

Válasz Sipiczki Mátyás, az MTA doktora, opponensi véleményében megfogalmazott kérdésekre

Köszönöm az értekezésem bírálatát. Az abban megfogalmazott kérdésekre az alábbi válaszokat adom.

1. Az INSDC-ben egymással összekapcsolt adatbázisok (melyekhez tartozik a szerző által használt GenBank is) egyike sem vizsgálja, hogy a behelyezett szekvenciák taxonómiai besorolása helytálló-e. A benyújtó által megadott fajnevet automatikusan elfogadják. A szekvenciák döntő többsége esetén szakirodalom sincs feltüntetve, amiben utána lehetne nézni, hogy mire alapul a besorolás. A szokásos gyakorlat az, hogy a behelyezésre kerülő szekvenciát olyan fajhoz sorolják, amely szekvenciájához leginkább hasonlít (ideális esetben azzal azonos). A megkívánt hasonlóság mértéke önkényes; akkora, amekkorát a benyújtó elegendőnek vél. Ahogy nő a behelyezett szekvenciák száma, úgy lesz egyre szélesebb egy faj szekvenciáinak tartománya, és egy idő után már átfed a rokon fajok egyikével vagy másikkal is. Így könnyen előfordulhat, hogy egy új izolátumot két vagy akár több fajhoz is be lehet sorolni a GenBank adatbázis alapján. Milyen lehetőséget lát a szerző az ebből származó hibák kiküszöbölésére? Találkozott-e ilyen problémával (ezen sorok írója igen), és amennyiben találkozott, azt hogyan oldotta meg?

Válasz: Sajnos Opponensemhez hasonlóan én is rendszeresen találkozom helytelen rendszertani besorolással felcímkézett DNS szekvenciákkal. Ennek egyik nyilvánvaló oka az, hogy nagyon sok nem taxonómus kutató helyez el szekvenciákat az INSDC-ben. A problémára részleges megoldást nyújthat, ha csak a típusörzsek DNS szekvencia adatait vesszük figyelembe. Némiképpen megkönnyíti a helyzetet, hogy a GenBank adatbázisban a szekvencia összehasonlítások esetén lehetőség van arra, hogy csak a típusanyagok szekvenciáival való összehasonlítást kérjünk, azonban a típusanyagok szekvenciáit tartalmazó adatbázis nem teljes. Segítséget jelenthet, az ellenőrzött (NCBI staff curated) szekvenciák használata is. Ez is sikeresen küszöböli ki a tévesen azonosított fajokat, de ez az adatbázis sem teljes. További problémát okoz, hogy a szekvenciák forrásául szolgáló mikroorganizmusok téves rendszertani besorolásán kívül, maguk a szekvenciák sem feltétlenül helyesek minden esetben. Még a kurátorok által ellenőrzött adatbázisban is „átcsúszik” a szűrőn néhány hibás szekvencia. A *Zygosaccharomyces favi* leírásakor pl. azzal szembesültünk, hogy az új fajhoz, nukleáris rRNS-ének nagy alegységét kódoló gén D1/D2 régiójának szekvencia hasonlósága alapján, legközelebb álló ismert faj, a *Z. gambellarensis*

típustörzsének két különböző D1/D2 szekvenciája található meg a GenBank-ban (JN874489 és FR725931). A két, különböző szerzők által deponált, szekvencia azonban nem kevesebb, mint 6 inszerciót/deléciót tartalmazó pozícióban különbözik egymástól. Itt a megoldást az jelentette, hogy beszereztük a *Z. gambellarensis* típustörzsét és a magunk által felszaporított D1/D2 régió bázissorrendjét határoztattuk meg. A *Cutaneotrichosporon suis* faj leírásakor az általunk izolált *C. suis* törzs D1/D2 szekvenciájával azonos szekvenciát találtunk a GenBank-ban, de azt a szekvencia deponálói *C. curvatum*-ból származónak gondolták. Ebben az esetben felvettük a kapcsolatot a szekvencia elhelyezőivel és a félreértés tisztázása után közösen írtuk le a *C. suis*-t.

Összefoglalva, a szekvenciák visszakeresését célszerű mindhárom (sztenderd, típusanyagok, kurátorok által ellenőrzött) adatbázisban elvégezni, és amennyiben lehetséges, a típustörzsek adatait figyelembe venni. Egy-egy típustörzsnek gyakran több azonos bare-code régiót képviselő szekvenciája is megtalálható, elsősorban a sztenderd adatbázisban. Ezek az esetek jelentős részében szerencsére azonosak, ha mégsem, célszerű az évtizedek óta ismert megbízható taxonómiai munkát végző műhelyek eredményeinek preferálása, abban az esetben, ha a szekvenciák sajátkezü ellenőrzésére nincs mód.

2. Az rRNS-eket kódoló kromoszómális szakaszok szekvenálása során időnként (egy fajoknál/fajcsoportoknál gyakran) előfordulnak di- és polimorf pozíciók. A szekvenciákban ezek N, Y, M, R stb betűk formájában jelennek meg. A jelenség mögött az áll, hogy a kromoszómában az rDNS egységek sokszor ismétlődnek, és az egyes egységek szekvenciái eltérnek egymástól (rossz hatékonyságú vagy elmaradó homogenizálás következtében). Az ilyen pozíciók nehezíti vagy akár tévessé is tehetik a taxonómiai és filogenetikai elemzéseket. Találkozott-e a szerző ezzel a problémával és ha találkozna, miként kezelné azt?

Válasz: Az előző kérdésre adott válasz során már említett *Zygosaccharomyces favi* leírásakor tapasztaltuk, hogy a *Z. favi* 5 vizsgált törzse esetén az ITS régiók különböző kópiái voltak jelen a törzsekben. A felszaporított ITS régiók bakteriális vektorba történt klónozását követő szekvenálás törzsenként 3-8 különböző ITS kópia jelenlétére derített fényt. Az egyes kópiák főleg a homopolimer régiók (poliA és poliT) hosszában tértek el egymástól, ami egyébként a PCR termék közvetlen szekvenálását lehetetlenné tette, mivel az egyes kópiák szekvenciáinál tapasztalt „megcsúszás” átfedő kromatogramokat eredményez. Szubsztitúciót mindössze egyet detektáltunk az 5 törzs ITS szekvenciái között. Ebben az esetben, mivel az ugyancsak polimorf pozíciókat tartalmazó *Z. gambellariensis* típustörzs ITS szekvenciáitól legalább 20

szubsztitúcióval eltértek a *Z. favi* ITS szekvenciái, a polimorf pozíciók ellenére az ITS szekvencia alkalmasnak bizonyult a két faj elválasztására. Amennyiben a polimorf pozíciók nem inszerciók/deléciók, hanem szubsztitúciók eredményeként jönnek létre, akkor szintén megoldás lehet a különböző kópiák klónozás utáni szekvenálása vagy a di- és polimorf pozíciók kizárása a filogenetikai elemzésből. Ha ez nem vezet kielégítő eredményre, akkor célszerű lehet hálózatelemzést végezni vagy további, polimorf pozícióktól mentes lokuszokat bevonni a filogenetikai elemzésbe.

3. Az új fajok leírása és a fajok filogenetikai viszonyainak vizsgálata során egyre gyakrabban használnak az rRNS-t kódoló kromoszómális szakaszoktól eltérő bar-code-okat is. A szerző szerint várható ezektől érdemi új információ? Vannak saját tapasztalatai a területen?

Válasz: Véleményem szerint általában elegendő az ITS régióinak és az rRNS-t kódoló gén D1/D2 régiójának az elemzése az új élesztőgomba fajok filogenetikai pozíciójának a meghatározásához. Ez alól azonban kivételt jelentenek pl. azok az esetek, amikor az ITS és D1/D2 régiók DNS bázissorrendjeinek összehasonlítása és együttes filogenetikai elemzése nem dönti el egyértelműen, hogy a vizsgált törzsek egy már ismert fajhoz tartoznak vagy egy leíratlan faj képviselői, illetve nem jelöli ki megbízhatóan a törzsek pozícióját a filogenetikai törzsfán. Ilyenkor az rRNS-t kódoló génszakaszoktól eltérő, pl. fehérje kódoló gének, vizsgálatba vonása hasznos plusz információval szolgálhat. Gyakran vizsgálják ilyen esetekben a transzlációs elongációs faktor 1 α (EF-1 α) gén szekvenciáját, de az aktin gén, valamint az RNS polimeráz II gén 1-es és 2-es alegységeinek szekvenciáit is rendszeresen bevonják a filogenetikai elemzésekbe. A fent említett *Zygosaccharomyces favi* leírásakor az EF-1 α gén szekvenciájának meghatározása segítségével erősítettük meg, hogy az általunk izolált törzsek nem tartoznak azonos fajba a *Z. gambellarensis* típus törzsével. Míg az 5 vizsgált *Z. favi* törzs azonos EF-1 α szekvenciával rendelkezik, 13 szubsztitúcióval térnek el a *Z. gambellarensis* típus törzsének EF-1 α szekvenciájától. A *Brettanomyces acidodurans* faj megbízható filogenetikai pozíciójának meghatározásakor, az rRNS-t kódoló gén egyes szakaszin kívül, ugyancsak elemzésbe vontuk a transzlációs elongációs faktor 1 α gén szekvenciáját. További érv az rRNS-t kódoló gén szekvenciáktól eltérő génszakaszok filogenetikai elemzésbe vonása mellett, hogy előbbieket filogenetikai elemzésekben való önálló alkalmazása esetén nem derül fény a vizsgált törzs hibrid mivoltára, amennyiben ezt a génszakaszt csak az egyik szülőttől örökölte. További bar-code-ok elemzésbe vonása ezt a

problémát kiküszöbölheti, valamint segít a filogenetikai pozíció megbízhatóbb meghatározásában.

4. A filogenetikai elemzéseknél gyakran csak a szubsztitúciókat veszik figyelembe, a deléciókat/inszerciókat figyelmen kívül hagyják. A szerző is élt ezzel a gyakorlattal. Mi indokolja, hogy az utóbbiakat irrelevánsnak kell tekinteni filogenetikai értelemben?

Válasz: Az inszerciók/deléciók fontos evolúciós események és figyelmen kívül hagyásuk információvesztéssel jár, tehát egyáltalán nem törvényszerű, hogy azokat irrelevánsnak kell tekinteni filogenetikai értelemben. Az inszerciót/deléciót tartalmazó nukleotid pozíciók kezelésének csupán egyik, de az élesztőgomba taxonómiában gyakran alkalmazott módja azok eltávolítása az összerendezett szekvenciákból. Ennek az a megfontolás képezi az alapját, hogy nehéz összehasonlítani olyan pozíciókat, amelyekben egyes szekvenciákban nukleotid található másokban viszont nem. Ezért a DNS szekvencia elemzés matematikai háttérének alapos ismeretének hiányában praktikusabb választásnak tűnik az inszerciót/deléciót tartalmazó nukleotid pozíciók figyelmen kívül hagyása. Dwivedi és Gadagkar (2009) beható matematikai elemzésük eredményeként arra a következtetésre jutottak, hogy amennyiben az inszerciót/deléciót tartalmazó pozíciók aránya kisebb 20%-nál, ahogy az általam is elemzett viszonylag kisszámú szekvenciát tartalmazó adathalmazok esetén is, akkor az alkalmazott módszerek, köztük az inszerciót/deléciót tartalmazó nukleotid pozíciók eltávolítása, egyformán jó eredményt adtak.

Dwived B., Gadagkar SR (2009) Phylogenetic inference under varying proportions of indel-induced alignment gaps. BMC Evolutionary Biology 9:211, doi:10.1186/1471-2148-9-211

Budapest, 2020. november 19.



Péter Gábor