

*RCLASS データベースおよび反応オントロジーを用いた
植物代謝産物の生合成に用いられる反応パターンの分類*

化学生命科学 時松 敏明

背景と目的

近年、植物メタボロームデータの蓄積が進み、トランスクリプトームデータなどの他のオミックスデータとの統合的な解析を行うためのデータリソースが整ってきた。このような統合的な解析では、代謝反応がオミックスデータ間に関連づける最も重要な情報の一つであり、良質なレファレンスデータベースの存在が重要である。このため、筆者が所属する研究室では既知の代謝反応からパターンを抽出する方法を開発し、抽出したパターンを詳細に整理・分類し RCLASS データベースと反応オントロジーを構築してきた。また、当研究室では代謝経路予測ツールなどのツール群の開発も行っているが、データベース内の反応情報のカバー率などのいくつかの問題のためか、データベースの整備状況に起因すると思われる、予測精度の問題があると推察された。道の植物代謝反応の予測精度の向上のためには、反応データベースのカバー率の向上や代謝産物グループごとに利用される反応タイプの特徴の把握が重要であると考えられる、そこで当研究室で構築した RCLASS データベースと反応オントロジーの反応データを用いて、植物代謝関連の反応を抽出し、生合成経路ごとに利用される反応パターンを比較・分析することにより、前述のような情報を取得し、植物代謝反応の予測のために有用な情報を整備することとした。

検討内容

KEGG BRITE の Phytochemical Compound(br08003)に骨格・生合成経路をもとに階層分類されリストされている代謝産物情報をもとに、LinkDB を用い Phytochemical Compound と KEGG PATHWAY, KEGG REACTION, KEGG orthology などとの対応関係を取得した。この情報をもとに、植物代謝に関わる生体反応を生合成経路ごとに分類して抽出して、当研究室で開発している GenomeNet で公開している RCLASS や Reaction Ontology の情報を用いて分類し、生合成経路と反応パターンの情報について比較・分析を行った。

結果

Table 1. に使用した植物代謝産物のリストおよび、これらのうち酵素反応がリンクしているものおよび代謝マップに登場するものがどの程度あるかを示した。代謝産物のうち、反応が付与されているものは 3 割強しかなく代謝マップに登場するものは 1/4 強しかない。また、代謝パスウェイのクラスによっては反応やパスウェイにほとんど情報がないクラスがあった。Table 2 にパスウェイ分類ごとの代謝産物にリンクしている反応数とそのうち代謝マップにある反応の数および KO の付与率を示す。900 の代謝産物に関与する 1500 強(KEGG 全反応の 1/6 弱)の代謝反応が抽出され、約半分に KO が付与されていた。また、これらの反応のうち 1200 弱が代謝マップに乗っており、マップにある反応については約 5/8 について KO が付与されていた。代謝産物が多く報告されている Terpenoids (TP), Alkaloids (AL), Flavonoids (FL), Phenylpropanoids (PP)については、反応数も多数報告され、代謝マップにも良好にカバーされていた。以下、これらとそれ以外の化合物群(OT)、の 5 つのグループに分け、反応パターンの違いを RCLASS と Reaction Ontology を利用して分類・比較した(Table 3,4.)。RCLASS では、2800 強の RCLASS のうち、659 の RCLASS が植物代謝

産物の反応経路に存在しており、そのうち 611 が他愛車クラスと悔いてきてあった。ただ、出現した RCLASS のうち約 2/3 の 411 については 1 回しか出現しない RCLASS であり、分類が細かすぎることが示唆された。そこで、官能基に基づく反応オントロジーを利用した分類で比較したところ、659 の RCLASS が 138 にまとまった。このうち、7 割弱の 94 のクラスは反応クラス特異的であり、特にアルカロイドとテルペノイドは反応の多様性が顕著であった。クラス特異的な反応は、ほとんどが出現頻度の少ない反応クラスであったが、中には 50 以上の反応が含まれるクラスでテルペノイド特異的な反応クラスのものがあった。

考察

KEGGDatabase 内の植物代謝反応を抽出し、生合成経路をもとにした化合物グループごとに反応の分類および比較を行った。KEGG 内の化合物群の中には、代謝反応および代謝経路情報のほとんどない代謝産物グループが存在し、あらためて網羅的な情報の収集が必要であることを確認した。また、反応が多く収集されている主要なグループの代謝反応を官能基の反応オントロジーを用いて分類してみたところ、代謝産物群ごとに用いられる変換パターンに特徴があることが示唆された。これらの特徴は、反応経路予測への利用に有用ではないかと思われる。

また、今後さらに反応情報の収集を行い、さらなる解析を行うことが必要であると考ええる。

Table 1. KEGG BRITE Phytochemical Compounds (br08003) の第一階層分類と、これらの植物代謝産物のうち Pathway および Reaction と関連付けられる化合物の個数

Compounds Class in br08003	Compounds in br08003	Compound in pathway	Coverage (%)	Compounds with Reaction	Coverage (%)
Terpenoids	1096	251	22.9	329	30.0
Phenylpropanoids	186	40	21.5	50	26.9
Skimate/acetate-malonate pathway derived compounds	43	5	11.6	5	11.6
Flavonoids	483	201	41.6	225	46.6
Polyketides	129	1	0.8	3	2.3
Alkaloids	710	206	29.0	238	33.5
Amino acid related compounds	104	31	29.8	27	26.0
Fatty acids related compounds	37	15	40.5	16	43.2
Others	46	5	10.9	7	15.2
Total	2834	755	26.6	900	31.8

Table 2. パスウェイ分類ごとの代謝産物にリンクしている反応数とそのうち代謝マップにある反応の数および KO の付与率

Class	Reactions	R with KO	Coverage	R in Map	R with KO	Coverage
Terpenoids	649	357	55	445	332	74.6
Phenylpropanoids	121	66	54.5	93	66	71
Skimate/acetate-malonate pathway derived compounds	8	4	50	6	4	66.7
Flavonoids	294	145	49.3	250	143	57.2
Polyketides	3	0	0	1	0	0
Alkaloids	400	163	40.8	326	162	49.7
Amino acid related compounds	45	21	46.7	38	21	55.3
Others	20	7	35	20	7	35
Total	1540	763	49.5	1179	735	62.3
KEGG Overall	9157	4755	51.9	6269	4516	72

Table 3. 代謝産物グループごとの RCLASS の出現パターンとそのクラスサイズ分布

Reaction # in RC	All	AL	TP	FL	PP	OT
50 -	5	5	4	5	5	5
10 - 49	18	10	13	6	7	8
5 - 9	41	17	18	12	6	6
4	30	11	13	6	1	6
3	50	16	27	6	3	2
2	99	29	49	8	8	10
1	416	140	225	15	15	21
Total	659	228	349	58	45	58
Class unique	611	193	319	40	26	33

Table 4. 代謝産物グループごとの Reaction ontology 第 4 階層の出現パターンとそのクラスサイズ分布

Reaction # in Class	All	AL	TP	FL	PP	OT
50 -	12	11	10	9	8	9
10 - 49	21	19	12	5	7	7
5 - 9	23	10	12	4	2	3
4	6	4	2	1	1	0
3	17	5	7	3	4	3
2	23	11	8	0	5	4
1	36	16	11	2	1	6
Total	138	76	62	24	28	32
Class unique	94	39	34	6	7	8