

TISKOVÁ ZPRÁVA

Praha 3. dubna 2023

Akademie věd ČR
Národní 1009/3, 110 00 Praha 1
www.avcr.cz

PŘI SESTŘIHU GENETICKÉ INFORMACE SE EVOLUCE NEOPAKUJE, ALE MŮŽE SE RÝMOVAT, ZJISTILI VĚDCI S POMOCÍ STROJOVÉHO UČENÍ

Díky alternativnímu sestřihu lze podle jednoho genu vytvořit několik pozměněných verzí stejného proteinu. Tým vedený badateli z České republiky ale zjistil, že u rostlin během evoluce některých genů vznikají shodné varianty proteinů několikrát nezávisle na sobě. Získané poznatky a vyvinutý software využívající strojové učení pomohou objasnit některé záhady spojené s významem alternativního sestřihu nejen u rostlin, ale i u živočichů včetně člověka. Výsledky výzkumu publikoval významný odborný časopis *New Phytologist*.

„Historie se neopakuje, ale často se rýmuje“ – citát připisovaný Marku Twainovi platí také v evoluci, a to mnohdy ve velmi nečekaných případech.

Gen je vlastně úsek DNA, v němž jsou zapsány instrukce pro tvorbu jedné bílkoviny neboli proteinu. Má-li buňka daný protein vyrobit, přepíše se pokyny z jeho genu do molekuly zvané mRNA a podle ní se pak sestaví molekula proteinu. Informace obsažená v genu však zahrnuje také úseky, které tyto pokyny nenesou a je nutné je z mRNA odstranit takzvaným sestřihem. Ten musí být proveden bezchybně. Poruchy v sestřihu konkrétních genů totiž mohou být i příčinou vážných dědičných chorob, u člověka například cystické fibrózy nebo spinální svalové atrofie.

„Spící kočka“ v podání alternativního sestřihu

Například v DNA je zapsáno „mámasdimáfdomakdudspícíšdhbkočku“. Buňka z toho musí vybrat smysluplnou informaci „máma má doma spící kočku“. Některé geny však lze sestřihat více způsoby a vyrobit tak dva či více poněkud odlišných proteinů – to zajistí právě alternativní sestřih. Po sestřihu může vzniknout například zpráva: „máma má...kočku“, která téměř zachovává původní význam, zatímco varianta „mám...doma...kočku“ už je podstatně jiná informace.

„Prakticky všechny geny jsou zpracovávány nějakou formou alternativního sestřihu a nikdo vlastně neví proč. Podstatná část takových úprav totiž zřejmě nemá žádnou významnější funkci, a opět se netuší, jak velká. Proto jsme chtěli na vybraných genech prozkoumat, do jaké míry se u nich alternativní sestřih zachovává během evoluce. Když totiž něco v evoluci dlouhodobě přetrvá, pak se to zřejmě osvědčilo a je

Kontakt pro média: **Eliška Zvolánková**
Divize vnějších vztahů AV ČR
press@avcr.cz
+420 739 535 007

Jan Kolář
Ústav experimentální botaniky AV ČR
kolar@ueb.cas.cz
+420 608 557 328

to pro organismy užitečné," říká vedoucí vědeckého týmu Kamil Růžička z Ústavu experimentální botaniky Akademie věd ČR (ÚEB AV ČR).

Pravá ruka biologů: strojové učení

Aby badatelé mohli provést takovou rozsáhlou analýzu, museli se vypořádat se zásadním problémem. Ve veřejně přístupných databázích je k dispozici mnoho experimentálních dat o vzhledu sestříhaných mRNA i výsledných proteinů u různých organismů. V minulosti se již samozřejmě objevily studie, které analyzovaly evoluční aspekty alternativního sestříhu. Bylo však těžké získané údaje bez počítačové průpravy rutinně prohlížet. Bioložka Ksenia Timofeyenko proto vzala na pomoc informatika Dzmitryho Kanavalaua. Společně během dvou let vyvinuli nový software určený přesně pro tento úkol.

Využili metody strojového učení a na malém vzorku nejdříve počítačový algoritmus „natrénovali“. Použili navíc netradiční postup. Nezkoumali mRNA, ale zaměřili se přímo na proteiny vzniklé coby důsledek alternativního sestříhu konkrétních genů. Poté nechali svůj software analyzovat dostupná data.

„*K našemu dalšímu překvapení funguje náš algoritmus trénovaný na rostlinných datech velmi dobře i u živočichů.*”

„Výsledky nás hodně překvapily. Ukázalo se, že u většiny zkoumaných rostlinných genů se alternativní sestříh vyvinul během evoluce hned několikrát. Přestože je genetická informace v těchto případech sestříhána různými způsoby, výsledkem jsou shodné proteiny – ke stejnému cíli tedy vede více cest. K našemu dalšímu překvapení funguje náš algoritmus trénovaný na rostlinných datech velmi dobře i u živočichů. Zde jsme však obdobnou rozmanitost pozorovali v mnohem menší míře. To patrně souvisí s jinými mechanismy evoluce jejich DNA,“ říká Kamil Růžička.

„Tento algoritmus umožní snáze vytipovat zajímavé geny pro budoucí podrobnější výzkum. Software, který jsme vyvinuli, jsme navíc zpřístupnili široké vědecké komunitě. Mohou jej velmi snadno využít běžní biologové: nejen rostlinní, ale i ti, kteří studují živočichy. Doufáme proto, že pomůže zlepšit porozumění alternativnímu sestříhu, o kterém se zatím ví jen velmi málo. Kolegové s humorem říkají, že se jedná o první smysluplný vědecký výstup pocházející od spící kočky,“ uzavírá Kamil Růžička, narážející tak na název algoritmu Catsnap, tedy Kočičí dřímání.

Kamil Růžička vede pracovní skupinu v Laboratoři hormonálních regulací u rostlin v ÚEB AV ČR. Ksenia Timofeyenko působí v ÚEB AV ČR a zároveň dokončuje postgraduální studium na Masarykově univerzitě. Na projektu se podíleli také vědkyně a vědci z univerzity BOKU ve Vídni a z výzkumného centra CEITEC v Brně. Zpřístupnění algoritmu vědecké komunitě proběhlo ve spolupráci se sdružením CESNET.

Více informací: **dr. Kamil Růžička**
Ústav experimentální botaniky AV ČR
kamil.ruzicka@ueb.cas.cz
+420 773 902 647

Citace článku:

Ksenia Timofeyenko, Dzmitry Kanavalau, Panagiotis Alexiou, Maria Kalyna, Kamil Růžička: [Catsnap: a user-friendly algorithm for determining the conservation of protein variants reveals extensive parallelisms in the evolution of alternative splicing](#). New Phytologist (2023), doi: 10.1111/nph.18799.