

BÍRÁLAT

Dr. Csősz Sándor

KVANTITATÍV MORFOLÓGIAI MÓDSZEREK FEJLESZTÉSE ÉS ALKALMAZÁSA HANGYA-TAXONÓMIAI VIZSGÁLATOKBAN

c. MTA doktori értekezéséről

Nagydoktori értekezések többé-kevésbé rutinos bírálójaként nem ért meglepetésként, hogy - alighanem kizárásos alapon - egy hangyákkal foglalkozó numerikus taxonómiai mű bírálatára lettem felkérve. Miután sem a hangyákhoz nem értek, sem taxonómus és statisztikus sem vagyok, így opponenciámat a józan paraszti észre és kritikai érzékre építem, abból kiindulva, hogy egy jó doktori műnek a tágabb szakma bármely képviselőjével meg kell tudnia értetnie munkájának lényegét. Ez a preambulum annyiban szabadkozás, hogy veszem a bátorságot, és szégyentelenül felteszem a létező legnaivabb kérdéseket, abban a meggyőződésben, hogy a szakmailag túlságosan involváltak pontosan az ilyen alapkérdéseket hajlamosak trivialisnak venni és nem feltenni. Hátha ezen kérdések megválaszolása mégsem lesz teljesen érdektelen, és közelebb hozza a munkát a szélesebb szakmai nyilvánossághoz.

Az értekezés megfelel a doktori művel szemben támasztott formai követelményeknek. Függelékkel együtt 161 oldalból áll, ebből a tartalmi rész 109 oldal hosszú. Igen részletesen, négy-szintű decimális számozással tagolt, még a tartalomjegyzék is külön számozást kapott (a cím nem). Zárójel. Mondjuk pl. a 4.3.4.2-es résznél csak tekergettem a nyakamat, hogy hol is tartok, míg ki nem derítettem, hogy valójában csak a bevezetésnél. Általános szubjektív kritikaként, a dolgozat olvashatóságát tekintve, többször az volt az érzésem, hogy ez egy meta-disszertáció, amelyik egy valahol létező tényleges disszertációról szól. Szinte minden rész és alrész kapott valami külön bevezetőt, ha mást nem, pár felvezető mondatot, újbóli nekifutást, amiből óhatatlanul ismétlődések adódtak; mindenütt meg volt magyarázva, hogy most éppen hol tartunk, olyannyira, hogy olvasóként azt hittem, hogy már valójában ott is tartunk, pedig nem, ez még csak a szerkezetbe behelyezés, a tényleges tartalmi rész csak utána jött (biztos, ami biztos, némi bevezetéssel). Végül is nem egy irodalmi mű, de nekem a dolgozatnak ez a sokszor önmagán való kívülállása és onnan való magyarázatok nem tették gördülékennyé a mondandó megemésztését. Zárójel bezárva.

Tartalmilag bevezetésnek tekinthetők a 4.2 és 4.3-as szakaszok, amelyek általánosságban szólnak taxonómiai problémákról, megközelítésekről, módszerekről. Szerintem egy jó arányérzékkel eltalált rész, amely összességében kellő felvezetést ad a dolgozat gerincét képző 5. részhez, amely rangos publikációkon alapulva öt saját tanulmány ismertetését adja, mindegyik saját bevezetéssel, módszerekkel, eredményekkel és diszkussziókkal. Ezek egy nagyon szép ívet alkotnak, kezdve a morfológiai adatok standardizálhatóságával; folytatva egy rájuk épített sokváltozós numerikus taxonómiai módszer fejlesztésének bemutatásával, továbbiakban erre épített két esettanulmánnyal, amelyek egyben a módszer továbbfejlesztését is bemutatják; végezetül pedig egy a módszer átadhatóságát, általános használhatóságát be-

mutató esettanulmány zárja a sort. A továbbiakban a dolgozat sorrendjében a felvetődött kérdéseimet, megjegyzéseimet osztom meg.

Nagyon tetszett az anatómiai struktúrák taxonómiája, vagyis egy ontologikus, hierarchikus szótár létrehozása, amely egyértelműsítheti és egységesíti a kvantitatív taxonómiában mérendő struktúrákat (= fenomikus adatbázis). Ez egy igazán modern kezdeményezés, mégis ezt a részt olvasva az a kérdés merült fel, hogy vajon nem túlságosan igényli-e ez a megközelítés egyes elkötelezett kutatók munkáját, hogy fenntartsák, fejlesszék az adatbázist. Mi történik, ha egymással ellentétes iskolák dolgoznak egy adott területen, és nincs egyetlen elfogadott szótár, illetve potenciális gond lehet, ha a felhasználók számára túlságosan meredek a tanulási görbe? Nekem megjegyzésként az a hasonlat jut eszembe, hogy olyan, mint a taxonómia „R nyelve”, de nem tudom van-e annyi felhajtó erő mögötte, hogy elérhesse az R nyelv sikerességét, általános elfogadottságát. Provokatívan azt is kérdezhetném, hogy vajon a saját taxonómiai munkáiban milyen arányban használta ezt a módszert a jelölt? További gondolat, hogy a jelenlegi adatbázis hártácsszárnyúakra vonatkozik – vajon milyen mértékig terjeszthető ki, illetve szélesíthető más taxonok felé, nem válik-e ettől kezelhetetlenül bonyolulttá a rendszer? De akár csak egyetlen (az eredeti taxonkörbe tartozó) új faj vagy génusz leírása is gondot okozhat, amennyiben egy jellemző tulajdonság új a szótár számára, minden ilyen esetben magát a szótárat is bővíteni kell (... igen értem, ez benne van a pakliban, de mégis ez a rendszer része, potenciálisan minden új fajnál előfordulhat). Ehhez a részhez végül egy idézet: „Az ontológia fejlesztése időigényes, különleges szakértelmet igényel, és az ismeretek felhalmozása közben is ellenőrizni kell az ontológia hatékonyságát és eredményességét. Ezért további új megközelítésekre van szükség...”; és egy kérdés: mik lehetnek ezek az új megközelítések?

A dolgozat hátralévő része valójában egy homogén témát dolgoz fel, a jelölt által kidolgozott NC-klaszterezés módszerének ismertetését, alkalmazásának és továbbfejlesztésének példáit. A rövidítés a „Nest-Centroid Clustering” névből adódik, amivel kapcsolatban az első sikerélmény akkor ért, amikor kiderült számomra, hogy a „nest” nem egy elvont statisztikai fogalom, hanem konkrét hangyafészek, darázsfészek, stb. Némi csalódás viszont akkor ért, amikor később kiderült számomra, hogy a fészek „centroidja” az gyakorta csak egy, másszor pedig 2-3 egyed alapján kerül számításra. Természetesen így is óriási számú minta lett le mérve, és a módszer hasznosságából, az ötlet eredetiségéből ez nem von le. Az a kérdés mégis felmerül bennem, hogy milyen alternatívák léteznek a módszerhez; szoliter fajokra mennyire értelmes a használata, hiszen itt is gyakorta csak egy egyed képviselt egy fészket; illetve mennyire lenne más az eredmény, ha egyszerűen PCA tengelyekre vonatkoztatott adatmátrixon végeznék a klaszterezést (ami intuícióm szerint egy szoliterekre alkalmazható alternatíva lehetne)?

Az NC-klaszterezés alkalmazását olvasva tetszett az iteratív módszer, a kiindulási fajhipotézis folyamatos javítása, a wild-card egyedek figyelembevétele, és az is, hogy többféle klaszterezési eljárás is alapszik. Ehhez szememben pozitívumként hozzájön a progresszív fajhipotézis kialakításának mintegy megerősítő jellege. Adva van tehát egy sok szempontot figyelem-

be vevő algoritmikus módszer, melynek objektivitása, algoritmikussága igen erősen hangsúlyozásra kerül a dolgozatban. Mégis, én úgy éreztem, hogy nem teljesen zárt a rendszer, szubjektív elemek több részen is beférkőzhetnek. Két fő rés amit látok – és ebben kérek kommentet, hogy mennyire jól látom – a különféle klaszterezési eljárások párhuzamos alkalmazása. Vajon filogenetikai mintára nincs módszer valamiféle konszenzusos fa kialakítására? A másik a független (pl. molekuláris) adatok figyelembevételének módja vajon mennyire standardizált és algoritmikus, vagy mennyi ebben a szubjektív elem? Ide kapcsolódik a módszer még egy kevésbé kontrollált eleme a (már említett) fészkenként vizsgált egyedszám. Ezzel kapcsolatban az 59. oldalról idézek: „...a mintázatfelismerés pontosságának javítására a fészeksorozatok méretében, azaz a fészeksorozatok átlagos egyedszámában [is] rejlik.” A kérdésem, hogy valójában mennyire a fészkenkénti mintanagyság *átlaga* a meghatározó, mi a helyzet, ha ez az átlag sok 1-es mintaszámból és néhány jóval számosabból, vagyis egy erősen ferde eloszlásból jön ki? A módszert szigorúan csak kriptikus fajpárokon tesztelték. Amennyiben nem kettő, hanem több elemű faj-komplexekeken történne az tesztelés, az ugyancsak működne, vagy módosítani kellene az eljárást? Volt-e a módszer ilyen faj-komplexekeken tesztelve?

A módszer alkalmazását bemutató első esettanulmány a *Temnothorax* génusz fajait vizsgálta, melyek, mint megtudjuk, monomorfok. Elég ezt az egy szót olvasni, és felmerül a kérdés, hogy olyan hangyafajoknál, ahol különböző dolgozói kasztok léteznek, vajon az NC-klaszterezés eredményét hogyan befolyásolják a kasztok közti, adott esetben nagy, morfológiai különbségek? Az esettanulmányban szerepel 2 fajpár, ahol (idézetek) egyrészt „...[a viszonylag alacsony] klasszifikációs siker már az a szint, ahol a fajhipotézisek újragondolása indokolt”, másrészt „A molekuláris módszer nem tudta teljes mértékben alátámasztani a [kérdéses] fajpárokat.” Vajon jogos-e ezek után különálló kriptikus fajokról beszélni, illetve van-e a módszernek olyan alkalmazása, amikor a végeredmény az, hogy a külön-faj hipotézist el kell vetni?

A második esettanulmány a Malagasi *Nesomyrmex* fauna feltárásán mutatja be az NC-klaszterezés továbbfejlesztését gap-analízissel, amit a jelölt NC-PART analízisnek hív. A jelölt a 85. oldalon írja: “A klaszterek optimális számának becslése nagyon fontos kérdés, amellyel meg tudjuk könnyíteni a taxonómiai munkákban állandóan felmerülő „*hány fajunk van?*” kérdésre való válaszadást. A *gap* statisztika formalizálja ezt a megközelítést, és egy könnyen alkalmazható algoritmus segítségével segít megtalálni a megfelelő klaszter számot.” A esettanulmány példájában felhozott nyilvánvaló csoportok számát jól meghatározza a gap módszer, de a kérdés az, hogy a biológiai folyamatok által gyakran létrejövő egymásba ágyazott csoportosulásokkal mit kezd, illetve mi a megoldás arra, hogy a megjósolt csoport szám az kongruens a biológiai fajok számával? A kérdésre később részben választ kapunk egy rekurzív módszer ismertetésével, azonban számomra továbbra sem világos, hogy statisztikailag felfedezett hasonlóság-alapú csoportok feltárásakor honnan tudjuk, hogy melyik szint a faji szint, vagyis mikor kapunk választ a jelölt által feltett, imént idézett kérdésre. Ugyanez a kérdés kissé később az esettanulmány diszkusszió részében a jelölt által is felvetődik: „... [a] dendrogramon megjelenített morfológiai mintázat megmutatja nekünk, hol vannak a na-

gyobb rések a homogénebb klaszterek között, azaz képet alkotunk arról, hogy vannak reprodukív akadályok által okozott morfológiai diszkontinuitások az adatsorban.” Egy kijelentő mondat, amely azonnal vitára sarkall, de a jelölt is érzi a problémát, elég csak tovább olvasnunk, hogy ő is rámutasson: „Természetesen nem csak reprodukív akadályok okozhatnak fenotípusbeli változásokat”. Úgy tűnik, hogy a dolgozat itt egyenlőségjelet tesz a reprodukív és földrajzi izoláció közé, és a helyzetet a földrajzi és morfológiai gap-ek korrelációjának keresésével próbálja feloldani: „A taxonómia szempontjából a fajképződés, azaz a reprodukív izoláció jelensége a legfontosabb, ezért fontos tisztázni, hogy a tapasztalt morfológiai diszkontinuitások milyen mértékben korrelálnak geográfiai mintázattal.” Hát igen, valószínűleg ha sok adatunk van és szép elkülönülő areákat látunk, akkor ez egy jó megoldás, bár nyilvánvaló hogy nem számol a reprodukív izoláció más lehetséges okaival (szimpatrikus fajképződés). És az is nyilvánvaló, hogy a taxonómiában igen gyakran nagyon kis számú példány és hiányos háttérinformáció alapján kell faji besorolásokat végezni, adott esetben új taxont leírni.

Bírálatom végén hadd utaljak vissza a dolgozat által megfogalmazott, talán legfontosabb célkitűzésre, megoldandó problémára (20. oldal): „... olyan feltáró jellegű elemzéstípus kidolgozásának és alkalmazásának igénye, mely képes nagyméretű, sokdimenziós adatsorban is megtalálni a biológiai jelentéssel bíró mintázatot.” A jelölt elismerésre méltóan küzdött célja eléréseért, de talán a biológiai jelentés problémája az egyik legnehezebben feltörhető dió az egész kérdéskörben, és ez az az egy pont, ahol objektív okok miatt véleményem szerint jelen dolgozat sem hozott áttörést, de talán előrelépést igen. Összegezve felismerésemet, megálapítom, hogy *a taxonómia egy nehéz favágó munka, ahol a nehézséget főként az okozza, hogy hol vágjuk el a fát - vagyis a dendrogramot.* A jelölt ezen munkában elismerésre méltó és nemzetközileg el is ismert eredményeket ért el, és a leíró taxonómiai munkásságán túl számos új módszer kidolgozásával és ismeretátadási kezdeményezéseivel az egész tudományterület fejlődését elősegítette.

A doktori munka tudományos eredményeit elegendőnek értékelem az MTA doktori cím elnyeréséhez és mindezek alapján javaslom az értekezés nyilvános vitára bocsájtását és sikeres védés esetén az MTA doktora cím odaítélését támogatom.

Apróbb, konkrét kérdések:

16. oldal Egy irodalomból átvett ábrába botlottam bele (2. ábra), ahol a felírat szerint „a halmazokban lévő pontok egy-egy elkülönítésre váró, önálló leszármazási vonalat (például fajokat) jelölnek” Kérdésem: hogy is van ez? Ha az egyik módszert egy élőlény csoportot 2 fajnak, egy másik pedig egy fajnak ismer fel, akkor hová kerülnek a pöttyök, honnan tudjuk, hogy a pöttyöknek van egymáshoz kapcsolatuk, az pöttyök egymástól való távolságnak mi a relevanciája?

27. oldal: Hogyan alakítsunk ki egy olyan környezetet, amelyben a kutatók könnyen vonhatnak le következtetéseket fenotípus adatok felhasználásával, például a környezet és a genetika közvetlen kapcsolatából? A fenotipikus plaszticitás jelenségét figyelembe véve, morfológiai adatok elemzésével alighanem sehogy, csak kísérletesen (common garden kísérletek).

39. oldal „Az eredményt egy dendrogram formájában lesz ábrázolva, amelynek az egyik végén az egyes elemek klaszterekbe rendezve találhatóak. Ezek a klaszterek alkotják számunkra az elsődleges taxonómiai hipotézist. „ – ha tudjuk, hogy melyik hasonlósági szinten vágjuk el a dendrogramot...

73. oldal Döntési fa: van olyan elágazás, ahol mindkét választás u.ahhoz a fajhoz (is) vezethet.

78. oldal „a parapatikus fajpár nyugati és keleti alkotóit más szociálpazita hangyafaj használja ki (Alfred Buschinger szóbeli közlése), ezáltal igazolva az általam morfológiai különbségek alapján elkülönített fajok biológiai validitását.” – Ezt én megkérdőjelezhető bizonyítéknak tekinteném, hiszen a paraziták is nyugodtan rendelkezhetnek diszjunkt elterjedéssel még ha egyazon szélesen elterjedt fajt parazitálnak, akkor is.

92. Jól látom, hogy a 29. ábrán a dendrogram alapján jó néhány rugosus-ként besorolható minta végül excelsior besorolást kapott?

Budapest, 2018.11.08



Samu Ferenc
MTA doktora