

Morfologiczny chaos?

W XVIII wieku Karol Linneusz w dziele *Systema naturae* opisał 7,7 tysiąca gatunków roślin oraz ponad 4 tysiące gatunków zwierząt, stosując po raz pierwszy binominalne nazewnictwo i tworząc tym samym podwaliny współczesnej taksonomii. Mimo iż w ciągu kolejnych 230 lat sklasyfikowano 1,7 miliona gatunków eukariontów, naukowcy wciąż mają przed sobą nie lada wyzwanie – szacuje się, że na świecie znaleźć można ponad 10 milionów organizmów!

Każdy taksonom musi zatem zmierzyć się z niesamowitą bioróżnorodnością, opierając się na specjalistycznych kluczach do oznaczania. Podczas identyfikacji gatunku przy użyciu takiego klucza brane są pod uwagę różne cechy morfologiczne, zarówno w budowie zewnętrznej, jak i wewnętrznej, często bardzo subtelne. W zależności od identyfikowanego osobnika analizuje się np. budowę liści, barwę czy wielkość ciała.

Klucze do oznaczania działają na zasadzie kolejnych zawężeń zbioru wynikowego – należy zapoznać się z następującymi po sobie stwierdzeniami, wybierając z dwóch możliwych opcji tę poprawną. Wybór odsyła użytkownika do kolejnego zestawu cech ze zbioru z odpowiednio zawężonymi informacjami, przy czym z każdym krokiem maleje liczba gatunków stanowiących wynik identyfikacji. Szczegółowe informacje i opis gatunku pojawiają się po zakończeniu analizy i dopiero wtedy jednoznacznie można stwierdzić, czy nigdzie po drodze nie została popełniona żadna pomyłka, a wynik oznaczenia faktycznie jest poprawny.

Niestety, oznaczanie gatunków metodą tradycyjną nadal pozostaje bardzo trudne w przypadku posiadania jedynie fragmentu organizmu lub jego formy młodocianej czy larwalnej.

Linneusz idzie w odstawkę?

Identyfikacja gatunków na podstawie cech morfologicznych często

Barkoding DNA – ideał czy idea?

Poprawna identyfikacja gatunkowa jest podstawą wszelkich badań biologicznych, ekologicznych i biogeograficznych. Umiejętność dokładnego oznaczenia przynależności organizmu do gatunku oraz znajomość biologii i ekologii znacznie ułatwiają badania aplikacyjne np. nad bakteriami (u których jednak termin gatunek stosowany musi być z niezwykłą ostrożnością), szybką diagnostykę użytej podczas ataków terrorystycznych broni biologicznej czy zwalczanie gospodarczo i ekonomicznie niebezpiecznych szkodników upraw. Rozwój biologii molekularnej, zwłaszcza badania nad rozszyfrowaniem kodu genetycznego i wyjaśnianiem mechanizmów dziedziczenia, dostarczył nowych metod identyfikacji gatunków, dzięki którym można w sposób pewny je zidentyfikować i to w bardzo krótkim czasie. Czy aby jednak na pewno?

jest zadaniem żmudnym i skomplikowanym, co może utrudnić nie tylko prowadzenie badań naukowych, ale także podejmowanie dalszych działań (np. w przypadku niepoprawnej identyfikacji groźnych patogenów opóźnione jest postawienie trafnej diagnozy i szybkie rozpoczęcie leczenia). Ułatwienie identyfikacji gatunkowej uznano zatem za kwestię priorytetową, angażując także inne dziedziny, w tym najprężniej rozwijającą się biologię molekularną.

Badania molekularne wykorzystują tzw. markery molekularne, czyli fragmenty DNA, stanowiące swoistą „metkę” specyficzną dla danego gatunku. Metodę identyfikacji gatunków z wykorzystaniem sekwencji markerowych po raz pierwszy wykorzystał w 2003 roku Paul Hebert z Instytutu Bioróżnorodności Uniwersytetu Guelph w Kanadzie. Umożliwiło mu to zbadanie setek dorosłych osobników motyli z gatunku *Astraptes fulgerator*, żyjących w Kostaryce. Motyle te okazały się bardzo do siebie podobne jedynie morfologicznie, w rzeczywistości stanowiąc kompleks dziesięciu różnych gatunków. Sekwencje markerowe mogą

pełnić funkcję swoistych „kodów paskowych” (ang. *barcodes*) ułatwiających szybką identyfikację osobników. Pomysł tworzenia katalogów genetycznych „metek” dla wszystkich organizmów żywych, kierowany obecnie przez konsorcjum CBOL (ang. *Consortium for the Barcode of Life*), okazał się początkiem prawdziwej rewolucji opartej na trzech podstawowych technikach biologii molekularnej: izolacji DNA, łańcuchowej reakcji polimerazy (PCR) i sekwencjonowaniu (tab. 1).

Zaskakujące rozwiązania

Początkiem szybkiego opracowywania „kodów paskowych” DNA było powstanie w 2004 roku konsorcjum zmierzającego do stworzenia baz danych unikatowych sekwencji DNA, jednoznacznie określających poszczególne gatunki. Równocześnie podjęto próbę stworzenia prostego, najlepiej przenośnego, przyrządu do identyfikacji „kodu paskowego” badanej próbki materiału biologicznego (tzw. barkodera).

Konsorcjum (<http://moc.barcodeoflife.net/>) zrzesza ponad 170 organizacji z 50 krajów (muzea, uni-

Tabela 1. Techniki biologii molekularnej

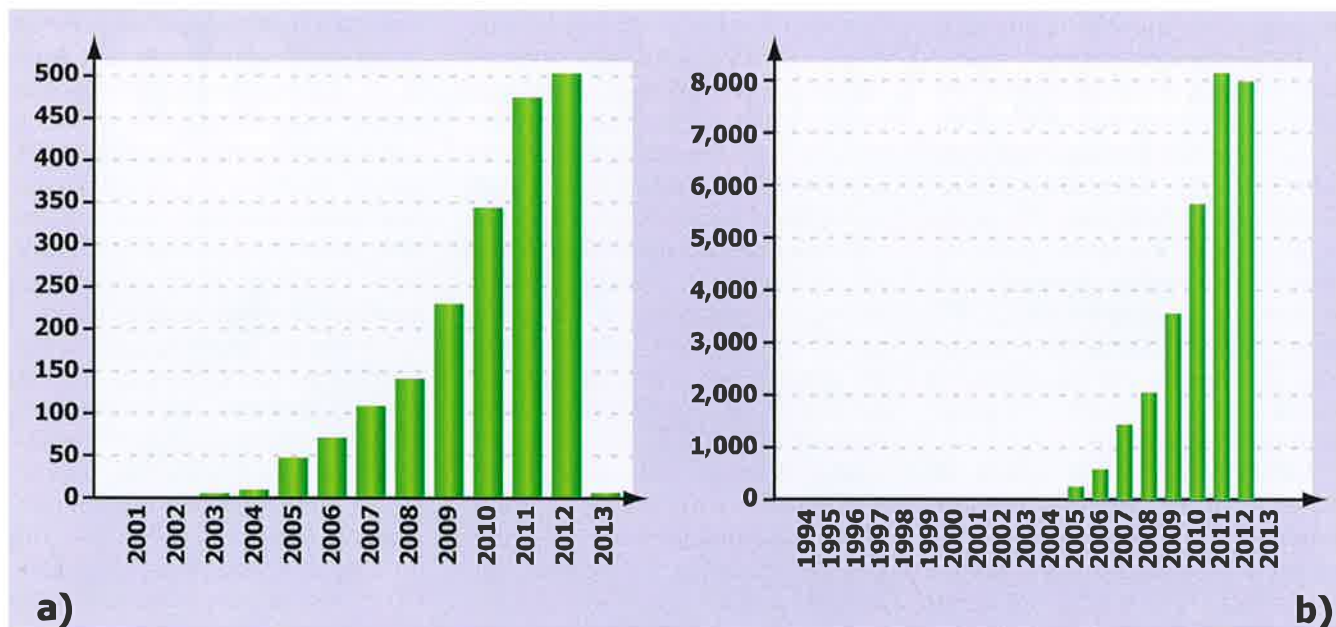
Izolacja DNA
Izolacja DNA jest niezwykle istotnym etapem pracy biologa molekularnego. Celem izolacji jest dostarczenie jak największej ilości DNA o wysokiej jakości. Procedura musi zostać dopasowana do typu komórki, z której izolowany jest DNA, np. w przypadku komórek zwierzęcych należy usunąć z materiału duże ilości białka, np. stosując enzym proteinazę K. Obecnie zazwyczaj podczas izolacji wykorzystuje się zestawy znacznie ułatwiające procedurę, przygotowane przez firmy biotechnologiczne.
Łańcuchowa reakcja polimerazy (PCR)
To metoda służąca do powielania fragmentów DNA w warunkach laboratoryjnych, prowadząca do uzyskania 2n cząsteczek dwuniciowego DNA, gdzie n stanowi liczbę przeprowadzonych cykli. Do reakcji wymagane są odpowiednio zaprojektowane, krótkie fragmenty DNA (tzw. startery), komplementarne do fragmentów nici znajdujących się na obu końcach powielanego fragmentu (tzw. forward i reverse), polimeraza DNA i przyłączane przez nią trifosforany deoksyrybonukleozydów, a także bufor o optymalnym stężeniu jonów Mg^{2+} niezbędnych dla polimerazy DNA. Najpopularniejszą stosowaną polimerazą jest Taq polimeraza zastępowana czasem przez inne polimerazy, np. Pwo lub Pfu, które mają zdolność natychmiastowego redagowania wprowadzonych przez siebie błędów (tzw. proof-reading). Reakcja zwykle przebiega w trzech etapach powtarzanych kilkadziesiąt razy: denaturacja dwuniciowego DNA, annealing (przyłączenie starterów) i elongacja nici.
Sekwencjonowanie
W 1980 roku trzech naukowców (Allan Maxam, Walter Gilbert i Frederick Sanger) otrzymało Nagrodę Nobla za wynalezienie metod umożliwiających odczytanie kolejności par nukleotydowych w cząsteczce DNA. Metoda Maxama-Gilberta wykorzystuje chemiczną degradację DNA wyznakowanych na 5'-końcu radioaktywnym izotopem fosforu ^{32}P . Degradacja przebiega w czterech oddzielnych reakcjach wykorzystujących odmienne zestawy związków chemicznych. W pierwszym etapie procesu następuje modyfikacja specyficznych dla danego odczynnika zasad azotowych nukleotydów oraz ich selektywne oderwanie od reszt cukrowych. Następnie nić DNA ulega przecięciu. Ponieważ modyfikacje, a w konsekwencji również cięcie DNA są tylko częściowe, w mieszaninie poreakcyjnej znajduje się zbiór wielu fragmentów DNA o różnej długości, zależnej od miejsca modyfikacji i cięcia DNA. Fragmenty te rozdziela się elektroforetycznie i uwidacznia, stosując autoradiografię.
Metoda dideoksy – metoda Sangera
Pierwotnie polegała ona na dodaniu do reakcji syntezy nowych, komplementarnych nici DNA niewielkiej ilości określonego trifosforanu dideoksynukleozydu (ddNTP). Reakcję prowadzono równocześnie w czterech oddzielnych probówkach. Do każdej próbówki dodawano tylko jeden rodzaj dideoksynukleotydu, którego wbudowanie zatrzymuje syntezę DNA. Ponieważ w probówkach znajdowały się równocześnie dezoksynukleotydy, w tym jeden wyznakowany ^{32}P , otrzymane, radioaktywne fragmenty DNA, o różnej długości, rozdzielane są elektroforetycznie (segregowane według wielkości). Sekwencję nukleotydową badanego fragmentu DNA odczytywano na podstawie otrzymanego autoradiogramu. Obecnie odczyt Sangera został zautomatyzowany dzięki wykorzystaniu fluorescencyjnie znakowanych trifosforanów dideoksynukleozydów.
Sekwencjonowanie wysokoprzepustowe
Obecnie stosowane są najnowocześniejsze metody sekwencjonowania wysokoprzepustowego (ang. deep sequencing), tzw. sekwencjonowanie następnego pokolenia (ang. next generation sequencing), m.in. pirosekwencjonowanie, sekwencjonowanie z ligacją czy sekwencjonowanie „mostkowe”. Pirosekwencjonowanie polega na przyłączeniu cząsteczek DNA do nośnika (np. kulek) i – po ich namnożeniu – umieszczeniu „kompleksów” w studzienkach. Do studzienek dodaje się kolejno cztery nukleotydy, z których jeden, komplementarny, przyłącza się do nici, uwalniając pirofosforan i powodując reakcję z wydzielaniem kwantu światła, który jest rejestrowany (sekwencjonowanie przez syntezę, około 800 000 odczytów po 250–500 pz). Sekwencjonowanie z ligacją wykorzystuje do identyfikacji nukleotydu w określonym miejscu w analizowanej sekwencji enzym ligazę DNA i wrażliwość katalizowanej przez nią reakcji na błędne parowanie zasad (metoda ta nie wymaga polimerazy DNA i tworzenia drugiej nici). Sekwencjonowanie „mostkowe” zaczyna się od losowego pofragmentowania DNA i przyłączenia, do obu końców otrzymanych fragmentów, elementu „przejściowego” (ang. adapter). Po denaturacji cząsteczki ssDNA przyłączane są do stałej powierzchni, na której znajdują się odpowiednie startery. Po dodaniu nieznakowanych nukleotydów i enzymu nić ugina się, tworząc „mostek” mający oba końce przyczepione do powierzchni i dobudowując drugą komplementarną nić. Nici ulegają rozszczepieniu po zakończeniu syntezy i „wyprostowaniu się”, każda w przeciwną stronę. Sekwencjonowanie zaczyna się dopiero po zakończeniu namnażania wielu takich pojedynczych nici. Sekwencjonowanie składa się z wielu cykli, z których każdy rozpoczyna się dodaniem czterech znakowanych, odwracalnych terminatorów, starterów i polimerazy DNA oraz umożliwia identyfikację jednej zasady (po wzbudzeniu laserem wbudowanego terminatora).

wersytety, ośrodki rządowe, prywatne firmy), które gromadzą sekwencje barkodowe charakterystyczne dla wszystkich gatunków motyli (*All-Leps*), ryb (*Fish Barcode of Life*), ptaków (*All Birds Barcoding Initiative*) oraz organizmów zamieszkujących regiony polarne (*Polar Barcode of Life Initiative*). Natomiast tworzona na Uniwersytecie w Guelph baza danych *The Bar-*

code of Life (BOLD) (www.barcodeoflife.org) służy nie tylko do archiwizacji uzyskanych w wyniku analiz markerowych sekwencji DNA. Dane molekularne połączone są z opisem morfologicznym i anatomicznym gatunku, informacjami o rozmieszczeniu oraz odnośnikami do literatury o nim traktującej.

W Polsce analizą z wykorzystaniem markerowych sekwencji

DNA (od 2006 roku) zajmuje się Krajowy Bank DNA Roślin, Grzybów i Zwierząt (www.bankdna.pl), będący inicjatywą kilku krajowych instytucji naukowych (Muzeum i Instytutu Zoologii PAN w Warszawie, Instytutu Botaniki PAN w Krakowie, Instytutu Hodowli i Aklimatyzacji Roślin w Radzikowie oraz Instytutu Genetyki i Hodowli Zwierząt PAN w Ja-



Rys. 1. Statystyki przedstawiające rosnącą popularność barkodingu DNA w latach 2003–2012: (a) liczba publikacji prezentujących przeprowadzone analizy barkodowe oraz (b) częstość cytowania tych publikacji (według webofknowledge.com)

strzębcu). Jego celem jest powiększenie posiadanej wiedzy na temat różnych gatunków roślin, grzybów i zwierząt poprzez gromadzenie materiału biologicznego i przeprowadzanie licznych badań taksonomicznych z zastosowaniem „kodów paskowych” DNA.

Obserwując działalność naukową największych ośrodków badawczych na świecie oraz analizując statystyki publikacji (rys. 1), można odnieść wrażenie, że naukowcy wręcz zachłystnęli się nową metodą, uznając ją za antidotum na wszelkie problemy z identyfikacją gatunkową. Powstaje jednak kluczowe pytanie – czy istotnie metoda ta ma szansę na zastąpienie metod tradycyjnych?

Potencjał barkodingu DNA umożliwia klasyfikację gatunków już znanych nauce oraz stanowi doskonałe narzędzie w odkrywaniu gatunków bliźniaczych. W przeciwieństwie do metod tradycyjnych identyfikacja możliwa jest także w przypadku fragmentów organizmu (np. pióra, kości) czy stadiów trudnych w oznaczaniu (np. larwy, nasiona). Również analiza gatunków kryptycznych, które morfologicznie różnią się nieznacznie, tak że ich określenie jest prawie niemożliwe, stała się łatwiejsza.

Barkodingu DNA okazuje się użyteczny w badaniu antropopresji

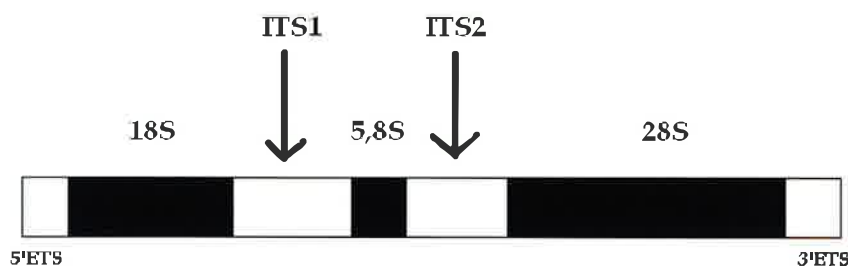
na bioróżnorodność, zarówno w skali regionalnej, jak i globalnej, znacznie ułatwiając monitoring środowiska. Niektórzy twierdzą, że może być przydatny przy odtwarzaniu filogenetycznych zależności oraz skutecznym prognozowaniu i zapobieganiu inwazji obcych gatunków. Diagnostyka z wykorzystaniem barkodingu DNA z powodzeniem stosowana jest w medycynie do badań alergologicznych oraz przez nauki sądowe do identyfikacji gatunków istotnych w określaniu czasu, przyczyny czy miejsca zgonu. Idea?

Genetyczne „kody paskowe”

Potencjalna sekwencja „kodu paskowego” DNA powinna charakteryzować się niską zmiennością w obrębie gatunku (ok. 2%) i równocześnie wysoką zmiennością między gatunkami (ok. 10%). Projektowanie starterów do reakcji

PCR oraz późniejszą standaryzację ułatwiają także duża liczba powtórzeń w genomie i oflankowanie konserwatywnymi domenami. Istotna jest również długość sekwencji „kodu paskowego” – krótkie fragmenty, z małą liczbą insercji/delecji, znacznie przyspieszają i ułatwiają identyfikację.

W przypadku identyfikacji gatunków zwierząt główną sekwencją barkodową jest 648-nukleotydowy fragment 5'-końca mitochondrialnego genu podjednostki I oksydazy cytochromowej (ang. *cytochrome c oxidase subunit I*, COI). Geny mitochondrialne zwierząt nie zawierają intronów, a kolista cząsteczka mitochondrialnego DNA (mtDNA) jest mniej podatna na degradację niż jądrowy DNA. Specyficzny sposób replikacji (bez rekombinacji seksualnej, dziedziczenie matczyne) zapewnia mniejszą liczbę insercji/delecji. Sekwencję COI łatwo powie-



Rys. 2. Fragment rDNA z sekwencjami kodującymi 18S, 5,8S i 28S rRNA oraz transkrybowane przez rywniki ITS1, ITS2 i ETS (5' i 3')

lić, jednak nie umożliwia ona identyfikacji i analizy hybryd, ponieważ wymaga to dodatkowych analiz. W tym przypadku wykorzystuje się bardziej informatywny genom jądrowy – sekwencje kodujące rRNA (18S, 5.8S oraz 28S zawierające wysoce konserwatywne domeny) oraz rozdzielające je transkrybowane przekładki ITS1 i ITS2 (ang. *Internal Transcribed Spacer*) (rys. 2). Obecnie ITS stosowane są jednak częściej jako sekwencje bar kodowe u grzybów.

Identyfikacja roślin opiera się na sekwencjach genów chloroplastowych, gdyż mitochondrialna sekwencja COI cechuje się niewystarczającą zmiennością, natomiast DNA genomu jądrowego (nuDNA) jest bardzo podatny na degradację. Co więcej, częste rearanżacje w mtDNA i wymiana genów z nuDNA oznaczają brak stabiliza-

cji i tym samym trudności w standaryzacji. Z kolei roślinne domeny ITS mogą występować u jednego organizmu w różnych kopiach (paralogi). Chloroplastowy DNA (cpDNA) charakteryzuje mała zmienność sekwencji kodujących i wysoka niekodujących. Stworzono dwa zestawy sekwencji markerowych stosowanych w barkodingu DNA. Pierwszy zestaw składa się tylko z sekwencji kodujących (rpoC1, matK, rpoB), drugi wykorzystuje sekwencje dwóch genów i jednej domeny niekodującej (rpoC1, matK, trnH-psbA). Genom chloroplastowy, podobnie jak mtDNA, dziedziczony jest w linii matczynej i analiza z jego wykorzystaniem nie umożliwia rozpoznania gatunków mieszańcowych i introgresantów.

Świetlana przyszłość

Wzmocniona praca naukowców na całym świecie dostarcza coraz większej ilości informacji, wzbogacając bazy danych o sekwencje bar kodowe kolejnych gatunków, przybliżając nas z każdym dniem do celu przedsięwzięcia. Przenośny barkoder będący – zdawać by się mogło – narzędziem na miarę filmów science fiction być może nie jest osiągnięciem aż tak odległym i w niedalekiej przyszłości każdy z nas będzie mógł za jego pomocą identyfikować gatunki na spacerze po parku.

Barkoding DNA zasługuje na miano pełnoprawnego narzędzia taksonomicznego, wymagającego wprowadzenia dopracowania, jednak stanowiącego potencjalne rozwiązanie konfliktu pomiędzy komercyjnym a naukowym wykorzystaniem w identyfikacji metod biologii molekularnej. Badacze udowadniają przecież, że analiza taka może być łatwa, tania i szybka przy uzyskaniu w większości przypadków wyników dokładnych i pewnych.

Niestety, nie można pominąć faktu, iż metoda ta posiada także ciemne strony. W dobie szybkiego rozwoju technik sekwencjonowania, gdy poznanie sekwencji całego organizmu staje się coraz tańsze, grozi metodzie barkodingu DNA „zestarczenie technologiczne”. Prace nad bazą sekwencji bar kodowych

wszystkich organizmów żywych zajmą naukowcom wiele lat i dopóki nie zostaną sfinalizowane, analiza danego fragmentu DNA w próbkę biologicznej pozwoli jedynie na wysnucie wniosku, że organizm ten najprawdopodobniej jest wskazanym przez bazę gatunkiem lub nietety innym z setek pozostałych, niezbar kodowanych do tej pory.

Co więcej, jak wykazano wyżej, „unikatowość” sekwencji również stanowi problem. Już na początku drogi stwierdzono, iż zastosowanie w barkodingu DNA jednej standardowej dla wszystkich organizmów żywych sekwencji DNA jest niemożliwe, a faktyczną „unikatowość” wybranych fragmentów będzie można potwierdzić dopiero po zakończeniu całego projektu identyfikowania sekwencji... unikatowych.

Plany na przyszłość

Koszt sekwencjonowania w ciągu 10 lat obniżył się 14 tysięcy razy przy wzroście prędkości odczytu sekwencji aż o 50 tysięcy razy! Jest to możliwe dzięki prawidłowości, według której im więcej jest znanych sekwencji danego organizmu, tym łatwiej poznać kolejne. Nic dziwnego zatem, że naukowcy i firmy komercyjne twierdzą, iż w ciągu kilku lat uda się im sekwencjonować genomy pojedynczych osób za 1000 dolarów.

Sekwencjonowanie wysokoprężne (ang. *deep sequencing*) umożliwia generowanie w jednym przebiegu milionów niezależnych krótkich odczytów (po 25–400 pz), znajdując zastosowanie m.in. w metagenomice (analiza sekwencji całości DNA wyizolowanego ze zbiorowiska organizmów, stanowiąca jedyny sposób badania mikroorganizmów, które nie dają się hodować).

Ostateczny wynik przesądzający o dalszym losie barkodingu DNA poznamy na pewno już niedługo. Obecnie trudno jednoznacznie stwierdzić, czy metodę tę czeka świetlana przyszłość, czy brutalne odesłanie do lamusa.

mgr Joanna Stojak
IBS PAN Białowieża

Barkodowanie DNA jest techniką, która wśród systematyków, a dokładniej – filogenetyków stosujących metody analizy DNA, znajduje tyluż gorących zwolenników, co zajadłych przeciwników. Ci ostatni zarzucają, że nie jest to metoda dostarczająca informacji dostatecznie precyzyjnych, by mogła być stosowana na poziomie gatunku. Jednak nie można zaprzeczyć, że poza etapem opracowania odpowiednich markerów jest to technika powtarzalna, wiarygodna, tania i wymagająca śladowych ilości materiału genetycznego z badanego organizmu. Niestety zwykle stosowanej w tego typu badaniach sekwencji obszaru COI mitochondrialnego DNA (mtDNA) nie da się wykorzystać w przypadku większości zwierząt, choć w przypadku kręgowców jest ona niezwykle użyteczna. Wiąże się to z polifiletycznym charakterem mtDNA wielu zwierząt. Autorka ma rację, twierdząc, że nie uda się znaleźć jednej, uniwersalnej sekwencji bar kodowej dla wszystkich znanych i nieznanych organizmów. Nie zmienia to faktu, że barkodowanie może być jeszcze bardzo długo użyteczne przykładowo do wstępnego klasyfikowania badanych, np. nowo odkrytych zwierząt. Jak widać, kody paskowe (kreskowe) znajdują zastosowanie zarówno w sklepach, jak i w biologii i to nie tylko przy katalogowaniu prób, np. przechowywanych w kriobankach, ale również do tworzenia swoistych etykiet na podstawie sekwencji nukleotydowej DNA, a w konsekwencji do identyfikacji przynależności gatunkowej badanych organizmów.

<red>