
MARCIN BIZUKOJC

Wydział Inżynierii Procesowej i Ochrony Środowiska, Politechnika Łódzka, Łódź

Automatyczna anotacja genomu jako narzędzie biologii systemów

Wprowadzenie

Określenie sekwencji nukleotydów (genomu) wybranego organizmu jest obecnie stosunkowo proste dzięki *high throughput technologies* i stanowi pierwszy etap rekonstrukcji jego sieci metabolicznej. Do wykonania funkcjonalnej anotacji genomu *in-silico*, zastosowanie znajdują algorytmy porównywania sekwencji. U podstaw ich działania jest zasada, że u wszystkich organizmów na Ziemi można znaleźć podobne geny zwane genami ortologicznymi. Są one funkcjonalnie zachowane u różnych gatunków ze względu na ich ewolucję od wspólnych przodków [1]. Najpopularniejszym ogólnie dostępnym narzędziem anotacji genomu jest KEGG *Automatic Annotation Server* (KAAS) dostępny w *Kyoto Encyclopedia of Genes and Genomes* (KEGG). Dysponując proteomem lub transkryptomem jakiegoś organizmu w formacie FASTA można przyporządkować transkrypty albo sekwencje aminokwasów do identyfikatorów ortologii KEGG (KO). Z kolei ortologie KEGG są związane z obiektami map KEGG. Kilkaset map KEGG obejmuje najważniejsze szlaki metaboliczne żywych organizmów wraz z pełną informacją na temat metabolitów w nich uczestniczących oraz enzymów katalizujących poszczególne przemiany [1].

Automatyczna anotacja pełnego genomu powinna teoretycznie pozwolić na odtworzenie kompletnej, składającej się z od kilkuset do kilku tysięcy reakcji sieci metabolicznej

całego organizmu, co jest pożądane przy holistycznym podejściu badawczym do organizmu żywego w biologii systemów. Taka sieć mogłaby być wykorzystana, np. do zaawansowanego modelowania metabolicznego na poziomie stechiometrii (*Metabolic Flux Analysis*). W praktyce jednak istnieją ograniczenia badania metabolizmu *in-silico* ze względu na trudność w identyfikacji funkcji genów

Celem pracy jest pokazanie możliwości zastosowania automatycznej anotacji genomu do badania metabolizmu mikroorganizmów na przykładzie grzybów nitkowych z rodzaju *Aspergillus*. Przedstawione zostaną możliwości oraz ograniczenia tej metody.

Materiały i metody

Analizę przeprowadzono dla siedmiu gatunków grzybów nitkowych rodzaju *Aspergillus*: *A. niger*, *A. nidulans*, *A. oryzae*, *A. fumigatus*, *A. terreus*, *A. flavus* i *A. clavatus*. Dla czterech pierwszych gatunków anotacja znajduje się w KEGG. Dla trzech ostatnich wykorzystano serwer KAAS z danymi wejściowymi w postaci plików FASTA zawierających transkrypt lub proteom danego gatunku. Pochodziły one z *Broad Institute (Massachusetts Institute of Technology)*. Do anotacji użyto algorytmu *single-directional best hit*, a jako materiał porównawczy genów ortologicznych użyto zestawu referencyjnego dla *Eukaryota* proponowanego przez KAAS [2].

Wyniki badań

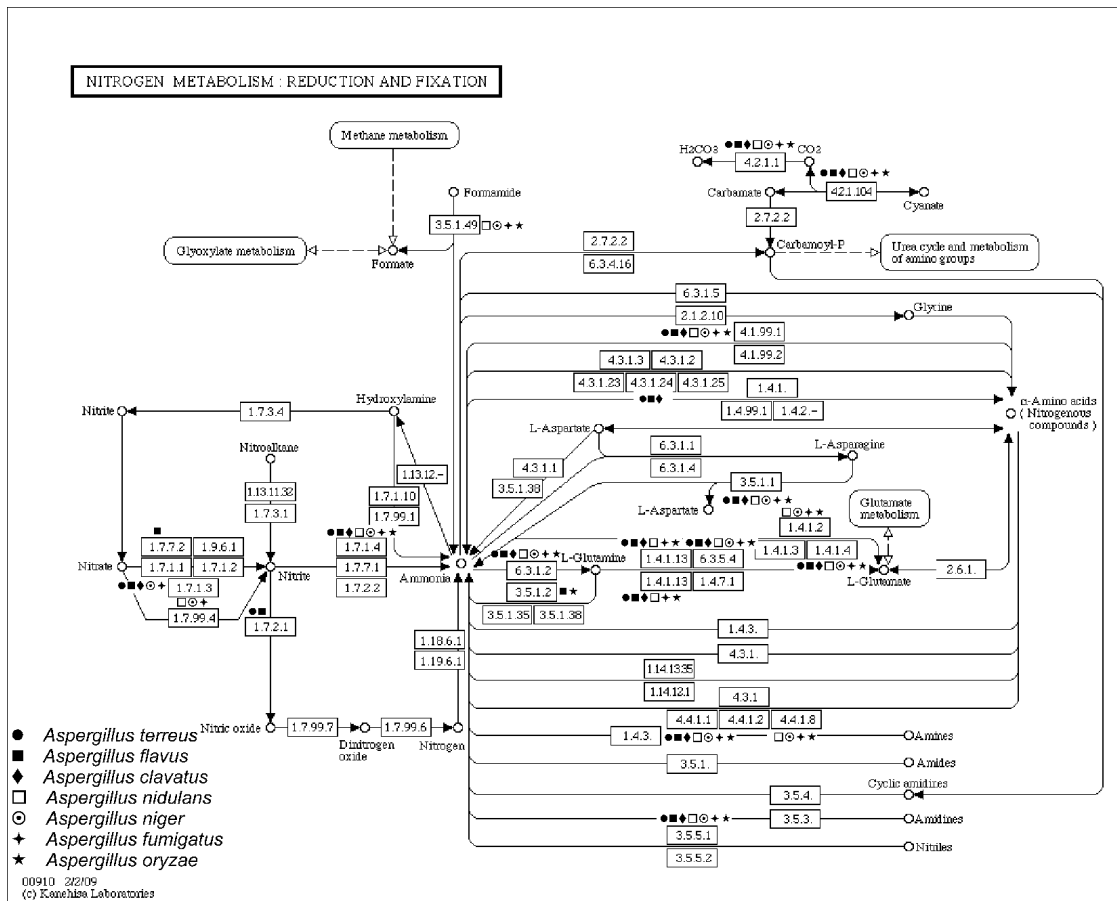
Najczęstszym zastosowaniem automatycznej anotacji genu jest porównanie zrekonstruowanych sieci metabolicznych dla analizowanych organizmów. Jako przykład omówiony zostanie metabolizm azotu i glicerolu dla siedmiu gatunków grzybów nitkowych z rodzaju *Aspergillus*.

Analizując sieć metabolizmu azotu (Rys. 1) można zadać pytanie, czy badane organizmy są zdolne do wykorzystywania azotanów jako źródła azotu. Reduktaza azotynowa (NAD/NADP) redukująca azotyny do jonów amonowych (EC 1.7.1.4) występuje u wszystkich siedmiu gatunków badanych grzybów. Zatem z punktu widzenia zdolności asymilacji azotanów ważniejszy jest etap redukcji azotanów do azotynów. Potencjalnie według KEGG aż sześć enzymów jest zdolnych do przeprowadzenia takiej reakcji. Jednakże reduktaza azotanowa sprzężona z NAD (EC 1.7.1.1) nie została odnaleziona ani u *A. nidulans*, ani u *A. oryzae*. *A. flavus* dodatkowo posiada

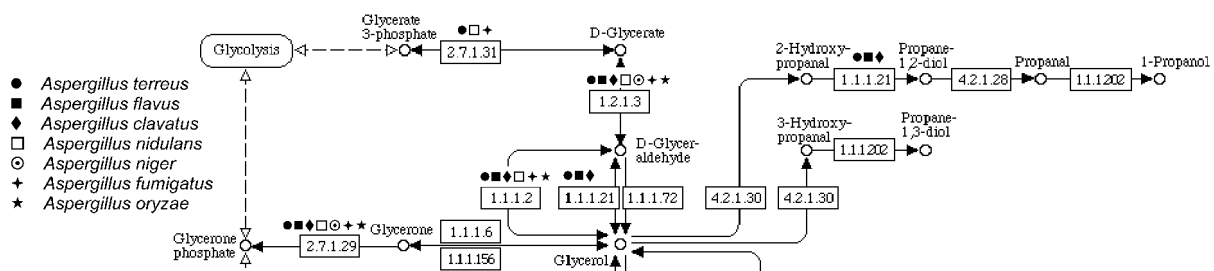
reduktazę azotanową EC 1.7.7.2, której kofaktorem jest ferredoksyna, zaś *A. fumigatus* i *A. niger* mają nietypową reduktazę EC 1.7.99.4 o nieznanym kofaktorze. Tenże enzym został także odnaleziony u *A. nidulans*. Zatem według anotacji tylko *A. oryzae* jest pozbawiony możliwości wzrostu na azotanach jako jedynym źródle azotu.

Analizując inne reakcje na rys. 1 można stwierdzić, że *A. nidulans*, *A. niger*, *A. fumigatus* i *A. oryzae* są zdolne do rozkładu formamidu do mrówczanu i amoniaku dzięki obecności hydrolazy formamidu (EC 3.5.1.49). Wszystkie badane gatunki mają syntetazę glutaminy (EC 6.3.1.2) odpowiedzialną za reakcję wiązania jonów amonowych do glutamianu, natomiast *A. terreus* oraz *A. oryzae* posiadają jeszcze glutaminazę (EC 3.5.1.2) rozkładającą L-glutaminę z powrotem do kwasu glutaminowego i jonów amonowych.

Na rys. 2 przedstawiono fragment mapy KEGG z metabolizmem glicerolu przedstawiający dwa szlaki włączania glicerolu do glikolizy. Te szlaki to redukcja glicerolu do gliceronu



Rys. 1. Metabolizm azotu. Oryginalna mapa KEGG z nałożonymi symbolami znalezionych anotacji dla badanych siedmiu gatunków *Aspergillus*



Rys. 2. Metabolizm glicerolu u *Aspergillus* sp. Fragment oryginalnej mapy KEGG z nałożonymi symbolami dla badanych siedmiu gatunków *Aspergillus* sp.

(dihydroksyacetonu) i jego fosforylacja oraz utlenienie glicerolu do aldehydu D-glicerynowego i D-glicerynianu oraz fosforylacja tego ostatniego. Żaden z badanych gatunków *Aspergillus* nie posiada ani dehydrogenazy glicerolowej (EC 1.1.1.6) ani dehydrogenazy 2-glicerolowej (EC 1.1.1.156). Te enzymy występują raczej w organizmach prokariotycznych, szczególnie w bakteriach anaerobowych i fakultatywnie aerobowych [4]. *A. niger* natomiast okazał się jedynym badanym szczepem niezdolnym do utleniania glicerolu do aldehydu D-glicerynowego, gdyż według anotacji nie posiada ani niespecyficznego dehydrogenazy alkoholowej NADP (EC 1.1.1.2) ani reduktazy aldehydowej (EC 1.1.1.21), katalizujących odwracalne reakcje. Natomiast do fosforylacji glicerynianu przy trzecim węglu (kinaza glicerynianowa, EC 2.7.1.31) zdolne są tylko *A. terreus*, *A. nidulans* i *A. fumigatus*.

Wyniki anotacji należy zawsze skonfrontować z danymi fizjologicznymi, ponieważ brak anotacji enzymu nie jest stu-procentowym dowodem na jego nieistnienie.

Wykorzystywanie azotanów jako źródła azotu to generalnie cecha większości grzybów nitkowych z rodzaju *Aspergillus* [3]. Na przykład *A. terreus* jest na pewno zdolny do wykorzystywania azotanów jako źródła azotu [5] oraz glicerolu jako źródła węgla [6]. Dehydrogenaza glicerolowa EC 1.1.1.72 nie została anotowana dla żadnego grzyba, ale była wcześniej wyizolowana z *A. niger* i *A. nidulans* [7]. Potwierdzone jest także jej występowanie u *A. oryzae* [4]. To jest więc przykład typowej „dziury” w anotacji sieci metabolicznej. Dla kinazy glicerynianowej EC 2.7.1.31 sytuacja jest trudna do określenia, gdyż tylko u *A. nidulans* ten gen został znaleziony i sklonowany *in-vitro* [8]. Brak reduktazy azotanowej u *A. oryzae* potwierdzony jest w [4].

Innym zastosowaniem map KEGG jest poszukiwanie konkretnego enzymu na podstawie anotacji sieci metabolicznej. Wiadomo na przykład, że grzyby nitkowe są zdolne do hydroksylacji węglowodorów aromatycznych, zarówno jednodobych i wielopierścieniowych [9]. Dla badanych *Aspergillus* przeanalizowano mapę KEGG biodegradacji toluenu, a dokładnie reakcje utleniania toluenu przez 2-monooksygenazę fenolową (EC 1.14.13.7) do 3-metylokatecholu. Spośród badanych gatunków grzybów nitkowych tylko w genomie *A. niger* i *A. fumigatus* znaleziona została odpowiednia anotacja. Zatem można potencjalnie przyjąć, że prawdopodobnie te dwa organizmy są w stanie hydroksylować toluen i należałoby szukać potwierdzenia tej właściwości eksperymentalnie.

Badanie metabolizmu wtórnego przy pomocy anotacji genomu jest utrudnione ze względu na specyficzność tego typu metabolitów. Są one syntezowane tylko przez jeden bądź kilka określonych gatunków szczepów mikroorganizmów, a w KEGG brakuje odpowiednich ortologii. W przypadku *A. terreus* w proteomie można znaleźć syntazę nonaketydową lowastatyny (EC 2.1.3.161), która nie ujawnia się na mapach KEGG, gdyż one nie obejmują biosyntezy tego metabolitu.

W KEGG zawarte jest kilka map biosyntezy metabolitów wtórnych, ale na przykład żaden z badanych *Aspergillus* nie posiada enzymów biosyntezy kwasu klawulonowego, ani poliketydowych makrolidów. To jest akurat zgodne z wiedzą na temat fizjologii grzybów, gdyż te wtórne metabolity są syntezowane przez prokariotyczne *Streptomyces* sp.

Natomiast u *A. flavus* KAAS znalazł anotację do mapy biosyntezy flawonoidów, potwierdzając obecność poliketydowej syntazy typu chalkonowego EC 2.3.1.74 której produktem jest między innymi chalkon naringeniny.

Podsumowanie

Na podstawie powyższych badań wyraźnie zarysowują się ograniczenia automatycznej anotacji genomu przy użyciu KAAS czy innych tego typu narzędzi. Po pierwsze taka anotacja jest zawsze mniej skuteczna w przypadku dysponowania częściowo zsekwencjonowanym genomem i wtedy może zabraknąć niektórych anotacji. Po drugie zestaw map KEGG jest nadal ograniczony i wiele szlaków metabolicznych szczególnie dotyczących metabolizmu wtórnego nie zostało w nich jeszcze uwzględnionych. Na przykład brakuje też w KEGG mapy metanogenezy. To powoduje, że nawet odnalezione sekwencje niektórych enzymów zawarte w pliku FASTA nie zostaną przyporządkowane do map KEGG. Zdecydowanie znacznie wyższa skuteczność anotacji genomu w KEGG jest w przypadku analizy metabolizmu pierwotnego.

Brak jakiegoś enzymu w szlaku metabolicznym nie oznacza, że z pewnością on nie istnieje, takie braki nazywa się *holes* („dziury”) i można je wypełnić dysponując danymi z eksperymentów *in-vitro*. Z powyższych powodów bardzo istotne jest porównywanie uzyskanych wyników tego modelowania szlaków metabolicznych *in-silico* na podstawie genomu (genotypu) z eksperymentalnymi danymi na temat fizjologii (fenotypem) danego mikroorganizmu oraz „ręczne” poszukiwanie danych na temat niektórych enzymów w literaturze (bibliomie).

LITERATURA

1. Y. Moriya, M. Itoh, S. Okuda, A.C. Yoshizawa, M. Kanehisa: *Nucleic Acids Research* 35 Web Server Issue doi: 10.1093/nar/gkm321 (2007).
2. KAAS – KEGG Automatic Annotation Server: www.kegg.jp/kaas
3. N. Kitamoto, T. Kimura, Y. Kito, K. Ohmiya, N. Tsukagoshi: *Bioscience, biotechnology and biochemistry*, **59**, nr 91, 795, (1995).
4. BRENDA Braunschweig Enzyme Database: www.brenda-enzymes.info
5. H. Hajjaj, P. Niederberger, P. Duboc: *Appl. Environ. Microbiol.* **67**, 2596 (2001).
6. J.L. Casas Lopez, J.A. Sanchez Perez, J.M. Fernandez Sevilla, F.G. Acien Fernandez, E. Molina Grima, Y. Chisti: *Enz. Microb. Technol.* **33**, 270 (2003).
7. R. Schuurink, R. Busink, D.H. Hondmann, C.F. Witteveen, J. Visser: *J. Gen. Microbiol.* **136**, 1043, (1990).
8. J.M. Clements, C.F. Roberts: *Curr. Gen.*, **9**, 293 (1985).
9. K. Lisowska, M. Bizukojć, J. Długoński: *Enz. Microb. Technol.* **39**, 1464, (2006).