

第6回 遺伝子の探索と遺伝子の機能予測

本日学ぶこと

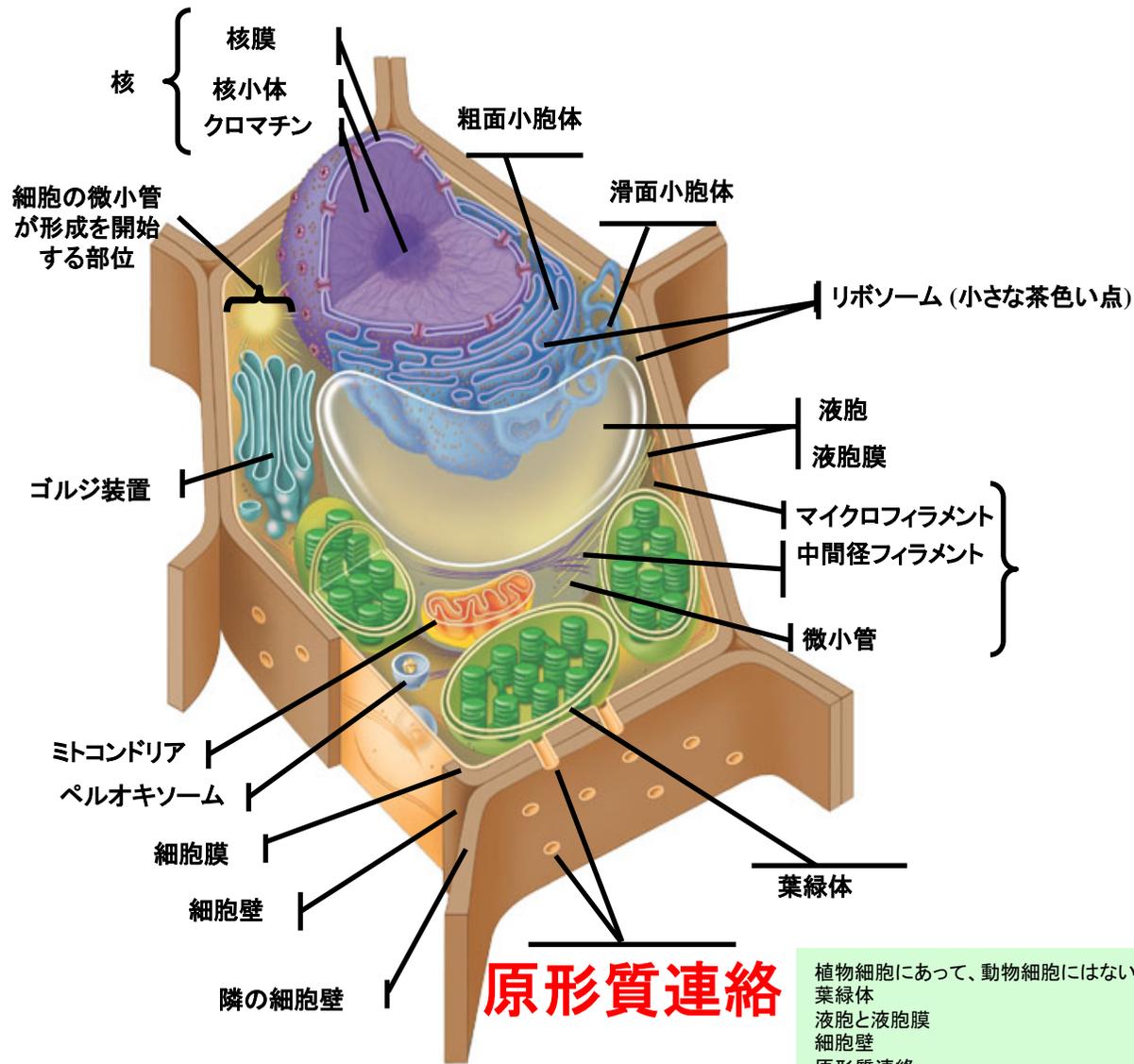
データベースを使って遺伝子を探して、
かつその遺伝子の機能を予測する。

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？

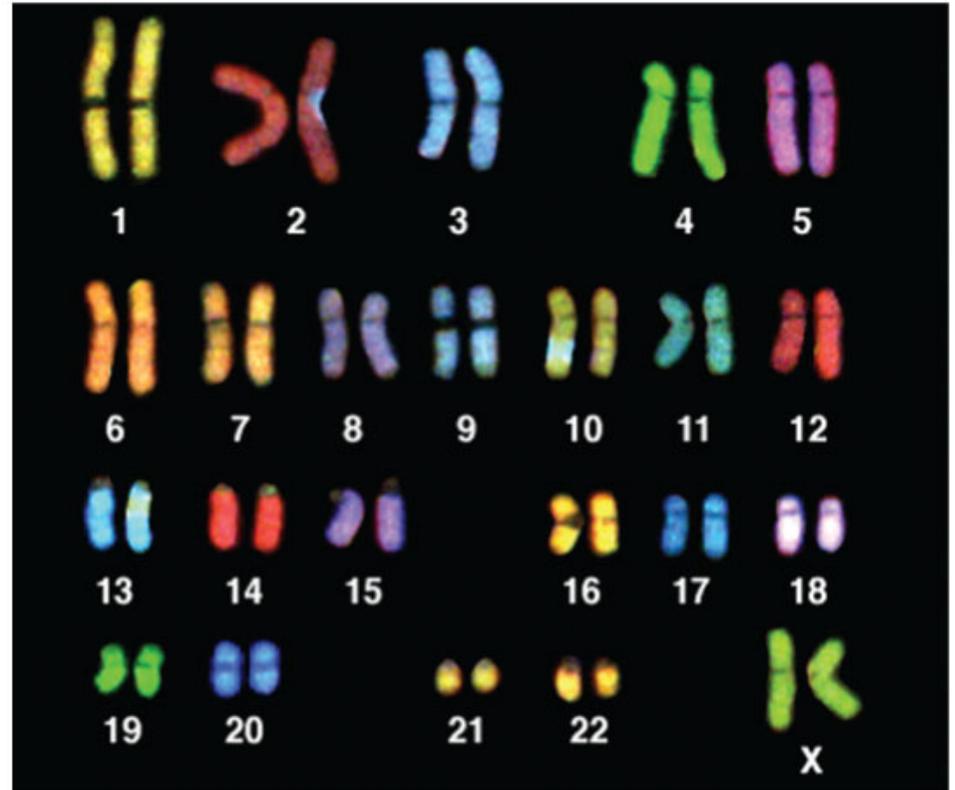
少し復習を試みよう。



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



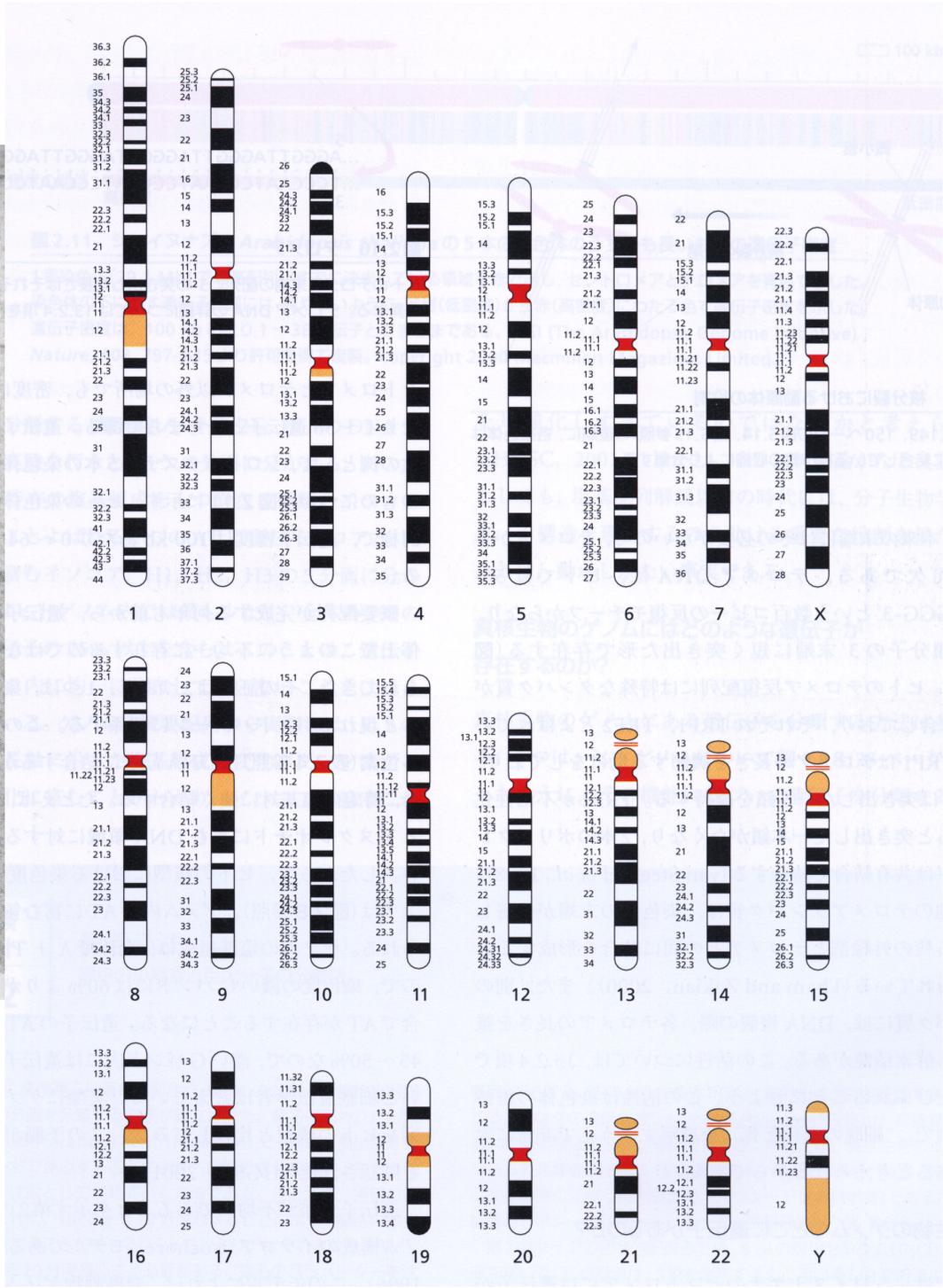
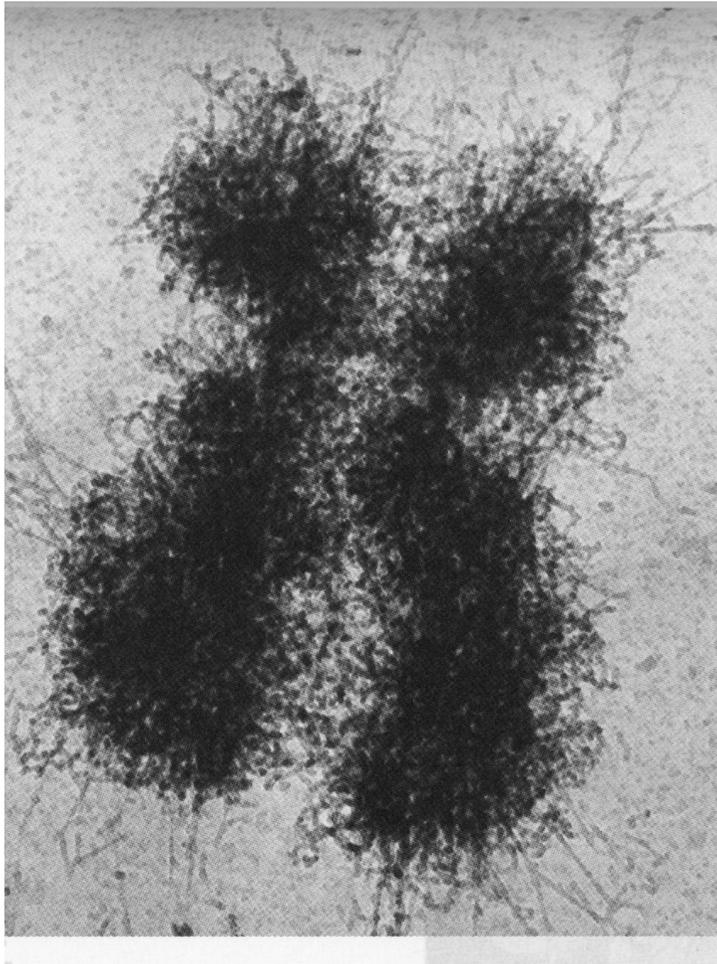
© 2011 Pearson Education, Inc.



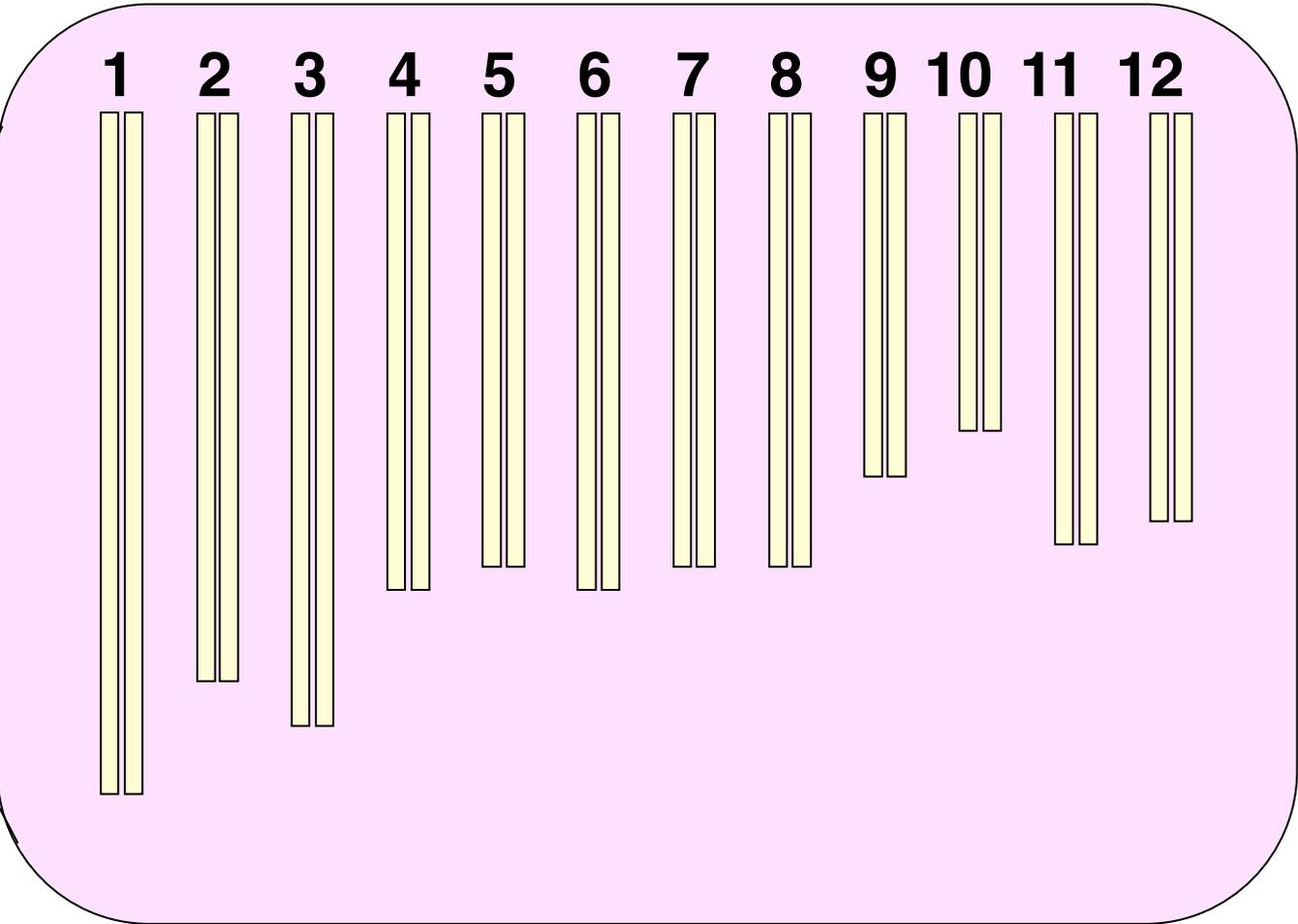
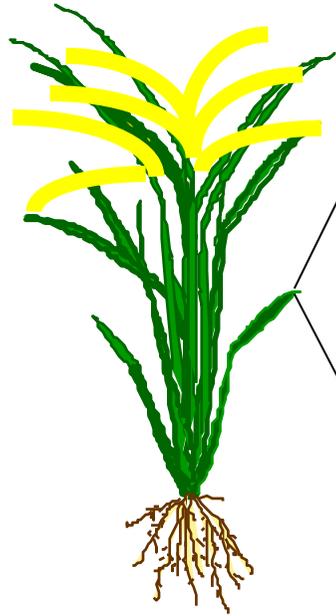
© 2011 Pearson Education, Inc.

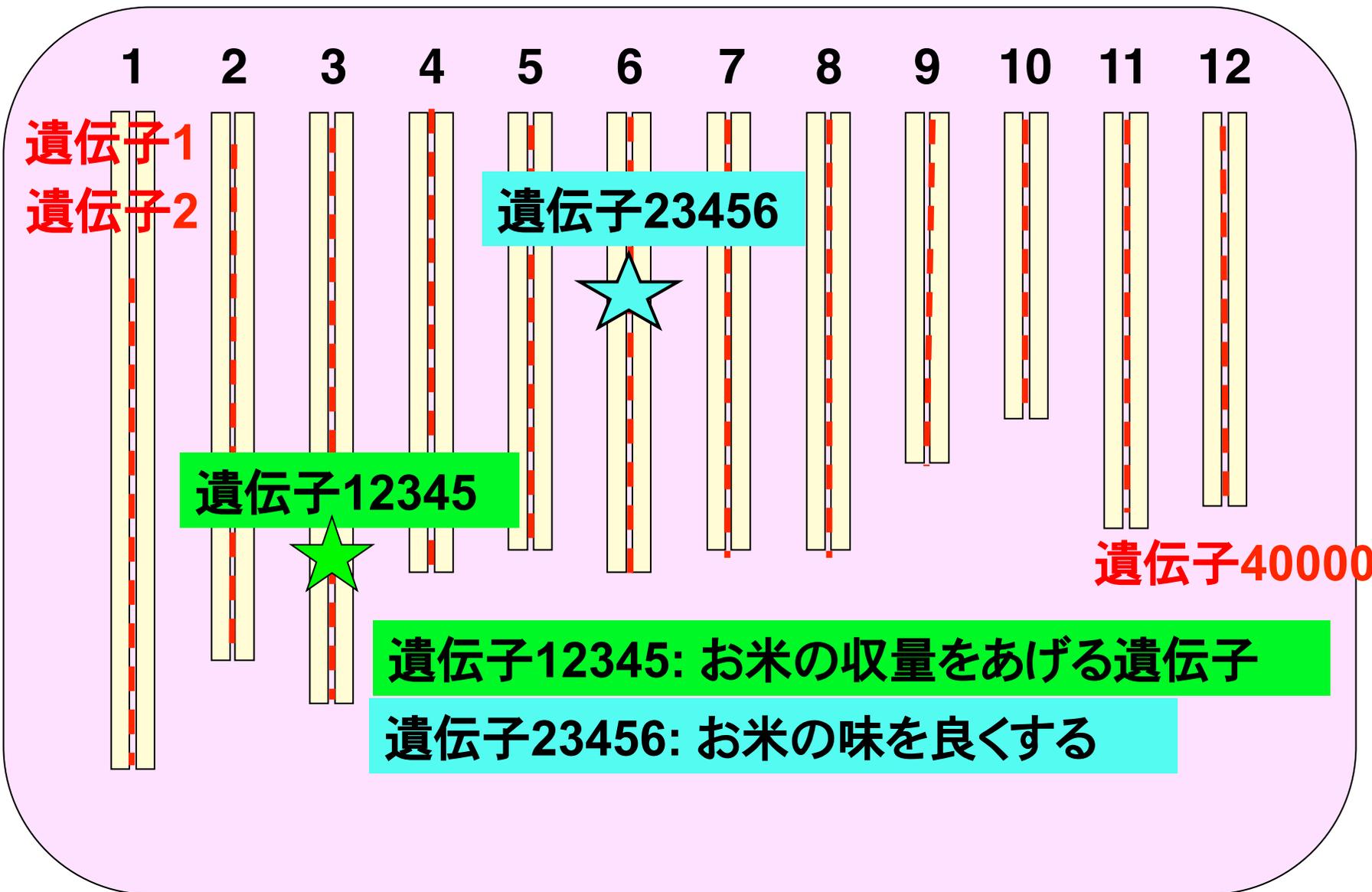
人間の染色体(核型)

キャンベル参照

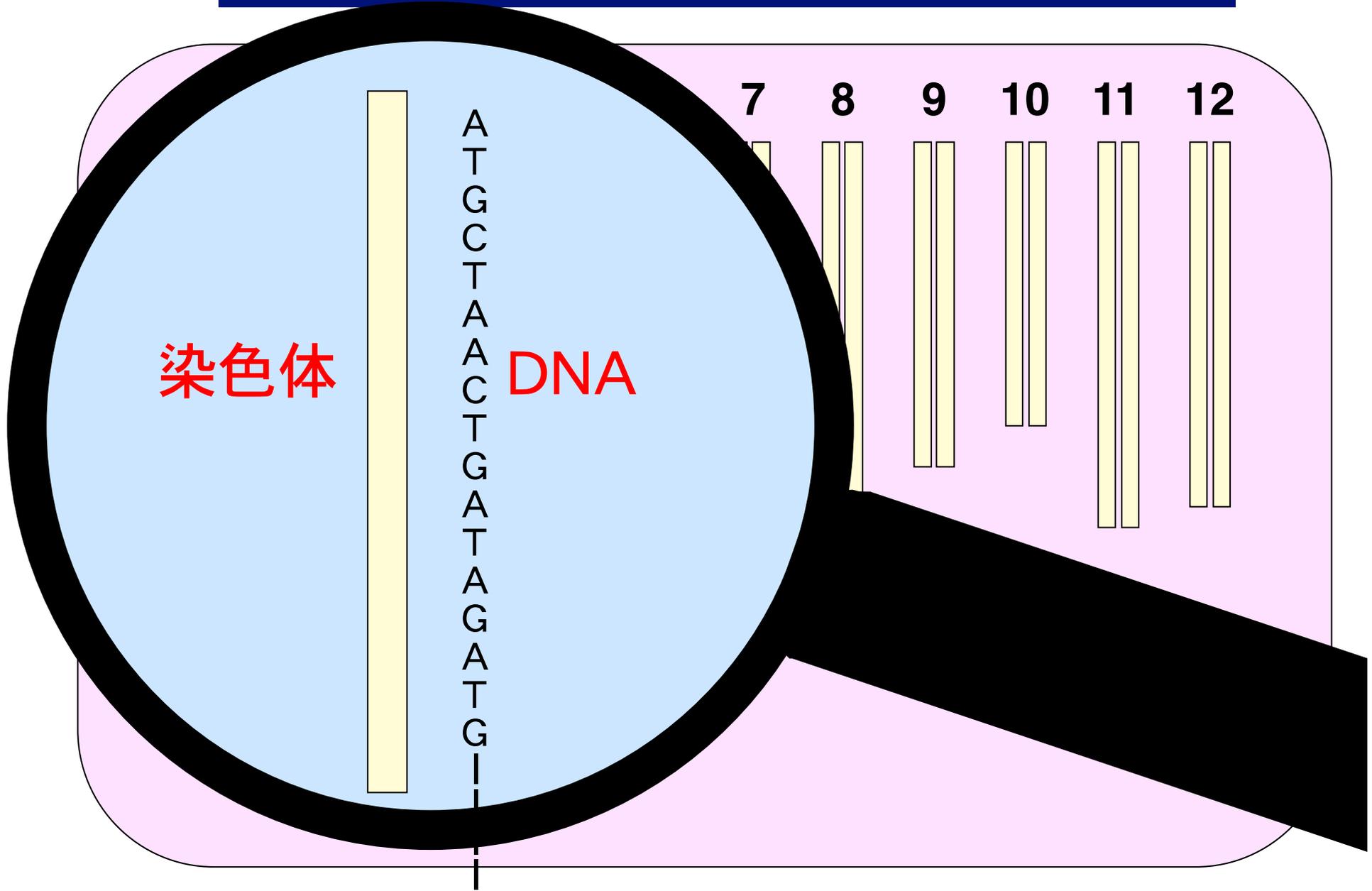


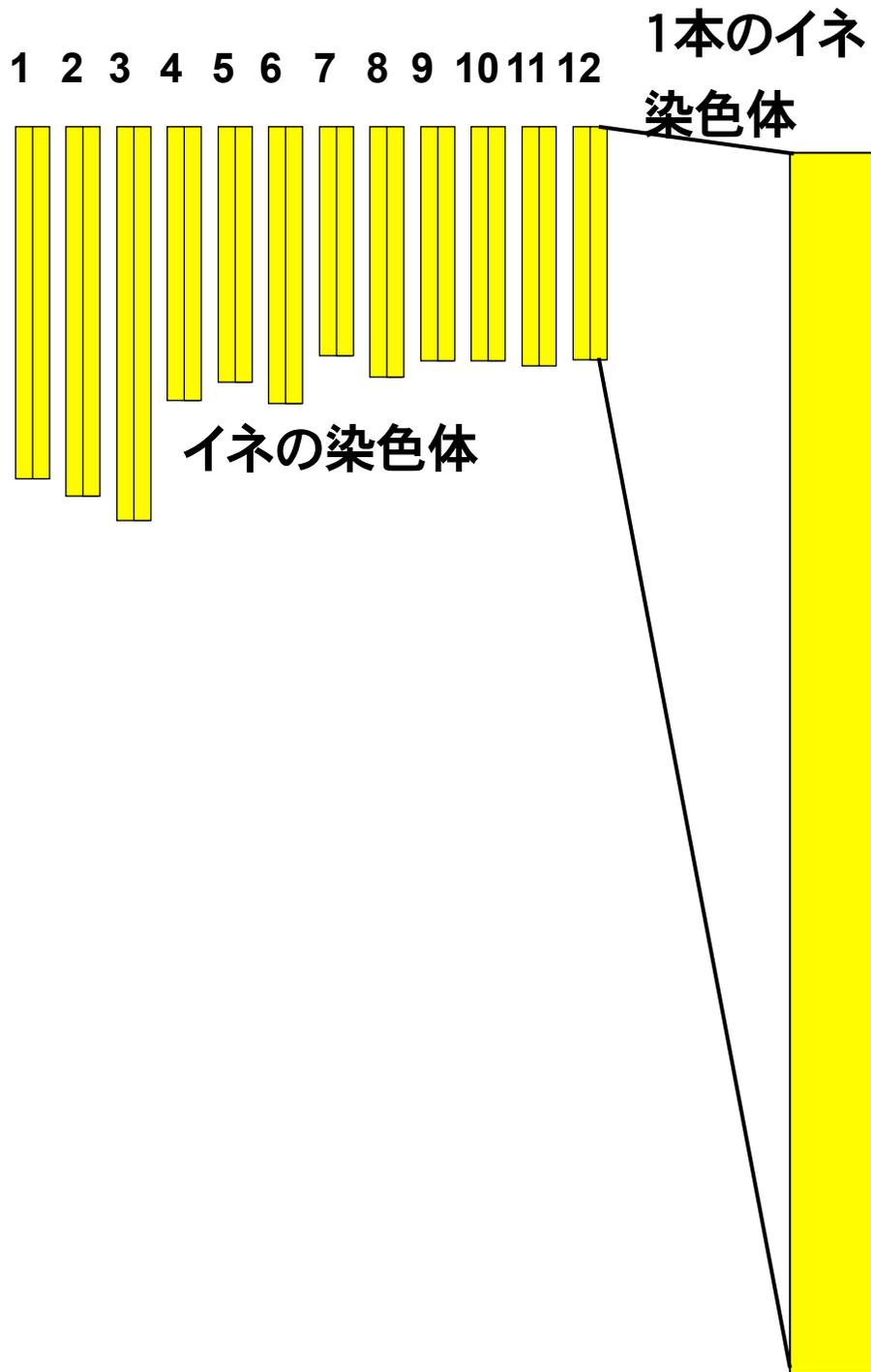
イネの細胞の中の染色体とDNA





イネの細胞の中の染色体とDNA



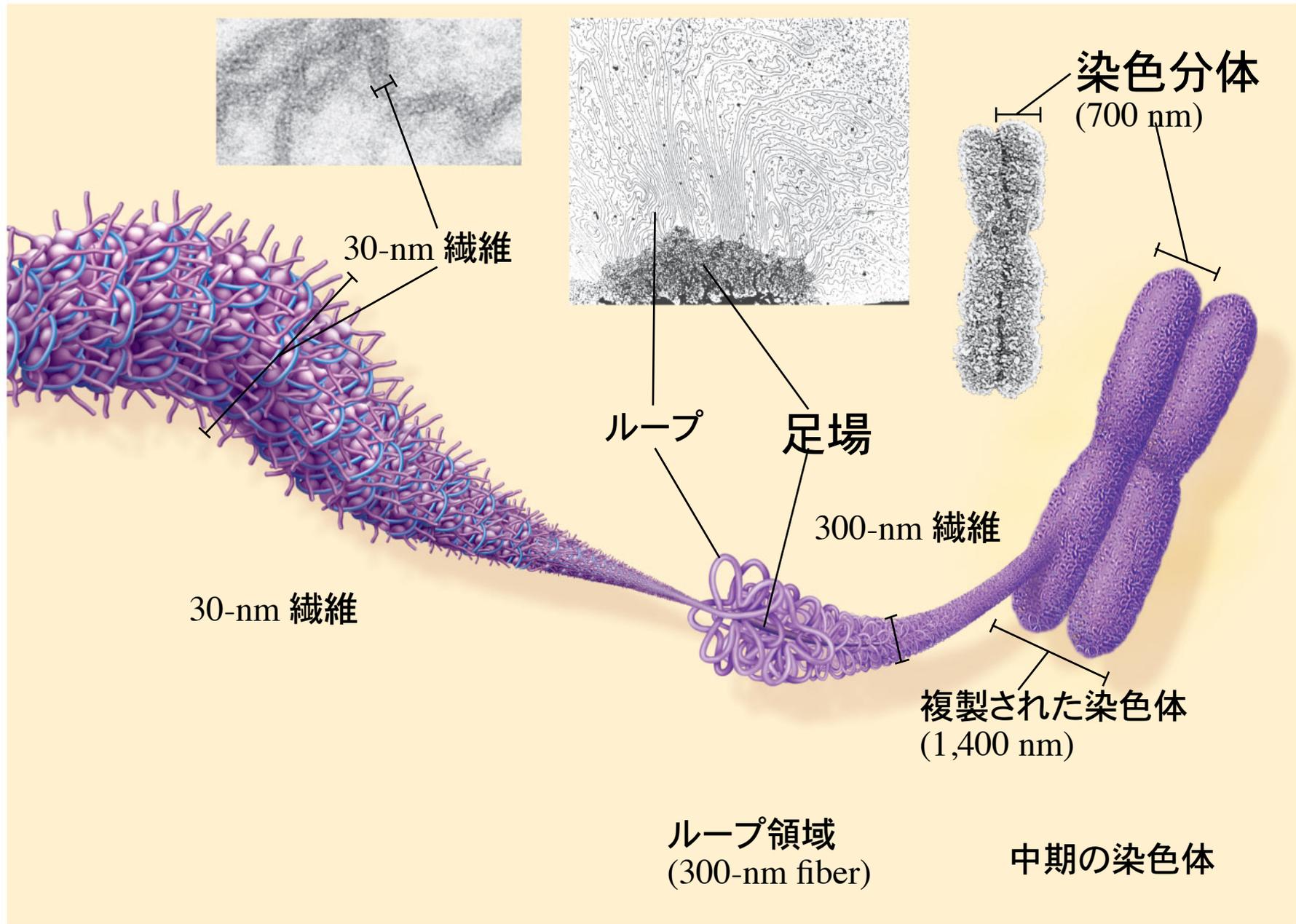


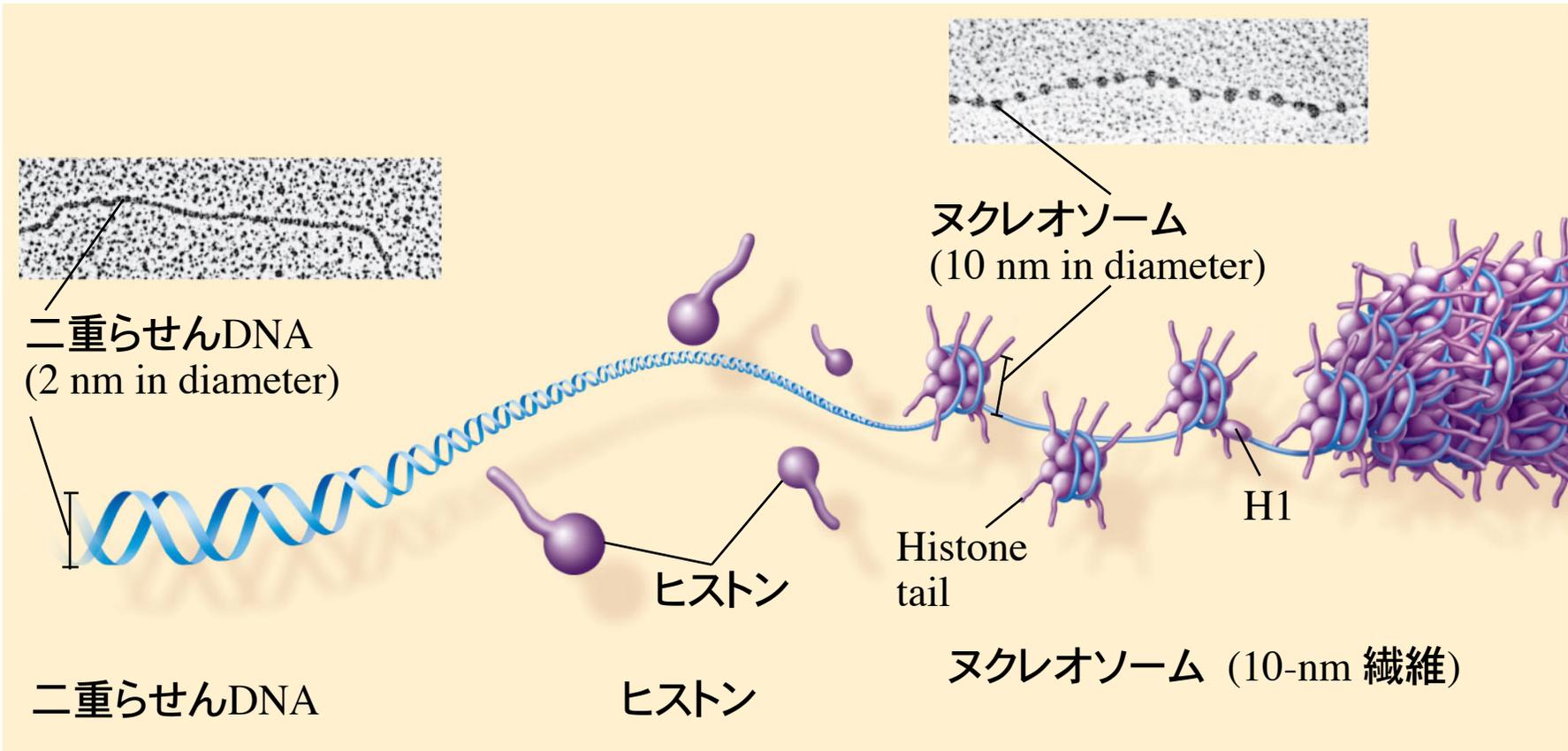
A
T
G
T
I
I
A
A
T
A
G
G
T
A
A
A
T
G
I
I
T
A
G
G
T



遺伝子1

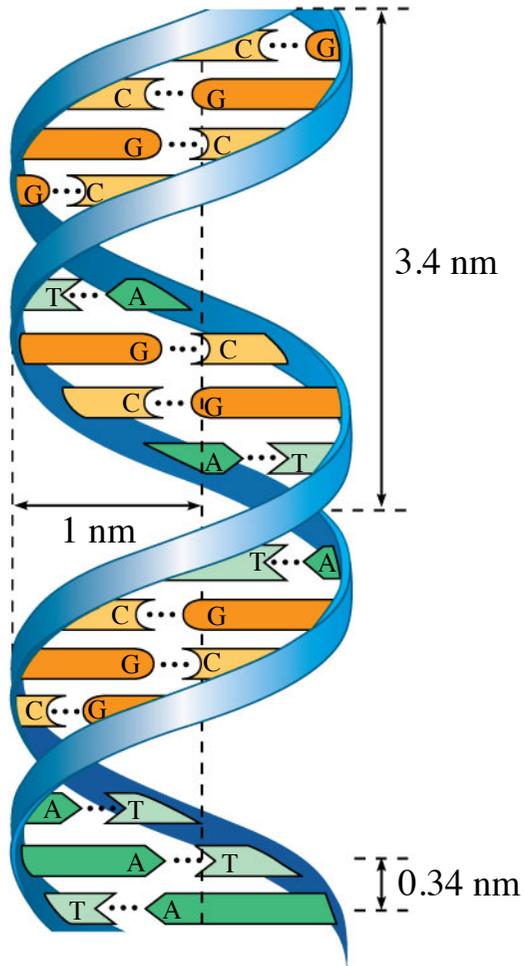
遺伝子2



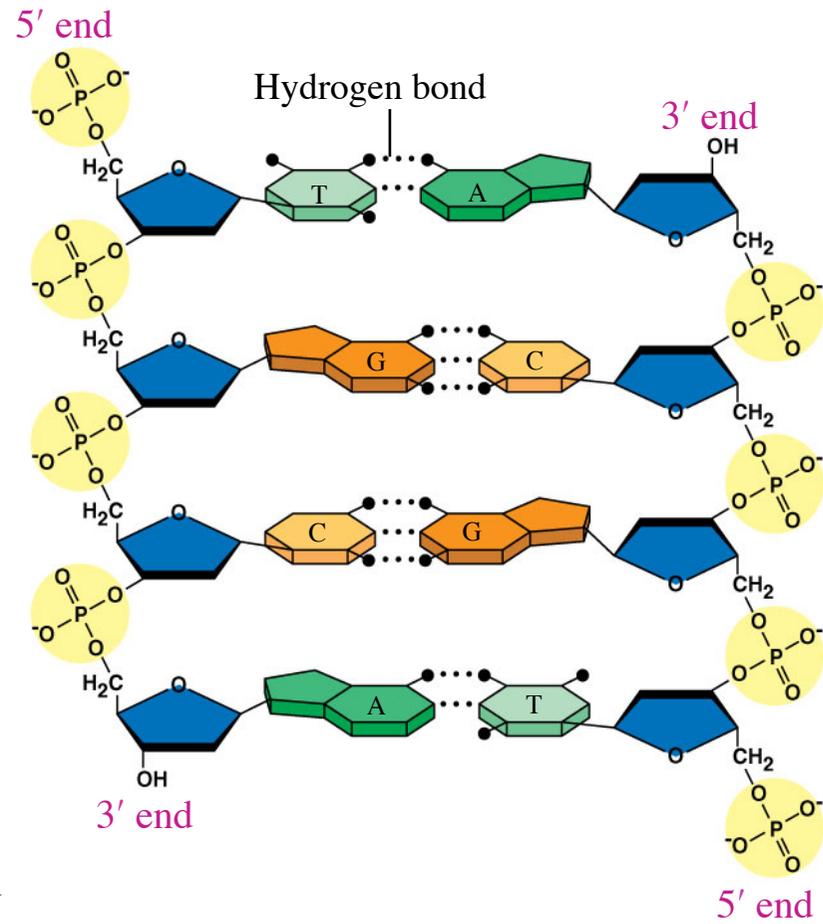


© 2011 Pearson Education, Inc.

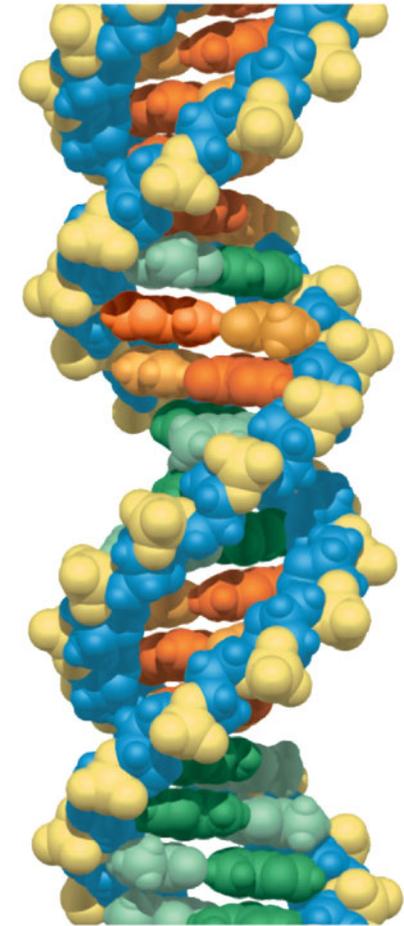
Figure 16.7



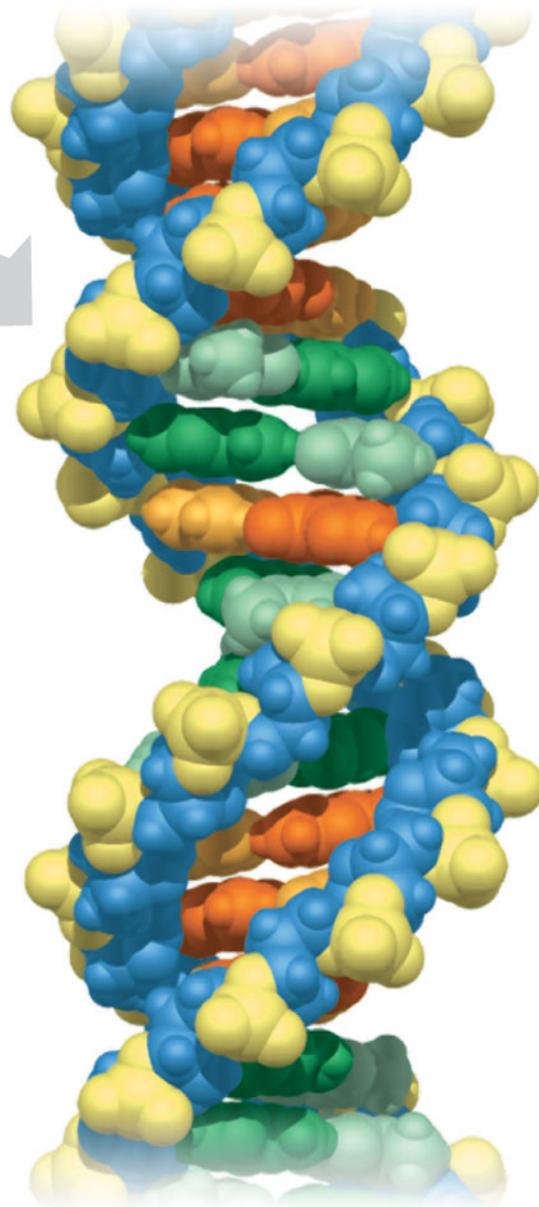
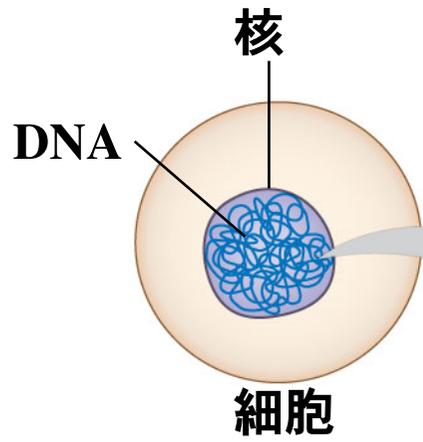
(a) Key features of DNA structure



(b) Partial chemical structure

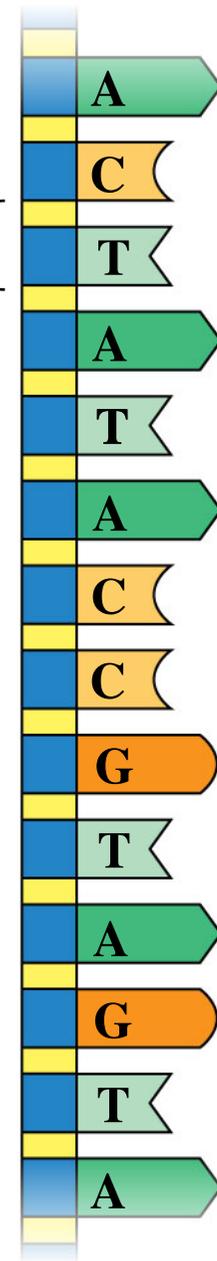


(c) Space-filling model

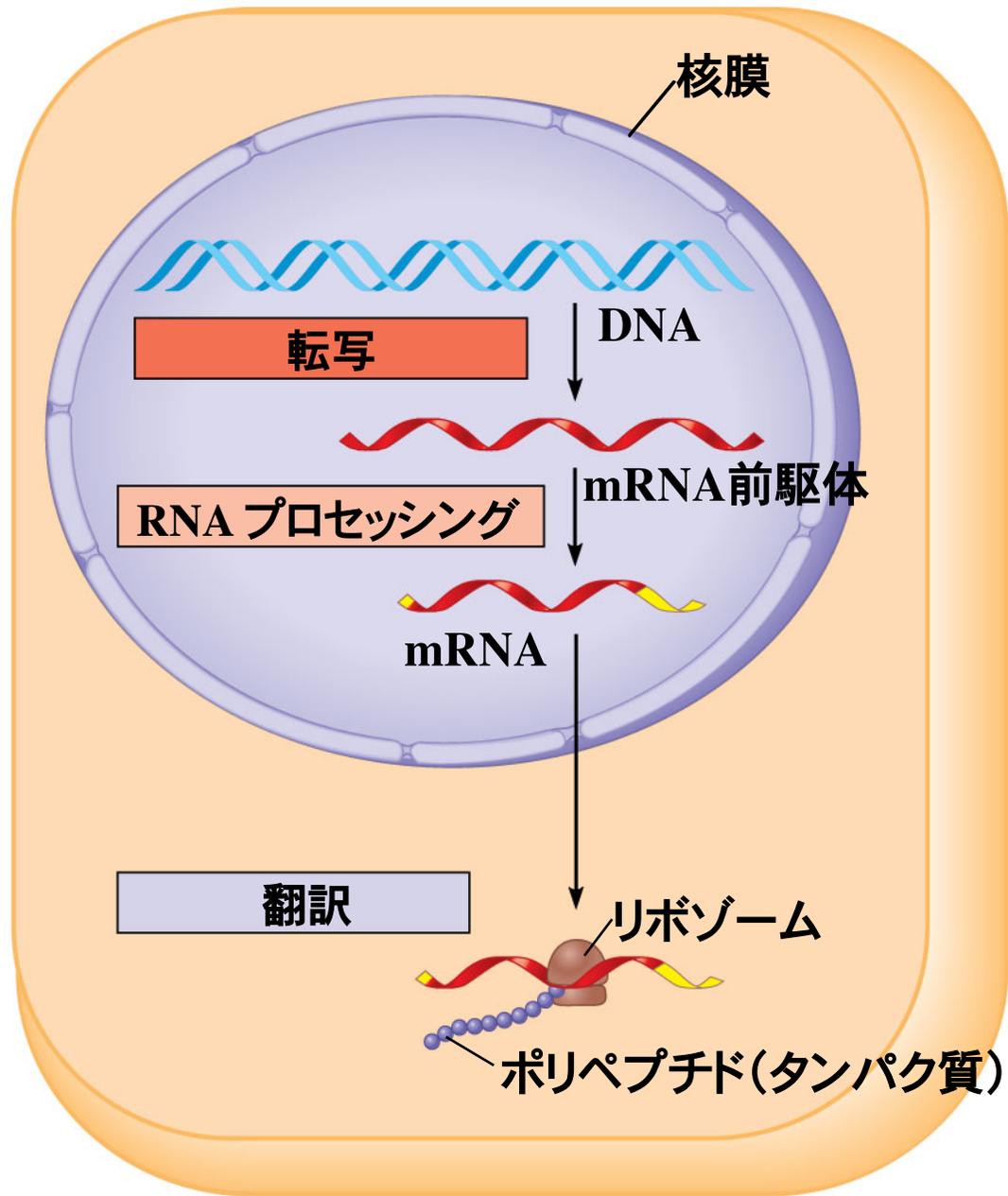


(a) DNA 二重らせん

ヌクレオチド



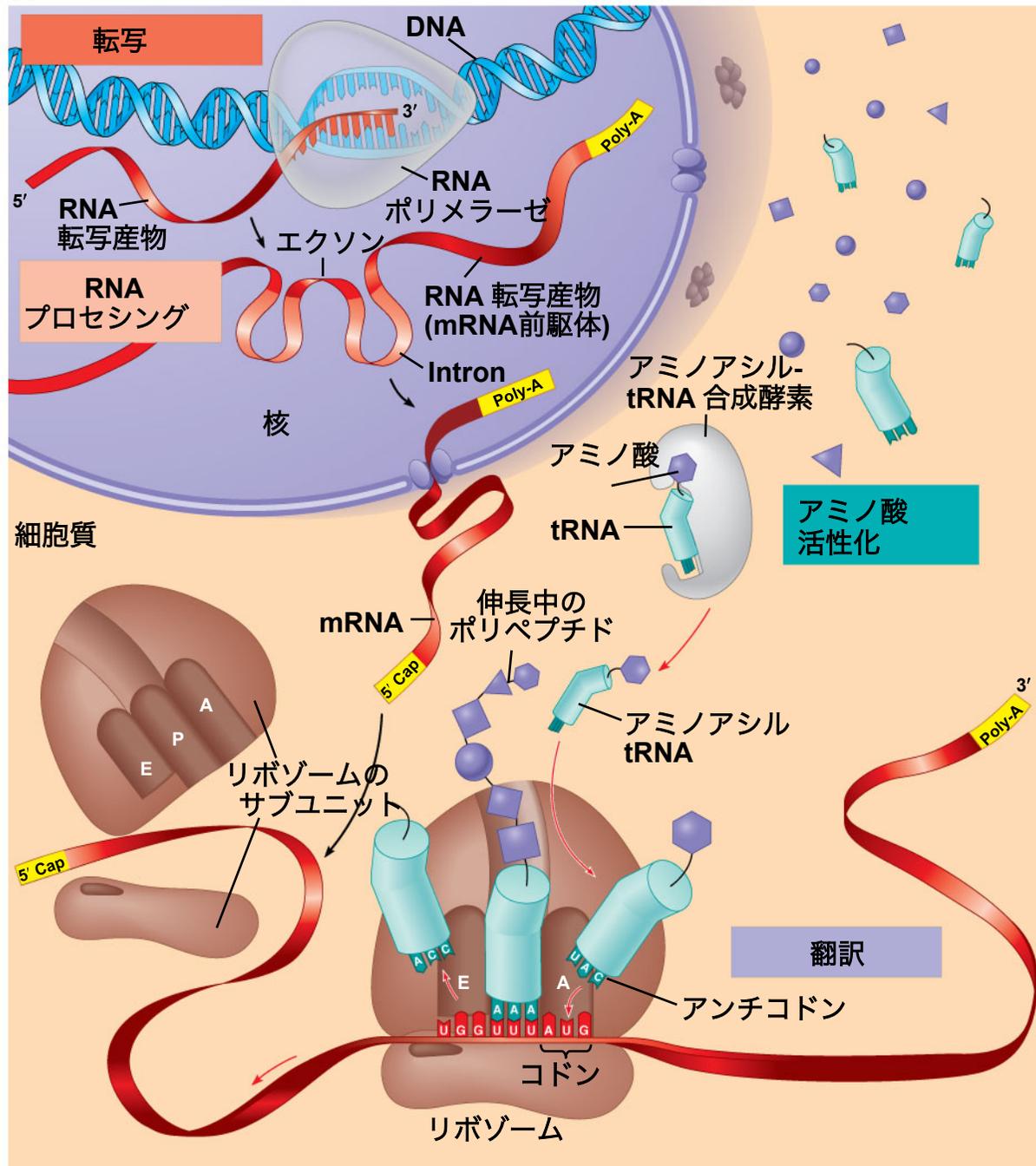
(b) 1本鎖DNA

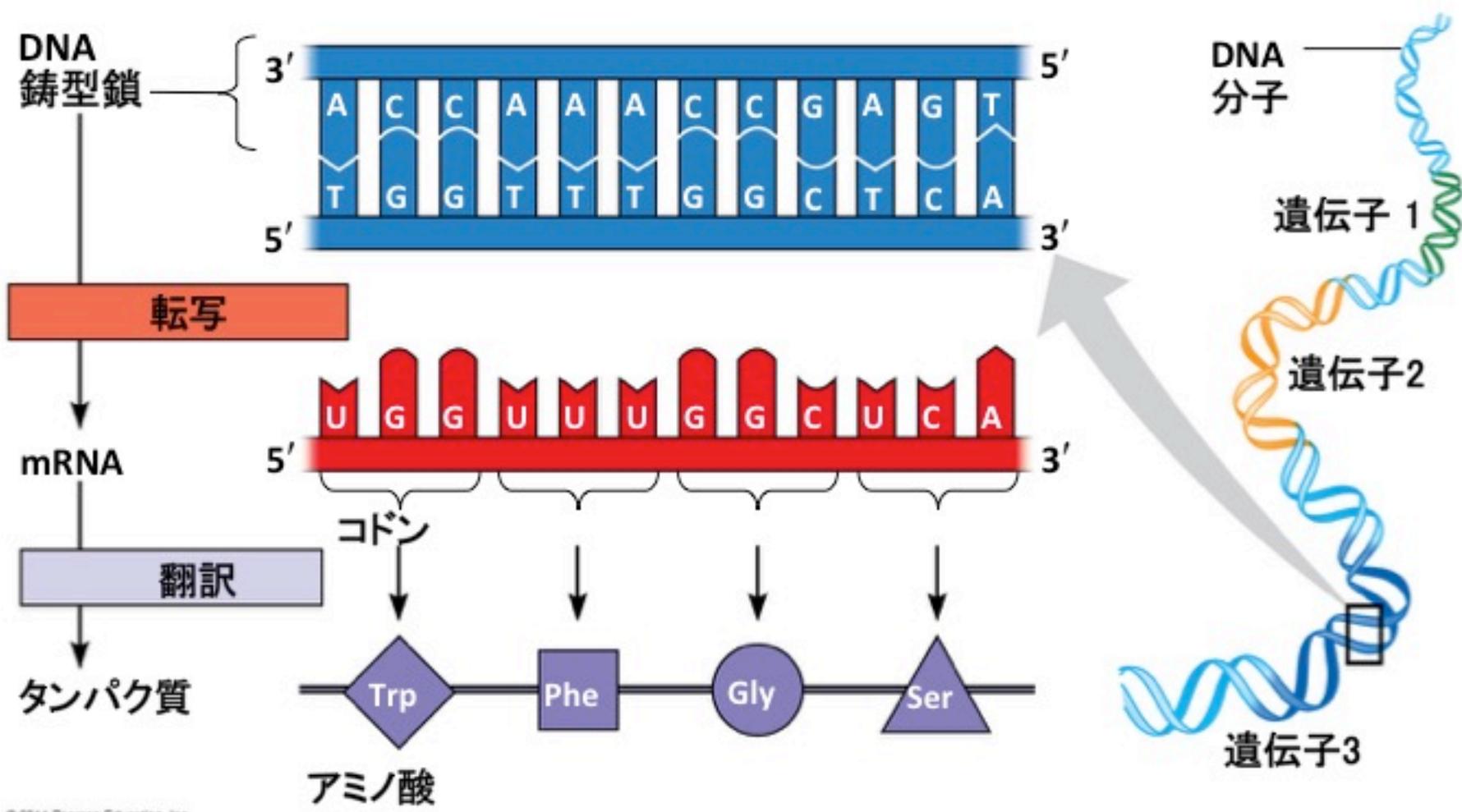


(b) 真核生物の細胞

© 2011 Pearson Education, Inc.

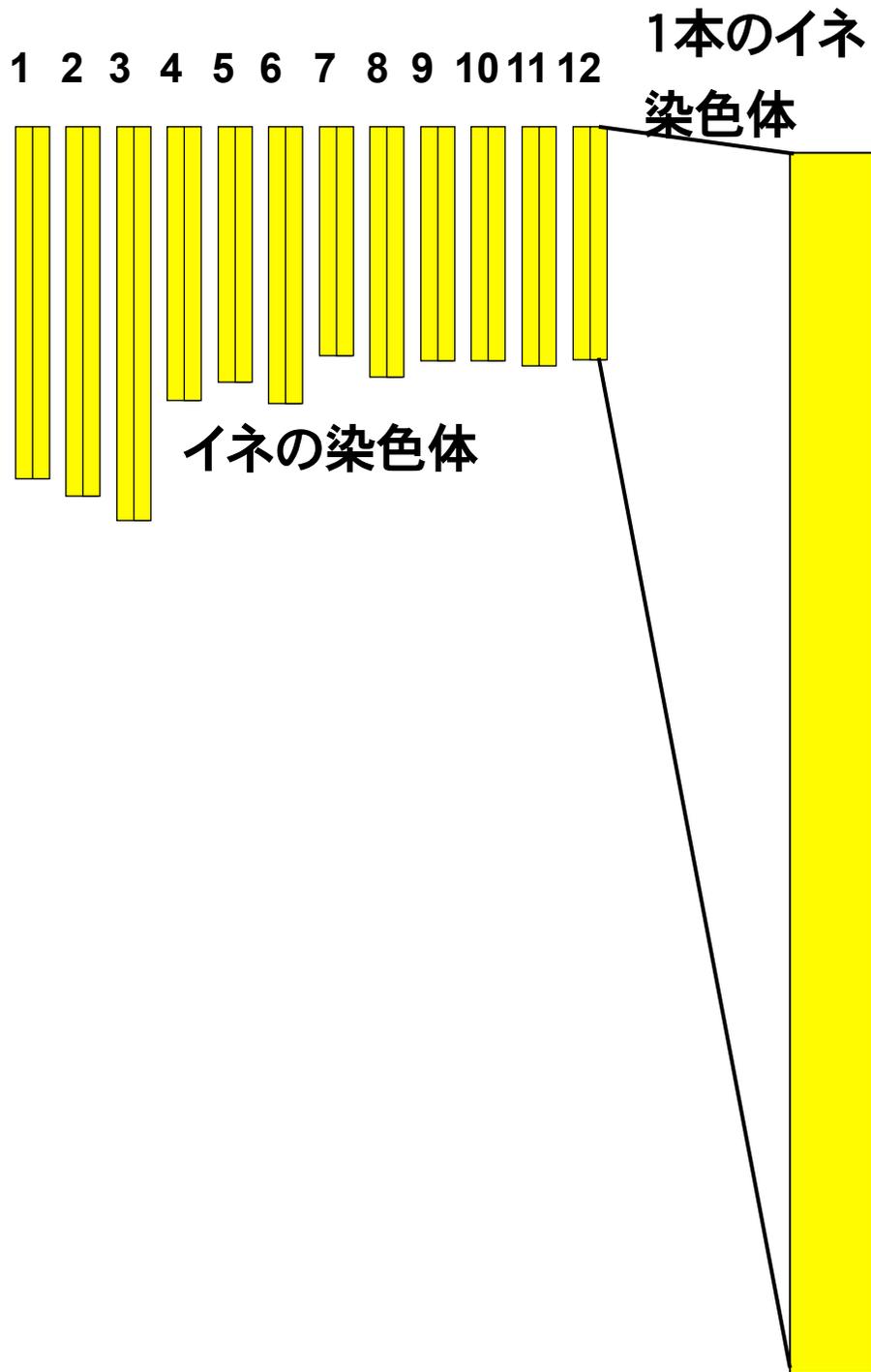
Figure 17





© 2011 Pearson Education, Inc.

キャンベル参照



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 GCCACGACAG 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 GCGCGCGGTG GCGACCTGAG GATGGAGCCC AGGATTCCTG GTGGCCGACG GCGCAGCGCA GCCCGCGCTC
310 320 330 340 350 360 370 380 390 400
GGCGCGCGAG CTGGACATGC CCGTGGTCCG CTGGGCGCTG CTCGCGACG GCGACGCCGA GGGGCTGCG CCGCGCCGCG CCGAGGTGGC CCGCCGCTGC
410 420 430 440 450 460 470 480 490 500
GCCACGCACG GGTCTTCCCA GGTGTCCGAG CACGGCGCTG ACGCCGCTCT GCGCGCGCGC GCCTCGACG GCGCCAGCAG CTCTTCCGCG CTCGCGCTCG
510 520 530 540 550 560 570 580 590 600
CCGAGAGCGG CCGCGCGCGC CCGCTCCGCG GCACCGTGTG CGGTACACC AGCGGCCACG CCGACCGCTT CGCTCCAGG TCCCATGGG AGGAGACCTT
610 620 630 640 650 660 670 680 690 700
CTCCTTCGGC TTCACAGCCG GCGCGCGCGC CCCGCTGCTC GCCGACTACT TCTCCAGCAC CCTCGGCCCG GACTTCGCGC CAATGGGGTA ATTAATACGA
710 720 730 740 750 760 770 780 790 800
TGGTGGACGA CATTGCATTT CAATTCAAA ACAATTCAAA AACACACCGA CCGAGATTAT GCTGATTTCA AACCGTITTG TCGCCGACGG AAGGTGTACC
810 820 830 840 850 860 870 880 890 900
AGAGACTGCG CAGAGAGATG AAGGAGCTGT CGCTGACGAT CATGGACTCT CTGGAGCTGA GCCTGGGCGT GAGCGGAGGG TACTACAGGG AGTTCTTCGC
910 920 930 940 950 960 970 980 990 1000
GGACAGCAGC TCARATCTGC GGTGCACTCA CTACCCGCCA TGCCCGAGC CCGAGCGGAC GCTCGGACCG GCGCCGACCT GCGACCCCGC CCGCTCACC
1010 1020 1030 1040 1050 1060 1070 1080 1090 1100
ATCCTCTCCG AGGACGACGT CCGCGCGCGC GAGGCTCTCG TCAGCGGCGA ATGGCGCCCG CTCAGCCCGC TCCCGCGCGC CATGGTATC ACATCGCGC
1110 1120 1130 1140 1150 1160 1170 1180 1190 1200
ACACCTTCAT GGTAAACCAT CTCTATTCTT CCTCTCTCTT GTTCTCCTCT GCTTCGAGAC AACAGACACA GTATTTCAAG CTTTTTTTTT TCTCTCGCGC
1210 1220 1230 1240 1250 1260 1270 1280 1290 1300
GAATTTGAGC AGAAAATTA GATCGTGTAA GGGCGCGCGC TTTCAGCTGA AAGCGGAGG AACCAGCTCT GACGTGTTTT CTCTGTTCCT ATCACAGCA
1310 1320 1330 1340 1350 1360 1370 1380 1390 1400
ATGGATGCCC CCACTCTCCG ATGTGTATAG ATTTATCTCA CATCTTATAG TTAATAGAGG TAAGTAAACA GACTATTTTT TCATATATA GTTCGTTTGA
1410 1420 1430 1440 1450 1460 1470 1480 1490 1500
TTTTTTTTTT TTAAGTTTT TTTAGTTTTA TCCAAITTA TTGAAARACT TAGCAGGTT TATATATCCA AATTAGTCTC ATTTAGTTTA ATATGTATTA
1510 1520 1530 1540 1550 1560 1570 1580 1590 1600
TATTTTGTAT ATATATTTAT GTTATATTTA AATATTTACT ATATTTTTCT ATAACATTA TTAAGGCCA TTTATATAT ATAAATGAGG GAGTAAITTA
1610 1620 1630 1640 1650 1660 1670 1680 1690 1700
TATGATCTC CCCCGACATG AATATTTTT CCGATGGTGT GACGACGCCA TGTAGCTTC GGTGGGCTG GAGCGCCAGG GGTCCACACA GCCACGTCCA
1710 1720 1730 1740 1750 1760 1770 1780 1790 1800
ACACCCCTCG GGTCCCGCCC TACACTCCCA AACAGTAGT AGTAGTGTCT CGTCGCGTTT TAGTATTTGA TGACAAACAA AGTGTGAGTT GAGTTAGCCA
1810 1820 1830 1840 1850 1860 1870 1880 1890 1900
CCACCAACTT GCACACGAGC ACATCACTTT GTGTCCATTC TCAGCAGTCA CTTCCTCTC TAGTCCATAC TCCTATCTAG CGATGTAGGC GATATTTTC
1910 1920 1930 1940 1950 1960 1970 1980 1990 2000
ATCATCGGTA TATAACCTG TTTGTATAG TTAATTTCTT ATATAACTT ATACAGTAT ACATTTTAAA AGAAAACAAA ATTAGGATTA ACAAGCCCTG
2010 2020 2030 2040 2050 2060 2070 2080 2090 2100
CTCCTATCCA TCCATGGCAC TTGGAGAGAC CAGACTCGGT CATGCCATGC CAGGCCAAGA TATGGGTTAT GAGAGGTAG AGAAGAGAGG AATGAGAGGA
2110 2120 2130 2140 2150 2160 2170 2180 2190 2200
TAGACATGG TTCTCTCTCT GTTGATGT GTATTTTGGG GGAATTTGTG TAGTAGTAGC AGCGGCGCGC CCGGACCGAA TCGGATGTT GCGCTTTTCG
2210 2220 2230 2240 2250 2260 2270 2280 2290 2300
GTGGCGTTTT CCGGGGGGGG TTTTGGTTTT GCGCTTGGGG GGAATGGCAT GCGCGGCGGT GCGGCTGAC GCCACACACA CCGCGCGCGA CCGACGTAGC
2310 2320 2330 2340 2350 2360 2370 2380 2390 2400
TCGTGCTGCG CCGCGCGCGA CCGTAGCTTA GGGTGGTGTG TTCCGCGCGC GCGCGCGGAT TGTCCATGC CGATCGATTT GCGCCACACC TCGCGCGCGC
2410 2420 2430 2440 2450 2460 2470 2480 2490 2500
TCTTGTGCGG TCGTGCGCCT CTCTCGCGCG GTTTGTCTCT GTCGGTTGCG TCAGCGCGCG ACAGGGGACG GACATTTGCG GATGTAGCCC TGACGTGTGC
2510 2520 2530 2540 2550 2560 2570 2580 2590 2600
GGCCTCTCCG TTGATGATG ATGATGTATG TATGTATTTT TTTTGTCTCG AAGGAAITTG TGGGAAITTG TTGTGTGTGC AGGCGCTGTC GACCGGGAGG
2610 2620 2630 2640 2650 2660 2670 2680 2690 2700
TATAGAGACT GCTCGCACG GCGGTGGTG AACACCGCG GAGAGCGGCG GTCCGCTGCG TTCTTCTGTG GCCCGCGGAA GACAGAGGTG GTCCGCGCGC
2710 2720 2730 2740 2750 2760 2770 2780 2790 2800
CGCCAGCGCG CCGCACGCGC CAGCACTACC CCGACTTCAC CTGGGCGGAC CTGATGCGCT TCACGACGCG CCACTACCGC CCGCACACCC GACAGCTCGA
2810 2820 2830 2840 2850 2860 2870 2880 2890 2900
CGCCTTCACG CGCTGCTCG CCGCGCGCGC CCGCAGCGCC GCGCGAGCG CCGAGGTGGA GCGCGCGCGC TGATCCCGAA ACGAGACGAA ACGAGACGAA
(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

10	20	30	40	50	60	70	80	90	100
TCTCCCTGT	TACAAATAC	CCACCCCTCT	GCCCAAGAC	CTCGCCCTGC	ACACACACAC	ACACTCACAC	TCACACACG	TCTCAACTC	CTCCCCTCA
110	120	130	140	150	160	170	180	190	200
ACACAGCGCT	CACCTTCTAT	CTCCAAATCT	ATGGTGGCCG	AGCACCCAC	GCCACCCAC	CCGACCCAC	CACCCGCCAT	GGACTCCAC	GCCGGCTCG
210	220	230	240	250	260	270	280	290	300
GCATTGCGCG	CCCGGCGCG	GCGCGGGTGT	GCGACTGAG	GATGGAGCCC	AGATCCCGG	AGCCATTCT	GTGGCCGAC	GCGACGCCA	GCCCGGGCTC
310	320	330	340	350	360	370	380	390	400
GCGCGCGAG	CTGGACATGC	CCGTGGTCTA	GCTGGGCGTG	CTCCGCGACG	GCGACGCCA	GGGGCTGCC	CGCCCGCGG	GCGAGGTGG	CGCCCGCTG
410	420	430	440	450	460	470	480	490	500
GCCACGACG	GGTCTTCCA	GGTGTCCGAG	CACGGCGCTC	ACGCCCTCT	GCGCGCGCC	GCGCTGACG	GCGCCAGCG	CTTCTCCCG	CTCCCCTCG
510	520	530	540	550	560	570	580	590	600
CCGAGAGCG	CCGCGCGCG	GCGGTCCCG	GCCCGCTCT	GCGTACACC	AGCGCCACG	CCGACCGCT	CGCTCCAGG	CTCCCATGG	AGGAGACCT
610	620	630	640	650	660	670	680	690	700
CTCCTTCCG	TTCCAGACC	GCGCGCGCG	CCCGCTGCT	GCGACTACT	TCTCCAGAC	CCTCGGCC	GACTTCCGC	CAATGGGTA	ATTAAACGA
710	720	730	740	750	760	770	780	790	800
TGGTGGACG	CATTGCATT	CAAAATCAA	ACAATTCAR	AACACACCG	CCGAGATTG	GCTGAATTC	ACCGGTTTG	TGGCGCAGG	AGGGTGTAC
810	820	830	840	850	860	870	880	890	900
AGAATGACT	CGAGGAGTG	AGGAGCTGT	CGCTGACAT	CATGGACTC	CTGGAGCTG	GCCTGGGCT	GGAGCGAGG	TACTACAGG	AGTTCTCCG
910	920	930	940	950	960	970	980	990	1000
GGACGACG	TCAATCATG	GGTGCACAT	CTACCCGCA	TGCCCGGAC	GCGAGCGAC	GCTCGGACG	GCCCGGACG	GCGACCCAC	GCCCGCTCC
1010	1020	1030	1040	1050	1060	1070	1080	1090	1100
ATCCTCTCT	AGGACGACT	GCGCGGCTC	GAGGTCTCT	TCGACGCCA	ATGGCGCCC	GTCAGCCCC	CATGGTCTC	ACACTCCGC	
1110	1120	1130	1140	1150	1160	1170	1180	1190	1200
ACACCTCAT	GTAARACAT	CTCCTATTCT	CCTCTCCTCT	GTTCTCCTCT	GCTTCGAGC	AACACACCA	GTAATTCAG	CTTTTTTTC	TCTCTCGGC
1210	1220	1230	1240	1250	1260	1270	1280	1290	1300
GAAATGACG	AGAAAATAG	GATCGTGTA	GGGGCGGGC	TTTCAGCTA	AGCGGGGAG	AAACCGACT	GCGGTATT	CTCTGTTCA	ATCACAAAC
1310	1320	1330	1340	1350	1360	1370	1380	1390	1400
ATGGAAATG	CCACTCTCC	ATGTGTATG	ATTATCTCA	CATCTTATG	TTAATAGAG	TAAATACAA	GCTACTTTT	TCATATTAT	GTTCTGTTG
1410	1420	1430	1440	1450	1460	1470	1480	1490	1500
TTTTTTTTT	TTAAAGTTT	TTTAGTTTA	TCCAAATTA	TTGAAARCT	TAGCAACGT	TATATACCA	AAATGATTC	ATTTAGTTA	ATATTGATA
1510	1520	1530	1540	1550	1560	1570	1580	1590	1600
TATTTTGTA	ATATATTTT	GTTATATTA	AAATATTTT	ATATTTTCT	ATAAACATTA	TTAAAGCCA	TTTATATTA	AAATGGAG	GAGTATTTA
1610	1620	1630	1640	1650	1660	1670	1680	1690	1700
TATGGATCT	CCCCGACATG	AGAAATTTT	CCGATGGTG	GACGACCCA	TGTAGCTTC	GTTGGGCTG	GACGGCCAG	GTTGCCACA	GCCACGCTC
1710	1720	1730	1740	1750	1760	1770	1780	1790	1800
ACACCCCTG	GGTCCCGCC	TAACTTCCA	ACAGTAGTG	AGTAGTGTCT	CGTCCGTTT	TAGTATTGA	TGACAAACA	AGTGTAGTT	GAGTTAGCA
1810	1820	1830	1840	1850	1860	1870	1880	1890	1900
CCACCAACT	GACACGAGC	ACATACATT	GTGTCATTC	TCGCCAGTC	CTTCCATCT	TAGCTTAC	TCCTATCTG	CGATGAGC	GGATATTTT
1910	1920	1930	1940	1950	1960	1970	1980	1990	2000
ATCATCGTA	TATAACCTG	TTTGTATAG	TTAATTTCT	ATATAACT	ATAACGAT	ACATTTTAA	AGAAACAAA	ATTAGGATA	ACAGCCCTG
2010	2020	2030	2040	2050	2060	2070	2080	2090	2100
CTCCTATCC	TCCATGGAC	TTGGAGGAC	CAGACTCGT	CATGCCATG	CAGGCCAGC	TATGGTTAT	GGAGGAGG	AGATGAGGA	
2110	2120	2130	2140	2150	2160	2170	2180	2190	2200
TAGCATGCG	TTCTCTCTC	CGTGGATGT	GATTTTGGG	GGGATTTGT	TAGTAGTAG	AGCGGCCCG	CGGGGACCA	TGCGATGGT	GCCGTTTCG
2210	2220	2230	2240	2250	2260	2270	2280	2290	2300
GTGGCGTTT	CCCGGGGGG	TTTTGGTTG	GCGCTTGGG	GGGATGGCAT	GCGCGCGCT	GCGCTGAC	GCGCGCGCG	GCGACGACG	GCGCGCTAG
2310	2320	2330	2340	2350	2360	2370	2380	2390	2400
TCGTGCTGC	GCGGGCGGA	CGTAGCTTA	GGTGGTGTG	TTCCGCGCG	GCGCGGGAT	TGTTCCATG	GATCGATT	GCGCCACCC	TCGCCGCGC
2410	2420	2430	2440	2450	2460	2470	2480	2490	2500
TCTGTGCGG	TGTTGCGCT	CTCTCGCGG	GTTTGTCTT	GTCGGTTGC	TCAGCCGCG	ACGGGGGAC	GACATTTGC	GATGTAGCC	TGACGTTGC
2510	2520	2530	2540	2550	2560	2570	2580	2590	2600
GCCCTTCCG	TTGATGATG	ATGATGATG	TATGATTTT	TTTTGTCTG	AGGAAATTT	TGGGAAATG	TTGTGTGTC	AGCGCTGTC	GACCGGGAG
2610	2620	2630	2640	2650	2660	2670	2680	2690	2700
TATAAGACT	GCCTGCACG	GCGCGGGTG	AACCGCGCG	GGAGCGGCG	GTCGCTGGC	TTCTTCTGT	GCCCGGGGA	GGACGGGTT	GTCGCGCGC
2710	2720	2730	2740	2750	2760	2770	2780	2790	2800
GCCCGACCG	GCCCGACCG	CAGCATTAC	CAGACTTAC	CTAGCGGAC	CTACCGGCT	TCACCGGAC	CCACTACCG	GCCGACCG	GACCGCTCA
2810	2820	2830	2840	2850	2860	2870	2880	2890	2900
GCCCTTCCG	CGCTGGCTG	GCCCGCGCG	GCCCGACCG	GCCCGACCG	GCGAGGTCG	GCGCGCCAC	TGATCCCGA	ACGGACCGA	ACGGACCGA

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

Untitled 3

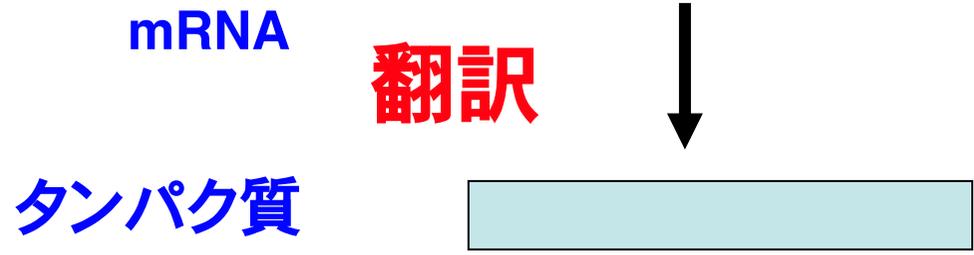
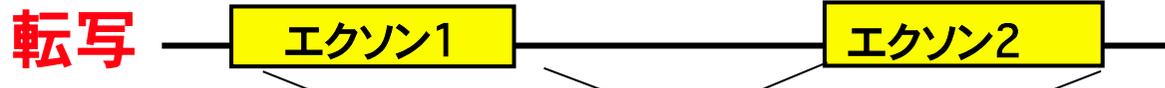
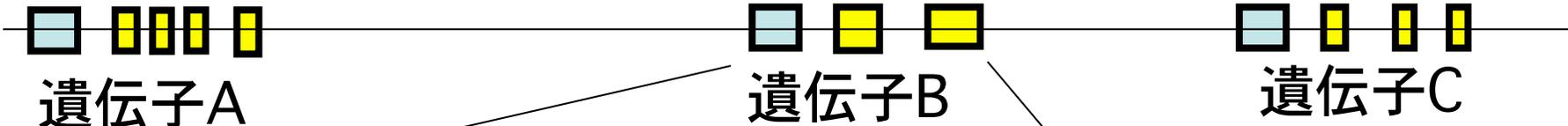
10	20	30	40	50	60	70	80	90	100
TCTCCCTGT	TACAAATAC	CCACCCCTCT	GCCCAAGAC	CTCGCCCTGC	ACACACACAC	ACACTCACAC	TCACACACG	TCTCAACTC	CTCCCCTCA
110	120	130	140	150	160	170	180	190	200
ACACAGCGCT	CACCTTCTAT	CTCCAAATCT	ATGGTGGCCG	AGCACCCAC	GCCACCCAC	CCGACCCAC	CACCCGCCAT	GGACTCCAC	GCCGGCTCG
210	220	230	240	250	260	270	280	290	300
GCATTGCGCG	CCCGGCGCG	GCGCGGGTGT	GCGACTGAG	GATGGAGCCC	AGATCCCGG	AGCCATTCT	GTGGCCGAC	GCGACGCCA	GCCCGGGCTC
310	320	330	340	350	360	370	380	390	400
GCGCGCGAG	CTGGACATGC	CCGTGGTCTA	GCTGGGCGTG	CTCCGCGACG	GCGACGCCA	GGGGCTGCC	CGCCCGCGG	GCGAGGTGG	CGCCCGCTG
410	420	430	440	450	460	470	480	490	500
GCCACGACG	GGTCTTCCA	GGTGTCCGAG	CACGGCGCTC	ACGCCCTCT	GCGCGCGCC	GCGCTGACG	GCGCCAGCG	CTTCTCCCG	CTCCCCTCG
510	520	530	540	550	560	570	580	590	600
CCGAGAGCG	CCGCGCGCG	GCGGTCCCG	GCCCGCTCT	GCGTACACC	AGCGCCACG	CCGACCGCT	CGCTCCAGG	CTCCCATGG	AGGAGACCT
610	620	630	640	650	660	670	680	690	700
CTCCTTCCG	TTCCAGACC	GCGCGCGCG	CCCGCTGCT	GCGACTACT	TCTCCAGAC	CCTCGGCC	GACTTCCGC	CAATGGGTA	ATTAAACGA
710	720	730	740	750	760	770	780	790	800
TGGTGGACG	CATTGCATT	CAAAATCAA	ACAATTCAR	AACACACCG	CCGAGATTG	GCTGAATTC	ACCGGTTTG	TGGCGCAGG	AGGGTGTAC
810	820	830	840	850	860	870	880	890	900
AGAATGACT	CGAGGAGTG	AGGAGCTGT	CGCTGACAT	CATGGACTC	CTGGAGCTG	GCCTGGGCT	GGAGCGAGG	TACTACAGG	AGTTCTCCG
910	920	930	940	950	960	970	980	990	1000
GGACGACG	TCAATCATG	GGTGCACAT	CTACCCGCA	TGCCCGGAC	GCGAGCGAC	GCTCGGACG	GCCCGGACG	GCGACCCAC	GCCCGCTCC
1010	1020	1030	1040	1050	1060	1070	1080	1090	1100
ATCCTCTCT	AGGACGACT	GCGCGGCTC	GAGGTCTCT	TCGACGCCA	ATGGCGCCC	GTCAGCCCC	CATGGTCTC	ACACTCCGC	
1110	1120	1130	1140	1150	1160	1170	1180	1190	1200
ACACCTCAT	GTAARACAT	CTCCTATTCT	CCTCTCCTCT	GTTCTCCTCT	GCTTCGAGC	AACACACCA	GTAATTCAG	CTTTTTTTC	TCTCTCGGC
1210	1220	1230	1240	1250	1260	1270	1280	1290	1300
GAAATGACG	AGAAAATAG	GATCGTGTA	GGGGCGGGC	TTTCAGCTA	AGCGGGGAG	AAACCGACT	GCGGTATT	CTCTGTTCA	ATCACAAAC
1310	1320	1330	1340	1350	1360	1370	1380	1390	1400
ATGGAAATG	CCACTCTCC	ATGTGTATG	ATTATCTCA	CATCTTATG	TTAATAGAG	TAAATACAA	GCTACTTTT	TCATATTAT	GTTCTGTTG
1410	1420	1430	1440	1450	1460	1470	1480	1490	1500
TTTTTTTTT	TTAAAGTTT	TTTAGTTTA	TCCAAATTA	TTGAAARCT	TAGCAACGT	TATATACCA	AAATGATTC	ATTTAGTTA	ATATTGATA
1510	1520	1530	1540	1550	1560	1570	1580	1590	1600
TATTTTGTA	ATATATTTT	GTTATATTA	AAATATTTT	ATATTTTCT	ATAAACATTA	TTAAAGCCA	TTTATATTA	AAATGGAG	GAGTATTTA
1610	1620	1630	1640	1650	1660	1670	1680	1690	1700
TATGGATCT	CCCCGACATG	AGAAATTTT	CCGATGGTG	GACGACCCA	TGTAGCTTC	GTTGGGCTG	GACGGCCAG	GTTGCCACA	GCCACGCTC
1710	1720	1730	1740	1750	1760	1770	1780	1790	1800
ACACCCCTG	GGTCCCGCC	TAACTTCCA	ACAGTAGTG	AGTAGTGTCT	CGTCCGTTT	TAGTATTGA	TGACAAACA	AGTGTAGTT	GAGTTAGCA
1810	1820	1830	1840	1850	1860	1870	1880	1890	1900
CCACCAACT	GACACGAGC	ACATACATT	GTGTCATTC	TCGCCAGTC	CTTCCATCT	TAGCTTAC	TCCTATCTG	CGATGAGC	GGATATTTT
1910	1920	1930	1940	1950	1960	1970	1980	1990	2000
ATCATCGTA	TATAACCTG	TTTGTATAG	TTAATTTCT	ATATAACT	ATAACGAT	ACATTTTAA	AGAAACAAA	ATTAGGATA	ACAGCCCTG
2010	2020	2030	2040	2050	2060	2070	2080	2090	2100
CTCCTATCC	TCCATGGAC	TTGGAGGAC	CAGACTCGT	CATGCCATG	CAGGCCAGC	TATGGTTAT	GGAGGAGG	AGATGAGGA	
2110	2120	2130	2140	2150	2160	2170	2180	2190	2200
TAGCATGCG	TTCTCTCTC	CGTGGATGT	GATTTTGGG	GGGATTTGT	TAGTAGTAG	AGCGGCCCG	CGGGGACCA	TGCGATGGT	GCCGTTTCG
2210	2220	2230	2240	2250	2260	2270	2280	2290	2300
GTGGCGTTT	CCCGGGGGG	TTTTGGTTG	GCGCTTGGG	GGGATGGCAT	GCGCGCGCT	GCGCTGAC	GCGCGCGCG	GCGACGACG	GCGCGCTAG
2310	2320	2330	2340	2350	2360	2370	2380	2390	2400
TCGTGCTGC	GCGGGCGGA	CGTAGCTTA	GGTGGTGTG	TTCCGCGCG	GCGCGGGAT	TGTTCCATG	GATCGATT	GCGCCACCC	TCGCCGCGC
2410	2420	2430	2440	2450	2460	2470	2480	2490	2500
TCTGTGCGG	TGTTGCGCT	CTCTCGCGG	GTTTGTCTT	GTCGGTTGC	TCAGCCGCG	ACGGGGGAC	GACATTTGC	GATGTAGCC	TGACGTTGC
2510	2520	2530	2540	2550	2560	2570	2580	2590	2600
GCCCTTCCG	TTGATGATG	ATGATGATG	TATGATTTT	TTTTGTCTG	AGGAAATTT	TGGGAAATG	TTGTGTGTC	AGCGCTGTC	GACCGGGAG
2610	2620	2630	2640	2650	2660	2670	2680	2690	2700
TATAAGACT	GCCTGCACG	GCGCGGGTG	AACCGCGCG	GGAGCGGCG	GTCGCTGGC	TTCTTCTGT	GCCCGGGGA	GGACGGGTT	GTCGCGCGC
2710	2720	2730	2740	2750	2760	2770	2780	2790	2800
GCCCGACCG	GCCCGACCG	CAGCATTAC	CAGACTTAC	CTAGCGGAC	CTACCGGCT	TCACCGGAC	CCACTACCG	GCCGACCG	GACCGCTCA
2810	2820	2830	2840	2850	2860	2870	2880	2890	2900
GCCCTTCCG	CGCTGGCTG	GCCCGCGCG	GCCCGACCG	GCCCGACCG	GCGAGGTCG	GCGCGCCAC	TGATCCCGA	ACGGACCGA	ACGGACCGA

(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

-----TGGCATATGGCATGCAGACTGGCATTGGCATGCGTCTGCATGCCATTGGTTTCTGAGGTCGCTGTGAAACTTATTTGGGCACAGACATGGCATGCAGACTGGCAC-----
-----ACCGTATACCGTACGTCTGACCGTAACCGTACGCAGACGTACGGTAACCAAAGAGTCCAGCACACTTTGAATAAACCCGTGTCTGTACCGTACGTCTGACCGTG-----



エキソンとイントロンの境界

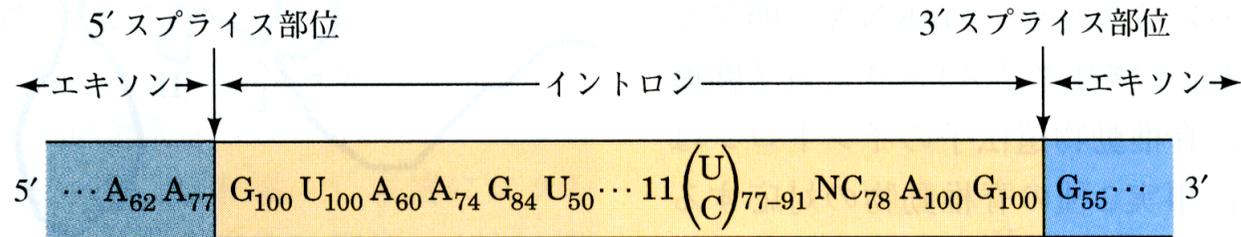


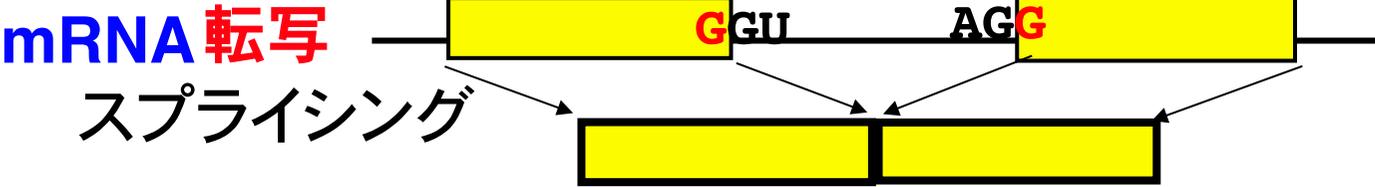
図 25・19 真核細胞 mRNA 前駆体のエキソン・イントロン接合部にあるコンセンサス配列 各塩基の添え数字はその塩基の出現頻度(%). 3' スプライス部位のすぐ上流にピリミジンに富む 11 nt 区域がある [R.A. Padgett, P.J. Grabowski, M.M. Konarska, S.S. Seiler, P.A. Sharp, *Annu. Rev. Biochem.* **55**, 1123 (1986) のデータに基づく]

DNAとcDNA



DNA ---GTCTGC**ATGCCATTGGTTTCTGAGGTCGTGTGA**AACA**ATAAAGGGCA**
 DNA ---CAGACG**TACGGTAACCAAAGACTCAGCACACT**TTGTTATTTCCCGT
PolyA認識サイト

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



成熟mRNA **CAP**GTCUGC**AUGCCAUUGGUCGUGUGA**AACAAUAAAGGGCAAAAAAA

cDNA CAGACG**TACGGTAACCAAAGACTCAGCACACT**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

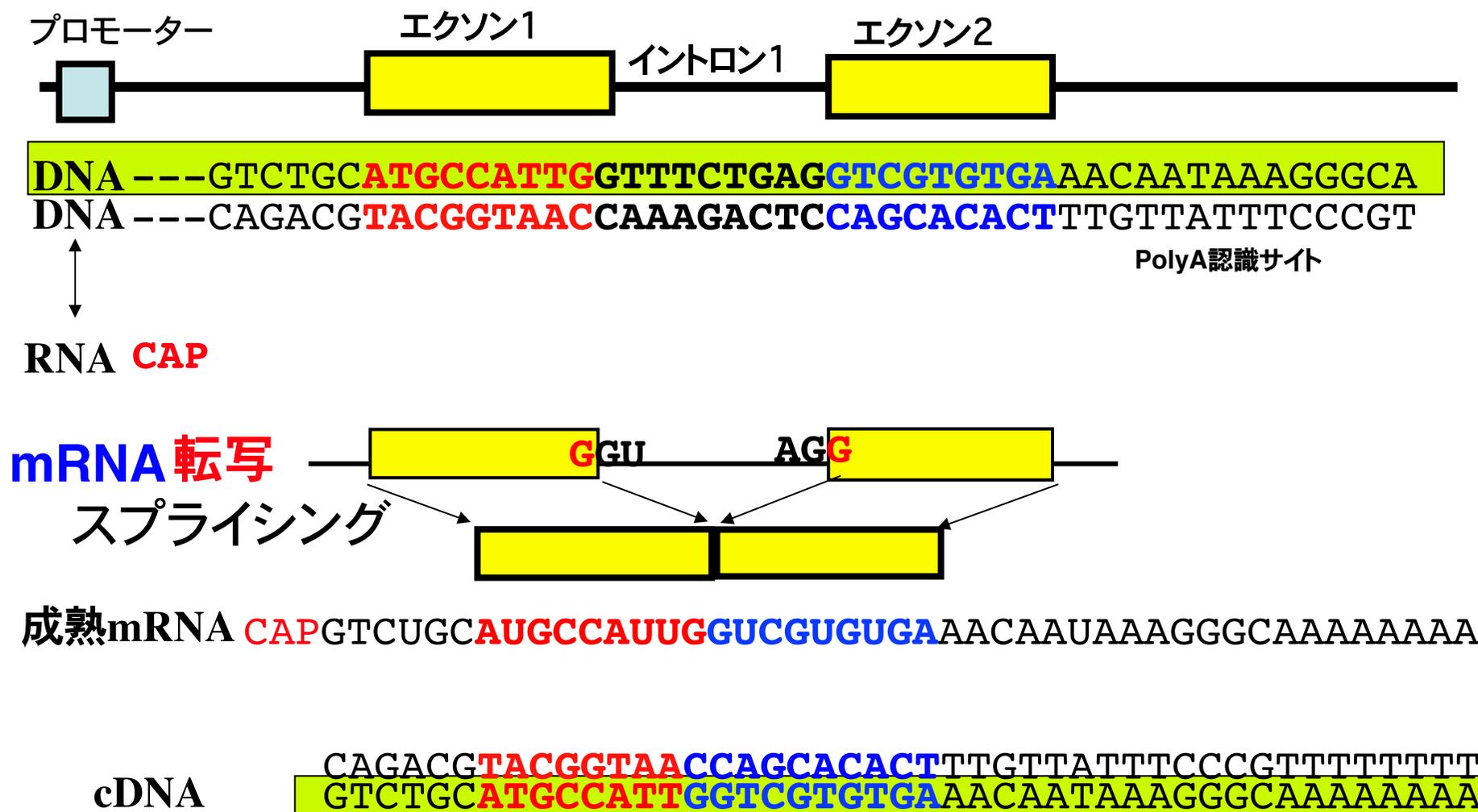
遺伝子予測ソフトで遺伝子領域を推定する

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。

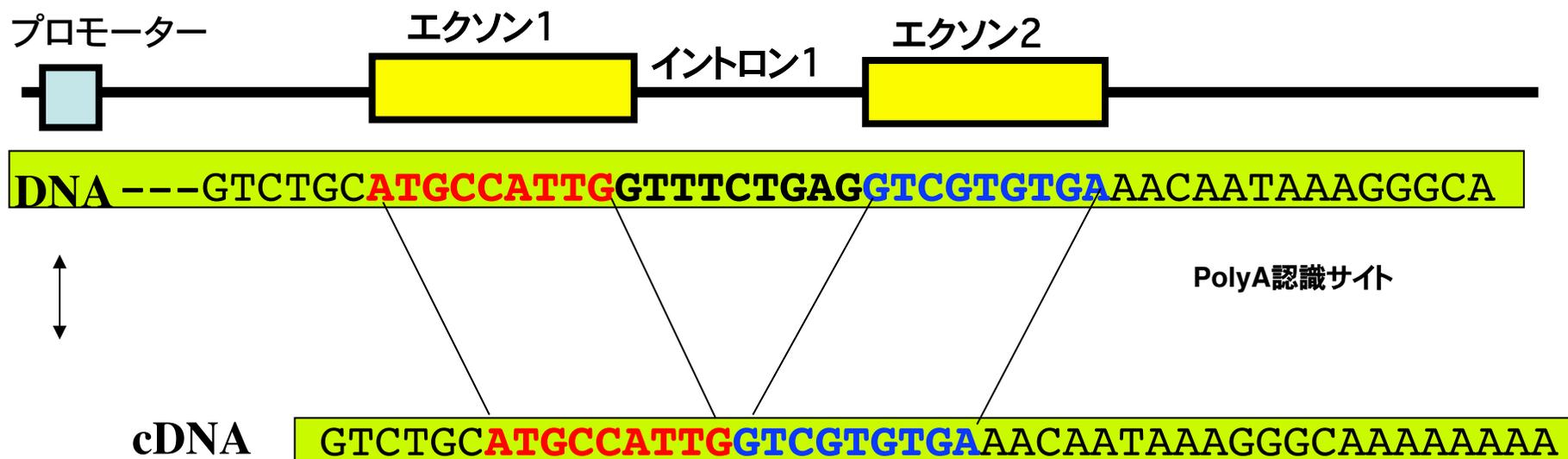
cDNA情報が有るとき.2

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
tctcccgtg taacaatac caaccctctc gccacagacg ctgcccctgc acacacacac acactaacac tcacacaacg tctaacatca ctcccgtca

110     120     130     140     150     160     170     180     190     200
acacagagct cacttctaat ctcaaatctc atggtggcgc agcaccaccac gccaccacag ccgcaccaac caaccgccat ggactaacac gccggctctg

210     220     230     240     250     260     270     280     290     300
gcattggcgc cccggcggcg gccggcggtg gcgaccagag gatggagccc aagatcccgg agccatttgt gtggcgaac gccgacgcga gcccgcgctc

310     320     330     340     350     360     370     380     390     400
ggcggcggag ctggacaatgc ccgtggtcga cgtggcgctg ctccgcgacg gcgacgcgca ggggcctgcg cgcgcgcggg cgcagctggc cgcgcgctgc

410     420     430     440     450     460     470     480     490     500
gccacgcacg ggttatttca ggtgtccgag cacggcctcg acgcccctct ggcgcgcgcc gcgctgacgc gcgccacgca cttcttcgcg ctcccgtctg

510     520     530     540     550     560     570     580     590     600
ccgagaagcg ccgcgcgcgc cgcgtcccgg gccaccgtgc cggctaacac agcgcgccac ccgaccgctt cgcctccaaq ctcccattga agggagacct

610     620     630     640     650     660     670     680     690     700
ctccttcgag ttccacgacc gcgcgcgcgc ccccgctgtc gccagctact tctccagcac cctgcgcccc gacttcgctc caatggggta attaaacaga

710     720     730     740     750     760     770     780     790     800
tgggtggcag cattgcattt caaatttcaa acaaattcaa aaacacacga ccgagattat gctgataatc aacggctttg tgcgcgcagg aggygttacc

810     820     830     840     850     860     870     880     890     900
agaagtactg ccgaggagat aaggagctgt cgcctgacat catggaactc ctggagctga gccctggcgt ggagcagagc tacctacagg actctctgc

910     920     930     940     950     960     970     980     990     1000
ggacagcagc tcaattatgc ggtgcaacta ctaccgcgca tgcgccggag ccggagcgac gctgcggcac ggcgccacct gccaccacac cgcctctaac

1010    1020    1030    1040    1050    1060    1070    1080    1090    1100
atctctctcc agaacgagct ccgcgcgcctc gaggctctcg tcgacggcga atggcgcgcc gtcagcccgc tcccggcgcc catgtctacc aacatggcg

1110    1120    1130    1140    1150    1160    1170    1180    1190    1200
acaccttcac ggtaaacctc ctctctctct cctctctctc gttctctctc gcttcgaagc aacagaacaa gtaattcaag ctttttttct tctctcgcgc

1210    1220    1230    1240    1250    1260    1270    1280    1290    1300
gaaattgacg agaaaataa gatctggtga ggggcggggc tttcagctga aagcgggaag aaaccgacct gacgtgattt ctctgttcca atcaacaaca

1310    1320    1330    1340    1350    1360    1370    1380    1390    1400
atggaaatgcc ccaactctcc atgtgttatg atttatctca catcttatag ttaataggag taagtaacaa gctacttttt tcaattata gttgtttga

1410    1420    1430    1440    1450    1460    1470    1480    1490    1500
tttttttttt ttaagttttt tttagtttta tcaaaattta ttgaaaactc tagcaacgtt tataatacca aattagctctc atttagttta atattgtata

1510    1520    1530    1540    1550    1560    1570    1580    1590    1600
tattttgata atatatattt gttatattaa aaatattact atatttttct ataacatta ttaaaagcca tttataat ataatggaaq gagtattaa

1610    1620    1630    1640    1650    1660    1670    1680    1690    1700
tatggatctc ccccgcagct agaataattt ccgatggctg gacgacgcca tghtaagcttc ggtggcccctg gacggccaga ggtgccaaca gccacgctcca

1710    1720    1730    1740    1750    1760    1770    1780    1790    1800
acaaccccctg ggtccccecc taacactcca aacagtagtg agtagtgtct cgtcgcgttt tagtatttga tgacaacaa agtgtgattt gagttagcca

1810    1820    1830    1840    1850    1860    1870    1880    1890    1900
ccaccaactt gccacagagc acaataattt ggttccatc tcgccagcca ctccatctc tagctcctac tctctctag cgatgtaacg ggaataattc

1910    1920    1930    1940    1950    1960    1970    1980    1990    2000
atcatccgta tataaacctg tttgttatg ttaatttctc atataaactc ataacagtat acattttaaa agaaaacaaa attaggataa caagcccctg

2010    2020    2030    2040    2050    2060    2070    2080    2090    2100
ctctatcca 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 cggggacgga tgcgatgtgt gccgctttcg

2210    2220    2230    2240    2250    2260    2270    2280    2290    2300
gtggcgtttt cccggggggg ttttggtttg gcgctgggg gggatggcat ggcgcgcgct gcggctgacg gccacacaca cgcgcgagca cgcacgctac

```

(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 agcaccaccac gccaccacag ccgcaccaac caaccgccat ggactaacac gccggctctg gcattggcgc cccggcggcg gccggcgtgt

110     120     130     140     150     160     170     180     190     200
gccaccctgag gatggagccc aagatcccgg agccatttgt gtggcgaac gccgacgcga gcccgcgctc gccggcggag ctggacatgc ccgtggtcga

210     220     230     240     250     260     270     280     290     300
cgtggcgctg ctccgcgacg gcgacgcgca ggggcctgcg ccgcgcggcg ccgagctggc cgcgcgctgc gccacgcacg ggttatttca ggtgtccgag

310     320     330     340     350     360     370     380     390     400
cacggcgtgc acgcgcctct ggcgcgcgcc gcgctgacgc gccgcagcga cttatttcgc ctcccgtctg ccgagagagc ccgcgcgcgc cgcgtctccg

410     420     430     440     450     460     470     480     490     500
gccaccgtgt ccgctaacac agcgcgccac ccgaccgctt cgcctccaaq ctcccattga agggagacct cctctctcgc tctccagcac cctgcgcccc

510     520     530     540     550     560     570     580     590     600
ccccctgtct gccagctact tctccagcac cctgcgcccc gacttcgctc caatggggag ggtgtaccag aagctactgc agggagatgaa ggagctgtctg

610     620     630     640     650     660     670     680     690     700
ctgacatgta tggaaactct ggagctgagc ctggcgctgg agcggagcta ctacaggag tctctcgcgc acagcagctc aatcatgagg tgcacactct

710     720     730     740     750     760     770     780     790     800
accgccatgt cccggagcgc gaggcggcgc tggcagctgc ccgcaccacg cctctccatc cctctccag gacgagctag gccggcctga

810     820     830     840     850     860     870     880     890     900
ggctctctct gacggcgaat ggcgccacct cagcccctgc cccggcgcga tggtaataaa catcggcagc accttaatgg cgtgtctgaa cgggaggtat

910     920     930     940     950     960     970     980     990     1000
aagagctgac tgcacaggc ggtagtgaa cagcggcggg agcggcgttc gctggcgttc tctctctcgc cgcggggaga cagggtggct gcgcgcgcgc

1010    1020    1030    1040    1050    1060    1070    1080    1090    1100
cagagcgcgc caagcgcagc cactaccggg acttcaactg ggccgacctc atgccttca ccgagcgcga ctaccgcgc gccaccgcca cgtctgagcgc

1110    1120    1130    1140    1150    1160    1170    1180    1190    1200
cttccagcgc ttgctcgcgc cgcgcgcgcc ccgaccgcgc gccagcggcg aggttagagg gccacgctga

```

(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 ctcceagta
110     120     130     140     150     160     170     180     190     200
acacagact cacttatac ctcaaatct atggtggcg agcaccacc gccaccacag ccgcaccac caccgccat ggcataacc gccgcctg
210     220     230     240     250     260     270     280     290     300
gcattggcg cccggcgcg gccggcggt ggcagctgag gatggagccc aagatcccg agccattgt gtggcgaac ggcagcgca ggcggcgct
310     320     330     340     350     360     370     380     390     400
ggcggcgag ctggacatg ccgtggtgca cgtggcgct ctcgcgacg gcgacgcga ggggcctgc ccgcgccgg ccaggtggc cgcgcgtgc
410     420     430     440     450     460     470     480     490     500
gccacagac ggttattca ggtgtccag cacggctcg accgccctt ggcgcgccc gcgctgacg gccaccacga cttctcgc ctcaccctg
510     520     530     540     550     560     570     580     590     600
ccgagaagc ccgcgcgcg ccgctcccg gccaccgtg ccgctacac agcgcacac ccgaccgctt ccctccaa cctccatga aggagacct
610     620     630     640     650     660     670     680     690     700
ctcttcgac ttccagacc gccgcgcgc cccgctgct gccgacctt tctccagac cctgcgccg gacttcgac caatgggta attaaagca

```

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 ccgcaccac caccgccat ggactcacc gccgctctg gcattggcg cccggcgcg gccgggtgt
110     120     130     140     150     160     170     180     190     200
gccaccctgag gatggagccc aagatcccg agccattgt gtggcgaac ggcagcgca ggcggcgct gccggcgag ctggacatg ccgtgtgca
210     220     230     240     250     260     270     280     290     300
cgtggcgct ctcgcgacg gcgacgcga ggggcctgc ccgcgccgg ccaggtggc cgcgcgtgc gccaccacg ggttattca ggtgtccag
310     320     330     340     350     360     370     380     390     400
cacggctgag accgcctct ggcgcgccc gcgctgacg gccaccacga cttctcgc ctcaccctg ccgagaagc ccgcgcgcg ccgctcccg
410     420     430     440     450     460     470     480     490     500
gccaccgtg ccgctacac agcgcacac ccgaccgctt ccctccaa cctccatga aggagacct cctcttcgac ttccagacc gccgcgcgc
510     520     530     540     550     560     570     580     590     600
ccccctgct gccacctt ctcaccacg cctgcgccg gacttcgac caatgggag ggtgtaccg aagctcgc aggagatgaa ggagctgtg
610     620     630     640     650     660     670     680     690     700

```

DNA ---GTCTGCATGCCATTGGTTTCTGAGGTCGTGTGAACAATAAAGGGCA

cDNA GTCTGCATGCCATTGGTCGTGTGAACAATAAAGGGCAAAAAAAAA

(EMBL) length: 2900 current pos: 401

```

1610     1620     1630     1640     1650     1660     1670     1680     1690     1700
tatggatct cccgcacatg agaataatt ccgatgggt gacgacgca tghtaagctt ggtggccctg gaccgcccga ggtgccaca gccacgtcca
1710     1720     1730     1740     1750     1760     1770     1780     1790     1800
acaaccccg ggtcccacc taacactca aacagtagt agtagtgtt cgtcgcgtt tagtattga tgacaacaa agtgtgatt gagttagca
1810     1820     1830     1840     1850     1860     1870     1880     1890     1900
ccaccaact gccaccagc aataaattt ggttccatt ccgcagctc cttccatct tagtccaac tctatctag cagttagc ggataattc
1910     1920     1930     1940     1950     1960     1970     1980     1990     2000
ataatccga tataaacct tttgtatg ttaattctt atataactt ataacgat acattitaaa agaaaacaa attagataa acagcccctg
2010     2020     2030     2040     2050     2060     2070     2080     2090     2100
ctctatcca tcatggcac ttggaaggac cagactcgt catgcatgc caagccaaga tatgggttat ggaagtagt agaagaggat agatgagaga
2110     2120     2130     2140     2150     2160     2170     2180     2190     2200
taagcatgag ttctctact cgttgaggt gtattttgga gggatttga tagtagtag agcggcgccg cggggacgga tgcgatggt gccgctttcg
2210     2220     2230     2240     2250     2260     2270     2280     2290     2300
gtggcgttt cccggggggg ttttggttt gccctgggg gggatggcat ggcgcggcgt gcgctgac gccacacaca ccgcagcga cgcagctac

```

(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
tcctccctgt tacaaaatacc caaccctctct gccacagacag ctcgcccctgc aacacacacac aaactaacac tcacaacagc tctcaactca ctcocagctca
110     120     130     140     150     160     170     180     190     200
acacagcgcct caactctctcat ctccaatctc atggtggccg agcaaccacc gccaccacag ccgaccacac caccgcccat ggaactcaacc gccgcctctg
210     220     230     240     250     260     270     280     290     300
gcattggcgc cccggcgggc gccggcggtgt gccagcttagc gatggagccc aagatcccgc agcaattctgt gttggccagac gccgacgcga gccacggcgtc
310     320     330     340     350     360     370     380     390     400
gccggcgggag ctggacaatgc ccgtgtgtgca cgtggcgctg ctcgcgacg gccagcccca gggctggcgc cgcgcgcggc cgcagggtgc cgcgcgctgc
410     420     430     440     450     460     470     480     490     500
gccacgcacg ggtttctcca ggtgtctcag caaggcctag acgcgcctct ggagcgcgcc gccctagacg gccgcagcga cttcttccgc ctcocagctag
510     520     530     540     550     560     570     580     590     600
ccgagaaagcg ccgcgcgcgc ccgctcccgc gccacgtgtc cggctaacac agcgcacca cgcgcgcctt cgcctccaag ctccaatgga agggagacct
610     620     630     640     650     660     670     680     690     700
ctcacttgcc ttccacgacc gccgcgcgc ccccgctgct gccagcact tctccagcac cctggccccc gacttcgcgc caatggggta attaaacaga
710     720     730     740     750     760     770     780     790     800
tggtagcaga cacttgcatc caaatctcaa acaaatctaa aacacacaga ccgagattat gctgaattca aacgcgtttg tgcgcgcagg aggggtgtcc
810     820     830     840     850     860     870     880     890     900
agaagctact cagaggagtg aaggagctgt ccgtgacgat catggaactc ctggagctga gccctggcgt ggagcagacc tacataaccg agttcttgc
910     920     930     940     950     960     970     980     990     1000
ggacagcaga taactactgc ggtgcaacta ctaccgcgca tgcccggagc ccgagccgac gctggccagc gccgcgcact gccaccacca cgcacctaac
1010    1020    1030    1040    1050    1060    1070    1080    1090    1100
atctctctcc agnacagcgt ccggcgctcc gaagtctctg tcgacggcga atggcgcgcc gtaaccgccc tcccagcgcc catggtctat aacatcggg
1110    1120    1130    1140    1150    1160    1170    1180    1190    1200
aacacttaat ggtaaacact ctactattct cctctctctc gttctctctc gcttcgaagc aacagaaaca gtaattcaag ctttttttcc tctctcgcgc
1210    1220    1230    1240    1250    1260    1270    1280    1290    1300
gaaattgacg agaaaaataa gatctgtgta gggcgggggc ttccagctga aagcgggaag aaaccgacct gacgtgattt cctctgtcca 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 ttaaagtttt tttagtttta tcaaaaatta ttgaaaaact tagcaacgtt tataatacca aattagctct atttagttta atattgtata
1510    1520    1530    1540    1550    1560    1570    1580    1590    1600
tattttgata atatatattat gttatattaa aatatattact atattttct ataacaatta ttaaaagcca tttataat ataaatggaag gagtantaat
1610    1620    1630    1640    1650    1660    1670    1680    1690    1700
tatggatctc ccccgacatg agaatttttt ccgatgggtg gacgacgcca tgttaagctc ggtagggcct gacggccaga ggtgccaaca gccacgtcca
1710    1720    1730    1740    1750    1760    1770    1780    1790    1800
acaaccctgt ggtccccccc taacactcca aacagtagtg agtagtgtct cgtcgcgttt tagtatttga tgacaacaac agtgtgattt gagtttagca
1810    1820    1830    1840    1850    1860    1870    1880    1890    1900
ccaccacatt gccacagcgc acatacattt gttgcaattc tcgcccagtea cttccactct tagtctaac tectatctag cagttaagc ggataatttc
1910    1920    1930    1940    1950    1960    1970    1980    1990    2000
atcatcctga tataaacctg tttgttatag ttaatttctc atataatact ataacagtat acattttaaa agaaaacaac attaggataa acagaccctg
2010    2020    2030    2040    2050    2060    2070    2080    2090    2100
ctctatctca tcaatggcac ttgnaaggac cagactcgtt catgccatgc caagccaaga tatgggttat ggaagagtag agaagaggag agatgagaga
2110    2120    2130    2140    2150    2160    2170    2180    2190    2200
taagcagcgc ttctactctc cgttggatgt gatttttggg gggatttttg tagtagtagc agcgcgcgca cggggacaga tgcggatggt gccgcctttc
2210    2220    2230    2240    2250    2260    2270    2280    2290    2300
gtggcgtttt cccggggggg ttttgatttg gccctggggg gggatggcat ggcgcgcgct gccgctgcaac gccaccacca 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 ggaactcaacc gccgcctctg gcattggcgc cccggcgggc gccggcggtgt
110     120     130     140     150     160     170     180     190     200
gccagcttagc gatggagccc aagatcccgc agcaattctgt gttggccagac gccgacgcga gccacggcgtc cgtgtgtgca cgtgtgtgca cgtgtgtgca
210     220     230     240     250     260     270     280     290     300
cgtgtgtgct ctcgcgacg gccagcccca gggctggcgc cgcgcgcggc cgcagggtgc cgcgcgctgc gccacgcacg ggtttctcca ggtgtctcag
310     320     330     340     350     360     370     380     390     400
cagcgcctag acgcgcctct ggagcgcgcc gccctagacg gccgcagcga cttcttccgc ctcocagctag ccgagaaagc ccgcgcgcgc cgcctccagg
410     420     430     440     450     460     470     480     490     500
gcaacgtgtc cggctaacac agcgcacca cgcgcgcctt agcaactcaag ctccaatgga agggagacct ctcacttgcc ttccacgacc gccgcgcgc
510     520     530     540     550     560     570     580     590     600
ccccgtgct gccagcact tctccagcac cctggccccc gacttcgcgc caatggggag ggtgtaccag aagctctgag agggagtgaa ggagcgtctg
610     620     630     640     650     660     670     680     690     700
ctgacagtaa tggaaactct ggagctgagc ctggggctgg agcagagcta ctacagggag ttctctgccc acagcagctc aaatctgccc tgcaactact
710     720     730     740     750     760     770     780     790     800
accgcgcctg cccggcgagg gacggcagca tggcgcaggg ccgcgcactc gccaccacag cctcaacctc cctctctcag gacgagctag gccggctaga
810     820     830     840     850     860     870     880     890     900
ggtctctgct gacggcgaa tggcgcagct cagccagctc cccggcgcca tggtaataca catcggcagc acctcaatgg ccctgtctga cgggaggtat
910     920     930     940     950     960     970     980     990     1000
aagagctgct tgcacagggc ggtgtgtgaa cagcgcggcg agcggcggtc gctggcgttc ttctgtgccc ccggggagga caggggtgtg ccgcgcgcgc
1010    1020    1030    1040    1050    1060    1070    1080    1090    1100
cgcgcgcgc cagcgcagc cactaccagg acttcaactg ggccgacct atgctctca ccgagcagca ctaccgcgcc gacaccagca ccctcagcgc
1110    1120    1130    1140    1150    1160    1170    1180    1190    1200
cttccagcgc tggctcgcgc ccgcgcgcgc ccagcgcgcc gccagcgcgc aggtcagagc ggccagctga

```

(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 bioinformatics software interface with two windows side-by-side. The left window is titled "GA20oxidase2-genomic DNA" and the right window is titled "GA20oxidase2-cDNA". Both windows display sequence alignments with nucleotide bases and position markers. The left window shows a sequence of 2900 BP, and the right window shows a sequence of 1170 BP. The interface includes a menu bar at the top with options like "Finder", "ファイル", "編集", "表示", "移動", "ウインドウ", and "ヘルプ". The status bar at the bottom of each window shows the current position and length of the sequence.

GA20oxidase2-genomic DNA
ID: XX
SQ: Sequence 2900 BP; 608 A; 864 C; 784 G; 644 T;

GA20oxidase2-cDNA
ID: C20 oxidase for a candidate gen of sd1
SQ: Sequence 1170 BP; 171 A; 455 C; 152 T; 392 G.

GA20oxidase2-genomic DNA Sequence (partial):

```
10 20 30 40 50 60 70 80 90 100
tcctccctgt tacaataacc ccaccctcct gccacagacg ctgcccctgc acacacacac acactcaaac tcacacacgc tcataactca ctcaccgctca
110 120 130 140 150 160 170 180 190 200
acacagcgcg caactctcat ctaacatctc atggtggcgc agcaaccacg gccaacacag ccgcacaacac caaccgcecat ggactcaacc gccgcctctg
210 220 230 240 250 260 270 280 290 300
gcaattgcgc ccagcgcgcg gccggcgggt gcgaccctgag gatggagccc aagatcccgg agccattctg tggccagaac ggcagcgcga gcccgcgctc
310 320 330 340 350 360 370 380 390 400
ggcggcggag ctggacaatgc ccgtggtcga cgtggcgctg ctcgcgcagc gcgacgcgca ggggcctgcg ccgcgcgcgc cgcaggtggc cgcgcgcgtgc
410 420 430 440 450 460 470 480 490 500
gccacgcacg ggttcttcca ggtgtccgag caaggcctgc acgcgcctct ggcgcgcgcg gcgcctgacg gccaccagca cttcttcgcg ctcaccgctgc
510 520 530 540 550 560 570 580 590 600
ccgagaagcg ccagcgcgcg ccgctcccgg gccaccctgct ccgctacacc agcgcaccaag ccagcgcctt ccgctccaag ctcccatgga aggagaccct
610 620 630 640 650 660 670 680 690 700
ctctctgcgc ttccacgacc gccgcgcgcg ccccgctgct gccgcctact tctccagcac cctcggcccc gactctgcgc caatggggta attaaacga
710 720 730 740 750 760 770 780 790 800
tggtggacga cattgcattt caaatcaaaa caaatcaaaa aacacacaga ccgagattat gctgaaatca aaccgctttg tgcgcgcgcg aggggtgtacc
810 820 830 840 850 860 870 880 890 900
agaagtactg ccagagagat gaaggactgt cgtgacgat catggaaact ctggagctga gccctggcgt ggagcggagc tactacaggg agttcttcgc
910 920 930 940 950 960 970 980 990 1000
ggacagcgcg tcaatcatgc ggtgcaacta ctaccgcgca tgcgcggagc ccgagcggag cctcggcacc ggcgcgcacc gccaccacac cgcctcaacc
1010 1020 1030 1040 1050 1060 1070 1080 1090 1100
atctctctcc aggaacagct ccggcgccct gaggctctcg tcgacggaga atggcgcgcc gtcagccccc tcccaggcgc catgtctatc aacatagcgg
1110 1120 1130 1140 1150 1160 1170 1180 1190 1200
aaccttcatg ggtaaacatc ctctctattc cctctctctt gttctctctt gcttgcagcc aacagaaaca gtaattcaag ctttttttct tctctgcgcg
1210 1220 1230 1240 1250 1260 1270 1280 1290 1300
gaaattgacg agaaaaataa gatctggtga gggcgggggc tttaagctga aagcgggaag aaaccagact gaactgtatt cctctgtcca atcaacaaca
1310 1320 1330 1340 1350 1360 1370 1380 1390 1400
atggaatgac caactctctc atgtgttatg atttatctca catcttatag ttaataggag taagtaacaa gctacttttt tcaattata gttctgttga
1410 1420 1430 1440 1450 1460 1470 1480 1490 1500
tttttttttt ttaaaatttt tttagtttta tcaaaattta ttgaaaaact tagcaacgctt tataatacca aattagctctc atttagttta atattgtata
1510 1520 1530 1540 1550 1560 1570 1580 1590 1600
tattttgata atatatattt gttatattaa aaatattact atattttctc ataaacotta ttaaaagcca ttataatat aaaaatgaag gagtaattaa
1610 1620 1630 1640 1650 1660 1670 1680 1690 1700
tatggatctc ccccgacatg agaataattt ccgatgggtg gacgacgcca tghtaagcttc ggtgggcctg gacgcccaga ggtgcccaca gccacgtcca
1710 1720 1730 1740 1750 1760 1770 1780 1790 1800
acaaccctcg ggtccccccc taaactccca aacagtagtg agtagtgtct cgtcgcgctt tagtatgta tgacaacaaa agtgtagatt gaggtagcca
1810 1820 1830 1840 1850 1860 1870 1880 1890 1900
ccaccaactt gcaacagcgc aatatacttt ggtgcaattc tgcgcagtea cttccatctc tagctcaaac tctatactag cagtgtgaag ggaataattc
1910 1920 1930 1940 1950 1960 1970 1980 1990 2000
atcatcagta tataaacctg ttgttatag ttaattctct atataactc ataacagtat acattttaaa agaaaaacaa attaggataa acaggccctg
2010 2020 2030 2040 2050 2060 2070 2080 2090 2100
ctctatctca tcaatgacac ttggaaggag cagactcggg catgccatgc caagccaaga tatgggttat ggaagatag agaagaggag agatgagaga
2110 2120 2130 2140 2150 2160 2170 2180 2190 2200
taagcatgag ttctctctct cgttgatgtg gtattttgga gggattttgt tagtagtagc agcgcgcgcg cggggacaga tgcggtggtt ggcgccttgc
2210 2220 2230 2240 2250 2260 2270 2280 2290 2300
gtggcgcttt ccgggggggg ttttggttgg gccttgggg gggatggcat ggcgcgcgcg ccgcctgcaac gccacaacaa ccgcgcgcga cgcacgtacg
```

GA20oxidase2-cDNA Sequence (partial):

```
10 20 30 40 50 60 70 80 90 100
atggtggcgc agcaaccacg gccacacagc ccgcacaacac caaccgcecat ggactcaacc gccgcctctg gcaattgcgc ccggcgggag gccggcgggt
110 120 130 140 150 160 170 180 190 200
gccacctgag gatggagccc aagatcccgg agccattctg tggccagaac ggcagcgcga gcccgcgctc gcccgcgctc cgtggcctgag ccgtggtcga
210 220 230 240 250 260 270 280 290 300
cgtggcctgag ctcaccgacg gcgacgcgca ggggcctgcg ccgcgcgcgc cgcaggtggc cgcagcgcgc gccaccagag ggtcttctca ggtgtccgag
310 320 330 340 350 360 370 380 390 400
caaggcctgc acgcgcctct ggcgcgcgcg gcgcctgacg gccaccagca cttcttcgcg ctcaccgctgc ccgagaagcg ccgcgcgcgc cgcctcccg
410 420 430 440 450 460 470 480 490 500
gccaccgtgc ccgctcaacc agcgcaccaag ccagcgcctt ccgctccaag ctcccatgga aggagaccct cctcttcgag ttccacgacc gccgcgcgcg
510 520 530 540 550 560 570 580 590 600
cccgctgctc gccgcctact tctccagcac cctcggcccc gactctgcgc caatggggag ggtgtaccag aagctctgcg aggagatgaa ggagctgtcg
610 620 630 640 650 660 670 680 690 700
ctgacgatac tggaaatctt ggagctgagc cttggcgtgg agcagggcga ctacagggag ttcttcgagg acagcgcgc caatactcgg tgcaactact
710 720 730 740 750 760 770 780 790 800
accgcgcaat ccagcgcgcg gacgagcgcg tcggcagcgg ccgcgcctgc gaaccacagc cctcaacatc cctctctcag gacgagctgc gccgcctcga
810 820 830 840 850 860 870 880 890 900
ggtctctgct gacgcgcaat ggcgcgcctg cagcgcctgc cccggcgcca ttgtaataca catcggcgcg accttcaatg ccgtgtcga cgggaggtat
910 920 930 940 950 960 970 980 990 1000
aagagctgac tgcacagggc ggtgtgtaac cagcgcgggg agcgcggctc gctggcgttc ttctctgccc ccgaggagga cagggtgggt ccgcgcgcgc
1010 1020 1030 1040 1050 1060 1070 1080 1090 1100
cgagcgcgcg ccgcgcgcag cactaccggg acctcaactg gccaccgcct atgcgttca agcagcgcga ctaccgcgcg gacaccgca cgtctgagc
1110 1120 1130 1140 1150 1160 1170 1180 1190 1200
cttcaacgcg tggctagcgc ccgcgcgcgc ccagcgcgcg gccagcgggc aggtcgaagg gccacagctga
```

Status: (EMBL) Length: 2900 current pos: 481 (Left window) / (EMBL) Length: 1170 current pos: 1171 (Right window)

Finder ファイル 編集 表示 移動 ウィンドウ ヘルプ

GA20oxidase2-genomic DNA

Untitled 3

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
ACACAGCGCT	CACCTTCTAT	CTCCAACTTC	ATGTTGGCCG	AGCACCCAC	GCCACCCAC	CCGACCCAC	CACCCGCCAT	GGACTCCAC	GCCGGCTCG
210	220	230	240	250	260	270	280	290	300
GCATTGCGCG	CCCGCGCGCG	GCGCGGGTGT	GCGACTGAG	GATGGAGCCC	AGATCCCGG	AGCCATTCTG	GTGGCCGAC	GCGACGCCA	GCCCGGGCTC
310	320	330	340	350	360	370	380	390	400
GCGCGCGAG	CTGGACATGC	CCGTGGTCGA	GCTGGGCGTG	CTCCGCGACG	GCGACGCCA	GGGGCTGCCG	CGCCCGCGG	GCGAGGTGG	CGCCCGCTGC
410	420	430	440	450	460	470	480	490	500
GCCACGACG	GGTCTTCCA	GGTGTCCGAG	CACGGCGCTC	ACGCCGCTC	GCGCGCGCC	GCGCTGACG	GCGCCAGCG	CTTCTCCCG	CTCCCCTGC
510	520	530	540	550	560	570	580	590	600
CCGAGAGCG	CCGCGCGCG	GCGTCCCGG	GCCCGCTC	GCGTACCC	AGCGCCACG	CCGACCGCT	CGCTCCAG	CTCCCATGG	AGGAGACCT
610	620	630	640	650	660	670	680	690	700
CTCCTTCGG	TTCCAGACC	GCGCGCGCG	CCCGCTGTC	GCGACTACT	TCTCCAGAC	CCTCGGCC	GACTTCGCG	CAATGGGTA	ATTAAACGA
710	720	730	740	750	760	770	780	790	800
TGGTGGACG	CATTGCATT	CAAAATTCAA	ACAATTCAR	AACACACCG	CCGAGATTG	GCTGAATTC	ACCGGTTTG	TGGCGGACG	AGGTTGTACC
810	820	830	840	850	860	870	880	890	900
AGAATGACT	CGAGGAGTG	AGGAGCTGT	CGCTGACAT	CATGGACTC	CTGGAGCTG	GCCTGGGCT	GGAGCGAGG	TACTACAGG	AGTTCTCCG
910	920	930	940	950	960	970	980	990	1000
GGACGACG	TCAATCATG	GGTGCACAT	CTACCCGCA	TGCCCGGAC	GCGAGCGAC	GCTCGGACG	GCCCGGAC	GCGACCCAC	GCCCGCTCC
1010	1020	1030	1040	1050	1060	1070	1080	1090	1100
ATCCTCTCT	AGGACGACT	GCGCGGCTC	GAGGTCCTG	TCGACGCCA	ATGGCGCCC	GTCAGCCCC	CATGGTCTC	ACACTCGCG	
1110	1120	1130	1140	1150	1160	1170	1180	1190	1200
ACACCTTCAT	GTAARACAT	CTCCTATTCT	CCTCTCCTCT	GTTCTCCTCT	GCTTCGAGC	AACACACCA	GTAATTCAG	CTTTTTTTC	TCTCTCGCG
1210	1220	1230	1240	1250	1260	1270	1280	1290	1300
GAAATGACG	AGAAAATAT	GATCGTGGT	GCGCGGGGC	TTTCACTCA	AGCGGGGAC	AAACCGACT	GCGGATTT	CTCTGTTCA	ATCACACCA
1310	1320	1330	1340	1350	1360	1370	1380	1390	1400
ATGGAAATG	CCACTCTCC	ATGTTGATG	ATTATCTCA	CATCTTATG	TTAATAGAG	TAAATACCA	GCTACTTTT	TCATATTAT	GTTCTGTTG
1410	1420	1430	1440	1450	1460	1470	1480	1490	1500
TTTTTTTTT	TTAAAGTTT	TTTAGTTTA	TCCAAATTA	TTGAAARCT	TAGCAACGT	TATATACCA	AAATGACTT	ATTTAGTTA	ATATTGATA
1510	1520	1530	1540	1550	1560	1570	1580	1590	1600
TATTTTGTA	ATATATTTT	GTTATATTA	AAATATTTT	ATATTTTCT	ATAAACATTA	TTAAAGCCA	TTTATATTA	AAATGGAG	GAGTATTTA
1610	1620	1630	1640	1650	1660	1670	1680	1690	1700
TATGGATCT	CCCCGACATG	AGAAATTTT	CCGATGGTG	GACGACCCA	TGTAGCTTC	GTTGGGCTG	GACGCGCAG	GTTGCCACA	GCCACTCCA
1710	1720	1730	1740	1750	1760	1770	1780	1790	1800
ACACCCCTG	GGTCCCGCC	TAACTTCCA	ACAGTAGTG	AGTAGTGTCT	CGTCCGTTT	TAGTATTTG	TGACAAACA	AGTGTAGTT	GAGTTAGCA
1810	1820	1830	1840	1850	1860	1870	1880	1890	1900
CCACCAACT	GACACGAGC	ACATACATT	GTGTCATTC	TCGCCAGTC	CTTCCATCT	TAGCTTAC	TCCTATCTG	CGATGAGC	GGATATTTT
1910	1920	1930	1940	1950	1960	1970	1980	1990	2000
ATCATCGTA	TATAACCTG	TTTGTATAG	TTAATTTCT	ATATAACT	ATAACGAT	ACATTTTAA	AGAAACAAA	ATTAGGATA	ACAGCCCTG
2010	2020	2030	2040	2050	2060	2070	2080	2090	2100
CTCCTATCCA	TCCATGGAC	TTGGAGGAC	CAGACTCGT	CATGCCATG	CAGGCCAGC	TATGGTTAT	GGAGGAGG	AGATGAGGA	
2110	2120	2130	2140	2150	2160	2170	2180	2190	2200
TAGCATGCG	TTCTCTCTC	CGTGGATGT	GATTTTTGA	GGGATTTGT	TAGTAGTAG	AGCGCGCCG	CGGGACCGA	TGCGATGGT	GCCGTTTCG
2210	2220	2230	2240	2250	2260	2270	2280	2290	2300
GTGGCGTTT	CCCGGGGGG	TTTTGGTTG	GCGCTTGGG	GGGATGGCAT	GCGCGCGCT	GCGCTGAC	GCCACACCA	GCGCGCGCG	GCGACGACG
2310	2320	2330	2340	2350	2360	2370	2380	2390	2400
TCGTGCTGC	GCGCGCGCG	CGTAGCTTA	GGTGGTGTG	TTCCGCGCG	GCGCGCGAT	TGTTCCATG	GATCGATTT	GCGCGCCCG	TCGCCCGCG
2410	2420	2430	2440	2450	2460	2470	2480	2490	2500
TCTTGTGCG	TGTTGCGCT	CTCTCGCGG	GTTTGTCTT	GTCGCGTTC	TCAGCCGCG	ACGGGGGAC	GACATTTGC	GATGTAGCC	TGACGTTGC
2510	2520	2530	2540	2550	2560	2570	2580	2590	2600
GCCCTTCCG	TTGATGATG	ATGATGATG	TATGATTTT	TTTTGTCTG	AGGAAATTT	TGGGAAATG	TTGTGTGTC	AGCGCTGTC	GACCGGGAG
2610	2620	2630	2640	2650	2660	2670	2680	2690	2700
TATAAGACT	GCCTGCACG	GCGCGGGTG	AACCGCGCG	GGAGCGCGG	GTCGCTGGC	TTCTTCTGT	GCCCGCGGA	GGACAGGGT	GTGCGCGCG
2710	2720	2730	2740	2750	2760	2770	2780	2790	2800
GCCCGACCG	GCCCGACCG	CAGCATTAC	CGGACTTAC	CTAGCGCGC	CTCAGCGCT	TCACCGCGC	CCACTACCG	GCCGACCGC	GACCGCTCA
2810	2820	2830	2840	2850	2860	2870	2880	2890	2900
GCCCTTCCG	CGCTGGCTG	GCCCGCGCG	GCGCGCGCG	GCGCGCGCG	GCGCGCGCG	TGATCCCGA	ACGGACCGA	ACGACCGAA	

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

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
ACACAGCGCT	CACCTTCTAT	CTCCAACTTC	ATGTTGGCCG	AGCACCCAC	GCCACCCAC	CCGACCCAC	CACCCGCCAT	GGACTCCAC	GCCGGCTCG
210	220	230	240	250	260	270	280	290	300
GCATTGCGCG	CCCGCGCGCG	GCGCGGGTGT	GCGACTGAG	GATGGAGCCC	AGATCCCGG	AGCCATTCTG	GTGGCCGAC	GCGACGCCA	GCCCGGGCTC
310	320	330	340	350	360	370	380	390	400
GCGCGCGAG	CTGGACATGC	CCGTGGTCGA	GCTGGGCGTG	CTCCGCGACG	GCGACGCCA	GGGGCTGCCG	CGCCCGCGG	GCGAGGTGG	CGCCCGCTGC
410	420	430	440	450	460	470	480	490	500
GCCACGACG	GGTCTTCCA	GGTGTCCGAG	CACGGCGCTC	ACGCCGCTC	GCGCGCGCC	GCGCTGACG	GCGCCAGCG	CTTCTCCCG	CTCCCCTGC
510	520	530	540	550	560	570	580	590	600
CCGAGAGCG	CCGCGCGCG	GCGTCCCGG	GCCCGCTC	GCGTACCC	AGCGCCACG	CCGACCGCT	CGCTCCAG	CTCCCATGG	AGGAGACCT
610	620	630	640	650	660	670	680	690	700
CTCCTTCGG	TTCCAGACC	GCGCGCGCG	CCCGCTGTC	GCGACTACT	TCTCCAGAC	CCTCGGCC	GACTTCGCG	CAATGGGTA	ATTAAACGA
710	720	730	740	750	760	770	780	790	800
TGGTGGACG	CATTGCATT	CAAAATTCAA	ACAATTCAR	AACACACCG	CCGAGATTG	GCTGAATTC	ACCGGTTTG	TGGCGGACG	AGGTTGTACC
810	820	830	840	850	860	870	880	890	900
AGAATGACT	CGAGGAGTG	AGGAGCTGT	CGCTGACAT	CATGGACTC	CTGGAGCTG	GCCTGGGCT	GGAGCGAGG	TACTACAGG	AGTTCTCCG
910	920	930	940	950	960	970	980	990	1000
GGACGACG	TCAATCATG	GGTGCACAT	CTACCCGCA	TGCCCGGAC	GCGAGCGAC	GCTCGGACG	GCCCGGAC	GCGACCCAC	GCCCGCTCC
1010	1020	1030	1040	1050	1060	1070	1080	1090	1100
ATCCTCTCT	AGGACGACT	GCGCGGCTC	GAGGTCCTG	TCGACGCCA	ATGGCGCCC	GTCAGCCCC	CATGGTCTC	ACACTCGCG	
1110	1120	1130	1140	1150	1160	1170	1180	1190	1200
ACACCTTCAT	GTAARACAT	CTCCTATTCT	CCTCTCCTCT	GTTCTCCTCT	GCTTCGAGC	AACACACCA	GTAATTCAG	CTTTTTTTC	TCTCTCGCG
1210	1220	1230	1240	1250	1260	1270	1280	1290	1300
GAAATGACG	AGAAAATAT	GATCGTGGT	GCGCGGGGC	TTTCACTCA	AGCGGGGAC	AAACCGACT	GCGGATTT	CTCTGTTCA	ATCACACCA
1310	1320	1330	1340	1350	1360	1370	1380	1390	1400
ATGGAAATG	CCACTCTCC	ATGTTGATG	ATTATCTCA	CATCTTATG	TTAATAGAG	TAAATACCA	GCTACTTTT	TCATATTAT	GTTCTGTTG
1410	1420	1430	1440	1450	1460	1470	1480	1490	1500
TTTTTTTTT	TTAAAGTTT	TTTAGTTTA	TCCAAATTA	TTGAAARCT	TAGCAACGT	TATATACCA	AAATGACTT	ATTTAGTTA	ATATTGATA
1510	1520	1530	1540	1550	1560	1570	1580	1590	1600
TATTTTGTA	ATATATTTT	GTTATATTA	AAATATTTT	ATATTTTCT	ATAAACATTA	TTAAAGCCA	TTTATATTA	AAATGGAG	GAGTATTTA
1610	1620	1630	1640	1650	1660	1670	1680	1690	1700
TATGGATCT	CCCCGACATG	AGAAATTTT	CCGATGGTG	GACGACCCA	TGTAGCTTC	GTTGGGCTG	GACGCGCAG	GTTGCCACA	GCCACTCCA
1710	1720	1730	1740	1750	1760	1770	1780	1790	1800
ACACCCCTG	GGTCCCGCC	TAACTTCCA	ACAGTAGTG	AGTAGTGTCT	CGTCCGTTT	TAGTATTTG	TGACAAACA	AGTGTAGTT	GAGTTAGCA
1810	1820	1830	1840	1850	1860	1870	1880	1890	1900
CCACCAACT	GACACGAGC	ACATACATT	GTGTCATTC	TCGCCAGTC	CTTCCATCT	TAGCTTAC	TCCTATCTG	CGATGAGC	GGATATTTT
1910	1920	1930	1940	1950	1960	1970	1980	1990	2000
ATCATCGTA	TATAACCTG	TTTGTATAG	TTAATTTCT	ATATAACT	ATAACGAT	ACATTTTAA	AGAAACAAA	ATTAGGATA	ACAGCCCTG
2010	2020	2030	2040	2050	2060	2070	2080	2090	2100
CTCCTATCCA	TCCATGGAC	TTGGAGGAC	CAGACTCGT	CATGCCATG	CAGGCCAGC	TATGGTTAT	GGAGGAGG	AGATGAGGA	
2110	2120	2130	2140	2150	2160	2170	2180	2190	2200
TAGCATGCG	TTCTCTCTC	CGTGGATGT	GATTTTTGA	GGGATTTGT	TAGTAGTAG	AGCGCGCCG	CGGGACCGA	TGCGATGGT	GCCGTTTCG
2210	2220	2230	2240	2250	2260	2270	2280	2290	2300
GTGGCGTTT	CCCGGGGGG	TTTTGGTTG	GCGCTTGGG	GGGATGGCAT	GCGCGCGCT	GCGCTGAC	GCCACACCA	GCGCGCGCG	GCGACGACG
2310	2320	2330	2340	2350	2360	2370	2380	2390	2400
TCGTGCTGC	GCGCGCGCG	CGTAGCTTA	GGTGGTGTG	TTCCGCGCG	GCGCGCGAT	TGTTCCATG	GATCGATTT	GCGCGCCCG	TCGCCCGCG
2410	2420	2430	2440	2450	2460	2470	2480	2490	2500
TCTTGTGCG	TGTTGCGCT	CTCTCGCGG	GTTTGTCTT	GTCGCGTTC	TCAGCCGCG	ACGGGGGAC	GACATTTGC	GATGTAGCC	TGACGTTGC
2510	2520	2530	2540	2550	2560	2570	2580	2590	2600
GCCCTTCCG	TTGATGATG	ATGATGATG	TATGATTTT	TTTTGTCTG	AGGAAATTT	TGGGAAATG	TTGTGTGTC	AGCGCTGTC	GACCGGGAG
2610	2620	2630	2640	2650	2660	2670	2680	2690	2700
TATAAGACT	GCCTGCACG	GCGCGGGTG	AACCGCGCG	GGAGCGCGG	GTCGCTGGC	TTCTTCTGT	GCCCGCGGA	GGACAGGGT	GTGCGCGCG
2710	2720	2730	2740	2750	2760	2770	2780	2790	2800
GCCCGACCG	GCCCGACCG	CAGCATTAC	CGGACTTAC	CTAGCGCGC	CTCAGCGCT	TCACCGCGC	CCACTACCG	GCCGACCGC	GACCGCTCA
2810	2820	2830	2840	2850	2860	2870	2880	2890	2900
GCCCTTCCG	CGCTGGCTG	GCCCGCGCG	GCGCGCGCG	GCGCGCGCG	GCGCGCGCG	TGATCCCGA	ACGGACCGA	ACGACCGAA	

(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 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

```
cggtgtgaaccagcggcgggagcggcggcgtcgtggcgttcttctgtgccg
cgggaggacaggggtgtgcggccgagcggccgacgcccagcagca
ctaccggacttcaactgggcccgaacctatgcttcaacgagccactaccg
cgccgacacccgacgctcgaacgcttcaacgctggtcgcggccggcggc
ccgacggccgagcggcaggtcgaagcggccagctgacgcaacg
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

```
gtataagagctgctgcacagggcgggtgtgaaccagcggcgggagcggcg
gtcgtggcgttcttctgtgcccggggagaggggtgtgcccggcggc
cgagcggccacggcggcagcactaccggacttcaactgggcccgaacctatg
cgcttcaacgagcggcactaccggcggcagcaccgacgctcgaacgcttcaacg
cgctgctcgcggccggccgacggccgacggcggcaggtcag
gcgccagctga
```

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
Exon 1	131-687	1-557	557	100.0%	0	0	d	
Exon 2	790-1112	558-880	323	100.0%	0	0	d	a
Exon 3	2584-2873	881-1170	290	100.0%	0	0		a

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

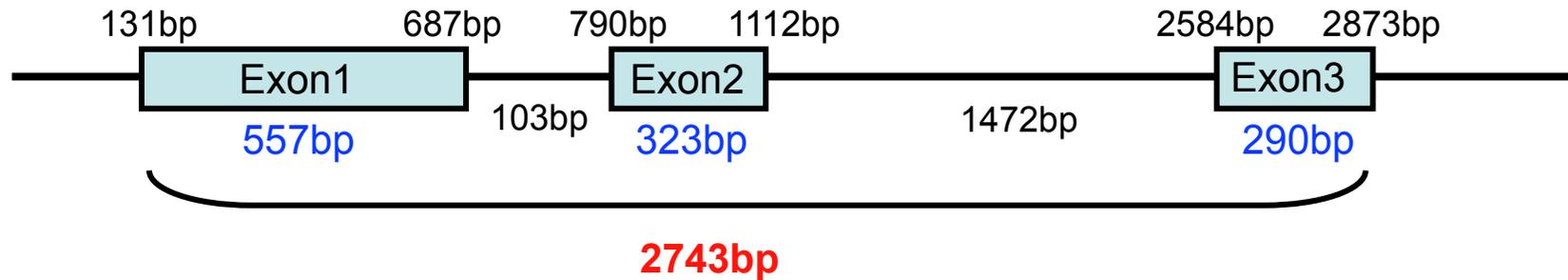
```

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
          N N G D T D M E R Y T R E R E V
    
```

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](#)

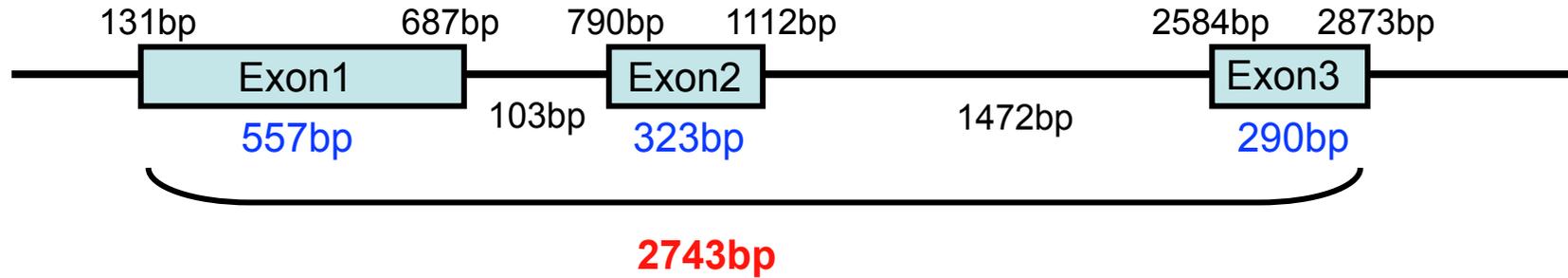
[ある遺伝子XのcDNA配列](#)

[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.

- | | |
|----------------------------------|--|
| nucleotide blast | Search a nucleotide database using a nucleotide query
<i>Algorithms:</i> blastn, megablast, discontinuous megablast |
| protein blast | Search protein database using a protein query
<i>Algorithms:</i> 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 nucleotide 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 Usir

A powerful feature 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

Job Title

Enter a descriptive title for your BLAST search

Blast 2 sequences

Choose Search Set

Database

Organism Optional

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.

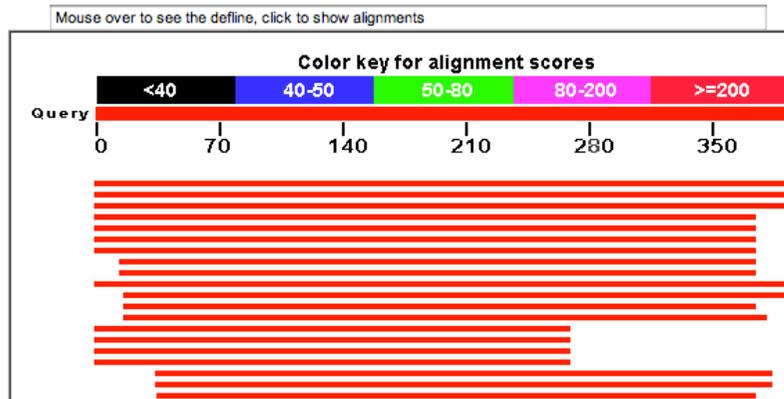
RID: GU22PYZU011

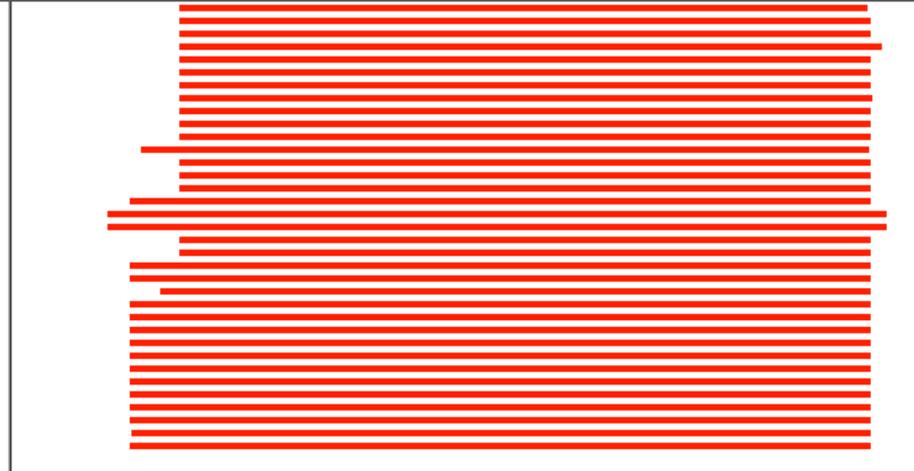
Database: All non-redundant GenBank CDS translations+PDB+SwissProt+PIR+PRF excluding environmental samples from WGS projects 7,256,259 sequences; 2,507,309,951 total letters

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



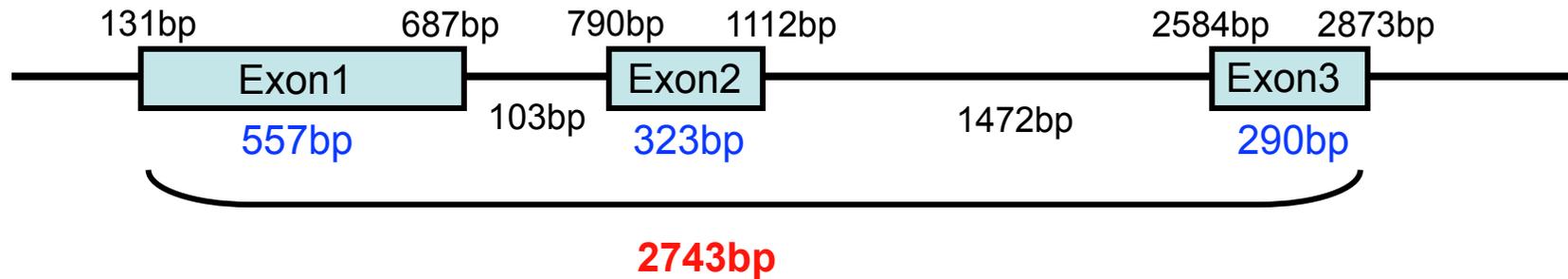


[Distance tree of results](#) **NEW**

Sequences producing significant alignments:

		Score (Bits)	E Value	
ref NP_001045014.1 	Os01g0883800 [Oryza sativa (japonica cult...	793	0.0	UG
qb AAN73384.1 	putative gibberellin 20 oxidase [Oryza rufipog...	788	0.0	
dbj BAG80954.1 	GA C20oxidase2 [Oryza glumipatula]	786	0.0	
dbj BAG80957.1 	GA C20oxidase2 [Oryza punctata]	671	0.0	
dbj BAG80956.1 	GA C20oxidase2 [Oryza longistaminata]	668	0.0	
dbj BAG80955.1 	GA C20oxidase2 [Oryza meridionalis]	668	0.0	
dbj BAG80960.1 	GA C20oxidase2 [Oryza alta]	659	0.0	
dbj BAG80959.1 	GA C20oxidase2 [Oryza officinalis]	644	0.0	
dbj BAG80958.1 	GA C20oxidase2 [Oryza minuta]	638	0.0	
dbj BAG80961.1 	GA C20oxidase2 [Oryza brachyantha]	637	0.0	
dbj BAG80962.1 	GA C20oxidase2 [Oryza granulata]	626	1e-177	
qb EAY76733.1 	hypothetical protein OsI_004580 [Oryza sativa ...	599	1e-169	
qb EAZ14396.1 	hypothetical protein OsJ_004221 [Oryza sativa ...	570	5e-161	
dbj BAG72281.1 	gibberellin-20 oxidase-2 [Oryza sativa Indica...	546	1e-153	
dbj BAE48438.1 	gibberellin-20 oxidase-2 [Oryza sativa Indica...	543	1e-152	
dbj BAG72282.1 	gibberellin-20 oxidase-2 [Oryza rufipogon]	542	2e-152	
dbj BAE48455.1 	gibberellin-20 oxidase-2 [Oryza glumipatula]	542	2e-152	
qb AAT44252.1 	putative gibberellin 20-oxidase [Oryza sativa ...	463	1e-128	
ref NP_001055584.1 	Os05g0421900 [Oryza sativa (japonica cult...	446	2e-123	UG
qb ABQ52488.1 	GA 20-oxidase [Paeonia suffruticosa]	441	7e-122	
qb ABF70102.1 	gibberellin 20-oxidase, putative [Musa balbisi...	440	1e-121	
qb ABF72027.1 	gibberellin 20-oxidase family protein [Musa ac...	436	1e-120	
qb AA998356.1 	GA 20-oxidase [Sesbania rostrata]	417	1e-114	
emb CAO61938.1 	unnamed protein product [Vitis vinifera]	415	3e-114	
dbj BAA37127.1 	gibberellin 20-oxidase [Lactuca sativa]	413	1e-113	
dbj BAG48318.1 	gibberellin 20-oxidase1 [Chrysanthemum x mori...	407	6e-112	
qb ABQ17965.1 	gibberellin 20-oxidase [Chrysanthemum x morifo...	406	2e-111	
dbj BAG12318.1 	gibberellin 20-oxidase [Chrysanthemum x morif...	405	2e-111	
dbj BAD30034.1 	gibberellin 20-oxidase2 [Daucus carota]	405	4e-111	
qb AAC49721.1 	GA 20-oxidase [Pisum sativum]	404	5e-111	
qb AAF29605.1 	AF138704.1 gibberellin c20-oxidase [Pisum sativum]	404	7e-111	
emb CAO43616.1 	gibberellin 20-oxidase [Helianthus annuus] >e...	403	1e-110	
emb CAO68523.1 	unnamed protein product [Vitis vinifera]	403	1e-110	
emb CAB96202.1 	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  
DFAPMGRVYQKYCEEMKELSLTIMELLELSLGVVERGYREFFADSSSIMRCNYYPPCPEP  
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。