

# 植物ゲノム統合データベースPlant GARDEN、微生物、メタボローム統合データベース間の連携検索システムの開発

○平川英樹<sup>1</sup>、藤澤貴智<sup>2</sup>、守屋勇樹<sup>3</sup>、信定知江<sup>4</sup>、長崎英樹<sup>1</sup>、森宙史<sup>2</sup>、福島敦史<sup>5</sup>、  
Andrea Ghelfi<sup>2</sup>、市原寿子<sup>1</sup>、中村保一<sup>2</sup>、金谷重彦<sup>6</sup>、有田正規<sup>2</sup>、黒川顕<sup>2</sup>、田畑哲之<sup>1</sup>、  
磯部祥子<sup>1</sup>

1. かずさDNA研究所
2. 国立遺伝学研究所
3. ライフサイエンス統合データベースセンター (DBCLS)
4. 科学技術振興機構NBDC事業推進部
5. 京都府立大学・生命環境科学研究科
6. 奈良先端科学技術大学院大学・情報科学領域・計算システムズ生物学

# 要旨

植物と微生物との相互作用や植物が産生する様々な二次代謝産物との関連性から新たな知見を見出すことを目的として、植物ゲノム統合データベースPlant GARDEN (<https://plantgarden.jp>)と微生物ゲノム統合化データベースMicrobeDB.jp、メタボローム統合データベースMetaboBank間でRDFによるデータ連携を進めており、2020年からトーゴーの日において進捗状況を発表してきた。2021年度の統合推進化プログラムの追加実施予算にて、一般ユーザにとって敷居が高いSPARQL言語を入力することなくSPARQL検索できるシステムの開発を検討していたところ、2021年10月にTogoDX/humanがDBCLSから公開された。そこで、TogoDXシステムに植物、微生物、メタボローム3者のコンテンツを格納したTogoDX-TPPを新たに構築し、さらに検索の効率化を図るため、TogoDX-TPPにおいてどのattributeにユーザが求めるデータが格納されているかを予め検索し、そこからTogoDX-TPPに検索を投入する連携検索システムを開発した。本発表では、現在の開発状況について発表する。

植物と微生物との相互作用は成長促進など有益なもの(PGPG)や病害など成長に影響を与えるものがある。また、植物は有用な二次代謝産物を産生し、多様な分野で広く用いられている。

## ①植物と微生物間の相互作用の理解

従来、植物と微生物の相互作用はマメ科植物にて知られていたが、近年、ブナ目やバラ目など様々な植物にて共生関係が見つかってきている。

### ・Plant Growth Promoting Bacteria (PGPB) の例

セイヨウアブラナ(アブラナ目) - *Xanthomonas* sp. RJ3: 成長刺激

ヒマワリ(キク目) - *Ochrobacterum intermedium*: 4価クロムの提供による成長促進

・病原性細菌、菌類、ウイルスによる感染: モザイク病、うどんこ病、さび病、べと病など多様な病害

植物と微生物との相互作用の理解は植物(作物)の生育の効率化に繋がると考えられる。

## ②有用な二次代謝産物の合成に関連する遺伝子の推定

植物はフラボノイドなど多様な二次代謝産物を産生し、染料や香辛料、医薬品など多様な方面に用いられている。植物の遺伝子と代謝産物などを関連付けることが重要である。

# 目的

植物のゲノム情報、微生物(菌叢)の情報、メタボロームによる化合物情報、さらに、国内外のデータとを検索することで新たな知識発見を目指すことを試みている。

Plant GARDENでは、植物ゲノム、遺伝子、アノテーション、DNAマーカー、QTLなどについてRDF化を行い、微生物ゲノム統合化データベースMicrobeDB.jp、メタボローム統合データベースMetaboBank間でRDFによるデータ連携を進めている。技術面では、DBCLSや遺伝研の方々にご協力頂きながらRDF化を進めている。

Plant GARDEN、MicrobeDB.jp、MetaboBankの各エンドポイントは構築しており、ユーザが各種データベースに対してVirtuosoを用いたSPARQL検索を行うことができるが、その際、SPARQL言語に関する知識が必要となる。このため、RDF化されたトリプルデータを揃えてもユーザにとっては敷居が高く、基礎研究から育種の現場における幅広いユーザに利用してもらうのは難しいと思われる。

この問題点を解決するため、RDFに慣れていない一般ユーザが各種統合化データベースに対してSPARQL検索を容易に実現できるユーザインタフェースを搭載した高度な連携検索システムを構築することを目的とした。

## Plant GARDEN

JA EN

Plant GARDEN ( Genome And Resource Database ENtry ) は  
さまざまな植物のゲノムやマーカー情報を集めたポータルサイトです。

植物ゲノム統合DB  
<https://plantgarden.jp>

例: Solanum

植物種からさがす

キクタニギク

ブドウ欧州種

ミヤコグサ

シロイヌナズナ

ダイコン

イチゴ

ラッカセイ

ダイズ

トマト

すべてみる ▶

他の方法でさがす

遺伝子から

配列から

データ一覧

マーカー

解析してみよう

SNP Detection

NEWS

● 25. Aug. 2022

8/19(金) 12:00 ~ 24(水)のシステムメンテナンスは終了致しました。

04. Aug. 2022

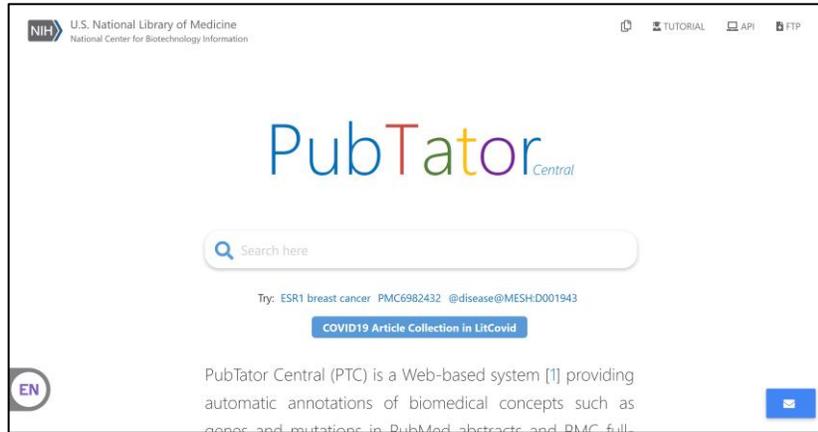
Triticum aestivum (パンコムギ)、Triticum turgidum (デュラムコムギ)、Hordeum v  
「ゲノム配列を見る」ページの塩基配列長が正しく表示されていませんでした。

- ・ゲノム配列(184種;データ件数:236)
- ・遺伝子配列(160種;件数:10,948,303)
- ・DNAマーカー(61種;件数:373,596)
- ・QTL(28種;件数:8,255)
- ・バリエーション情報(106種)
- ・機能(アノテーション)情報
- ・解析アプリケーションの開発

Plant GARDENの詳細は15番のポスターで発表

# 植物の生物種間相互作用、病気、遺伝子、化合物情報の収集、RDF化

https://www.ncbi.nlm.nih.gov/research/pubtator/



Plant GARDENで公開している植物種名で検索 (API)

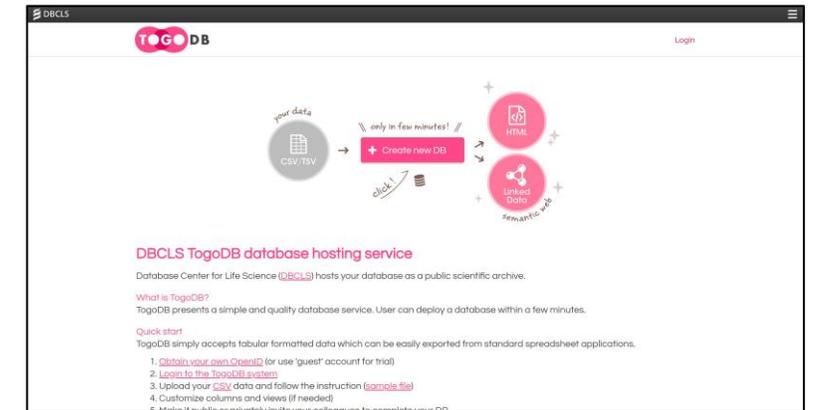


31848360	t				Solving the controversy of healthier organic fruit: Leaf wounding triggers distant gene expression response of polyphenol biosynthesis in
31848360	a				The claim that organic agriculture produces higher levels of phytochemicals has been controversial for decades. Using strawberries as a
31848360	111	121	polyphenol	Chemical	MESH:D05980
31848360	500	506	sugars	Chemical	MESH:D00224
32252456	t				Priming of Defense Systems and Upregulation of MYC2 and JAZ1 Gene
32252456	a				Several attempts have been made to study the effects of methyl jasmonate
32252456	73	89	Botrytis cinerea	Species	
32252456	105	121	Methyl Jasmonate	Chemical	MESH:C07223
32252456	886	897	anthocyanin	Chemical	MESH:D00087
32647129	t				A new functional JAZ1 degron sequence in strawberry JAZ1 revealed by
32647129	a				The phytohormone jasmonoyl-isoleucine (JA-Ile) regulates fundamental
32647129	51	55	JAZ1	Gene	
32647129	325	329	COI1	Gene	
32709075	t				Effects of Processing on Polyphenolic and Volatile Composition and Fr
32709075	a				Strawberries belonging to cultivar Clery (Fragaria x ananassa) (Duches
32709075	820	829	flavonoid	Chemical	MESH:D00541
32709075	1149	1159	polyphenol	Chemical	MESH:D05980

## Bioconcept

- CellLine (細胞株)
- Chemical (化合物)
- DNAMutation (塩基置換)
- Gene (遺伝子)
- Species (種)
- Strain (株)
- ProteinMutation (アミノ酸残基置換)
- SNP
- Genus (属)

TogoDBへの格納、RDF化



pg\_pubtator database (格納済)

pg\_pubtator database:

This message can be edited in the "Customize HTML" > "Title" > "Body".  
Title (i can be set in the "General settings" > "Metadatas")  
Description (i can be set in the "General settings" > "Metadatas")

Entry	Taxonomy_id	Organism	Pmid	Start	End	Term	Bio_concepts	id	Sentence
Show	13747	Fragaria x ananassa	31848360	111	121	polyphenol	Chemical	MESH:D05980	Solving the controversy of healthier organic fruit: Leaf wounding
Show	13747	Fragaria x ananassa	31848360	138	148	strawberry	Species	3747	Solving the controversy of healthier organic fruit: Leaf wounding
Show	13747	Fragaria x ananassa	31848360	156	175	Fragaria x ananassa	Species	3747	Solving the controversy of healthier organic fruit: Leaf wounding
Show	13747	Fragaria x ananassa	31848360	500	506	sugars	Chemical	MESH:D00224	As a result phenolic compounds (PCs) and total soluble sugar
Show	13747	Fragaria x ananassa	31848360	547	563	phenylpropanoids	Chemical	MESH:C5204C	As a result phenolic compounds (PCs) and total soluble sugar
Show	13747	Fragaria x ananassa	31848360	642	647	sugar	Chemical	MESH:D00224	As a result phenolic compounds (PCs) and total soluble sugar
Show	13747	Fragaria x ananassa	31848360	853	859	carbon	Chemical	MESH:D00224	It was observed that the accumulation of PCs on fruits can be
Show	13747	Fragaria x ananassa	31849062	46	56	Strawberry	Species	3747	Determination of Pesticide Residual Levels in Strawberry (Fra
Show	13747	Fragaria x ananassa	31849062	279	289	strawberry	Species	3747	BACKGROUND: In this study, the development of IR based-p
Show	13747	Fragaria x ananassa	31849062	291	310	Fragaria x ananassa	Species	3747	BACKGROUND: In this study, the development of IR based-p
Show	13747	Fragaria x ananassa	31849062	815	829	Pyraclostrobin	Chemical	MESH:C5134Z	A total of 60 batch sample groups, each consisting of 8 straw
Show	13747	Fragaria x ananassa	31849062	876	886	strawberry	Species	3747	A total of 60 batch sample groups, each consisting of 8 straw
Show	13747	Fragaria x ananassa	31849062	900	910	strawberry	Species	3747	The strawberry samples with pesticide residuals were used by
Show	13747	Fragaria x ananassa	31849062	1264	1272	boscalid	Chemical	MESH:C5500E	RESULTS AND CONCLUSION: PLSR (Partial Least Squares
Show	13747	Fragaria x ananassa	31849062	1277	1291	pyraclostrobin	Chemical	MESH:C5134Z	RESULTS AND CONCLUSION: PLSR (Partial Least Squares

センテンスの抽出  
共起表現で情報抽出できる

# データベース間の連携

RDF化を進めているが、SPARQL検索には言語に関する知識が必要であり、一般ユーザにとっては検索することが難しい

## Plant GARDEN

- ・ゲノム配列(184種;データ件数:236)
- ・遺伝子配列(160種;件数:10,948,303)
- ・DNAマーカー(61種;件数:373,596)
- ・QTL(28種;件数:8,255)
- ・高次アノテーション情報  
有用遺伝子に関するセンテンス情報

連携

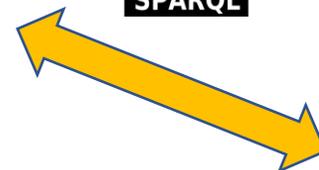


## ● MetaboBank ●

- ・実験で得られた代謝産物に関する生データや実験に関するメタデータ
- ・データ数
- ・KNASAcKのRDF  
(57,919の化合物)



連携



- ・系統・遺伝子・環境の3つの軸に沿って整理・統合されたフルRDFのデータベース
- ・菌叢メタゲノムデータ(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

外部データベースのRDF



Medical Subject Headings RDF



統合化データベース連携検索システムの構築

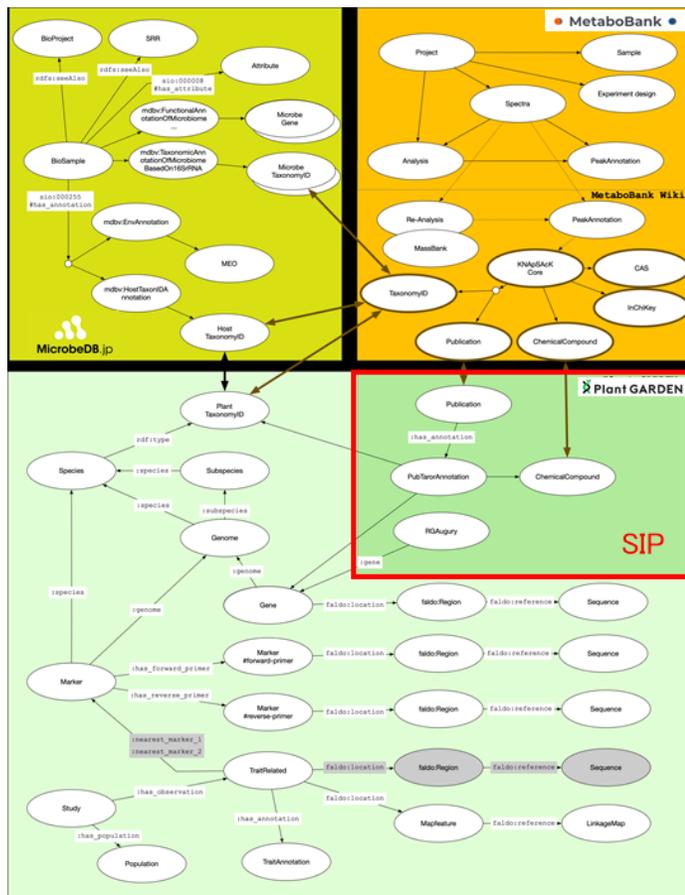
一般ユーザが利用しやすい連携検索システムを構築し、基礎研究から育種などの応用研究に至る幅広い分野での利用を図る



# データベース間の連携

植物、微生物、メタボローム統合データベースにおいて構築されたRDFを元にTogoDX上でSPARQL検索できるかを確認するため、以下のデータセットを対象とし、RDFスキーマ図を作成し、データを格納したTogoDX-TPPを構築した

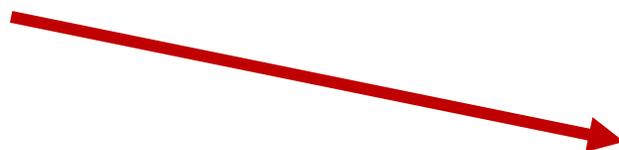
選出したattributeに関するRDFスキーマ図



遺伝研: 藤澤

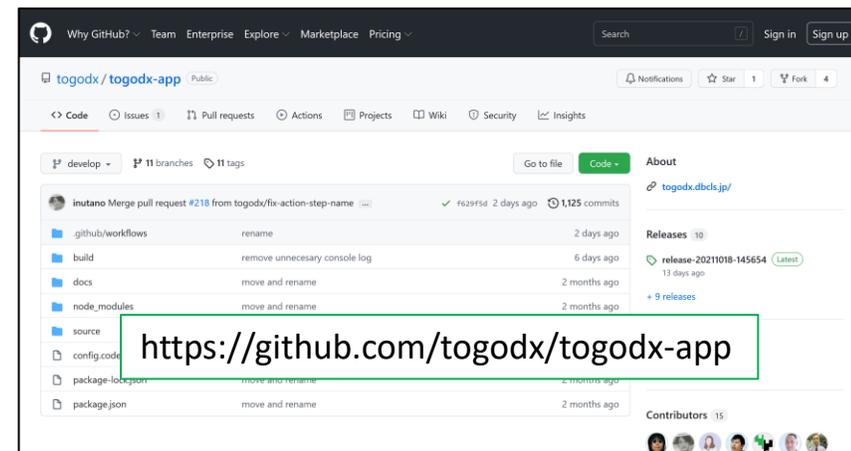
RDF/tsvをTogoDX-TPPに格納

- Plant GARDEN  
植物種、ゲノム配列  
センテンスキュレーション  
(Plant GARDENから公開予定)
- MicrobeDB.jp  
系統、BioSample



- MetaboBank  
KNAPSAcKの化合物、生物種  
文献情報
- DDBJ  
系統、BioProject、BioSample
- TogoGenome: 系統、ゲノム
- 化合物データ (EBI CheBI、  
NCBI PubChem)

TogoDX-TPPの構築




# TogoDX-TPP

インフォ・ラウンジ社のサーバー上で限定公開(開発版)

現時点では全データセットは格納できておらず、以下の項目のみ格納できている。

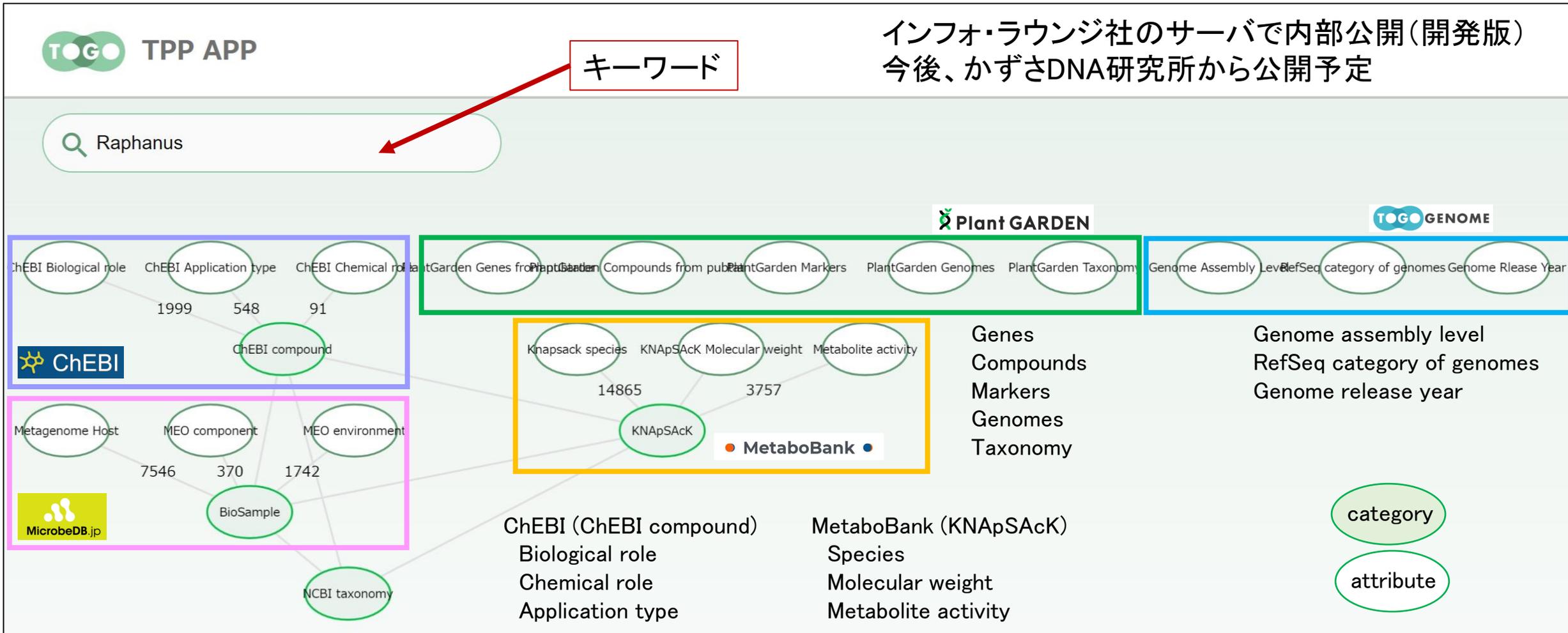
- Plant GARDEN: Taxonomy、DNA marker
- MicrobeDB.jp: MEO (component、environment)
- MetaboBank: KNApSAcK(chemical role、application type、biological role、molecular weight、metabolite activity)

かずさDNA研究所から公開予定

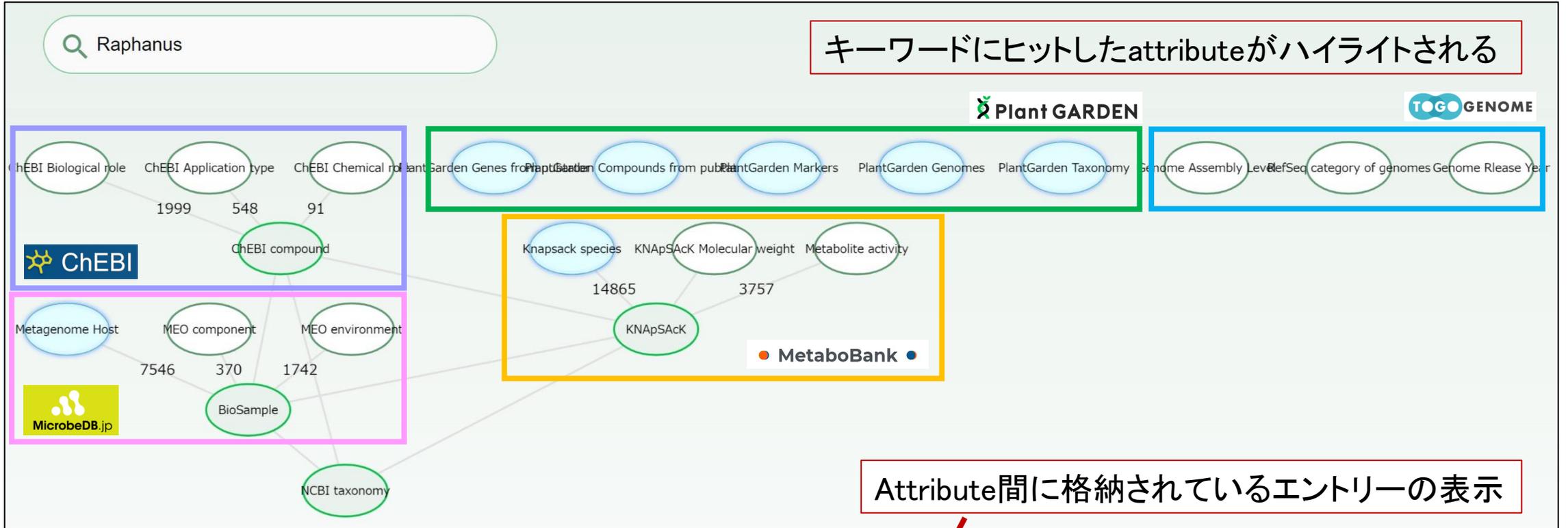
The screenshot displays the TogoDX-TPP interface. The main content area shows a tree structure of attribute keys and values. The 'MetaboBank (TPP)' section is highlighted with an orange box, containing 'Chemical role', 'Application type', 'Biological role', 'Molecular weight', and 'Metabolite activity'. The 'Plant GARDEN' section is highlighted with a green box, containing 'Taxonomy' and 'PlantGarden genome marker'. The 'MicrobeDB.jp' section is highlighted with a pink box, containing 'MEO component' and 'MEO environment'. On the right, the 'Condition builder' sidebar is visible, with 'MetaboBank' highlighted in an orange box. The sidebar includes a 'Select target dataset' dropdown, a 'Map your IDs' section, and a 'View results' button.

# 連携検索システムの構築(開発版)

キーワードを入力すると、予めTogoDX-TPPに格納されているデータに対して検索が実行される。  
各データベースがNCBI Taxonomyを介して繋がるようになっている。



# 検索結果の例



marker to PlantGarden Markers    pubtator\_gene to PlantGarden Genes from pubtator    knapsack to Knapsack species    ncbi\_taxonomy to PlantGarden

ID	ATTRIBUTE VALUE	
3726	Raphanus sativus	Launch TogoDX TPP
109996	Raphanus raphanistrum	Launch TogoDX TPP

Attribute間に格納されているエントリーの表示

TogoDX-TPPで表示

# TogoDX-TPP での表示

The screenshot shows the TogoDX TPP interface. On the left, there are several attribute keys categorized by color: MetaboBank (green), Plant GARDEN (yellow), MicrobeDB.jp (red), TogoGenome (purple), and RDI-Portal (blue). Each key has a corresponding attribute value bar chart. On the right, there is a 'Condition builder' panel with a 'Select target dataset' dropdown set to 'ChEBI compound', a 'Map your IDs' section with a 'Try' button, and an 'Add filters' section where 'Raphanus sativus' is selected. Below the condition builder is a 'View results' button, which is highlighted by a red arrow pointing to the right.

View resultsをクリック

The screenshot shows the search results page. At the top, there are radio buttons for 'Hits/All' (selected), 'Hits/All (%)', and 'Hits only'. Below this, there is a list of ChEBI compounds and their associated hits. The hits are displayed as a vertical bar chart on the right side of the page, with a green bar for 'Raphanus sativus'. The list of hits includes:

- C00044171 Gibberellin A112;GA112
- C00052344 m-Coumaric acid
- C00007549 Zeatin 7-glucoside;trans-Zeatin 7-glucoside
- C00007802 n-Hexyl glucosinolate
- C00007828 C00007828
- C00007835 C00007835
- C00007874 Gibberellin A112;GA112
- C00007801 C00007801
- C00007827 C00007827
- C00053075 Caffeoylmalic acid
- C00053173 Feruloylmalic acid
- C00053805 Sulfuraphene
- C00057534 Raphanusol A
- C00057549 Raphanusol B



## まとめ

連携検索システムUIを開発することで、ユーザがSPARQL言語を入力しなくても、SPARQL相当の検索を実現することができた。

SPARQL検索ではコンテンツのデータ量が増えると検索速度が遅くなるという問題点があるため、今回は、3者のデータベースからRDFが特に整備されているコンテンツを選出し、UIに格納した。

連携した3者のデータベースにおけるattributeの関係性についてRDF化し、TaxonomyとBioSample、PubChem、PubMedがのりしろとなることで、連携検索が可能になった。

今回構築した連携検索システムにより得られた結果では、コンテンツの内容やデータ量が十分でなかったため、知識発見に役立てる程のデータを得ることはできなかった。

今回構築した連携検索システムは、複数のデータベースから知識発見に役立てることができるデータ基盤として活用することができ、データベース統合技術としてのRDFの有効性を示すことができたと考えている。これにより、MicrobeDB.jpとMetaboBankに限らず、他のデータベースにおいてRDFで定義されたトリプルもしくは二項関係で繋がれたデータセットがあれば、さらなる連携の拡充を図ることができ、UIのネットワーク図を見ながら、検索の幅をより広げることができると考えている。

## 今後の取り組み

- ・ 新たな知識発見を目指すためには、コンテンツの内容やエントリー数を増やすことで検索結果を充実させ、検索の幅を広げる必要がある。
- ・ 現在、連携検索システムUIは開発中であるため、インフォ・ラウンジ社のテストサーバ上にて公開しているが、システムが完成し次第、かずさDNA研究所のウェブサーバから公開する。
- ・ その後、かずさDNA研究所から公開しているPlant GARDENのトップページにアイコンを設置し、そこからUIに向けてリンクを張る予定である。

### \* 問題点

- ・ データベース間の連携を図る際、遺伝子IDが重要になることが分かった。Plant GARDENが格納している遺伝子IDは、各実験グループが独自に定義したものであり、INSDCやNCBI RefSeq、Ensembl plantが扱う遺伝子IDとは異なるため、データ接続に大きな課題があった。このため、生物リソースやゲノムのメタデータに関しても、BioProjectやBioSample、Assembly DBとの接続が不完全になった。
- ・ 今後、遺伝子IDを定義することで、検索の幅をより広げることができると考えられた。

\* 謝辞： 本連携検索システムの構築は統合化推進プログラムの追加実施予算を受けて実施しました。