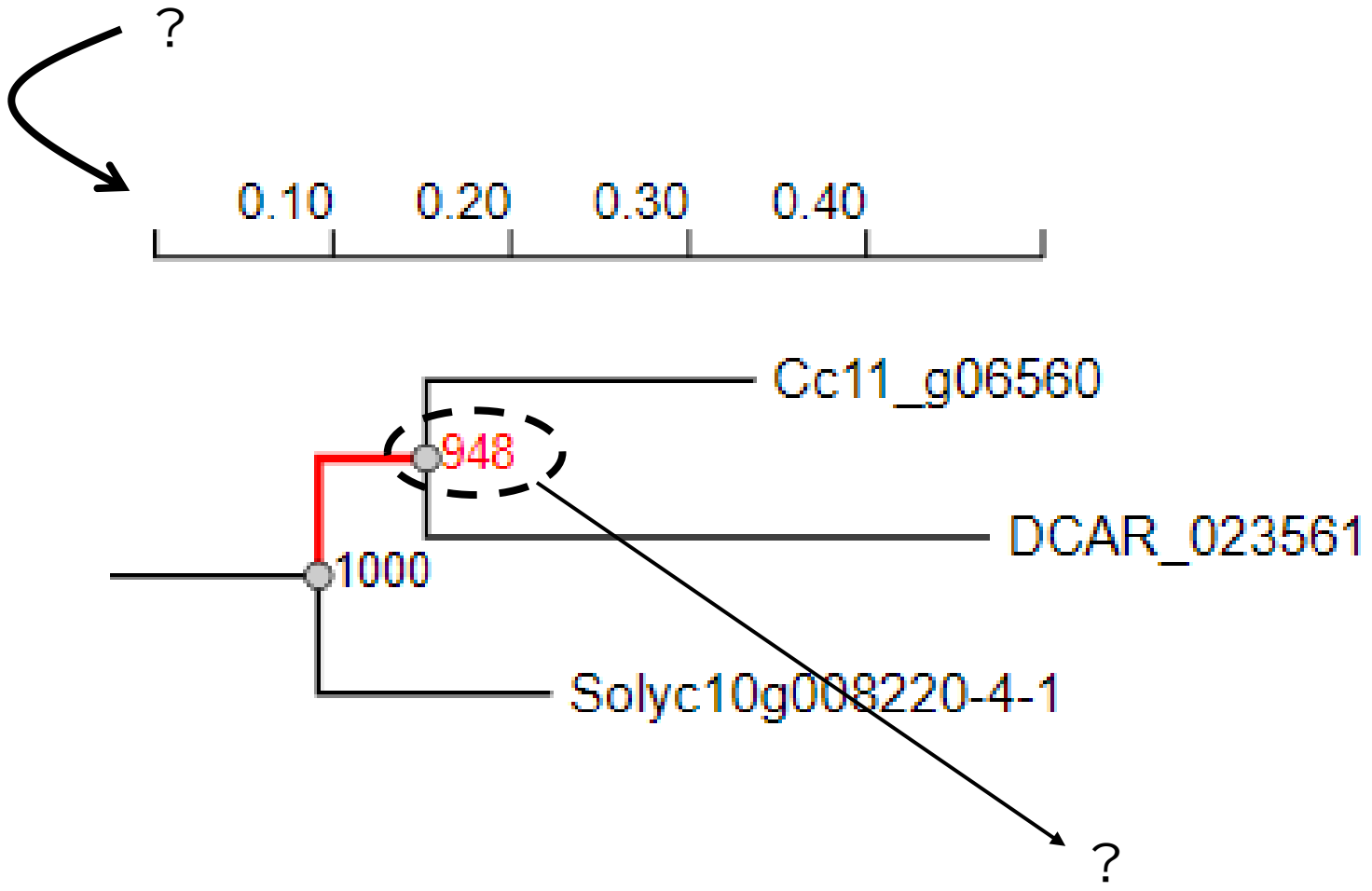
系統樹について
&
NCBIデータベースについて

2025/04/22

栽培植物起原学分野　助教

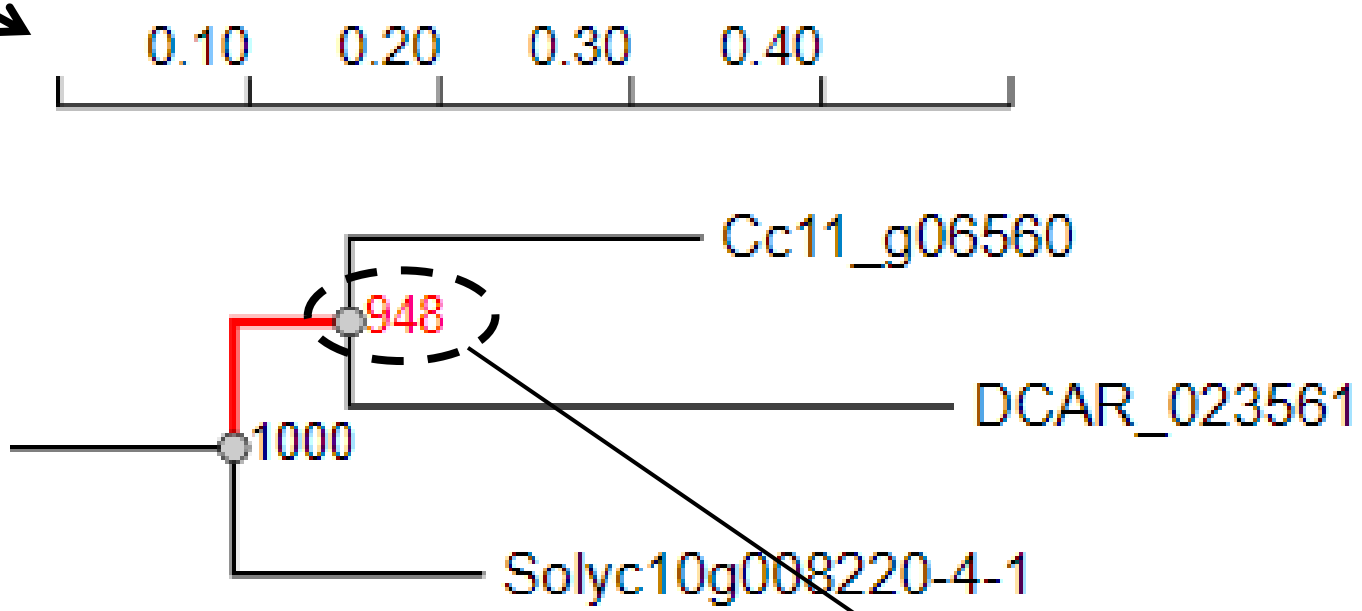
堺　俊之

系統樹の数値の意味


?
0.10
0.20
0.30
0.40
Cc11_g06560
948
DCAR_023561
1000
Solyc10g008220-4-1
?

各数値は系統的な距離や分岐の信頼度を示す

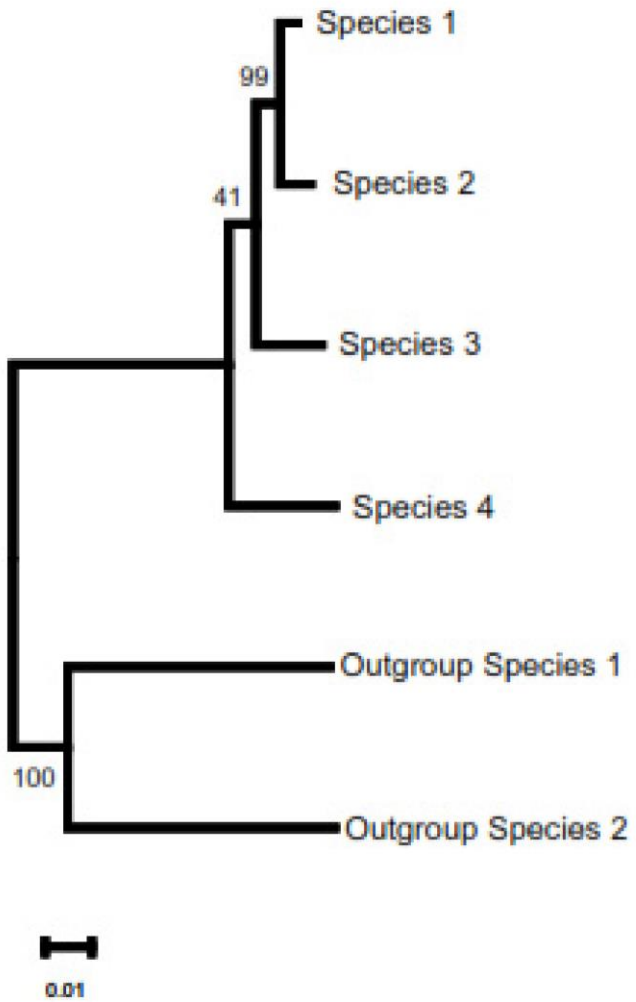
Phylogenetic distance: 系統的な距離、
どれだけ配列に違いがあるかの基準



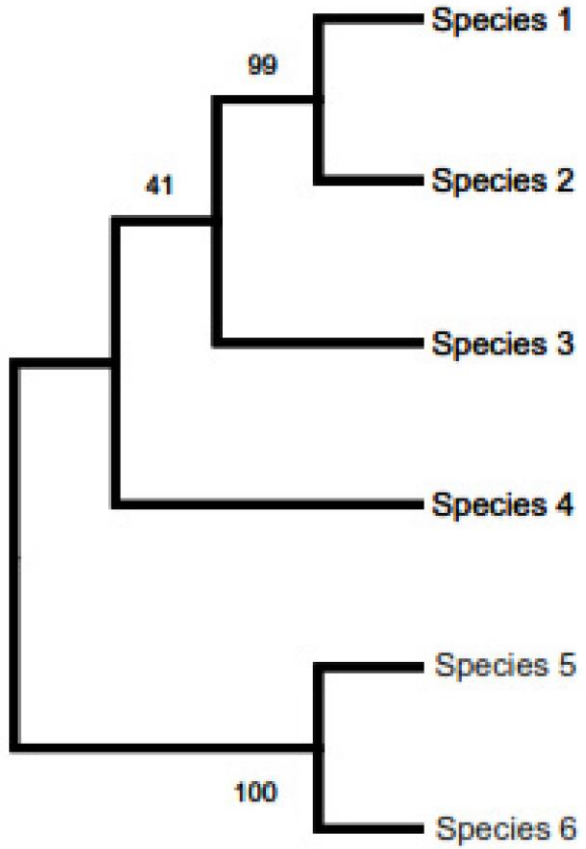
Branch support value
分岐の信頼度。
反復数に近いほど信頼度が高い。
（分岐していると考えて良い。）

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

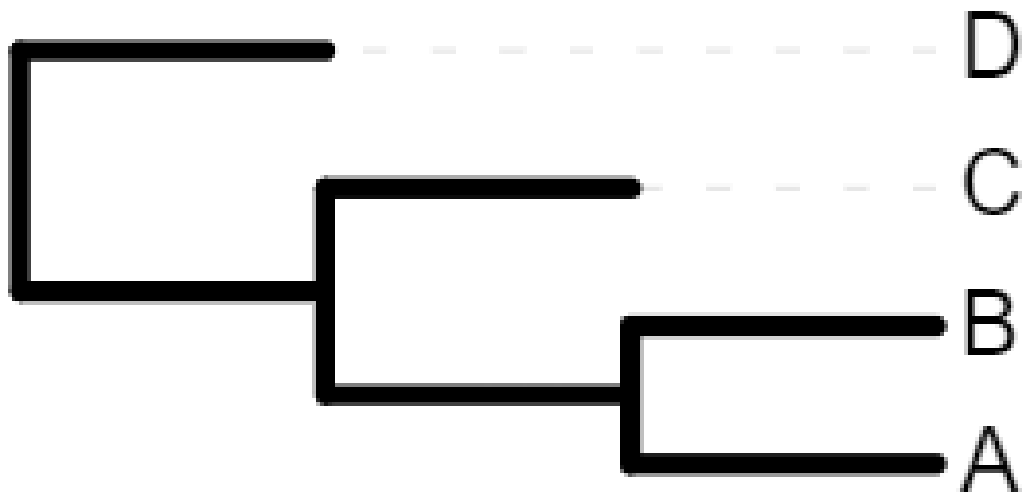


×Cladogram



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…？

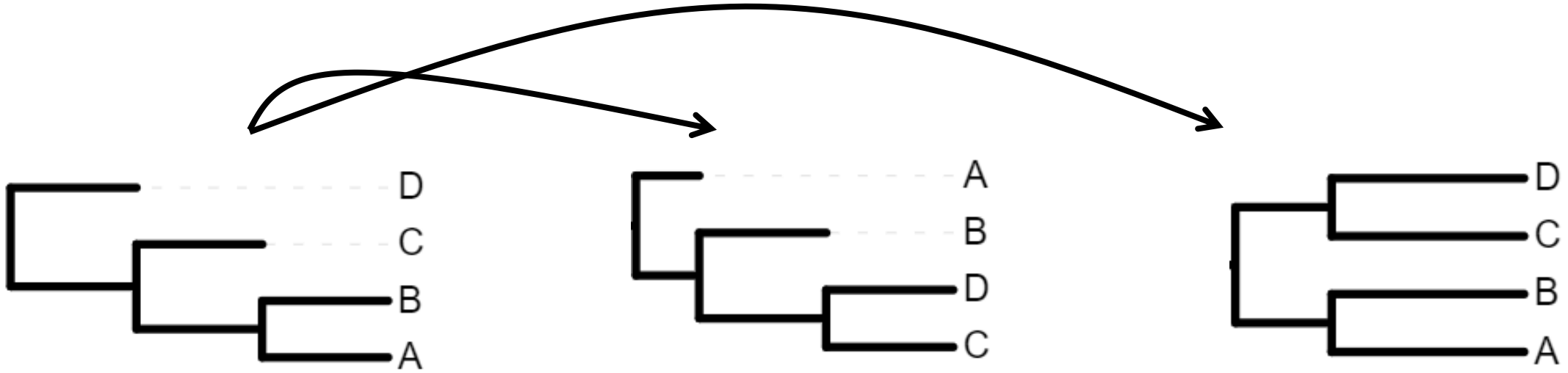


系統樹に基づいた進化的な関係の推定

A. 言えない

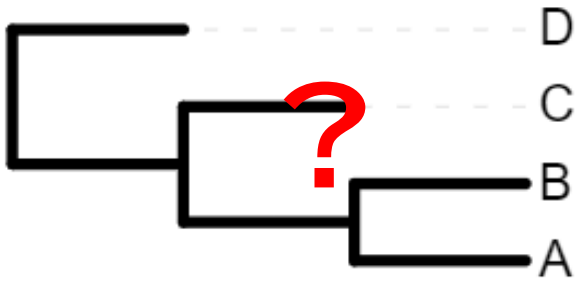
どの遺伝子/種が最初（Rootと呼ぶ）か分からないので、

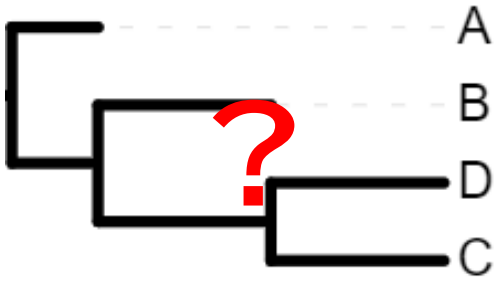
系統樹の関係からだけでは起源や進化的な関係は判断できない

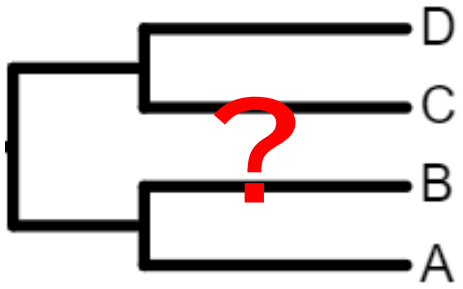


系統樹に基づいた進化的な関係の推定

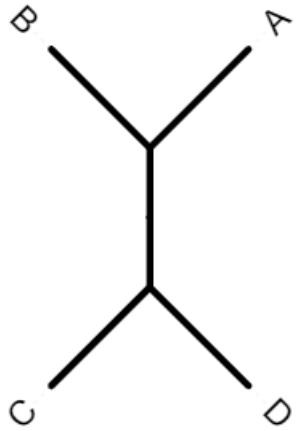
起源や進化的な関係が判断できない場合: その①







系統的な関係性
（系統的に近いか遠いか）
のみ示す



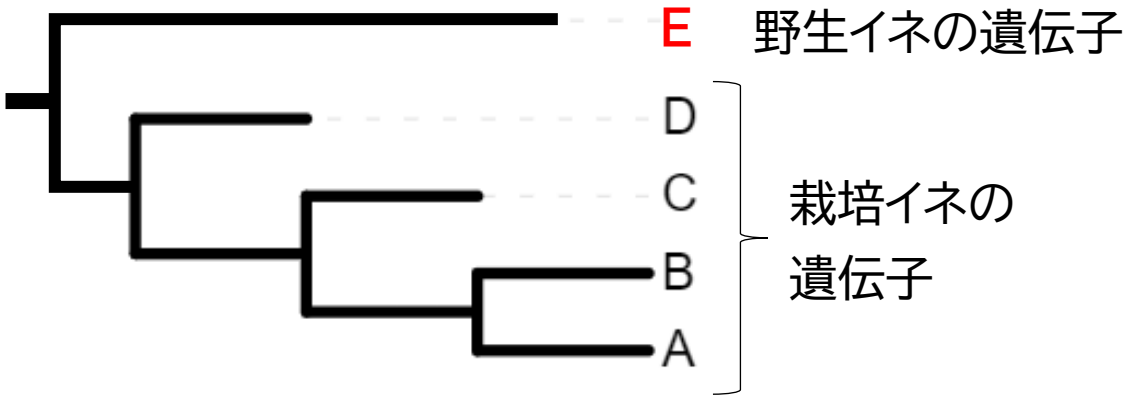
Unrooted tree と呼ぶ

系統樹に基づいた進化的な関係の推定

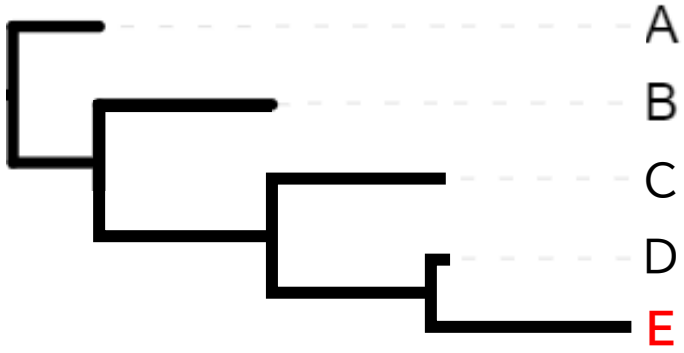
起源や進化的な関係が判断できない場合: その②

Outgroup(外群)を入れRootを決める

○



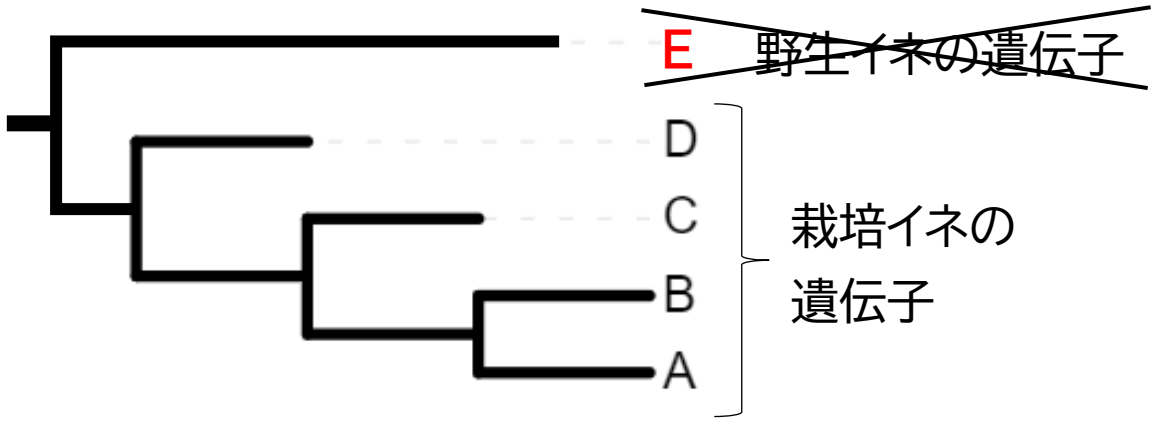
×



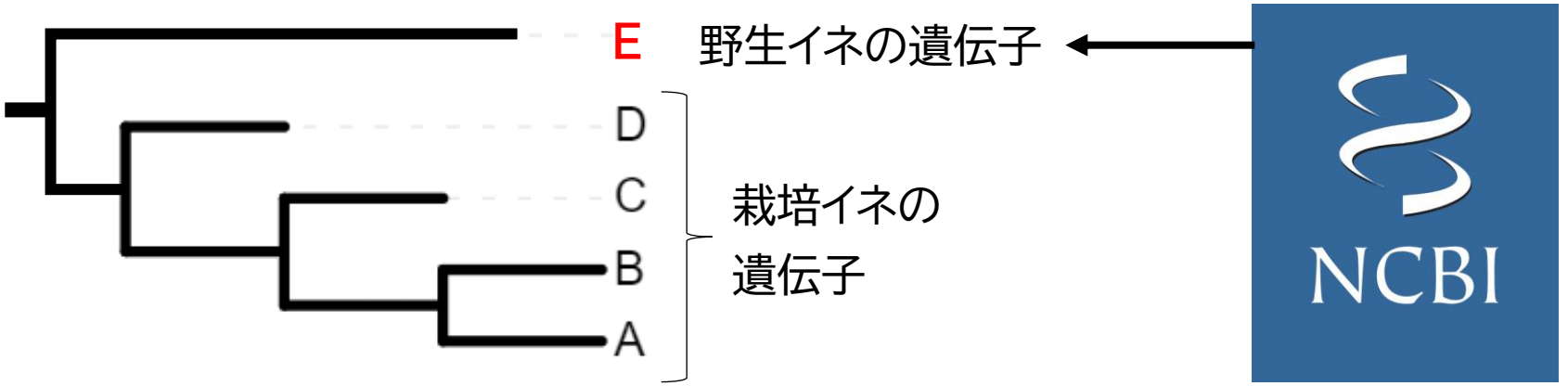
外群の配列をどう手に入れるのか？

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そ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

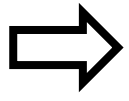
NCBIのBLASTで外群となる類似配列を検索する

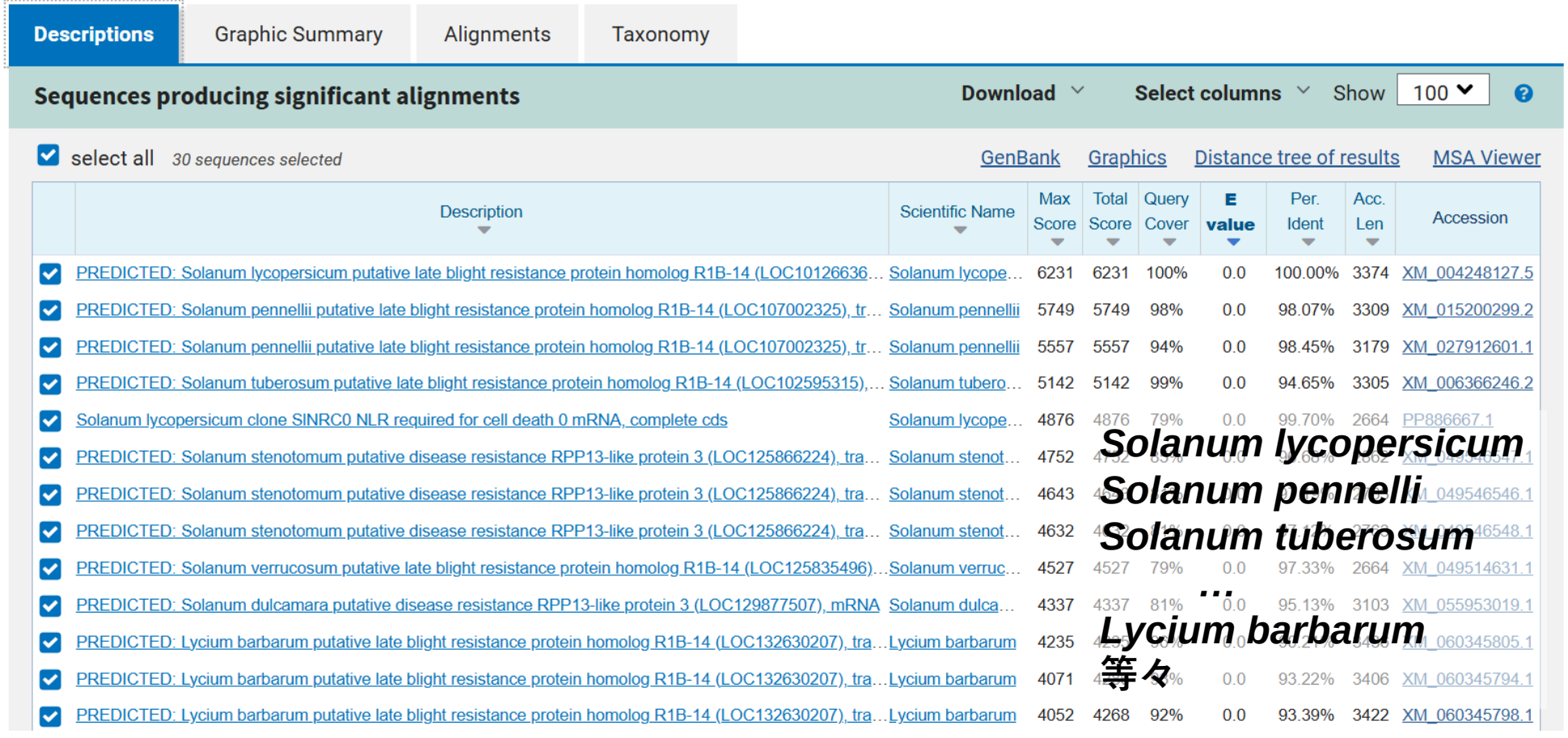
BLAST … NCBIに保存されているデータの中で類似する配列を検索

ex) どんな種が配列Aの類似配列を持っているか？

配列A

GGGAAAAAAACATTTTA
AAAAAATCTATAAAAAAA
AATAGAATTCATAAAAAT
TCTTGCTTTTCCATATAT
TGAGCTTCAATTATTTTT
TTTTTCTCATAAAAAGTT
GCTTTTGGGAAAAAAAA
AAAAAAACTCATTTCTAA
TACCAAACAACTGATTT
CTTCAGGGAGAGTGC…





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

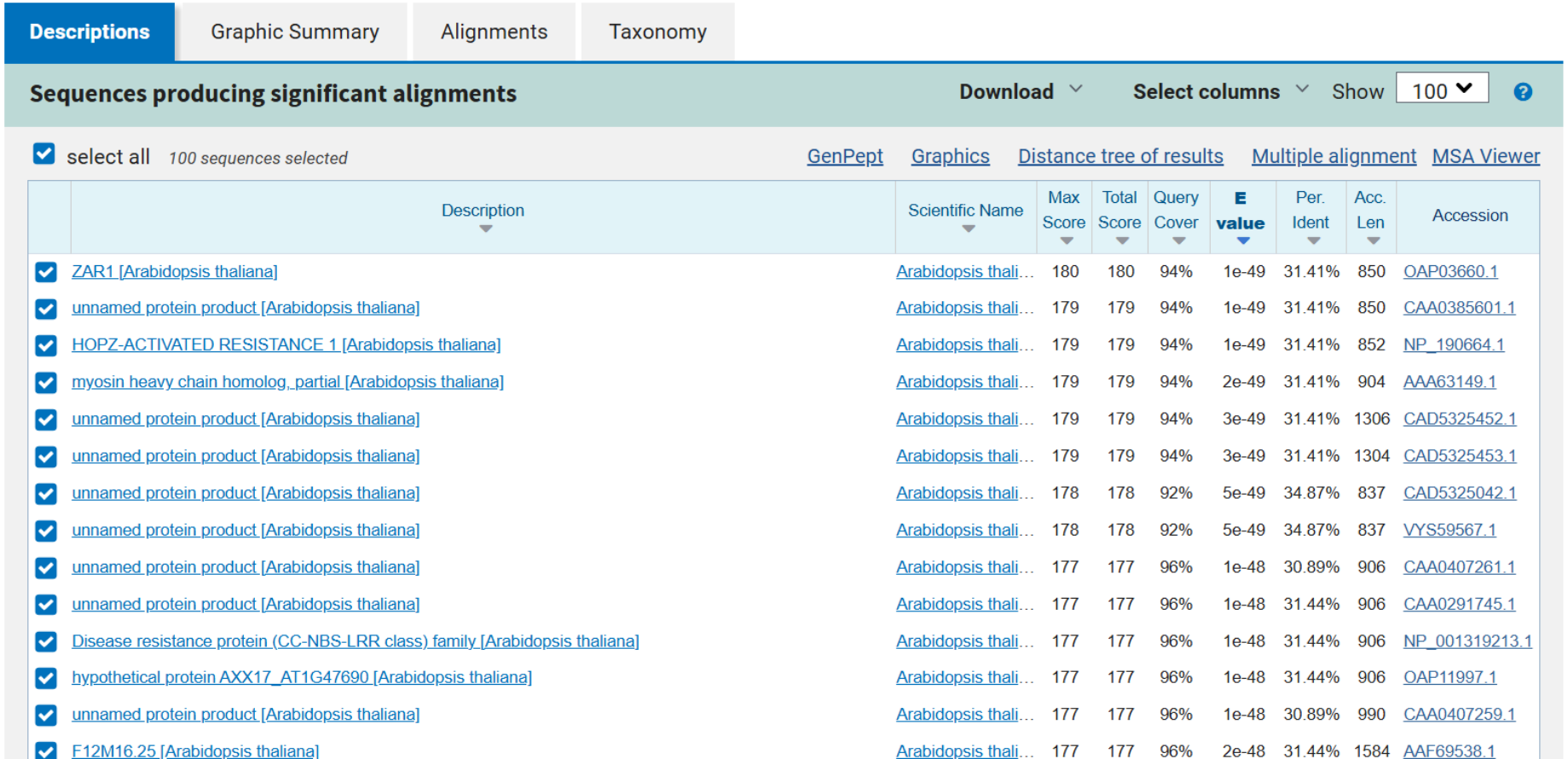
ex) *Arabidopsis thaliana*のゲノム中に似た配列があるか？

配列A

GGGAAAAAAACATTTTA
AAAAAATCTATAAAAAAA
AATAGAATTCATAAAAAT
TCTTGCTTTTCCATATAT
TGAGCTTCAATTATTTTT
TTTTTCTCATAAAAAGTT
GCTTTTGGGAAAAAAAA
AAAAAAACTCATTTCTAA
TACCAAACAACTGATTT
CTTCAGGGAGAGTGC…

⇨

Arabidopsis Thaliana のみ対象



Descriptions | Graphic Summary | Alignments | Taxonomy

Sequences producing significant alignments Download Select columns Show 100

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
ZAR1 [Arabidopsis thaliana]	Arabidopsis thali…	180	180	94%	1e-49	31.41%	850	OAP03660.1
unnamed protein product [Arabidopsis thaliana]	Arabidopsis thali…	179	179	94%	1e-49	31.41%	850	CAA0385601.1
HOPZ-ACTIVATED RESISTANCE 1 [Arabidopsis thaliana]	Arabidopsis thali…	179	179	94%	1e-49	31.41%	852	NP_190664.1
myosin heavy chain homolog, partial [Arabidopsis thaliana]	Arabidopsis thali…	179	179	94%	2e-49	31.41%	904	AAA63149.1
unnamed protein product [Arabidopsis thaliana]	Arabidopsis thali…	179	179	94%	3e-49	31.41%	1306	CAD5325452.1
unnamed protein product [Arabidopsis thaliana]	Arabidopsis thali…	179	179	94%	3e-49	31.41%	1304	CAD5325453.1
unnamed protein product [Arabidopsis thaliana]	Arabidopsis thali…	178	178	92%	5e-49	34.87%	837	CAD5325042.1
unnamed protein product [Arabidopsis thaliana]	Arabidopsis thali…	178	178	92%	5e-49	34.87%	837	VYS59567.1
unnamed protein product [Arabidopsis thaliana]	Arabidopsis thali…	177	177	96%	1e-48	30.89%	906	CAA0407261.1
unnamed protein product [Arabidopsis thaliana]	Arabidopsis thali…	177	177	96%	1e-48	31.44%	906	CAA0291745.1
Disease resistance protein (CC-NBS-LRR class) family [Arabidopsis thaliana]	Arabidopsis thali…	177	177	96%	1e-48	31.44%	906	NP_001319213.1
hypothetical protein AXX17_AT1G47690 [Arabidopsis thaliana]	Arabidopsis thali…	177	177	96%	1e-48	31.44%	906	OAP11997.1
unnamed protein product [Arabidopsis thaliana]	Arabidopsis thali…	177	177	96%	1e-48	30.89%	990	CAA0407259.1
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

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

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