A-21

SALAD database (http://salad.dna.affrc.go.jp/salad/)

～アミノ酸配列の類似性を基準にした比較ゲノムデータベース～

○三原基広、伊藤剛、井澤毅 （(独)農業生物資源研究所）

要旨

我々は、植物ゲノムアノテーション情報(アミノ酸配列)の比較解析を効率的に行うためのデータベース、SALAD database を作成した。SALAD databaseは、ゲノム内・ゲノム間で進化上保存された様々なアミノ酸配列(モチーフ)を期待値最大化法(MEME)で抽出し、その抽出モチーフの類似性・組み合わせによりアノテーションを分類した比較ゲノムデータベースである(特許出願済)。任意のモチーフの組み合わせからアライメントを作成し、NJ樹(アミノ酸、核酸ともに)も作成可能である。クレードごとに、アミノ酸配列のロゴを作成し、アミノ酸座位の保存性も比較できる。現在のデータセットは7つの生物種(イネ、ソルガム、シロイヌナズナ、ヒメツリガネゴケ、紅藻、緑藻、酵母)の全ゲノムアノテーション情報を内包している。これらを利用すれば種を超えて似た機能を持つ遺伝子を推定することもでき、シロイヌナズナ等からの先行知見を、他の植物・作物の研究に効率的に反映させられる。また、この類似性クラスタリングとマイクロアレイデータを組み合わせた新しいViewer(SALAD on ARRAYs)を開発しており(特許出願済)、マイクロアレイデータも公開している。これを使えばアレイデータ上のパラログ間の発現比較が簡単になる。

SALADの作り方

Blastサーチによりホモログスなアテーショングループを作る

MEME (http://meme.sdsc.edu/meme/intro.html)を使ってグループごとにモチーフ(8-50 a.a.)を抽出する

モチーフの有無・類似度によるクラスタリングを行う(特開2008-225689)

SALAD databaseとは？

●SALAD database (**S**urveyed conserved motif **AL**ignment diagram and the **A**ssociating **D**endrogram)

●進化的に保存されたアミノ酸配列(モチーフ)パターンに基づいて、植物種のタンパク質アノテーションを分類した比較ゲノムデータベース

●ブラウザ上で操作可能なwebアプリケーションにより、グラフィカルに結果を確認できる

●ver.2.0では、7つの生物種(イネ、ソルガム、シロイヌナズナ、ヒメツリガネゴケ、紅藻、緑藻、酵母)の比較が行える

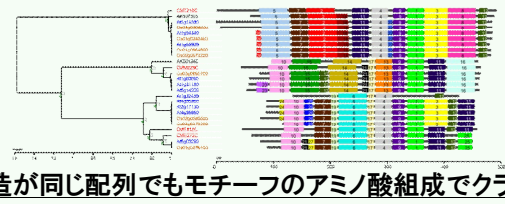
Ver. 2.0 (7 species)

Species	Annotated genes	Database	Version
Oryza sativa	52214	RAP-DB	RAP2
Sorghum bicolor	34496	JGI	Sbi1_4
Arabidopsis thaliana	27025	TAIR	TAIR8
Physcomitrella patens	35938	JGI	v1.1
Ostreococcus tauri	7725	JGI	v2.0
Cyanidioschyzon merolae	4773	Genome Project	2008/6/4
Saccharomyces cerevisiae	6698	SGD	SGD1.01.50
	Total 168869		

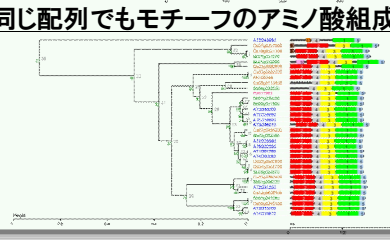
クラスタリングの特徴

SALAD databaseでは、モチーフの有無・類似度によるクラスタリングを行うためアミノ酸置換マトリクスをスコア計算に利用したクラスタリング方法を開発した(特開2008-225689)。そのクラスタリング方法には次のような特徴がある。

1)モチーフ構造が似ている配列をクラスタリングする

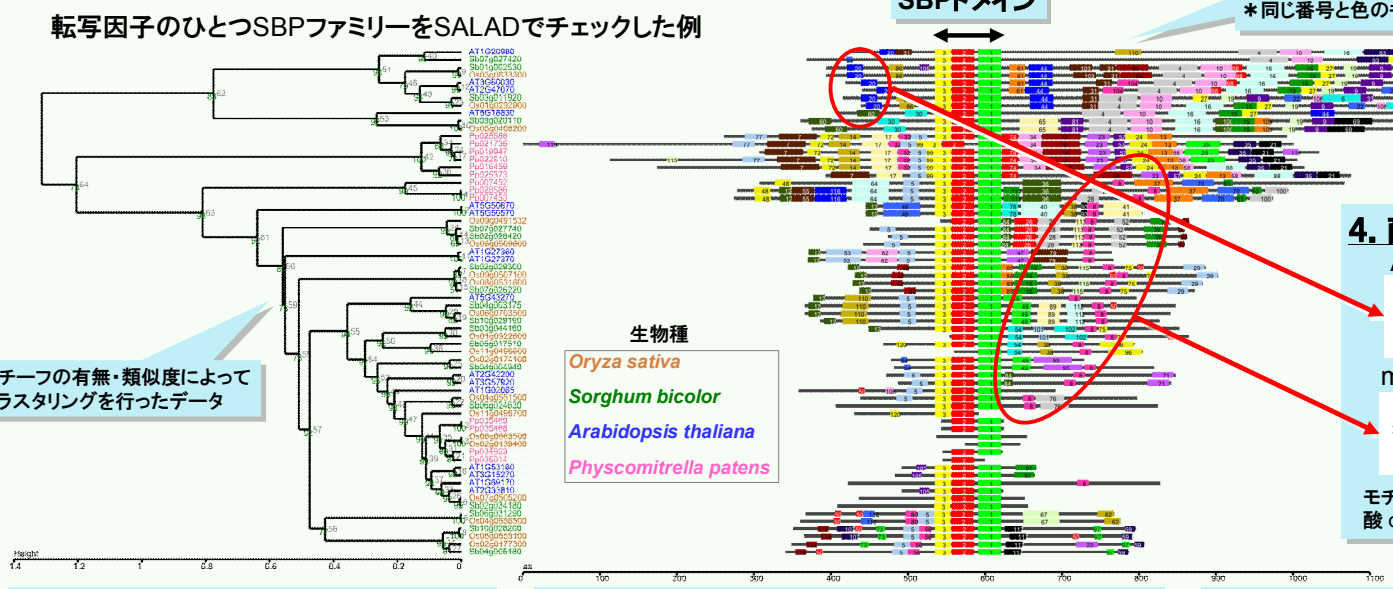


2)モチーフ構造が同じ配列でもモチーフのアミノ酸組成でクラスタリングできる



SALAD databaseのデータと機能

転写因子のひとつSBPファミリーをSALADでチェックした例

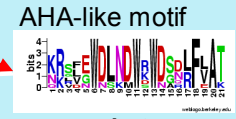


モチーフの位置をグラフィカルに表示することで、構造が似た遺伝子が一目瞭然である
＊同じ番号と色のものは、同じモチーフ

モチーフの有無・類似度によってクラスタリングを行ったデータ

4. 配列ロゴ作成機能

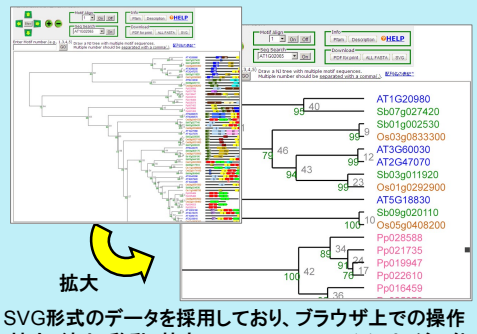
AHA-like motif



motif8にあるmiR156結合領域

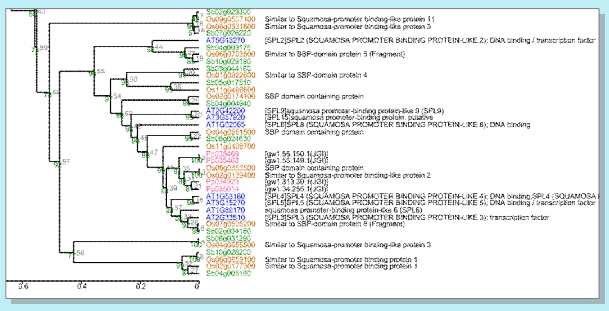
モチーフ配列で構築した進化系統樹のクレードごとに、アミノ酸 or 核酸の保存部位を示す配列ロゴの作成ができる

1. グラフィックデータの操作



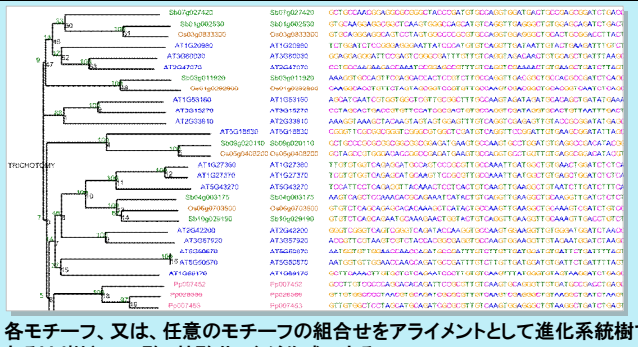
SVG形式のデータを採用しており、ブラウザ上での操作(拡大・縮小・移動・特定のモチーフでアライメント)が可能

2. アノテーション情報の表示



遺伝子情報を表示することが可能なので、モチーフ構造の近い遺伝子から機能類推も可能

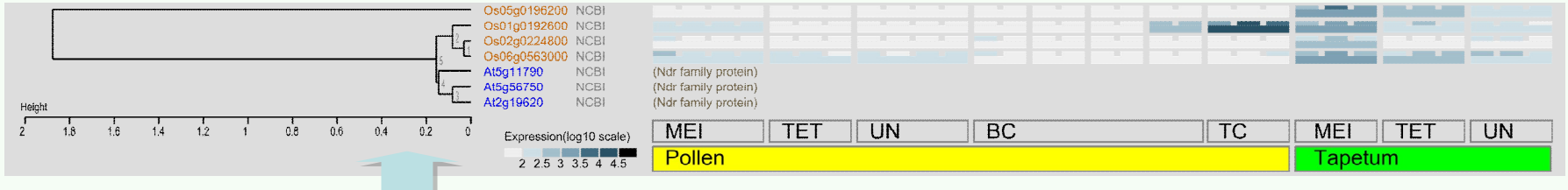
3. モチーフ配列で進化系統樹の作成（アミノ酸、核酸）



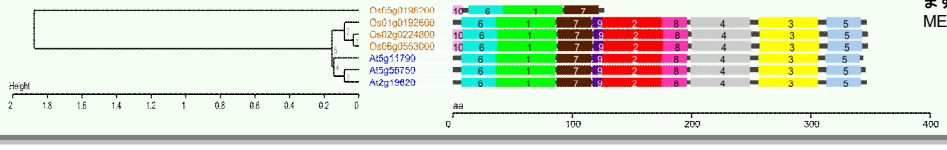
各モチーフ、又は、任意のモチーフの組合せをアライメントとして進化系統樹であるNJ樹(アミノ酸、核酸共に)が作成できる

マイクロアレイの新規ビュワー ～発現量をグラデーション表示～

SALAD databaseのクラスタリングとマイクロアレイのデータを組み合わせた新しいViewer（SALAD on ARRAYs）を開発し公開している(特願2008-252353)。



上記と同じデータのモチーフ構造をチェックしてみると



現在、SALAD on ARRAYs(Mihara et al. 2008)では、"LM microarray analysis of pollen development in rice" のデータが利用できます(Suwabe et al. 2008; Hobo et al. 2008, Hirano et al. 2008)。
MEI: meiotic phase, TET: tetrad phase, UN: unicellular microspore, BC: bicellular, TC: mature tricellular pollen

"SALAD on ARRAYs"では、モチーフの類似性によってクラスタリングされた遺伝子の発現量が色の濃淡で示されている。これを使えば、パラログなものや関係する遺伝子間で特異的に発現しているものが簡単に分かる。

参考

発表論文 Mihara, M, Itoh, T, and Izawa, T. (2008) *In silico* identification of short nucleotide sequences associated with gene expression of pollen development in rice. Plant Cell Physiol. 49:1451-1464.

関連特許 遺伝子クラスタリング装置、遺伝子クラスタリング方法およびプログラム (特開2008-225689)
遺伝子クラスタリング装置およびプログラム (特願2008-252353)

問合わせ先: 029-838-7446、mihara@ml.affrc.go.jp