

天然物代謝のコンピュータ予測

L1ゼミ 2015年11月19日 15:05～16:35 B222 小寺研究室



天然物代謝のコンピュータ予測

L1ゼミ 2015年11月19日(木)15:05～16:35 B222 小寺研究室

来年度配属の学生と取り組みたいテーマ(例) 2～3名募集します。

- ・ 中間代謝物予測法の改良
- ・ 開環・閉環反応予測法の改良
- ・ 連続反応予測法の改良
- ・ 酵素反応アノテーション戦略の改良

緑が丘。コアタイム11時～17時。実験はしません。生き物には触りません。

プログラミングや数学は得意でなくていいです。配属後にゼミを通じて勉強しましょう。

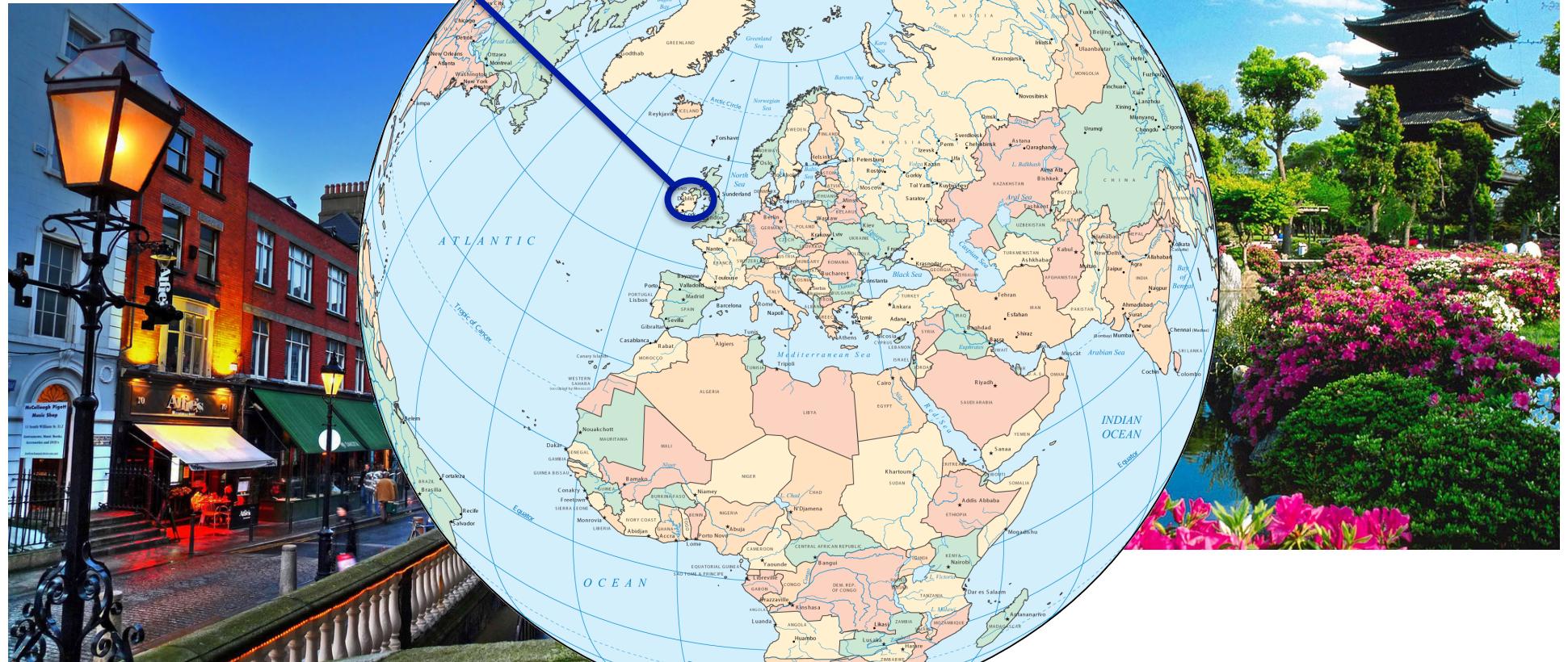
課題

- ・ 次の論文の要点を「落合先生フォーマットで」まとめて下さい(ただし、図は要りません)。
 - Masaaki Kotera, Yasuo Tabei, Yoshihiro Yamanishi, Toshiaki Tokimatsu and Susumu Goto. **Supervised de novo reconstruction of metabolic pathways from metabolome-scale compound sets.** *Bioinformatics* 2013, 29(13), i135-i144.
<http://bioinformatics.oxfordjournals.org/content/29/13/i135.full.pdf>
- ・ 提出方法:学籍番号と氏名を明記の上、maskot@bio.titech.ac.jp にメールで提出。
- ・ 提出期限:11月26日(木)15:00

職歴

2005-2008 アイルラン
ド共和国・ダブリン大
学トリニティカレッジ
ポスドク

2014- 東京工業大学 講師
2008-2014 京都大学 助教



国際的命名法を決める団体で活動しています

- **IUPAC = International Union of Pure and Applied Chemistry**
 - 国際純正・応用化学連合。化合物の体系名の命名法(IUPAC命名法)を決めている国際団体。
 - たとえば糖鎖など、まだ系統的命名法の整っていない化合物群は存在する。新規化合物の報告数の増加に伴い、新しい命名法を追加することは、今後も起こりうる。
- **IUBMB = International Union of Biochemistry and Molecular Biology**
 - 国際生化学・分子生物学連合。生化学命名法や、酵素を含む種々の命名法を制定している。
 - IUBMB's Enzyme List (Enzyme Commission (EC)番号とも呼ばれる)で酵素を分類・命名しており、現在も新規酵素は増加し続けている。
- **JCBN = IUPAC-IUBMB Joint Commission on Biochemical Nomenclature**
 - IUPACとIUBMBとの共同で、生化学分野のあらゆる国際標準命名法について議論している団体。**←酵素と糖鎖について議論に参加しています。**

EC番号とは何か？

- EC番号(酵素番号、Enzyme Commission numbers)は酵素を整理すべく反応形式に従ってECに続く4組の数字で表したもの(1961年～)。
- 分類は階層的でありECの接頭辞にピリオドで区切った続けた4個の番号“EC X.X.X.X”(Xは数字)による表記がなされる。
 - EC 1.X.X.X – オキシドレダクターゼ(酸化還元酵素)、酸化還元反応を触媒
 - EC 2.X.X.X – トランスフェラーゼ(転移酵素)、原子団(官能基など)のある分子から別の分子へ転移
 - EC 3.X.X.X – ヒドロラーゼ(加水分解酵素)、加水分解反応を触媒
 - EC 4.X.X.X – リアーゼ(脱離酵素)、原子団を二重結合あるいは、結合の解離の触媒
 - EC 5.X.X.X – イソメラーゼ(異性化酵素)、分子の異性体を作る
 - EC 6.X.X.X – リガーゼ(合成酵素)、ATPの加水分解エネルギーを利用して、2つの分子を結合

博士論文「酵素反応分類アルゴリズムの開発と
ゲノムスケール解析への応用」
(2005年3月、京都大学大学院理学研究科)



Published on Web 11/26/2004

**Computational Assignment of the EC Numbers for
Genomic-Scale Analysis of Enzymatic Reactions**

Masaaki Kotera, Yasushi Okuno,[†] Masahiro Hattori, Susumu Goto, and
Minoru Kanehisa*

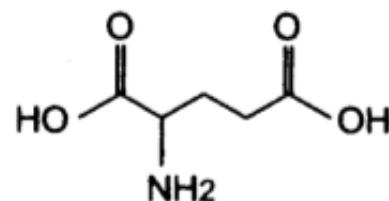
*Contribution from the Bioinformatics Center, Institute for Chemical Research, Kyoto University,
Uji, Kyoto 611-0011, Japan*

Received June 8, 2004; E-mail: kanehisa@kuicr.kyoto-u.ac.jp

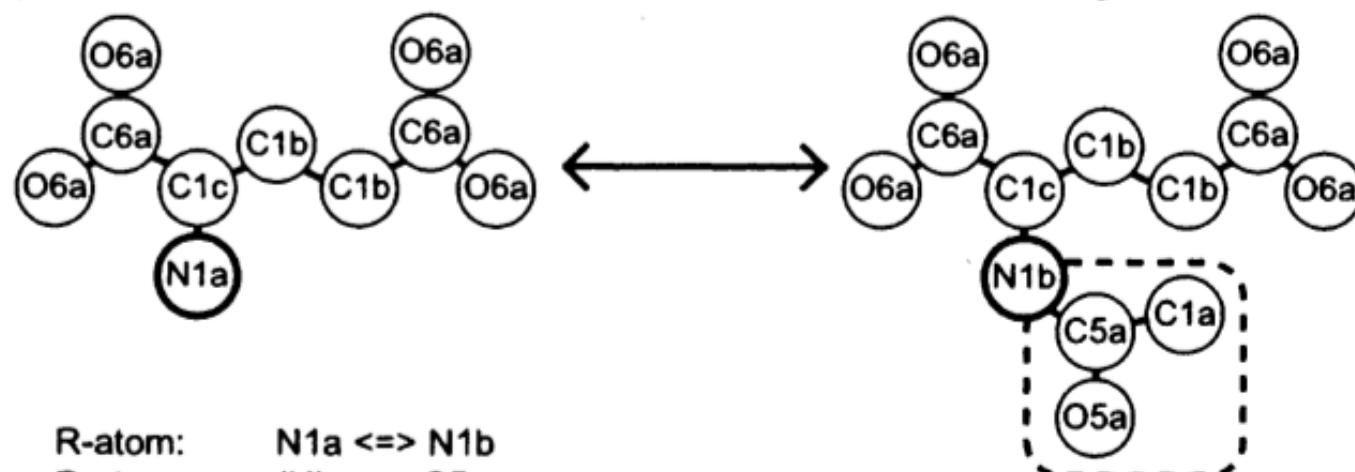
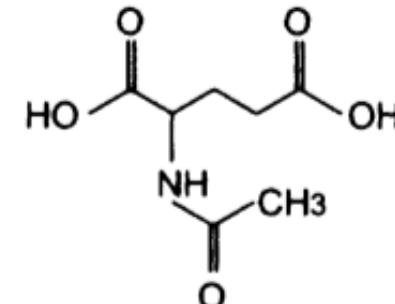
Abstract: The EC (Enzyme Commission) numbers represent a hierarchical classification of enzymatic reactions, but they are also commonly utilized as identifiers of enzymes or enzyme genes in the analysis of complete genomes. This duality of the EC numbers makes it possible to link the genomic repertoire of enzyme genes to the chemical repertoire of metabolic pathways, the process called metabolic reconstruction. Unfortunately, there are numerous reactions known to be present in various pathways, but they will never get EC numbers because the EC number assignment requires published articles on full characterization of enzymes. Here we report a computerized method to automatically assign the EC numbers up to the sub-subclasses, i.e., without the fourth serial number for substrate specificity, given pairs of substrates and products. The method is based on a new classification scheme of enzymatic reactions, named the RC (reaction classification) number. Each reaction in the current dataset of the EC numbers is first decomposed

化学構造変換パターンの文字列表現

C00025



C00624



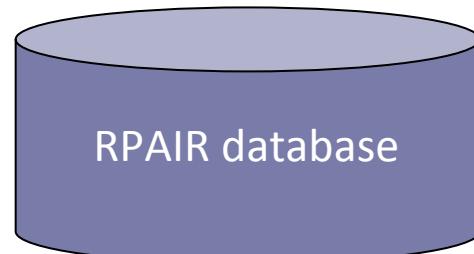
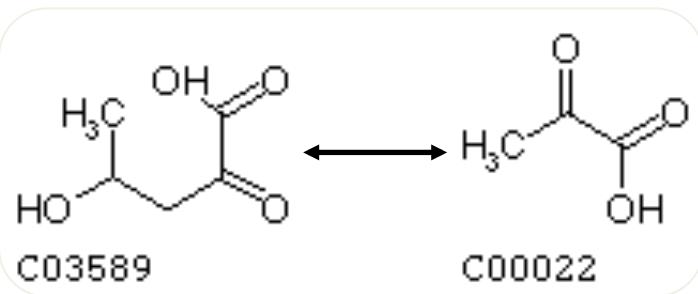
R-atom: N1a <=> N1b
D-atom: (H) <=> C5a
M-atom: C1c <=> C1c

RDM (C00025, C00624) = “N1a-N1b: *-C5a:C1c-C1c”



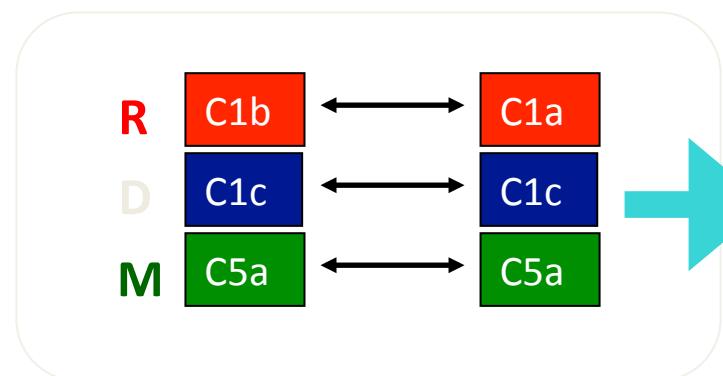
化学構造比較からの酵素予測

[Kotera et al, 2004; Yamanishi et al, 2009; Moriya et al, 2016]

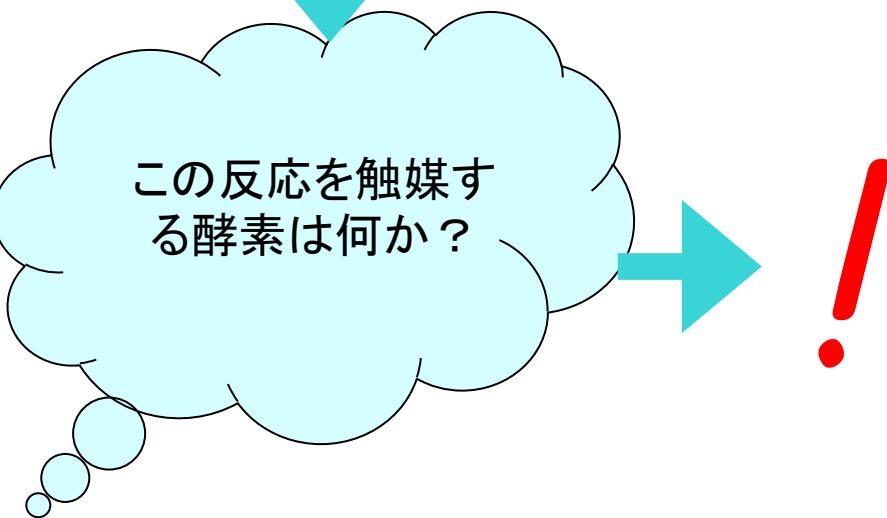


RPAIR database

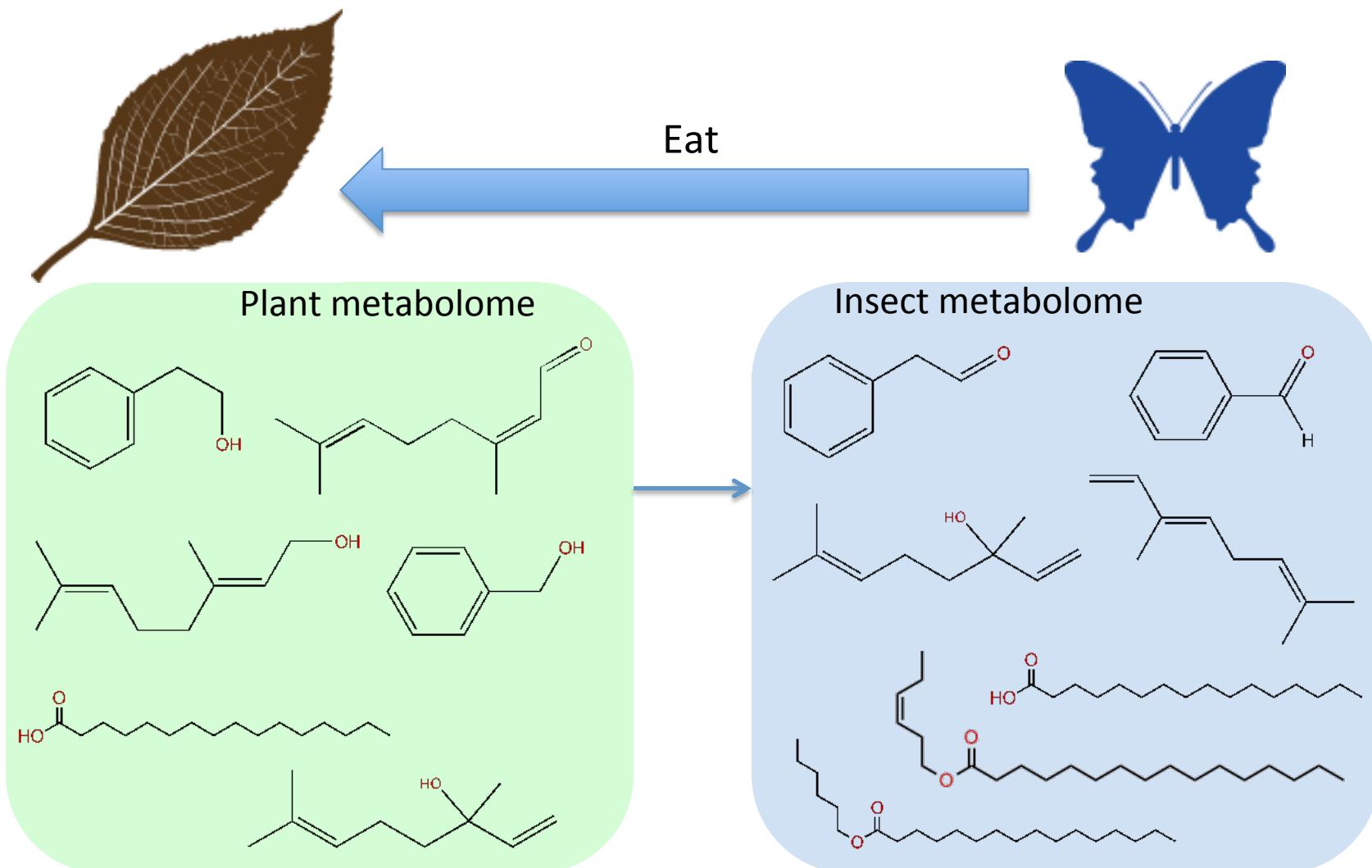
既知酵素反応の化学変換
パターンデータベース



RDM chemical transformation pattern



最近のマイブーム: 草食系



ところで、



動かない植物が、
どうやって身を守っているか、
考えたことはありますか？



日本の毒草の中で有名なもののだけピックアップ

トリカブト



ドクゼリ



ドクウツギ

(子どもが食べて中毒を起こしやすい)



ヒガンバナ



ハシリドコロ

(フキノトウなどと間違われやすい)



チョウセンアサガオ
(別名:エンジェルトランペット)



オナモミ
(別名:くつつきむし)



キョウチクトウ



スズラン

(これを活けた水でも中毒)



オトギリソウ



あなたがリツイート



ultraviolet @raurublock · 3月10日

タマネギ/チョコレート/コーヒーは犬猫に有毒だが、犬猫が弱いのではなく、ヒトが哺乳類の中でも例外的に毒に強いのだ。そもそも植物は哺乳類から身を守る毒として硫化アリル/テオプロミン/カフェインを身につけたのに、このろくでもないサルの子孫は「うまいうまい」と言って好んで食べ始める始末



13,156

8,456

...

あなたがリツイート



FujiiACC @FujiiACC · 3月8日

犬猫にテーブルから少しだけ、おやつを与える飼い主さんは多いですが、与えてはいけないもの、として玉ねぎ、にんにくは有名です。カフェイン、キシリトール、チョコレート、ぶどう、レーズン、アポカド、マカダミアナッツ、ナッツ類、熟していないトマト、ほうれん草、キノコ類には注意が必要です。



260

234

...

あなたがリツイート



kyouko_sugisita @kyouko_sugisita · 3月10日

ネギ、チョコレート、コーヒーは犬猫に限らず殆どの動物がダメですよ。

そういうや、チョコレートソフトをおいしそうに食べる猫が可愛い、とかTV番組で放送されてたな。最近TV番組で「適切な情報」ってお目にかかる気がする。



1

2

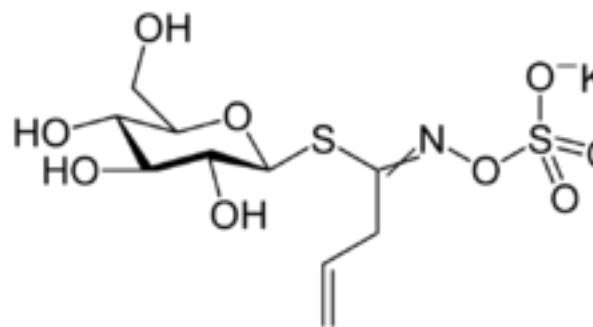
...

Chemical ecology

化学生態学

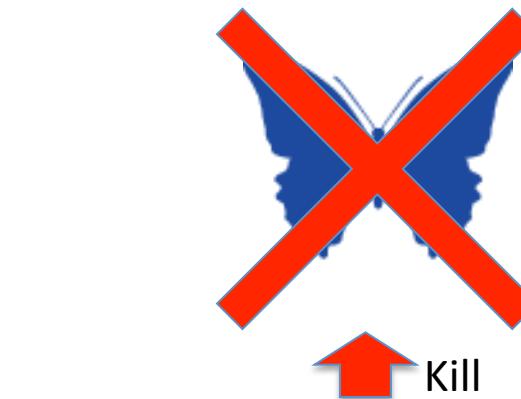


Biosynthesis and store

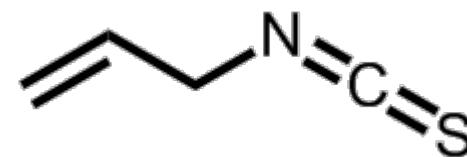


Precursor of a toxin

Degradation upon eating



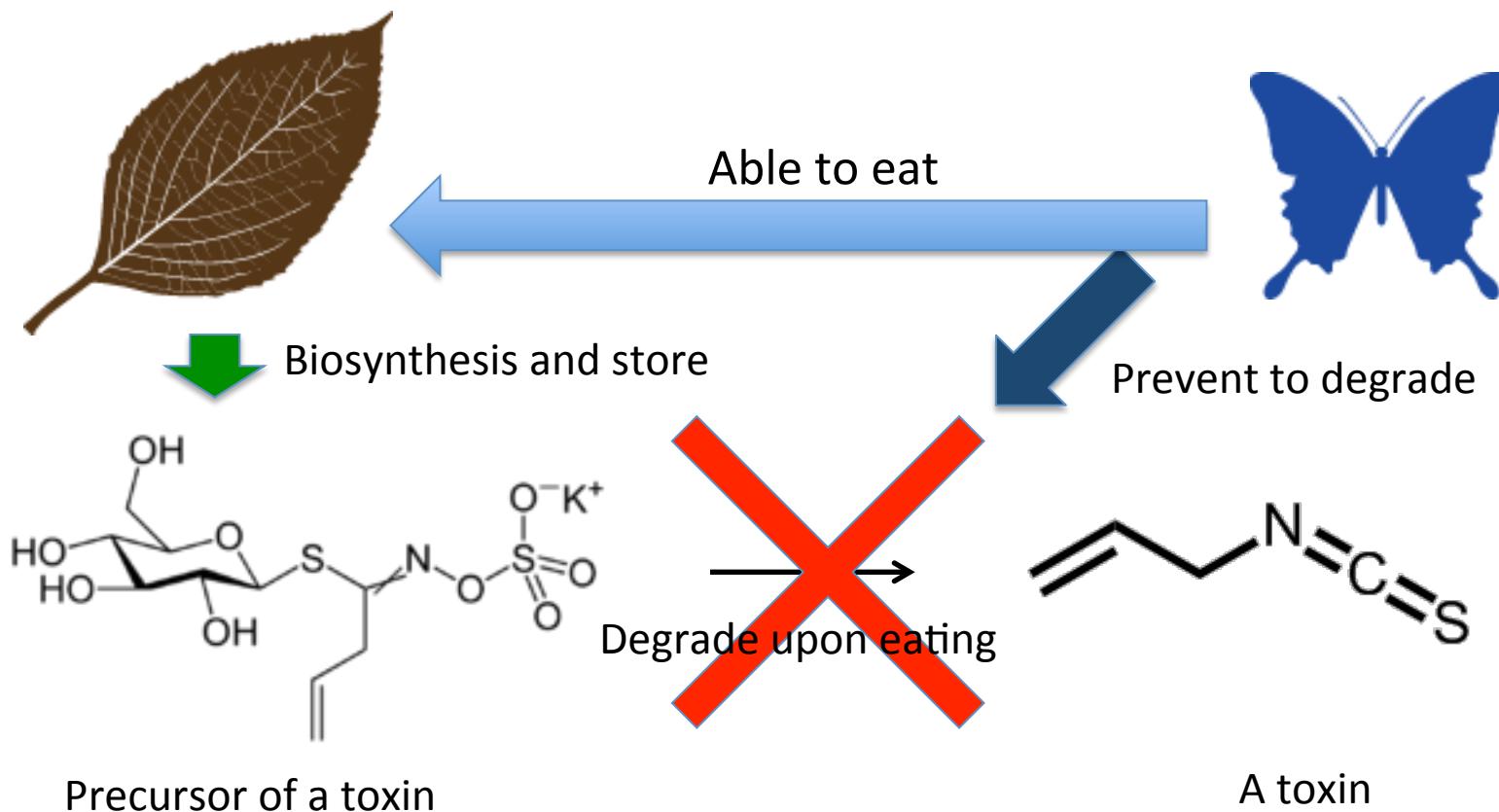
Kill



A toxin

Chemical ecology

化学生態学

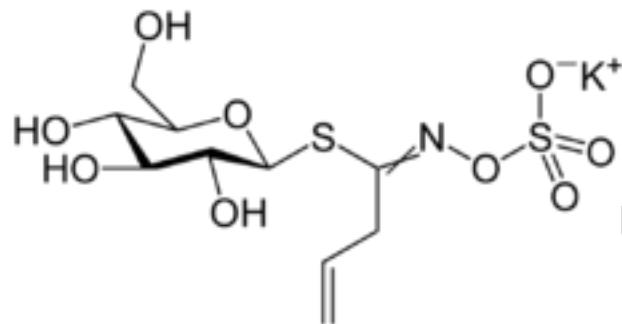


Chemical ecology

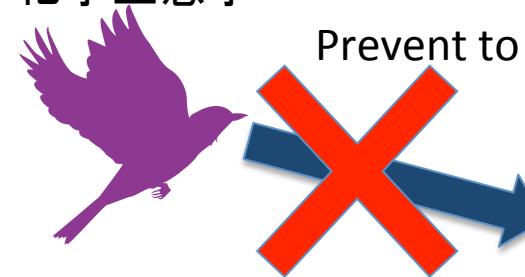
化学生態学



Biosynthesis and store



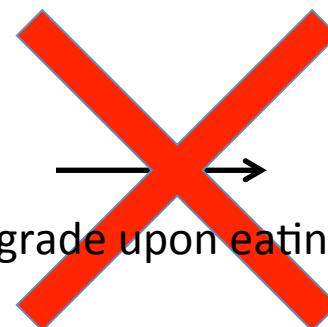
Precursor of a toxin



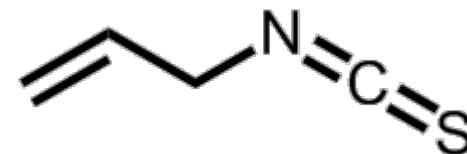
Prevent to be eaten



Prevent to degrade and store

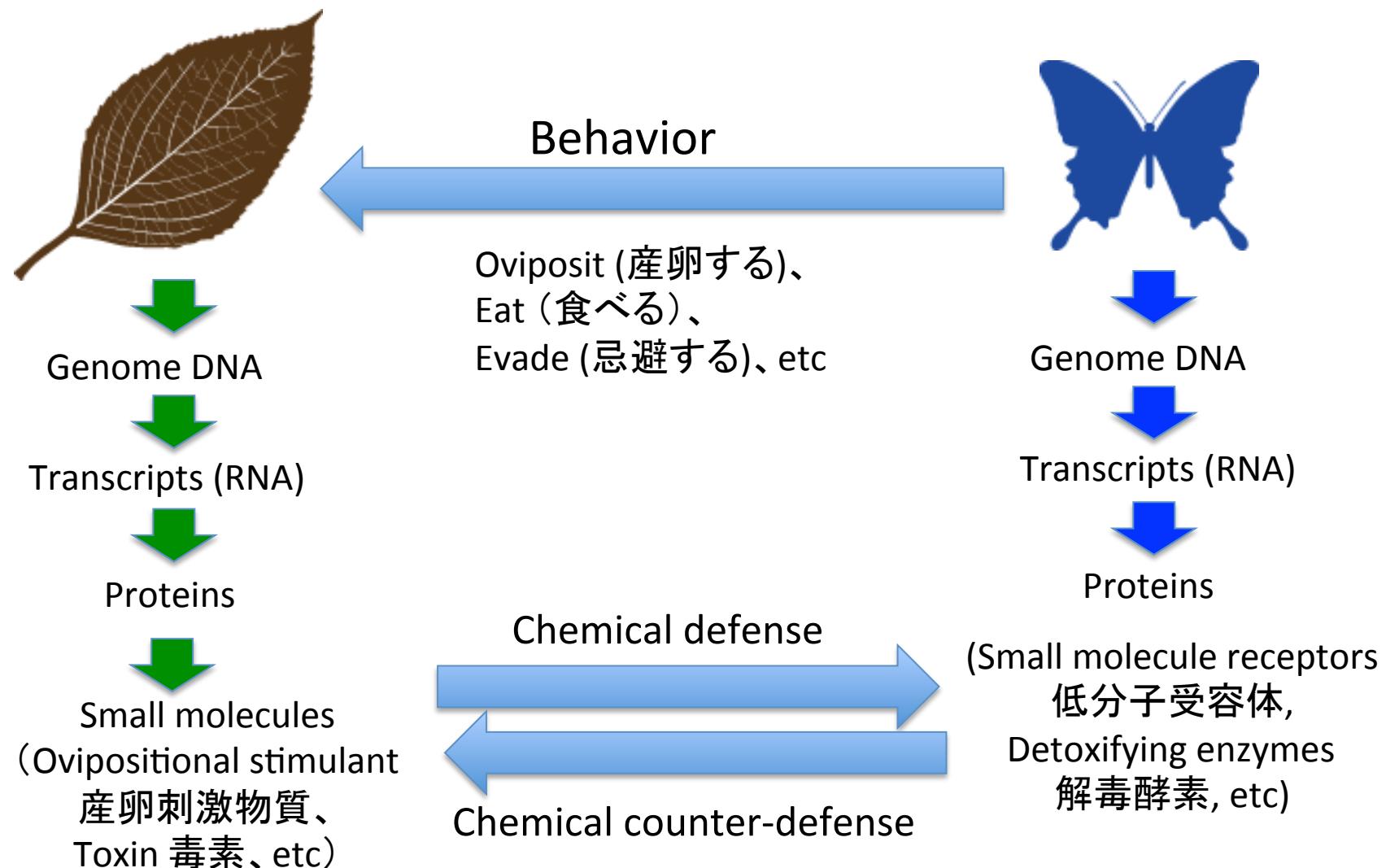


Degradation upon eating



A toxin

化学的軍拠競争の分子基盤



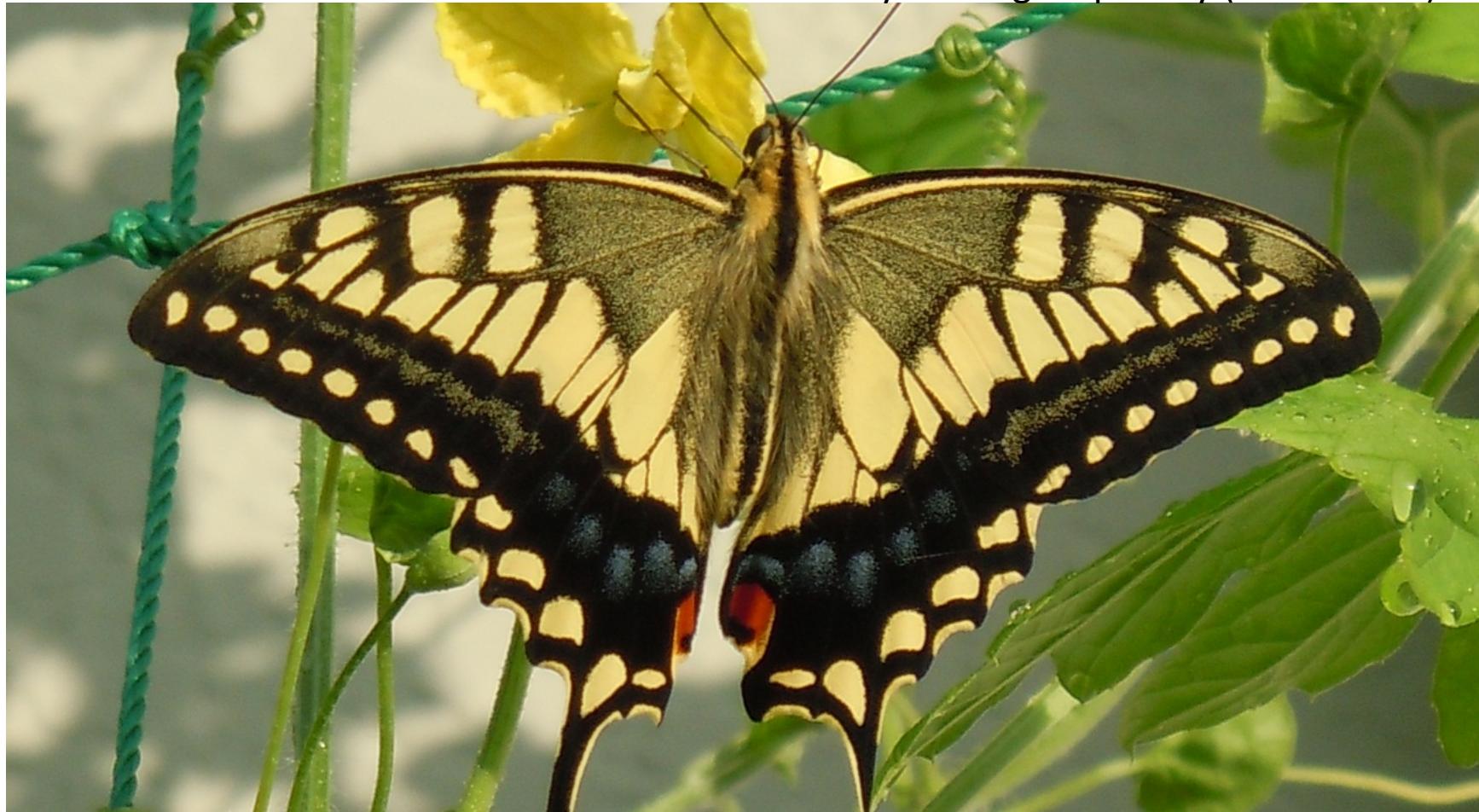
ナミアゲハ (*Papilio xuthus*)

Mainly feeding on citrus (ミカン科植物)



キアゲハ (*Papilio machaon*)

Mainly feeding on parsley (セリ科植物)



モンシロチョウ (*Pieris rapae*)

Mainly feeding on cabbages (アブラナ科植物)



Community ecology

群集生態学



Eat, oviposit, evade, etc.



Eat, oviposit, evade, etc.



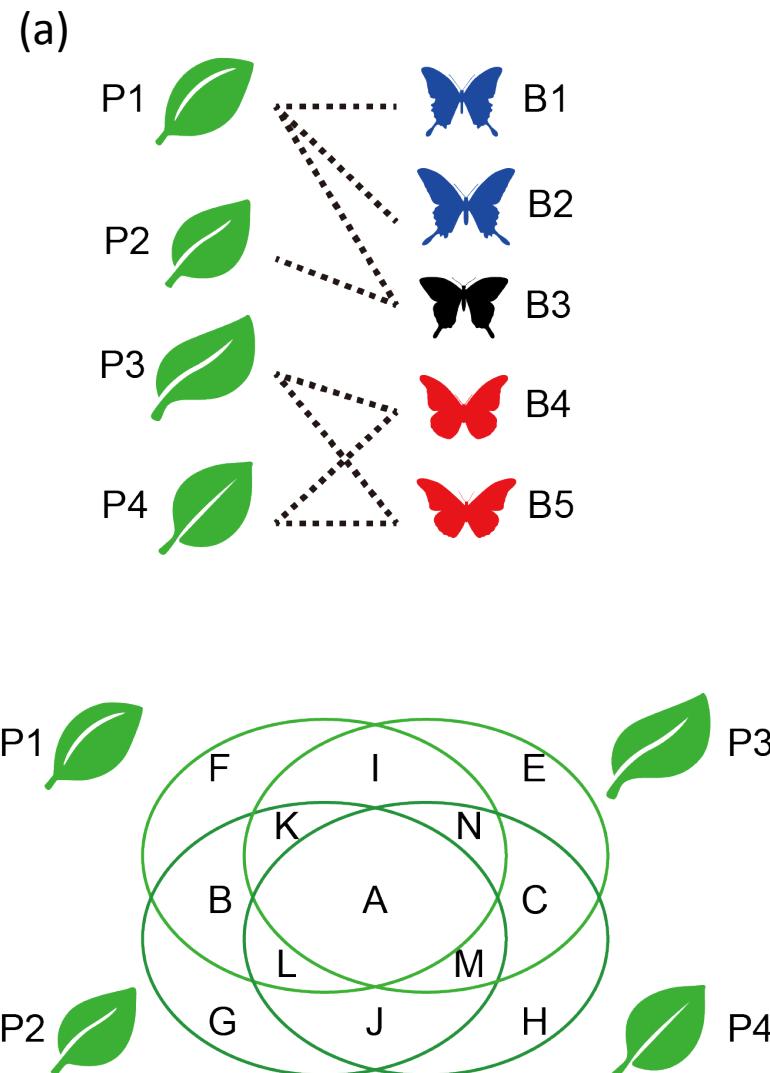
Eat, oviposit, evade, etc.



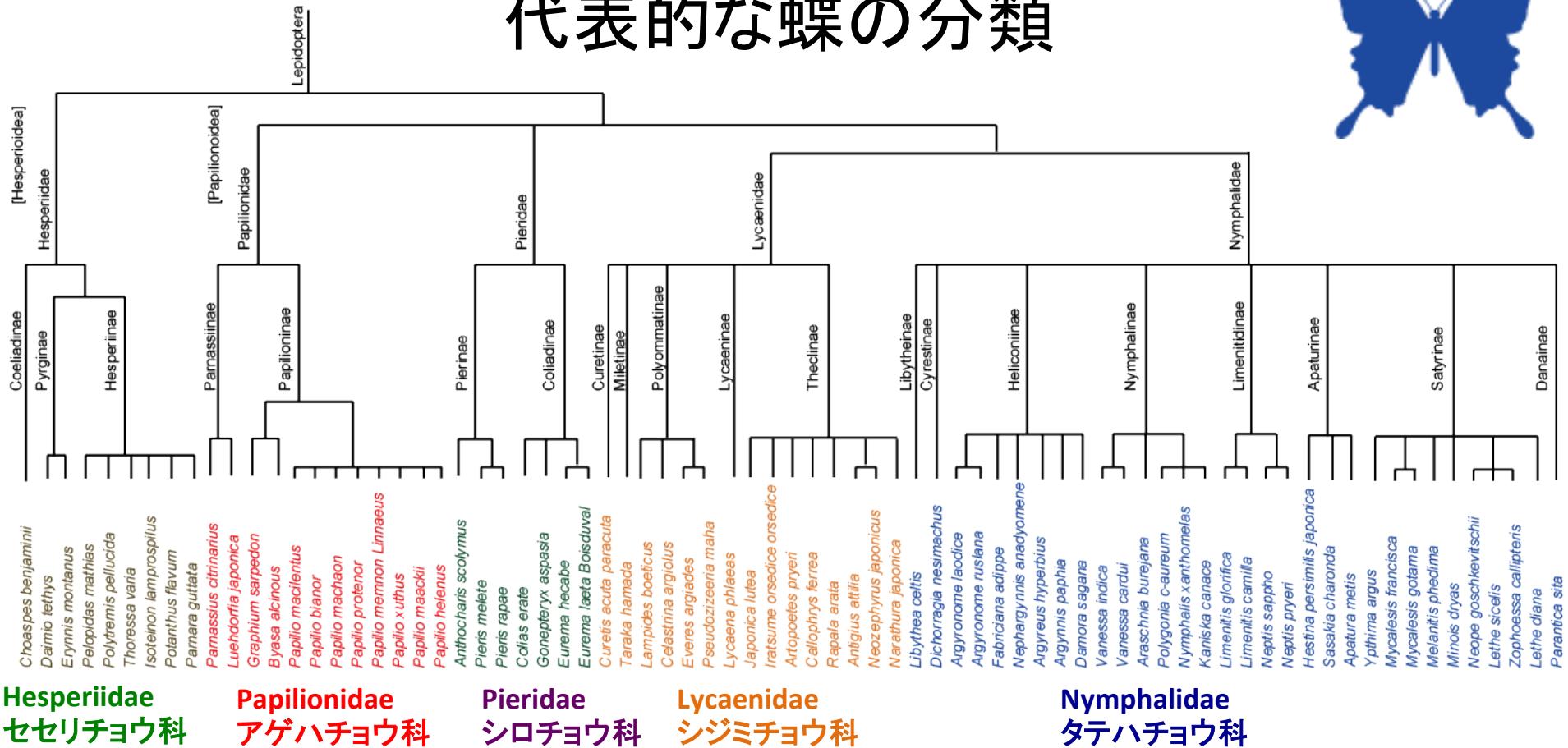
Eat, oviposit, evade, etc.



(a) 草食ネットワーク、(b) 食草マトリクス、(c) 共通化合物



代表的な蝶の分類



コチャバネセセリ
Thoressa varia



ナミアゲハ
Papilio xuthus



モンシロチョウ
Pieris rapae

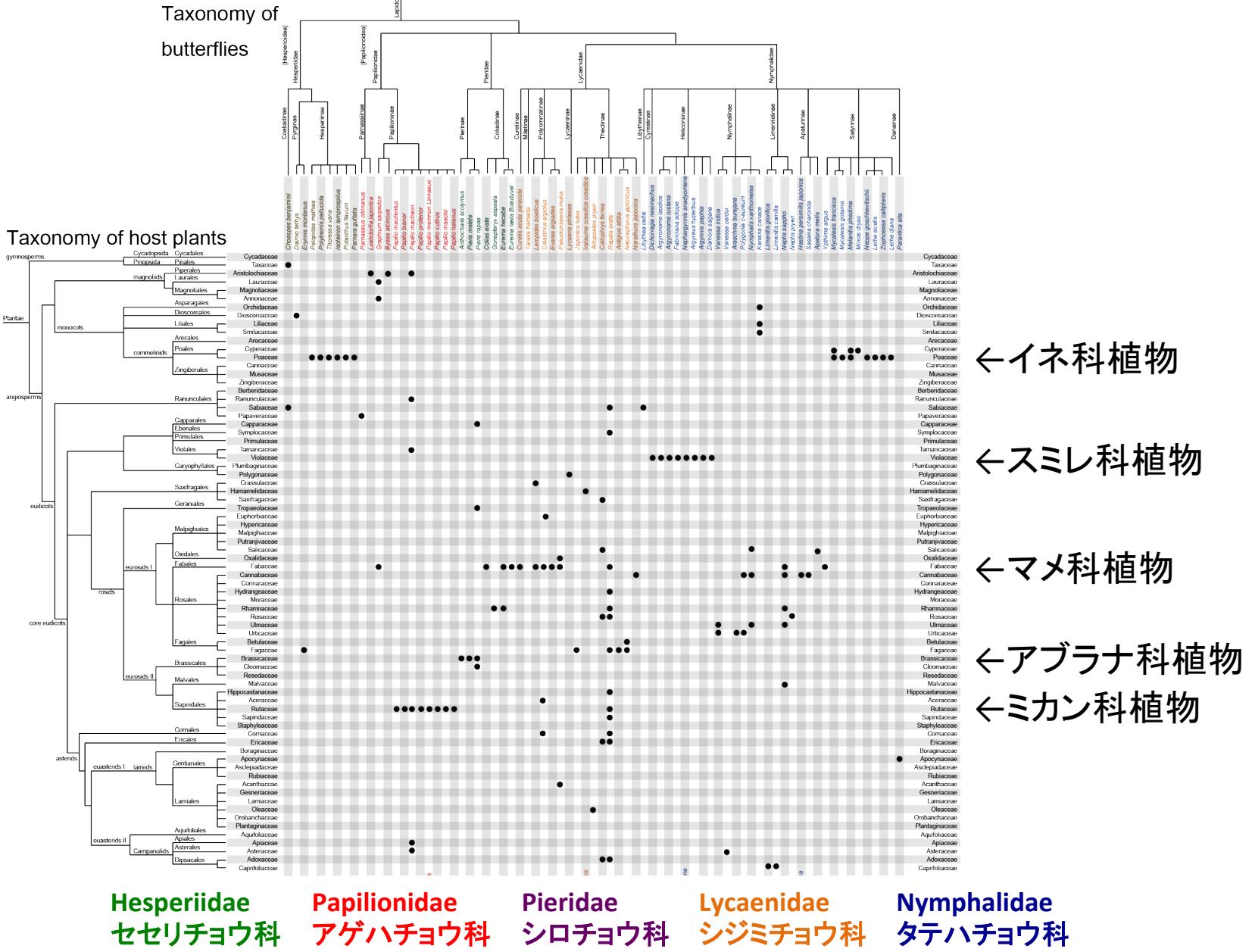


ベニシジミ
Lycaena phlaeas



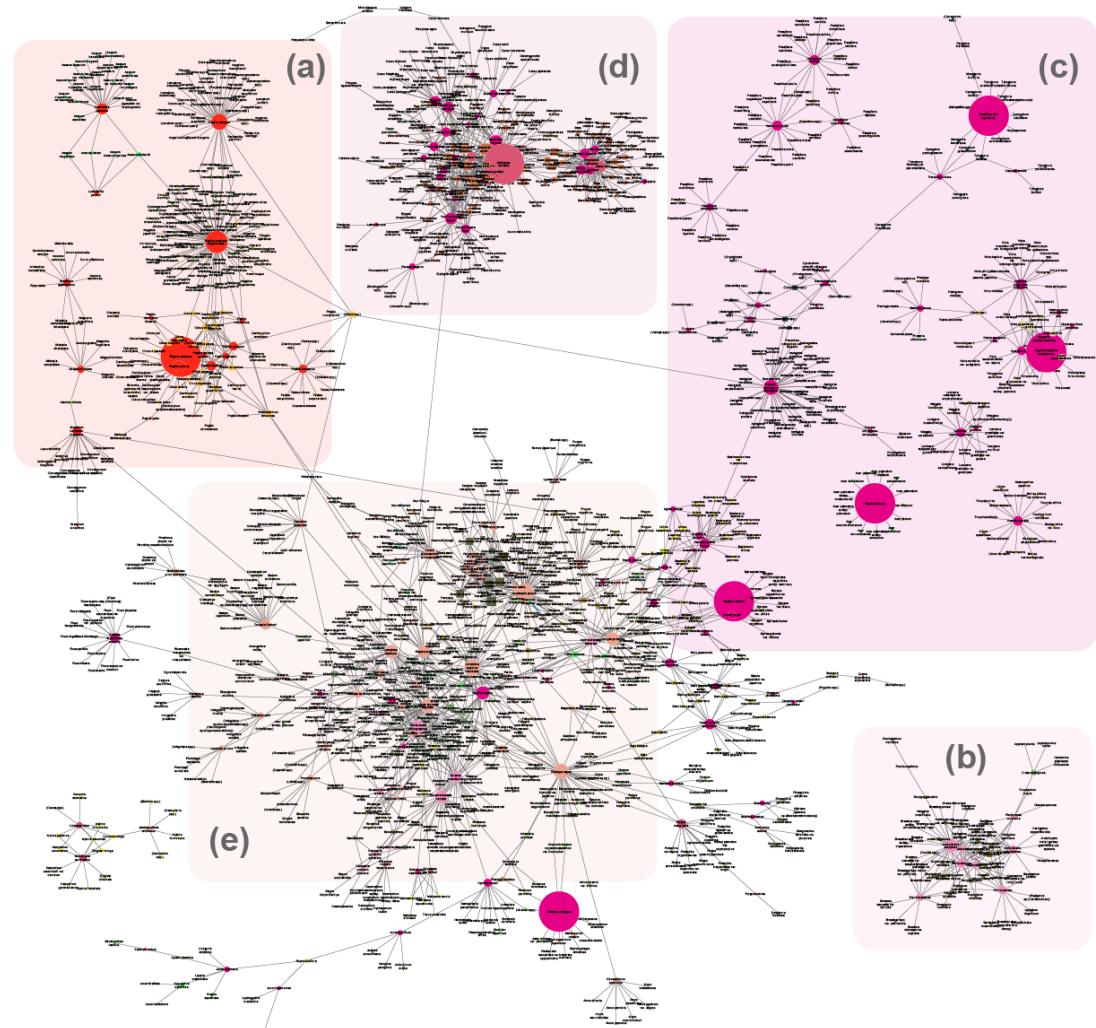
アサギマダラ
Parantica sita

代表的な食草関係だけ描いてみると

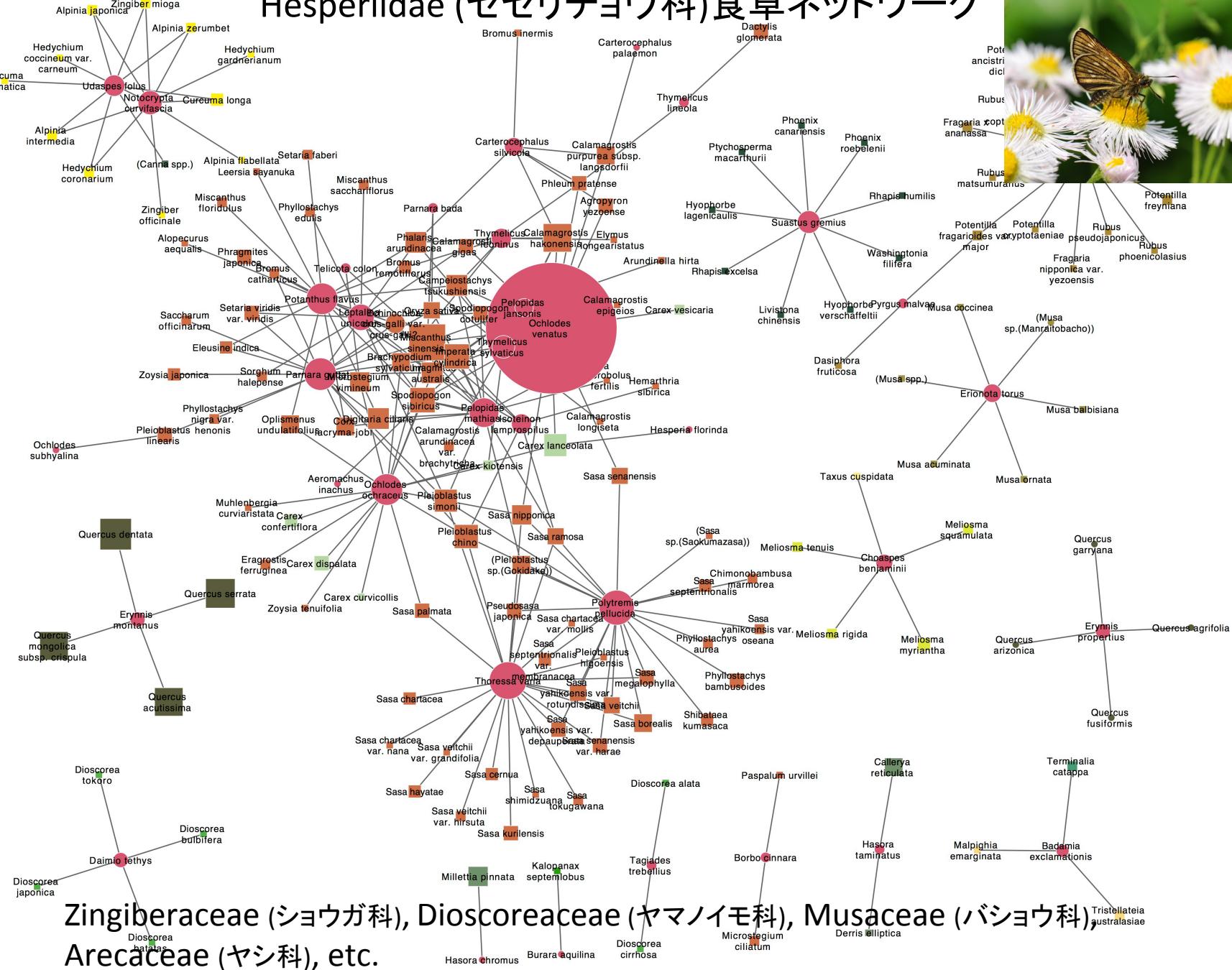


本気で全部集めると

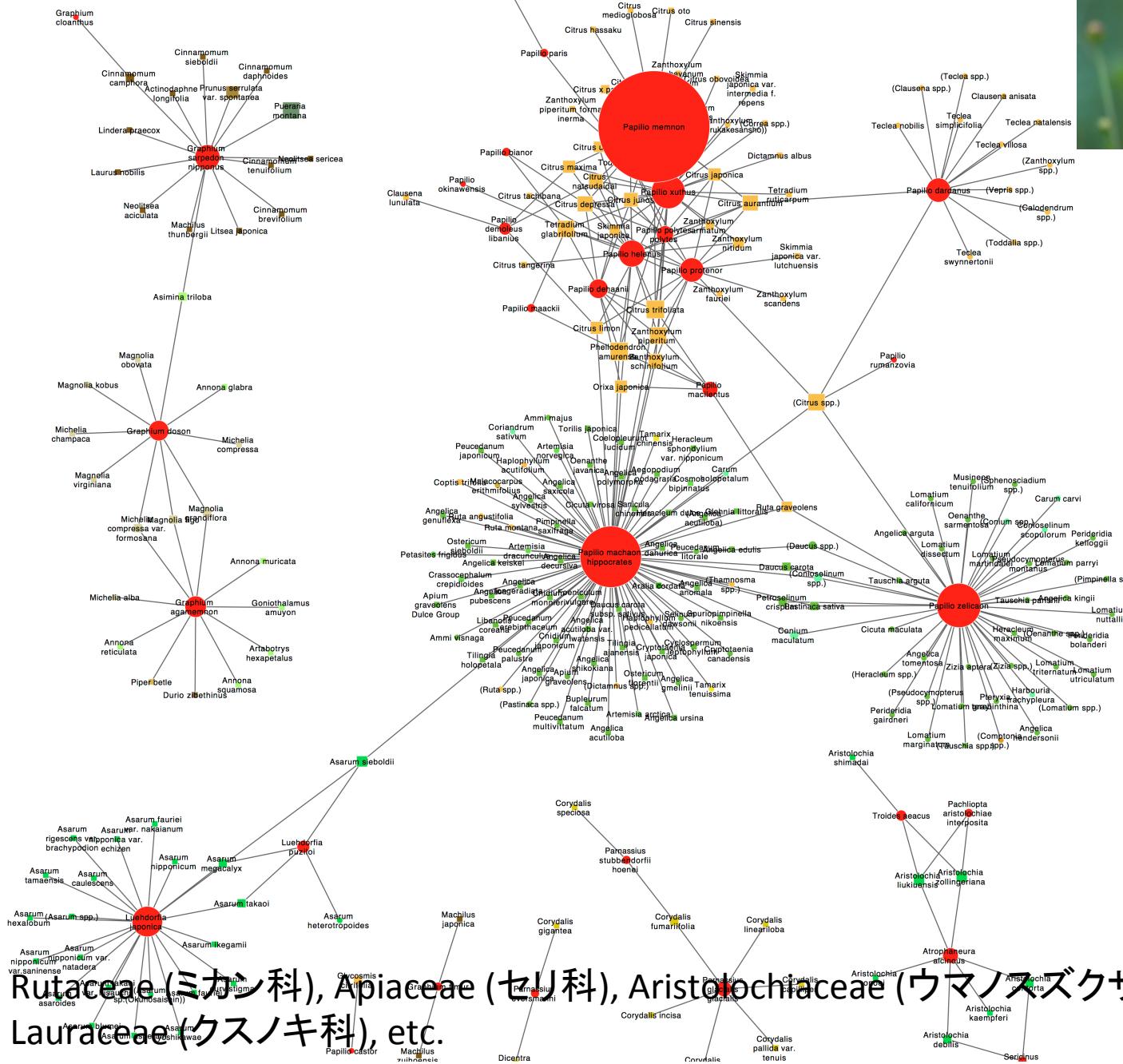
- チョウ 545 種
 - Hesperiidae セセリチョウ科
 - Papilionidae アゲハチョウ科
 - Pieridae シロチョウ科
 - Lycaenidae シジミチョウ科
 - Nymphalidae タテハチョウ科
- 食草・食樹 1922 種
 - Poaceae イネ科
 - Rutaceae ミカン科
 - Brassicaceae アブラナ科
 - Fabaceae マメ科
 - Violaceae スミレ科
 - etc. など
- 3435 捕食・被食関係



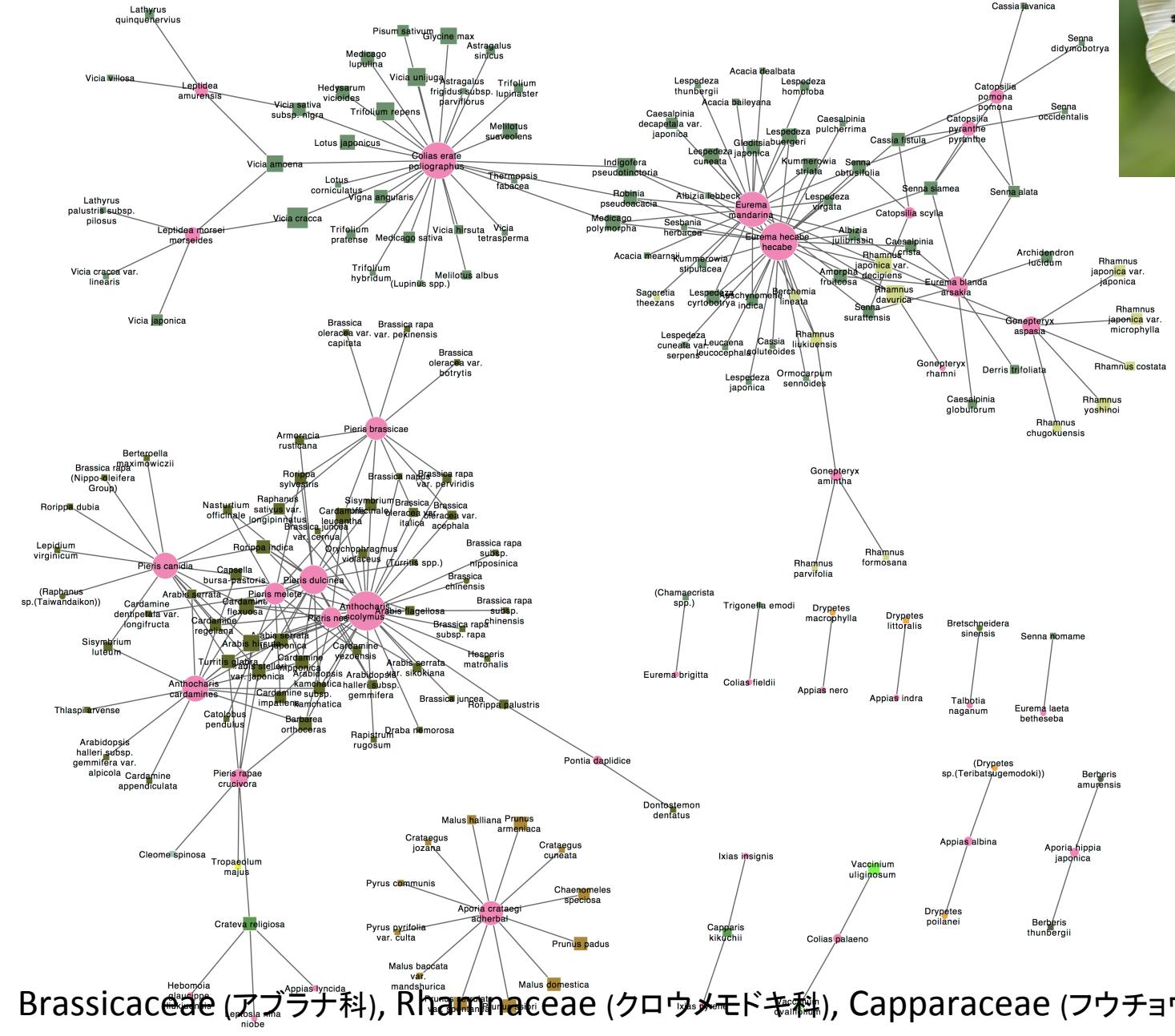
Hesperiidae (セセリチョウ科) 食草ネットワーク



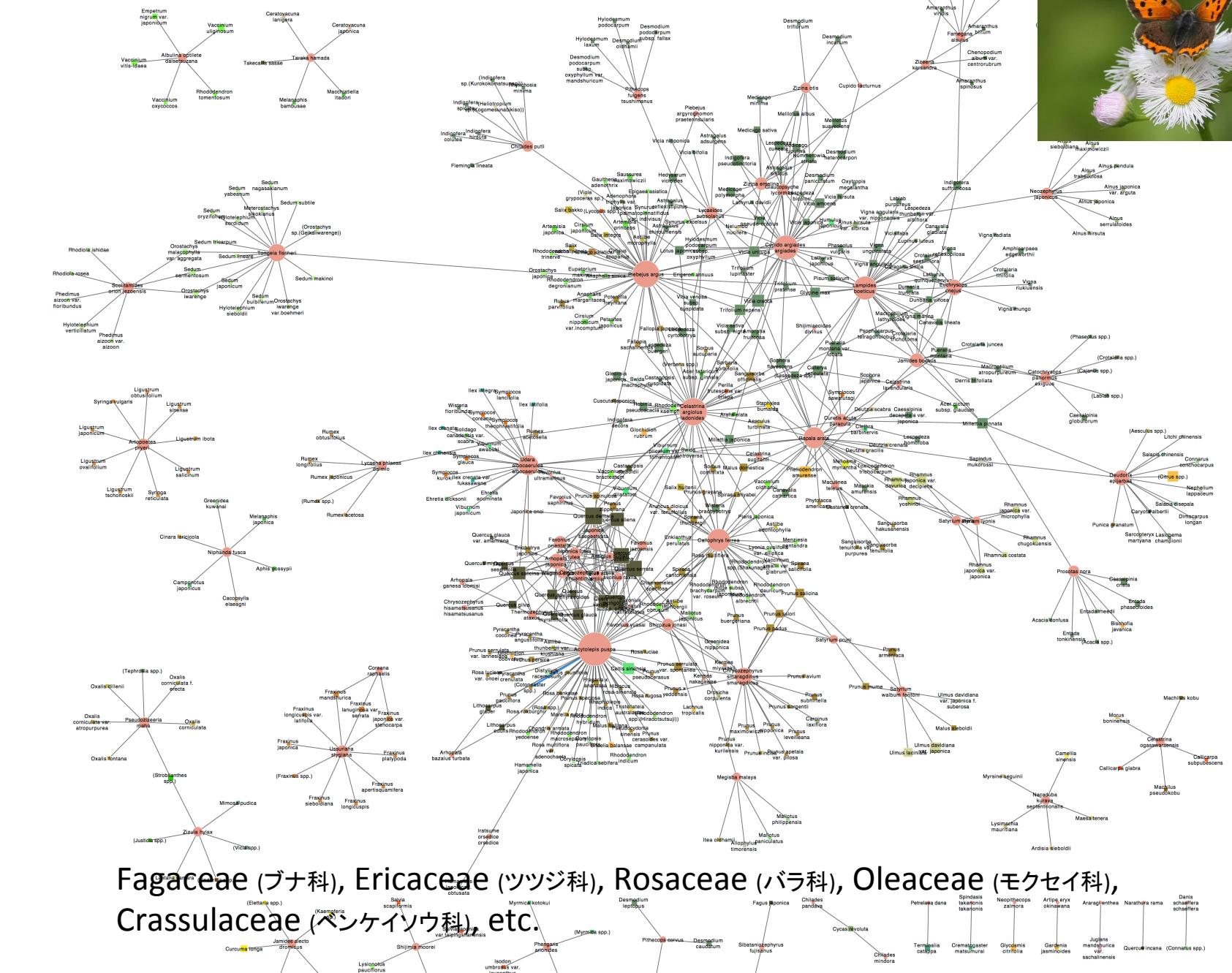
Papilionidae (アゲハチョウ科) 食草ネットワーク



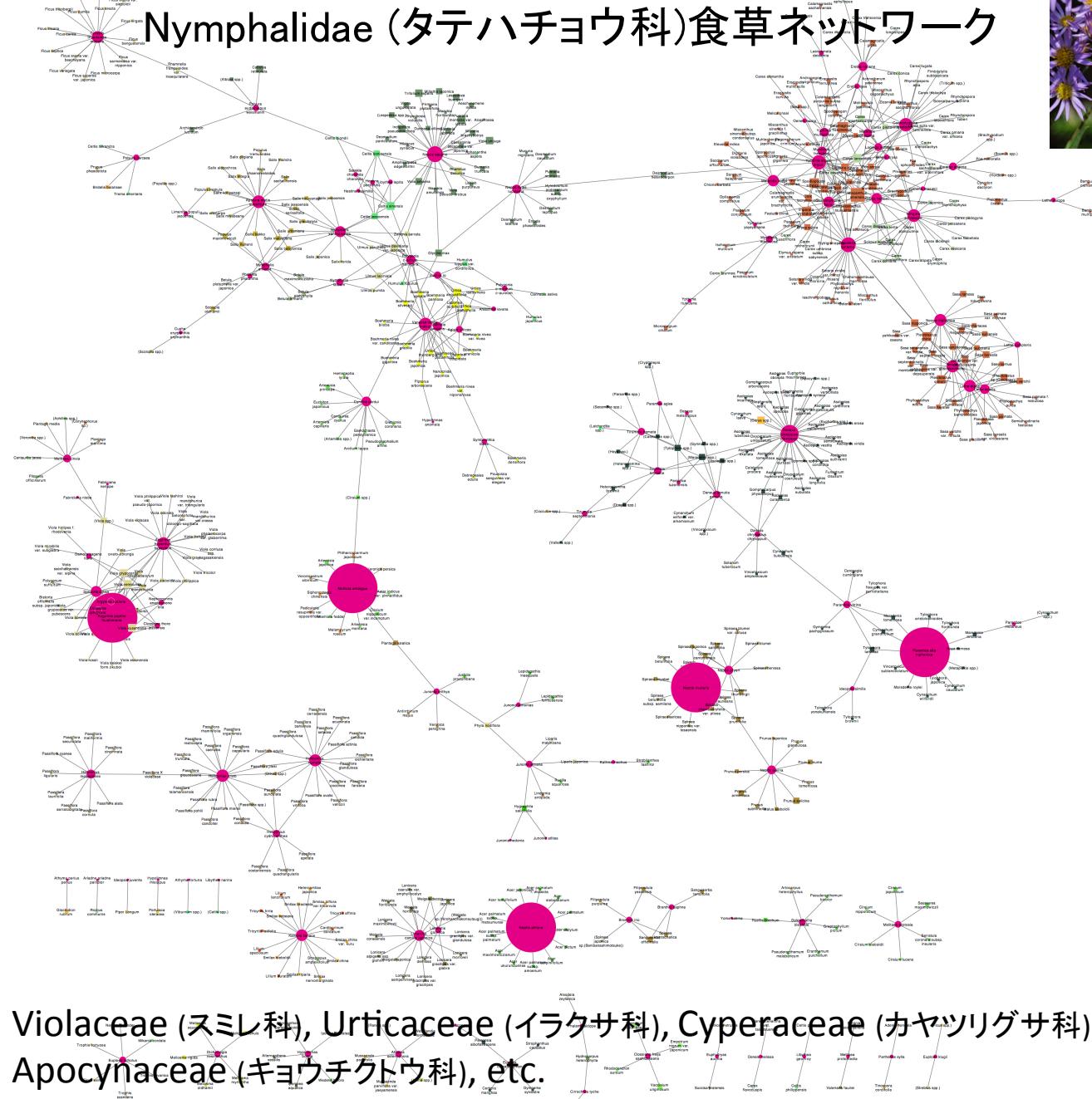
Pieridae (シロチョウ科) 食草ネットワーク

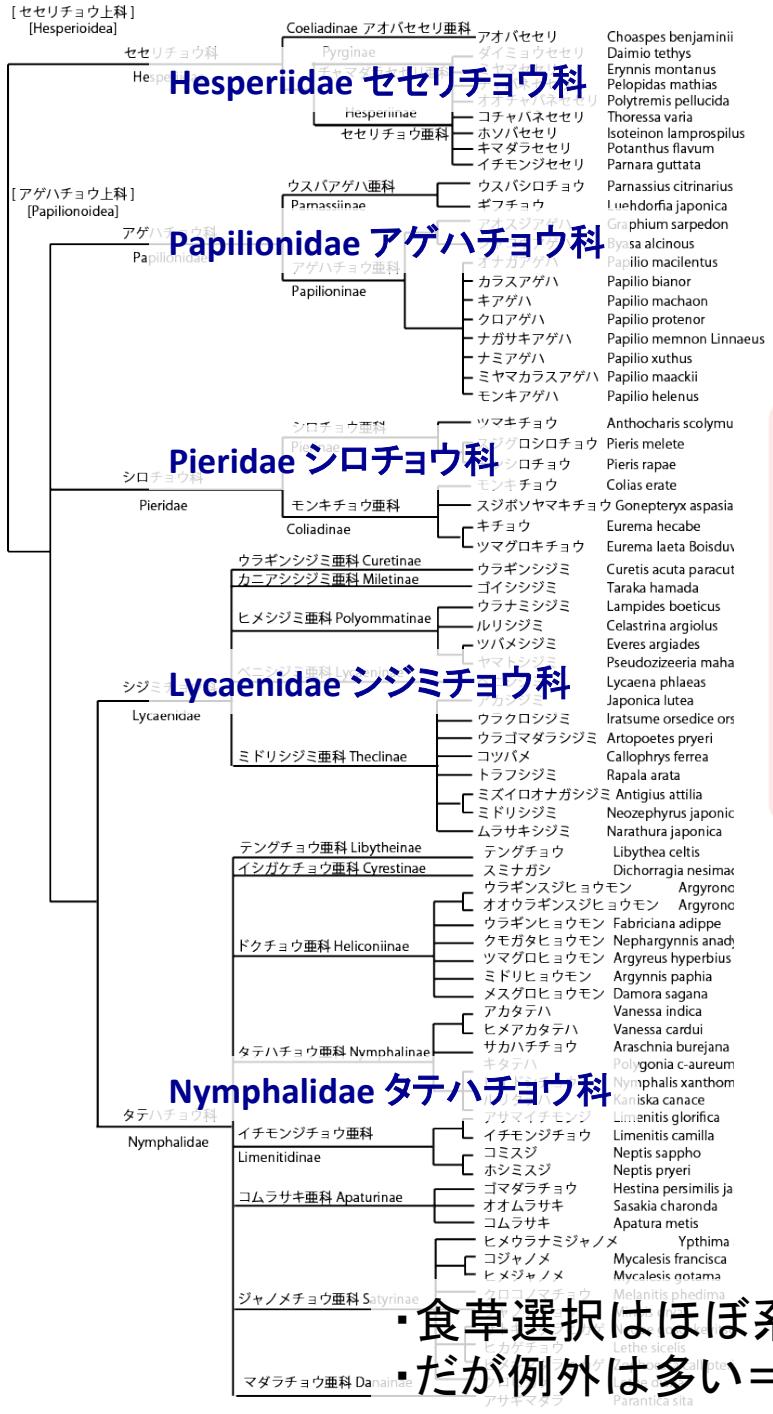


Lycaenidae (シジミチョウ科) 食草ネットワーク



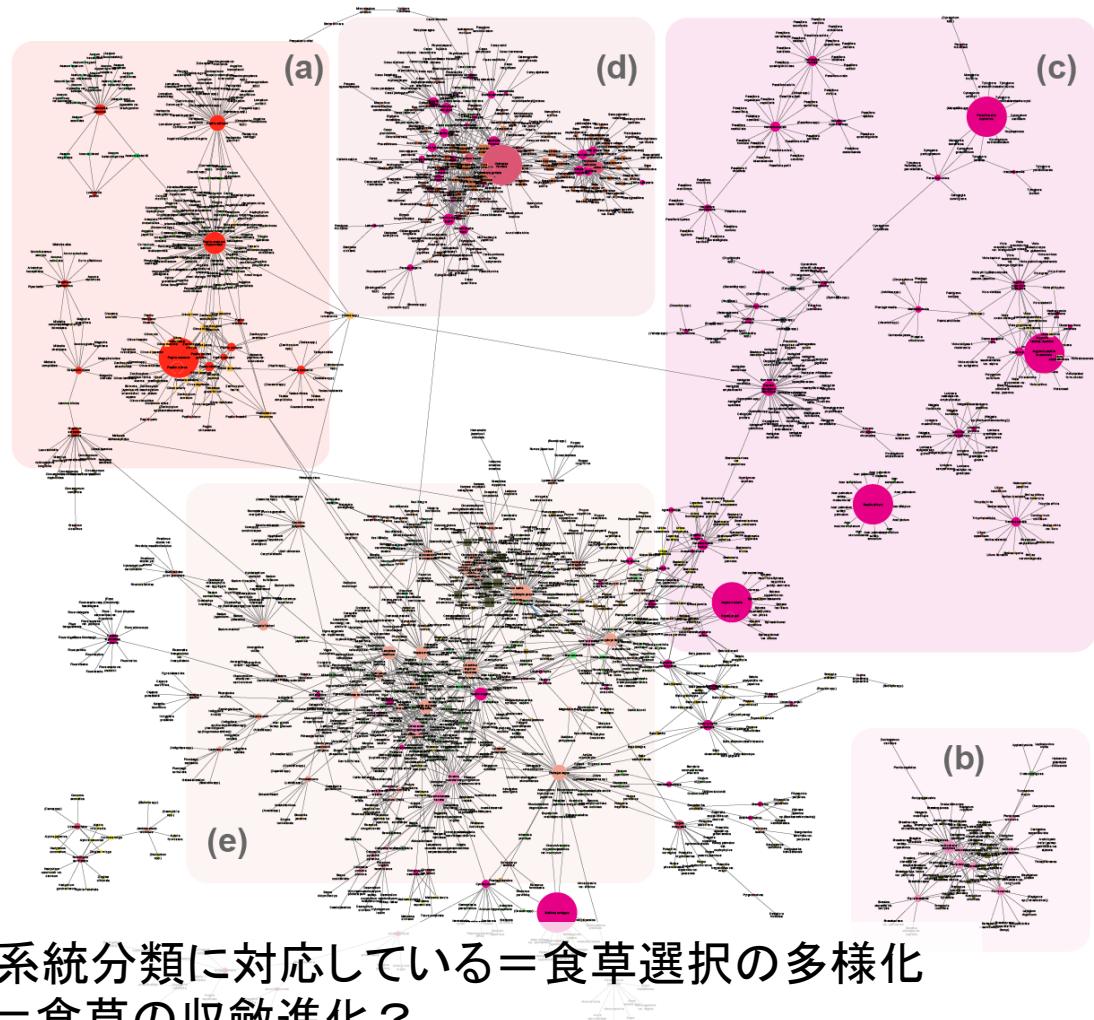
Nymphalidae (タテハチョウ科)食草ネットワーク

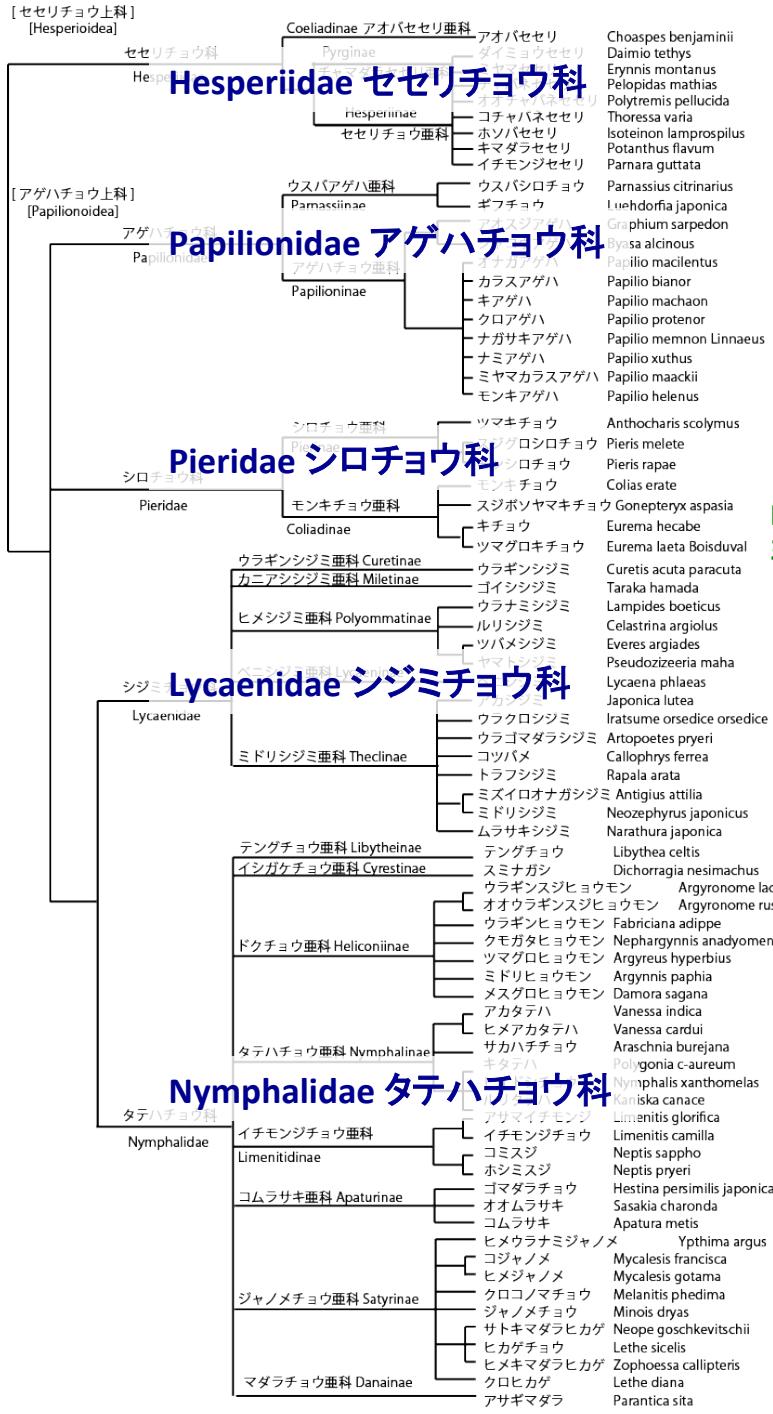




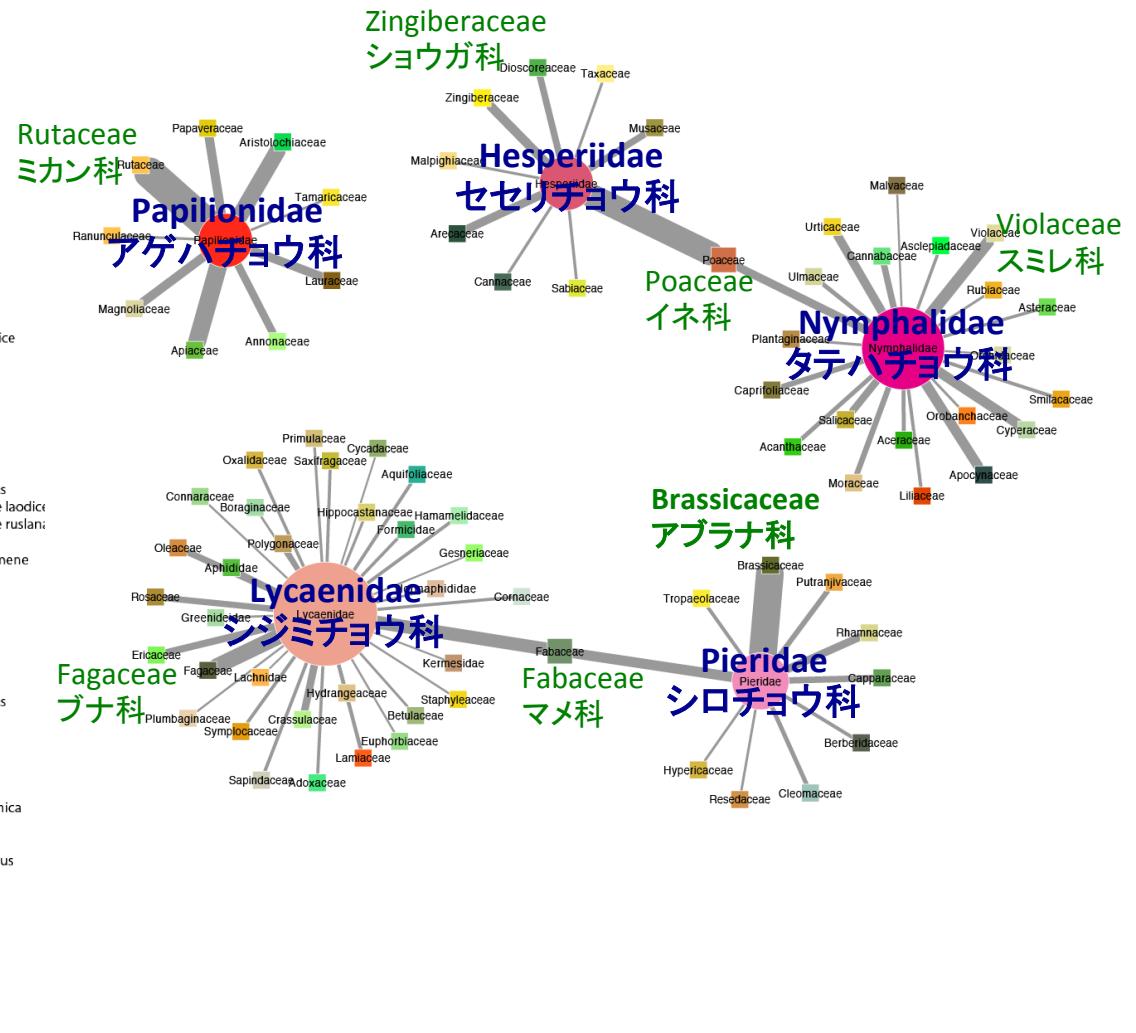
統合すると：

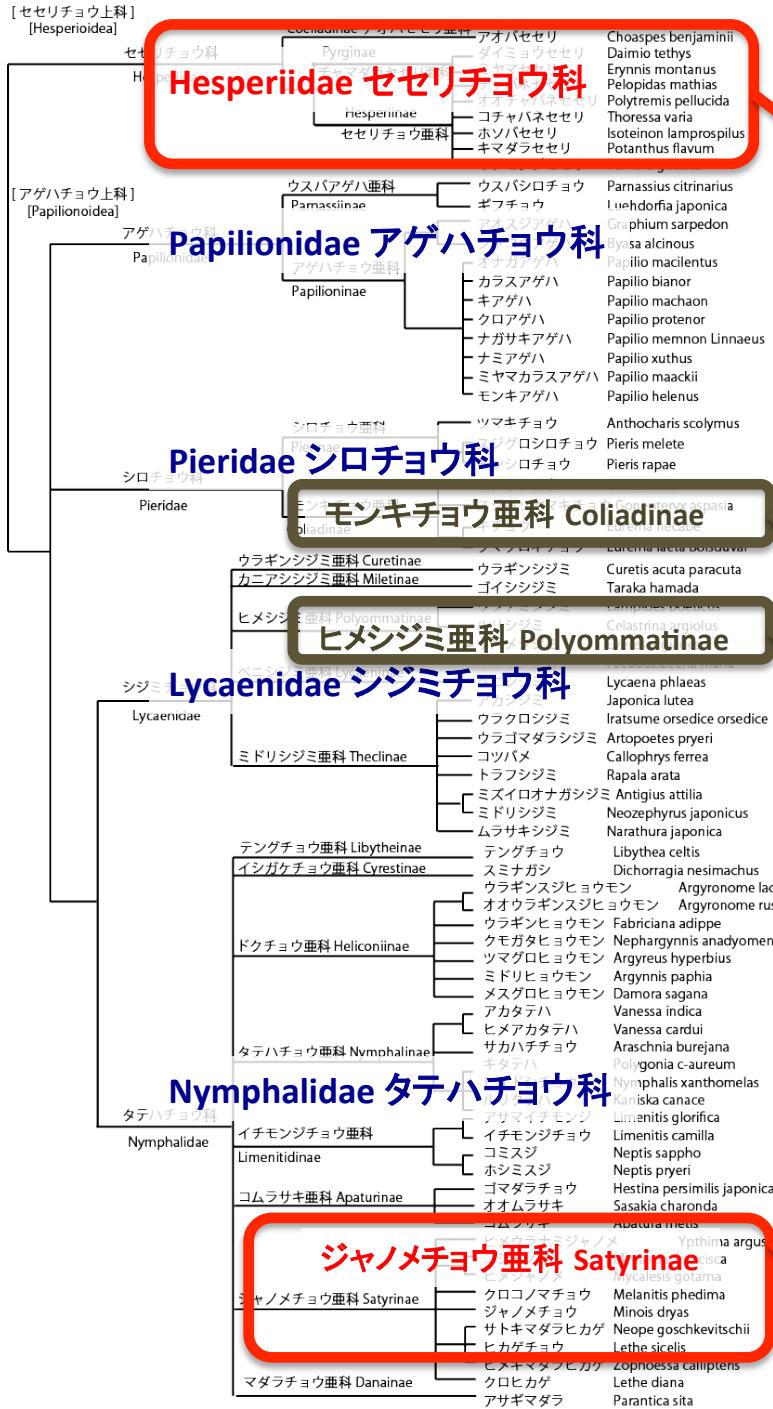
- a) アゲハチョウ科特異的ネットワーク
 - b) シロチョウ科特異的ネットワーク
 - c) タテハチョウ科特異的ネットワーク
 - d) セセリチョウ科+タテハチョウ科ネットワーク
 - e) シジミチョウ科+シロチョウ科ネットワーク





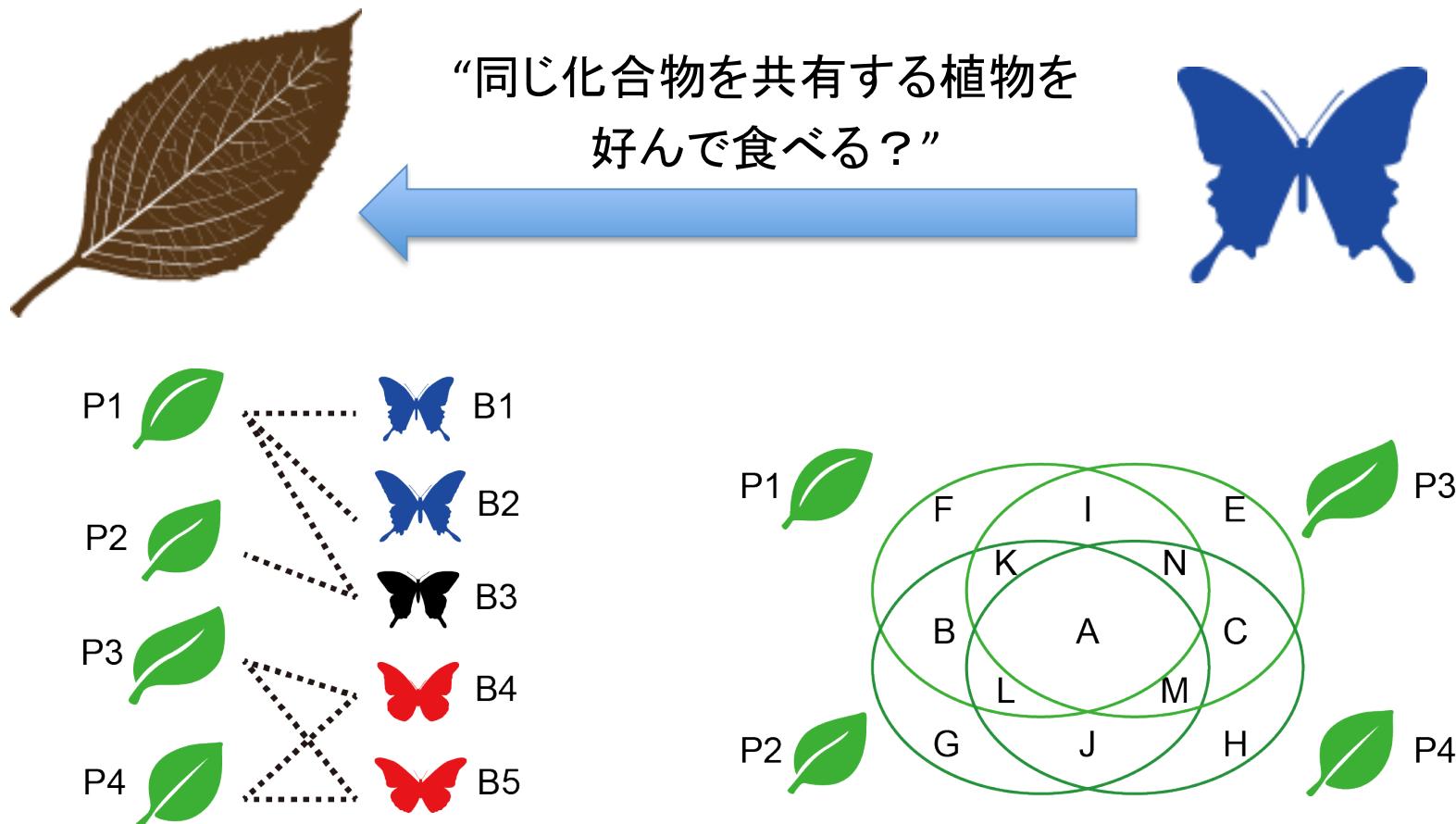
科レベルで眺める





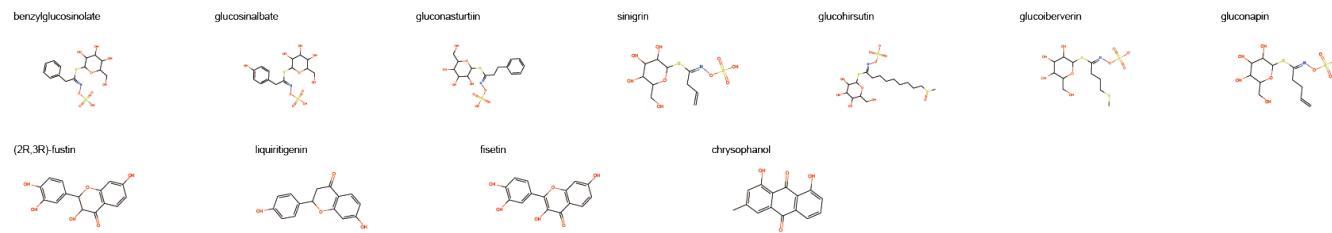
~~食草の収斂進化(?)が亞科レベルで起こっている(約550万年前)~~

食草選択に対する植物化合物の影響



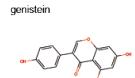
KNApSAcKデータベースと食草関係データを使った仮説検定の結果、微妙に支持する結果。
(ただし植物種によって影響度は異なる)

(a) Pieridae-specific シロチョウ科食草特異的化合物



シジミチョウ科食草特異的 アゲハチョウ科食草特異的

(b) Lycaenidae-specific

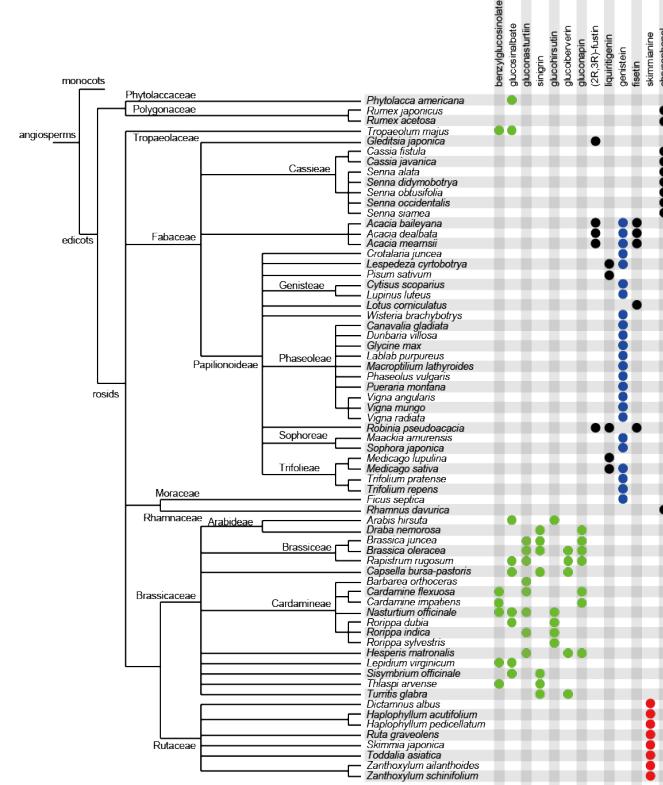


(c) Papilionidae-specific



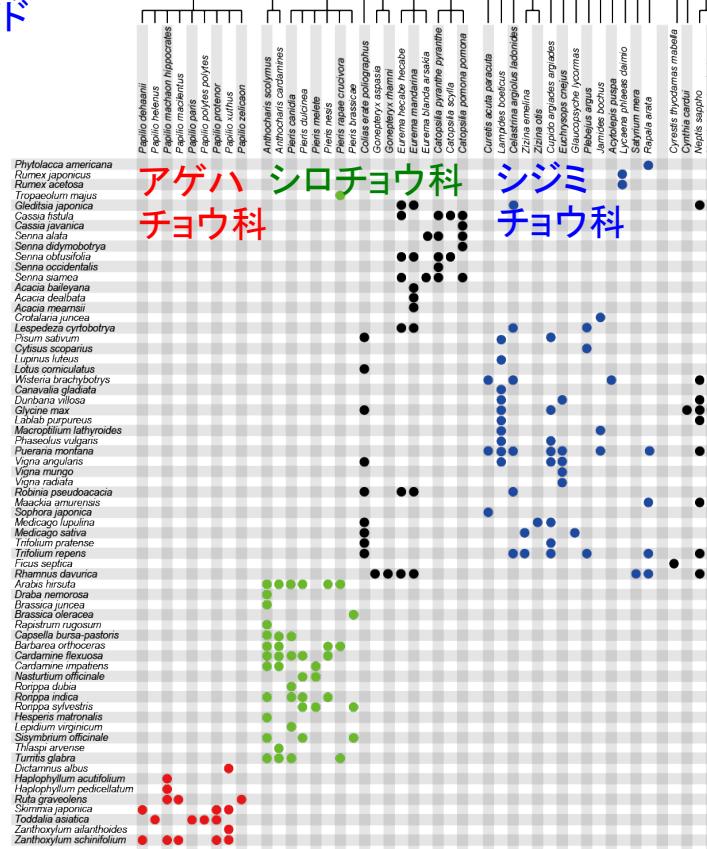
(d) Plant-compound

辛子油配糖体 フラボノイド



(e) Food web

アゲハ シロチョウ科
シジミ チョウ科



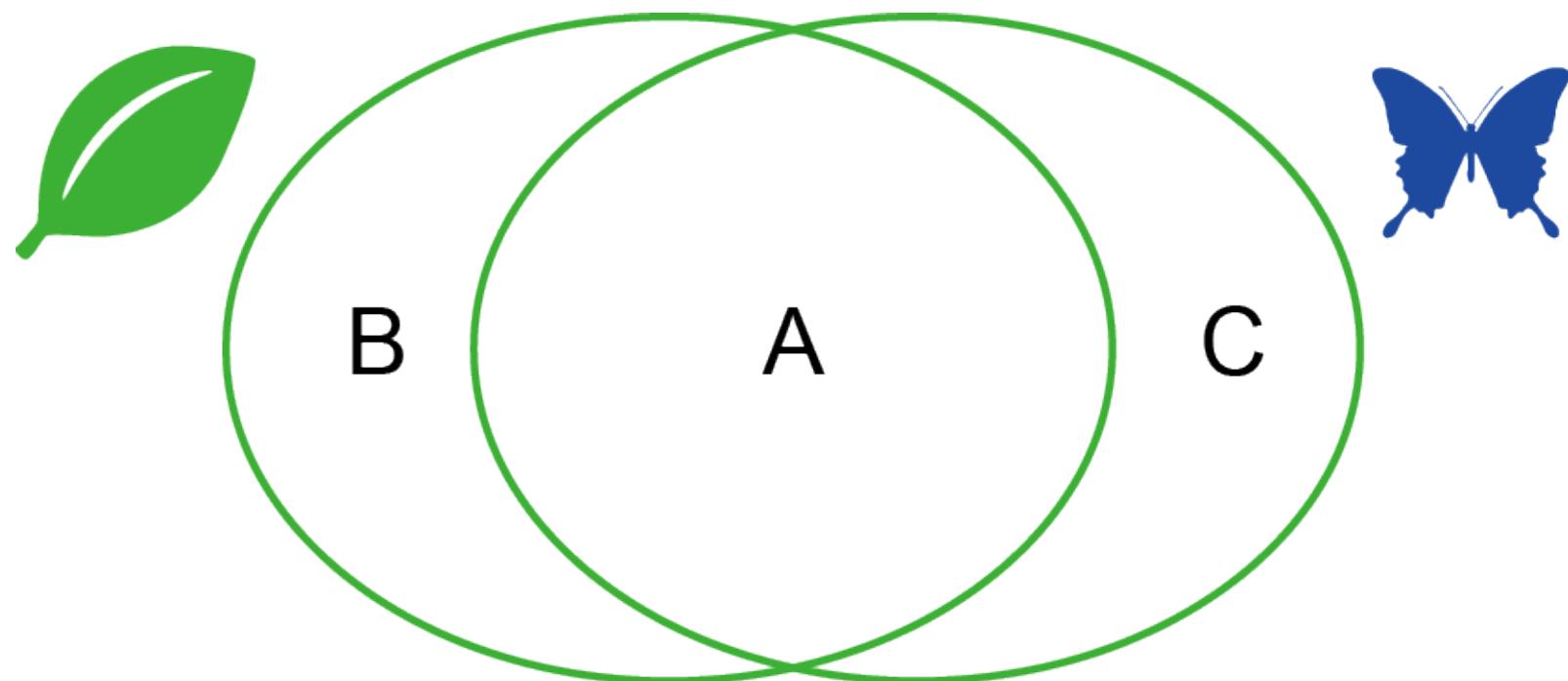
科レベルの捕食・被食関係に特徴的な植物化合物の抽出

マメ科植物

アブラナ科植物

ミカン科植物

草食昆虫と天然物代謝



全ての代謝パスウェイが明らかになったわけではありません。

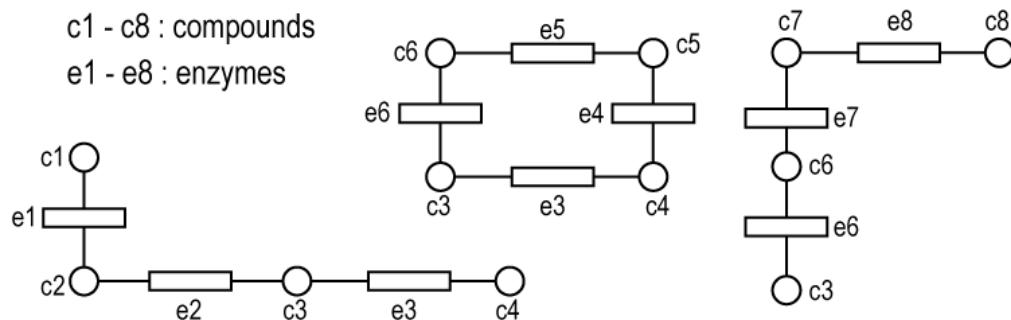
- ・異なる生物種は異なる代謝パスウェイを持ち
- ・代謝パスウェイは条件によって変わったりします。

代謝パスウェイをモデリングするには？

→ 多くの研究で、**定義済みパスウェイ**をテンプレートとして使っています。

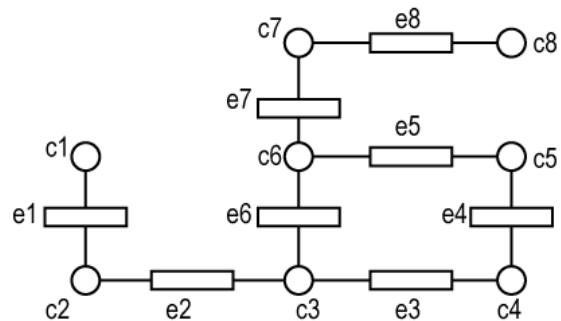
テンプレート依存の代謝パスウェイモデリング

Known pathways in literatures



テンプレート依存の代謝パスウェイモデリング

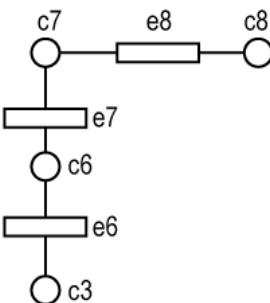
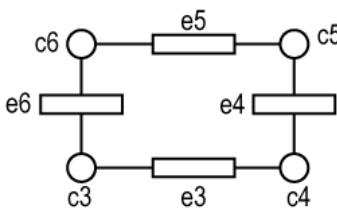
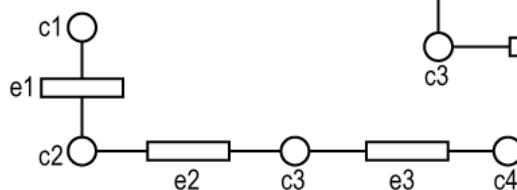
Reference pathway



Known pathways in literatures

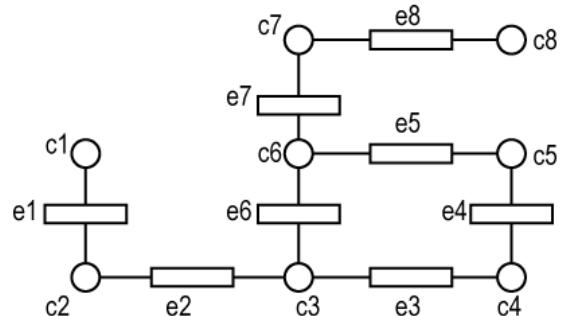
$c_1 - c_8$: compounds

$e_1 - e_8$: enzymes



テンプレート依存の代謝パスウェイモデリング

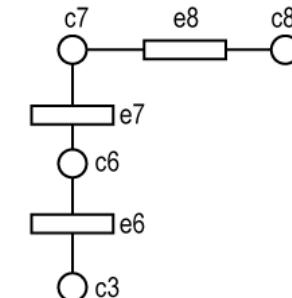
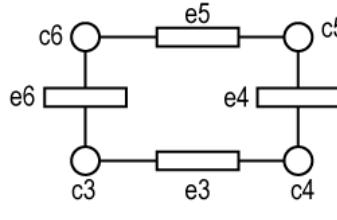
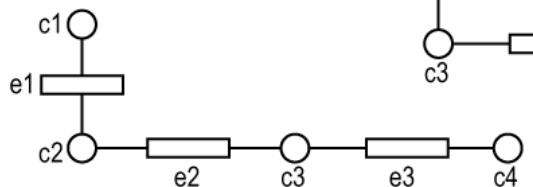
Reference pathway



Known pathways in literatures

c1 - c8 : compounds

e1 - e8 : enzymes



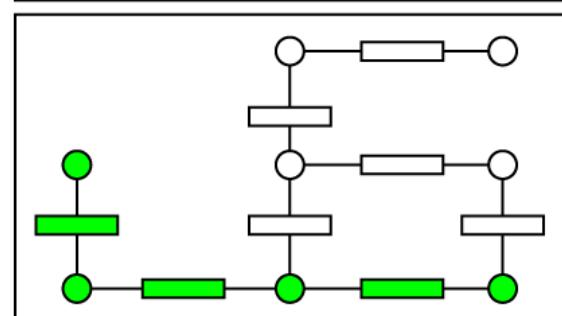
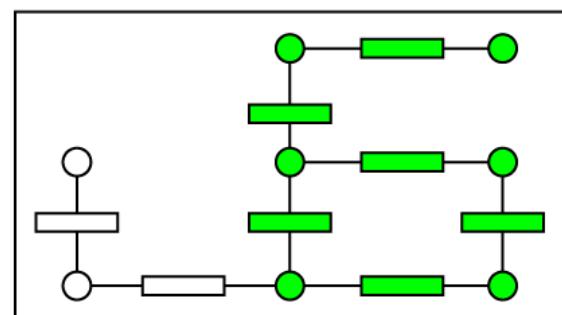
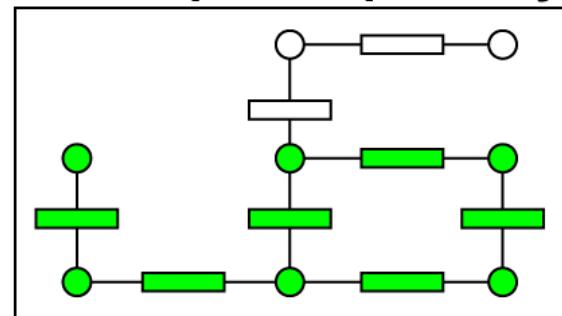
Genome annotation

genome A

genome B

genome C

Species-specific pathways



多様な生物種は多様な天然物を生合成しています。



テンプレートのない
代謝パスウェイだけ
なんだけど、どうすんの？

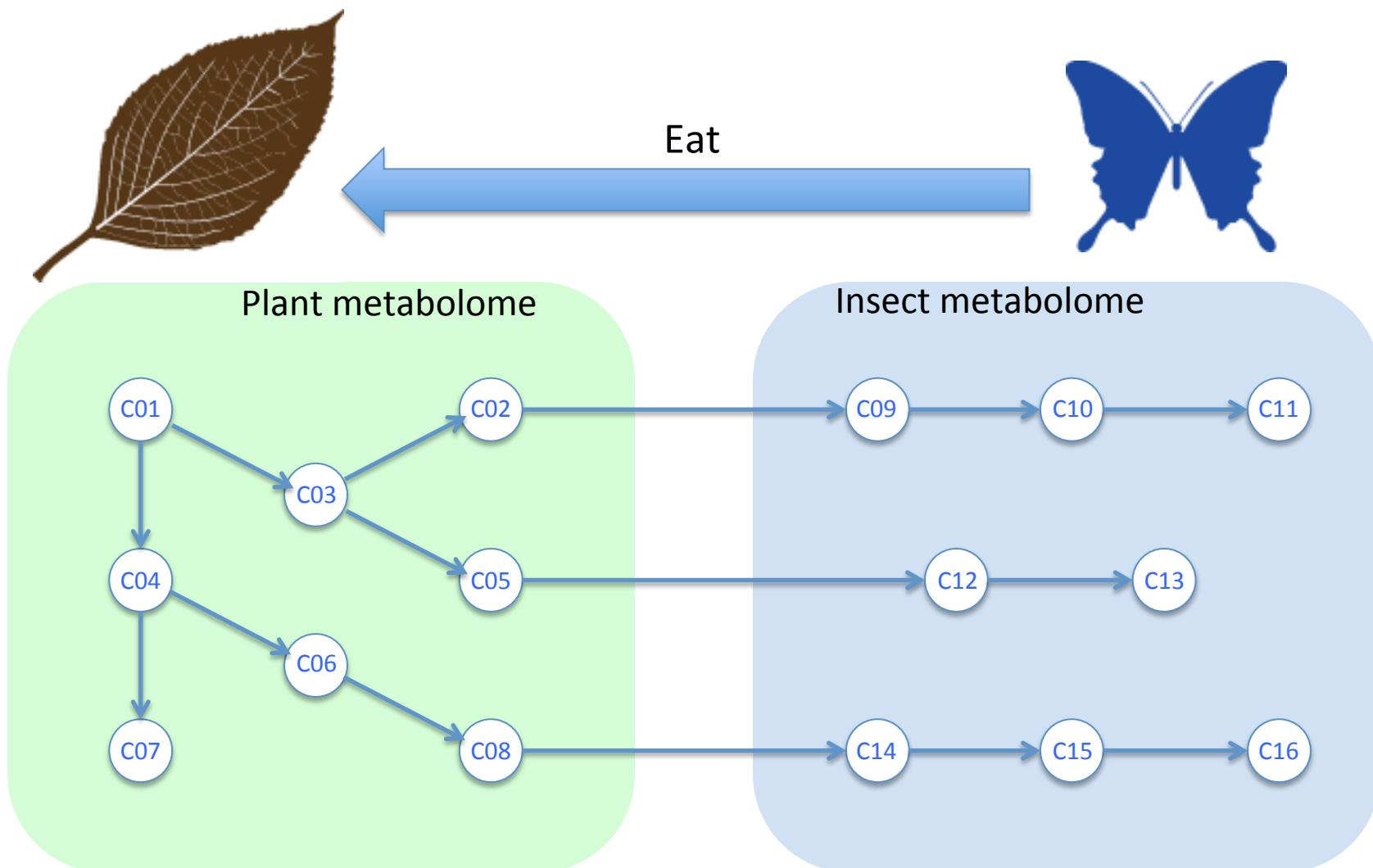


De novo 代謝パスウェイ予測



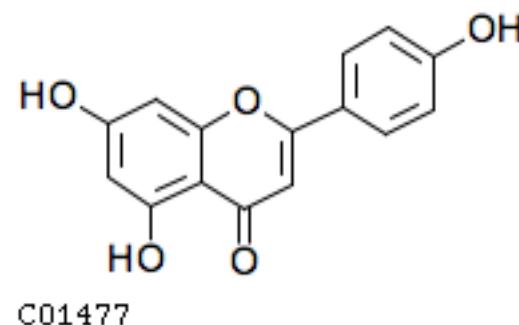
Methods:

天然物代謝予測: Phase 1

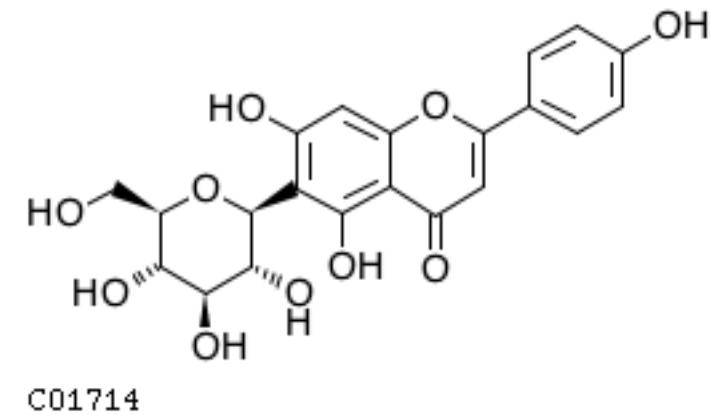


Materials & Methods:

Does this look like an enzymatic reaction?

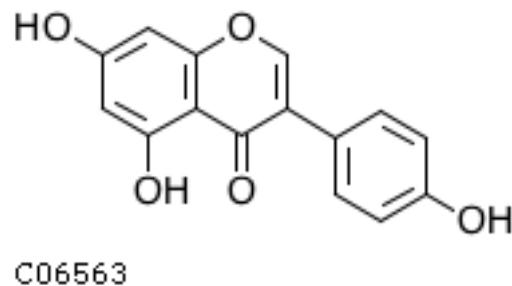


Positive example
(known)

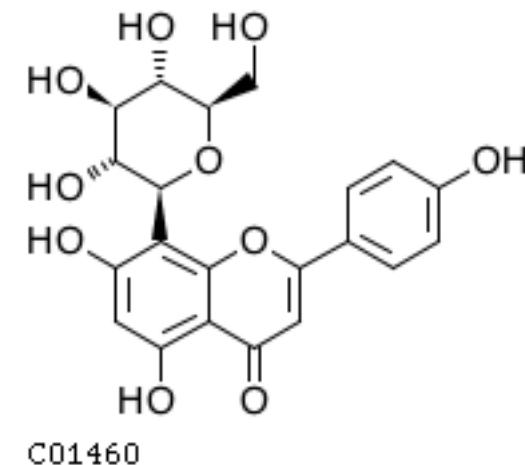


They must be distinguished.

How about this?



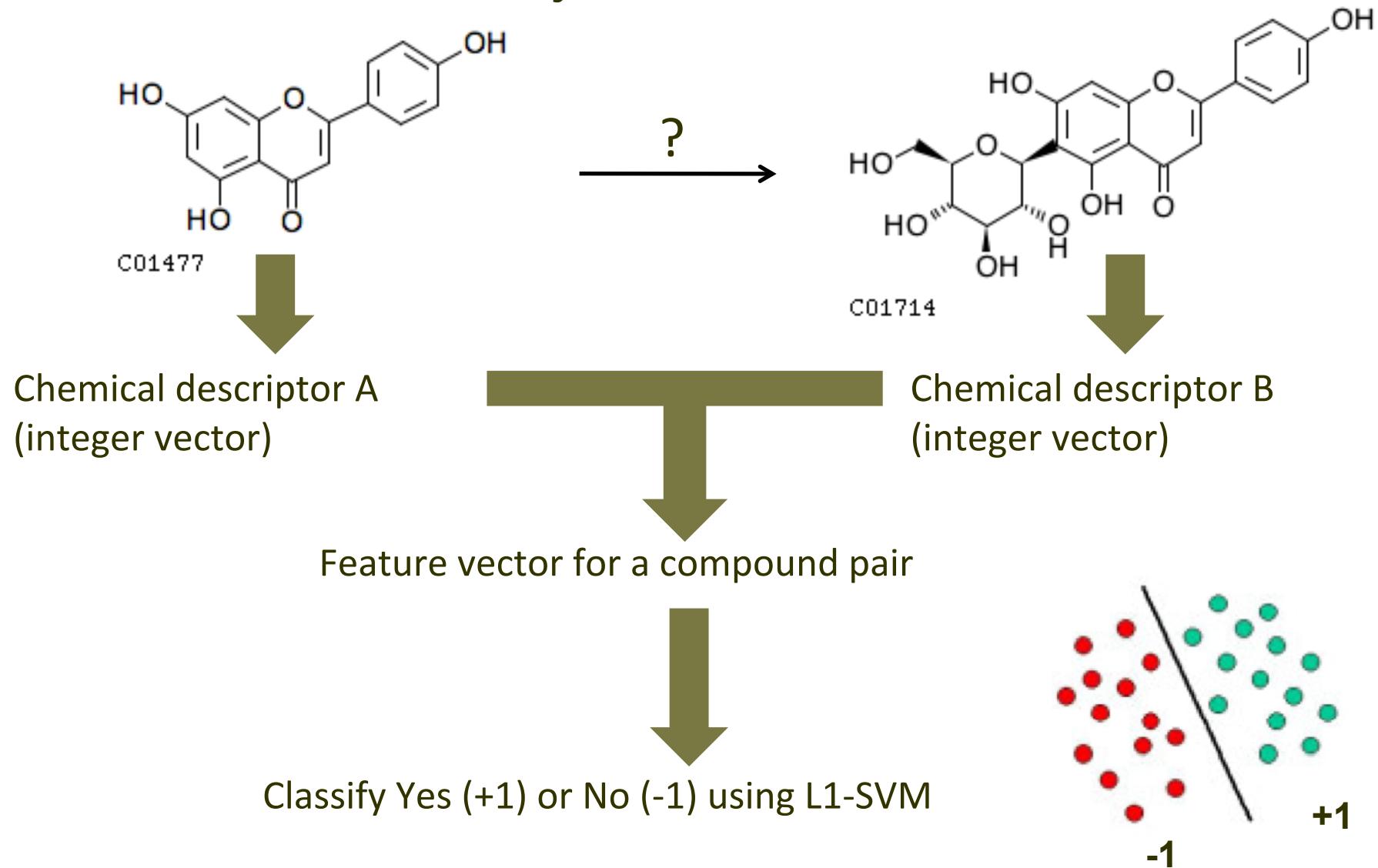
Negative example
(unknown, but no guarantee
whether it is really negative)



Materials & Methods:

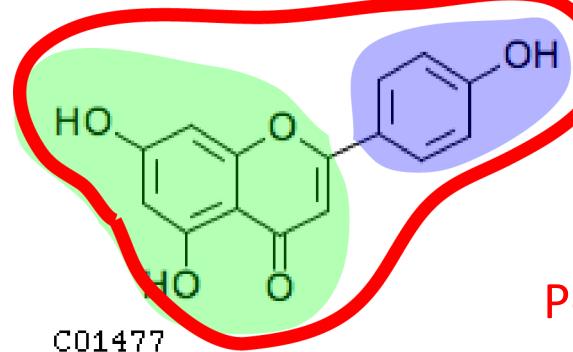
提案手法1 (descriptor-based)

Does this look like an enzymatic reaction?

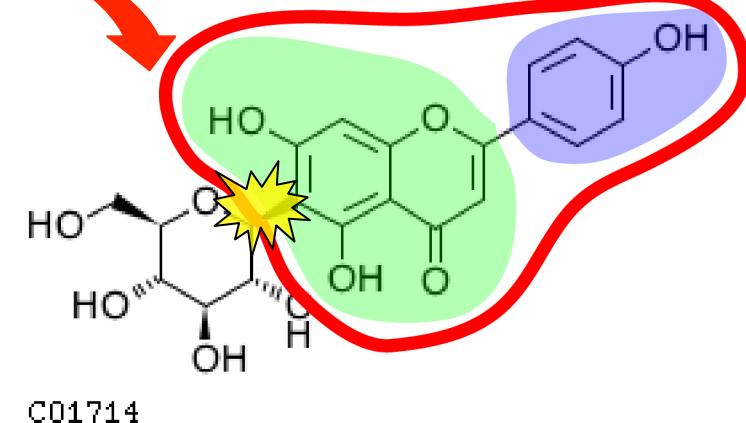


Materials & Methods:

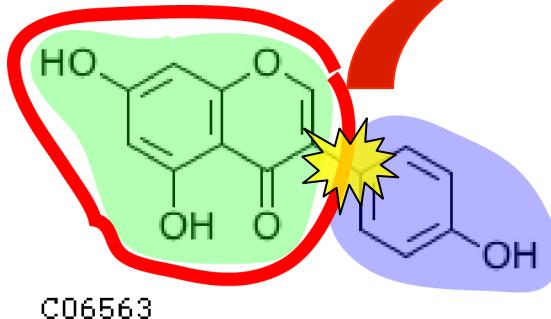
提案手法2 (MCS-based)



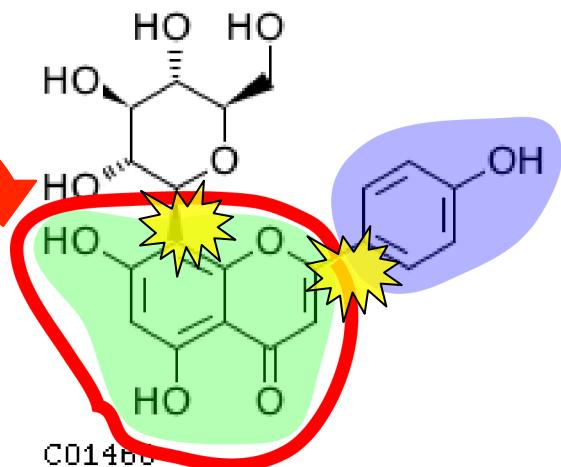
?
Positive example



★ represents the positions of chemical changes.



?
Negative example

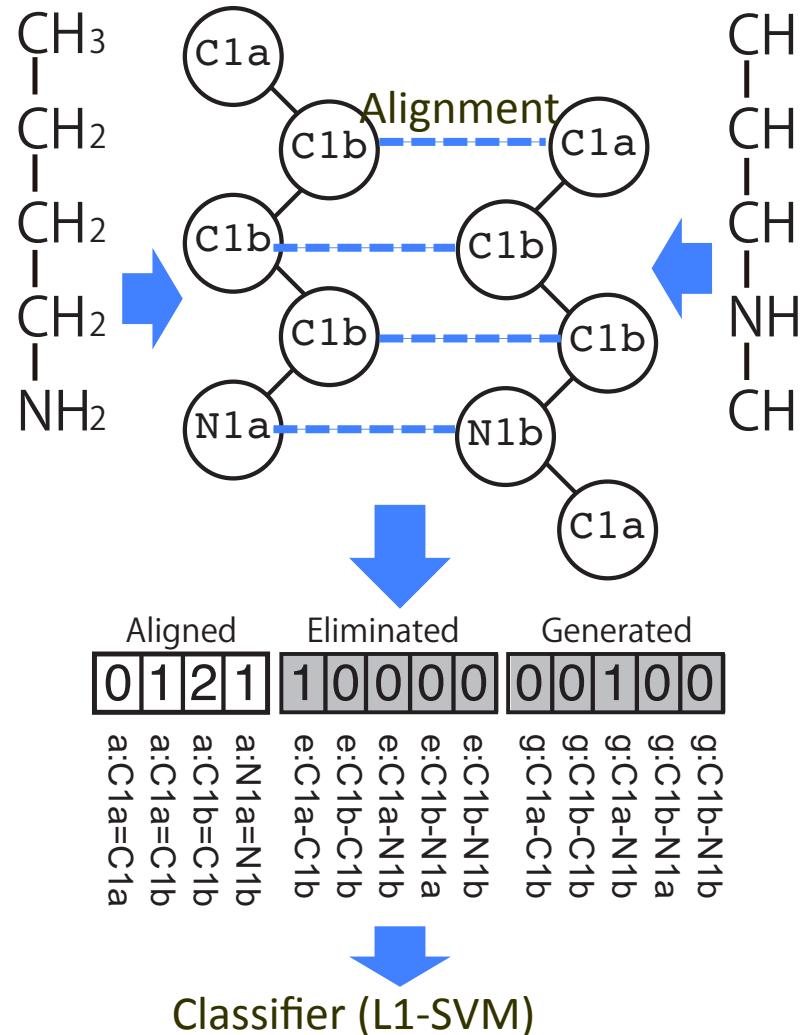


Materials & Methods:

提案手法2 (MCS-based)

Aligned:

- The numbers of atom pairs.
- The pair with the same label represents the conserved substructure.
- The pair with the different labels represents the chemical change.

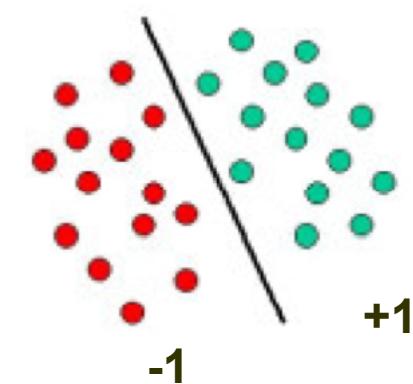


Eliminated:

- The numbers of chemical bonds that were eliminated during the reaction.

Generated:

- The numbers of chemical bonds that were generated during the reaction.



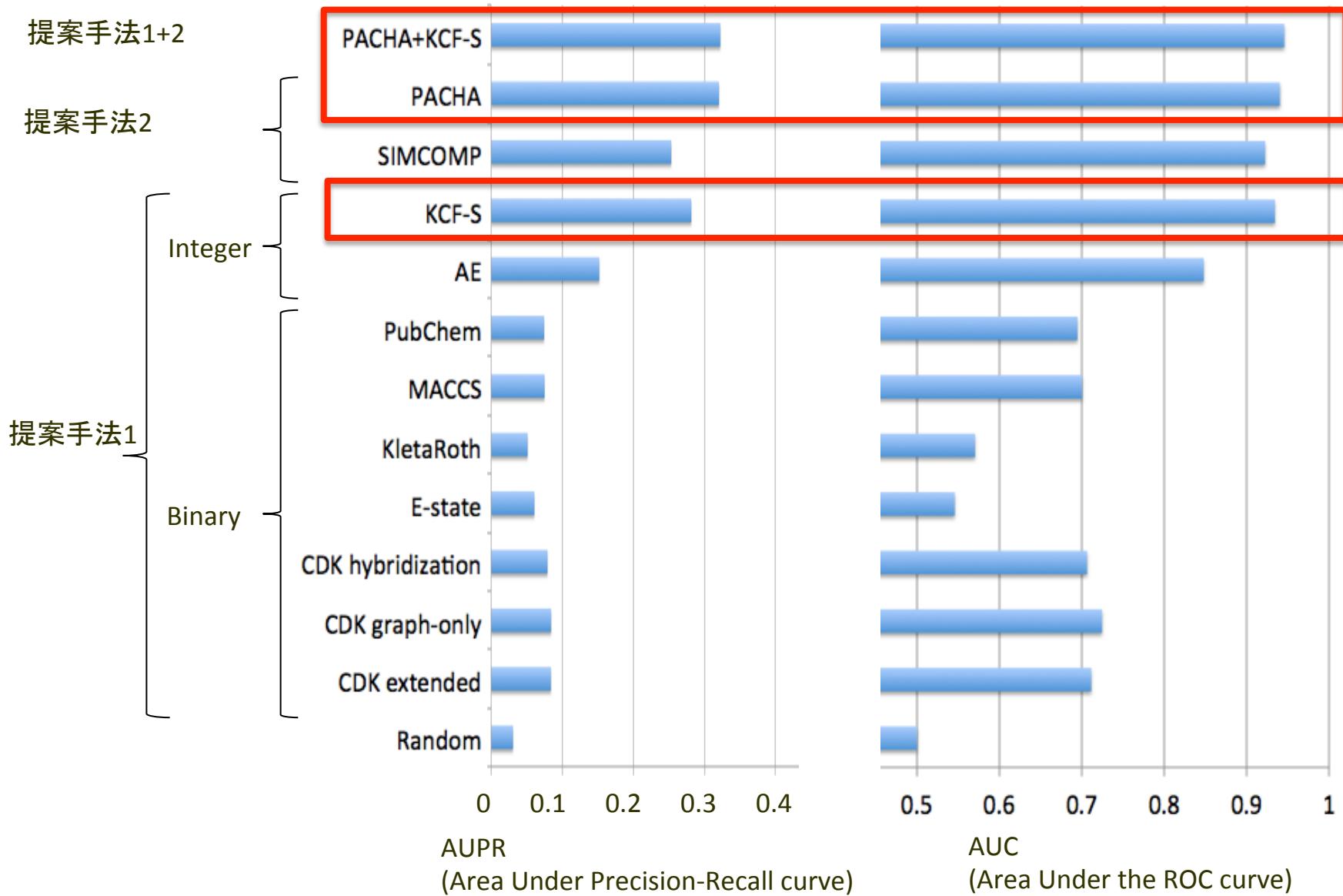
Materials & Methods:

手法比較

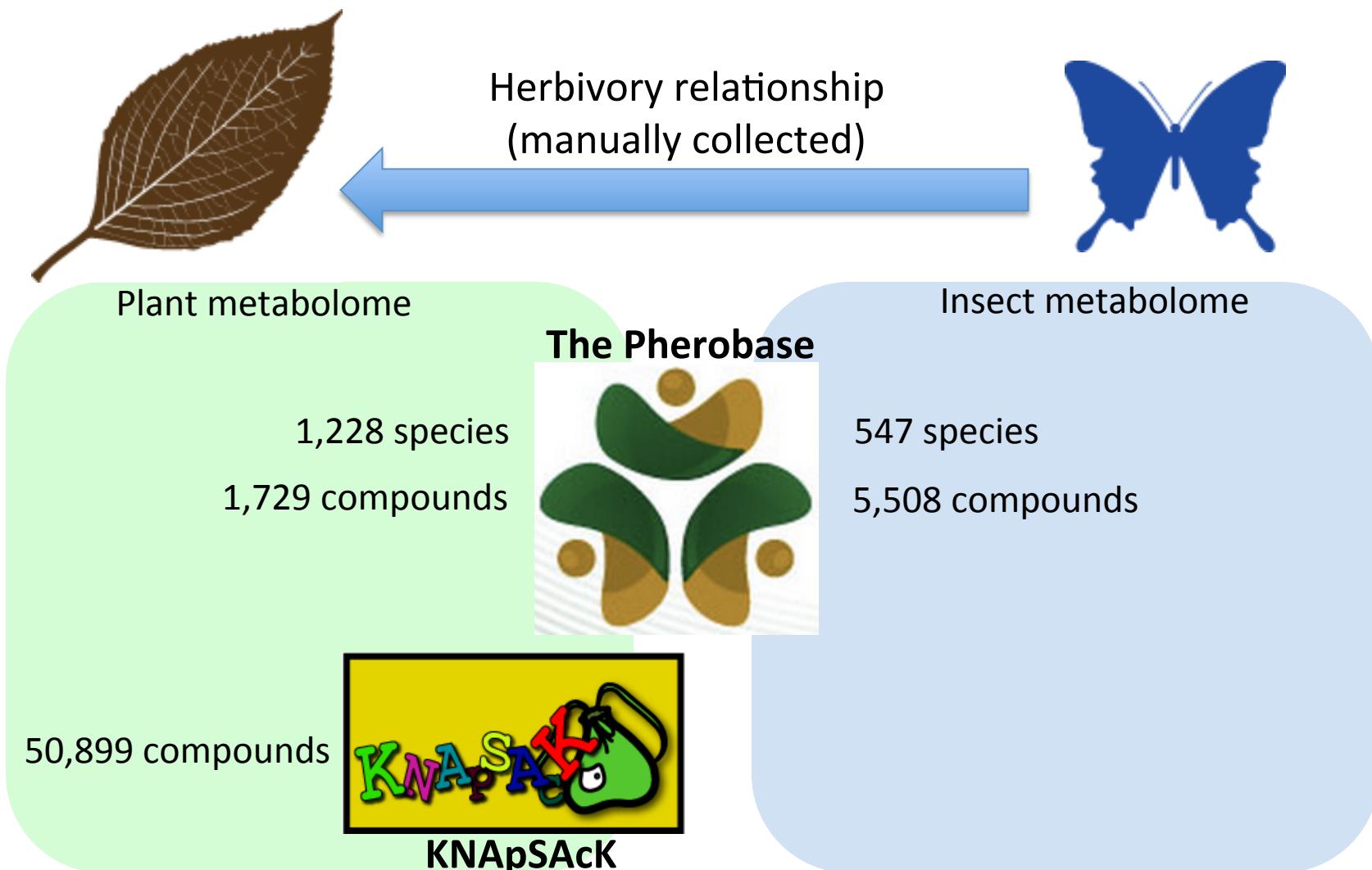
- 提案手法1 (descriptor-based)
 - Binary fingerprints
 - PubChem
 - CDK
 - MACCS
 - KlettaRoth
 - E-state
 - Integer vectors
 - AE
 - Resembles Extended-Connectivity Fingerprints
 - KCF-S (Kotera et al, 2013)
 - Represents the numbers of biochemically defined substructures
- 提案手法2 (MCS-based)
 - Clique-finding algorithm
 - SIMCOMP
 - Priority-search algorithm
 - PACHA (Kotera et al, 2015)
 - Finds a single connected common subgraph between two compounds using atom environment similarity.

Results:

手法比較

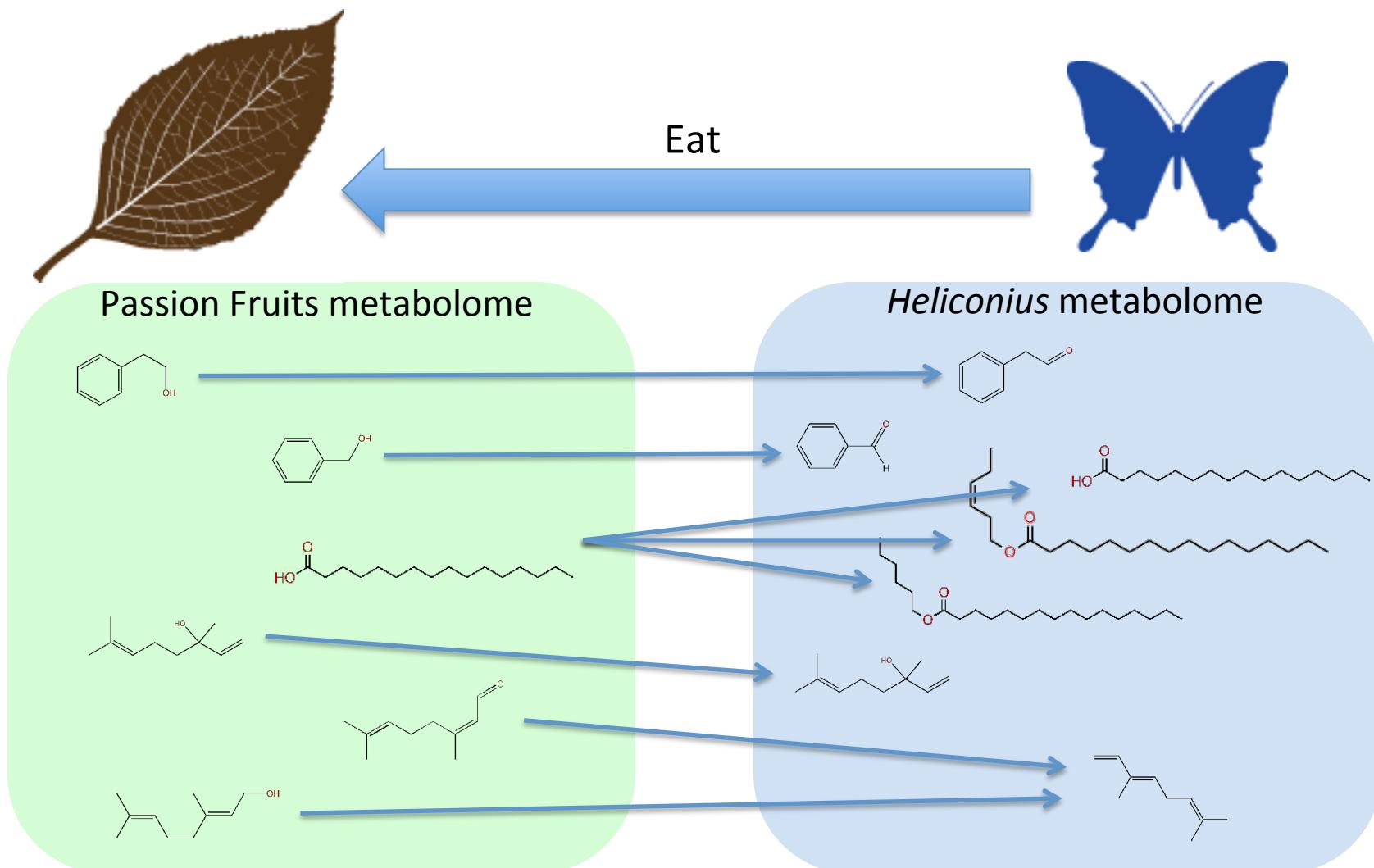


食草ネットワークへの応用



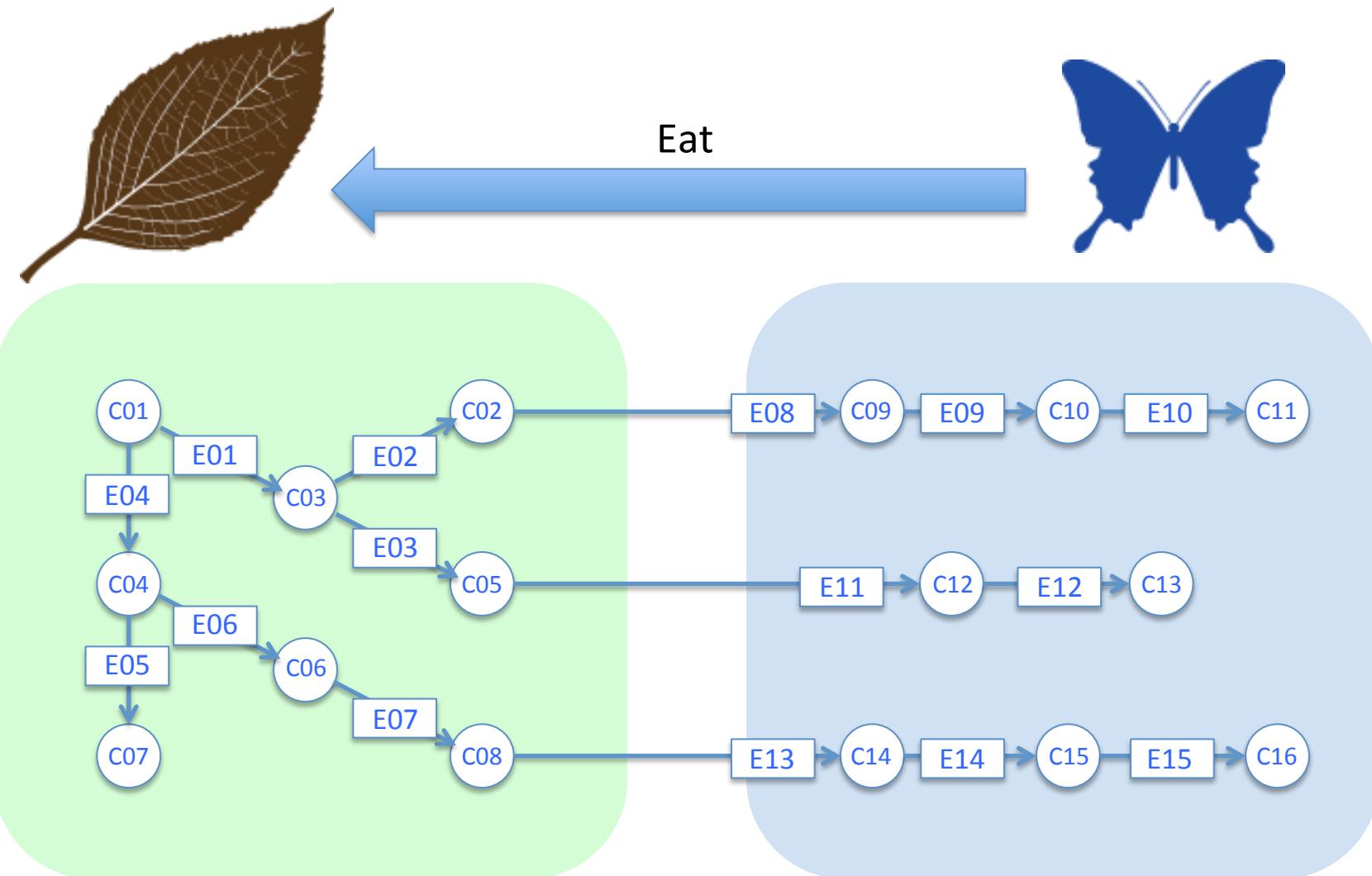
Results:

食草ネットワークへの応用



Methods:

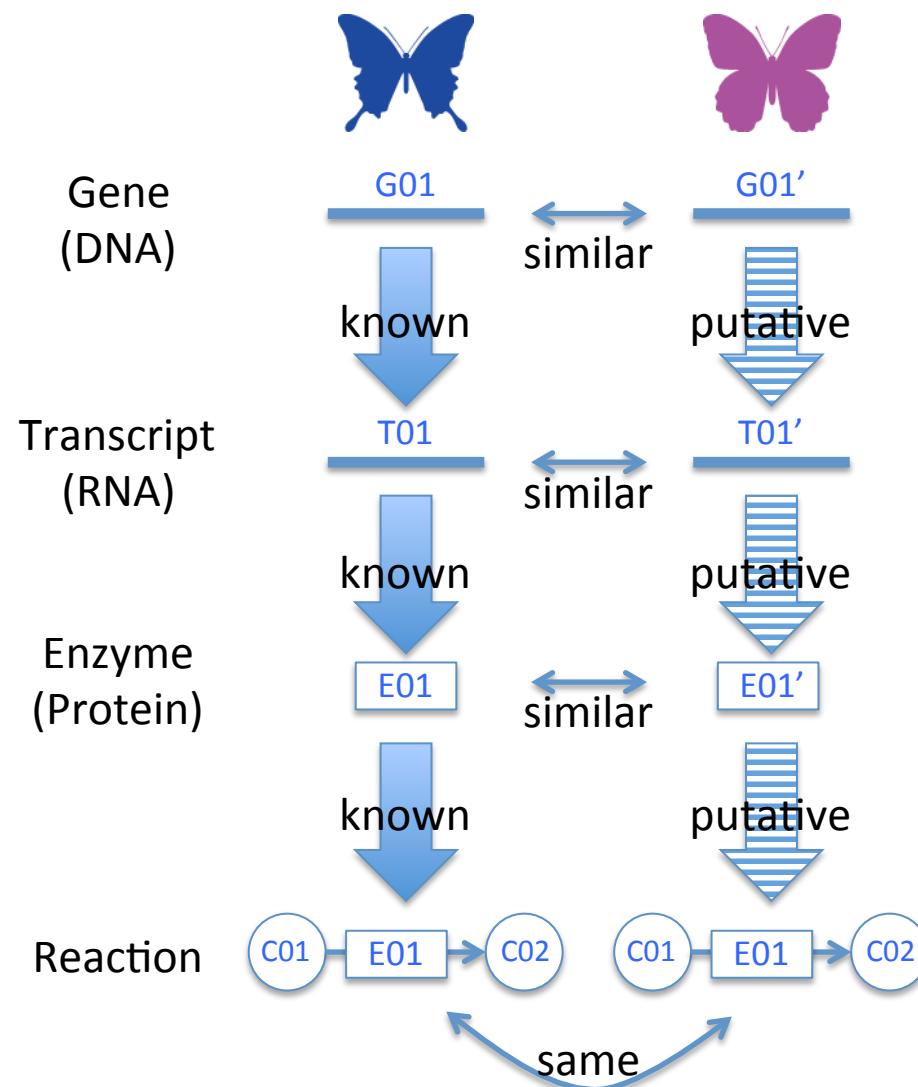
天然物代謝予測: Phase 2



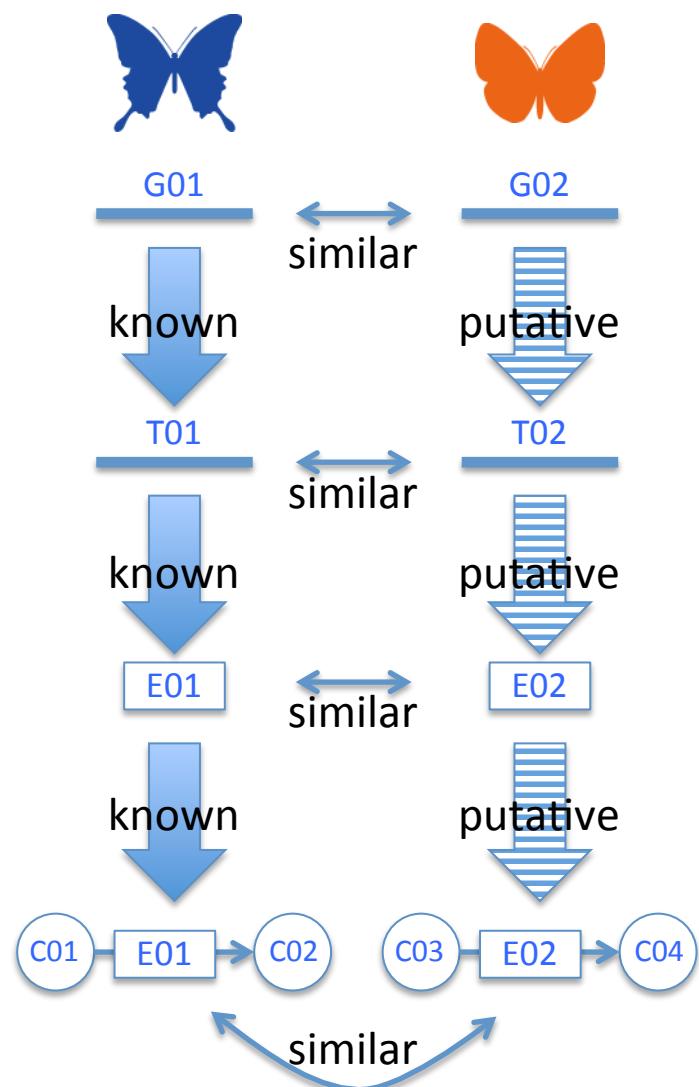
Methods:

酵素反応アノテーション戦略

普通に用いられている戦略
(中心代謝などに有効)



提案する戦略[Kotera et al, 2004]
(天然物代謝予測には、これしかない?)



“似ている”とは何か？

- Enzyme “proteins” are similar
- Enzyme “reactions” are similar

	Sequence	3D
Globally	Full-length	Fold
Locally	Motif	Cavity

	Reaction	Substrates
Globally	?	?
Locally	?	?

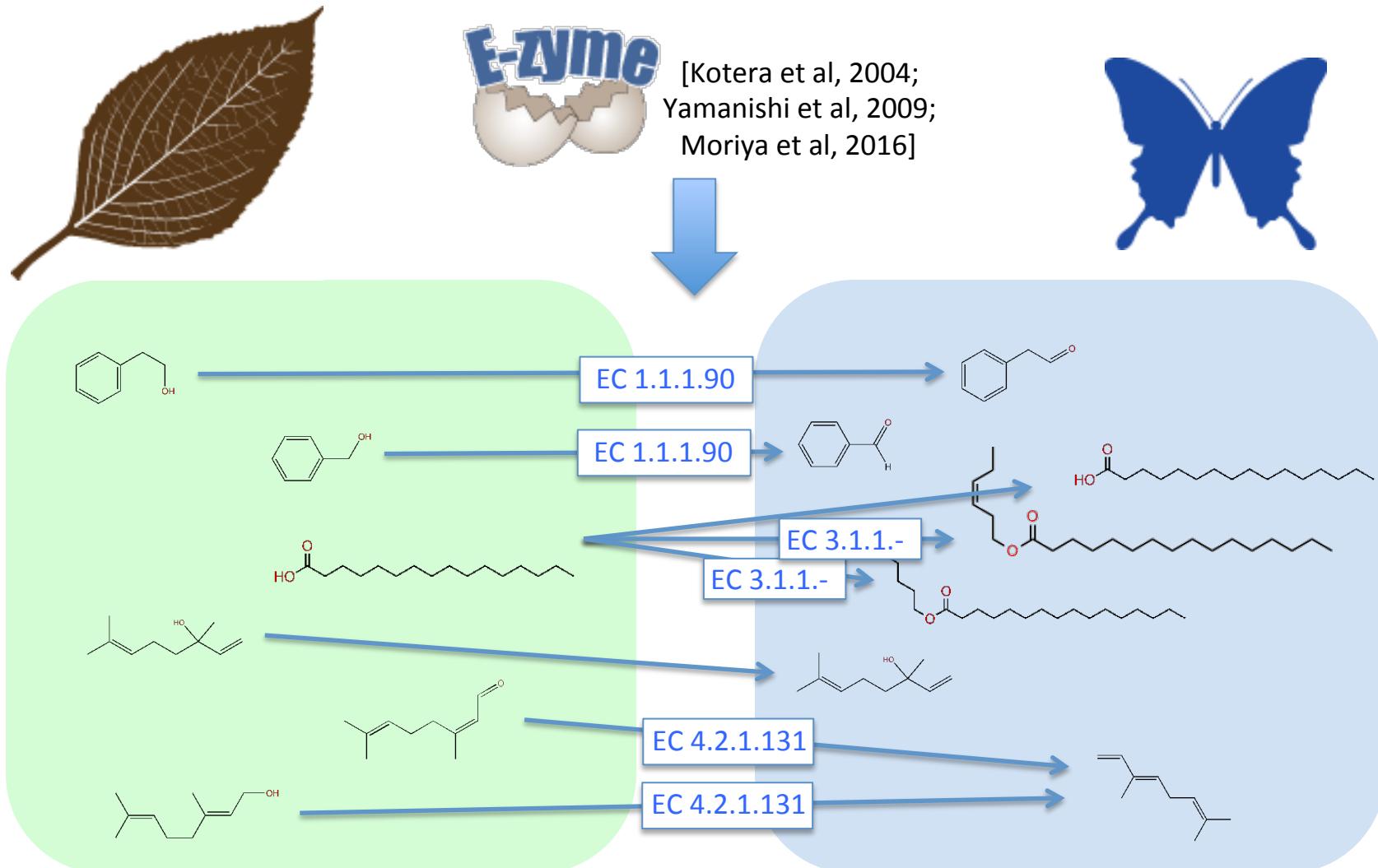
類似性指標とアノテーションの整備が必要 ↑

EC番号とは何か？

- EC番号(酵素番号、Enzyme Commission numbers)は酵素を整理すべく反応形式に従ってECに続く4組の数字で表したもの(1961年～)。
- 分類は階層的でありECの接頭辞にピリオドで区切った続けた4個の番号“EC X.X.X.X”(Xは数字)による表記がなされる。
 - EC 1.X.X.X – オキシドレダクターゼ(酸化還元酵素)、酸化還元反応を触媒
 - EC 2.X.X.X – トランスフェラーゼ(転移酵素)、原子団(官能基など)のある分子から別の分子へ転移
 - EC 3.X.X.X – ヒドロラーゼ(加水分解酵素)、加水分解反応を触媒
 - EC 4.X.X.X – リアーゼ(脱離酵素)、原子団を二重結合あるいは、結合の解離の触媒
 - EC 5.X.X.X – イソメラーゼ(異性化酵素)、分子の異性体を作る
 - 分子内酸化還元反応、分子内転移反応、分子内脱離反応など
 - EC 6.X.X.X – リガーゼ(合成酵素)、ATPの加水分解エネルギーを利用して、2つの分子を結合

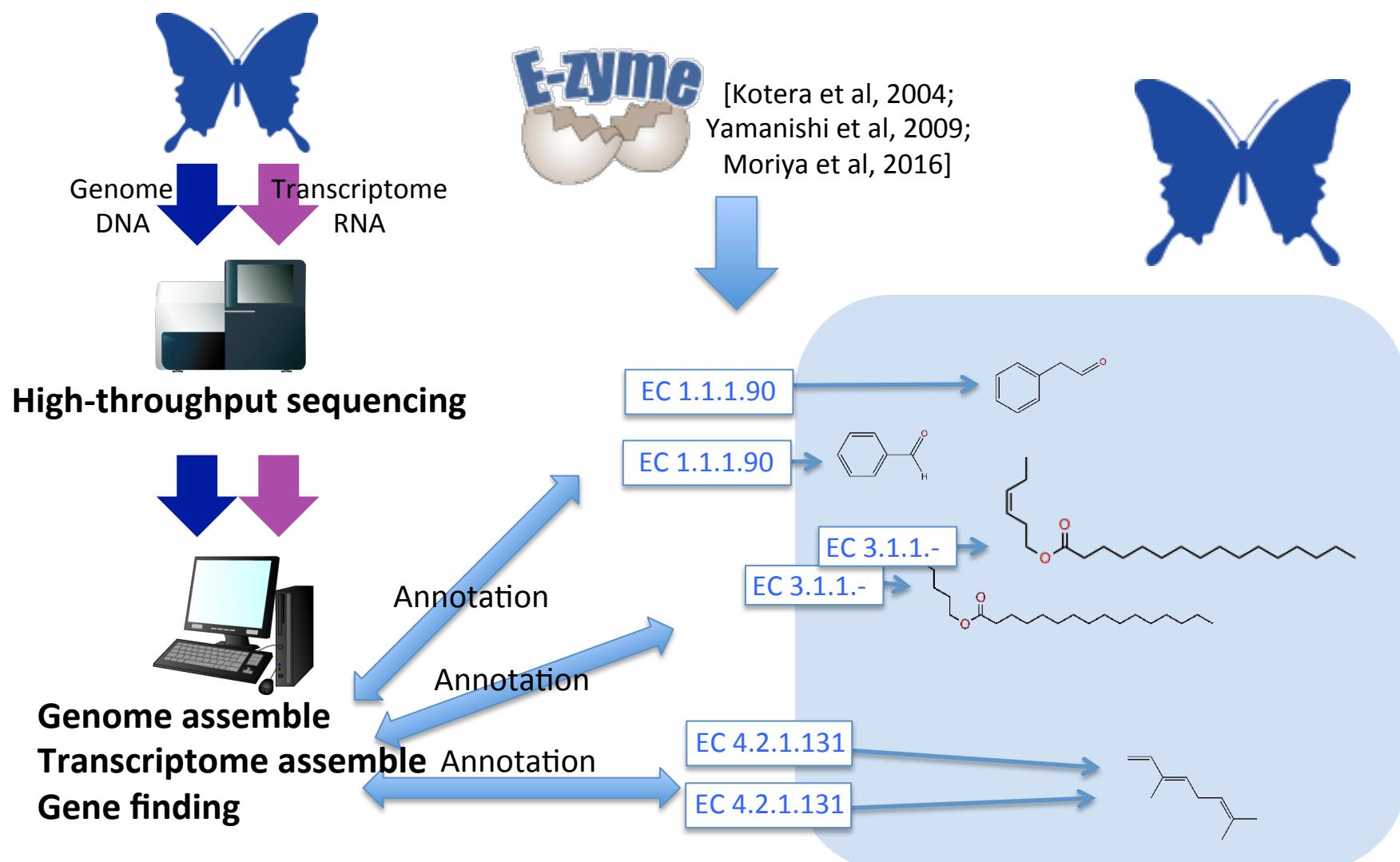
Methods & Results:

最も類似した反応の検索

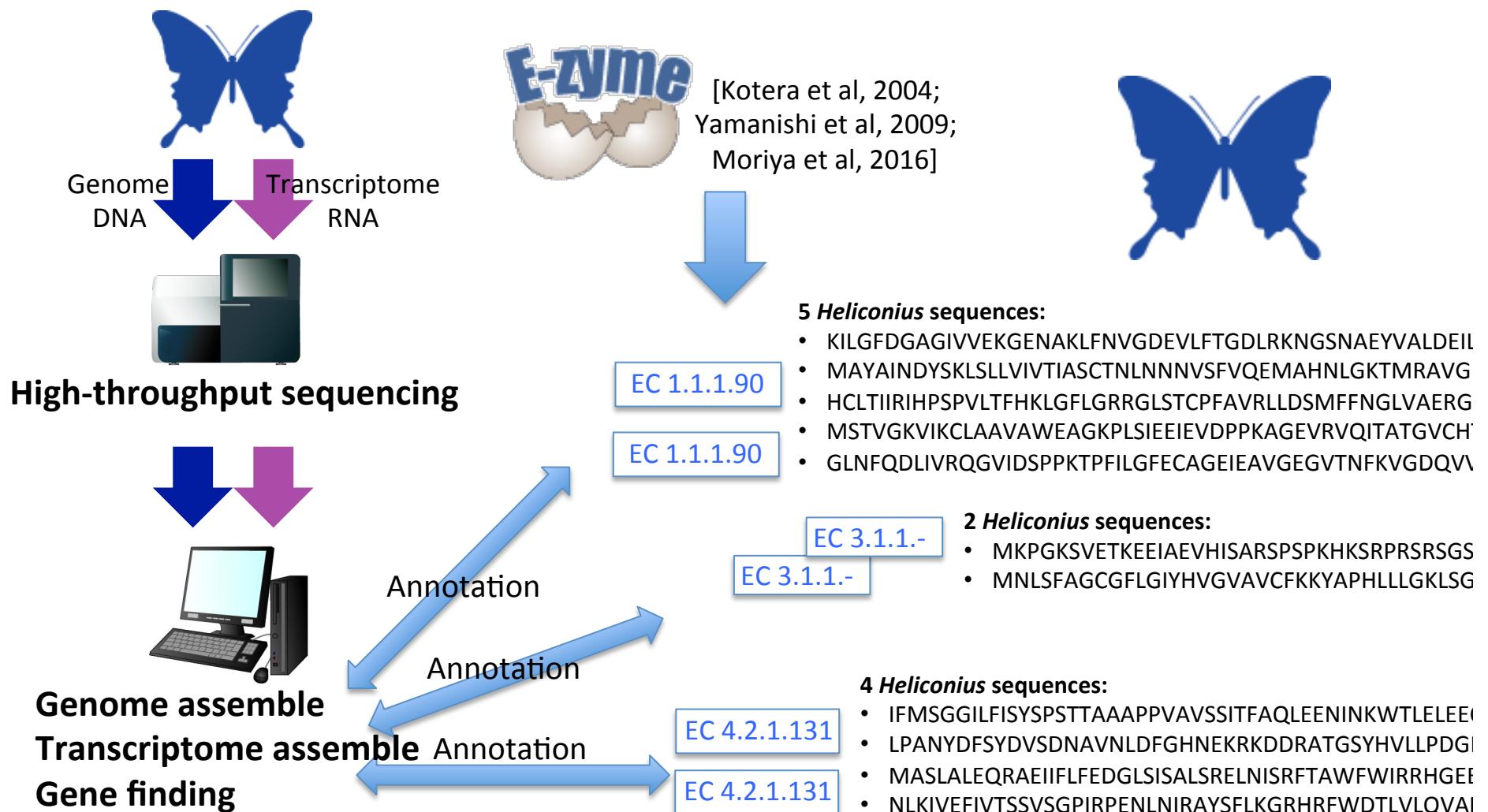


Methods:

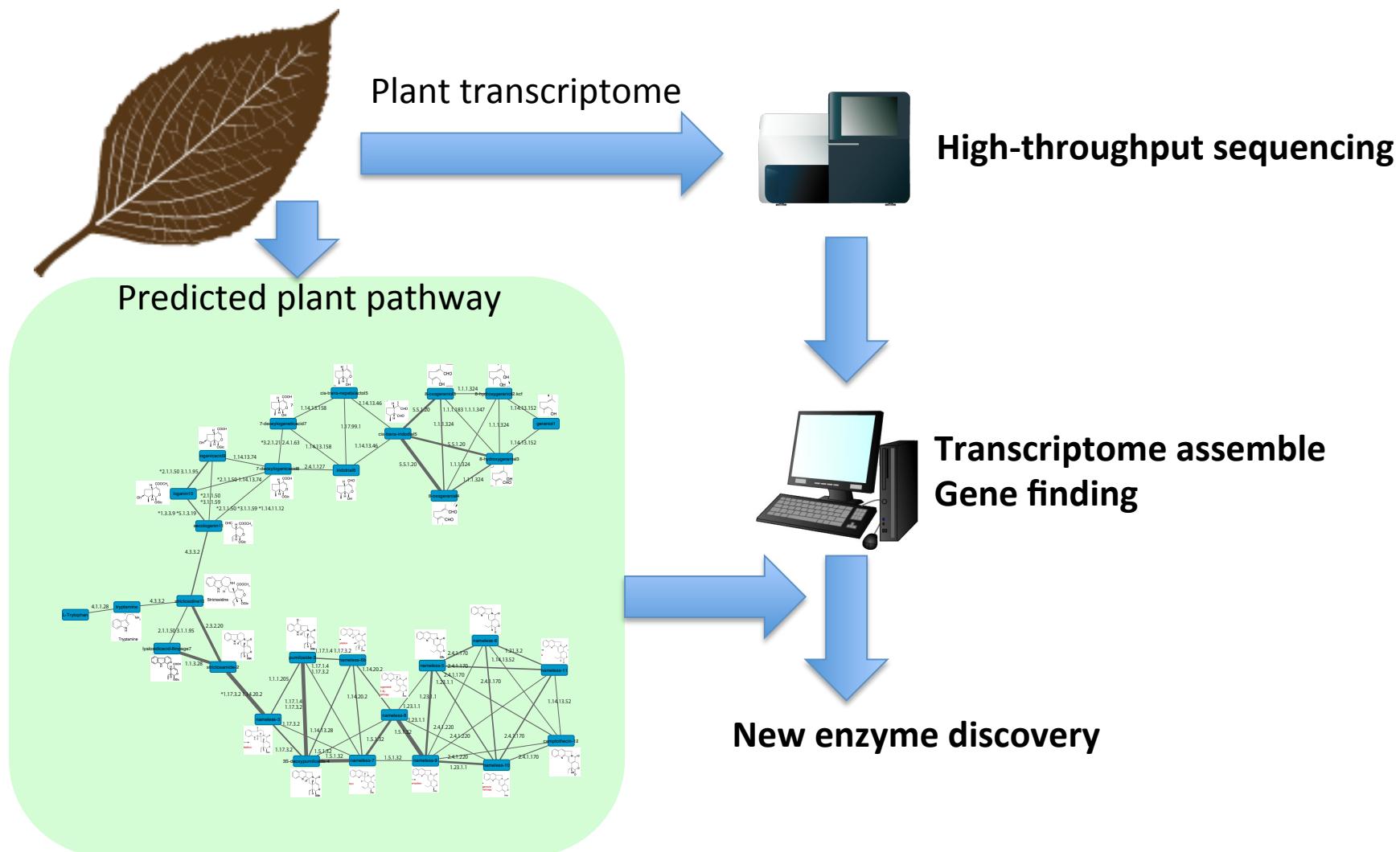
次世代シーケンサーを用いた候補遺伝子予測



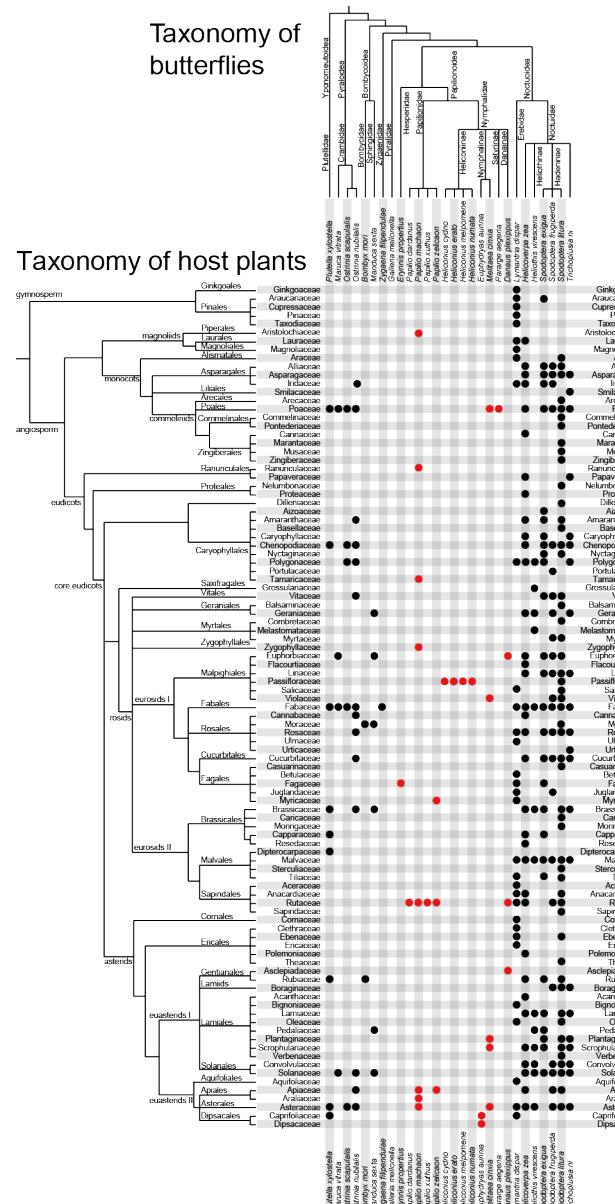
次世代シークエンサーを用いた候補遺伝子予測



次世代シークエンサーを用いた植物化合物代謝予測



Sequence Read Archive (SRA) で GenomeまたはTranscriptomeデータの ある昆虫の食草マトリクス



次世代シークエンサーを活用した
群集生態学へ

- ・昆虫代謝予測
- ・植物代謝予測
- ・昆虫の食性とその分子メカニズム予測
消化酵素・合成酵素など
- ・生物間相互作用とその分子メカニズム予測
味覚受容体・嗅覚受容体など

天然物代謝のコンピュータ予測

L1ゼミ 2015年11月19日(木)15:05～16:35 B222 小寺研究室

来年度配属の学生と取り組みたいテーマ(例) 2～3名募集します。

- ・ 中間代謝物予測法の改良
- ・ 開環・閉環反応予測法の改良
- ・ 連続反応予測法の改良
- ・ 酵素反応アノテーション戦略の改良

緑が丘。コアタイム11時～17時。実験はしません。生き物には触りません。

プログラミングや数学は得意でなくていいです。配属後にゼミを通じて勉強しましょう。

課題

- ・ 次の論文の要点を「落合先生フォーマットで」まとめて下さい(ただし、図は要りません)。
 - Masaaki Kotera, Yasuo Tabei, Yoshihiro Yamanishi, Toshiaki Tokimatsu and Susumu Goto. **Supervised de novo reconstruction of metabolic pathways from metabolome-scale compound sets.** *Bioinformatics* 2013, 29(13), i135-i144.
<http://bioinformatics.oxfordjournals.org/content/29/13/i135.full.pdf>
- ・ 提出方法:学籍番号と氏名を明記の上、maskot@bio.titech.ac.jp にメールで提出。
- ・ 提出期限:11月26日(木)15:00

高速で論文がバリバリ読める「落合先生フォーマット」

<http://lafrenze.hatenablog.com/entry/2015/08/04/120205>

The Visual Microphone: Passive Recovery of Sound from Video

Abel Davis¹ Michael Rubinstein^{2,1} Neal Wadhwa³ Gautham J. Mysore³ Frédéric Durand¹ William T. Freeman¹

High speed camera → Feature extraction → Motion vectors → Sound recovery → Ground truth audio

Figure showing recovered sounds for various speakers and environments:

- Female speaker - (right, sd)
- Male speaker - (mico3, sd)

Figure showing regression coefficients (R^2) for different methods:

- 0.95 (Visual Microphone)
- 0.93 (LSDV)
- 0.92 (LSDV)
- 0.91 (LSDV)
- 0.89 (LSDV)
- 0.88 (LSDV)
- 0.87 (LSDV)
- 0.86 (LSDV)
- 0.85 (LSDV)
- 0.84 (LSDV)
- 0.83 (LSDV)
- 0.82 (LSDV)
- 0.81 (LSDV)
- 0.80 (LSDV)
- 0.79 (LSDV)
- 0.78 (LSDV)
- 0.77 (LSDV)
- 0.76 (LSDV)
- 0.75 (LSDV)
- 0.74 (LSDV)
- 0.73 (LSDV)
- 0.72 (LSDV)
- 0.71 (LSDV)
- 0.70 (LSDV)
- 0.69 (LSDV)
- 0.68 (LSDV)
- 0.67 (LSDV)
- 0.66 (LSDV)
- 0.65 (LSDV)
- 0.64 (LSDV)
- 0.63 (LSDV)
- 0.62 (LSDV)
- 0.61 (LSDV)
- 0.60 (LSDV)
- 0.59 (LSDV)
- 0.58 (LSDV)
- 0.57 (LSDV)
- 0.56 (LSDV)
- 0.55 (LSDV)
- 0.54 (LSDV)
- 0.53 (LSDV)
- 0.52 (LSDV)
- 0.51 (LSDV)
- 0.50 (LSDV)
- 0.49 (LSDV)
- 0.48 (LSDV)
- 0.47 (LSDV)
- 0.46 (LSDV)
- 0.45 (LSDV)
- 0.44 (LSDV)
- 0.43 (LSDV)
- 0.42 (LSDV)
- 0.41 (LSDV)
- 0.40 (LSDV)
- 0.39 (LSDV)
- 0.38 (LSDV)
- 0.37 (LSDV)
- 0.36 (LSDV)
- 0.35 (LSDV)
- 0.34 (LSDV)
- 0.33 (LSDV)
- 0.32 (LSDV)
- 0.31 (LSDV)
- 0.30 (LSDV)
- 0.29 (LSDV)
- 0.28 (LSDV)
- 0.27 (LSDV)
- 0.26 (LSDV)
- 0.25 (LSDV)
- 0.24 (LSDV)
- 0.23 (LSDV)
- 0.22 (LSDV)
- 0.21 (LSDV)
- 0.20 (LSDV)
- 0.19 (LSDV)
- 0.18 (LSDV)
- 0.17 (LSDV)
- 0.16 (LSDV)
- 0.15 (LSDV)
- 0.14 (LSDV)
- 0.13 (LSDV)
- 0.12 (LSDV)
- 0.11 (LSDV)
- 0.10 (LSDV)
- 0.09 (LSDV)
- 0.08 (LSDV)
- 0.07 (LSDV)
- 0.06 (LSDV)
- 0.05 (LSDV)
- 0.04 (LSDV)
- 0.03 (LSDV)
- 0.02 (LSDV)
- 0.01 (LSDV)
- 0.00 (LSDV)

Sequence	Method	SNR	LLR Mean	Intelligibility
Female speaker - (right, sd)	Vid	24.3	1.67	0.73
	LSDV	26.8	1.82	0.74
Female speaker - (right, sd)	Vid	28.7	1.53	0.63
	LSDV	26.5	1.82	0.70
Female speaker - (right, sd)	Vid	29.4	1.54	0.76
	LSDV	24.4	1.83	0.73
Male speaker - (mico3, sd)	Vid	23.2	1.59	0.67
	LSDV	23.8	1.96	0.68
Male speaker - (mico3, sd)	Vid	23.3	1.68	0.71
	LSDV	28.2	1.74	0.76
Male speaker - (mico3, sd)	Vid	25.5	1.81	0.72
	LSDV	26.0	1.88	0.74

どんなもの？

高速カメラの映像からその場所でかかっていた音を復元する。一眼レフを使った例も検証した。

先行研究と比べてどこがすごい？

レーザー手法（レーザードッパー）は80年代からあったが、本手法ではレーザーを当てなくてもハイスピード動画から音を復元できる。

技術や手法のキモはどこ？

動画から微細な変化を検出する手法
[Wu et al 2012]や他を音声に応用

どうやって有効だと検証した？

レーザードッパーと比べても、有用なデータが出た。統計的誤差を比較した。
スピーチを復元してみて周波数分布を比べてみた。
音響解析してみて振動範囲を調べた。

議論はある？

軽くて硬いものは精度よく復元できる。
光を当てないでよい分レーザーより便利かもしれない。

次に読むべき論文は？

Wu, et al 2012, かな

200711510 落合陽一 #1 (仏コース)