

# 第6回 遺伝子の探索と遺伝子の機能予測

## 本日学ぶこと

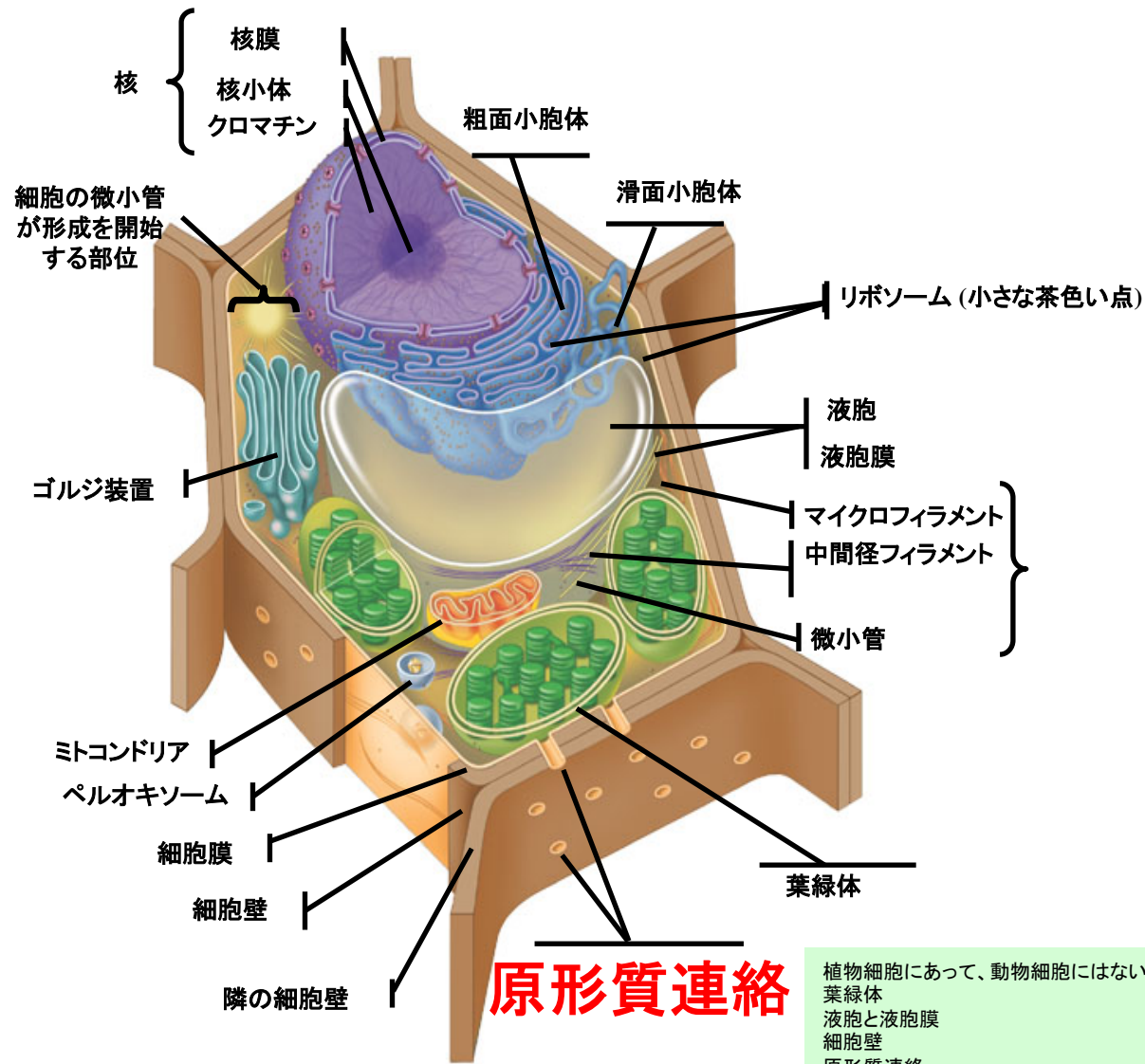
データベースを使って遺伝子を探して、  
かつその遺伝子の機能を予測する。

これまで大学に入学して2年半の授業の中で、DNA、RNA、アミノ酸、蛋白質について構造や特性などを学習してきた。

また、DNAの一部が転写され、スプライシングを受けた後、蛋白質に翻訳されることを学習してきた。

しかし、実際長いDNA配列のどの部分が遺伝子なのだろうか？

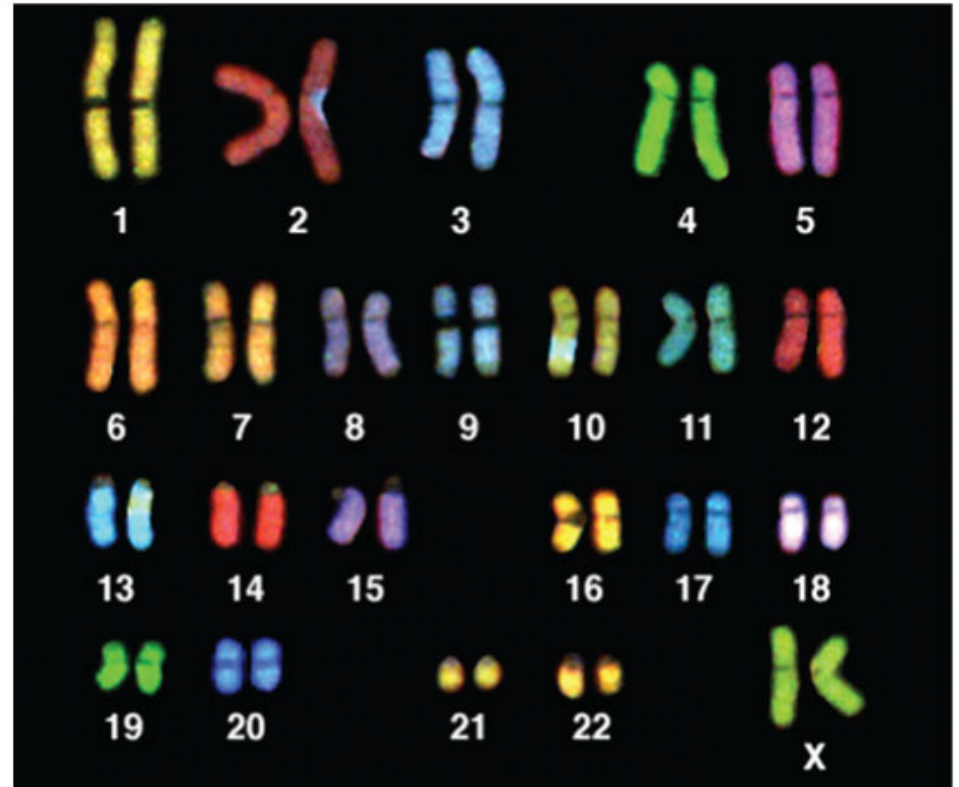
少し復習を試みよう。



細胞質基質(サイトゾル)  
細胞質から細胞小器官を除いたもの



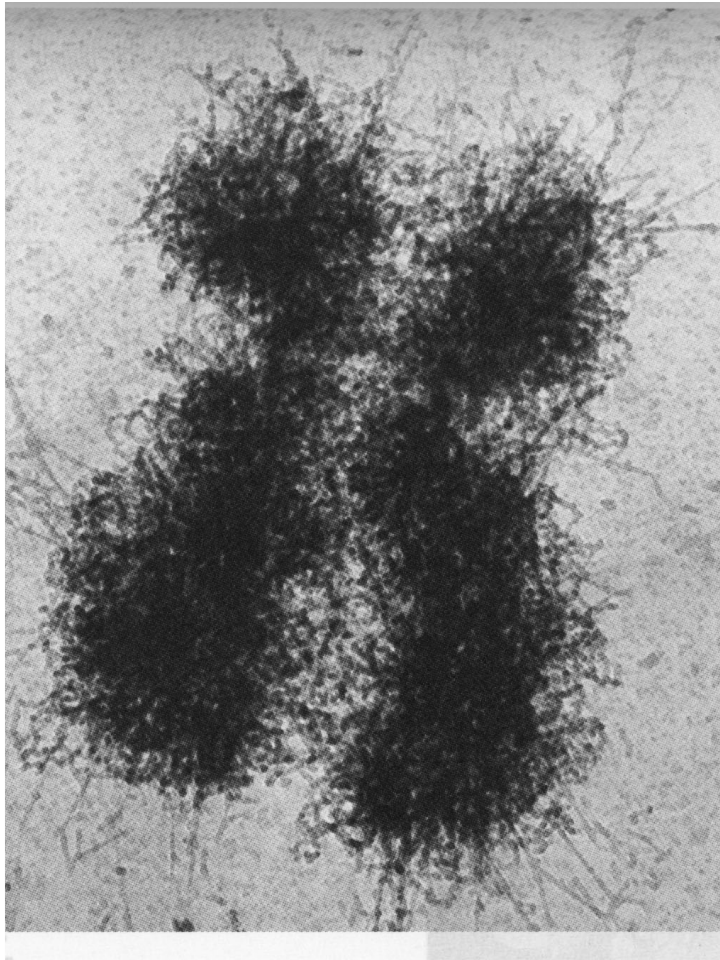
© 2011 Pearson Education, Inc.



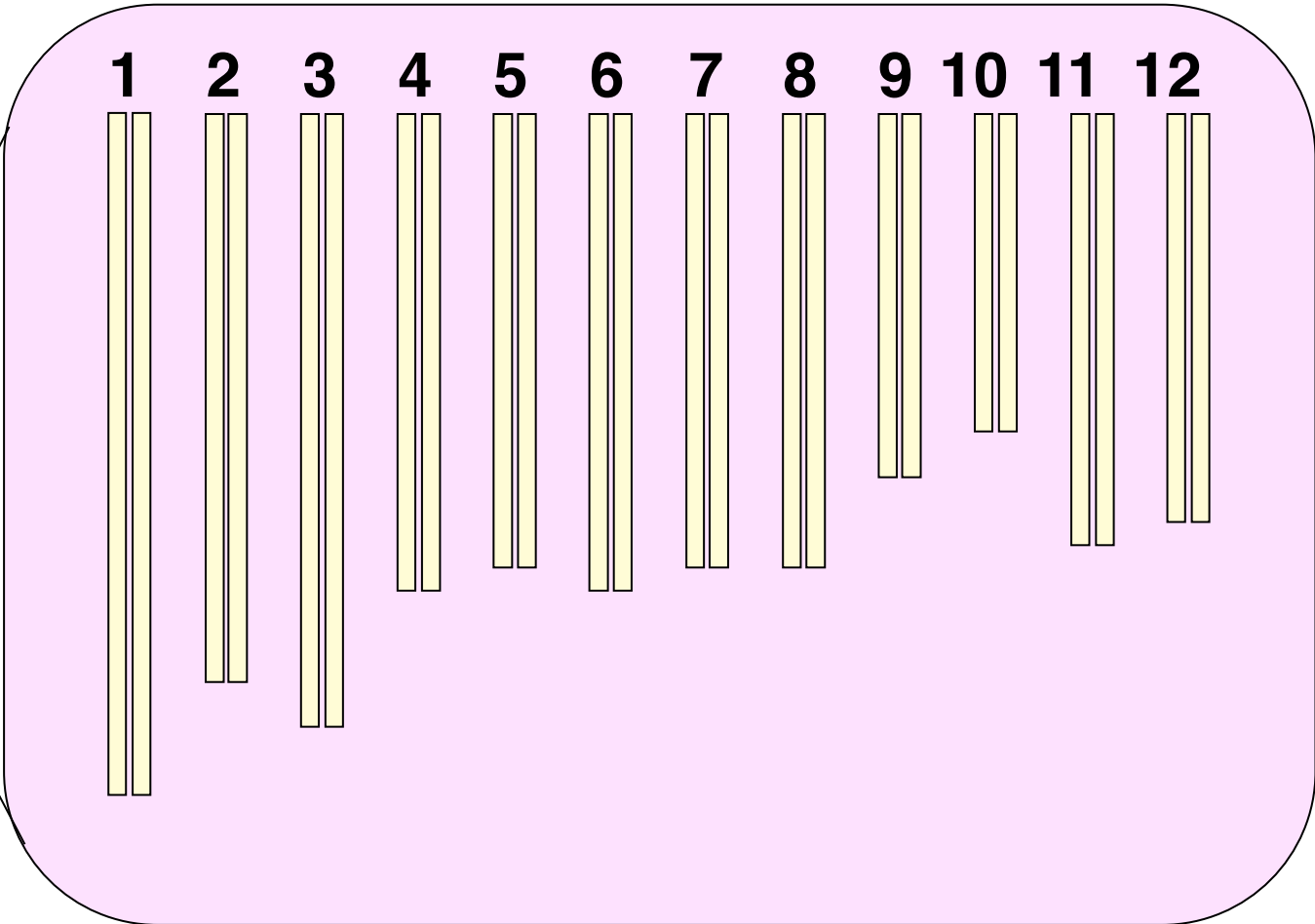
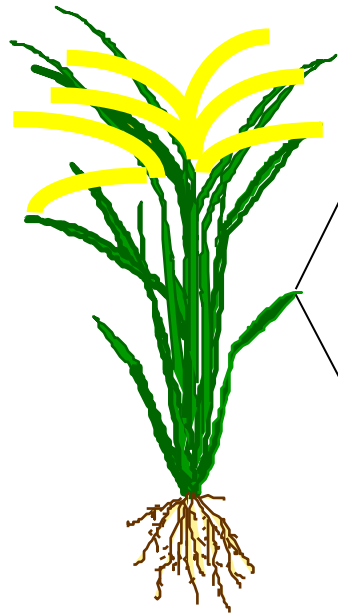
© 2011 Pearson Education, Inc.

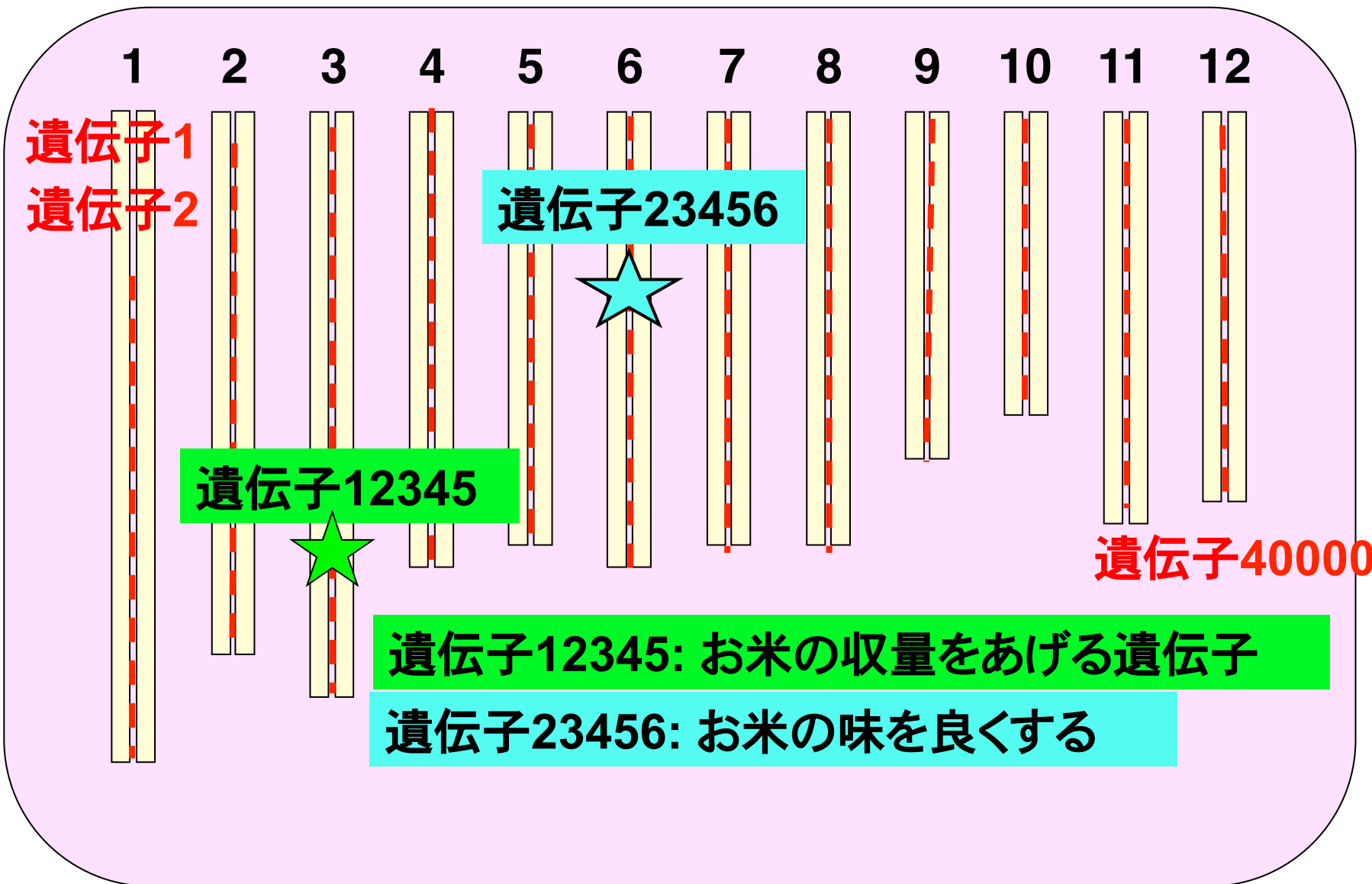
## 人間の染色体(核型)

キャンベル参照

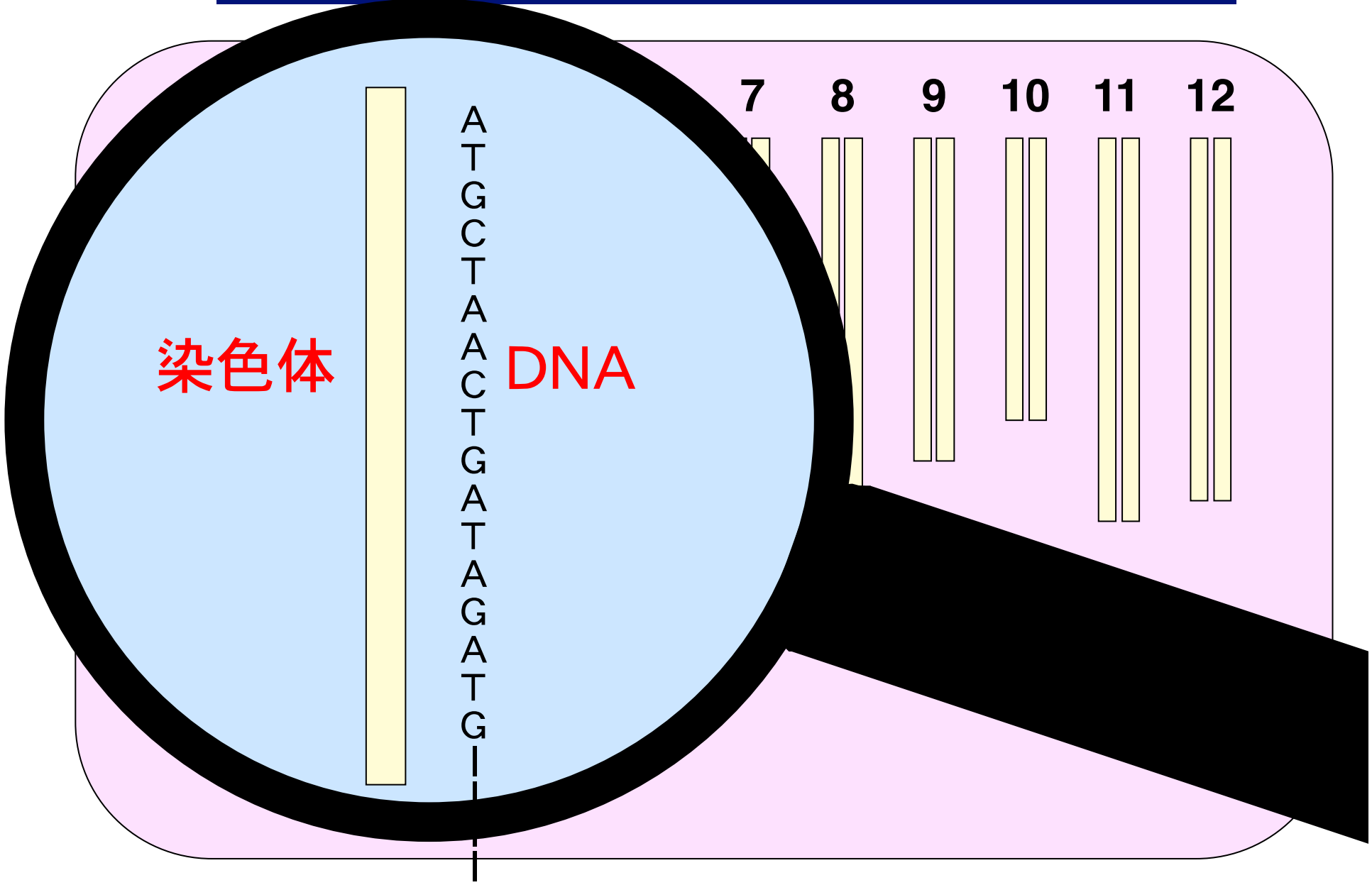


# イネの細胞の中の染色体とDNA





# イネの細胞の中の染色体とDNA



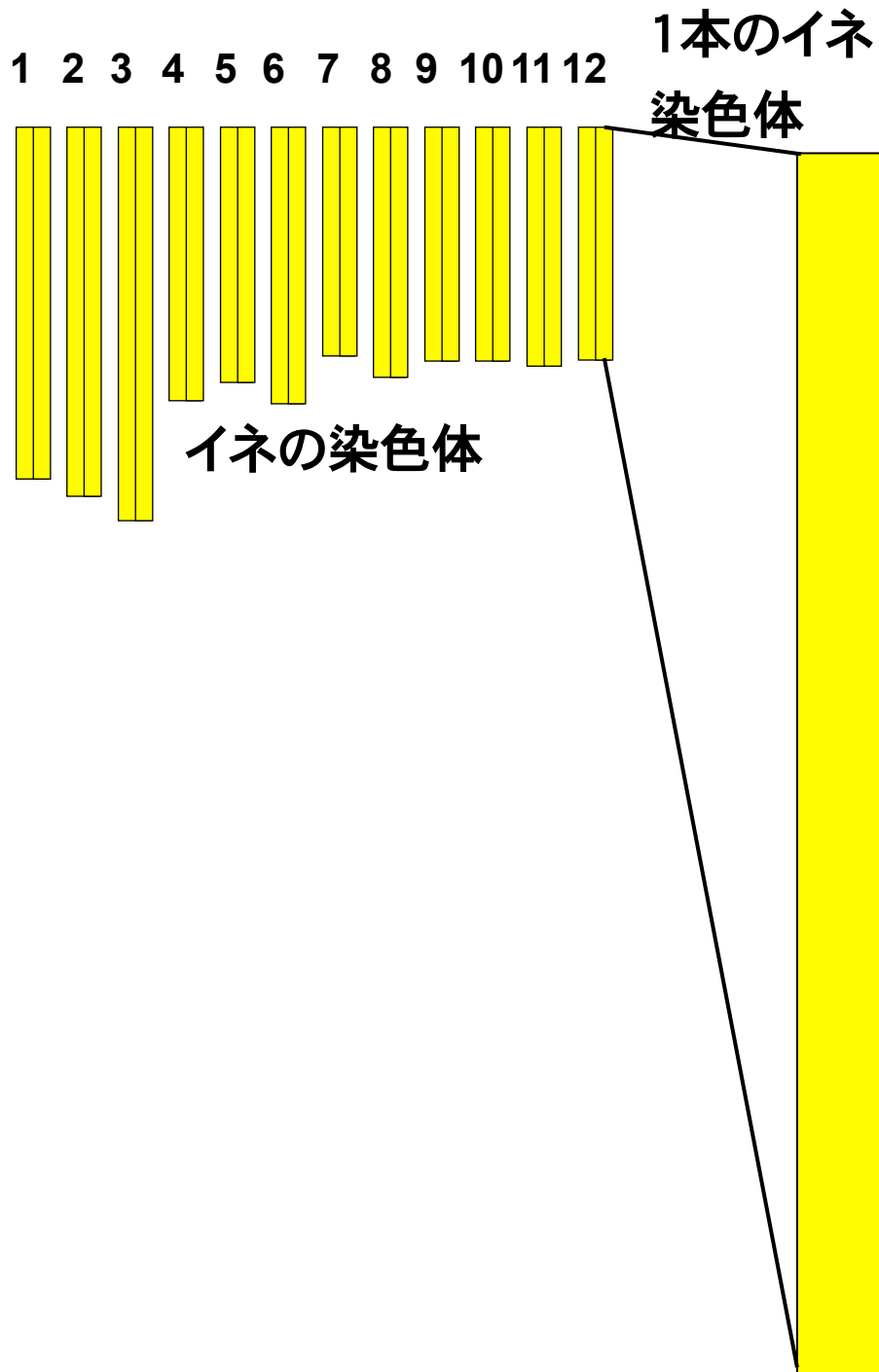
染色体

DNA

A  
T  
G  
C  
T  
A  
A  
C  
T  
G  
A  
T  
A  
G  
A  
T  
G

7 8 9 10 11 12





DNA

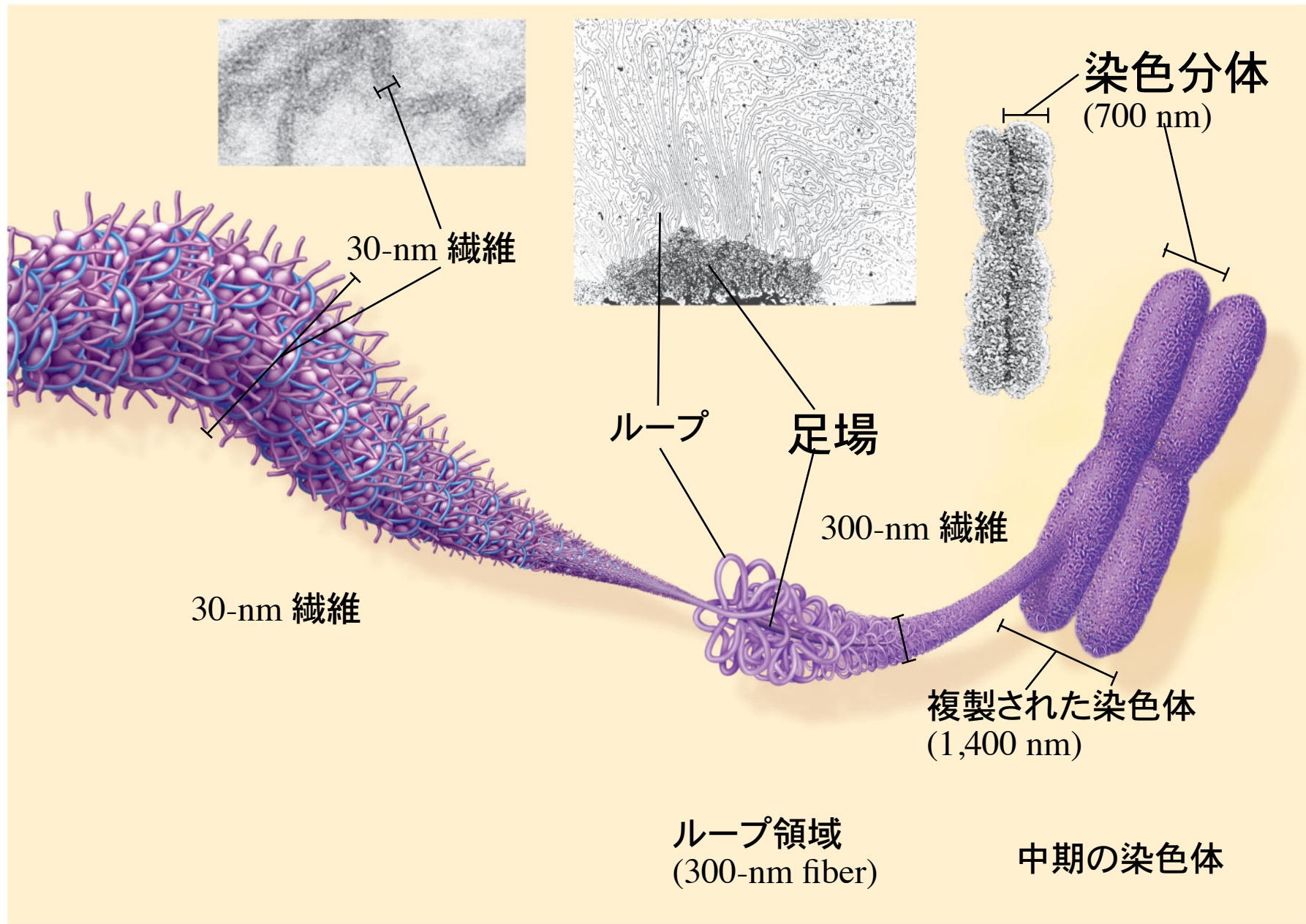
A  
T  
G  
T  
I  
I  
A  
A  
T  
A  
G  
G  
T  
A  
A  
A  
T  
G  
I  
I  
T  
A  
G  
G  
T

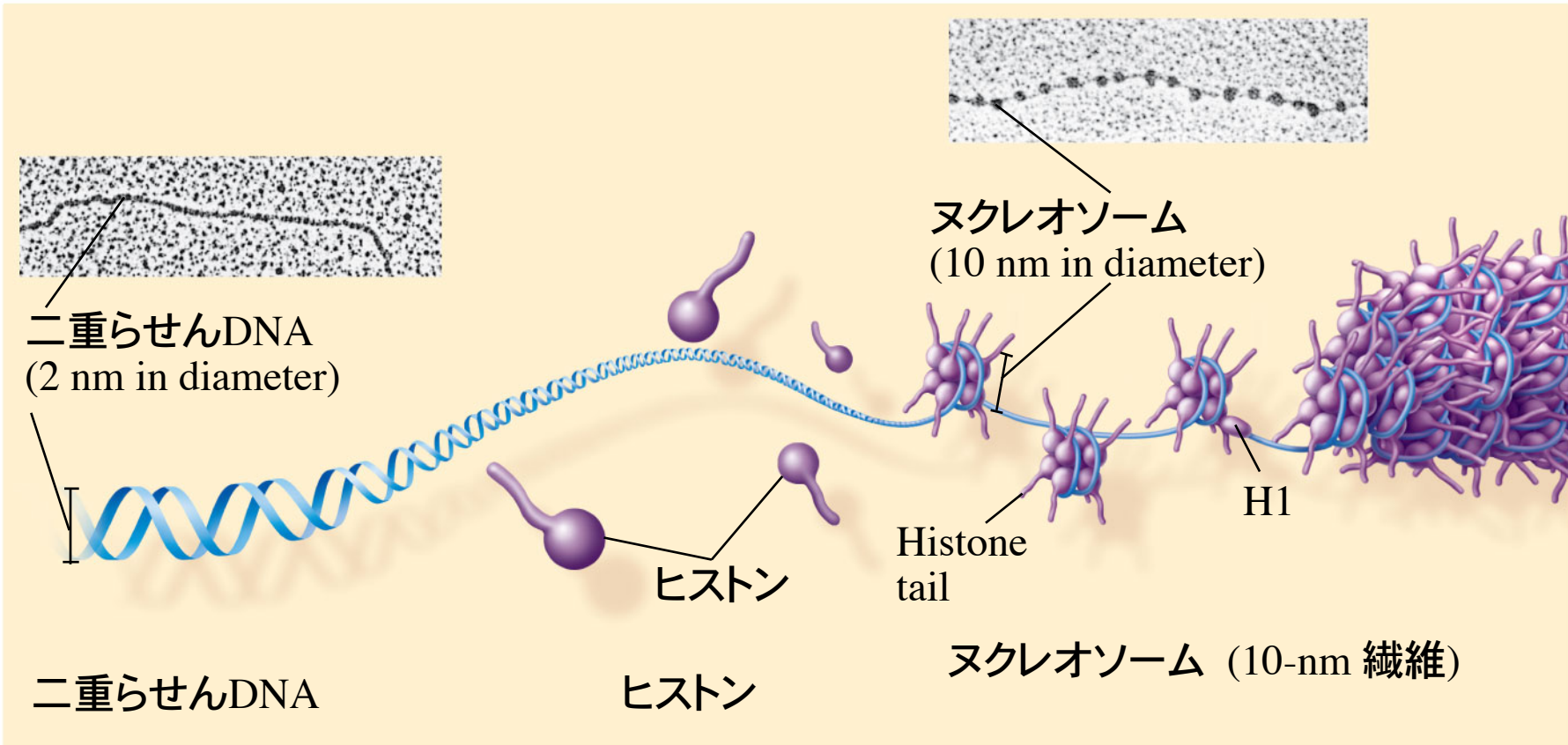


遺伝子1



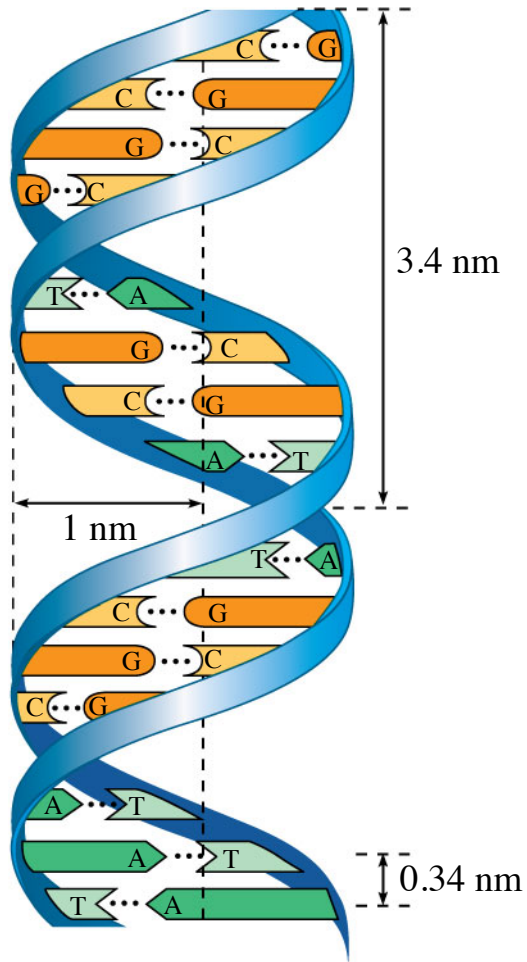
遺伝子2



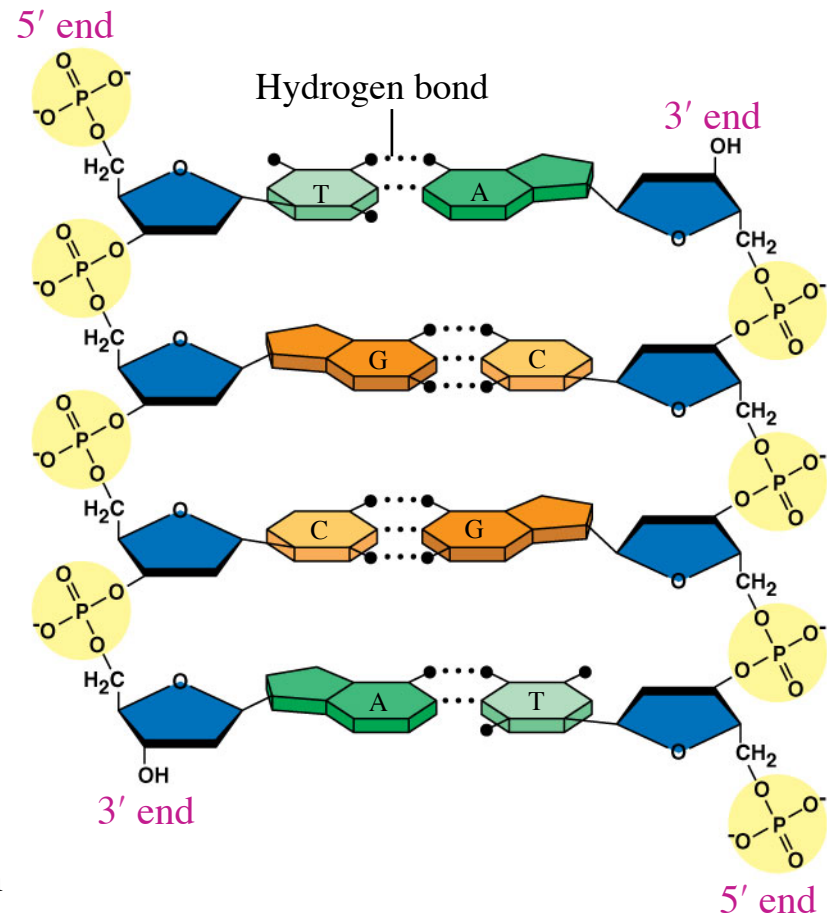


© 2011 Pearson Education, Inc.

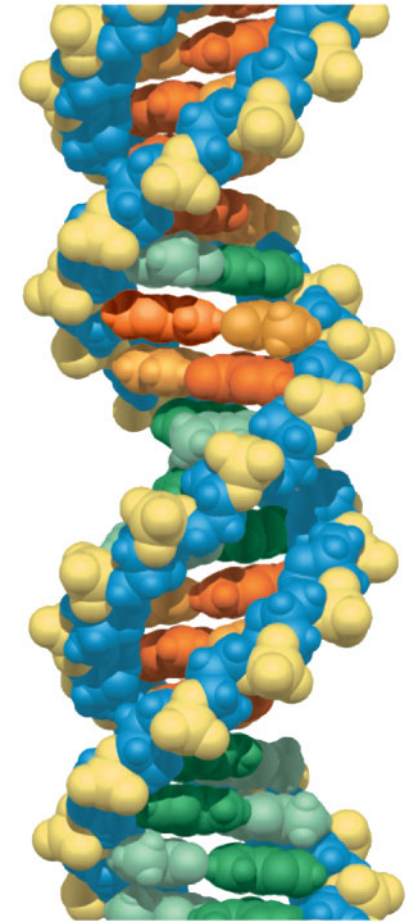
Figure 16.7



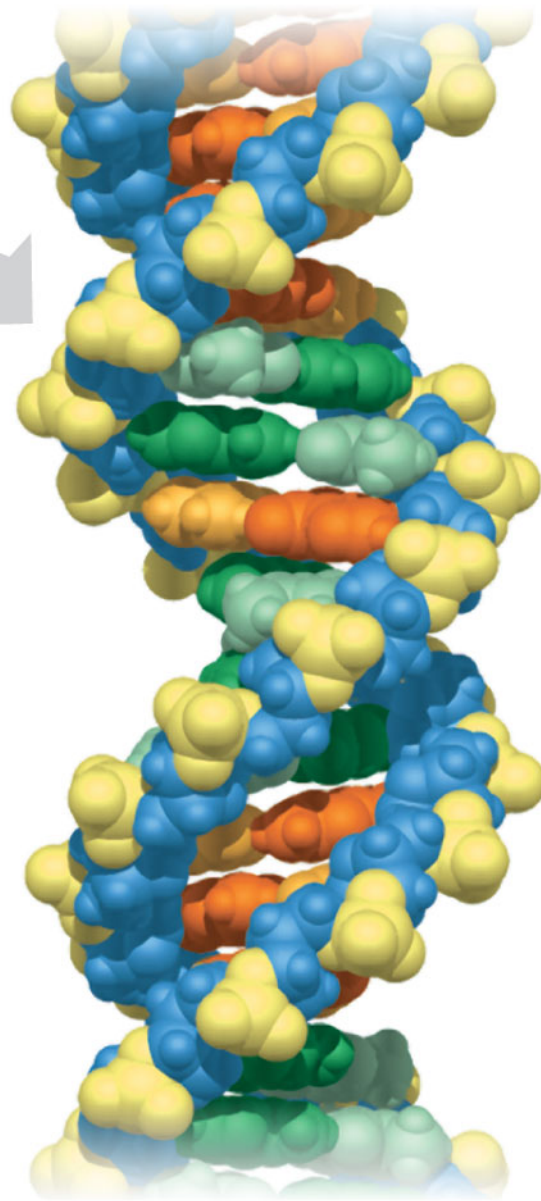
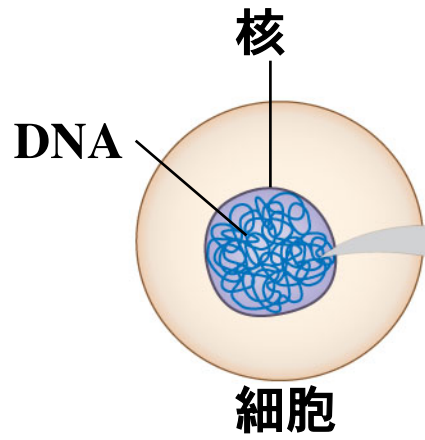
(a) Key features of DNA structure



(b) Partial chemical structure

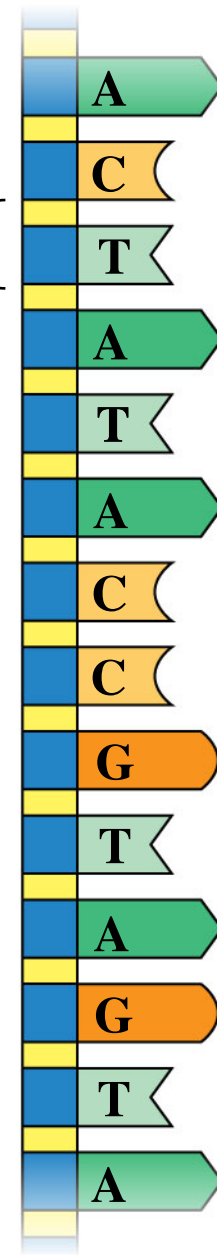


(c) Space-filling model

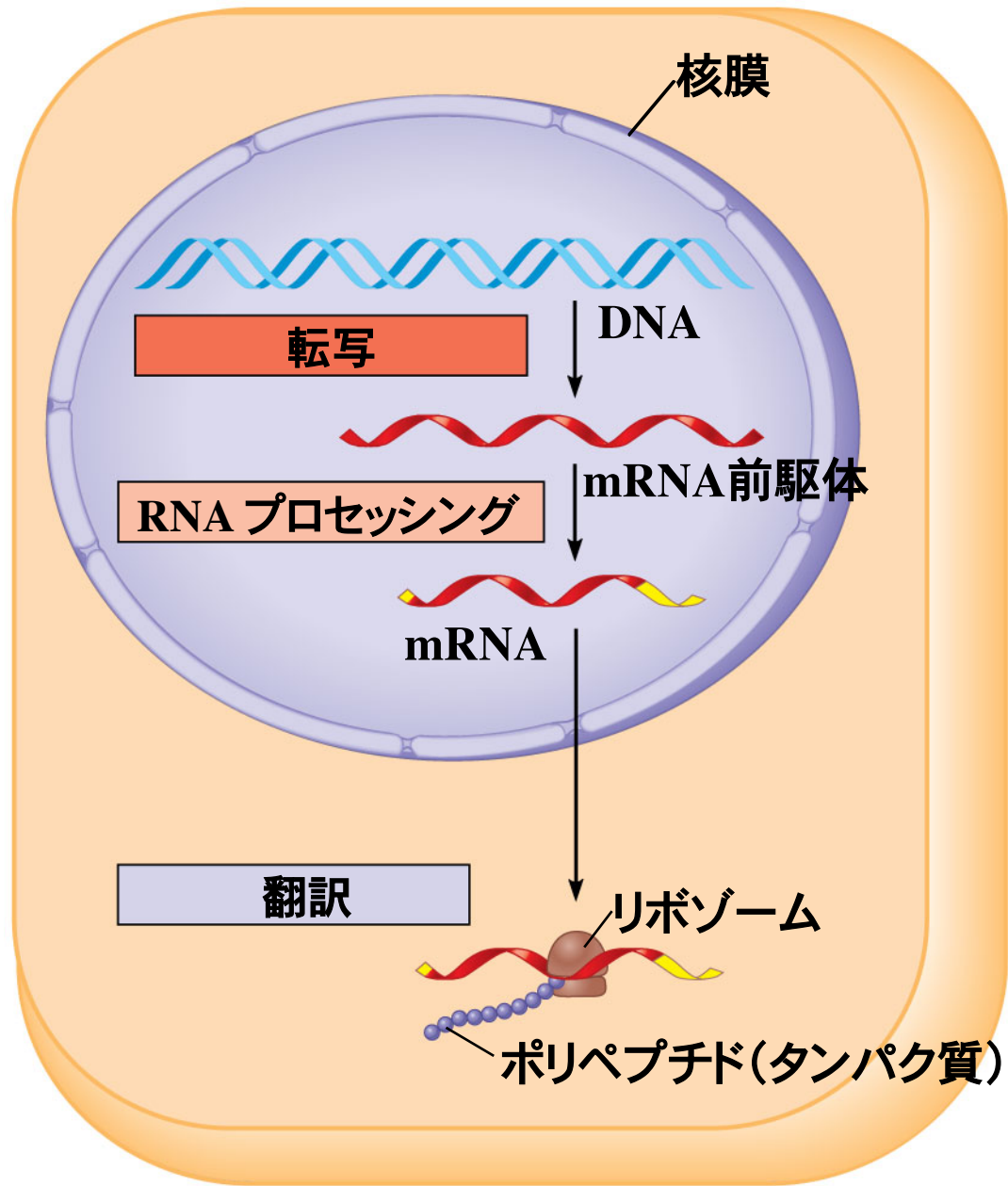


(a) DNA 二重らせん

ヌクレオチド



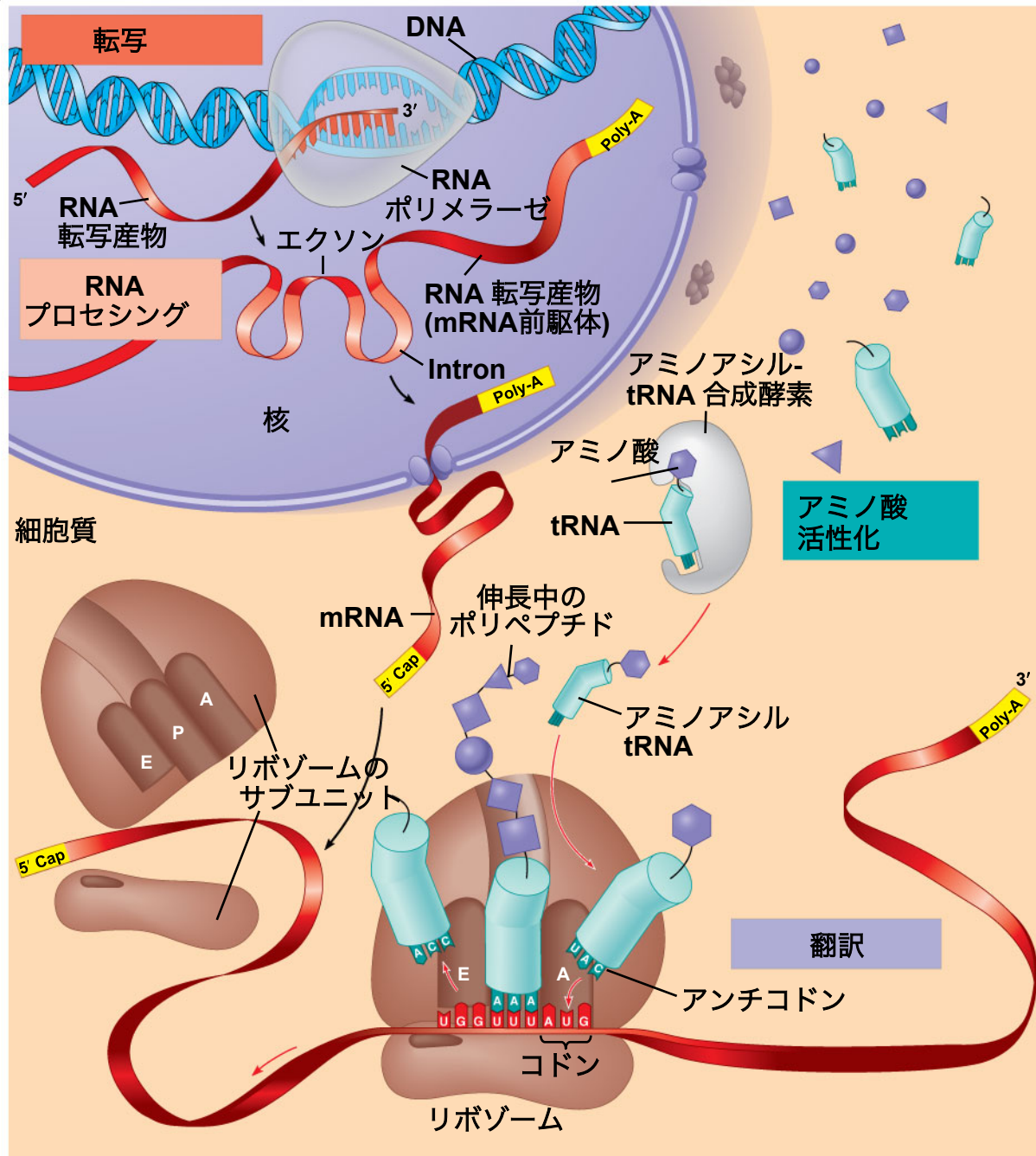
(b) 1本鎖DNA

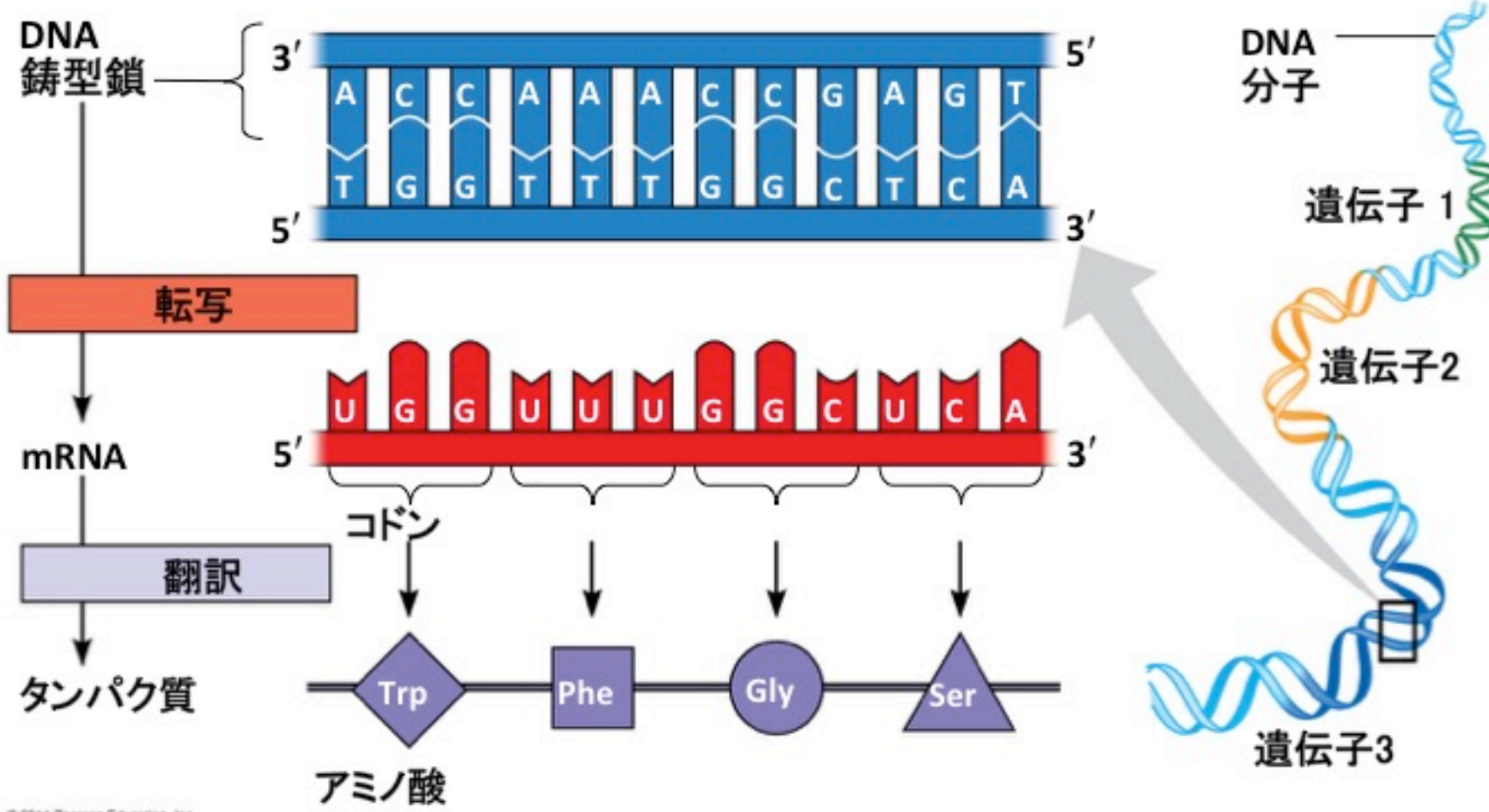


(b) 真核生物の細胞

© 2011 Pearson Education, Inc.

Figure 17

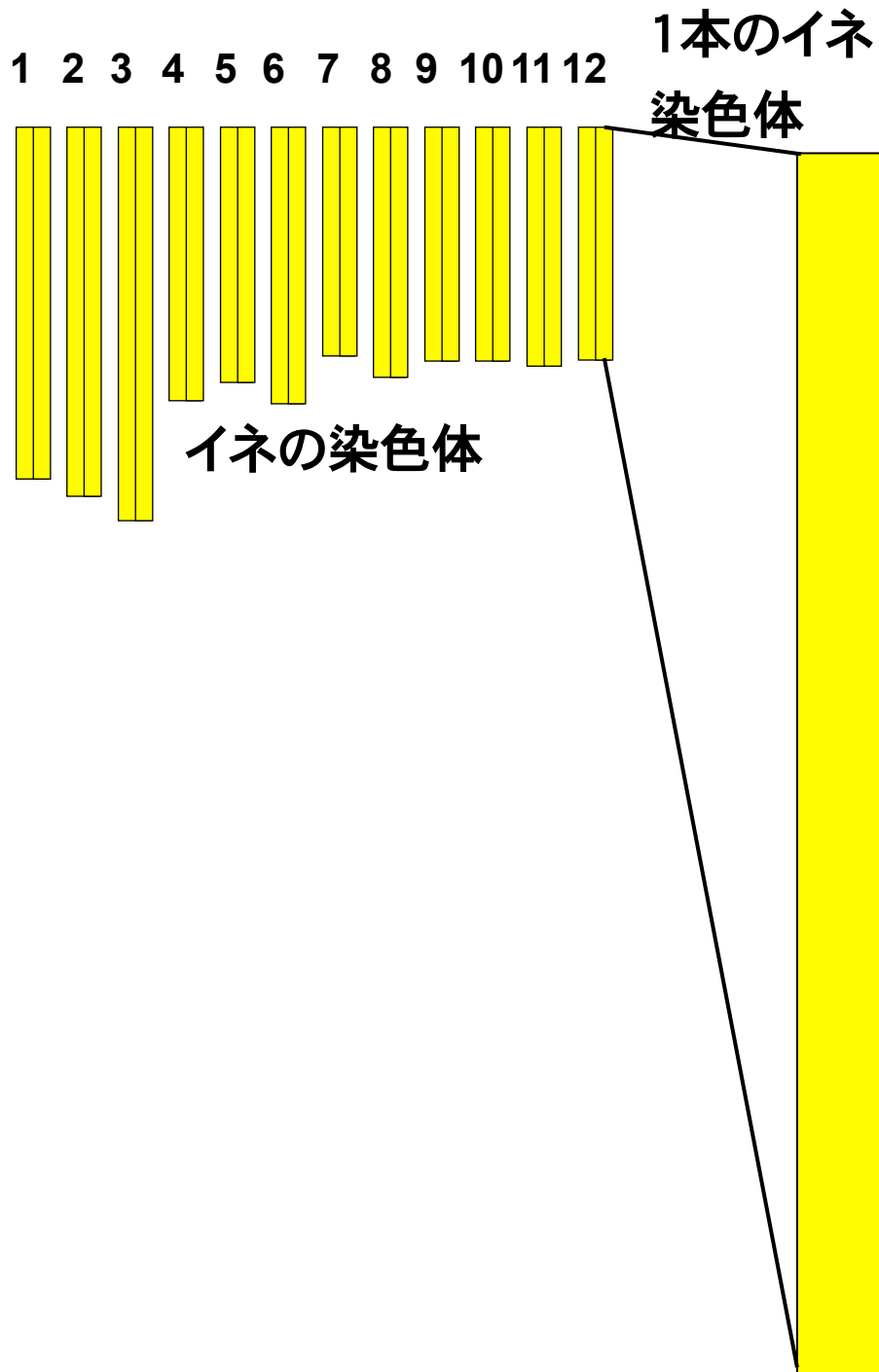




© 2011 Pearson Education, Inc.

キャンベル参照





DNA

A  
T  
G  
T  
I  
I  
A  
A  
T  
A  
G  
G  
T  
A  
A  
A  
T  
G  
I  
I  
T  
A  
G  
G  
T



遺伝子1



遺伝子2

```
GENETYX-MAC File Edit Find Nucleotide Peptide Font Windows
GA20oxidase2-genomic DNA
10 20 30 40 50 60 70 80 90 100
TCTCCCGTGT TACAAATACC CCACCTCCTT GCCAGACAGC CTCGCCCTGC ACACACACAC ACACCTCACAC TCACACACGC TCTCACTACT CTCCCGCTCA
110 120 130 140 150 160 170 180 190 200
ACACAGCGCT CACTTCTCAT CTCCACTCTC ATGGTGCCCG AGACDCCAC GCCACCCACG CCGACACACG CACCGCCCAT GCACTCACC GCCGGCTCTG
210 220 230 240 250 260 270 280 290 300
GCATTGCCGC CCCGGCGCGC GCGCGCGTGT GCGACCTGAG GATGGAGCCC AGGCACTTCGT GTGGCCGACG GCGCAGCGCA GCCCGCGCTC
310 320 330 340 350 360 370 380 390 400
GGCGCGCGAG CTGGACATGC CCGTGGTCCG CTGGGCGGTG CTCGCCGACG GCGACCGCGA GGGGCTGCG CCGCGCCGCG CCGAGGTGGC CCGCCGCTGC
410 420 430 440 450 460 470 480 490 500
GCCACGCACG GGTCTCTCCA GGTGTCCGAG CACGGCGTCC AGCCCGCTCT GCGCGCGCGC GCCTCGACG GCGCCAGCGA CTCTTCCGCG CTCGCCGCTG
510 520 530 540 550 560 570 580 590 600
CCGAGAGCGG CCGCGCGCGC CCGCTCCGCG GCACCGTGTG CGGTACACC AGCGCCCGCG CCGACCGCTT CGCTCCAGG TCCCATGGG AGGAGCCCT
610 620 630 640 650 660 670 680 690 700
CTCCTTCGGC TTCACAGCCG GCGCCGCGCG CCCCGTCTCT GCCGACTACT TCTCCAGCAC CCTCGGCCCG GACTTCGCGC CAATGGGGTA ATTAATACGA
710 720 730 740 750 760 770 780 790 800
TGGTGGACGA CATTGCATTT CAATTCAAA ACAATTCARA AACACACCGA CCGAGATTAT GCTGATTTCA AACCGTITTG TCGCCGACGG AGGGTGTACC
810 820 830 840 850 860 870 880 890 900
AGAGACTGCG CAGAGAGATG AAGGAGTGTG CGCTGACGAT CATGGACTCT CTGGAGCTGA GCCTGGGCGT GAGCGGAGCG TACTACAGGG AGTTCTTCGC
910 920 930 940 950 960 970 980 990 1000
GGACAGCAGC TCARATCTGC GGTGCACTCA CTACCCGCCA TGCCCGAGC CCGAGCGGAC GCTCGGACCG GCGCCGACCT GCGACCCCGC CCGCTCACCC
1010 1020 1030 1040 1050 1060 1070 1080 1090 1100
ATCCTCTCCG AGGACGAGCT CCGCGCGCTC GAGGTCTCTG TCAGCGGCGA ATGGCGCCCG CTCAGCCCGC TCCCGCGCGC CATGGTATC ACATCGGGC
1110 1120 1130 1140 1150 1160 1170 1180 1190 1200
ACACCTTCAT GGTAAACCAT CTCTATTCTT CCTCTCTCTT GTTCTCTCTT GCTTCGAGCG AACAGACARA GTATTCAGG CTTTTTTTTT TCTCTCGCGC
1210 1220 1230 1240 1250 1260 1270 1280 1290 1300
GAATTTGAGC AGAAATATTA GATCGTGTAA GGGCGCGGCG TTTCAGCTGA AAGCGGAGAG AACCAGCTCT GACGTGATTT CTCTGTTCCT ATCACAGCA
1310 1320 1330 1340 1350 1360 1370 1380 1390 1400
ATGGATGCCC CCAGCTCTCC ATGTGTATAG ATTTATCTCA CATCTTATAG TTAATAGAGG TAAGTAAACA GACTACTTTT TCATATATA GTTCGTTTGA
1410 1420 1430 1440 1450 1460 1470 1480 1490 1500
TTTTTTTTTT TTAAGTTTTT TTTAGTTTTA TCCAAATTTA TTGAAARACT TAGCAGGTTT TATATATCCA AATTAGTCTC ATTTAGTTTA ATATGTGATA
1510 1520 1530 1540 1550 1560 1570 1580 1590 1600
TATTTTGTAT ATATATTTAT GTTATATTTA AATATTTACT ATATTTTTCT ATAACATTA TTAAGAGCCA TTTATATATAT AATATGGAGG GAGTAAATTA
1610 1620 1630 1640 1650 1660 1670 1680 1690 1700
TATGGATCTC CCCCGACATG AGAATATTTT CCGATGGTGT GACGACGCCA TGTAGCTTC GGTGGGCGTG GAGCGCCAGG GGTCCACACA GCCACGTCCA
1710 1720 1730 1740 1750 1760 1770 1780 1790 1800
ACACCCCTCG GGTCCCGCCC TACACCTCCA AACAGTAGTG AGTAGTGTCT CGTCGCGTTT TAGTATTTGA TGACAAACAA AGTGTGATTT GAGTTAGCCA
1810 1820 1830 1840 1850 1860 1870 1880 1890 1900
CCACCAACTT GCACACGAGC ACATCACTTT GTGTCCATTC TCGCCAGTCA CTTCCTCTC TAGTCCATAC TCCTATCTAG CGATGTAGGC GGTAAATTTT
1910 1920 1930 1940 1950 1960 1970 1980 1990 2000
ATCATCGGTA TATAACCTG TTTGTTATAG TTAATTTCTT ATATAACTT ATACAGTAT ACATTTTAAA AGAAACAAA ATTAGGATTA ACAAGCCCTG
2010 2020 2030 2040 2050 2060 2070 2080 2090 2100
CTCCTATCCA TCCATGGCAC TTGGAGAGAC CAGACTCGGT CATGCCATGC CAGGCCAAGA TATGGGTTAT GGAAGGTAG AGAAGAGAGG AGATGAGAGA
2110 2120 2130 2140 2150 2160 2170 2180 2190 2200
TAGACATGCG TTCTCTCTCT GTTGGATGT GTATTTTGGG GGAATTTGTG TAGTAGTAGC AGCGGCGCGG CCGGAGCGGA TCGGATGTT GCGCTTTTGG
2210 2220 2230 2240 2250 2260 2270 2280 2290 2300
GTGGCGTTTT CCCGGGGGGG TTTTGGTTTT GCGCTTGGGG GGAATGGCAT GCGCGGCGGT GCGGCTGAC GCCACACACA CCGCGCGCGA CCGACGTAGC
2310 2320 2330 2340 2350 2360 2370 2380 2390 2400
TCGTGCTGCG CCGCGCGCGA CCGTAGCTTA GGGTGGTGTG TTCCGCGCGC GCGCGCGGAT TGTTCATGC CGATCGATTT GCGCCACACC TCGCGCGCGC
2410 2420 2430 2440 2450 2460 2470 2480 2490 2500
TCTTGTGCGG TCGTGCGCCT CTCTCGCGCG GTTTGTCTCT GTCGGTTGCG TCAGCGCGCG ACAGGGGACG GACATATGGC GATGTAGCCC TGACGTGTGC
2510 2520 2530 2540 2550 2560 2570 2580 2590 2600
GGCCTCTCCG TTGATGATG ATGATGTATG TATGTATTTT TTTTGTCTG AAGGAAATTT TGGGAAATTT TTGTGTGTGC AGGCGCTGTC GACCGGGAGG
2610 2620 2630 2640 2650 2660 2670 2680 2690 2700
TATAGAGACT GCTCGCACG GCGGTGGTG AACACCGCG GAGAGCGGCG GTCCGCGCGC TTCTTCTGT GCGCGCGGGA GACAGAGGTG GTCCGCGCGC
2710 2720 2730 2740 2750 2760 2770 2780 2790 2800
CGCCAGCGCG CCGCACGCGC CAGACTTACC CCGACTTCAC CTGGGCGGAC CTGATGCGCT TCACGACGCG CCACTACCGC CCGCAGACCC GACAGCTCGA
2810 2820 2830 2840 2850 2860 2870 2880 2890 2900
CGCCTTCACG CGCTGCTCG CCGCGCGCGC CCGCAGCGCC GCGCGAGCG CCGAGGTGGA GCGCGCGCGC TGATCCCGA ACGAGACAAA ACGAGACAAA
(EMBL) length: 2900 current pos: 697
```

人間の場合 - 32億塩基対  
イネの場合 - 4億塩基対

どこからどこまでが遺伝子なのか？

Finder ファイル 編集 表示 移動 ウィンドウ ヘルプ
GA20oxidase2-genomic DNA
Untitled 3

N → [Icons]
[Icons]

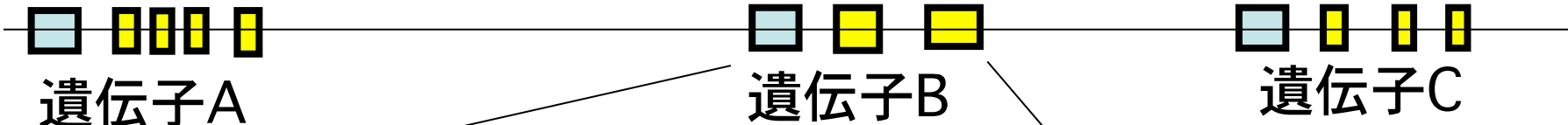
10	20	30	40	50	60	70	80	90	100
TCTCCCTGT	TACAAATAC	CCACCCCTCT	GCCCAAGAC	CTCGCCCTGC	ACACACACAC	ACACTCACAC	TCACACACG	TCTCACTCA	CTCCCCTCA
110	120	130	140	150	160	170	180	190	200
ACACAGCGT	CACCTTCTAT	CTCCAACTTC	ATGGTGCCCG	AGCACCACAC	GCCACCCAC	CCGACCCAC	CACCCGCAAT	GGACTCCAC	GCCGCGCTG
210	220	230	240	250	260	270	280	290	300
GCATTGCCG	CCCGCGCGC	GCGCGGGTGT	GCGACTGAG	GATGGAGCCC	AGATCCCGG	AGCCATTCT	GTGGCCGAC	GGCGACCGA	GCCGCGGCT
310	320	330	340	350	360	370	380	390	400
GCGCGCGAG	CTGGACATG	CCGTGGTCGA	GCTGGGCGT	CTCCGCGAC	GCGACGCCA	GGGGCTGCC	CGCCCGCGG	CGCAGGTGG	CGCCCGGTC
410	420	430	440	450	460	470	480	490	500
GCCACGACG	GGTCTTCCA	GGTGCCGAG	CACGGCGCT	ACGCCGCTT	GCGCGCGCC	GCGCTGACG	GCGCCAGCG	CTTCTCCCG	CTCCCCTGC
510	520	530	540	550	560	570	580	590	600
CCGAGAGCG	CCGCGCGCG	GCGCTCCCG	GCCCGCTTC	GCGTACACC	AGCGCCACG	CCGACCGCT	CGCCTCCAG	CTCCCATGG	AGGAGACCT
610	620	630	640	650	660	670	680	690	700
CTCCTTCCG	TTCCACGAC	GCGCGCGCG	CCCGCTGTC	GCGACTACT	TCTCCAGAC	CCTCGGCCG	GACTTCCGC	CAATGGGTA	ATTAAACGA
710	720	730	740	750	760	770	780	790	800
TGGTGGACG	CATTGCATT	CAAAATCARR	ACAAATCARR	ACACACCCG	CCGAGATTG	GCTGAATTC	ACCCGTTTG	TGGCGCAGG	AGGGTGTAC
810	820	830	840	850	860	870	880	890	900
AGAACTACT	CGAGGAGTG	AGGAGCTGT	CGCTGCAGT	CATGGACTC	CTGGAGCTG	GCCTGGCGT	GGAGCGAGG	TACTACAGG	AGTTCTCCG
910	920	930	940	950	960	970	980	990	1000
GGACGACCG	TCAATCATG	GGTGACATC	CTACCCGCA	TGCCCGGAC	GCGAGCGAG	GCTCGGACG	GGCCCGCAC	GCGACCCAC	GCCCTCCAC
1010	1020	1030	1040	1050	1060	1070	1080	1090	1100
ATCCTCTCC	AGGACGAGT	CSCGCGCTC	GAGGTCTCT	TCGACGCCA	ATGGCGCCC	GTCCGCGCC	CATGGTCTC	ACACTCCGC	
1110	1120	1130	1140	1150	1160	1170	1180	1190	1200
ACACCTTCAT	GTAARACAT	CTCCTATTCT	CCTCTCCTCT	GTTCCTCCT	GCTTCGAGC	ACACAGACAA	GTAATTCAG	CTTTTTTTC	TCTCTCGGC
1210	1220	1230	1240	1250	1260	1270	1280	1290	1300
GAAATTCAG	AGAAARATG	GATCGTGGT	GGGCGGGGC	TTTCACTGA	AGCGGGGAG	AAACCGACT	GCGGTATT	CTCTGTTCG	ATCACAAAC
1310	1320	1330	1340	1350	1360	1370	1380	1390	1400
ATGGAAATG	CCACTCTCC	ATGTGTATG	ATTATCTCA	CATCTTATG	TTAATAGAG	TAAATACAA	GCTACTTTT	TCATATTAT	GTTCGTTTG
1410	1420	1430	1440	1450	1460	1470	1480	1490	1500
TTTTTTTTT	TTAAAGTTT	TTTAGTTTA	TCCAAATTT	TTGAAARCT	TAGCAACGT	TATATACCA	AAATAGTCT	ATTTAGTTA	ATATTGTAT
1510	1520	1530	1540	1550	1560	1570	1580	1590	1600
TATTTTGAT	ATATATTTT	GTATATTTT	AAATATTTT	ATATTTTCT	ATAAACATTA	TAAARAGCA	TTTATATAT	AAATGGAGG	GAGTAAATTA
1610	1620	1630	1640	1650	1660	1670	1680	1690	1700
TATGGATCT	CCCCGACATG	AGAAATTTT	CCGATGGTG	GACGACCCA	TGTAAGCTT	GTTGGGCTG	GACGGCCAG	GTTGCCACA	GCCACTCCA
1710	1720	1730	1740	1750	1760	1770	1780	1790	1800
ACACCCCTG	GGTCCCGCC	TAACTTCCA	ACAGTAGTG	AGTAGTGTCT	CGTCGCGTT	TAGTATTGA	TGACAAACA	AGTGTAGTT	GAGTTAGCA
1810	1820	1830	1840	1850	1860	1870	1880	1890	1900
CCACCAACT	GACACGAGC	ACATACATT	GTGTCCATC	TCGCCAGTA	CTTCCATCT	TAGCTTACG	TCCTATCTG	CGATGAGC	GGATATTTT
1910	1920	1930	1940	1950	1960	1970	1980	1990	2000
ATCATCCGA	TATAACCTG	TTTGTTATG	TAAATTTCT	ATATAACT	ATACAGTAT	ACATTTTAA	AGAAACAAA	ATTAGGATA	ACAGCCCTG
2010	2020	2030	2040	2050	2060	2070	2080	2090	2100
CTCCTATCA	TCCATGGAC	TTGGAGGAC	CAGACTCGT	CATGCCATG	CAGGCCAGG	TATGGTTAT	GGAGGAGG	AGATGAGGA	
2110	2120	2130	2140	2150	2160	2170	2180	2190	2200
TAAAGTGGC	TTCTCTCCT	CGTGGATGT	GTATTTTGA	GGGATTTGT	TAGTAGTAG	AGCGCGGCC	CGGGACCGA	TGCCGATGT	GCCGCTTTC
2210	2220	2230	2240	2250	2260	2270	2280	2290	2300
GTGGCGTTT	CCCGGGGGG	TTTTGGTTG	GCGCTTGGG	GGGATGGCAT	GCGCGCGCT	GCGCTGACG	CSCGCGCGC	GCGACGATC	
2310	2320	2330	2340	2350	2360	2370	2380	2390	2400
TCGTCTCCG	CSCGCGCGA	CGTAGCTTA	GGTGGTGTG	TTCCGCGCC	GGCGCGGAT	TGTTCCATG	GATCGATT	GCGCCACCC	TCGCCCGGC
2410	2420	2430	2440	2450	2460	2470	2480	2490	2500
TCTGTGCGG	TGTTGCGCT	CTCTCGCGG	GTTTGTCTT	GTCCGCTTG	TCAGCCGCG	ACGGGGCGC	GACATTTGG	GATGTAGCC	TGACGTTGC
2510	2520	2530	2540	2550	2560	2570	2580	2590	2600
GCCCTTCCG	TTGATGATG	ATGATGATG	TATGATTTT	TTTTGTCTG	AGGAAATTT	TGGGAAATG	TTGTGTGTC	AGCGCGTCT	GACCGGGAG
2610	2620	2630	2640	2650	2660	2670	2680	2690	2700
TATAAGACT	GCCTGCACG	GCGCGGGTG	AAACCGCGC	GGAGCGGGC	GTCCGTGGC	TTCTTCTGT	GCCCGCGGA	GGACAGGTT	GTCCGCGCC
2710	2720	2730	2740	2750	2760	2770	2780	2790	2800
GCCCGACCG	GCCCGACCG	CAGCACTAC	CAGCACTAC	CTAGCGCGC	CTAGCGCGC	TCACCGCGC	CCACTACCG	GCCGACCGC	GACCGCTCG
2810	2820	2830	2840	2850	2860	2870	2880	2890	2900
GCCCTTCCG	CGCTGGCTG	CSCGCGCGC	GCGCGCGCC	GCCCGACCG	CSCAGGTCA	GCGCGCGCC	TGATCCCGA	ACGGACCGA	ACGGACCGA

(EMBL) length: 2900 start: 1 - end: 2900 ( 2900 )
(EMBL) length: 2900 current pos: 2901

# 真核生物ではゲノムDNAの一部が遺伝子をコードしている

DNA

-----TGGCATATGGCATGCAGACTGGCATTGGCATGCGTCTGCATGCCATTGGTTTCTGAGGTCGCTGTGAACCTTATTTGGGCACAGACATGGCATGCAGACTGGCAC-----  
-----ACCGTATACCGTACGTCTGACCGTAACCGTACGCAGACGTACGGTAACCAAAGAGTCCAGCACACTTTGAATAAACCCGTGTCTGTACCGTACGTCTGACCGTG-----



mRNA

翻訳

タンパク質



# エキソンとイントロンの境界

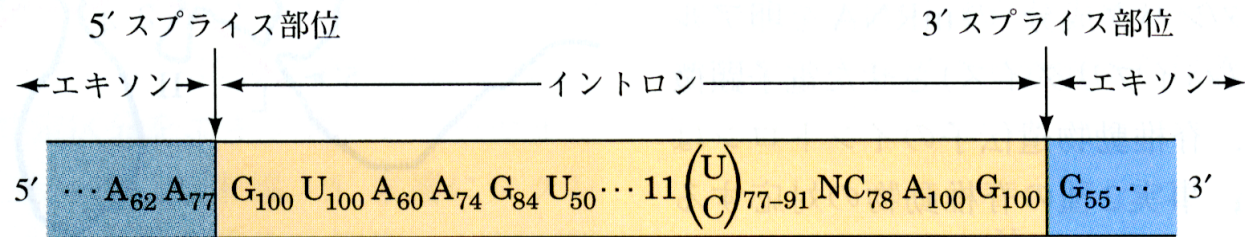
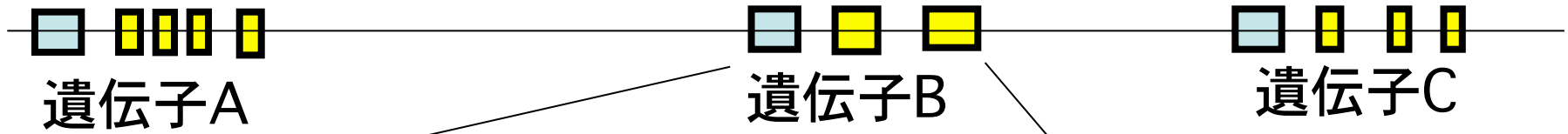


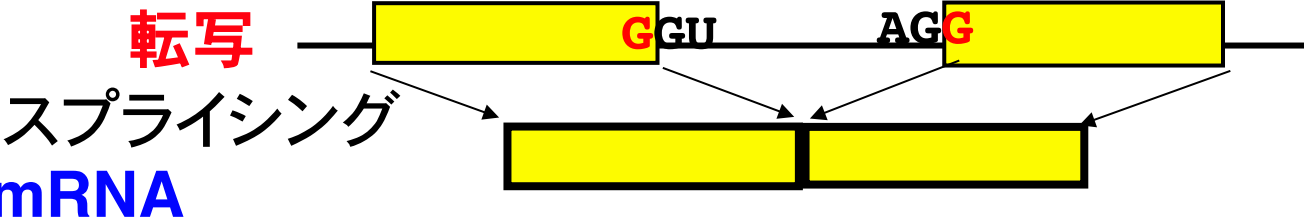
図 25・19 真核細胞 mRNA 前駆体のエキソン・イントロン接合部にあるコンセンサス配列 各塩基の添え数字はその塩基の出現頻度(%). 3' スプライス部位のすぐ上流にピリミジンに富む 11 nt 区域がある [R.A. Padgett, P.J. Grabowski, M.M. Konarska, S.S. Seiler, P.A. Sharp, *Annu. Rev. Biochem.* **55**, 1123 (1986) のデータに基づく]

# DNAからRNAへそしてタンパク質へ



DNA ---GTCTGC**ATGCCATTGGTTTCTGAGGTCGTGTGA**AACA**ATAA**AGGGCA  
 DNA ---CAGACG**TACGGTAACCAAAGACTC**CAGCACACTTTGTTATTTCCCGT  
 PolyA認識サイト

RNA  
**CAP** GTCUGC**AUGCCA**UUGGUUUCUGAG**GUCGUGUGA**AACAAUAAAGGGCAAAAAA



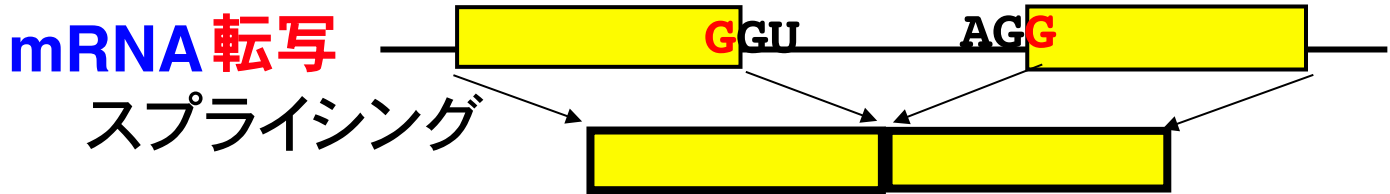
成熟mRNA **CAP**GTCUGC**AUGCCA**UUG**GUCGUGUGA**AACUUAUUUGGGCAAAAAAA  
 Met Pro Leu Val Val Stop

# DNAとcDNA



DNA ---GTCTGC**ATGCCATTGGTTTCTGAGGTCGTGTGA**AACA**AATAA**AGGGCA  
 DNA ---CAGACG**TACGGTAACCAAAGACTC**CAGCACACTTTGTTATTTCCCGT  
PolyA認識サイト

RNA **CAP**GTCUGC**AUGCCAUUG**GUUUCUGAG**GUCGUGUGA**AACAAUAAAGGGCAAAAAA



**成熟mRNA** **CAP**GTCUGC**AUGCCAUUGGUCGUGUGA**AACAAUAAAGGGCAAAAAAAA

**cDNA**      CAGACG**TACGGTAACCAAAGACTC**TTGTTATTTCCCGTTTTTTTTT  
 GTCTGC**ATGCCATTGGTCGTGTGA**AACAATAAAGGGCAAAAAAAA

DNA配列のどこが遺伝子なのか？ORFを検出してみる

(1) cDNA情報が有るとき

DNA配列とcDNA配列を比較する

(2) cDNA情報が無いとき

遺伝子予測ソフトで遺伝子領域を推定する

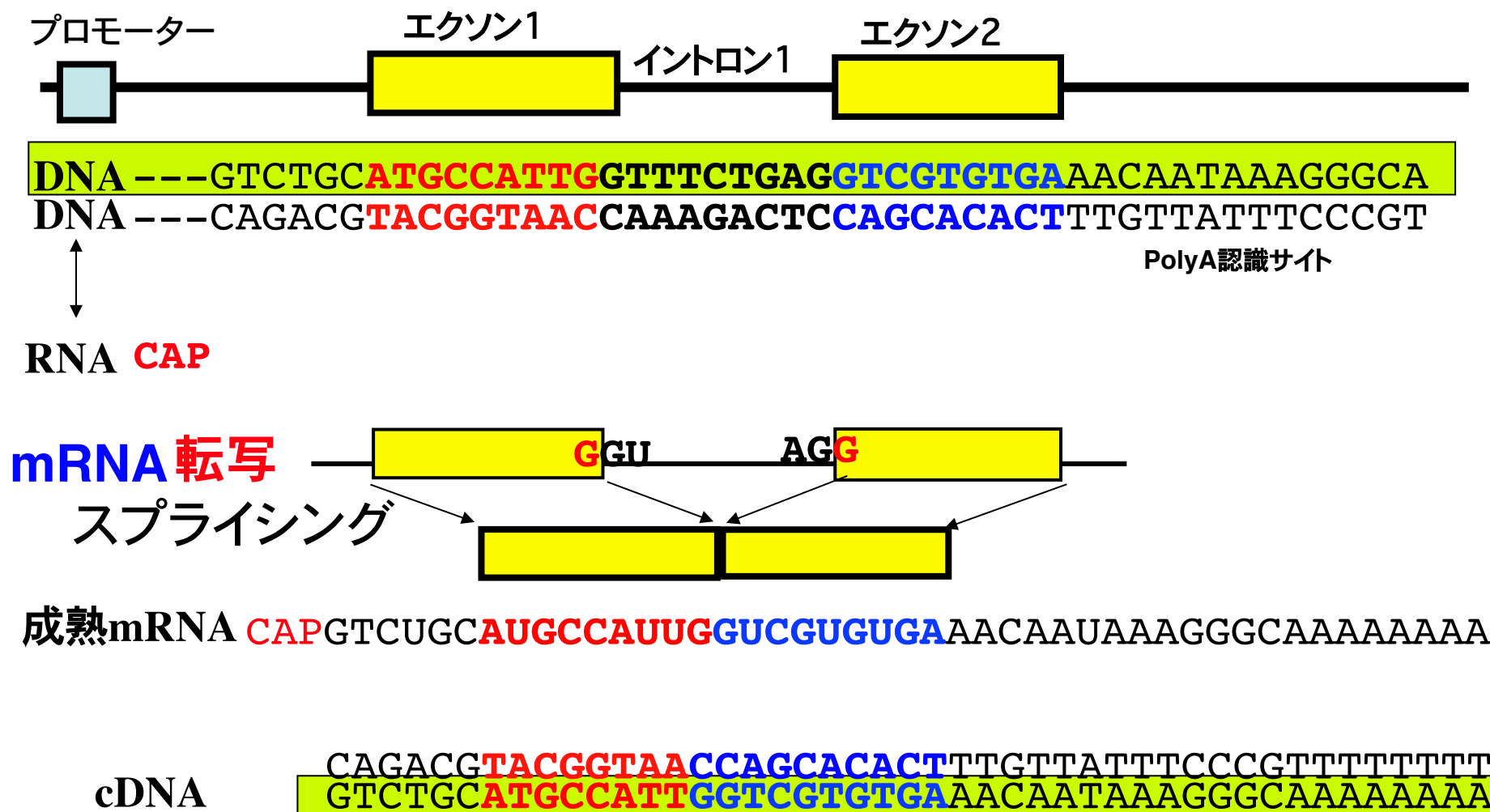


## 実際に遺伝子を探す

- ・マニュアルで遺伝子がどこにあるのか調べてみる。

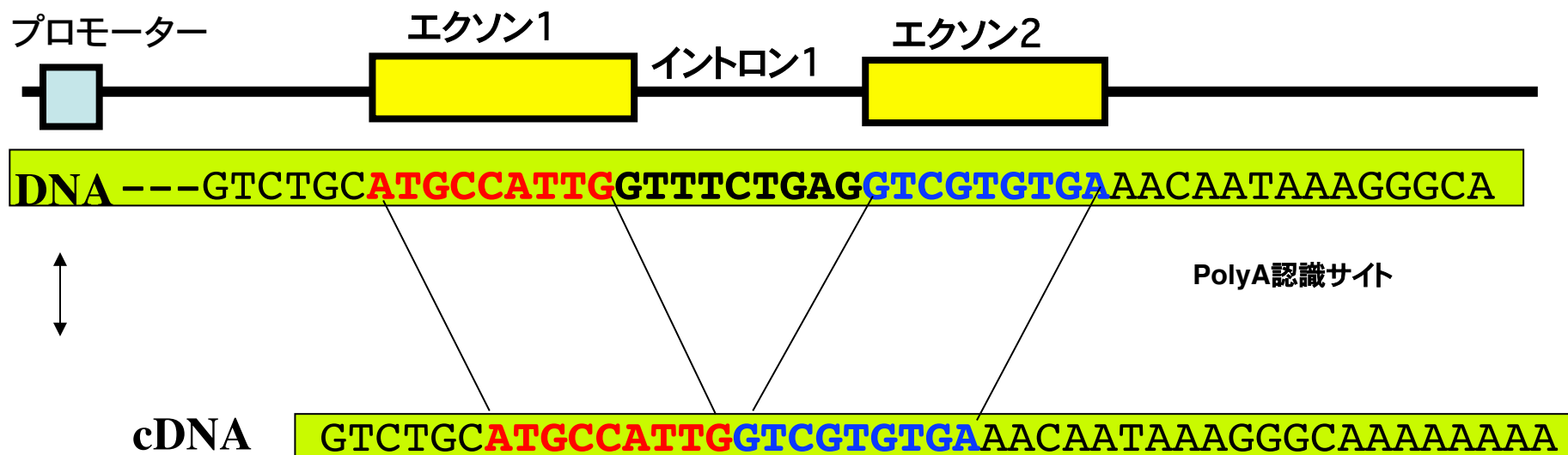
## cDNA情報が有るとき.2

### DNA配列とcDNA配列を比較する



# cDNA情報が有るとき.3

## DNA配列とcDNA配列を比較する



GA20oxidase2-genomic DNA

ID XX  
SQ Sequence 2900 BP; 608 A; 864 C; 784 G; 644 T;

## ある配列のGenomicDNA

```

10      20      30      40      50      60      70      80      90      100
tctcccctgt taacaatacc caaccctctc gccacagacg ctgcccctgc acacacacac acactaacac tcacacaacg tctaacatca ctcccagctc
110     120     130     140     150     160     170     180     190     200
acacagagct cacttctaat ctcaaatctc atggtggcgc agcaccaccg gccaccacag ccgcaccaac caaccgccat ggactaacac gccgcctctg
210     220     230     240     250     260     270     280     290     300
gcattggcgc cccggcggcg gccggcggtt gcgaccagag gatggagccc aagatcccgg agccatttgt gtggcgaac gccgacgcga gccgcggcct
310     320     330     340     350     360     370     380     390     400
ggcggcggag ctggacaatgc ccgtggtcga cgtggcgctg ctcgcgacg gcgacgcaga ggggcctgac cgcgcgcggg cgcagctggc cgcgcgctgc
410     420     430     440     450     460     470     480     490     500
gccacgcacg ggttattcca ggtgtccgag cacggcctcg acgcccctct ggcgcgcgcc gcgctgacg gcgcccacga cttcttcgcg ctcccgcctg
510     520     530     540     550     560     570     580     590     600
ccgagaagcg ccgcgcgcgc cgcgtcccgg gccaccgtgc ccgctacacc agcgcgccac ccgaccgctt cgcctccaaq ctcccattga agggagacct
610     620     630     640     650     660     670     680     690     700
ctccttcgag ttccacgacc gcgcgcgcgc ccccgctgtc gccacacctc tctccagcac cctgcgcccc gacttcgagc caatggggta attaaaacga
710     720     730     740     750     760     770     780     790     800
tgggtggcag cattgcattt caaattccaa acaaatccaa aaacacacga ccgagattat gctgaatca aacggctttg tgcgcgcagg aggygttacc
810     820     830     840     850     860     870     880     890     900
agaagtactg cagaggagtg aaggagctgt cgcctgacgt catggaacct ctggagctga gccctggcgt ggagcgagac tacctacagg actctctgac
910     920     930     940     950     960     970     980     990     1000
ggacagcagc taactaatgc ggtgcaacta ctaccgcgca tgccccggag ccggagcgac gctgcgacag gcccccagct gccaccacac cgcctctaac
1010    1020    1030    1040    1050    1060    1070    1080    1090    1100
atctctctcc aggacgagct ccgcgcgcctc gaggctctcg tcgacggcga atggcgcgcc gtcagcccgc tcccgggcgc catgtctacc aacatggcgc
1110    1120    1130    1140    1150    1160    1170    1180    1190    1200
acaccttcac ggtaaacctc ctccattctt cctctcctct gttctctctt gcttcgaagc aacagaacaa gtaattcaag cttttttttc tctctcgcgc
1210    1220    1230    1240    1250    1260    1270    1280    1290    1300
gaaattgacg agaaaataaa gatctgtgta ggggcggggc tttcagctga aagcgggaaq aaaccgacct gacgtgattt ctctgttcca atcaacaaca
1310    1320    1330    1340    1350    1360    1370    1380    1390    1400
atggaaatgcc caactctctc atgtgttatg atttatctca catcttatag ttaataggag taagtaacaa gctacttttt teattatata gttgttttga
1410    1420    1430    1440    1450    1460    1470    1480    1490    1500
tttttttttt ttaaagtttt tttagtttta tccaaattta ttgaaaacct tagcaacgtt tataatacca aattagctctc atttagttta atattgtata
1510    1520    1530    1540    1550    1560    1570    1580    1590    1600
tattttgata atatatattt gttatattaa aaatattact atatttttct ataacatta ttaaagccca tttataatat aaaaatggaq gagtatttaa
1610    1620    1630    1640    1650    1660    1670    1680    1690    1700
tatggatctc ccccgcacatg agaataatttt ccgatgggtt gacgacgcca tghtaagcttc ggtggcccctg gacggccaga ggtgccaaca gccacgtcca
1710    1720    1730    1740    1750    1760    1770    1780    1790    1800
acaaccccctg ggtccccecc taacactcca aacagtagtg agtagtgctc cgtcgcgttt tagtatttga tgacaacaa agtgtgattt gagttagcca
1810    1820    1830    1840    1850    1860    1870    1880    1890    1900
ccaccaactt gccacagcgc aataaatttt ggttccattc tcgcccagta cttccatctc tagctcctaac tctctctag cgatgtaacg ggaataattc
1910    1920    1930    1940    1950    1960    1970    1980    1990    2000
atcatccgta tataaacctg ttgttatag ttaatttctc atataaactc ataacagtat acattttaaa agaaaacaaa attaggataa caagcccctg
2010    2020    2030    2040    2050    2060    2070    2080    2090    2100
ctctatccca tcaatggcac ttggaaggac cagactcgtt catgccatgc caagccaaga tatggtttat ggaagatag agaagaggag agatgagaga
2110    2120    2130    2140    2150    2160    2170    2180    2190    2200
taagcatgag ttctctctct cgttgagatg gtattttgga gggatttgtg tagtagtagc agcggcgcgc cggggacaga tgcgatgtgt gccgccttcg
2210    2220    2230    2240    2250    2260    2270    2280    2290    2300
gtggcgtttt cccggggggg ttttggtttt gcccttgggg gggatggcat gccgcgcgct gccgcctgac gccacacaca cgcgcgagca cgcacgtacc

```

(EMBL) length: 2900 current pos: 401

GA20oxidase2-cDNA

ID C20 oxidase for a candidate gen of sd1  
XX  
SQ Sequence 1170 BP; 171 A; 455 C; 152 T; 392 G.

## ある配列のcDNA

```

10      20      30      40      50      60      70      80      90      100
atggtggcgc agcaccaccg gccaccacag ccgcaccaac caaccgccat ggactaacac gccgcctctg gcattggcgc cccggcggcg gccgcggctg
110     120     130     140     150     160     170     180     190     200
gccaccctgag gatggagccc aagatcccgg agccatttgt gtggcgaac gccgacgcga gccgcggcctc ggcggcggag ctggacatgc ccgtggtcga
210     220     230     240     250     260     270     280     290     300
cgtggcgctg ctcgcgacg gcgacgcaga gggcctgagc cgcgcgcggg ccgaggtggc cgcgcgctgc gccacgcacg ggttattcca ggtgtccag
310     320     330     340     350     360     370     380     390     400
caaggcctg acgcgcctc ggcgcgcgc gcgctgacg gccgcagcga cttatttcgc ctcccgcctg ccgagagcgc ccgcgcgcgc cgcgtcccgg
410     420     430     440     450     460     470     480     490     500
gccaccgtgc ccgctaacac agcgcgccac ccgaccgctt cgcctccaaq ctcccattga agggagacct cctctctgag ttccacgacc gccgcgcgcg
510     520     530     540     550     560     570     580     590     600
ccccctgtgc gccacacctc tctccagcac cctgcgcccc gacttcgagc caatggggta ggtgtaccag aagctctgag agggagatgaa ggcgctgtctg
610     620     630     640     650     660     670     680     690     700
ctgacatgta tggaaacctt ggagctgagc ctggcgctgg agcggagcta ctacaggag tctctcggcg acagcagctc aatctatgag tgcacacctc
710     720     730     740     750     760     770     780     790     800
accgccatgc cccggagcgc gaggcggcgc tcggcagcgc ccgcacagtc gcccccacgc cctccacctc cctctccag gacgagctag gccgcctgta
810     820     830     840     850     860     870     880     890     900
ggctctctct gacggcgaat ggcgcgccct cagcccagtc cccggcgcca ttgtatcaaa catcggcagc accttcattg ccgctgtcga cgggaggtat
910     920     930     940     950     960     970     980     990     1000
aagagctgac tgcaacaggc ggtgtgtaac cagcggcggg agcggcggtc gctggccttc tctctctgag ccgaggagga cagggtgtgt ccgcgcgcgc
1010    1020    1030    1040    1050    1060    1070    1080    1090    1100
cagagcgcgc caagcgcagc cactaccggg acttcaactg ggcccagctc atgccttca ccgagcgcga ctaccgcgc gccaccgcca cgtctgagcgc
1110    1120    1130    1140    1150    1160    1170    1180    1190    1200
cttccagcgc ttgctcgcgc cgcgcgcgc ccgacgcgcc gccagcgggc aggtataggc ggccagctga

```

(EMBL) length: 1170 current pos: 1171

GA20oxidase2-genomic DNA

ID XX  
SQ Sequence 2900 BP; 608 A; 864 C; 784 G; 644 T;

### ある配列のGenomicDNA

```

10      20      30      40      50      60      70      80      90      100
tctcccgtg tacaaatac caaccctct gccacagac ctgcccctg acacacacac acactaacac tcacacaagc tctaacata ctcccagca
110     120     130     140     150     160     170     180     190     200
acacagact cacttatac ctcaaatct atggtggcg agcaccacc gccaccacag ccgaccacac caccgccat ggactaac gccgcctg
210     220     230     240     250     260     270     280     290     300
gcattggcg cccggcggg gcggcggtg ggcactgag gatggagcc aagatcccg agccattgt gtggcgaac ggcgacgca ggcggcgct
310     320     330     340     350     360     370     380     390     400
ggcggcgag ctggacatg ccgtggtca cgtggcgct ctcgcgacg gcgacgcga ggggcctgc cgcgccggg cgcaggtgc cgcgcgtgc
410     420     430     440     450     460     470     480     490     500
gccacagac ggttattca ggtgtccag cacggctcg accgccctt ggcgcgccc gcgctgacg gcgaccagca cttctcgc ctcccctg
510     520     530     540     550     560     570     580     590     600
ccgagaagc ccgcgcgcg ccgctccgg gccactgac cggctacac agcgcacac ccgaccgctt cgcctcaag ctcccattg agggagcct
610     620     630     640     650     660     670     680     690     700
ctcttcgac ttccagacc gcgcgcgac ccgcctgct gcgcactt ctccagacc ccttcgccc gacttcgac caatgggta attaaagca

```

EMBL length: 2900 current pos: 401

GA20oxidase2-cDNA

ID C20 oxidase for a candidate gen of sd1  
XX  
SQ Sequence 1170 BP; 171 A; 455 C; 152 T; 392 G.

### ある配列のcDNA

```

10      20      30      40      50      60      70      80      90      100
atggtggcg agcaccacc gccaccacag ccgaccacac caccgccat ggactaac gccgcctg gcattggcg cccggcggg gcggcggtg
110     120     130     140     150     160     170     180     190     200
gcgactgag gatggagcc aagatcccg agccattgt gtggcgaac ggcgacgca ggcggcgct gcgggggag ctggacatg ccgtggtca
210     220     230     240     250     260     270     280     290     300
cgtggcgct ctcgcgacg gcgacgcga gggcctgc cgcgccggg cgcaggtgc cgcgcgtgc gccacgacg ggttattca ggttcagc
310     320     330     340     350     360     370     380     390     400
caccgctg accgcctt ggcgcgccc gcgctgacg gcgaccagca cttctcgc ctcccctg ccgagaagc ccgcgcgac cgcctccgg
410     420     430     440     450     460     470     480     490     500
gcacgtgt ccgctaac accgcgccg ccgaccgctt cgcctcaag ctcccattg agggagcct cctctcgc ttccagacc gcgcgcgac
510     520     530     540     550     560     570     580     590     600
ccccctgt gccactac tctccagac cctgcgccc gacttcgac caatgggag ggtgtacc aagtacgac agggatgaa ggagctgtg
610     620     630     640     650     660     670     680     690     700

```

EMBL length: 1170 current pos: 1171

DNA ---GTCTGC**ATGCCATTG**GTTTCTGAG**GTCGTGTGA**AACAATAAAGGGCA

cDNA GTCTGC**ATGCCATTG**GTCGTGTGA**AACAATAAAGGGCA**AAAAAAAAA

GA20oxidase2-genomic DNA

ID XX  
SQ Sequence 2900 BP; 608 A; 864 C; 784 G; 644 T;

## ある配列のGenomicDNA

```

10      20      30      40      50      60      70      80      90      100
tcctccctgt tacaaatacc caaccctctc gccacagcag ctcgccctgc aacacacac aaactaacac tcaacaagc tctcaactca ctcaccgtca
110     120     130     140     150     160     170     180     190     200
aacacagcgt caactctcat ctccaatctc atggtggcag agcaaccacc gccaccacag ccgaccacac caccgcccat ggaactcaac gccgcctcgt
210     220     230     240     250     260     270     280     290     300
gcattgcgcg cccggcgggc gccggcggtg gccacgttag gatggagccc aagatcccag agcattctgt gtcggcagac gccgacgcga gccacggcgtc
310     320     330     340     350     360     370     380     390     400
gccggcgggg ctggacaatg ccgtgtgtga cgtggcgctg ctcgcgagc gccagcccca gggcgtgcgc cgcgcgcggg cgcagggtgc cgcgcggtgc
410     420     430     440     450     460     470     480     490     500
gccacgcacg ggtttctcca ggtgtaccag caaggcctag acgcgcctct ggagcgcgcc gccctagcag gccgcagcga cttcttccgc ctcaccgtag
510     520     530     540     550     560     570     580     590     600
ccgagaaagc ccgcgcgcgc cgcgtcccgc gccaccgtgc cggctaacac agcgcacca cgcaccgctt cgcctccaag ctccaatgga agggagacct
610     620     630     640     650     660     670     680     690     700
ctcacttgcc ttccacgacc gccgcgcgcg ccccgtctgc gccacgact tctaccgacc cctggccccc gacttcgcgc caatggggta attaaagca
710     720     730     740     750     760     770     780     790     800
tggtagcaga cattgcatct caaatccaaa acaaatccaa aacacacaga ccgagattat gctgaattca aacgcgtttg tgcgcgcagg aggggtgtcc
810     820     830     840     850     860     870     880     890     900
agaagctact cagaggagtg aaggagctgt cgtgacgat catggaactc ctygagctga gccctggcgt ggagcagacc tacataagcc agttcttgc
910     920     930     940     950     960     970     980     990     1000
ggacagcaga taactactgc ggtgcaacta ctaccgcgca tgcccggagc ccgagcagac gctggcacag gccccgcact gccaccacac cgcacctaac
1010    1020    1030    1040    1050    1060    1070    1080    1090    1100
atctctctcc agnacagcgt ccggcgccct gaggtccctg tcgacggcga atggcgcgcc gtaaccgcgc tcccgcgcgc catggtctat aacatggcg
1110    1120    1130    1140    1150    1160    1170    1180    1190    1200
aacacttaat ggtaaacctc ctctctact ctctctact gttctctct gttctctct gcttcgaagc aacagaaaca gtaattcaag ctttttttcc tctctgcgc
1210    1220    1230    1240    1250    1260    1270    1280    1290    1300
gaaattgacg agaaaaataa gatctgtgta gggcgggggc ttccagctga aagcgggaag aaaccgacct gacgtgattt ctcctgtcca atcaacaaca
1310    1320    1330    1340    1350    1360    1370    1380    1390    1400
atggaaatgcc ccactctctc atgtgttatg atttatctca catcttatag ttaataggag taagtacaac gctacttttt teatattata gttctgttga
1410    1420    1430    1440    1450    1460    1470    1480    1490    1500
tttttttttt ttaagatttt tttagtttta tcaaaaatta ttgaaaaact tagcaacgtt tataatacca aattagcttc atttagttta atattgtata
1510    1520    1530    1540    1550    1560    1570    1580    1590    1600
tattttgata atatatattat gttatattaa aatatattact atattttct ataacaatta ttaaaagcca tttataat ataaatggaag gagtattaa
1610    1620    1630    1640    1650    1660    1670    1680    1690    1700
tatggacttc ccccgacatg agaatttttt ccgatgggtg gacgacgcca tghtaagctc ggtagggcct gacggcagca ggtgccacaa gccacgtcca
1710    1720    1730    1740    1750    1760    1770    1780    1790    1800
aacaccctgt ggtccccccc taacactcca aacagtagtg agtagtgtct cgtcgcgttt tagtatttga tgcaaacaaa agtgtgattt gagtttagca
1810    1820    1830    1840    1850    1860    1870    1880    1890    1900
ccaccacatt gccacagcag acatacattt gttctcaatc tcgcccagtea cttccatctc tagtctaac tectatctag cagttaagc ggataatttc
1910    1920    1930    1940    1950    1960    1970    1980    1990    2000
atcatcctga tataaacctg tttgttatag ttaatttctc atataatact ataacagtat acattttaaa agaaaacaaa attaggataa acagaccctg
2010    2020    2030    2040    2050    2060    2070    2080    2090    2100
ctctatccca tcaatggcac ttgnaaggac cagactcggt catgccatgc caagccaaga tatgggttat ggaagagtag agaagaggag agatgagaga
2110    2120    2130    2140    2150    2160    2170    2180    2190    2200
taagcattgc ttctctctct cgttggatgt gatttttggg gggattttgt tagtagtagc agcgcgcgca cggggacaga tgcggatggt gccgcctttc
2210    2220    2230    2240    2250    2260    2270    2280    2290    2300
gtggcgtttt cccggggggg ttttgatttg gccgtgggg gggatggcat ggcgaggcgt gccgctgca cccaccacaa cgcgcgcgca agcaagctag

```

(EMBL) Length: 2900 current pos: 401

GA20oxidase2-cDNA

ID C20 oxidase for a candidate gen of sd1  
XX  
SQ Sequence 1170 BP; 171 A; 455 C; 152 T; 392 G;

## ある配列のcDNA

```

10      20      30      40      50      60      70      80      90      100
atggtggcgc agcaaccacc gccaccacag ccgaccacac caccgcccat ggaactcaac gccgcctcgt gcattgcgcg cccggcgggc gccggcggtg
110     120     130     140     150     160     170     180     190     200
gccacgttag gatggagccc aagatcccag agcattctgt gtcggcagac gccgacgcga gccacggcgtc cgtgtgtcga cgtgtgtcga
210     220     230     240     250     260     270     280     290     300
cgtgtgtcgt ctcgcgagc gccagcccca gggcgtgcgc cgcgcgcggg cgcagggtgc cgcgcggtgc gccacggcgt gcattctcca ggtgtaccag
310     320     330     340     350     360     370     380     390     400
caccgcgtag acgcgcctct ggagcgcgcc gccctagcag gccaccgaga cttcttccgc ctcaccgtag ccgagaaagc ccgcgcgcgc cgcctccagg
410     420     430     440     450     460     470     480     490     500
gcaacgtgta cggctaacac agcgcacca cgcaccgctt agactccaag ctccaatgga agggagacct ctcacttgcc ttccacgacc gccgcgcgc
510     520     530     540     550     560     570     580     590     600
ccccgtctgc gccacgact tctaccgacc cctggccccc gacttcgcgc caatggggag ggtgtaccag aagctactgc agggagtgaa ggagcgtctg
610     620     630     640     650     660     670     680     690     700
ctgacagtaa tggaaacctc ggagctgagc ctgggcgctg agcagagcta ctacagggag ttctctgcgc acagcagcct aaatctgcgc tgcaactact
710     720     730     740     750     760     770     780     790     800
accgcagctg cccggcgagc gacgagcaga tcggcagcgg ccgcacatgc gaccaccacg cctcaacctc ctctctcag gacgagctag gccggctaga
810     820     830     840     850     860     870     880     890     900
ggtctctctc gacggcgaa cggcgcacct caccgcagc cccggcgcca tggtaataca catcggcagc acctcaatgg cgcctctaga cgggaggtat
910     920     930     940     950     960     970     980     990     1000
aagagctgct tgcacaggcc ggtgtgtgaa cagcgcggcg agcggcggtc gctggcgttc ttctgtgccc cgcgggagga caggggtgtg ccgcgcgcgc
1010    1020    1030    1040    1050    1060    1070    1080    1090    1100
cgcgcgcgcg cagcgcagag caactaccgg acttcaactg ggccgacctc atgctctca cgcagcagca ctaccgcgcc gacaccgcaa cgcctagacc
1110    1120    1130    1140    1150    1160    1170    1180    1190    1200
cttcaagcgc tggctgcgc cgcgcgcgcg ccagcgcgcc gccagcgcgc aggtcgagcc ggccaagcga

```

(EMBL) Length: 1170 current pos: 1171

# マニュアルで配列を解析してみよう

The image shows a software interface for sequence analysis, likely a BLAST or alignment tool. It features a menu bar at the top with options: "Finder", "ファイル", "編集", "表示", "移動", "ウインドウ", "ヘルプ". Below the menu is a toolbar with various icons for navigation and editing.

There are two main windows displayed side-by-side:

- GA20oxidase2-genomic DNA:** This window shows a sequence of 2900 base pairs. The sequence is displayed in a grid format with position markers every 100 bases (10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 260, 270, 280, 290, 300, 310, 320, 330, 340, 350, 360, 370, 380, 390, 400, 410, 420, 430, 440, 450, 460, 470, 480, 490, 500, 510, 520, 530, 540, 550, 560, 570, 580, 590, 600, 610, 620, 630, 640, 650, 660, 670, 680, 690, 700, 710, 720, 730, 740, 750, 760, 770, 780, 790, 800, 810, 820, 830, 840, 850, 860, 870, 880, 890, 900, 910, 920, 930, 940, 950, 960, 970, 980, 990, 1000, 1010, 1020, 1030, 1040, 1050, 1060, 1070, 1080, 1090, 1100, 1110, 1120, 1130, 1140, 1150, 1160, 1170, 1180, 1190, 1200, 1210, 1220, 1230, 1240, 1250, 1260, 1270, 1280, 1290, 1300, 1310, 1320, 1330, 1340, 1350, 1360, 1370, 1380, 1390, 1400, 1410, 1420, 1430, 1440, 1450, 1460, 1470, 1480, 1490, 1500, 1510, 1520, 1530, 1540, 1550, 1560, 1570, 1580, 1590, 1600, 1610, 1620, 1630, 1640, 1650, 1660, 1670, 1680, 1690, 1700, 1710, 1720, 1730, 1740, 1750, 1760, 1770, 1780, 1790, 1800, 1810, 1820, 1830, 1840, 1850, 1860, 1870, 1880, 1890, 1900, 1910, 1920, 1930, 1940, 1950, 1960, 1970, 1980, 1990, 2000, 2010, 2020, 2030, 2040, 2050, 2060, 2070, 2080, 2090, 2100, 2110, 2120, 2130, 2140, 2150, 2160, 2170, 2180, 2190, 2200, 2210, 2220, 2230, 2240, 2250, 2260, 2270, 2280, 2290, 2300). The sequence is shown in a single line with line wrapping.
- GA20oxidase2-cDNA:** This window shows a sequence of 1170 base pairs. The sequence is displayed in a grid format with position markers every 100 bases (10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 260, 270, 280, 290, 300, 310, 320, 330, 340, 350, 360, 370, 380, 390, 400, 410, 420, 430, 440, 450, 460, 470, 480, 490, 500, 510, 520, 530, 540, 550, 560, 570, 580, 590, 600, 610, 620, 630, 640, 650, 660, 670, 680, 690, 700, 710, 720, 730, 740, 750, 760, 770, 780, 790, 800, 810, 820, 830, 840, 850, 860, 870, 880, 890, 900, 910, 920, 930, 940, 950, 960, 970, 980, 990, 1000, 1010, 1020, 1030, 1040, 1050, 1060, 1070, 1080, 1090, 1100, 1110, 1120, 1130, 1140, 1150, 1160, 1170, 1180, 1190, 1200). The sequence is shown in a single line with line wrapping.

At the bottom of each window, there is a status bar. For the genomic DNA window, it shows "(EMBL) Length: 2900 current pos: 481". For the cDNA window, it shows "(EMBL) Length: 1170 current pos: 1171".

Finder ファイル 編集 表示 移動 ウィンドウ ヘルプ

GA20oxidase2-genomic DNA

10 20 30 40 50 60 70 80 90 100

TCTCCCTGT TACAAATACC CCACCCCTCT GCCCAGACAG CTCGCCCTGC ACACACACAC ACACCTCACAC TCACACACGC TCTCAACTCA CTCCCCTCTA

110 120 130 140 150 160 170 180 190 200

ACACAGCGCT CACTTCTCAT CTCCAATCTC ATGTGGTCCG AGCACCCACG GCCACCCACG CCGCACCCAC CACCCGCCAT GGACTCCACC GCCGGCTCTG

210 220 230 240 250 260 270 280 290 300

GCATTGCCGC CCCGGCGCGC GCGCGGTGT GCGACTGAG GATGGAGCCC AAGATCCCGC AGCCATTCTG GTGCCCGAAC GCGCACCCGA GCCCGCGCTC

310 320 330 340 350 360 370 380 390 400

GCGCGCGAG CTGGACATGC CCGTGGTCTG CGTGGCGGTG CTCGGCGACG GCGACGCCGA GGGGCTGCCG CCGCCCGCGG CGCAGGTGGC CCGCCGCTGC

410 420 430 440 450 460 470 480 490 500

GCCACGCACG GGTTCCTCCA GGTGCCGAGC CACGGCGCTC ACGCCGCTCT GGCGCGCGCC GCGCTCGACG GCGCCAGCGA CTTCTCCCGC CTCCCCTCTG

510 520 530 540 550 560 570 580 590 600

CCGAGAGCGC CCGCGCGCGC GCGGTCCCGG GCACCCTGCT CGGTACACCC AGCGCCACCG CCGACCGCTT CGCCTCCAGC CTCCCATGGA AGGAGACCTT

610 620 630 640 650 660 670 680 690 700

CTCCTTCGGC TTCACGACCC GCGCGCGCGC CCGCGTGCCT GCGGACTACT TCTCCAGCAC CCTCGGCCCC GACTTCGCGC CAATGGGGTA ATTAARACGA

710 720 730 740 750 760 770 780 790 800

TGTTGGACGA CATTGCATT CAARATTCARA ACAARATTCARA AACACACCCGA CCGAGATTAT GCTGATTTCA AACCGTITTG TGCGCGCAGG AAGGTGTACC

810 820 830 840 850 860 870 880 890 900

AAGATGACTG CAGAGAGATG AAGAGGCTGT CGCTGACGAT CATGGAACTC CTGGAGCTGA GCCTGGGGCT GGAGCGAGGC TACTACAGGG AGTTCTTCGC

910 920 930 940 950 960 970 980 990 1000

GGACAGCAGC TCARATCATG GGTGCACATC CTACCCCGCA TGCCCGGAGC CCGAGCGGAC GCTCGGACG GGGCCGACG GCGACCCACG CGCCCTCAAC

1010 1020 1030 1040 1050 1060 1070 1080 1090 1100

ATCCTCTCTC AGGACGACGT CCGCGGCTCT GAGGTCTCTG TCGAGCGGCA ATGGCGCCCC GTCAGCCCGC CATGGTCATC AACATCGCGC

1110 1120 1130 1140 1150 1160 1170 1180 1190 1200

ACACCTTCAT GGTAAACCAT CTCCTATTCT CCTCTCCTCT GTTCTCCTCT GCTTCGAGAC AACACAGACA GTARITTCAG CTTTTTTTTT TCTCTCGCGC

1210 1220 1230 1240 1250 1260 1270 1280 1290 1300

GAAATGACG AAAAAATATG ATCTGTGTA GGGCGGGGCG TTTAGCTGA AAGCGGGAGG AARCCGACCT GCGGTATT CTCTGTTCRA ATCAACACCA

1310 1320 1330 1340 1350 1360 1370 1380 1390 1400

ATGGATGCCC CCACCTCTCC ATGTGTTATG ATTTATCTCA CATCTTATAG TTAATAGGAG TAAGTARACGA GCTACTTTTT TCATATTATA GTTCTGTTGA

1410 1420 1430 1440 1450 1460 1470 1480 1490 1500

TTTTTTTTTT TTAAGTTTTT TTAGTTTTTA TCCAAATTTA TTGAARAAT TAGCAACGTT TATATATCCA ARTTAGTCTC ATTTAGTTTA ATATTGTATA

1510 1520 1530 1540 1550 1560 1570 1580 1590 1600

TATTTTGTAT ATATATTTAT GTTATATTAR AARATTTACT ATATTTTCT ATAAACATTA TTAARAGCCA TTTATATAT AARATGGAG GAGTARITTA

1610 1620 1630 1640 1650 1660 1670 1680 1690 1700

TATGGATCTC CCCGACATG AAGATATTTT CCGATGGTGT GACGACGCCA TGTAGCTTC GGTGGGCTGC GACGGCCAGA GGTGCCARCA GCCACGTCRA

1710 1720 1730 1740 1750 1760 1770 1780 1790 1800

ACACCCCTCG GGTCCCGCCC TACACCTCCA AACAGTAGTG AGTAGTGTCT CGTCGCGTTT TAGTATTTGA TGACAAACAA AGTGTAGTT GAGTTAGCCA

1810 1820 1830 1840 1850 1860 1870 1880 1890 1900

CCACCAACTT GCACAGAGGC ACATACATTT GTGTCCATTC TCGCCAGTCA CTTCCATCTC TAGCTTACG TCCTATCTAG CGATGTAGC GGATARTTTC

1910 1920 1930 1940 1950 1960 1970 1980 1990 2000

ATCATCCGTA TATAACCTG TTTGTTATG TTAATTTCT ATATAACT ATACAGTAT ACATTTTAAA AAAAAACAA ATTAGATARA ACAGCCCTG

2010 2020 2030 2040 2050 2060 2070 2080 2090 2100

CTCCTATCCA TCCATGGCAC TTGGAGGAC CAGACTCGGT CATGCCATGC CAAGCCAGCA TATGGTTTAT GGAGAGTAG AAAAAAGAGG AAGTARAGGA

2110 2120 2130 2140 2150 2160 2170 2180 2190 2200

TAGCATGCG TTCTCTCTCT CGTGGATGT GTATTTTGGG GGGATTTGTG TAGTAGTAGC AGCGCGCCCG CCGGGACCGA TCGGATGGT GCGCTTTTCG

2210 2220 2230 2240 2250 2260 2270 2280 2290 2300

GTGGCGTTT CCCGGGGGGG TTTTGGTTTG GCGCTTGGGG GGGATGGCAT GGCGCGCGCT GCGCGCGCAC GCGCGCGCGC GCGCAGTAGC

2310 2320 2330 2340 2350 2360 2370 2380 2390 2400

TCGTCTGCG CCGCGCGCGA CCGTAGCTTA GGGTGGTGTG TTCCGCGCGC GGGCGGGAT TGTTCATGCT GCGTCAATTT GCGCGCACCC TCGCCCGCGC

2410 2420 2430 2440 2450 2460 2470 2480 2490 2500

TCTTGTGCG TCGTGGGCT CTCTCGCGCG GTTTGTCTCT GTCGGTTGCG TACGCCGCGC ACGGGGCGAC GACATTTGCG GATGTAGCCC TGACGTTGTC

2510 2520 2530 2540 2550 2560 2570 2580 2590 2600

GCCCTCTCCG TTGATGATG ATGATGATG TATGATTTT TTTTGTCTG AAGGATTTG TGGGATTTG TTGTGTGTC AGCGCGTCTG GACCGGGAGG

2610 2620 2630 2640 2650 2660 2670 2680 2690 2700

TATAGAGCT GCCTGCACAG GCGCGGGTGG AACCGCGCGC GGGAGCGGCG GTCGCTGGCG TTTCTCTCTG GCCCGCGGGA GGACAGGGTG GTGCGCGCGC

2710 2720 2730 2740 2750 2760 2770 2780 2790 2800

GCGCGAGCG CCGCACGCCG CAGCATTACC CAGGATTTAC CTATGGCGAC CTCATCGGCT TCACCGCGAC CCACTACCGC GCCGACRCC GCACGCTCGA

2810 2820 2830 2840 2850 2860 2870 2880 2890 2900

GCCCTCTCAC GCGTGGCTCG CCGCGCGCGC CCGCGAGGCC GCGCGACGCG CCGAGGTGCG GCGCGCACCG TGATCCCGCA ACGGACCGAA ACGGACCGAA

(EMBL) Length: 2900 start: 1 - end: 2900 ( 2900 )

Untitled 3

10 20 30 40 50 60 70 80 90 100

TCTCCCTGT TACAAATACC CCACCCCTCT GCCCAGACAG CTCGCCCTGC ACACACACAC ACACCTCACAC TCACACACGC TCTCAACTCA CTCCCCTCTA

110 120 130 140 150 160 170 180 190 200

ACACAGCGCT CACTTCTCAT CTCCAATCTC ATGTGGTCCG AGCACCCACG GCCACCCACG CCGCACCCAC CACCCGCCAT GGACTCCACC GCCGGCTCTG

210 220 230 240 250 260 270 280 290 300

GCATTGCCGC CCCGGCGCGC GCGCGGTGT GCGACTGAG GATGGAGCCC AAGATCCCGC AGCCATTCTG GTGCCCGAAC GCGCACCCGA GCCCGCGCTC

310 320 330 340 350 360 370 380 390 400

GCGCGCGAG CTGGACATGC CCGTGGTCTG CGTGGCGGTG CTCGGCGACG GCGACGCCGA GGGGCTGCCG CCGCCCGCGG CGCAGGTGGC CCGCCGCTGC

410 420 430 440 450 460 470 480 490 500

GCCACGCACG GGTTCCTCCA GGTGCCGAGC CACGGCGCTC ACGCCGCTCT GGCGCGCGCC GCGCTCGACG GCGCCAGCGA CTTCTCCCGC CTCCCCTCTG

510 520 530 540 550 560 570 580 590 600

CCGAGAGCGC CCGCGCGCGC GCGGTCCCGG GCACCCTGCT CGGTACACCC AGCGCCACCG CCGACCGCTT CGCCTCCAGC CTCCCATGGA AGGAGACCTT

610 620 630 640 650 660 670 680 690 700

CTCCTTCGGC TTCACGACCC GCGCGCGCGC CCGCGTGCCT GCGGACTACT TCTCCAGCAC CCTCGGCCCC GACTTCGCGC CAATGGGGTA ATTAARACGA

710 720 730 740 750 760 770 780 790 800

TGTTGGACGA CATTGCATT CAARATTCARA ACAARATTCARA AACACACCCGA CCGAGATTAT GCTGATTTCA AACCGTITTG TGCGCGCAGG AAGGTGTACC

810 820 830 840 850 860 870 880 890 900

AAGATGACTG CAGAGAGATG AAGAGGCTGT CGCTGACGAT CATGGAACTC CTGGAGCTGA GCCTGGGGCT GGAGCGAGGC TACTACAGGG AGTTCTTCGC

910 920 930 940 950 960 970 980 990 1000

GGACAGCAGC TCARATCATG GGTGCACATC CTACCCCGCA TGCCCGGAGC CCGAGCGGAC GCTCGGACG GGGCCGACG GCGACCCACG CGCCCTCAAC

1010 1020 1030 1040 1050 1060 1070 1080 1090 1100

ATCCTCTCTC AGGACGACGT CCGCGGCTCT GAGGTCTCTG TCGAGCGGCA ATGGCGCCCC GTCAGCCCGC CATGGTCATC AACATCGCGC

1110 1120 1130 1140 1150 1160 1170 1180 1190 1200

ACACCTTCAT GGTAAACCAT CTCCTATTCT CCTCTCCTCT GTTCTCCTCT GCTTCGAGAC AACACAGACA GTARITTCAG CTTTTTTTTT TCTCTCGCGC

1210 1220 1230 1240 1250 1260 1270 1280 1290 1300

GAAATGACG AAAAAATATG ATCTGTGTA GGGCGGGGCG TTTAGCTGA AAGCGGGAGG AARCCGACCT GCGGTATT CTCTGTTCRA ATCAACACCA

1310 1320 1330 1340 1350 1360 1370 1380 1390 1400

ATGGATGCCC CCACCTCTCC ATGTGTTATG ATTTATCTCA CATCTTATAG TTAATAGGAG TAAGTARACGA GCTACTTTTT TCATATTATA GTTCTGTTGA

1410 1420 1430 1440 1450 1460 1470 1480 1490 1500

TTTTTTTTTT TTAAGTTTTT TTAGTTTTTA TCCAAATTTA TTGAARAAT TAGCAACGTT TATATATCCA ARTTAGTCTC ATTTAGTTTA ATATTGTATA

1510 1520 1530 1540 1550 1560 1570 1580 1590 1600

TATTTTGTAT ATATATTTAT GTTATATTAR AARATTTACT ATATTTTCT ATAAACATTA TTAARAGCCA TTTATATAT AARATGGAG GAGTARITTA

1610 1620 1630 1640 1650 1660 1670 1680 1690 1700

TATGGATCTC CCCGACATG AAGATATTTT CCGATGGTGT GACGACGCCA TGTAGCTTC GGTGGGCTGC GACGGCCAGA GGTGCCARCA GCCACGTCRA

1710 1720 1730 1740 1750 1760 1770 1780 1790 1800

ACACCCCTCG GGTCCCGCCC TACACCTCCA AACAGTAGTG AGTAGTGTCT CGTCGCGTTT TAGTATTTGA TGACAAACAA AGTGTAGTT GAGTTAGCCA

1810 1820 1830 1840 1850 1860 1870 1880 1890 1900

CCACCAACTT GCACAGAGGC ACATACATTT GTGTCCATTC TCGCCAGTCA CTTCCATCTC TAGCTTACG TCCTATCTAG CGATGTAGC GGATARTTTC

1910 1920 1930 1940 1950 1960 1970 1980 1990 2000

ATCATCCGTA TATAACCTG TTTGTTATG TTAATTTCT ATATAACT ATACAGTAT ACATTTTAAA AAAAAACAA ATTAGATARA ACAGCCCTG

2010 2020 2030 2040 2050 2060 2070 2080 2090 2100

CTCCTATCCA TCCATGGCAC TTGGAGGAC CAGACTCGGT CATGCCATGC CAAGCCAGCA TATGGTTTAT GGAGAGTAG AAAAAAGAGG AAGTARAGGA

2110 2120 2130 2140 2150 2160 2170 2180 2190 2200

TAGCATGCG TTCTCTCTCT CGTGGATGT GTATTTTGGG GGGATTTGTG TAGTAGTAGC AGCGCGCCCG CCGGGACCGA TCGGATGGT GCGCTTTTCG

2210 2220 2230 2240 2250 2260 2270 2280 2290 2300

GTGGCGTTT CCCGGGGGGG TTTTGGTTTG GCGCTTGGGG GGGATGGCAT GGCGCGCGCT GCGCGCGCAC GCGCGCGCGC GCGCAGTAGC

2310 2320 2330 2340 2350 2360 2370 2380 2390 2400

TCGTCTGCG CCGCGCGCGA CCGTAGCTTA GGGTGGTGTG TTCCGCGCGC GGGCGGGAT TGTTCATGCT GCGTCAATTT GCGCGCACCC TCGCCCGCGC

2410 2420 2430 2440 2450 2460 2470 2480 2490 2500

TCTTGTGCG TCGTGGGCT CTCTCGCGCG GTTTGTCTCT GTCGGTTGCG TACGCCGCGC ACGGGGCGAC GACATTTGCG GATGTAGCCC TGACGTTGTC

2510 2520 2530 2540 2550 2560 2570 2580 2590 2600

GCCCTCTCCG TTGATGATG ATGATGATG TATGATTTT TTTTGTCTG AAGGATTTG TGGGATTTG TTGTGTGTC AGCGCGTCTG GACCGGGAGG

2610 2620 2630 2640 2650 2660 2670 2680 2690 2700

TATAGAGCT GCCTGCACAG GCGCGGGTGG AACCGCGCGC GGGAGCGGCG GTCGCTGGCG TTTCTCTCTG GCCCGCGGGA GGACAGGGTG GTGCGCGCGC

2710 2720 2730 2740 2750 2760 2770 2780 2790 2800

GCGCGAGCG CCGCACGCCG CAGCATTACC CAGGATTTAC CTATGGCGAC CTCATCGGCT TCACCGCGAC CCACTACCGC GCCGACRCC GCACGCTCGA

2810 2820 2830 2840 2850 2860 2870 2880 2890 2900

GCCCTCTCAC GCGTGGCTCG CCGCGCGCGC CCGCGAGGCC GCGCGACGCG CCGAGGTGCG GCGCGCACCG TGATCCCGCA ACGGACCGAA ACGGACCGAA

(EMBL) Length: 2900 current pos: 2901



# 解析ソフトを用いて遺伝子を探す

## cDNAが有る場合

NCBI(National Center for Biotechnology Information)のToolsのSpideyを使ってORFを探す。

<http://www.ncbi.nlm.nih.gov/IEB/Research/Ostell/Spidey/>

第10回目 (11/26) [復習](#)

## リンク集

データベース検索：

1. [PubMed](#): 論文検索
2. [Entrez](#) : 総合データベース
3. [Google Scholar](#) : 文献データベース
4. [特許データベース](#)

ホモロジー検索：

1. [BLAST](#) [GenomeNET]
2. [FASTA](#) [GenomeNET]

配列解析：

1. [DNA → AA](#) : DNA配列をアミノ酸配列に変換
2. [Spidey](#) : DNAとcDNAのマッチング
3. [SSPN](#) : スプライシングの予測
4. [ORF Finder](#) : ORFの予測
5. [clustalW](#) : アライメント
6. [TFSEARCH](#) : 転写因子結合部位の予測
7. [MOTIF](#) : タンパク質モチーフ検索
8. [pI/Mw](#) : 等電点、分子量の予測

解析用配列

- [ある遺伝子AのgenomeDNA配列](#)
- [ある遺伝子AのcDNA配列](#)
- [ある遺伝子XのgenomeDNA配列](#)
- [ある遺伝子XのcDNA配列](#)
- [Saccharomyces cerevisiae actin gene](#)



- Spidey FAQ
- Spidey documentation
- Spidey executables
- Help/Contact

Spidey is an mRNA-to-genomic alignment program. For a complete description of how Spidey works, click [here](#). For an example, click [here](#).

**Genomic sequence (FASTA or GI/Accession):**

Upload file:  ファイルが選...ていません

From:  To:

**mRNA sequence(s) (One or more FASTA or GI/Accession) ?:**

Upload file:  ファイルが選...ていません

divergent sequences ?  
 Use large intron sizes ?

Minimum mRNA-genomic identity ?  %  
Minimum length of mRNA covered ?  %

**Genomic sequence is: ?**

- Vertebrate
- Drosophila
- C. elegans

**Output options:**

- Text/summary
- Summary only
- ASN.1
- Print multiple alignment

第10回目 (11/26) [復習](#)

## リンク集

データベース検索：

1. [PubMed](#): 論文検索
2. [Entrez](#) : 総合データベース
3. [Google Scholar](#) : 文献データベース
4. [特許データベース](#)

ホモロジー検索：

1. [BLAST](#) [GenomeNET]
2. [FASTA](#) [GenomeNET]

配列解析：

1. [DNA → AA](#) : DNA配列をアミノ酸配列に変換
2. [Spidey](#) : DNAとcDNAのマッチング
3. [SSPN](#) : スプライシングの予測
4. [ORF Finder](#) : ORFの予測
5. [clustalW](#) : アライメント
6. [TFSEARCH](#) : 転写因子結合部位の予測
7. [MOTIF](#) : タンパク質モチーフ検索
8. [pI/Mw](#) : 等電点、分子量の予測

解析用配列

 [ある遺伝子AのgenomeDNA配列](#)

 [ある遺伝子AのcDNA配列](#)

[ある遺伝子XのgenomeDNA配列](#)

[ある遺伝子XのcDNA配列](#)

[Saccharomyces cerevisiae actin gene](#)

# Spidey

PubMed Entrez BLAST OMIM Taxonomy Structure

- Spidey FAQ
- Spidey documentation
- Spidey executables
- Help/Contact

Spidey is an mRNA-to-genomic alignment program. For a complete description of how Spidey works, click [here](#). For an example, click [here](#).

### Genomic sequence (FASTA or GI/Accession):

Upload file:  ファイルが選...ていません

```
cgggtgtgaaccagcggcggggagcgggtcgtcgtggtttcttctgtgcccg  
cgggagggacagggtgtgctgcccgcgagcgcgcacgcccagca  
ctaccggacttcactgggcccgaacctatgctctcacgacggccactaccg  
cgccgacaccccacgcctgcagccttcaagcctggctcgcgcccggcgcg  
cggagcccgcgagcggcgaggtcggggggccacgtactgcccgaacg  
gaacgaaacggaacgaa
```

From:  To:

genomicDNA

### mRNA sequence(s) (One or more FASTA or GI/Accession):

Upload file:  ファイルが選...ていません

```
gtataagagctgctgacagggcggtgtgtaaccagcggcgggagcggcg  
gtcgtggcggtttcttctgctcggcggggagaggggtgtgccccggc  
cgagcggccacggcgcgactaccggacttcactggcgccgacctatg  
cgcttcaagcagcgcactaccgcccgaacccgcagctcgagccttcaag  
cgctgctcgcccggcccgcgaccccgcgagcggcgaggtcag  
ggcgaccagctga
```

cDNA

クリック

- divergent sequences ?
  - Use large intron sizes?
- Minimum mRNA-genomic identity ?  %  
Minimum length of mRNA covered ?  %

### Genomic sequence is: ?

- Vertebrate
- Drosophila
- C. elegans

### Output options:

- Text/summary
- Summary only
- ASN.1
- Print multiple alignment



[Spidey Home](#)

Genomic: lcltmpseq\_0 No definition line found  
 mRNA: lcltmpseq\_1 No definition line found

Alignment is on plus strand of genomic sequence and on plus strand of mRNA sequence  
 mRNA coverage: 100%  
 Overall percent identity: 100.0%

131 2873

	Genomic coordinates	mRNA coordinates	length	identity	mismatches	gaps	Donor site	Acc. site
<a href="#">Exon 1</a>	131-687	1-557	557	100.0%	0	0	d	
<a href="#">Exon 2</a>	790-1112	558-880	323	100.0%	0	0	d	a
<a href="#">Exon 3</a>	2584-2873	881-1170	290	100.0%	0	0		a

**Exon 1: 131-687 (genomic); 1-557 (mRNA)**

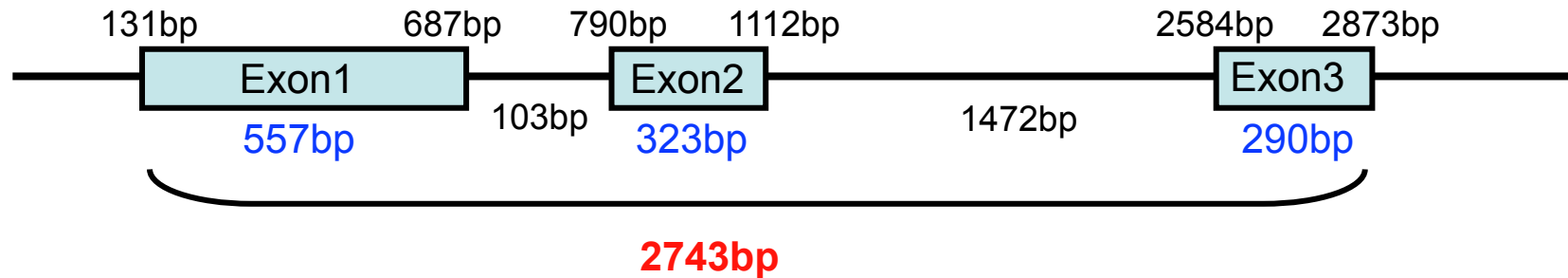
```

131      CTCCAATCTCATGGTGGCCGAGCACCCACGCCACCACAGCCGCACCAAC
          |||
1         ATGGTGGCCGAGCACCCACGCCACCACAGCCGCACCAAC
          M V A E H P T P P Q P H Q

171      CACCGCCCATGGACTCCACCGCCGGCTCTGGCATTGCCGCCCGGGGGCG
          |||
41       CACCGCCCATGGACTCCACCGCCGGCTCTGGCATTGCCGCCCGGGGGCG
          P P P M D S T A G S G I A A P A A

221      GCGGCGGTGTGCGACCTGAGGATGGAGCCCAAGATCCCGGAGCCATTTCGT
          |||
91       GCGGCGGTGTGCGACCTGAGGATGGAGCCCAAGATCCCGGAGCCATTTCGT
    
```

## ある遺伝子Aの構造



ある遺伝子

3つのエクソン（第1エクソン 557bp, 第2エクソン 323bp, 第3エクソン 290bp）、  
2つのイントロン（第1イントロン 103bp, 第2イントロン 1472bp）からなる。  
遺伝子の長さは、2743bp(但しプロモーター領域は含まず)

第10回目 (11/26) [復習](#)

## リンク集


データベース検索：

1. [PubMed](#): 論文検索
2. [Entrez](#): 総合データベース
3. [Google Scholar](#): 文献データベース
4. [特許データベース](#)

ホモロジー検索：

1. [BLAST](#) [GenomeNET]
2. [FASTA](#) [GenomeNET]

配列解析：

- 
1. [DNA → AA](#): DNA配列をアミノ酸配列に変換
  2. [Spidey](#): DNAとcDNAのマッチング
  3. [SSPN](#): スプライシングの予測
  4. [ORF Finder](#): ORFの予測
  5. [clustalW](#): アライメント
  6. [TFSEARCH](#): 転写因子結合部位の予測
  7. [MOTIF](#): タンパク質モチーフ検索
  8. [pI/Mw](#): 等電点、分子量の予測

解析用配列

[ある遺伝子AのgenomeDNA配列](#) [ある遺伝子AのcDNA配列](#)[ある遺伝子XのgenomeDNA配列](#)[ある遺伝子XのcDNA配列](#)[Saccharomyces cerevisiae actin gene](#)

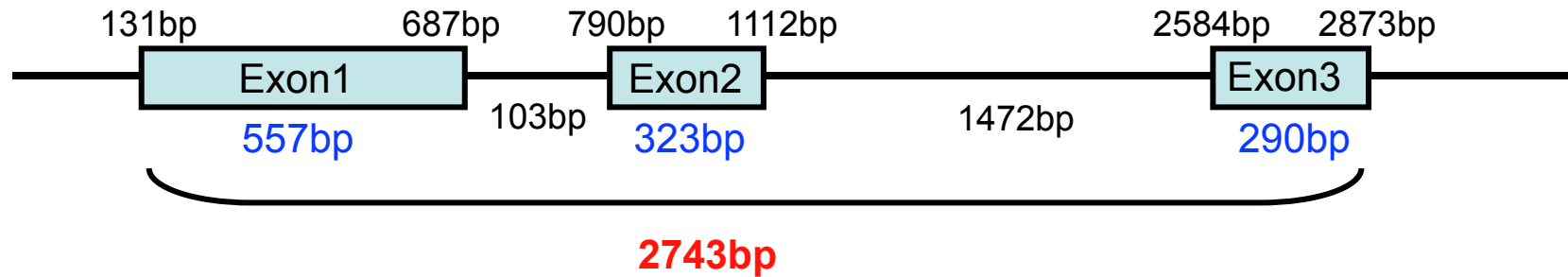
ある遺伝子Aは何をコードしているか？  
ホモロジー検索をやってみよう。

まずcDNAをアミノ酸に変換してみよう





## ある遺伝子の構造



### ある遺伝子

3つのエクソン(第1エクソン 557bp, 第2エクソン 323bp, 第3エクソン 290bp)、  
2つのイントロン(第1イントロン 103bp, 第2イントロン 1472bp) からなる。  
遺伝子の長さは、2743bp(但しプロモーター領域は含まず)

### アミノ酸配列

```
MVAEHPTPPQPHQPPMDSTAGSGIAAPAAAAVCDLRMEPKIPEPFVWPNGDARPASAAE  
LDMPVVDVGVLRDGAEGLRRAAAQVAAACATHGFFQVSEHGVDAAALARAALDGASDFFR  
LPLAEKRRARRVPGTVSGY TSAHADRFASKLPWKETLSFGFHDRAAAPVVADYFSSTLGP  
DFAPMGRVYQKYCEEMKELSLTIMELLELSLGVERGYREFFADSSSIMRCNYYPPCPEP  
ERTLGTGPHCDPTALTILLQDDVGGLEVLVDGEWRPVSPVPGAMVINIGDTFMALSNGRY  
KSCLHRAVNVNQRRERRSLAFFLCPREDRVVRPPPSAATPQHYPDFTWADLMRFTQRHYRA  
DTRTLDAFTRWLAPPAADAAATAQVEAAS*
```

## 2. 何の遺伝子をコードしているのか？

(1) ホモロジーサーチによって遺伝子を推測する

- Blast searchを行う



BLAST

Basic Local Alignment Search Tool

[Home](#)[Recent Results](#)[Saved Strategies](#)[Help](#)▶ [NCBI/BLAST Home](#)

BLAST finds regions of similarity between biological sequences. [more...](#)

**New** Designing or Testing PCR Primers? Try your search in [Primer-BLAST](#). [Go](#)

## BLAST Assembled Genomes

Choose a species genome to search, or [list all genomic BLAST databases](#).

- [Human](#)
- [Mouse](#)
- [Rat](#)
- [Arabidopsis thaliana](#)
- [Oryza sativa](#)
- [Bos taurus](#)
- [Danio rerio](#)
- [Drosophila melanogaster](#)
- [Gallus gallus](#)
- [Pan troglodytes](#)
- [Microbes](#)
- [Apis mellifera](#)

## Basic BLAST

Choose a BLAST program to run.

- |                                  |                                                                                                                                      |
|----------------------------------|--------------------------------------------------------------------------------------------------------------------------------------|
| <a href="#">nucleotide blast</a> | Search a <b>nucleotide</b> database using a <b>nucleotide</b> query<br><i>Algorithms:</i> blastn, megablast, discontinuous megablast |
| <a href="#">protein blast</a>    | Search <b>protein</b> database using a <b>protein</b> query<br><i>Algorithms:</i> blastp, psi-blast, phi-blast                       |
| <a href="#">blastx</a>           | Search <b>protein</b> database using a <b>translated nucleotide</b> query                                                            |
| <a href="#">tblastn</a>          | Search <b>translated nucleotide</b> database using a <b>protein</b> query                                                            |
| <a href="#">tblastx</a>          | Search <b>translated nucleotide</b> database using a <b>translated nucleotide</b> query                                              |

## Specialized BLAST

Choose a type of specialized search (or database name in parentheses.)

- Make specific primers with [Primer-BLAST](#)
- Search [trace archives](#)
- Find [conserved domains](#) in your sequence (cds)
- Find sequences with similar [conserved domain architecture](#) (cdart)
- Search sequences that have [gene expression profiles](#) (GEO)
- Search [immunoglobulins](#) (IgBLAST)

### News

[Align Sequences](#)

A new BL2seq func  
added to the stand  
that allows you to  
a set of subject se  
2008-09-04 12:56:

[More BLAST](#)

### Tip of the Day

How to Search C  
in Web-Blast Usr

A powerful featur  
interface is the abil  
searches to a sub  
using a standard E  
use of Entrez quer  
equivalent of on-th  
databases of exac

[More tips...](#)

BLAST finds regions of similarity between biological sequences. [more...](#)

**New** Designing or Testing PCR Primers? Try your search in **Primer-BLAST**.

#### Align Sequences with BLAST

A new BL2seq functionality has been added to the standard BLAST pages that allows you to align a query against a set of subject sequences.  
2008-09-04 12:56:52

[More BLAST news...](#)

#### Tip of the Day

How to Search Custom Databases  
Entrez Queries.

## BLAST Assembled Genomes

Choose a species genome to search, or [list all genomic BLAST databases](#).

- [Human](#)
- [Mouse](#)
- [Rat](#)
- [Arabidopsis thaliana](#)
- [Oryza sativa](#)
- [Bos taurus](#)
- [Danio rerio](#)
- [Drosophila melanogaster](#)
- [Gallus gallus](#)
- [Pan troglodytes](#)
- [Microbes](#)
- [Apis mellifera](#)

## Basic BLAST

Choose a BLAST program to run.

- [nucleotide blast](#) Search a **nucleotide** database using a **nucleotide** query  
*Algorithms: blastn, megablast, discontinuous megablast*
- [protein blast](#) Search **protein** database using a **protein** query  
*Algorithms: blastp, psi-blast, phi-blast*
- [blastx](#) Search **protein** database using a **translated nucleotide** query
- [tblastn](#) Search **translated nucleotide** database using a **protein** query
- [tblastx](#) Search **translated nucleotide** database using a **translated nu**

Query

データベース

DNA塩基配列	→	DNA塩基配列
アミノ酸配列	→	アミノ酸配列
DNA塩基配列をアミノ酸配列に変換	→	アミノ酸配列
アミノ酸配列	→	DNA塩基配列をアミノ酸配列に変換
DNA塩基配列をアミノ酸配列に変換	→	DNA塩基配列をアミノ酸配列に変換

## Specialized BLAST

Choose a type of specialized search (or database name in parentheses.)

- Make specific primers with [Primer-BLAST](#)
- Search [trace archives](#)
- Find [conserved domains](#) in your sequence (cds)
- Find sequences with similar [conserved domain architecture](#) (cdart)
- Search sequences that have [gene expression profiles](#) (GEO)
- Search [immunoglobulins](#) (IgBLAST)
- Search for [SNPs](#) (snp)
- Screen sequence for [vector contamination](#) (vecscreen)
- [Align](#) two sequences using BLAST (bl2seq)
- Search [protein](#) or [nucleotide](#) targets in PubChem BioAssay

1. TATGGCTTA----  
T G L

2. TATGGCTTA----  
M A -

3. TATGGCTTA----  
W L -

**BLAST** Basic Local Alignment Search Tool

Home Recent Results Saved Strategies Help

NCBI/BLAST/blastp suite: BLASTP programs search protein databases using a protein query. [more...](#) [Reset page](#) [Bookmark](#)

Enter Query Sequence

Enter accession number, gi, or FASTA sequence [Clear](#)

```
LPLAEKRRARRVPGTVSGYTSAHADRFASKLPWKETLSFGFHDRAAAPVVADYFSSTLGP
DFAPMGRVYQKYCEEMKELSLTIMELLELSLGVVERGYREFFADSSSIMRCNYYPPCPEP
ERTLGTGPHCDPTALTILLQDDVGGLEVLVDGEWRPVPVPGAMVINIGDTFMALSNGRY
KSCLHRAVNVNQRERRSLAFFLCPREDRVVRPPSAATPQHYPDFTWADLMRFTQRHYRA
DTRTLDAFTRWLAPPAADAAATAQVEAAS
```

Query subrange

From

To

Or, upload file  ファイル... いません

Job Title

Enter a descriptive title for your BLAST search

Blast 2 sequences

Choose Search Set

Database

Non-redundant protein sequences (nr)

Organism  
Optional

Enter organism name or id--completions will be suggested

Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown.

Entrez Query  
Optional

Enter an Entrez query to limit search

Program Selection

Algorithm

- blastp (protein-protein BLAST)
- PSI-BLAST (Position-Specific Iterated BLAST)
- PHI-BLAST (Pattern Hit Initiated BLAST)

Choose a BLAST algorithm

**BLAST**

Search **database nr** using **Blastp (protein-protein BLAST)**

Show results in a new window

[Algorithm parameters](#)

Job Title: Protein Sequence (389 letters)

[Show Conserved Domain](#)

BLASTP 2.2.18+

Reference:

Stephen F. Altschul, Thomas L. Madden, Alejandro A. Schäffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997), "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", Nucleic Acids Res. 25:3389-3402.

Reference for compositional score matrix adjustment:

Stephen F. Altschul, John C. Wootton, E. Michael Gertz, Richa Agarwala, Aleksandr Morgulis, Alejandro A. Schäffer, and Yi-Kuo Yu (2005) "Protein database searches using compositionally adjusted substitution matrices", FEBS J. 272:5101-5109.

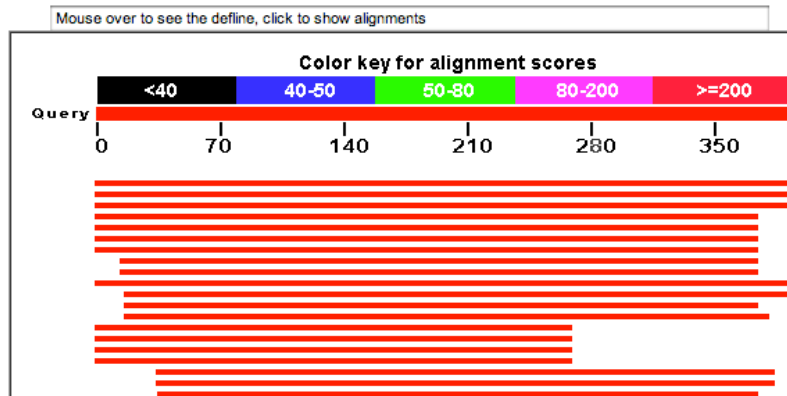
RID: GU22PYZU011

Database: All non-redundant GenBank CDS translations+PDB+SwissProt+PIR+PRF excluding environmental samples from WGS projects  
7,256,259 sequences; 2,507,309,951 total letters

If you have any problems or questions with the results of this search please refer to the [BLAST FAQs](#) [Taxonomy reports](#)

Query=  
Length=389

Distribution of 101 Blast Hits on the Query Sequence

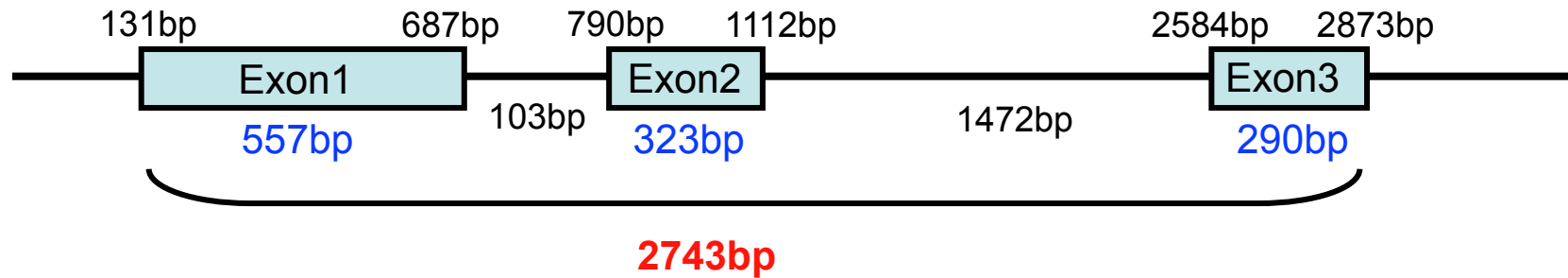



[Distance tree of results](#) **NEW**

Sequences producing significant alignments:		Score (Bits)	E Value	
<a href="#">ref NP_001045014.1 </a>	Os01g0883800 [Oryza sativa (japonica cult...	793	0.0	<b>UG</b>
<a href="#">qb AAN73384.1 </a>	putative gibberellin 20 oxidase [Oryza rufipog...	788	0.0	
<a href="#">dbj BAG80954.1 </a>	GA C20oxidase2 [Oryza glumipatula]	786	0.0	
<a href="#">dbj BAG80957.1 </a>	GA C20oxidase2 [Oryza punctata]	671	0.0	
<a href="#">dbj BAG80956.1 </a>	GA C20oxidase2 [Oryza longistaminata]	668	0.0	
<a href="#">dbj BAG80955.1 </a>	GA C20oxidase2 [Oryza meridionalis]	668	0.0	
<a href="#">dbj BAG80960.1 </a>	GA C20oxidase2 [Oryza alta]	659	0.0	
<a href="#">dbj BAG80959.1 </a>	GA C20oxidase2 [Oryza officinalis]	644	0.0	
<a href="#">dbj BAG80958.1 </a>	GA C20oxidase2 [Oryza minuta]	638	0.0	
<a href="#">dbj BAG80961.1 </a>	GA C20oxidase2 [Oryza brachyantha]	637	0.0	
<a href="#">dbj BAG80962.1 </a>	GA C20oxidase2 [Oryza granulata]	626	1e-177	
<a href="#">qb EAY76733.1 </a>	hypothetical protein OsI_004580 [Oryza sativa ...	599	1e-169	
<a href="#">qb EAZ14396.1 </a>	hypothetical protein OsJ_004221 [Oryza sativa ...	570	5e-161	
<a href="#">dbj BAG72281.1 </a>	gibberellin-20 oxidase-2 [Oryza sativa Indica...	546	1e-153	
<a href="#">dbj BAE48438.1 </a>	gibberellin-20 oxidase-2 [Oryza sativa Indica...	543	1e-152	
<a href="#">dbj BAG72282.1 </a>	gibberellin-20 oxidase-2 [Oryza rufipogon]	542	2e-152	
<a href="#">dbj BAE48455.1 </a>	gibberellin-20 oxidase-2 [Oryza glumipatula]	542	2e-152	
<a href="#">qb AAT44252.1 </a>	putative gibberellin 20-oxidase [Oryza sativa ...	463	1e-128	
<a href="#">ref NP_001055584.1 </a>	Os05g0421900 [Oryza sativa (japonica cult...	446	2e-123	<b>UG</b>
<a href="#">qb ABQ52488.1 </a>	GA 20-oxidase [Paeonia suffruticosa]	441	7e-122	
<a href="#">qb ABF70102.1 </a>	gibberellin 20-oxidase, putative [Musa balbisi...	440	1e-121	
<a href="#">qb ABF72027.1 </a>	gibberellin 20-oxidase family protein [Musa ac...	436	1e-120	
<a href="#">qb AA998356.1 </a>	GA 20-oxidase [Sesbania rostrata]	417	1e-114	
<a href="#">emb CAO61938.1 </a>	unnamed protein product [Vitis vinifera]	415	3e-114	
<a href="#">dbj BAA37127.1 </a>	gibberellin 20-oxidase [Lactuca sativa]	413	1e-113	
<a href="#">dbj BAG48318.1 </a>	gibberellin 20-oxidase1 [Chrysanthemum x mori...	407	6e-112	
<a href="#">qb ABQ17965.1 </a>	gibberellin 20-oxidase [Chrysanthemum x morifo...	406	2e-111	
<a href="#">dbj BAG12318.1 </a>	gibberellin 20-oxidase [Chrysanthemum x morif...	405	2e-111	
<a href="#">dbj BAD30034.1 </a>	gibberellin 20-oxidase2 [Daucus carota]	405	4e-111	
<a href="#">qb AAC49721.1 </a>	GA 20-oxidase [Pisum sativum]	404	5e-111	
<a href="#">qb AAF29605.1 </a>	<a href="#">AF138704.1</a> gibberellin c20-oxidase [Pisum sativum]	404	7e-111	
<a href="#">emb CAQ43616.1 </a>	gibberellin 20-oxidase [Helianthus annuus] >e...	403	1e-110	
<a href="#">emb CAO68523.1 </a>	unnamed protein product [Vitis vinifera]	403	1e-110	
<a href="#">emb CAB96202.1 </a>	gibberellin 20-oxidase [Citrus sinensis x Pon...	402	3e-110	



## ある遺伝子Aの構造



### ある遺伝子A

3つのエクソン(第1エクソン 557bp, 第2エクソン 323bp, 第3エクソン 290bp)、  
2つのイントロン(第1イントロン 103bp, 第2イントロン 1472bp) からなる。  
遺伝子の長さは、2743bp(但しプロモーター領域は含まず)

### アミノ酸配列

```
MVAEHPTPPQPHQPPMDSTAGSGIAAPAAAACDLRMEPKIPEPFVWPNGDARPASAAE  
LDMPVVDVGVLRDGAEGLRRAAAQVAAACATHGFFQVSEHGVDAAALARAALDGASDFFR  
LPLAEKRRARRVPGTVSGYTSAHADRFASKLPWKETLSFGFHDRAAAPVVADYFSSTLGP  
DFAPMGRVYQKYCEEMKELSLTIMELLELSLGVERGYREFFADSSSIMRCNYYPPCPEP  
ERTLGTGPHCDPTALTILLQDDVGGLEVLVDGEWRPVSPVPGAMVINIGDTFMALSNGRY  
KSCLHRAVVNQRRERRSLAFFLCPREDRVVRPPPSAATPQHYPDFTWADLMRFTQRHYRA  
DTRTLDAFTRWLAPPAADAAATAQVEAAS*
```

予想される遺伝子: ジベレリンの生合成酵素遺伝子  
予想される機能: ジベレリンの合成を担っている。