

ポリケチド合成酵素ファミリーの包括的解析

Comprehensive analysis of polyketide synthase families

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 清水 祐吾

背景と目的

ポリケチドはバクテリアや菌類、植物等によって産生される天然化合物(二次代謝物質)であり、多様な化学構造及び生物学的機能を持つだけでなく、抗生物質、抗癌剤、抗寄生虫剤、免疫抑制剤、コレステロール低下剤等の薬理活性を持つ臨床的に重要な薬物が多数含まれていることから非常に注目されている化合物群である。ポリケチドは生体内においてはポリケチド合成酵素(PKS)によって触媒されることで合成されている。その過程は脂肪酸合成酵素(FAS)による脂肪酸生合成経路と類似しており、マロニル CoA やアセチル CoA を始め様々なアシル CoA をスターター基質として用いマロニル CoA やその誘導体を伸長鎖基質として繰り返し縮合、修飾を行うというものである。縮合の回数、各伸長鎖における修飾の種類と有無、そして他の関連酵素による反応後修飾によってポリケチドの広範な多様性が生み出されている。縮合反応や修飾反応に対応する多くの配列ドメインが知られており、1 回の C-C 伸長ステップに用いられる一連のドメインはモジュールと呼ばれている。このモジュール内におけるドメインの組成と反応様式の違いに基づいて PKS は大きく分けてタイプ I, II, III の 3 種類に分類されており、全てのタイプが縮合反応に関わる KS ドメインを保持しているが、タイプ I, II の KS ドメインとタイプ III の KS ドメインの配列間には大きな隔たりが見られる。

現在、ゲノム決定や配列解析によってポリケチドと推定されるタンパク質の遺伝子配列が大量に発見されているが、その機能は未知のものも多く、タンパク質の配列から計算機的に生成物ポリケチドの化学構造及びその性質を予測するという試みが行われている。生成物予測に利用できるような包括的な分類体系を整えるため、我々はゲノムが解明された生物種の遺伝子・タンパク質データベースである KEGG GENES に対し配列類似度検索法である PSI-BLAST によって PKS 候補遺伝子を抽出した後、タンパク質ドメイン情報データベース PFAM 及び HMMER プログラムを用いてそのドメインの組み合わせを決定し、UniProtKB に含まれる機能既知の PKS を含めて KS ドメインの系統樹を作成することでその PKS のタイプを決定し、既知 PKS 配列を含むクラスターの機能アノテーションを行ったが、既知 PKS のうちのいくつかはゲノムプロジェクトから得られたものの PKS 候補ともクラスターを形成しないものがあった。そこで、本研究では、近年蓄積されてきている様々な環境における培養困難な生物のメタゲノムデータやトランスクリプトームデータを用いることで既知 PKS の独立クラスターを補完し、ゲノムの決まっていない生物も含めた PKS の多様性を探索することを目的として解析を行った。

検討内容

既知の PKS の KS ドメインをクエリーとし、KEGG GENES と Uniref90 データベースを併合し、冗長性を除いた(CD-HIT100%)データベース(以下、リファレンス NRDB と呼ぶ)を作成し、これに対し BLASTP を用いて検索(閾値 $e\text{-value} 10^{-10}$)を行い、高類似度配列(アライメント長 80%以上、非フラグメント、CD-HIT95%によりフィルタリング)を取得し、これと元クエリー配列から PSI-BLAST を用いて KS ドメイン様配列のプロファイル(PSSM)を作成した。クエリーとしてはタイプ I/II から 3 種類(eryAI, rifC, pikAIII)とタイプ III から 4 種類(CH5, STS, PhID,

RppA)を採用し、それぞれに対してプロファイルを作成した。これらのプロファイルをそれぞれクエリーとし、リファレンス NRDB に対して PSI-BLAST で検索(閾値 $e\text{-value}10^{-5}$)を行うことで PKS 候補配列を取得し、タイプ I/II クエリー、タイプ III クエリーから得られたアラインメント領域をそれぞれ併合しタイプ I/II KS ドメイン候補配列とタイプ III KS ドメイン候補配列を得た。この KS ドメイン候補配列のマルチプルアラインメントを行い(MAFFT)、系統樹を作成する(QuickTree, Neighbor-Joining)ことにより PKS と PKS 以外(主に FAS からなる)への分離を行った。同様にメタゲノム、トランスクリプトームデータベース(以下 MG/TDB と呼ぶ)に対して KS ドメイン候補配列を取得し、リファレンス NRDB から得られた KS ドメイン候補配列への BLASTP ベストヒットを取ることで PKS 配列を取得した。MG/TDB は、Human Microbiome Project (HMP, 約 7,600 万配列)によるヒトの腸内細菌叢を始めとした様々なヒト体内、表面の細菌叢メタゲノムデータ、Global Ocean Sampling Expedition (GOS, 約 4,100 万配列)と TARA Oceans (約 1 億 1,100 万配列)による海洋環境における微生物群のメタゲノムデータ、Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP, 約 1,500 万配列)による海洋環境における真核生物トランスクリプトームデータの 4 つを使用した。最後に、リファレンス NRDB から得られた PKS 配列の KS ドメインと MG/TDB から得られた PKS 配列の KS ドメインを CD-HIT によりクラスタリングし、クラスター数を調べることで未知の PKS 候補のバリエーションを探った。

結果と考察

リファレンス DB に含まれるタイプ I/II、タイプ III の PKS 配列の数はそれぞれ 22,341、2,473 となった。MG/TDB に含まれるタイプ I/II の PKS 配列の数は HMP で 8,958、GOS で 2,259、MMETSP で 9,061、TARA で 11,551 となった。また、タイプ III の PKS 配列の数は HMP で 1,021、GOS で 316、MMETSP で 974、TARA で 2,265 となった。データベース中のタイプ I/II の PKS 配列の数とタイプ III の PKS 配列の比を取ると、TARA は他のデータベースに比べてタイプ III の割合が多かった。得られた PKS の配列数をデータベース全体の配列数と比較するとトランスクリプトームデータベースである MMETSP がメタゲノムデータベースに比べて高い割合を示した。クラスタリングの結果(図 1)、海洋環境からタイプ I/II、タイプ III 共に多くの新奇 PKS 様クラスターが見つかったが、ヒト細菌叢からはそれほど多く見つからなかった。TARA は特にタイプ III 新奇クラスターを多く持っており、MMETSP は特にタイプ I/II 新奇クラスターを持っていることがわかった。また、タイプ I/II では MG/TDB の配列とリファレンス DB の配列が同じクラスターに入るケースが多く見つかったが、タイプ III ではそれほど多く見つからなかった。このことからタイプ I/II の MG/TDB の配列はある程度既存のリファレンス配列からアノテーションが可能であると考えられるが、タイプ III については難しいといえる。

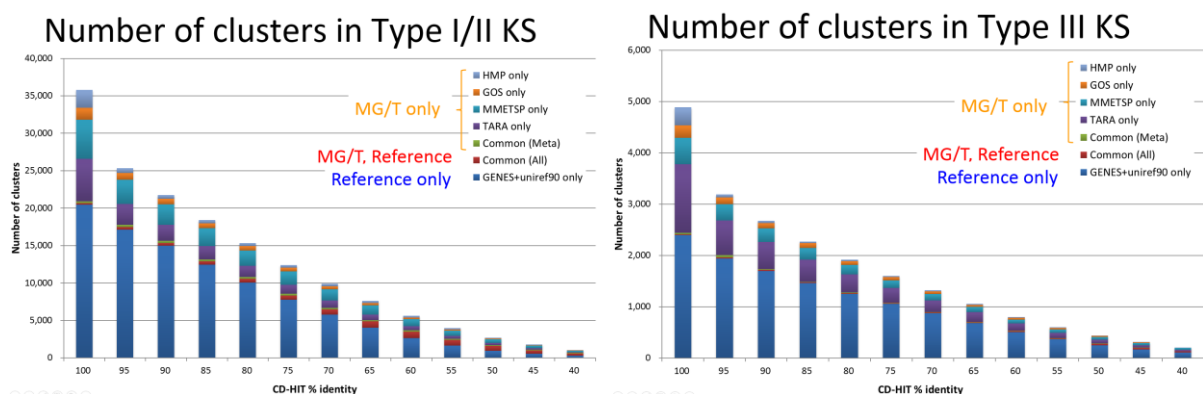


図 1. KS ドメインの様々な配列類似度の CD-HIT クラスタリングによる PKS クラスター数の変化