

Dr hab. Robert Rutkowski, prof. MiIZ PAN

Zastępca Dyrektora ds. Naukowych

Muzeum i Instytut Zoologii PAN

00-679 Warszawa

Ul. Wilcza 64

Warszawa, 11.12.2023

Recenzja rozprawy doktorskiej mgr inż. Karola Puchały

Uwagi ogólne

Do recenzji przedstawiono rozprawę doktorską pt. 'Kompleksowa analiza zmienności genetycznej polskiej populacji sokoła wędrownego (*Falco peregrinus*)', autorstwa mgr inż. Karola Puchały. Praca została wykonana pod kierunkiem prof. dr hab. Wandy Olech-Piaseckiej w Katedrze Genetyki i Ochrony Zwierząt SGGW w Warszawie.

Praca składa się z dwóch zasadniczych części: (i) obszernego, polskojęzycznego streszczenia pracy doktorskiej oraz, (ii) dwóch publikacji w czasopiśmie *Genes*. Część obszernego streszczenia składa się z wyraźnie wyodrębnionych rozdziałów: **wstępu**, w którym jeden podrozdział poświęcony jest badanemu gatunkowi, a drugi markerom mikrosatelitarnym; **opisowi celów i zakresu pracy; sformułowaniu hipotezie badawczej; opisowi materiałów i metod**, oraz **syntetycznemu omówieniu dwóch publikacji**. Wszystko spięte jest rozdziałem **'Podsumowania i wnioski'**. Całość, wraz ze stroną tytułową, oświadczeniami, spisem treści i wykazem cytowanej literatury, obejmuje 34 strony.

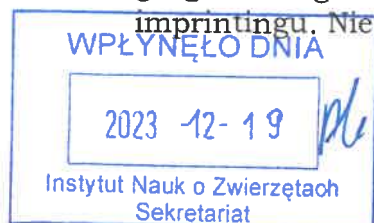
Załącznikami do powyższego tekstu są dwie publikacje oraz oświadczenia współautorów o wkładzie w powstanie dwóch publikacji naukowych, które – jak rozumiem - stanowią podstawę dysertacji.

Praca doktorska opiera się na analizie 465 osobników sokoła wędrownego (prawdopodobnie w znacznej części spokrewnionych) w 10 loci mikrosatelitarnych, z czego jedno locus, ze względu na brak zidentyfikowanych heterozygot, wyeliminowano z niektórych analiz (o czym dowiadujemy się dopiero z publikacji II), co jest bardzo słuszne, natomiast uwzględniano ten marker w ocenie zmienności genetycznej (publikacja I), co jest poważnym błędem.

Część pierwsza – obszerne streszczenie

Wykonana jest bardzo niestarannie. Mnóstwo jest niezręczności językowych, co oczywiście nie może decydować o zawartości merytorycznej i wartościowości badań, ale jednak świadczy o zaangażowaniu doktoranta.

Polskojęzyczny wstęp jest chaotyczny, doktorant przeskakuje od rozmieszczenia geograficznego do różnic dymorficznych, wplatając w to kwestię podgatunków i imprintingu. Nieco płynniej zaczyna wyglądać część poświęcona problemowi DDT i



ginięciu sokoła wędrownego w XX wieku. Niemniej, całość nie prezentuje się na miarę rozprawy doktorskiej, a nawet syntetycznego streszczenia osiągnięć opublikowanych w czasopismach naukowych. Jednym z przykładów są tabele, składające się z dwóch kolumn - wydaje mi się, że tę część danych można było ująć w o wiele bardziej obrazowe przedstawienie liczby osobników wsiedlanych w poszczególnych latach. Pytanie też, na ile podawanie dokładnych danych w każdym roku na przestrzeni 11 lat wnosi coś istotnego do odczytu wyników i ich interpretacji. Owszem, zbiorcza tabela pokazująca liczbę wsiedlanych ptaków i zbadanych ptaków z podziałem na hodowle, liczby spokrewnionych osobników itd. byłaby bardzo potrzebna. Brakuje tu zaangażowania i wyobraźni, a doktorant powinien tym nadrobić bardzo słabą rozprawę doktorską. To samo dotyczy Tabeli 3 – są to dane techniczne, mało ważne z punktu widzenia interpretacji wyników, a zajmują prawie pół strony.

Podrozdział o mikrosatelitach, moim zdaniem, jest zbędny. Od ponad trzech dekad są to podstawowe markery genetyczne i nie ma już potrzeby podkreślania ich znaczenia i możliwości wykorzystywania w badaniach z zakresu genetyki populacyjnej. To że były wykorzystywane w badaniach sokołów, jak i innych ptaków drapieżnych jest oczywiste. Oczywiście, fajna byłaby jakaś synteza tych danych, pokazująca generalne trendy, na przykład różnorodności allelicznej lub heterozygotyczności w różnych populacjach, ale tego niestety próżno szukać w tej dysertacji.

Wymieniony 'Cel i zakres pracy' to określenie zmienności genetycznej polskiej populacji sokoła wędrownego. W tytule jest „kompleksowa analiza...”. Jako kompleksową rozumiałbym analizę zmienności genetycznej w jej różnych frakcjach, tymczasem w dysertacji mamy raptem do czynienia z określeniem polimorfizmu w 10 (9?) loci mikrosatelitarnych, czyli bardzo ograniczonej liczbie najpewniej neutralnych selekcyjnie loci w genomie. Zaproponowany tytuł absolutnie nie przystaje do zawartości dysertacji.

Nie wiadomo, co stoi za stwierdzeniem 'Określenie (czasami pada stwierdzenie 'ocena') zmienności genetycznej...'. Z dalszej części tekstu wynika, że jest to oszacowanie polimorfizmu badanych markerów mikrosatelitarnych u kilkuset (w tym osobników blisko-spokrewnionych) ptaków.

Na stronie 18 przedstawiono główne hipotezy badawcze: (i) Niska zmienność genetyczna populacji; (ii) Dobór hodowlany może zwiększyć zmienność genetyczną. Niestety, hipotezy nie są ani klarowne ani nie zostały zweryfikowane.

Właściwie jedynym ciekawym stwierdzeniem, opisanym dokładniej w pierwszej publikacji – dziwie się, że cała pierwsza praca nie została skoncentrowana wokół tego tematu – jest wykazanie różnic między dwoma ekotypami sokoła wędrownego. Wydaje się, że uzyskane wyniki sugerują klarowny podział badanych osobników na dwie grupy genetyczne, odpowiadające miejscom gniazdowania. Niestety, z niewiadomych względów, nie podporządkowano temu schematu analiz. Po pierwsze, osobniki przypisane do ekotypów powinny być, A PRIORI, pogrupowane w dwie grupy genetyczne i na tych grupach powinny być prowadzone analizy, np. wyznaczone różnicowanie genetyczne, choćby za pomocą Fst (do tego wróć jeszcze później omawiając tzw. 'indeks fiksacji'). Niestety, tę ciekawą obserwację dyskredytuje nieco

fakt, że z opisanej metodyki (w publikacji I, bo nie jest to opisane w streszczeniu) wynika, że analiza STRUCTURE została przeprowadzona w oparciu o spokrewnione osobniki (wszystkie pisklęta z kilku gniazd). Jest to bardzo poważny błąd, który mógł zaważyć na wyniku.

Rysunek 4 jest niezrozumiały. Jak mniemam, celem analizy było wykazanie różnic między dwoma ekotypami. Czemu więc nie stworzono dwóch grup i nie przeprowadzono na nich porównania. Wtedy byłoby wyraźnie widać, że dwa ekotypy są przypisane do odrębnych grup genetycznych. Obecnie z pokazanych wyników to zupełnie nie wynika. Obiecujące wydaje się drzewo dystansów genetycznych, choć oczywiście stosowanie tego typu analizy przy mikrosatelitach, badanych w obrębie jednego gatunku, jest bardzo dyskusyjne. Z pewnością w analizach ekotypów powinno zastosować się kilka innych metod, choćby standardowe Fst czy AMOVA – uwzględniające podział geograficzny i ekotypowy.

Pewne nieścisłości językowe: ‘Indeks fiksacji’ – sprawdziłem słowo ‘fiksacja’ w słownikach. Fiksacja określa zaburzenie psychiczne. Nie rozumiem, czemu użyto tego terminu. Już w tej chwili jest kilka polski opracowań z zakresu genetyki populacyjnej, gdzie wprowadzono podstawowe polskie słownictwo dotyczące wskaźników zmienności genetycznej. Przypuszczam, że indeks fiksacji jest kalką językową z angielskiego, i dotyczy ‘*fixation indices*’, komponentów statystyki F Wrighta. Problem w tym, oprócz błędnego tłumaczenia, bo *fixation index* należy tłumaczyć raczej jako ‘indeks utrwalenia’, że moim zdaniem pomyłono tu dwie miary Wrighta, gdyż *fixation index* to Fst, pokazujące poziom zróżnicowania genetycznego między populacjami, natomiast doktorant najwyraźniej używa tego terminu do określania Fis, czyli inbreeding coefficient – współczynnika/wskaźnika inbredu, zazwyczaj wykorzystywanego do oszacowania poziomu odchylenia heterozygotyczności oczekiwanej od obserwowanej (za *Population Genetics Glossary*). Rozróżnienie komponentów statystyki F Wrighta powinno być wyraźnie zaznaczone w dysertacji, gdyż są one stosowane wymiennie.

Publikacja I.

Oświadczenia o współautorstwie wskazują na bardzo znaczny udział doktoranta w powstaniu pracy.

Praca opublikowana w piśmie Genes. Na temat jakości pism MDPI i etyki publikowania w nich oczywiście trwa obecnie gorąca dyskusja, ale w tej recenzji odcinam się zupełnie od tych rozważań. Faktem jest, że to co miało być przedmiotem dysertacji zostało opublikowane w czasopiśmie naukowym o znacznym wskaźniku IF. Natomiast czytając tekst trudno oprzeć się wrażeniu, że redaktor prowadzący przeoczył mnóstwo niedociągnięć, natomiast co skłoniło recenzentów do napisania pozytywnej recenzji – tego nie wiem, a na pewno takiej decyzji nie rozumiem.

Wstęp

Nie jestem ‘native speaker’ i wolałbym uniknąć dywagacji dotyczących poprawności języka, niemniej uważam, że publikacja napisana jest bardzo złym angielskim. Dotyczy to nie tylko wstępu, ale całej publikacji.

Schemat wstępu oceniam jako poprawny. Natomiast zaskakujące są pewne niedociągnięcia, świadczące – ponownie - o dość niestarannym podejściu do publikowanego tekstu. Przykładów jest wiele, podam jeden: ostatni akapit wstępu i dwa identyczne zdania powtórzone: „Some peregrine falcon populations...” w pierwszym zdaniu; i pod koniec akapitu – identyczne zdanie. Poza tym, ostatni akapit miał być prawdopodobnie poświęcony przeglądowi badań genetycznych nad sokołem wędrownym, i tak się dzieje, aż do ostatniego zdania, gdy do akapitu wskakuje pustułka... Trudno dostrzec tu jakąkolwiek koncepcję. Nie widzę żadnego związku między cytowanymi wcześniej pracami a genetycznym zróżnicowaniem między miejskimi i pozamiejskimi populacjami pustułki – zupełnie innego gatunku. Bardzo rażącym błędem jest także pominięcie w tego typu publikacji wyników badań Clytona White'a i współpracowników, dotyczących genetycznych relacji między podgatunkami sokoła wędrownego (White et al. 2013, *The Auk*, <https://doi.org/10.1525/auk.2012.11173>). Moim zdaniem, w tej pracy cała dyskusja powinna się zacząć właśnie od omówienia tych wyników. Podobnie jeśli chodzi o streszczeniową część dysertacji (strony 7–35) ale to już pomijam, bo moja ocena tej części jest bardzo krytyczna. Liczyłem, że poziom publikacji zrekompensuje niedociągnięcia w podsumowaniu ogólnym. Niestety, tak nie jest.

Materiał i metody, Wyniki (najważniejsze uwagi)

Mapa – ukazuje miejsce zbioru prób z populacji nazywanej 'dziką' na tle rozmieszczenia na terenie Polski główny cieków wodnych. Dlaczego akurat to pokazano na mapie? Czy nie bardziej właściwa byłaby mapa pokazująca rozmieszczenie terenów zalesionych i zurbanizowanych? Po raz kolejny pojawia się wątpliwość, dlaczego redakcji czasopisma (recenzentom) taka mapa nie przeszkadzała? Dla mnie jest ona mało czytelna i nic nie wnosząca.

Materiał: w przypadku populacji dzikich wiadomo, że 91 prób pochodziło od ptaków z 19 gniazd, czyli znaczna część osobników porównywanych genetycznie była spokrewniona. To ma istotny wpływ na wyniki analiz, np. w analizie STRUCTURE czy w ocenie odchylenia od równowagi Hardy'ego-Weinberga (H-E). Tak naprawdę, do porównania ekotypu miejskiego i leśnego powinno być brane pod uwagę tylko 19 osobników, czyli po jednym z gniazda. Ile było osobników niespokrewnionych z grupy hodowlanej – nie wiadomo, nie są podane informacje na ten temat. Ale najpewniej też było dużo osobników spokrewnionych, uwzględnionych w analizach populacyjnych, co zaburzyło wyniki, i zakłamuje rzeczywistość. W związku z tym, jedyny ciekawy wynik przedstawiony w dysertacji – różnice genetyczne między grupą gniazdującą na drzewach i na budynkach - też nie ma solidnych podstaw metodycznych. Jestem bardzo zdziwiony, że czasopismo naukowe zdecydowało się opublikować takie dane.

Analizy: PCA powinno być zrobione w programie ADEGENET (środowisko R) i pokazywać nie rozproszenie grup, a rozproszenie osobników na wykresie. Pokazywanie tego wyniku, na wykresie generowanym przez Excel, nie jest na poziomie rozprawy doktorskiej. Nie jest także zrozumiałe, dlaczego doktorant wybrał sędziwy dystans Nei do prowadzenia analiz zróżnicowania genetycznego między populacjami. Obecnie, od wielu lat zresztą, standardowo wykorzystuje się F_{st} Wrighta i jego standaryzowane wartości, także R_{st} ... Niestety mam wrażenie, że doktorant opanował tylko analizę w programie GENALEX, natomiast nie sięgnął do bardziej zaawansowanych, i bardziej adekwatnych, narzędzi analitycznych. Zarówno w publikacji, jak i w części streszczeniowej, razi także niezrozumienie istoty

statystyki F Wrighta. Ogólnie, brakuje mi zastosowania w analizach jakiegokolwiek programu z pakietów funkcjonujących w środowisku R.

Testowanie równowagi Hardy'ego-Weinberga (HWE): Według metodyki zastosowano podstawowe testy (należy nadmienić, że standardem jest stosowanie testów w programie GenePopOnTheWeb), natomiast w wynikach nie mogłem znaleźć żadnej wzmianki o wynikach tej analizy. Ile loci było w odchyleniu/równowadze? Jak wygląda ogólny test dla całej badanej populacji? Czy odchylenia w poszczególnych loci są istotne statystycznie? Tych informacji w pracy nie ma. Jest to ważne z punktu widzenia choćby wyników analizy w STRUCTURE. W tabeli 1 w ostatniej kolumnie pokazano F_{st} . W jaki sposób wyliczono ten wskaźnik? W metodyce nie jest to opisane. W opisie tabeli F_{st} jest opisane jako 'F-statistics' bez podania odpowiednich cytacji, więc nie wiadomo, której statystyki F dotyczy ta wartość. Obawiam się, że nie jest to F_{st} tylko F_{is} . Ponownie, bardzo dziwię się, że recenzenci nie zwrócili na to uwagi i nie poprosili o uszczegółowienie.

Podsumowując: opis metod jest niewystarczający, a przedstawienie wyników nieprzystające do poziomu rozprawy doktorskiej. Na przykład, czy opisano i przedyskutowano później wartości 'effective number of alleles' (tabela 1)? Gdzie pokazane są wyniki 'pairwise F_{st} '? W metodyce jest napisane, że taką analizę przeprowadzono. Powinna być tabela pokazująca wyniki dla par porównywanych grup. Takich przykładów można wymienić znacznie więcej.

Dyskusja

Znaczna część poświęcona jest porównaniu liczby alleli w poszczególnych loci, zidentyfikowanych w różnych pracach o sokole wędrownym. Jest to oczywiście jakiś sposób, ale zazwyczaj stosuje się wartości średnie, takie jak różnorodność alleliczna (allelic diversity) czy *allelic richness*. Tę część dyskusji oceniam negatywnie. Brakuje mi tabel, zestawień, które by lepiej obrazowały uzyskane wyniki, nie wspominając o testach statystycznych. Podsumowując – zupełnie nie wiadomo, jaki jest wniosek z badania zmienności genetycznej w grupie osobników sokoła wędrownego, będących obiektem analiz doktoranta.

Podrozdział dyskusji: *Bird Release Point* - znaczna część dyskusji jest dla mnie niezrozumiała lub bardzo kontrowersyjna w świetle faktów naukowych (argumentacja byłaby bardzo obszerna). Przedstawione wnioski nie są poparte uzyskanymi wynikami. Przykładem jest fragment poświęcony korelacji między dystansem genetycznym i, jak przypuszczam – geograficznym, bo w publikacji jest opisany tylko jako 'distance of the studied nests'. Według metodyki żadna analiza korelacji dystansu genetycznego i geograficznego nie została wykonana. Powinien być wykonany chociażby Test Mantela. Poza tym, doktorant postuluje, że według jego wyników grupa gniazdująca na drzewach ma niższą zmienność genetyczną niż grupa gniazdująca na budynkach, czyli na obszarach miejskich. Byłoby to niezmiernie ciekawe, tylko że nie jest poparte żadnymi wynikami uzyskanymi w tej dysertacji. Gdzie porównano te dwie grupy pod względem zmienności genetycznej? Jaki test zastosowano?

Dalsza część dyskusji: zwracam uwagę, że podziękowania i uwydatnianie wkładu organizacji typu „Stowarzyszenie na Rzecz dzikich Zwierząt” nie powinno mieć miejsca w dyskusji, tylko w podziękowaniach.

Ogólnie, **publikację I** oceniam jako słabą merytorycznie, nie wnoszącą nic istotnego do obecnej wiedzy, i o wnioskowaniu nie popartym wynikami zaprezentowanych badań.

Publikacja II

Oświadczenia o współautorstwie wskazują na bardzo znaczny udział doktoranta w powstaniu pracy.

Podobnie jak pierwsza publikacja, także i ta została opublikowana w Genes.

Tytuł sugeruje, że oceniany będzie wpływ przebiegu reintrodukcji na hodowlę sokoła wędrownego. Jest to zaskakujące, choć – należy podkreślić – bardzo ciekawe podejście do tematu. Wykazania takich zależności niewątpliwie miałyby ogromny wpływ na planowanie reintrodukcji wielu gatunków.

Wstęp

Rozpoczyna się niezwykle obiecująco od historii sokolnictwa i metod oraz osiągnięć hodowli sokołów. Niestety ten wątek – jak najbardziej pasujący do tytułu pracy – zostaje bardzo szybko zarzucony. Natomiast podane są dokładne liczebności osobników z hodowli wsiedlanych do środowiska naturalnego w poszczególnych latach. Nie rozumiem, po co podawane są aż tak szczegółowe dane. Jest to trudne w odbiorze dla czytelnika, a przede wszystkim mało wnoszące do późniejszego odbioru wyników. Moim zdaniem powinien być opisany jakiś ogólny trend, który od razu zasygnalizuje konieczność podjęcia tematu badań. Wstęp jest ogólnie bardzo krótki. Jest to zaskakujące, biorąc pod uwagę podjęcie niezwykle trudnego tematu. Brakuje mi wprowadzenia w wyniki badań podobnego typu, tzn. wykazania, że przebieg reintrodukcji może rzeczywiście wpływać na to, co dzieje się w hodowlach. Spora (proporcjonalnie do całej objętości, bo ogólnie, jak wspominałem, ta część jest bardzo krótka) część wstępu jest poświęcona imprintingowi i ekotypom sokoła wędrownego, co nie ma żadnego związku z podjętą tematyką.

Na zakończenie rozdziału proponowany jest **cel pracy**: analiza zmian w strukturze genetycznej osobników (*the genetic structure of individuals*) na przestrzeni lat. Przyznam, że pojęcie struktury genetycznej osobników jest dla mnie zaskakujące. Zazwyczaj w badaniach genetyki populacyjnej mamy do czynienia ze strukturą genetyczną populacji, czyli podziałem populacji na podjednostki, wyróżnione na podstawie poziomu przepływu genów. Podobnie jak w poprzedniej pracy, także i w publikacji II wyznaczony cel nie został osiągnięty w następstwie przeprowadzonych analiz.

Materiał i metody

W obydwu pracach procedury laboratoryjne są opisane w sposób odpowiedni. Dowiadujemy się przede wszystkim, że w porównaniu do materiału z poprzedniej pracy, liczba osobników z hodowli w różnych krajach została nieco zwiększona. Nie rozumiem po co wymieniany jest podział hodowców, typu 'breeder 3' czy 'breeder 1'. Nie wnosi to nic istotnego.

W przypadku opisu analizy markerów mikrosatelitarnych zostały podane pewne sprzeczne informacje (podrozdział 2.2. Microsatellite Genotyping). Otóż do analiz wybrano markery charakteryzujące się wysokim polimorfizmem na podstawie

wcześniejszych badań doktoranta, natomiast w opisie metodyki tej pracy okazuje się, że jeden z markerów (NVHfp79_4) zostaje wykluczony ze względu na zerową heterozygotyczność obserwowaną. Jest to oczywiście słuszne, ale chciałbym zaznaczyć, że zerowa heterozygotyczność w tym locus była już wykazana w publikacji I (Genes 2021). Należało więc od razu wykluczyć ten marker z analiz i nie pisać w streszczeniu, że badano 10 mikrosatelitów.

Analizy statystyczne

Ponownie, tak jak w publikacji I, główne analizy zostały przeprowadzone w programie GenAlEx. W pracy zaznaczono, że oszacowano spokrewnienie i poziom zimbredowania osobników. Niestety, w wynikach próżno szukać tych danych. Ponownie pada sformułowanie, że wyliczono 'indeks fiksacji' – fixation index – nie wiadomo, o jaka miarę z zakresu genetyki populacyjnej chodzi, nie ma odpowiedniej cytacji. Te uwagi umieściłem już we wcześniejszej części recenzji. Nie wiadomo, czy mamy do czynienia z Fis czy Fst... Jeśli chodzi o analizy 'relatedness' i 'inbreeding' – gdzie są wyniki tych analiz? Gdzie opisano ich metodykę? Moim zdaniem ani jednego, ani drugiego w pracy nie ma. Do wyliczania spokrewnienia między osobnikami na podstawie markerów mikrosatelitarnych służy parametr *r-relatedness coefficient*, który można oszacować w wielu programach i wieloma różnymi algorytmami. Także w programie GenAlEx. Spokrewnienie powinno być oszacowane między poszczególnymi osobnikami. Podobnie ze współczynnikiem zimbredowania. Według mnie powinien on być przedstawiony jako poziom homozygotyczności poszczególnych osobników.

Analizy w STRUCTURE – ponownie, jak w poprzedniej publikacji, nie jest wyjaśnione, w jaki sposób uwzględniono spokrewnienie osobników? Ten program opiera się na identyfikacji najbardziej wiarygodnej liczby grup genetycznych w badanej puli osobników poprzez optymalizację równowagi Hardy'ego-Weinberga i sprzężeń między loci. Jeśli od razu w analizie uwzględniamy osobniki spokrewnione (pisklęta z jednego gniazda) to bardzo mocno wpływamy (negatywnie) na jakość wyników.

Wizualizacja wyników w programie Excel to nie jest poziom rozprawy doktorskiej. Domyślam się, że chodzi o wykresy 3 i 5 (Fig 3, Fig 5 w publikacji). Po pierwsze, jak wspominałem wcześniej w aspekcie Fis/Fst, nie wiadomo co one przedstawiają. Co przedstawiają zakresy na tych wykresach? Czy chodzi o odchylenie standardowe? Jak je wyliczano? Uwzględniano różnice między osobnikami w danym roku? Dlaczego postępowano w ten sposób? Dlaczego nigdzie nie jest to opisane? Z opisu wykresu nie wynika, co przedstawiają poszczególne linie i punkty.

W analizach statystycznych brakuje podstawowych informacji: porównanie liczby alleli w poszczególnych latach z korektą na różnice w wielkości próby, test statystyczny na równowagę Hardy'ego-Weinberga, testowanie statystycznej istotności współczynnika inbredu (jeśli takowy rzeczywiście wyliczono).

Wyniki

Na początku pojawia się informacja o 10 loci, choć już z metodyki, i z poprzedniej pracy, wiadomo, że jedno locus jest nieprzydatne w analizach. Natomiast najbardziej zaskakująca informacja pojawia się w trzecim zdaniu – dowiadujemy się, że wśród blisko 350 osobników tylko 85 (zaledwie 25% badanej próby) miało unikatowe genotypy w badanych loci. O czym to świadczy – wiele osobników z

badanej grupy sokołów wędrownych jest nierozróżnialnych genetycznie na podstawie zastosowanej metodyki. W analizach multiplikowane są więc takie samego genotypy, ponieważ dany genotyp może posiadać kilka niezależnych osobników. Jest to bardzo poważny problem, sprawiający, że analizy w STRUCTURE, szacunki dystansu genetycznego, a przede wszystkim obliczanie współczynnika zimbredowania (czy testowanie HWE) nie ma żadnego sensu. Jest to problem mający swoje podstawy w niedociągnięciach z pierwszej publikacji. To w badaniach pilotowych powinno wyliczyć się wskaźniki polimorfizmu dla poszczególnych loci i kombinacji różnych loci, które określiłyby prawdopodobieństwo rozróżnienia poszczególnych osobników na podstawie wybranego panelu markerów mikrosatelitarnych. Mam tu na myśli indeksy polimorficzności czy prawdopodobieństwo wykluczenia. Jest to na tyle istotny błąd, że dalsza dyskusja dotycząca wyników właściwie jest bez sensu. W pracy w ogóle nie przedyskutowano kwestii identycznych genotypów. Gdzie się one pojawiały? W tych samych lęgach? W tych samych hodowlach? Ogólnie – brak informacji dotyczących liczby osobników spokrewnionych w badanej próbie sprawia, że trudno jest ocenić uzyskane wyniki. Doktorant powinien być świadomy wagi takich informacji.

Tabela 1 przedstawia podstawowe dane dla badanych loci mikrosatelitarnych. Nie są porównane żadne grupy ptaków, czy to z różnych hodowli, czy też z różnych lat wsiedlania. Nie przeprowadzono testu na istotność odchylenia heterozygotyczności oczekiwanej od obserwowanej. Nie podano jaką metodę dokładnie wykorzystano do testowania obecności alleli zerowych. Jeśli w tabeli 1 w ostatniej kolumnie (F) podano wartości Fis (ich istotność również powinna być przetestowana, na przykład w programie Fstat), to nie wierzę, że Cervus nie wykazał obecności alleli zerowych. Dodatkowo – do analiz powinien być wykorzystany program MicroChecker, lub PopGenReport w środowisku R – szczególnie ten drugi oferuje szeroką gamę różnych testów na obecność alleli zerowych w zależności od różnych czynników, wpływających na charakterystykę próby.

Analiza STRUCTURE – stwierdzono, że najbardziej prawidłowy podział, sugerowany przez metodę ΔK to dwie grupy genetyczne. Rzeczywiście tak wynika z Fig 1. Natomiast na wykresie widzimy także 'piki' przy $K = 7$ i $K = 12$. Te podziały również powinny być przedyskutowane. Najpewniej nie odzwierciedlają one rzeczywistości i wynikają z uwzględnienia w analizie osobników spokrewnionych i zastosowania małej liczby markerów mikrosatelitarnych. Poza tym – podział na dwa klady jest bardzo często sugerowany jako najbardziej prawdopodobny w metodzie ΔK w przypadku schematu izolacji-przez-dystans, jak i w innych modelach zróżnicowania genetycznego populacji. Zarówno w prezentacji wyników, jak i w dyskusji, ta kwestia powinna być uwzględniona.

Zupełnie niezrozumiałe jest dlaczego w przypadku jednego z hodowców, nazywanego breeder 6, przedstawiono wynik dla trzech grup genetycznych. Owszem, w metodyce zaznaczono, że oddzielna analiza została przeprowadzona dla pojedynczego hodowcy, natomiast nie jest to w żaden sposób uzasadnione. Nie wiadomo więc o co chodzi w tej analizie, ani jak interpretować wynik. Nie są też pokazane wartości ΔK dla tego hodowcy, więc trudno ocenić czy rzeczywście prezentacja dla $K = 3$ jest właściwa.

Fig 3 i 5 – pokazują zmiany wartości 'fixation index' w różnych latach. W tekście, opisując wykres, zaznaczono, że najwyższą wartość zaobserwowano w 2016 roku. Niestety, akurat tego roku na wykresie nie ma. Ponieważ doktorant wiąże ten wskaźnik z inbredem należy przypuszczać, że chodzi o Fis. Moim zdaniem jest to duże

uogólnienie. Fis przedstawia wielkość odchylenia heterozygotyczności obserwowanej od oczekiwanej. Oczywiście, jednym z powodów takiego odchylenia może być kojarzenie się osobników w bliskim pokrewieństwie, ale jest to tylko jedna z możliwych przyczyn. Inną doktorant też wymienia – jest nią niska liczebność próby, przez co nie można właściwie ocenić heterozygotyczności obserwowanej w populacji. Natomiast nie można w prosty sposób interpretować wskaźnika Fis jako inbredu. Ciekawa wydaje się obserwacja przedstawiona w Fig 5 natomiast brakuje jednej ważnej informacji – jakie były wielkości próby (czyli liczby osobników od poszczególnych dostawców), z ilu lęgów pochodziły osobniki od poszczególnych dostawców, a przede wszystkim jaka była liczba osobników niespokrewnionych genetycznie. Jeśli w analizie Fis wykorzystywane są osobniki o identycznych genotypach to wskaźnik ten musi osiągać wysoką wartość. W związku z tym wyliczenia dla takiej próby nie mają sensu. Już właściwsze, moim zdaniem, byłoby porównanie obserwowanej heterozygotyczności w różnych latach, ale tak naprawdę powinno się określać heterozygotyczność osobniczą dla ptaków z poszczególnych hodowli i to mogłoby dać rzeczywiście jakieś wyobrażenie o zmianach w poziomie spokrewnienia. Natomiast należy zastrzec, że nawet przybliżona ocena zimbredowania danego osobnika na podstawie kilku loci mikrosatelitarnych jest niezwykle trudna, jeśli w ogóle możliwa. W dyskusji powinno to być uwzględnione.

Individuals from the six main providers – najprawdopodobniej ten ustęp ma sens, natomiast trudno ocenić uzyskane wyniki ponieważ nie podano w pracy ile dokładnie osobników było badanych z poszczególnych ‘main providers’. Dodatkowo, z tekstu wynika, że 5 osobników w ogóle nie zgenotypowano. Nie wyjaśniono, co się z nimi stało w poprzednich analizach (dla wszystkich dostawców). Poza tym, jeśli jest to prawda, to nie zgadzają się wartości w tabeli 2, w kolumnie N.

Ogólnie opis wyników STRUCTURE w obu pracach wskazuje, że doktorant nie rozumie na czym polega przeprowadzana analiza. Używanie stwierdzeń typu: ‘individuals were assigned to group’, czy ‘the genetic affiliation to groups’ jest błędne w przypadku STRUCTURE – w analizie tej określana jest wiarygodność pochodzenia danego osobnika z ‘K’ grup genetycznych. Do wykonywania testów przypisania (assignment tests) służą inne programy (np. GeneClass) i ta analiza polega zupełnie na czym innym. Nie jestem przekonany, że doktorant rozumie działanie stosowanych programów – STRUCTURE HARVESTER jest programem pomocniczym, ułatwiającym porządkowanie i obrazowanie olbrzymiej ilości danych, generowanych przez STRUCTURE. Przy okazji należy wspomnieć o niewłaściwym generowaniu wykresów pokazujących wyniki STRUCTURE. Uporządkowane dane symulacyjne ze STRUCTURE HARVESTER powinny być obrazowane za pomocą Distruct lub podobnego programu. Dopiero wtedy wykres przedstawia w odpowiedni sposób (graficznie) wyniki z kilku powtórzeń analizy w STRUCTURE. W przypadku obu prac nie wiadomo jak wygenerowane zostały wykresy, nie wiadomo także co opisuje oś rzędnych. Podejrzewam, że użyto tu wyniku pojedynczej symulacji w STRUCTURE. Rzeczywiście, zaobserwowana zmiana w pochodzeniu osobników jest interesująca, ale na podstawie opisanej metodyki trudno potwierdzić wiarygodność tych wyników. W dyskusji ta kwestia nie jest wyjaśniona. Z zaprezentowanych wyników można wyciągnąć wniosek, że w roku 2013/2014 hodowcy pozbyli się dotychczasowych par rozrodczych i sprowadzili zupełnie odrębną linię genetyczną. I nie dotyczy to pojedynczej hodowli, ale, jak rozumiem, wszystkich hodowców dostarczających osobniki do reintrodukcji w danych latach. Byłoby to bardzo dziwne.

Dyskusja

Bardzo trudno komentować ten rozdział, wiedząc, że w analizach nie zastosowano odpowiednich metod, ani nie podano ważnych informacji dotyczących badanych prób. W pierwszych akapitach opisowo porównano liczbę alleli w loci z wynikami z poprzedniej pracy oraz innymi pracami na sokołach wędrownych. Ponieważ w pracy nie zastosowano żadnych metod statystycznych, by ocenić istotność zaobserwowanych różnic, trudno komentować ten fragment dyskusji. Podobnie postępuje doktorant, nie odnosząc się w żaden sposób do opisanych danych.

Zaskoczył mnie czwarty akapit. Doktorant stwierdza, że hodowlana populacja sokoła wędrownego wykazuje oznaki zmienności genetycznej ('signes of genetic diversity'). Niestety, zdanie jest dla mnie zupełnie niezrozumiałe. O co chodzi? Zmienność genetyczna (genetic diversity) jest zawsze obecna w populacjach, czy to hodowlanych czy naturalnych. Owszem, może być różna – wysoka, niska itd. W następnym zdaniu pojawia się nowy wskaźnik – F-ratio, który nie był opisany ani w metodyce, ani w wynikach. Można się tylko domyślać, o co chodzi. Podana jest informacja, że na Fig. 3 umieszczono dane o odchyleniu standardowym (powinno to być opisane w metodyce i wynikach) natomiast jest dla mnie niezrozumiałe, jak na tej podstawie wyciągnięto wnioski o bardzo zróżnicowanym pochodzeniu badanych osobników.

Nie wiadomo też na jakiej podstawie doktorant stwierdza, że jego wyniki świadczą, że hodowcy chętnie pozyskują ptaki z zagranicznych hodowli. Zastanawia mnie, czy takie informacje nie są dostępne u hodowców. Wydaje mi się, że w przypadku sokoła wędrownego hodowcy powinni posiadać certyfikaty pochodzenia. Na tej podstawie powinno być wiadome pochodzenie par rodzicielskich, a co za tym idzie genetyczna kompozycja potomstwa. Przyznaje, że nie sprawdziłem tej informacji. Być może powinna się ona pojawić, w którejś z publikacji.

Niestety w dyskusji jest jeszcze więcej podobnego typu nieścisłości i nie popartych uzyskanymi danymi stwierdzeń. Na przykład w konkluzjach doktorant stwierdza, że na podstawie jego wyników, można stwierdzić, że dzięki działaniom hodowców powstała reintrodukowana populacja sokoła, charakteryzująca się dobrymi parametrami genetycznymi. Problem w tym, że doktorant nie badał populacji reintrodukowanej, lecz wsiedlane ptaki. Czy przetrwały, przystąpiły do rozrodu i utworzyły populację? Takich informacji w dysertacji ani w publikacjach nie ma. Moim zdaniem z opisanej metodyki i wyników nie można wyciągnąć takiego wniosku. Po drugie: co oznacza stwierdzenie: „dobre parametry genetyczne”? Ani wyniki, ani dyskusja nie przekonują do takiego stwierdzenia.

Ogólnie, **publikację I** oceniam jako słabą merytorycznie, nie wnoszącą nic istotnego do obecnej wiedzy, i o wnioskowaniu nie popartym wynikami zaprezentowanych badań.

Podsumowanie recenzji

Praca wykonana w dużej mierze samodzielnie, o czym świadczą oświadczenia współautorów. Nie wiem, jaki był udział doktoranta w zbieraniu materiału. Prawdopodobnie doktorant opanował w stopniu zadowalającym procedury laboratoryjne: izolacja DNA, analiza markerów mikrosatelitarnych. Natomiast nie opanował metod analizy wyników. Stosowanych jest 5 bardzo podstawowych

programów do analizy danych mikrosatelitarnych. Z testu prac wynika, że doktorant nie rozumie, czym są wyznaczane wskaźniki, oraz jakie jest sedno analizy, na przykład w programie STRUCTURE. Wyniki nie są zaprezentowane w czytelnych sposób. Brakuje szacowania istotności statystycznej różnic poszczególnych parametrów. W opisie wyników doktorant posługuje się ogólnikami. W pracy naukowej takie stwierdzenia powinny być poparte wynikami analiz statystycznych. Nie zastosowano podstawowych testów, np. żeby określić istotność odchylenia od równowagi H-E, czy istotności współczynnika Fis. W metodyce nie podano bardzo ważnych danych dotyczących liczebności prób od poszczególnych hodowców, liczby lęgów, liczby spokrewnionych/niespokrewnionych osobników. Dyskusje w obu pracach nie są przeprowadzone na poziomie rozprawy doktorskiej. Większość wniosków jest nieuprawniona na podstawie zastosowanej metodyki i uzyskanych wyników.

W związku z tym stwierdzam, że przedstawiona mi do recenzji rozprawa mgr inż. Karola Puchały nie spełnia warunków określonych w art. 187 ust. 1 i 2 Ustawy z dnia 20 lipca 2018 r. Prawo o szkolnictwie wyższym i nauce (Dz. U. 2018 poz. 1668 z późn. zm. Tekst jednolity Dz. U. 2023, poz. 742) oraz §9 Regulaminu przeprowadzania postępowań w sprawie nadania stopnia doktora Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie (Uchwała Senatu nr 89 – 2022/2023 z dnia 26.06.2023), a co za tym idzie, wnioskuję o niedopuszczanie mgr inż. Karola Puchały do dalszych etapów przewodu doktorskiego.

Zastępca Dyrektora ds. Naukowych

dr hab. Robert Rutkowski

