

1 STRESZCZENIE

Geny głównego układu zgodności tkankowej (MHC) zajmują istotne miejsce w badaniach z pogranicza immunologii i biologii ewolucyjnej. MHC jest ważnym elementem układu odpornościowego i posiada szereg cech, dzięki którym od lat cieszy się dużym zainteresowaniem naukowców. MHC jest efektywną bronią wszystkich kręgowców przeciwko patogenom, gdyż koduje cząsteczki prezentujące antygeny. Jest to duża rodzina genowa, obejmująca wysoce polimorficzne klasyczne geny MHC. Geny te stanowią wzorcowy przykład genów ewoluujących pod działaniem doboru równoważącego, tzn. takiego, który utrzymuje zmienność w populacji. W teorii geny takie powinny ulegać również łatwiejszej introgresji międzygatunkowej, która wprowadza nowe allele do puli genetycznej gatunku, gdyż takie allele, jako rzadkie, powinny mieć przewagę selekcyjną. Adaptacyjna introgresja jest bardzo ciekawym, lecz trudnym przedmiotem badań, gdyż pozostawia w genomie podobne wzory zmienności jak niektóre inne procesy, na przykład ancestralny polimorfizm. Kluczowe w takim przypadku jest znalezienie systemu gatunków, który pozwoli na ich rozróżnienie. Jednym z takich systemów są badane w niniejszym projekcie traszki z kompleksu *Lissotriton vulgaris*.

Kompleks *Lissotriton vulgaris* obejmuje traszkę karpacką (*Lissotriton montandoni*; Lm) oraz 8 linii ewolucyjnych grupowanych do niedawna pod nazwą traszki zwyczajnej (*L. vulgaris*; Lv), z których 4 ostatnio zyskały rangę osobnych gatunków. Lm i Lv są grupami siostrzanymi, które rozdzieliły się jeszcze przed plejstocenem. Traszka karpacka jest genetycznie dość jednorodnym gatunkiem zamieszkującym jedynie Karpaty i wschodnią część Sudetów. Linie ewolucyjne Lv można podzielić na podstawie rozmieszczenia geograficznego na 2 podgrupy: północne Lv (Nvl), zajmujące duży obszar środkowej i północnej Europy oraz południowe Lv (Svl), obejmujące stare, izolowane genetycznie linie, które zamieszkują Bałkany i Półwysep Anatolijski.

Celem pracy w pierwszej kolejności była charakterystyka genów MHC klasy I i II u traszek z kompleksu *L. vulgaris*. Wykazano bardzo dużą zmienność w obu klasach genów. Dla uzyskanych 1383 alleli MHC klasy I i 271 klasy II przeprowadzono analizy zmienności i testy doboru. Analizowanymi sekwencjami był egzon 2, który (w MHC klasy I wraz z egzonom 3) jest najbardziej zmienny, gdyż koduje miejsca wiążące antygen. Stwierdzono wysoką zmienność niesynonimową w miejscach wiążących antygen i wykryto silny sygnał doboru pozytywnego. Dla alleli MHC klasy I przeprowadzono również analizę ekspresji na poziomie transkrypcji, dzięki czemu zakwalifikowano je do dwóch kategorii: o wysokiej (HEX) oraz o niskiej ekspresji (LEX). Te drugie mogą świadczyć o tym, że są to allele нефункционалне lub nieklasyczne, dlatego część analiz została przeprowadzona osobno dla alleli HEX i LEX. Drzewa filogenetyczne zawierające dodatkowe allele pochodzące od innych gatunków sugerują polimorfizm ponadgatunkowy zarówno dla MHC klasy I jak i II. Dla obu klas opisano także pełne sekwencje kodujące pochodzące od dwóch osobników traszki zwyczajnej wraz z identyfikacją poszczególnych domen oraz miejsc konserwatywnych. Sekwencje umieszczone na drzewie filogenetycznym wraz z allelami pochodzącymi od kilku innych gatunków płazów odzwierciedliły ich filogenezę, sugerując, że linie alleliczne i ewentualne duplikacje genów MHC powstały już po podziale tych taksonów.

Ze względu na wysoki polimorfizm i duplikacje, geny MHC nie są łatwe do analizy. Osobniki pochodzące nawet z tej samej populacji mogą posiadać różne liczby genów, a allele z różnych genów mogą być do siebie bardziej podobne niż do innych alleli z tego samego *locus*, dlatego często nie jest możliwe nawet przypisanie alleli do *loci*. Nie da się więc określić liczby genów, którą posiada dany osobnik. W takiej sytuacji można jednak określić przekazywane przez rodziców potomstwu haplotypy oraz minimalną liczbę genów na danym haplocyfie. W niniejszej

pracy, na podstawie analizy segregacji w dwóch rodzinach określono haplotypy MHC. W ten sposób scharakteryzowano 8 haplotypów, które zawierały od 5 do 9 alleli MHC klasy I oraz od 1 do 3 alleli MHC klasy II. Geny MHC zostały również umieszczone na mapie genetycznej – znajdują się one na końcu grupy sprzężeniowej (LG) 1. Ponieważ nie stwierdzono występowania rekombinantów między MHC klasy I i II, uznano że obie klasy genów są ze sobą silnie sprzężone.

W drugiej części pracy przetestowano hipotezę o łatwiejszej introgresji genów MHC klasy I w porównaniu do innych genów w genomie w dużej skali geograficznej. Test opierał się na analizie zróżnicowania genów MHC klasy I między 35 populacjami z 3 grup geograficznych: Nvul, Svul i Lm, oraz porównaniu ze 154 innymi, losowymi genami. Oczekiwano, że jeśli MHC ulega łatwiejszej introgresji, to najmniejsze zróżnicowanie powinna wykazać para Nvul-Lm, większe para Nvul-Svul i największe Lm-Svul. W przypadku innych genów ta kolejność powinna być odwrócona i odzwierciedlać pokrewieństwo grup, czyli $Nvul-Svul < Nvul-Lm < Svul-Lm$. I rzeczywiście wyniki badań potwierdziły, że w genach MHC zróżnicowanie jest najmniejsze tam, gdzie spodziewano się największego przepływu genów, a zatem dla pary Nvul-Lm, a nie między grupami najbliższej spokrewnionymi (Nvul-Svul). Analiza ta pozwoliła na pośrednie wnioskowanie o tempie introgresji genów MHC klasy I w stosunku do średniej ogólnogenomowej.

Badania nad introgresją w dużej skali geograficznej i czasowej mają swoje zalety, ale też ograniczenia i charakter wnioskowania pośredniego. Nie pozwalają stwierdzić, czy introgresja była wyjątkowym, epizodycznym procesem, którego konsekwencje obserwujemy obecnie, czy też występuje regularnie lub ciągle. Dlatego też przeprowadzono dokładną analizę introgresji genów MHC klasy I i II w dwóch strefach mieszańcowych (IN i OUT, położone, odpowiednio wewnątrz Basenu Karpackiego i na zewnątrz łuku Karpat, w każdej z nich przebadano dwa transekty) *L. vulgaris* i *L. montandoni* na obszarze Karpat. Badania stref mieszańcowych pozwalają badać konsekwencje obecnej lub niedawnej introgresji i jej wpływ na skład genetyczny populacji. W tym przypadku oczekiwano, że jeśli introgresja MHC jest szybsza/silniejsza niż innych genów, to dla genów MHC wewnątrzgatunkowe zróżnicowanie między transektami będzie wyższe niż międzygatunkowe zróżnicowanie wewnątrz transektu. Dla tak wysokiego podobieństwa genetycznego w strefie kontaktu ciężko znaleźć inne wytłumaczenie niż introgresja. Analizy wykonane wcześniej na > 1000 losowych genach kodujących stanowiących średnią genomową wykazały bardzo silne zróżnicowanie międzygatunkowe. Natomiast w przypadku genów MHC analizy PCA oraz współdzielenia alleli w strefie IN pokazały, że zróżnicowanie między osobnikami z różnych gatunków w obrębie transektu było zdecydowanie niższe niż między osobnikami tego samego gatunku z różnych transektów. W strefie OUT uzyskano inne wyniki, zróżnicowanie międzygatunkowe było silniejsze w obu transektach. Tam introgresja MHC była wykrywalna tylko w teście współdzielenia alleli w syntopii (transekt S), oraz w syntopii i parapatryi (transekt T). Wyniki te sugerują, że w strefie OUT introgresja genów MHC jest słabsza, co z kolei może wynikać z tego, że strefa kontaktu jest tam młodsza niż w strefie IN. Analiza klin nie wykazała istotnych różnic między genami MHC a resztą genomu.

Silna introgresja MHC w obliczu silnej izolacji całego genomu sugeruje proces adaptacyjny. Hybrydyzacja i introgresja mogą mieć więc kluczowe znaczenie dla długoterminowego utrzymania zmienności oraz stanowić potężny mechanizm jej przywracania. Introgresja jako mechanizm utrzymujący zmienność adaptacyjną w genach pod doborem równoważącym może zwiększać potencjał adaptacyjny hybrydujących taksonów, co może mieć duże znaczenie w ochronie zagrożonych gatunków.



2 SUMMARY

Genes of the major histocompatibility complex (MHC) occupy an important place in research combining immunology and evolutionary biology. MHC is an important element of the immune system and has a number of features that make it of interest for many researchers. MHC is an effective weapon of vertebrates against pathogens because it encodes molecules presenting antigen to the immune T cells. MHC is a large gene family, including highly polymorphic classical MHC genes. These genes are a model example of genes evolving under balancing selection, i.e. the mode of selection that maintains variation in a population. In theory, such genes should also undergo easier interspecies introgression, which introduces new alleles into the genetic pool of the species, since such alleles should have a selective advantage. Adaptive introgression is a very interesting but difficult subject of study, as it leaves similar patterns of variation in the genome as some other processes. The key in this case is to find a species system that will allow them to be distinguished. One of such systems are the newts from the *Lissotriton vulgaris* complex studied in this project.

The *Lissotriton vulgaris* complex includes the Carpathian newt (*Lissotriton montandoni*; Lm) and 8 evolutionary lineages known until recently as one species – smooth newt (*L. vulgaris*; Lv), of which four have recently been elevated to the rank of separate species. Lm and Lv are sister groups that split before the Pleistocene. Carpathian newt is a genetically homogeneous species inhabiting only the Carpathians and the eastern part of the Sudetes. The evolutionary lineages of Lv can be divided based on geographical distribution into 2 subgroups: northern Lv (Nvl), inhabiting a large area of central and northern Europe, and southern Lv (Svul), old, genetically isolated lineages that inhabit the Balkans and the Anatolian Peninsula.

The first aim of the study was to characterize MHC class I and II genes in newts from *L. vulgaris* complex. High variation was demonstrated in genes of both MHC classes. For 1383 MHC class I and 271 class II alleles, variation was analysed and selection tests were performed. They were based on sequences of the 2nd exon, which (in MHC class I together with 3rd exon) is the most variable, because it encodes antigen binding sites. High non-synonymous variation was found at the antigen binding sites and strong signal of positive selection was detected. To estimate expression level of MHC class I alleles, analysis of transcriptome sequences was performed. Alleles were classified into two categories: high (HEX) and low expression (LEX). The latter may indicate that they are non-functional or non-classical alleles, therefore some of the analyzes were carried out separately for the HEX and LEX alleles. Phylogenetic trees containing additional alleles from other species suggest transspecies polymorphism for both MHC class I and II. For both classes, complete coding sequences from two individuals of the smooth newt were described, along with the identification of domains and conservative sites. The sequences placed on the phylogenetic tree along with alleles from several other amphibian species reflected their phylogenesis, suggesting that allelic lineages and possible MHC gene duplications arose after the splitting of these taxa.

Due to their polymorphism and duplications, MHC genes are not easy to analyze. Even individuals from the same population may have different numbers of genes, and alleles from different genes may be more similar to each other than to other alleles from the same *locus*, so it is often not possible to assign alleles to *loci*. In consequence it is usually impossible to determine the number of genes per individual. In this situation, however, one can determine haplotypes passed by parents to offspring and the minimum number of genes on a given haplotype. In this work, MHC haplotypes were determined based on segregation analysis in two families. Eight haplotypes containing from 5 to 9 MHC class I alleles and from 1 to 3 MHC class II alleles were



characterized. Also the position of MHC genes on the genetic map was determined - they are situated at the end of the linkage group 1 (LG1). The complete lack of recombinants in > 200 meioses indicates tight linkage of both MHC classes.

In the second part of the thesis, the hypothesis of easier introgression of MHC class I genes compared to other genes in the genome was tested at a large geographical scale. The test was based on the analysis of class I MHC gene differentiation between 35 populations from 3 geographical groups: Nvul, Svul and Lm, and a comparison with 154 other random genes. It was expected that if MHC introgresses more easily, then the lowest differentiation should occur in the Nvul-Lm pair, the larger in the Nvul-Svul pair and the largest between Lm and Svul. For other genes, this order should be reversed and reflect the relationship of the groups, i.e. Nvul-Svul < Nvul-Lm < Svul-Lm. Indeed, the results confirmed that in MHC genes, differentiation is the smallest where the highest gene flow was expected, and therefore for the Nvul-Lm pair, and not between the closely related groups (Nvul-Svul). This analysis supports stronger introgression of MHC class I genes relative to the genome average.

Research on introgression at a large geographical and temporal scale has its advantages, but also the limitations and the inference is indirect. It is not possible to determine whether introgression was a unique, episodic process, the consequences of which are currently observed, or whether it occurs regularly or continuously. Therefore, a direct analysis of introgression of MHC class I and II genes was carried out in two hybrid zones (IN and OUT, located within the Carpathian Basin and outside the range of the Carpathians, respectively, two transects were tested in each) between *L. vulgaris* and *L. montandoni*. Studies of hybrid zones allow to study the consequences of current or recent introgression and its impact on the genetic composition of the population. In this case, it was expected that if MHC introgression is faster/stronger than that of other genes, then for MHC genes intraspecific variation between transects will be higher than interspecific variation within transect. For such high genetic similarity in the contact zone it is difficult to find an explanation other than introgression. Analyses performed earlier on > 1000 random coding genes representing the genomic average showed very strong interspecies differentiation. However, in the case of MHC genes, Principal Component Analysis and allele sharing in the IN zone showed that the differentiation between individuals from different species within the transect was lower than between individuals of the same species from different transects. Other results were obtained in the OUT zone, where interspecies differentiation was stronger in both transects. MHC introgression in the OUT zone was only detectable as increased allele sharing in syntopy (S transect), or in syntopy and parapatry (T transect). These results suggest that in the OUT zone, the introgression of MHC genes is weaker, which in turn may be due to the fact that the contact zone is younger there than in the IN zone. Cline analysis showed no significant differences between the MHC genes and the rest of the genome. Clines similarity may be apparent, however, because their origin may be different in both cases, so comparing them directly may not be justified.

Strong MHC introgression in the face of strong isolation of the entire genome suggests an adaptative process. Hybridization and introgression can therefore be crucial for long-term maintenance of variability and provide a powerful mechanism to restore it. Introgression as a mechanism that maintains adaptive variation in genes under balancing selection can increase the adaptive potential of hybridizing taxa, which can be important in conservation of endangered species.

