

科学研究費助成事業 研究成果報告書

平成 27 年 6 月 24 日現在

機関番号：14603

研究種目：新学術領域研究(研究領域提案型)

研究期間：2010～2014

課題番号：22108009

研究課題名(和文) 悉皆的二次代謝経路推定に向けたデータベースおよび要素技術の研究開発

研究課題名(英文) Development of database toward an estimation of secondary metabolic pathways

研究代表者

金谷 重彦(Kanaya, Shigehiko)

奈良先端科学技術大学院大学・情報科学研究科・教授

研究者番号：90224584

交付決定額(研究期間全体)：(直接経費) 26,900,000円

研究成果の概要(和文)：二次代謝経路に関わる酵素のペプチド配列と反応特異性を体系的に関連づけることを目的に、生物-酵素-酵素反応の関係からなるデータベース(Motorcycle DB)を設計した。本データベースからは、酵素反応、基質と生成物、反応の詳細情報、ペプチド配列の情報を検索することができる。現在までに、文献情報をもとに、モノテルペン合成酵素、セスキテルペン合成酵素、ジテルペン合成酵素、トリテルペン合成酵素、P450酵素、アルカロイド合成、フラボノイド合成に関わる酵素の情報をDBへの蓄積し、<http://kanaya.naist.jp/motorcycle/top2.html>より無償公開した。

研究成果の概要(英文)：In order to systematize between relationships between peptide sequences and enzyme functions including reaction mechanism, we constructed enzyme database (KNAPsACK Motorcycle DB, <http://kanaya.naist.jp/motorcycle/top2.html>) involving secondary metabolites in the scientific literature. The DB comprises 2,421 secondary metabolic reactions. Enzyme reactions can be retrieved using keywords of the enzymes, species, genes, metabolites and peptide sequences obtained from a BLASTP search. In the metabolite search using its keywords, we obtain information on enzyme name, reaction equation, compound class and subclass of metabolic reactions and reaction mechanisms. In the BLASTP search, we can comprehensively predict reaction equations for a targeted peptide sequence using information on the class and subclass of metabolic pathways. Thus, the Motorcycle DB makes it possible to predict enzyme reactions based on the class and subclass of metabolic reactions evidenced by experiments.

研究分野：バイオインフォマティクス

 キーワード：バイオインフォマティクス ケモインフォマティクス データベース 二次代謝物 生合成情報 生物
代謝 ゲノム メタボローム

1. 研究開始当初の背景

地球上全体の顕花植物で生合成される二次代謝物(約106万種)の多様性を遺伝子レベルで理解することが、このように多様な二次代謝物を創出する進化のメカニズムを解明することへとつながる。そこで、これらの二次代謝に関わる酵素の情報を生物種、酵素反応、酵素のアミノ酸配列と関連づけたデータベース(Motorcycle DB)を設計し、データを格納し、二次代謝経路に関わる酵素のつかさどる化学反応とアミノ酸配列情報を関係づけるバイオインフォマティクス技術の開発することは、本分野を体系的に理解し、さらに新規反応に関わるアミノ酸配列におけるメカニズムを解明するために非常に重要な課題である。

2. 研究の目的

生物・酵素・酵素反応の関係からなるデータベース(Motorcycle DB)を設計し、データの体系化を目指す。また、本DBをもとに、酵素のペプチド配列特異性を、ペプチド配列により特徴づける要素技術を開発する。

3. 研究の方法

(1) 二次代謝反応データベース KNApSAcK Motorcycle DBの設計について以下に述べる。まずはじめに、生物種・代謝物・酵素の関係を体系化するためのデータベースを設計を行った。具体的には、フラボノイド、テルペノイド、ポリケタイド、ペプチドなど生物種を考慮に入れた既知生合成酵素についてアミノ酸配列、反応様式に関する情報および二次代謝物の反応情報を関連づけたデータベースの設計を行った。さらに設計されたデータベースに実際にデータを格納した。具体的には、16項目(酵素名、KEGG番号、EC番号、反応、生物種名、遺伝子名、文献、アミノ酸配列、Accession No.、反応分類-1、反応代謝物-2、最終生成物、酵素分類、反応メカニズム、反応経路図、登録者名)からなる収集データの入力型式を定義した。この定義に従ってデータを格納し <http://kanaya.naist.jp/motorcycle/top2.html> より公開した。

生合成システム(バイオ・マシナリー)の体系化をめざし、生合成反応の分類体系を構築する必要がある。そのために酵素反応を3階層からなる分類体系を構築した。その概略について以下に示す。第一階層においては、2つの大分類、[1]代謝物の骨格を構築する酵素グループ、[2]骨格に対して化学修飾を行う酵素グループにより酵素反応を分類した。[1]については、さらに10種のサブ分類と、それぞれの第3階層に分類した。[1-1]脂肪酸、[1-2]ポリケタイド([1-2-1]Type-III PKS(フラボノイドを除く)、[1-2-2]Type-II PKS(単ドメイン酵素)、[1-2-3]Type-I PKS(複合ドメイン酵素)、[1-2-4]Cis-AT-type(ATドメイン酵素)、[1-2-5]Trans-AT-type(単ドメイン

酵素)、[1-2-6]Iterative type-I PKS(single module)、[1-3]アミノ酸重合によるポリペプチド合成酵素、[1-4]リボソームによる合成されるポリペプチド合成酵素(RPS)、[1-5]リボソームを用いない合成酵素(NRPS)、[1-6]PKS-NRPS ハイブリッド合成酵素、[1-7]テルペン合成酵素([1-7-1]モノテルペン合成酵素、[1-7-2]セスキテルペン合成酵素、[1-7-3]ジテルペン合成酵素、[1-7-4]トリテルペン合成酵素、[1-7-3]ステロイド合成酵素、[1-7-4]メロテルペン合成酵素)、[1-8]アルカロイド合成酵素([1-8-1]テルペンアルカロイド合成酵素、[1-8-2]イソキノナルカロイド、[1-8-3]インドールアルカロイド合成酵素)、[1-9]フェニルプロパノイド合成酵素([1-9-1]フラボノイド合成酵素、[1-9-2]イソフラボノイド合成酵素、[1-9-3]フェニルプロパン合成酵素、[1-9-4]クマリン合成酵素、[1-9-5]スチルベン合成酵素)、[1-10]その他([1-10-1]アミノグリコシド合成酵素、[1-10-2]スクレオシド型抗生物質合成酵素)。

一方、[2]骨格に対して化学修飾を行う酵素グループについては、[2-1]酸化酵素、([2-1-1]P-450モノオキシゲナーゼ、[2-1-2]aKGジオキシゲナーゼ、[2-1-3]FADモノオキシゲナーゼ、[2-1-4]デヒドロゲナーゼ、[2-1-5]ハロゲン化酵素、[2-1-6]その他)、[2-2]アシル酵素、[2-3]プレニル化酵素、[2-4]メチル化酵素、[2-5]グリコシル化酵素、[2-6]アミノ化酵素、[2-7]その他に分類した。図1はKNApSAcK Motorcycle DBのメインウインドウである。

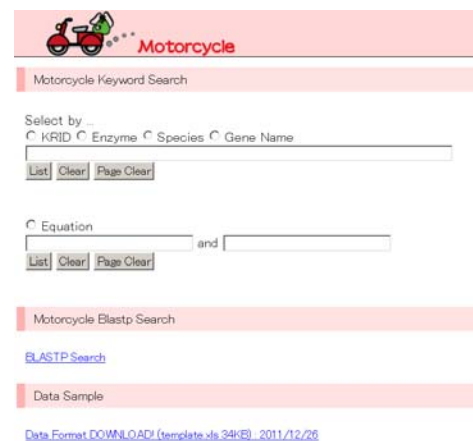


図1 KNApSAcK Motorcycle DBのメインウインドウ

図1によるメインウインドウから二次代謝経路に関わる酵素を、酵素番号、酵素名、酵素を有する生物、遺伝子名から検索することができる(図1のKRID, Enzyme, Species, Gene Nameと対応)。図1におけるEquationを選択し、テキストボックスに基質と生成物を入力することにより、酵素を検索することができる。また、Motorcycle BLASTP Searchにより配列情報から酵素を検索することが

できる。図2は、Geranyldiphosphateを基質とした場合の検索結果である。

input word = Geranyl diphosphate and -
Number of matched data: 100

KRID	Enzyme	KEGG ID	EC	Equation
KR000413	Dimethylallyl-diphosphate isopentenyl-diphosphate dimethylallyltransferase	R01408	2.5.1.1	Dimethylallyl diphosphate + Isopentenyl diphosphate <=> Diphosphate + Geranyl diphosphate
KR000443	Geranyl-diphosphate isopentenyl-diphosphate geranyltransferase	R02003	2.5.1.10	Geranyl diphosphate + Isopentenyl diphosphate <=> Diphosphate + trans-trans-Farnesyl diphosphate
KR0001658	Linalool synthase	--	--	Geranyl diphosphate -> (-)-(3R)-Linalool (100) + Pyrophosphate
KR0001659	Linalool synthase	--	--	Geranyl diphosphate -> (-)-(3R)-Linalool (100) + Pyrophosphate
KR0001660	Linalool synthase	--	--	Geranyl diphosphate -> (-)-(3R)-Linalool (96) + Pyrophosphate
KR0001661	Linalool synthase	--	--	Geranyl diphosphate -> (-)-(3R)-Linalool (97) + Pyrophosphate
KR0001662	Linalool synthase	--	--	Geranyl diphosphate -> (-)-(3R)-Linalool (100) + Pyrophosphate
KR0001663	Linalool synthase	--	--	Geranyl diphosphate -> (-)-(3R)-Linalool (100) + Pyrophosphate
KR0001664	Linalool synthase	--	--	Geranyl diphosphate -> (-)-(3R)-Linalool (100) + Pyrophosphate
KR0001665	(-)-4S-beta-phellandrene synthase	--	--	Geranyl diphosphate -> (-)-4S-beta-Phellandrene (52) + Pyrophosphate

図2 検索結果

図2の検索結果において例えば酵素反応番号KR0001661を選択すると図3における酵素反応の詳細が出力される。ここで、CclassとEclassは、それぞれ、生成物の構造分類ならびに酵素反応分類である。また、反応式(Equation)のアイコンをクリックすることにより、KNAPSAcK Core DBによる基質ならびに生成物を生産する生物を検索することができる。また、Reaction MechanismにおけるME000001.gifをクリックすることにより、反応メカニズムにおける情報(図4)を得ることができる。

Select Keyword = KRID
input word = KR0001661



KRID	KR0001661
Enzyme	Linalool synthase
KEGG ID	--
EC	--
Equation	 Geranyl diphosphate -> (-)-(3R)-Linalool (97) +  Pyrophosphate
C-class	Terpene
C-subclass	Monoterpene
FinalProduct	(-)-(3R)-Linalool
Eclass	Monoterpene synthase
Reaction Mechanism	ME000001.gif
Pathway	--
Curator	Shigehiko KANAYA

図3 酵素反応の詳細情報

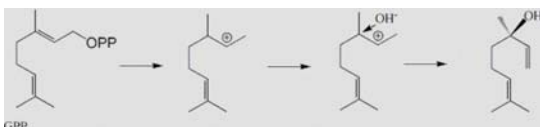


図4 反応メカニズムの一例。

このようにして、酵素反応データベースを設計し、データの蓄積を行った。KNAPSAcK MotorcycleはKNAPSAcK Family DBの情報と密にリンクづけられており、二次代謝物質に関わる活性、生物情報などの多面的情報が提供できる。KNAPSAcK Family DBにおける月あたりのアクセス数は、おおよそ15万件であり、本データベースは世界の多くの研究者から活用され、世界標準としての認知

されるに至っている。

4. 研究成果

(1) KNAPSAcK Motorcycle DBにおいてはペプチド配列から酵素反応を検索することも可能である。BLASTP Searchにより、酵素のペプチド配列をもとに反応検索を行うことができる。出力結果において、酵素大分類、代謝物種の大分類と小分類が得られる。一方で、ペプチド配列が非常に高いレベルで類似であってもその代謝反応は非常に多様であることがMotorcycle DBを使うことで把握することが可能である。このことは、NGSによる配列データをもとに、代謝経路を高精度で検討するときに非常に有益な情報となると期待される。現在までに、文献情報をもとに植物を中心に、モノテルペン合成酵素、セスキテルペン合成酵素、ジテルペン合成酵素、トリテルペン合成酵素、P450 (CYP) 酵素、アルカロイド合成、フラボノイド合成に関わる酵素の反応を整理し、DBへの蓄積がほぼ完了し、現在までに、全体で2421反応種についての整理が終わった。

(2)KNAPSAcK Motorcycle DBを活用した要素技術の開発を進め、セスキテルペン合成酵素、ジテルペン合成酵素、トリテルペン合成酵素、P450 (CYP) 酵素、アルカロイド合成、フラボノイド合成経路に関わる酵素のペプチド配列特異性を、ペプチド配列における2ペプチド頻度により特徴づけることにも成功した。まずはじめに、植物(59165種)と微生物(代表的66種)の71万種のペプチド配列をもとに、ペプチド配列を2ペプチド頻度のベクトルで表現し、自己組織化法によりデータ構造を把握した。このようにして得られたデータ分布図を自己組織化地図という。図5に例として、ペプチド配列頻度による4種のテルペンサイクラーゼ(モノテルペン、セスキテルペン、ジテルペン、トリテルペン・サイクラーゼ)の自己組織化地図を示す。この図から、4種のテルペンサイクラーゼには固有の配列特性が見られ、さらに、モノテルペンとセスキテルペンサイクラーゼは非常に共通性が高いことを示している。さらに、おおまかにではあるがそれぞれのテルペンサイクラーゼグループは、2,3個の共通配列により構成されていることも自己組織化地図から把握できた。さらに、この地図上で、P450 (CYP) 酵素、アルカロイド合成、フラボノイド合成経路にかかわる酵素の分布により、それぞれの酵素の配列特性を特徴づけることができた(図5A)。また、モノテルペン・サイクラーゼとセスキテルペン・サイクラーゼは、混合してt1ならびにt2に分布した。ジテルペンサイクラーゼはt3とt4に、またトリテルペン・サイクラーゼt5とt6に分布した。

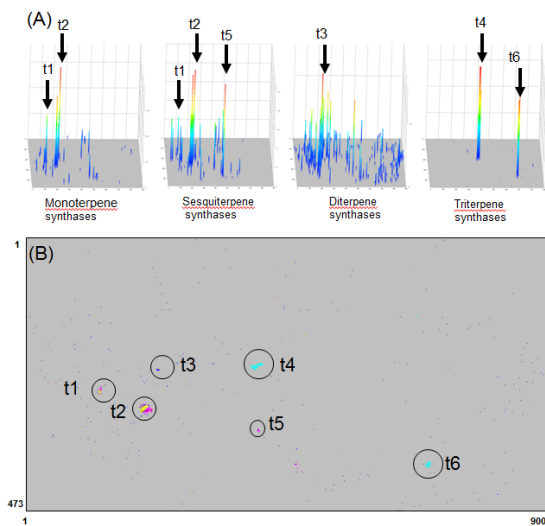


図 5 2 ペプチド配列頻度による 4 種のテルペンサイクラーゼの自己組織化地図。

今後、ゲノム、トランスクリプトーム、プロテオーム、メタボロームに代表されるオーム情報をいかに効率よく解釈するかということは、医薬・薬学・栄養学・生物学などの分野の大きな課題である。情報科学では、たとえば世界のウェブ上での情報の集約と情報の流れといった、さらに大きな情報をビッグ・データとよび、そこから、統一解釈を導くためのマイニング研究が盛んに行われている。本研究ではまさにビッグ・データを用いて酵素のペプチド配列特徴を把握することに成功した。今後、さらにデータベースの充実を図るとともに、さらに、ビッグ・データ・バイオロジーに向けたマイニング技術の研究開発を展開したい。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計 6 件)

- ① Ohtana Y, Abdullah AA, Amin MA, Huang M, Ono N, Sato T, Sugiura T, Horai H, Nakamura Y, Morita AH, Lange KW, Kibinge NK, Katsuragi T, Shirai T, Kanaya S, Clustering of 3D-Structure Similarity Based Network of Secondary Metabolites Reveals Their Relationships with Biological Activities, *Mol. Inf.* 33, 2014, 790-801 査読有
DOI: 10.1002/minf.201400123
- ② Kibinge N, Ikeda S, Ono N, Amin MA, Kanaya S, Integration of residue attributes for sequence diversity characterization of terpenoid enzymes, *BioMed Res. International*, 2014, 2014, 753428.1-10 査読有
DOI: 10.1155/2014/753428

- ③ 池田俊, 桂樹哲雄, 小野直亮, 中谷淳至, 中村由紀子, 森田晶, 金谷重彦, オミックス・プラットフォーム: バイオ・ビッグ・データに挑む, *生物工学会誌*, 90 巻, 2014, 777-781、査読有
- ④ Ikeda S, Abe T, Nakamura Y, Kibinge N, Hirai Morita A, Nakatani A, Ono N, Ikemura T, Nakamura K, Altaf-UI-Amin M, Kanaya S., Systematization of the protein sequence diversity in enzymes related to secondary metabolic pathways in plants, in the context of big data biology inspired by the KNApSAcK motorcycle database, *Plant Cell Physiol.* 54, 2013, 711-27 査読有、DOI: 10.1093/pcp/pct041
- ⑤ Afendi FM, Okada T, Yamazaki M, Hirai-Morita A, Nakamura Y, Nakamura K, Ikeda S, Takahashi H, Altaf-UI-Amin M, Darusman LK, Saito K, Kanaya S., KNApSAcK family databases: integrated metabolite-plant species databases for multifaceted plant research, *Plant Cell Physiol.* 53, 2012, e1、査読有
DOI: 10.1093/pcp/pcr165.
- ⑥ Wada M, Takahashi H, Altaf-UI-Amin M, Nakamura K, Hirai MY, Ohta D, Kanaya S., Prediction of operon-like gene clusters in the *Arabidopsis thaliana* genome based on co-expression analysis of neighboring genes, *Gene*, 503, 2012, 56-64、査読有、DOI: 10.1016/j.gene.2012

[学会発表] (計 4 件)

- ① Yuki Otana, Shigehiko Kanaya, Tsuyoshi Shiraiishi, Takaaki Nishioka, MD. ALTAF-UL-AMIN, Tadao Sugiura, Naoaki Ono, Tetsuo Sato, Ming Huang, Tetsuo Katsuragi and Yukiko Nakamura, Clustering of 3D structure similarity based network of secondary metabolites to reveal their relationships with biological activities, GIW ISCB-Asia 2014, Dec. 2014、東京
- ② 金谷重彦, 森田 晶, 中村 由紀子, 佐藤 哲大, 小野 直亮, 杉浦 忠男, 桂樹 哲雄, 植物における二次代謝物の効能と立体構造の関係に関する研究, 第 37 回情報化学討論会 2014.11.28、豊橋市
- ③ 金谷重彦, 第 4 回生物起源微量ガスワークショップ, 謝物データベース KNApSAcK Family DB、2014.11.20、つくば市
- ④ Shigehiko Kanaya, Naoaki Ono, Tetsuo Sato, Aki Morita-Hirai, Yukiko Nakamura, Tadao Sugiura,

Ming Huang and MD.
ALTAF-UL-AMIN, KNApSAcK
Family DB towards Big Data
Biology, 12th Int'l Symposium on
Cytochrome P450 Biodiversity and
Biotechnology, Sep. 2014,京都

〔図書〕（計 0 件）

〔産業財産権〕

○出願状況（計 0 件）

○取得状況（計 0 件）

〔その他〕

ホームページ等

6. 研究組織

(1) 研究代表者

金谷 重彦 (KANAYA Shigehiko)

研究者番号：90224584

(2) 研究分担者 なし

(3) 連携研究者 なし