

生物間相互作用の解明に向けた植物・微生物統合データベース間の連携

○平川英樹¹、藤澤貴智²、森宙史²、ゲルフィ アンドレア¹、市原寿子¹、中村保一²、磯部祥子¹、田畑哲之¹、黒川顕²

1. かずさDNA研究所、2. 国立遺伝学研究所

要旨

統合化推進プログラムにおいて、植物ではPlant GARDEN(<https://plantgarden.jp>)、微生物ではMicrobeDB.jp(<https://microbedb.jp>)がそれぞれ構築されている。植物・微生物間相互作用にはマメ科植物と根粒菌との共生といった植物にとって有益なものがある。一方、細菌や菌類、ウイルスに感染することで様々な病害を受けるといった問題も生じる。植物・微生物間相互作用を明らかにするため、統合データベース間での連携として、これまでにゲノムおよびメタゲノム情報を軸に統合化を進めてきた。互いのデータを有効に活用し、植物と微生物の全体像を解明するホロゲノム研究の情報基盤を構築することを目的として、Plant GARDENにおける植物ゲノム情報、DNAマーカー、QTL情報のRDF化を実施した。MicrobeDB.jpにおいては、植物ホストを中心にメタゲノムサンプルおよび解析データを更新しRDFを拡充させた。本発表では、植物・微生物間の相互作用に関する情報のRDF化により植物と微生物統合データベース間の更なる統合化の試みについて述べる。

背景

生物間の相互作用

植物は微生物(細菌、真菌)、ウイルスなどと相互作用している。共生など植物の成長を促進させるPlant Growth Promoting Bacteria(PGPB)も明らかになりつつある。一方、病気を引き起こすものもある。

＊共生:根粒菌(ミヤコグサなど)、菌根菌

＊病原性

- ・細菌:トマト斑点細菌病菌、黒斑細菌病菌など
- ・真菌:べと病菌、かび病菌など
- ・ウイルス:タバコモザイクウイルスによるモザイク病など

ゲノム解析手法の進展

次世代シーケンサーにより多様な植物や微生物のゲノム解読が行われている。様々な環境の菌叢を調べるメタゲノム解析やメタトランスクリプトーム解析が行われている。

・メタゲノム解析

菌叢を解析→有用微生物の探索など

・メタトランスクリプトーム解析

環境中に存在する遺伝子を探索
→有用遺伝子の探索など



以下の様な環境全体における生物間の相互作用を明らかにするホロゲノム解析が注目されている。

・ウイルス、微生物叢(土壌細菌など)、植物(共生菌、内生菌)、糸状菌、線虫、原虫、節足動物、昆虫(腸内細菌)など

植物、微生物統合データベース



- ・118植物種(133ゲノム)のゲノム情報
ゲノム配列、CDS、PEP、Transcript配列
- ・機能(アノテーション)情報
- ・品種間SNP
- ・DNAマーカー
- ・QTL
- ・解析アプリケーションの開発

ユーザが有益な植物に関するゲノム情報を効率よく取得できることを目指し開発



- ・微生物に関するデータを系統・遺伝子・環境の3つの軸に沿って整理・統合したフルRDFのデータベース
- ・メタゲノムデータ(メタ16S rRNA、遺伝子)
- ・ゲノム・ドラフトゲノムデータ
- ・単細胞の真菌類、藻類のゲノムデータ
- ・解析プロトコルの標準化
- ・解析パイプラインの開発

微生物学の専門家のみならず非専門家も、微生物のゲノムやメタゲノム情報を容易に利活用できるDBを目指し開発

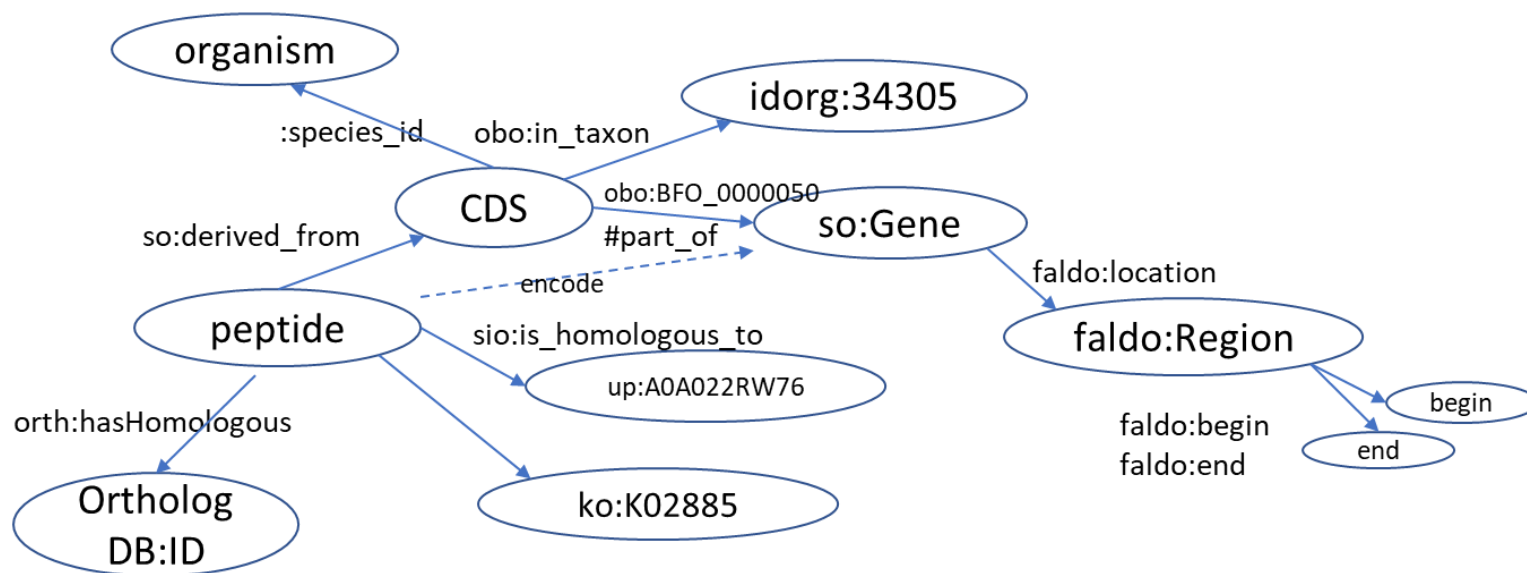
DNAマーカーを用いて遺伝子と形質との関連性を見出すことを目的とした。
他DBとの連携を図るため、公開情報のRDF化を行った。

登録情報

Plant GARDENに格納されている
植物ゲノム情報(gffファイル)

- ・植物種名(organism)
- ・Taxonomy ID(idorg:)
- ・遺伝子配列(CDS、pep配列)
- ・遺伝子の位置(faldo:)
- ・アノテーション情報
 - UniProtKB(up:)
 - KEGG KO(ko:)
- ・オルソログ情報
 - OrthoDB(OrthologDB:)

RDFスキーマ図



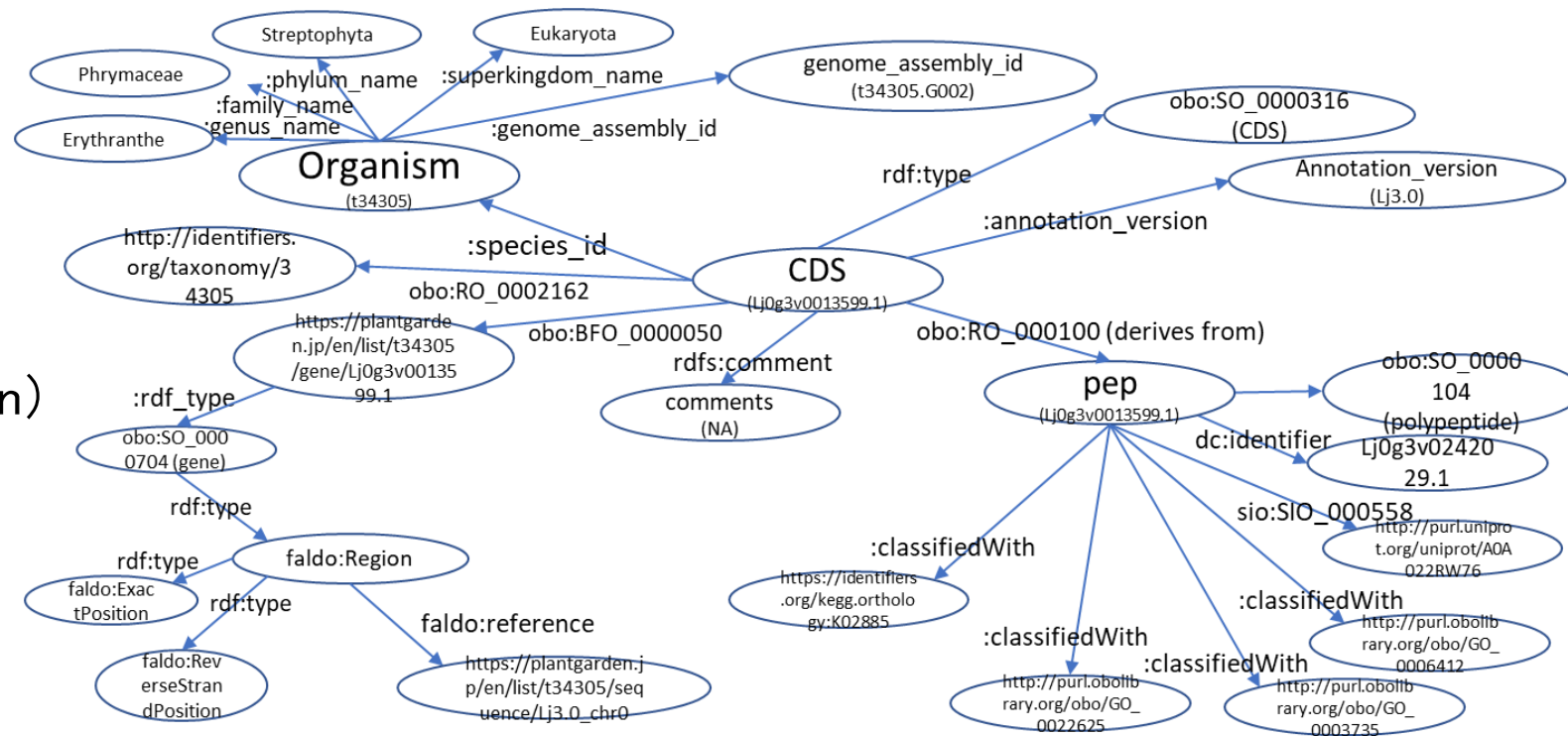
登録情報

Plant GARDENではHayai-Annotationに基づき再アノテーションを実施。
(Ghelfi et al., 2019, *Bioinformatics*)

Hayai-Annotationの出力結果をRDF化

- ・植物種名 (Organism)
- ・ゲノムアセンブルID (Annotation version)
- ・遺伝子配列 (CDS、pep配列)
- ・遺伝子の位置 (faldo:)
- ・アノテーション情報
GO (BP、CC、MP)
UniProtKB (up:)
KEGG KO (ko:)

RDFスキーマ図



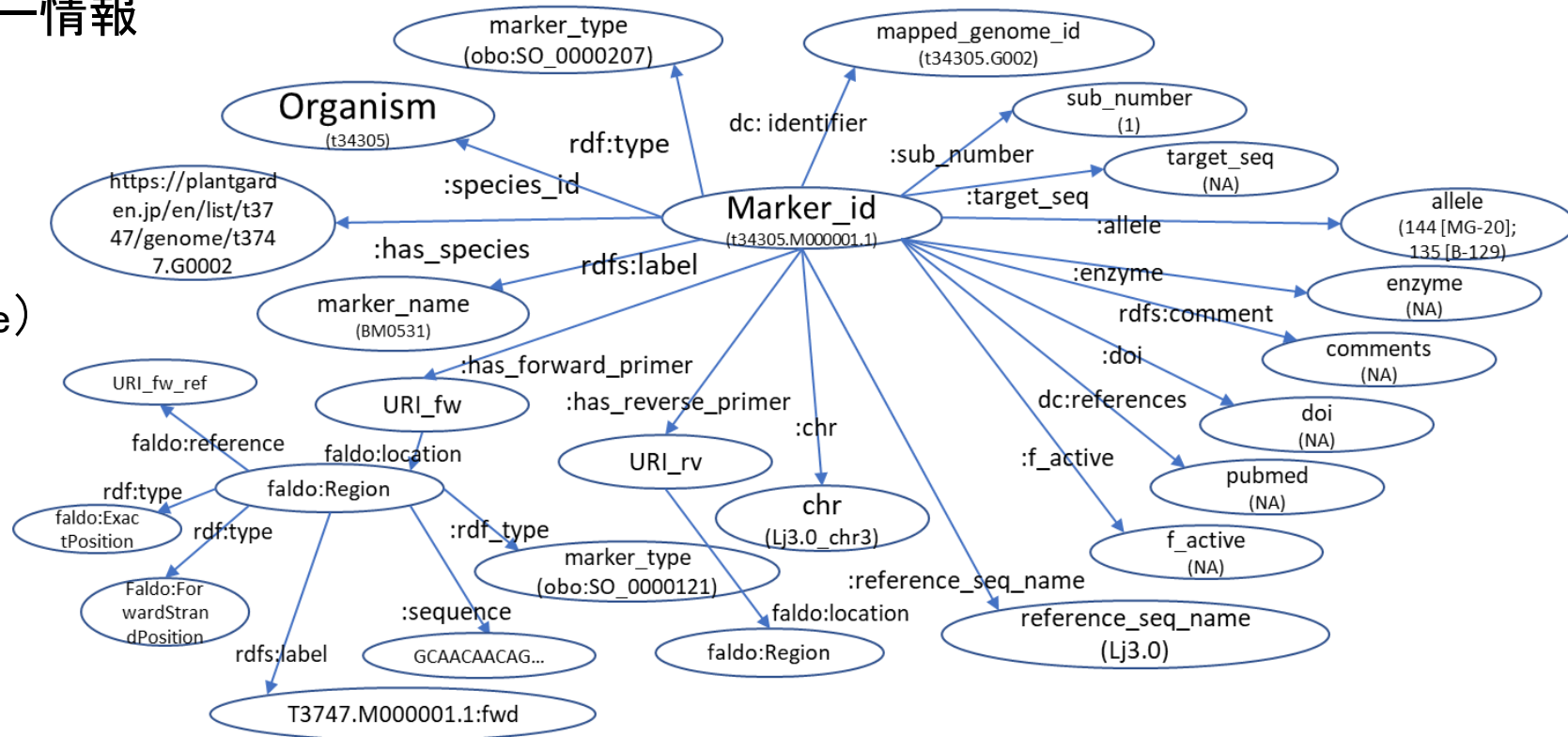
登録情報

文献から手動で収集したDNAマーカー情報

掲載したDNAマーカー情報をRDF化

- ・植物種名 (Organism)
- ・マーカー名 (Marker_id)
- ・ゲノムバージョン (reference_seq_name)
- ・染色体番号 (chr)
- ・プライマー情報
ゲノム上の位置 (faldo)
プライマー配列
- ・アレル情報 (allele)
- ・文献情報 (pubmed)
- など

RDFスキーマ図



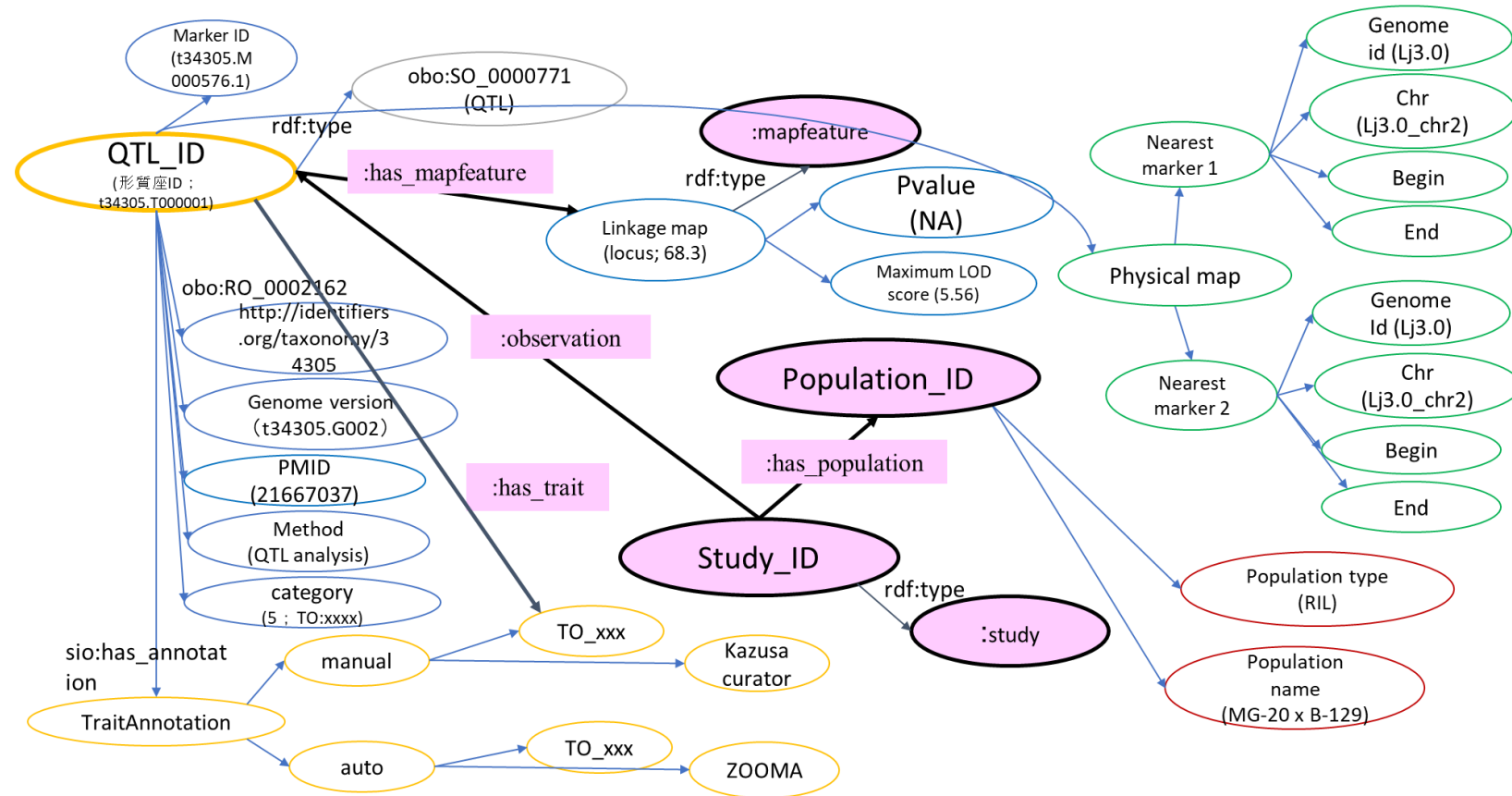
登録情報

文献から手動で収集したQTL情報

掲載したDNAマーカー情報をRDF化

- ・植物種名
- ・遺伝子配列(CDS、pep配列)
- ・遺伝子の位置
- ・アノテーション情報
 - UniProtKB
 - KEGG KO
- ・オルソログ情報
 - OrthoDB
- など

RDFスキーマ図



AgroLDのデータモデルと合わせたもの



MicrobeDB.jpで現在公開されている植物・微生物の関係情報



Plant GARDENに格納されている植物ホストについてのメタゲノム関連情報(一部)

植物種	学名	BioSample数	メタ16S解析数
シロイヌナズナ	<i>Arabidopsis thaliana</i>	3,241	1,023
イネ	<i>Oryza sativa</i>	4,939	426
テンサイ	<i>Beta vulgaris</i>	125	116
イチゴ	<i>Fragaria x ananassa</i>	187	45
トウモロコシ	<i>Zea mays</i>	6,119	17
Brassica rapa	<i>Brassica rapa</i>	177	15
ソルガム	<i>Sorghum bicolor</i>	1,357	12
キュウリ	<i>Cucumis sativus</i>	47	10
トマト	<i>Solanum lycopersicum</i>	1,293	8
リョクトウ	<i>Vigna radiata</i>	17	7
ペポカボチャ	<i>Cucurbita pepo</i>	6	5
レタス	<i>Lactuca sativa</i>	104	4
コムギ	<i>Triticum aestivum</i>	2,578	1
ブドウ	<i>Vitis vinifera</i>	1,313	1
リンゴ	<i>Malus x domestica</i>	317	1
Arabis alpine	<i>Arabis alpina</i>	134	1
ラッカセイ	<i>Arachis hypogaea</i>	123	1
デュラムコムギ	<i>Triticum turgidum</i>	73	1
モモ	<i>Prunus persica</i>	38	1

BioSample数の
合計: 36,594



INSDC DRA/ERA/SRAで公開されたマイクロバイオームのサンプル数
(2020年8月時点)

	サンプル数
マイクロバイオーム	約 2,300,000 (※)
ヒト	約 650,000
マウス	約 120,000
土壌	約 280,000
水環境	約 240,000
人工環境	約 160,000
植物(共生関連)	約 150,000

植物(共生関連)の
BioSample数の増加

(※)ヒトや土壌等マイクロバイオームサンプルの総数

植物・微生物間相互作用の関連性(PGPB(共生))(一部)

植物	細菌	働き
<i>Lotus japonicus</i>	<i>Mesorhizobium loti</i>	Increased plant growth (Nitrogen fixation)
<i>Brassica juncea</i>	<i>Azotobacter chroococcum</i> HKN-5	Stimulated plant growth (ref)
<i>Brassica napus</i>	<i>Xanthomonas</i> sp. RJ3, etc	Stimulated plant growth, cadmium accumulation (ref)
<i>Glycine max</i> (ダイズ)	<i>Pseudomonas fluorescens</i>	Increased plant growth (ref)
<i>Helianthus annuus</i> (ヒマワリ)	<i>Ochrobactrum intermedium</i>	Increased plant growth, decreased Cr (VI) uptake (ref)
<i>Trifolium repens</i> (シロクローバ)	<i>Brevibacillus</i>	Enhanced plant growth & nutrition of plants etc (ref)

Reference: Khan MS et al, *Environ. Chem. Lett.* (2008)

植物	真菌	働き
<i>Ipomoea batatas</i> (サツマイモ)	<i>Glomus</i> spp.	Drought (Proline & soluble sugars adjust osmotic potential)
<i>Fragaria x ananassa</i> (イチゴ)	<i>Funneliformis mosseae</i>	Drought (Increased shoot & root fresh weights, WUE, and plant survival)
<i>Solanum lycopersicum</i> (トマト)	<i>Rhizophagus irregularis</i>	High temperature (Enhanced photosynthetic capacity etc)

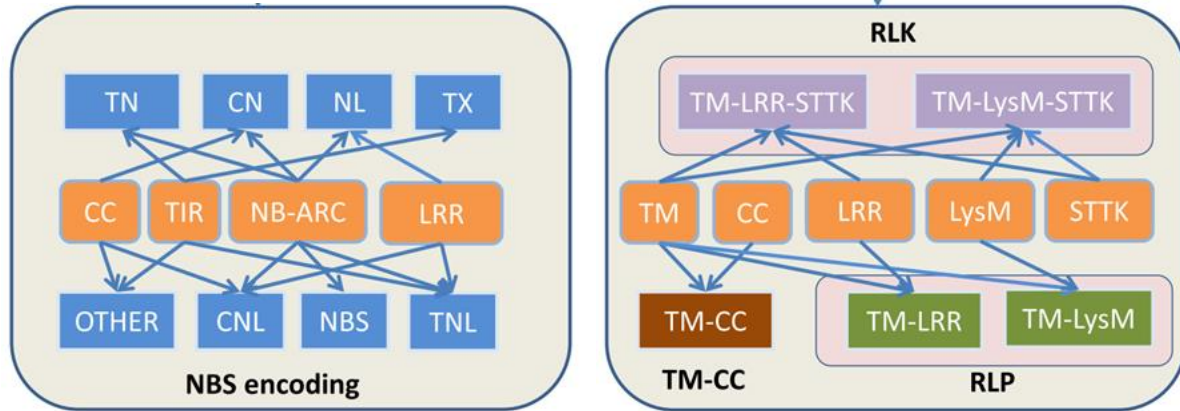
Reference: Begum N et al, *Frontiers in Plant Sci.* (2019)

PGPBの植物に対する働きが調べられている。菌叢と微生物との関連性を明らかにすることを目指す。

植物・微生物間相互作用の関連性(病原性)

植物(ホスト)側

病害抵抗性遺伝子候補(Resistance Gene Analog)の探索



ドメインの組み合わせの個数

ITAG4.0

NBS	NBS	54
CNL	CC-NBS-LRR	59
TNL	TIR-NGS-LRR	18
CN	CC-NBS	8
TN	TIR-NBS	6
NL	NBS-LRR	79
TX	TIR-unknown domain	12
Others	Others	1

RLP	Receptor like protein	85
RLK	Receptor like kinase	512
TM-CC	Transmembrane CC	147
Total		981



トマト
(*Solanum lycopersicum*)

CC: coiled-coil
LRR: leucine rich repeat
LysM: lysin motif
NB-ARC: nucleotide binding site-
activity regulated cytoskeleton

NBS: nucleotide-binding site
RGA: resistance gene analog
STTK: serine/threonine and tyrosine
kinase
TIR: Toll/Interleukin-1 receptor

RGAugury

細菌側

例) トマトかいよう病
(*Clavibacter michiganensis*)

NZ_CP033724.1

Subsp.: michiganensis

Strain: UF1

Country: USA Florida

Host: Tomato

Other strains:

61-1 DOAB 397
7580 CFBP 7577
HF4 NCPPB 2581
DOAB 395

ゲノム(遺伝子)情報

- ・病原性細菌のエフェクタータンパク質
- ・III型分泌装置など

RDFを用いたデータ統合により菌叢における病原性
細菌と植物種との関連性を明らかにすることを目指す

今後の連携への取り組み

- Plant GARDENにおいて、ユーザおよび連携データベースにRDFを提供するためのSPARQLエンドポイントやAPIなどの新しいインターフェースを提供する
- Plant GARDENとMicrobeDB.jpが公開している情報基盤を発展させ、植物・微生物間の相互作用を理解するためのデータ統合をさらに拡充する
- 各植物のPGPBや病原性に関わる微生物情報の収集とキュレーションを進める
- 各種オミックスデータとの組み合わせることで他データベースとの連携を図る

まとめ

植物と微生物のゲノム統合データベースを連携することを試みるため、まずは、Plant GARDENに格納されているゲノム情報についてRDF化を行った。植物ゲノムデータ、アノテーションデータ、DNAマーカー、QTLに関するRDFスキーマを構築した。QTLに関しては、AgroLDとの連携を図るため、AgroLDで使われているデータモデルに合わせた。

今回、連携するための準備として、Plant GARDENで公開されている植物のゲノム情報、アノテーション、DNAマーカー、QTLに関するRDF化を試みた。MicrobeDB.jpでは植物に関するメタ16Sに関する調査を行った。MicrobeDB.jpのRDF化、Stanzaの作成は行われており、今後の植物・微生物間の連携のための情報基盤の構築を引き続き進める。

微生物統合データベースMicrobeDB.jpに格納されている各植物ホストについてのメタゲノム関連情報のRDFを確認した。今後、ゲノム配列が解読されている微生物（細菌、真菌）、ウイルスについて、宿主（植物）と菌種（株）などに関する情報を収集し、キュレーションする予定である。

謝辞

本研究は、JST・NBDCのDatabase Integration Coordination Program(DICP)の支援を受けて実施しています。