

„IDENTYFIKACJA POLIMORFIZMU LICZBY KOPII DNA JAKO ISTOTNEGO SKŁADNIKA KSZTAŁTUJĄCEGO ZMIENNOŚĆ GENETYCZNĄ *ARABIDOPSIS THALIANA*”

Zmienność liczby kopii (CNV) polega na występowaniu dużych zmian o charakterze duplikacji lub delecji, różnicujących genomy osobników należących do jednego gatunku. Z uwagi na swoje duże znaczenie ewolucyjne oraz wpływ na kształtowanie zmienności fenotypowej, ten typ polimorfizmu jest intensywnie badany od kilkunastu lat. Pomimo to, informacje na temat CNV w genomie rzodkiewnika (*Arabidopsis thaliana*) były dotąd stosunkowo nieliczne, a międzynarodowa inicjatywa poznania zmienności strukturalnej genomu tej modelowej rośliny (Projekt 1001 Genomów) nie obejmowała CNV. Dlatego jako cel moich badań przyjąłem identyfikację i charakterystykę wariantów CNV w genomie rzodkiewnika oraz określenie, w jakim stopniu polimorfizm liczby kopii wpływa na organizację informacji genetycznej i czy może być powiązany ze zmiennością obserwowaną na poziomie fenotypu.

Do identyfikacji CNV wykorzystałem całogenomowe dane z wysokoprzepustowego sekwencjonowania ponad 1000 ekotypów rzodkiewnika. Poprawność stworzonego atlasu regionów CNV sprawdziłem poprzez liczne analizy porównawcze, zarówno w skali całej populacji jak i pojedynczych genomów. Zaowocowało to detekcją blisko 20 tysięcy wariantów CNV, którym nadałem zbiorczą nazwę wariantów AthCNV. Następnie skupiłem się na charakterystyce obszarów genomowych pokrywających się z regionami CNV. Pokazałem, że wzajemne ułożenie genów i transpozonów oraz stopień ich zmienności są ze sobą mocno powiązane. To wpływa na globalny rozkład tych elementów w genomie, który w przypadku genów jest powiązany również z ich funkcją. Obrazuje to jak współistniejące ze sobą i często przeciwnie działające mechanizmy wpływają na strukturę i kierunek ewolucji genomu.

Następnie połączyłem podejście bioinformatyczne oraz eksperymentalne, aby określić liczbę kopii ponad 27 tysięcy genów w 1060 ekotypach rzodkiewnika. Wyniki genotypowania zostały udostępnione społeczności naukowej poprzez stronę internetową (<http://athcnv.ibch.poznan.pl/>).

Ponadto podjęłem się oceny zasadności zastosowania wariantów CNV jako markerów w analizach genetycznych i populacyjnych, zwyczajowo bazujących na polimorfizmach pojedynczego nukleotydu (SNP). Pokazałem, że w porównaniu z danymi SNP, warianty CNV są cennym źródłem dodatkowych informacji na temat historii demograficznej rzodkiewnika, a także mogą służyć jako markery genetyczne w badaniach asocjacyjnych na skalę całego genomu (tzw. GWAS). Pokazałem również istnienie bezpośredniej zależności pomiędzy liczbą kopii genu, a jego ekspresją, a co za tym idzie – zmiennością tej ekspresji w populacji, obserwowaną poziomem transkryptu oraz białka, na przykładzie genu *SEC10*.

Równolegle zoptymalizowałem do badań na rzodkiewniku oraz wykonałem porównanie dwóch eksperymentalnych metod genotypowania CNV: emulsyjnego PCR oraz multipleksowej zależnej od ligacji amplifikacji sond (MLPA). Opracowałem szczegółowy protokół zastosowania MLPA do badania genów multiallelicznych (tj. występujących w wielu wariantach liczby kopii w populacji) u roślin. Scharakteryzowałem również złożony strukturalnie region CNV obejmujący geny *MSH2*, *AT3G18530* i *AT3G18535*, wykazując rolę nieallelicznej rekombinacji homologicznej w generowaniu duplikacji i delecji w jego obrębie.

Powyzsze wyniki, stanowiące moje osiągnięcie naukowe, opisałem w czterech wieloautorskich publikacjach, opublikowanych w latach 2014-2020 z których jedna jest pracą przeglądową, a pozostałe trzy przedstawiają wyniki oryginalnych badań. We wszystkich tych publikacjach jestem pierwszą i/lub korespondencyjną autorką. Sumaryczny pięcioletni wskaźnik Impact Factor tych publikacji wynosi 22,851, a ich łączna liczba cytowań wynosi 118 (113 bez cytowań własnych).