# SigHunt: Nová metoda pro hledání kandidátních genomických ostrovů v DNA sekvencích

Jaron K. S.1, Moravec J. C.2, Martínková N.1,2

1 Masarykova univerzita, Institut biostatistiky a analýz, Kamenice 3, 625 00 Brno, 376090@mail.muni.cz

2 Akademie věd ČR, Ústav biologie obratlovců, v.v.i., Květná 8, 603 65 Brno

Genomické ostrovy vznikají horizontálním přenosem genů a jsou v současnosti jedněmi z nejčastěji zkoumaných genomických úseku. Význam horizontálního přenosu genů na evoluci prokaryot byl v minulosti popsán, teprve však nedávno byli objeveny genomické ostrovy nesoucí klíčové geny pro životní styl eukaryotních organismů, jako například u plísní *Aspergillus* nebo *Pyrenophora.* Představujeme zde první metodu optimalizovanou pro detekci genomických ostrovů v eukaryotických genomech a nekompletních genomických sekvencích. SigHunt je založený na variabilitě tetranukleotidů mezi organismy, ale zároveň bere v úvahu variabilitu uvnitř sekvence. Lokální odchylky od domovské signatury SigHunt hodnotí podle toho, nakolik se frekvence tetranukleotidu v krátké sekvenci DNA odlišuje od lokální hustoty jeho frekvence. Metoda byla testována na sekvencích s uměle vloženými genomickými ostrovy v rámci široké škály organismů. SigHunt prokázal vysokou sensitivitu a specificitu určenou pomocí ROC (průměrná plocha pod křivkou byla 0.77). Při testování na biologických datech, sestávající z publikovaných a fylogeneticky verifikovaných ostrovů u rodů *Aspergillus*, *Saccharomyces*, *Pyrenophora*, *Galdieria* a *Cryptosporidium*, byl SigHunt schopen detekovat bezmála 70% popsaných genomických ostrovů v krátkém výpočetním čase. Tato metoda byla vytvořena s ohledem na velikost a heterogenitu eukaryotických genomů. Díky své výpočetní nenáročnosti umožňuje zpracovávat celé eukaryotické genomy na osobních počítačích a v relativně krátké době naleznout kandidátní genomické ostrovy pro následnou verifikaci.