➤ **生物情報工学（バイオインフォマティクス）ってなに？**

分子生物学研究の結果、大量のDNA配列データが蓄積されてきた。

Gene A: ATGAAGTTTAGACCGATCAGTACT….
Gene B: ATGCAGTACGAAGTACGATAGACAAGT…

新奇遺伝子が見つかった。さて、これは既知のものと同じ機能か否か？

Novel Gene: ATGCAGTACTAAGTTTAGACCGAT….

さらに、ゲノムプロジェクトの進展で大量のDNA配列データが手に入った。
さて、配列が分かってもどこが遺伝子か分からない…。

じゃあ、過去の知見をもとに遺伝子領域を予測しよう！
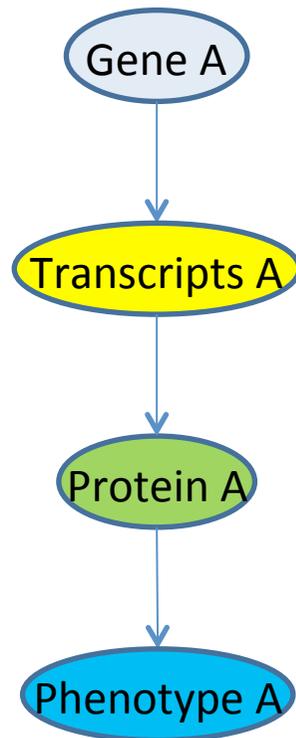BLAST、ORF finder、Softberry/FGENESH

DNAシーケンサーの発達により、ゲノム配列情報がさらに充実。
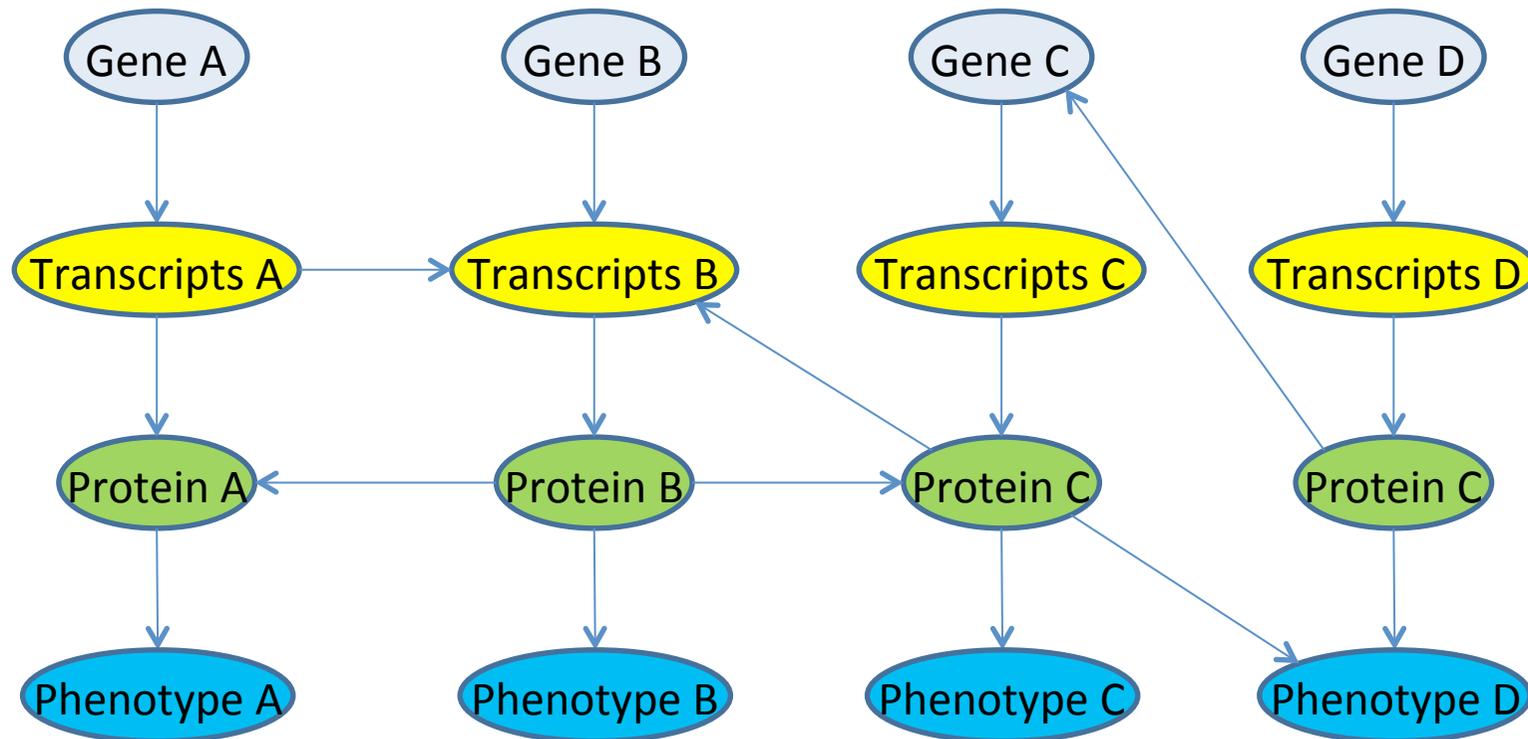ゲノムを種間比較して、進化学的な研究をしよう⇒比較ゲノミクス

➢ 生物情報工学（バイオインフォマティクス）ってなに？

古典的な分子生物学

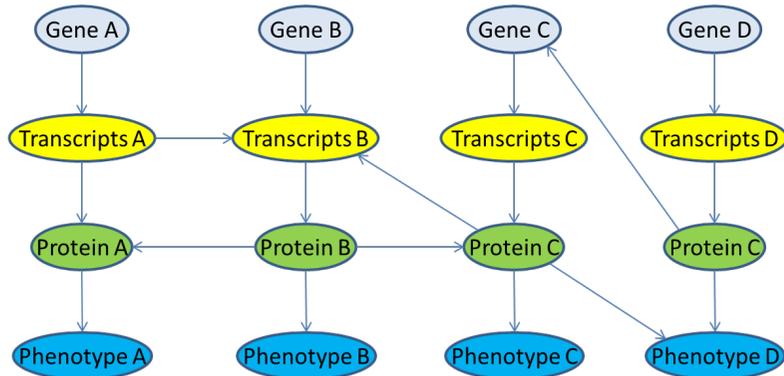# ちょっと整理しておきたいのですが...

➢ 生物情報工学（バイオインフォマティクス）ってなに？

分子生物学の発展

ゲノム、転写産物、たんぱく質、表現型の各階層内・階層間での複雑な相互作用

一つ一つの要素だけに注目していても、生命現象の本質的な理解には至らない？
木を見て森を見ず？

ちょっと整理しておきたいのですが...

➢ 生物情報工学（バイオインフォマティクス）ってなに？

分子生物学の発展

ゲノム、転写産物、たんぱく質、表現型の各階層内・階層間での複雑な相互作用

オミクス(Omics)
- Genomics （ゲノム）
- Transcriptomics （転写産物）
- Proteomics （タンパク質）
- Metabolomics （代謝産物）
- Glycomics （糖鎖付加）
- Epigenomics （ゲノム修飾）
- Phenomics （表現型）

複雑な相互作用ネットワークから成る生命システムを分解せず複雑なネットワークのまま理解する。

データを数理モデルやネットワークモデルに落とし込んで、生命システムを理解したい。
　⇒　システムバイオロジー
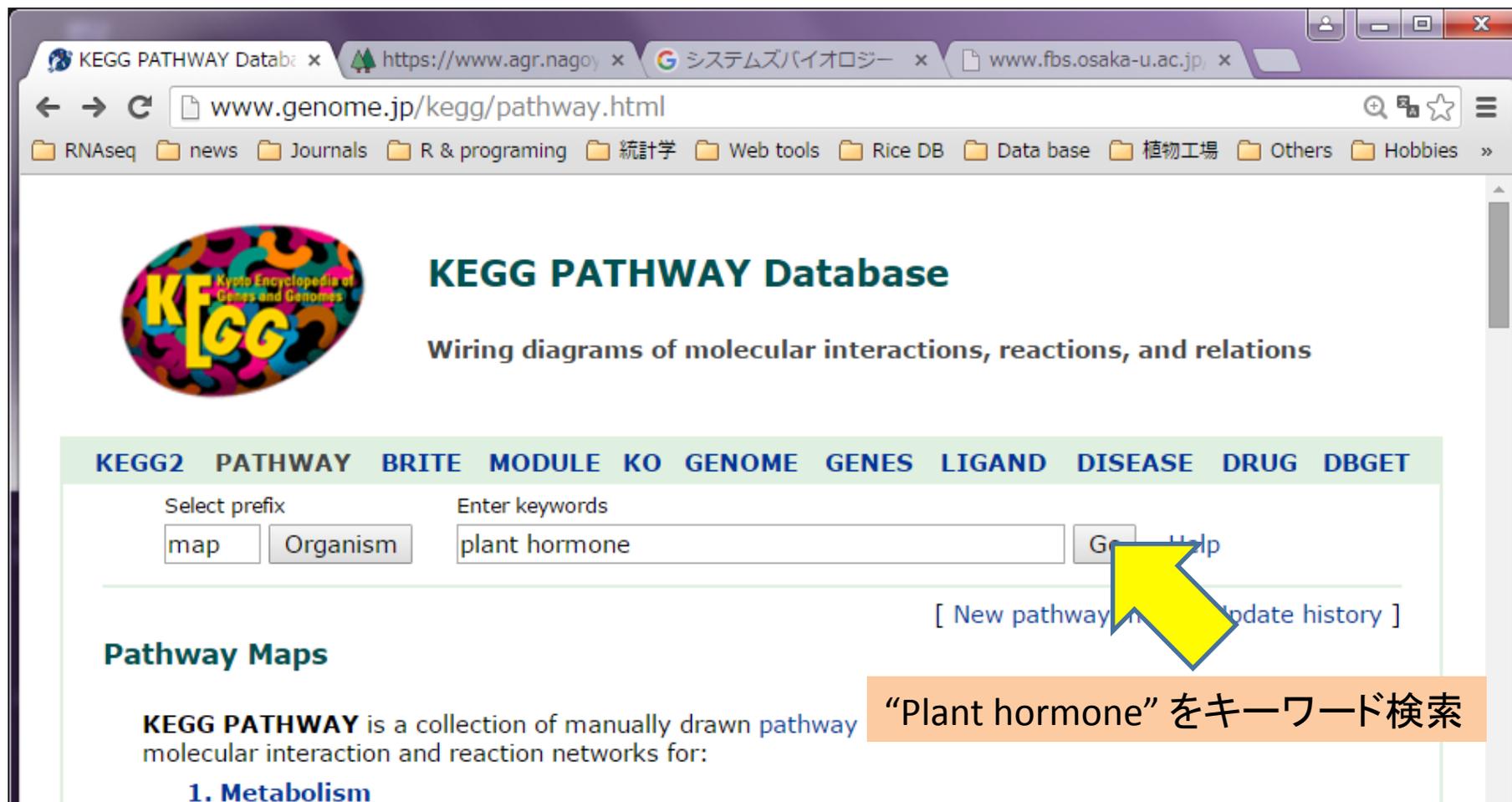
システムバイオロジーの申し子「KEGG」

Googleで"KEGG"を検索

システムバイオロジーの申し子「KEGG」

KEGG

京都大学バイオインフォマティクスセンターが管理しているバイオインフォマティクス研究用データベース。
システムバイオジー的なコンセプトを基にしており、遺伝子間、分子間の相互作用ネットワークに関する情報をデータベース化してある。

KEGGのコアデータベースKEGG PATHWAY

"Plant hormone" をキーワード検索

KEGGで植物ホルモンシグナル伝達系を俯瞰する

"Plant hormone signal transduction" のthumbnail imageをクリック

イネにおける植物ホルモンシグナル伝達系遺伝子

"Reference pathway"のプルダウンメニューから
"Oryza sativa japonica (Japanese rice)(RAPDB)"を
選択し、"GO"をクリック
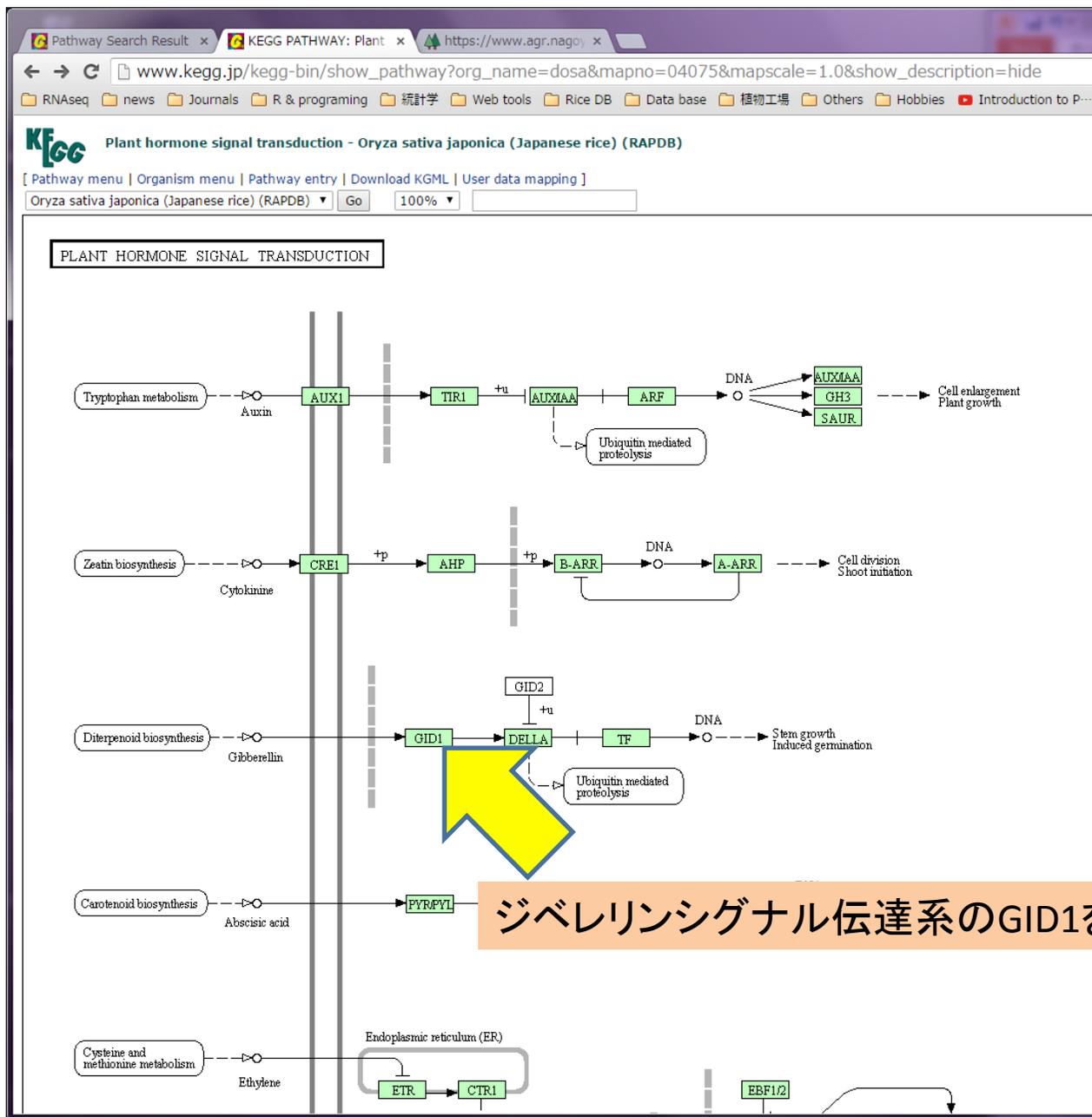
イネのジベレリン受容体

ジベレリンシグナル伝達系のGID1をクリック

10

ジベレリン受容体配列を用いてモチーフサーチ

アミノ酸配列をコピー

ジベレリン受容体配列を用いてモチーフサーチ

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

"MOTIF"をクリック

解析用配列

1. ある遺伝子AのgenomeDNA配列

ジベレリン受容体配列を用いてモチーフサーチ

アミノ酸配列をペーストして、"Compute"をクリック

ジベレリン受容体配列を用いてモチーフサーチ

KEGG からモチーフサーチ結果への直接リンク

"Motif"ボタンをクリック

KEGG からモチーフサーチ結果への直接リンク
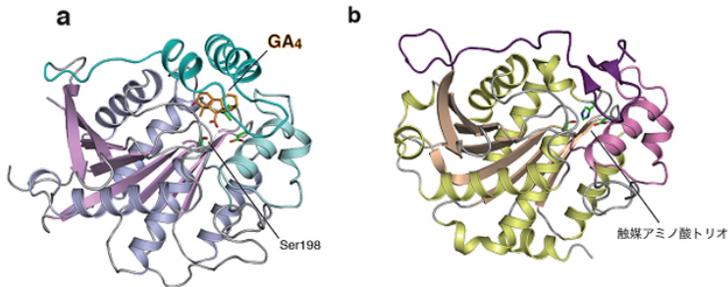
ジベレリン受容体GID1は、加水分解酵素！？

16

図1　リボンモデルで表示したジベレリン受容体とリパーゼの立体構造の比較
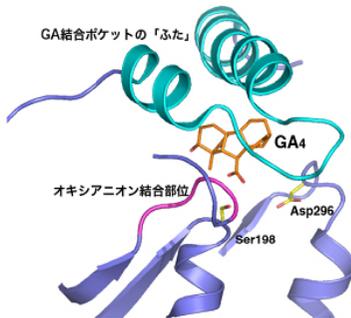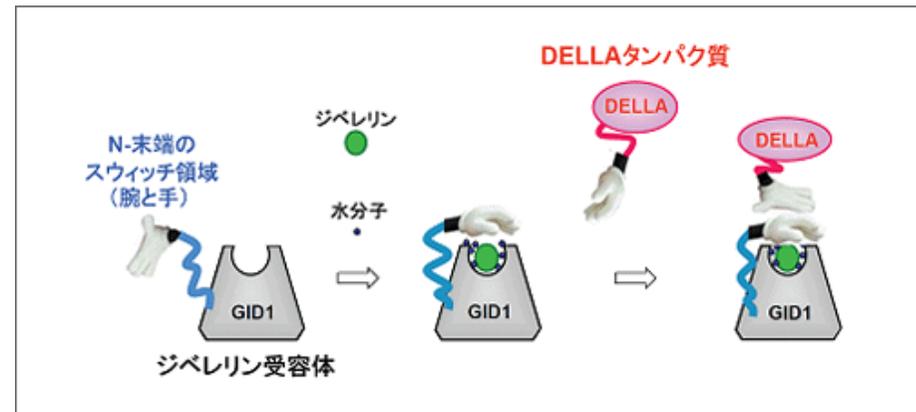a ジベレリン受容体、b リパーゼ。両者の基本構造は非常によく似ている。a、bにおいてそれぞれ薄紫色、金色で示したのがα/β水解酵素型構造。a、bにおいてそれぞれ濃緑色、濃紫色で示したのがN末端のリッド（ふた）。

図2　ジベレリン受容体のジベレリン（GA）結合ポケットの拡大図
活性型ジベレリン(GA4)は炭素原子をオレンジ、酸素原子を赤で表示した。リパーゼ触媒アミノ酸トリオに相当するSer198とAsp296は、炭素原子を黄色、酸素原子を赤で表示した。オキシアニオン結合部位は紫色、ジベレリン結合ポケットの「ふた」は水色で表示した。

http://www.kyoto-u.ac.jp/static/ja/news_data/h/h1/news6/2008/081127_1.htm

ジベレリンを分解することなく、結合による構造変化でシグナルを伝達。

脂質加水分解酵素リパーゼの基質結合部位に変異が起こり、ジベレリンと結合できるようになった。

17

バイオインフォマティクスの真骨頂：比較ゲノミクス

Googleで"Phytozome"を検索

Phytozome
数十の植物種のゲノム情報を統合し、そのゲノム情報を比較する目的に特化したデータベース。

植物種間横断的にBLAST解析を行ったり、とある遺伝子のホモログやパラログを容易に検索できる。

通常のNCBIのBLASTなどに比べて、検索される情報が整理されているので扱いが容易。

Phytozomeでお手軽比較ゲノミクス的解析

"Tools"プルダウンメニューから"BLAST"を選択

Phytozomeでお手軽比較ゲノミクス的解析

Phytozomeでお手軽比較ゲノミクス的解析

一番上の"G"ボタンをクリック

Phytozomeでお手軽比較ゲノミクス的解析

"Homolog filter"をクリック

Phytozomeでお手軽比較ゲノミクス的解析

Embryophyte
Physcomitrella patens v3.0
Tracheophyte
Selaginella moellendorffii v1.0
Grass
Brachypodium distachyon v2.1
Oryza sativa v7.0
Zea mays 6a
Pentapetalae
Solanum tuberosum v3.4
Malvidae
Populus trichocarpa v3.0
Brassicaceae
Arabidopsis thaliana TAIR10
Citrus
Citrus sinensis v1.1
Fabidae
Glycine max Wm82.a2.v1

23

Phytozomeでお手軽比較ゲノミクス的解析

"Protein Homologs"タブをクリック

Phytozomeでお手軽比較ゲノミクス的解析

Phytozomeでお手軽比較ゲノミクス的解析

Phytozomeでお手軽比較ゲノミクス的解析

"Download"ボタンをクリック

Phytozomeでお手軽比較ゲノミクス的解析

ダウンロードされたファイルを開く

Phytozomeでお手軽比較ゲノミクス的解析

ダウンロードした配列データを
GenomeNetのClustalWでmultiple alignment

Phytozomeでお手軽比較ゲノミクス的解析

CLUSTALW Result

[clustalw.aln][clustalw.dnd][readme]

| Select tree menu ▼ | Exec |

Select tree menu
Rooted phylogenetic tree (UPGMA)
Rooted phylogenetic tree with branch length (UPGMA)
Unrooted phylogenetic tree (N-J)
Unrooted phylogenetic tree with branch length (N-J)

"Rooted phylogenetic tree (UPGMA)"を選択し、
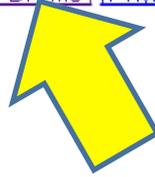"Exec"をクリック
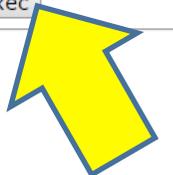
```
Sequence type explicitly set to Protein
Sequence format is Pearson
Sequence 1: 271143            371 aa
Sequence 2: 97885             355 aa
Sequence 3: AT3G05120         345 aa
Sequence 4: AT3G63010         358 aa
Sequence 5: AT5G27320         344 aa
Sequence 6: Bradi2g25600      355 aa
Sequence 7: Glyma.02G151100   342 aa
Sequence 8: Glyma.03G148300   346 aa
Sequence 9: Glyma.10G022900   343 aa
Sequence 10: Glyma.10G158000   344 aa
Sequence 11: Glyma.20G230600   344 aa
Sequence 12: GRMZM2G173630     351 aa
Sequence 13: LOC_Os05g33730    354 aa
Sequence 14: orange1.1g019193m.g   344 aa
Sequence 15: orange1.1g019235m.g   344 aa
Sequence 16: PGSC0003DMG400003849  345 aa
Sequence 17: PGSC0003DMG400021991  345 aa
Sequence 18: PGSC0003DMG400028559  348 aa
Sequence 19: Potri.002G213100  344 aa
Sequence 20: Potri.005G040600  344 aa
Sequence 21: Potri.013G028700  344 aa
Sequence 22: Potri.014G135900  346 aa
Start of Pairwise alignments
Aligning...


Sequences (1:2) Aligned. Score: 56.0563
Sequences (1:3) Aligned. Score: 47.8261
Sequences (1:4) Aligned. Score: 42.7374
Sequences (1:5) Aligned. Score: 50
Sequences (1:6) Aligned. Score: 43.0986
Sequences (1:7) Aligned. Score: 46.1988
Sequences (1:8) Aligned. Score: 46.2428
Sequences (1:9) Aligned. Score: 46.3557
Sequences (1:10) Aligned. Score: 48.2558
Sequences (1:11) Aligned. Score: 49.1279
Sequences (1:12) Aligned. Score: 44.4444
Sequences (1:13) Aligned. Score: 44.6328
Sequences (1:14) Aligned. Score: 49.1279
Sequences (1:15) Aligned. Score: 45.9302
Sequences (1:16) Aligned. Score: 46.9565
Sequences (1:17) Aligned. Score: 46.9565
Sequences (1:18) Aligned. Score: 47.1264
```
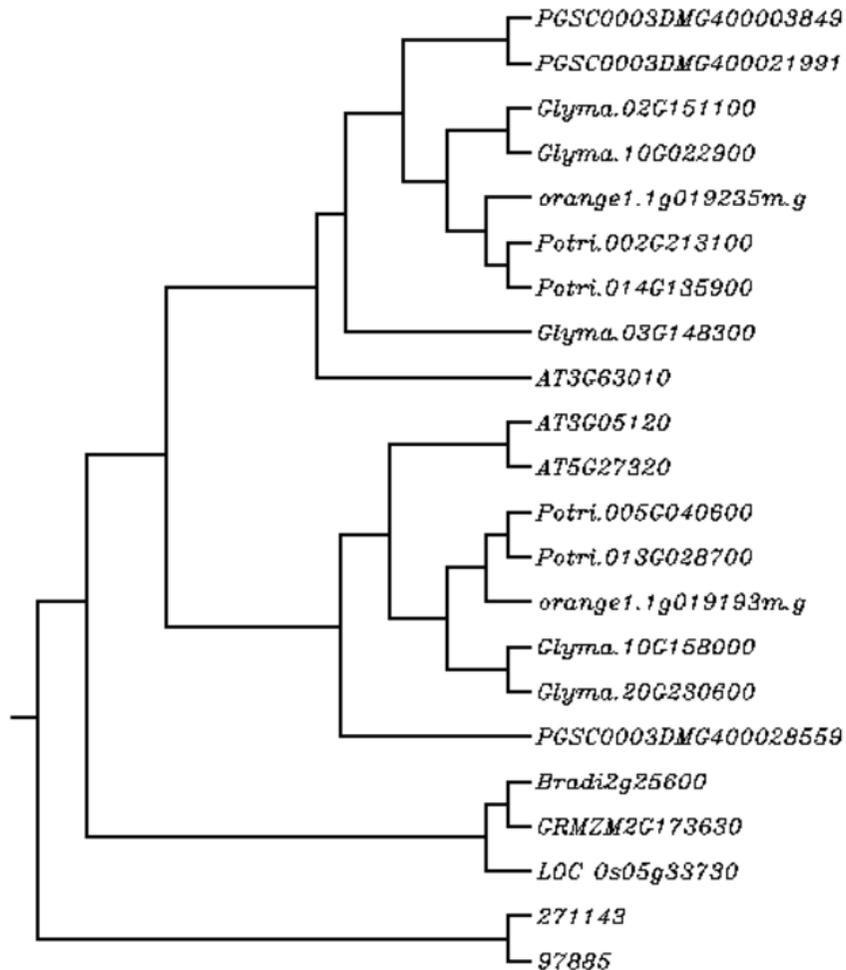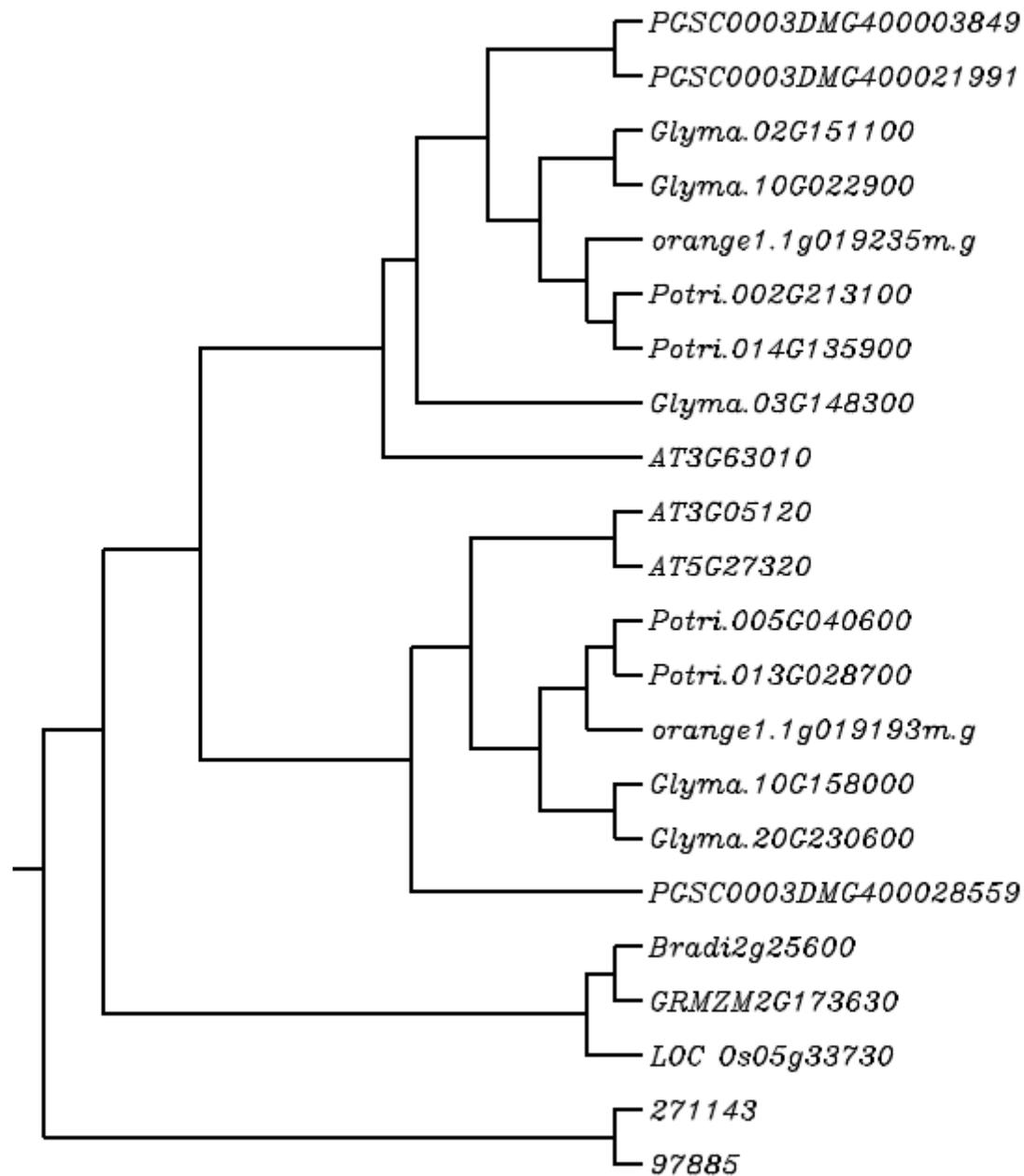
Phytozomeでお手軽比較ゲノミクス的解析

Phytozomeでお手軽比較ゲノミクス的解析

PGSC0003DMG400003849	ジャガイモ
PGSC0003DMG400021991	ジャガイモ
Glyma.02G151100	ダイズ
Glyma.10G022900	ダイズ
orange1.1g019235m.g	オレンジ
Potri.002G213100	ポプラ
Potri.014G135900	ポプラ
Glyma.03G148300	ダイズ
AT3G63010	シロイヌナズナ
AT3G05120	シロイヌナズナ
AT5G27320	シロイヌナズナ
Potri.005G040600	ポプラ
Potri.013G028700	ポプラ
orange1.1g019193m.g	オレンジ
Glyma.10G158000	ダイズ
Glyma.20G230600	ダイズ
PGSC0003DMG400028559	ジャガイモ
Bradi2g25600	ミナトカモジグサ
GRMZM2G173630	トウモロコシ
LOC_Os05g33730	イネ
271143	シダ
97885	シダ

Phytozomeでお手軽比較ゲノミクス的解析

Phytozomeでお手軽比較ゲノミクス的解析

"Species display"をクリック

Phytozomeでお手軽比較ゲノミクス的解析

Phytozomeでお手軽比較ゲノミクス的解析

Synteny欄に、各ゲノム間でのシンテニー（染色体上での遺伝子の並び順の類似性）が表示される