

WERONIKA WRZESIŃSKA, TOMASZ MAMOS, MICHAŁ GRABOWSKI

*Katedra Zoologii Bezkręgowców i Hydrobiologii  
Uniwersytet Łódzki  
Banacha 12/16, 90-237 Łódź  
E-mail: weronika.wrzesinska@biol.uni.lodz.pl  
tomasz.mamos@biol.uni.lodz.pl  
michal.grabowski@biol.uni.lodz.pl*

## ZEGAR MOLEKULARNY W TEORII I PRAKTYCE: PRZEWODNIK DLA POCZĄTKUJĄCYCH\*

### TŁO HISTORYCZNE

Zegar molekularny to metoda szacowania chronologii rozdzielania się poszczególnych linii ewolucyjnych. Podstawą tej koncepcji są różnice między sekwencjami nukleotydowymi (lub aminokwasowymi) u różnych rozpatrywanych grup organizmów, na podstawie których oblicza się czas występowania ich ostatniego wspólnego przodka (ang. most recent common ancestor, MRCA; ostatni wspólny przodek, od którego bezpośrednio pochodzą wszystkie organizmy w danej grupie). Ten rewolucyjny pomysł uważany jest za „jeden z najprostszych i najpotężniejszych pomysłów w dziedzinie ewolucji” (LEWIN 1997) oraz „najbardziej znaczące odkrycie w badaniach nad ewolucją molekularną” (WILSON i współaut. 1977) (patrz MORGAN 1998).

Jakkolwiek koncepcja zegara molekularnego została sformułowana ponad pół wieku temu w oparciu o proste założenia, to kolejne dziesięciolecia badań pokazały złożoność tego zagadnienia. Postęp w dziedzinie technik obliczeniowych zaowocował opracowaniem algorytmów pozwalających na tworzenie coraz bardziej zaawansowanych modeli ewolucji na poziomie sekwencji DNA oraz metod kalibracji zegara molekularnego. Wszystko to sprawia, że zarówno współczesna podbudowa teoretyczna koncepcji zegara molekularnego, jak i stosowanie jej w prak-

tyce wydaje się być, bez mała, wiedzą tajemną dla wielu, zwłaszcza początkujących, badaczy i studentów. Problem pogłębiany jest przez mnogość programów komputerowych i metod tworzenia drzew filogenetycznych. Nasz artykuł ma za zadanie zaznajomić czytelnika z podstawowymi koncepcjami, metodami i programami stosowanymi w datowanej rekonstrukcji filogenezy, jak również zaproponować literaturę, która pomoże mu poszerzyć swoją wiedzę. Pomijając ogólne wzmianki w podręcznikach takich jak *Ekologia Molekularna* (FREELAND 2008), *Ewolucja* (FUTUYMA 2008), *Markery Molekularne, Historia Naturalna i Ewolucja* (AVISE 2008), czy krótki opis w przeglądowej pracy SPALIKA i PIWCZYŃSKIEGO (2009), brak jest na rynku krajowym przystępnych opracowań przybliżających koncepcję zegara molekularnego szerszej grupie odbiorców. Mamy nadzieję, że nasz artykuł chociaż w części wypełni tę lukę.

Nazwę „ewolucyjny zegar molekularny” zawdzięczamy dwóm naukowcom, Emile Zuckerkandlowi i Linusowi Paulingowi, którzy na przełomie lat 50. i 60. XX w. porównali cząsteczki hemoglobiny pochodzące od kilku gatunków ssaków. Zestawili liczbę podstawień w łańcuchach aminokwasowych z czasem rozejścia się linii ewolucyjnych w badaniach gatunków, oszacowanym na podstawie danych kopalnych. Odnotowali, że

**Słowa kluczowe:** zegar molekularny, tempo ewolucji, tempo substytucji, metody kalibracji, modele zegara molekularnego, dywergencja, markery molekularne

\*Publikacja artykułu została sfinansowana ze środków Grantu NCN OPUS8 nr 2014/15/B/NZ8/00266 „Neogeńskie i czwartorzędowe wahania poziomu morza a pochodzenie epigeicznej słodkowodnej fauny pancierzowców (Malacostraca) regionu peryadriatyckiego”.

tempo ewolucji tego białka jest różne w różnych liniach ewolucyjnych. Ta sama zależność dotyczyła innych, testowanych białek, tj. histonów, cytochromu c czy fibrynopeptydów. Zauważyli również, że każdy rodzaj białka posiada charakterystyczne dla niego tempo ewolucji (np. cytochrom c ma szybsze tempo od histonów) (ZUCKERKANDL i PAULING 1962, 1965; MORGAN 1998; BROMHAM i PENNY 2003).

Ważnym etapem rozumienia zegara molekularnego były obserwacje Motoo Kimury, który zauważył istotną dysproporcję pomiędzy niższym tempem substytucji aminokwasów w białkach a szacowanym wyższym tempem substytucji nukleotydów w łańcuchach DNA. Wskazywało to, że znacząca część mutacji na poziomie nukleotydowym nie ujawnia się w łańcuchu białkowym. Mutacje tego rodzaju Kimura określił jako synonimiczne, czyli nie dające efektów fenotypowych i prawdopodobnie neutralne z punktu widzenia doboru naturalnego. W konsekwencji, mutacje dające efekty fenotypowe (niesynonimiczne) korzystne lub szkodliwe dla gatunku muszą podlegać selekcji naturalnej. Według tak sformułowanej Teorii Mutacji Neutralnych (TMN) (KIMURA 1968) liczba mutacji podlegających doborowi jest znikoma, wskutek czego tempo utrwalania nowych mutacji jest zasadniczo równe tempu ich powstawania. Dziś, dzięki licznym eksperymentom „podglądającym” ewolucję w czasie rzeczywistym wiemy, że wiele, a nawet większość nowych mutacji jest neutralnych albo prawie neutralnych (BARRICK i LENSKI 2013).

Przyjęto, że skoro dla danego genu tempo ewolucji jest stałe pomiędzy liniami filogenetycznymi oraz znane jest tempo substytucji, można oszacować czas dywergencji tych linii. Wiedza zastosowana w praktyce pozwoliła na konstruowanie chronogramów, czyli drzew filogenetycznych, w których długość gałęzi jest wyskalowana za pomocą czasu (HO i LO 2013, HO 2014, HO i DUCHENE 2014).

Wczesne badania wykorzystujące koncepcję zegara molekularnego zakładały jego uniwersalność, a więc stałość tempa ewolucji poszczególnych genów u bardzo różnych organizmów. Jednak dość szybko wykazano, że tempo i dynamika ewolucji molekularnej różni się nawet między blisko spokrewnionymi liniami taksonomicznymi (GILLESPIE 1984, WELCH i BROMHAM 2005).

#### UNIWERSALNOŚĆ ZEGARA MOLEKULARNEGO?

Teoria Mutacji Neutralnych podkreśla znaczenie dryfu genetycznego, czyli przypad-

kowych wahań częstości występowania genów w małych populacjach, wynikających z losowego charakteru przekazywania genów z pokolenia na pokolenie. TMN przewiduje, że tempo ewolucji zależy od tempa powstawania spontanicznych mutacji, niezależnie od wielkości populacji, a bezwzględne tempo ewolucji (w odniesieniu do czasu) jest negatywnie skorelowane z długością cyklu życiowego. Ponieważ większość dziedzicznych mutacji zachodzi podczas replikacji DNA linii zarodkowej, gatunki o długim cyklu życiowym mają tendencję do gromadzenia mniejszej liczby mutacji w przeliczeniu na jednostkę czasu (HO 2014). Badania empiryczne wskazują, że tempo ewolucji białek jest prawdopodobnie niezależne od długości cyklu życiowego, ale zależy od tego, czy powstająca mutacja jest synonimiczna czy niesynonimiczna. Substytucje skutkujące zamianą aminokwasów podlegają ostrzejszej selekcji niż mutacje synonimiczne, które nie wpływają na dostosowanie i kumulują się w ciągu życia (HO i LO 2013). W latach 90. Tomoko Ohta potwierdziła, że długość cyklu życiowego silniej wiąże się z utrwaleniem mutacji synonimicznych niż niesynonimicznych. Założenie to jest podstawą „Nearly Neutral Theory”, modyfikacja TMN, która wskazuje na wielkość populacji jako na ważny czynnik warunkujący tempo ewolucji, zgodnie z zasadą, że selekcja silniej wpływa na duże populacje, natomiast w małych populacjach zaznaczony jest efekt działania dryfu genetycznego (OHTA 1995).

Od czasu sformułowania TMN liczne badania wykazały istnienie wielu różnych czynników wpływających na tempo ewolucji poszczególnych taksonów i zaprzeczających uniwersalności zegara molekularnego. Efekty ich działania można generalnie zaklasyfikować do trzech grup: efekty genozależne, efekty taksonozależne i efekty zależne od pozostałych czynników.

Efekty genozależne wynikają z różnej szybkości ewolucji poszczególnych fragmentów tego samego genomu. Zwykle różnorodność tego tempa wiąże się z tym, jak bardzo kluczowa dla funkcjonowania organizmu jest dana sekwencja DNA. Na przykład, fragmenty kodujące charakteryzują się niższym tempem zmian niż niekodujące, a eksony niższym niż introny. Tempo mutacji różni się nie tylko pomiędzy konkretnymi loci, ale także pomiędzy genomem jądrowym i organelowym (BULMER i współaut. 1991, HO i DUCHENE 2014). Kluczową rolę odgrywają tu m.in. reaktywne formy tlenu, powstające jako produkt uboczny reakcji zachodzących w mitochondriach. Ich produkcja jest pozytywnie skorelowana z wiekiem i tempem metabolizmu. Reaktywne formy tlenu uszkadza-

ją DNA, a więc przyczyniają się do zwiększenia tempa mutacji. Kolejną przyczyną jest różna wydajność systemów naprawczych genomu mitochondrialnego (mtDNA) i jądrowego (nDNA). W mtDNA systemy naprawy są znacznie mniej wydajne, przez co mutacje są częstsze, co pociąga za sobą szybsze tempo ewolucji w stosunku do genomu jądrowego. Na przykład u ssaków tempo ewolucji mtDNA jest nawet pięciokrotnie wyższe niż nDNA (BALLARD i WHITLOCK 2004). Inaczej jest u roślin, które mają zupełnie odmienną strategię naprawy mtDNA, przez co ewoluuje ono co najmniej kilka razy wolniej niż genom jądrowy (GALTIER 2011). Tempo ewolucji molekularnej DNA chloroplastowego (chlDNA) jest nieco wyższe niż w mtDNA,

jednak wciąż o połowę niższe w porównaniu z nDNA. Ponadto, w chloroplastowym DNA istnieją fragmenty niekodujące, które ewoluują znacznie wolniej niż to się zwykle zakłada (WOLFE i współaut. 1987).

Wiedzę o zmienności tempa ewolucji różnych regionów genomu naukowcy wykorzystali w praktyce, stosując je jako markery molekularne służące różnorodnym celom. W badaniu określonego poziomu zmienności genetycznej należy zastosować taki fragment DNA, którego tempo ewolucji odzwierciedli analizowane relacje. Ogólnie rzecz ujmując, markery molekularne ewoluujące wolniej należy wykorzystywać dla wyższych rang taksonomicznych, podczas gdy te ewoluujące najszybciej np. w badaniach na poziomie

Tabela 1. Porównanie właściwości przykładowych markerów molekularnych używanych w badaniach filogenetycznych z wykorzystaniem zegara molekularnego.

	Rodzaj markera	Ranga taksonomiczna	Przykładowe tempo substytucji
Markery mitochondrialne	12S rDNA	Wyższa, np. typy i podtypy.	U ssaków $3,4 \pm 1,2 \times 10^{-3}$ substytucji/pozycję nukleotydu/mln lat (PESOLE i współaut. 1999)
	16S rDNA	Niska i średnia, np. rodzaje czy rodziny.	U ssaków $4,9 \pm 1,4 \times 10^{-3}$ substytucji/pozycję nukleotydu/mln lat (PESOLE i współaut. 1999)
	Sekwencje kodujące białka	Niska i średnia, np. rodziny, rodzaje i gatunki.	COI u stawonogów $1,15 \times 10^{-2}$ substytucji/na pozycję nukleotydową/mln lat (BROWER 1994)
		Np. COI stosowany u wielu zwierząt jako marker gatunkowo specyficzny (DNA barcoding).	Cytochrom b u ryb $2,2-2,3 \times 10^{-2}$ substytucji/na pozycję nukleotydową/mln lat (ROCHA i współaut. 2011)
	Region kontrolny genomu mitochondrialnego	Niska, np. analizy wewnątrzgatunkowe lub blisko spokrewnionych gatunków.	U ludzi od 0,94 do 4,3 substytucji/pozycję nukleotydu/mln lat (PARSONS i współaut. 1997)
Markery jądrowe	18S rDNA	Wyższa, np. królestwa, typy, gromady, rzędy.	U promienic (Radiolaria) $3,3 \times 10^{-4}$ substytucji/pozycję nukleotydu/mln lat (GROUSSIN i współaut. 2011)
	28S rDNA	U niektórych taksonów (np. Gastrotricha i Radiolaria) stosowany jako marker gatunkowo specyficzny.	Średnia, np. rzędy, rodziny. U skorupiaków obunogich $3 \times 10^{-3}$ substytucji/pozycję nukleotydu/mln lat (MAMOS i współaut. 2016)
Markery chloroplastowe	rbcl	Średnia i wysoka.	U glików $5,0 \times 10^{-4}$ substytucji/pozycję nukleotydu/mln lat (VILLARREAL i RENNER 2012)
	matK	Różne pozycje taksonomiczne (głównie gatunki).	U sosnowatych $1,0-1,9 \times 10^{-4}$ substytucji/pozycję nukleotydu/mln lat (LOCKWOOD i współaut. 2013)

międzygatunkowym lub wewnątrzgatunkowym. Różnorodne markery DNA, zarówno mitochondrialne jak i jądrowe, są obecnie stosowane w analizach filogenetycznych (Tabela 1).

Efekty taksonozależne odnoszą się do czynników mających wpływ na cały genom. Każdy takson posiada własną, potencjalnie unikatową historię ewolucyjną i specyficzny profil fizjologiczny, które warunkują dynamikę ewolucji jego genomu. Uważa się, że podłożem tych efektów jest: czas trwania pokolenia, rozmiar ciała, tempo metabolizmu i różnice w efektywności mechanizmów naprawy DNA (HO 2014).

Badania wykazały, że rozmiar ciała jest ściśle skorelowany z tempem ewolucji molekularnej. Na przykład małe ssaki, takie jak gryzonie, charakteryzują się szybszym tempem ewolucji molekularnej niż ich więksi krewniacy. Jest to związane zarówno z tempem powstawania mutacji, jak i z tempem ich utrwalania (tempo substytucji). Mniejsze ssaki mają zwykle szybszą wymianę pokoleń. W krótkim cyklu życiowym dany osobnik wytwarza więcej komórek płciowych na rok, niż ten z dłuższym cyklem życiowym, a więc częściej kopiuje swój genom, gromadząc więcej mutacji związanych z błędami pojawiającymi się w czasie replikacji. To z kolei wpływa dodatnio na tempo ewolucji. Ponadto, małe ssaki charakteryzują się szybkim tempem metabolizmu, którego produkty uboczne są czynnikami mutagennymi. W związku z tym relacja pomiędzy tempem ewolucji a rozmiarem ciała nie jest wcale tak prosta jak się wydaje, ponieważ nie wiadomo, który z wymienionych czynników ma największe znaczenie (MARTIN i PALUMBI 1993, BROMHAM 2009).

Powszechność negatywnej zależności pomiędzy rozmiarem ciała a szybkością ewolucji, zweryfikowano w badaniach obejmujących 50 grup zwierząt (Metazoa), w oparciu o niezależne analizy zestawów genów mitochondrialnych i jądrowych. Należy tu zaznaczyć, że dla genów jądrowych nie zaobserwowano żadnej istotnej korelacji pomiędzy rozmiarem ciała a tempem ewolucji molekularnej (FONTANILLAS i współaut. 2007). THOMAS i współaut. (2010) wykazali, że jakkolwiek u bezkręgowców brak jest wspomnianej relacji między wielkością ciała a tempem ewolucji, to zależność pomiędzy długością trwania pokolenia a tempem ewolucji ma ten sam charakter u bezkręgowców co u kręgowców (szybsza wymiana pokoleń przekłada się na szybsze tempo ewolucji).

Sekwencje nukleotydowe są nieustająco narażone na uszkodzenia fizykochemiczne spowodowane m.in. promieniowaniem UV czy substancjami genotoksycznymi. Istnie-

ją różne mechanizmy naprawy DNA w komórkach. Jeden z nich polega na wycięciu uszkodzonego fragmentu sekwencji i zastąpieniu go nowym, przy czym może on być identyczny z wersją sprzed uszkodzenia. Oznacza to, że tempo mutacji jest wypadkową działania mutagenów i wydajności systemów naprawczych. Oba te czynniki w różny sposób działają na różne grupy organizmów w zależności od (i) środowiska życia (np. stopnia ekspozycji na promieniowanie UV), (ii) tempa metabolizmu (wyższe tempo metabolizmu generuje szybsze tempo mutacji) oraz (iii) intensywności replikacji DNA komórek zarodkowych (BROMHAM 2009).

Z korelacją pomiędzy tempem metabolizmu a tempem ewolucji związana jest teoria PUURTINENA i współaut. (2016) próbująca wyjaśnić, dlaczego ogólna liczba gatunków jest wyższa w pobliżu równika niż w strefach klimatów zwrotnikowych i podbiegunowych. Teoria ta podkreśla, że u większości gatunków stabilność białek i innych cząsteczek katalitycznych zależy od temperatury środowiska i zdolności termoregulacyjnych organizmu. Ze względu na naturę oddziaływań fizyko-chemicznych pomiędzy biocząsteczkami a wodą, struktury i funkcje tych cząsteczek są najbardziej stabilne w temperaturze około 20°C. U gatunków przystosowanych do niskich temperatur wyewoluowały swoiste, lecz niestabilne, enzymy zdolne katalizować reakcje w takim środowisku, a materiał genetyczny organizmów żyjących w chłodnych strefach klimatycznych (podbiegunowych) jest bardziej podatny na mutacje. Podobnie, ekstremalnie wysokie temperatury i duże amplitudy dobowe w strefach klimatów zwrotnikowych przyczyniają się do zwiększenia podatności na mutacje, ponieważ struktury biologiczne mogą nie być dość odporne na denaturację. Dlatego u organizmów żyjących w warunkach ekstremalnych pojawia się wiele mutacji letalnych, a silnie działająca presja selekcyjna skutkuje spowolnieniem ewolucji molekularnej. Taka argumentacja może wyjaśniać dlaczego ewolucja molekularna i tempo specjacji są szybsze w strefie równikowej o klimacie ciepłym, ale stabilnym (PUURTINEN i współaut. 2016). Może to po części tłumaczyć większą różnorodność biologiczną w lasach tropikalnych niż na obszarach narażonych na większe wahania temperatur, takich jak np. sawanny czy lasy klimatów umiarkowanych.

Pozostałe czynniki. Do tej grupy należą czynniki niejednorodnie wpływające na genom i będące wynikiem wzajemnego oddziaływania efektów genozależnych i taksonozależnych, m.in. wielkość populacji czy dobór naturalny (HO 2014).



Efektywna wielkość populacji to liczebność idealnej populacji, w której wszystkie osobniki przystępują do rozrodu i przekazują geny do następnego pokolenia. Jest ona jednym z najważniejszych czynników determinujących tempo substytucji. Przeprowadzono wiele badań w celu zrozumienia na czym dokładnie polega związek efektywnej wielkości populacji z tempem ewolucji molekularnej. Dobór naturalny i dryf genetyczny to dwie siły, których równowaga warunkuje liczbę i rodzaje mutacji ulegających utrwaleniu w populacji. Skuteczność doboru naturalnego może różnić się pomiędzy poszczególnymi fragmentami genomu i pomiędzy gatunkami. Rozbieżność działania selekcji na różne regiony genomu, wynika m.in. z różnego tempa rekombinacji; np. nieznacznie szkodliwe mutacje utrwalają się częściej w regionach o niskim tempie rekombinacji. Ogólna reguła jest taka, że na małe populacje znacznie mocniej oddziałuje dryf genetyczny niż dobór naturalny. Jeśli znaczna część mutacji jest łagodnie lub umiarkowanie szkodliwa, to gatunki tworzące małe populacje będą miały wyższe tempo substytucji niż te tworzące duże populacje. Dzieje się tak dlatego, że w mniejszych populacjach istnieje wyższe prawdopodobieństwo utrwalenia się tego rodzaju mutacji przez dryf genetyczny niż ich usunięcie wskutek selekcji. Należy pamiętać, że ze względu na niestabilność wielkości populacji, wartości tempa substytucji mogą znacząco różnić się pomiędzy nawet bardzo blisko spokrewnionymi gatunkami (BROMHAM 2009, WOOLFIT 2009).

Podsumowując, oryginalna koncepcja uniwersalnego zegara molekularnego nie wytrzymała próby czasu. Obecnie wiadomo, że istnieje szereg czynników wpływających na tempo ewolucji molekularnej, chociaż ich natura i wzajemne interakcje są wciąż słabo poznane. Nasuwa się więc pytanie o sens stosowania koncepcji zegara molekularnego w badaniach pokrewieństw między organizmami. Specjaliści starają się oszacować niepewności związane ze wspomnianymi wyżej czynnikami oraz akceptują fakt, że stosowanie tej koncepcji opiera się na subiektywnych założeniach opartych na aktualnej wiedzy i zgromadzonych doświadczeniach (ang. educated guesses). Przyjmują, że nawet niedoskonały zegar, który się spieszy lub spóźnia, jest lepszy niż zegar, którego w ogóle nie da się zsynchronizować z czasem rzeczywistym (GIBBS 2016). Postępowanie według odpowiedniego schematu oraz świadomość niedoskonałości metodologicznych pozwala przynajmniej częściowo doprecyzować oszacowanie czasu dywergencji linii filogenetycznych.

## ZEGARY MOLEKULARNE W PRAKTYCE

Pomijając obszerną problematykę związaną z samą rekonstrukcją filogenezy i wnioskowaniem filogenetycznym, omówioną przystępnie w wielu publikacjach (np. AVISE 2008, HALL 2008, SPALIK i PIWCZYNSKI 2009), praktyczne zastosowanie zegara molekularnego wiąże się z kilkoma ogólnymi kwestiami metodologicznymi: (i) wyborem markerów molekularnych, (ii) wyborem modelu substytucji, (iii) wyborem modelu zegara molekularnego, (iv) wyborem odpowiedniej metody kalibracji zegara molekularnego, (v) określeniem liczby, położenia i wieku punktów kalibracyjnych.

### WYBÓR MARKERÓW MOLEKULARNYCH

W pierwszym etapie najważniejsze jest, aby dobrać odpowiedni marker molekularny do badanej grupy organizmów i poziomu taksonomicznego warunkującego skalę czasową filogenezy. W badaniach różnych grup organizmów stosuje się różne grupy markerów. Na przykład, markery plastydowe i jądrowe wykorzystuje się w badaniach roślin i glonów, a markery mitochondrialne i jądrowe do badań zwierząt. Różnice te wynikają z obecności/braku określonych markerów (np. chlDNA) u konkretnej grupy organizmów, a po części z odmiennego schematu ewolucji tych samych fragmentów genomu u różnych taksonów (np. znacznie wolniejsza ewolucja genomu mitochondrialnego u roślin niż u zwierząt). Użycie konkretnego markera jest często warunkowane dostępnością specyficznych starterów oligonukleotydowych, bibliotek sekwencji porównawczych lub znajomości jego tempa ewolucji. Na przykład, spośród markerów mitochondrialnych wykorzystywanych w filogeografii stawonogów najczęściej używa się fragmentu genu kodującego podjednostkę I oksydazy cytochromowej (COI), a w przypadku kręgowców, np. ryb i płazów, fragmentu genu kodującego cytochrom *b* (Cyt *b*). Ze względu na odmiennie tempo ewolucji różnych genów, nie należy stosować tego samego markera w rekonstrukcji filogenezy na poziomie gatunku oraz gromady lub typu (patrz Tabela 1).

### WYBÓR MODELU SUBSTYTUCJI

Najprostszą miarą dystansu genetycznego pomiędzy dwoma sekwencjami DNA jest liczba miejsc, w których sekwencje te różnią się między sobą. Liczba tych różnic może być podana w proporcji do długości sekwencji i ten rodzaj dystansu oznaczany jest często jako *p*. Z TMN wynika, że *p* może być użyte jako prosta miara odległości ewolucyjnej pomiędzy sekwencjami. Jednak po upływie odpowiednio długiego czasu, nu-

kleotydy zajmujący konkretną pozycję może ulec wielokrotnym podstawieniom, czyli wysyceniu mutacjami (ang. saturation), co powoduje zagłuszenie autentycznego sygnału filogenetycznego. Ponieważ kod genetyczny składa się zaledwie z czterech „liter” (A, T, G, C), podstawienia te doprowadzą w końcu do stanu pierwotnego (rewersja mutacji) lub do utrwalenia się w tej samej pozycji, w dwóch sekwencjach, tego samego nukleotydu, lecz w wyniku odmiennej mutacji (paralelizm). Wskutek tego część mutacji będzie niezauważona, a tempo substytucji, i tym samym dystans genetyczny, pozostaną niedoszacowane. Aby zniwelować ten efekt, dystanse genetyczne są poddawane różnego rodzaju korekcjom, które wymagają dopasowania odpowiedniego modelu ewolucji molekularnej do analizowanego zestawu danych. Modele te różnią się względnym prawdopodobieństwem wystąpienia określonego typu mutacji np. większą częstością wystąpienia tranzykcji niż transwersji oraz proporcją pozycji nukleotydowych, które nie podlegają mutacjom. Dobór niewłaściwego modelu ewolucji może spowodować błędne oszacowanie długości gałęzi i wsparcia topologii drzewa filogenetycznego (prawdopodobieństwa wystąpienia określonego układu węzłów i gałęzi). Dopasowanie modelu substytucji do zestawu danych wykorzystuje najczęściej metodę największej wiarygodności (ang. maximum likelihood) lub inferencji bayesowskiej (ang. Bayesian inference) i może być wykonane w programach, takich jak jModelTest, MEGA czy BEAST. Metoda największej wiarygodności polega na oszacowaniu prawdopodobieństwa obserwowanych danych, kiedy parametry modelu są znane. Natomiast inferencja bayesowska to metoda wnioskowania statystycznego, w której korzysta się z twierdzenia Bayesa do aktualizowania prawdopodobieństwa zdarzeń założonego *a priori* przez dane empiryczne uzyskane *a posteriori*. Założenia teoretyczne i proces wyboru właściwego modelu są bardzo szerokimi zagadnieniami. Bardziej szczegółowe podejście do tematu czytelnik znajdzie m.in. w pracach SULLIVANA i JOYCEA (2005), POSADY (2009) oraz DRUMMONDA i BOUCKAERTA (2015).

#### WYBÓR MODELU ZEGARA MOLEKULARNEGO

##### Modele zegara molekularnego

Tempo ewolucji jest zwykle wyrażone jako liczba substytucji/pozycja nukleotydu/jednostka czasu. Jednostka czasu zależy od skali czasowej analizy; w badaniach dotyczących głębszych poziomów filogenezy będzie to milion lat, podczas gdy np. w epidemiologii molekularnej jest to najczęściej rok.

Najprostszy model zegara molekularnego zakłada jednorodność tempa ewolucji w obrębie wszystkich gałęzi drzewa filogenetycznego. Zegar o stałym tempie „tykania” (ang. strict molecular clock) jest wiarygodny tylko w przypadku jednorodnego zestawu danych, to jest sekwencji reprezentujących organizmy należące do jednego lub grupy blisko spokrewnionych gatunków. W przypadku organizmów o znacząco odmiennej historii ewolucyjnej, efekt taksonu (np. różna długość życia, tempo metabolizmu) różnicuje tempo ewolucji poszczególnych gałęzi na tyle, że odbiega ono znacząco od modelu „strict clock”. Opracowano szereg alternatywnych modeli zegara molekularnego zakładających, że jakkolwiek tempo ewolucji jest stałe dla poszczególnych gałęzi, to może różnić się między nimi. Istnieje wiele metod obliczania wariacji tempa substytucji pomiędzy gałęziami drzewa filogenetycznego. Można je podzielić na trzy główne grupy: metody oparte na dystansie genetycznym (ang. distance based), metody parametryczne (np. maximum likelihood, Bayesian inference) i metody nieparametryczne (np. rate smoothing). Obecnie za najbardziej godne zaufania, zarówno podczas obliczania wariacji tempa substytucji, jak i rekonstrukcji topologii drzewa filogenetycznego, zostały uznane metody największej wiarygodności i inferencji bayesowskiej (BAUM i SMITH 2013). Główne modele zegara, opracowane celem rozwiązania kwestii zróżnicowania tempa ewolucji w obrębie tego samego drzewa filogenetycznego, zostały przedstawione w Tabeli 2 (WELCH i BROMHAM 2005, HO 2009, HO i DUCHENE 2014). Ogólnie, popularnością cieszą się zegary o tzw. rozluźnionym, czyli zmiennym tempie „tykania” dla poszczególnych gałęzi (ang. relaxed clocks), a najlepszym rozwiązaniem jest jednocześnie ustalenie topologii drzewa i oszacowanie czasu rozejścia się poszczególnych linii filogenetycznych. Umożliwia to kilka programów implementujących wspomniane już podejście bayesowskie (Tabela 3, 4) (DRUMMOND i współaut. 2006).

##### Selekcja modelu zegara molekularnego

Wybór odpowiedniego modelu zegara molekularnego wymaga odnalezienia równowagi pomiędzy złożonością modelu a jego dopasowaniem do danych (SULLIVAN i JOYCE 2005). Warto zaznaczyć, że nie ma idealnego modelu w pełni odwzorowującego historię ewolucyjną badanej grupy organizmów, gdyż każdy obarczony jest błędem wynikającym ze zbyt dużej liczby zmiennych i zdarzeń pozostających poza naszą wiedzą. Dlatego selekcję modelu należy postrzegać jedynie jako mniej lub bardziej niedoskonałe uproszczenie rzeczywistości. Wieloparametrowe modele mogące wier-

Tabela 2. Cztery podstawowe metody kalibracji oraz ich charakterystyka (wg WILKE i współaut. 2009).

Metoda kalibracji	Najważniejsze zalety	Wady	Przykłady zastosowania
Znane tempo ewolucji (kalibracja wtórna lub pośrednia)	<ul style="list-style-type: none"> <li>– nie wymaga całego procesu wybierania miejsca i liczby punktów kalibracyjnych</li> </ul>	<ul style="list-style-type: none"> <li>– jeśli pierwotnie nieprawidłowo ją oszacowano, to będzie źródłem błędów dla kolejnych analiz</li> <li>– dla zestawów danych charakteryzujących się stałym tempem ewolucji molekularnej (ang. strict molecular clock)</li> <li>– relatywnie niewiele dostępnych, wiarygodnych zewnętrznych wartości tempa substytucji</li> </ul>	<ul style="list-style-type: none"> <li>– ptasi zegar molekularny (WEIR i SCHLUTER 2008)</li> <li>– trait-specific Protostomia COI clock (WILKE i współaut. 2009)</li> </ul>
Dane biogeograficzne	<ul style="list-style-type: none"> <li>– możliwe analizy porównawcze dla różnych taksonów</li> </ul>	<ul style="list-style-type: none"> <li>– często występują wątpliwości co do datowań wydarzeń biogeograficznych</li> <li>– potencjalnie nieprawidłowe połączenie wydarzeń w filogenezie z wydarzeniami biogeograficznymi</li> <li>– zwykle dostarczają tylko jeden lub kilka punktów kalibracyjnych</li> <li>– wydarzenia biogeograficzne dostarczają często tylko górną lub dolną granicę wieku węzła</li> </ul>	<ul style="list-style-type: none"> <li>– pojawienie się Przesmyku Panamskiego</li> <li>– Kryzys Messyński</li> <li>– formacja Rowu Transegejskiego (PAPADOPOULOU i współaut. 2010)</li> </ul>
Dane kopalne	<ul style="list-style-type: none"> <li>– w wielu przypadkach łatwo dostępne</li> </ul>	<ul style="list-style-type: none"> <li>– trudne do właściwego umieszczenia w filogenezie</li> <li>– dane kopalne nie mogą być traktowane jako bezpośredni przodekowie</li> <li>– skamieniałości mogą dostarczyć jedynie minimalną granicę dla wieku węzła</li> <li>– wątpliwości związane z datowaniem zapisów kopalnych</li> <li>– często trudne do zidentyfikowania i zaklasyfikowania</li> <li>– często poważne problemy z brakującymi taksonami</li> </ul>	<ul style="list-style-type: none"> <li>– użycie szczególnie bogatego, dobrze poznanego zapisu kopalnego grupy Nothofagus (SAUQUET i współaut. 2012)</li> </ul>
aDNA/RNA	<ul style="list-style-type: none"> <li>– pozwala na bezpośrednią kalibrację</li> </ul>	<ul style="list-style-type: none"> <li>– rzadko dostępne</li> <li>– trudności z umiejscowieniem w filogenezie</li> <li>– rzadko istnieje możliwość założenia, że antyczny DNA pochodzi od przodka</li> <li>– błędy spowodowane intruzjami w próbce</li> <li>– często problem z degradacją antycznego materiału genetycznego</li> </ul>	<ul style="list-style-type: none"> <li>– zestaw wydatowanych kości należących do hatterii (od 650 do 9000 lat) (HAY i współaut. 2008)</li> </ul>

Tabela 3. Charakterystyka i przykłady użycia poszczególnych modeli zegara molekularnego. Wyjaśnienie użytych symboli literowych:  $k$  – liczba (różnych) temp ewolucji,  $n$  – liczba gałęzi (wg HO i DUCHENE 2014).

Strict clock	
Stale tempo substytucji na wszystkich gałęziach drzewa.	– analiza sekwencji z osobników tego samego gatunku lub kiedy spodziewamy się niewielkiej różnorodności tempa pomiędzy gałęziami
Opisuje go tylko jeden parametr – tempo ewolucji.	– model zerowy w testowaniu dopasowania innych modeli zegara molekularnego
$k=1$	– przykłady: wiele wczesnych badań nad zegarem molekularnym (np. DOOLITTLE i współaut. 1996, DICKERSON 1971, RUNNEGAR 1982)
Multi-rate molecular clocks	
Ograniczona wariacja tempa ewolucji wzdłuż drzewa filogenetycznego.	
Liczba odrębnych temp substytucji jest większa niż 1, ale mniejsza niż liczba gałęzi.	
$1 < k < n$	
local multi-rate clocks	
Gałęzie, które współdziela to samo tempo są zgrupowane a każda grupa ma odrębne tempo substytucji.	– pozwala na definiowanie klastrow na podstawie charakterystycznych cech biologicznych które je odróżniają, np. różne taksony mają odrębne tempa ewolucji
	– przykłady zastosowania: HASEGAWA i współaut. 1989, YODER i YANG 2000
'discrete' multi-rate clocks	
Gałęzie, współdzielące tempo niekoniecznie sąsiadują ze sobą (mogą być w zupełnie innych miejscach drzewa).	– pozwala na definiowanie klastrow na podstawie charakterystycznych cech biologicznych które je odróżniają, np. różne taksony mają odrębne tempa ewolucji
	– przykłady zastosowania: HASEGAWA i współaut. 1989, YODER i YANG 2000
Relaxed molecular clock	
Zrelaksowane zegary molekularne pozwalają na to, aby każda gałąź na drzewie mogła mieć odrębne tempo substytucji	
Jedyny model, który zakłada, że liczba temp ewolucji może się równać liczbie gałęzi.	
$k=n$	
Relaxed molecular-clock z wykorzystaniem metody "rate-smoothing"	
Celem tych metod jest zminimalizowanie zmian tempa pomiędzy sąsiadującymi gałęziami w filogenezie.	– łatwo je przystosować do genomowych zbiorów danych.
Korzystają z innych metod w celu ustalenia topologii drzewa i długości gałęzi.	– metody te są stosunkowo szybkie nawet dla zbiorów danych z dużą liczbą taksonów
	– np. BRITTON i współaut. 2007, SMITH i O'MEARA 2012
Wymagają ustalonej topologii drzewa.	
Relaxed molecular-clock z wykorzystaniem metod baysowskich	
Bayesowskie zrelaksowane zegary molekularne oceniają tempo dla każdej gałęzi korzystając z już ustalonej topologii albo szacując filogenezę i wiek węzłów jednocześnie.	Bayesowskie zrelaksowane zegary molekularne można podzielić na dwie klasy:
	*Autoskorelowane – zakładają że sąsiadujące gałęzie mają podobne tempa ewolucji.
Pozwalają na odpowiednie wykorzystanie wiedzy na temat danych oraz użycie parametrycznie bogatych modeli ewolucji. Podejście takie, przy zastosowaniu odpowiedniej metody próbkowania (Łańcuchy Markowa) pozwala na uzyskanie bardzo wiarygodnych wyników.	*Nieskorelowane – brak założenia o korelacji tempa pomiędzy sąsiadującymi gałęziami



Tabela 4. Wybrane pakiety oprogramowania do rekonstrukcji chronogramów przy użyciu różnych modeli zegarów molekularnych opisanych w Tabeli 3 (wg HO i DUCHENE 2014, TAMURA i współaut. 2012).

Oprogramowanie	Model zegara molekularnego	Podejście statystyczne	Topologia drzewa	Referencje
DAMBE	Strict clock, discrete multi-rate clock, autocorrelated relaxed clock	Metody oparte na dystansie genetycznym	Ustalona wcześniejszej	XIA i YANG 2011
PATHD8	Autocorrelated relaxed clock	Metody oparte na dystansie genetycznym	Ustalona wcześniejszej	BRITTON i współaut. 2007
R8S	Strict clock, local multi rate clock, discrete multi-rate clock, autocorrelated relaxed clock	Metody nieparametryczne – rate smothering, największej wiarygodności	Ustalona wcześniejszej	SANDERSON 2003
PAML-BASEML	Strict clock, local multi rate clock, discrete multi-rate clock	Metody największej wiarygodności	Ustalona wcześniejszej	YANG 2007
TREEPL	Strict clock, autocorrelated relaxed clock	Metody największej wiarygodności	Ustalona wcześniejszej	SMITH i O'MEARA 2012
BEAST	Strict clock, local multi rate clock, autocorrelated relaxed clock, uncorrelated relaxed clock	Metody Bayesowskie	Nie ma konieczności ustalania topologii wcześniejszej	DRUMMOND i współaut. 2012, DRUMMOND i BOUCKAERT 2015
MRBAYES	Strict clock, autocorrelated relaxed clock, uncorrelated relaxed clock	Metody Bayesowskie	Nie ma konieczności ustalania topologii wcześniejszej	RONQUIST i współaut. 2012b
MULTIDIVTIME	Strict clock, autocorrelated relaxed clock	Metody Bayesowskie	Ustalona wcześniejszej	THORNE i współaut. 1998
PAML-MCMCTREE	Strict clock, autocorrelated relaxed clock, uncorrelated relaxed clock	Metody Bayesowskie	Ustalona wcześniejszej	YANG 2007
PHYLOBAYES	Strict clock, autocorrelated relaxed clock, uncorrelated relaxed clock	Metody Bayesowskie	Ustalona wcześniejszej	LARTILLOT i współaut. 2009
REVBAYES	Strict clock, local multi rate clock, discrete multi-rate clock, autocorrelated relaxed clock, uncorrelated relaxed clock	Metody Bayesowskie	Nie ma konieczności ustalania topologii wcześniejszej	HÖHNA i współaut. 2016
MEGA	Relaxed clock - RelTime	Metody największej wiarygodności	Ustalona wcześniejszej	TAMURA i współaut. 2012

nie odtworzyć przebieg ewolucji są bardziej atrakcyjne od modeli prostych. Posiadają one jednak wady, które czynią analizę trudną obliczeniowo i bardziej czasochłonna, a ponadto, ze względu na użycie tych samych danych do oszacowania wielu parametrów, zwiększają prawdopodobieństwo pojawienia się błędów. Dlatego selekcja modelu zegara molekularnego musi być oparta na tzw. brzytwie Ockhama – zasadzie ekonomii myślenia, według której do wyjaśniania zjawisk należy użyć możliwie najmniejszej liczby założeń i pojęć (POSADA i BUCKLEY 2004).

Na początku każdej analizy powinno się przetestować, czy wariancja tempa substytucji pomiędzy analizowanymi liniami taksonomicznymi spełnia założenia zegara molekularnego typu „strict clock”, a tym samym uzasadnia zastosowanie bardziej złożonego modelu. Jeżeli hipoteza o „strict clock” nie zostanie odrzucona, preferowane jest zastosowanie zegara o stałym tempie ewolucji, czyli modelu z mniejszą liczbą parametrów, prowadzącego do możliwie najprecyzyjniejszych oszacowań czasów dywergencji (HO i DUCHENE 2014).

Pierwsze testy, tzw. „relative rate tests”, polegały na porównaniu, czy różnica między tempem ewolucji jednego kladu (monofiletycznej grupy sekwencji) a kladem referencyjnym [grupą zewnętrzną (ang. outgroup) – sekwencją lub grupą sekwencji, która jest równie odległa od wszystkich sekwencji w analizowanym zestawie danych] jest statystycznie różna od różnicy pomiędzy tempem ewolucji drugiego kladu a tą samą grupą referencyjną (SCHWARTZ i MUELLER 2010). Ich zaletą był brak konieczności znajomości topologii drzewa filogenetycznego, a jedynie wymóg, aby dwa brane pod uwagę klady były ze sobą bliżej spokrewnione niż z grupą zewnętrzną. Zaawansowane metody, oparte na topologii całych drzewach filogenetycznych, są zazwyczaj znacznie bardziej użyteczne, ponieważ pozwalają na przetestowanie większej liczby modeli zakładających różnorodność tempa pomiędzy gałęziami, a nie tylko zgodności z modelem „strict clock”. Takimi metodami są np. Likelihood Ratio Test (LRT), Akaike Information Criteria through Markov Chain Monte Carlo (AICM), Path Sampling (PS) czy Stepping-Stone Sampling (SSS). Służą one do wyboru najprostszego modelu spośród dostępnych, który jednocześnie jest najlepiej dopasowany do zrekonstruowanej topologii drzewa filogenetycznego (SULLIVAN i JOYCE 2005, SCHWARTZ i MUELLER 2010, HO i DUCHENE 2014). Na przykład, selekcja modelu zegara za pomocą najbardziej zaawansowanych metod takich jak AICM, PS czy SSS może być przeprowadzona z wykorzystaniem pakietów BEAST, MrBayes, RevBayes oraz programu Tracer (RAMBAUT i współaut. 2018) (Tabela 4). Wszystkie te pakiety mają rozbudowane instrukcje obsługi oraz aktywne fora internetowe (<https://groups.google.com/forum/#!forum/beat-users> oraz [.../revbayes-users](https://groups.google.com/forum/#!forum/revbayes-users)), gdzie użytkownik może uzyskać szczegółowe informacje na temat konkretnych analiz.

#### KALIBRACJA ZEGARA MOLEKULARNEGO

##### Metody kalibracji

Dystans genetyczny pomiędzy sekwencjami w zestawie danych jest pochodną ich tempa ewolucji i czasu jaki upłynął od ostatniego wspólnego przodka. By precyzyjnie określić czas rozejścia się poszczególnych gałęzi drzewa filogenetycznego i właściwie oszacować tempo ewolucji molekularnej, konieczna jest prawidłowa kalibracja zegara. Polega ona na oszacowaniu liczby mutacji, które utrwały się wzdłuż poszczególnych linii filogenetycznych od momentu ich dywergencji, w przeliczeniu na jednostkę czasu (np. milion lat), używając przy tym niezależnych danych źródłowych. Wybór

odpowiedniej kalibracji jest krytycznym etapem w badaniach filogenetycznych, ale niestety przez wiele lat był zaniebdywany, a jego znaczenie trywializowane (LEE 1999, FOREST 2009, DUCHENE i współaut. 2014). Zegar molekularny można skalibrować: (i) pośrednio, używając znanej już wartości tempa substytucji dla danego markera lub (ii) bezpośrednio, poprzez określenie wieku przynajmniej jednego węzła drzewa filogenetycznego w oparciu o niezależne dane biogeograficzne, kopalne lub aDNA (ang. ancient DNA) (HO 2014). Warto pamiętać, że dane kopalne czy biogeograficzne nie pozwalają na konkretne, precyzyjne określenie wieku węzła. Umożliwiają one jedynie wyznaczenie minimalnych i/lub maksymalnych granic wieku węzła (ang. node age constraint), które mogą być zastosowane do kalibracji zegara molekularnego (WARNOCK i współaut. 2014).

##### Kalibracja na podstawie znanego tempa substytucji

Metoda ta zakłada stałość tempa ewolucji danego fragmentu DNA pomiędzy badanymi liniami filogenetycznymi (model „strict clock”). Przykładem szeroko stosowanym w kalibracji zegara molekularnego może być tempo ewolucji fragmentu genomu mitochondrialnego (COI, tRNA Leu, COII) stawonogów równe  $1,15 \times 10^{-8}$  substytucji/na pozycję nukleotydową/na milion lat (BROWER 1994). Do 2017 r. praca ta była cytowana ponad 1000 razy, z czego 60 razy tylko w 2016 r., co świadczy, że wartość ta stała się swego rodzaju standardem w badaniach genetycznych tej grupy zwierząt (Web of Science, dostęp 12.04.2017). W rzeczywistości założenie o stałości tempa ewolucji w obrębie tak różnorodnej grupy organizmów jak Arthropoda jest błędne. Kalibracja zegara za pomocą znanego stałego tempa substytucji powinna być stosowana jedynie w badaniach relacji filogenetycznych blisko spokrewnionych gatunków oraz markerów o dobrze poznanej, stabilnej dynamice ewolucji (np. geny kodujące białka) (HO i LO 2013). W tym przypadku model „strict clock” pozwala uzyskać bardziej precyzyjne szacowanie czasów dywergencji niż pozostałe modele.

##### Kalibracja na podstawie wydarzeń biogeograficznych

Przyczyną różnicowania się gatunków czy populacji jest najczęściej powstanie barier geograficznych (np. ruchy górotwórcze i tektoniczne, zlodowacenia, zmiany układu sieci hydrograficznej) prowadzące do specjacji. Jeśli tego typu wydarzenie jest datowane (np. radiometrycznie), to można je przypisać do konkretnego epizodu dywergencji badanych organizmów reprezentowanego przez węzeł

na drzewie filogenetycznym. Ten typ kalibracji jest narażony na mniejsze nieścisłości niż zapis kopalny, jednak należy stosować go z rozważą, szczególnie jeśli chcemy użyć zjawisk zachodzących w sposób ciągły, przez dziesiątki milionów lat. Na przykład, trudno jest wykorzystać pęknięcie płyty tektonicznej czy orogenezy do ścisłego określenia wieku danego węzła.

Ponadto, w przypadku rozejścia się dwóch mas lądowych, jeszcze przez miliony lat może zachodzić przepływ genów między populacjami zamieszkujących je organizmów, co jest uzależnione od możliwości, sposobu i tempa ich dyspersji. Kalibracja na podstawie wydarzeń biogeograficznych jest szczególnie przydatna, kiedy dane kopalne są bardzo skąpe, co ma miejsce u wielu grup bezkręgowców (WARNOCK i współaut. 2014).

Przykładem zastosowania kalibracji biogeograficznej jest pojawienie się Przesmyku Panamskiego, które może określić minimalną granicę wieku ostatniego wspólnego przodka organizmów morskich zamieszkujących Morze Karaibskie i Ocean Spokojny. Przesmyk ten stanowi równocześnie pomost lądowy określający maksymalną granicę wieku blisko spokrewnionych organizmów lądowych zamieszkujących Amerykę Północną i Południową (FOREST 2009, HO i współaut. 2011).

#### Kalibracja na podstawie danych kopalnych

Jedną z najpowszechniejszych metod kalibracji szczególnie w badaniach dotyczących kręgowców jest użycie danych kopalnych. Skamieniałość dostarcza informację o najwcześniejszym pojawieniu się możliwego do zidentyfikowania przedstawiciela danej linii rodowej. Tym samym, pozwala ona na sprecyzowanie dolnej (minimalnej) granicy wieku dywergencji, która dała początek tej linii. Należy jednak pamiętać, że pojawienie się taksonu w zapisie kopalnych reprezentuje raczej czas jego licznego występowania niż powstania. Przyczyną, dla której zapis kopalny zwykle nie może nam wskazywać górnej (maksymalnej) granicy wieku węzła jest fakt, że istnieje prawdopodobieństwo odnalezienia starszych skamieniałości, które przesuną wstecz czas pojawienia się linii taksonomicznej (DONOGHUE i BENTON 2007, FOREST 2009).

Kalibracja na podstawie danych kopalnych może być związana z szeregiem błędów, które możemy podzielić na kilka grup: (i) błędy w topologii drzewa filogenetycznego, (ii) błędy wynikające ze stopnia zachowania danych kopalnych (ich niekompletność lub fragmentaryczność), (iii) błędna i/lub błędnie zastosowana metodologia użyta do identyfikacji skamieniałości i jej przynależności taksonomicznej, (iv) błędy związane

z samym procesem pozyskiwania prób, (v) błędnie oceniony wiek skamieniałości, jako efekt bezpośredniego utożsamienia go z wiekiem formacji skalnej, w której została znaleziona (DONOGHUE i BENTON 2007). W celu poprawienia wiarygodności tego rodzaju kalibracji w 2015 r. stworzono międzynarodową internetową bazę danych, Fossil Calibration Database, udostępniającą informacje o ponad 150 zweryfikowanych punktów kalibracyjnych dla różnych grup Eukaryota (KSEPKA i współaut. 2015).

Doskonałym przykładem użycia szeregu skamieniałości do kalibracji zegara molekularnego na różnych poziomach filogenezy są badania nad radiacją pingwinów, które wykazały, że linie współcześnie żyjących gatunków różnicowały się w miocenie i pliocenie, znacznie później niż zakładano do tej pory. Badania te wykazały dodatkowo, że użycie większej liczby głęboko położonych fosylnych punktów kalibracyjnych wpłynęło pozytywnie na precyzję i wiarygodność oszacowania czasów dywergencji współczesnych gatunków (GAVRYUSHKINA i współaut. 2017).

#### Kalibracja w oparciu o kopalne antyczne DNA (aDNA)

Jeżeli mamy dostęp do kopalnego DNA i znamy jego wiek, to możemy je użyć do bezpośredniej kalibracji zegara molekularnego. Jest to możliwe w przypadku dobrze zakonserwowanych tkanek lub nawet szybko ewoluujących wirusów, których próbki pobierane były w seriach czasowych. Kalibrowanie zegara molekularnego za pomocą aDNA może być bardzo efektywne, ponieważ zwykle nie ma wątpliwości co do jego wieku i stosunkowo łatwo można przypisać ten wiek do konkretnego węzła w filogenezie (HO 2014).

Mimo niewątpliwych zalet, istnieje kilka poważnych problemów związanych z tym rodzajem kalibracji. Antyczne DNA jest zwykle trudno dostępne i relatywnie zbyt młode, aby nadawało się do szacowania wieku węzłów na głębszych poziomach filogenezy. Jednak największym wyzwaniem tego rodzaju kalibracji jest pośmiertna degradacja kwasów nukleinowych. Po śmierci organizmu, cząsteczki DNA ulegają fragmentacji oksydacyjnej, hydrolizie i wiążą się krzyżowo, co ogranicza wydajność reakcji PCR. Skutkuje to przede wszystkim częstymi błędami podczas wstawiania zasad przez polimerazę, krótszymi produktami amplifikacji lub ich brakiem (HO i współaut. 2007, WILKE i współaut. 2009).

Przykładem udanego zastosowania aDNA w celu ustalenia tempa substytucji jest praca dotycząca dynamiki ewolucji hatterii, uznawanej za żywą skamieniałość i relik



ery mezozoicznej. Hatteria charakteryzuje się bardzo wolnym metabolizmem i wzrostem, długim cyklem życiowym i wolnym tempem reprodukcji, sugerując niskie tempo ewolucji molekularnej. Porównano fragment sekwencji mitochondrialnego regionu kontrolnego z dobrze zakonserwowanych subfosylnych próbek tkanki kostnej o znanym wieku oraz z tkanek współcześnie żyjących hatterii. Wbrew założeniom okazało się, że tempo substytucji w badanym fragmencie jest jednym z najwyższych zarejestrowanych wśród kręgowców. Brak więc w tym przypadku potwierdzenia dla korelacji pomiędzy tempem substytucji a takimi czynnikami jak tempo metabolizmu czy długość trwania pokolenia, opisanych w poprzednim rozdziale niniejszego artykułu. To pokazuje, że nie należy generalizować i zakładać, że jeśli u wielu grup organizmów jakiś czynnik ma wpływ na heterogeniczność tempa ewolucji, to można tę zależność przełożyć na wszystkie organizmy (HAY i współaut. 2008).

Podsumowanie opisanych powyżej metod kalibracji znajduje się w Tabeli 2.

#### OKREŚLENIE LICZBY, POŁOŻENIA I WIEKU PUNKTÓW KALIBRACYJNYCH

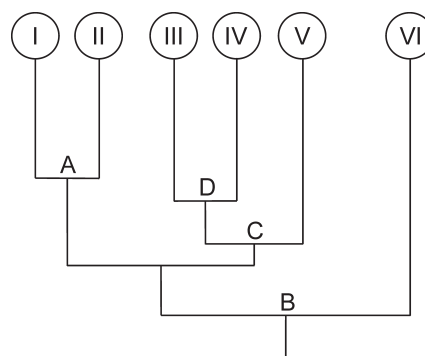
##### Miejsce i liczba punktów kalibracyjnych

Punkt kalibracyjny to węzeł o znanym czasie dywergencji linii filogenetycznych, powiązany z konkretnym wydarzeniem z przeszłości datowanym za pomocą niezależnych metod np. stratygraficznych czy radiometrycznych. Takim punktem może być na przykład początek izolacji linii ewolucyjnych żab zielonych (*Pelophylax*) zamieszkujących Cypr i zachodnią Anatolię, na skutek wypełnienia basenu Morza Śródziemnego przez wody oceaniczne po Kryzysie Messyńskim, ok. 5,5–5,3 miliona lat temu (AKIN i współaut. 2010). Innym przykładem może być powstanie starożytnego Jeziora Ochrydzkiego ok. 2–3 miliony lat temu, zapoczątkowujące radiację zamieszkujących je skorupiaków obunogich (WYSOCKA i współaut. 2013).

Umieszczenie i liczba kalibrowanych węzłów powinny być dobrze przemyślane. Nawet jeśli mamy prawidłowo przyrównany zestaw danych, wiarygodną kalibrację i właściwie dobrany model zegara molekularnego, to niewłaściwa pozycja i niedostateczna liczba punktów kalibracyjnych może przesądzić o prawidłowości wyniku analizy. Wielu badaczy wskazuje, że dla wiarygodności analizy, nawet przy homogeniczności tempa ewolucji, warto użyć więcej niż jednego punktu kalibracyjnego. Z drugiej strony, niedobór sprawdzonych punktów kalibracyjnych dla wielu grup taksonomicznych wiąże się z trudnym wyborem, czy użyć jednego,

ale pewnego, czy kilku, ale mniej wiarygodnych punktów. Wybór nie jest prosty. Zastosowanie jednego punktu kalibracyjnego skutkuje większą liczbą błędów i mniejszą precyzją niż zastosowanie kilku kalibracji, zaś jeśli chodzi o pozycję punktu kalibracyjnego, to lepsze wyniki daje umieszczenie go na głębszych węzłach drzewa filogenetycznego. Wpływ położenia punktu kalibracyjnego na datowanie filogenezy wyjaśniamy w oparciu o Ryc. 1. Użycie tylko jednego węzła, np. A jako punktu kalibracyjnego przy założeniu stałego tempa ewolucji na wszystkich gałęziach drzewa, nie jest problematyczne, jednak stałość tempa, to sytuacja rzadka. Przy heterogeniczności tempa ewolucji, może być ono znacznie wyższe dla linii filogenetycznych I i II, pochodzących od przodka A, niż u pozostałych (III–VI). Wówczas, opierając się tylko o tę kalibrację otrzymamy fałszywą, przeszacowaną ogólną wartość tempa substytucji. Dlatego znacznie bardziej wiarygodna będzie kalibracja węzła B, bliżej korzenia drzewa, co spowoduje mniejszą podatność szacowanego tempa substytucji na czynniki związane z efektem taksonu. Podobny efekt będzie miała kalibracja kilku punktów jednocześnie (np. węzły B, C i D), co pozwoli na zbalansowanie oceny tempa substytucji na drzewie (WILKE i współaut. 2009, DUCHENE i współaut. 2014).

Ponieważ szacowanie „tempa tykania” zegara molekularnego jest wrażliwe na pozycję punktu kalibracyjnego, trzeba mieć świadomość, że tempo ewolucji może być nieskorelowane lub autoskorelowane pomiędzy liniami taksonomicznymi mającymi bliskiego wspólnego przodka. Modele o nieskorelowanym tempie wyróżniają się tym, że tempo ewolucji dowolnie wybranej gałęzi



Ryc. 1. Schemat obrazujący przykładowe drzewo filogenetyczne.

Węzły, które można wykorzystać do kalibracji zegara molekularnego oznaczone są literami A–D, natomiast poszczególne linie taksonomiczne cyframi rzymskimi I–VI. Węzły płytkie to te położone najbliżej korony drzewa np. A, natomiast głębokie (węzły starsze) to np. B.



drzewa jest niezależne od innych. Modele, gdzie tempo jest skorelowane zakładają, że jest ono powiązane na sąsiadujących gałęziach. Użycie pojedynczej lub/i płytkiej kalibracji jest obarczone większym ryzykiem błędu w przypadku zegara molekularnego o tempie autoskorelowanym. Dzieje się tak, ponieważ para gałęzi składająca się z gałęzi ancestralnej i potomnej będzie miała zwykle bardziej podobne tempo substytucji niż para losowo wybranych gałęzi z dowolnych części drzewa. Stąd wiarygodność szacowania wieku węzła zależy od tego czy jest on zagnieżdżony w kładzie o korzeniu będącym punktem skalibrowanym. Patrząc na drzewo filogenetyczne przedstawione na Ryc. 1, kalibracja węzła C bardziej nadaje się do określenia wieku węzła D niż A. Również w takim przypadku wskazana jest kalibracja węzłów na głębszych poziomach filogenezy (WILKE i współaut. 2009, DUCHENE i współaut. 2014).

Określenie wieku dywergencji, które wydarzyły się stosunkowo niedawno, wymaga kalibracji na płytszych poziomach filogenezy, a kalibracja na poziomie korzenia lub głębokich węzłów, prowadzi do poważnych błędów i obniżonej precyzji w szacowaniu wartości tempa ewolucji. W tym przypadku tradycyjna kalibracja na podstawie wydarzeń geologicznych (mających zwykle co najmniej kilka mln lat) jest mało przydatna. Przykładem zwrócenia uwagi na ten problem są badania filogenezy populacji słodkowodnych ryb z rodzajów *Gobiomorphus* i *Galaxias*, zamieszkujących różne systemy rzeczne Nowej Zelandii. Tu najlepszymi punktami kalibracyjnymi okazały się niedawne plejstocenijskie zmiany geomorfologiczne terenu powodujące reorganizację sieci hydrologicznej, a w konsekwencji prowadzące do izolacji populacji zamieszkujących je organizmów (WATERS i współaut. 2007).

Nieadekwatna kalibracja może być brzemienne w skutkach, a ich charakter bywa trudny do przewidzenia. Jeżeli tempo ewolucji w danej linii jest bardzo wysokie (np. w przypadku wirusów), dodatkowym problemem jest szczególnie wysokie wysycenie mutacjami, zwłaszcza w głębszych partiach drzewa filogenetycznego. Nawet użycie bogatego w parametry modelu substytucji może być niewystarczające, bowiem substytucja w jednej pozycji mogła zajść wielokrotnie. Kalibracja płytkich węzłów w drzewie zrekonstruowanym na podstawie sekwencji z problemem saturacji, prowadzi do jego kompresji. Terminem tym określa się zjawisko niedoszacowania tempa mutacji na głębszym poziomie filogenezy. Skutkiem jest skrócenie długości podstawowych gałęzi drzewa, podczas gdy w rzeczywistości

powinny być one znacząco dłuższe. Kompresja ta jest dodatkowo niejednorodna i większa w przypadku długich gałęzi. Istnieje kilka strategii służących złagodzeniu efektu kompresji drzewa. Jedną z nich jest bardzo rygorystyczne podejście do selekcji modelu zegara molekularnego, ponieważ szacunki długości gałęzi są bardzo wrażliwe na parametry w nim zawarte. Również zmiana na wspomnianą już kalibrację głębokich węzłów, a nawet korzenia drzewa może być w takiej sytuacji pomocna (SULLIVAN i JOYCE 2005, DUCHENE i współaut. 2014). LANFEAR i współaut. (2010) zaproponowali nieco inne rozwiązanie problemu saturacji. Pierwsze zakłada wyłączenie z analizy informacji danych dotkniętych saturacją, co jest rozwiązaniem kłopotliwym, ponieważ zubaża zestaw analizowanych sekwencji. Drugie rozwiązanie szacuje tempa substytucji jedynie w oparciu o rzadziej zachodzące typy substytucji, skutkujące mniejszą liczbą wysycień. Przykładowo, można uzyskać tempo substytucji tylko na podstawie podstawień niesynonimicznych (używając do analizy np. sekwencji kodujących białka), które zachodzą znacznie rzadziej niż synonimiczne. Inni badacze w takiej sytuacji szacują tempo ewolucji tylko na podstawie transwersji, które zachodzą rzadziej niż tranzycje, używając metody zwanej RY-coding (ang. purine-pyrimidine coding). LANFEAR i współaut. (2010) wskazują, że mając do czynienia z problemem saturacji, lepiej jest ponownie przeprowadzić analizę, tyle że używając rzadszej kategorii substytucji, niż usuwać problematyczne dane. Z kolei przeszacowanie wieku płytkich węzłów doprowadzić może do rozszerzenia/nadbudowy drzewa. Jest to szczególnie problematyczne w analizie grup taksonomicznych, dla których brak jest wiarygodnych punktów kalibracyjnych lub jest ich bardzo mało. Powszechnie używanym rozwiązaniem jest wówczas włączenie do rekonstrukcji filogenezy grupy zewnętrznej, o znanym wieku rozejścia się z grupą analizowaną (WILKE i współaut. 2009, DUCHENE i współaut. 2014).

Określanie wieku węzła kalibracyjnego w latach

Istnieją trzy główne podejścia do określenia wieku kalibracji węzłów filogenetycznych.

Kalibracja punktowa. Polega ona na przypisaniu konkretnego wieku w latach, do co najmniej jednego węzła. Na przykład ZUCKERKANDL i PAULING (1965), badając czas pojawiania się duplikacji ssaczego genu globiny założyli, że ostatni wspólny przodek ssaków łożyskowych żył 80 mln lat temu. Tego rodzaju zamiana dowodu kopalnego w punkt kalibracyjny o ściśle określonym wieku jest jednak możliwa tylko wtedy,

gdy takson kopalny reprezentuje wspólnego przodka dwóch ciągle istniejących linii. Taka sytuacja należy do wyjątkowo rzadkich. Kalibracja punktowa jest najstarszą i najprostszą techniką kalibracji, której podstawową wadą jest niepewność wieku tak skalibrowanych węzłów, co może dać złudne wrażenie precyzji. Pomimo swoich wad metoda ta jest wciąż popularna, ze względu na łatwość jej stosowania.

Nałożenie minimalnych i/lub maksymalnych granic wieku węzła. Metoda ta nie ogranicza wieku węzła do jednej konkretnej wartości, lecz umożliwia przypisanie mu zakresu czasu, co jest podejściem znacznie rozsądniejszym niż poprzednie. Na przykład, SANDERSON (1997) w badaniach nad dywersyfikacją roślin nasiennych, na podstawie danych kopalnych założył, że minimalny wiek dywergencji pomiędzy roślinami szpilkowymi a miłorzębami wynosi 320 mln lat, a minimalny wiek dwuliściennych określił na 125 mln lat temu.

Rozkłady parametryczne. Rozwijane w ostatnich latach metody bayesowskie pozwalają na ograniczenie dolnej i górnej granicy wieku węzła poprzez zastosowanie, odzwierciedlających przedział ufności, parametrycznych rozkładów prawdopodobieństwa posiadania przez węzeł określonego wieku w zadanym zakresie czasu. W niektórych przypadkach możliwe jest określenie tylko jednej z granic wieku węzła. Wiek węzła w metodach bayesowskich może zostać wyrażony przy pomocy dowolnego rozkładu prawdopodobieństwa, zdefiniowanego przez badacza w zależności od potrzeb i wiedzy. Jednak najczęściej używane są rozkłady: normalny, logarytmiczny, wykładniczy i gamma (GRAUR i MARTIN 2004, HO i PHILIPS 2009, NOWAK i współaut. 2013). Dla przykładu, w swoich badaniach nad przyczynami i ramami czasowymi dywersyfikacji jednego z morfogatunków słodkowodnych skorupiaków obunogich w Europie, MAMOS i współaut. (2016) nałożyli rozkład logarytmiczny na węzeł reprezentujący wspólnego przodka populacji zamieszkujących obecnie Karpaty Południowe i Góry Bałkan. Żył on we wczesnym plejstocenie, między 2,6 mln lat temu, kiedy rozpoczęła się regresja Morza Moezyjskiego rozdzielającego te dwa łańcuchy górskie, a 1,3 mln lat temu, kiedy rozwinęło się dorzecze dolnego Dunaju, łączące spływające z nich rzeki.

Inne istotne kwestie związane z oszacowaniem tempa ewolucji przy pomocy zegara molekularnego

Ocena tempa substytucji na podstawie zestawu sekwencji o małej liczbie podstawień, wynikających z niskiego tempa sub-

stytucji, niedawnej dywergencji lub zbyt małej liczby porównywanych sekwencji, może być obarczona błędem. Dzieje się tak, ponieważ nawet niewielkie różnice w liczbie podstawień między sekwencjami, spowodują nieproporcjonalnie duże przeszacowanie tempa ewolucji. Jest to szczególnie problematyczne w filogenetyce porównawczej, kiedy zmiany tempa substytucji są podstawową analizowaną zmienną. Istnieje jednak test pozwalający zidentyfikować takie pozornie istotne zmiany tempa ewolucji. Porównuje on wielkości wariancji oszacowanych zmian tempa podstawień i wychwytuje te, które są wynikiem zbyt małej liczby substytucji w zestawie danych. Po rozpoznaniu, problematyczne oszacowania powinny zostać wykluczone z analizy. Trudno pogodzić się z taką utratą niekiedy sporej części danych. Jednak zarówno teoria, jak i praktyka naukowa pokazują, że jest to często jedyny sposób na polepszenie mocy kolejnych analiz statystycznych (LANFEAR i współaut. 2010).

Również próba określenia tempa ewolucji wśród blisko spokrewnionych gatunków wiąże się z niewielką liczbą podstawień w zestawie danych. Warto podkreślić, że nie wszystkie różnice pomiędzy sekwencjami to substytucje, które utrwały się po rozpoczęciu izolacji reprodukcyjnej pomiędzy gatunkami. Mogą to być wewnątrzgatunkowe polimorfizmy, pojawiające się spontanicznie u niektórych osobników lub utrwalone wcześniej polimorfizmy ancestralne. Mylenie takich substytucji z intraspecyficznymi polimorfizmami może utrudniać uzyskanie wiarygodnych wyników. Na przykład, w przeciwieństwie do wcześniej utrwalonych substytucji, częstotliwość występowania polimorfizmów wewnątrzgatunkowych zależy od wielkości populacji. W związku z powyższym, najlepiej jest stosować analizę tempa ewolucji do dywergencji na tyle starych linii, że liczba polimorfizmów wewnątrzgatunkowych będzie nieistotna w porównaniu z liczbą substytucji różnicujących gatunki (WILKE i współaut. 2009, LANFEAR i współaut. 2010).

Następnym problemem związanym z analizami tempa ewolucji jest efekt „gęstości węzłów” (ang. node-density effect), w którym liczba substytucji często jest niedoszacowana w przypadku długich gałęzi drzewa. Wynika to między innymi z wysycenia substytucjami, braku wiedzy na temat wymarłych linii filogenetycznych lub ich faktycznej różnorodności w części drzewa zawierającej długie gałęzie. Można wówczas dojść do mylnego wniosku, że tempo ewolucji jest wyższe w części drzewa, gdzie znajduje się większa liczba węzłów. Szczególnie ważne jest zwrócenie uwagi na

to zjawisko w badaniach porównujących tempo dywersyfikacji różnych grup organizmów. Istnieją jednak testy statystyczne sprawdzające, czy filogeneza pozostaje pod wpływem efektu „gęstości węzłów” (LANFEAR i współaut. 2010).

#### PRZYKŁADOWE ZASTOSOWANIE KONCEPCJI ZEGARA MOLEKULARNEGO W BADANIACH EWOLUCYJNYCH

Znaczenie filogenetyki molekularnej zaczęło rosnąć w latach 90. XX w., ze względu na opracowanie i popularyzację techniki PCR umożliwiającej szybkie namnażanie fragmentów DNA, powstawanie coraz bardziej rygorystycznych metod konstruowania drzew oraz rosnące zainteresowanie projektami genomowymi. Zegar molekularny, będący jednym z ważniejszych owoców filogenetyki, znalazł wiele zastosowań w różnych dziedzinach nauk biologicznych i medycznych, takich jak: biologia ewolucyjna, biologia systematyczna, ochrona środowiska (np. monitorowanie bioróżnorodności czy tempa powstawania i wymierania gatunków), epidemiologia molekularna czy archeologia i prehistoria człowieka.

#### ARCHEOLOGIA I POCHODZENIE CZŁOWIEKA

W badaniach odtwarzających historię ewolucyjną człowieka zawsze były ważne metody zarówno fizyczne (np. datowanie izotopowe), jak i biologiczne. Archeolodzy sięgali najczęściej do danych z dziedziny anatomii, patologii czy palinologii. Jednak od lat 80. XX w. coraz chętniej wykorzystują w tym celu biologię molekularną, a wraz z nią zegar molekularny. Przykładem zastosowania tej koncepcji w antropologii są badania nad pochodzeniem człowieka i jego pokrewieństwem z innymi zwierzętami.

Karol Darwin by pierwszym biologiem, który zasugerował, że istnieje bliskie pokrewieństwo pomiędzy człowiekiem a innymi naczelnymi. Przez kilka dekad wielu naukowców, w tym ewolucjonistów, odrzucało tę ideę jako zbyt kontrowersyjną. Do lat 60. XX w. badania nad skamieniałościami potwierdziły co prawda teorię, że to szympanś i goryl są naszymi najbliższymi krewniakami, jednak pokrewieństwo to uważano za odległe w czasie. Więcej światła na tę sprawę rzuciły liczne badania molekularne z wykorzystaniem zegara molekularnego, które polegały na porównaniu całych genomów oraz wzoru metylacji wysp CpG u człowieka i innych wybranych naczelnych. Analiza zebranych danych pokazała, że relacje z szympansem i gorylem są znacznie bliższe niżby to wynikało z danych paleontologicznych.

Szympanś okazał się najbliższym spokrewnionym z człowiekiem (rozdzielił linię ewolucyjną nastąpił około 9,3–6,5 mln lat temu), natomiast goryl jest naszym nieco dalszym krewnym (rozdzielił nastąpił około 12,2–9,4 mln lat temu) (MOORJANI i współaut. 2016).

Wciąż prowadzone są intensywne badania nad pochodzeniem współczesnych naczelnych. Dotychczasowe badania molekularne i morfologiczne wspierają monofiletyczność kładu Euarchonta, do którego należą współcześnie żyjące trzy grupy ssaków: naczelne, wiewióreczniki i skóroskrzydłe (czasem błędnie nazywane latającymi lemurami). Siostrzaną grupą dla tego kładu są siekaczowce (Glires – zajęczaki i gryzonie). Przez długi czas brakowało wiedzy na temat relacji filogenetycznych w obrębie Euarchonta. W ciągu ostatnich 30 lat powstały trzy hipotezy: (1) grupą siostrzaną dla naczelnych są wiewióreczniki, (2) grupą siostrzaną dla naczelnych są skóroskrzydłe, (3) grupą siostrzaną dla naczelnych są zarówno wiewióreczniki, jak i skóroskrzydłe. JANECKA i współaut. (2007) użyli „relaxed molecular clock”, aby odtworzyć filogenezę grupy Euarchonta. Kład Glires wykorzystano jako grupę zewnętrzną. Wyniki potwierdziły jednoznacznie, że to skóroskrzydłe są najbliższymi żyjącymi krewnymi naczelnymi, od których oddzieliły się ok. 80 milionów lat temu. Natomiast grupa skóroskrzydłe-naczelne oddzieliła się od wiewióreczników ok. 88 milionów lat temu.

#### BIOLOGIA SYSTEMATYCZNA

Ostatnie lata badań z wykorzystaniem markerów molekularnych zrewolucjonizowały poglądy na temat różnorodności biologicznej w przypadku wielu grup organizmów. Okazało się, że wiele gatunków zwierząt i roślin definiowanych konwencjonalnie w oparciu o cechy fenotypowe, to w rzeczywistości kompleksy odrębnych linii filogenetycznych, nierozróżnialnych morfologicznie lecz dywergentnych genetycznie i posiadających własne historie ewolucyjne. Zastosowanie zegara molekularnego ujawniło np., że w przypadku słodkowodnych skorupiaków obunogich z morfogatunku *Gammarus fossarum* zamieszkujących Karpaty, linie te zaczęły się rozdzielać nawet 12 milionów lat temu, a większość z nich ewoluuje odrębnie już od ponad 5 milionów lat (COPILAȘ-CIOCIANU i PETRUSEK 2015). Datowanie pozwoliło również na powiązanie historii ewolucyjnej poszczególnych linii z historią geologiczną Karpat. Dla odmiany, w przypadku strzykw z rodzaju *Holothuria* występujących w Morzu Śródziemnym i północno-wschodnim Atlantyku, poszczególne gatunki mają jeszcze dłuższą historię, sięgającą zwykle 20–60 milionów lat, a



niekiedy nawet 170 milionów lat (BORRERO-PÉREZ i współaut. 2010).

#### EPIDEMIOLOGIA MOLEKULARNA

Jednym z najbardziej praktycznych zastosowań zegara molekularnego są badania nad powstawaniem i rozprzestrzenianiem się chorób zakaźnych. Oszacowanie czasu powstania linii filogenetycznego patogenu może być kluczem do odkrycia miejsca i przyczyny pojawienia się choroby. Patogeny takie jak wirusy nie pozostawiają po sobie żadnych danych kopalnych, przez co ich geneza jest niejasna i jedynie dane molekularne mogą pozwolić na rekonstrukcję historii epidemii, głównie na podstawie zachowanych próbek (np. krwi czy tkanek pacjentów). Wirusowe zegary molekularne są najczęściej kalibrowane na podstawie wcześniej zgromadzonych próbek, np. jeśli próbki krwi pobrano seryjnie od tego samego zainfekowanego pacjenta albo jest dostępny zbiór próbek archiwalnych, to możliwe jest przetestowanie stałości tempa ewolucji wzdłuż filogenezy wirusa (BOMHAM i PENNY 2003).

Bardzo znanym przykładem jest ustalenie historii pandemii AIDS spowodowanej przez wirus zespołu nabytego braku odporności (HIV). Dokładne datowanie transmisji tego wirusa jest krytyczne dla określenia ryzykownych zachowań i rozwoju odpowiedzi immunologicznej organizmu człowieka. Wielokrotnie podejmowano próby zrekonstruowania historii ewolucyjnej HIV. Badania filogenetyczne z użyciem metody zegara molekularnego pozwoliły na wyciągnięcie co najmniej kilku istotnych wniosków. Najprawdopodobniej wirus HIV pochodzi od małego wirusa nabytego niedoboru odporności (SIV) i pojawił się on w ludzkich populacjach Zachodniej i Centralnej Afryki w wyniku wielokrotnych odzwierzęcych transmisji. Początkowo pojawiła się teoria jakoby źródłem wirusa SIV była zanieczyszczona doustna szczepionka przeciwko polio, którą podawano afrykańskim dzieciom w ramach programów szczepień ochronnych prowadzonych na przełomie lat 50 i 60 (szczepionka była produkowana na tkankach szympansa). Hipotezę tę jednak odrzucono. Międzygatunkowe infekcje mogły zajść w wyniku polowań na małpy, jako źródło mięsa, oraz popularnego w tym regionie trzymania ich jako zwierzęta domowe. Datowanie na podstawie danych molekularnych wskazuje na to, że główny szczep, HIV-1 M, jest najstarszą linią atakującą ludzi, której ostatni wspólny (z innymi ludzkimi szczepami) żyjący przodek istniał na początku XX w. Szacuje się, że ostatni wspólny żyjący przodek HIV-1 i SIV żył najprawdopodobniej w drugiej połowie XIX w. (KORBER i współaut. 2000, HEMELAAR 2012).

Przykłady przywołane w tej pracy pokazują, jak bardzo przydatny jest zegar molekularny przy rozwiązywaniu problemów badawczych z bardzo różnych specjalności nauk biologicznych. Dlatego, pomimo obszernej podbudowy teoretycznej tej koncepcji i dość skomplikowanej metodologii rekonstrukcji chronogramów, warto sięgnąć po to narzędzie w praktyce badawczej.

#### PODZIĘKOWANIA

Serdecznie dziękujemy Magdalenie Błażewicz, Piotrowi Gadawskiemu, Aleksandrze Jakiel, Małgorzacie Koliczkiej, Krzysztofowi Pabisowi i Agnieszce Soszyńskiej-Maj, którzy jako czytelnicy pierwotnej wersji manuskryptu przekazali nam szereg uwag i poprawek, które pozwoliły udoskonalić tekst i uczynić go bardziej przystępnym dla odbiorcy. Pragniemy również podziękować anonimowemu recenzentowi za cenne uwagi merytoryczne.

#### Streszczenie

Koncepcja zegara molekularnego powstała ponad pół wieku temu i od tamtej pory odegrała ważną rolę w badaniach biologicznych, zajmujących się zarówno opisem procesów ewolucyjnych, jak i szacowaniem wieku dywergencji. Zegar molekularny odmierza czas pomiędzy poszczególnymi wydarzeniami ewolucyjnymi, opierając się na założonym modelu ewolucji (czyli tempie substytucji). Dziesiątki lat badań pokazały, że tempo ewolucji nie jest wartością uniwersalną dla wszystkich organizmów i zależy od szeregu czynników, takich jak: wielkość ciała, długość trwania cyklu życiowego czy tempo metabolizmu. Spowodowało to opracowanie wielu modeli zegara molekularnego biorących pod uwagę zmienność tempa wzdłuż gałęzi drzewa filogenetycznego. Odmianą rolę w odtwarzaniu filogenezy przy użyciu zegara molekularnego jako narzędzia, stanowi jego kalibracja. Istnieje kilka metod kalibracyjnych, których zastosowanie zależy od dostępności punktów kalibracyjnych dla badanej grupy organizmów. Głównym celem artykułu jest zapoznanie czytelnika zarówno z teoretycznymi, jak i z praktycznymi zagadnieniami dotyczącymi zegara molekularnego.

#### LITERATURA

- AKIN C., BILGIN C., BEERLI P., WESTAWAY R., OHST T. i współaut., 2010. *Phylogeographic patterns of genetic diversity in eastern Mediterranean water frogs have been determined by geological processes and climate change in the Late Cenozoic*. J. Biogeogr. 37, 2111-2124.
- AVISE J. C., 2008. *Markery molekularne historia naturalna i ewolucja*. Wydawnictwo Uniwersytetu Warszawskiego. Warszawa.
- BALLARD J. W. O., WHITLOCK M. C., 2004. *The incomplete natural history of mitochondria*. Mol. Ecol. 13, 729-744.
- BARRICK J. E., LENSKI R. E., 2013. *Genome dynamics during experimental evolution*. Nat. Rev. Genet. 14, 827-839.
- BAUM D. A., SMITH S. D., 2013. *Tree thinking: An introduction to phylogenetic biology*. Roberts and Company Publishers, Inc., Greenwood Village, Colorado.



- BORRERO-PÉREZ G. H., GÓMEZ-ZURITA J., GONZÁLEZ-WANGÜEMERT M., MARCOS C., PÉREZ-RUZAF A., 2010. *Molecular systematics of the genus Holothuria in the Mediterranean and Northeastern Atlantic and a molecular clock for the diversification of the Holothuriidae (Echinodermata: Holothuroidea)*. Mol. Phylog. Evol. 57, 899-906.
- BRITTON T., ANDERSON C. L., JACQUET D., LUNDQVIST S., BREMER K., 2007. *Estimating divergence times in large phylogenetic trees*. Syst. Biol. 56, 741-752.
- BROMHAM L., 2009. *Why do species vary in their rate of molecular evolution?* Biol. Lett. 5, 401-404.
- BROMHAM L., PENNY D., 2003. *The modern molecular clock*. Nat. Rev. Genet. 4, 216-224.
- BROWER A. V. Z., 1994. *Rapid morphological radiation and convergence among races of the butterfly Heliconius erato inferred from patterns of mitochondrial DNA evolution*. Proc. Natl. Acad. Sci. USA 91, 6491-6495.
- BULMER M., WOLFE K. H., SHARP P. M., 1991. *Synonymous nucleotide substitution rates in mammalian genes: Implications for the molecular clock and the relationship of mammalian orders*. Proc. Natl. Acad. Sci. USA 88, 5974-5978.
- COPILAŞ-CIOCIANU D., PETRUSEK A., 2015. *The southwestern Carpathians as an ancient centre of diversity of freshwater gammarid amphipods: insights from the Gammarus fossarum species complex*. Mol. Ecol. 24, 3980-3992.
- DICKERSON R. E., 1971. *The structures of cytochrome c and the rates of molecular evolution*. J. Mol. Evol. 1, 26-45.
- DONOGHUE P. C. J., BENTON M. J., 2007. *Rocks and clocks: calibrating the Tree of Life using fossils and molecules*. Trends Ecol. Evol. 22, 424-431.
- DOOLITTLE R. F., FENG D-F., TSANG S., CHO G., LITTLE E., 1996. *Determining divergence times of the major kingdoms of living organisms with a protein clock*. Science 271, 470-477.
- DRUMMOND A. J., BOUCKAERT R. R., 2015. *Bayesian evolutionary analysis with BEAST 2*. Cambridge University Press, United Kingdom.
- DRUMMOND A. J., HO S. Y. W., PHILLIPS M. J., RAMBAUT A., 2006. *Relaxed phylogenetics and dating with confidence*. PLoS Biol. 4, 0699-0710.
- DRUMMOND A. J., SUCHARD M. A., XIE D., RAMBAUT A., 2012. *Bayesian phylogenetics with BEAUti and the BEAST 1.7*. Mol. Biol. Evol. 29, 1969-1973.
- FONTANILLAS E., WELCH J. J., THOMAS J. A., BROMHAM L., 2007. *The influence of body size and net diversification rate on molecular evolution during the radiation of animal phyla*. BMC Evol. Biol. 7, 1-12.
- FOREST F., 2009. *Calibrating the tree of life: fossils, molecules and evolutionary timescales*. Ann. Bot. 104, 789-794.
- FREELAND J. R., 2008. *Ekologia molekularna*. Wydawnictwo Naukowe PWN. Warszawa.
- FUTUYMA D., 2005. *Ewolucja*. Wydawnictwo Uniwersytetu Warszawskiego. Warszawa.
- GALTIER N., 2011. *The intriguing evolutionary dynamics of plant mitochondrial DNA*. BMC Biol. 9, 61.
- GAVRYUSHKINA A., HEATH T. A., KSEPKA D. T., STADLER T., WELCH D. i wsp. 2017. *Bayesian total-evidence dating reveals the recent crown radiation of penguins*. Syst. Biol. 66, 57-73.
- GIBBS G., 2016. *Ghosts of Gondwana: The history of life in New Zealand*. Potton & Burton Limited.
- GILLESPIE J. H., 1984. *The molecular clock may be an episodic clock*. Proc. Natl. Acad. Sci. USA 81, 8009-8013.
- GRAUR G., MARTIN W., 2004. *Reading the entrails of chickens: molecular timescales of evolution and the illusion of precision*. Trends Genet. 20, 80-86.
- GROUSSIN M., PAWLOWSKI J., YANG Z. H., 2011. *Bayesian relaxed clock estimation of divergence times in foraminifera*. Mol. Phylogenet. Evol. 61, 157-166.
- HALL B. G., 2008. *Łatwe drzewa filogenetyczne*. Wydawnictwo Uniwersytetu Warszawskiego. Warszawa.
- HASEGAWA M., KISHINO H., YANO T., 1989. *Estimation of branching dates among primates by molecular clocks of nuclear DNA which slowed down in Hominoidea*. J. Human Evol. 18, 461-476.
- HAY J. M., SUBRAMANIAN S., MILLAR C. D., MOHANDASAN E., LAMBERT D. M., 2008. *Rapid molecular evolution in a living fossil*. Trends Genet. 24, 106-109.
- HEMELAAR J., 2012. *The origin and diversity of the HIV-1 pandemic*. Trends Mol. Med. 18, 182-192.
- HO S. Y. W., 2009. *An examination of phylogenetic models of substitution rate variation among lineages*. Biol. Lett. 5, 421-424.
- HO S. Y. W., 2014. *The changing face of the molecular evolutionary clock*. Trends Ecol. Evol. 29, 496-503.
- HO S. Y. W., PHILLIPS M. J., 2009. *Accounting for calibration uncertainty in phylogenetic estimation of evolutionary divergence times*. Syst. Biol. 58, 367-380.
- HO S. Y. W., LO N., 2013. *The insect molecular clock*. Austr. J. Entomol. 52, 101-105.
- HO S. Y. W., DUCHENE S., 2014. *Molecular-clock methods for estimating evolutionary rates and timescales*. Mol. Ecol. 23, 5947-5965.
- HO S. Y. W., LANFEAR R., BROMHAM L., PHILLIPS M. J., SOUBRIER J. i wsp. 2011. *Time-dependent rates of molecular evolution*. Mol. Ecol. 20, 3087-3101.
- HO S. Y. W., SHAPIRO B., PHILLIPS M. J., COOPER A., DRUMMOND A. J., 2007. *Evidence for time dependency of molecular rate estimates*. Syst. Biol. 56, 515-522.
- HÖHNA S., LANDIS M. J., HEATH T. A., BOUSSAU B., LARTILLOT N., MOORE B. R., HUELSENBECK J. P. i wsp. 2016. *RevBayes: Bayesian phylogenetic inference using graphical models and an interactive model-specification language*. Syst. Biol. 65, 726-736.
- JANEČKA J. E., MILLER W., PRINGLE T. H., WIENS F., ZITZMANN A. i wsp. 2007. *Molecular and genomic data identify the closest living relative of primates*. Science 318, 792-794.
- KIMURA M., 1968. *Evolutionary rate at the molecular level*. Nature 217, 624-626.
- KORBER B., MULDOON M., THEILER J., GAO F., GUPTA R. i wsp. 2000. *Timing the ancestor of the HIV-1 pandemic strains*. Science 288, 1789-1796.
- KSEPKA D. T., PARHAM J. F., ALLMAN J. F., BENTON M. J., CARRANO M. T. i wsp. 2015. *The fossil calibration database – a new resource for divergence dating*. Syst. Biol. 64, 853-859.
- LANFEAR R., WELCH J., BROMHAM L., 2010. *Watching the clock: Studying variation in rates of*

- molecular evolution between species. *Trends Ecol. Evol.* 25, 495-503.
- LARTILLOT N., LEPAGE T., BLANQUART S. 2009. *PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating*. *Bioinformatics* 25, 2286-2288.
- LEE M. S. Y., 1999. *Molecular clock calibrations and metazoan divergence dates*. *J. Mol. Evol.* 49, 385-391.
- LEWIN R., 1997. *Patterns in Evolution: The new molecular view*. "Scientific American" Library. New York.
- LOCKWOOD J.D., ALEKSIC J., ZOU J., WAN, J., LIU J. i współprac., 2013. *A new phylogeny for the genus Picea from plastid, mitochondrial, and nuclear sequences*. *Mol. Phylog. Evol.* 96, 717-727.
- MAMOS T., WATTIER R., BURZYŃSKI A., GRABOWSKI M., 2016. *The legacy of a vanished sea: a high level of diversification within a European freshwater amphipod species complex driven by 15 My of Paratethys regression*. *Mol. Ecol.* 25, 795-810.
- MARTIN A. P., PALUMBI S. R., 1993. *Body size, metabolic rate, generation time, and the molecular clock*. *Proc. Natl. Acad. Sci. USA* 90, 4087-4091.
- MOORJANI P., AMORIM C. E., ARNDT P., PRZEWORSKI M., 2016. *VARIATION IN THE MOLECULAR CLOCK OF PRIMATES*. *Proc. Natl. Acad. Sci. USA* 113, 10607-10612.
- MORGAN G. J., 1998. *Emile Zuckerkandl, Linus Pauling, and the Molecular Evolutionary Clock, 1959-1965*. *J. Hist. Biol.* 31, 155-178.
- NOWAK M. D., SMITH A. B., SIMPSON C., ZWICKL D. J., 2013. *A simple method for estimating informative node age priors for the fossil calibration of molecular divergence time analyses*. *PLoS One* 8, e66245.
- OHTA T., 1995. *Synonymous and nonsynonymous substitutions in mammalian genes and the nearly neutral theory*. *J. Mol. Evol.* 40, 56-63.
- PAPADOPOULOU A., ANASTASIOU I., VOGLER A. P., 2010. *Revisiting the insect mitochondrial molecular clock: the mid-Aegean trench calibration*. *Mol. Biol. Evol.* 27, 1659-1672.
- PARSONS T. J., MUNIEC D. S., SULLIVAN K., WOODYATT N., ALLISTON-GREINER R. i współprac., 1997. *A high observed substitution rate in the human mitochondrial DNA control region*. *Nat. Genet.* 15, 363-368.
- PESOLE G., GISSI C., DE CHIRICO A., SACCONI C., 1999. *Nucleotide substitution rate of mammalian mitochondrial genomes*. *J. Mol. Evol.* 48, 427-34.
- POSADA D., 2009. *Selection of models of DNA evolution with jModelTest*. *Meth. Mol. Biol.* 537, 93-112.
- POSADA D., BUCKLEY T. R., 2004. *Model selection and model averaging in phylogenetics: advantages of akaike information criterion and bayesian approaches over likelihood ratio tests*. *Syst. Biol.* 53, 793-808.
- PUURTINEN M., ELO M., JALASVUORI M., KAHILAINEN A., KETOLA T., KOTIAHO J. S. i współprac., 2016. *Temperature-dependent mutational robustness can explain faster molecular evolution at warm temperatures, affecting speciation rate and global patterns of species diversity*. *Ecography* 39, 1025-1033.
- RAMBAUT A., DRUMMOND A. J., XIE D., BAELE G., SUCHARD M. A., 2018. *Posterior summarisation in Bayesian phylogenetics using Tracer 1.7*. *Syst. Biol.* (in press).
- ROCHA L. A., ROCHA C. R., ROBERTSON D. R., BOWEN B. W., 2008. *Comparative phylogeography of Atlantic reef fishes indicates both origin and accumulation of diversity in the Caribbean*. *BMC Evol. Biol.* 8, 157.
- RONQUIST F., TESLENKO M., VAN DER MARK P. i współprac., 2012. *MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space*. *Syst. Biol.* 61, 539-542.
- RUNNEGAR B., 1982. *A molecular-clock date for the origin of the animal phyla*. *Lethaia* 15, 199-205.
- SANDERSON M. J., 1997. *A nonparametric approach to estimating divergence times in the absence of rate constancy*. *Mol. Biol. Evol.* 14, 1218-1231.
- SANDERSON M. J., 2003. *r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock*. *Bioinformatics* 19, 301-302.
- SAUQUET H., HO S. Y., GANDOLFO M. A., JORDAN G. J., WILF P. i współprac., 2012. *Testing the impact of calibration on molecular divergence times using a fossil-rich group: the case of Nothofagus (Fagales)*. *Syst. Biol.* 61, 289-313.
- SCHWARTZ R. S., MUELLER R., 2010. *Variation in DNA substitution rates among lineages erroneously inferred from simulated clock-like data*. *PLoS One* 5, e9649.
- SMITH S. A., O'MEARA B. C., 2012. *treePL: divergence time estimation using penalized likelihood for large phylogenies*. *Bioinformatics* 28, 2689-2690.
- SPALIK K., PIWCZYŃSKI M., 2009. *Rekonstrukcja filogenezy i wnioskowanie filogenetyczne w badaniach ewolucyjnych*. *Kosmos* 58, 485-498.
- SULLIVAN J., JOYCE P., 2005. *Model selection in phylogenetics*. *Ann. Rev. Ecol. Evol. Syst.* 36, 445-66.
- TAMURA K., BATTISTUZZI F. U., BILLING-ROSS P., MURILLO O., FILIPSKI A. i współprac., 2012. *Estimating divergence times in large molecular phylogenies*. *Proc. Natl. Acad. Sci. USA* 109, 19333-19338.
- THOMAS J. A., WELCH J. J., LANFEAR R., BROMHAM L., 2010. *A generation time effect on the rate of molecular evolution in invertebrates*. *Mol. Biol. Evol.* 5, 1173-1180.
- THORNE J. L., KISHINO H., PAINTER I. S., 1998. *Estimating the rate of evolution of the rate of molecular evolution*. *Mol. Biol. Evol.* 15, 1647-1657.
- VILLARREAL J. C., RENNER S. S., 2012. *Hornwort pyrenoids, a carbon-concentrating mechanism, evolved and were lost at least five times during the last 100 million years*. *Proc. Natl. Acad. Sci. USA* 109, 18873-18878.
- WARNOCK R. C. M., PARHAM J. F., JOYCE W. G., LYSON T. R., DONOGHUE P. C. J., 2015. *Calibration uncertainty in molecular dating analyses: there is no substitute for the prior evaluation of time priors*. *Proc. Royal Soc. B* 282, 20141013.
- WATERS J. M., ROWE D. L., APTE S., KING T. M., WALLIS G. P. i współprac., 2007. *Geological dates and molecular rates: rapid divergence of rivers and their biotas*. *Syst. Biol.* 56, 271-282.
- WEIR J. T., SCHLUTER D., 2008. *Calibrating the avian molecular clock*. *Mol. Ecol.* 17, 2321-2328.
- WELCH J. J., BROMHAM L., 2005. *Molecular dating when rates vary*. *Trends Ecol. Evol.* 20, 620-627.
- WILKE T., SCHULTHEISS R., ALBRECHT C., 2009. *As time goes by: A simple fool's guide to mo-*

- lecular clock approaches in invertebrates.* Am. Malacol. Bull. 27, 25-45.
- WILSON A. C., CARLSON S. S., WHITE T. J., 1977. *Biochemical evolution.* Ann. Rev. Biochem. 46, 573-639.
- WOLFE K. H., LI W. H., SHARP P. M., 1987. *Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs.* Proc. Natl. Acad. Sci. USA 84, 9054-9058.
- WOOLFIT M., 2009. *Effective population size and the rate and pattern of nucleotide substitutions.* Biol. Lett. 5, 417-420.
- WYSOCKA A., GRABOWSKI M., SWOROBOWICZ L., BURZYŃSKI A., KILKOWSKA A i współaut. 2013. *A tale of time and depth: intralacustrine radiation in endemic Gammarus species flock from the ancient Lake Ohrid.* Zool. J. Linnean Soc. 167, 345-359.
- XIA X., YANG Q., 2011. *A distance-based least-square method for dating speciation events.* Mol. Phylogenet. Evol. 59, 342-353.
- YANG Z., 2007. *PAML 4: phylogenetic analysis by maximum likelihood.* Mol. Biol. Evol. 24, 1586-1591.
- YODER A. D., YANG Z., 2000. *Estimation of primate speciation dates using local molecular clocks.* Mol. Biol. Evol. 17, 1081-1090.
- ZUCKERKANDL E., PAULING L., 1962. *Molecular disease, evolution and genic heterogeneity.* Horiz. Biochem. Albert Szent-Györgyi Dedicatory Vol., 189-225.
- ZUCKERKANDL E., PAULING L., 1965. *Evolutionary divergence and convergence in proteins.* Evolv. Genes Proteins, 97-166.

**KOSMOS Vol. 67, 3, 475–493, 2018**

WERONIKA WRZESIŃSKA, TOMASZ MAMOS, MICHAŁ GRABOWSKI

*Department of Invertebrate Zoology and Hydrobiology, University of Lodz, Banacha 12/16, 90-237 Lodz,  
E-mail: weronika.wrzesinska@biol.uni.lodz.pl, tomasz.mamos@biol.uni.lodz.pl, michal.grabowski@biol.uni.lodz.pl*

#### MOLECULAR CLOCK IN THEORY AND PRACTICE: A GUIDE FOR BEGINNERS

##### Summary

The concept of molecular clock was established more than half a century ago. Since then it has played an important role in studies dealing both with the description of evolutionary processes and with the estimation of time periods at which evolutionary lineages diverged. The molecular clock measures the time passed between subsequent evolutionary events, based on the assumed model of evolution (substitution rate). Decades of studies have shown that the substitution rate is not universal for all organisms. It differs among taxa and this heterogeneity is related to a number of factors, such as: body size, generation time, and metabolic rate. Uncovering causes and correlates of the rate variation have induced development of many new models of molecular clock, taking into account the variation of substitution rate along the branches in the phylogenetic tree. Proper calibration of molecular clock plays a major role in inferring time-calibrated phylogeny. There are several calibration methods and their application depends on the availability of calibration points for the studied organisms. The main aim of our paper is to familiarize the reader with both the theoretical and practical issues related to the idea of molecular clock and its practical application.

Key words: molecular clock, rate of evolution, substitution rate, calibration methods, molecular clock models, divergence, molecular markers