

植物の効能メカニズムを解明するためのデータベース PCIDB の開発

development of PCIDB: PhytoChemical Interactions DB to elucidate

therapeutic effect of plants on human

京都大学化学研究所 バイオインフォマティクスセンター 化学生命科学 西村陽介

背景と目的

植物は生薬として解毒、鎮痛、病気の治癒といった様々な目的のために利用されているだけでなく、多様な二次代謝産物群を持ち、創薬資源として重要な価値を持っている。処方せん医薬品の 25%は植物代謝産物に由来すると言われており、特に抗癌剤の多くが植物等によって産生される天然物またはその誘導体である。

近年の大規模解析技術の発達に伴って植物メタボロームの情報や、植物代謝産物のヒトに対する生理活性についてのアッセイ情報が蓄積されており、生薬や二次代謝産物の効能メカニズムの解明基盤として徐々に整備されつつある。しかし現状では化合物のデータベース間の対応などがなされておらず、これらの情報を繋ぎ合わせて解釈することが容易ではない。そこで我々はこれらの情報を統合して活用し、植物の効能メカニズムを解明するためのデータベースである PCIDB を開発した。

検討内容・結果

PCIDB の開発にあたり、以下の項目について検討した(図1)。

1. KNApSAcK 化合物と、ChEMBL, CTD のヒト相互作用タンパク質データのリンク作成

約 4 万の植物代謝産物とその由来となる生物種が収載されている KNApSAcK データベースの化合物情報を取得し、タンパク質と化合物の相互作用に関するアッセイ情報が収載されている ChEMBL、環境物質と遺伝子や疾患との関連性を文献から抽出した情報が収載されている CTD との化合物リンクを作成した。ChEMBL に関しては化合物の記述フォーマットである Standard InChI を利用してリンクを作成した結果、4,712 化合物についてヒト相互作用タンパク質の情報を得た。CTD に関しては CAS 登録番号を利用し、1,123 化合物について関連するヒトタンパク質や疾患情報のリンクを作成した。

2. ヒト疾患に関する情報の関連付け

上記の相互作用するヒトタンパク質との関連が認められる疾患についての情報を OMIM や KEGG DISEASE から取得して、上記データと関連付けた。また、CTD の化合物と疾患の関連性の情報を利用して、各化合物と関連のある疾患情報についても取得可能とした。

3. KNApSAcK 代謝産物の(a)構造類似性に基づく分類、(b)母核分類を利用した分類の活用

(a) 部分構造ディスクリプタ KCF-S を利用して KNApSAcK 化合物のクラスタリングを行い(KCF-S cluster)、8,955 個の KCF-S cluster に分類した。各クラスタのヒトに対する機能情報として、相互作用タンパク質情報や効能について検索可能となった。

(b) 上記 KCF-S cluster を、植物代謝産物の母核分類である、KEGG BRITE の Phytochemical Compounds の階層分類を利用して、さらに大きなクラスタにまとめた。結果として 84 個の化合物クラスタ(phytochemical cluster)を得て、これらについて相互作用タンパク質情報や効能についての情報が関連付けられた。

PCIDB: to elucidate (1) & (2)

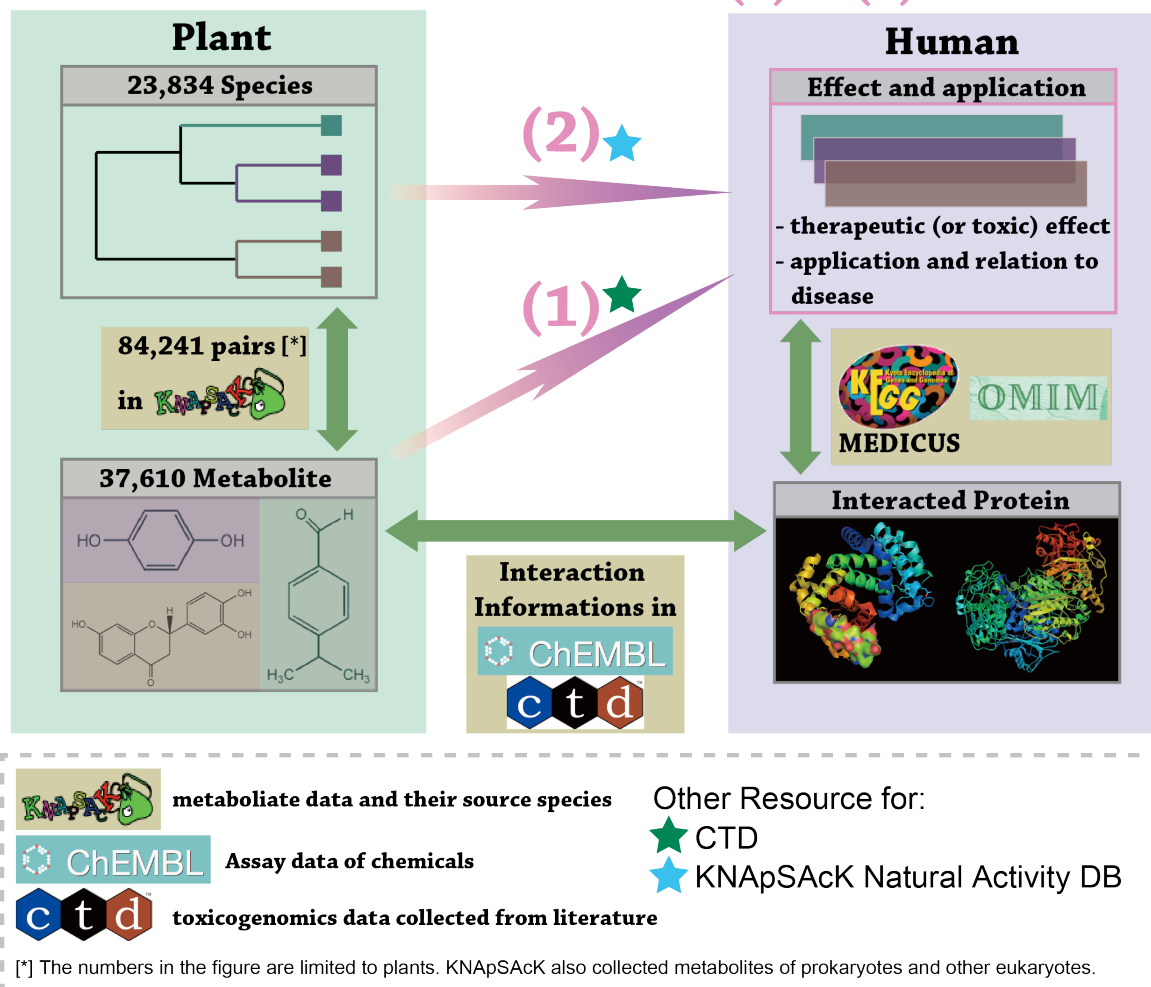


図1: PCIDB の概念図

まとめ・今後の課題

KNApSackの植物代謝産物情報と、ChEMBLやCTDのヒトタンパク質に対するアッセイ・文献情報を関連付けて、植物化合物クラスタが持つヒトに対する効能[図1-(1)]や、植物が持つヒトに対する効能[図1-(2)]に関する情報を取得できるデータベース PCIDB を開発した。これにより植物やその化合物群に関して、ヒトに対する効能についての分子メカニズムを参照することが可能となった。

PCIDB は <http://www.genome.jp/db/pcidb> にて公開されている。

今後の課題として、(1)ChEMBLのアッセイについての種類別のフィルタリングや閾値を設定して、より高い活性を示す相互作用のみを示すことができるようにすること、(2)効能や、関係のある疾患についてのグループ分けを行うことにより、各植物や化合物クラスタが持つ効能についての情報をわかりやすく提示することについて取り組む予定である。