

Válasz Dr. Samu Ferenc opponensi véleményére

Először is, szeretném megköszönni Dr. Samu Ferenc opponensi munkáját, dolgozatommal kapcsolatos észrevételeit, köszönöm kérdéseit, és hogy az értekezést nyilvános vitára javasolja.

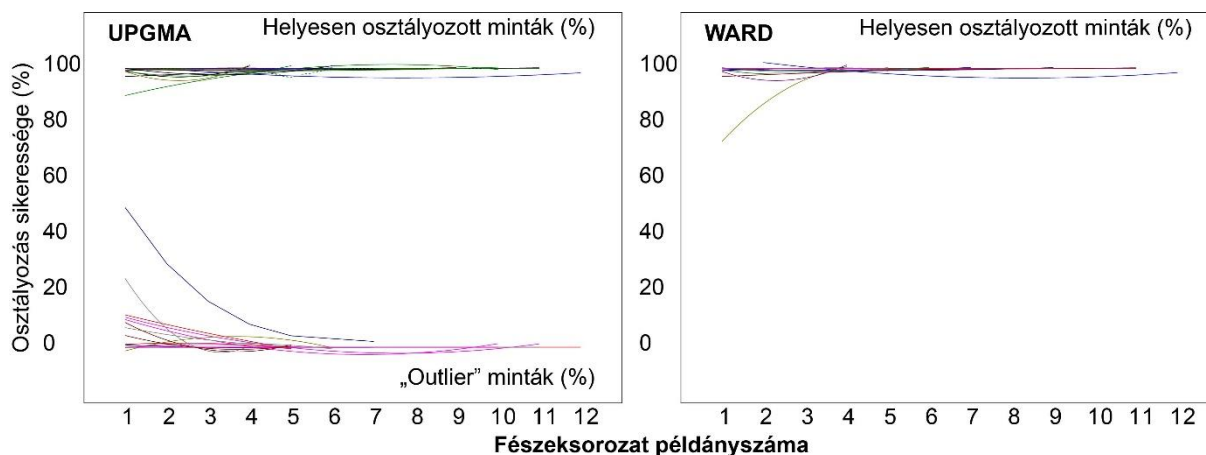
A bírálóm formai tartalmú építő jellegű megjegyzéseit köszönöm és azokkal egyetértek. Opponensem felvetése az ontológiai adatbázis létrehozásával és kezelésével kapcsolatos humán erőforrás igényére teljesen jogos. Ezeket a szemantikus fogalmakat definiálni, adatbázisban rögzíteni, ontológiai nyelvi elemekké formázni kell és az adatbázis folyamatos karbantartása is rendkívül sok munkát igényel. Bírálómnak arra a kérdésére, hogy – idézem *„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.”* azt tudom válaszolni, hogy a hasonlat nagyon találó, és valószínűleg a kezdeményezés hasonló utat fog bejárni, mint a ma ismert programozási nyelvek mindegyike. Lesz olyan, amelyeket sokan kezdenek használni, mások a feledés homályába merülnek. Hogy melyik nyelv fog végül érvényesülni, az ma nem tudjuk megmondani.

Bírálóm következő kérdésére, nevezetesen *„Provokatívan azt is kérdezhetném, hogy vajon a saját taxonómiai munkáiban milyen arányban használta ezt a módszert a jelölt?,”* azt tudom válaszolni, hogy minden cikkemben használtam a módszert, ahol lehetőségem volt a szemantikus fenotípus leírásokat is publikálni. Az említett ontológiai munka megjelenése, azaz 2015 óta mindössze két olyan taxonómiai munkát közöltem, ahol mellőztem az ontológiai morfológiai leírásokat. Egyik esetben sem volt lehetőség az URI tábla melléklésére a publikációban. Az egyik ilyen munkámban (Csősz et al. 2018), mely az európai *Temnothorax*-ok egy fajcsoportjával foglalkozik valamelyest redundáns is lett volna a szemantikus fenotípus leírásokat publikálni, hiszen a *Temnothorax*-ok karakterállapot-mátrixának leírását már egy korábbi munkámban (Csősz et al. 2015) leközöltem.

Egyetértek opponensemvel abban, hogy az ontológiai kifejezések adatbázisának hártácsszárnyúakon túli kiterjesztése nehézségekbe ütközhet. Az ontológiai anatómiai leírásokra viszont általánosan igaz, hogy egészen addig terjeszthetők ki, amíg homológ struktúrákat találunk. Nyilván vannak hártácsszárnyú-specifikus karakterek, de rovarok között általánosan elterjedt struktúrákra a rovarvilágon belül általánosan használható. Tovább menve, például a lábak leírásánál az ízeltlábúak között széleskörben alkalmazható lehet.

Némileg megértem opponensem csalódottságát arra vonatkozóan, hogy a „Nest-Centroid Clustering” tulajdonképpen csak a fészekből vett mintára, s nem a teljes fészekre vonatkozik. A fészeksorozat valóban csak jellemzően kettő és kilenc közötti egyedszámot takar fészkenként. Gondolhatnánk, hogy a minél nagyobb egyedszám lemérése lenne az ideális megoldás, de gyakorlati okokból a fészkenként vizsgált egyedszámot érdemes lesoorítani. Az általam végzett összehasonlító elemzések alapján azt lehet elmondani, hogy

fészeksorozatonként kb. 3-4 dolgozóval a módszer már teljesen megbízható eredményeket produkál (lásd a mellékelt ábrát lejjebb), s a megbízhatósága a dolgozók számának növekedésével tovább már nem növekszik.



A további dolgozók lemerése nagyon kevés gyakorlati haszon mellett indokolatlanul sok állat megölésével, preparálásával és lemerésével járna. Egy hangyadolgozó 20 morfometriás karakterének lemerése a húsz éves tapasztalatommal is legalább 5 percet vesz igénybe, tehát érdemes optimalizálni a befektetett időt és egy fészek szükségtelenül hosszadalmas vizsgálata helyett inkább több hangyafészek optimális mennyiségű egyedének vizsgálatába invesztálni. A mintánkénti egyedszám meghatározó az elemzés pontosságának szempontjából. A módszer ugyanis csak fészeksorozatokból (egynél nagyobb mintaszámú egységekből) képes modellt építeni. Bár a dendrogramon a "singletonok" (magányosan gyűjtött egyedek) is szerepelnek, azok nem vesznek részt az algoritmus által alkotott döntéshozatalban, úgymond passzívan kerülnek elhelyezésre a fészeksorozatokból alkotott modell alapján. A bírálóm kérdésére számszerűen is válaszolva, a legutóbbi tesztelési eredményem szerint 70% singleton mellett 30% 2-3 dolgozóból álló fészeksorozat még értékelhetően jó képet adott egy két-fajos rendszerben. Olyan esetekben, ahol az adatsorban nincsenek fészeksorozatok, esetleg kifejezetten szoliter fajokkal dolgozunk, akkor az NC-klaszterezés nem működik, ezért valóban alternatív utakat kell keresnünk. Megoldásként a morfológiai adatok elemzésére például a PCA, vagy annak egy speciális változata, az sPCA (Baur & Leuenberger 2011) jöhet szóba. Ez utóbbi módszer, ellentétben a hagyományos PCA-val, képes kiszűrni a méretváltozókat is, ami nagy előnyt jelent a természetes populációk vizsgálatában.

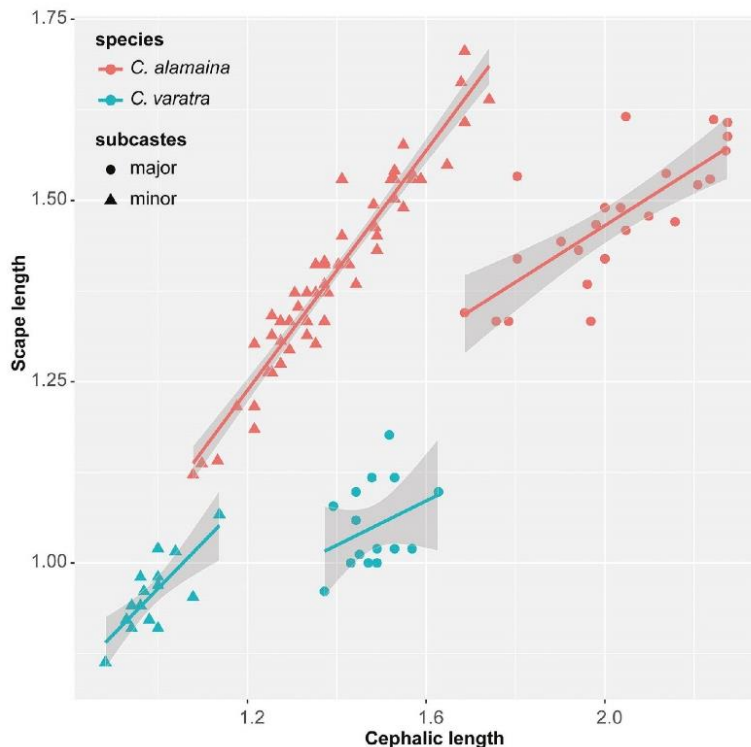
Maradéktalanul egyetértek bírálóm abbéli észrevételével, hogy az igyekezetem ellenére marad tere némi szubjektivitásnak, és a rendszer „nem teljesen zárt”. Még a legobjektívebbnek nevezett eljárásokba is beférkőzhet ugyanis a szubjektivitás, még hozzá azért, mert mi emberek, azaz a vizsgálók is a rendszer részét képezzük (ld. még Kant – Tiszta ész kritikája). A vizsgáló személyén keresztül a szubjektivitás kockázatát hordozza magában minden, amiben mi magunk döntünk: milyen karakterek halmazát vizsgáljuk, milyen módszert választunk az elemzéshez, vagy milyen statisztikai *threshold* értéket tekintünk mérvadónak a következtetéseink levonásához. Egy bizonyos fokú szubjektivitás tehát nem csak az én módszeremnél létező jelenség, hanem ezzel minden tudományterület

képviselőinek számolnia kell. Ez alól nincs kivétel, de mégis úgy érzem, ennek ellenére törekednünk kell a lehető legnagyobb objektivitás elérésére. Ezt a vizsgálat teljes tartama alatt azonos adat-, és elemzés-készlettel tudjuk elérni, amit a tudományossági kritériumok között is gyakran említenek, s az *uniformitás* elvének neveznek. Az én javaslatom itt az, hogy azonos elvek alapján húzzuk meg a természetben található varianciában felismert határokat. Ennek eldöntéséhez egy általánosa(bba)n alkalmazható, statisztikai alapú határértéken alapuló eljárást próbáltam ki, ami eddigi tapasztalataim alapján a hangyákra nagyszerűen működik. Nincs kételyem afelől, hogy legalább néhány egyéb csoport esetében is hasonló sikerrel alkalmazható. Ha valaki tehát szeretné saját taxonómiai döntéseit „uniformizálni”, akkor számára már létezik egy általam kidolgozott és kipróbált lehetőség, de mindenkit bátorítok egy saját maga számára jól alkalmazható saját rendszer kidolgozására.

A bírálóm kérdése az NC-klaszterezés többfajos rendszerekben mutatott hatékonyságával kapcsolatban nagyon releváns, miután a módszert bemutató cikkben kifejezetten kriptikus fajpárokon teszteltük. Felmerülhet tehát a kérdés, hogy több fajos rendszerekben hogyan viselkedik a módszer. Az NC-klaszterezés bevezetésekor azért ragaszkodtunk a kétfajos teszt adatsorokhoz, mert az NC-klaszterezéssel szembe állított K-means klaszterezés összetettebb adatsorokon mutatott működéséről, s annak konzisztenciájáról nem voltunk meggyőződve és igyekeztük elkerülni a klaszterszám növeléséből fakadó ismeretlen hatásokat. Magyarul, kifejezetten a hagyományos módszerek kedvéért teszteltük a módszerünket kizárólag kriptikus fajpárokon. Az NC-klaszterezés számára az adatsorok összetettsége nem jelent problémát. Ezért a karakterek száma, vagy/és az esetek száma is igen tág határok között növelhető. Ugyanígy nem érzékeny a módszer a fajsáma sem. Persze a nagyobb adatsor nagyobb számítási kapacitást és/vagy számítási időtartamot igényel, de ez az algoritmikus módszereknél általános jelenség. Tapasztalataim alapján azt mondhatom, hogy a hagyományos K-means, vagy PCA módszerekkel összehasonlítva az NC-klaszterezés kiemelkedő hatékonysága a több fajos rendszerekben még inkább megmutatkozik.

A módszerem gyakorlati alkalmazását bemutató első esettanulmány a *Temnothorax* génuszon történt, ezzel kapcsolatban a bírálóm két kérdést tett fel. Szeretném ezeket külön megválaszolni. Az első kérdés így szól: *„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?”*

A polimorfizmus két módjáról beszélhetünk, melyeket a kérdés megválaszolása érdekében meg kell különböztessük egymástól. Az első csoportba azok a populációk tartoznak, melyek allometrikus változása lineáris. Ezt a fajta polimorfizmust angol terminusban "reaction norm"-nak nevezik, s kialakulását a folytonos környezeti változókra adott folytonos fiziológiai, morfológiai eltérésekkel szokták magyarázni. Az ilyen lineáris eltéréseket az NC-klaszterezés tudja kezelni, ezért az ilyen populációk elemzése nem okoz semmi féle problémát. A másik csoportba a polifénikus fajok tartoznak, melyek környezeti változókra nem lineáris morfológiai változásokkal reagálnak. Az ilyen fajok egyedeinek egyes testrészei fejlődésük során egy lárvális határérték méretet elérve egy másik fejlődési pályára állnak át (Emlen & Nijhout 2000). Ennek hatására bizonyos testrészek meggyorsult növekedéssel reagálnak, ami az imágó új típusú morfológiáját vonja maga után.



Ilyenek például a *Pheidole*, vagy egyes *Camponotus* fajok, ahol a dolgozók mellett megjelennek a katonák is (lásd a mellékelt ábrát). Ezeket a populációkat csak ugyanazon alkaszatok összehasonlításával lehet vizsgálni, például csak dolgozókat, vagy csak katonákat elemezhetünk, a kettőt egybe nem vegyíthetjük. Az utóbbi korrekcióra láthatunk példát a *Camponotus* genusz néhány fajcsoportjának revízióiban, melyeket a helyi kollégákkal együtt végeztünk el (Rakotonirina, Csősz S & Fisher 2016 stb.).

Az ide vonatkozó második kérdés pedig így szól: „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 faji határ meglétét a morfológiai mintázatban tapasztalható kontinuitás megszakadása mutatja, ami a cumulatív LDA megerősítő tesztelés során alacsonyabb volt, mint a kívánatos 95% klasszifikációs siker. Ennek oka mindenképpen a fajpárok közötti kis morfológiai távolságban keresendő. Mindkét említett fajpár parapatrikus elterjedésű, azaz a fajpárok tagjai egy szűk, 20-50 km széles kontaktzónában találkoznak. Vizsgálataink során ezért külön hangsúlyt fektettünk a kontaktzónákból gyűjtött minták vizsgálatára, hogy kiderítsük, vajon előfordul-e keveredés a két faj között, avagy sem. Az eredmények azt mutatták, hogy ezeken a közösen elfoglalt területen sem a fajok intraspecifikus polimorfizmusa, sem pedig az átmeneti formák gyakorisága nem növekedett meg, vagyis morfológiai módszerekkel nem mutatható ki fajok közötti keveredés (Csősz et al. 2014, Seifert & Csősz 2015). Az a tény, hogy a *T. laconicus* - *T. lichtensteini*, valamint a *T. crasecundus* - *T. crassispinus* fajpárokat az általunk használt mitokondriális COI szakaszok vizsgálata nem tudta megerősíteni óvatosságra int, de önmagában nem perdöntő, mert bár ez utóbbi közkedvelt és olcsó megoldás, annak sok hiányossága (Collins & Cruickshank 2013) miatt többen támadják. Egyértelmű választ a feltett kérdésre nukleáris gének vizsgálata tudna adni. A későbbiekben a kérdésben szereplő fajpárok tagjainak hasonló problémájára mindenképpen szeretnék megnyugtató választ találni például UCE - Ultra Conservative Elements segítségével.

Bírálom egy másik esettanulmány kapcsán felmerülő kérdése is hasonló témában született.

A *gap* statisztika, illetve az azt használó NC-PART algoritmus diszkontinuitásokat keres az morfológiai adatsorban, melyek a genetikai információáramlásra utalhatnak. Éppen ez, azaz a szaporodásbeli elkülönülés a biológiai fajfogalom alapja. A fajok közötti morfológiai eltérések leginkább az elszigetelődésük során beálló reprodukzív izoláció melléktermékének tekinthetők. Többváltozós módszerekkel ezek a morfológiában adódó különbségek jól detektálhatók. A faji határ meglétét a morfológiai mintázatban tapasztalható kontinuitás megszakadása jelentheti. Nehéz tehát egy morfológiai klaszter párról teljes bizonyossággal kijelenteni, hogy azok biológiai fajokat reprezentálnak, de az általam megállapított, morfológiailag jól elkülönült klaszterek integratív módszerekkel (molekuláris filogenetikai, korológiai adatok, fenológia, ökológiai modellek stb.) történő vizsgálatai rendre megerősítették a morfológiai klaszterek nagy többségének faji szintű besorolását.

Bírálom a dolgozat végén feltett „apróbb, konkrét” kérdéseire pontokban kívánok válaszolni.

1. *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 módszer 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, a pöttyök egymástól való távolságnak mi a relevanciája?*

A 2. ábrán látható pontok a különböző megközelítések (mtDNA, morphology, ecological niche) alapján meghozott döntéseket jelzik. Ha a döntés a különböző megközelítések metszéspontjában helyezkedik el, akkor tekinthető elfogadhatónak. Koordináta tengelyeket képzelhetünk el a három halmaz alatt és bal oldalán. Az x tengely például jelenthet fajszámot, így ha az egyik módszer 1, a másik 2 fajt állapít meg, az nem esik a három megközelítés metszéspontjába, tehát nem fogadjuk el.

2. *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).*

A jelenleg elérhető ontológiákon alapuló rendszerek csak közvetetten kötik össze a fenotípusokat példányokkal és ezért, bár tükrözik a fajon belüli variabilitást, valóban nem teszik lehetővé, hogy példány szinten lehessen modellezni a környezet-fenotípus-genotípus kapcsolatokat. A legújabb ontológián alapuló rendszerek azonban egy olyan példány alapú szemantikus reprezentációra adnak lehetőséget, amiben a fenotípus-példány-környezet összeköttetések konkrétak és ezért ezekben a rendszerekben nem okoz gondot a fenotipikus plaszticitás, sőt lehetőséget ad arra, hogy megállapítsuk korrelál-e a fenotípus a környezettel olyan tesztelni való hipotéziseket szolgáltatva, amik a fent említett környezet genotípus kapcsolatokat írják le. A fenotípusok szemantikus reprezentálása egy ilyen rendszer alapköve.

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

A klaszterek egy vonallal való „elvághatása” kezdetleges módja a csoportok formázásának. Valójában több távolságmódszernél, mint a Ward módszer, nem is ad jó eredményt. Az UPGMA esetében – a módszer nagy fokú konzisztenciája miatt – lehet az „elvághatásnak” relevanciája, de ott sem ajánlanám. Nem minden biológiai tartalommal bíró klaszter kezdődik ugyanolyan magasságnál. A dendrogram egy tetszőleges pontját elvágva tehát hamis következtetésekre juthatunk. A NC-klaszterezés PART algoritmussal való kombinálása pontosan erre a problémára kínál egy kényelmes és széleskörű megoldást: ez ugyanis nem elvágja a fát (mint mondjuk azt a „cut.dendrogram” parancs tenné) hanem a klasztereket az adatsorban létező diszkontinuitás alapján definiálja és jelöli ki annak elemeit.

4. 73. oldal *Döntési fa: van olyan elágazás, ahol mindkét választás ugyanahhoz a fajhoz (is) vezethet.*

Ez valóban így van, a határozókulcsok néha alkalmazzák az (in part) kifejezést, ami olyan esetekben alkalmazható, mikor egy jelleg akár a kulcs további lépései során az egyik (vagy mindkét) ág fajai között visszatérhet. Ilyenkor egy fajhoz akár két úton is el lehet jutni. Ez nem szabad, hogy problémát okozzon, s a további lépések alapján ezek a fajok is meghatározhatók.

5. 78. oldal *„a parapatikus fajpár nyugati és keleti alkotóit más szociálp parazita 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.*

Köszönöm bírálóm észrevételét. Valóban, egy fajt akár több szociálp parazita faj is kihasználhat. A fenti példa önmagában valóban nem lenne elégséges a metapopulációk elkülönülésére, de alátámaszthatja a morfológiai különbségeket.

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

Igen, a megerősítő elemzés ezeket a fészekmintákat az *Nesomyrmex excelsior* nevű fajba sorolta és a protokoll szerint ezt kell elfogadni.

Végezetül szeretném még egyszer megköszönni opponensem részletes bírálatát, fontos észrevételeit és kérdéseit, és eddigi munkám pozitív értékelését.

Budapest, 2019. 02. 18.

Csósz Sándor

Felhasznált irodalom

- Baur H & Leuenberger C (2011) Analysis of ratios in multivariate morphometry. *Systematic Biology*, 60: 813–825. doi: 10.1093/sysbio/syr061 PMID: 21828084
- Collins, R. A., & Cruickshank, R. H. (2013). The seven deadly sins of DNA barcoding. *Molecular ecology resources*, 13(6), 969-975.
- Csősz S, Heinze J & Mikó I (2015) Taxonomic Synopsis of the Ponto-Mediterranean ants of *Temnothorax nylander* species-group. *PloS one*, 10(11), e0140000.
- Csősz S., Seifert B., Müller B., Trindl A., Schulz A. & Heinze J. (2014). Cryptic diversity in the Mediterranean *Temnothorax lichtensteini* species complex (Hymenoptera: Formicidae). *Organisms Diversity & Evolution*, 14(1), 75-88.
- Csősz, S., Salata, S., & Borowiec, L. (2018). Three Turano-European species of the *Temnothorax interruptus* group (Hymenoptera: Formicidae) demonstrated by quantitative morphology. *Myrmecological News*, 26, 101-119.
- Emlen, D. J., & Nijhout, H. F. (2000). The development and evolution of exaggerated morphologies in insects. *Annual review of entomology*, 45(1), 661-708.
- Kant, I. (1981). A tiszta ész kritikája. Akadémiai Kiadó.
- Rakotonirina JC, Csősz S & Fisher BL (2016) Revision of the Malagasy *Camponotus ed-mondi* species group (Hymenoptera, Formicidae, Formicinae): integrating qualitative morphology and multivariate morphometric analysis. *ZooKeys*, 572: 81–154.
- Seifert B. & Csősz S. (2015) *Temnothorax crasecundus* sp. n.–a cryptic Eurocaucasian ant species (Hymenoptera, Formicidae) discovered by Nest Centroid Clustering. *ZooKeys*, 479: 37.