

W genomach eukariotycznych wydzielić można elementy takie jak geny kodujące białka, sekwencje telomerowe czy sekwencje centromerowe. Te ostatnie niezbędne są przy podziale komórkowym, gdyż to tam przyłączają się włókna wrzeciona podziałowego odpowiedzialne za prawidłowy podział materiału genetycznego. Sekwencje centromerowe często składają się z tysięcy tandemowo powtórzonych odcinków DNA, każdy z nich bardzo podobny do pozostałych. Ta właściwość uniemożliwiła poprawne odczytywanie centromerowych sekwencji powtórzonych, i mimo że Human Genome Project już w 2001 roku opublikował „pełną” sekwencję genomu ludzkiego, sekwencje centromerowe musiały czekać na uzupełnienie aż do 2022 roku, gdy nowe technologie umożliwiły odczytanie tych skomplikowanych regionów. Z tego powodu, centromery nadal pozostają „czarną dziurą” genomu, a ich badanie jest mocno ograniczone brakiem narzędzi i doświadczenia. Jest to jednocześnie niezwykle ważne zagadnienie, gdyż nieprawidłowości w sekwencjach centromerowych związane są z rozwojem niektórych nowotworów czy z nieprawidłowościami w liczbie chromosomów, jak przy zespole Downa. Poza znaczeniem medycznym, obserwowana zmienność w strukturze centromerów stanowi ważny punkt w badaniu wielu podstawowych dla biologii zjawisk jak wyodrębnianie nowych gatunków, a celowana modyfikacja centromerów może być zastosowana w produkcji nowych odmian w hodowli roślin.

Problemem nadal pozostaje analiza sekwencji centromerowych, podobnie jak problematyczne było samo uzyskanie tych sekwencji. Klasyczne metody analiz bioinformatycznych często zupełnie pomijają sekwencje powtórzone, gdyż nie były w tym celu projektowane, a same sekwencje powtórzone nie były do końca wiarygodne. Mimo że sekwencje i podstawowe analizy pojedynczych gatunków są już obecne, nadal brak odpowiedniej klasyfikacji i organizacji obserwowanej zmienności wśród organizmów eukariotycznych, a wiele potencjalnych mechanizmów funkcjonowania i ewolucji centromerów zaproponowanych było na podstawie niepełnych informacji na temat ich sekwencji. Niezbędne jest więc uporządkowanie rodzajów organizacji i analiza porównawcza sekwencji centromerowych oraz związane z nią analizy białek centromerowych czy zmian epigenetycznych charakterystycznych dla centromerów.

W tym projekcie, rozwinięte będą narzędzia bioinformatyczne umożliwiające analizę sekwencji centromerowych, mając za punkt wyjścia wcześniej opublikowane oprogramowanie do mapowania sekwencji powtórzonych, TRASH, które będą zastosowane do przeprowadzenia wielkoskalowej analizy sekwencji centromerowych. Sekwencje do analizy będą pochodzić z danych udostępnianych przez międzynarodowe konsorcja zajmujące się pozyskaniem wysokiej jakości pełnych sekwencji genomowych wielu gatunków.

Poza dostarczeniem istotnych wyników na temat funkcji i organizacji centromerów, oraz rozwiązań bioinformatycznych dla badaczy opisujących struktury genomowe, projekt ten może przyczynić się do lepszego zrozumienia patologicznych zjawisk zachodzących w procesach nowotworzenia czy segregacji chromosomów.