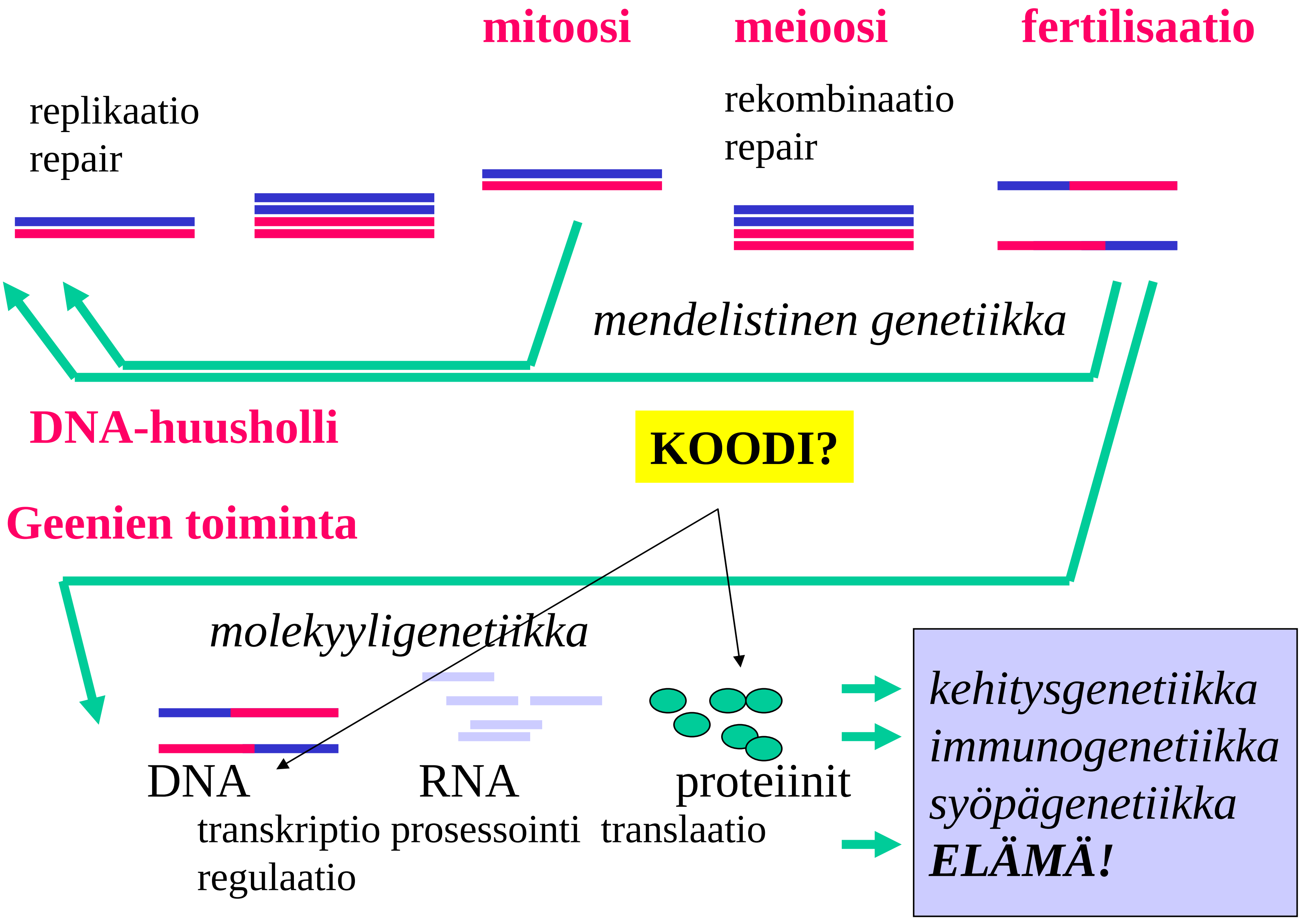


Genetiikan perusteiden miellekartta: geneettinen triplettikoodi

Translaatiossa mRNA:ssa tarvitaan seuraavat *sanomat* :

m^7G - pipo

Ehkä internal ribosome entry site **IRES**

poly-A -häntä sirkulaarisen polysomin muodostamiseksi

AUG aloituskodoni ensimmäistä metioniinia varten

UAA, **UAG** tai **UGA** lopetuskodoni RF:ää varten

Sekä ennen AUG-kodonia että STOP-kodonin jälkeen voi olla säätelyyn osallistuvia **UTR-** jaksoja (untranslated region)

UTR koodi **UTR**

m^7G ---AUG----(IRES**) -----AUG -----...-----UAA------p(AAA)**

Translaatiossa mRNA:ssa tarvitaan seuraavat *sanomat* :

m^7G - pipo

Ehkä internal ribosome entry site **IRES (**CELL 448)

poly-A -häntä sirkulaarisen polysomin muodostamiseksi

AUG aloituskodoni ensimmäistä metioniinia varten

UAA, **UAG** tai **UGA** lopetuskodoni RF:ää varten

Sekä ennen AUG-kodonia että STOP-kodonin jälkeen voi olla säätelyyn osallistuvia **UTR-** jaksoja (untranslated region)

UTR koodi **UTR**

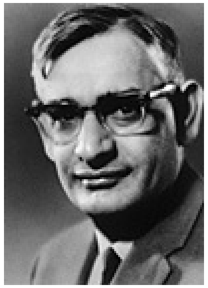
m^7G ---AUG----(IRES**) -----AUG -----...-----UAA------p(AAA)**

Tässä tulikin jo pari triplettikoodin sanaa

Aloitus (= metioniini) **AUG**

STOP: **UAA**, **UGA** tai **UAG**

Koodin murtaminen oli molekyyligenetiikan historian suuria juttuja: nobelistit Holley (tRNA), Khorana ja Nirenberg (**1968**)



CELL 335-

The Genetic Code

1st position (5′ end)	2nd Position U	C	A	G	3rd Position (3′ end)
U	Phe	Ser	Tyr	Cys	U
	Phe	Ser	Tyr	Cys	C
	Leu	Ser	**STOP**	**STOP**	A
	Leu	Ser	**STOP**	Trp	G
C	Leu	Pro	His	Arg	U
	Leu	Pro	His	Arg	C
	Leu	Pro	Gln	Arg	A
	Leu	Pro	Gln	Arg	G
A	Ile	Thr	Asn	Ser	U
	Ile	Thr	Asn	Ser	C
	Ile	Thr	Lys	Arg	A
	Met	Thr	Lys	Arg	G
G	Val	Ala	Asp	Gly	U
	Val	Ala	Asp	Gly	C
	Val	Ala	Glu	Gly	A
	Val	Ala	Glu	Gly	G

Amino Acids Codons

A	Ala	Alanine	GCA GCC GCG GCU
C	Cys	Cysteine	UGC UGU
D	Asp	Aspartic acid	GAC GAU
E	Glu	Glutamic acid	GAA GAG
F	Phe	Phenylalanine	UUC UUU
G	Gly	Glycine	GGA GGC GGG GGU
H	His	Histidine	CAC CAU
I	Ile	Isoleucine	AUA AUC AUU
K	Lys	Lysine	AAA AAG
L	Leu	Leucine	UUA UUG CUA CUC CUG CUU
M	Met	Methionine	AUG
N	Asn	Asparagine	AAC AAU
P	Pro	Proline	CCA CCC CCG CCU
Q	Gln	Glutamine	CAA CAG
R	Arg	Arginine	AGA AGG CGA CGC CGG CGU
S	Ser	Serine	AGC AGU UCA UCC UCG UCU
T	Thr	Threonine	ACA ACC ACG ACU
V	Val	Valine	GUA GUC GUG GUU
W	Trp	Tryptophan	UGG
Y	Tyr	Tyrosine	UAC UAU

DNA words are three letters long. - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Search Favorites Media

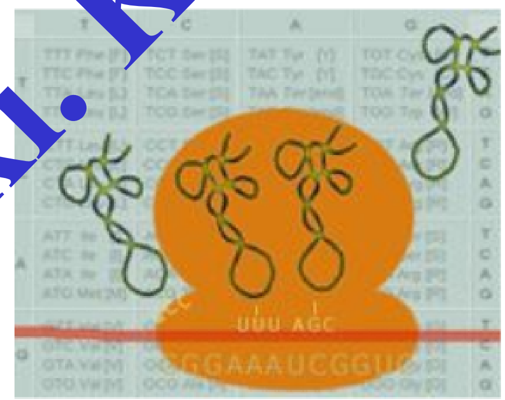
Address http://www.dnaftb.org/dnaftb/22/concept/index.html Go Links

DNA words are three letters long.

22

DNA FROM THE BEGINNING

The genetic code had to be a "language" – using the DNA alphabet of A, T, C, and G – that produced enough DNA "words" to specify each of the 20 known amino acids. Simple math showed that only 16 words are possible from a two-letter combination, but a three-letter code produces 64 words. Operating on the principle that the simplest solution is often correct, researchers assumed a three-letter code called a codon.



Research teams at University of British Columbia and the National Institutes of Health laboriously synthesized different RNA molecules, each a long strand composed of a single repeated codon. Then, each type of synthetic RNA was added to a cell-free translation system containing ribosomes, transfer RNAs, and amino acids. As predicted, each type of synthetic RNA produced a polypeptide chain composed of repeated units of a single amino acid. Several codons are "stop" signals and many amino acids are specified by several different codons, accounting for all 64 three-letter combinations.

CONCEPT | ANIMATION | GALLERY | AUDIO/VIDEO | BIO | PROBLEM | LINKS

Done Internet

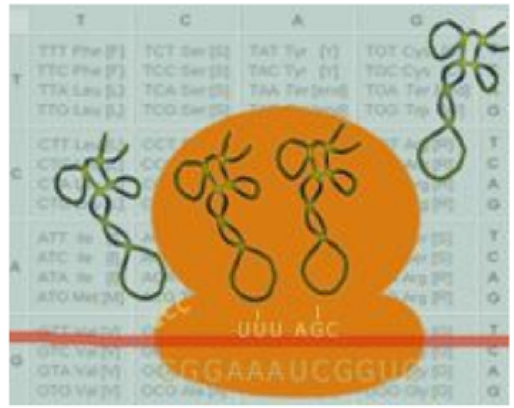
Kuva on linkki. Klikkaa!

DNA words are three letters long.

22

DNA FROM THE BEGINNING

The genetic code had to be a "language" – using the DNA alphabet of A, T, C, and G – that produced enough DNA "words" to specify each of the 20 known amino acids. Simple math showed that only 16 words are possible from a two-letter combination, but a three-letter code produces 64 words. Operating on the principle that the simplest solution is often correct, researchers assumed a three-letter code called a codon.



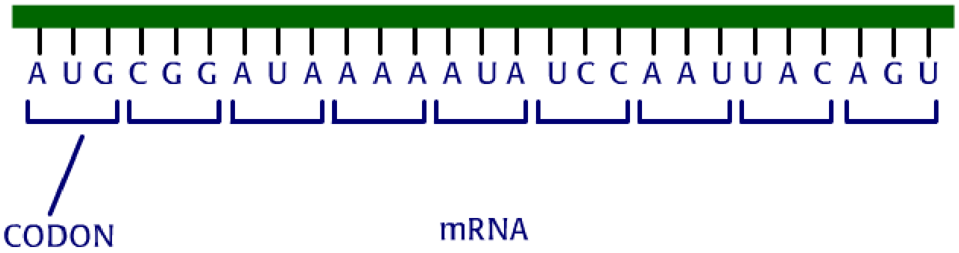
Research teams at University of British Columbia and the National Institutes of Health laboriously synthesized different RNA molecules, each a long strand composed of a single repeated codon. Then, each type of synthetic RNA was added to a cell-free translation system containing ribosomes, transfer RNAs, and amino acids. As predicted, each type of synthetic RNA produced a polypeptide chain composed of repeated units of a single amino acid. Several codons are "stop" signals and many amino acids are specified by several different codons, accounting for all 64 three-letter combinations.

DNA words are three letters long.

22

DNA FROM THE BEGINNING

I'm Marshall Nirenberg. Har Khorana's group and mine "cracked" the genetic code. We figured out how the nucleotide language of mRNA is "translated" into the amino acid language of proteins. Genetic data from Crick and others showed that three nucleotides form a "codon" – an mRNA word that specifies one amino acid.



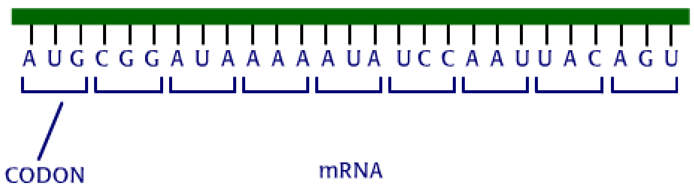
CONCEPT ANIMATION GALLERY AUDIO/VIDEO BIO PROBLEM LINKS

DNA words are three letters long.

22

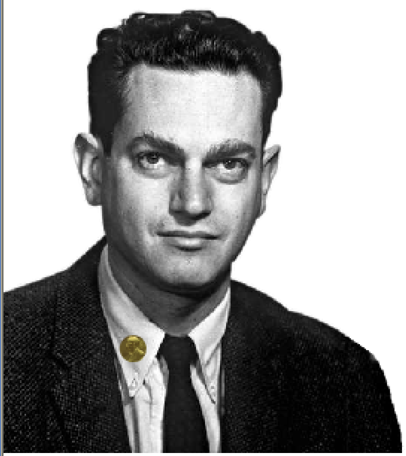
DNA FROM THE BEGINNING

This made sense, because a codon made from only one or two nucleotides would not produce enough combinations (words) to code for all 20 of the known amino acids.



1 NUCLEOTIDE = 4 POSSIBLE CODONS

2 NUCLEOTIDES = 4 X 4 POSSIBLE CODONS

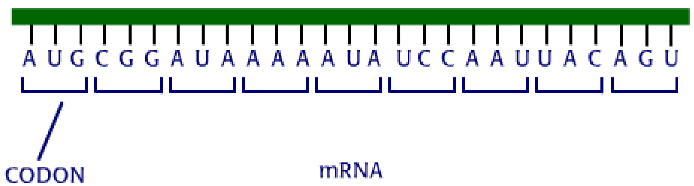


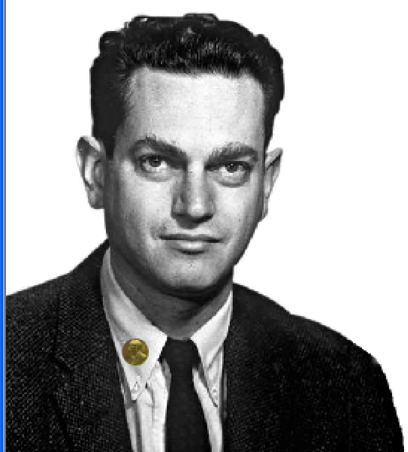
DNA words are three letters long.

22

DNA FROM THE BEGINNING

However, a three-nucleotide codon produces 64 combinations. This would produce a redundant, or degenerate, code where several different codons specify the same amino acid. The parsimony principle – that the simplest solution is often right – ruled out a four-nucleotide codon.





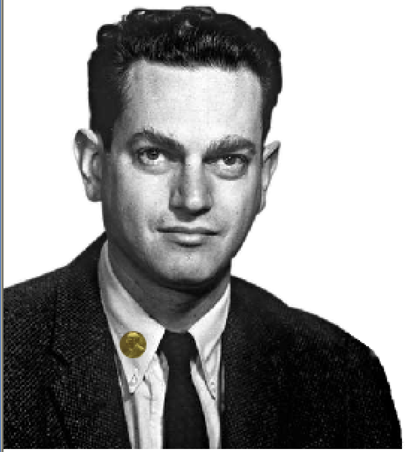
3 NUCLEOTIDES = 4 X 4 X 4 POSSIBLE CODONS
3 NUCLEOTIDES = 64 codons for 20 amino acids

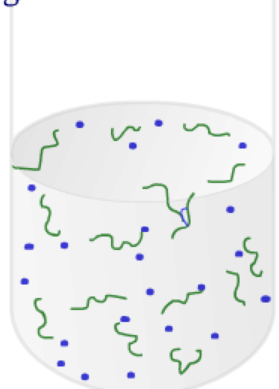
DNA words are three letters long.

22

DNA FROM THE BEGINNING

In 1961, Heinrich Matthaei, a visiting post-doc, and I started experiments to test the triplet codon hypothesis. We used a cell-free extract of *E. coli* because we believed that this extract should contain all the components needed to translate mRNA into proteins. The extract was treated with DNase to destroy any remaining *E. coli* DNA – so there was no template from which messenger RNA could be made.





CELL-FREE EXTRACT

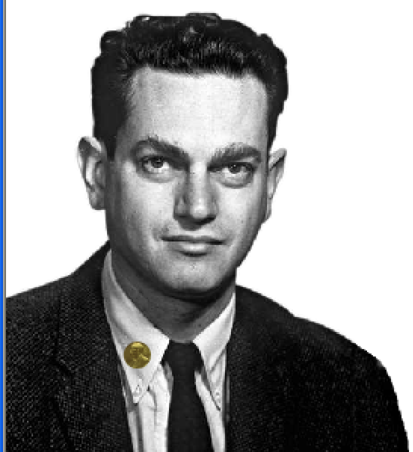
CONCEPT ANIMATION GALLERY AUDIO/VIDEO BIO PROBLEM LINKS

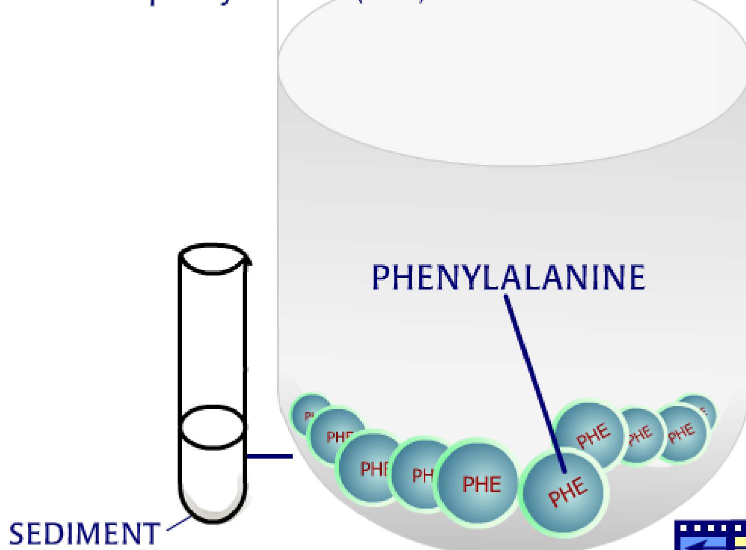
DNA words are three letters long.

22

DNA FROM THE BEGINNING

When we examined the products produced in this cell-free system, we found polypeptides composed entirely of the amino acid phenylalanine (PHE).



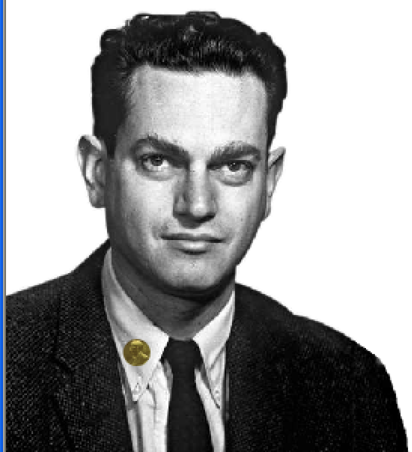


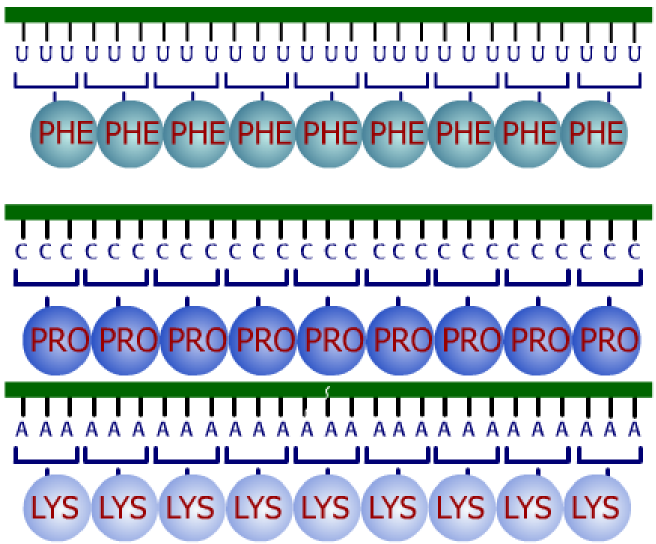
DNA words are three letters long.

22

DNA FROM THE BEGINNING

Matthaei and I tried other polynucleotide chains. Poly-C made a proline (PRO) chain; poly-A made a lysine (LYS) chain. Interestingly, no protein was made with the poly-G chain.





CONCEPT ANIMATION GALLERY AUDIO/VIDEO BIO PROBLEM LINKS

NA words are three etters long.

22

DNA FROM THE BEGINNING

her researchers and I quickly recognized the power of this approach. Using RNA nplates containing different nucleotide combinations, we assigned amino acids to out 50 triplet codons.

SECOND LETTER

FIRST LETTER	U	C	A	G	THIRD LETTER
U	U U U U U C } PHE U U A U U G } LEU	U C U U C C U C G } SER ?	U A U U A C } TYR ? ?	U G U – CYS ? ? U G G – TRP	U C A G
C	C U U C U C } LEU ? ?	C C U C C C C C A C C G } PRO	C A U C A C } HIS C A A – GIN ?	C G U C G C C G A } ARG ?	U C A G
A	A U U A U C } IIE A U A A U G } MET	A C U A C C A C A } THR ?	A A U A A C } ASN A A A – LYS ?	? A G C – SER A G A – ARG ?	U C A G
G	G U U G U A } VAL ? ?	G C U G C C G C A } ALA ?	G A U G A C } ASP G A A – GLU ?	G G U G G C G G A } GLY ?	U C A G

CONCEPT ANIMATION GALLERY AUDIO/VIDEO BIO PROBLEM LINKS

DNA words are three letters long. - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Search Favorites Media

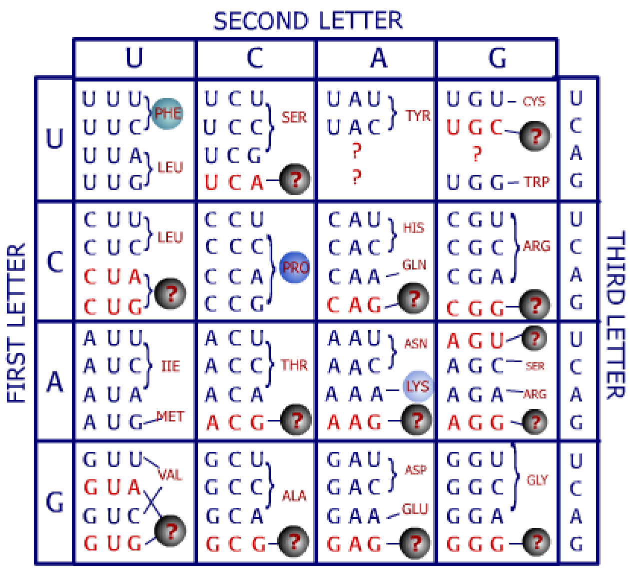
Address http://www.dnaftb.org/dnaftb/22/concept/index.html Go Links

DNA words are three letters long.

22

DNA FROM THE BEGINNING

Hi, I'm Phil Leder. I helped Marshall Nirenberg with the rest of the genetic code.



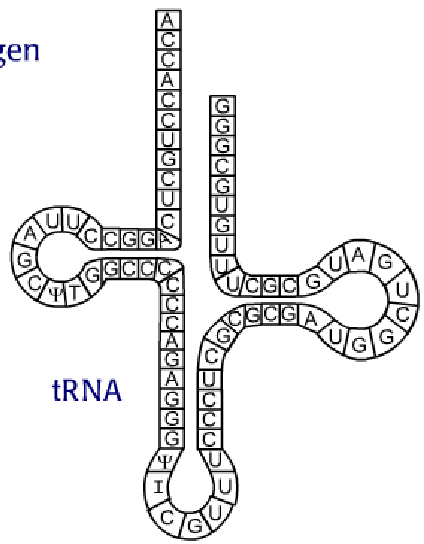
CONCEPT ANIMATION GALLERY AUDIO/VIDEO BIO PROBLEM LINKS

DNA words are three letters long.

22

DNA FROM THE BEGINNING

tRNA is the molecule that carries amino acids to the ribosomes for protein synthesis. In 1962, Robert Holley solved the structure of tRNA. Although tRNA is single stranded, stretches of complementary nucleotides hydrogen bond to form short double-stranded regions, which bend the tRNA into a characteristic cloverleaf shape.

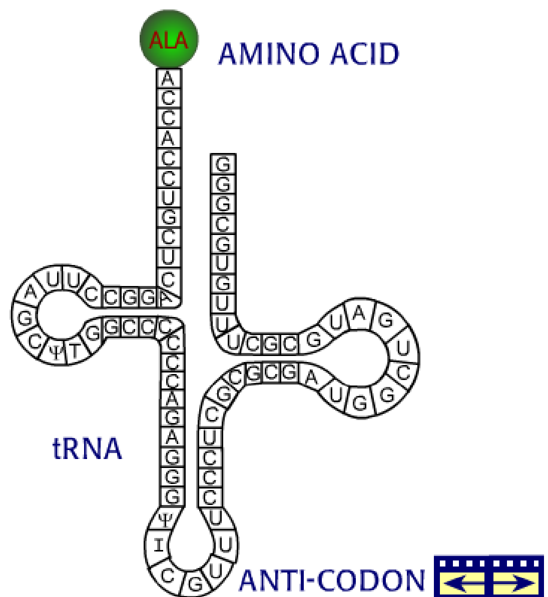


DNA words are three letters long.

22

DNA FROM THE BEGINNING

Using their cell-free translation system, Zamecnik and his group showed that tRNA becomes activated when the amino acid binds to the tRNA's stem. This step requires chemical energy in the form of ATP.



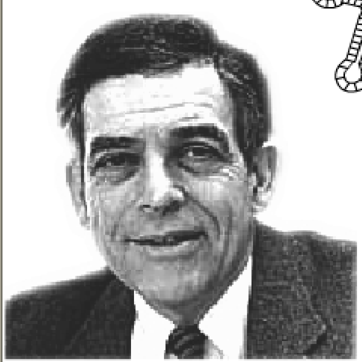
CONCEPT ANIMATION GALLERY AUDIO/VIDEO BIO PROBLEM LINKS

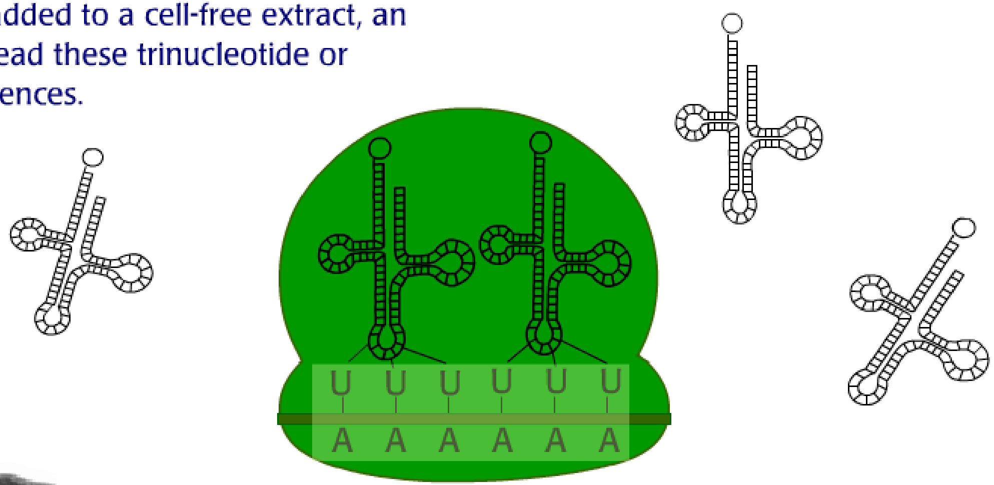
DNA words are three letters long.

22

DNA FROM THE BEGINNING

I first made short RNA chains composed of three or six nucleotides. When added to a cell-free extract, an activated tRNA will read these trinucleotide or hexanucleotide sequences.





HEXANUCLEOTIDE SEQUENCE

DNA words are three letters long.

22

DNA FROM THE BEGINNING

After that, all I had to do was make specific trinucleotide and hexanucleotide chains to confirm the genetic code.

SECOND LETTER

FIRST LETTER	U	C	A	G	THIRD LETTER
U	UUU, UUC } PHE; UUA, UUG } LEU	UCU, UCC, UCG, UCA } SER	UAU, UAC } TYR	UGU, UGC } CYS; UGG – TRP	U C A G
C	CUU, CUC, CUA, CUG } LEU	CCU, CCC, CCA, CCG } PRO	CAU, CAC } HIS; CAA, CAG } GLN	CGU, CGC, CGA, CGG } ARG	U C A G
A	AUU, AUC, AUA } ILE; AUG – MET	ACU, ACC, ACA, ACG } THR	AAU, AAC } ASN; AAA, AAG } LYS	AGU, AGC } SER; AGA, AGG } ARG	U C A G
G	GUU, GUA, GUC, GUG } VAL	GCU, GCC, GCA, GCG } ALA	GAU, GAC } ASP; GAA, GAG } GLU	GGU, GGC, GGA, GGG } GLY	U C A G

GENETIC CODE TABLE

CONCEPT ANIMATION GALLERY AUDIO/VIDEO BIO PROBLEM LINKS

DNA words are three letters long.

22

DNA FROM THE BEGINNING

We also found that translation starts with a specific codon AUG – the only unique codon – and there are stop codons that end translation. So, in addition to words, the genetic code also has punctuation points.

SECOND LETTER

FIRST LETTER	U	C	A	G	THIRD LETTER
U	UUU PHE UUC PHE UUA LEU UUG LEU	UCU SER UCC SER UCG SER UCA SER	UAU TYR UAC TYR UAA STOP UAG STOP	UGU CYS UGC CYS UGA STOP UGG TRP	U C A G
C	CUU LEU CUC LEU CUA LEU CUG LEU	CCU PRO CCC PRO CCA PRO CCG PRO	CAU HIS CAC HIS CAA GLN CAG GLN	CGU ARG CGC ARG CGA ARG CGG ARG	U C A G
A	AUU ILE AUC ILE AUA ILE AUG MET	ACU THR ACC THR ACA THR ACG THR	AAU ASN AAC ASN AAA LYS AAG LYS	AGU SER AGC SER AGA ARG AGG ARG	U C A G
G	GUU VAL GUA VAL GUC VAL GUG VAL	GCU ALA GCC ALA GCA ALA GCG ALA	GAU ASP GAC ASP GAA GLU GAG GLU	GGU GLY GGC GLY GGA GLY GGG GLY	U C A G

GENETIC CODE TABLE
(Mouseover table to enlarge)

CONCEPT ANIMATION GALLERY AUDIO/VIDEO BIO PROBLEM LINKS

Geneettinen koodi ja avoimet lukukehykset: polypeptidiä koodaavan geenin “ennustamisen” päävälineet

On tapana kirjoittaa ja tallettaa geenipankkiin vain toinen juoste, alkaen 5’ päästä, esimerkiksi näin:

`5′-TTATTTTATTTCGATAATTCGACCTTAAACGCGAAACTTCACTTAAC-3′`

Geenin viesti voi kumminkin olla kummassa tahansa juosteessa, ja sen jako tripletteihin voi alkaa molemmissa kolmesta eri paikasta

Tässä voisi siis olla peräti kuusi erilaista lukutapaa, joista vain yksi lienee oikea (aloituskodonia ei etsitä, vaan pelkästään stoppeja)

5′-TTATTTTATTTCGATAATTCGACCTTAAACGCGAAACTTCACTTAAC-3′
3′-AATAAAATAAAGCTATTAAGCTGGAATTTGCGCTTTGAAGTGAATTG-5′

Ensimmäinen *kodoni* voi siis olla

>>> UUA, UAU tai AUU >> TEMPLAATTINA ALEMPI JUOSTE

tai

<<<GUU, UUA tai UAA <<< TEMPLAATTINA YLEMPI JUOSTE

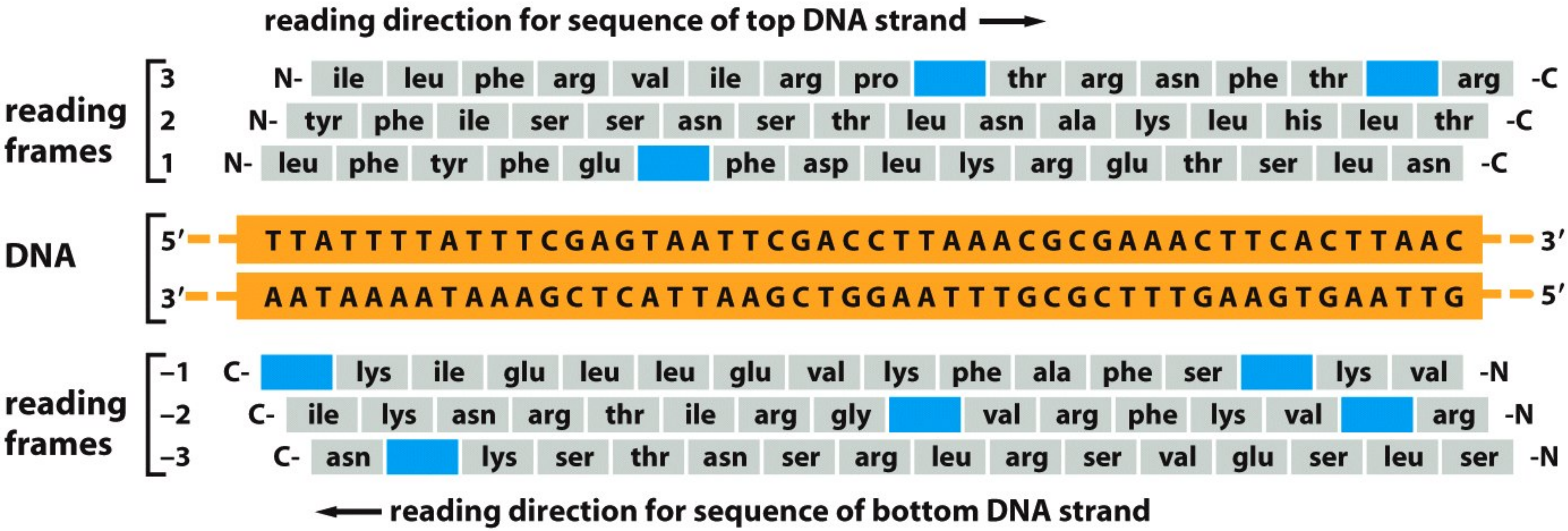


Figure 8-52a Molecular Biology of the Cell 5/e (© Garland Science 2008)

STOP on merkitty sinisellä laatikolla (UAA, UAG, UGA).

Avain lukemiseen:
leu UUA, UUG, CUU, CUC, CUA, CUG

Mikä siis on geeni? Alempi vai ylempi juoste?

CELL 550

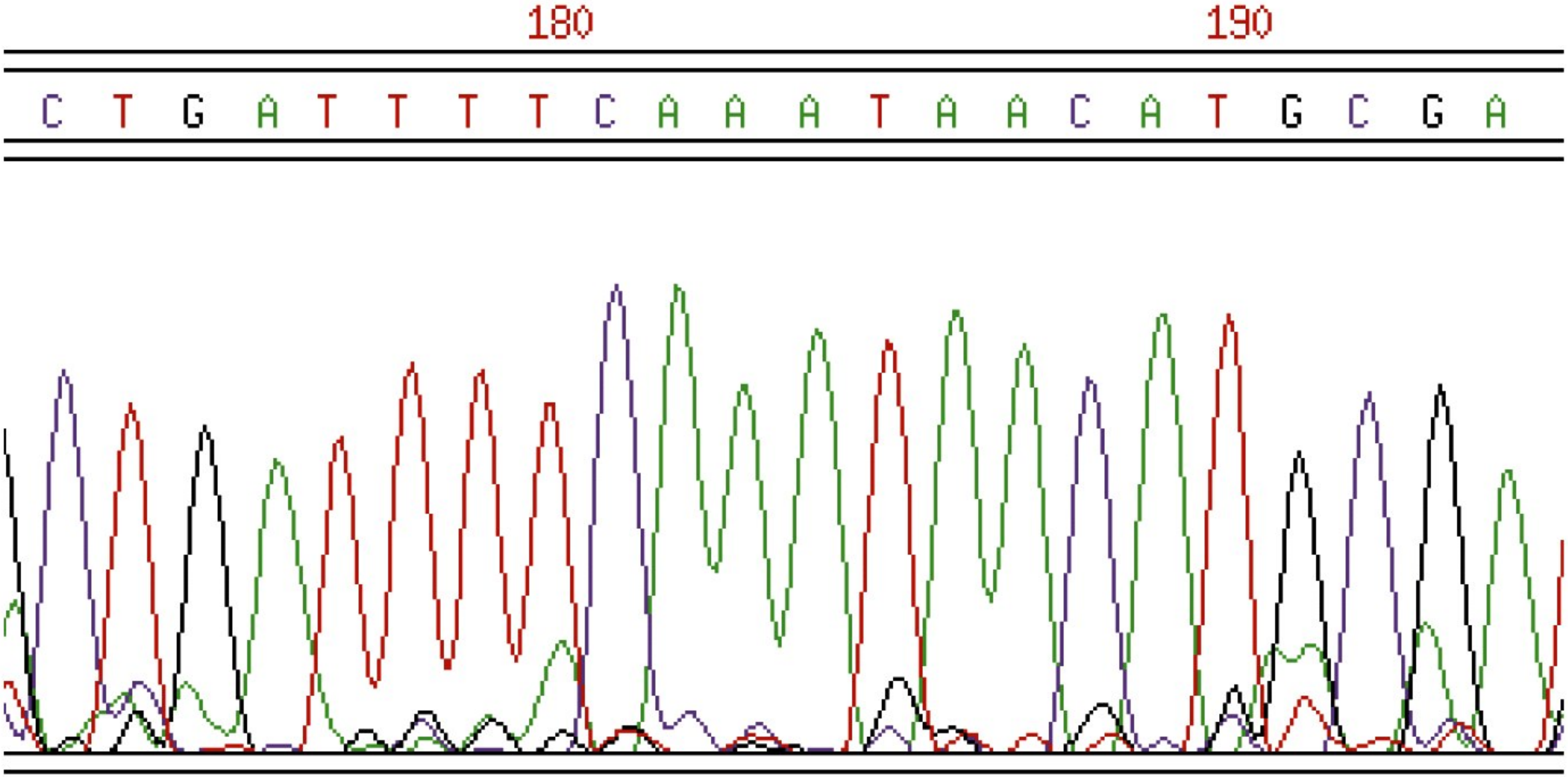


Figure 8-51 Molecular Biology of the Cell 5/e (© Garland Science 2008)

CELL 550

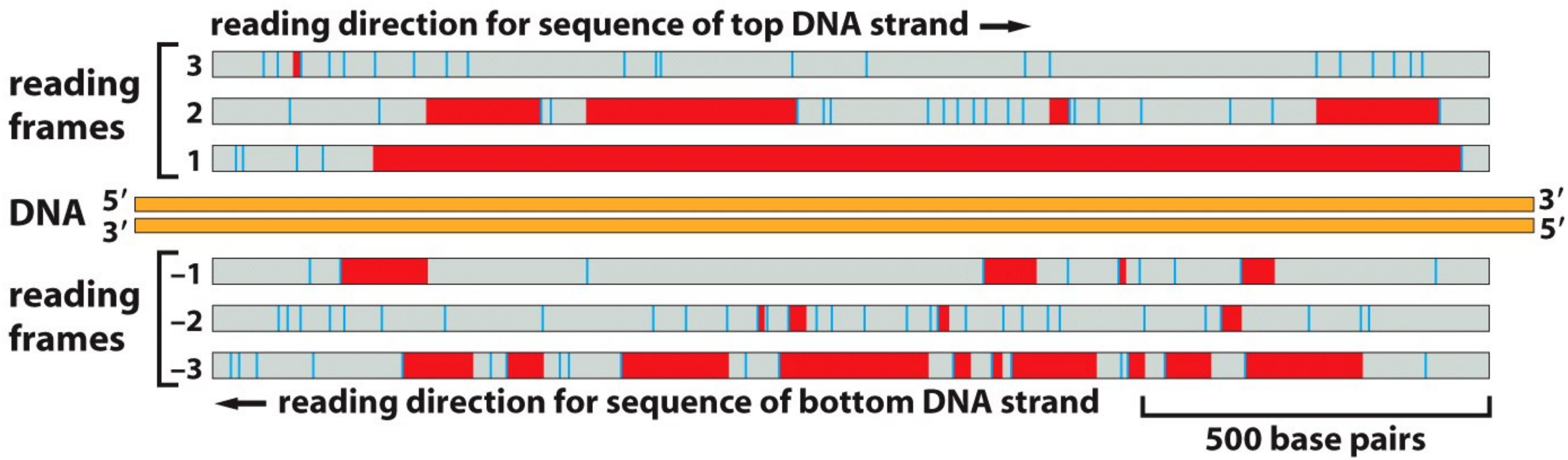


Figure 8-52b Molecular Biology of the Cell 5/e (© Garland Science 2008)

Tässä esimerkissä on tarkasteltu 1700 nukleotidia. STOP on merkitty vihreällä viivalla, ja mahdolliset START-STOP -välit on väritetty punaisella.

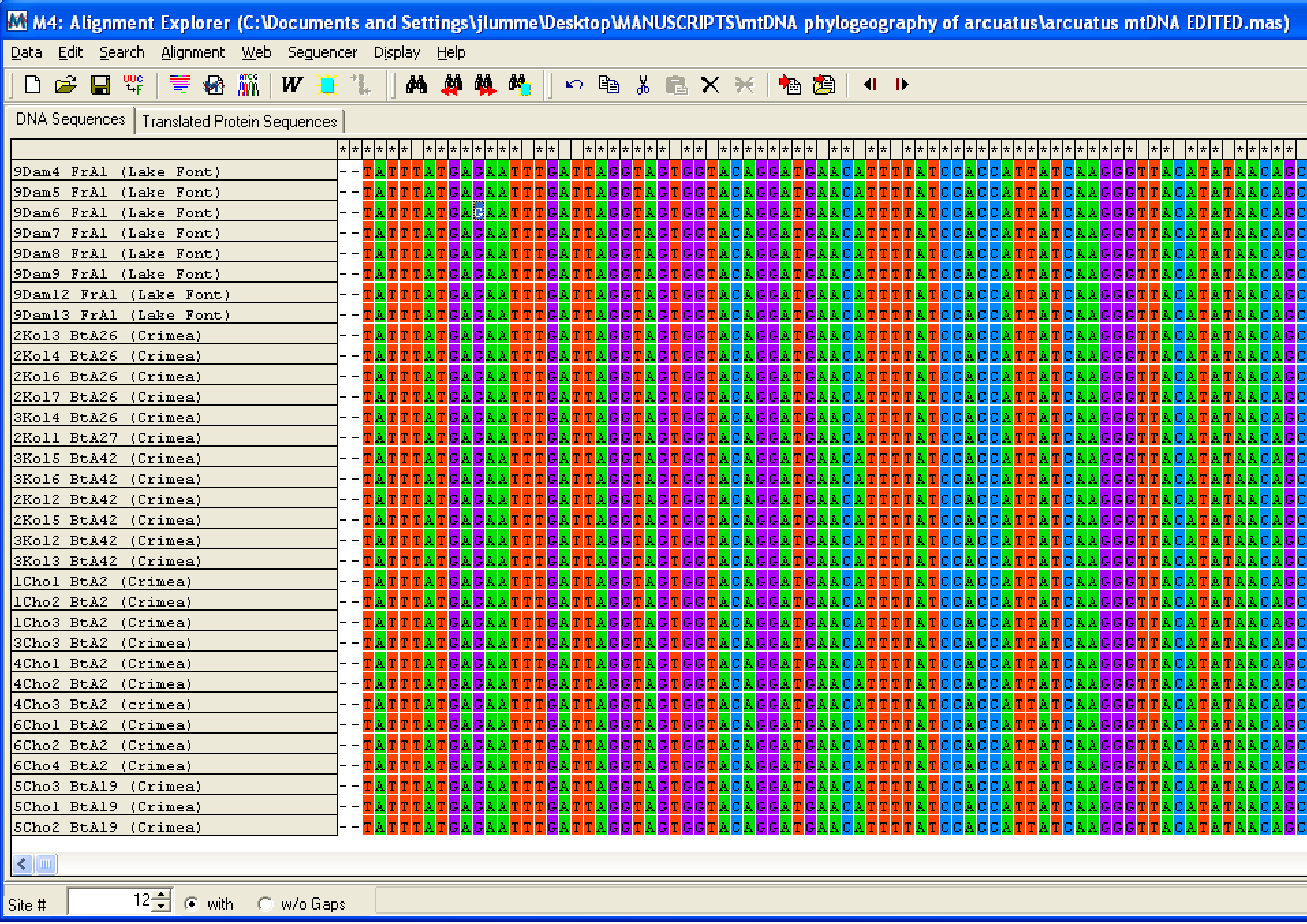
Ainoastaan FRAME 1 näyttää sisältävän pitkän jakson, joka ***voi*** olla polypeptidiketjun rakennusohjetta.

Tätä sanotaan **ORF**: iksi (open reading frame) ja se on geenin ennustamisen ensimmäinen askel.

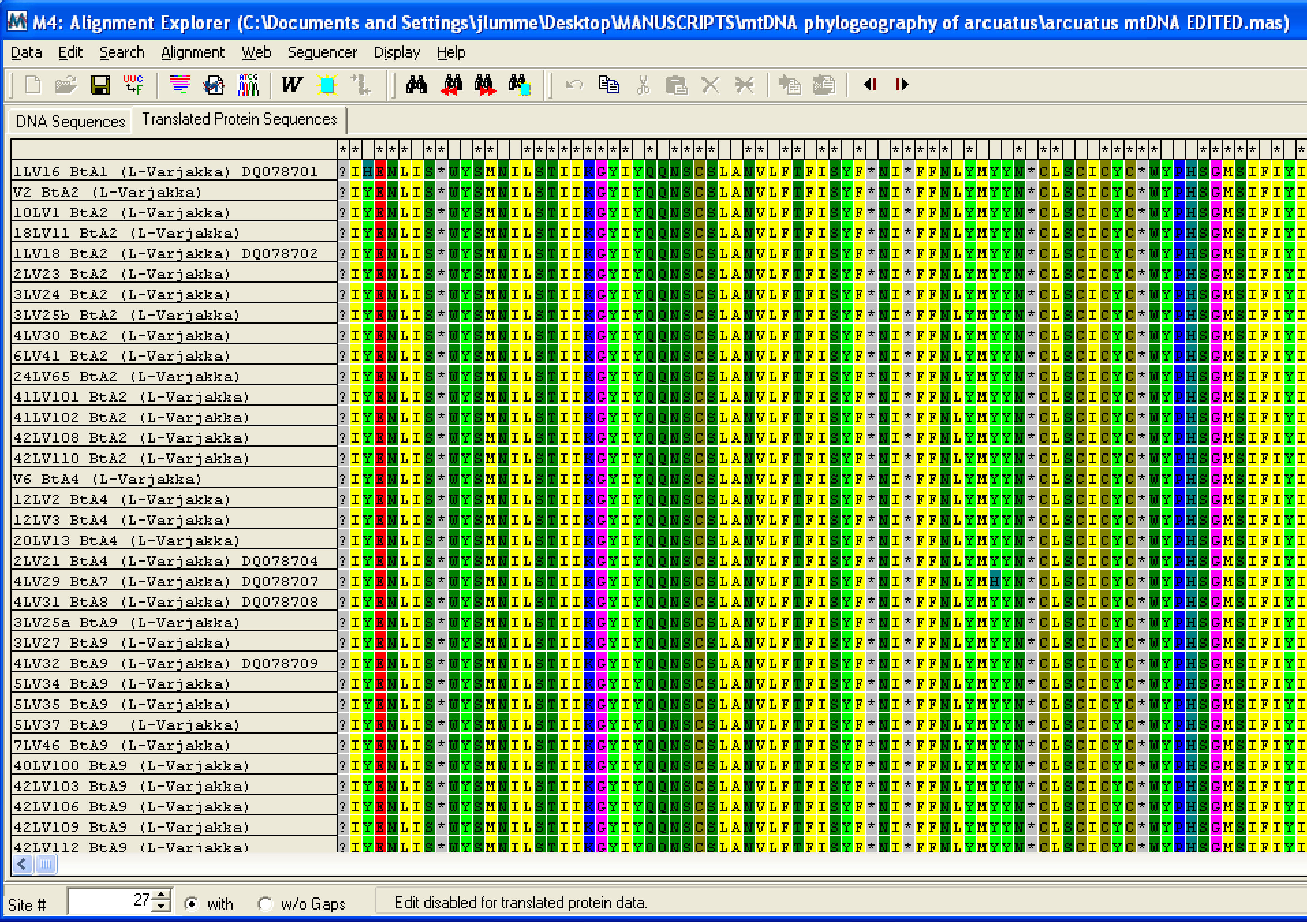
CELL 550

ORF:in ennustaminen jos matskua on vähän enemmän

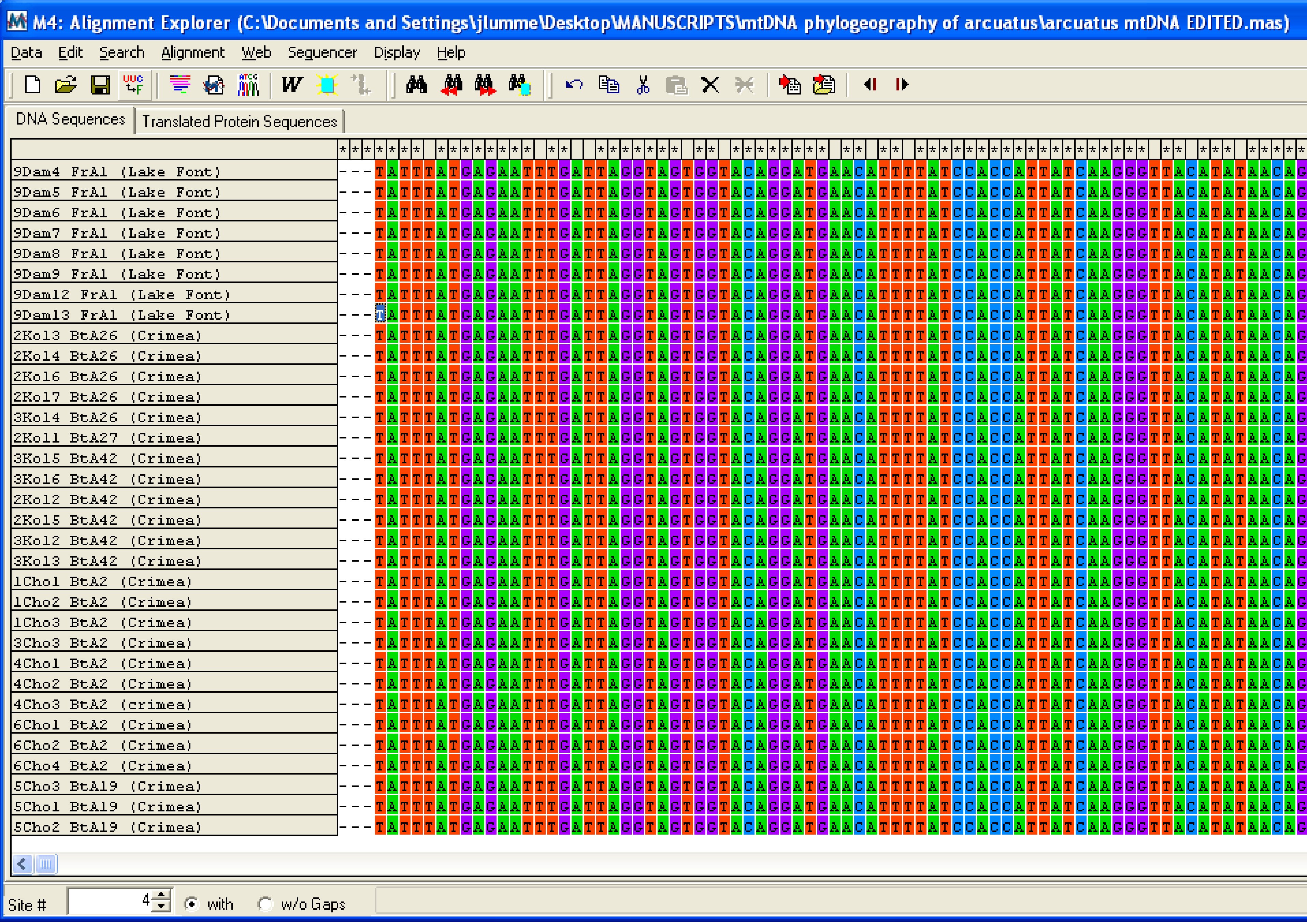
Gyrodactylus arcuatuksen **mtDNA (CO1)**



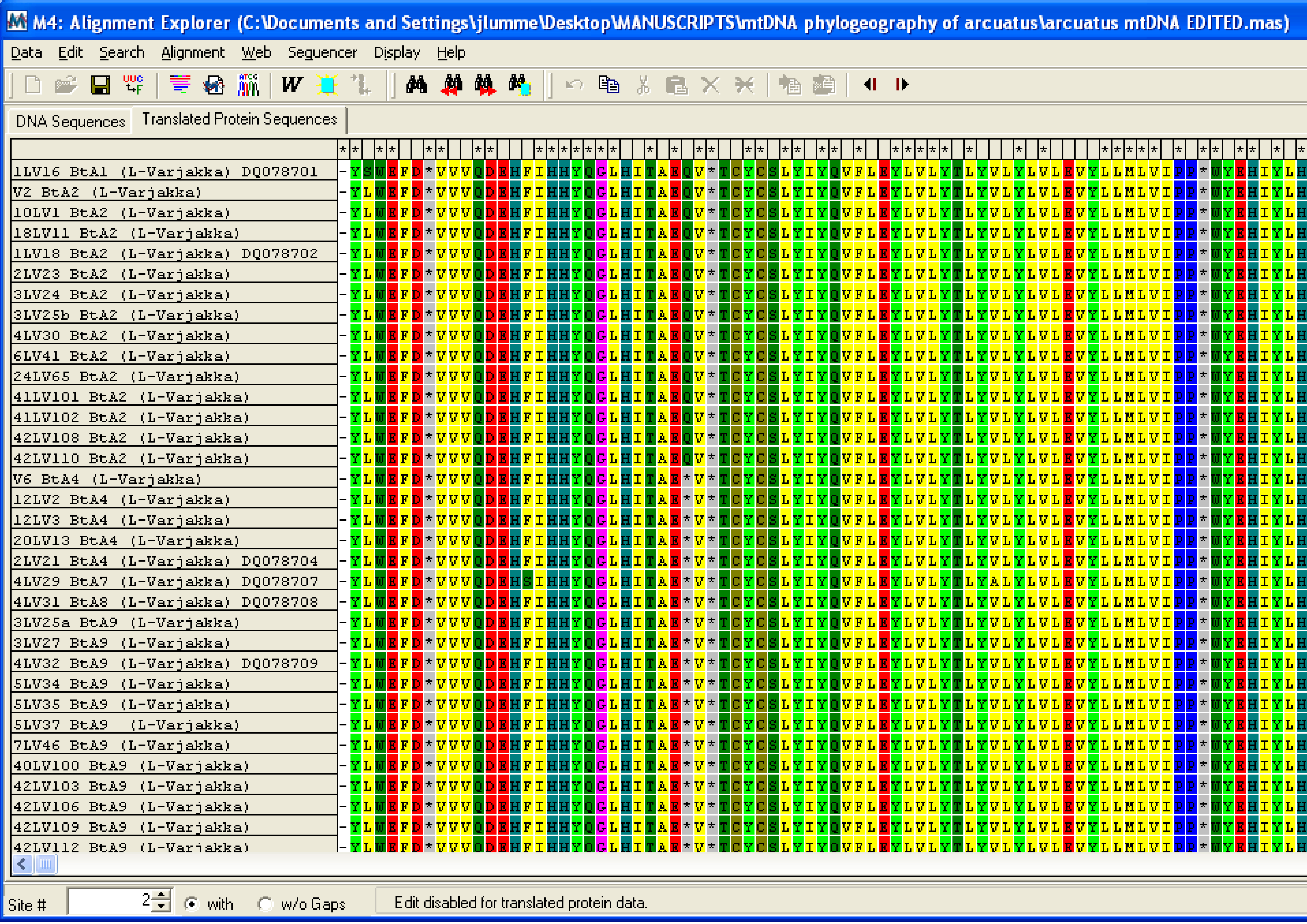
Ensimmäinen TÄYSI kodoni ATT



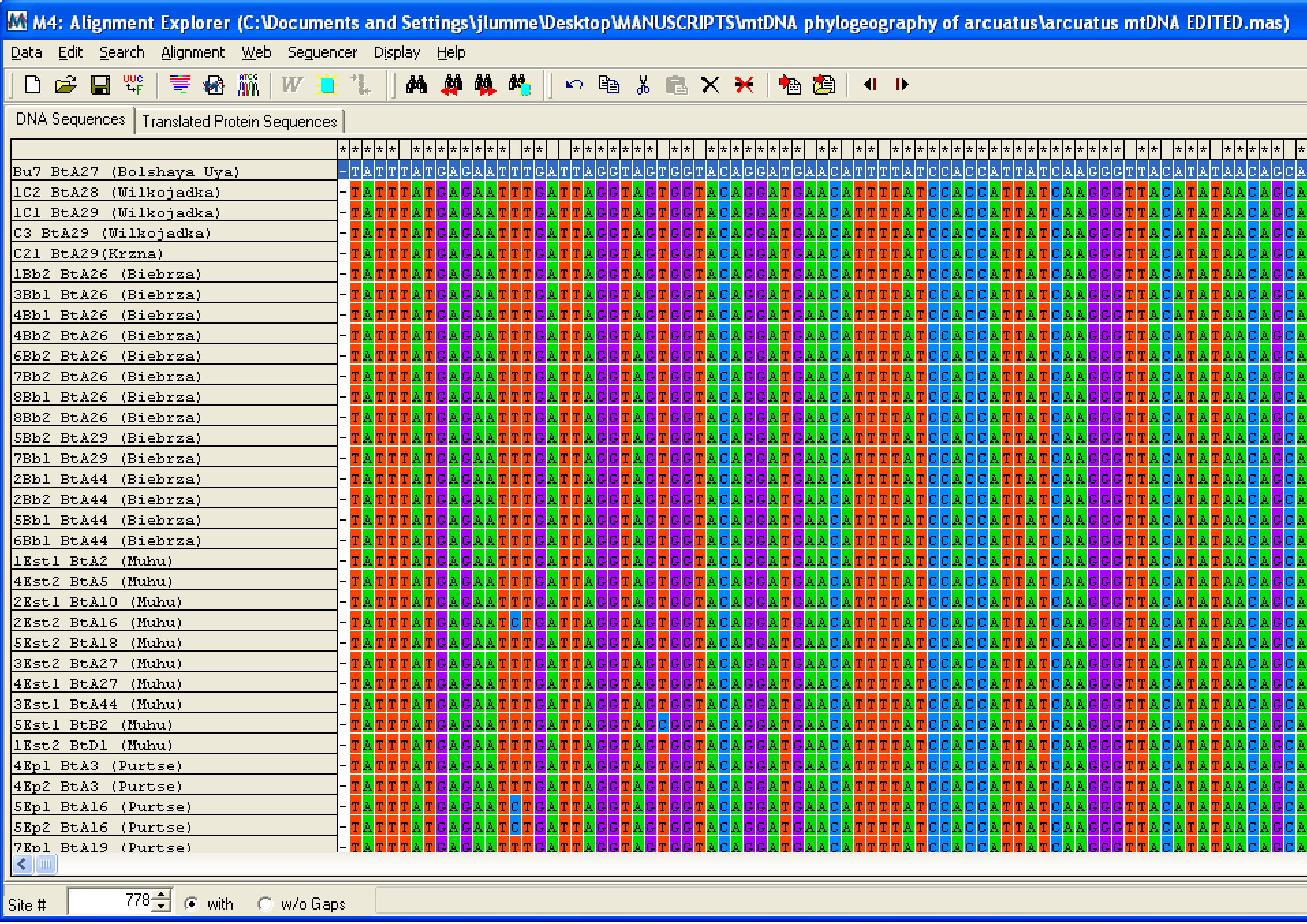
Huonolta näyttää. Harmaat tähdet on stoppeja



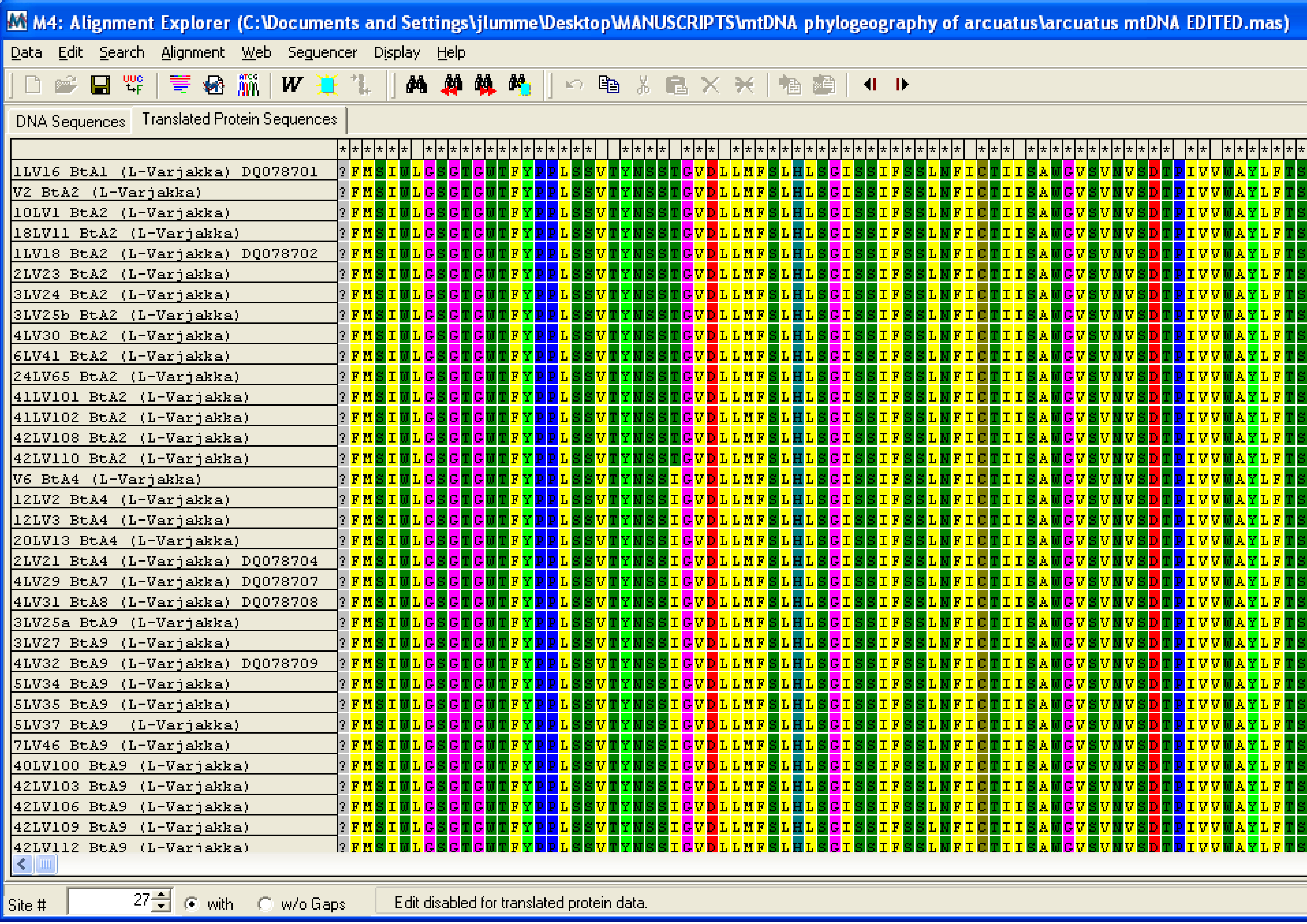
Ensimmäinen luettu TAT, mites se toimisi?



Huono huono. Stoppeja runsaasti, myöskin variaatiota liikaa



Ensimmäinen kokonainen TTT?



Fenylalaniinilla alkaa, stoppeja ei ole, variaatiota vähän. Tämä kehys OK.

tagctatgactaccaccctaattacccacattattaaccccctcgcatatatcgtccccg ttcttttagcagtcgctttcctca
ccttac tcgaacgaaa agtccttgga tatatgcaacttcggaaagggccaaacatcgtcggtccatacggattgcttca
acctatcgcggacggcctaaaactattcattaaagaaccagttcgaccgtccacctcctccccttcctatttctcgctaca
cccatacttgcccttacgcttgcactaaccctgtgagcccccatgcccatcccttaccccattacagacctaaatctcggg
gtactatttgtcctcgcactttccagcctagccgtgtattctattttaggctcagggtgagcttcaaattctaaatatgccctaat
tggagctctacgagcagtggcacaaaccatttcctacgaagtcagccttggactaatcttactcagcgtaattatctttacgg
ggggatttacactacaaaccttcaatgtagcccaagaaagcatctgactactcgtaccagcctgaccccttgccgccatat
ggta tatctctact ctagctgaaa

4441 caaaccgtgc accctttgac cttacagaag gagaatcaga attagtctcc ggatttaatg
4501 tagaatacgc cggagggccc ttcgccctct tctttctagc cgaatacgct aatatccttc
4561 taatgaacac actctcggcc attctatttt taggcgcatc ccacatcccc gccttccccg
4621 aattaacagc cgtaaaccta ataacaaagg ccgccctcct ctccgttgta tttttatgag
4681 tacgagcctc ctacccacga tttcgctacg accaactcat acatttagtt tgaaaaagct
4741 tcttaccttt gacactagcc cttgtcctat gacacctagc acttccaacc gcaatggcag
4801 gcctccctcc ccaactttaa gccccaagga attgtgcctg aatgtttaag gaccaccttg

NCBI homepage

Jonkun otuksen jonkun geenin aminohapposekvenssiä??

MTTTLITHIINPLAYIVPVLLAVAFLTLLERKVLGYMQLRKG
PNIVGPYGLLQPIADGLKLFIKEPVRPSTSSPFLFLATPMLALT
LALTLWAPMPIPYPITDLNLGVLFVLALSSLAVYSILGSGWAS
NSKYALIGALRAVAQTISYEVSLGLILLSVIIFTGGFTLQTFN
VAQESIWLLVPAWPLAAMWYISTLAETNRAPFDLTEGESEL
VSGFNVEYAGGPFALFFLAEYANILLMNTLSAILFLGASHI
PAFPELTAVNLMTKAALLSVVFLWVRASYPRFRYDQLMHLV
WKSFLPLTLALVLWHLALPTAMAGLPPQ L

????????????????

NCBI homepage

Geneettinen koodi on “universaali”

Oleellisesti samaa koodia, *pienin* muutoksin, käyttävät arkit, bakteerit, eukaryootit ja vielä mitokondriot ja kloroplastitkin

Eukaryootin tuman geenit ja mitokondrion geenit luetaan eri koodilla, siksi mitokondriossa on oma rRNA (2) ja 22 tRNA:ta (**TRANSLAATIOKONEISTO**)

Esim. UGA on universaalisti STOP, mutta mitokondriossa se on tryptofaanin koodi, ja AGA ja AGG taas ovat yleensä arginiinin koodina, mutta nisäkäsmitokondriossa ne luetaan STOP.

CELL 862

The Genetic Code

1st position (5′ end)	2nd Position U	C	A	G	3rd Position (3′ end)
U	Phe	Ser	Tyr	Cys	U
	Phe	Ser	Tyr	Cys	C
	Leu	Ser	**STOP**	**STOP**	A
	Leu	Ser	**STOP**	Trp	G
C	Leu	Pro	His	Arg	U
	Leu	Pro	His	Arg	C
	Leu	Pro	Gln	Arg	A
	Leu	Pro	Gln	Arg	G
A	Ile	Thr	Asn	Ser	U
	Ile	Thr	Asn	Ser	C
	Ile	Thr	Lys	Arg	A
	Met	Thr	Lys	Arg	G
G	Val	Ala	Asp	Gly	U
	Val	Ala	Asp	Gly	C
	Val	Ala	Glu	Gly	A
	Val	Ala	Glu	Gly	G

Phe	171	UUU	AAA	0	Ser	147	UCU	AGA	10	Tyr	124	UAU	AUA	1	Cys	99	UGU	ACA	0
Phe	203	UUC	GAA	14	Ser	172	UCC	GGA	0	Tyr	158	UAC	GUA	11	Cys	119	UGC	GCA	30
Leu	73	UUA	UAA	8	Ser	118	UCA	UGA	5	stop	0	UAA	UUA	0	stop	0	UGA	UCA	0
Leu	125	UUG	CAA	6	Ser	45	UCG	CGA	4	stop	0	UAG	CUA	0	Trp	122	UGG	CCA	7
Leu	127	CUU	AAG	13	Pro	175	CCU	AGG	11	His	104	CAU	AUG	0	Arg	47	CGU	ACG	9
Leu	187	CUC	GAG	0	Pro	197	CCC	GGG	0	His	147	CAC	GUG	12	Arg	107	CGC	GCG	0
Leu	69	CUA	UAG	2	Pro	170	CCA	UGG	10	Gln	121	CAA	UUG	11	Arg	63	CGA	UCG	7
Leu	392	CUG	CAG	6	Pro	69	CCG	CGG	4	Gln	343	CAG	CUG	21	Arg	115	CGG	CCG	5
Ile	165	AUU	AAU	13	Thr	131	ACU	AGU	8	Asn	174	AAU	AUU	1	Ser	121	AGU	ACU	0
Ile	218	AUC	GAU	1	Thr	192	ACC	GGU	0	Asn	199	AAC	GUU	33	Ser	191	AGC	GCU	7
Ile	71	AUA	UAU	5	Thr	150	ACA	UGU	10	Lys	248	AAA	UUU	16	Arg	113	AGA	UCU	5
Met	221	AUG	CAU	17	Thr	63	ACG	CGU	7	Lys	331	AAG	CUU	22	Arg	110	AGG	CCU	4
Val	111	GUU	AAC	20	Ala	185	GCU	AGC	25	Asp	230	GAU	AUC	0	Gly	112	GGU	ACC	0
Val	146	GUC	GAC	0	Ala	282	GCC	GGC	0	Asp	262	GAC	GUC	10	Gly	230	GGC	GCC	11
Val	72	GUA	UAC	5	Ala	160	GCA	UGC	10	Glu	301	GAA	UUC	14	Gly	168	GGA	UCC	5
Val	288	GUG	CAC	19	Ala	74	GCG	CGC	5	Glu	404	GAG	CUC	8	Gly	160	GGG	CCC	8

Figure 34 The human genetic code and associated tRNA genes. For each of the 64 codons, we show: the corresponding amino acid; the observed frequency of the codon per 10,000 codons; the codon; predicted wobble pairing to a tRNA anticodon (black lines); an unmodified tRNA anticodon sequence; and the number of tRNA genes found with this anticodon. For example, phenylalanine is encoded by UUU or UUC; UUC is seen more frequently, 203 to 171 occurrences per 10,000 total codons; both codons are expected to be decoded by a single tRNA anticodon type, GAA, using a G/U wobble; and there are 14 tRNA genes found with this anticodon. The modified anticodon sequence in the mature tRNA is not shown, even where post-transcriptional modifications can be confidently predicted (for example, when an A is used to decode a U/C third position, the A is almost certainly an inosine in the mature tRNA). The Figure also does not show the number of distinct tRNA species (such as distinct sequence families) for each anticodon; often there is more than one species for each anticodon.

Kaikki kodonit eivät ole yhtä suosittuja