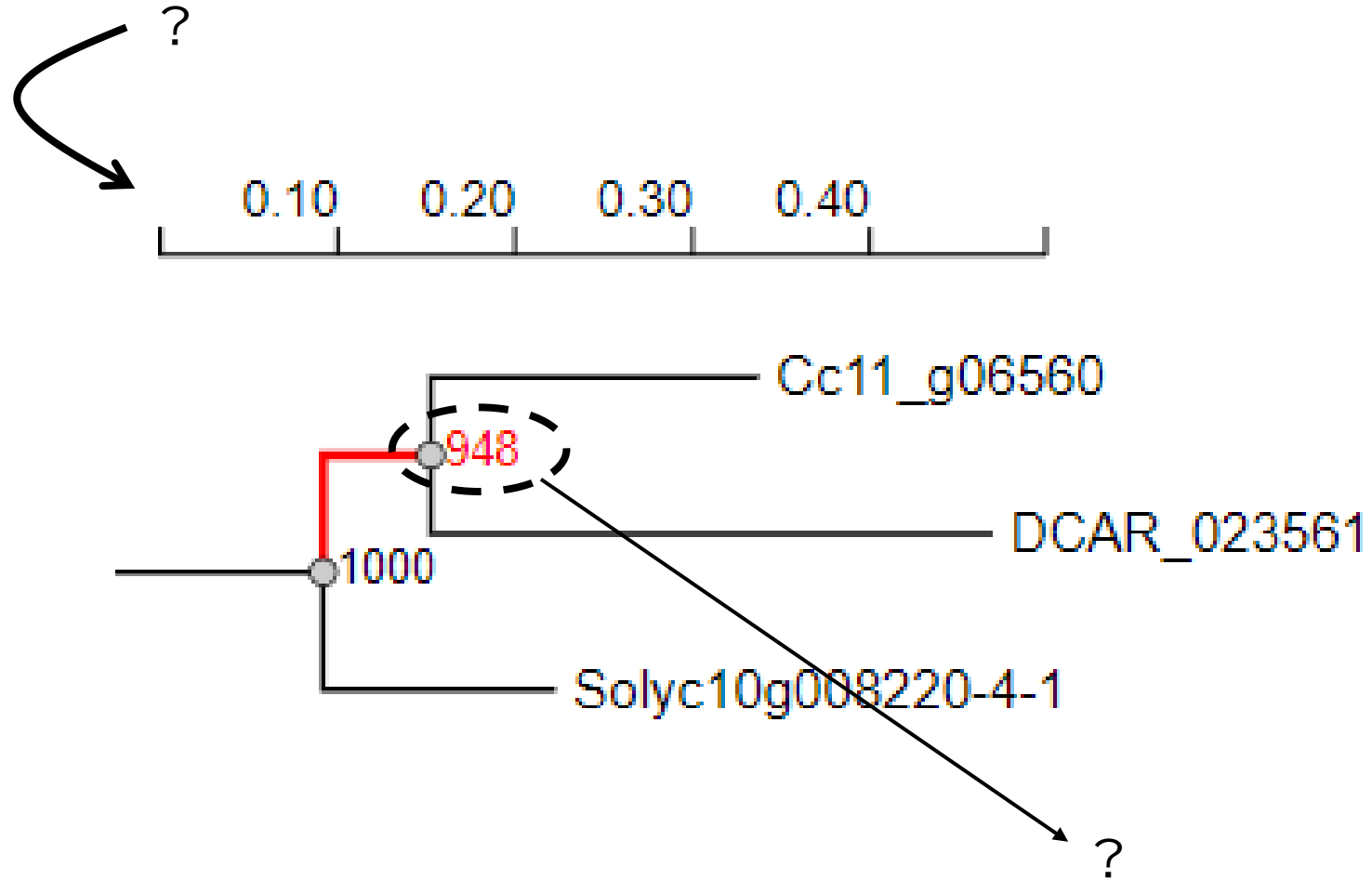

系統樹について & NCBIデータベースについて

2025/04/22

栽培植物起原学分野 助教

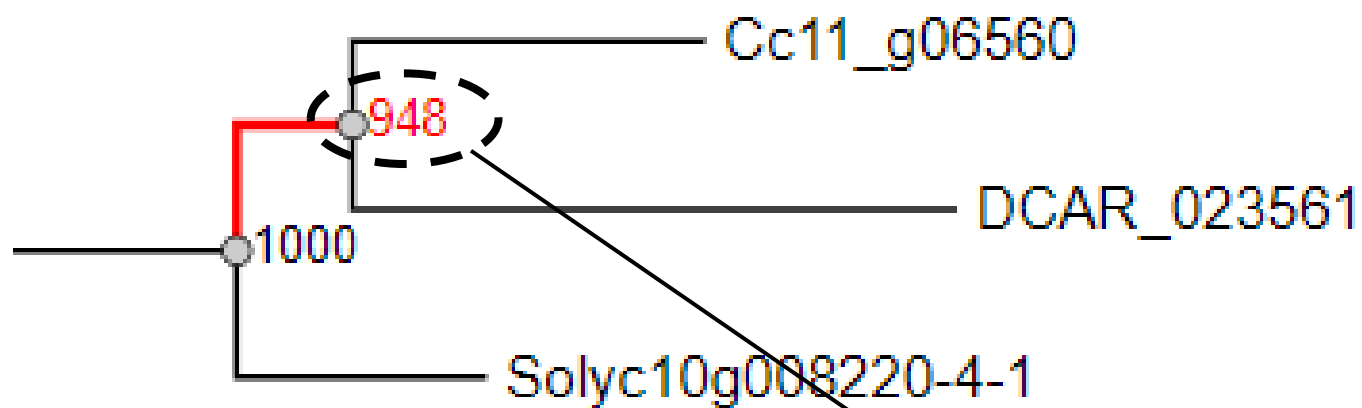
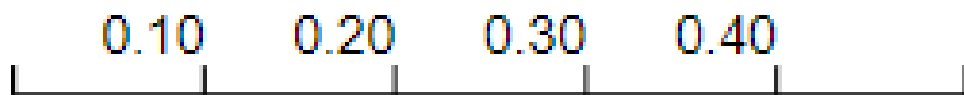
堺 俊之

系統樹の数値の意味



各数値は系統的な距離や分岐の信頼度を示す

Phylogenetic distance: 系統的な距離、
どれだけ配列に違いがあるかの基準



Branch support value

分岐の信頼度。

反復数に近いほど信頼度が高い。

(分岐していると考えて良い。)

Bootstrap branch supports

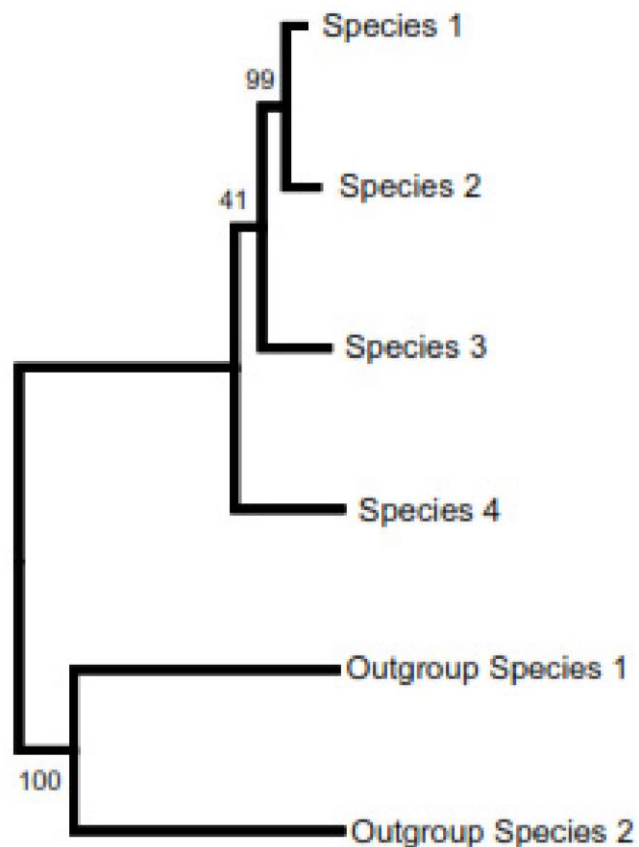
Yes

Number of replicates

1000

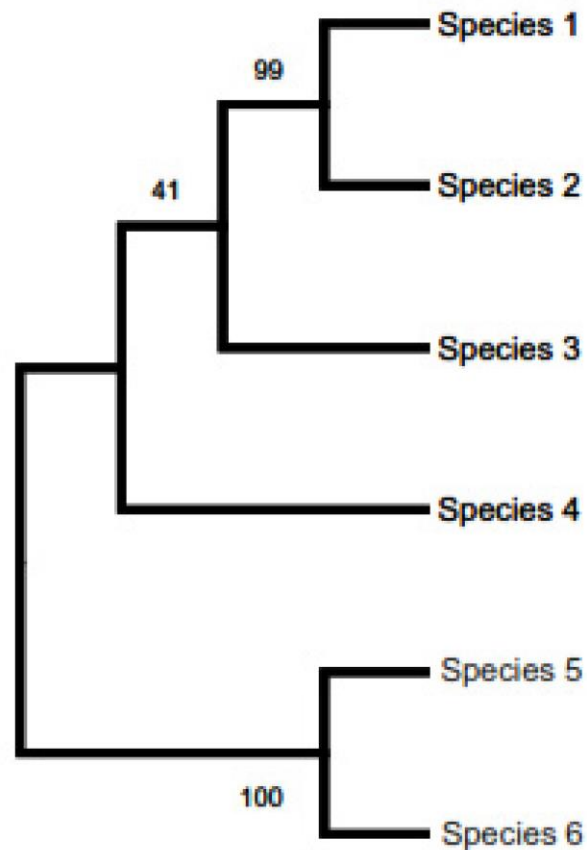
系統樹は系統的な距離の情報を含んでいるものを示す

○ Phylogenetic tree



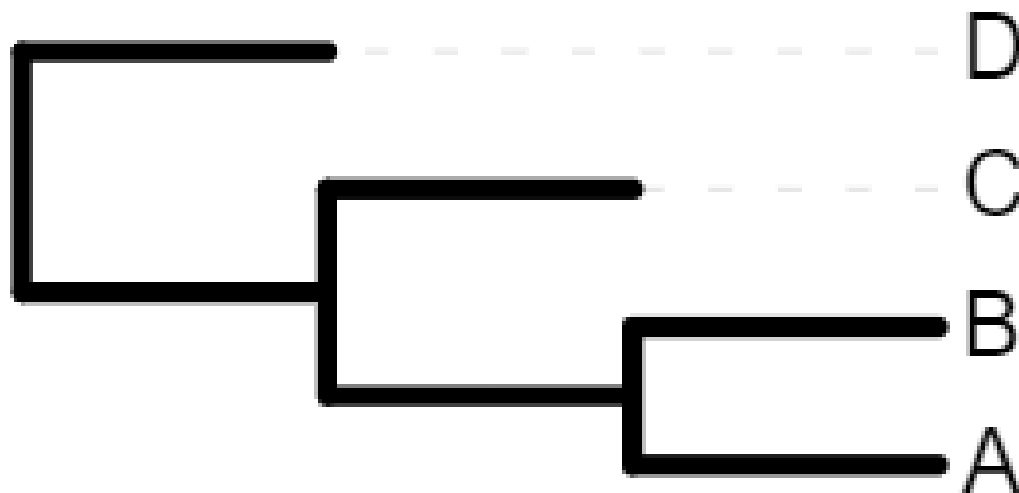
0.01

✗ Cladogram



系統樹に基づいた進化的な関係の推定

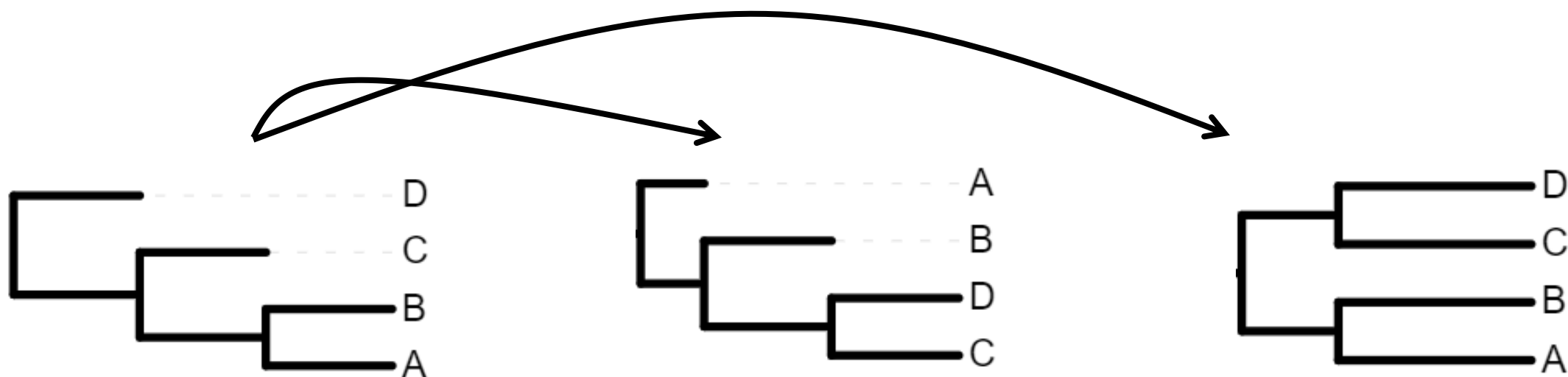
Q. 配列から以下の様な系統樹が得られた。進化の過程で DからCに分化し、B,Aに更に分化したと言えるか…？



系統樹に基づいた進化的な関係の推定

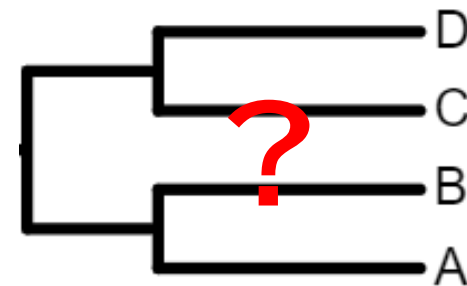
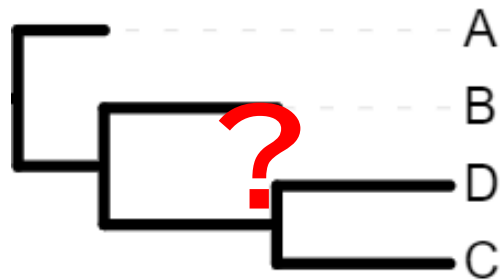
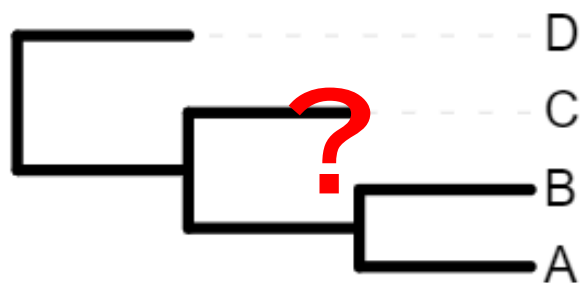
A. 言えない

どの遺伝子/種が最初か分からないので、
系統樹の関係からだけでは起源や進化的な関係は判断できない

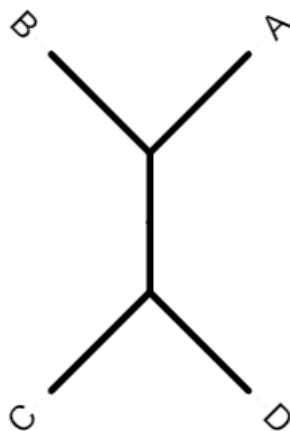
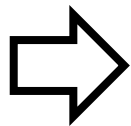


系統樹に基づいた進化的な関係の推定

起源や進化的な関係が判断できない場合: その①



系統的な関係性
(系統的に近いか遠いか)
のみ示す

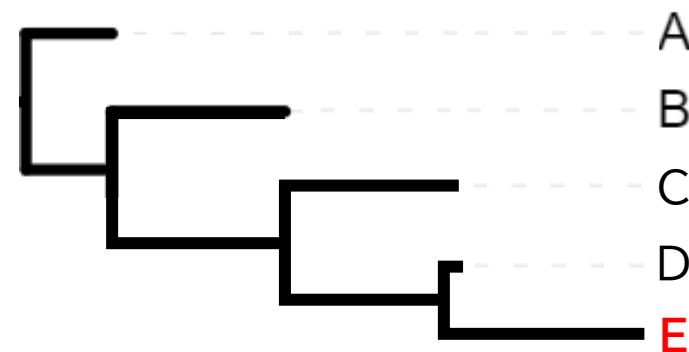
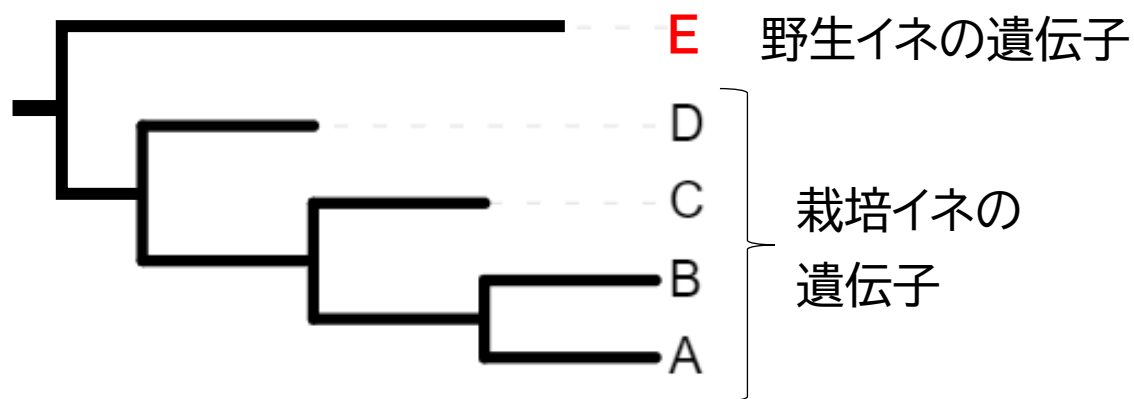
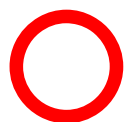


Unrooted tree と呼ぶ

系統樹に基づいた進化的な関係の推定

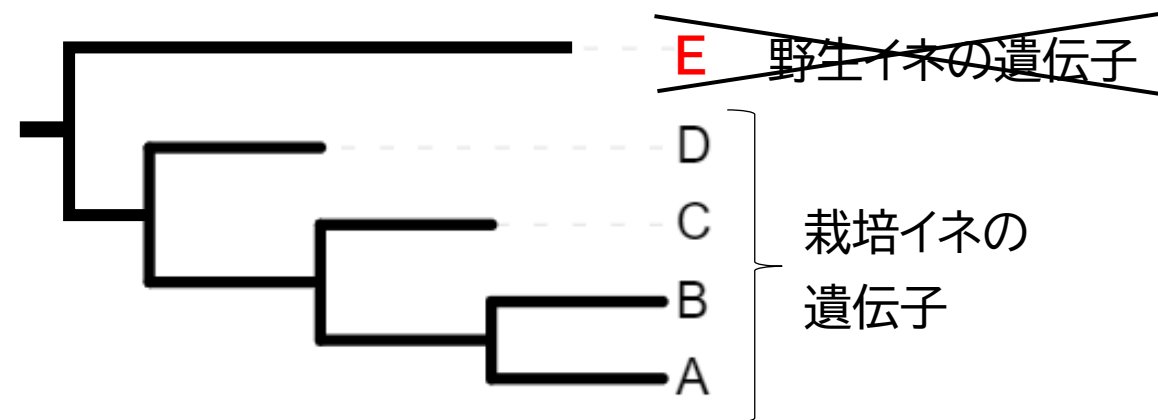
起源や進化的な関係が判断できない場合: その②

Outgroup(外群)を入れRootを決める

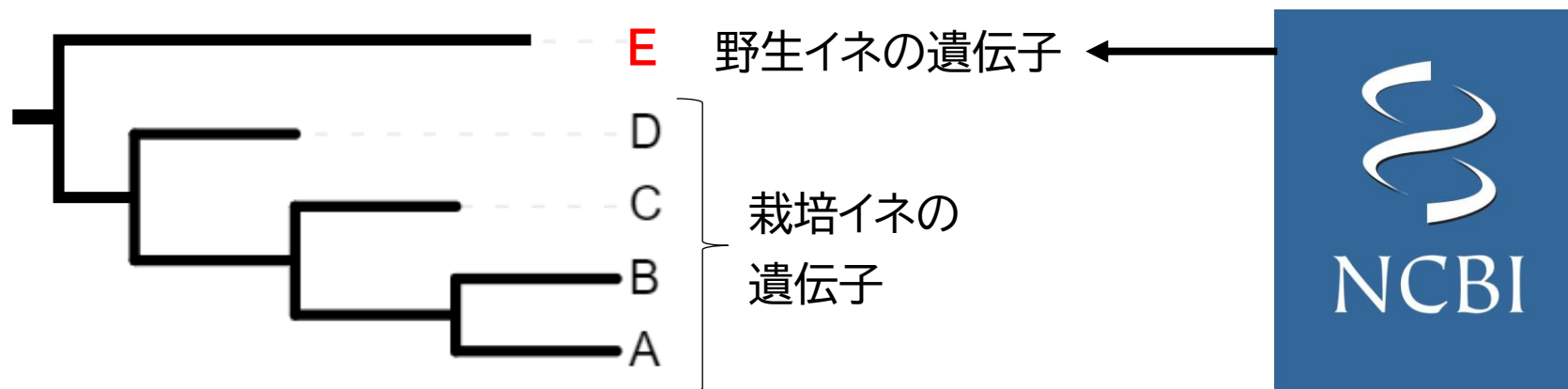


外群の配列をどう手に入れるのか？

- ① 外群のサンプルを準備しておく
→ 今回は無い



- ② NCBI等のシーケンスデータベースで使えるような遺伝子配列を探す



代表的な生物学のデータベース

- NCBI (National Center of Biotechnology Information)



- EBI (European Bioinformatics Institute)



- SIB (Swiss Institute of Bioinformatics)



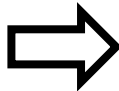
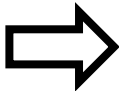
塩基配列・アミノ酸配列・タンパク質の構造・文献情報等、
様々な生物学のデータを上記のプラットフォームを通じて検索可能

NCBIのBLASTで類似配列の検索が可能

 配列A

GGGAAAAAACATTTTA
AAAAAATCTATAAAAAA
AATAGAATTCATAAAAT
TCTTGCTTTTCCATATAT
TGAGCTTCAATTATTTT
TTTTTCTCATAAAAAGTT
GCTTTTGGGAAAAA
AAAAAACTCATTTCTAA
TACCAAACAAGTATTT
CTTCAGGGAGAGTGC...

BLASTで
NCBIのデータベースを検索



様々な生物種の持つ
配列Aの類似配列を検出



GGGAAAAAACATTTTA
AAAAAATCTATAAAAAA
AATAGAATTCATAAAAT
TCTTGCA^ATTTCCATATAT
TGAGCTTCAATTATTTT
TTTTTCTCATAAAAAGTT
GCTTTTGGGAAAAA
AAAAAACTC^TTTTCTAA
TACCAAACAAGTATTT
CTTCAGGGAGAGTGC...



GGGAAAAAACATTTTA
AAAAAATCTATAAAAAA
AATAGAATTCATAAAAT
TCTT^CCTTTTCCATATAT
TGAGCTTCAATTATTTT
TTTTTCTCATAAAAAGTT
GCTTTTGGGAA^ATAAAA
AAAAAACTCATTTCTAA
TACCAAACAAGTATTT
CTTCAGGGAGAGTGC...



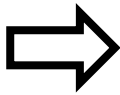
GGGAAAAAACATTTTA
AAAA^CATCTATAAA^TAA
AATAGAATTCATAAAAT
TCTTGCTTTTCCATATAT
TGAGCTTCAATTATTTT
TTTTTCTCATAAAAAGTT
GCTTTTGGGAAAAA
AAAAAACTCATTTCTAA
TACCA^GACAAGTATTT
CTTCAGGGAGAGTGC...

NCBIのBLASTで外群となる類似配列を検索する

ex) どんな種が配列Aの類似配列を持っているか？

配列A

GGGAAAAAACATTTTA
AAAAATCTATAAAAAA
AATAGAATTCATAAAAT
TCTTGCTTTTCCATATAT
TGAGCTTCAATTATTTTT
TTTTTCTCATAAAAAGTT
GCTTTTGGGAAAAAAA
AAAAAACTCATTTCTAA
TACCAAACAAGTATTT
CTTCAGGGAGAGTGC...



Descriptions										Graphic Summary	Alignments	Taxonomy		
Sequences producing significant alignments										Download	Select columns	Show	100	
<input checked="" type="checkbox"/> select all 30 sequences selected										GenBank	Graphics	Distance tree of results	MSA Viewer	
	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession					
<input checked="" type="checkbox"/>	PREDICTED: Solanum lycopersicum putative late blight resistance protein homolog R1B-14 (LOC10126636...)	Solanum lycopersicum	6231	6231	100%	0.0	100.00%	3374	XM_004248127.5					
<input checked="" type="checkbox"/>	PREDICTED: Solanum pennellii putative late blight resistance protein homolog R1B-14 (LOC107002325), tr...	Solanum pennellii	5749	5749	98%	0.0	98.07%	3309	XM_015200299.2					
<input checked="" type="checkbox"/>	PREDICTED: Solanum pennellii putative late blight resistance protein homolog R1B-14 (LOC107002325), tr...	Solanum pennellii	5557	5557	94%	0.0	98.45%	3179	XM_027912601.1					
<input checked="" type="checkbox"/>	PREDICTED: Solanum tuberosum putative late blight resistance protein homolog R1B-14 (LOC102595315),...	Solanum tuberosum	5142	5142	99%	0.0	94.65%	3305	XM_006366246.2					
<input checked="" type="checkbox"/>	Solanum lycopersicum clone SINRC0 NLR required for cell death 0 mRNA, complete cds	Solanum lycopersicum	4876	4876	79%	0.0	99.70%	2664	PP886667.1					
<input checked="" type="checkbox"/>	PREDICTED: Solanum stenotomum putative disease resistance RPP13-like protein 3 (LOC125866224), tra...	Solanum stenotomum	4752	4752	85%	0.0	99.15%	3455	XM_049546546.1					
<input checked="" type="checkbox"/>	PREDICTED: Solanum stenotomum putative disease resistance RPP13-like protein 3 (LOC125866224), tra...	Solanum stenotomum	4643	4643	85%	0.0	99.15%	3455	XM_049546546.1					
<input checked="" type="checkbox"/>	PREDICTED: Solanum stenotomum putative disease resistance RPP13-like protein 3 (LOC125866224), tra...	Solanum stenotomum	4632	4632	81%	0.0	97.15%	3753	XM_016546548.1					
<input checked="" type="checkbox"/>	PREDICTED: Solanum verrucosum putative late blight resistance protein homolog R1B-14 (LOC125835496),...	Solanum verrucosum	4527	4527	79%	0.0	97.33%	2664	XM_049514631.1					
<input checked="" type="checkbox"/>	PREDICTED: Solanum dulcamara putative disease resistance RPP13-like protein 3 (LOC129877507), mRNA	Solanum dulcamara	4337	4337	81%	0.0	95.13%	3103	XM_055953019.1					
<input checked="" type="checkbox"/>	PREDICTED: Lycium barbarum putative late blight resistance protein homolog R1B-14 (LOC132630207), tra...	Lycium barbarum	4235	4235	96%	0.0	93.21%	3406	XM_060345805.1					
<input checked="" type="checkbox"/>	PREDICTED: Lycium barbarum putative late blight resistance protein homolog R1B-14 (LOC132630207), tra...	Lycium barbarum	4071	4071	96%	0.0	93.22%	3406	XM_060345794.1					
<input checked="" type="checkbox"/>	PREDICTED: Lycium barbarum putative late blight resistance protein homolog R1B-14 (LOC132630207), tra...	Lycium barbarum	4052	4268	92%	0.0	93.39%	3422	XM_060345798.1					

Solanum lycopersicum
Solanum pennellii
Solanum tuberosum
Lycium barbarum
等々


NCBIのBLASTで外群となる類似配列を検索する

特定の種が持つ遺伝子の中で類似する配列を検索等も可能

ex) *Arabidopsis thaliana*のゲノム中に似た配列があるか？

配列A

GGGAAAAAACATTTTA
AAAAAATCTATAAAAAA
AATAGAATTCATAAAAT
TCTTGCTTTTCCATATAT
TGAGCTTCAATTATTTTT
TTTTTCTCATAAAAAGTT
GCTTTTGGGAAAAAAA
AAAAAACTCATTTCTAA
TACCAAACAAGTATTT
CTTCAGGGAGAGTGC...



Arabidopsis
Thaliana
のみ対象

Descriptions		Graphic Summary		Alignments		Taxonomy															
Sequences producing significant alignments						Download ▾		Select columns ▾		Show 100 ▾		?									
<input checked="" type="checkbox"/> select all 100 sequences selected						GenPept		Graphics		Distance tree of results		Multiple alignment		MSA Viewer							
Description ▾						Scientific Name ▾		Max Score ▾		Total Score ▾		Query Cover ▾		E value ▾		Per. Ident ▾		Acc. Len ▾		Accession	
<input checked="" type="checkbox"/> ZAR1 [Arabidopsis thaliana]						Arabidopsis thali...		180		180		94%		1e-49		31.41%		850		OAP03660.1	
<input checked="" type="checkbox"/> unnamed protein product [Arabidopsis thaliana]						Arabidopsis thali...		179		179		94%		1e-49		31.41%		850		CAA0385601.1	
<input checked="" type="checkbox"/> HOPZ-ACTIVATED RESISTANCE 1 [Arabidopsis thaliana]						Arabidopsis thali...		179		179		94%		1e-49		31.41%		852		NP_190664.1	
<input checked="" type="checkbox"/> myosin heavy chain homolog, partial [Arabidopsis thaliana]						Arabidopsis thali...		179		179		94%		2e-49		31.41%		904		AAA63149.1	
<input checked="" type="checkbox"/> unnamed protein product [Arabidopsis thaliana]						Arabidopsis thali...		179		179		94%		3e-49		31.41%		1306		CAD5325452.1	
<input checked="" type="checkbox"/> unnamed protein product [Arabidopsis thaliana]						Arabidopsis thali...		179		179		94%		3e-49		31.41%		1304		CAD5325453.1	
<input checked="" type="checkbox"/> unnamed protein product [Arabidopsis thaliana]						Arabidopsis thali...		178		178		92%		5e-49		34.87%		837		CAD5325042.1	
<input checked="" type="checkbox"/> unnamed protein product [Arabidopsis thaliana]						Arabidopsis thali...		178		178		92%		5e-49		34.87%		837		VYS59567.1	
<input checked="" type="checkbox"/> unnamed protein product [Arabidopsis thaliana]						Arabidopsis thali...		177		177		96%		1e-48		30.89%		906		CAA0407261.1	
<input checked="" type="checkbox"/> unnamed protein product [Arabidopsis thaliana]						Arabidopsis thali...		177		177		96%		1e-48		31.44%		906		CAA0291745.1	
<input checked="" type="checkbox"/> Disease resistance protein (CC-NBS-LRR class) family [Arabidopsis thaliana]						Arabidopsis thali...		177		177		96%		1e-48		31.44%		906		NP_001319213.1	
<input checked="" type="checkbox"/> hypothetical protein AXX17_AT1G47690 [Arabidopsis thaliana]						Arabidopsis thali...		177		177		96%		1e-48		31.44%		906		OAP11997.1	
<input checked="" type="checkbox"/> unnamed protein product [Arabidopsis thaliana]						Arabidopsis thali...		177		177		96%		1e-48		30.89%		990		CAA0407259.1	
<input checked="" type="checkbox"/> F12M16.25 [Arabidopsis thaliana]						Arabidopsis thali...		177		177		96%		2e-48		31.44%		1584		AAF69538.1	

NCBIのBLASTで外群となる類似配列を検索する

特定の種が持つ遺伝子の中で類似する配列を検索等も可能

より祖先の種に類似配列があれば、外群として利用できる可能性がある

