

Kraków, 17 marca 2014

dr hab. Wiesław Babik
Instytut Nauk o Środowisku
Uniwersytet Jagielloński
ul. Gronostajowa 7
30-387 Kraków
email: wieslaw.babik@uj.edu.pl
tel. 12 664 51 71

**Recenzja rozprawy doktorskiej Pani mgr Joanny Ciomborowskiej
"Functional retrogens in the human genome"**

Retrotranspozycja jest jednym z istotnych procesów kształtujących ewolucję eukariotycznych. Jednym z jej "skutków ubocznych" jest powstawanie retrokopii, czyli produktów retropozycji genów wieloegzonowych. Uważa się, iż większość tak powstałych sekwencji jest niefunkcjonalna, lecz okazuje się, iż funkcjonalne retrogeny nie są wcale wielką rzadkością, natomiast ich liczba w genomach i ewolucja są dość słabo poznane. Właśnie badania nad funkcjonalnymi retrogenami są przedmiotem ocenianej rozprawy doktorskiej.

Zasadniczą częścią rozprawy są trzy publikacje: jedna z nich to rozdział o charakterze artykułu przeglądowego w książce *Evolutionary Biology – Mechanisms and Trends*, a dwie pozostałe to oryginalne artykuły w bardzo dobrym czasopiśmie *Molecular Biology and Evolution*. Poza tym rdzeniem merytorycznym praca zawiera streszczenie w językach polskim i angielskim, oświadczenie doktorantki o jej udziale w poszczególnych publikacjach oraz oświadczenia współautorów, określające ich udział w powstaniu publikacji. Z oświadczeń tych wynika, iż udział doktorantki w przypadku każdej z prac był fundamentalny, co uzasadnia wykorzystanie publikacji jako podstawy rozprawy doktorskiej. W jednej z prac oryginalnych p. Ciomborowska dzieli pierwsze autorstwo z dr. M. Szcześniakiem. Praktyka ta jest powszechna w przypadku szeroko zakrojonych, wielowątkowych projektów; moim zdaniem nie ma żadnej przeszkody we włączaniu takich prac w skład rozprawy doktorskiej. W streszczeniu autorka omawia zwięźle zawartość trzech artykułów stanowiących rdzeń rozprawy. Wersja angielska napisana jest zrozumiałym językiem, lecz można w niej odnaleźć sporo kalek językowych i kilka niezbyt szczęśliwych sformułowań. Rezygnacja ze wstępu, często przygotowywanego w przypadku rozpraw mających charakter cyklu publikacji, jest tutaj całkowicie uzasadniona, gdyż rolę wstępu odgrywa publikacja 1.

Publikacja ta to artykuł przeglądowy omawiający funkcjonalne retrogeny w genomach zwierząt. Krótkie wprowadzenie zajmuje się znaczeniem duplikacji w ewolucji genomów i rolą retropozycji w generowaniu duplikacji. Następnie omówiono dane dotyczące liczby funkcjonalnych pseudogenów. Duże wahania w oszacowaniach wynikają z różnorodności kryteriów stosowanych przez różnych autorów oraz braku jasności, które – jeżeli w ogóle

którekolwiek – kryteria są uniwersalnie użyteczne. Z kryteriami ściśle wiążą się metody identyfikacji omawiane w kolejnym podrozdziale. Podkreślono dominację różnorodnych metod bioinformatycznych, znacznie rzadziej wnioskowanie o funkcji jest weryfikowane metodami eksperymentalnymi. Kolejnym omawianym zagadnieniem jest rola retrogenów w kształtowaniu różnic między gatunkami, zwłaszcza w kontekście roli w specjacji, która to rola pojawia się również w podsumowaniu. W przypadku tego podrozdziału, być może z racji własnych zainteresowań, mam najwięcej wątpliwości. W szczególności uważam, że nie przedstawiono argumentów za bezpośrednią rolą retrogenów w specjacji. Natomiast zasadnicze różnice w składzie retrogenów między gatunkami, których dywergencja nastąpiła dość dawno, tak jak w przypadku myszy i człowieka ok. 90 mln lat temu, nie mogą być używane jako argument, że ich rola w specjacji jest duża. Uważam, że wskazuje to raczej na to, iż proces retropozycji zachodzi ciągle (o czym zresztą autorzy piszą), jednak czas życia powstałych w ten sposób retrogenów jest często krótki, co prowadzi do dywergencji składu retrogenów między liniami ewolucyjnymi, jednak nie dowodzi ich roli w specjacji, rozumianej jako powstanie izolacji rozrodczej między populacjami. Mam również pewną wątpliwość co do użyteczności przykładu dotyczącego retrogenu *fgf4*, występującego u ras psów posiadających krótkie nogi. Wysoka częstość lub utrwalenie tego retrogenu w krótkonogich rasach wydaje się być wynikiem doboru sztucznego i w związku z tym może nie być reprezentatywna dla ewolucyjnych losów retrogenów. Jest prawdopodobne, iż w naturalnych populacjach wariant taki byłby usunięty lub utrzymywany w niskiej częstości przez dobór oczyszczający. Następny podrozdział omawia ewolucję retrogenów, zachowywanie przez nie funkcjonalności i nabywanie nowych funkcji. Ostatni podrozdział zajmuje się wzorcami ekspresji funkcjonalnych pseudogenów, a zwłaszcza ich tkankowo-specyficzną ekspresją, która ogranicza się często do jąder. Podjęto też próbę wyjaśnienia obserwacji, iż u gatunków o systemie determinacji płci XX/XY funkcjonalne retrogeny często pochodzą z genów na chromosomie X. Przedstawiono aktualne hipotezy oraz obserwacje przemawiające za nimi. Pracę kończą krótkie wnioski. Artykuł jest interesujący i daje dobry przegląd zagadnień związanych z identyfikacją, charakterystyką, ewolucją i znaczeniem funkcjonalnych retrogenów w genomach zwierząt. Odnoszę wrażenie, iż angielski jest w niektórych miejscach chropawy.

Publikacja 2 to praca oryginalna, dotycząca powstania intronów w retrogenach oraz alternatywnego składania transkryptów retrogenów zawierających introny. Praca opiera się na bioinformatycznej analizie porównawczej genomów, zawiera jednak również – co zasługuje na podkreślenie – elementy analizy eksperymentalnej przeprowadzonej metodami biologii molekularnej. Doktorantka skupiła się na analizach bioinformatycznych. Praca dotyczy dwu genów, *RNF113* i *DCAF12*. Pierwszy z nich występuje u ssaków tylko w formie bezegzonowej, a więc stanowi prawdopodobnie przykład zastąpienia genu rodzicielskiego przez retrogen. W obu tych genach u niektórych ssaków doszło do intronizacji sekwencji egzonowej poprzez mutacje wytwarzające miejsca donorowe i akceptorowe. Uzyskanie nowych intronów jest bardzo rzadkim zjawiskiem u kręgowców. Powstanie intronów doprowadziło w obu przypadkach do alternatywnego składania, przy czym forma z wyciętym intronem wykazuje w obu przypadkach niską oraz bardziej tkankowo-specyficzną ekspresję.

Zwraca uwagę, podobnie jak w przypadku wielu innych nowopowstałych retrogenów, ekspresja w jądrach.

Pierwsze zdanie podrozdziału "Evolution of introns" jest mylące, gdyż sugeruje, że retrogen występuje u wszystkich kręgowców, podczas gdy chodzi tu (jak rozumiem) o białko kodowane przez gen *RNF113A*. Nie jestem też przekonany, że termin "exaptation" został użyty poprawnie. Ważna jest zasygnalizowana w konkluzjach możliwość, że retrogeny z intronami mogą być częstsze niż przypuszczano, lecz po prostu nie szukano ich w odpowiedni sposób, uznając, że – niejako z definicji – retrogeny intronów nie mają.

O ile w przypadku publikacji 2 można by powiedzieć, że stanowi ona interesujący opis genomowej "ciekawostki", publikacja 3 ma moim zdaniem zasadnicze znaczenie dla genomiki porównawczej. Dotyczy ona identyfikacji "osieroconych" retrogenów (dalej OR) w genomie człowieka. Mianem OR określa się takie retrogeny, dla których nie można odnaleźć w genomie badanego organizmu wieloegzonowego genu rodzicielskiego. Omawiana publikacja jest pierwszą badającą systematycznie to zjawisko. Jak podkreślono we wstępie, oprócz znaczenia poznawczego w zakresie ewolucji genomów, badania te mogą mieć spore znaczenie biomedyczne. A to dlatego, że często uznaje się, iż badania na modelach zwierzęcych odzwierciedlają funkcjonowanie danego mechanizmu u człowieka. Jeżeli jednak sekwencje kodujące określonych genów u człowieka i zwierzęcego modelu są podobne, ale ich regulacja odbywa się w odmienny sposób, czego można się spodziewać w porównaniach genów rodzicielskich i OR, wnioski wyciągane z badań na modelach zwierzęcych będą nieadekwatne do sytuacji u człowieka, a nawet mylące. We wstępie omówiono również duplikacje jako źródło nowych genów oraz duże rozbieżności w oszacowaniach liczby funkcjonalnych retrogenów w genomach. Na koniec wstępu autorzy podkreślają, że dotychczasowe badania, oparte na założeniu, że rodzicielski gen wieloegzonowy oraz retrogen współwystępują w genomie, nie pozwalały na identyfikację OR.

Materiały i metody opisują metodykę badań, która oparta była na porównaniu zbioru ludzkich jednoegzonowych transkryptów do ludzkich transkryptów wieloegzonowych, kurzych transkryptów wieloegzonowych oraz sekwencji ortologów niciania i człowieka. Nie wyjaśniono, dlaczego wykorzystano tylko ok. 4600 mRNA kury. Na podstawie tych danych zidentyfikowano OR trzema metodami. Dwie pierwsze oparte były na porównaniu sekwencji człowieka i kury oraz kontekstu genomowego, w którym występują potencjalne OR. Kandydatami na OR były więc takie jednoegzonowe geny człowieka, które wykazywały większe podobieństwo sekwencji do wieloegzonowych genów kury niż do wieloegzonowych genów człowieka. Dodatkowym warunkiem było występowanie w ich sąsiedztwie w genomie człowieka innych genów niż w sąsiedztwie genów wieloegzonowych (w domyśle rodzicielskich) w genomie kury (konsekwencja retropozycji). Co ciekawe, jak wynika z Materiałów i Metod, stwierdzono takie pary jednoegzonowy gen ludzki-wieloegzonowy gen kury, które występowały w takim samym lub podobnym kontekście genomowym. Sugeruje to, iż możliwe są inne mechanizmy utraty intronów niż retropozycja. Zasygnalizowano to taką możliwość również w dyskusji. Szkoda, że w całej pracy doktorskiej nie ma informacji o znanych lub potencjalnych mechanizmach utraty intronów innych niż retropozycja. Trzecia metoda oparta była na identyfikacji w kolekcji ortologów człowiek-nicień takich genów, które

mają jeden egzon u człowieka, a co najmniej dwa u nicienia. Kandydaci zidentyfikowani przez zastosowanie omówionych wyżej metod zostali dalej zweryfikowani manualnie, celem odróżnienia rzeczywistych OR od fałszywie pozytywnych wyników. Aby sprawdzić, kiedy powstały zidentyfikowane w ten sposób "osierocone" retrogeny oraz kiedy doszło do utraty genów rodzicielskich, przeszukano genomy kilku gatunków kręgowców oraz *Drosophila melanogaster* pod kątem występowania OR oraz ich genów rodzicielskich. Dodatkowe analizy obejmowały określenie tkankowej specyficzności ekspresji zidentyfikowanych OR metodą PCR czasu rzeczywistego oraz identyfikację miejsc wiązania microRNA i czynników transkrypcyjnych, a także obliczenie stosunku tempa podstawień niesynonimowych do tempa podstawień synonimowych.

Zidentyfikowano łącznie 25 OR, z czego najwięcej wykryła metoda oparta na porównaniu ortologów człowieka i nicienia. Zwraca uwagę wysoka częstość wyników fałszywie pozytywnych w metodach opartych na porównaniu genów człowieka i kury. Te fałszywie pozytywne wyniki przypisano w większości błędom w anotacji genomu kury i występowaniu niezsekwencjonowanych lub źle złożonych regionów w tym genomie. Stwierdzono, iż większość OR powstała dawno, a do utraty genu rodzicielskiego doszło często na tej samej gałęzi ewolucyjnej. Trudno się jednak zgodzić ze stwierdzeniem o niemal natychmiastowej utracie genu rodzicielskiego w takiej sytuacji, gdyż długości gałęzi często wynoszą po kilkadziesiąt lub więcej milionów lat, tak więc analiza nie ma wystarczającej rozdzielczości, by uzasadnić takie stwierdzenie.

Wyniki omówione w podrozdziale "Disease Association" są moim zdaniem uspokajające w kontekście badań biomedycznych z wykorzystaniem modeli zwierzęcych, ponieważ sytuacje, w których potencjalne ortologi między gatunkami ssaków są tak naprawdę parami gen wieloegzonowy – "osierocony" retrogen są stosunkowo rzadkie. Stwierdzono wysoki stopień ewolucyjnej konserwacji OR. Wydaje mi się, że ta obserwacja może być konsekwencją metody identyfikacji OR. Otóż większość (16 z 25) OR zidentyfikowano w porównaniu człowiek-nicień, co sugeruje, że już na tym etapie OR były wzbogacone w geny o wolnym tempie ewolucji, gdyż oczekiwałbym, iż zestaw ortologów człowiek-nicień będzie zawierał przewagę wolno ewoluujących genów w stosunku do średniej ogólnogenomowej. Stąd nie jest niespodzianką obserwacja, że tempo ewolucji mierzone stosunkiem K_A/K_S tych genów między myszą a człowiekiem jest niższe niż średnie tempo w porównaniach wszystkich par ortologów mysz-człowiek. Kolejną częścią wyników jest szczegółowa analiza przypadku genu *CHMP1B*, dla którego znaleziono ślady genu rodzicielskiego w genomie człowieka. Analiza ekspresji wykazała, iż wszystkie zidentyfikowane OR charakteryzują się szerokim spektrum ekspresji, a w żadnym wypadku ich ekspresja nie jest ograniczona do jąder, co jest częstą obserwacją w przypadku funkcjonalnych retrogenów.

Dyskusja omawia znaczenie uzyskanych wyników w szerokim kontekście, podkreślając, iż zaprezentowana analiza jest pierwszą tego rodzaju, co czyni ją pionierską, lecz jednocześnie utrudnia interpretację ze względu na brak możliwości porównania z wynikami innych badań. Mam tutaj dwa komentarze, w których nie staram się podważyć wniosków autorów, a tylko sugeruję możliwość innego rozłożenia akcentów. Otóż po pierwsze, jak autorzy słusznie zauważają, OR są wyjątkowe wśród funkcjonalnych pseudogenów, gdyż w genomie brak jest

genu rodzicielskiego. Dlatego funkcja genu rodzicielskiego, jeżeli została zachowana, w całości przeszła na OR. Z tego powodu trudno się spodziewać u OR takich cech, jak ograniczona do jąder ekspresja czy osłabienie doboru oczyszczającego. Po drugie, wnioski dla stosowania kręgowców, a szczególnie ssaków jako modeli w biomedycynie, jakie wynikają z przedstawionych wyników, są moim zdaniem krzepiące (patrz też wyżej), gdyż sytuacje, w których retrogen zastąpił gen rodzicielski, są bardzo rzadkie, mimo wysokiego tempa retropozycji u ssaków. Większość takich przypadków zaszła na dawnych etapach ewolucji zwierząt, a przynajmniej przed powstaniem owodniowców.

W rozprawie brakuje odniesienia do debat na temat pochodzenia, pojawiania się i utraty intronów w kontekście ewolucji genomów. Myślę, że tego typu rozważania, na które raczej nie ma miejsca w pracach oryginalnych, mogły znaleźć się w artykule przeglądowym lub przynajmniej w streszczeniu.

Jak widać z treści recenzji, nie znalazłem w ocenianej pracy wielu uchybień, a większość uwag dotyczy interpretacji i ma charakter polemiczny. Nie powinno to dziwić, gdyż w przypadku doktoratów składających się z opublikowanych prac, przeszły one wcześniej przez proces rygorystycznych recenzji przeprowadzanych przez specjalistów. Nie zwalnia to oczywiście recenzenta od wnikliwego zapoznania się z publikacjami i wychwycenia ewentualnych błędów, lecz w omawianym przypadku z przyjemnością stwierdzam, że nie było ich wiele. Podsumowując, uważam że oceniana rozprawa jest przykładem badań z zakresu genomiki porównawczej na bardzo dobrym międzynarodowym poziomie. Prace wchodzące w jej skład mają szansę na stałe wejść do światowej literatury w tej dziedzinie.

Stwierdzam, że przedstawiona do recenzji praca spełnia wymagania stawiane rozprawom doktorskim przez ustawę z dnia 14 marca 2003 o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki (wraz z późniejszymi zmianami) i wnioskuję o dopuszczenie Pani mgr Joanny Ciomborowskiej do dalszych etapów przewodu doktorskiego; jednocześnie wnioskuję do Rady Wydziału Biologii UAM o nagrodzenie pracy.

Wiesław Babik