# SALAD database

## -美味しくSALADを食べるには？-

○三原基広、伊藤剛、井澤毅　（(独) 農業生物資源研究所）

http://salad.dna.affrc.go.jp/salad/

## 要旨

ゲノム情報は膨大で、種間で同じ機能を持つと推測されるタンパク質をコンピューター上で整理整頓することは、非常に手間がかかります。そこで、我々はゲノム情報からのプロテオーム情報がそろっている植物種を対象に、推定アミノ酸配列の類似性を指標にした比較ゲノムデータベースSALAD databaseを開発してきました。今回は、(1) イネ・シロイヌナズナを含む10種の植物種の比較解析が簡単に行えるSALAD database(Ver.3)、(2) 類似遺伝子間の発現比較が一画面で簡単にできるという、他には無い特徴をもつマイクロアレイビュワーSALAD on ARRAYs、(3) ユーザの興味ある配列を使って幅広い生物種 (phytozome等のデータベースを利用して17種以上の生物種を利用可能) から、相同性のある50個のアミノ酸配列セットをBLAST検索の結果から自動生成し、SALADと同じ比較解析ができるBlasting SALAD 解析について紹介する。

## SALAD database

**進化的に保存されたモチーフパターンに基づいて、植物種の蛋白質アノテーションを分類したデータベース**

SALAD databaseのデータセット (ver. 3.0)

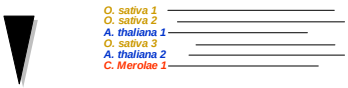
| Species | Genome size | From | #seq |
|---|---|---|---|
| *Oryza sativa* (representative) | 372Mb | RAP2[rep] | 30192 |
| (prediction) | | RAP2[prediction] | 22022 |
| *Sorghum bicolor* | 770Mb | JGI v1.4 | 34496 |
| *Arabidopsis thaliana* | 157Mb | TAIR8 | 27025 |
| *Vitis vinifera* | 500Mb | Vitis project | 30434 |
| *Selaginella moellendorffii* | 100Mb | JGI v1.0 | 34697 |
| *Physcomitrella Patens* | 480Mb | JGI v1.1 | 35938 |
| *Chlamydomonas reinhardtii* | 120Mb | JGI v4.0 | 16687 |
| *Ostreococcus tauri* | 9.7 Mb | JGI v2.0 | 7725 |
| *Cyanidioschyzon merolae* | 16.5Mb | genome project | 4773 |
| *Saccharomyces cerevisiae* | 12 Mb | SGD | 6698 |
| | | Total | 250687 |

画像引用:生物アイコン, Copyright© 2009 ライフサイエンス統合データベースセンター

SALADデータの作成手順

**1. Blastによるグルーピング**

Blastサーチによりホモログガスなアテーションググループを作る

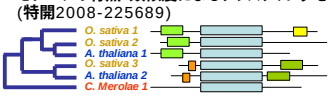



**2. MEMEによるモチーフ抽出**

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

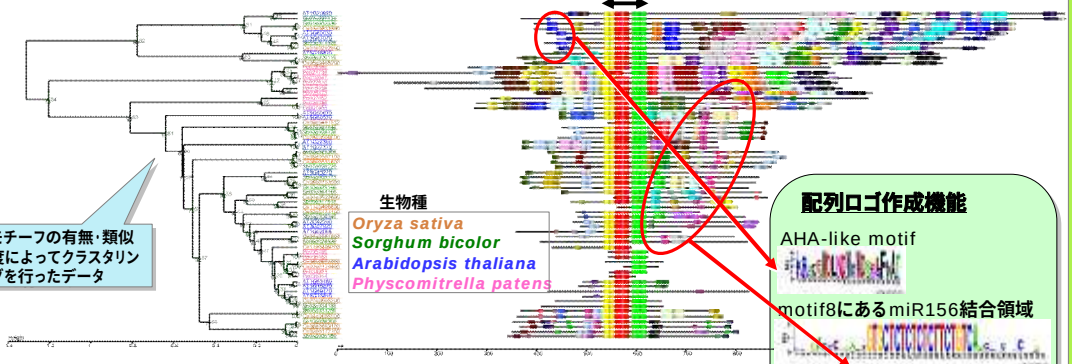
**3. 類似性クラスタリング**

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のビュワー

**モチーフ構造をビジュアルに比較可能！**

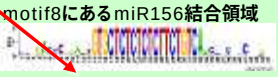



**配列ロゴ作成機能**

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

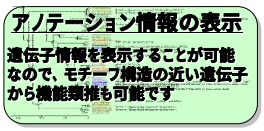
様々な機能

**グラフィクデータの操作**

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・特定のモチーフでのアライメント) が可能

**モチーフ配列で系統樹**

共通モチーフをアライメントとして系統樹 (NJ法) も簡単にかけます

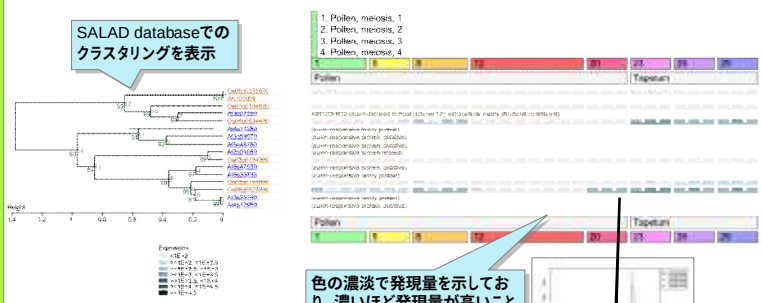


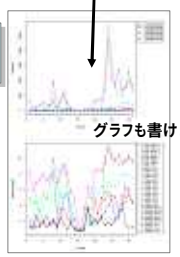
**アノテーション情報の表示**

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です

## SALAD on ARRAYs

**発現量を色の濃淡で表示し、ユーザーが注目する遺伝子とパラログガスな (もしくは類似性が高い) 遺伝子の発現パタン比較が一画面で簡単に行える！**




データ及び機能

・閲覧可能なデータは1000以上

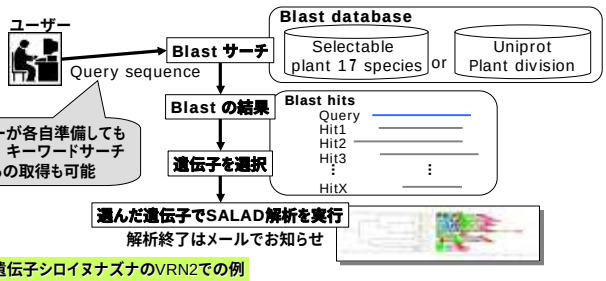
| データ | 生物種 | プラットフォーム | アレイ数 |
|---|---|---|---|
| LM アレイの花粉、タペータム | イネ | Agilent | 31 |
| AtGenExpress | シロイヌナズナ | Affymetrix | 783 |
| GEO登録データ (AtGenExpress 以外) | シロイヌナズナ | Affymetrix | 337 |

・表示アレイの順番や組み合わせを変えることがweb上で可能

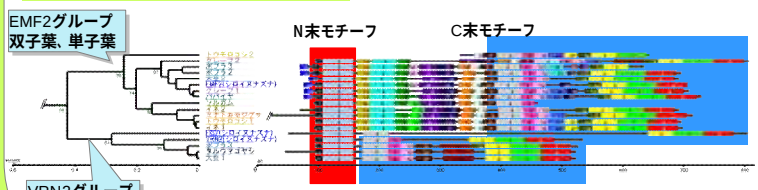
・ログインすることで、独自のデータセットの作成・登録が可能

## Blasting SALAD Analysis

**自分が比較したい配列を、自分で集めて解析可能！**
**特定のパラログ・オーソログの探索等を効率的に行える。**




春化経路遺伝子シロイヌナズナのVRN2での例




VRN2の仲間は双子葉のみにみられる。一方で、配列の相同性の高い開花抑制因子EMF2の仲間は双子葉にも単子葉にも存在する。VRN2のグループと、EMF2のグループは、N末とC末のモチーフ構造は共通なので、このデータからはEMF2の真ん中のモチーフの欠損でVRN2が誕生した可能性が示唆される。

**発表論文**

Mihara et al. (2010) SALAD database: a motif-based database of protein annotations for plant comparative genomics. NAR Database issue D835–D842. doi:10.1093/nar/gkp831, PMID: 19854933

Mihara et al. (2008) *In silico* identification of short-nucleotides sequences associated with gene expression of pollen development in rice. Plant & Cell Physiol. doi: 10.1093/pcp/pcn129, PMID: 18835840

**関連特許**

遺伝子クラスタリング装置およびプログラム (特願2008-252353)

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)

**助成** 農林水産省 新農業展開ゲノムプロジェクト GIR1002

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