

**Udział polimorfizmu liczby kopii w kształtowaniu wewnątrzgatunkowej
zmienności strukturalnej metabolicznych klastrów genów
u *Arabidopsis thaliana***

mgr inż. Małgorzata Marszałek-Zeńczak

Sekwencja genomowego DNA jest charakterystyczną cechą każdego żywego organizmu. Nawet genomy indywidualnych osobników w obrębie gatunku różnią się między sobą. Zmienność genetyczna obejmuje różnorodne warianty, od zmian pojedynczych nukleotydów (SNP) aż po duże zmiany strukturalne (SV). Niezbalansowane SV, w których fragment sekwencji DNA jest tracony lub powielony, nazywane są zmiennością liczby kopii (CNV). Większość CNV nie ma znaczącego wpływu na fenotyp osobnika. Jednak niektóre z nich mogą mieć wyraźny - szkodliwy efekt, np. związany z rozwinięciem się choroby, podczas gdy inne ujawniają się lepszą adaptacją osobnika do warunków środowiska. U eukariontów geny zaangażowane we wspólny szlak metaboliczny są zazwyczaj rozproszone w genomie. Jednak w ostatnich latach zidentyfikowano metaboliczne klastry genów (MGC) czyli skupiska niehomologicznych genów zaangażowanych we wspólny szlak metaboliczny. Zarówno CNV jak i MGC często występują w dynamicznych regionach genomu takich jak okolice centromerów czy regiony bogate w transpozony. Oba te zjawiska, choć słabo poznane, wydają się mieć istotny wpływ na kształtowanie genomów roślinnych. W regionach podatnych na rearanżacje strukturalne możliwość połączenia korzystnych zestawów genów jest większa niż w pozostałej części genomu, promując tym samym tworzenie MGC. CNV i MGC mają wiele elementów wspólnych i wydaje się, że udział CNV w kształtowaniu i ewolucji MGC jest duży i znaczący.

Celem prowadzonych badań w ramach mojej rozprawy doktorskiej było określenie wewnątrzgatunkowej zmienności liczby kopii dla rośliny modelowej *Arabidopsis thaliana*, a także analiza, w jakim stopniu CNV wpływają na strukturę i stabilność metabolicznych klastrów genów. Wykorzystując dane z wysokoprzepustowego sekwencjonowania nowej generacji dla ponad 1000 naturalnych linii *A. thaliana*, opracowałam autorskie podejście do integracji wyników z siedmiu programów opartych o trzy główne metody detekcji CNV: analizę głębokości pokrycia, analizę mapowania odczytów sparowanych i analizę mapowania odczytów rozłącznych. Stworzyłam katalog dużych indeli (50-499 pz) oraz wariantów CNV (≥ 500 pz). Wykazałam, że CNV są ważnymi markerami, które mogą być wykorzystywane w analizach populacyjnych i asocjacji z fenotypem. Następnie skupiłam się na zbadaniu zmienności strukturalnej (głównie CNV) w czterech MGC, w populacji *A. thaliana*. Zaobserwowałam ogromną różnorodność w obrębie badanych klastrów. Klaster genów marneralu wydaje się ustalony na poziomie gatunku. Klaster biosyntezy thalianolu istnieje w dwóch wersjach. W tym MGC zidentyfikowałam inwersję, obecną aż w 65% badanej populacji, która skutkowałą bardziej zwartym klastrem centralnym. Kompaktowa wersja klastra biosyntezy thalianolu była dominująca i bardziej konserwatywna niż wersja nieciągła. Największy, wysoce zmienny i zróżnicowany w populacji jest klaster biosyntezy arabidiolu/baruolu. W tym klastrze zidentyfikowałam dużą (21-27 kpz) insercję genomową obecną w około jednej trzeciej analizowanej populacji. Wewnątrz insercji występuje nowa para genów: *CYP705A2a-BARS2*, przy czym *BARS2* stanowi niereferencyjny gen kodujący nową, dotychczas niescharakteryzowaną syntazę oksydoskwalenu (OSC) w genomie *A. thaliana*. Analiza asocjacji z fenotypem (GWAS) wskazała, że linie z insercją miały

wolniejszą dynamikę wzrostu korzeni i były związane z cieplejszym klimatem w porównaniu do linii z referencyjnym układem genów. W korzeniach i liściach, profil ekspresji genów w klastrze arabidiolu/baruolu był różny dla linii z insercją i bez. Dodatkowo, analiza par genów: OSC-oksydazy cytochromu P450 wykazała, że pary genów są bardziej zmienne niż ich niesparowane odpowiedniki.

Uzyskane wyniki badań podkreślają istotny wpływ zmienności genetycznej w formowaniu i kształtowaniu MGC. Duża różnorodność klastrów u *A. thaliana* wskazuje na ich dynamiczną ewolucję zaś wyniki GWAS potwierdzają ich możliwą rolę w różnorodności fenotypowej i adaptacji roślin. Zrozumienie zmienności genetycznej i jej wpływu na organizację genomu jest kluczowe w lepszym zrozumieniu ich funkcji.