

<https://www.agr.nagoya-u.ac.jp/~bioinfo/>

## 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。

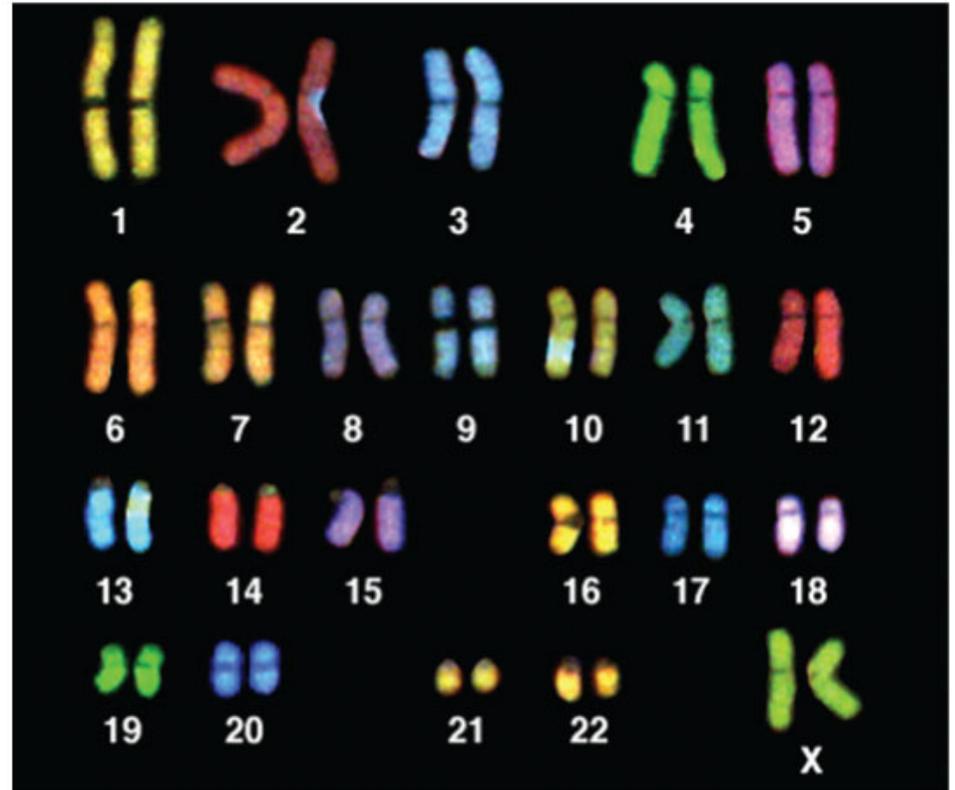
また、DNAの一部が転写され、スプライシングを受けた後、  
蛋白質に翻訳されることを学習してきた。

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。



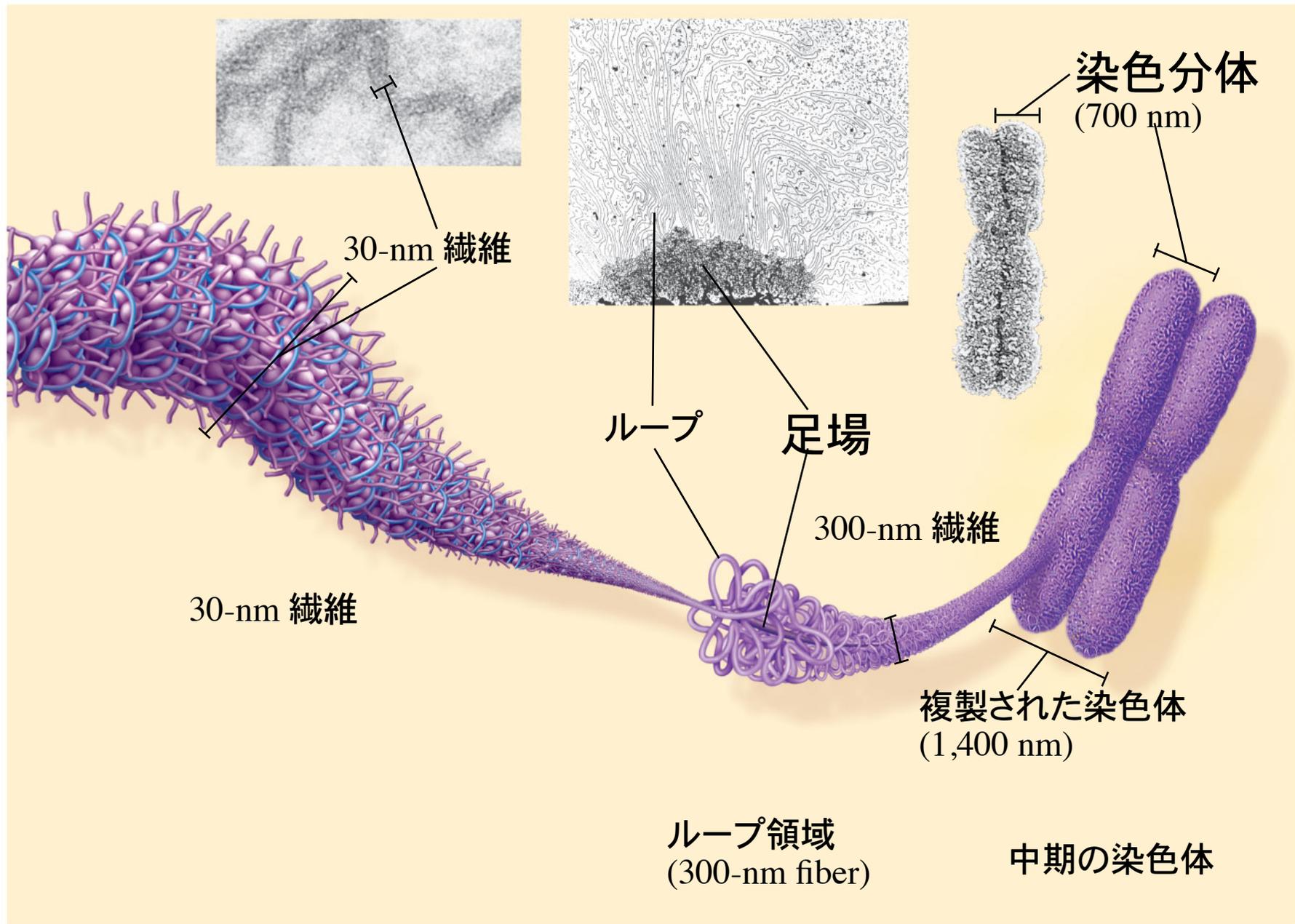
© 2011 Pearson Education, Inc.

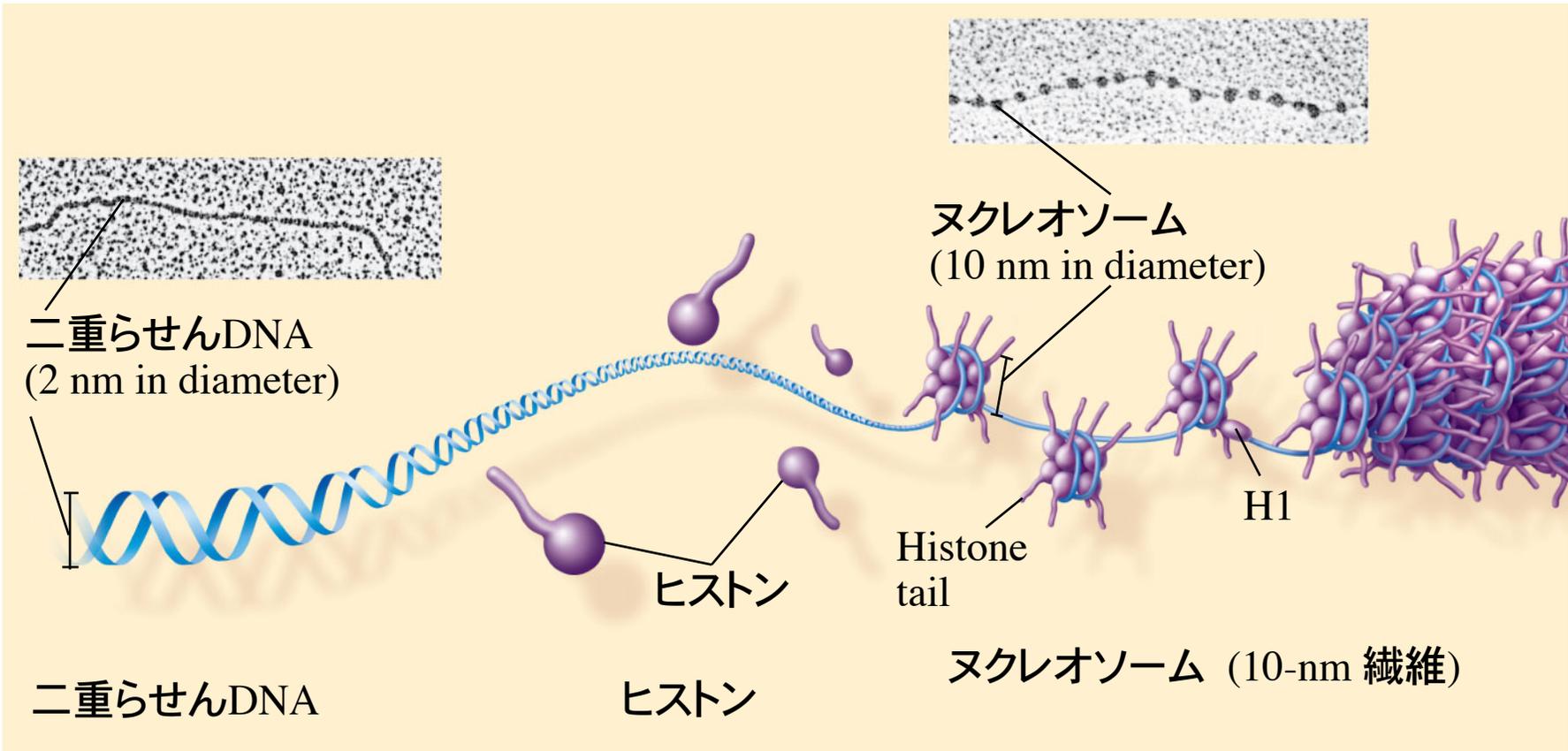


© 2011 Pearson Education, Inc.

## 人間の染色体(核型)

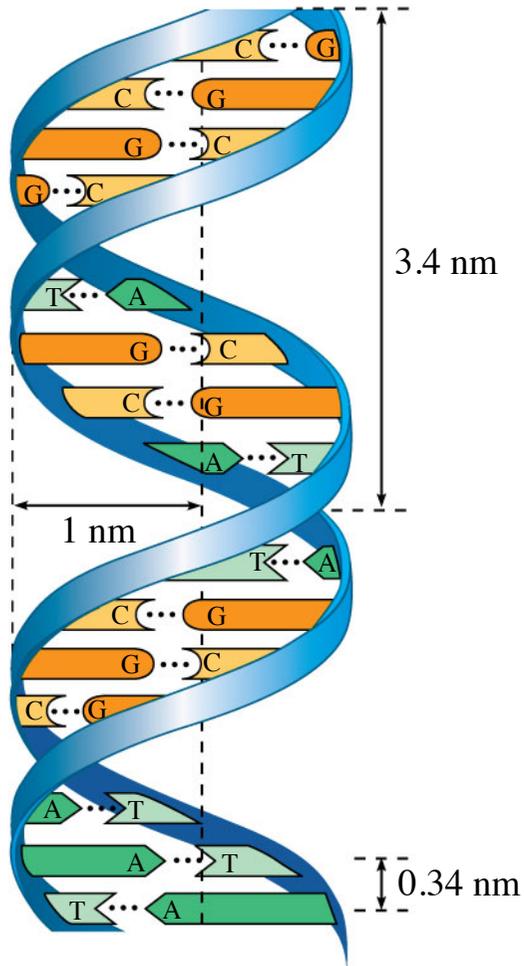
キャンベル参照





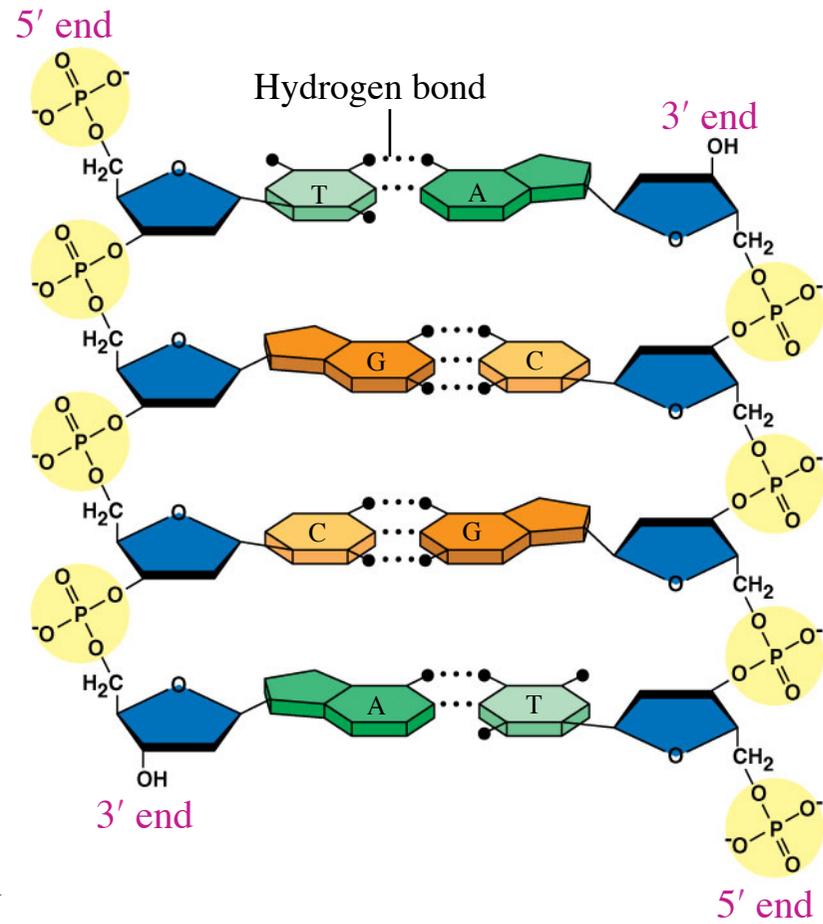
© 2011 Pearson Education, Inc.

Figure 16.7

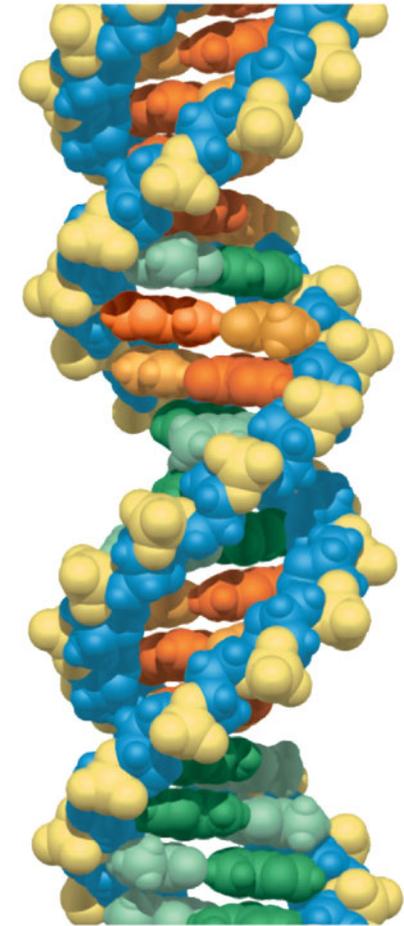


(a) Key features of DNA structure

© 2011 Pearson Education, Inc.

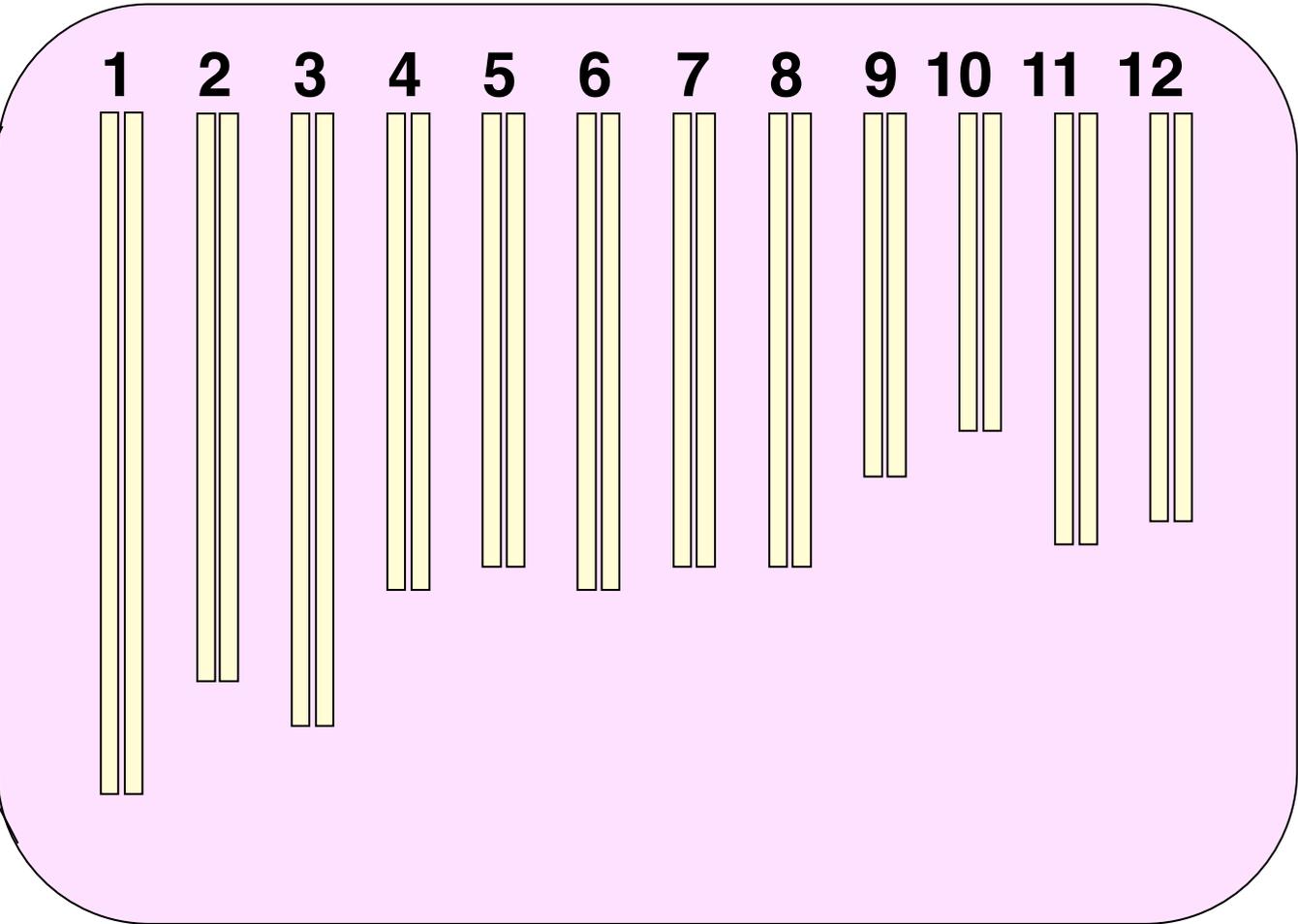
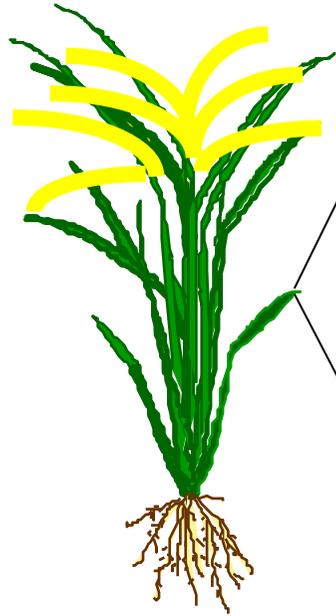


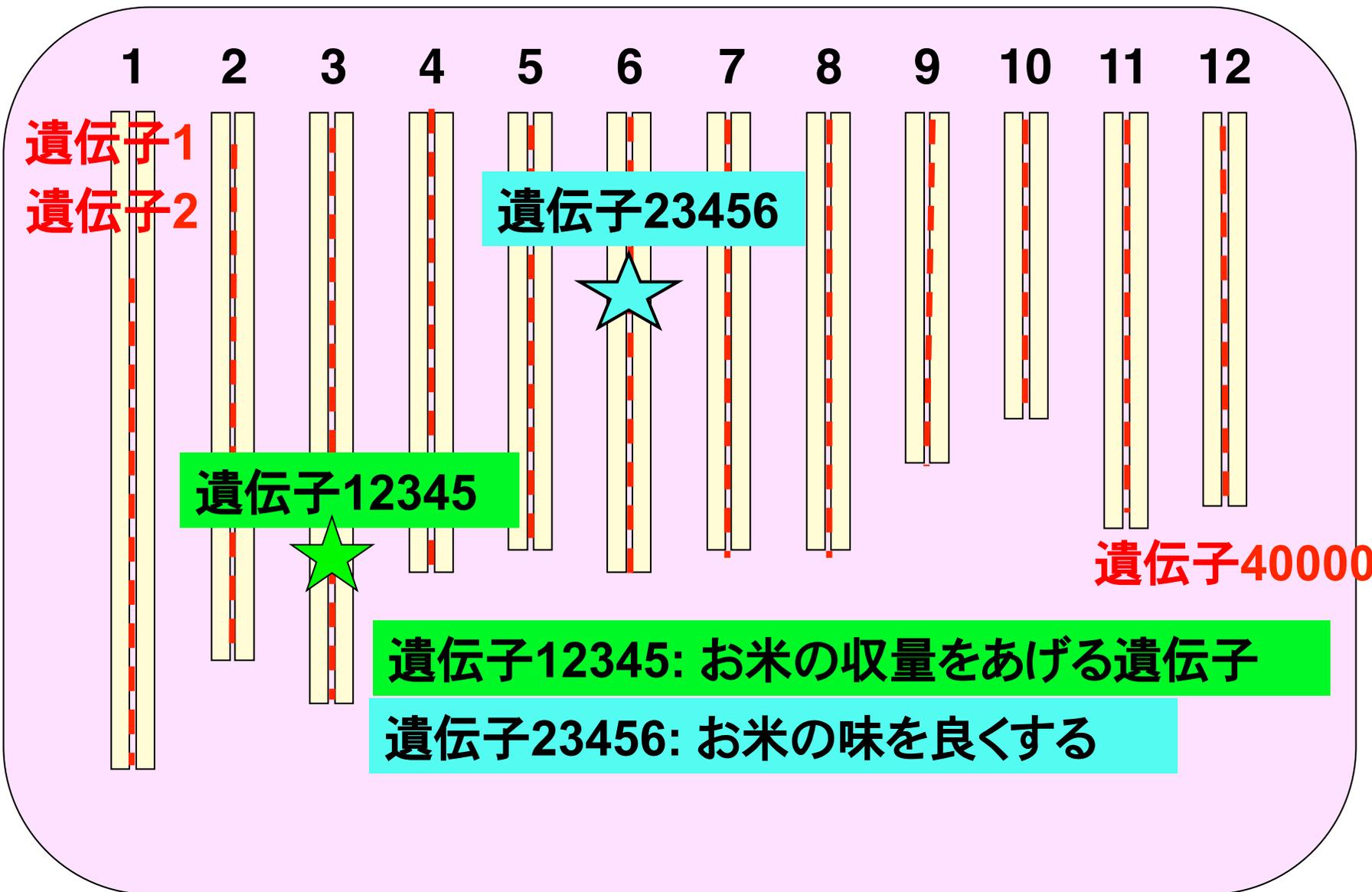
(b) Partial chemical structure



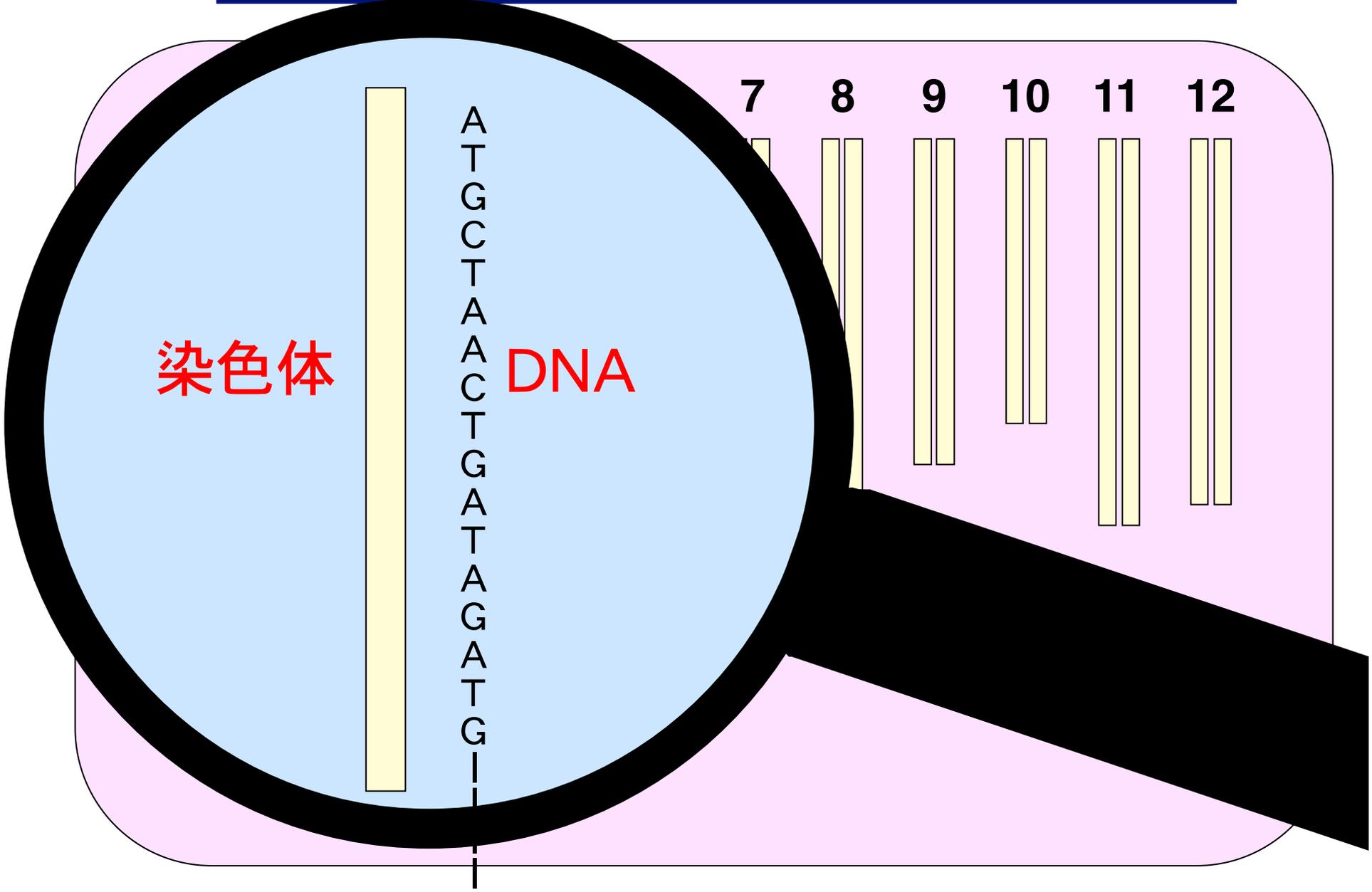
(c) Space-filling model

# 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

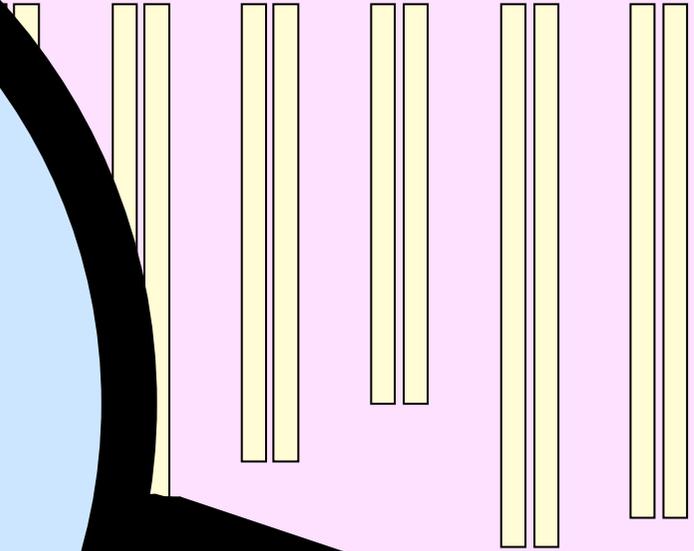


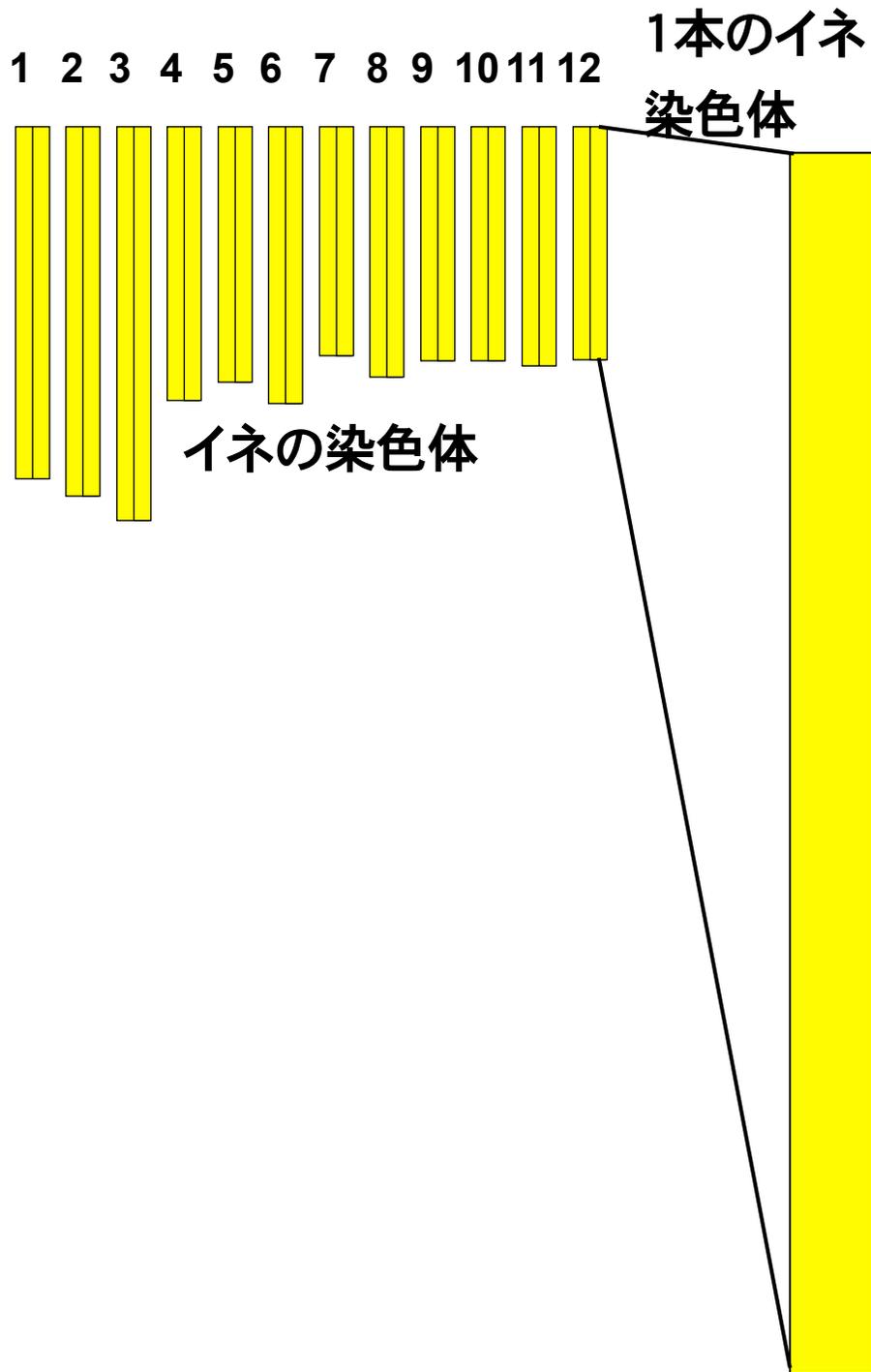
染色体

DNA

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

7 8 9 10 11 12





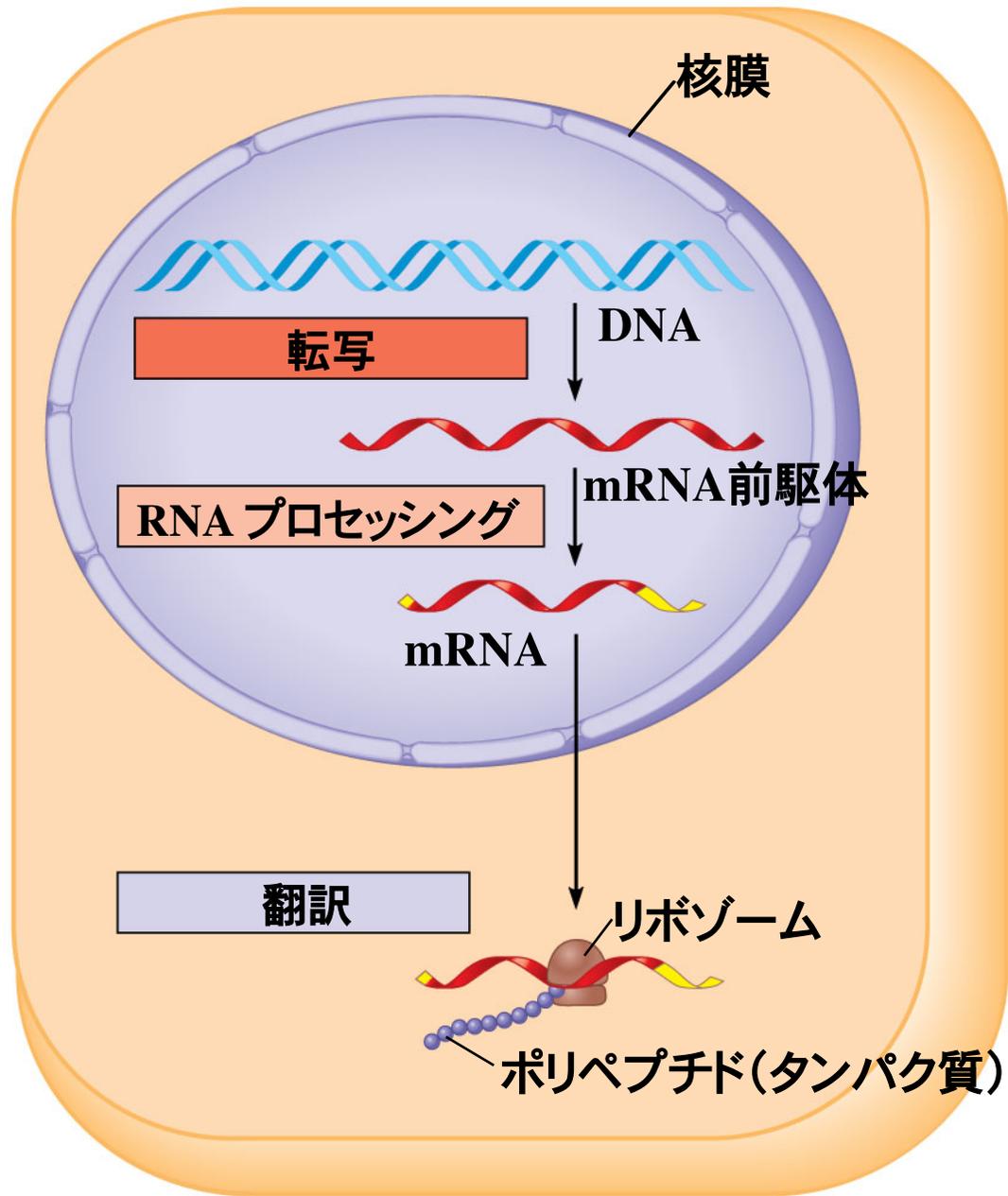
A  
T  
G  
T  
I  
I  
A  
A  
T  
A  
G  
G  
T  
A  
A  
A  
T  
G  
I  
I  
T  
A  
G  
G  
T



遺伝子1



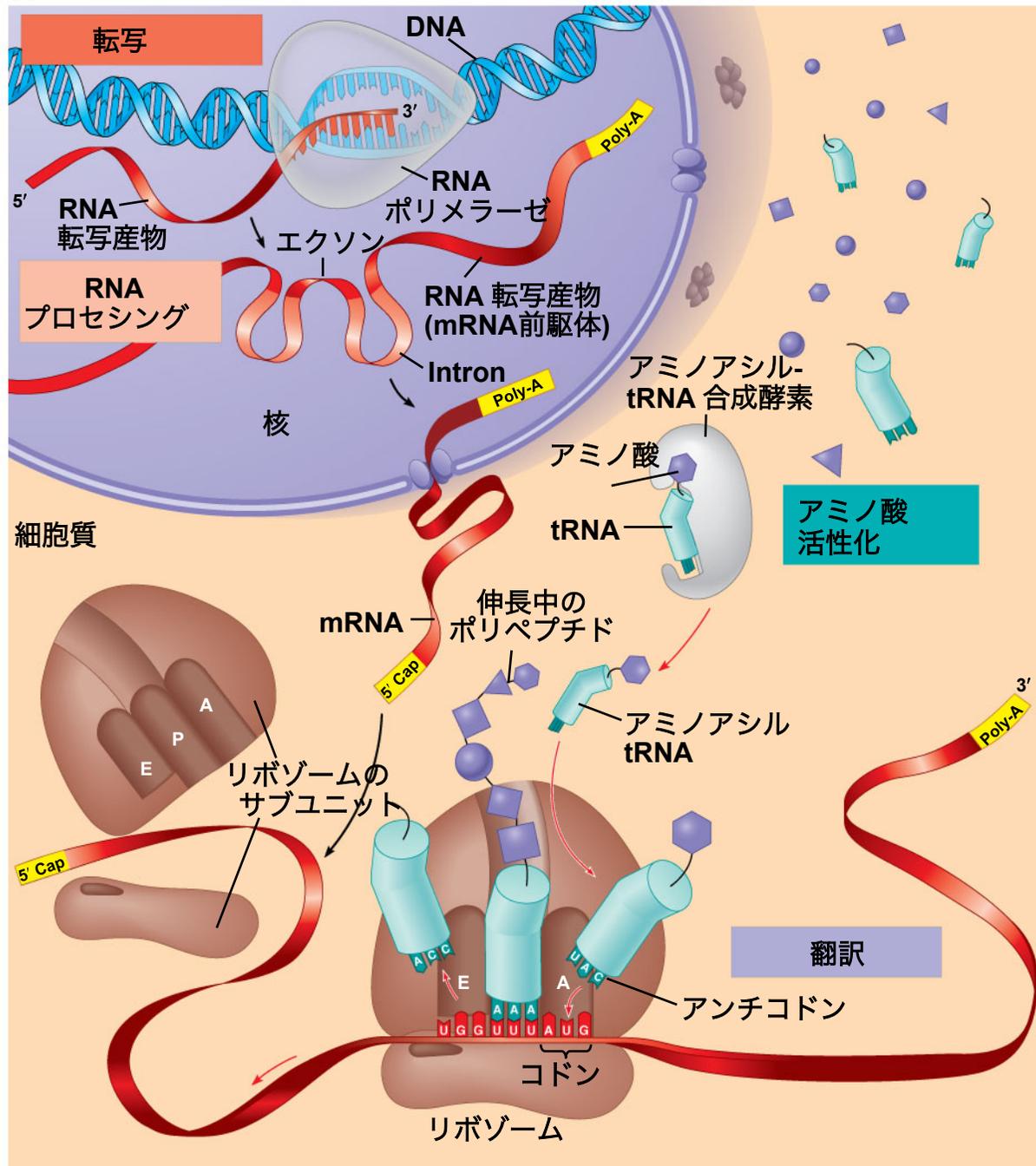
遺伝子2

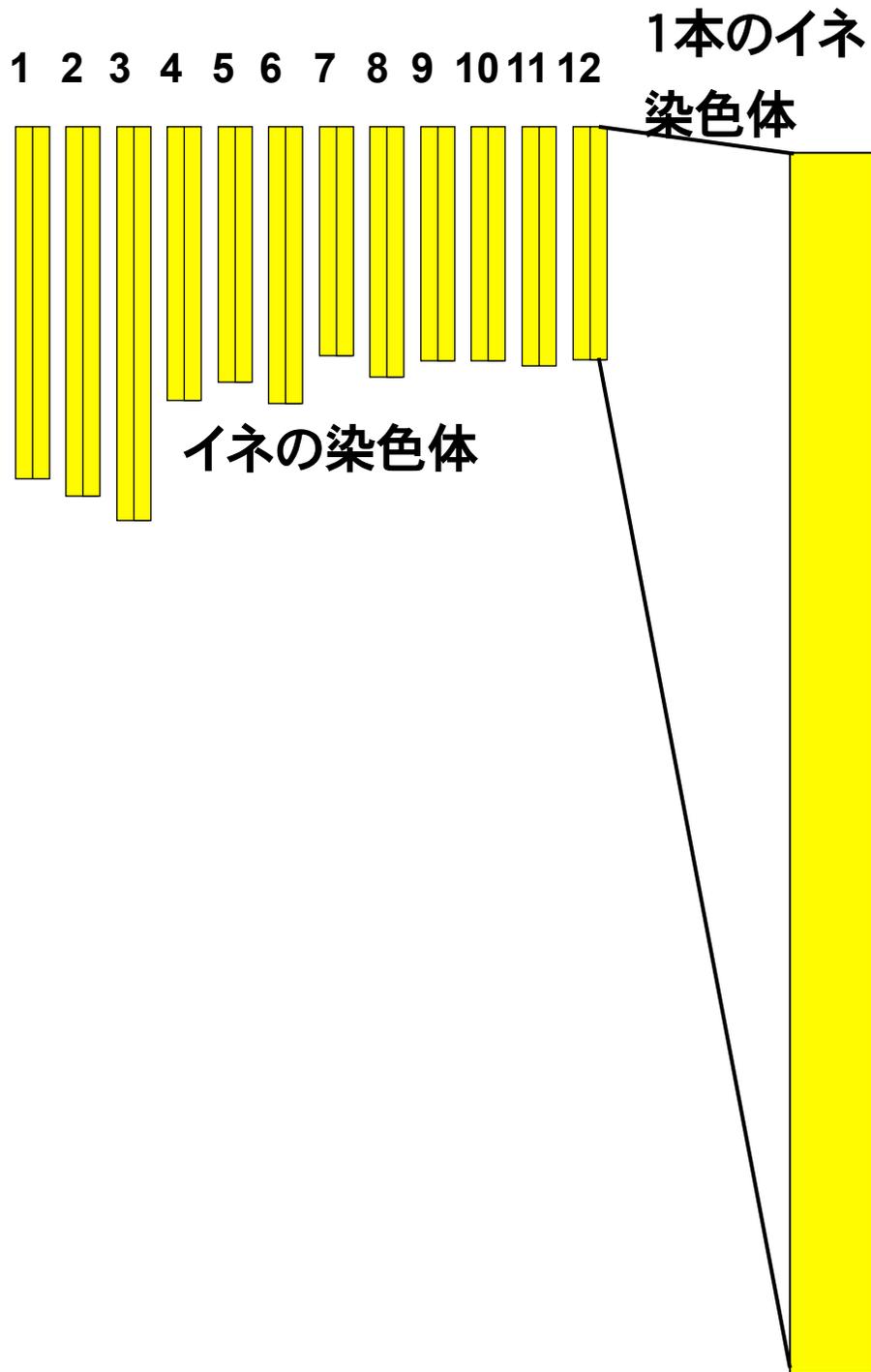


(b) 真核生物の細胞

© 2011 Pearson Education, Inc.

Figure 17





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 GCGCGCGGTG GCGACCTGAG GATGGAGCCC AGGCACTTCGT GTGGCCGACG GCGCAGCGCA GCCCGCGCTC
310 320 330 340 350 360 370 380 390 400
GGCGCGCGAG CTGGACATGC CCGTGGTCCG CTGGGCGGTG CTCGCCGACG GCGACGCCGA GGGGCTGCG CCGCCCGCGG CCGAGGTGGC CCGCCGCTGC
410 420 430 440 450 460 470 480 490 500
GCCACGCACG GGTCTTCCCA GGTGTCCGAG CACGGCGTCC AGCCCGCTCT GCGCGCGCGC GCGCTCGACG GCGCCAGCGA CTCTTCCCGC CTCGCCGCTG
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 GCGCCGCGCG CCCCGTCTCT GCCGACTACT TCTCCAGCAC CCTCGGCCCG GACTTCCGCG 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 AAGGAGCTGT CGCTGACGAT CATGGACTCT CTGGAGCTGA GCCTGGGGGT GAGCGGAGGG TACTACAGGG AGTTCTTCGG
910 920 930 940 950 960 970 980 990 1000
GGACAGCAGC TCARATCTGC GGTGCACTCA CTACCCGCCA TGCCCGGAGC CCGAGCGGAC GCTCGGACCG GCGCCGACAT GCGACCCCGC CCGCTCACCC
1010 1020 1030 1040 1050 1060 1070 1080 1090 1100
ATCCTCTCCG AGGACGACGT CCGCGCGCTC GAGGTCTCTG TCAGCGGCGA ATGGCGCCCG CTCAGCCCGG TCCCGCGCGG CATGGTATC ACATCGGGC
1110 1120 1130 1140 1150 1160 1170 1180 1190 1200
ACACCTTCAT GGTAAACCAT CTCTATTCTT CCTCTCTCTT GTTCTCTCTT GCTTCGAGAC AACAGACACA GTATTTCAAG CTTTTTTTTT TCTCTCGCGC
1210 1220 1230 1240 1250 1260 1270 1280 1290 1300
GAATTTGAGC AGAAATATTA GATCGTGTA GGGCGCGGGC TTTCAGCTGA AAGCGGGAGG AATCCGACTT GACGTGATTT CTCTGTTCCA ATCAGACACA
1310 1320 1330 1340 1350 1360 1370 1380 1390 1400
ATGGATGCCC CCACTCTCCG ATGTGTATAG ATTTATCTCA CATCTTATAG TTAATAGAGG TAAGTAAACA GACTATTTTT TCATATATTA GTTCGTTTGA
1410 1420 1430 1440 1450 1460 1470 1480 1490 1500
TTTTTTTTTT TTAAGTTTTT TTTAGTTTTA TCCAAATTTA TTGAAARACT TAGCAGCGTT TATATATCCA AATTAGTCTC ATTTAGTTTA ATATGTATTA
1510 1520 1530 1540 1550 1560 1570 1580 1590 1600
TATTTTGTAT ATATATTTAT GTTATATTTA AATATTTACT ATATTTTTCT ATAAACATTA TTAAGAGCCA TTTATATATAT AATATGGAGG GAGTAAATTA
1610 1620 1630 1640 1650 1660 1670 1680 1690 1700
TATGGATCTC CCCCGACATG AGATATTTT CCGATGGTGT GACGACGCCA TGTAGCTTTC GGTGGGCGTG GACGGCCAGG GGTGCCACCA GCCACGTCCA
1710 1720 1730 1740 1750 1760 1770 1780 1790 1800
ACACCCCGCT GGTCCCGCCC TACACCTCCA AACAGTAGTG AGTAGTGTCT CGTCGCGTTT TAGTATTTGA TGACAAACAA AGTGTGATTT GAGTTAGCCA
1810 1820 1830 1840 1850 1860 1870 1880 1890 1900
CCACCAACTT GCACACGAGC ACATCACTTT GTGTCCATTC TCGCCAGTCA CTTCCTACTC TAGTCCATAC TCCTATCTAG CGATGTAGGC GGTAAATTTT
1910 1920 1930 1940 1950 1960 1970 1980 1990 2000
ATCATCGGTA TATAACCTG TTTGTTATAG TTAATTTCTT ATATAACTAT ATACAGTAT ACATTTTAAA AGAAACAAA ATTAGGATTA ACAAGCCCTG
2010 2020 2030 2040 2050 2060 2070 2080 2090 2100
CTCCTATCCA TCCATGGCAC TTGGAGAGAC CAGACTCGGT CATGCCATGC CAGGCCAAGA TATGGGTTAT GAGAGGTAG AGAAGAGAGG AGATGAGAGA
2110 2120 2130 2140 2150 2160 2170 2180 2190 2200
TAGACATGGC TTCTCTCTCT GTTGGATGT GTATTTTGGG GGGATTTGTG TAGTAGTAGC AGCGGCGCGG CCGGGACGGA TCGGATGGT GCGCTTTCCG
2210 2220 2230 2240 2250 2260 2270 2280 2290 2300
GTGGCGTTTT CCGGGGGGGG TTTTGGTTTT GCGCTTGGGG GGGATGGCAT GCGCGGCGGT GCGGCTGAC GCCACACACA CCGCGCGCGA CCGACGTACG
2310 2320 2330 2340 2350 2360 2370 2380 2390 2400
TCGTGCTGCG CCGGGCGCGA CCGTAGCTTA GGGTGGTGTG TTCCGCGCGC GCGCGCGGAT TGTTCATGC CGATCGATTT GCGCCACACC TCGCGCGCGC
2410 2420 2430 2440 2450 2460 2470 2480 2490 2500
TCTTGTGCGG TCGTGCGCCT CTCTCGCGCG GTTTGTCTCT GTCGGTTGCG TCAGCGCGCG ACGGGGGACG GACATTTGCG GATGTAGCCC TGACGTGTGC
2510 2520 2530 2540 2550 2560 2570 2580 2590 2600
GGCCTCTCCG TTGATGATG ATGATGTATG TATGTATTTT TTTTGTCTCG AAGGAAATTT TGGGAAATTT TTGTGTGTGC AGGCGCTGTC GACCGGGAGG
2610 2620 2630 2640 2650 2660 2670 2680 2690 2700
TATAGAGACT GCTCGCACG GCGGTGGTG AACACCGCG GAGAGCGGCG GTCCGTCGCG TTCTTCTGTG GCCCGCGGGA 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 GCGGCGCAGC TGATCCCGCA 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	GCCCGAGCG	CTCGCCCTGC	ACACACACAC	ACACTCACAC	TCACACACG	TCTCACTCA	CTCCCCTCA
110	120	130	140	150	160	170	180	190	200
ACACAGCGCT	CACCTTCTAT	CTCCAACTTC	ATGTTGGCCG	AGCACCCTAC	GCCACCCAC	CCGACCCAC	CACCCGCAT	GGACTCCAC	GCCGGCTCG
210	220	230	240	250	260	270	280	290	300
GCATTGCGCG	CCCGGCGCG	GCGCGGTGT	GCAGCTGAG	GATGGAGCCC	AGATCCCGG	AGCCATTCT	GTGGCCGAC	GCGACCCGA	GCCCGCGTC
310	320	330	340	350	360	370	380	390	400
GCGCGCGAG	CTGGACATG	CCGTGGTCG	CGTGGCGTG	CTCCGCGAC	GCGACCCGA	GGGGCTGCC	CGCCCGCGG	CGCAGGTGG	CGCCCGCTG
410	420	430	440	450	460	470	480	490	500
GCCACGACG	GGTCTTCCA	GGTGTCCG	CACGGCTGC	ACGCCCTCT	GCGCGCGCC	GCGCTGACG	GCGCCAGCG	CTTCTCCCG	CTCCCCTCG
510	520	530	540	550	560	570	580	590	600
CCGAGAGCG	CCGCGCGCG	CGCGTCCCG	GCACCCTGC	CGGTACACC	AGCGCCACG	CCGACCCGT	CGCCTCCAG	CTCCCATGG	AGGAGACCT
610	620	630	640	650	660	670	680	690	700
CTCCTTCCG	TTCCAGACC	GCGCGCGCG	CCCGCTGTC	GCGACTACT	TCTCCAGAC	CCTCGCCCG	GACTTCGCG	CAATGGGTA	ATTAAACGA
710	720	730	740	750	760	770	780	790	800
TGTTGGACG	CATTGCATT	CAAAATCRA	ACAATTCRA	AACACCCGA	CCGAGATTG	GCTGAATTC	ACCCGTTTG	TGGCGCAGG	AGGTGTACC
810	820	830	840	850	860	870	880	890	900
AGAACTACT	CGAGAGATG	AGAGGCTGT	CGCTGACAT	CATGGACTC	CTGGAGCTG	GCCTGGGCT	GGAGCAGCG	TACTACAGG	AGTTCTCCG
910	920	930	940	950	960	970	980	990	1000
GGACAGACG	TCAATCATG	GGTGACATC	CTACCCGCA	TGCCCGGAC	GCGAGCGAC	GCTCGGACG	GGCCCGCAC	GCGACCCAC	CGCCCTCAC
1010	1020	1030	1040	1050	1060	1070	1080	1090	1100
ATCCTCTCC	AGGACGACT	CGCGGCGCT	GAGGTCTCG	TCGACGCCG	ATGGCGCCC	GTCAGCCCG	CATGGTCTC	ACACTCGCG	
1110	1120	1130	1140	1150	1160	1170	1180	1190	1200
ACACCTTCAT	GTAARACAT	CTCCTATTCT	CCTCTCCTCT	GTTCCTCCT	GCTTCGAGC	AACAGACAA	GTAATTCAG	CTTTTTTTC	TCTCTCGCG
1210	1220	1230	1240	1250	1260	1270	1280	1290	1300
GAATTTGCG	AGAAAATTA	GATCGTGTA	GCGCGGGCG	TTTCAGCTG	AGCGGGGAG	AAACCGACT	CGCGATTTC	CTCTGTTCA	ATCACAAAC
1310	1320	1330	1340	1350	1360	1370	1380	1390	1400
ATGGAAATG	CCACTCTCC	ATGTTGATG	ATTATCTCA	CATCTTATG	TTAATAGAG	TAAATACAA	GCTACTTTT	TCATATTAT	GTTCTGTTG
1410	1420	1430	1440	1450	1460	1470	1480	1490	1500
TTTTTTTTT	TTAAAGTTT	TTTAGTTTA	TCCAAATTA	TTGAAARCT	TAGCAACGT	TATATACCA	AATTAGTTC	ATTTAGTTA	ATATTGTATA
1510	1520	1530	1540	1550	1560	1570	1580	1590	1600
TATTTTGTA	ATATATTAT	GTTATATT	AAATATTCT	ATATTTTCT	ATAAACATTA	TTAAAGCCA	TTTAAATTA	AAATGGAG	GAGTAAATTA
1610	1620	1630	1640	1650	1660	1670	1680	1690	1700
TATGGATCT	CCCCGACATG	AGAAATTTT	CCGATGGTG	GACGACCCA	TGTAAGCTC	GTTGGGCGT	GACGGCCGA	GGTCCACCA	GCCACGTTCA
1710	1720	1730	1740	1750	1760	1770	1780	1790	1800
ACACCCCTG	GGTCCCGCC	TAAACATCA	ACAGTAGTG	AGTAGTGTCT	CGTCGCGTT	TAGTATTGA	TGACAAACA	AGTGTAGTT	GAGTTAGCCA
1810	1820	1830	1840	1850	1860	1870	1880	1890	1900
CCACCAACT	GACACGAGC	ACATACATT	GTGTCATTCT	TCGCCAGTC	CTTCCATCT	TAGCTTAC	TCCTATCTG	CGATGAGC	GGATATTTT
1910	1920	1930	1940	1950	1960	1970	1980	1990	2000
ATCATCCGA	TATAACCTG	TTTGTATAG	TTAATTTCT	ATATAACT	ATACAGTAT	ACATTTTAA	AGAAAACAA	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
TAGCATGCG	TTCTCTCCT	CGTTGGATG	GATTTTGGG	GGGATTGTT	TAGTAGTAG	AGCGGCGCG	CGGGGACGA	TCCGATGGT	GCCGTTTTC
2210	2220	2230	2240	2250	2260	2270	2280	2290	2300
GTGGCGTTT	CCCGGGGGG	TTTTGGTTG	GCGCTTGGG	GGGATGGCAT	GCGCGCGCT	GCGCTGAC	GCGCGCGCG	GCGACGACG	CAGCATGACG
2310	2320	2330	2340	2350	2360	2370	2380	2390	2400
TCGTGCTCG	CAGCGGCGG	CGTAGCTTA	GGTGGTGGT	TTCCGCGCG	GCGCGCGGAT	TGTTCCATG	GATCGATTT	GCGCGCCCG	TCGCCGCGG
2410	2420	2430	2440	2450	2460	2470	2480	2490	2500
TCTGTGCGG	TCGTGCGCT	CTCTCGCGG	GTTTGTCTT	GTCGGTTGC	TCAGCCGCG	ACGGGGGAC	GACATTGCG	GATGAGCCC	TGACGTTCT
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	GCCTGACAG	GCGCGGGTG	AACAGCGCG	GGAGCGGCG	GTCGCTGCG	TTCTTCTGT	GCCCGCGGA	GGACAGGTT	GTCGCGCGG
2710	2720	2730	2740	2750	2760	2770	2780	2790	2800
GCCCGAGCG	GCCCGAGCG	CAGCACTAC	CGGACTTAC	CTAGGCGAC	CTCAGCGCT	TCACCGCG	CCACTACCG	GCCGACCG	GACCGCTCG
2810	2820	2830	2840	2850	2860	2870	2880	2890	2900
GCCCTTCCG	CGCTGGCTG	GCCCGCGCG	GCCGACCGC	GCCGCGCGG	CAGAGGTCA	GCGCGCGCG	TGATCCCGA	ACGGACGAA	ACGGACGAA

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

Untitled 3

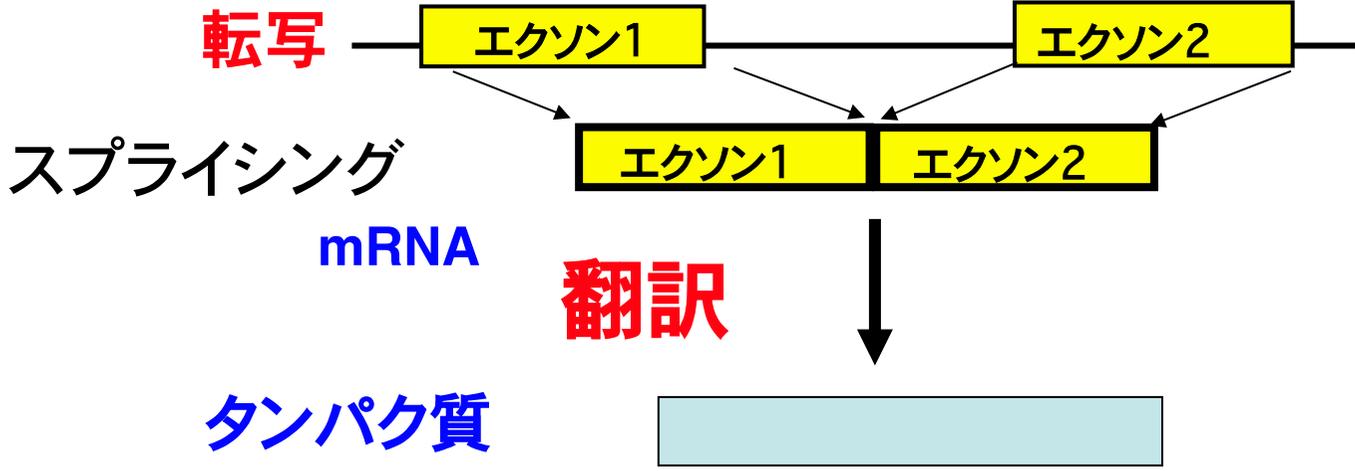
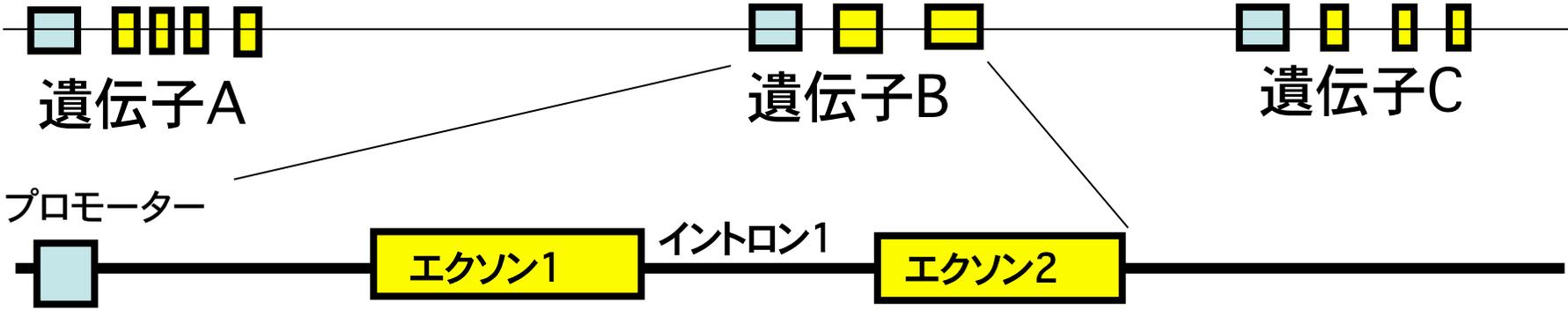
10	20	30	40	50	60	70	80	90	100
TCTCCCTGT	TACAAATAC	CCACCCCTCT	GCCCGAGCG	CTCGCCCTGC	ACACACACAC	ACACTCACAC	TCACACACG	TCTCACTCA	CTCCCCTCA
110	120	130	140	150	160	170	180	190	200
ACACAGCGCT	CACCTTCTAT	CTCCAACTTC	ATGTTGGCCG	AGCACCCTAC	GCCACCCAC	CCGACCCAC	CACCCGCAT	GGACTCCAC	GCCGGCTCG
210	220	230	240	250	260	270	280	290	300
GCATTGCGCG	CCCGGCGCG	GCGCGGTGT	GCAGCTGAG	GATGGAGCCC	AGATCCCGG	AGCCATTCT	GTGGCCGAC	GCGACCCGA	GCCCGCGTC
310	320	330	340	350	360	370	380	390	400
GCGCGCGAG	CTGGACATG	CCGTGGTCG	CGTGGCGTG	CTCCGCGAC	GCGACCCGA	GGGGCTGCC	CGCCCGCGG	CGCAGGTGG	CGCCCGCTG
410	420	430	440	450	460	470	480	490	500
GCCACGACG	GGTCTTCCA	GGTGTCCG	CACGGCTGC	ACGCCCTCT	GCGCGCGCC	GCGCTGACG	GCGCCAGCG	CTTCTCCCG	CTCCCCTCG
510	520	530	540	550	560	570	580	590	600
CCGAGAGCG	CCGCGCGCG	CGCGTCCCG	GCACCCTGC	CGGTACACC	AGCGCCACG	CCGACCCGT	CGCCTCCAG	CTCCCATGG	AGGAGACCT
610	620	630	640	650	660	670	680	690	700
CTCCTTCCG	TTCCAGACC	GCGCGCGCG	CCCGCTGTC	GCGACTACT	TCTCCAGAC	CCTCGCCCG	GACTTCGCG	CAATGGGTA	ATTAAACGA
710	720	730	740	750	760	770	780	790	800
TGTTGGACG	CATTGCATT	CAAAATCRA	ACAATTCRA	AACACCCGA	CCGAGATTG	GCTGAATTC	ACCCGTTTG	TGGCGCAGG	AGGTGTACC
810	820	830	840	850	860	870	880	890	900
AGAACTACT	CGAGAGATG	AGAGGCTGT	CGCTGACAT	CATGGACTC	CTGGAGCTG	GCCTGGGCT	GGAGCAGCG	TACTACAGG	AGTTCTCCG
910	920	930	940	950	960	970	980	990	1000
GGACAGACG	TCAATCATG	GGTGACATC	CTACCCGCA	TGCCCGGAC	GCGAGCGAC	GCTCGGACG	GGCCCGCAC	GCGACCCAC	CGCCCTCAC
1010	1020	1030	1040	1050	1060	1070	1080	1090	1100
ATCCTCTCC	AGGACGACT	CGCGGCGCT	GAGGTCTCG	TCGACGCCG	ATGGCGCCC	GTCAGCCCG	CATGGTCTC	ACACTCGCG	
1110	1120	1130	1140	1150	1160	1170	1180	1190	1200
ACACCTTCAT	GTAARACAT	CTCCTATTCT	CCTCTCCTCT	GTTCCTCCT	GCTTCGAGC	AACAGACAA	GTAATTCAG	CTTTTTTTC	TCTCTCGCG
1210	1220	1230	1240	1250	1260	1270	1280	1290	1300
GAATTTGCG	AGAAAATTA	GATCGTGTA	GCGCGGGCG	TTTCAGCTG	AGCGGGGAG	AAACCGACT	CGCGATTTC	CTCTGTTCA	ATCACAAAC
1310	1320	1330	1340	1350	1360	1370	1380	1390	1400
ATGGAAATG	CCACTCTCC	ATGTTGATG	ATTATCTCA	CATCTTATG	TTAATAGAG	TAAATACAA	GCTACTTTT	TCATATTAT	GTTCTGTTG
1410	1420	1430	1440	1450	1460	1470	1480	1490	1500
TTTTTTTTT	TTAAAGTTT	TTTAGTTTA	TCCAAATTA	TTGAAARCT	TAGCAACGT	TATATACCA	AATTAGTTC	ATTTAGTTA	ATATTGTATA
1510	1520	1530	1540	1550	1560	1570	1580	1590	1600
TATTTTGTA	ATATATTAT	GTTATATT	AAATATTCT	ATATTTTCT	ATAAACATTA	TTAAAGCCA	TTTAAATTA	AAATGGAG	GAGTAAATTA
1610	1620	1630	1640	1650	1660	1670	1680	1690	1700
TATGGATCT	CCCCGACATG	AGAAATTTT	CCGATGGTG	GACGACCCA	TGTAAGCTC	GTTGGGCGT	GACGGCCGA	GGTCCACCA	GCCACGTTCA
1710	1720	1730	1740	1750	1760	1770	1780	1790	1800
ACACCCCTG	GGTCCCGCC	TAAACATCA	ACAGTAGTG	AGTAGTGTCT	CGTCGCGTT	TAGTATTGA	TGACAAACA	AGTGTAGTT	GAGTTAGCCA
1810	1820	1830	1840	1850	1860	1870	1880	1890	1900
CCACCAACT	GACACGAGC	ACATACATT	GTGTCATTCT	TCGCCAGTC	CTTCCATCT	TAGCTTAC	TCCTATCTG	CGATGAGC	GGATATTTT
1910	1920	1930	1940	1950	1960	1970	1980	1990	2000
ATCATCCGA	TATAACCTG	TTTGTATAG	TTAATTTCT	ATATAACT	ATACAGTAT	ACATTTTAA	AGAAAACAA	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
TAGCATGCG	TTCTCTCCT	CGTTGGATG	GATTTTGGG	GGGATTGTT	TAGTAGTAG	AGCGGCGCG	CGGGGACGA	TCCGATGGT	GCCGTTTTC
2210	2220	2230	2240	2250	2260	2270	2280	2290	2300
GTGGCGTTT	CCCGGGGGG	TTTTGGTTG	GCGCTTGGG	GGGATGGCAT	GCGCGCGCT	GCGCTGAC	GCGCGCGCG	GCGACGACG	CAGCATGACG
2310	2320	2330	2340	2350	2360	2370	2380	2390	2400
TCGTGCTCG	CAGCGGCGG	CGTAGCTTA	GGTGGTGGT	TTCCGCGCG	GCGCGCGGAT	TGTTCCATG	GATCGATTT	GCGCGCCCG	TCGCCGCGG
2410	2420	2430	2440	2450	2460	2470	2480	2490	2500
TCTGTGCGG	TCGTGCGCT	CTCTCGCGG	GTTTGTCTT	GTCGGTTGC	TCAGCCGCG	ACGGGGGAC	GACATTGCG	GATGAGCCC	TGACGTTCT
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	GCCTGACAG	GCGCGGGTG	AACAGCGCG	GGAGCGGCG	GTCGCTGCG	TTCTTCTGT	GCCCGCGGA	GGACAGGTT	GTCGCGCGG
2710	2720	2730	2740	2750	2760	2770	2780	2790	2800
GCCCGAGCG	GCCCGAGCG	CAGCACTAC	CGGACTTAC	CTAGGCGAC	CTCAGCGCT	TCACCGCG	CCACTACCG	GCCGACCG	GACCGCTCG
2810	2820	2830	2840	2850	2860	2870	2880	2890	2900
GCCCTTCCG	CGCTGGCTG	GCCCGCGCG	GCCGACCGC	GCCGCGCGG	CAGAGGTCA	GCGCGCGCG	TGATCCCGA	ACGGACGAA	ACGGACGAA

(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

-----TGGCATATGGCATGCAGACTGGCATTGGCATGCGTCTGCATGCCATTGGTTTCTGAGGTCGTGTGAAACTTATTTGGGCACAGACATGGCATGCAGACTGGCAC-----  
-----ACCGTATACCGTACGTCTGACCGTAACCGTACGCAGACGTACGGTAACCAAAGAGTCCAGCACACTTGAATAAACCCGTGTCTGTACCGTACGTCTGACCGTG-----



# エキソンとイントロンの境界

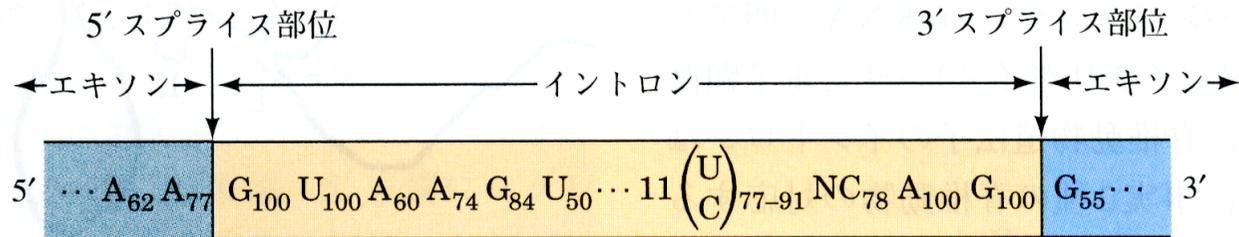
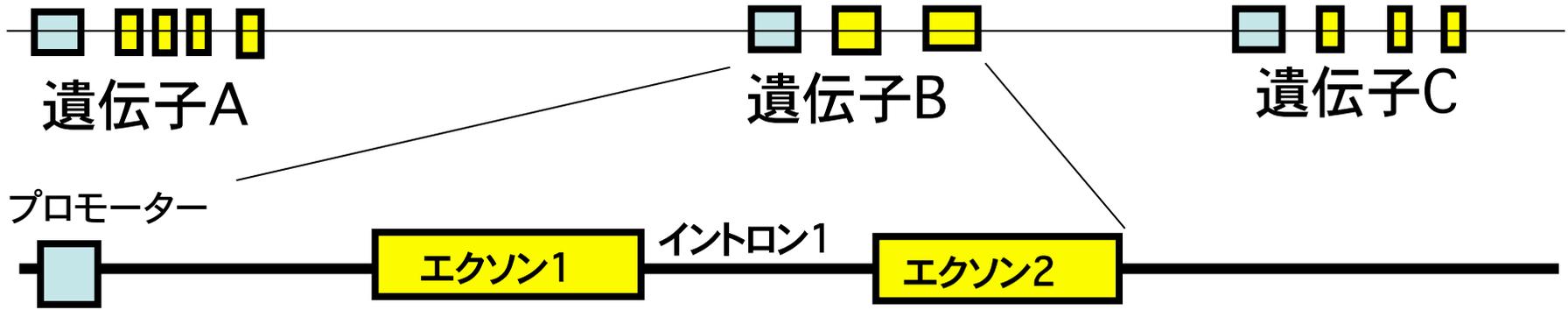


図 25・19 真核細胞 mRNA 前駆体のエキソン・イントロン接合部にあるコンセンサス配列 各塩基の添え数字はその塩基の出現頻度(%). 3' スプライス部位のすぐ上流にピリミジンに富む 11 nt 区域がある [R.A. Padgett, P.J. Grabowski, M.M. Konarska, S.S. Seiler, P.A. Sharp, *Annu. Rev. Biochem.* **55**, 1123 (1986) のデータに基づく]

# DNAからRNAへそしてタンパク質へ



コード鎖

DNA 5' --GTCTGC**ATGCCATTGGTTTCTGAGGTCGTGTGA**AACA**AATAAAGGGCA**  
 DNA 3' --CAGACG**TACGGTAACCAAAGACTCCAGCACACT**TTGTTATTTCCCGT

鋳型鎖

PolyA認識サイト

RNAポリメラーゼ

mRNA前駆体

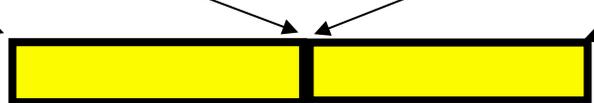
**CAP** GTCUGC**AUGCCA**UUGGUUUCUGAG**GUCGUGUGA**AACAAUAAAGGGCAAAAAA

転写



スプライシング

成熟mRNA



成熟mRNA **CAP**GTCUGC|**AUG**|**CCA**|**UUG**|**GUC**|**GUGUGA**|AACUUAUUUGGGCAAAAAAA

Met Pro Leu Val Val Stop

# DNAとcDNA (complementary DNA)



コード鎖

DNA 5' --GTCCTGC**ATGCCATTG**GTTTCTGAG**GTCGTGTGA**AACAATAAGGGCA  
 DNA 3' --CAGACGT**TACGGTAAC**CAAAGACTC**CAGCACACT**TTGTTATTTCCCGT

鋳型鎖

PolyA認識サイト

RNAポリメラーゼ

mRNA前駆体

**CAP** GTCUGC**AUGCCA**UUGG**UUUCUGAGGUCGUGUGA**AACAAUAAAGGGCAAAAAA

転写



スプライシング

成熟mRNA

成熟mRNA **CAP**GTCUGC**AUGCCA**UUG**GUCGUGUGA**AACUUAUUUGGGCAAAAAAAA

Met Pro Leu Val Val Stop

逆転写酵素

cDNA

GTCTGC**ATGCCATTGGTCGTGTGA**AACAATAAGGGCAAAAAAAA  
 CAGACGT**TACGGTAACCAGCACACT**TTGTTATTTCCCGTTTTTTTTT

## DNA配列のどこが遺伝子なのか？ORFを検出してみる

### (1) CDS or cDNA (EST) 情報が有るとき

EST (expressed sequence tag)とは、ある組織において発現しているタンパク質のmRNA 配列に由来するcDNA配列を部分的に決定したもののことを指す

DNA配列とCDS or cDNA配列を比較する

### (2) CDS or cDNA (EST) 情報が無いとき

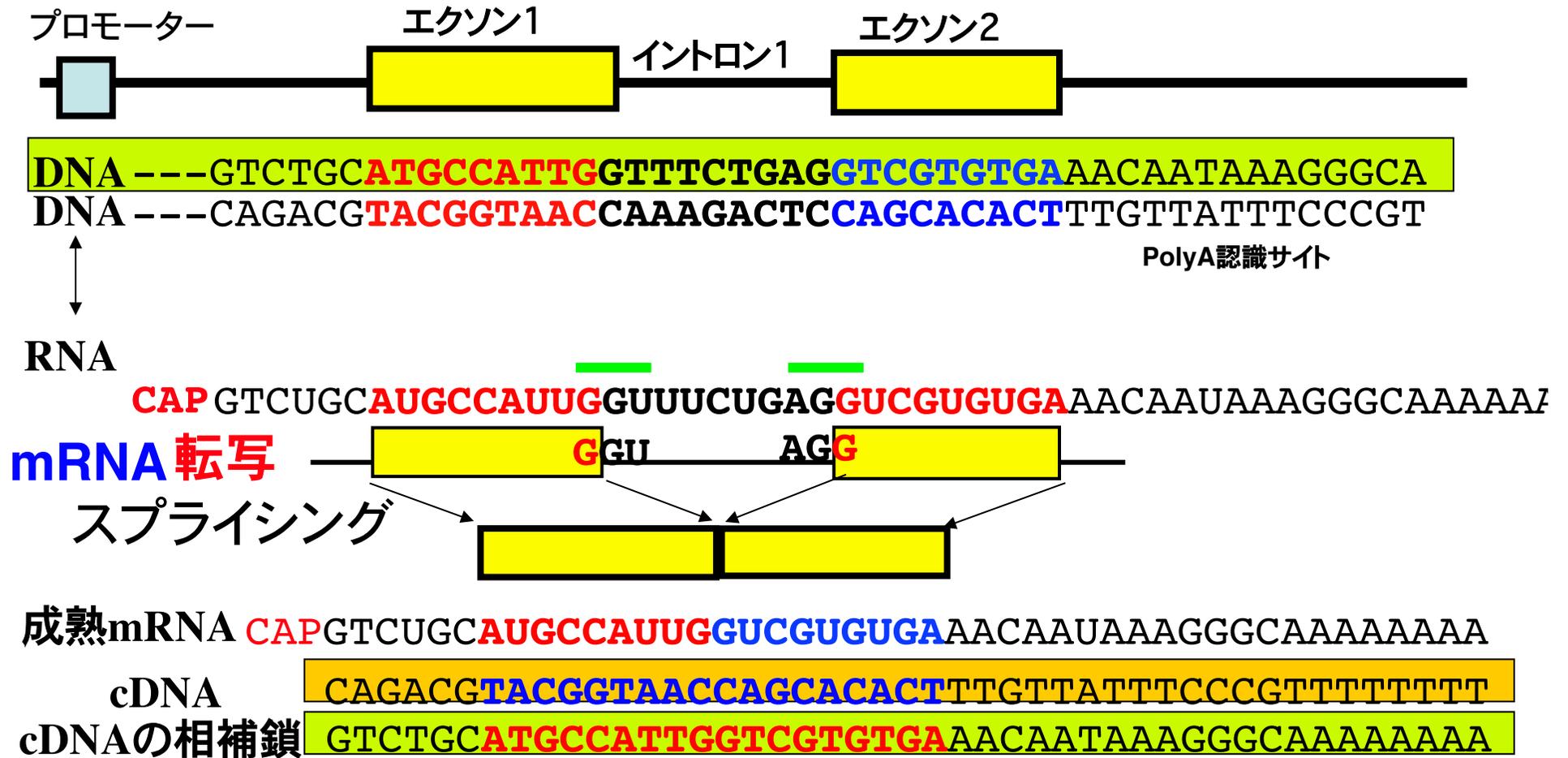
遺伝子予測ソフトで遺伝子領域を推定する

## 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。

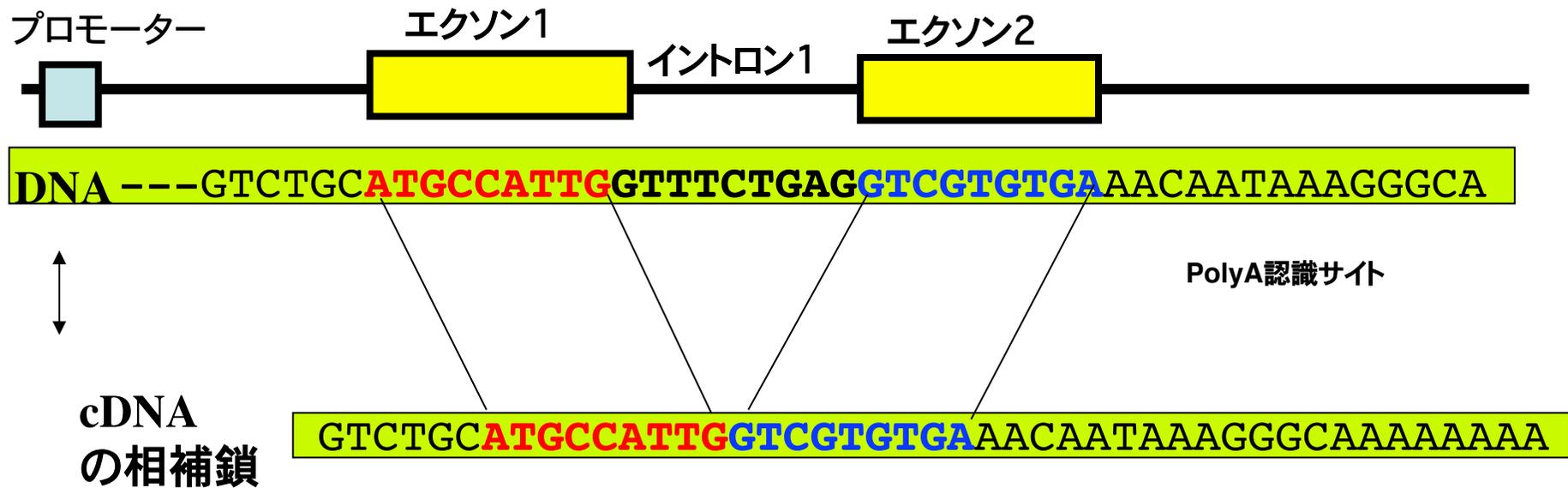
# cDNA情報が有るとき.1

## DNA配列とcDNA配列を比較する



## cDNA情報が有るとき.2

DNA配列とcDNA配列(の相補鎖)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
tcctcccgtg	tacaataacc	caaccctctc	gccacagacg	ctgcacctgc	acacacacac	acacttaaac	tcacacaacg	tctaacatca	ctccagctca
110	120	130	140	150	160	170	180	190	200
acacagagct	cacttctaat	ctcaaatctc	atggtggcgc	agcaacccac	gcccaaccag	cgcacacaac	caaccgccat	ggactaacac	gcggcctctg
210	220	230	240	250	260	270	280	290	300
gcattggcgc	cccgccggcg	gcggcggtgt	gcgacctgag	gatggagccc	aagatccagg	agccattctg	gtggccgaac	ggcgacgcga	ggccggcctc
310	320	330	340	350	360	370	380	390	400
ggcggcggag	ctggacaatgc	cctggctcga	cgctggcctg	ctccgcgacg	gcgacgcaga	ggggctgcgc	cgcgcgccgg	cgcagctggc	cgcgcgctgc
410	420	430	440	450	460	470	480	490	500
gccacgcacg	ggttcttcca	ggtgtccgag	caacggctcg	acgcccctct	ggcgcgcgcc	gcgcctgcgc	gcgccacaga	cttctctcgc	ctccgcctcg
510	520	530	540	550	560	570	580	590	600
ccgagaagcg	cgcgcgcgcg	cgcgctccgg	gcaccgtgtc	cggtacacac	agcgcaccaag	ccgaccgctt	cgcctccaag	ctcccattga	aggagacct
610	620	630	640	650	660	670	680	690	700
ctctctcgcg	ttccacgacc	gcgcgcgcgc	cccgcctgtc	gcgcgacct	ctctcagcac	ccctgcgcc	gactctgcgc	caatggggta	attaaaacga
710	720	730	740	750	760	770	780	790	800
tgtgtggcag	cattgcattt	caaatctaac	acaaatctaa	aaacacacga	ccgagattat	ctgtaataca	aacggctttg	tgcgcgcgag	aggygttacc
810	820	830	840	850	860	870	880	890	900
agaagtactg	cgaggagatg	aaggagctgt	cctctgacat	catggaaact	ctggagctga	gccctggcgt	ggagcgagcg	tactacaggg	agttctcgc
910	920	930	940	950	960	970	980	990	1000
ggacagcagc	taaatatcgc	ggtgcaacta	ctaccgcgca	tgcccggagc	cggagcggac	gctgcggcag	ggcccagcct	gcgcacccac	cgcctctaac
1010	1020	1030	1040	1050	1060	1070	1080	1090	1100
atctctctcc	agagcagcgt	cgcgcgcctc	gaggtctctg	tcgacggcga	atggcgcgcc	gtaccgcccg	tcaccggcgc	catgtctatc	aacatggcgc
1110	1120	1130	1140	1150	1160	1170	1180	1190	1200
acaccttcac	ggtaaacacat	ctctctctct	ctctctctct	gtctctctct	gcttccgaagc	aacagaacaa	gtaattcaag	cttttttttc	tctctcgcgc
1210	1220	1230	1240	1250	1260	1270	1280	1290	1300
gaaattgacg	agaaaataaa	gatctgtgta	ggggcggggc	tttcagctga	aagcgggaag	aaaccgacct	gacgtgattt	ctctgttcca	atacaaacaa
1310	1320	1330	1340	1350	1360	1370	1380	1390	1400
atggaaatgcc	caactctctc	atgtgttatg	atttatctca	catcttatag	ttaataggag	taagtaacaa	gctacttttt	taattattata	gttctgttga
1410	1420	1430	1440	1450	1460	1470	1480	1490	1500
tttttttttt	ttaagttttt	tttagtttta	tcaaaattta	ttgaaaact	tageaacgtt	tataatacca	aattagctct	atttagttta	atattgtata
1510	1520	1530	1540	1550	1560	1570	1580	1590	1600
tattttgata	atatatttat	gttatattaa	aaatattact	atatttttct	ataaacatta	ttaaaagcca	tttataatat	aaaatggaa	gagtaattaa
1610	1620	1630	1640	1650	1660	1670	1680	1690	1700
tatggatctc	cccgcacatg	agaatatttt	ccgatggctg	gacgacgcca	tgttaagcttc	ggctggccctg	gacggccaga	ggtgccacaa	gccacgctcca
1710	1720	1730	1740	1750	1760	1770	1780	1790	1800
acaaccccctg	ggtccccecc	taaacctcca	aacagtagtg	agtagtgtct	cgtcgcgttt	tagtatttga	tgacaacaaa	agtgtagttt	gagttagcca
1810	1820	1830	1840	1850	1860	1870	1880	1890	1900
ccaccaactt	gcacacgagc	acaataattt	gtgtccatc	tcgccagcca	cttcacatc	tagtctaac	tactctctag	cgatgttaagc	ggataatttc
1910	1920	1930	1940	1950	1960	1970	1980	1990	2000
atcatccgta	tataaacctg	ttgttatag	ttaatttctc	atataactat	ataacagtat	acattttaaa	agaaaacaaa	atttaggata	caagccctctg
2010	2020	2030	2040	2050	2060	2070	2080	2090	2100
ctctatccaa	tcattggcac	ttggaaggac	cagactcggg	catgccatgc	caagccaaga	tatgggttat	ggaagagtag	agaagaggag	agatgagaga
2110	2120	2130	2140	2150	2160	2170	2180	2190	2200
taagcatgag	ttctctctct	cgttgagatg	gtattttgga	gggatttctg	tagtagtagc	agcggcgcgc	cggggacaga	tgcgatgtgt	ggcgccttcc
2210	2220	2230	2240	2250	2260	2270	2280	2290	2300
gtggcgtttt	cccggggggg	ttttgttttg	gcgctggggg	gggatggcat	ggcgcggcgt	gcggctgcac	gccacacaca	cgcgcgagca	cgcacgctacg

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の相補鎖

10	20	30	40	50	60	70	80	90	100
atggtggcgc	agcacccecc	gccaccacag	cgcacacaac	caaccgccat	ggactccacc	gccggctctg	gcattggcgc	cccgccggcg	gcggcggtgt
110	120	130	140	150	160	170	180	190	200
gcgacctgag	gatggagccc	aagatccagg	agccattctg	gtggccgaac	ggcgacgcga	ggcggcgtct	ggcgggggag	ctggacatgc	cctgtgtcga
210	220	230	240	250	260	270	280	290	300
cgtggcgctg	ctccgcgacg	gcgacgcaga	ggggctgcgc	cgcgcccgcg	cgcaggtggc	cgcgcgctgc	gccacgcacg	ggttcttcca	ggtgtccag
310	320	330	340	350	360	370	380	390	400
caaggcgtgc	acgcgcctct	ggcgcgcgcg	gcgctgcagc	gcgcacagca	cttctctcgc	ctccgcctgc	ccgagagacc	cgcgcgcgcg	cgcgtctccg
410	420	430	440	450	460	470	480	490	500
gcaccgtgtc	cggctaacac	agcgcaccaag	cgcacgcctt	cgcctccaag	ctcccattga	aggagacctc	ctctctcgcg	ttccaacgac	gcgcgcgcgc
510	520	530	540	550	560	570	580	590	600
ccccctgtct	gcgcgacctc	ctctcagcac	ccctgcgcc	gactctgcgc	caatggggag	ggtgtaccag	aagtactgcg	aggagatgaa	ggagctgtctg
610	620	630	640	650	660	670	680	690	700
ctgacatca	tggaaactct	ggagctgcgc	ctggcgctgg	agcagagctc	ctacagggg	ttctctcgcg	acagcagctc	aactatgcgg	tgcacactct
710	720	730	740	750	760	770	780	790	800
accgcccatg	cccggagcgc	gagcggcgcg	tcggcagcgc	cccgcacctc	gacccacccg	cctctaccat	cctctccag	gacgagctcg	gcggcctgca
810	820	830	840	850	860	870	880	890	900
ggtctctctc	gacggcgaat	ggcgcgccct	cagcccccct	cccggcgcca	tggtaataaa	catcgccgac	accttaatgg	cgcgtctgaa	cgggaggtat
910	920	930	940	950	960	970	980	990	1000
aagagctgac	tgcaacaggc	ggtgtgtaac	cagcggcggg	agcggcgttc	gctggccttc	ttctctctgc	cgcggggaga	caggtgtgtg	cgcgcgcgcg
1010	1020	1030	1040	1050	1060	1070	1080	1090	1100
cagagcgcgc	caagccgcag	cactaccggg	acttcaactg	ggccgacctc	atgccttcca	cgcagcgcca	ctaccgcgcc	gacaccgcca	cgcctcagcgc
1110	1120	1130	1140	1150	1160	1170	1180	1190	1200
cttccagcgc	tgctctcgcg	cgcgcgcgcg	cgcgcgcgcc	gcgacgcgcg	aggctagagg	ggccagctga			

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 ctagccctg acacacacac acactaacac tcacacaagc tctaacata ctcacgctc
110     120     130     140     150     160     170     180     190     200
acacagagct cacttctcat ctcaaatct atggtggcg agcacccac gccacacag ccgaccacac caaccgccat ggactaacg gccgcctg
210     220     230     240     250     260     270     280     290     300
gcattggcgc cccggcggg gccggcggg gggacctgag gatggagccc aagatccagg agccattgt gtggcgaac ggcgacgaga ggcggcgtc
310     320     330     340     350     360     370     380     390     400
ggcggcggg ctagaacatg ccgtggtcga cgtggcgtg cctccgagc gcgacgcaga ggggcctgac cgcgccggg ccgaggtgac cgcgcgtgc
410     420     430     440     450     460     470     480     490     500
gccacagcag ggttattcca ggtgctcag cacggcctg acgccctct ggcgcgcgc ccgctcagc cttctccgc ctcaccgtg
510     520     530     540     550     560     570     580     590     600
ccgagagagc ccgcgcgcgc ccgctccagg gccaccgtgc ccgctacacc agcgcgcagc ccgaccgctt ccgctccacc ctcccatgga agggagacc
610     620     630     640     650     660     670     680     690     700
ctctctggc ttccacgacc ggcgcgcgc ccgctctgc gccacactt tctccagcacc cctggccccc gacttcagc caatggagta 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の相補差

```

10      20      30      40      50      60      70      80      90     100
atggtggcgc agcacccac gccacacag ccgaccacac caaccgccat ggactaacg gccgcctg gcattggcgc cccggcggg gccggcgtg
110     120     130     140     150     160     170     180     190     200
ggcaccctgag gatggagccc aagatccagg agccattgt gtggcgaac ggcgacgaga ggcggcgtc ggcggcggg ctggacatgc ccgtggtcga
210     220     230     240     250     260     270     280     290     300
cgtggcgtg ctcgcgcagc gcgacgcaga gggcctgac cgcgccggg ccgaggtgac cgcgcgtgc gccacagcag ggttattcca ggtgctcag
310     320     330     340     350     360     370     380     390     400
cacggcgtg acgcgcctt ggcgcgcgc gcgctcagc ggcgcgcaga cttctccgc ctcaccgtg ccgagagcag ccgcgcgcgc ccgctccagg
410     420     430     440     450     460     470     480     490     500
gccaccgtgc ccgctacacc agcgcgcagc ccgaccgctt ccgctccacc agcgcgcagc ccgaccgctt ccgctccacc ctcccatgga agggagacc
510     520     530     540     550     560     570     580     590     600
ccccctgct gccacactt tctccagcacc cctggccccc gacttcagc caatggagta ggtgctcagc aagctacgag agggagatgaa ggagcctgct
610     620     630     640     650     660     670     680     690     700
ggctctgct ggcgcgcgc ccgaccgctt ccgaccgctt ccgaccgctt ccgaccgctt ccgaccgctt ccgaccgctt ccgaccgctt ccgaccgctt

```

(EMBL) length: 1170 current pos: 1171

**DNA ---GTCTGCATGCCATTGGTTTCTGAGGTCGTGTGAACAATAAAGGGCA**

**cDNA GTCTGCATGCCATTGGTCGTGTGAACAATAAAGGGCAAAAAAAAA**

例

# ある配列のGenomicDNA

GA20oxidase2-genomic DNA

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

```

10      20      30      40      50      60      70      80      90     100
tc tccactgt tacaataacc caaccctctc gccacagacg ctgcacctgc aacacacac acactcaacc tcacaacaga tctcaactca ctcaccgtca
110     120     130     140     150     160     170     180     190     200
acacagcgt caactctcat ctccaatctc atggtggcgc agcaaccacc gccaccacag ccgacacac caccgcccat ggaactcaac gccgcctcgt
210     220     230     240     250     260     270     280     290     300
gcattggcgc cccggcgggc gcggcggtgt gccacgttag gatggagccc aagatcccag agcattctgt gttggccagac gccgacgcga gccacggcgtc
310     320     330     340     350     360     370     380     390     400
ggcggcgggc ctggacaatg ccgtgtgtga cgtggcgctg ctcgcgacg gccagcccca gggcctgcgc cgcgcgcggc cgcagggtgc cgcgcgctgc
410     420     430     440     450     460     470     480     490     500
gccacgcacg ggtttctcca ggtgtccagg caaggcctag acgcgcctct ggagcgcgcc gccctagaag gccgcacaga cttcttccgc ctcaccgtcg
510     520     530     540     550     560     570     580     590     600
ccgagaaagc ccgcgcgcgc cgcgtcccgc gccacgtgtc cggctacacc agcgcacca ccgaccgctt cgcctccaag ctccaatgga agggaccctt
610     620     630     640     650     660     670     680     690     700
ctcacttgcc ttccaacacc gccgcgcgc ccccgctgtc gccagctact tctccagcac cctggcctcc gacttcgcgc caatggggta attaaacaga
710     720     730     740     750     760     770     780     790     800
tggtagcaga cacttgcatc caaatcca aaacattcaa aacacacaga ccgagattat gctgaattca aacgcgtttg tgcgcgcagg aggggtgtcc
810     820     830     840     850     860     870     880     890     900
agaagctact cagaggatgt aaggagctgt cgtgacgat catggaactc ctggagctga gccctggcgt ggagcagacc tacataaccg agttcttccg
910     920     930     940     950     960     970     980     990     1000
ggacagcaga taactactgc ggtgcaacta ctaccgcgca tgcccggagc ccgagccgac gctggcacag gccccgcact gccaccaccc cgcctcaacc
1010    1020    1030    1040    1050    1060    1070    1080    1090    1100
atctctctcc agnacagcgt ccggcgctcc gaagtctctg tcgacggcga atggcgcctc gtaagccccc tcccgcgcgc catggtctat aacatggcgc
1110    1120    1130    1140    1150    1160    1170    1180    1190    1200
aacactcaat ggtaaacact ctactattct cctctctctc gttctctctc gcttcgaagc aacagaaaca gtaattcaag ctttttttcc tctctcgcgc
1210    1220    1230    1240    1250    1260    1270    1280    1290    1300
gaaattgacg agaaaaataa gatcgtggtc gggcgggggc ttccagctga aagcgggaag aaaccgacct gacgtgattt cctctgtcca atcaacaaca
1310    1320    1330    1340    1350    1360    1370    1380    1390    1400
atggaaatgcc ccactctctc atgtgttatg atttatctca catcttatag ttaataggag taagtaacaa 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 gttatatcaa 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
acaacccctg ggtccccccc taacactcca aacagtagtg agtagtgtct cgtcgcgttt tagtatttga tgacaacaac agtgtgattt gagtttagca
1810    1820    1830    1840    1850    1860    1870    1880    1890    1900
ccaccaactt gccacagcgc acatacattt gtgtcaattc tcgcccagtea cttccatctc tagtcttaac tectatctag cgtgttaagc 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
ctctatccca tcaatggcac ttgnaaggac cagactcgtt catgccatgc caagccaaga tatgggttat ggaagagtag agaagaggag agatgagaga
2110    2120    2130    2140    2150    2160    2170    2180    2190    2200
taagcattgc ttctactctc cgttggatgt gatttttggc gggattttgt tagtagtagc agcgcgcgcc cggggacaga tgcggatggt gccgcctttc
2210    2220    2230    2240    2250    2260    2270    2280    2290    2300
gtggcgtttt cccggggggg ttttgatttg gccctggggg gggatggcat ggcgagcgtt gccgctgcac gccaccacaca cgcgcgcgca gcaacgttag

```

(EMBL) Length: 2900 current pos: 401

# ある配列のcDNAの相補鎖

GA20oxidase2-cDNA

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

```

10      20      30      40      50      60      70      80      90     100
atggtggcgc agcaaccacc gccaccacag ccgacacac caccgcccat ggaactcaac gccgcctcgt gcattggcgc cccggcgggc gcggcggtgt
110     120     130     140     150     160     170     180     190     200
gccacgttag gatggagccc aagatcccag agcattctgt gttggccagac gccgacgcga gccacggcgtc cgcgcgctgc cgcgcgctgc
210     220     230     240     250     260     270     280     290     300
cgtggcgctg ctcgcgacg gccagcccca gggcctgcgc cgcgcgcggc cgcagggtgc cgcgcgctgc gccacgcacg ggtttctcca ggtgtccagg
310     320     330     340     350     360     370     380     390     400
caccgcgtcg acgcgcctct ggagcgcgcc gccctagaag cttcttccgc ctcaccgtcg ccgagaaagc ccgcgcgcgc cgcctcaacc cgcctcaacc
410     420     430     440     450     460     470     480     490     500
gcaacgtgtc cggctacacc agcgcacca ccgaccgctt agactccaag ctccaatgga agggaccctt ctcacttgcc ttccaacacc gccgcgcgc
510     520     530     540     550     560     570     580     590     600
ccccgtcttc gccacactct tctccagcac cctggcctcc gacttcgcgc caatggggag ggtgtaccag aagctactgc agggagtgaa ggagctgtcg
610     620     630     640     650     660     670     680     690     700
ctgacagaca tggaaactct ggagctgagc ctggggctgg agcagagcta ctacagggag ttcttgcggc acagcagctc aatcatcgcc tgcaactact
710     720     730     740     750     760     770     780     790     800
accgcagctg cccggcgagg gacggcagca tggcgcaggg ccgcacatgc gaccaccaag cctcaacctc cctctccag gacgactctg gccgctcaga
810     820     830     840     850     860     870     880     890     900
ggtctctctc gacggcgaaat ggccgccctg cagcccagct cccggcgcca tggtaataca catcggcagc acctcaatgg cgcctctgaa 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 cagcgcagag caactaccgg acttcaactg ggccgacctc atgctctca cgcagcagca ctaaccgcgc gacaccagca cgcctgagcc
1110    1120    1130    1140    1150    1160    1170    1180    1190    1200
cttcaacgcg ttgcctgcgc cgcgcgcgc ccagcgcgcc gccagcgcgc aggtcgagcc gcccagctga

```

(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 screenshot of a bioinformatics software interface with two windows side-by-side. The top window is titled "GA20oxidase2-genomic DNA" and the bottom window is titled "GA20oxidase2-cDNA". Both windows display DNA sequence alignments with position markers and sequence data.

**GA20oxidase2-genomic DNA**

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

Sequence alignment (Genomic DNA):

10	20	30	40	50	60	70	80	90	100
tcctccctgt	tacaaatacc	ccaccctctt	gccacagacg	ctgcccctgc	acacacacac	acactcaaac	tcacacacgc	tcacaaacta	ctcccgcctc
110	120	130	140	150	160	170	180	190	200
acacagcctg	caactctcat	ctcaaatctc	atggtggcgc	agcaaccacg	gcccaaccacg	ccgcacaaca	caaccgccaat	ggactcaacc	gccgcctctg
210	220	230	240	250	260	270	280	290	300
gcattggcgc	ccagcgcgcg	gcggcgggtg	gcgaccctag	gatggagccc	aagatcccgc	agccattctg	gtggccgaac	ggcgacgcga	ggcggcgctc
310	320	330	340	350	360	370	380	390	400
ggcggcggag	ctggacaatg	ccgtggtcga	ctctggcgct	ctccgcgacg	gcgacgcgca	ggggctgcgc	cgccgcgcgc	cgccggctgc	cgccgcgcgc
410	420	430	440	450	460	470	480	490	500
gccacgcacg	ggttcttcca	gggtctcgag	caaggcctgc	acgcgcctct	ggcgcgcgcg	gcgcctgcgc	gcgcgcgcgc	cttctccgcg	ctcccgcctc
510	520	530	540	550	560	570	580	590	600
ccgagagcgc	ccgcgcgcgc	ccgcctccgc	gcaccctgtc	cggtctaccc	agcgcgcacg	ccgacgcctc	cgccctcaag	ctcccatgga	aggagaccct
610	620	630	640	650	660	670	680	690	700
ctctctgcgc	ttccacgcac	gcgcgcgcgc	ccccgtctgc	gcgcgactct	ctccacgcac	ccctggcccc	gaactctgcg	caatggggta	attaaacaga
710	720	730	740	750	760	770	780	790	800
tggtggcaga	cattgcattt	caaatcaaaa	acaaattcaa	aacacacaga	ccgagattat	gctgaattca	aaccgctttg	tgccgcgcgc	agggtgtacc
810	820	830	840	850	860	870	880	890	900
agaagtactg	cgagagatg	aaggagctgt	cgtctacgat	catggaaact	ctggagctga	gccctggcgt	ggagcggagc	tactacaggg	agttctctgc
910	920	930	940	950	960	970	980	990	1000
ggacagcgcg	taaatcatgc	ggtagcaact	ctaccgcaca	tgcccggagc	cggaagcgac	gctcggacgc	ggcgcgcacg	cgccctcaac	cgccctcaac
1010	1020	1030	1040	1050	1060	1070	1080	1090	1100
atctctctcc	aggaacagct	cgccggcctc	gaggtctctg	tcgacggaga	atggcccccg	gtacccccgc	tcccaggcgc	catgtctatc	aacatagcgc
1110	1120	1130	1140	1150	1160	1170	1180	1190	1200
acacctcat	ggtaaacat	ctctctattt	ccctctctct	gttctctctt	gcttctctct	gcttctctct	gcttctctct	gcttctctct	gcttctctct
1210	1220	1230	1240	1250	1260	1270	1280	1290	1300
gaaattgacg	agaaaaataa	gatctgggta	ggggcggggc	tttcaactga	aagcgggaag	aaaccgacct	gaactgtatt	ctctgttcca	atcacaacaa
1310	1320	1330	1340	1350	1360	1370	1380	1390	1400
atggaatgac	caactctctc	atgtgttatg	atttatactc	catcttatag	ttaataggag	taagtaacaa	gctacttttt	tcattattata	gttctgttga
1410	1420	1430	1440	1450	1460	1470	1480	1490	1500
tttttttttt	ttaaagtttt	tttagtttta	ttcaaattta	ttgaaaaact	tggcaacgct	tataatacca	aattagctct	atttagttta	atattgtata
1510	1520	1530	1540	1550	1560	1570	1580	1590	1600
tattttgata	atataattat	gttatattaa	aaatattact	atattttctt	ataaacotta	ttaaagccca	tttataatat	aaaatgaag	gagttaattaa
1610	1620	1630	1640	1650	1660	1670	1680	1690	1700
tatggatctc	ccccgacatg	agaatatttt	ccgatgggtg	gacgacgcca	tgtaagcttc	ggtagggcct	gacgcccaga	ggtgcccaca	gccacgtcca
1710	1720	1730	1740	1750	1760	1770	1780	1790	1800
acaaccctct	ggctcccccc	taaacctcca	aacagtagtg	agtagtgctt	cgctgcgctt	tagtatgtta	tgacaacaaa	agtgtagatt	gagttagcca
1810	1820	1830	1840	1850	1860	1870	1880	1890	1900
ccaccaactt	gcaacagcgc	acaatacttt	gtgtccattc	tcgcagctca	cttccatctc	tagctctaac	ctctctctag	cgatgttaag	ggaataattc
1910	1920	1930	1940	1950	1960	1970	1980	1990	2000
atcatctcta	tataaacctg	ttgttatag	tttaattctt	atataactat	ataacagtat	acaattttaa	agaaaacaaa	attaggataa	acaggccctg
2010	2020	2030	2040	2050	2060	2070	2080	2090	2100
ctctctctca	tcctatgcac	ttggaaggac	cagactcggg	catgccatgc	caagcccaaga	tatgggttat	ggaagatag	agaagaggag	agatgagaga
2110	2120	2130	2140	2150	2160	2170	2180	2190	2200
taagcatgag	ttctctctct	cgctggatgt	gtattttgga	gggattttgt	tagtagtagc	agcgcgcgcg	cggggacaga	tgccgatgtt	ggcgccttct
2210	2220	2230	2240	2250	2260	2270	2280	2290	2300
gtggcgtttt	ccccgggggg	ttttggtttg	gcgctggggg	gggatggcat	ggcgcgcgcg	gcgcctgcac	gccacaacaa	cgccgcgcga	cgccgcctag

(EMBL) Length: 2900 current pos: 481

**GA20oxidase2-cDNA**

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

Sequence alignment (cDNA):

10	20	30	40	50	60	70	80	90	100
atggtggcgc	agcaaccacg	gccacacacg	ccgcacaaca	caaccgccaat	ggactcaacc	gccgcctctg	gcattggcgc	cccgccggag	gcggcgggtg
110	120	130	140	150	160	170	180	190	200
gcgaccctgag	gatggagccc	aagatcccgc	agccattctg	gtggccgaac	ggcgacgcga	ggcggcgctc	ggcggcggag	ctggacaatg	ccgtggtcga
210	220	230	240	250	260	270	280	290	300
cgtagggcgt	ctccgcgacg	gcgacgcgca	ggggctgcgc	cgccgcgcgc	cgcaagctgg	cgccgcgcgc	ggctcttcca	gggtctccag	gggtctccag
310	320	330	340	350	360	370	380	390	400
caaggcctgc	acgcgcctct	ggcgcgcgcg	gcgcctgcgc	gcgcgcgcgc	cttctccgcg	ctcccgcctc	ccgagaagcg	cccgccgcgc	cgctctccgc
410	420	430	440	450	460	470	480	490	500
gcaccgtctc	cggtctaccc	agcgcgcacg	ccgacgcctc	cgccctcaag	ctcccatgga	aggagaccct	ctctctctgc	ttccacgcac	gcgcgcgcgc
510	520	530	540	550	560	570	580	590	600
ccccgtctgc	gcgcgactct	ctccacgcac	ccctggcccc	gaactctgcg	caatggggta	gggtctccag	aagtagctgc	aggagatgaa	ggagctgtct
610	620	630	640	650	660	670	680	690	700
ctgacgatac	tggaactctt	ggagctgagc	ctgggcgtgg	agcagggcga	ctacagggag	ttctctccgc	acagcagctc	aaatctcggg	tgcaacttact
710	720	730	740	750	760	770	780	790	800
accgcgcaat	ccgagcgcgc	gagcggagcgc	tcggcagcgg	cccgccctgc	gaaccacacc	ccctcaacat	ccctctccag	gaagcagctg	ggcgcctcga
810	820	830	840	850	860	870	880	890	900
ggctctctgc	gaagcagcgc	ggcgcgcgct	cagcccccgc	cccgccgcca	tggtctcaaa	catcggcgac	accctcaatg	cgctgtcga	cgggaggtat
910	920	930	940	950	960	970	980	990	1000
aagagctgac	tgcaacaggg	ggtagtgacg	cagcggcggg	agcggcggct	gctggcgttc	ttctctctgc	cgccgggaga	cagggtgggt	cgccgcgcgc
1010	1020	1030	1040	1050	1060	1070	1080	1090	1100
cgagcgcgcg	caagcgcgcg	caactaccgc	acttcaactg	ggcgcgcctc	atgccttca	agcagcgcga	ctaccgcgcg	gaacacgcga	cgctctcagc
1110	1120	1130	1140	1150	1160	1170	1180	1190	1200
cttcaacgcg	tggtctcgcg	cgccgcgcgc	cgaccgcgcg	gcgcgcgcgc	aggctcaggg	ggccagcctg			

(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

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	AGCACCCTAC	GCCACCCAC	CCGACCCAC	CACCCGCCAT	GGACTCCACC	GCCGGCTCG
210	220	230	240	250	260	270	280	290	300
GCATTGCCGC	CCCGCGCGCG	GCGCGGGTGT	GCAGCTCAG	GATGGAGCCC	AGATCCCGG	AGCCATTCTG	GTGGCCGAC	GGCGACCGA	GCCCGGGTC
310	320	330	340	350	360	370	380	390	400
GCCGGCGGAG	CTGGACATGC	CCGTGGTCGA	CGTGGGCGTG	CTCCGCGACG	GCGACGCCGA	GGGGCTGCCG	CGCCCGCGG	GCGAGGTGG	CGCCCGGTG
410	420	430	440	450	460	470	480	490	500
GCCACGACGC	GGTTCTTCCA	GGTGTCCGAG	CACGGCGCTG	ACGCCGCTCT	GGCGCGCGCC	GCGCTGACG	GCGCCACGCA	CTTCTCCCG	CTCCCCTCG
510	520	530	540	550	560	570	580	590	600
CCGAGGAGCG	CCGCGCGCGC	GCGGTCCCGG	GCACCCTGCT	GCGGTACACC	AGCGCCACAC	CCGACCGCTT	CGCTCCAGG	CTCCCATTGA	AGGAGACCT
610	620	630	640	650	660	670	680	690	700
CTCCTTCGGC	TTCCAGGACC	GCGCGCGCGC	CCCGCTGCTG	GCGACTACT	TCTCCAGCAC	CCTCGGCCCC	GACTTCGGC	CAATGGGTA	ATTAAACGA
710	720	730	740	750	760	770	780	790	800
TGGTGGACGA	CATTGCATT	CAAAATTCAR	ACAAATTCAR	AACACACCGA	CCGAGATTAT	GCTGAATTC	AACCGTTCG	TGGCGGACG	AGGTGTACC
810	820	830	840	850	860	870	880	890	900
AGAACTACTG	CGAGGAGATG	AGAGGCTGT	CGCTGACGAT	CATGGACTC	CTGGAGCTGA	GCCTGGGCT	GGAGCGAGG	TACTACAGG	AGTTCTTCG
910	920	930	940	950	960	970	980	990	1000
GGACGACGCG	TCAATCATCG	GGTGACATCA	CTACCCGCGA	TGCCCGGACG	GCGAGCGGAC	GCTCGGACG	GCCCGCGAC	GCGACCCCA	GCCCTTCAC
1010	1020	1030	1040	1050	1060	1070	1080	1090	1100
ATCCTCTCTC	AGGACGACGT	CAGCGGCTCT	GAGGTCTCTG	TGAGCGGCGA	ATGGCGCCG	GTCAGCCCG	CATGGTCTC	AGCATCGCG	
1110	1120	1130	1140	1150	1160	1170	1180	1190	1200
ACACCTTCAT	GTAARACCAT	CTCCTATTCT	CCTCTCCTCT	GTTCTCCTCT	GCTTCGAGC	AACACACARA	GTAATTCAG	CTTTTTTTC	TCTCTCGGC
1210	1220	1230	1240	1250	1260	1270	1280	1290	1300
GAATTTGACG	AGAAAATATA	GATCGTGGTA	GGGGCGGGCG	TTTCACTGA	AGCGGGGAG	AAACCGACT	GCGGTATT	CTCTGTTC	ATCACARCA
1310	1320	1330	1340	1350	1360	1370	1380	1390	1400
ATGGAAATGCC	CCACTCTCTC	ATGTGTATAG	ATTATCTCA	CATCTTATAG	TTAATAGGAG	TAAATACARA	GCTACTTTTT	TCATATTATA	GTTCTGTTG
1410	1420	1430	1440	1450	1460	1470	1480	1490	1500
TTTTTTTTTT	TTAAAGTTTT	TTTAGTTTTT	TCCAAATTTA	TTGAAARAAT	TAGCAACGTT	TATATATCCA	AAATGATCTC	ATTTAGTTTA	ATATTGTATA
1510	1520	1530	1540	1550	1560	1570	1580	1590	1600
TATTTTGATA	ATATATTTAT	GTTATATTAA	AAATATTACT	ATATTTTCT	ATARAACATA	TTAAARACCA	TTTATTAAT	AAATGGAG	GAGTAAATTA
1610	1620	1630	1640	1650	1660	1670	1680	1690	1700
TATGGATCTC	CCCCGACATG	AGAAATTTTT	CCGATGGTGT	GACGACGCCA	TGTAAGCTTC	GTTGGGCTGT	GACGGCCAGA	GTTGCCARCA	GCCACGTC
1710	1720	1730	1740	1750	1760	1770	1780	1790	1800
ACACCCCTCG	GGTCCCGCCC	TAACTACCA	ACAGATAGTG	AGTAGTGTCT	CGTCCGTTT	TAGTATTTGA	TGACAAACAA	AGTGTAGTT	GAGTTAGCCA
1810	1820	1830	1840	1850	1860	1870	1880	1890	1900
CCACCAACTT	GACACGAGCG	ACATACATTT	GTGCTCATTC	TCGCCAGTCA	CTTCCATCTC	TAGCTTACG	TCCTATCTAG	CGATGTAGC	GGATAAITTC
1910	1920	1930	1940	1950	1960	1970	1980	1990	2000
ATCATCCGTA	TATAAACCTG	TTTGTATAG	TAAATTTCT	ATATAACT	ATACAGTAT	ACATTTTAAA	AGAAAACAAA	ATTAGGATTA	ACAGCCCTG
2010	2020	2030	2040	2050	2060	2070	2080	2090	2100
CTCCTATCCA	TCCATGGCAC	TTGGAGGAC	CAGACTCGGT	CATGCCATGC	CAGGCCAGCA	TATGGTTTAT	GGAGAGTAG	AGAAAGGAG	AGATGAGAA
2110	2120	2130	2140	2150	2160	2170	2180	2190	2200
TAAAGATGCG	TTCTCTCTCT	CGTGTGATGT	GTAATTTTGA	GGGATTTGTG	TAGTAGTAGC	AGCGCGGCCG	CGGGGACGGA	TGCGATGGT	GCCGCTTTCG
2210	2220	2230	2240	2250	2260	2270	2280	2290	2300
GTGGCGTTTT	CCCGGGGGGG	TTTTGGTTTT	GCGCTTGGGG	GGGATGGCAT	GGCGCGCGT	GCGCTGCGAC	GCCACGACCA	GCGCGCGCGA	CGCAGCTACG
2310	2320	2330	2340	2350	2360	2370	2380	2390	2400
TCGTCTGTCG	CAGCGGGCGA	CGTAGCTTAA	GGTGGTGTGT	TTCCGCGCGC	GGCGCGGGAT	TGTTCCATGC	GATCGATTT	GCGCCACCC	TCGCCGCGC
2410	2420	2430	2440	2450	2460	2470	2480	2490	2500
TCTGTGCGCG	TCGTGCGCCT	CTCTCGCGCG	GTTTGTCTCT	GTCGCGTTGC	TCAGCCGGCG	ACGGGGCGAC	GACAAATGGC	GATGTAGCCC	TGACGTGTG
2510	2520	2530	2540	2550	2560	2570	2580	2590	2600
GCCCTCTCCG	TTGATGATG	ATGATGTATG	TATGTATTTT	TTTTGTCTG	AGGAAATTTG	TGGGAAATG	TTGTGTGTG	AGCGCGTCT	GACCGGGAG
2610	2620	2630	2640	2650	2660	2670	2680	2690	2700
TATAAGACT	GCCTGACAG	GCGCGGGTGT	AACAGCGGCG	GGAGCGGGCG	GTCGCTGGCG	TTCTTCTCTG	GCCCGCGGGA	GGACAGGGT	GTGCGCGCG
2710	2720	2730	2740	2750	2760	2770	2780	2790	2800
CGCCGACGCG	CGCCGACGCG	CAGCACTACC	CAGACTTCC	CTAGGCGGAC	CTCATGCGCT	TCACCGGCG	CCACTACCG	GCCGACACC	GACGCTCGA
2810	2820	2830	2840	2850	2860	2870	2880	2890	2900
GCCCTTCCAG	CGCTGGCTCG	CAGCGCGCGC	CAGCGCGGCG	GCCGCGACG	CAGAGGTGCA	GCGCGCCAC	TGATCCCGCA	ACGGACGCA	ACGAAACGA

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

解析ソフトを用いて遺伝子を探す

cDNAが有る場合

# 生物情報工学II (担当 芦莉) 遺伝子の配列解析

第6回目-1 (11/1) [遺伝子の予測と機能解析](#)

-2 (11/1) [演習](#)

第7回目-1 (11/8) [DNA配列から遺伝子を探す](#)

-2 (11/8) [DNA、アミノ酸のアラインメント、系統樹解析](#)

-3 (11/8) [演習](#)

第8回目-1 (11/15) [Genetyxを用いた解析](#)

-2 (11/15) [演習](#)

第9回目-1 (11/22) [KEGGを用いた解析](#)

[連鎖解析](#)

出席と宿題の提出 [提出方法](#)

## 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](#) : 何でも検索
4. [Google Scholar](#) : 科学に関する検索
5. [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. [Splign](#) : DNAとcDNAのマッチング 
2. [Blast align](#) : DNAとcDNAのマッチング
3. [EMBOSS program"](#) : DNAとcDNAのマッチング
4. [EMBOSS program"](#) : DNA配列をアミノ酸配列に変換
5. [BLAST\(NCBI\)](#) : 遺伝子のホモロジー検索
6. [BLAST\(DDBJ\)](#) : 遺伝子のホモロジー検索
7. [Softberry/FGENESH](#) : 遺伝子の予測
8. [GeneMark](#) : 遺伝子の予測
9. [GENESCAN](#) : 遺伝子の予測
10. [clustalW](#) : アラインメント
11. [InterProScan](#) : ドメイン解析
12. [PSORT](#) : 細胞内局在性予測
13. [GEO](#) : 遺伝子発現解析
14. [Gene Ontology](#) : 用語を用いた検索
15. [SSPN](#) : スプライシングの予測

Splign is a utility for computing cDNA-to-Genomic, or spliced sequence alignments. At the heart of the program is a compartmentization algorithm which identifies possible gene duplications, and a refined alignment algorithm recognizing introns and splice signals.

Please follow one of the links below or navigate using the menu bar at the top of this page.

- [Download a stand-alone application](#) if you need to compute alignments for a large number of sequences.
- If you have a small number of sequences to align or just want to try out, [click here](#) to submit an online job.
- [Read the documentation](#).

**References:**

- Y.Kapustin, A.Souvorov, T.Tatusova and D.Lipman. *Splign: algorithms for computing spliced alignments with identification of paralogs*. *Biology Direct*, May 2008.
- Y.Kapustin, A.Souvorov and T.Tatusova. *Splign - a Hybrid Approach To Spliced Alignments*. RECOMB 2004 - Currents in Computational Molecular Biology. p.741.

**Related links:** [ProSplign](#), the tool for alignment of proteins against genomic sequences which is aware of splice signals and possible frameshifts.

**Help:** for questions, comments, or bug reporting, please visit [NCBI Support Center](#)

Please specify input sequences by GI/Accession or in FASTA format.

Examples (click to select):

- [NM\\_214647 / NW\\_732498 \(one model\)](#)
- [AF238306 / NT\\_033777 \(one model, frameshifts\)](#)
- [NM\\_020978 / NG\\_004750 \(multiple models\)](#)

**cDNA:**

**Genomic:**

From:  To:

- Lower quality query sequence (e.g. EST)
- Reverse and complement the query
- More partial alignments
- Use discontinuous megablast (e.g. for cross-species)

**Whole genome:**

**Help:** for questions, comments, or bug reporting, please visit [NCBI Support Center](#)

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

配列解析：

1. [Splign](#) : DNAとcDNAのマッチング
2. [Blast align](#) : DNAとcDNAのマッチング
3. [EMBOSS program](#)" : DNAとcDNAのマッチング
4. [EMBOSS program](#)" : DNA配列をアミノ酸配列に変換
5. [BLAST\(NCBI\)](#) : 遺伝子のホモロジー検索
6. [BLAST\(DDBJ\)](#) : 遺伝子のホモロジー検索
7. [Softberry/FGENESH](#) : 遺伝子の予測
8. [GeneMark](#) : 遺伝子の予測
9. [GENESCAN](#) : 遺伝子の予測
10. [clustalW](#) : アライメント
11. [InterProScan](#) : ドメイン解析
12. [PSORT](#) : 細胞内局在性予測
13. [GEO](#) : 遺伝子発現解析
14. [Gene Ontology](#) : 用語を用いた検索
15. [SSPN](#) : スプライシングの予測
16. [ORF Finder](#) : ORFの予測
17. [TFSEARCH](#) : 転写因子結合部位の予測
18. [MOTIF](#) : タンパク質モチーフ検索
19. [pI/Mw](#) : 等電点、分子量の予測
20. [RAP-DB](#) : イネのゲノムデータベース

解析用配列

1. [ある遺伝子AのgenomeDNA配列](#)



2. [ある遺伝子AのcDNA配列](#)



3. [ある遺伝子XのgenomeDNA配列](#)

4. [ある遺伝子XのcDNA配列](#)

5. [ある遺伝子YのgenomicDNA](#)

6. [ある遺伝子YのcDNA](#)

7. [Z1遺伝子のgenomicDNA](#)

8. [Z1遺伝子のCDS](#)

9. [Z1遺伝子の予想アミノ酸配列](#)

10. [Z2遺伝子のgenomicDNA](#)

11. [Z2遺伝子のCDS](#)

12. [Z2遺伝子のアミノ酸配列](#)

13. [野生型SD1-genomicDNA配列](#)

14. [野生型SD1-cDNA配列](#)

15. [カルロース-cDNA](#)

Please specify input sequences by GI/Accession or in FASTA format.

Examples (click to select):

- [NM\\_214647 / NW\\_732498 \(one model\)](#)
- [AF238306 / NT\\_033777 \(one model, frameshifts\)](#)
- [NM\\_020978 / NG\\_004750 \(multiple models\)](#)

## FASTA format

**cDNA:**

>AcDNA  
atggtggccgagcaccaccgcccaccacagccgacccaaccaccgccc  
ggactccaccgcccggctctggcattgcccggccggcgggcggtg  
cgacctgaggatggagcccaagatcccggagccattcgtgtggcgaacg  
cgacgcgagccggcgtcgggcgagctggacatgccctgtgtcgacg  
tggcggtgctccgacggcgacccgagggctgcccggcggcggc

**Genomic:**

From:  To:

>Genome  
tctccctgttacaataccccccctcctgcccagacagctgcctgcaca  
cacacacactcaactcacacagctctcaactactcccgtcaacacag  
cgctcacttctcatctccaatctcatggtggccgagcaccaccgcccaca  
gccgaccaaccaccgcccattggactccaccgcccgtctggcattgccg  
cccggcggcggcggtgtcgacctgaggatggagcccaagatcccgg

- Lower quality query sequence (e.g. EST)
- Reverse and complement the query
- More partial alignments
- Use discontinuous megablast (e.g. for cross-species)

**Whole genome:**

**Help:** for questions, comments, or bug reporting, please visit [NCBI Support Center](#)

#	Query	Subject	Span(bp)	Coverage(%)	Overall(%)	Exon(%)	CDS(%)	In-frame(%)
1	AcDNA(+)	Genome(+)	131-2873	100.00	100.00	100.00	0.00	0.00

<b>Model 1</b>	Coverage 100.00%	CDS 0.00%	Mismatches and indels 0
	Overall 100.00%	In-frame 0.00%	Exons (min/max/ave), bp 291 / 557 / 390
	Exon 100.00%	Primary transcript 1170 bp	Introns (min/max/ave), bp 103 / 1472 / 787

Graphics|Text



クリック

Segments Alignment

```

1 2 3
1 M V A E H P T P P Q P H Q P P P M D S T A G S
ATGGTGGCCGAGCACCCACGCCACACAGCGCCACCAACACCGCCCATGGACTCCACGCGCGGCTCTG
131 ATGGTGGCCGAGCACCCACGCCACACAGCGCCACCAACACCGCCCATGGACTCCACGCGCGGCTCTG
G I A A P A A A A V C D L R M E P K I P E P F V
71 GCATTGCCGCCCGCGGGCGGGCGGGCGGTGTGCGACCTGAGGATGGAGCCCAAGATCCCGGAGCCATTCGT
201 GCATTGCCGCCCGCGGGCGGGCGGGCGGTGTGCGACCTGAGGATGGAGCCCAAGATCCCGGAGCCATTCGT
W P N G D A R P A S A A E L D M P V V D V G V
141 GTGGCCGAACGGCGACGCGAGGGCCGGCGTGGCGGGCGGAGCTGGACATGCCCGTGGTTCGACGTGGGCGTG
271 GTGGCCGAACGGCGACGCGAGGGCCGGCGTGGCGGGCGGAGCTGGACATGCCCGTGGTTCGACGTGGGCGTG
L R D G D A E G L R R A A A Q V A A A C A T H
211 CTCGCGACGGCGACGCGAGGGGCTGCGCCGCGCCGCGGCGCAGGTGGCGCGCCGCTGCGCCACGCACG
341 CTCGCGACGGCGACGCGAGGGGCTGCGCCGCGCCGCGGCGCAGGTGGCGCGCCGCTGCGCCACGCACG
G F F Q V S E H G V D A A L A R A A L D G A S D
281 GGTTCCTCCAGGTGTCCGAGCACGGCGTTCGACGCGCTCTGGCGCGCGCCGCTCGACGGCGCCAGCGA
411 GGTTCCTCCAGGTGTCCGAGCACGGCGTTCGACGCGCTCTGGCGCGCGCCGCTCGACGGCGCCAGCGA
F F R L P L A E K R R A R R V P G T V S G Y T
351 CTTCTTCGCTCCCGCTCGCCGAGAAGCGCGCGCGCGCGCTCCCGGGCACCGTGTCCGGCTACACC
481 CTTCTTCGCTCCCGCTCGCCGAGAAGCGCGCGCGCGCGCTCCCGGGCACCGTGTCCGGCTACACC
S A H A D R F A S K L P W K E T L S F G F H D
421 AGCGCCACGCGGACCGCTTCGCTCCAAGCTCCCATGGAAGGAGACCTCTCCTTCGGCTTCCACGACC
551 AGCGCCACGCGGACCGCTTCGCTCCAAGCTCCCATGGAAGGAGACCTCTCCTTCGGCTTCCACGACC
R A A A P V V A D Y F S S T L G P D F A P M G
491 GCGCGCGCGCCCGTTCGCGCGACTACTTCTCAGCACCTCGGCCCGACTTCGCGCCAATGGG...
621 GCGCGCGCGCCCGTTCGCGCGACTACTTCTCAGCACCTCGGCCCGACTTCGCGCCAATGGGGTA

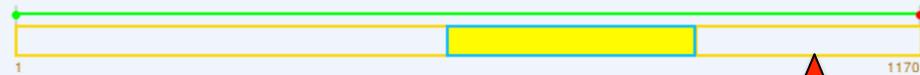
558 ..
691 AT
    
```

#	Query	Subject	Span(bp)	Coverage(%)	Overall(%)	Exon(%)	CDS(%)	In-frame(%)
1	A_cDNA(+)	A_genomic_DNA_ATG(+)	1-2743	100.00	100.00	100.00	0.00	0.00

Graphics|Text

Model 1	Coverage	CDS	Mismatches and indels
	100.00%	0.00%	0
	Overall 100.00%	In-frame 0.00%	Exons (min/max/ave), bp 291 / 557 / 390
	Exon 100.00%	Primary transcript 1170 bp	Introns (min/max/ave), bp 103 / 1472 / 787

A\_cDNA (+)



A\_genomic\_DNA\_ATG (+)



Segments Alignment

```

1 2 3
558 . . . . . R V Y Q K Y C E E M K E L S L T I M E L L
      | | | | |
655 CGCAGGAGGGTGTACCAGAACTACTGCGAGGAGATGAAGGAGCTGTCGCTGACGATCATGGAACCTCTGG
      | | | | |
      E L S L G V E R G Y Y R E F F A D S S S I M R C
623 AGCTGAGCCTGGGCGTGGAGCGAGGCTACTACAGGGAGTTCTTCGCGGACAGCAGCTCAATCATGCGGTG
      | | | | |
725 AGCTGAGCCTGGGCGTGGAGCGAGGCTACTACAGGGAGTTCTTCGCGGACAGCAGCTCAATCATGCGGTG
      | | | | |
      N Y Y P P C P E P E R T L G T G P H C D P T A
693 CAACTACTACCCGCCATGCCCGGAGCCGGAGCGGACGCTCGGCACGGGCCCCGCACTGCGACCCACCGCC
      | | | | |
795 CAACTACTACCCGCCATGCCCGGAGCCGGAGCGGACGCTCGGCACGGGCCCCGCACTGCGACCCACCGCC
      | | | | |
      L T I L L Q D D V G G L E V L V D G E W R P V
763 CTCACCATCCTCCTCCAGGACGACGTCGGCGGCCTCGAGGTCCTCGTCGACGGCGAATGGCGCCCGTCA
      | | | | |
865 CTCACCATCCTCCTCCAGGACGACGTCGGCGGCCTCGAGGTCCTCGTCGACGGCGAATGGCGCCCGTCA
      | | | | |
      S P V P G A M V I N I G D T F M
833 GCCCGTCCCGGGCGCCATGGTCATCAACATCGGCGACACCTTCATG . . . . .
      | | | | |
935 GCCCGTCCCGGGCGCCATGGTCATCAACATCGGCGACACCTTCATGGTAAA
  
```

#	Query	Subject	Span(bp)	Coverage(%)	Overall(%)	Exon(%)	CDS(%)	In-frame(%)
1	A_cDNA(+)	A_genomic_DNA_ATG(+)	1-2743	100.00	100.00	100.00	0.00	0.00

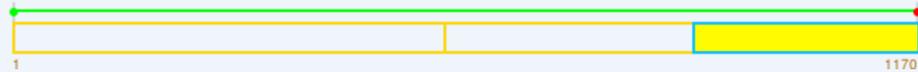
Graphics|Text

Model 1	Coverage	100.00%	CDS	0.00%	Mismatches and indels	0
	Overall	100.00%	In-frame	0.00%	Exons (min/max/ave), bp	291 / 557 / 390
	Exon	100.00%	Primary transcript	1170 bp	Introns (min/max/ave), bp	103 / 1472 / 787



クリック

A\_cDNA (+)



A\_genomic\_DNA\_ATG (+)



Segments Alignment

1	2	3
880	.....GCGCTGTCGAACGGGAGGTATAAGAGCTGCCTGCACAGGGCGGTGGTGAACCAGCGGGAGCG	A L S N G R Y K S C L H R A V V N Q R R E R
2448	TGCAGGCGCTGTCGAACGGGAGGTATAAGAGCTGCCTGCACAGGGCGGTGGTGAACCAGCGGGAGCG	
945	GCGGTCGCTGGCGTTCTTCCTGTGCCCGCGGGAGGACAGGGTGGTGCGGCCGCCGAGCGCCGCCACG	R S L A F F L C P R E D R V V R P P P S A A T
2518	GCGGTCGCTGGCGTTCTTCCTGTGCCCGCGGGAGGACAGGGTGGTGCGGCCGCCGAGCGCCGCCACG	
1015	CCGCAGCACTACCCGGACTTCACCTGGGCCGACCTCATGCGCTTCACGCAGCGCCACTACCGCGCCGACA	P Q H Y P D F T W A D L M R F T Q R H Y R A D
2588	CCGCAGCACTACCCGGACTTCACCTGGGCCGACCTCATGCGCTTCACGCAGCGCCACTACCGCGCCGACA	
1085	CCCGCACGCTCGACGCCTTCACGCGCTGGCTCGCGCCGCCGCGCCGACGCGCCGCGACGGCGCAGGT	T R T L D A F T R W L A P P A A D A A A T A Q V
2658	CCCGCACGCTCGACGCCTTCACGCGCTGGCTCGCGCCGCCGCGCCGACGCGCCGCGACGGCGCAGGT	
1155	CGAGGCGGCCAGCTGA	E A A S *
2728	CGAGGCGGCCAGCTGA	

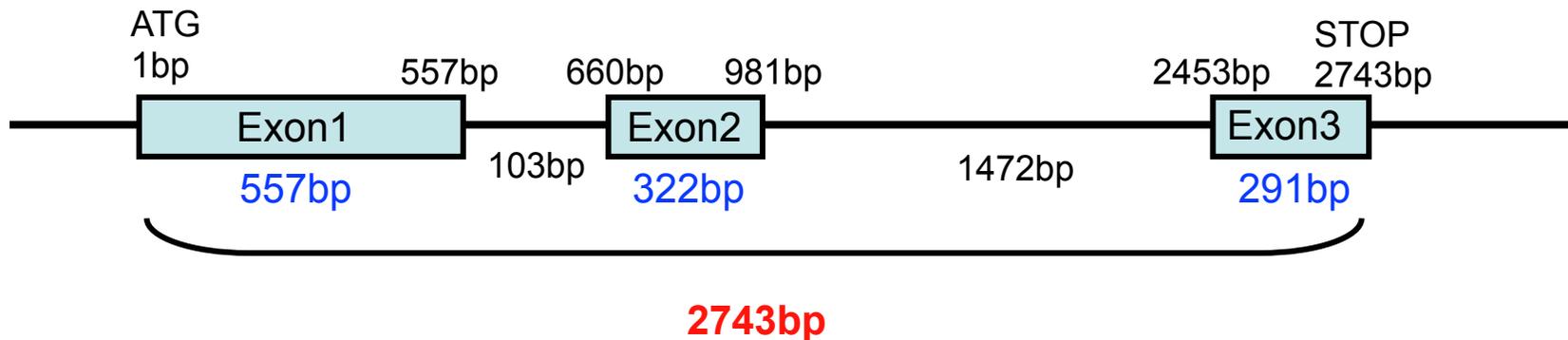
#	Query	Subject	Span(bp)	Coverage(%)	Overall(%)	Exon(%)	CDS(%)	In-frame(%)
1	A_cDNA(+)	A_genomic_DNA_ATG(+)	1-2743	100.00	100.00	100.00	0.00	0.00

[Graphics](#) | [Text](#)

#	Query	Subject	Idty	Len	Q.Start	Q.Fin	S.Start	S.Fin	Type	Details
+1	A_cDNA	A_genomic_DNA_ATG	1	557	1	557	1	557	<exon>GT	M557
+1	A_cDNA	A_genomic_DNA_ATG	1	322	558	879	660	981	AG<exon>GT	M322
+1	A_cDNA	A_genomic_DNA_ATG	1	291	880	1170	2453	2743	AG<exon>	M291

**Help:** for questions, comments, or bug reporting, please visit [NCBI Support Center](#)

## 遺伝子構造



# 生物情報工学II (担当 芦莉) 遺伝子の配列解析

第6回目-1 (11/1) [遺伝子の予測と機能解析](#)

-2 (11/1) [演習](#)

第7回目-1 (11/8) [DNA配列から遺伝子を探す](#)

-2 (11/8) [DNA、アミノ酸のアラインメント、系統樹解析](#)

-3 (11/8) [演習](#)

第8回目-1 (11/15) [Genetyxを用いた解析](#)

-2 (11/15) [演習](#)

第9回目-1 (11/22) [KEGGを用いた解析](#)

[連鎖解析](#)

出席と宿題の提出 [提出方法](#)

## 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](#) : 何でも検索
4. [Google Scholar](#) : 科学に関する検索
5. [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. [Splign](#) : DNAとcDNAのマッチング
2. [Blast align](#) : DNAとcDNAのマッチング
3. [EMBOSS program"](#) : DNAとcDNAのマッチング 
4. [EMBOSS program"](#) : DNA配列をアミノ酸配列に変換
5. [BLAST\(NCBI\)](#) : 遺伝子のホモロジー検索
6. [BLAST\(DDBJ\)](#) : 遺伝子のホモロジー検索
7. [Softberry/FGENESH](#) : 遺伝子の予測
8. [GeneMark](#) : 遺伝子の予測
9. [GENESCAN](#) : 遺伝子の予測
10. [clustalW](#) : アライメント
11. [InterProScan](#) : ドメイン解析
12. [PSORT](#) : 細胞内局在性予測
13. [GEO](#) : 遺伝子発現解析
14. [Gene Ontology](#) : 用語を用いた検索
15. [SSPN](#) : スプライシングの予測

# EMBOSS Programs

[Feedback](#)[Share](#)

Tools > EMBOSS Programs

## EMBL-EBI to be HTTPS by default from 1st October

On the 1st October the majority of services hosted on [www.ebi.ac.uk](http://www.ebi.ac.uk) will be served over HTTPS by default. Services that are becoming HTTPS by default will automatically redirect users accessing the site on insecure HTTP URLs to secure HTTPS URLs.

Users of EMBL-EBI services may wish to update links, bookmarks or API clients to use the HTTPS URLs.

## Selected EMBOSS tools for sequence analysis

### Pairwise Sequence Alignment

#### Needle

Create an optimal global alignment of two sequences using the Needleman-Wunsch algorithm

[Protein](#) [Nucleotide](#)

#### Stretcher

Improved version of the Needleman-Wunsch algorithm that allows larger sequences to be globally aligned

[Protein](#) [Nucleotide](#)

#### Water

Use the Smith-Waterman algorithm to calculate the local alignment of two sequences

[Protein](#) [Nucleotide](#)

#### Matcher

Identify local similarities between two sequences using a rigorous algorithm based on the LALIGN application

[Protein](#) [Nucleotide](#)

### Sequence Translation

#### Transeq

Translate nucleic acid sequences to the corresponding peptide sequences

[Launch Transeq](#)

#### Sixpack

Display DNA sequences with 6-frame translation and ORFs

[Launch Sixpack](#)





# EMBOSS Needle

Protein alignment

Nucleotide alignment

Web services

Help & Documentation

Also in this section

Feedback

Share

Tools > Pairwise Sequence Alignment > EMBOSS Needle

## Results for job emboss\_needle-I20181031-080452-0441-43335280-p2m

Alignment

Submission Details

View Alignment File

```
#####
# Program: needle
# Rundate: Wed 31 Oct 2018 08:04:55
# Commandline: needle
# -auto
# -stdout
# -asequence emboss_needle-I20181031-080452-0441-43335280-p2m.asequence
# -bsequence emboss_needle-I20181031-080452-0441-43335280-p2m.bsequence
# -datafile EDNAFULL
# -gapopen 10.0
# -gapextend 0.5
# -endopen 10.0
# -endextend 0.5
# -aformat3 pair
# -snucleotide1
# -snucleotide2
# Align_format: pair
# Report_file: stdout
#####
```

```
#=====
#
# Aligned_sequences: 2
# 1: EMBOSS_001
# 2: EMBOSS_001
# Matrix: EDNAFULL
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 2770
# Identity: 1170/2770 (42.2%)
# Similarity: 1170/2770 (42.2%)
# Gaps: 1600/2770 (57.8%)
# Score: 5044.5
#
#
#=====
```

```
EMBOSS_001 1 atggtggccgagcaccacgcccaccacagccgaccacaccaccgccc 50
```

EMBOSS_001	1	atggtggccgagcaccgccaccacagccgaccaaccaccgccc	50
<b>EMBOSS Needle - Alignment</b>			
EMBOSS_001	1	atggtggccgagcaccgccaccacagccgaccaaccaccgccc	50
EMBOSS_001	51	ggactccaccgccggtcttgccattgccgccccggcgggcggtgt	100
EMBOSS_001	51	ggactccaccgccggtcttgccattgccgccccggcgggcggtgt	100
EMBOSS_001	101	gcgacctgaggatggagccaagatcccgagccattcgtgtgcccgaac	150
EMBOSS_001	101	gcgacctgaggatggagccaagatcccgagccattcgtgtgcccgaac	150
EMBOSS_001	151	ggcgacgcgagccggcgtcggcggggagctggacatgcccgtggtcga	200
EMBOSS_001	151	ggcgacgcgagccggcgtcggcggggagctggacatgcccgtggtcga	200
EMBOSS_001	201	cgtggcggtgctcccgacggcgacgccgaggggctgcgccgcgccg	250
EMBOSS_001	201	cgtggcggtgctcccgacggcgacgccgaggggctgcgccgcgccg	250
EMBOSS_001	251	cgaggtggcccccgtgcgccacgcacgggttcttccaggtgtccgag	300
EMBOSS_001	251	cgaggtggcccccgtgcgccacgcacgggttcttccaggtgtccgag	300
EMBOSS_001	301	cacggcgtcgacgccgctctggcgccgcccgcctcgacggcggcagcga	350
EMBOSS_001	301	cacggcgtcgacgccgctctggcgccgcccgcctcgacggcggcagcga	350
EMBOSS_001	351	cttctccgcctcccgcctgcgagagaagcggcgccgcccgtcccgg	400
EMBOSS_001	351	cttctccgcctcccgcctgcgagagaagcggcgccgcccgtcccgg	400
EMBOSS_001	401	gcaccgtgtccggctacaccagcggcggcggcggcggcggcggcggc	450
EMBOSS_001	401	gcaccgtgtccggctacaccagcggcggcggcggcggcggcggcggc	450
EMBOSS_001	451	ctcccatggaaggagaccctctccttcggcttccacgaccgcccggc	500
EMBOSS_001	451	ctcccatggaaggagaccctctccttcggcttccacgaccgcccggc	500
EMBOSS_001	501	cccggttcctccggactacttctccagcaccctcggccccgacttcg	550
EMBOSS_001	501	cccggttcctccggactacttctccagcaccctcggccccgacttcg	550
EMBOSS_001	551	caatgggtaattaaaacgatggtggacgacattgcattcaaatcaaa	600
EMBOSS_001	551	caatgg-----	556
EMBOSS_001	601	acaaattcaaaacacaccgaccgagattatgctgaattcaaacgcgttg	650
EMBOSS_001	557	-----	556
EMBOSS_001	651	tgcgcccaggagggtgtaccagaagtactgaggagatgaaggagctgt	700
EMBOSS_001	557	-----ggagggtgtaccagaagtactgaggagatgaaggagctgt	598
EMBOSS_001	701	cgctgacgatcatggaactcctggagctgagcctggcggtggagcgaggc	750
EMBOSS_001	599	cgctgacgatcatggaactcctggagctgagcctggcggtggagcgaggc	648
EMBOSS_001	751	tactacaggagtcttctcgggacagcagctcaatcatgcggtgcaacta	800
EMBOSS_001	649	tactacaggagtcttctcgggacagcagctcaatcatgcggtgcaacta	698

# エキソンとイントロンの境界

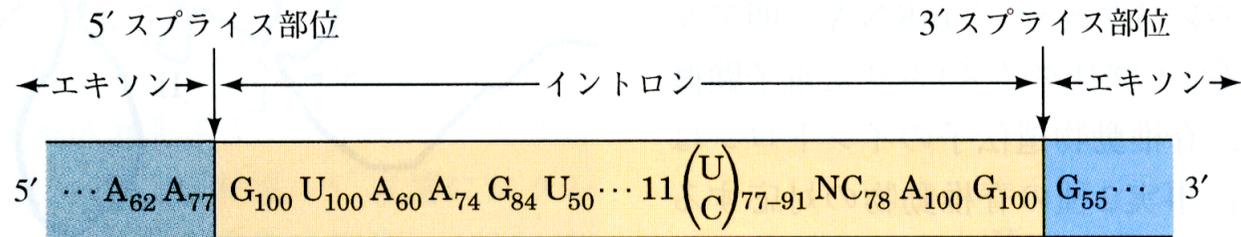


図 25・19 真核細胞 mRNA 前駆体のエキソン・イントロン接合部にあるコンセンサス配列 各塩基の添え数字はその塩基の出現頻度(%). 3' スプライス部位のすぐ上流にピリミジンに富む 11 nt 区域がある [R.A. Padgett, P.J. Grabowski, M.M. Konarska, S.S. Seiler, P.A. Sharp, *Annu. Rev. Biochem.* **55**, 1123 (1986) のデータに基づく]

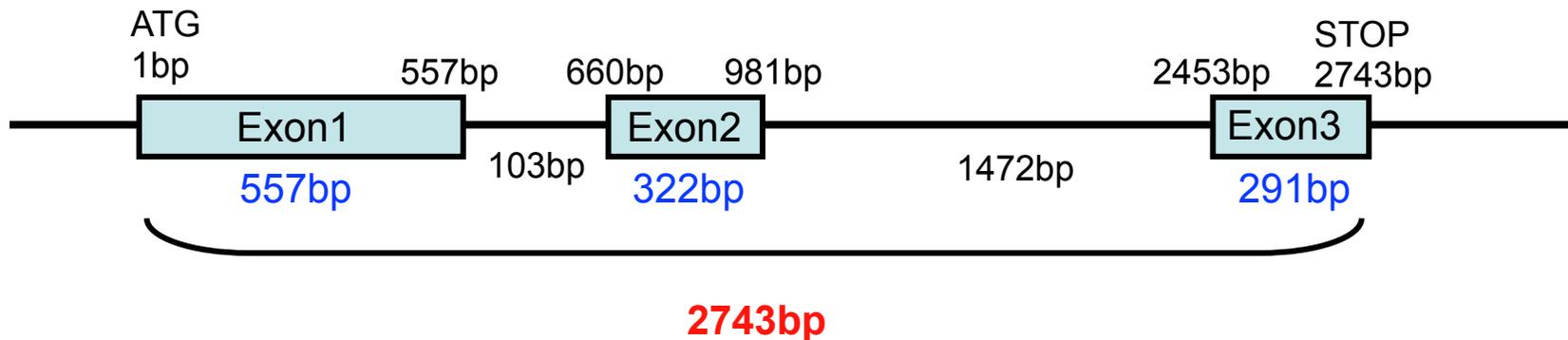
#	Query	Subject	Span(bp)	Coverage(%)	Overall(%)	Exon(%)	CDS(%)	In-frame(%)
1	A_cDNA(+)	A_genomic_DNA_ATG(+)	1-2743	100.00	100.00	100.00	0.00	0.00

[Graphics](#) | [Text](#)

#	Query	Subject	Idty	Len	Q.Start	Q.Fin	S.Start	S.Fin	Type	Details
+1	A_cDNA	A_genomic_DNA_ATG	1	557	1	557	1	557	<exon>GT	M557
+1	A_cDNA	A_genomic_DNA_ATG	1	322	558	879	660	981	AG<exon>GT	M322
+1	A_cDNA	A_genomic_DNA_ATG	1	291	880	1170	2453	2743	AG<exon>	M291

**Help:** for questions, comments, or bug reporting, please visit [NCBI Support Center](#)

## 遺伝子構造



# 生物情報工学II (担当 芦莉) 遺伝子の配列解析

第6回目-1 (11/1) [遺伝子の予測と機能解析](#)

-2 (11/1) [演習](#)

第7回目-1 (11/8) [DNA配列から遺伝子を探す](#)

-2 (11/8) [DNA、アミノ酸のアラインメント、系統樹解析](#)

-3 (11/8) [演習](#)

第8回目-1 (11/15) [Genetyxを用いた解析](#)

-2 (11/15) [演習](#)

第9回目-1 (11/22) [KEGGを用いた解析](#)

[連鎖解析](#)

出席と宿題の提出 [提出方法](#)

## 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](#) : 何でも検索
4. [Google Scholar](#) : 科学に関する検索
5. [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. [Salign](#) : DNAとcDNAのマッチング
2. [Blast align](#) : DNAとcDNAのマッチング
3. [EMBOSS program](#)" : DNAとcDNAのマッチング
4. [EMBOSS program"](#) : DNA配列をアミノ酸配列に変換
5. [BLAST\(NCBI\)](#) : 遺伝子のホモロジー検索
6. [BLAST\(DDBJ\)](#) : 遺伝子のホモロジー検索
7. [Softberry/FGENESH](#) : 遺伝子の予測
8. [GeneMark](#) : 遺伝子の予測
9. [GENESCAN](#) : 遺伝子の予測
10. [clustalW](#) : アライメント
11. [InterProScan](#) : ドメイン解析
12. [PSORT](#) : 細胞内局在性予測
13. [GEO](#) : 遺伝子発現解析
14. [Gene Ontology](#) : 用語を用いた検索
15. [SSPN](#) : スプライシングの予測

エクソン-イントロンを正確に知りたいとき  
ただ配列比較をしたい場合

# 生物情報工学II (担当 芦苺) 遺伝子の配列解析

第6回目-1 (11/1) [遺伝子の予測と機能解析](#)

-2 (11/1) [演習](#)

第7回目-1 (11/8) [DNA配列から遺伝子を探す](#)

-2 (11/8) [DNA、アミノ酸のアラインメント、系統樹解析](#)

-3 (11/8) [演習](#)

第8回目-1 (11/15) [Genetyxを用いた解析](#)

-2 (11/15) [演習](#)

第9回目-1 (11/22) [KEGGを用いた解析](#)

[連鎖解析](#)

出席と宿題の提出 [提出方法](#)

## 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](#) : 何でも検索
4. [Google Scholar](#) : 科学に関する検索
5. [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. [Splign](#) : DNAとcDNAのマッチング
2. [Blast align](#) : DNAとcDNAのマッチング
3. [EMBOSS program](#) : DNAとcDNAのマッチング
4. [EMBOSS program](#) : DNA配列をアミノ酸配列に変換
5. [BLAST\(NCBI\)](#) : 遺伝子のホモロジー検索
6. [BLAST\(DDBJ\)](#) : 遺伝子のホモロジー検索
7. [Softberry/FGENESH](#) : 遺伝子の予測
8. [GeneMark](#) : 遺伝子の予測
9. [GENESCAN](#) : 遺伝子の予測
10. [clustalW](#) : アラインメント
11. [InterProScan](#) : ドメイン解析
12. [PSORT](#) : 細胞内局在性予測

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

# EMBOSS Programs

[Feedback](#)[Share](#)

Tools &gt; EMBOSS Programs

## EMBL-EBI to be HTTPS by default from 1st October

On the 1st October the majority of services hosted on [www.ebi.ac.uk](http://www.ebi.ac.uk) will be served over HTTPS by default. Services that are becoming HTTPS by default will automatically redirect users accessing the site on insecure HTTP URLs to secure HTTPS URLs.

Users of EMBL-EBI services may wish to update links, bookmarks or API clients to use the HTTPS URLs.

## Selected EMBOSS tools for sequence analysis

### Pairwise Sequence Alignment

#### Needle

Create an optimal global alignment of two sequences using the Needleman-Wunsch algorithm

[Protein](#) [Nucleotide](#)

#### Stretcher

Improved version of the Needleman-Wunsch algorithm that allows larger sequences to be globally aligned

[Protein](#) [Nucleotide](#)

#### Water

Use the Smith-Waterman algorithm to calculate the local alignment of two sequences

[Protein](#) [Nucleotide](#)

#### Matcher

Identify local similarities between two sequences using a rigorous algorithm based on the LALIGN application

[Protein](#) [Nucleotide](#)

### Sequence Translation

#### Transeq

Translate nucleic acid sequences to the corresponding peptide sequences

[Launch Transeq](#)

#### Sixpack

Display DNA sequences with 6-frame translation and ORFs

[Launch Sixpack](#)

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

配列解析：

1. [Splign](#) : DNAとcDNAのマッチング
2. [Blast align](#) : DNAとcDNAのマッチング
3. [EMBOSS program](#)" : DNAとcDNAのマッチング
4. [EMBOSS program](#)" : DNA配列をアミノ酸配列に変換
5. [BLAST\(NCBI\)](#) : 遺伝子のホモロジー検索
6. [BLAST\(DDBJ\)](#) : 遺伝子のホモロジー検索
7. [Softberry/FGENESH](#) : 遺伝子の予測
8. [GeneMark](#) : 遺伝子の予測
9. [GENESCAN](#) : 遺伝子の予測
10. [clustalW](#) : アライメント
11. [InterProScan](#) : ドメイン解析
12. [PSORT](#) : 細胞内局在性予測
13. [GEO](#) : 遺伝子発現解析
14. [Gene Ontology](#) : 用語を用いた検索
15. [SSPN](#) : スプライシングの予測
16. [ORF Finder](#) : ORFの予測
17. [TFSEARCH](#) : 転写因子結合部位の予測
18. [MOTIF](#) : タンパク質モチーフ検索
19. [pI/Mw](#) : 等電点、分子量の予測
20. [RAP-DB](#) : イネのゲノムデータベース

解析用配列

1. [ある遺伝子AのgenomeDNA配列](#)
2. [ある遺伝子AのcDNA配列](#)
3. [ある遺伝子XのgenomeDNA配列](#)
4. [ある遺伝子XのcDNA配列](#)
5. [ある遺伝子YのgenomicDNA](#)
6. [ある遺伝子YのcDNA](#)
7. [Z1遺伝子のgenomicDNA](#)
8. [Z1遺伝子のCDS](#)
9. [Z1遺伝子の予想アミノ酸配列](#)
10. [Z2遺伝子のgenomicDNA](#)
11. [Z2遺伝子のCDS](#)
12. [Z2遺伝子のアミノ酸配列](#)
13. [野生型SD1-genomicDNA配列](#)
14. [野生型SD1-cDNA配列](#)
15. [カルロース-cDNA](#)



## EMBL-EBI to be HTTPS by default from 1st October

On the 1st October the majority of services hosted on www.ebi.ac.uk will be served over HTTPS by default. Services that are becoming HTTPS by default will automatically redirect users accessing the site on insecure HTTP URLs to secure HTTPS URLs.

Users of EMBL-EBI services may wish to update links, bookmarks or API clients to use the HTTPS URLs.

# EMBOSS Transeq

EMBOSS Transeq translates nucleic acid sequences to their corresponding peptide sequences. It can translate to the three forward and three reverse frames, and output multiple frame translations at once.

### STEP 1 - Enter your input sequence

Enter or paste a DNA/RNA sequence in any supported format:

```
atggtggccgagcaccaccacgcccacacagccgcccaccaaccaccgcccattgactccaccgcccgtctggcattgccgccccggcggcggcggtgtgacacagctgagctgagcccaagatcccggagc  
cattcgtgtggcgaacggcgagcgagggcggcgctggcggggagctggacatgccggtgtgacgtggcgctcgcgacggcgagccgaggggctcgcgcgccgcccggcgagtggtgccgccc  
cgtgcgccacgcacgggttctccaggtgtccgagcacggcgtcgcgcccgtctggcgccgcccgcgctcgcgagccgcccagcacttctccgctcccgctcggcgagaagcgccgccccgctcccg  
gcaccgtgtccggtacaccagcggccacgcccgcacgctcgcctccaagctccatggaaggagacacctctccttccgcttccacgaccgcccgcggcccccgtcgtcggcactacttccagaccctcggc  
cccgactcgcgccaatggggagggtgtaccagaagtactgcgaggagatgaaggagctgtcgtgacgatggaactcctggagctgagcctggcggtggagcgaggctactacagggagtcttcggcgacag  
cagctcaatcatcggtgcaactactaccgcatgcccggagcggagcggagcgtcggcacgggccgcactgcgaccccaccgcccctcctccaggacgacgtcggcgccctcgaggtcctcgt  
cgacggcgaaatggcggcccgcagccccgtccccggcgcctggtcatcaacatcggcgacacctcatggcgtgtcgaacgggaggtataagagctgcctgcacagggcggtggtaaccagcggcggggagc
```

Or, upload a file: ファイルを選択 ファイル未選択

### STEP 2 - Select Parameters

FRAME

CODON TABLE

F (Forward three frames)

Standard Code

The default settings will fulfill the needs of most users.

More options... (Click here, if you want to view or change the default settings.)

### STEP 3 - Submit your job

Be notified by email (Tick this box if you want to be notified by email when the results are available)

Submit

If you plan to use these services during a course please [contact us](#).

# EMBOSS Transeq

[Input form](#)[Web services](#)[Help & Documentation](#)[Feedback](#)[Share](#)[Tools](#) > [Sequence Translation](#) > [EMBOSS Transeq](#)

## EMBL-EBI to be HTTPS by default from 1st October

On the 1st October the majority of services hosted on [www.ebi.ac.uk](http://www.ebi.ac.uk) will be served over HTTPS by default. Services that are becoming HTTPS by default will automatically redirect users accessing the site on insecure HTTP URLs to secure HTTPS URLs.

Users of EMBL-EBI services may wish to update links, bookmarks or API clients to use the HTTPS URLs.

## Results for job [emboss\\_transeq-l20170929-090805-0838-23778789-pg](#)

[Tool Output](#)[Submission Details](#)[Download](#)[Show Colors](#)

```
>EMBOSS_001_1
MVAEHPTFPQPHQPPMDSTAGSGIAAPAAAACVCDLRMEPKIPEPFVWPNGDARPASAAE
LDMPVVDVGVLRDGDAGLRRAAAQVAAACATHGFFQVSEHGVDAAALARAALDGASDFFR
LPLAEKRRRARRVPGTVSGYTSAHADRFASKLPWKETLSFGFHDRAAAPVVADYFSSTLGP
DFAPMGRVYQKYCEEMKELSLTIMELLELSLGVVERGYREFFADSSSIMRCNYPPCPEP
ERTLGTGPHCDPTALTILLQDDVGGLEVLVDGEWRPVSPVPGAMVINIGDTFMALSNTRY
KSCLEHRAVVNQRRERRSLAFFLCPREDRVVPPPSAATPQHYPDFTWADLMRFTQRHYRA
DTRTLDAFTRWLAPPAADAAATAQVEAAS*
>EMBOSS_001_2
WFPSTFRHHSRTNHRPWTPPALALPPRRRRRCAT*GWSPRSRSWSCGRTATRGRRRRRS
WTCFWSWACSAATATPRGCAAPRRRWPPRAPRTGSSRCPSTASTPLWRAPRSTAPATSSA
SRSPRSARAASRAPCPATPAPTPTASPPSSHGRRPSPASSTAPPFPSSPTTSPAPSAP
TSRQGGCTRSTARR*RSR*RSWNSW*AWASEATTGSSRTAAQSCGATTRHARSR
SGRSARARTATPPSPSSRTTSAASRSSSTANGAPSPAPWSSSATPSWRCRTGGI
RAACTGRW*TSGGSGGRWRSSCARGRTGWCGRRRAPRRSTTRTSPGPTSCASRSATTAP
TPARSTPSRAGSRRRPPTPRRRRSRRPAX
>EMBOSS_001_3
CGRAPHATTAAPTAAHGLHRLWHCRPGGGGVREDGAQDPGAIRVAERRREAGVGGGA
GHARGRRGAPRRRRRGAAPRRGAGRRVRHARVLPVGRARRRRRSGARRARRRQLLPP
PARREAPRAPRPHRVRHLHQRPRRRLRLQAPMEGDPDLLRPLRRRRRRLQLHPRPR
LRANGEGVEVLRDEGAVADDHCTPGAEPGRGARLLQCVLRGQLNHAVQLLPAMPAGG
ADARHGAPALRPHRPHPPGRRRRRPRGRRRRMAPRQPRRRRHGHQRRRHLHGAVEREV*
ELPAQGGGEPAAAGAAVAGVLPVAGGGGAAAAERRHAALPGLHLGRPHALHAAALPRR
HPHARRLHALARAAGRRRRRDGAGRGQLX
```

コピー

>EMBOSS\_001\_1

MVAEHPTPPQPHQPPMDSTAGSGIAAPAAAACDLRMEPKIPEPFVWPNGDARPASAAE  
LDMPVVDVGVLDRDGAEGLRRAAAQVAAACATHGFFQVSEHGVDAAALARAALDGASDFFR  
LPLAEKRRARRVPGTVSGYTSAHADRFASKLPWKETLSFGFHDRAAAPVVADYFSSSTLGP  
DFAPMGRVYQKYCEEMKELSLTIMELLELSLGVVERGYREFFADSSSIMRCNYYPPCPEP  
ERTLGTGPHCDPTALTILLQDDVGGLEVLVDGEWRPVSPVPGAMVINIGDTFMALSNGRY  
KSCLHRAVVNQRRERRSLAFFLCPREDRVVRPPPSAATPQHYPDFTWADLMRFTQRHYRA  
DTRTLDAFTRWLAPPAADAAATAQVEAAS\*

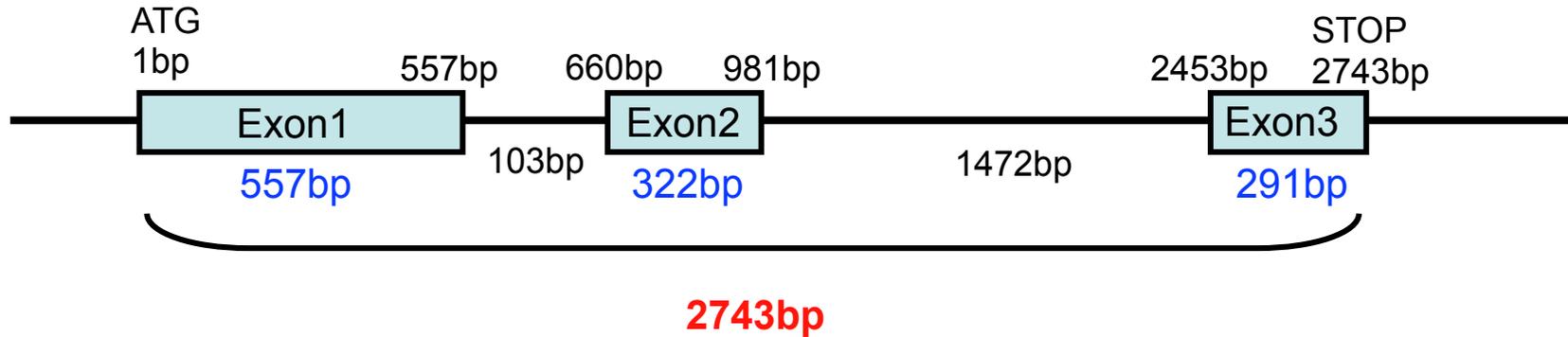
>EMBOSS\_001\_2

WWPSTPRHHSRTNHRPWTPPPALALPPRRRRRCAT\*GWSPRSRSRSHSCGRTATRGRRRRRS  
WTCPWSTWACSATATPRGCAAPRRRWPPRAPRTGSSRCPSTASTPLWRAPRSTAPATSSA  
SRSPRSAARAASRAPCPATPAPTPTASPPSSHGRRPSPSASTTAPPPSSPTTSPAPSAP  
TSRQWGGCTRSTAR\*RSCR\*RSWNSW\*AWAWSEATTGSSSRATAQSCGATTTRHARSR  
SGRSARARTATPPSPSSSRTTSAASRSSSTANGAPSAPSPAPWSSTSATPSWRCRTGGI  
RAACTGRW\*TSGGSGGRWRSSCARGRTGWCGRRRAPRRSTTRTSPGPTSCASRSATTAP  
TPARSTPSRAGSRRRPPTPPRRRRSRRPAX

>EMBOSS\_001\_3

GGRAPHATTAAPTTHAHLHRLWHCRPGGGGGVVRPEDGAQDPGAI RVAERRREAGVGGGA  
GHARGRRGRAPRRRRRGAAPRRGAGGRRVRHARVLPGVRARRRRRSGARRARRRQRLLP  
PARREAPRAPRPGHRVRLHQRRRPLRLQAPMEGDPLLRLPRRRRPRRRRLLQHPRPR  
LRANGEGVPEVLRGDEGAVADDHGTPGAEPGRGARLLQGVLRGQQLNHAVQLLPAMPAG  
ADARHGPAALRPHRPHHPPPGRRRRRPRGPRRRRMAPRQPRRRRHGHQHRRLHGAVEREV\*  
ELPAQGGGEPAAAGAAVAGVLPVPAGGQGGAAAERRHAAALPGLHLGRPHALHAAPLPRR  
HPHARRLHALARAAGRRRRRDGAGRGGQLX

## 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 322bp, 第3エクソン 291bp)、  
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

```
MVAEHPTPPQPHQPPMDSTAGSGIAAPAAA VCDLRMEPKIPEPFVWPNGDARPASAAE  
LDMPVVDVGVLRDGAEGLRRAAAQVAAACATHGFFQVSEHGVDAAALARAALDGASDFFR  
LPLAEKRRARRVPGTVSGY TSAHADRFASKLPWKETLSFGFHDRAAAPVVADYFSSTLGP  
DFAPMGRVYQKYCEEMKELSLTIMELLELSLGVERGYREFFADSSIMRCNYYPPCPEP  
ERTLGTGPHCDPTALTILLQDDVGGLEVLVDGEWRPVSPVPGAMVINIGDTFMALSNGRY  
KSCLHRAVNVQRRERRSLAFFLCPREDRVVRPPPSAATPQHYPDFTWADLMRFTQRHYRA  
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

# 生物情報工学II (担当 芦莉) 遺伝子の配列解析

第6回目-1 (11/1) [遺伝子の予測と機能解析](#)

-2 (11/1) [演習](#)

第7回目-1 (11/8) [DNA配列から遺伝子を探す](#)

-2 (11/8) [DNA、アミノ酸のアラインメント、系統樹解析](#)

-3 (11/8) [演習](#)

第8回目-1 (11/15) [Genetyxを用いた解析](#)

-2 (11/15) [演習](#)

第9回目-1 (11/22) [KEGGを用いた解析](#)

[連鎖解析](#)

出席と宿題の提出 [提出方法](#)

---

## 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](#): 何でも検索
4. [Google Scholar](#): 科学に関する検索
5. [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. [Splign](#) : DNAとcDNAのマッチング
2. [Blast align](#) : DNAとcDNAのマッチング
3. [EMBOSS program"](#) : DNAとcDNAのマッチング
4. [EMBOSS program"](#) : DNA配列をアミノ酸配列に変換
5. [BLAST\(NCBI\)](#) : 遺伝子のホモロジー検索 
6. [BLAST\(DDBJ\)](#) : 遺伝子のホモロジー検索
7. [Softberry/FGENESH](#) : 遺伝子の予測
8. [GeneMark](#) : 遺伝子の予測
9. [GENESCAN](#) : 遺伝子の予測
10. [clustalW](#) : アライメント
11. [InterProScan](#) : ドメイン解析
12. [PSORT](#) : 細胞内局在性予測

## Basic Local Alignment Search Tool

**BLAST** finds regions of similarity between biological sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance. [Learn more](#)

NEWS

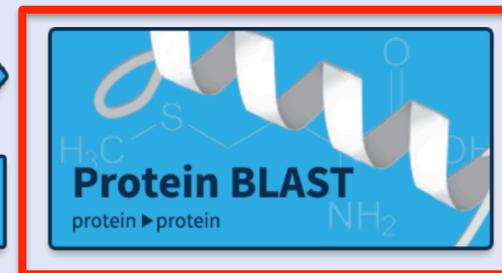
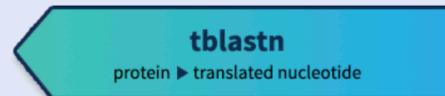
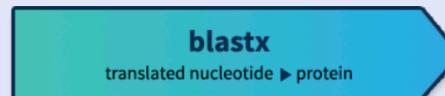
### QuickBLASTP

Try **QuickBLASTP** for a fast protein search of nr.

Tue, 23 May 2017 13:00:00 EST

[More BLAST news...](#)

## Web BLAST



## BLAST Genomes

Enter organism common name, scientific name, or tax id.

[Human](#)[Mouse](#)[Rat](#)[Microbes](#)

## Standalone and API BLAST



### Download BLAST

Get BLAST databases and executables



### Use BLAST API

Call BLAST from your application



### Use BLAST in the cloud

Start an instance at a cloud provider

## Specialized searches

### SmartBLAST



Find proteins highly similar to your query

### Primer-BLAST



Design primers specific to your PCR template

### Global Align



Compare two sequences across their entire span

### CD-search



Find conserved domains in your sequence

## Standard Protein BLAST

blastn **blastp** blastx tblastn tblastxBLASTP programs search protein databases using a protein query. [more...](#)[Reset page](#) [Bookmark](#)

## Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) [Clear](#)Query subrange [?](#)From To 

Or, upload file

ファイルを選択 ファイル未選択 [?](#)

Job Title

  
Enter a descriptive title for your BLAST search [?](#) Align two or more sequences [?](#)

## Choose Search Set

Database

Non-redundant protein sequences (nr) [?](#)

Organism

Optional

 Enter organism name or id--completions will be suggested Exclude [+](#)Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown. [?](#)

Exclude

Optional

 Models (XM/XP)  Uncultured/environmental sample sequences

Entrez Query

Optional

 [YouTube](#) [Create custom database](#)Enter an Entrez query to limit search [?](#)

## Program Selection

Algorithm

 Quick BLASTP (Accelerated protein-protein BLAST) **New** blastp (protein-protein BLAST) PSI-BLAST (Position-Specific Iterated BLAST) PHI-BLAST (Pattern Hit Initiated BLAST) DELTA-BLAST (Domain Enhanced Lookup Time Accelerated BLAST)Choose a BLAST algorithm [?](#)**BLAST**Search database **Non-redundant protein sequences (nr)** using **Blastp (protein-protein BLAST)** Show results in a new window[+ Algorithm parameters](#)

BLAST is a registered trademark of the National Library of Medicine

[Support center](#) [Mailing list](#) [YouTube](#)

NCBI

National Center for Biotechnology Information, U.S. National Library of Medicine  
8600 Rockville Pike, Bethesda MD, 20894 USA[Policies and Guidelines](#) | [Contact](#)

**Standard Protein BLAST**[blastn](#) **[blastp](#)** [blastx](#) [tblastn](#) [tblastx](#)BLASTP programs search protein databases using a protein query. [more...](#)[Reset page](#) [Bookmark](#)**Enter Query Sequence**Enter accession number(s), gi(s), or FASTA sequence(s) [Clear](#)Query subrange 

```
>Aminoacid
MVAEHPTPPQPHQPPMDSTAGSGIAAPAAAAVCDLRMEPKIPEPFVWPNGDARPASAAELDM
PVVDVGV
LRDGDAGLRRAAQVAAACATHGFFQVSEHGVDAAALARAALDGASDFRFLPLAEKRRARRVP
GTVSGYT
```

From To 

Or, upload file

Job Title

Enter a descriptive title for your BLAST search  **Align two or more sequences****Choose Search Set**

Database

Organism

Optional

 **Exclude** 

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

Exclude

Optional

 Models (XM/XP)  Uncultured/environmental sample sequences

Entrez Query

Optional

[YouTube](#) [Create custom database](#)**Program Selection**

Algorithm

 Quick BLASTP (Accelerated protein-protein BLAST) **New** **blastp** (protein-protein BLAST) PSI-BLAST (Position-Specific Iterated BLAST) PHI-BLAST (Pattern Hit Initiated BLAST) DELTA-BLAST (Domain Enhanced Lookup Time Accelerated BLAST)Choose a BLAST algorithm **BLAST**Search database **Non-redundant protein sequences (nr)** using **Blastp** (protein-protein BLAST) Show results in a new window[+ Algorithm parameters](#)

BLAST is a registered trademark of the National Library of Medicine

[Support center](#) [Mailing list](#) [YouTube](#)**NCBI**National Center for Biotechnology Information, U.S. National Library of Medicine  
8600 Rockville Pike, Bethesda MD, 20894 USA[Policies and Guidelines](#) | [Contact](#)

## BLAST Results

[Edit and Resubmit](#) [Save Search Strategies](#) [Formatting options](#) [Download](#)[YouTube](#) [How to read this page](#) [Blast report description](#)

## Job title: A aminoacid

RID [WW1SFB77015](#) (Expires on 09-30 16:40 pm)

**Query ID** |cl|Query\_230675  
**Description** Aminoacid  
**Molecule type** amino acid  
**Query Length** 389

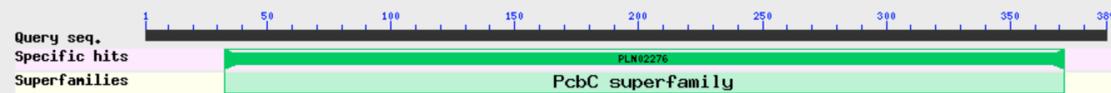
**Database Name** nr  
**Description** All non-redundant GenBank CDS translations+PDB+SwissProt+PIR+PRF excluding environmental samples from WGS projects  
**Program** BLASTP 2.7.0+ [Citation](#)

Other reports: [Search Summary](#) [Taxonomy reports](#) [Distance tree of results](#) [Multiple alignment](#) [MSA viewer](#)**New** Analyze your query with [SmartBLAST](#)

## Graphic Summary

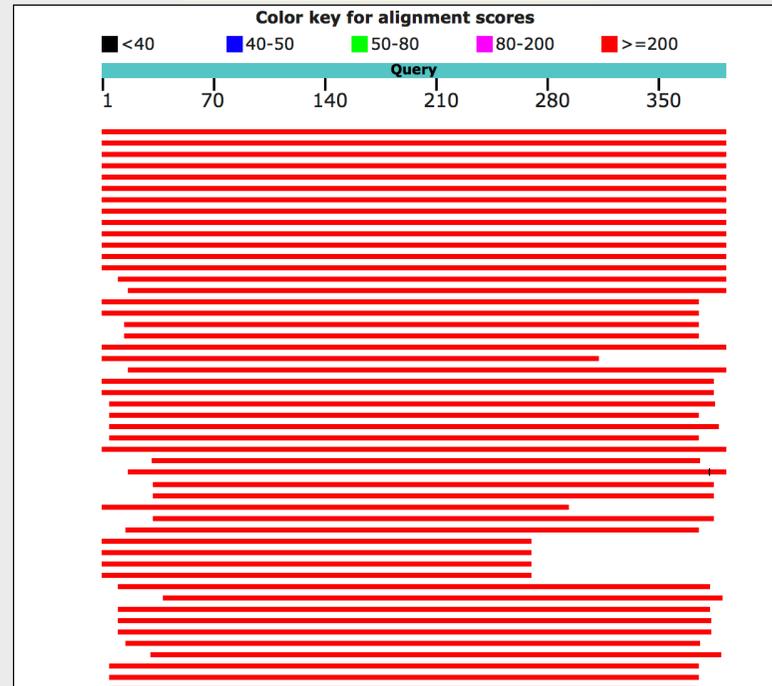
 Show Conserved Domains

Putative conserved domains have been detected, click on the image below for detailed results.



## Distribution of the top 101 Blast Hits on 100 subject sequences

Mouse over to see the title, click to show alignments



Sequences producing significant alignments:

Select: [All](#) [None](#) Selected:0

[Alignments](#) [Download](#) [GenPept](#) [Graphics](#) [Distance tree of results](#) [Multiple alignment](#)

	Description	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/>	PREDICTED: gibberellin 20 oxidase 2 [Oryza sativa Japonica Group]	793	793	100%	0.0	100%	<a href="#">XP_015627721.1</a>
<input type="checkbox"/>	SD1 [Oryza sativa Japonica Group]	790	790	100%	0.0	99%	<a href="#">AER45907.1</a>
<input type="checkbox"/>	putative gibberellin 20 oxidase [Oryza rufipogon]	788	788	100%	0.0	99%	<a href="#">AAN73384.1</a>
<input type="checkbox"/>	SD1 [Oryza rufipogon]	788	788	100%	0.0	99%	<a href="#">AER45865.1</a>
<input type="checkbox"/>	GA C20oxidase2 [Oryza glumipatula]	786	786	100%	0.0	99%	<a href="#">BAG80954.1</a>
<input type="checkbox"/>	gibberellin-20 oxidase-2 [Oryza rufipogon]	785	785	100%	0.0	99%	<a href="#">BAK39011.1</a>
<input type="checkbox"/>	SD1 [Oryza rufipogon]	785	785	100%	0.0	99%	<a href="#">AER45849.1</a>
<input type="checkbox"/>	gibberellin-20 oxidase-2 [Oryza sativa Indica Group]	785	785	100%	0.0	99%	<a href="#">BAL03272.1</a>
<input type="checkbox"/>	GA C20oxidase2 [Oryza longistaminata]	783	783	100%	0.0	98%	<a href="#">BAG80956.1</a>
<input type="checkbox"/>	GA C20oxidase2 [Oryza meridionalis]	778	778	100%	0.0	98%	<a href="#">BAG80955.1</a>
<input type="checkbox"/>	gibberellin-20 oxidase-2 [Oryza sativa Indica Group]	774	774	100%	0.0	98%	<a href="#">BAK39040.1</a>
<input type="checkbox"/>	gibberellin-20 oxidase-2 [Oryza sativa Indica Group]	774	774	100%	0.0	98%	<a href="#">BAK39033.1</a>
<input type="checkbox"/>	gibberellin-20 oxidase-2 [Oryza sativa Indica Group]	773	773	100%	0.0	98%	<a href="#">BAK39038.1</a>
<input type="checkbox"/>	SD1 [Oryza sativa]	767	767	97%	0.0	99%	<a href="#">AER45797.1</a>
<input type="checkbox"/>	hypothetical protein Osl_04688 [Oryza sativa Indica Group]	687	687	95%	0.0	93%	<a href="#">EAY76733.1</a>
<input type="checkbox"/>	GA C20oxidase2 [Oryza punctata]	671	671	95%	0.0	95%	<a href="#">BAG80957.1</a>
<input type="checkbox"/>	GA C20oxidase2 [Oryza alta]	659	659	95%	0.0	89%	<a href="#">BAG80960.1</a>
<input type="checkbox"/>	GA C20oxidase2 [Oryza officinalis]	644	644	92%	0.0	95%	<a href="#">BAG80959.1</a>
<input type="checkbox"/>	GA C20oxidase2 [Oryza minuta]	638	638	92%	0.0	94%	<a href="#">BAG80958.1</a>
<input type="checkbox"/>	GA C20oxidase2 [Oryza brachyantha]	637	637	100%	0.0	88%	<a href="#">BAG80961.1</a>
<input type="checkbox"/>	SD1 [Oryza rufipogon]	631	631	79%	0.0	99%	<a href="#">AER45851.1</a>
<input type="checkbox"/>	GA C20oxidase2 [Oryza meyeriana var. granulata]	626	626	95%	0.0	87%	<a href="#">BAG80962.1</a>
<input type="checkbox"/>	gibberellin 20 oxidase 2 isoform X1 [Sorghum bicolor]	615	615	97%	0.0	83%	<a href="#">XP_002456751.1</a>
<input type="checkbox"/>	hypothetical protein SORBI_3003G379500 [Sorghum bicolor]	614	614	97%	0.0	83%	<a href="#">OQU87995.1</a>
<input type="checkbox"/>	uncharacterized protein LOC107521947 [Zea mays]	609	609	96%	0.0	85%	<a href="#">NP_001308615.1</a>
<input type="checkbox"/>	gibberellin 20 oxidase 2-like [Aegilops tauschii subsp. tauschii]	607	607	94%	0.0	83%	<a href="#">XP_020187952.1</a>
<input type="checkbox"/>	unnamed protein product [Triticum aestivum]	605	605	97%	0.0	81%	<a href="#">CDM85079.1</a>
<input type="checkbox"/>	GA20ox-D2 [Triticum aestivum]	604	604	94%	0.0	83%	<a href="#">CFV04355.1</a>
<input type="checkbox"/>	PREDICTED: gibberellin 20 oxidase 2-like [Oryza brachyantha]	603	603	100%	0.0	80%	<a href="#">XP_015690591.1</a>
<input type="checkbox"/>	PREDICTED: gibberellin 20 oxidase 2-like [Setaria italica]	602	602	87%	0.0	88%	<a href="#">XP_004970813.1</a>

## SD1 [Oryza sativa Japonica Group]

Sequence ID: [AER45907.1](#) Length: 389 Number of Matches: 1

Range 1: 1 to 389 GenPept Graphics Next Match Previous Match

Score	Expect	Method	Identities	Positives	Gaps
790 bits(2041)	0.0	Compositional matrix adjust.	388/389(99%)	388/389(99%)	0/389(0%)
Query 1	MVAEHP <sup>T</sup> PPQHP <sup>Q</sup> PPMDSTAGSGIAAPAAAACV <sup>D</sup> LRMEPKI <sup>E</sup> PEFFVW <sup>F</sup> PNGDARPASAAE				60
Sbjct 1	MVAEHP <sup>T</sup> PPQHP <sup>Q</sup> PPMDSTAGSGIAAPAAAACV <sup>D</sup> LRMEPKI <sup>E</sup> PEFFVW <sup>F</sup> PNGDARPASAAE				60
Query 61	LDMPVVDV <sup>G</sup> VL <sup>R</sup> DGDAEGLRRAAAQVAAACATHG <sup>F</sup> FQVSEHGVDAA <sup>L</sup> ARAALDGASDFFR				120
Sbjct 61	LDMPVVDV <sup>G</sup> VL <sup>R</sup> DGDAEGLRRAAAQVAAACATHG <sup>F</sup> FQVSEHGVDAA <sup>L</sup> ARAALDGASDFFR				120
Query 121	LPLAEKRRAR <sup>R</sup> VP <sup>G</sup> TVSGY <sup>T</sup> SAHADRFASKL <sup>P</sup> WKETLSFGFHDRAA <sup>A</sup> PVVADYFSSTLGP				180
Sbjct 121	LPLAEKRRAR <sup>R</sup> VP <sup>G</sup> TVSGY <sup>T</sup> SAHADRFASKL <sup>P</sup> WKETLSFGFHDRAA <sup>A</sup> PVVADYFSSTLGP				180
Query 181	DFAPMGRVY <sup>Q</sup> KYCEEMKEL <sup>S</sup> LIMELLESLGVERGY <sup>R</sup> REFFADSSIMRCNY <sup>Y</sup> PPCPEP				240
Sbjct 181	DFAPMGRVY <sup>Q</sup> KYCEEMKEL <sup>S</sup> LIMELLESLGVERGY <sup>R</sup> REFFADSSIMRCNY <sup>Y</sup> PPCPEP				240
Query 241	ERTLTGPHCD <sup>P</sup> TAL <sup>T</sup> ILLQDDVGGLE <sup>V</sup> LVDGEWRP <sup>V</sup> SPVPGAMVINIG <sup>D</sup> TFMALSNGRY				300
Sbjct 241	ERTLTGPHCD <sup>P</sup> TAL <sup>T</sup> ILLQDDVGGLE <sup>V</sup> LVDGEWRP <sup>V</sup> SPVPGAMVINIG <sup>D</sup> TFMALSNGRY				300
Query 301	KSCLHRAV <sup>N</sup> QRRERRSLA <sup>F</sup> FLCPREDRV <sup>R</sup> PPPSAAT <sup>P</sup> QHYPDF <sup>T</sup> WADLMR <sup>F</sup> TQRHYRA				360
Sbjct 301	KSCLHRAV <sup>N</sup> QRRERRSLA <sup>F</sup> FLCPREDRV <sup>R</sup> PPPSAAT <sup>P</sup> QHYPDF <sup>T</sup> WADLMR <sup>F</sup> TQRHYRA				360
Query 361	DTRTLDA <sup>F</sup> TRWLAP <sup>P</sup> AADAATAQVEAAS 389				
Sbjct 361	DTRTLDA <sup>F</sup> TRWLAP <sup>P</sup> AADAATAQVEAAS 389				

## Related Information

[Gene](#) - associated gene details

## putative gibberellin 20 oxidase [Oryza rufipogon]

Sequence ID: [AAN73384.1](#) Length: 389 Number of Matches: 1[See 127 more title\(s\)](#)

Range 1: 1 to 389 GenPept Graphics Next Match Previous Match

Score	Expect	Method	Identities	Positives	Gaps
788 bits(2035)	0.0	Compositional matrix adjust.	387/389(99%)	388/389(99%)	0/389(0%)
Query 1	MVAEHP <sup>T</sup> PPQHP <sup>Q</sup> PPMDSTAGSGIAAPAAAACV <sup>D</sup> LRMEPKI <sup>E</sup> PEFFVW <sup>F</sup> PNGDARPASAAE				60
Sbjct 1	MVAEHP <sup>T</sup> PPQHP <sup>Q</sup> PPMDSTAGSGIAAPAAAACV <sup>D</sup> LRMEPKI <sup>E</sup> PEFFVW <sup>F</sup> PNGDARPASAAE				60
Query 61	LDMPVVDV <sup>G</sup> VL <sup>R</sup> DGDAEGLRRAAAQVAAACATHG <sup>F</sup> FQVSEHGVDAA <sup>L</sup> ARAALDGASDFFR				120
Sbjct 61	LDMPVVDV <sup>G</sup> VL <sup>R</sup> DGDAEGLRRAAAQVAAACATHG <sup>F</sup> FQVSEHGVDAA <sup>L</sup> ARAALDGASDFFR				120
Query 121	LPLAEKRRAR <sup>R</sup> VP <sup>G</sup> TVSGY <sup>T</sup> SAHADRFASKL <sup>P</sup> WKETLSFGFHDRAA <sup>A</sup> PVVADYFSSTLGP				180
Sbjct 121	LPLAEKRRAR <sup>R</sup> VP <sup>G</sup> TVSGY <sup>T</sup> SAHADRFASKL <sup>P</sup> WKETLSFGFHDRAA <sup>A</sup> PVVADYFSSTLGP				180
Query 181	DFAPMGRVY <sup>Q</sup> KYCEEMKEL <sup>S</sup> LIMELLESLGVERGY <sup>R</sup> REFFADSSIMRCNY <sup>Y</sup> PPCPEP				240
Sbjct 181	DFAPMGRVY <sup>Q</sup> KYCEEMKEL <sup>S</sup> LIMELLESLGVERGY <sup>R</sup> REFFADSSIMRCNY <sup>Y</sup> PPCPEP				240
Query 241	ERTLTGPHCD <sup>P</sup> TAL <sup>T</sup> ILLQDDVGGLE <sup>V</sup> LVDGEWRP <sup>V</sup> SPVPGAMVINIG <sup>D</sup> TFMALSNGRY				300
Sbjct 241	ERTLTGPHCD <sup>P</sup> TAL <sup>T</sup> ILLQDDVGGLE <sup>V</sup> LVDGEWRP <sup>V</sup> SPVPGAMVINIG <sup>D</sup> TFMALSNGRY				300
Query 301	KSCLHRAV <sup>N</sup> QRRERRSLA <sup>F</sup> FLCPREDRV <sup>R</sup> PPPSAAT <sup>P</sup> QHYPDF <sup>T</sup> WADLMR <sup>F</sup> TQRHYRA				360
Sbjct 301	KSCLHRAV <sup>N</sup> QRRERRSLA <sup>F</sup> FLCPREDRV <sup>R</sup> PPPSAAT <sup>P</sup> QHYPDF <sup>T</sup> WADLMR <sup>F</sup> TQRHYRA				360
Query 361	DTRTLDA <sup>F</sup> TRWLAP <sup>P</sup> AADAATAQVEAAS 389				
Sbjct 361	DTRTLDA <sup>F</sup> TRWLAP <sup>P</sup> AADAATAQVEAAS 389				

## Related Information

[Gene](#) - associated gene details[Identical Proteins](#) - Identical proteins to AAN73384.1

## SD1 [Oryza rufipogon]

Sequence ID: [AER45865.1](#) Length: 389 Number of Matches: 1

Range 1: 1 to 389 GenPept Graphics Next Match Previous Match

Score	Expect	Method	Identities	Positives	Gaps
788 bits(2034)	0.0	Compositional matrix adjust.	387/389(99%)	387/389(99%)	0/389(0%)

## Related Information

Protein

Protein

Search

Advanced

Help

GenPept

Send to

Change region shown

Customize view

# SD1 [Oryza sativa Japonica Group]

GenBank: AER45907.1

[Identical Proteins](#) [FASTA](#) [Graphics](#)

Go to

LOCUS AER45907 389 aa linear PLN 25-JUL-2016  
 DEFINITION SD1 [Oryza sativa Japonica Group].  
 ACCESSION AER45907  
 VERSION AER45907.1  
 DBSOURCE accession [JN541538.1](#)  
 KEYWORDS .  
 SOURCE Oryza sativa Japonica Group (Japanese rice)  
 ORGANISM [Oryza sativa Japonica Group](#)  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; BOP  
 clade; Oryzoideae; Oryzeae; Oryzinae; Oryza; Oryza sativa.  
 REFERENCE 1 (residues 1 to 389)  
 AUTHORS Reagon,M., Thurber,C.S., Olsen,K.M., Jia,Y. and Caicedo,A.L.  
 TITLE The long and the short of it: SD1 polymorphism and the evolution of  
 growth trait divergence in U.S. weedy rice  
 JOURNAL Mol. Ecol. (2011) In press  
 PUBMED [21854475](#)  
 REMARK Publication Status: Available-Online prior to print  
 REFERENCE 2 (residues 1 to 389)  
 AUTHORS Reagon,M., Thurber,C.S., Olsen,K.M., Jia,Y. and Caicedo,A.L.  
 TITLE Direct Submission  
 JOURNAL Submitted (01-AUG-2011) Department of Biology, University of  
 Massachusetts, Amherst, MA 01003, USA  
 COMMENT Method: conceptual translation.  
 FEATURES Location/Qualifiers  
 source 1..389  
 /organism="Oryza sativa Japonica Group"  
 /cultivar="Kotobuki Mochi"  
 /specimen\_voucher="2545"  
 /db\_xref="taxon:39947"  
 /chromosome="1"  
 /country="Japan"  
 Protein 1..389  
 /product="SD1"  
 Region 33..372  
 /region\_name="PLN02276"  
 /note="gibberellin 20-oxidase"  
 /db\_xref="CDD:215156"  
 Region 64..169  
 /region\_name="DIOX\_N"  
 /note="non-haem dioxygenase in morphine synthesis  
 N-terminal; pfam14226"  
 /db\_xref="CDD:290926"  
 Region 227..324  
 /region\_name="2OG-FeII\_Oxy"  
 /note="2OG-Fe(II) oxygenase superfamily; pfam03171"  
 /db\_xref="CDD:281202"  
 CDS 1..389  
 /coded\_by="join(JN541538.1:1522..2078,  
 JN541538.1:2181..2502,JN541538.1:3935..4225)"  
 ORIGIN  
 1 mvaehptppq phqppmdst agsgiaapaa aavcdlrmp kipepfvwpn gdarpsaae  
 61 ldmpvvdvgy lrdgdaegl raaqvaaac athgffqvsx hgvdaalara aldgasdffr  
 121 lplaekrrar rvpgtvsgyt sahadrfask lpwketlsfg fhdraapvv adyfsstlpg  
 181 dfapmgrvyq kyceemkels ltimellels lgvergyyre ffadsssimr cnypppcpep

Analyze this sequence

Run BLAST

Identify Conserved Domains

Highlight Sequence Features

Find in this Sequence

Articles about the LOC4325003 gene

Control of grain protein contents through SEMIDWARF1 mute [Mol Genet Genomics. 2015]

Green revolution: a mutant gibberellin-synthesis gene in rice. [Nature. 2002]

See all...

Reference sequence information

RefSeq protein

See the reference protein sequence for PREDICTED: gibberellin 20 oxidase 2 (XP\_015627721.1).

More about the gene LOC4325003

LOC4325003 gene  
 Also Known As: OSNPB\_010883800, 20ox2,...

Related information

Similar protein sequences using SmartBlast

Related Sequences

CDD Search Results

Conserved Domains (Concise)

Conserved Domains (Full)

Domain Relatives

Gene

GeneView in dbSNP

Nucleotide

PopSet

Proteins with Similar Sequences

PubMed

Sequences producing significant alignments:

Select: All None Selected:0

Alignments Download GenPept Graphics Distance tree of results Multiple alignment

	Description	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/>	PREDICTED: gibberellin 20 oxidase 2 [Oryza sativa Japonica Group]	793	793	100%	0.0	100%	<a href="#">XP_015627721.1</a>
<input type="checkbox"/>	SD1 [Oryza sativa Japonica Group]	790	790	100%	0.0	99%	<a href="#">AER45907.1</a>
<input type="checkbox"/>	putative gibberellin 20 oxidase [Oryza rufipogon]	788	788	100%	0.0	99%	<a href="#">AAN73384.1</a>
<input type="checkbox"/>	SD1 [Oryza rufipogon]	788	788	100%	0.0	99%	<a href="#">AER45865.1</a>
<input type="checkbox"/>	GA C20oxidase2 [Oryza glumipatula]	786	786	100%	0.0	99%	<a href="#">BAG80954.1</a>
<input type="checkbox"/>	gibberellin-20 oxidase-2 [Oryza rufipogon]	785	785	100%	0.0	99%	<a href="#">BAK39011.1</a>
<input type="checkbox"/>	SD1 [Oryza rufipogon]	785	785	100%	0.0	99%	<a href="#">AER45849.1</a>
<input type="checkbox"/>	gibberellin-20 oxidase-2 [Oryza sativa Indica Group]	785	785	100%	0.0	99%	<a href="#">BAL03272.1</a>
<input type="checkbox"/>	GA C20oxidase2 [Oryza longistaminata]	783	783	100%	0.0	98%	<a href="#">BAG80956.1</a>
<input type="checkbox"/>	GA C20oxidase2 [Oryza meridionalis]	778	778	100%	0.0	98%	<a href="#">BAG80955.1</a>
<input type="checkbox"/>	gibberellin-20 oxidase-2 [Oryza sativa Indica Group]	774	774	100%	0.0	98%	<a href="#">BAK39040.1</a>
<input type="checkbox"/>	gibberellin-20 oxidase-2 [Oryza sativa Indica Group]	774	774	100%	0.0	98%	<a href="#">BAK39033.1</a>
<input type="checkbox"/>	gibberellin-20 oxidase-2 [Oryza sativa Indica Group]	773	773	100%	0.0	98%	<a href="#">BAK39038.1</a>
<input type="checkbox"/>	SD1 [Oryza sativa]	767	767	97%	0.0	99%	<a href="#">AER45797.1</a>
<input type="checkbox"/>	hypothetical protein Osl_04688 [Oryza sativa Indica Group]	687	687	95%	0.0	93%	<a href="#">EAY76733.1</a>
<input type="checkbox"/>	GA C20oxidase2 [Oryza punctata]	671	671	95%	0.0	95%	<a href="#">BAG80957.1</a>
<input type="checkbox"/>	GA C20oxidase2 [Oryza alta]	659	659	95%	0.0	89%	<a href="#">BAG80960.1</a>
<input type="checkbox"/>	GA C20oxidase2 [Oryza officinalis]	644	644	92%	0.0	95%	<a href="#">BAG80959.1</a>
<input type="checkbox"/>	GA C20oxidase2 [Oryza minuta]	638	638	92%	0.0	94%	<a href="#">BAG80958.1</a>
<input type="checkbox"/>	GA C20oxidase2 [Oryza brachyantha]	637	637	100%	0.0	88%	<a href="#">BAG80961.1</a>
<input type="checkbox"/>	SD1 [Oryza rufipogon]	631	631	79%	0.0	99%	<a href="#">AER45851.1</a>
<input type="checkbox"/>	GA C20oxidase2 [Oryza meyeriana var. granulata]	626	626	95%	0.0	87%	<a href="#">BAG80962.1</a>
<input type="checkbox"/>	gibberellin 20 oxidase 2 isoform X1 [Sorghum bicolor]	615	615	97%	0.0	83%	<a href="#">XP_002456751.1</a>
<input type="checkbox"/>	hypothetical protein SORBI_3003G379500 [Sorghum bicolor]	614	614	97%	0.0	83%	<a href="#">OQU87995.1</a>
<input type="checkbox"/>	uncharacterized protein LOC107521947 [Zea mays]	609	609	96%	0.0	85%	<a href="#">NP_001308615.1</a>
<input type="checkbox"/>	gibberellin 20 oxidase 2-like [Aegilops tauschii subsp. tauschii]	607	607	94%	0.0	83%	<a href="#">XP_020187952.1</a>
<input type="checkbox"/>	unnamed protein product [Triticum aestivum]	605	605	97%	0.0	81%	<a href="#">CDM85079.1</a>
<input type="checkbox"/>	GA20ox-D2 [Triticum aestivum]	604	604	94%	0.0	83%	<a href="#">CFV04355.1</a>
<input type="checkbox"/>	PREDICTED: gibberellin 20 oxidase 2-like [Oryza brachyantha]	603	603	100%	0.0	80%	<a href="#">XP_015690591.1</a>
<input type="checkbox"/>	PREDICTED: gibberellin 20 oxidase 2-like [Setaria italica]	602	602	87%	0.0	88%	<a href="#">XP_004970813.1</a>

約 1,010 件 (0.29 秒)

**GA20OX2 - Gibberellin 20 oxidase 2 - Arabidopsis thaliana (Mouse ...**[www.uniprot.org/uniprot/Q39111](http://www.uniprot.org/uniprot/Q39111) ▾ [このページを訳す](#)

Key oxidase enzyme in the biosynthesis of gibberellin that catalyzes the conversion of GA12 and GA53 to GA9 and GA20 respectively, via a three-step oxidation at C-20 of the GA skeleton. GA53 is less effectively oxidized than GA12, and ...

**RiceXPro**[ricexpro.dna.affrc.go.jp/GGEP/gene-search.php?keyword...](http://ricexpro.dna.affrc.go.jp/GGEP/gene-search.php?keyword...) ▾ [このページを訳す](#)

(Gibberellin C-20 oxidase 2) (**GA 20-oxidase 2**) (Os20ox2) (Semidwarf-1 protein). LOC\_Os01g66100 · 18955, BD187781, S-12114 (non-unique), Gibberellin 20 oxidase 2 (EC 1.14.11.-) (Gibberellin C-20 oxidase 2) (**GA 20-oxidase 2**) ...

**KEGG REACTION: R06322**[www.genome.jp/dbget-bin/www\\_bget?R06322+R06323+R06326](http://www.genome.jp/dbget-bin/www_bget?R06322+R06323+R06326) ▾ [このページを訳す](#)

Name. (gibberellin-44),2-oxoglutarate:oxygen oxidoreductase. Definition. Gibberellin A24 <=> Gibberellin A9 + CO2. Equation. C11861 <=> C11863 + C00011. Comment. **GA 20-oxidase**(2-oxoglutarate dependent dioxygenase), GA20ox

**KEGG REACTION: R07184**[www.genome.jp/dbget-bin/www\\_bget?reaction+R07184...](http://www.genome.jp/dbget-bin/www_bget?reaction+R07184...) ▾ [このページを訳す](#)

Name. (gibberellin-44),2-oxoglutarate:oxygen oxidoreductase. Definition. Gibberellin A24 <=> Gibberellin A9 + CO2. Equation. C11861 <=> C11863 + C00011. Comment. **GA 20-oxidase**(2-oxoglutarate dependent dioxygenase), GA20ox

**Gene - Detail - SHIGEN**<https://shigen.nig.ac.jp/rice/oryzabase/gene/detail/470> ▾ [このページを訳す](#)

2017/07/07 - Gene Name Synonym, dee-geo-woo-gen dwarf(sd1-d), green revolution gene, semidwarf-1, GA C20oxidase2, semidwarf-1, reimei dwarf(sd1-r), Calrose76(sd1-3), Jikkoku(sd1-2). rice GA 20-oxidase2, **GA 20-oxidase 2**, ...

**Biosafety of Forest Transgenic Trees: Improving the Scientific Basis ...**<https://books.google.co.jp/books?isbn=9401775311> - [このページを訳す](#)

Cristina Vettori, Fernando Gallardo, Hely Häggman - 2016 - Nature (2015) 35S::GA2oxidase1,6::OCS 35S::GA2oxidase3::OCS SH1::SH1::SH1 PHOR1::PHOR1::PHOR1 GA2ox1::GA2oxidase2::NOS RGL1-1::**GA20oxidase2**::NOS Empty vector Phytochrome trial 717 35S::PHYB1::OCS 2004– 2013 220 ...

**oxidoreductase, 2OG-Fe(II) oxygenase family protein - Protein Result**<https://www.ncbi.nlm.nih.gov/protein?term=oxidoreductase...> - [このページを訳す](#)

RecName: Full=Gibberellin 20 oxidase 2; AltName: Full=**GA 20-oxidase 2**; AltName: Full=Gibberellin C-20 oxidase 2; AltName: Full=Os20ox2; AltName: Full =Protein semidwarf-1 389 aa protein P0C5H5.1 GI:158705787 2. RecName: ...

**EARLY FLOWERING3 Regulates Flowering in Spring Barley by ...**<https://www.ncbi.nlm.nih.gov/pubmed/24781117> - [このページを訳す](#)

# UniProtKB - Q39111 (GAOX2\_ARATH)



## Display

[Entry](#)
[Publications](#)
[Feature viewer](#)
[Feature table](#)

All None

 Function

 Names & Taxonomy

 Subcellular location

 Pathology & Biotech

 PTM / Processing

 Expression

 Interaction

 Structure

 Family & Domains

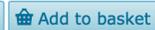
 Sequence

 Similar proteins

 Cross-references

 Entry information

 Miscellaneous

[Top](#)

**Protein** | **Gibberellin 20 oxidase 2**
**Gene** | **GA20OX2**
**Organism** | *Arabidopsis thaliana* (Mouse-ear cress)

**Status** | Reviewed - Annotation score: ●●●●● - Experimental evidence at transcript level<sup>i</sup>

## Function<sup>i</sup>

Key oxidase enzyme in the biosynthesis of gibberellin that catalyzes the conversion of GA12 and GA53 to GA9 and GA20 respectively, via a three-step oxidation at C-20 of the GA skeleton. GA53 is less effectively oxidized than GA12, and GA25 is also formed as a minor product. Involved in the promotion of the floral transition, fertility and silique elongation, but plays only a minor role in elongation of seedling organs. Acts redundantly with GA20OX1. [1 Publication](#)

## Cofactor<sup>i</sup>

Protein has several cofactor binding sites:

 Fe<sup>2+</sup>

**Note:** Binds 1 Fe<sup>2+</sup> ion per subunit.

L-ascorbate

2-oxoglutarate

## Pathway<sup>i</sup>: gibberellin biosynthesis

This protein is involved in the pathway gibberellin biosynthesis, which is part of Plant hormone biosynthesis.

View all proteins of this organism that are known to be involved in the pathway [gibberellin biosynthesis](#) and in [Plant hormone biosynthesis](#).

## Sites

Feature key	Position(s)	Description	Actions	Graphical view	Length
Metal binding <sup>i</sup>	245	Iron <a href="#">PROSITE-ProRule annotation</a>			1
Metal binding <sup>i</sup>	247	Iron <a href="#">PROSITE-ProRule annotation</a>			1
Metal binding <sup>i</sup>	301	Iron <a href="#">PROSITE-ProRule annotation</a>			1
Active site <sup>i</sup>	311	<a href="#">Sequence analysis</a>			1

## GO - Molecular function<sup>i</sup>

- [gibberellin 20-oxidase activity](#) [Source: TAIR](#)
- [metal ion binding](#) [Source: UniProtKB-KW](#)

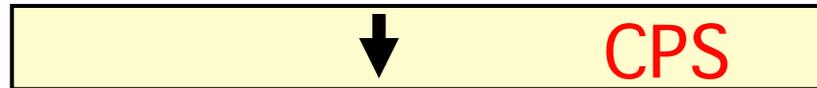
[Complete GO annotation...](#)

## GO - Biological process<sup>i</sup>

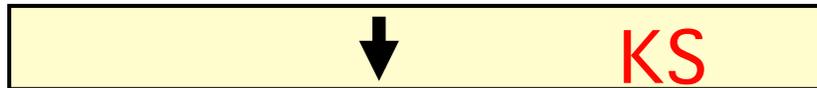
- [flower development](#) [Source: TAIR](#)
- [gibberellic acid mediated signaling pathway](#) [Source: GO\\_Central](#)
- [gibberellin biosynthetic process](#) [Source: TAIR](#)
- [response to gibberellin](#) [Source: TAIR](#)
- [response to karrikin](#) [Source: TAIR](#)
- [response to red or far red light](#) [Source: TAIR](#)

# 植物におけるGA合成経路

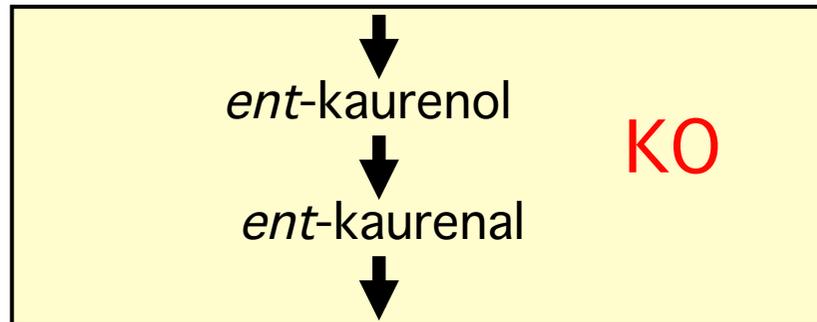
Geranylgeranyl diphosphate(GGDP)



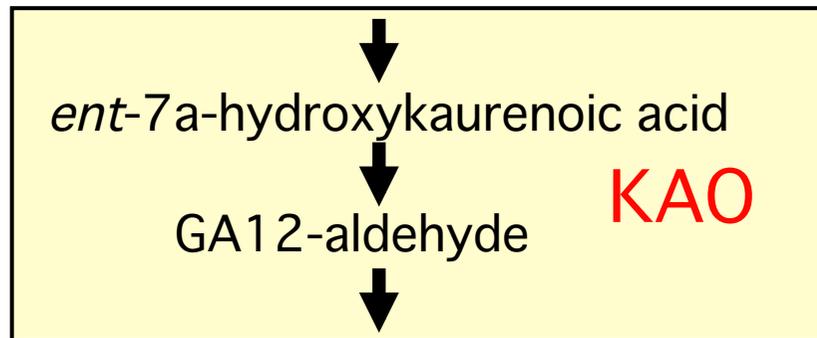
*ent*-Copalyl diphosphate(CDP)



*ent*-kaurene Plastids

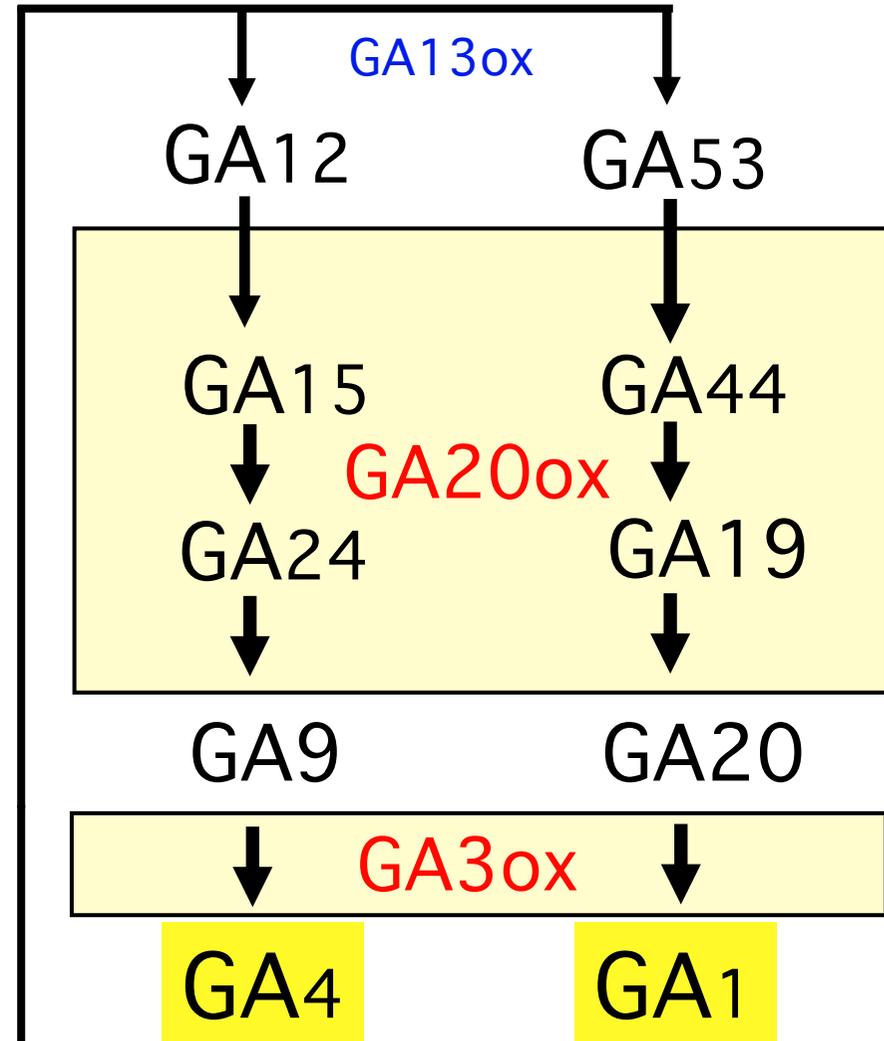


*ent*-kaurenoic acid



GA12

Endoplasmic reticulum



Cytoplasm

約 1,710 件 (0.38 秒)

## 名古屋大学生物機能開発利用研究センター植物分子育種分野

[bbc.agr.nagoya-u.ac.jp/~pmb/HANWAISEI.htm](http://bbc.agr.nagoya-u.ac.jp/~pmb/HANWAISEI.htm)

GA20酸化酵素遺伝子はその5番目の反応を触媒する。sd1変異体が、ジベレリン合成不全によりイネの生長が阻害され矮性になったとしたら、それ以外の合成反応がダメになっても、同じようにイネの矮性が引き起こされるはずで、それらの矮性変異体を用いれ...

## 半矮性遺伝子 - Wikipedia

<https://ja.wikipedia.org/wiki/半矮性遺伝子>

半矮性遺伝子（はんわいせいいでんし、semidwarf gene）とは、植物の背の高さ（草丈）を低くする遺伝子。主に植物育種学で用いられる用語。緑の革命においてイネ、コムギ品種の改良に用いられ、その子孫に当たる品種に引き継がれている。... しかしながら、先述のイネの半矮性遺伝子sd1および小麦農林10号の半矮性遺伝子Rht1,Rht2の作用機構には植物ホルモンの一種であるジベレリンが関与することが明らかになっている。

[矮性と半矮性](#) · [歴史](#) · [生理](#) · [脚注](#)

## <sup>[PDF]</sup> 古代人の 緑の革命 イネにおける 緑の革命 遺伝子の人類への ... - 神...

[www.kobe-u.ac.jp/documents/research/news/H23/pp2011\\_05\\_06-02.pdf](http://www.kobe-u.ac.jp/documents/research/news/H23/pp2011_05_06-02.pdf)

2011/06/20 - 1940 1960年代にイネやコムギで 緑の革命が起こり、収量が飛躍的に増加しました。イネで大きく貢献した遺伝子はSD1 (Semi-dwarf1、半矮性遺伝子、この変異体の草丈は低くなる) であり、現在でも世界中の多くの品種でこの遺伝子が ...

## <sup>[PDF]</sup> イネの収量を決定する重要遺伝子を同定 - 理化学研究所

[www.riken.jp/~media/riken/pr/press/2005/20050624\\_1/20050624\\_1.pdf](http://www.riken.jp/~media/riken/pr/press/2005/20050624_1/20050624_1.pdf)

2005/06/24 - 一総長) は、イネの収量増加に関わる重要な遺伝子が第一染色体上腕部の「OsCKX2」であることを ... さらにこの知見を分子育種に利用し、Gn1 と背丈を低くする QTL(sd1) ... し、イネの緑の革命の原因遺伝子であった sd1※4 と同一でした。

## イネの半矮性遺伝子sd-1と 脱粒性遺伝子の連鎖 緒言 実験材料 ... - J-St...

[https://www.jstage.jst.go.jp/article/jsta1957/37/2/37\\_2\\_115/\\_pdf](https://www.jstage.jst.go.jp/article/jsta1957/37/2/37_2_115/_pdf)

S OBA 著 - 1993 - 被引用数: 3 - 関連記事

1993. イネの半矮性遺伝子sd-1と 脱粒性遺伝子の連鎖. 第2報. インド型半矮性在来品種低脚尖の持つ半矮性遺伝子 sd-1と 脱粒性遺伝子の連鎖分析. 大場 伸 哉・菊 池 文 雄\*. 岐阜大学農学部. 〒501-11岐阜市柳戸1-1,. \*筑波大学農林学系, 〒305つ.

## 作物ゲノム育種研究センター:Sd1 | 農研機構

[www.naro.affrc.go.jp](http://www.naro.affrc.go.jp) > ... > ゲノム育種支援 > ゲノム育種マーカー情報(イネ) >

農研機構は食料・農業・農村に関する研究開発を行う機関です。マーカー名(遺伝子名、遺伝子座名) Sd1 形質 耐倒伏性 特性 収量性 活用...

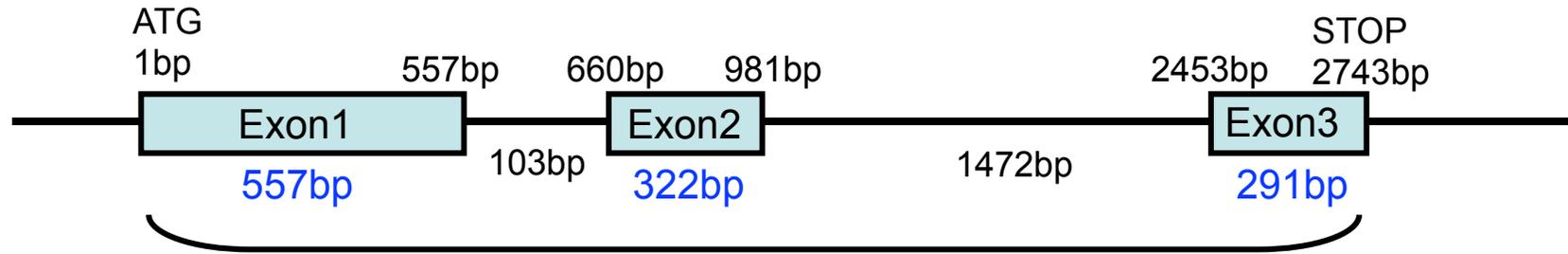
## 短稈遺伝子sd1のアリルにより、イネ玄米のタンパク質含量がコントロ...

[www.naro.affrc.go.jp](http://www.naro.affrc.go.jp) > ... > 中央農業総合研究センター 2015年の成果情報 >

農研機構は食料・農業・農村に関する研究開発を行う機関です。玄米のタンパク質含量のQTLが第1染色体に有り、インディカ型劣性で高タンパク質含量を示す。この原因遺伝子は短稈遺伝子 <em>sd1</em>であり、低脚尖およびCalrose 76由来 ...

## 第1回 組み換えと異なるもうひとつの遺伝子技術 - ナショナルジオグラ

## ある遺伝子Aの構造



**2743bp**

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 322bp, 第3エクソン 291bp)、  
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

```
MVAEHPTPPQPHQPPPMDS TAGSGIAAPAAA AVCDLRMEPKIPEPFVWPNGDARPASAAE  
LDMPVVDVGVLRDGA EGLRRAAAQVAAACATHGFFQVSEHGVDAA LARAALDGASDFFR  
LPLAEKRRARRVPGTVSGY TSAHADRFASKLPWKETLSFGFHDR AAAPVVADYFSSTLGP  
DFAPMGRVYQKYCEEMKELSLT IMELLELSLGVERGYREFFADSS SIMRCNYYPPCPEP  
ERTLGTGPHCDPTALTILLQDDVGGLEVLVDGEWRPVSPVPGAMVINIGDTFMALSNGRY  
KSCLHRAVNVNQRRERRSLAFFLCPREDRVVRPPPSAATPQHYPDFTWADLMRFTQRHYRA  
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。

## Basic Local Alignment Search Tool

**BLAST** finds regions of similarity between biological sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance. [Learn more](#)

NEWS

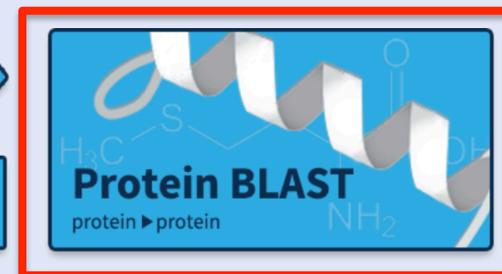
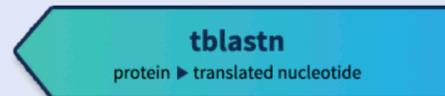
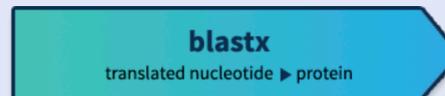
### QuickBLASTP

Try **QuickBLASTP** for a fast protein search of nr.

Tue, 23 May 2017 13:00:00 EST

[More BLAST news...](#)

## Web BLAST



## BLAST Genomes

Enter organism common name, scientific name, or tax id.

[Human](#)[Mouse](#)[Rat](#)[Microbes](#)

## Standalone and API BLAST



### Download BLAST

Get BLAST databases and executables



### Use BLAST API

Call BLAST from your application



### Use BLAST in the cloud

Start an instance at a cloud provider

## Specialized searches

### SmartBLAST



Find proteins highly similar to your query

### Primer-BLAST



Design primers specific to your PCR template

### Global Align



Compare two sequences across their entire span  
(Needleman, 1998)

### CD-search



Find conserved domains in your sequence

## Standard Protein BLAST

blastn blastp blastx tblastn tblastx

BLASTP programs search protein databases using a protein query. [more...](#)[Reset page](#) [Bookmark](#)

## Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s)

[Clear](#)

Query subrange

From

To

Or, upload file

ファイルを選択 ファイル未選択

Job Title

Enter a descriptive title for your BLAST search

 Align two or more sequences

## Choose Search Set

Database

Organism

Optional

Exclude

Optional

Entrez Query

Optional

## Program Selection

Algorithm

blastn

blastp

blastx

tblastn

tblastx

Choose a BLAST algorithm

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

BLAST

Search database Non-redundant protein sequences (nr) using BLASTP

 Show results in a new window[+ Algorithm parameters](#)

BLAST is a registered trademark of the National Library of Medicine

NCBI

National Center for Biotechnology Information, U.S. National Library of Medicine  
8600 Rockville Pike, Bethesda MD, 20894 USA[Policies and Guidelines](#) | [Contact](#)

1. TATGGCTTA----  
T G L

2. TATGGCTTA----  
M A -

3. TATGGCTTA----  
W L -

[Support center](#)[Mailing list](#)[YouTube](#)