaVguaKaaa		[GaT]	GggaTacc	
[VT]	VguaTaca		[GRV]	GggcRcguVgua	
[VKu]	VguaKuaa		[GL]	GggcLuua	
[TY]	TacaYuac		[GI]	GggaIauc	
[TP]	TacaPcca		[GC]	GggcCugc	
[T5]	Taca5S--		[ED]	EgaaDgac	
[SuEV]	SuccEgaaVgua		[DW]	DgacWugg	
[SE]	SagcEgaa		[DK]	DgacKaaa	
[RRRR]	RcguRcguRcguRcgu		[7S16A]	7S--Sagc16S-Agca	
[RPA]	RcguPccaAgca		[5N]	5S--Naac	
[RH]	RcggHcac		[5C]	5S--Cugc	
[PHP]	PccaHcacPcca		[5-23]	5S--23S-	
[NE]	NaacEgaa		[235-]	23S-5S--	
[MM]	MaugMaug		[16I23]	[16I]23S-	
[ML]	MaugLcua		[16I]	16S-Iauc	
[MJS]	M/IaugM/IaugSuca		[16E23]	16S-Ega23S	
[ME]	MaugEgaa		[16AI]	16S-AgcaIauc	
[MDF]	[MD]Fuuc		[16A23]	16S-Agca23S-	
[MD]	(Maug/Xaug)Dgac		[16235-]	[1623]5S--	
[LP]	LcugPcca		[1623]	16S-23S-	
[KLu]	KaaaLuua				
[KLg]	KaaaLcug				
[KLc]	KaaaLcua				

(Figure. 2) A gene cluster (an owner's name code) Accession Nos.(GenBank/EMBL/DDBJ)

```
                 Tacu  #[YQ][KLu]     (M-mycca) D00548  M18050
                       #[YQ][KLu]Ggga (M-mycpn)   X17113

                         #Yuac[GaT] (N-pseae)X07950
                         #Yuac[16I] (M-mylsp)  M86340
                         #YuacYuac (N-escco)  K01197 J01720 K01198 K01217 K01300 M100704
                    [TP] #YuacKaaa5S--[DK]Dgac (A-metva)     X00916
                    Yuac #Yuac (N-escco)  K01197 J01720 K01198 K01217 K01300 M100704
                     23S-[5N][SuEV][MDF][TY] #Wugg[HQ][GC]LuuaLuug  (P-bacsu)K01389
5S--[VT][KLc][GL][RPA][MJS]DgacSuca[MDF][TY] #Wugg[HQ]CugcGggaGggaGggaLuug (P-staau) L11530

                              #WuggGggc             (M-mycpn)  L15242
                     [MM][TY] #Wugg                  (N-thema) Z11839
                         Tacc #Wugg                  (N-chltr)      L22216

                 #[WW]Sucg (M-spici)   M91385
                 #[WW] (M-mycca) K02974 X07691 D00551

     [GC] #[VcVc]Vguc (P-strli)  X52072
          #[VcVc]   (N-escco)      X52788

         [GC] #VgucGggc (P-miclu)  X55099
   [GC][VcVc] #Vguc (P-strli)  X52072

    Vgua #[VVKa]  (N-escco)        M13687  X63976 X55757 X63977
         #[VVKa]  (N-escco)         X52796

  [NE]Tacg #[VT]KaagLcua                       (M-mycpn)   L15240

    [23-5] #[VT][KLu]Agca[MJS][MDF]   (M-achla)    X61068
  [16235-] #[VT][KLg][GL][RPA][MJS][MDF]Hcac[GI]Naac[SE] (P-bacsu)K00637 M10606 X00007
    [23-5] #[VT][KLg][GL][RPA][MJS][MDF]Hcac[GI]Naac[SE] (P-bacsu) K01390
           #[VT][KLg][GL]Rcgu (P-bacsu)   M23210

      5S-- #[VT][KLc][GL][RPA][MJS]DgacSuca[MDF][TY]Wugg[HQ]CugcGggaGggaGggaLuug(P-staau)L11530
  [16235-] #[VT][KLc][GL][RPA][16235-]  (P-bacsu)  D26185
      5S-- #[VT][KLc][GL][RPA]16S-             (P-bacsu)         M87888
      5S-- #[VT]Kaaa[GL][RPA][16I23]5S--[1623]  (P-staau) L36472
      5S-- #[VT]Kaaa                            (N-azoli)    X75615        Exception
      [NE] #[VT]                                (M-mycmy)  M16450   (M-mycca)  D00549

     VguaVgua #[VKu] (N-escco)          X17321
         Kaaa #[VKu] (N-escco)     X04171

              #VguaVgua[VKu] (N-escco)          X17321
              #Vgua[VVKa]  (N-escco)        M13687  X63976 X55757 X63977
         Vgua #Vgua[VKu] (N-escco)          X17321
              #VguaDgacDgac (N-pseae)X00651 X56183
         Egag #Vgua (N-theaq) U06642
              #V---[MD]F---  {[MDF]}(P-staho)X66090  S42072

       [NE] #Tacg[VT]KaagLcua (M-mycpn)   L15240

            #TaccWugg (N-chltr)      L22216
            #Tacc[TY][GaT]        (N-escco)     X04181
            #TaccMaug (P-miclu) X55100
   23S-[5N] #TaccGggc[RPA]16S-  (P-bacsu)     K01986

                     23S-[5N][SuEV][MDF] #[TY]Wugg[HQ][GC]LuuaLuug  (P-bacsu)     K01389
5S--[VT][KLc][GL][RPA][MJS]DgacSuca[MDF] #[TY]Wugg[HQ]CugcGggaGggaGggaLuug (P-staau) L11530
                                    [MM] #[TY]Wugg (N-thema) Z11839
             #[TY][GgT]        (N-theth)        X51824
        Tacc #[TY][GaT]        (N-escco)     X04181
             #[TY][GaT](N-stiau)X82820  (N-escco)M29107 J01717 X04176
             #[TY]             (N-escco)          V00362
             #[TP]YuacKaaa5S--[DK]Dgac (A-metva)     X00916
             #[TP][DK]                 (A-metfe)    M26977
             #[TP]                      (A-thece)    X68397 S55983 S55986 S55987 S55988

            [23-5] #[T5] (N-escco)        V00334 J01693
        [16IA235-] #[T5]  (N-escco)    U18997
              5S-- #[T5] (N-escco)      D12500

                                #SucgLcug (A-sulso)  V01548
                           Succ #Sucg (M-mycpn)   L15235
                                #SuccSucg (M-mycpn)   L15235

  23S-[5N] #[SuEV][MDF][TY]Wugg[HQ][GC]LuuaLuug  (P-bacsu)    K01389
  rRNANaac #[SuEV][MD]  (P-bacsp)   X60981
                                #SucaSagc  (P-cloac)     L24102
5S--[VT][KLc][GL][RPA][MJS]Dgac #Suca[MDF][TY]Wugg[HQ]CugcGggaGggaGggaLuug (P-staau) L11530
                            Egaa #SucaMaugFuuc[GI][5N]      (P-lacla)X75364

                  #Sagc[RRRR]  (N-escco)      X52797
                  #SagcRcgu (P-strli)   X70689
                  #SagcLcug   (A-halma)      M8783
   [16235-][VT][KLg][GL][RPA][MJS][MDF]Hcac[GI]Naac #[SE]  (P-bacsu)K00637 M10606 X00007
     [23-5][VT][KLg][GL][RPA][MJS][MDF]Hcac[GI]Naac #[SE]  (P-bacsu) K01390
                                               tRNA #[SE]  (P-bacsu)       Z25795
                                                    #[SE]     (M-achla)    X61065
               Suca #Sagc      (P-cloac)L24102
                    #S---[RRRR]  (N-salty)  X64175 X70226

    Sagc #[RRRR]  (N-escco)      X52797
    S--- #[RRRR]  (N-salty)  X64175 X70226

  [16235-][VT][KLg][GL] #[RPA][MJS][MDF]Hcac[GI]Naac[SE]  (P-bacsu)K00637 M10606 X00007
    [23-5][VT][KLg][GL] #[RPA][MJS][MDF]Hcac[GI]Naac[SE]  (P-bacsu) K01390
                   Cugc #[RPA][MJS][MDF]                  (M-spime)   X03715
                        #[RPA][MJS][MDF]   (M-mycmy)X03154  (M-mycca)  D00547
      5S--[VT][KLc][GL] #[RPA][MJS]DgacSuca[MDF][TY]Wugg[HQ]CugcGggaGggaGggaLuug (P-staau)L11530
       5S--[VT]Kaaa[GL] #[RPA][16I23]5S--[1623]  (P-staau) L36472
  [16235-][VT][KLc][GL] #[RPA][16235-]  (P-bacsu)  D26185
      5S--[VT][KLc][GL] #[RPA]16S-  (P-bacsu)          M87888
       23S-[5N]TaccGggc #[RPA]16S-  (P-bacsu)      K01986


                     Sagc #Rcgu (P-strli)   X70689
            [VT][KLg][GL] #Rcgu (P-bacsu)   M23210

        #[RH][PHP]  (N-vibha)  M29762  M31832
   [DW] #[RH][LP]  (N-escco)     M87049
        #[RH][LP]  (N-escco)K01994   (N-salty) X00066  M10156 M95047  (N-aerhy)    X12977
        #[RH]        (N-phoph) M31833 M29761
                      rrr #Rcgg  (P-mycle)  U15186

                        #QcagEgagEgagQcagEgag (P-strli)  X58873
           QcagEgagEgag #QcagEgag (P-strli)  X58873
       [ML]QcaaQcaaMaug #QcagQcag  (N-escco)      J01713
   [ML]QcaaQcaaMaugQcag #Qcag  (N-escco)      J01713

               [ML] #QcaaQcaaMaugQcagQcag  (N-escco)      J01713
           [ML]Qcaa #QcaaMaugQcagQcag  (N-escco)      J01713

         PccaPcca #[PHP][PHP]  (N-)X12976
             [RH] #[PHP]  (N-vibha)  M29762  M31832
    PccaPcca[PHP] #[PHP]  (N-phoph)X12976
                  #[PHP]  (N-phoph) X12975

             Pcca #Pcca[PHP][PHP]  (N-phoph)X12976
                  #PccaPcca[PHP][PHP]  (N-phoph)X12976
            Cugc #Pcca[MJS][MDF] (M-mycpn)  L15234
            [5N] #Pcca[GRV]Dgac (P-lacde) X15245
                 #P---tRNA      (N-escco)    U00339

           #[NE][VT] (M-mycmy)  M16450   (M-mycca)  D00549
           #[NE]Tacg[VT]KaagLcua (M-mycpn)   L15240

                                        rRNA #Naac[SuEV][MD]  (P-bacsp)   X60981
[16235-][VT][KLg][GL][RPA][MJS][MDF]Hcac[GI] #Naac[SE] (P-bacsu)K00637 M10606 X00007
  [23-5][VT][KLg][GL][RPA][MJS][MDF]Hcac[GI] #Naac[SE] (P-bacsu) K01390
                                             #NaacNaac (P-strli)    X52070
                                             #Nauc[ME]LcuaHcac (A-metfe)    M26978
                                        Naac #Naac (P-strli)   X52070

        #[MM][TY]Wugg         (N-thema) Z11839
        #[MM]                 (N-escco)        M21681

                      #[ML]QcaaQcaaMaugQcagQcag  (N-escco)     J01713
                      #[ML]                      (N-phole) X52969

  [16235-][VT][KLg][GL][RPA] #[MJS][MDF]Hcac[GI]Naac[SE] (P-bacsu)K00637 M10606 X00007
    [23-5][VT][KLg][GL][RPA] #[MJS][MDF]Hcac[GI]Naac[SE] (P-bacsu) K01390
                   Cugc[RPA] #[MJS][MDF]                 (M-spime)   X03715
                       [RPA] #[MJS][MDF]           (M-mycmy)  X03154    (M-mycca)  D00547
                    CugcPcca #[MJS][MDF]           (M-mycpn)  L15234
         [23-5][VT][KLu]Agca #[MJS][MDF]   (M-achla)    X61068
      5S--[VT][KLc][GL][RPA] #[MJS]DgacSuca[MDF][TY]Wugg[HQ]CugcGggaGggaGggaLuug (P-staau) L11530

                  Nauc #[ME]LcuaHcac (A-metfe)   M26978
                       #[ME]         (P-bacsu)  D26185 L23497

 5S--[VT][KLc][GL][RPA][MJS]DgacSuca #[MDF][TY]Wugg[HQ]CugcGggaGggaGggaLuug   (P-staau) L11530
                      23S-[5N][SuEV] #[MDF][TY]Wugg[HQ][GC]LuuaLuug  (P-bacsu)    K01389
     [16235-][VT][KLg][GL][RPA][MJS] #[MDF]Hcac[GI]Naac[SE] (P-bacsu)K00637 M10606 X00007
       [23-5][VT][KLg][GL][RPA][MJS] #[MDF]Hcac[GI]Naac[SE] (P-bacsu) K01390
            [23-5][VT][KLu]Agca[MJS] #[MDF]                 (M-achla)    X61068
                      Cugc[RPA][MJS] #[MDF]                 (M-spime)   X03715
                          [RPA][MJS] #[MDF]    (M-mycmy)  X03154    (M-mycca)  D00547
                       CugcPcca[MJS] #[MDF]                (M-mycpn)  L1523

                 #X---DgacF---   (P-stawa) X66089 S42075  (P-staau)  X66088 S42081
            V--- #[MD]F--- (P-staho)X66090  S42072
  rRNANaac[SuEV] #[MD]  (P-bacsp)   X60981
          [23-5] #[MD]  (P-bacsu)     L08236

      EgaaSuca #MaugFuuc[GI][5N]  (P-lacla)      X75364
               #MaugF---          (P-strpy) M95072

                   Tacc #Maug (P-miclu) X55100
         16S-[IA][23-5] #Xaug (N-rhosp) X53853   X53854 X53855
                 [1623] #Maug (A-thepe) X14835

   [DW][RH] #[LP] (N-escco)    M87049
       [RH] #[LP] (N-escco)K01994  (N-salty)X00066 M10156 M95047  (N-aerhy)X12977


                             23S-[5N][SuEV][MDF][TY]Wugg[HQ][GC]Luua #Luug (P-bacsu)K01389
5S--[VT][KLc][GL][RPA][MJS]DgacSuca[MDF][TY]Wugg[HQ]CugcGggaGggaGgga #Luug (P-staau)L11530

    23S-[5N][SuEV][MDF][TY]Wugg[HQ][GC] #LuuaLuug  (P-bacsu)    K01389
                                     [GC] #Luua (N-escco)        X52789

                               #LcugLcugLcug     (N-escco)    U14003  V00363
                          Lcug #LcugLcug     (N-escco)    U14003  V00363
                           Sucg #Lcug (A-sulso)  V01548
                          Sagc #Lcug   (A-halma)       M8783
                      LcugLcug #Lcug     (N-escco)    U14003  V00363
           Nauc[ME] #LcuaHcac (A-metfe)   M26978
   [NE]Tacg[VT]Kaag #Lcua (M-mycpn)   L15240
               [HQ] #Lcua   (M-achla)    X61067

           [NE]Tacg[VT] #KaagLcua (M-mycpn)   L15240

                 #Kaaa[VKu] (N-escco)     X04171

                   [YQ] #[KLu]Ggga            (M-mycpn)   X17113
             [23-5][VT] #[KLu]Agca[MJS][MDF]   (M-achla)    X61068
               Tacu[YQ] #[KLu]                (M-mycca) D00548    M18050

16235-][VT] #[KLg][GL][RPA][MJS][MDF]Hcac[GI]Naac[SE] (P-bacsu)K00637 M10606 X00007
 [23-5][VT] #[KLg][GL][RPA][MJS][MDF]Hcac[GI]Naac[SE] (P-bacsu) K01390
       [VT] #[KLg][GL]Rcgu (P-bacsu)   M23210

    5S--[VT] #[KLc][GL][RPA][MJS]DgacSuca[MDF][TY]Wugg[HQ]CugcGggaGggaGggaLuug (P-staau)L11530
[16235-][VT] #[KLc][GL][RPA][16235-]  (P-bacsu)  D26185
    5S--[VT] #[KLc][GL][RPA]16S-  (P-bacsu)         M87888
             #[KLc]16S-            (M-mycca) X06727
             #[KLc]               (M-mycsp)X05660    (M-mycca) D00550
        5S--[VT] #Kaaa[GL][RPA][16I23]5S--[1623]  (P-staau) L36472
                 #Kaaa[ED]Fuuc (P-bacsu)         D26185    (P-bacsu)    X00889 K00968
        [TP]Yuac #Kaaa5S--[DK]Dgac               (A-metva)     X00916
        5S--[VT] #Kaaa                           (N-azoli)    X75615
            [GL] #Kaaa  (N-halin)        X53782

16S- #[IA]23S-5S-- (N-escco)D12649 (N-thife) U18989   (P-bacsu)D26185 [(O-braja) Z35330]
16S- #[IA]23S- (N-thife)X07395 (N-psepi)L28163 (N-psefl)L06304 L06305 L06306 L06307 (N-psegl) L28156
    L28157 (N-psema)  L28158 (N-neime)L31411 (N-psece) L28151 L28152 L28153  L28145  (N-caucr)L00194
   (N-escco)J01702 M87049 (N-pseps)L28164 (N-pseae)L28148 L28149 L28150  L28165  (N-barba)L26364
   (O-artsp)X70769 (O-synsp)X00346 (P-staau)U11781 U11781 U11784 U11786 (P-bacsu)J01551 (M-achla)D13259

       #[IA]23S- (N-escco) K00763  [(O-trisp) X72871 ]
  16S- #[IA] (O-anani)      K01983    (O-rhoma)  X80994
       #[IA]   (N-coxbu)X78986 (N-pseae)X06693 (M-mycpn)  L15238

                        23S-[5N][SuEV][MDF][TY]Wugg #[HQ][GC]LuuaLuug (P-bacsu)K01389
   5S--[VT][KLc][GL][RPA][MJS]DgacSuca[MDF][TY]Wugg #[HQ]CugcGggaGggaGggaLuug (P-staau)L11530

[16235-][VT][KLg][GL][RPA][MJS][MDF] #Hcac[GI]Naac[SE] (P-bacsu)K00637 M10606 X00007
  [23-5][VT][KLg][GL][RPA][MJS][MDF] #Hcac[GI]Naac[SE] (P-bacsu) K01390
                        Nauc[ME]Lcua #Hcac (A-metfe)   M26978

            5S-- #[GRV][ED] (P-lacde)      X15246
        [5N]Pcca #[GRV]Dgac (P-lacde)       X15245

    5S--[VT][KLc] #[GL][RPA][MJS]DgacSuca[MDF][TY]Wugg[HQ]CugcGggaGggaGggaLuug (P-staau)L11530
     5S--[VT]Kaaa #[GL][RPA][16I23]5S--[1623]       (P-staau) L36472
[16235-][VT][KLc] #[GL][RPA][16235-]              (P-bacsu)  D26185
    5S--[VT][KLc] #[GL][RPA]16S-                  (P-bacsu)  M87888
        [VT][KLg] #[GL]Rcgu                 (P-bacsu)   M23210
                  #[GL]Kaaa              (N-halin)        X53782

            Fuuc #Gggg (A-sulso) X06053

    23S-[5N]Tacc #Gggc[RPA]16S-  (P-bacsu)     K01986
                 #GggcGggcGggc (N-escco)  X53236   U14003
            Gggc #GggcGggc (N-escco)  X53236   U14003
                                 #[GC][VcVc]Vguc (P-strli)  X52072
                                 #[GC]VgucGggc (P-miclu)   X55099
 23S-[5N][SuEV][MDF][TY]Wugg[HQ] #[GC]LuuaLuug  (P-bacsu)    K01389
                                 #[GC]Luua (N-escco)        X52789
            Wugg #Gggc (M-mycpn)  L15242
        [GC]Vguc #Gggc (P-miclu)   X55099
        GggcGggc #Gggc (N-escco)  X53236   U14003
```

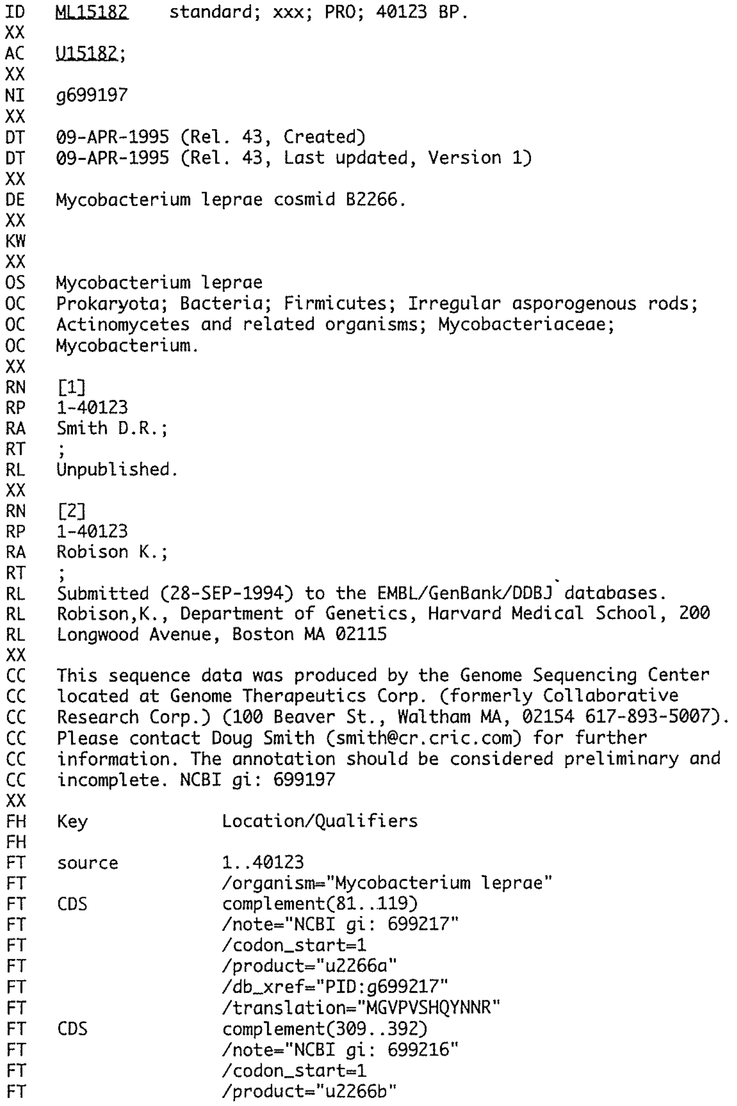
```
               Yuac #[GaT] (N-pseae)X07950
           Tacc[TY] #[GaT]        (N-escco)     X04181
               [TY] #[GaT]        (N-stiau) X82820 (N-escco)    M29107 J01717   X04176

 5S--[VT][KLc][GL][RPA][MJS]DgacSuca[MDF][TY]Wugg[HQ]CugcGggaGgga #GggaLuug(P-staau) L11530
[16235-][VT][KLg][GL][RPA][MJS][MDF]Hcac #[GI]Naac[SE] (P-bacsu)K00637 M10606 X00007
  [23-5][VT][KLg][GL][RPA][MJS][MDF]Hcac #[GI]Naac[SE] (P-bacsu) K01390
                        EgaaSucaMaugFuuc #[GI][5N]  (P-lacla)      X75364
     5S--[VT][KLc][GL][RPA][MJS]DgacSuca[MDF][TY]Wugg[HQ]CugcGgga #GggaGggaLuug
         5S--[VT][KLc][GL][RPA][MJS]DgacSuca[MDF][TY]Wugg[HQ]Cugc #GggaGggaGggaLuug

     EgaaSucaMaug #Fuuc[GI][5N] (P-lacla)      X75364
                  #FuucGggg (A-sulso) X06053
             Maug #F---         (P-strpy) M95072
         V---[MD] #F--- (P-staho)X66090  S42072
         X---Dgac #F--- (P-stawa) X66089 S42075 (P-staau)  X66088 S42081
         Kaaa[ED] #Fuuc (P-bacsu)         D26185   (P-bacsu)    X00889 K00968

                           #EgagVgua (N-theaq) U06642
                  QcagEgag #EgagQcagEgag (P-strli)  X58873
                      Qcag #EgagEgagQcagEgag (P-strli)  X58873
          QcagEgagEgagQcag #Egag (P-strli)  X58873

                       #EgaaSucaMaugFuuc[GI][5N] (P-lacla)      X75364
                  Kaaa #[ED]Fuuc (P-bacsu)  D26185   (P-bacsu)    X00889 K00968
             5S--[GRV] #[ED] (P-lacde)       X15246
                       #Egaa[23-5]   (N-bucap)     U09230
                  16S- #Egaa                    (N-actac) U07776    U07777

                 #[DW][RH][LP] (N-escco)    M87049
          [23-5] #[DW]  (N-escco)       J01696
            5S-- #[DW]   (N-escco)   V00349
                 #[DW]  (N-escco)      K02846

      [TP]YuacKaaa5S-- #[DK]Dgac (A-metva)     X00916
                  [TP] #[DK]     (A-metfe)    M26977

 5S--[VT][KLc][GL][RPA][MJS] #DgacSuca[MDF][TY]Wugg[HQ]CugcGggaGggaGggaLuug (P-staau) L11530
                V---Maug #DgacF--- (P-staho)X66090  S42072
                      X--- #DgacF--- (P-stawa)X66089 S42075 (P-staau) X66088 S42081
                        Vgua #DgacDgac (N-pseae)X00651 X56183
          rRNANaac[SuEV]Maug #Dgac  (P-bacsp)   X60981
               [5N]Pcca[GRV] #Dgac (P-lacde)       X15245
                  [23-5]Maug #Dgac  (P-bacsu)     L08236
        [TP]YuacKaaa5S--[DK] #Dgac (A-metva)     X00916
                    VguaDgac #Dgac (N-pseae)X00651 X56183
                      [23-5] #Dgac       (N-escco)      V00336  J01698

                                                   #Cugc[RPA][MJS][MDF] (M-spime)X03715
                                                   #CugcPcca[MJS][MDF] (M-mycpn)L15234
  [VT][KLc][GL][RPA][MJS]DgacSuca[MDF][TY]Wugg[HQ] #CugcGggaGggaGggaLuug (P-staau) L11530
                                                   #Cuga[16E23]5S-- (N-escco) L10328

             #AgccAgcc (N-escco)M13687  X63976 X55757 X63977  X52795
        Agcc #Agcc (N-escco)M13687  X63976 X55757 X63977  X52795

     [23-5][VT][KLu] #Agca[MJS][MDF]   (M-achla)    X61068
                rRNA #Agca23S--  (N-Xanma) L28166
                rRNA #AgcarRNA  (P-enthi)  L00925
                     #A---[23-5]    (O-synsp)    X00512

                #[7S16A][23-5]  (A-metth)     X15364 X05482
                #[7S16A] (A-metfe)   M32222
                #7S--5S--  (N-escco)         X00162
```

```
                      23S- #[5N]TaccGggc[RPA]16S-  (P-bacsu)     K01986
                      23S- #[5N][SuEV][MDF][TY]Wugg[HQ][GC]LuuaLuug  (P-bacsu)    K01389
                           #[5N]Pcca[GRV]Dgac (P-lacde)       X15245
                   [16A23] #[5N]   (P-lacla)       X64887
                      23S- #[5N] (P-strsa)  S43413     (M-achla)     X61066
      EgaaSucaMaugFuuc[GI] #[5N](P-lacla)      X75364

   [16A23] #[5C] (A-halmo)       X72588
      23S- #[5C]  (A-halcu)      K02972
           #[5C] (A-halvo)        X02128

      #[5-23]  (N-borbu)    X57791  S37072

 23S- #[23-5]  (N-coxbu)     X79704
      #[23-5]   (A-metva)    X02729   (N-legbr) Z30536  (N-legmo) Z30457  (N-legoa) Z30546 (N-escco)
                   X56780 X57090   (P-mycle)   X56657    (O-halma) X13738

 #[16I23] (P-staau)U11773 U11776-7 U11779-80 U11782-3 U11787 U11789 (M-mylor)  U15442
 #[16I] (N-caucr)  X63363  (N-escco)  K01057 (N-caucr)X63362   (O-rhoma) X77140
                             (M-mycsp)  X76560     (M-molsp) X83431  X83438 X83432
 #16S-I--- {[16I]} (N-escco)  V00333
 #[16E23]5S--  (N-escco)       U00006    J01695
 #[16E23] (N-halin)L31410 (N-escco)V00348 X12420 (N-actac) U077768  (N-aersa)   L25601

            #[16AI][23-5]  (N-camje)     Z29326
            #[16AI]  (N-camje)     X04108
 #[16A235-][23-5]  (N-borbu)     U03396
 #[16A235-](A-halha)    X03407 X00872 X01699 (A-natma)  X72495
 #[16A23] (A-halcu)K02971 (A-metfr)X69874  (A-metco)X51423 (A-metva)X00083 (A-pyrfu)
        L36458  (N-coxbu) L16515 (P-strpn)  L31413   M60763   (P-strag) L31412

 #[16235-]  (P-frasp)M88466   (P-frasp)M55343 M58598 (P-stram)  M27245 (P-strgr) M76388 X55435 X61478
            (P-strri)X62884  (P-cloac)L08062  (M-mycga) L08897
 #[1623] (A-pyrbr)L36461 (P-strco) Y00411 M35377 (P-staau)U11774 U11775 U11778 U11785 U11788 (Pmyctu)X58890
       S45391 (P-mycph)X74493  (P-mycpa)X74495 (P-mycav)X74494   (M-mycsp) M94728    (M-mycfl)   L22210

 #[1623]interval  210 (A-pyrbr)L36461 276(P-strco) Y00411  M35377  (P-staau)469 U11774   302 U11775 362
U11778  319  U11785  364 U11788    274 (P-myctu)X58890 S45391  359 (P-mycph)  X74493  274(P-mycpa) X74495
274 (P-mycav)X74494  mavt others  226  (M-mycsp)226 M94728    395(M-mycfl)   L22210
```

Figure 2. The current KWIC index-like gene cluster list（# : a heading）

4. Results

Figure 3 shows three parts of the 40123-base pair DNA fragment (Locus ML15182, Accession No. U15182) of *M. leprae* (P-mycle) : (a) the heading part, (b) two parts covering the four tRNA genes, and (c) the relevant regions of the DNA sequence. As shown in Figure 3 (b), the four known tRNA genes are "Asp (ATC) [or Dcau]" (37219 to 37291, 73 bp), "Phe (GGC) [or Fgcc]" (37322 to 37395, 74 b), Lys (TTT) [or Kaaa] (complement, 37095 to 37023, 73 bp), and "Glu (ATT) [or Eaau]" (complement, 3315 to 3233, 83 bp). Attempting to confirm their cloverleaf strucutres, we found two new putative genes, Egaa (37225 to 37296, 72 bp) and Dgac (37328 to 37400, 73bp), which can form one of the popular gene clusters [ED].



```
ID   ML15182    standard; xxx; PRO; 40123 BP.
XX
AC   U15182;
XX
NI   g699197
XX
DT   09-APR-1995 (Rel. 43, Created)
DT   09-APR-1995 (Rel. 43, Last updated, Version 1)
XX
DE   Mycobacterium leprae cosmid B2266.
XX
KW
XX
OS   Mycobacterium leprae
OC   Prokaryota; Bacteria; Firmicutes; Irregular asporogenous rods;
OC   Actinomycetes and related organisms; Mycobacteriaceae;
OC   Mycobacterium.
XX
RN   [1]
RP   1-40123
RA   Smith D.R.;
RT   ;
RL   Unpublished.
XX
RN   [2]
RP   1-40123
RA   Robison K.;
RT   ;
RL   Submitted (28-SEP-1994) to the EMBL/GenBank/DDBJ databases.
RL   Robison,K., Department of Genetics, Harvard Medical School, 200
RL   Longwood Avenue, Boston MA 02115
XX
CC   This sequence data was produced by the Genome Sequencing Center
CC   located at Genome Therapeutics Corp. (formerly Collaborative
CC   Research Corp.) (100 Beaver St., Waltham MA, 02154 617-893-5007).
CC   Please contact Doug Smith (smith@cr.cric.com) for further
CC   information. The annotation should be considered preliminary and
CC   incomplete. NCBI gi: 699197
XX
FH   Key             Location/Qualifiers
FH
FT   source          1..40123
FT                   /organism="Mycobacterium leprae"
FT   CDS             complement(81..119)
FT                   /note="NCBI gi: 699217"
FT                   /codon_start=1
FT                   /product="u2266a"
FT                   /db_xref="PID:g699217"
FT                   /translation="MGVPVSHQYNNR"
FT   CDS             complement(309..392)
FT                   /note="NCBI gi: 699216"
FT                   /codon_start=1
FT                   /product="u2266b"
```

(a) The heading part

Figure 3. The 40123-bp DNA fragment of *Mycobacterium leprae* (part). This was gotten from GBDET integranted database retrieval system, GenomeNet (Kyoto Center)

```
FT                   RTGIPTVDVGAAQLAMHSARELMGAHDVAVYSAALQAFFSADLF"
FT   CDS             2870..5134
FT                   /note="NCBI gi: 699222"
FT                   /codon_start=1
FT                   /product="purL"
FT                   /db_xref="PID:g699222"
FT                   /translation="MIDTVEYAATTPDQPQPFAELGLREDEYQRVREILGRRPTDTELA
FT                   MYSVMWSEHCSYKSSKVHLRYFGETTTEEMRTGMLAGIGENAGVVDIGDGWAVTFKVES
FT                   HNHPSYVEPYQGAATGVGGIVRDIMAMGARPVAVMDQLRFGAADALDTRRVLDGVVRGI
FT                   GGYGNSLGLPNIGGETVFDSCYDGNPLVNALCVGVLRQEDLHLAFASGAGNKIILFGAC
FT                   TGLDGIGGVSVLASDTFDAEGARKKLPSVQVGDPFMEKVLIECCLELYAGGLVIGIQDL
FT                   GGAGLSCATSELASAGDVGMAIQLDTVPRRAKDMTPAEVFCSESQERMCAVVAPENVDA
FT                   FLAVCRKWEVLATVIGEVTDGDRLRITWHGETVVDVPPRTVAHEGPVYQRPVSRPESQE
FT                   ALNADSSKGLPRPVSGDELRATLLALLGSPHLCSRAFITEQYDRYVRGNTVLAEHADAG
FT                   VLRIDESTGRGIALSTDASGRYTRLDPYAGAQLALAEAYRNVAVTGATPVAVTNCLNFG
FT                   SPEDPGVMWQFAQAVRGLADGCAALKIPVTGGNVSFYNQTGAVAILPTPVVGVLGVLDN
FT                   VARRIHTSLGTEPGEILMLLGDTYDEFDGSVWAQVMAGHLGGLPPMVDLAREKLLAEVL
FT                   SSASRDELVSAAHDLSEGGLAQAIVESALAGETGCRIALPEDADPFVMLFSESAGRVLV
FT                   AVPRPEESRFRSMCEARGLPAMRIGVVDQGSDSIEVRGQFTVSLAELRMTFEAVLPRFF
FT                   G"
FT   tRNA            complement(3233..3315)
FT                   /note="codon recognized: ATT; aa: Glu"
FT   CDS             5267..6220
FT                   /note="NCBI gi: 699231"
FT                   /codon_start=1

FT                   /translation="MRTPDKLKEIERSVHSTTCDWGCPAGSYKVPCSPVFEGYAKDELA
FT                   SRCGR"
FT   CDS             complement(36865..36972)
FT                   /note="NCBI gi: 699201"
FT                   /codon_start=1
FT                   /product="u2266q"
FT                   /db_xref="PID:g699201"
FT                   /translation="MQQLVCQWSTGVVPDLLARRAAKRRVTAISLVMLS"
FT   tRNA            complement(37023..37095)
FT                   /note="codon recognized: TTT; aa: Lys"
FT   tRNA            37219..37291
FT                   /note="codon recognized: ATC; aa: Asp"
FT   tRNA            37322..37395
FT                   /note="codon recognized: GGC; aa: Phe"
FT   CDS             37596..37745
FT                   /note="NCBI gi: 699200"
FT                   /codon_start=1
FT                   /product="u2266r"
FT                   /db_xref="PID:g699200"
FT                   /translation="MWSQVRSRFVATQPSLSVYEIAVDIGGASWCGSSVRSVRSAPDTR
FT                   ARLV"
FT   CDS             38211..38435
FT                   /note="NCBI gi: 699199"
FT                   /codon_start=1
FT                   /product="u2266s"
FT                   /db_xref="PID:g699199"
```

(Figure 3) (b) Two parts describing the four tRNA genes.

```
cgttgaatat gctgctacca cccccgacca gccgcaaccg tttgctgagc tgggcctcag      2940
agaggacgag taccagcgag ttcgcgaaat tctaggtcgg cgacctacgg acaccgagct      3000
ggcgatgtat tcggtgatgt ggagtgagca ctgttcgtac aagtcttcca aagttcacct      3060
gcgttacttc ggtgaaacca ccaccgagga gatgcgtacc ggaatgctgg ccggtatcgg      3120
tgaaaatgct ggtgttgtcg acatcggtga cggctgggcc gtcactttca aggtagaatc      3180
gcacaaccac ccttcttatg tcgagccgta ccagggggcc gcaaccggcg tcggcggcat      3240
cgtacgtgac atcatggcga tgggcgcgcg acctgtcgct gtgatggacc agcttcggtt      3300
cggtgctgcc gatgcgctcg atacccgccg cgtgctggat ggcgtggtcc gcggtattgg      3360
cggctacggc aactcgctgg gcttgcccaa catcggcggc gagactgtct tcgactcgtg      3420
ctacgacggc aatcctctgg tgaatgcgtt gtgtgtcggg gtgttacgcc aggaggatct      3480
tcatttggcg tttgcctcgg gtgcgggtaa caagatcatt ctgttcggcg cgtgtaccgg      3540
gctggacggt atcggcgggg tctcggtgtt ggcgtcggac accttcgatg ctgaaggagc      3600
gcggaagaaa ctcccgtcgg ttcaggtggg tgacccettt atggaaaagg tactcatcga      3660
gtgctgtctg gaattgtatg ctggcgggct ggtgatcggc atccaagact taggaggtgc      3720
```

```
tggtgagcaa gtcctgaaat actgcacgaa ctggtgtgta caaatcgtga taattgtagc      36480
cagcatatcc ctgtcgtact tcggaatagg cttgttgcgt gtccagaagc tggggtcgat      36540
atccgttagg aaaaatggtc taggtttggg ccctatgacc cggtgcagat gggaattaga      36600
atcgaccgcc tgaccgatat tcctgggccg taacggagaa tgagtgattt gcctcgctgc      36660
gtttcgcggg tggcatcggt gctgtgccag agccaccatc ttgcaagcca cgttttgagt      36720
ctgacgtttg attggcaact gatggcaaac cggtgttgca gctcctagcc ggatgattcg      36780
gttcgggaat cagttgccgg gtgacggatt tccagtgtga ccagggcaac gggcctggtt      36840
aaggttgact ccatcgccac ttattcagga aagcataacc aacgatatgg cggtaacacg      36900
tcgtttagcc gcgcgccgtg ccagtaagtc cggcaccacc ccggtagacc attggcaaac      36960
gagttgctgc acggcgctat cgtcgaaggc gccaacatcg gataacgtaa tgatagcagg      37020
gggatcagtg cccccactag gactcgaacc taggacctgc ggattaaaag tccgtagctc      37080
taccaactga gctatagggg cgtggatatt taggataccg gtctcgggaa ctgagctcgt      37140
ttgaggattg ggcccacggt gacctaagct gactgggctc caacggtaac tgcgttgtga      37200
gtaccccgga gagatgcggt tctggccccc ttcgtctagt ggcctaggac gccgcccttt      37260
caaggcggta gcgcgggttc gaatcccgta gggggtactg ctacgcggtg acgtggagca      37320
caagcaaggc cctgtggcgc agttggttag cgcgccgccc tgtcacggcg gaggtcgcgg      37380
gttcgagtcc cgtcagggtc gctagcgcag cgaggcacac acgctgcctt ctggccaggt      37440
agctcagtcg gtatgagcgt ccgcctgaaa agcggaaggt cggcggttcg atcccgcccc      37500
tggctaccgg ttctgagctg tgcggacact gtgatggcag ctatggacgt ttcagtgtct      37560
ccggtttcta tccgttcgcc tgggttgctg cggagatgtg gagccaggtc cggtcgaggt      37620
tcgtggctac gcagccgagt ttgtcggtgt atgagatcgc tgtagatatt ggtggtgcca      37680
gttggtgtgg atcttctgta cgtagcgtaa ggtcggcgcc ggacactcgg gccaggctgg      37740
tatgaggttg tggatggtca caggtgcgcc aagtgggttt tttgaggaat ttgttgcagt      37800
tggtttccgg cgtcagttgc ttgatctgag taaagccgcc ggtcgggtga ggtagaccac      37860
ttgttcttga ggtgcatggt tgtcgatgcg tggattgagg gtatcgataa cgattggcgt      37920
gcagtgggac ggtgcggacg gctgcgcgcg ttttggcgag ccaactatga tccggctttt      37980
gactgccggc gcggcttggc ggacgttgag cctgctcgcc gttaggtaga tgtcgccgac      38040
gctcaggccg acgagttctg accagtacaa ttcggtacag gtcaggatag ttacgacgtt      38100
gccttggtcg ccacacgccg ctgcgagcgt ggccacttgg gcgtccgtga ggtagcggtg      38160
gcgttctcag gacgcatcca ctcggctgat acgtcctgat ctgagtttat gtgaatccga      38220
ccattttagc aggctatgtc gagaatcgac cgcagcagcc gaagcgtggt gaatttagct      38280
cacgggccga cagtgagggt gtcgacgaat tgctatacgc ccttgcggca gattttgtcg      38340
gtcggttagg agtttccaat gctgggtgta accactccac gtcttgggcg acacggctgg      38400
gtgttttgac gcgctgtatt tgggtccaga tttaggcgag tgtgatgcag cccagtcggg      38460
ggtcgaaccg cctggtgcca agtgtcgctt cactgttccg ctcggctttg aaaatttacg      38520
cttatagcaa taaggcacag gcggatggcg tttagaccca tctcgacgag aaggctggat      38580
```

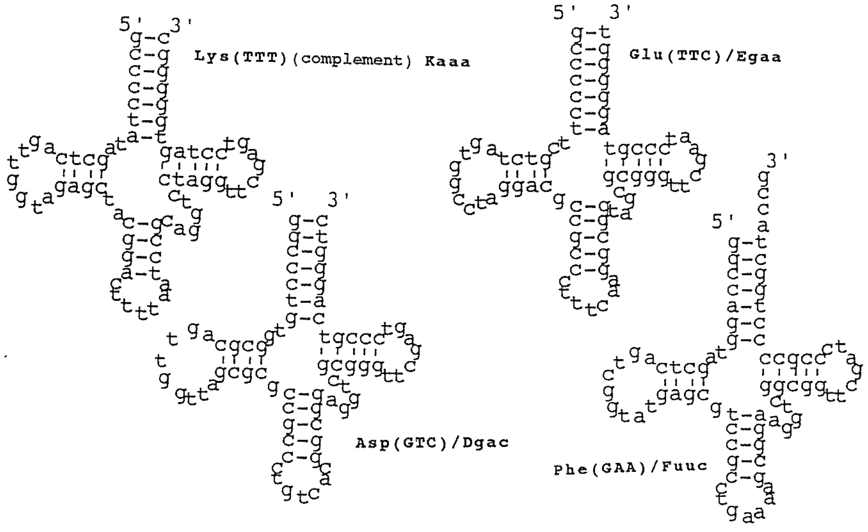
(Figure 3) (c) Two parts of the DNA sequences covering the four known tRNA genes.

The gene cluster list currently has only two entries whose headwords are [ED] as follws:

Kaaa #[ED] Fuuc (P-bacsu)D26185 X00889 K00968
5S- [GRV] #[ED] (P-lacde)X15246.

Because *M. leprae* (P-mycle) also is Gram-positive, we searched sequences near this putative [ED] for new putative Kaaa (upstream), Fuus (downstream), and/or Vgua (upstream), and fortunately found Fuuc (37432 to 37509, 78 bp). Moreover, one more putative gene, Laaa (37101 to 37029, complement, 73 bp), came in sight. Figure 4 shows some data on the known and the putative tRNA genes: (a) DNA sequences, (b) their proposed cloverleaf structures, and (c) relationship of positions.

```
  <-------------the known Lys(TTT)(complement)   Kaaa (37023-37095)---------
                  .......                                                             . ......
 gatcagt|gccccca ctaggactcgaacctaggacctgcggattaaaagtccgtagctctaccaactgagctat|aggggc
          <-------the putative Lys(TTT)(complement) Kaaa (37029-37101)----------

  ------------- the known  Asp(ATC) / Dgau (37210-37291)------------------->
                .......                                                          .. .....
 gttctg|gcccccttcgtctagtggcctaggacgccgcccttcaaggcggtagcgcgggttcgaatcccgtag|ggggt
        ---------the putative Glu(TTC)/Egaa(37225-37296) --------------------->

 --------------- the known   Phe(GGC) / Fgcc (37322-37395)   --------------->
               .......                                                           .. .....
aagcaa|ggccctgtggcgcagttggttagcgcgccgccctgtcacggcggaggtcgcgggttcgagtcccgtca|gggtc
        ---------the putative Asp(GTC)/Dgac (37328-37400)-------------------->

  .......                                                            .......
 ggccaggtagctcagtcggtatgagcgtccgcctgaaaagcggaaggtcggcggttcgatcccgccccctggctaccg
 --------------A putative Phe(GAA)/Fuuc  (37342-37509)-------------------->
(a)  Assignment of DNA sequences for the putative tRNA genes
```

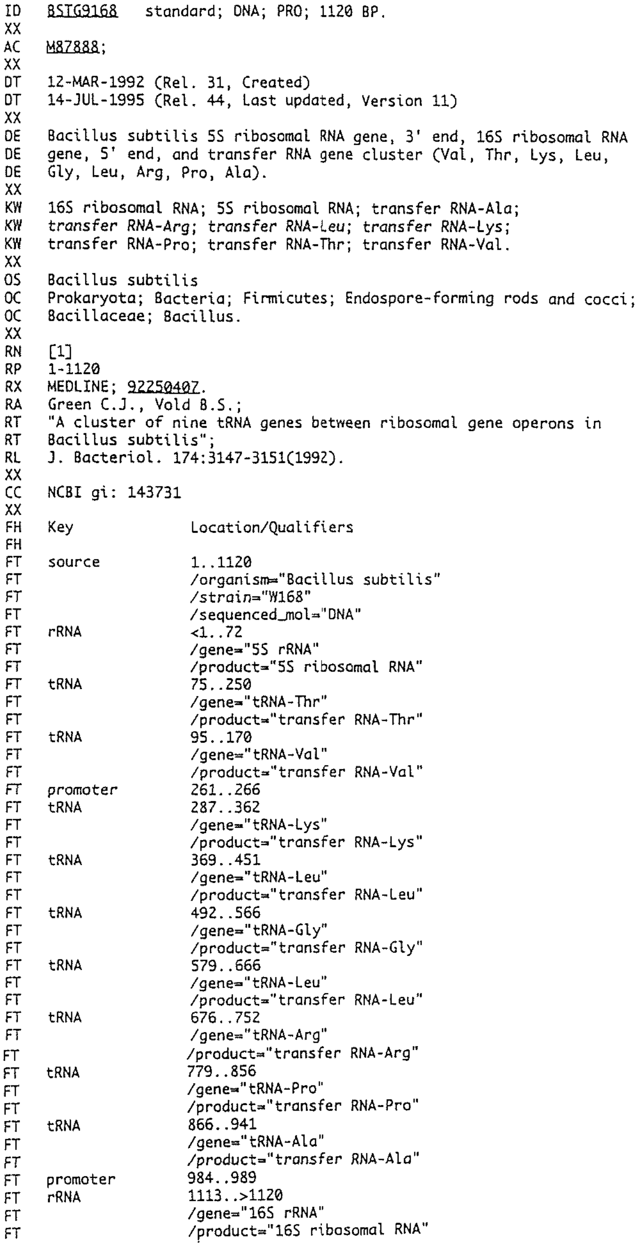


(b) Proposed cloverleaf structures for the four putative tRNA genes.

```
        37000        100  200       300       400       500   37600
            |      |    |  |  |     |     |     |     |     |    |  |    |
The known    <--Kaaa--         --Dgau->   --Fgcc->                (CDS 596-
Putative      <--Kaaa-         --Egaa->   -Dgac->    --Fuuc->     (CDS 596-

(c)  Relationship of positions
```

Figure 4 . Relationship between the three known and the four new putative tRNA genes.

In order to reveal one more relevant fact, Figure 5 shows the 9-tRNA gene cluster of the 1120-bp DNA fragment (Locus BSTG9168, Accession No. M87888) of *B. subtilis* (P-bacsu) consisting of Taca (75 to 250, 176 bp), Vgua (95 to 170, 76 bp) and other seven ones. The most upstream part (75 to 174) of the Taca perfectly covers the Vgua (95 to 170).



```
ID   BSTG9168   standard; DNA; PRO; 1120 BP.
XX
AC   M87888;
XX
DT   12-MAR-1992 (Rel. 31, Created)
DT   14-JUL-1995 (Rel. 44, Last updated, Version 11)
XX
DE   Bacillus subtilis 5S ribosomal RNA gene, 3' end, 16S ribosomal RNA
DE   gene, 5' end, and transfer RNA gene cluster (Val, Thr, Lys, Leu,
DE   Gly, Leu, Arg, Pro, Ala).
XX
KW   16S ribosomal RNA; 5S ribosomal RNA; transfer RNA-Ala;
KW   transfer RNA-Arg; transfer RNA-Leu; transfer RNA-Lys;
KW   transfer RNA-Pro; transfer RNA-Thr; transfer RNA-Val.
XX
OS   Bacillus subtilis
OC   Prokaryota; Bacteria; Firmicutes; Endospore-forming rods and cocci;
OC   Bacillaceae; Bacillus.
XX
RN   [1]
RP   1-1120
RX   MEDLINE; 92250407.
RA   Green C.J., Vold B.S.;
RT   "A cluster of nine tRNA genes between ribosomal gene operons in
RT   Bacillus subtilis";
RL   J. Bacteriol. 174:3147-3151(1992).
XX
CC   NCBI gi: 143731
XX
FH   Key             Location/Qualifiers
FH
FT   source          1..1120
FT                   /organism="Bacillus subtilis"
FT                   /strain="W168"
FT                   /sequenced_mol="DNA"
FT   rRNA            <1..72
FT                   /gene="5S rRNA"
FT                   /product="5S ribosomal RNA"
FT   tRNA            75..250
FT                   /gene="tRNA-Thr"
FT                   /product="transfer RNA-Thr"
FT   tRNA            95..170
FT                   /gene="tRNA-Val"
FT                   /product="transfer RNA-Val"
FT   promoter        261..266
FT   tRNA            287..362
FT                   /gene="tRNA-Lys"
FT                   /product="transfer RNA-Lys"
FT   tRNA            369..451
FT                   /gene="tRNA-Leu"
FT                   /product="transfer RNA-Leu"
FT   tRNA            492..566
FT                   /gene="tRNA-Gly"
FT                   /product="transfer RNA-Gly"
FT   tRNA            579..666
FT                   /gene="tRNA-Leu"
FT                   /product="transfer RNA-Leu"
FT   tRNA            676..752
FT                   /gene="tRNA-Arg"
FT                   /product="transfer RNA-Arg"
FT   tRNA            779..856
FT                   /gene="tRNA-Pro"
FT                   /product="transfer RNA-Pro"
FT   tRNA            866..941
FT                   /gene="tRNA-Ala"
FT                   /product="transfer RNA-Ala"
FT   promoter        984..989
FT   rRNA            1113..>1120
FT                   /gene="16S rRNA"
FT                   /product="16S ribosomal RNA"
```

(a) The heading part, where Thr (75-250) and Val (95-170) are described.

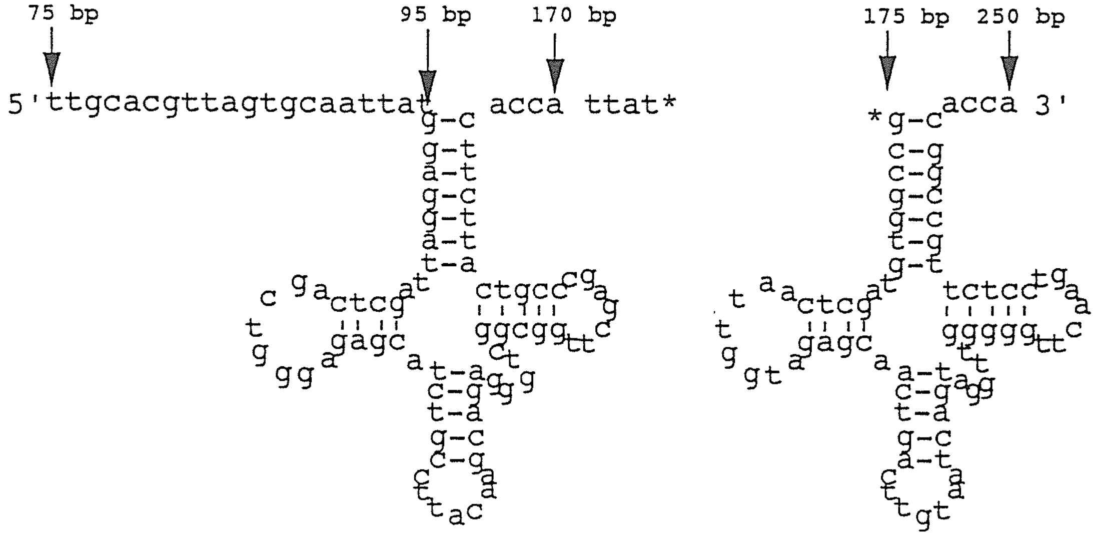
Figure 5. The 1120-bp DNA fragment of B. *subtilis* obtained through DBGET integrated database retrieval system, GenomeNet (Kyoto center).

(Figure 5 . Continued)

```
XX
SQ   Sequence 1120 BP; 278 A; 252 C; 296 G; 294 T; 0 other;
     cacggaagtt aagctcttca gcgccgatgg tagtcggggg tttcccctg tgagagtagg        60
     acgccgccaa gcaattgcac gttagtgcaa ttatggagga ttagctcagc tgggagagca       120
     tctgccttac aagcagaggg tcggcggttc gagcccgtca tcctccacca ttatgccggt       180
     gtagctcaat tggtagagca actgacttgt aatcagtagg ttgggggttc aagtcctctt       240
     gccggcacca cttttatatg atataatact caagtctctt gtagaagagc cattagctca       300
     gttggtagag catctgactt ttaatcagag ggtcgaaggt tcgagtcctt catggctcac       360
     cattttacgc gggtgtggcg gaattggcag acgcgctaga cttaggatct agtgtcttta       420
     tgacgtgggg gttcaagtcc cttcacccgc gttatatacc ataacagtta gaaaaactgg       480
     acatcctgtc tgcggaagta gttcagtggt tgaacaccac cttgccaagg tgggggtcgc       540
     gggttcgaat cccgtcttcc gctccaacta taccatccac gccggggtgg tggaattggc       600
     agacacacag gacttaaaat cctgcggtag gtgactaccg tgccggttca agtccggccc       660
     tcggcattat gttttgcgcc cgtagctcaa ttggatagag cgtttgacta cggatcaaaa       720
     ggttaggggt tcgactcctc tcgggcgcgc catatctttt aatagaatag ataggaaatc       780
     gggaagtagc tcagcttggt agagcacatg gtttgggacc atggggtcgc aggttcgaat       840
     cctgtcttcc cgaccatttt tttatggggc cttagctcag ctgggagagc gcctgctttg       900
     cacgcaggag gtcagcggtt cgatcccgct aggctccacc aaaagttttt aaaaaagttg       960
     ttgactttga agaagtgacg ttgtatacta ataaagttgc tttaacaaag cggacaaaca      1020
     aaatgatctt tgaaaactaa acaagacaaa acgtacctgt taattcagtt tttaaaaacg      1080
     ctacagcgat gtgcgtagtc agtcaaacta ctttatcgga                            1120
//
```

DBGET integrated database retrieval system, GenomeNet (Kyoto center)

(b) The DNA sequence.



(c) The cloverleaf structures of Thr (75-250) and Val (95-170), which is included in the Thr.

5. Discussion

Our examples may be too few to be used for reliable deductive inference. Nevertheless, the few examples appear to demonstrate that this approach is useful as a means of finding some clue for the detection of tRNA gene clusters. We hope that this suggestion is supported with much more cases. When we began this study, we found this rule of thumb on the relation of bacterial categories and tRNA gene clusters [3], and correctly predicted classes, which were unknown for us at that time, of three Gram-nagative bacteria *Rhodebacter spaheroides* (N-rhosp), *Thermotoga maritina* (N-thema), and *Thermas thermophilus* (N-theth). Addtion of information on codon/ANTICODON [4] enhanced the resolution with few exceptions. For example, Ser-Leu of Archaebacterium was decomposed into SagcLcug from *Halobacterium marismortui* (A-halma) and SucgLcug from *Salfolobus solfataricus* (A-sulso). But at the same time, this kind of enhancement of the resolution resulted in advantageous division of Leu-Leu into LcugLcug from *E. coli* (N-escco) and LuuaLuug from *B. subtilis* (P-bacsu) [5]. Characteristic analysis of tRNA and rRNA gene clusters is on going from a viewpoint of phylogenetic relationship. This concept also can lead to the chance detection of erroneous operations as exemplified with the description of Thr (75 to 250) -Val (90 to 170). The gene cluster list keeps many entries whose headwords are [VT] (i. e., VguaTaca), but there is no entry of Thr-Val because we revised even this only one entry itself to [VT].

According to e-mail from the DDBJ (kfukami@genes.nig.ac.jp) dated on 22 Oct 1995, they informed GenBank and EMBL of our questions on the facts described here and some other topics which were sent to the DDBJ through e-mail dated on 10 Oct 1995. According to GenomeNet (Kyoto center), the description of the Thr (Locus BACTG9168, Accession No. M87888) has been updated on 060-FEB-1997, and the data of the Thr has been revised from (75 to 250) to (175 to 250) as described in the original paper [6] without any comment.

References

[1] Patric Keeling, Robert L. Cherlebois, and W. Ford Doolittle : Current Opinion in Genetics and Development, 4, 816-822 (1994).

[2] (a) M. J. Rogers, A. A. Steinmetz and R. T. Walker : Nucleic Acids Res., 14, 3145 (1986).

(b) Akira Muto, Yoshiki Andachi, Harumi Yuzawa, Fumiaki Yamao, and Shozo Osawa : Nucleic Acids Res. 18, 5037 (1990).

(c) Reiji Tanaka, Yoshiki Andachi, and Akira Muto : Nucleic Acids Res. 19, 6787-6792 (1991).

[3] KUDO Yoshihiro, KANAYA Shigehiko, and KONNO Tomiharu: Bull. Yamagata Univ. (Eng.), 23, 113-126 (1995).

[4] KUDO Yoshihiro, and KANAYA Shigehiko : Proceedings Genome Informatics Workshop 1995, 90-91 (1995).

[5] Yoshihiro KUDO and Shigehiko KANAYA : 15th Intl. CODATA Conf., 203-204; Supplement, 13-18 (1996).

[6] Christopher J. Green, and Barbara S. Vold: J. Bacteriol. 174, 3147-3151 (1992).