

## Streszczenie

Istotny wpływ na rozwój genomiki w ostatnich latach miały badania nad niekodującymi RNA (ncRNA), które zaangażowane są w regulację na niemal wszystkich etapach ekspresji genów, również na poziomie epigenetycznym. Do dwóch najbardziej licznych klas ncRNA zidentyfikowanych w roślinach zaliczają się małe interferujące RNA (siRNA) oraz mikroRNA (miRNA), które regulują ekspresję genów na poziomie transkrypcyjnym i post-transkrypcyjnym. Inne dwie dobrze opisane grupy ncRNA obejmują małe jądrowe RNA (snRNAs), które biorą udział w procesie składania RNA, oraz małe jąderkowe RNA (snoRNAs) i związane z nimi scaRNAs (small Cajal bodies RNA), głównie zaangażowane w modyfikację RNA. W oparciu o wysokoprzepustowe techniki sekwencjonowania i metody przewidywania *in silico* w ostatnich czasach odkryto szereg nowych ncRNA. Głównym wyzwaniem w chwili obecnej pozostaje jednak rozróżnienie biologicznie funkcjonalnych ncRNAs od tych powstałych w wyniku degradacji lub będących efektem szumu transkrypcyjnego.

Ekspresja genów regulowana jest przez tzw. cis-regulatorowe elementy (CRE), których specyficzna lokalizacja w genomowym DNA definiuje regiony promotorowe. Dlatego wykorzystanie konserwatywnych motywów CRE jako wskaźników aktywnie transkrybowanych genów powinno pozwolić na detekcję regionów promotorowych i tym samym identyfikację i adnotację funkcjonalnych cząsteczek ncRNA. Regiony promotorowe roślinnych genów snoRNA charakteryzują się obecnością dobrze zachowanych motywów regulatorowych typu Telo-box oraz typu Site II. Ponadto, prawie wszystkie sekwencje promotorowe genów snRNA zawierają konserwatywny motyw USE oraz kasetę TATA, które zostały także znalezione w genach kodujących U3, U13 oraz MRP (mitochondrial RNA processing complex). Cząsteczki U3, U13 i MRP uczestniczą w składaniu prekursorów rybosomalnego RNA i dlatego kombinacja motywów regulatorowych USE-TATA może być odpowiedzialna za koordynację ekspresji genów uczestniczących w składaniu transkryptów (splicing).

W celu identyfikacji nowych genów kodujących ncRNA, genom *Arabidopsis* został przeszukany przy pomocy konserwatywnych motywów Telo-SiteII i USE-TATA. Nowe, znalezione geny zostały zweryfikowane w oparciu o dostępne dane sekwencyjne na temat poziomu ekspresji oraz na drodze badań eksperymentalnych. W rezultacie zidentyfikowano 25 nowych snoRNA oraz 46 nowych ncRNA, które nie wykazują podobieństwa do żadnej opisanej do tej pory rodziny RNA. Ciekawym przykładem nowo odkrytych transkryptów są

geny dicystronowe, które kodują prekursorsy z których powstają dojrzałe cząsteczki zarówno snoRNA jak i miRNA. Podsumowując, przedstawione wyniki wskazują, że analiza oparta o identyfikację zachowanych regionów promotorowych jest skuteczną metodą eksploracji i poznawania nowych cząsteczek ncRNA.

**Słowa kluczowe:** *Telo-box*, *Site II*, *USE*, *TATA-box*, *snoRNA*, *miRNA*, *snRNA*, *Niekodujące RNA*, *Arabidopsis thaliana*, *Oryza sativa*