

MTA DOKTORI ÉRTEKEZÉS

**KVANTITATÍV MORFOLÓGIAI MÓDSZEREK FEJLESZTÉSE ÉS
ALKALMAZÁSA HANGYA-TAXONÓMIAI VIZSGÁLATOKBAN**

Dr. Csősz Sándor

2017

1 Tartalomjegyzék

1 Tartalomjegyzék.....	2
2 Előszó.....	6
3 A dolgozat felépítése	8
4 Bevezetés és célkitűzések.....	10
4.1 Motivációim.....	10
4.2 A taxonómia modern kori kihívásai.....	10
4.3 Az alkalmazott megközelítések	14
4.3.1 A fenotípusos adatok fontossága	14
4.3.2 A »bizonyíték alapú taxonómia«	15
4.3.3 A morfológiai vizsgálatok használata az integratív taxonómiában	16
4.3.4 A kvantitatív módszerek előnyei és alkalmazásának korlátai	17
4.3.4.1 Miért a kvantitatív morfológiában látom a taxonómia fejlődésének kulcsát?	18
4.3.4.2 A kvantitatív morfológiai megközelítések taxonómiai alkalmazásának korlátai.....	18
4.4 Célkitűzések: kvantitatív morfológiai megközelítések alkalmazása és népszerűsítése a taxonómiában.....	20
5 Saját eredmények és megvitatásuk	23
5.1 A fenotípusos jellegek adatsorokká alakítása és szabványosítása – a Hymenoptera Anatomy Ontology portál	23
5.1.1 Egy fenotípus regisztráló környezet kialakítása	27
5.1.2 Az ontológiai projektek jelen kori kihívásai	29
5.1.3 Az adat integrálás elérése.....	30
5.1.4 A szemantikus fenotípus megközelítés gyakorlati taxonómiai alkalmazásának példái	32
5.2 NC-klaszterezés – Egy új alkalmazás biodiverzitás morfológiai mintázatainak felismerésére	34
5.2.1 A mintázatfelismerő módszerek fejlesztésének szükségessége	36
5.2.2 Az új megközelítés gyakorlati kivitelezése.....	38
5.2.3 Anyag és módszerek	40
5.2.3.1 A tesztelésre kiválasztott csoportok és a kiválasztás kritériumai	40
5.2.3.2 Az adat mátrix előkészítése.	41
5.2.3.3 Agglomerative Nesting (AGNES)	41
5.2.4 A kvantitatív morfológiai fajhipotézisek generálásának protokollja	43
5.2.4.1 A »végleges morfológiai fajhipotézis« megalkotása	43
5.2.4.2 »A progresszív fajhipotézis«	44
5.2.5 Eredmények és diszkusszió.....	45
5.2.5.1 Az eredményül kapott progresszív fajhipotézis összehasonlítása korábban publikált taxonómiai eredményekkel	45
5.2.5.2 A klaszterező algoritmusok eredményeinek összehasonlítása	47
5.2.5.3 Hibrid esetek vizsgálata	55
5.2.6 Konklúzió és ajánlások	57

5.2.6.1	Módszerek kombinált használata	57
5.2.6.2	A minták mérete az NC-klaszterezés sikerességét befolyásoló tényező.....	58
5.2.6.3	A négy vizsgált módszer nem javasolt hibrid esetek vizsgálatára	60
5.3	Az új módszertani elemek gyakorlati alkalmazása a biodiverzitás kutatásában – NC- klaszterezés alkalmazása Nyugat Palearktikus hangya fauna taxonómiai feltárásában	61
5.3.1	A munka aktualitása	62
5.3.2	Anyag és módszerek	64
5.3.2.1	A morfometrikus karakterek mérésének protokollja.....	64
5.3.2.2	A statisztikai feldolgozás	64
5.3.2.3	Morfometrikus adatok feltáró elemzése	65
5.3.3	Eredmények	66
5.3.3.1	Morfológiai eredmények	66
5.3.3.2	Molekuláris eredmények	69
5.3.4	Konklúzió.....	71
5.4	A Malagasi <i>Nesomyrmex</i> fauna taxonómiai feltárása	79
5.4.1	Automatizált taxonómiai döntéshozási eljárások bevezetése a Madagaszkár hangya faunájának taxonómiai feltárásának meggyorsítása céljából	81
5.4.2	A munka aktualitása	87
5.4.3	Anyag és módszer	89
5.4.3.1	A morfometrikus karakterek mérésének protokollja.....	91
5.4.3.2	Statisztikai feldolgozás	91
5.4.4	Eredmények és konklúzió	95
5.4.4.1	A Malagasi <i>Nesomyrmex</i> fauna diverzitása	95
5.4.4.2	Kriptikus <i>Nesomyrmex</i> fajok aránya a régióban	95
5.4.4.3	A <i>Nesomyrmex</i> fajok környezeti és geográfiai szegregációja	98
5.4.4.4	Malagasi endemizmusok a fauna vizsgált szegmensében	98
5.5	Malagasi kollégák kvantitatív taxonómiai munkára való felkészítése és a módszertan meghonosítása helyi szinten.....	101
6	Az eredmények összefoglalása	105
6.1	A szemantikus fenotípus megközelítés alkalmazása hangyákon	105
6.2	NC-klaszterezés és NC-PART morfológiai mintázat felismerő rendszerek.....	105
6.3	Taxonómiai munkásságom	107
6.4	Taxonómus-képzés Madagaszkáron – módszertani tréning helyi kollégák számára	108
7	Köszönetnyilvánítás	109
8	Irodalomjegyzék	110
8.1	Az értekezéshez felhasznált publikációim listája	110
a)	A doktori fokozat megszerzése utáni (2007-2017) publikációk listája	110
b)	A doktori fokozat megszerzése előtti (1998-2007) publikációk listája	112
8.2	Az értekezés elkészítéséhez fel nem használt publikációim listája	113
8.3	Az értekezéshez felhasznált egyéb publikációk listája	115

9	Mellékletek.....	124
9.1	Melléklet. A leírásokban és a kulcsokban használt morfológiai állítások terminológiája, valamint a fenotípuson alapuló diagnózisok, ontológiák URI táblázatban rögzített reprezentációja.	124
9.2	Melléklet. A Semantic phenotype approach gyakorlati alkalmazásának bemutatása.	151
9.3	Melléklet. Az 5.3. fejezetben bemutatott integratív taxonómiai munkák molekuláris filogenetikai módszertanának kivonata.....	153
9.4	Melléklet. Az 5.3. fejezetben bemutatott integratív taxonómiai munkákban használt morfológiai karakter felvételek pontos protokolljának kivonata.	154
9.5	Melléklet. A NC-PART klaszterezés és a 'mark.dendrogram' alkalmazás R scriptje.	157
9.6	Melléklet. Az 5.4. fejezetben bemutatott taxonómiai munkák morfológiai karaktereinek és mérési pontos protokolljának kivonata.....	160

*“Science has taught us, that because we have a talent for
deceiving ourselves, subjectivity may not freely reign.”*

(Carl Sagan: A Universe Not Made For Us)

2 Előszó

A minket körülvevő világ megismerésére való törekvés egyidős az emberiséggel, amely az emberi természet egyik alapvető jellemvonásából, a biztonságra való törekvésből fakad. A folyton változó környezetünk ugyanis drasztikus hatással van életünkre, kiszámíthatatlanságával folyamatos kihívásokat jelent számunkra. Érthető tehát az igény, hogy tudni akarjuk, mi vesz minket körül, milyen folyamatok zajlanak körülöttünk, milyen hatása lehet az életünkre, valamint hogyan tudjuk befolyásolni ezek kimenetelét. Az emberre jellemző megismerés vágyát tehát – véleményem szerint – nem pusztán a világra való rácsodálkozás fűti, hanem legalább ugyanilyen mértékben a biztonságra törekvésünkből fakad. A környezetünkben zajló folyamatok előrejelzésének szükséges előfeltétele a minket körülölelő világ alapvető törvényszerűségeinek felismerése és megértése. Nem csak a közvetlen veszélyeket rejtő időjárási, geológiai vagy egészségügyi jelenségek felfedezése érthető ez alatt. Ugyanilyen fontos a biodiverzitás megismerése, azaz tudatában lenni annak, hogy az élővilág milyen változatossága vesz minket körül. Az emberi mivoltunk alapeleme az a fajta tudatosság és előrelátás, amely képessé tesz minket arra, hogy a folyamatokat fölülről vizsgáljuk, és magunkra a rendszer egyik alkotóelemeként tekintsünk. Ez a fajta szemlélet azonban nem engedi meg, hogy különbséget tegyünk a különféle információk fontosságát illetően, függetlenül attól, hogy az mely tudományágból származik.

Őseink tapasztalati úton próbálták megérteni a világ működését, a világ számukra nem volt sokkal nagyobb annál, mint amit láttak belőle, a természetes folyamatok megértése rendezetlen információk alapján sok téves felismeréshez vezetett. A tudomány tanított meg minket a körülöttünk lezajló történések mélyreható elemzésének és értelmezésének módjára. A tudomány tanított meg minket a bizonyítékok keresésére, rendszerezésére és

a tesztelések fontosságára, legfőképpen annak okán, hogy *“mivel tehetségünk van önmagunk becsapására, szubjektivitásunk nem uralkodhat szabadon¹”*. A számszerű formában gyűjtött adatok azonos elvek alapján történő elemzésének legfőbb érdeme ugyanis az, hogy a következtetések megfogalmazásában minimális teret enged a szubjektivitásnak. A dolgozatom jelmondata is a tudománynak ezt az attribútumát igyekszik tömören megragadni.

A tudományon belül a taxonómia az a tudományág, amely igyekszik választ találni a globális fajgazdagság kérdésére, valamint arra, hogy honnan kezdődik az egyik faj, és hol végződik egy másik. A standardizált módszerekkel gyűjtött mennyiségi adatok tesztelésével operáló, »bizonyíték alapú taxonómia« létjogosultsága éppen abban rejlik, hogy lehetőséget teremt a természetbeni mintázatok azonos ismérvek alapján történő felismerésére, valamint arra, hogy a döntésünkben ne a saját, szubjektív elvárásaink játsszák a döntő szerepet.

Karrierem során ezen megközelítés – azaz a taxonómia kvantitatív irányzatának alkalmazása – alapjaiban befolyásolta, hogy milyen célok motiváltak. Egy olyan, modern módszertanon alapuló taxonómia művelését és terjesztését tűztem ki célul, amely teret ad a taxonómiai eredmények széleskörű magyarázásának és az ismeretek átadásának. Egy ilyen alapon működő taxonómia segíthet a biológiai sokféleség gyors és a lehetőségekhez mérten objektív feltárásában.

Mit nyerhetünk a körülöttünk élő sokféleség megismeréséből? Sok millió példát² találhatunk arra, hogy más fajok hogyan élnek a környezet adta lehetőségekkel és küzdenek meg a kihívásokkal. A hangya kolóniák szociális szerveződésének vizsgálatából (rokoni konfliktusok, rokonszelekció, önszerveződés stb.) levezetett analógiák segíthetnek jobban megérteni az emberi közösségeket működtető legmélyebb összefüggéseket. Az önös érdeken túlmenően azonban a biodiverzitás feltárásának van még egy prózai oka: felelősek vagyunk a környezetünk – beleértve a biológiai sokféleség – épségéért és ennek első lépéseként muszáj megismernünk azt, amit meg akarunk védeni. A megszerzett tudásunkkal felvértezve, mi emberek vagyunk a legnagyobb hatással a bioszféra állapotára, és ezzel párhuzamosan mi magunk tehetjük a legtöbbet annak megőrzéséért.

¹ A dolgozat jelmondataul választott idézet a szerző fordításában. (Carl Sagan: A Universe Not Made For Us. www.wimp.com/carl-sagan-a-universe-not-made-for-us/)

² A földi fajszámot ma 4 és 30 millió közé teszik (Godfray 2002, Zhang 2008, Padial és mtsai. 2010)

3 A dolgozat felépítése

Az értekezés a karrierem eddigi szakaszának célkitűzéseit és eredményeit mutatja be rangos, nemzetközi folyóiratokban megjelent publikációimon keresztül. A dolgozat felépítéséhez kiválasztott munkák segítségével az elért eredményeken túl igyekszem bemutatni azt a szemléletmódot is, amely az eddigi tudományos tevékenységemet átszövi. A doktori mű úgy lett felépítve, hogy a fő csapásirányként megfogalmazott kvantitatív morfológiai elemekkel működő taxonómia alkalmazásával elért eredményeim, illetve az általam elvégzett módszertani fejlesztések háttere és folyamata jól nyomon követhető legyen. Az értekezés alapjául szolgáló tudományos munkák különböző megközelítésben, eltérő típusú (módszertani, taxonómiai, biogeográfiai) kérdésekre keresnek válaszokat, amelyek mindegyike, meglehetősen eltérő megközelítést igényel. Az említett heterogenitás miatt ezért az egyes fejezetekben egyenként, külön mutatom be az aktuális alfejezetekhez tartozó tudományos hátteret, a módszertant és az eredményeket.

A következő, 4. fejezet („Bevezetés és célkitűzések”) alfejezeteiben fel fogom vázolni, miért látok komoly potenciált a taxonómia modern, XXI. századi információ alapú tudományrá alakításában, és mi motivál az ehhez elengedhetetlen módszertani fejlesztésekkel való hozzájárulásra. Ugyanezen fejezet 4.2. alfejezetében tárgyalom a taxonómia jelen kori kihívásait, melyek meghatározzák a céljaim megalkotását (4.4. alfejezet), valamint szó esik a kvantitatív módszerek előnyeiről és hátrányairól is (4.3.4. alfejezet).

A doktori mű az 5. fejezet („Saját eredmények és megvitatásuk”) alfejezeteiben mutatja be az eredményeket. Az alfejezetek rendszerint egy vagy néhány közlemény adataira épülnek (ez alól kivétel az 5.4. fejezet, amely összesen öt, nagyon hasonló elvek mentén elkészült cikk eredményeit ötvözi).

A morfológiai jellegek adatsorokká alakításának, szabványosításának fontosságáról és módjáról, azaz a szemantikus fenotípus megközelítésről 5.1. fejezetben számolok be. A morfológiai adatokban rejlő mintázatok felismerésére általam készített módszer, az NC-klaszterezés (5.2. fejezet) bemutatása a módszertani leírásokon és tesztelési eredményeken túl azt a filozófiát is részletesen bemutatja, amellyel dolgozom.

Az ezt követő, 5.3. és 5.4. fejezetekben az új módszertani elemek biodiverzitás kutatásban adódó gyakorlati alkalmazásának lehetőségeit két, nagyon különböző ökoszisztéma (Európa és Madagaszkár) jellegzetes, de nem domináns hangyacsoportjának (*Temnothorax* és *Nesomyrmex* genuszok) taxonómiai feldolgozásán keresztül fogom bemutatni. Ez a két hangya genusz szembeszökő példáját adja az olyan közepesen fajgazdag csoportoknak, melyekben a morfológiai különbségek nem túlzottan nagyok, sőt, az esetek közel felében

beszélhetünk fajokról, melyeket a taxonómia hagyományos módszertanával nagyon nehéz felismerni.

Az 5.4.1. alfejezetben részletesen bemutatom az automatizált taxonómiai döntéshozásra is alkalmas NC-PART klaszterezési eljárást, amelyet kifejezetten morfológiai adatsorokban rejlő rejtett mintázatok felismerésére hoztam létre. Gyakorlati bemutatását a Malagasi hangyafauna példáján fogom megtenni.

Az eredmények összefoglalására a 6. fejezetben kerül sor, alfejezetenként részletezve a főbb komponensek jellemzőit.

Az irodalomjegyzékben külön kezelem az értekezéshez felhasznált, a PhD fokozat (2007) megszerzése után készült munkáimat, valamint a korábbi cikkeimet.

A mellékletekben a doktori műben tárgyaltak megértéséhez közvetlenül nem szükséges, de a tudományos teljesség miatt szükségesnek talált anyagokat közlöm a források pontos megadása mellett.

4 Bevezetés és célkitűzések

4.1 Motivációim

A motivációim között az első helyen szerepel a taxonómia modern, XXI. századi információ alapú tudománnyá alakításához elengedhetetlen módszertani fejlesztésekkel való hozzájárulás. Vízióm szerint a taxonómia – ez az élettudományok területén alapvető fontosságú tudományág – olyan virágzó tudományággá válhat, amely képes források nagyobb mértékű bevonására, és alkalmas lehet az új kutató generáció tagjait komoly tudományos perspektívát nyújtva megszólítani. Karrierem során a taxonómia kvantitatív irányzatának alkalmazása alapjaiban befolyásolta a célkitűzéseim kijelölését és végrehajtását. Egy olyan, modern módszertanon alapuló taxonómia jobbítása és művelésének terjesztése motivált, amely teret ad a taxonómiai eredmények széleskörű bemutatathatóságának és az ismeretek átadásának.

4.2 A taxonómia modern kori kihívásai

A taxonómia, az élő dolgok megismerésének, leírásának és rendszerezésének tudománya alapvető fontosságú a biológiai tudományokban. Eredete az ókori görög tudományokig nyúlik vissza, modern formájában azonban 1758 után beszélhetünk róla, amikor Linné bevezette³ azt a binomiális nomenklátúra alapú osztályozást, amelyet – bár apró változtatásokkal – napjainkig használunk.

A taxonómia hosszú története egyben a válságok és azokra sikeresen kidolgozott válaszok története is. Az első komoly kihívás a mindjárt a XIX. században jelentkezett, felszínre hozva az addig gyűjtött hatalmas információ mennyisége és minősége, azaz rendezettsége körüli hiányosságokat. A mai zsargonban ezt a jelenséget az első bioinformatikai válságként emlegetjük (Godfray 2002). Erre a problémára a megoldást egy olyan szabályrendszer bevezetése jelentette, amely megalapozta a modern taxonómiai munkák alapjait⁴, s melynek elvei a mai napig meghatározóak a taxonómiában.

³ Bár a kettős nevezéktan bevezetését és széles körű elterjesztését ma Linné munkásságának tulajdonítjuk, pusztán a történeti hűség kedvéért meg kell jegyezzük, hogy egy évszázaddal őt megelőzte egy svájci botanikus-anatómus, Gaspard Bauhin. Az említett tudós 1626-ban a *Pinax Theatri Botanici* című művében 6,000 növény fajt írt le, és rendszerezett (<http://edb.kulib.kyoto-u.ac.jp/exhibit-e/b07/image/1/b07l0016.html>). Ezt a kettős nevezéktant (genus, species) vette át Linné később a munkáiban. (forrás: https://en.wikipedia.org/wiki/Gaspard_Bauhin).

⁴ Ez a szabályrendszer képezi alapjait a ma elfogadott állattani (International Code of Zoological Nomenclature – ICZN, <http://iczn.org/>) és egyéb biológiai nevezéktan nemzetközi kódexeinek (International Code of Nomenclature for algae, fungi, and plants – ICBN, <http://icbn.org/>; The International Code of Nomenclature of Bacteria – ICNB, <http://www.ictvonline.org/>).

Napjainkban a taxonómia újabb kihívásokkal kell szembenézzen. Míg a régebbi korok biodiverzitás kutatásának ez a – hagyományosan, alapvetően fenotípusos adatokkal dolgozó – tudományág szinte kizárólagos diszciplínájaként szerepelt, a modern időkben számos újonnan létrehozott adatforrás (molekuláris filogenetika, ökológia, enzim vizsgálatok, bioakusztika, kutikuláris hidrokarbonát vizsgálatok) vesz részt a biológiai sokféleség leírásában. Ha összehasonlítjuk a fent említett tudományágakat a taxonómia hagyományos gyakorlatával, a legnagyobb különbségeket a kvantitatív adatgyűjtésben és kifinomult adatelemzésekben, illetve azok hiányában láthatjuk. Ezek a „fiatalabb” diszciplínák ugyanis döntően nagyméretű adatsorok robusztus analíziseiből vonnak le következtetéseket. A következőkben a taxonómia modern kori kihívásai közül négy – a taxonómia jobb integrációja szempontjából alapvető fontosságú – igényt vázolok fel, melyekre való megfelelés fajsúlyos szerepet kapott karrierem során. Vallom ugyanis, hogy ezen igényekre jó megoldást találva a taxonómia és egyéb szupraindividuális diszciplínák (evolúció biológia, ökológia stb.) közötti átjárhatósági akadályok nagymértékben elháríthatók, és ezáltal a taxonómiai eredmények jobb integrációja megvalósítható.

A taxonómiai eredmények és tudományos igényű közlésük két fő összetevőből áll. Az egyik alkotóelem maga a taxonómia, azaz a taxon, vagy taxonok határainak felismerése, amelynek alapján az egységeket körülhatároljuk és leírjuk. A másik komponens a nevezéktan, aminek segítségével felcímkézzük az általunk talált mintázat egyes elemeit, a taxonokat. Mindkét összetevő jobbítására, vagy akár gyökeres átalakítására léteznek törekvések, azonban, mivel a karrierem során én magam a taxonok felismerésének fejlesztésére fókuszáltam, a jelen munka a taxonómia a diverzitás felismerő részét veszi górcső alá. A dolgozat a nevezéktani problémákat, javaslatokat, valamint az ide vonatkozó vitákat a továbbiakban nem tárgyalja részletesen.

Az alábbiakban azokat a biodiverzitási mintázatok felismerését érintő módszertani jellegű kihívásokat foglalom össze, melyekre egy modern és sikeres XXI. századi, információ alapú taxonómia megteremtéséhez meg kell találnunk a válaszokat.

- a) Kvantitatív fenotípus adatok igénye. A taxonómiai eredmények más diszciplínákkal való integrálásának igénye felveti az adatok közvetlen összehasonlíthatóságának szükségességét. A modern taxonómia módszertanának, ezért, illeszkednie kell más, „fiatalabb” tudományágak (molekuláris filogenetika, ökológia, enzim vizsgálatok, bioakusztika, kutikuláris hidrokarbonát vizsgálatok stb.) eszközrendszeréhez, mert a diszciplínák megállapításainak közvetlen összehasonlítását csak ezen a módon lehet biztosítani. A nagy hagyománnyal bíró konvencionális taxonómia adatainak struktúrájának számára ez azonban komoly kihívásokat hordoz, mert bár a taxonómiai munkák során értékelt morfológiai karakterek skálája rendkívül

széles, az ilyen adatok, főként a kvalitatív jellegüknek köszönhetően, kvantitatív adatsorokkal nehezen integrálhatók. Egy olyan folyamat kidolgozására van tehát szükség, amely a fenotípusok pontos, kvantitatív alapú leírására képes. Ennek megvalósításához elsődleges fontosságú a jelenlegi adatforrások és a használt adatkezelési módok felülvizsgálata, jobbítása és a kifejezések szinkronizálása, valamint új elemzési technikák kidolgozása.

- b) Hagyomány és modernitás egyben. A fent említett tudományágak legtöbbje mindössze néhány évtizedet tudhat maga mögött, amely azt jelenti, hogy az őket körülvevő teljes infrastruktúra eredendően modern. A taxonómia ezzel szemben az egyik legrégebbi biológiai diszciplína, a XXI. század taxonómiai eredményei mellett ugyanúgy megtalálhatók az akár 250 éve leírt fajok és megállapítások. A taxonómiát övező szemlélet, valamint az infrastruktúra ennek megfelelően tükrözi annak a kornak a szellemét, amelyben kialakult. Ez a jellegzetesség komoly nehézségeket hordoz magában. Amikor a taxonómia megújításának igényéről beszélünk, akkor mindenképpen figyelembe kell vegyünk, hogy a taxonómia módszertanának új alapokra helyezését úgy kell megvalósítsuk, hogy az egyéb, fiatalabb tudományágakkal való legjobb integrálhatóság elérése mellett a korábbi, több mint 200 év hagyományos módszertanával elért tudományos eredményekkel való összeférhetőségét teljes mértékben meg tudjuk megőrizni. Ennek az örökségnek, azaz az eddig elért eredményeknek a megőrzése ugyanis elsődleges fontosságú (Godfray 2002), amennyiben nem akarjuk a biodiverzitás kutatását teljesen előről kezdeni⁵.
- c) A tudás transzferálhatóságának igénye. Egy másik probléma, amivel mi, taxonómusok gyakran szembesülünk, az új generáció képzési nehézségeiben és a gyakorlott szakemberek csoportok közötti váltásában, azaz a megszerzett tudás transzferálhatóságában jelenik meg. Egy taxonómus szakember képzése sok-sok évet, esetleg évtizedet igényel. Ebbe beletartozik a gyűjteménnyel való munka elsajátítása, a nomenklaturai szabályok megismerése, az új generáció evolúciós biológiai, ökológiai, biogeográfiai ismeretekkel való felvértezése és – amely a jelen gondolatmenet szempontjából meghatározó – a kiszemelt rovarcsoport taxonómiai döntéshozási mechanizmusához szükséges ismeretének elsajátítása. Ez a tudás rendkívül csoport-specifikus, amely taxonómiai munka széleskörű interpretálhatóságának és az ismeretek átadásának legnagyobb akadályát is jelenti egyben. A tudásnak ezen utóbbi szegmense csak mérsékelten transzferálható abban az esetben, ha a

⁵ Pusztán a teljesség kedvéért említem meg, de a jelen dolgozat keretei között részleteiben nem térek ki azokra a létező törekvésekre, melyek a taxonómia modernizációját a korábbi évszázadok eredményeinek felhasználása nélkül képzelik el. Ezen törekvések körüli vita tárgyalása nem tartozik a jelen munka témájához.

szakember a karrierjét egy másik csoport kutatásában kívánja folytatni. Az új csoportra vonatkozó releváns ismereteket újra meg kell szerezni, s addig, amíg azt magas szinten nem sikerült elsajátítani, csak behatárolt taxonómiai döntések meghozatalára képes. Kétségtelen tény, hogy egy taxonómus számára a kiszemelt csoport morfológiáját, ökológiáját, biogeográfiáját minden esetben ismerni kell. Azonban az általa közvetített tudás átadásának határfoka szabványos vagy széles körben elfogadott – és ez által sokak által érthető – minőségi (kvalitatív) módszerek alkalmazásával nagymértékben javítható.

- d) A taxonómiai munka meggyorsításának igénye. A taxonómia alapvető célkitűzése a biodiverzitás megismerése, ami a fajok felfedezésében, leírásában és rendszerezésében nyilvánul meg. A teljes földi fajszámot illetően jelenleg csupán becslések állnak rendelkezésre, melyek ezt a számot egy rendkívül széles intervallumban 4 és 30 millió közé teszik (Godfray 2002, Padial és mtsai. 2010, Zhang 2008). A jelenlegi ütemben zajló taxonómiai munka mellett, évente leírt átlagosan 18,500 állatfajjal (Zhang 2008) kalkulálva a legóvatosabb becslés által sugallt 4 milliós teljes fajszám tudományos igényű leírásához is több mint száz évre lenne szükségünk. Ha azonban figyelembe vesszük, hogy egyes becslések ennek teljes fajszámnak többszörösét valószínűsítik, akkor megállapíthatjuk, hogy a taxonómia fő célkitűzésének teljesítése a jelenlegi technikák nyújtotta lehetőségek között távolról sem tűnik megvalósíthatónak. Ezért, ha eltökéltek vagyunk a biodiverzitás teljes körű leírását illetően, akkor ennek végrehajtásához új koncepciókra, új módszerek kidolgozására van szükség (Yeates és mtsai. 2011). A technológia folyamatos fejlődésével, mint például a gyorsabb DNS szekvenálási technikák, komputer tomográfia, térinformatikai rendszerek és az egyre szélesebb körben elérhető internet elterjedésével ez a célunk reálisabbnak tűnik (Padial és mtsai. 2010). Ahhoz, azonban, hogy az alapvetően morfológiai módszerekkel operáló taxonómia lépést tudjon tartani az új technikák segítségével leírt taxonok nagy számával, olyan egységesen elfogadott és viszonylag könnyen elsajátítható módszertanra van szükség, amely az eredmények interpretálásának megkönnyítésén túl képes jelentősen felgyorsítani a biodiverzitás kutatását.

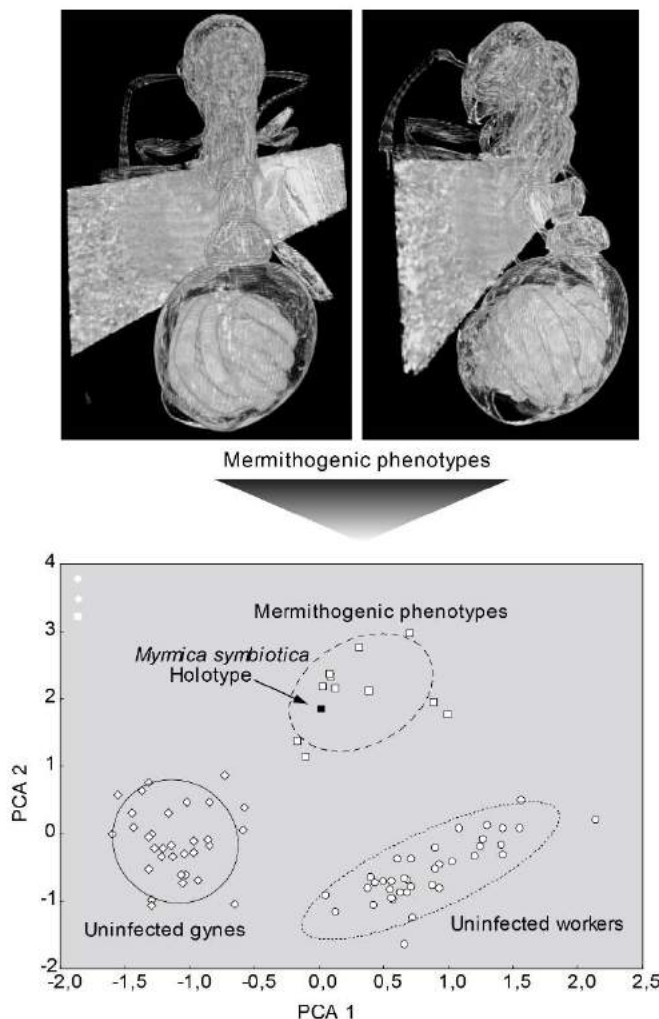
A taxonómia megújításának folyamatát a fent vázolt pontok figyelembevételével látom szükségesnek megvalósítani. A fentebb felsorolt igények mentén haladva igyekszem olyan választ találni, amelyek a fent említett kihívásoknak egyszerre képes megfelelni, és amely egy lehetséges alternatívaként elvezethet egy modern, információ alapú taxonómiához.

4.3 Az alkalmazott megközelítések

4.3.1 A fenotípusos adatok fontossága

Fenotípusnak nevezzük a genetikailag kódolt információ fizikai – más szóval mérhető módon – megjelenő formáját, mint például anatómiai jellegek, viselkedés, ökológiai preferenciák. A fenotípusos adatokból levont következtetések sokáig meghatározó súllyal vettek részt a biológiai ismereteink felhalmozásában, így az élettudományok megannyi területének fejlődésében alapvető érdeme van. Összességében, ezek az adatok képezik az életről alkotott tudásunk nagy részét.

A fenotípusos jellegek alkalmazásának a mai napig dominás szerep jut a biológiában. Többek között elsődleges adatként szolgálnak fajok elkülönítéséhez a taxonómiában, esetleg a törzsfejlődési mintázatok jobb megértésében, és nélkülözhetetlenek nagyszabású adatbázisok, például a *Tree of Life Web Project* (Burleigh és mtsai. 2013) létrehozásában is. A különféle szervezetek fenotípusainak vizsgálata komoly ösztönöző erővel bír olyan, interdiszciplináris tudományágak esetén, melyek az élőlények egyedfejlődését vagy a környezettel való kölcsönhatások jobb megértését tűzik ki célul.



1. ábra. Parazitogén *Myrmica* fajok morfometrikus és röntgen-mikrotomográfiás vizsgálata. A hangyák morfometrikus vizsgálata (alsó kép) külső alaktani jellegek alapján képes roncsolás mentesen és hiba nélkül osztályozni, és megtalálni a parazitogén fenotípusokat (*Mermithogenic phenotypes*). Valamint képes bizonyítani egyes típuspéldányok hovatartozását boncolás, vagy drága röntgen-mikrotomográfiás vizsgálatok (felső képpár) szükségessége nélkül is. (Csősz 2012 nyomán)

A mai napig a fenotípusos adatok szolgálnak a biodiverzitás kutatás alapvető forrásául, de az ilyen adatok nem csak a taxonómia, vagy az anatómia művelőinek hozhatnak új felfedezéseket, hanem az élettudomány legszétágazóbb interdiszciplináris területeinek is hasznos információ forrása lehet. Többek között fenotípusos vizsgálatok képezik a gerincét olyan úttörő kutatásoknak, melyek alapjául szolgáltak új antimikrobiális anyagok létrehozásának (Pogodin és mtsai. 2013), öregedési folyamatok megismerésének (Salmon és mtsai. 2009) vagy gabonanövények jellegeinek javításának (*National Plant Genome Initiative*: 2009–2013).

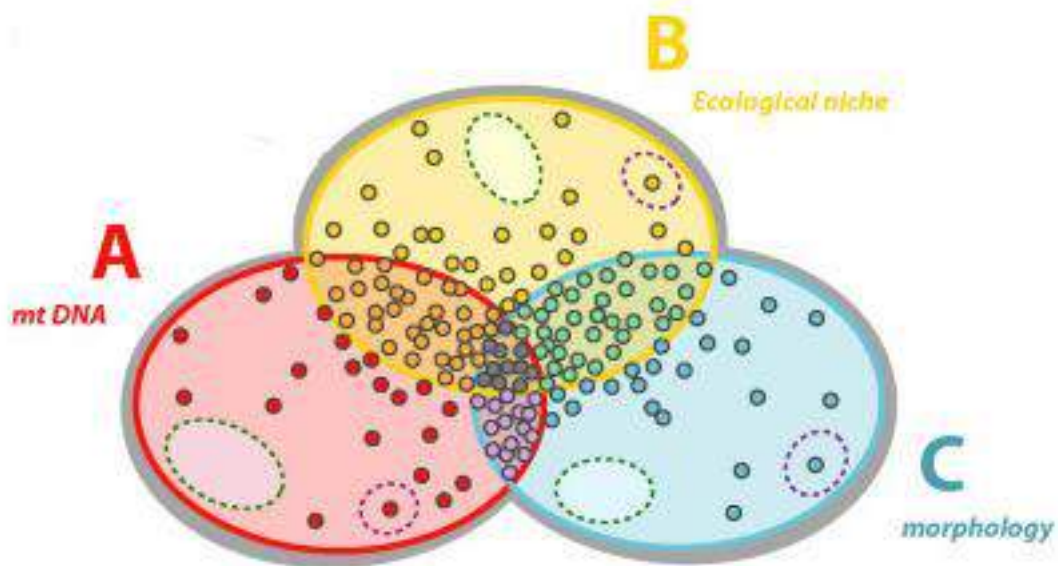
A betegségek által okozott fenotípusos változások gyakran genomikus vagy környezeti változó utáni kutatásra készítetnek minket, hogy akár kezelési vizsgálatok útján bepillantást nyerjünk a háttér folyamatokba. Azonban a parazitogén fenotípusok vizsgálata gyors és átfogó képpel szolgálhat (**1. ábra**) a betegség vagy a gazda-parazita kölcsönhatások természetét tekintve (ld. **Csősz & Majoros 2009**). Ezek a vizsgálatok sokszor tartogatnak nem várt eredményeket, amelyek előbbre vihetnek az egyedfejlődés jobb megértésében (Mollet és mtsai. 2012), vagy – visszakanyarodva a jelen dolgozat témájához, a taxonómiához – nem várt taxonómiai következtetésekre vezetnek (**Csősz 2012**).

4.3.2 A »bizonyíték alapú taxonómia«

Az alfa-taxonómia alapvető és egyben a legnagyobb kihívást jelentő feladata a biodiverzitásban rejlő mintázatok felismerése, azaz a fajok közötti határok felismerése és meghatározása. Ez a mintázatfelismerés hosszú évszázadokon keresztül pusztán kvalitatív morfológiai karakterek – gyakran szubjektív – elemzése útján valósult meg. A molekuláris technikák, mint a fehérje alapú (izoenzim) vizsgálatok (O'Brien & MacIntyre 1969, Pecsénye 1989, Pecsénye & Meglécz 1995) és az örökítőanyag, a DNS szekvenciák elemzéseinek megjelenése komoly változásokat indítottak el a taxonómiai munkafolyamatokban. Ezen modern technikák alkalmazásának elterjedésével megalakult egy napjainkban meghatározó irányzat, amit összefoglaló néven bizonyíték alapú taxonómiának (*evidence based taxonomy*) nevezünk. Ennek az új irányzatnak egy elterjedt formája az ún. integratív taxonómia (*integrative taxonomy*), amely összetetten értékeli a különböző megközelítések, mint morfológia, molekuláris filogenetika, ökológia, kutikuláris hidrokarbonát vizsgálatok, bioakusztikai adatok eredményeit és így alkot egy biodiverzitási mintázatokra vonatkozó egyesített hipotézist. A kifejezést egymástól függetlenül, szinte egyidejűleg vezette be Dayrat (2005) és Will és mtsai. (2005), természetesen némileg mást értve a kifejezés alatt. Az integratív taxonómiai megközelítések napjainkban erősen összetett taxonómiai, rendszertani és filogeográfiai problémák megoldásához is képesek átfogó és alapos megoldásokkal szolgálni (Cremer al. 2008, Schlick-Steiner és mtsai. 2010, Galimberti és mtsai 2012, Ward & Sumnicht 2012).

4.3.3 A morfológiai vizsgálatok használata az integratív taxonómiában

A morfológiai adatok sokáig a taxonómiai következtetések kizárólagos forrásaként szerepeltek. Az alaktani eltérések megfigyelésén alapuló megközelítés alapvetően arra a tapasztalatra épül, hogy a morfológiai különbségek és a génáramlás megszakadása között szoros – de nem kizárólagos – kapcsolat áll fenn. Az esetek döntő többségében a biológiai fajok határai pusztán morfológiai vizsgálatokkal is jól megállapíthatóak, viszont ismeretes, hogy a külső megjelenés fejlődéstani folyamatai környezeti hatások alatt is állnak, ezért célszerű lehet a morfológia mellett más megközelítések segítségével generált adatokat is bevonni a vizsgálatokba (**2. ábra**). Az integratív taxonómia leggyakrabban alkalmazott adatforrása, a morfológia mellett, a molekuláris filogenetika, azaz a DNS adatok.



2. ábra. Az integratív taxonómia sematikus megjelenítése Padial et al (2010) nyomán. A sárga (ökológiai niche), piros (mitokondriális DNS) és kék (morfológiai karakterek) színnel az alkalmazott megközelítési módokat látjuk. A halmazokban lévő pontok egy-egy elkülönítésre váró, önálló leszármazási vonalat (például fajokat) jelölnek. A megközelítések által alkotott halmazok átfedéseibe kerülő fajok esetében a független megközelítések az osztályozást tekintve egybehangzó állítást fogalmaznak meg. A különböző módszerek metszésében találhatóak azok a minták, melyek esetében az integratív megközelítés segítségével a legbiztosabb taxonómiai döntést hozhatjuk.

A molekuláris módszerek eredményeinek a taxonómiai döntéshozásban elfoglalt szerepe és figyelembe vétele hamar dominánssá vált a hagyományos morfológiai megközelítések megállapításaival szemben, és ez a trend napjainkban is észrevehető. Történik mindez annak ellenére, hogy az integratív taxonómia módszertanát Will és mtsai. (2005) elsődlegesen a *DNS barcoding*-gal szembeni kritikaként vezette be, és amelyben egy olyan taxonómiai folyamatot vizionált, amely a pusztán mitokondriális COI szekvenciáinak elemzéseiből kapott eredményekkel szemben képes több, független forrásból származó bizonyítékok együttes elemzéséből következtetéseket levonni. Will és mtsai. (2005) számára az alaktani jellegek vagy jelleg-kombinációk jelentetik az elsődleges adatforrást. Nézőpontjuk szerint a morfológiai karakterek összetettebb felépítése és feltételezhetően többgének öröklődésmenete sokkal biztosabb alapot nyújt fajhatárok megállapítására, mint – az akkoriban elérhető – rövid mitokondriális DNS szakaszok elemzése (ld. még Yeates és mtsai. 2011).

Napjainkra azonban az alaktani vizsgálatok jelentős teret veszítettek, aminek egyik okaként a módszertani fejlesztések elmaradását említhetjük. Éppen ezért, modern módszertani alkalmazások kidolgozása alapvető fontosságú a taxonómia jövője szempontjából. Új eljárások bevezetésével azonban lehetőségünk nyílik a morfológiai adatok széleskörűbb integrálására, ezzel képesek lehetünk egy értékes, függetlenül értékelhető adatsorhoz jutni, amely nélkülözhetetlennek bizonyulhat olyan, gyakorlatban előforduló esetekben (pl. igen régi példányok vizsgálata, subfosszilis, vagy fosszilis leletek értékelése), amikor molekuláris vizsgálatok nem, vagy nehezen alkalmazhatóak (Schlick-Steiner és mtsai. 2007a). Végül említsük meg a legfontosabb, morfológiai taxonómia mellett szóló érvet. A taxon nevek alkalmazását determináló elsődleges típus példányok egyedisége miatti szabályozások sok esetben kizárólag nem invazív vizsgálatokat engednek meg, ami kizárja a molekuláris mintavételezés lehetőségét. Ilyen esetekben csakis a morfológiai adatok teremthetik meg a közvetlen kapcsolatot a gyakorlati fajhipotézisek valamint az tudományos igényű névalkalmazás között (Steiner és mtsai. 2009, Bagherian és mtsai. 2012).

4.3.4 A kvantitatív módszerek előnyei és alkalmazásának korlátai

Mint minden módszertannak, a kvantitatív morfológia taxonómiai alkalmazásának is megvannak a maga előnyei és hátrányai, melyeket feltétlenül ismernünk kell ahhoz, hogy sikerrel alkalmazzuk őket. A jelen alfejezetben igyekszem rendszerezni azokat az általam fontosnak ítélt jellemzőket, amiket a közel két évtizedes tapasztalatom alapján fontosnak tartok megemlíteni.

4.3.4.1 *Miért a kvantitatív morfológiában látom a taxonómia fejlődésének kulcsát?*

A taxonómia modernizációjának több útja is lehetséges, melyek egymást kiegészítve vezetnek el a tudományág megújulásához. A számtalan lehetőség közül a magam részéről a fenotípusos jellegek mennyiségi interpretációjának jobbítását, és a biodiverzitás morfológiai adatokban rejlő mintázatokat felismerő módszerek fejlesztését választottam, mivel ez az irányvonal képes választ találni az 4.2. fejezet a-d pontjaiban vázolt kihívásokra:

- a) A fenotípusos adatok számszerűsíthetők, a folytonos skálán mért adatsorok – saját tapasztalatból is megerősített általános vélemény szerint – a legtöbb esetben normál eloszlást mutatnak, így a statisztikai elemzésük is egyszerűbb, mint egy nem normál eloszlású adatsornál.
- b) A fenotípusok kvantitatív alapokra helyezett vizsgálatának eredményei egyszerűen összehangba hozhatók a korábbi, kvalitatív morfológiai módszertannal elért eredményekkel. A kvantitatív morfológiai vizsgálatok közvetlen kapcsolatot jelenthetnek a gyakorlati fajhipotézisek valamint a zoológiai nevezéktan között (Steiner és mtsai. 2009, Bagherian és mtsai. 2012).
- c) Egy csoporton (rend, család, génusz) megtanult módszertan főbb elemei más csoportok esetében is alkalmazhatóak lehetnek, és a tesztek eredményei más tudományágaknál megszokott módon interpretálhatók. Mindez növeli a tudás transferálhatóságát, emellett hozzájárul az eredmények közlésének széleskörűbb megértéséhez és elfogadottságához.
- d) A kvantitatív morfológiai módszerek kitűnő alapul szolgálhatnak a taxonómia alapvető célkitűzéséhez, a biodiverzitás teljes megismeréséhez vezető folyamat meggyorsításához. Egy jól megválasztott elemzés ugyanis nagyban meggyorsíthatja a taxonómiai döntéshozás – hagyományos módszerekkel hosszadalmas és nagy tudást igénylő – folyamatát. Így a kvantitatív adatgyűjtés időigényessége ellenére a gyors döntéshozás eszköze lehet, mellyel összességében sok időt takaríthatunk meg.

A kvantitatív módszerek jelentős segítséget nyújthatnak abban, hogy a nagy diverzitású csoportokban gyors és pontos eredményeket érhessünk el. Bizonyos élőlénycsoportokban viszonylag gyorsan és viszonylag olcsón kivitelezhetőek, ezért olyan térségekben is számításba jöhetnek, ahol molekuláris vagy egyéb, drága módszerek elérhetők.

4.3.4.2 *A kvantitatív morfológiai megközelítések taxonómiai alkalmazásának korlátai*

A fentiekben részletesen tárgyaltam a morfometrikus módszerek alkalmazásának előnyeit, azonban a biztonságos alkalmazás érdekében szükséges megvizsgálunk a módszer korlátait is. Ezek a módszerek nyilvánvalóan nem nyújtanak gyógyírt a taxonómia összes

problémájára. Az elmúlt évszázadok során a taxonómia olyan mennyiségű adósságot halmozott fel, amelyek a teljes eddigi ismeretanyag használhatóságát, megbízhatóságát és magát a binomiális nomenklatúrát is veszélyeztetni kezdik. Az ilyen, főként nomenklatúrai természetű problémákra a kvantitatív módszertan nem ad válaszokat. A kvantitatív módszerek korlátozottan vagy nem alkalmazhatók nehezen kvantifikálható tulajdonságok, vagy nem homológ struktúrák vizsgálatára. Ilyen esetekben csak a tapasztalat és a konvencionális taxonómia módszertana nyújthat megoldást. A morfometrikus módszerek elemzése statisztikai tesztek révén valósul meg, tehát egy kritikus tömeget elérő mintaelemszám elérése létfontosságú. A kvantitatív taxonómia, tehát nem használható kis példányszámú anyag esetén sem.

A kvantitatív morfológiai eljárások elsősorban olyan esetekben jelentenek megoldást, ahol sok, egymástól alig különböző taxonnal dolgozunk, ahol az anatómiai struktúrák homológiája kellőképpen tisztázott. Olyan esetekben viszont, ahol a taxonok között nagy alaktani különbségek miatt (ez utóbbit a paleontológiában diszparitásnak nevezik) a struktúrák homológiája nem vizsgált vagy kérdéses, a kvantitatív módszerek téves következtetésekre vezethetnek. Ennek az az oka, hogy a kvantitatív vizsgálatok során a számszerűsített morfológiai struktúrákról fel kell tételezzük a homológiát, vagyis ugyanazon struktúra méretbeli változásait vizsgáljuk. Szélsőséges példa erre, hogy a cápafajok kvantitatív morfológiai elemzéséhez a hátszony méretei felhasználhatók, de magasabb szintű rendszer-tani vizsgálatoknál (pl. csontshalakat is bevonva) a homológiák hiánya miatt ugyanez a karakter már nem használható.

A mennyiségi morfológiai módszereknek tehát ugyanúgy, mint a molekuláris technikáknak is, megvan a helyük a világban, és ha a tudomány a megfelelő kérdéseket a megfelelő módszerekkel próbálja megválaszolni, akkor nagy segítséget jelenthetnek. Az általam javasolt optimális megoldást a 4.3.3. alfejezetben részletezett *integratív taxonómia* jelentheti, amely a módszerek, megközelítések kombinációjából von le következtetéseket, s egyben a módszerek keresztbe tesztelhetik egymást. Ilyen kombinált vizsgálatok egy alap-eleme lehet a morfometria. Összefoglalva, ezen előnyök és korlátok figyelembe vétele mellett, a kvantitatív módszerek a taxonómia valóban sok problémájára kínálnak megoldást. Komoly segítséget jelentenek a diverzitás gyors feltárása terén azokban az esetekben, ahol a fenotípusos jellegek jól számszerűsíthetők, a karakterek homológiája tisztázott, és megfelelő mennyiségű anyag áll rendelkezésre.

4.4 Célkitűzések: kvantitatív morfológiai megközelítések alkalmazása és népszerűsítése a taxonómiában

Én magam – csatlakozva ahhoz a taxonómusi közösséghez, amelyet *kvantitatív iskolaként* tartunk számon – a karrierem egy hosszabb szakaszát annak az elhatározásnak szenteltem, hogy gyakorlatban hasznosítható fejlesztéseket tegyek a morfológiai taxonómiai módszerekben, ezzel segítve a taxonómia megújítását.

Nézőpontom szerint egy morfológiai vizsgálat is lehet modern, ha modern megoldásokat alkalmazunk a megvalósítás során. A korai, a doktori (PhD) fokozatom megszerzése előtti időszakára jellemző munkáim (Csősz & Seifert 2003, Csősz & Markó 2004, Steiner és mtsai. 2006, Csősz, Schulz & Radchenko 2007) hagyományos morfometriás alkalmazásait felváltotta egy olyan feltáró jellegű elemzéstípus (Seifert, Ritz & Csősz 2014, Csősz & Fisher 2016a, 2016b) kidolgozásának és alkalmazásának igénye, mely képes nagyméretű, sokdimenziós adatsorban is megtalálni a biológiai jelentéssel bíró mintázatot (Csősz és mtsai. 2014, Csősz, Mikó & Heinze 2015, Csősz & Fisher 2015, Seifert & Csősz 2015). Mindezt kvantitatív morfológiai adatok sokváltozós analízisekkel történő elemzés segítségével.

A feltáró jellegű elemzések (*Exploratory Data Analyses – EDA*) családja pontosan ilyen esetekre lett kidolgozva, azaz hogy ismerjék fel az olyan adatsorokban rejlő mintázatokat, melyeket gyengén elkülönülő objektumok alkotnak, valamint, hogy előzetes hipotézis megalkotása nélkül adjanak betekintést többváltozós adatstruktúrákba. Napjainkban sok olyan algoritmus elérhető, melyeket feltáró jellegű elemzések céljára fejlesztettek ki, és sokat közülük kifejezetten morfológiai adatokban rejlő komplex mintázatok felismerésére dolgoztak ki (Goloboff és mtsai. 2006, Baur & Leuenberger 2011, Klingenberg 2011). A legtöbb módszer azonban vagy nem ábrázolja vizuálisan az eredményeket, így nem nyernék betekintést a finomabb adatstruktúrákba (pl. Expectation-Maximization algoritmusok), vagy az ábrázolható dimenziószám túl kevés egy összetett kép értékeléséhez (PCA), vagy az allometrikus összefüggéseket nem képes megfelelően kezelni, így a méret varianciából adódó természetes testarány eltolódás elfedi a létező különbségeket.

Alapvető célkitűzéseim közé tartozott a taxonómiai döntéshozás kvantitatív morfológiai adatokra helyezése. Egy olyan eljárás kidolgozása, amely amellet, hogy jelentősen növeli a felismert mintázat pontosságát és megbízhatóságát, a morfológiai taxonómia döntéshozatali mechanizmusát is nagyban meggyorsítja. Ez ideálisan egy olyan, algoritmus alapú módszer kidolgozását jelentette, amely képes automatizált morfológiai fajhipotézis alkotására, akár szakember előzetes hipotézisalkotása nélkül is.

Első lépésben egy olyan, meglévő elemek kombinációból alkotott eljárás kidolgozását céloztam meg, amely a felsorolt problémákat kezelni képes. Ennek a törekvésnek eredményeként egy új eljárást „*Nest-Centroid clustering*” (Fészek-Súlypont klaszterezés, melyet a továbbiakban a dolgozatban rövidítve NC klaszterezés-ként fogok említeni) néven vezettem be, saját elgondolás és kezdeményezés alapján (Seifert, Ritz & Csősz⁶ (2014) *Myrmecol News* 19: 1-15.). A módszer extenzív tesztelési folyamatainak eredményei alapján (ld. részletesen később, az 5.2.5. alfejezetben) kijelenthető, hogy elméletileg bármely euszociális szervezet populációira – úgy mint hártváscsárnyúak (hangyák, méhek, darazsak), természetes kolóniái, gubacsképző tetvek, stb. – alkalmazható. Ezeken felül az NC-klaszterezés általánosságban minden olyan összetartó rendszernél működik, melyekben igazolhatóan azonos fajú elemek – például ugyanazon a növény levelei és virágai, a korall "fej" genetikailag azonos polipjai, levéltetű telepek által termelt egyetlen fundatrix – az elemzés számára kielégítő mennyiségben előfordulnak. A módszert megjelenése óta rangos folyóiratokban publikált, integratív taxonómiai munkákban használják morfológiai adatok elemzésére (Derkarabetian & Hedin (2014) *PLoS ONE* 9: e104982.; Gratiashvili (2014) *Org. Divers. Evol.*, 14: 259-267.; Wachter és mtsai. (2015) *Mol. Ecol.* 24: 863-889.), de morfológiai alkalmazásokon túl kutikuláris hidrokarbonát vizsgálatoknak is megbízható módszere lehet (Guillem és mtsai. (2014) *Current Zool.*, 60: 62-75.; Guillem és mtsai. (2016) *J. Chem. Ecol.*, 42: 1052-1062.). Én magam a módszer 2014-es bevezetése óta elvégzett revízióimban 4 euópai és afrikai genusz 73 valid fajtát kezeltem faji rangon, melyekből 50 bizonyult tudományra új taxonnak (ezek többségéről az 5.3. és az 5.4. fejezetekben számolok be részletesen). Ezek közül becsléseim alapján 25-35 taxon tartozhat tipikusan a *kriptikus fajok* (Seifert 2009) kategóriájába⁷, tehát amely taxonokat az NC-klaszterezés alkalmazása nélkül nagy valószínűséggel pusztán külső morfológiai vizsgálatok alapján nem tudtam volna felismerni.

Az NC klaszterezés továbbfejlesztéseként terveztem egy újramintázásos alkalmazást, amelynek működése, legfőképpen a drámaian megnövekedett számítási kapacitásigény miatt, nem váltotta be reményeket. Később, a módszer további automatizálása felé egy váratlan helyről érkezett kérés. A *California Academy of Sciences*-ben elnyert pályázatom idejének első szakaszában merült fel egy olyan, könnyen kezelhető, gyors és egyszerű hipotézis-alkotó algoritmus igénye, amellyel kevésbé képzett – esetünkben madagaszkári – kollégák gyors és megbízható következtetést tudnak levonni a fajok számára vonatkozóan.

⁶ Levelező szerzőként.

⁷ A kriptikus fajok olyan, egyéb biológiai információ által megerősített fajok, melyek morfológiai elkülönítése hagyományos kvalitatív vizsgálatokkal nagyon nehéz vagy lehetetlen. Kriptikus fajok jelenlétére és azonosítására molekuláris, ökológiai, viselkedési vagy morfometrikus adatsorok elemzéséből következtethetünk.

A feladat végrehajtásához egy R-ben már meglévő és az NC-klaszterezéssel együtt kombinálható ún. rekurzív partíciónáló alkalmazás (PART, Nilsen & Lingjaerde 2013) nyújt segítséget. Ezzel kiküszöböltem az NC-klaszterezésnek azt a hiányosságát is, hogy bár látjuk a mintázatot, az adatsorban rejlő klaszterek számára nem kapunk becsléseket. Ezt a kombinációt szintén extenzív tesztelési folyamat után – a hozzá tartozó R-script-tel együtt – publikáltam (Csősz & Fisher (2016a) *PeerJ* 4:e1796. 35 p., <https://doi.org/10.7717/peerj.1796>; Csősz & Fisher (2016b) *PLoS ONE* 11(4): e0152454. 31 p., <http://dx.doi.org/10.1371/journal.pone.0152454>). Az eljárás még új, de alkalmazása egyre elterjedtebb (Rakotonirina, Csősz & Fisher 2016, 2017), és várhatóan a madagaszkári hangya diverzitás kutatás következő éveinek fontos eszköze lesz.

A módszertani fejlesztések mellett komoly hangsúlyt helyeztem magára a taxonómiai munkára is. Taxonómiai munkáimban a karrierem során összesen 121 Palearktikus (69 faj) és trópusi (52 faj) taxont kezeltem faji rangon, ebből 14 taxon státuszát változtattam meg és 59 tudományra új fajt írtam le (ebből 17 Palearktikus, 42 trópusi), ezen felül 20 új szinonímiát javasoltam munkáimban. A biodiverzitásra vonatkozó legfőbb célkitűzésem a szélesebb régióink, a Nyugat Palearktis hangya faunájának jobb megismerése (Csősz & Seifert 2003, Csősz & Markó 2004, Schlick-Steiner és mtsai. 2006, Csősz és mtsai. 2007, Csősz & Schulz 2010). Ezen belül is kitüntetett helyen szerepel a Kárpát-medence faunájának a Ponto-Kaspi régióval való kapcsolatának kutatása (Csősz és mtsai. 2014a, 2014b, 2015). A Nyugat Palearktikus taxonok közül jellemzően – de nem kizárólag – Myrmicinae hangyák taxonómiai, (mint a *Leptothorax*, *Messor*, *Myrmica*, *Temnothorax*, *Tetramorium* generumok) feltárásával foglalkoztam. A karrierem későbbi szakaszában lehetőségem nyílt komolyan hozzájárulni a biodiverzitás egy trópusi szegmensének, a madagaszkári hangyafaunának a feltárásához (Csősz & Fisher 2015, 2016a, 2016b, 2016c, 2016d, Rakotonirina, Csősz & Fisher 2016, 2017, Rasoamanana, Csősz & Fisher 2017).

Az általam revideált csoportok taxonómiai tisztázásán túl a *Madagascar Biodiversity Centre* (<http://www.madagascarbio.org/>) munkatársainak morfometrikus és többváltozós statisztikai képzését tekinthetem fontos hozzájárulásomnak a taxonómiai célkitűzések mihamarabbi megvalósulásához ott, ahol arra a legnagyobb szükség mutatkozik.

5 Saját eredmények és megvitatásuk

Ebben a fejezetben azon protokollok vagy kvantitatív morfológiai statisztikai alkalmazás bevezetését fogom alfejezetenként bemutatni, melyekkel hozzájárultam a biodiverzitás kutatásának fejlesztéséhez, a fenotípusos adatok jobb integrálhatóságához, és a taxonómiai döntéshozási mechanizmus javításához.

Ezek nagyobb része saját kezdeményezés, és saját elgondolás alapján született munkákon alapulnak (Seifert, Ritz, **Csősz** (2014) *Myrmecological News*, 19: 1–15., **Csősz** & Fisher (2016a) *PeerJ* 4:e1796. 35 p., <https://doi.org/10.7717/peerj.1796>; **Csősz** & Fisher (2016b) *PLoS ONE* 11(4): e0152454. 31 p., <http://dx.doi.org/10.1371/journal.pone.0152454>), vagy meglévő, széleskörű nemzetközi kezdeményezéshez csatlakoztam résztvevőként (Deans AR, Lewis SE, Huala E, Anzaldo SS, Ashburner M, Balhoff JP, Blackburn DC, Blake JA, Burleigh JG, Chanet B, Cooper LD, Courtot M, **Csősz** S, és mtsai. (2015) *PLoS Biol* 13(1): e1002033. <http://dx.doi.org/10.1371/journal.pbio.1002033>).

A dolgozatban a protokollok elméleti bemutatásán túl saját, kifejezetten taxonómiai témájú munkáimból merített gyakorlati alkalmazásokon keresztül fogom alátámasztani azok alkalmazhatóságát. Ezzel együtt be fogom mutatni azt a kvantitatív taxonómiai felfogást, (pl. **Csősz**, Mikó & Heinze 2015, **Csősz** & Fisher 2015.) amelynek segítségével, sikerrel bírkóztam meg a sok hangya taxonómus által a legnehezebbnek tartott Palearktikus *Tetramorium* és *Temnothorax*, valamint a madagaszkári *Nesomyrmex* genuszok taxonómiai feldolgozásával.

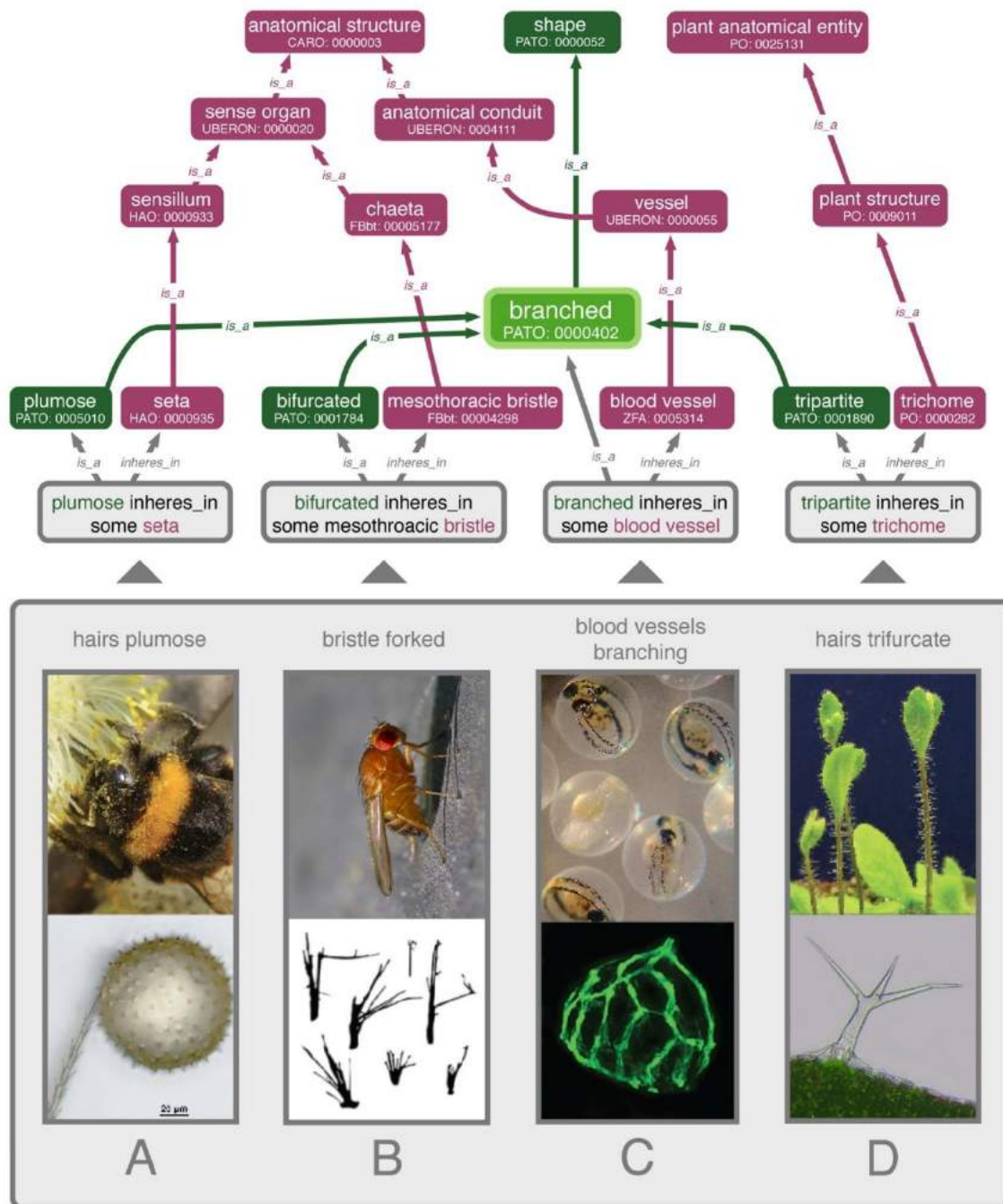
5.1 A fenotípusos jellegek adatsorokká alakítása és szabványosítása – a Hymenoptera Anatomy Ontology portál

A jelen alfejezet gerincét a Deans AR, Lewis SE, Huala E, Anzaldo SS, Ashburner M, Balhoff JP, Blackburn DC, Blake JA, Burleigh JG, Chanet B, Cooper LD, Courtot M, **Csősz** S, és mtsai. (72 szerző). (2015) *PLoS Biol* 13(1): e1002033. doi:10.1371/journal.pbio.1002033, <http://dx.doi.org/10.1371/journal.pbio.1002033>; **Csősz**, Mikó & Heinze (2015) *PLoS ONE* 10: e0140000. 62 p., <http://dx.doi.org/10.1371/journal.pone.0140000/> és **Csősz** & Fisher (2016a) *PeerJ* 4:e1796. 35 p., <https://doi.org/10.7717/peerj.1796> tudományos cikkek adják. Napjainkban a fenotípusos adatok szerkezetének egységesítésére, így a tudományterületek közötti átjárhatóság felé, komoly és széleskörű erőfeszítések történnek. Az adatok jelenleg elfogadott formája, szerkezete, közlésének és tárolásának módja drasztikusan csökkenti a produktív adatelemzés hatékonyságát. A manapság széleskörűen elterjedt gyakorlat szerint a szűkebb tudományterületek normái alapján leírt fenotípusok homológ egységei között komoly szóhasználatbeli vagy értelmezésbeli eltérések lehetnek. Ezért sok

esetben fennáll a veszélye annak, hogy a szerző és az olvasó más struktúrát ért azonos terminusokkal leírt jellegek alatt, vagy éppen ellenkezőleg, homológ jellegeket más szak kifejezésekkel definiál. Ez azokban az esetekben lehet különösen szembeötlő, amikor a szerző és az olvasó nagyon eltérő tudományterületen (például hangya morfológia – humán morfológia) szerezték addigi tapasztalataikat. Ez sokat ront a fenotípusos adatsorok általános értelmezhetőségén és az adatsorok integrálhatóságán, röviden fogalmazva a tudományos értékén.

Erre a problémára igyekszem nemzetközi együttműködésben modern válaszokat találni. Egy olyan megoldás kifejlesztésén dolgozom, amely döntően hozzájárulhat a – legnagyobb részt fenotípusos adatokon alapuló – taxonómia XXI. századi, információ alapú tudomány nyá emeléséhez. Nézőpontom szerint a fenotípusos adatok integrálhatósága terén tapasztalt leginkább szűk keresztmetszetet egy olyan, ember és algoritmusok által is értelmezhető, konszenzus alapú rendszer hiánya jelenti, amellyel a fenotípusos adatokat széleskörű biológiai (genetikai, környezeti, fejlődéstani, orvostani, evolúciós, ökológiai és rendszertani) kontextusba tudnánk helyezni. Napjaink felgyorsult információáramlása lehetővé teszi, de egyben meg is követeli a különféle tudományterületekről származó adatok együttes értelmezését és felhasználását. A különböző diszciplínákból származó eredmények jobb integrálhatóságának eléréséhez azonban elsődleges fontosságú a jelenlegi adatforrások és a használt adatkezelési módok felülvizsgálata, jobbítása és a kifejezések szinkronizálása. Egy olyan folyamat kidolgozására van tehát szükség, amely a fenotípusok széles körben értelmezhető, és szabatos leírására képes.

A jelen alfejezetben egy olyan interdiszciplináris adatkezelési elképzelést mutatok be, amely képes ezeknek a kihívásoknak megfelelni, és a fenotípusos adatokat képes megfelelő formában rögzíteni. Az említett fejlesztés már gyakorlatban működik, számos taxonómiai munkában (Mikó és mtsai. 2014, **Csősz**, Mikó & Heinze 2015, **Csősz & Fisher** 2015, stb.) felhasználásra került szabványosítva a fenotípusos karakterek leírását. Ez a fajta adatkezelési mód vonzó lehet széleskörű biológiai felhasználás számára, áthidalva a nehézségeket megkönnyíti az adatok elemzését. A fenotípusok kutatásának skálája rendkívül széles, de az ilyen adatok, főként a kvalitatív jellegüknek köszönhetően (**3. ábra**, alsó panel) tudományágak közötti kölcsönhatással viszonylag kevészer találkozunk. Tehát a fenotípusos adatok döntően kvalitatív jellegét lehet megnevezni a diszciplínák közötti átjárás legfőbb akadályának. Amennyiben a fenotípus adatok modernizálásának fejlesztése elmarad, a diszciplínák közötti átjárhatóság sem fog javulni, és ezzel rendkívüli felfedezések lehetőségétől eshetünk el a továbbiakban. Éppen ezért a fenotípusos adatok számszerűsítése és szabványosítása felé tett lépések a jövőre nézve rendkívül fontosak.



3. ábra. Hogyan ismerjük fel az elágazó fenotípusokat?

(Alsó panel) Természetes nyelvi eszközökkel nehezen megfogalmazható, különböző módon elágazó fenotípusok. (A) A méhek tollas mikroszkopikus felépítésű szőrszálai, melyek a pollengyűjtést teszik lehetővé, a legtöbb rovar-szőrszáltól könnyen elkülöníthetők. (B) Egy mutáns *Drosophila melanogaster* elágazó sörtéi. (C) A zebra dánió (*Danio rerio*) lárvák véredény elágazódásai az angiogenesis kezdetén. (D) Növényi „trichomák” számos fajtája megtalálható. [Folytatás a következő oldalon]

[A 3. ábra magyarázatának folytatása] (Felső panel) Az említett fenotípusok ontológiai eszközökkel leírt meghatározásai. A szemantikus gráfban a szabad megfogalmazások pontos anatómiai elnevezések konzisztens használatával fenotípusos leírássá lettek konvertálva. A leírások két komponensből állnak: egy anatómiai terminusból (lila) és egy ún. minőségi terminusból (zöld). A cédulák alján látható a terminusok ontológiai azonosító száma (pl. HAO 0000935). Deans AR, Lewis SE, Huala E, Anzaldo SS, Ashburner M, Balhoff JP, Blackburn DC, Blake JA, Burleigh JG, Chanet B, Cooper LD, Courtot M, Csősz S, és mtsai. (2015) *PLoS Biol* 13(1): e1002033. doi:10.1371/journal.pbio.1002033 nyomán.

A fenotípusos vizsgálatokkal ellentétben a genomikus adatok gyűjtéséhez, vizsgálatához és átláthatóságához világosan megfogalmazott stratégiák, jól fejlett közös módszerek, eszközök, szintaxisok állnak rendelkezésére. Szinte minden fajta genetikai elem szabványai pontos megfogalmazással, mindenki számára elérhető oldalakon, szabadon elérhetők (National Center for Biotechnology Information (NCBI) <http://www.ncbi.nlm.nih.gov>, European Bioinformatics Institute (EBI) databases <http://www.ebi.ac.uk/services>, DNA Data Bank of Japan (DDBJ) <http://www.ddbj.nig.ac.jp/>, Ensembl Plants <http://plants.ensembl.org>, stb). A genomikus adatokat emellett nagy adatbázisok gyűjtik és kezelik. Ehhez képest a fenotípusok számára sem a genomok vizsgálatához hasonló szabályozottság és szabványok, sem pedig egy hasonlóan nagy és konszolidált fenotípusos adatbázis nem állt rendelkezésünkre.

Annak ellenére, hogy a fenotípusos adatok tömegesen elérhetőek, és számuk egyre gyakorszik – ld. fajleírások elemei, digitális kép adatbázisok, fiziológiás mérések, teljes genom kiegészítés tanulmányok, nagy áteresztőképességű vizsgálatok, elektronikus egészségügyi nyilvántartások – ezek rendkívül heterogén adatsorok, és strukturáltságuk (leginkább szöveges forma miatt) nem felel meg a modern adat elemzési munkák céljainak. Ebből kifolyólag a fenotípusos adatokat – a manapság gyakran elérhető formájukban – nehéz összegyűjteni és interpretálni.

Összefoglalva, míg a fenotípus adatok lehetnek ugyanannyira komplexek, ugyanolyan izgalmasak és ugyanúgy értékes információval szolgálhatnak, mint a genomikus adatok, a fenotípusoknál nem rendelkezünk a molekuláris adatoknál megszokott, és hasonlóan szabványosított adatkezelési stratégiákkal. Érdemes megemlíteni, hogy a fajleírások fenotípus adatainak egy jelentős része valóban megfelel a formai követelményeknek, megbízható, kereshető és számítástechnikailag elemezhető formában állnak rendelkezésre. Azonban még itt is sok az eltérő felfogású fenotípus megközelítés – beleértve a különböző definiálási stratégiákat és a terminusok eltérő használatát – ezért sajnos ezeket az adatokat sem mindig vagyunk képesek teljesen egységesen kezelni.

Mivel azonban a fenotípus adatok rendkívül sok, egyelőre kiaknázatlan lehetőséget rejtenek magukban, és a modern számítástechnikai kapacitások is lehetővé teszik komplex fenotípus adatok feldolgozását, eljött az ideje annak egy egész rendszerre kiterjedő fejlesztésnek, amely magában foglalja a szükséges eszközök és szttenderdek kialakítását. Számos – egymással együttműködő vagy független – projekt kezdett el dolgozni a területen, és fektette le a szükséges alapokat. Napjainkban egy nagy lehetőség előtt állunk, kihasználva a fenotípusos adatok elemzésére és kezelésére alkalmas új technológiák előnyeit, megalkothatunk egy átfogó, különféle tudományterületek által széles körben használható ún. *fenomikus*⁸ adatbázist, amennyiben készek vagyunk konzisztens módon, az újonnan lefedett alapokra helyezve megújítani a fenotípusos adatok kezelését és felhasználását.

Ez az a cél, ami életre hívta azt a közös együttműködést, amelyről a jelen fejezetben beszámolok: a *Hymenoptera Anatomy Ontology Portal*-t⁹, amely célkitűzésként ezt a hiányosságot igyekszik pótolni.

5.1.1 Egy fenotípus regisztráló környezet kialakítása

Hogyan alakítsunk ki egy olyan környezetet, amelyben a kutatók könnyen vonhatnak le következtetéseket fenotípus adatok felhasználásával, például a környezet és a genetika közvetlen kapcsolatából? Ahhoz, hogy ez az elképzelés ilyen nagy mennyiségű adat kezelése mellett valósággá váljon, négy követelménynek kell teljesülnie.

- a) A fenotípus leírásokat elemezhető formába kell alakítani és rögzíteni, amely általában magában foglalja az ontológiai szempontoknak megfelelő terminusok alkalmazását. Más szavakkal, az eredetileg szabad szöveg vagy adatforrások formájában található fenotípusos leírásokat szabványos formába kell önteni (ld. Uniform Resource Identifier [URI], vagy magyarul Egységes Erőforrás Azonosító). Minden szövegrészletet ez által tulajdonságokkal és egyéb terminusokhoz való kapcsolatokkal töltünk meg (**3. ábra**, felső panel).

⁸ A phenomics a biológia egyik új területe, amely magában foglalja a fenotípusos jellegek és jelenségek mérését. A kutatás egyik fontos területe mind minőségileg, mind kvantitatív módon javítja a fenotípusos jelenségek mérési képességét. Egyre gyakrabban alkalmazzák a funkcionális genomikában, a gyógyszerészeti kutatásokban, az anyagcsere-technikában, de leggyakrabban a filogenetikában találkozhatunk vele.

⁹ A National Science Foundation (NSF) által támogatott Hymenoptera Anatomy Ontology Project-nek kezdetektől fogva résztvevője vagyok.

Egy verbatim fenotípus rdf/xml lekérdezhető (queriable) reprezentációja a hozzá tartozó linkekkel.

```
"<!-- http://www.semanticweb.org/imiko/ontologies/2015/3/temnothorax\_semantic\_statements#phenotype\_28143 ->
```

```
<owl:Class      rdf:about="http://www.semanticweb.org/imiko/ontologies/2015/3/temnothorax_semantic_statements#phenotype_28143">
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="http://purl.obolibrary.org/obo/BFO_0000051"/>
      <owl:someValuesFrom>
        <owl:Class>
          <owl:intersectionOf rdf:parseType="Collection">
            <rdf:Description rdf:about="http://purl.obolibrary.org/obo/HAO_0001077"/>
            <owl:Restriction>
              <owl:onProperty rdf:resource="http://purl.obolibrary.org/obo/RO_0000053"/>
              <owl:someValuesFrom rdf:resource="http://purl.obolibrary.org/obo/PATO_0002295"/>
            </owl:Restriction>
            <owl:Restriction>
              <owl:onProperty rdf:resource="http://purl.obolibrary.org/obo/RO_0000053"/>
              <owl:someValuesFrom>
                <owl:Class>
                  <owl:intersectionOf rdf:parseType="Collection">
                    <rdf:Description rdf:about="http://purl.obolibrary.org/obo/PATO_0001359"/>
                    <owl:Restriction>
                      <owl:onProperty rdf:resource="http://purl.obolibrary.org/obo/BFO_0000051"/>
                      <owl:someValuesFrom>
                        <owl:Class>
                          <owl:intersectionOf rdf:parseType="Collection">
                            <rdf:Description rdf:about="http://purl.obolibrary.org/obo/HAO_0000904"/>
                            <owl:Restriction>
                              <owl:onProperty rdf:resource="http://purl.obolibrary.org/obo/RO_0000053"/>
                              <owl:someValuesFrom rdf:resource="http://purl.obolibrary.org/obo/PATO_0000402"/>
                            </owl:Restriction>
                          </owl:intersectionOf>
                        </owl:Class>
                      </owl:someValuesFrom>
                    </owl:Restriction>
                  </owl:intersectionOf>
                </owl:Class>
              </owl:someValuesFrom>
            </owl:Restriction>
          </owl:intersectionOf>
        </owl:Class>
      </owl:someValuesFrom>
    </owl:Restriction>
  </owl:intersectionOf>
</owl:Class>
```

- b) Ezeket a jelentéstanilag egységesített fenotípus adatokat (**3. ábra**, alsó panel) úgy kell tárolni, hogy az az interneten mindenki számára elérhető legyen, (pl. a jellemzően metaadatokra kialakított *Resource Description Framework (RDF)*).

- c) A harmadik feltétel egy sor olyan új, algoritmusos eljárás fejlesztése, amelyek segítik az adatok kereshető formában történő tárolását, valamint lehetővé teszik a felhasználók számára az adatok elemzését. A fenti szövegdobozban egy kereshető, verbatim formában tárolt fenotípus rdf/xml lekérdezhető reprezentációját mutatom be, amelyben minden egyes állítás egy általam létrehozott ontológiához van hozzárendelve. A verbatim fenotípusok Csősz és mtsai (2015) által közölt, teljes és részletes lekérdezhető, rdf/xml (queriable) reprezentációja a **9.1. mellékletben** megtalálható.
- d) Végezetül olyan algoritmusokra van szükség, melyek képesek értékelni az ontológiákban rejlő logikai kapcsolatokat, például képesek azonosítani (és kulcsszavak után kikeresni) hasonló fenotípusokat és/vagy azok specifikus genetikai vagy környezeti tényezőkkel való összefüggéseit.

5.1.2 Az ontológiai projektek jelen kori kihívásai

Számos példát ismerünk olyan rendszerekről, amelyek képesek voltak átalakítani a tudomány területüket. Például a Phenoscape Tudásbázis projekt (*Phenoscape Knowledgebase*, <http://kb.phenoscape.org/>) keretében természetes populációk fenotípusainak és génin-aktivált (*Gene disruption*) modell szervezetek fenotípusainak összehasonlításával kutatják a gének fenotípus meghatározó jellegének kísérleti vizsgálatra való alkalmasságát. Ezen két, korábban nem kapcsolódó, evolúciós és orvosi biológiai területről származó adatok egyesítésével olyan új lehetőségek nyíltak meg, amelyek gyakorlatilag automatizálják az evolúciós fejlődési (*Evo-Devo*) hipotézisek megalkotását. Hasonlóképpen, a Gramene Project (Younens-Clark et al 2010) kidolgozta a növény Jelleg Ontológiát (*Trait Ontology – TO*) a Mennyiségi Tulajdonság Lókusz (*Quantitative Trait Locus – QTL*) rögzítése céljából több haszonnövény, beleértve a rizs, kukorica, búza, esetében (Ni és mtsai. 2009).

A különböző projektek egymástól jelentősen eltérő céljainak ellenére figyelemre méltó az a tény, hogy a fenotípusos adatok, melyek összegyűltek a projektek során – miután minden kutatócsoport közös ontológiát (Id. szemantika) és adatkezelési stratégiát használt – teljes mértékben összehasonlíthatóak. Ez egy olyan eredmény, amely a szemantikus fenotípusos megközelítés előtti időszakban nem, vagy nagyon nehezen lett volna elérhető. Az általuk használt rendszerek tehát logikailag értelmezhetők, átjárhatók és a fenotípusos adatokat további beavatkozás nélkül össze lehet hasonlítani, vagy egyetlen adatsorrá alakítani. A szemantikus fenotípusos megközelítés még szélesebb elterjedéséhez azonban még le kell küzdeni néhány akadályt.

Először is, jelenleg a fenotípusos adatoknak csak egy nagyon kis szelete van ontológiai formában rögzítve. 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, amelyek meggyorsítják az ontológia fejlesztését.

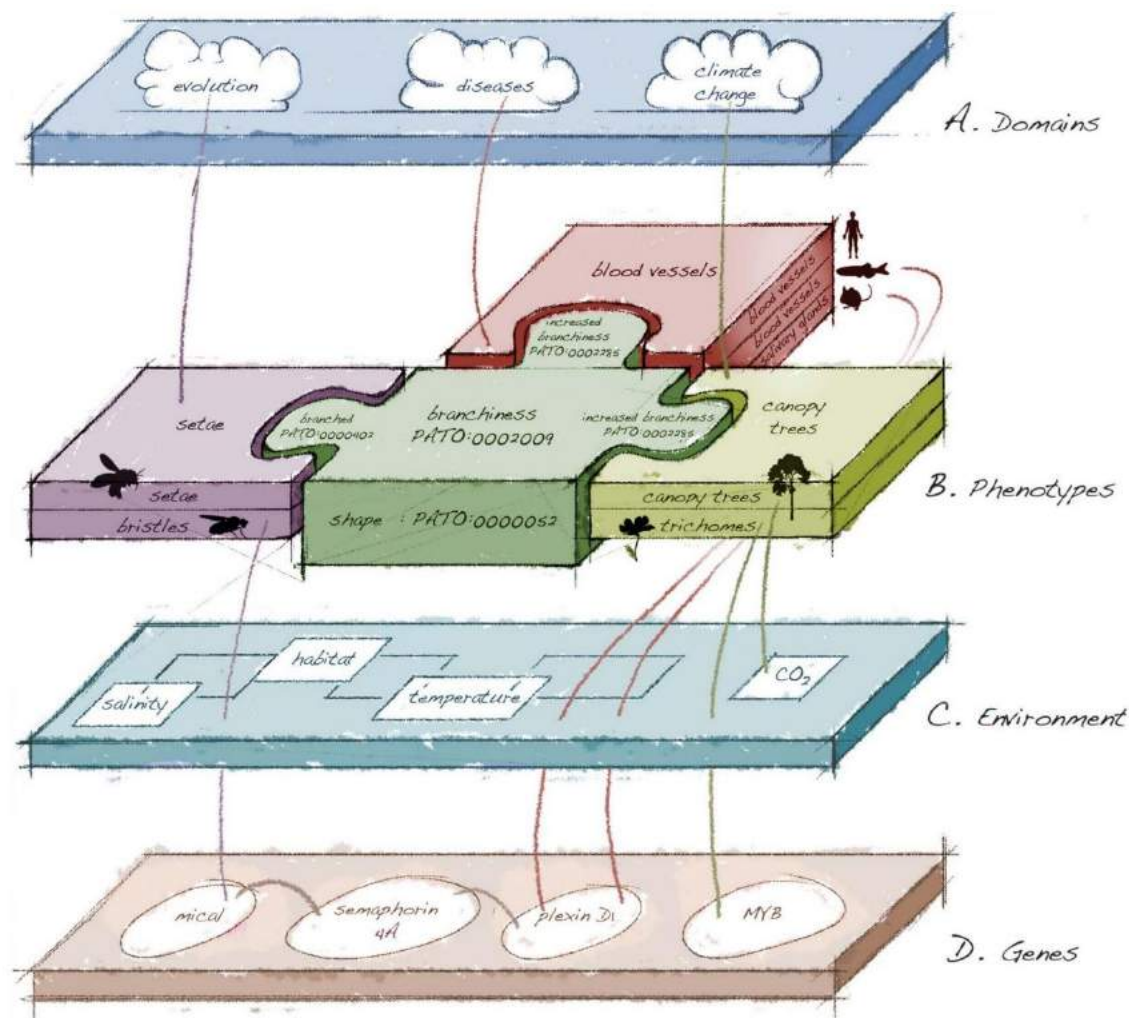
Másodszor, a fenotípusos adatok rögzítése – a jelenlegi módszerekkel – nagyrészt kézi erővel történik, így jelentős humán erőforrás ráfordítása szükséges az adatok szabványosított formában történő rögzítéséhez. Ezért a jelenleg rendelkezésre álló fenotípusok – és egyéb szöveg adatok – kinyerésére szolgáló félig automatizált stratégiákat tovább kell fejleszteni.

Legutolsó sorban, a régi adatok átalakítását is el kell végezni, amely bár időrabló, de egy-szeri befektetést igényel. Azonban ezt csak akkor lehetséges hatékonyan elvégezni, ha a jelenlegi és a jövőbeli projektek és közösen elfogadott szabványok együttesen fejlődnek, és aktívan hozzájárulnak a folyamatos fejlesztésekhez és karbantartáshoz. Így szükséges további szerzők, szerkesztők, kiadók és a finanszírozó szervezetek, valamint források bevonása a cyber-infrastruktúra fejlesztésének folyamatába.

5.1.3 Az adat integrálás elérése

Azok a kutatók, akik dolgoztak már multidiszciplináris megközelítésből származó biológiai adatokkal, tisztában vannak azzal, hogy több fajból és különböző publikációkból gyűjtött változó struktúrájú és koncepciójú adatok összevonása rendkívül nehéz feladat. A múltbeli próbálkozások nagyrészt korlátozott méretű adatsorok kézi integrálásával oldották meg ezt a problémát. A jelen helyzetben, a biológia sok területe támaszkodik genomikai és fenotípusos adatok integrálására, de az ilyen fajta egyesítés az elérhető és elemezhető fenotípusos adatok hiánya miatt zsákutcában van. A **4. ábra** egy művészi elgondolást mutat arról (illusztrátor: Anya Broverman-Wray, Id. Deans és mtsai. 2015) , hogy a logikailag strukturált fenotípus leírások különböző kutatási területek milyen milyen széles köréhez juttathatnak információkat – más szavakkal a diszciplínák milyen széles skálájában lehet a fenotípusos adatok és a diszciplínákra jellemző adatainak integrációjával felfedezéseket tenni. A jövőben a fenotípusos tulajdonságok tanulmányozása során egyre inkább szükség lesz nagy és gyorsan növekvő adattárakra, amelyekben csak automatizált vagy félig automatizált módszerekkel lehet hatékonyan keresni.

A taxonómia XXI. századi informatikai tudományként való megújításának egyik fontos eszköze lehet a szemantikus fenotípusos megközelítés, azaz a kalkulálható fenotípusos adatok bevezetése és széles körű elterjedése, amelynek a taxonómián túl akár az evolúciobiológia területén is nagy hasznát vennénk.



4. ábra. Látszólag egymáshoz közvetlenül nem kapcsolódó tudományágak kapcsolódási lehetőségei. Ahogy fentebb vázoltam, a fenotípus adatok kvantitatív formában történt rögzítésével a diszciplínák összekapcsolása és az adatok integrálása a legkönnyebben megvalósítható. (A) Diszciplínák: Különböző területek, mint az evolúciós biológia, az emberi betegségek és az orvostudomány, valamint az éghajlatváltozás melyek fenotípusos adatokat alkalmaznak. (B) A fenotípus: rovarok, gerincesek, növények, sőt az erdők mind tartalmaznak valamilyen formában elágazó jellegeket, melyek különböző sémák alapján vannak leírva, és így nehezen összehasonlíthatóak. A fenotípusos adatok egy számítógép számára csak akkor válnak alkalmazhatóvá, ha azokat egyedi azonosítóval, logikailag kapcsolt ontológiákkal látjuk el. Az ábrán a "shape" a "PATO" kód alatti "branchiness" fogalom egy iránymutató gyűjtőnév ami a "branched" és "increased branchiness" kategóriákat is magában foglalja. (C) Környezet: Változatos környezeti behatások befolyásolják a szervezetek fenotípusát. (D) Genetikai szint: különböző típusú elágazásos fenotípus karakterek ("branchiness") és a gének (pl. a semaphorinplexin jelátviteli hálózat) asszociációi. Forrás: Deans és mtsai. (2015) PLoS Biol 13(1): e1002033. doi:10.1371/journal.pbio.1002033 nyomán.

Ennek az irányvonalnak egyik úttörője a hártýásszárnyúak szemantikus fenotípusos megközelítésére fókuszáló Hymenoptera Anatomy Ontology projekt (Yoder és mtsai. 2010), amely, a kihasználva a gyorsan fejlődő számítógépes technikai hátteret, a fajok leírásakor használt fenotípus definíciókat explicit és kereshető formákban rendezi (ld. Csősz, Mikó & Heinze 2015, Csősz & Fisher 2015).

5.1.4 A szemantikus fenotípus megközelítés gyakorlati taxonómiai alkalmazásának példái

A szemantikai kijelentések hangyákra történő kialakítása az ontológiák fejlesztése mellett a világon elsőként a saját munkáimhoz kötődik. Ezért a megközelítés gyakorlati felhasználásának alapjait két, rangos folyóiratban közölt, szabad hozzáférésű munkáim (Csősz, Mikó & Heinze (2015) *PLoS ONE* 10: e0140000. 62 p., <http://dx.doi.org/10.1371/journal.pone.0140000/> és Csősz & Fisher (2016a) *PeerJ* 4:e1796. 35 p., <https://doi.org/10.7717/peerj.1796>) példáján fogom röviden bemutatni a következőkben. Munkáim során folyamatosan fejlesztem a hangyák alaktani adatainak ontológiai alkalmazását az 5.1.1. alfejezet a-c pontjaiban leírt szempontok figyelembevételével.

Abbr.	Label	Class genus differentia definition	Comments	uri
CL	maximum cephalic length in median view	The median anatomical line that extends between the posterior margin of the cranium and the distal margin of the clypeus in frontal view.	The maximum cephalic length in median view is not equivalent to the maximum cephalic size that extends between the posterior cranial margin and the distal clypeal line. The head must be carefully tilted to the position with the true maximum. Excavations of hind vertex and/or clypeus reduce CL (Fig. 1A).	http://purl.obolibrary.org/obo/HAO_0002331
CW	head width	The anatomical line that is the longest horizontal diameter of the cranium in frontal view.	The head width is the largest distance between the lateral margins of the compound eyes measured in frontal view (Fig. 1A).	http://purl.obolibrary.org/obo/HAO_0002268

1. táblázat. URI tábla részlete. Morfometrikus karakterek és morfológiai állítások hártýásszárnyú-specifikus terminológiái fenotípus-releváns ontológiákhoz kötve. A működő linkekkel ellátott teljes táblázat a **9.1. mellékletben** található. Csősz & Fisher (2016a) *PeerJ* 4:e1796. 35 p., <https://doi.org/10.7717/peerj.1796> alapján.

A szemantikus megközelítés kifejezetten a fenotípusos adatok felhasználásának széleskörű elősegítését szolgálja, és minél több csoport bevonása alapvető fontosságú. Az első hártácsszárnyúaknál megjelent munka, amelyben a taxonómiai leírásokra az újonnan kifejlesztett jelentési fenotípus modell került alkalmazásra, egy lapospotrohú darázs (Hymenoptera: Evaniidae) revízió során Balhoff és mtsai. (2013) nevéhez fűződik. E modell lehetővé teszi a fenotípus adatok széleskörűbb felhasználását, növelve a taxonómiai munka értékét.

Munkám során folyamatosan igyekszem finomítani és tisztázni a korábbi taxonómiai revíziók során használt fenotípus megközelítéseket, és megvizsgálom a meglévő adatokkal való újítások lehetőségeit (Csősz & Fisher 2016a). A leírásokban és kulcsokban használt morfológiai állítások hártácsszárnyú-specifikus terminológiája, valamint a fenotípuson alapuló diagnózisok ontológiák (Hymenoptera Anatómia Ontológia (HAO) URI táblázatban kerülnek csoportosításra (**1. táblázat**).

A Csősz & Fisher (2016a) működő linkekkel ellátott teljes eredményét eredeti, táblázatos formában a **9.1. mellékletben** csatolom. A taxonómiai nevezéktan, az *OTUk* (*Operational Taxonomic Unit*, Kezelhető Taxonómiai Egységek) definiálása és a természetes nyelvi fenotípusok összeállítása az *mx* adatbázis (<http://purl.org/NET/mx-database>) felületén történik, amelynek egyik tárgyi szerkesztőjeként feladatom a hangyák alaktani jellegeinek egységesítése szemantikus irányelvek alapján.

A taxon kezelések (latin név, leírók, illetve a verbatim állítások) lekérdezhető leírásai megtalálhatóak az *mx* web alapú könyvtár (<http://purl.org/NET/mx-database>) *taxonomic treatment* filejában. A munkám során a fenotípusok kezelését az *owl* nyelv egy leegyszerűsített szintaxával, a *manchester syntax*-szal állítottam össze.

A fent említett fenotípus *manchester syntax* által egyszerűsített reprezentációja (a hozzá tartozó linkekkel) a következő képpen néz ki:

„[has part](#) some ([vertex](#) and ([bearer of](#) some [areolate](#)) and ([bearer of](#) some ([rugose](#) and ([has part](#) some ([ruga](#) and ([bearer of](#) some [branched](#)))))))) ,,

A szemantikus fenotípus megközelítés alkalmazását és szerkezetét a részletesen a **9.2. mellékletben** Csősz, Mikó & Heinze (2015) alapján mutatom be.

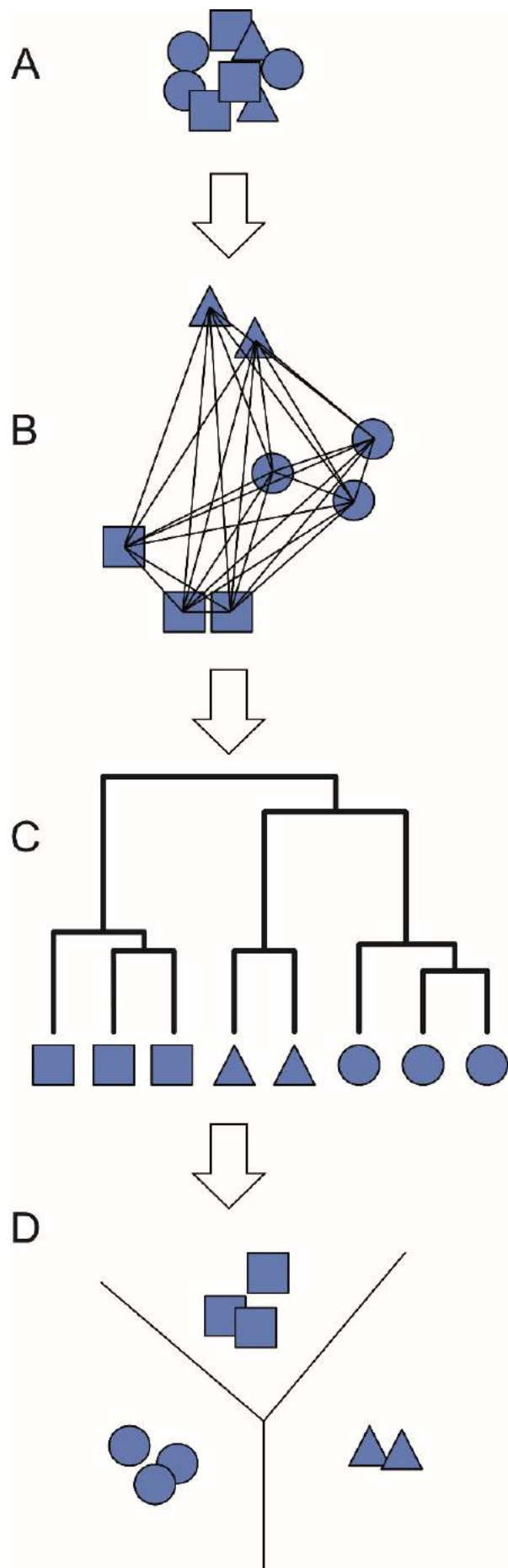
5.2 NC-klaszterezés – Egy új alkalmazás biodiverzitás morfológiai mintázatainak felismerésére

A karrierem egy hosszabb szakaszát annak az elhatározásnak szenteltem, hogy taxonómusok számára széles körben hozzáférhető és alkalmazható morfológiai módszereken alapuló fejlesztéseket tegyek. A következő fejezetben ennek a törekvésnek egyik eredményeként kidolgozott és közzétett Feltáró Adat Elemzés (*Exploratory Data Analysis – EDA*) alkalmazást mutatom be részletesen. A fejezetben leírtak legnagyobb részben Seifert, Ritz & Csősz (2014) cikkén alapulnak, helyenként új ismeretekkel, magyarázatokkal kiegészítve. Az új módszer saját kezdeményezéséből két, szakmai körökben elismert kolléga bevonásával született. B. Seifert a hangya morfometrikus vizsgálatok tapasztalt képviselője, a karrierje során lemért, több tízezer hangya morfometrikus adatával járult hozzá a vizsgálatokhoz, a teljes tesztelési adatsor közel 90%-a az ő anyagából állt össze; M. Ritz különféle startisztikai kérdések megoldásáért felelt (NMDS, K-means és ezek kombinálása), jómagam pedig az ötlet és a kivitelezés mellett a tesztelési adatsorok 10%-át adtam, valamint a statisztikai tesztelések NC-klaszterezési részéért feleltem¹⁰.

Az új eljárást „Nest-Centroid Clustering” (Fészek-Súlypont klaszterezés) elnevezéssel vezettettem be, vagy rövidítve *NC-klaszterezés*, utalva az eljárás alapelveire. A módszer elméletileg bármely euszociális szervezet – úgy mint hártáásszárnyúak (hangyák, méhek, darazsak), termeszek kolóniái, gubacsképző tetvek, stb. – populációira alkalmazható. Ezeken felül az NC-klaszterezés általánosságban minden olyan összetartó rendszernél működik, melyekben igazolhatóan azonos fajú elemek – például ugyanazon a növény levelei és virágjai, a korall „fej” genetikailag azonos polipjai, levéltetű telepek által termelt egyetlen fundatrix – az elemzés számára kielégítő mennyiségben előfordulnak. A módszer a kifejezett taxonómiai alkalmazásokon túl alkalmazható intraspecifikus állatföldrajzi mintázatok vizsgálatára is, viszont nem ajánlható filogenetikai következtetések levonására és hibridizációs esetek vizsgálatára. A bevezetett módszer legalapvetőbb újdonsága az egyik alkotóelemének, a diszkriminancia-analízis(LDA) rendhagyó alkalmazása: nem a fajhipotézis alkotja a csoportosítási faktort, hanem minden egyes hangya fészekminta tagjai vagy bármely egyéb egybefüggő rendszer elemei képeznek külön csoportot. Az eljárás alapelve egyszerű, morfológiai adatok traszformálása után keletkező távolságmátrixból történő klaszterezés során a hasonló morfológiájú minták egyazon klaszterbe kerülnek.

¹⁰ Levelező szerzőként

5. ábra. A bemutatott NC-clustering módszer lépéseinek sematikusan ábrázolt lépései. Nincsenek felismerhető fajhatárok az adatsorban a szélesen átfedő morfológiai jellegeknek köszönhetően (A). A Lineáris Diszkriminancia Analízis elkülöníti a fészekmintákat, mint csoportokat és megadja a csoportokra jellemző lineáris diszkrimináns értékeket (B), amely alapján a csoportok közötti távolságmátrix kiszámolható. Hierarchikus klaszterezés klaszterekre bontva ábrázolja az adatsorban található mintázatot: a legkevésbé eltérő (leginkább hasonló) esetek egyazon klaszterekbe kerülnek, a különböző morfológiájú esetek elkülönülnek (C). Azok a minták, melyek egyazon klaszterben találhatóak alkotják az elsődleges morfológiai faj-hipotézis alapját. Az NC-klaszterezés utáni ellenőrző lépésként (D) az elsődleges faj-hipotézis alapján felállított csoportosításra egy megerősítő elemzést (LDA) végzünk. (Seifert, Ritz és Csősz 2014 nyomán).



Az egész módszer motorja a kumulatív alkalmazott Lineáris Diszkrimináns Analízis (LDA) (5. ábra), ami minden fészekmintát (ugyanazon fészkekből gyűjtött egyedek, melyekről feltételezhetjük, hogy rokonok, gyakran testvérei egymásnak) külön csoportként ismer fel. A diszkriminancia-analízis a statisztikában, mintázat felismerésben és gépi tanulásban (*machine learning*) széles körben használt módszer, amely a független változók olyan lineáris kombinációját képes megtalálni, amely a függő változó alapján kialakított csoportokat a lehető legjobban elkülöníti. Az LDA tehát szétbontja az egymással gyakran korreláló – esetünkben morфомetrikus – változókat egy nem korreláló adatsorrá, amit lineáris diszkriminánsnak (*linear discriminant score*) nevezünk. Ez a fészek mintákra számolt nemsztenderdizált lineáris diszkriminánsok adattömegéből kalkulált euklideszi távolság-mátrix szolgál bemeneti adatként hierarchikus klaszterezéshez. A módszerhez egy összevonó (*agglomerative*) hierarchikus klaszterező algoritmust választottam, mely az irodalomban Agglomerative Nesting (AGNES) néven szerepel (Kaufman & Rousseeuw 1990). Az NC-klasztterezés előnye, hogy alkalmazásával nagymértékben lerövidíthetjük a taxonómiai döntéshozás hosszadalmas folyamatát, mivel az elsődleges fajhipotézis folytonos morфомetrikus változókból történő megalkotása teljesen algoritmosos alapokra van helyezve.

5.2.1 A mintázatfelismerő módszerek fejlesztésének szükségessége

Az itt bemutatott eljárás amellett, hogy jelentősen növeli a felismert mintázat pontosságát és megbízhatóságát, a morfológiai taxonómia döntéshozatali mechanizmusát is nagyban meggyorsítja. Először is vizsgáljuk meg, miben nyújt az eljárás mást, mint az eddig használatos módszerek. Itt szükséges visszautalni a korábban részleteiben tárgyalt kvantitatív taxonómiai módszerek általánosan jellemző előnyeire (4.3.4.1. alfejezet) és hátrányaira (4.3.4.2. alfejezet) is.

A konvencionális, morfológiai alapú alfa taxonómiai megközelítések során a taxonómusok kvalitatív morfológiai karakterek vizsgálatából alkotnak fajhipotéziseket, melyek során hasonló jellegekkel rendelkező példányok egyazon csoportba kerülve elkülönülnek más, eltérő jellemvonásokkal rendelkező esetektől. Ezek a csoportok széles körű feltételezések szerint megfeleltethetők biológiai fajoknak, vagy jól elkülönült populációknak, metapopulációknak. A fent leírt módszertan rendszerint morfológiai jegyek szubjektív kiválasztásán alapul, azaz a szakember – korábban megszerzett tapasztalatai alapján – maga dönt arról, hogy mely karakterek vagy karakterkombinációkat tartja döntőnek a csoportosításhoz. Ez a szubjektív alapokon történő csoportosítási stratégia jól működik olyan esetekben, mikor egyes morfológiai jellegekben található különbségek nyilvánvalóak. Ismertek azonban olyan helyzetek, mikor egyetlen tisztán diszkriminatív morfológiai karakter vagy karakterkombináció sem található, s ilyenkor az átfedő skálájú kvalitatív karakterek olyan bonyolult mátrixával vagyunk kénytelenek dolgozni, amelyek nagy tapasztalatú szakemberek

számára is komoly nehézséget okoznak. Nem beszélve az ún. *kriptikus* fajok létezéséről, mikor két vagy több – molekuláris módszerekkel igazoltan elszigetelt – faj elkülönítéséhez gyakorlatilag nem rendelkezünk értékelhető morfológiai támponttal. A kriptikus hangya fajok száma sokkal magasabb, mint ahogy azt néhány évtizede gondoltuk. A jelen ismereteink szerint megközelíti a teljes fajszám 50%-át. Néhány közepesen fajgazdag genusból példát merítve: az ismert 94 Palearktikus *Lasius* faj 46%-a, a 67 Palearktikus *Formica* faj 43%-a, és a világon ismert 77 *Cardiocondyla* faj 52%-a bizonyult *kriptikus* fajnak (Seifert 2009). A *kriptikus* fajok megjelenése ugyanilyen mértékben jellemző a trópusi területeken élő csoportokban is, a legutóbbi projektem során feldolgozott Malagasi *Nesomyrmex* fauna 42%-a esett ebbe a kategóriába (bővebben ld. később, a 5.4.4.2. alfejezetben). A *kriptikus* fajok hasonlóan magas arányával szembesülhetünk egyéb hangyagénuszok, vagy akár más rovarrendek esetében is (Ferreira és mtsai. 2010), habár a probléma különösen erősen jelentkezik a hangyák redukált morfológiájú dolgozó kasztjainak esetében.

A probléma megoldásának egyik széles körben alkalmazott módja lehet a kvantitatív taxonómia, amely morfológiai karakterek számszerű formában történő rögzítésével képes kezelni a problémát. Ennek során nagy adatsorok generálásával és statisztikai eljárásokkal tesztelhetjük a biodiverzitásról alkotott hipotéziseket. Ezzel lehetőséget kapunk statisztikailag elemezni a csoportok elkülönülését morfológiai jellegek skáláján, a többváltozós statisztikák – jellemzően Lineáris Diszkrimináns Analízis – használatával pedig több változó egyidejű elemzéséből tudunk a csoportok elkülönülésére következtetni. Ezek a megerősítő statisztikai algoritmusok nagyon megbízható módszerek, de hátrányuk, hogy egy megelőző, ún. *a priori* csoportosítást igényelnek, amely a taxonómus általi előzetes fajhipotézisen alapul. A hipotézis-vezérelt (*hypothesis-driven*) algoritmusok e csoportosítás alapján tesztelik az előzetes fajhipotézist. A probléma az, hogy annak függvényében, hogy mely jellegeket vagy jelleg kombinációkat tekintjük döntőnek az elkülönítésben, számos egészen eltérő csoportosítás lehetséges, amelyek valamilyen szintű, de statisztikailag értelmezhető megerősítést nyerhetnek. Ilyen esetekben – hacsak nem áll rendelkezésre egyéb, megbízható információ, pl. DNS, biokémiai analízis, bioakusztikai szignálok – a taxonómus számára komoly nehézséget okozhat megtalálni a legvalószínűbb megoldást.

Hogyan generálhatunk morfológiai alapon fajhipotézist szakember beavatkozása nélkül pusztán algoritmus segítségével? Az ilyen esetek megoldásához egy olyan analízis típusra van szükségünk, amely pontosan azon problémák megoldására fejlesztettek ki, ahol a kérdés nincs tisztán definiálva. A feltáró jellegű elemzések pontosan ilyen esetekre lettek kidolgozva, azaz ismerjék fel az olyan adatsorokban rejlő mintázatokat, melyeket gyengén elkülönülő objektumok alkotnak, valamint hogy előzetes hipotézis nélkül adjanak betekintést többváltozós adatstruktúrákba. Napjainkban sok olyan algoritmus elérhető, melyeket

feltáró jellegű elemzések céljára fejlesztettek ki, és sokat közülük kifejezetten morfológiai adatokban rejlő komplex mintázatok felismerésére dolgoztak ki (Goloboff és mtsai. 2006, Baur & Leuenberger 2011, Klingenberg 2011).

5.2.2 Az új megközelítés gyakorlati kivitelezése

A legkézenfekvőbb megoldás valamilyen klaszterelemzést használni az elemzésekhez. A klaszterezés egy olyan dimenziócsökkentő eljárás, amellyel objektumok sorát tudjuk homogén csoportokba sorolni, azaz osztályozni. Ezeket a csoportokat nevezzük klasztereknek. Az egyes klasztereken belüli esetek valamilyen paraméterüket tekintve jobban hasonlítanak egymáshoz, és ugyanezen paraméter mentén különböznek a többi klaszter elemeitől. A csoportosítás alapját különböző távolság- vagy hasonlóságmértékek képezik. Számos megjelenítési mód használható az EDA algoritmusok eredményeinek megértéséhez, így mint hisztogramok, szórásdiagramok vagy dendrogramok. Az első két megjelenítési mód nagyszerűen használható kevés csoport esetén, de a csoportok és a tengelyek növekvő számával ezek használhatósága, vagyis az adatsorokban rejlő struktúrák ábrázolásának értelmezhetősége rohamosan csökken. A dendrogramok azonban képesek könnyen érthető formában megjeleníteni bonyolult adatstruktúrákat, és alkalmasak gyors és átlátható csoportosítási hipotézisek kinyerésére (Rousseeuw 1986) akár nagy adatsorokból is.

A klaszterelemzés algoritmusai lehet hierarchikus vagy nem hierarchikus. A hierarchikus algoritmus az új klasztereket az előzőleg kialakított klaszterek alapján keresi meg, míg a nem hierarchikus algoritmus egyszerre határozza meg az összes klasztert. A hierarchikus klaszterező algoritmusok alkalmasak két klaszter egyesítéséről, jó döntéseket hozni mivel az összes pont páronkénti hasonlóságával kapcsolatos információt fel tudják használni. További előnyük, hogy ezek az algoritmusok tipikusan kis számítástechnikai kapacitást igényelnek, és sok statisztikai csomagban megtalálhatók. Hátrányuk, hogy egyes lépések véglegesek, azok nem vonhatóak vissza.

Általánosságban elmondható, hogy a hierarchikus klaszterezés jó eredményeket ad biológiai adatokon (Brock 2008). Talán ez is közrejátszik abban, hogy hierarchikus klaszterezés széles körben elterjedt feltáró módszer a legkülönbözőbb biológiai vizsgálatokban, pl. növényi genetikai diverzitás vizsgálatában (Odong és mtsai. 2011), különféle fűfajok morfológiai jellegeinek és molekuláris markereinek vizsgálata esetében (Tabacchi és mtsai. 2006), különféle gerinces és rovarcsoportok enzim adatainak összehasonlító vizsgálatában (Wiens 2000), skorpiók méreg molekuláinak genealogikus kapcsolatainak értékelésében (Nascimento 2006), vagy *Drosophila* legyek populációi közötti fehérje mintázatok megértésében (Feltens 2010).

A saját módszerünkhöz a mintázat értékelésének céljára a hierarchikus klaszterezési eljárások egyik típusát az *Agglomerative Nesting*-et (AGNES) választottuk ki, amelyet Kaufman és Rousseeuw (1990) dolgozott ki. Ez az összevonáson alapuló algoritmus először minden egyes elemet külön klaszternek tekint, és ezek összekapcsolásával az objektumokat egyre nagyobb klaszterekbe csoportosítja, míg a végén egyetlen, az összes elemet tartalmazó klasztert kapunk. 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. A hierarchikus klaszterező módszerek számára az objektumok távolságából álló mátrix jelenti a bemenő adatsort. Egy megfelelő távolság függvény alkalmazásával, ami segít az objektumokat optimálisan párosítani, az AGNES képes a leghasonlóbb eseteket klaszterekbe, azaz csomópontok alá rendezni egészen addig, amíg minden objektum klaszterezésre kerül (Day & Edelsbrunner 1984, R Development Core Team 2015). A módszertani cikkünkben (Seifert, Ritz & Csősz 2014) áttekintésre került az újonnan bevezetett NC-clustering morfológiai mintázatfelismerésre, azaz taxonómiai munkára történő alkalmazhatósága.

Az R »cluster« statisztikai csomagjában (Rousseeuw és mtsai. 2012) található hat távolság-módszerből két alkalmazást választottunk tesztelés céljára. Ezek a Ward módszer (Ward's minimum variance method, a hazai szakirodalomban Ward-féle eljárásnak nevezzük) (Ward 1963) és az UPGMA (*Unweighted Pair Group Method Arithmetic averages*, a hazai szakirodalomban a módszer teljes neve "súlyozás nélküli pár csoport módszer számtani átlaggal"). A Ward módszer használata nagyon elterjedt a klaszterezési eljárásokban. Lényege, hogy a pontoknak a klaszter középpontjaiktól vett négyzetes távolságaik összegét minimalizálja. Az UPGMA, a kifejezett stabilitás és következetesség (Struyf és mtsai. 1997) miatt, szintén népszerű módszernek számít. Az ilyen algoritmussal összevont objektum-párok együtthatói mindegyik klaszter méretétől egyaránt függenek, ez nagyfokú kiegyensúlyozottságot eredményez. A súlyozatlan eljárásokról, mint az UPGMA, általánosságban elmondható, hogy előnyt élveznek a súlyozott (pl. *WPGMA – Weighted Pair Group Method Arithmetic averages*) megközelítésekkel szemben, hacsak nincs okunk feltételezni, hogy a különböző csoportból származó objektumokból eltérő módon vettünk mintát, s ekkor a pontokat súlyoznunk kell. Esetünkben erről nincs szó, tehát súlyozatlan módszert használtunk. A tesztelési folyamat során a bevezetett módszer eredményeit 48 fajpárba rendezett 74 kriptikus hangya faj morfológiai adatsorán vizsgáltuk meg. Az eredményeket összehasonlítottuk két teljesen eltérő, de szintén népszerű eljárással, az oordináló nem lineáris sokdimenziós skálázás (*NMDS – Non-metric Multi Dimensional Scaling*) (Kruskal 1964, Legendre & Legendre 1998), és egy másik, nem hierarchikus ún, particionáló klaszterező módszerrel, a K-közép módszerrel (Lloyd 1982).

5.2.3 Anyag és módszerek

A tesztelés során 4844 fészekmintába tartozó több mint 10,000 dolgozó átlagosan 14 morfológiai karakterkészletéből, összesen 120,000 cella – saját magunk által gyűjtött – adatait elemeztük ki. Az általunk bevezetett és alább leírt módszer R statisztikai környezetben működik. Az általam közölt R script az alábbi oldalról szabadon letölthető: <http://sourceforge.net/projects/agnesclustering/>. A tesztelés során alkalmazott protokollt az alábbiakban részletezem.

5.2.3.1 A tesztelésre kiválasztott csoportok és a kiválasztás kritériumai

A tesztelésre kiválasztott faj-pároknak a következő szabályoknak kellett szigorúan megfelelni:

- a) A fajok morfológiai elkülönítésének nehéznek kell legyen. Minden rendelkezésre álló morfológiai karakter kimutathatóan átfedésben kell legyen a két faj populációi között. Ugyanígy a testméretben is hasonló átfedést várunk el.
- b) Az adatoknak meg kell feleljenek a normalitási követelményeknek.
- c) Minden fajpár esetében a teszteléshez legalább 40 fészek minta szükséges.
- d) Csak két faj van jelen a fajpárban. Azok az adatsorok, melyekben esetleg több faj is jelen lehet – a K-közép-el, illetve az NMDS-el való legjobb összevethetőség kedvéért –nem tekintetők megfelelő teszt-adatsornak.
- e) A fészek minták nagysága is lényeges szempont lehet, ezért az egyetlen dogozót tartalmazó „minták” számát az adatsor 50%-ában maximáltuk. (Későbbi vizsgálatok alapján kijelenthető, hogy a módszer a teljes minta számra vonatkoztatott 70%-os „egyke” aránynál is megfelelő eredményt ad – de ezzel a tesztelések ideje alatt még nem voltunk tisztában.)

Az olyan eseteket, melyeknél a hibrid minták esélye megvolt elővigyázatosságból kimaradtak az analízisből. Összesen 42 fajpár felelt meg ezeknek a kritériumoknak, melyeket 74 ún. szigorúan *kriptikus faj* alkotott. Hat olyan fajpár esetében, ahol nagyobb volt a klaszterezési hiba a morfológiai karakterek csökkentett számával egy kiegészítő elemzést is végeztünk. A csökkent karakter számot SPSS 15.0 programban Lineáris Diszkriminációs Analízisnél választható lépésenkénti regresszióval határoztuk meg. Az így kapott adatsorok az ide vágó referenciákkal az **2. táblázat**ban találhatók. A fent leírt tesztelési eljárásokon túl ugyanezt az eljárást került megismétlésre öt olyan esetben is, melyben a *Formica*, *Lasius*, *Temnothorax* és *Messor* genuszokba tartozó fajpárok közötti hibridizálódás jelensége ismert és dokumentált volt, és közzé is tettük.

5.2.3.2 Az adat mátrix előkészítése.

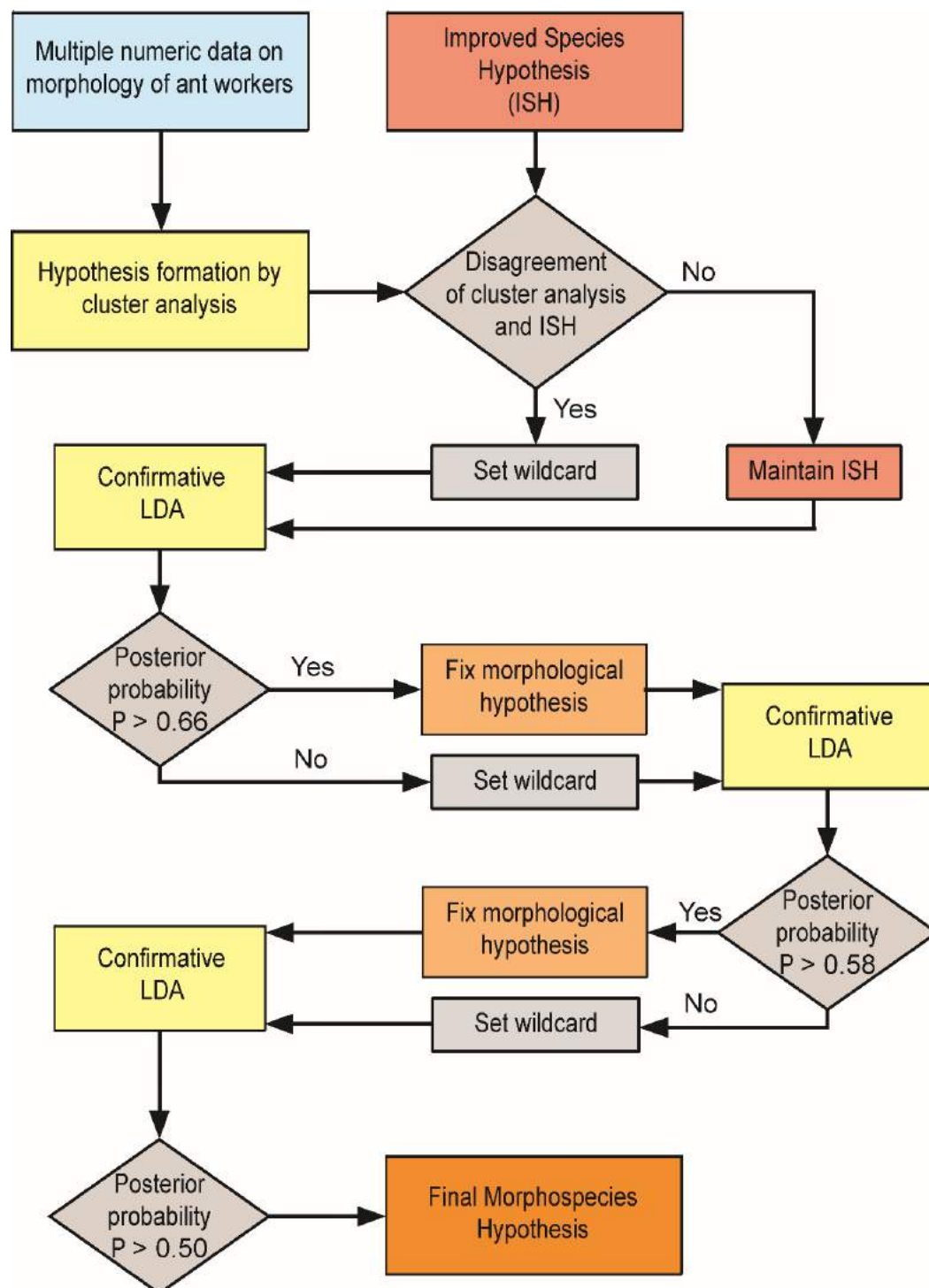
Ahhoz, hogy a Ward módszer, az UPGMA valamint az NMDS és a K-közép módszerrel végzett klaszterezés által kapott eredményeket összehasonlíthassuk, a bemenő adatokat ugyanolyan módon készítettük elő. Ennek megfelelően az NC-klaszterezés első lépéséből keletkező adatmátrix (azaz a fészekminták, mint csoportok, lineáris diszkriminánsainak mátrixa) képezte mind a négy elemzés típus bemenő adatsorát¹¹. Mind a négy módszer egy-egy fészekmintát jellemző lineáris diszkriminánsokból számolt Euklideszi távolság mátrixszal operál (5. ábra). Figyelembe véve, hogy egy fészekminta egytől tetszőleges számú dolgozóból állhat, kérdésesnek tűnhet ezek egyetlen objektumként való kezelésének helyénvalósága.

Ez a megközelítés azonban biológiailag könnyen magyarázható: elméletben (és legtöbbször a gyakorlatban is) minden hangya, melyet egyazon fészekben találunk ugyanannak a fajnak a képviselője, sőt, mivel sok esetben egy vagy néhány királynő utódairól van szó, testvérek vagy rokonok. Kivételt képeznek e szabály alól olyan szociális parazita, ún. *parabiotikus* esetek, mikor valóban több faj él egy fészekben. Ezeket a ritka eseteket azonban könnyedén, akár mikroszkópos vizsgálat nélkül fel lehet ismerni, mivel az ilyen közös fészekben nagyon eltérő fenotípusú egyedek élnek együtt.

5.2.3.3 Agglomerative Nesting (AGNES)

Az *Agglomerative Nesting* (AGNES) egy jól ismert, robusztus hierarchikus klaszterező algoritmus (Struyf és mtsai. 1997), amely az objektumok sorozatos összevonásával dolgozik. A folyamat egy olyan zéró ponton kezdődik, ahol minden objektum saját maga formál egy „klasztert”, majd az algoritmus összevonja a két legkevésbé különböző esetet egy klaszterbe, majd az összevonás lépései addig ismétlődnek, amíg végül minden objektum klaszterezésre nem kerül (Day & Edelsbrunner 1984, Legendre & Legendre 1998). Az általunk használt AGNES algoritmus a »cluster« csomagban (Maechler és mtsai. 2012) található és R környezetben futtatható. A »cluster« csomagban összesen hat távolság-módszer választható: "average" (*Unweighted Pair Group Method with Arithmetic Mean – UPGMA*), "single" (*single linkage*), "complete" (*complete linkage*), "ward" (*Ward's method*), "weighted" (*weighted average linkage*) és annak általánosított verziójaként megadható "flexible". Az alapértelmezett módszer az "average". Az elemzéseink során kétféle összevonási módszer eredményeit teszteltük, ezek a Ward módszer (Ward 1963) és az UPGMA (Sokal & Michener 1958).

¹¹ Ez alól csak a K-közép módszer a kivétel, ahol a transzformált adatokon végzett elemzés mellett illusztrációként szerepel az alapadatokkal végzett elemzés is, ahogy az a 2. táblázatban látható.



6. ábra. Folyamatábra a végleges morfológiai fajhipotézis megalkotásához. A tökéletesített fajhipotézisből (Improved Species Hypothesis) több lépésben alakítjuk ki a végleges morfológiai fajhipotézist. További magyarázat a szövegben. (Seifert, Ritz és Csősz 2014 nyomán).

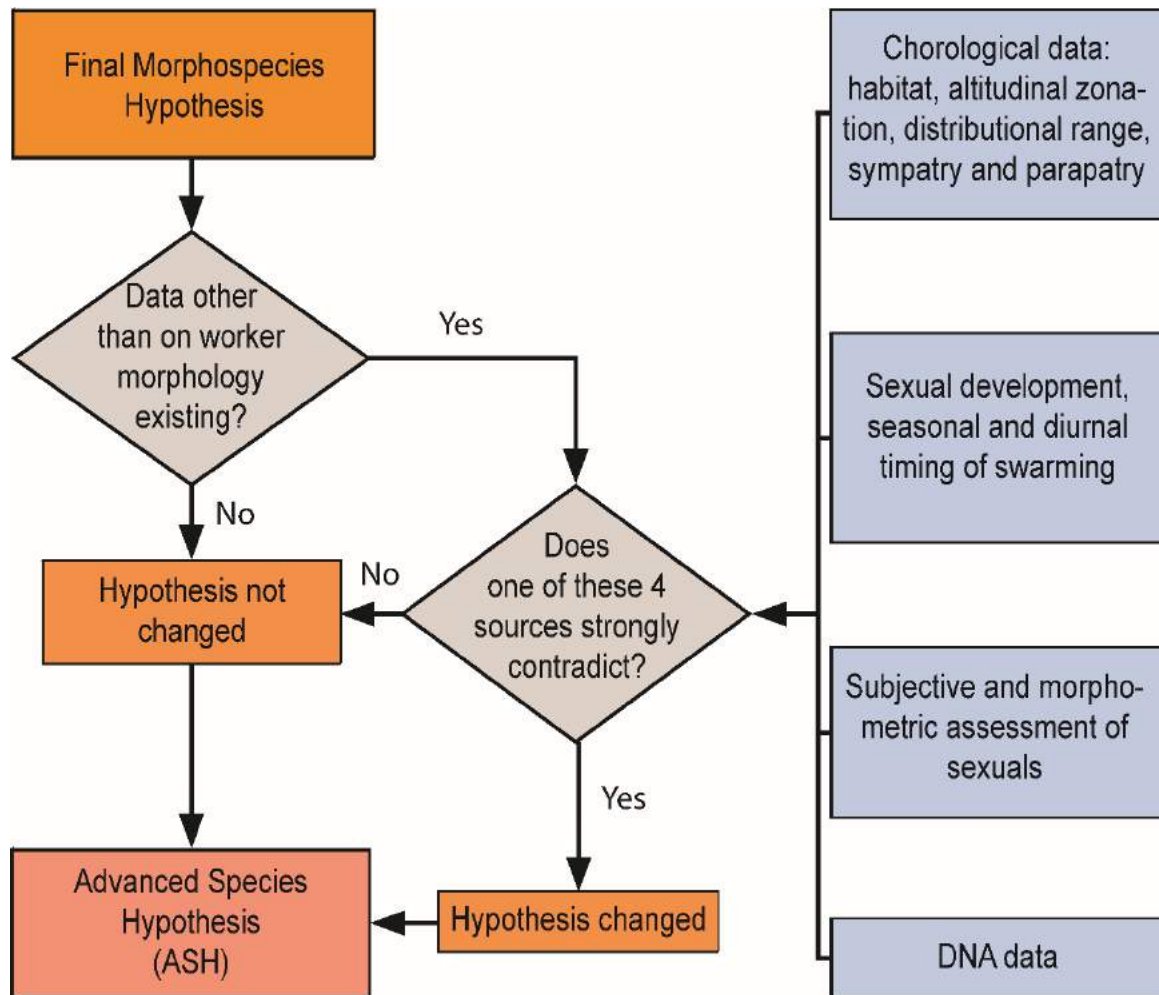
5.2.4 A kvantitatív morfológiai fajhipotézisek generálásának protokollja

5.2.4.1 A »végleges morfológiai fajhipotézis« megalkotása

Az NC klaszterezésről szóló módszertani munkában szereplő kriptikus hangyafajok gyengén differenciált dolgozói alapján felállított faji hipotézisalkotás fő forrásául a morfometrikus karakterek szolgáltak, de emellett az általános fenotípusos megjelenés szubjektív értékelése is szerepet játszott. A hipotézisalkotás lépéseinek egységesítése miatt megalkottuk a *végleges morfológiai fajhipotézist* (**6. ábra**), melynek során a szubjektív és statisztikailag támogatott elemek váltakozva egymást erősítik. Eredményül egy olyan fajhipotézist kapunk, ahol szubjektíven megalkotott morfológiai fajhipotézisek tesztelési szakaszokkal – a morfometrikus karakterek alkotta adatsorból diszkriminancia-analízissel – módosulhatnak.

Azok a minták, melyek nagyfokú bizonytalanságot hordoznak magukban a statisztikai elemzés esetleges torzításait elkerülendő elsődleges faji feltételezés nélkül (angol terminusban „wild-card”-ként) kerülnek az LDA megerősítő elemzéseibe. Ezen esetek végső osztályozásának prediktálása az LDA feladata marad. Ez az eljárás segít elkerülni a hibásan definiált csoportokat, vagyis a megerősítő elemzések eredményeinek eltorzításával kapott műtermékeket. Az LDA elemzések során elemzett esetek száma alapvető fontosságú annak eredményeit és megbízhatóságát tekintve. Nagyobb mintaelemszám jobban biztosítja a vezetve tanulós módszerek osztályozási pontosságát. Az irodalomban javaslatként ideálisnak tartott esetszám többszörösen kell, meghaladja a vizsgált példányok számát csoportonként (Lachenbruch & Mickey 1968, Lesaffre és mtsai. 1989) LDA esetében.

A hangya taxonómiában több szerző által követett gyakorlat a fészeksorozatok átlagát venni egy esetnek. A módszer bevezetéséhez tartozó teszteléseink során ettől a gyakorlattól eltértünk és egyedeket tekintettünk különálló eseteknek (ami széleskörű megoldás szoliter rovarfajok esetében), mivel így könnyebben biztosítani tudtuk a javasolt karakter-szám/esetszám arányt. Az így kapott megerősítő eredmény igen megbízhatónak tekinthető. Az első megerősítő LDA elemzés eredményei után, amelyek azon elsődleges hipotézis nélküli minták, melyek posterior p értékei magasabbak, mint posterior $p > 0.66$. Azon minták, melyek nem érik el ezt a küszöböt (posterior $p \leq 0.66$), elsődleges hipotézis nélküli minták maradnak a következő körben is. A második körben ez a határérték elméletileg lecsökken posterior $p > 0.66$ -ról posterior $p > 0.58$ -ra, de a gyakorlatban az összes eset osztályozásának követelménye miatt néha nem tudjuk elkerülni, hogy a határérték posterior $p > 0.5$ -re lecsökkenjen. Így a *végleges morfológiai fajhipotézis* megalkotásakor már nem marad osztályozatlan minta az adatsorunkban (**6. ábra**).



7. ábra. Folyamatábra a progresszív fajhipotézis megalkotásához. A végleges morfológiai fajhipotézisből (*Final Morphospecies Hypothesis*) kiindulva és egyéb diszciplínákból származó információkkal egyetértésben, több lépésben alakítjuk ki a végleges morfológiai fajhipotézist. További magyarázat a szövegben. (Seifert, Ritz és Csősz 2014 nyomán).

5.2.4.2 »A progresszív fajhipotézis«

Azonban nem csak morfológiai adatok játszanak szerepet fajhipotézisünk kialakításában. Az előzetesen kialakított fajhipotézisünk a *végleges morfológiai fajhipotézis* megalkotása során kialakult véleményünk akkor is megváltozhat, és ún. *progresszív fajhipotézissé* alakulhat (7. ábra), amennyiben ehhez erős, nem morfológiai jellegű bizonyíték szolgál alappul. Ilyen független adatsorok a következők lehetnek például:

- a) A hangyafészkekből gyűjtött ivaros egyedek (szűz királynők, hímek), melyek a differenciáltabb morfológiájuk miatt néhány hangya faj esetében a dolgozóknál nyilvánvalóbb faji eltéréseket képesek mutatni. Az ilyen adatok a hangyafajok kevesebb mint 10%-ánál voltak a módszertan tesztelésének során felhasználhatók.

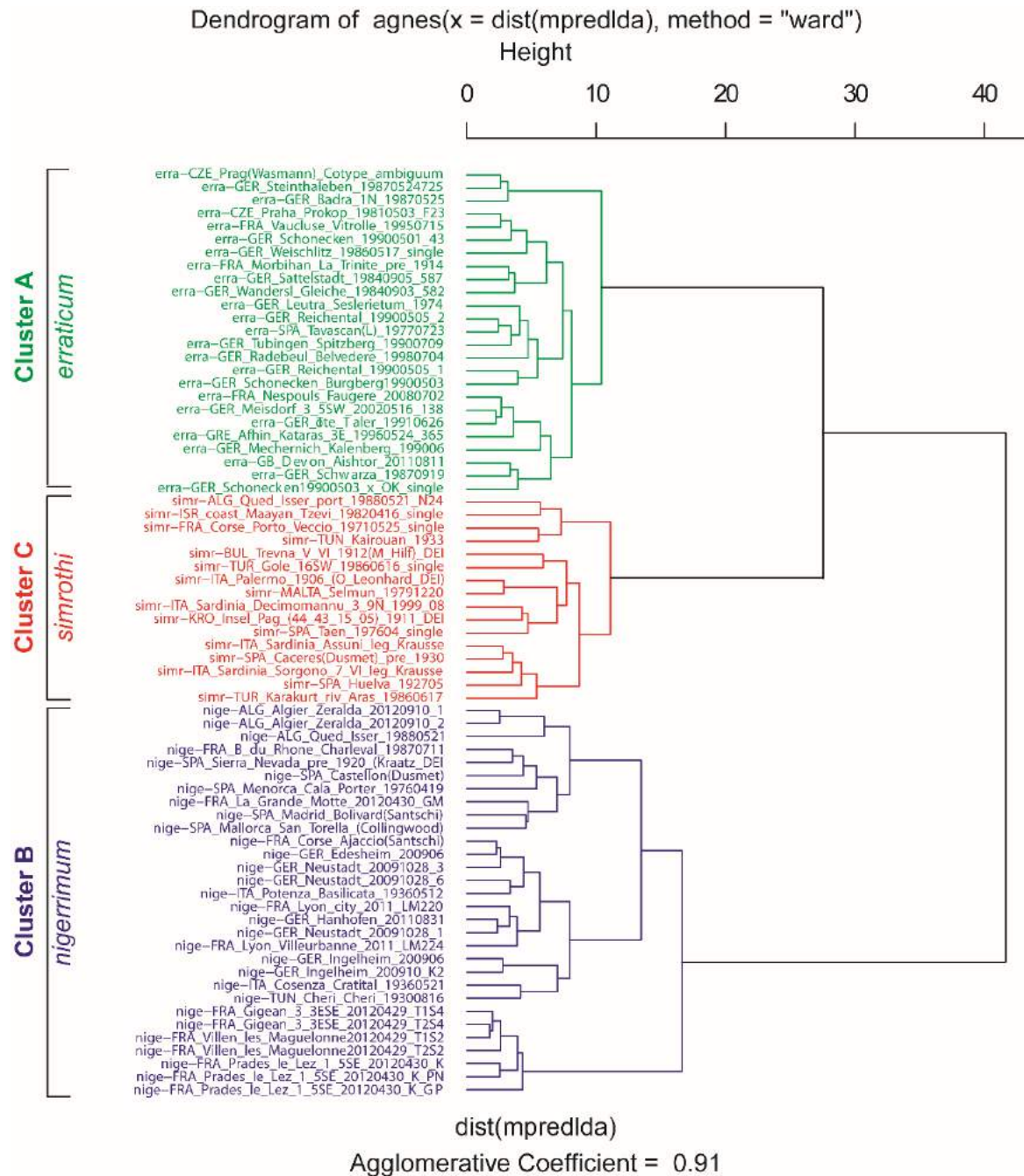
- b) Korológiai adatok (zoogeográfiai adatok, tengerszint feletti magasság, zonáció, elterjedésbeli átfedés, együttes előfordulás stb.). Ezek az adatok a fajok mindegyikénél rendelkezésre álltak.
- c) Habitat preferenciák, fészkelési szokások. Az ilyen adatok a teljes vizsgált anyag mennyiségének kevesebb mint 50%-ánál álltak rendelkezésre.
- d) Szezonális, napi aktivitásbeli adatok, nászrepülési időszak adatok. Az ilyen adatok a minták kevesebb mint 10%-ánál voltak jelen.
- e) DNS markerek, molekuláris filogenetikai adatok és elemzések. Ezek alatt leggyakrabban mitokondriális DNS értendő (COI, barcoding). A szekvencia adatok a fajok 20%-ánál elérhetőek voltak, a legnagyobb részt a kevésbé megbízható COI adatok adták, melyek bár alkalmazhatóak hangyáknál taxonómiai döntéshozásra (ld. Csősz, Heinze & Mikó 2015), a néha előforduló *parafilia* miatt (ld. még Seifert 2009) tesztelési célokra nem alkalmaztuk ezeket az adatsorokat.

Az a–e pontokban felsorolt adatok hiányos információ forrásainak ellenére, általában minden mintára jutott valamely információból annyi, amellyel meg lehetett erősíteni azt a feltételezést, hogy a morfológiai elkülönülés mögött egyéb biológiai információ is rejlik, azaz nagy valószínűséggel a morfológiai faj hipotézisek biológiai fajt reprezentálnak. A *progresszív fajhipotézis* magalkotása során ezeket az egyéb információkat szintén figyelembe vesszük (**7. ábra**).

5.2.5 Eredmények és diszkusszió

5.2.5.1 Az eredményül kapott progresszív fajhipotézis összehasonlítása korábban publikált taxonómiai eredményekkel

Ahogy azt a **2. táblázat** mutatja a 48 vizsgált fajpáron elvégzett tesztek során mind a négy módszer, melyek az NC klaszterezésből származó lineáris diszkriminánsok felhasználásával működött, 2.09 és 5.25 százalék közötti eltérést állapított meg (NC-UPGMA: 5.25%, NMDS: 2.58%, NC-Ward: 2.40% és K-közép-NC: 2.09%) a korábban publikált progresszív fajhipotézishez képest. Ez az eredmény szokatlanul jónak mondható egy pusztán morfológiai megközelítésen alapuló módszertől. Az átalakítatlan, azaz alapadatokkal végzett „K-közép-rawdata” (ld. **2. táblázat**) elemzéstípusnál megfigyelhető drámaian kiugró (21.50%) átlagos hiba, szemben a transzformált adatokkal végzett „K-közép-NC” 2.09%-os átlagos hibájával (ld. **2. táblázat**) a bemenő adatok különbözőségén alapul, mivel a K-közép módszer beállításain nem változtattunk.



8. ábra. Ward agglomerációs módszerrel kombinált NC klaszterezés dendrogramja. Három kriptikus Tapinoma faj (Tapinoma erraticum, T. nigerrimum és T. subboreale) készített dendrogram mutatja a fajok morfológiai elkülönülését. Színek: T. erraticum: zöld, T. nigerrimum: kék, T. subboreale: piros. A cédulák információs sorrendje a következő: progresszív fajhipotézis – Ország_lelőhelyhez legközelebbi város_gyűjtés dátuma_minta kód (amennyiben rendelkezésre állt). A Ward módszerrel alkotott NC klaszterezés eredményei és a progresszív fajhipotézis között teljes (100%) egybeesést találunk.

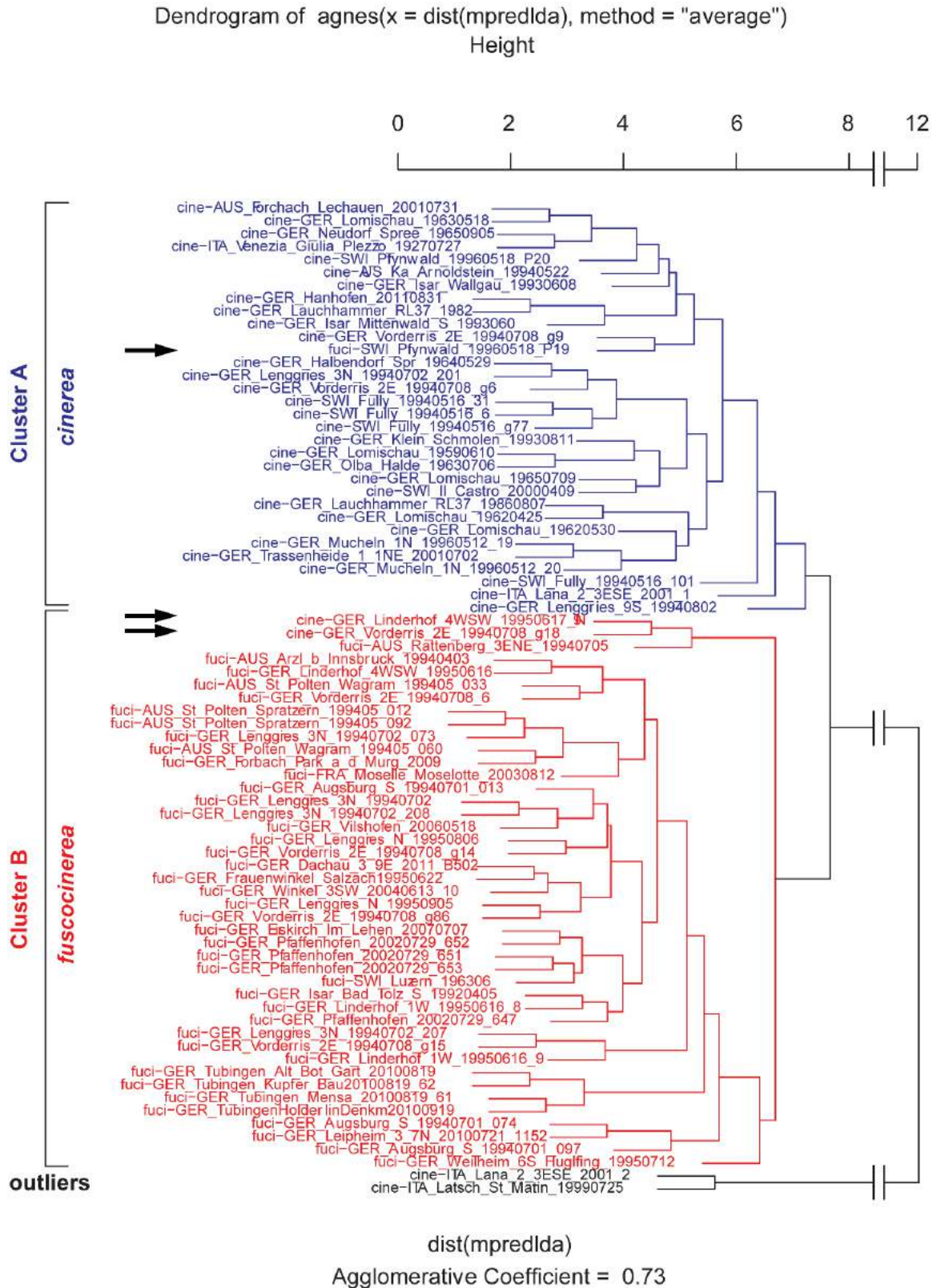
Tulajdonképpen azt látjuk, hogy az NC klaszterezés első lépése során, a fészeksorozati csoportokon végzett kumulatív LDA adat transzformációs lépése a morfológiai mintázatfelismerés megbízhatóságának szempontjából döntő fontosságú. Az NC klaszterezéssel kapott osztályozás 25 korábbi publikációban leközölt 35 fajpár (**2. táblázat**) esetén egybevág a korábban közzétett *progresszív fajhipotézissel*.

Kisebbségi eltérések (< 6%) tapasztalhatók a jelen tesztek során felállított, valamint a korábbi munkák *progresszív fajhipotézisei* között, gyakran az eredeti közlésben használt adatsor bővítése miatt, vagy, mert a biológiai háttérinformációk (ld. az 5.2.4.2. alfejezet a–e pontjait) bővülése módosította a korábbi fajhipotézist. Összesen hét olyan fajpár létezik a vizsgálatban, ahol akkoriban még publikálatlan adatsoron teszteltük a módszer eredményességét. Ezekből 5 esetben a módszer alkalmazása faji szintre emelést javasolt, és 2 további esetben új taxon (*Lasius paralienus* sp. 2 és *Temnothorax crassispinus* sp. 2) leírását tette szükségessé¹². Fontosnak tartom hangsúlyozni, hogy az itt bemutatott munka elsődleges célja egy módszer tesztelése volt, és így semmilyen taxonómiai, vagy nevezéktani célkitűzést nem fogalmazott meg.

5.2.5.2 A klaszterező algoritmusok eredményeinek összehasonlítása

Ahogy az 5.2.3.1. „A tesztelésre kiválasztott csoportok és a kiválasztás kritériumai” alfejezet a–e pontjainál felsoroltak is jelzik, a tesztelési eljárásra kiválasztásnál fontos szempont volt, hogy a teszt adatsorban rejlő morfológiai mintázat felismerhetősége a lehető legnehezebb legyen (ld. *kriptikus fajok*, a kiválasztott fajpárok morfológiai karaktereinek és méreteinek átfedése). Egy szélsőséges példaként a *Tetramorium alpestre* / *T. impurum* fajpár – amely egyike volt azoknak a fajpároknak, ahol a fajok elkülönítését molekuláris filogenetika is támogatta (Steiner et al 2010) – legmegbízhatóbb karaktere a példányok 2/3-át (68%) a fajok 95%-os konfidencia intervallumánkívül, az intraspecifikus térben helyezte el. A második legjobb karakter még nagyobb (78%) átfedést mutat. Az ilyen adatsorokon végzett tesztelések nagyon jól megmutatják azt, hogy egy módszer alkalmas-e nehezebb feladatok elvégzésére. A fent vázolt nehézségek ellenére teljes (100%) egybeesést találunk például az NC-Ward dendrogramok eredményei és a több, korábban megállapított *progresszív fajhipotézis* között három kriptikus *Tapinoma* faj, a *T. erraticum*, *T. nigerrimum* és *T. simrothi* esetében (**8. ábra**). Ugyanilyen tökéletes egyezés található az adatsorok nagy részében, ahol egy vagy több módszer is teljes egyezést adott.

¹² Megj.: a módszertani cikk megjelenése (2014) óta mind a két tudományra új faj leírásra került: a *Lasius paralienus* sp. 2 = *Lasius cazevitzi* Seifert & Galkowski, 2016 (Zootaxa, 4132 (1): 044–058.) és *Temnothorax crassispinus* sp. 2 = *Temnothorax crasecundus* Seifert & Csósz, 2015 (ZooKeys 479: 37–64 (29 Jan 2015) <https://doi.org/10.3897/zookeys.479.8510>).



9. ábra. UPGMA agglomerációs módszerrel kombinált NC klaszterezés dendrogramja. Két kriptikus *Formica* faj (*F. cinerea* és *F. fuscocinerea*) készített dendrogram mutatja a fajok morfológiai elkülönülését. Színekódok: *F. cinerea*: kék, *F. fuscocinerea*: piros. A prog-

resszív fajhipotézistől eltérő klasszifikációt a képeken nyilakkal jelöltem. A cédulák információs sorrendje a következő: progresszív fajhipotézis – Ország_ leőhelyhez legközelebbi város_ gyűjtés dátuma_ minta kód (amennyiben rendelkezésre állt). Az UPGMA-val alkotott NC klaszterezés eredményei 6.4%-al térnek el a progresszív fajhipotézistől.

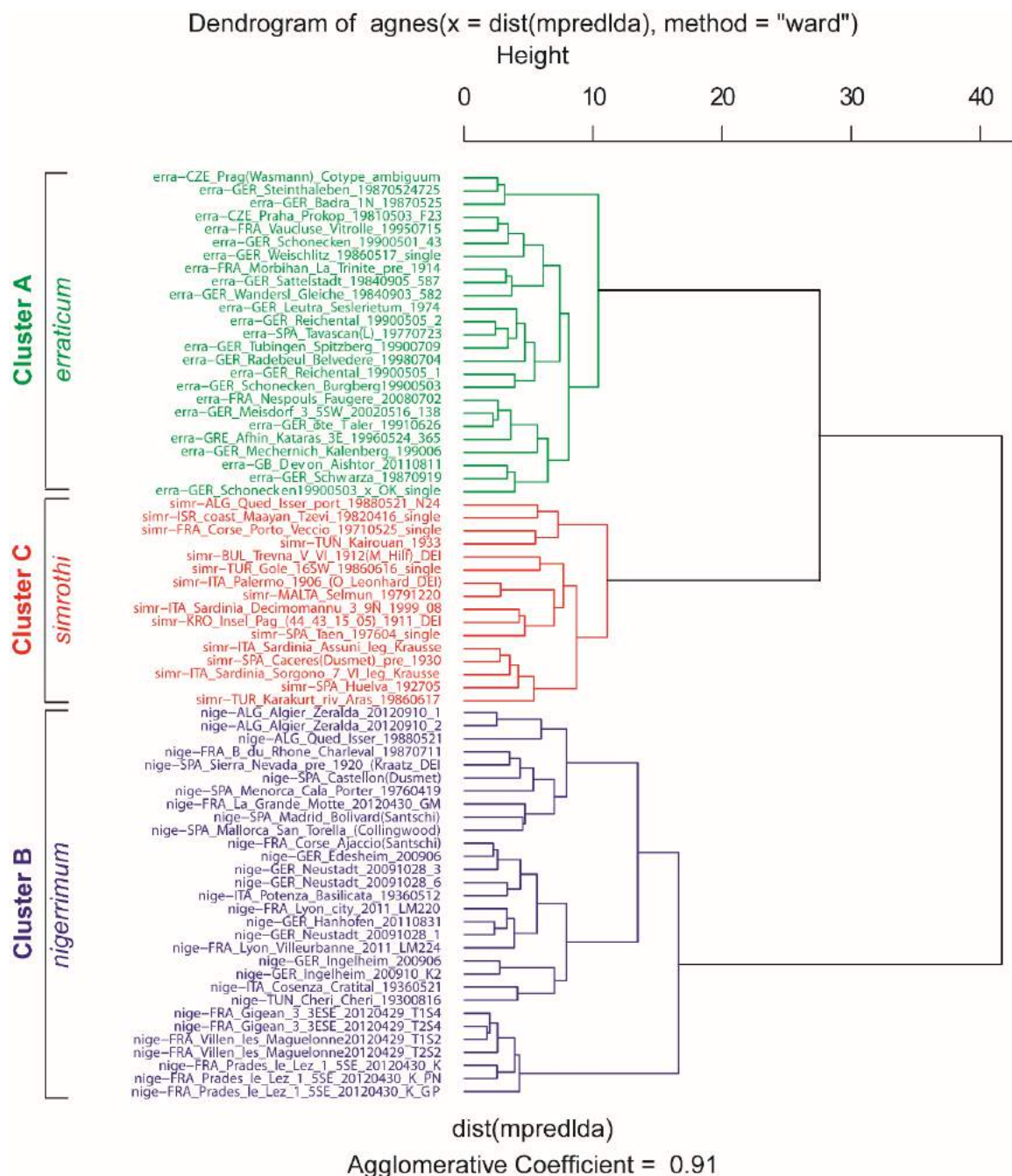
A módszerek sikeressége módszerenkénti lebontásban így szerepel: NC-Ward's módszer az adatsorok 46%, a NC-UPGMA 21%, az NMDS 31%, és a K-közép-NC 35%. Az módszerek osztályozási eredményeinek mind a 48 adatsorra lebontott sikeressége részletesen a **2. táblázatban** látható.

Bár valóban sok adatsornál találunk teljes egyezést az NC-klasztterezés és a *progresszív fajhipotézis* között, valójában – ha az elemzések átlagát nézzük – egy kisebb (2-5%) eltérésre számítani lehet a két fajhipotézis között. Ez nem meglepő, hiszen a természetben gyűjtött adatok elemzésénél 5% alatti hiba általános vélekedés szerint is elfogadottnak számít. A **9. ábra** és a **10. ábra** két, közép-európai kriptikus *Formica* faj, a *F. cinerea* és *F. fuscocinerea* NC klaszterezésének eredményét mutatja. Az ábrákon nyilak mutatják azokat a mintákat, melyek klaszterezése a *progresszív fajhipotézistől* eltérő eredményt hozott: A Ward módszerrel egyetlen rosszul klasszifikált egyedet találunk, az eltérések aránya 1.3%, UPGMA esetében a három hibás osztályozás 3.8%-os hibát jelent. Az UPGMA agglomerációs módszernél fontos megemlítenünk az ún. *outlierek*, vagy kiugró értékek előfordulását.

Az *outlierek* fogalmának nincs egységesen elfogadott definíciója a szakirodalomban. A gyakorlatban a nem magyarázható kiugró értékeket nevezzük az angol nyelvű szakirodalomban elterjedt kifejezés szerint *outlierek*nek. Hogy mit tekintünk kiugró értéknek, illetve kevésbé hihető vagy a modellünkbe nem jól illeszkedő adatnak, az nehezen fejezhető ki egzakt módon (Csereháti 2004). Mégis, ezek az esetek nagyon fontos információt hordoznak, illetve azonosításuk esszenciális a további elemzés helyénvalóságának szempontjából. Adatokban előfordulhatnak olyan értékek, melyek nem tűnnek hihetőnek, mintha „ki-lógnának” az adatok „tömegéből”. Amennyiben ez a gyanúnk alaposnak bizonyul, azt a mintát ellenőrizni kell, akár újramérés is szükséges lehet. A szakirodalomban vannak utalások e minták eltávolításának szükségességére (Csereháti 2004), vagy legalábbis más módon történő kezelésére, melyekre szélsőséges esetekben valóban szükség lehet, ha nem akarjuk, hogy a későbbi elemzések eredményeit eltorzítsák. Az *outlierek* problémája egyike a statisztika legnehezebben kezelhető kérdéseinek, elvileg megoldhatatlan, és nem is létezik olyan módszer, mellyel általánosan alkalmazható megoldást nyújtana minden esetre. Az *outlierek* problémájának kezelésére viszont valamilyen gyakorlati megoldásra

mindenképpen szükségünk van. Az általam alkalmazott eljárás modell alapú, azaz bizonyos eloszlást vesz figyelembe, és az *outlierek* így csoportosíthatóak valamely ismert klaszter egyikébe, feltéve, hogy az ellenőrzés során nem találunk mérési hibára utaló jelet, mert ez esetben az újramérés jelenti a megoldást. Természetesen nem állíthatjuk biztosan, hogy a detektált *outlier* nem egy külön álló – általunk még nem ismert – klaszter egyik eleme lenne, s pusztán a kis mintaelemszám miatt tűnik *outlier*nek. Valamilyen kockázatot azonban muszáj felvállalnunk a kérdés eldöntésére. Az UPGMA – bár a *progresszív fajhipotézistől* némileg nagyobb eltérést produkál, mint a Ward módszer – nagy előnye, hogy az *outliereket* felismerhetővé teszi, így alkalmunk nyílik ezek felismerésére és alaposabb vizsgálatára. A 9. ábraán látható UPGMA dendrogram 2 *outliert* mutat, melyeket az általam leírt módon kezeltem, majd besoroltam a megfelelő fajba.

Ahogy az 2. táblázatban látható a 48 vizsgált fajpáron elvégzett tesztek során mind a négy módszer, melyek az NC klaszterezésből származó lineáris diszkriminánsok felhasználásával működött, a korábban publikált *progresszív fajhipotézishez* képest 2.09 és 5.25 százalék közötti eltérést állapított meg (NC-UPGMA: 5.25%, NMDS: 2.58%, NC-Ward: 2.40% és K-közép-NC: 2.09%). A négy módszer (plusz az alapadatokkal K-közép módszer) eredményeinek összehasonlítását Generalized Linear Modellel (GLM) végeztük.



10. ábra. Ward agglomerációs módszerrel kombinált NC klaszterezés dendrogramja. Két kriptikus *Formica* faj (*F. cinerea* és *F. fuscocinerea*) készített dendrogram mutatja a fajok morfológiai elkülönülését. Színek: *F. cinerea*: kék, *F. fuscocinerea*: piros. A progresszív fajhipotézistől eltérő klasszifikációt a képeken nyílal jelöltem. A cédulák információs sorrendje a következő: progresszív fajhipotézis – Ország_lelőhelyhez legközelebbi város_gyűjtés dátuma_minta kód (amennyiben rendelkezésre állt). A Ward módszerrel alkotott NC klaszterezés eredményei 1.3%-kal térnek el a progresszív fajhipotézistől.

A vizsgálataink során minden módszer közül a Ward módszer adta a legjobb osztályozást. A Ward módszer általi osztályozott esetek szignifikánsan magasabb arányban álltak közelebb *progresszív fajhipotézis*hez, mint az UPGMA és az alapadatokkal futtatott K-közép módszer (mindkét esetben $p < 0.0001$) mintázatai, ugyanakkor nem találunk szignifikáns különbséget a Ward módszer és az NMDS, valamint a lineáris diszkrimináns adatok használatával futtatott K-közép módszerek eredményei között.

Az alapadatokkal futtatott K-közép módszer a lineáris diszkrimináns adatok használatával (vagyis az NC klaszterezéssel transzformált adatokkal) futtatott K-közép módszernél drámaian rosszabb osztályozási eredményei külön figyelmet érdemelnek. A klaszterező algoritmus beállításai ugyanis változatlanok maradtak, pusztán a bemenő adatok változtak. Az eredmények szerint az NC-klaszterezés osztályozási ereje és nagyfokú sikeressége sokkal inkább az NC klaszterezés adat transzformálásának köszönhető, mintsem a klaszterező algoritmusok kiválasztásának.

A Ward módszer által használt négyzetes euklideszi távolság úgy tűnik, hogy az euklideszi távolságot használó UPGMA-nál némileg pontosabb klaszterezést tesz lehetővé. Az a tény, hogy a Ward módszernél tapasztaljuk a *progresszív fajhipotézistől* való legkisebb átlagos eltérést (amely mindössze 2.09% a teljes adatsorra vetítve) azt mutatja, hogy ez egy nagyon biztonságosan használható klaszterezési eljárás. Azonban nem jelenthetjük ki egyértelműen, hogy ez lenne a legjobb módszer, hiszen a négyzetes euklideszi távolság alkalmazása a különbségeket néha túlbecsüli, valamint a módszer működéséből kifolyólag különálló alkasztereket, valamint a fentebb tárgyalt *outliereket* (ld. 5.2.5.2 alfejezet) az UPGMA-nál jóval nagyobb valószínűséggel von össze egyetlen klaszteré, így kétséges biológiai jelentőséggel bíró csoportok alkotásának is nagyobb a veszélye. Ennek az a legfőbb oka, hogy a Ward módszer a „négyzetes hibaminimalizálási” kritériumot használja annak eldöntésére, hogy mely klasztereket vonja össze. Az egyes szintek klaszterei nem lesznek a teljes stabilak abban az értelemben, hogy egy pont közelebb lehet egy másik klaszter középpontjához, mint a sajátjához.

Az UPGMA-nál hasonló probléma nem lép fel, egyrészt mert együtthatói függenek mind-egyik összevont klaszter méretétől, másrészt az objektumok mindig a saját klaszterük középpontjához esnek legközelebb, mivel minden egyes klaszteralkotásnál (lépésnél) a két legközelebbi klaszter olvad össze egy klaszterré. Ezért úgy tartják, hogy az UPGMA kiegyensúlyozott klasztereket képes alkotni (Odong és mtsai. 2011) annak árán, hogy egyes kérdéses eseteket *outlierekként* határoz meg. A mi vizsgálataink során megerősítést nyert ez a tulajdonság. Általánosságban elmondható, hogy az UPGMA dendrogramjait nehezebb kész hipotézissé fordítani, mint a Ward módszer klasztereit, de az *outlierek* kieme-

lése és a stabilabb klaszteralkotási képesség ellensúlyozza ezt a hátrányt. Ennek megfelelően, elkerülve az egyes módszerek hiányosságaiból adódó hibás hipotézisalkotást, két vagy több módszer együttes alkalmazása javasolható egy taxonómiai munkában.

Az NMDS és K-közép módszerek a Ward módszerhez hasonlóan magas *progresszív fajhipotézissel* való egyezést adtak. Ezen két, utóbbi módszer legnagyobb hiányossága, hogy meg kell adni a várt klaszterek számát, amelyek szerint az algoritmusok partícionálják az eseteket. A másik hátulütőjük a partícionáló módszereknek, hogy nem ábrázolják a struktúrát, és ilyen módon nem alkothatunk képet a klasztereken belüli mintázatról. A dendrogramok ezzel szemben azonnali betekintést engednek a klaszterek felépítésébe, melyek nagy segítséget jelentenek a biológiai relevanciával is bíró mintázat felismerésében. Egy további, gyakorlati nehézség, amellyel az NMDS és K-közép módszereknél számolnunk kell, az a hosszú számolási idő és a megnövekedett számítástechnikai kapacitás igénye. Ezek a módszerek ugyanis sokszoros ismétlésszámmal működnek (*iterative* módszerek), ugyan megadhatunk kisebb ismétlésszámot, a jó eredmény érdekében azt érdemes magasan tartani. A módszertani munka tesztelési folyamataihoz használt számítógépek kapacitása (2013-ban 2MB RAM és 3GHz CPU) mellett három perc volt szükséges egy 156 mintát tartalmazó adatsor elemzéséhez, viszont 230 mintás adatsor elemzése már 2 teljes órát igényelt a script¹³ alapbeállításai mellett. Hasonló problémával az NC-UPGMA és NC-Ward esetében nem találkozunk: egy 500 mintás adatsor számításához szükséges idő mindössze néhány másodperc.

Amikor a tesztelt módszerek a *progresszív fajhipotézissel* való szokatlanul nagymértékű egyezésről beszélünk (**2. táblázat**), akkor szót kell ejtenünk a másik oldalról is, azaz meg kell említenünk a módszertant, amely alapján a *progresszív fajhipotézist* felállítottuk. A döntő adattípus, amely minden esetben rendelkezésre állt a morfológia volt, amelyből kinyert mintázatokat, azaz elsődleges fajhipotézist, valamilyen egyéb forrásból származó adatokkal, például ivaros alakok morfológiája, elterjedési adatok, fenológia, ökológiai vagy molekuláris adatok, megerősítettük. Annak ellenére, hogy majdnem minden fajpárnál rendelkezésre állt a morfológia mellett legalább egy járulékos adatsor, amellyel a fajhipotézis tesztelését el tudtuk végezni, az ilyen irányú adatgyűjtés összetettségének hiánya miatt (mint amilyen az integratív taxonómiai vizsgálatok sajátja) az általunk használt *progresszív fajhipotézist* érdemes alapvetően morfológiai fajhipotézisként kezelni.

¹³ Az R-ben írt script a következő oldalról szabadon letölthető: <http://sourceforge.net/projects/agnesclustering/>.

Példák	Nk	Nm	UPGMA	Ward	NMDS	K-means-NC	K-Means	Hivatkozások
<i>Bothriomymex communistus / corsicus</i>	16	111	0	0	0	0	1,8	[1]
<i>Camponotus atricolor / piceus</i>	11	94	4,25	0	3,19	2,12	14,89	[BS], [2]
<i>Camponotus herculeanus / ligniperda</i>	6	49	0	0	0	0	22,45	[3]
<i>Cardiocondyla bulgarica/ sahlbergi</i>	17	65	0	0	3,08	3,08	20	[BS], [4]
<i>Cardiocondyla dalmatica / elegans</i>	13	78	10,26	2,56	2,56	2,56	17,95	[BS]
<i>Cardiocondyla mauritanica / kagutsuchi</i>	14	140	0,71	0,71	0,71	0,71	27,14	[BS], [4]
<i>Crematogaster schmidtii / scutellaris</i>	10	69	0	0	1,45	1,45	5,8	[BS], [2]
<i>Formica cinerea / fuscocinerea</i>	13	78	6,41	1,28	2,56	2,56	11,54	[BS], [5]
<i>Formica clara / cunicularia</i>	18	121	4,96	4,13	4,13	2,48	8,26	[6]
<i>Formica clarissima / litoralis</i>	17	143	5,59	0	2,8	1,4	35,66	[7]
<i>Formica exsecta / fennica</i>	14	126	0,79	0	3,17	3,17	50	[BS], [8]
<i>Formica foreli / pressilabris</i>	7	229	1,74	1,74	0	0	0,87	[BS], [8]
<i>Formica litoralis / pamirica</i>	17	120	3,33	3,33	3,33	3,33	44,17	[7]
<i>Formica lugubris / pratensis (Panpalaeartctic)</i>	5	316	1,27	0,63	0,32	0,32	0,32	[9], [BS]
<i>Hypoponera punctatissima / schauinslandi</i>	10	54	1,85	1,85	0	0	37,04	[10], [BS]
<i>Lasius barbarus / lasioides</i>	13	100	25	6	6	6	17	[BS]
<i>Lasius barbarus / lasioides</i>	→8	100	26	9	7	3	28	[BS]
<i>Lasius emarginatus / illyricus</i>	15	85	4,71	3,53	0	1,18	7,06	[BS]
<i>Lasius gebaueri / psammophilus</i>	16	77	0	0	1,3	1,3	42,86	[BS], [11]
<i>Lasius japonicus / platythorax</i>	14	70	2,86	2,86	0	0	18,57	[BS], [11]
<i>Lasius mixtus / sabularum</i>	14	66	0	0	0	0	33,33	[BS], [12]
<i>Lasius niger / platythorax</i>	14	114	0	0	0	0	9,65	[BS], [13]
<i>Lasius paralienus / paralienus 2</i>	16	62	1,61	4,84	3,23	4,84	3,23	[BS]
<i>Lasius paralienus / psammophilus</i>	16	123	1,63	0	3,25	2,44	32,52	[BS], [11]
<i>Lasius paralienus 2 / psammophilus</i>	16	91	6,59	1,1	1,1	7,69	23,08	[BS]
<i>Lasius piliferus / psammophilus</i>	16	67	7,46	0	0	0	43,28	[BS], [11]
<i>Lasius sabularum / umbratus</i>	14	150	1,33	0	0	0	26	[BS], [12]
<i>Myrmica constricta / hellenica</i>	16	91	7,69	3,3	1,1	1,1	9,89	[14]
<i>Myrmica lobicornis / lobulicornis</i>	16	97	0	7,22	1,03	1,03	7,22	[15], [BS]
<i>Myrmica lobicornis / lobulicornis</i>	→10	97	2,06	0	2,06	2,06	7,22	[15], [BS]
<i>Myrmica salina / specioides</i>	16	161	9,32	9,94	2,48	3,73	30,13	[16], [BS]
<i>Myrmica salina / specioides</i>	→6	161	3,11	9,32	3,11	3,11	5,59	[16], [BS]
<i>Ponera coarctata / testacea</i>	8	141	1,42	0,71	0	0	10,64	[BS], [17]
<i>Tapinoma erraticum / nigerrimum</i>	14	55	0	0	0	0	3,64	[18]
<i>Tapinoma erraticum / simrothi</i>	14	41	2,44	0	0	0	2,44	[BS], [19]
<i>Tapinoma erraticum / subboreale</i>	14	52	0	0	0	0	3,84	[18]
<i>Tapinoma nigerrimum / simrothi</i>	14	48	0	0	0	0	18,75	[BS], [19]
<i>Temnothorax crassispinus / nylanderi sp.2</i>	18	100	5	2	12	3	50	[BS]
<i>Temnothorax lichtensteini / parvulus</i>	18	113	6,19	0,88	7,96	6,19	39,82	[BS]
<i>Temnothorax lichtensteini / parvulus</i>	→8	113	1,77	1,77	1,77	1,77	4,42	[BS]
<i>Temnothorax luteus / racovitzai</i>	17	64	14,06	3,12	3,12	3,12	17,18	[BS]
<i>Temnothorax nigriceps / tuberum</i>	18	89	3,37	0	1,12	1,12	33,71	[BS]
<i>Temnothorax saxonicus / sordidulus</i>	18	96	17,71	17,71	2,08	1,04	43,75	[20]
<i>Temnothorax saxonicus / sordidulus</i>	→10	96	7,29	4,17	0	0	14,58	[20]
<i>Tetramorium alpestre / impurum</i>	26	103	33,98	8,74	33,98	20,39	48,54	[21]
<i>Tetramorium alpestre / impurum</i>	→10	103	13,59	2,91	2,91	2,91	38,83	[21]
<i>Tetramorium cheketi / moravicum</i>	17	62	3,23	0	0	0	29,03	[SCS], [22]
<i>Tetramorium diomedaeum / ferox</i>	21	63	1,59	0	0	0	30,16	[23]
A 48 adatsorból számolt átlag			5,25	2,4	2,58	2,09	21,5	

2. táblázat. Magyarázat a következő oldalon.

2. táblázat. Különböző módszerekkel (NC-klaszterezés+UPGMA, NC-klaszterezés+Ward módszer, NMDS, és K-közép módszer) elért osztályozás az ún. progresszív fajhipotézistől való százalékos eltérései. A progresszív fajhipotézis kialakításának folyamatát ld. a szövegben (5.2.4.2. alfejezet). Magyarázat: „K-közép-NC” = a K-közép módszer bemeneti adata az NC-klaszterezés során transzformált adatsor volt, a „K-közép” = bemeneti adata megegyezett fészeksorozatok alapadatával. „Nm” = a minták száma. „Nk” = a karakterek száma az elemzésben (egy nyíl jelzi, ahol az elemzést megismételtük csökkentett karakter számmal). A „Hivatkozások” oszlopban található rövidítések magyarázata: [BS] és [SCS] a szerzők, Bernhard Seifert [BS] és Csősz Sándor [SCS] publikálatlan adatai. A további referenciák [1 - 23] megtalálhatóak a referencialistában és a következő cikkekre utalnak: [1] Seifert (2012b), [2] Seifert (2007), [3] Seifert (2008), [4] Seifert (2003a), [5] Seifert (2003b), [6] Seifert & Schultz (2009a), [7] Seifert & Schultz (2009b), [8] Seifert (2000), [9] Seifert & Goropashnaya (2004), [10] Seifert (2004), [11] Seifert (1992), [12] Seifert (1988), [13] Seifert (1991), [14] Seifert és mtsai. (2009), [15] Seifert (2005), [16] Seifert (2011), [17] Csősz & Seifert (2003), [18] Seifert (2012a), [19] Seifert (1984), [20] Seifert (2006a), [21] Steiner és mtsai. (2010), [22] Csősz és mtsai. (2007), [23] Csősz & Schulz (2010). A táblázat Seifert, Ritz, Csősz (2014) *Myrmecological News*, 19: 1–15. nyomán készült.

Azonban egyes kutatások a morfológiai módszerek nagyfokú taxonómiai stabilitására utalnak. Schlick-Steiner és mtsai. (2010) 184 olyan, ízeltlábúakon végzett integratív taxonómiai munkát vizsgáltak meg, melyek faji szintű leírásokat eredményeztek, és azt találták, hogy a különféle módszerek közül (morfológia, DNS, ökológia, enzim vizsgálatok, viselkedés, citogenetika, kémia, pl. kutikuláris hidrokarbonátok) a morfológiai vizsgálatok eredményei mutatták a legnagyobb stabilitást, valamint a végső fajhipotézissel való legnagyobb egyezést. Ismerve a morfológiai vizsgálatok ízeltlábúak esetében való alkalmazhatóságát – kombinálva a saját tapasztalatomból származó ismeretekkel – elmondható, hogy a módszerek teszteléseinek során összehasonlító adatsorokként alkalmazott *progresszív fajhipotézisek* nagyon közel állnak – az elméleti – biológiai fajhipotézisekhez.

5.2.5.3 Hibrid esetek vizsgálata

Az eddig bemutatott eredmények olyan adatsorok összehasonlításával jöttek létre, melyekben a fajok közötti hibridizációs esetekkel – ismereteink szerint – nem találkozhatunk, vagy jelenlétük nem volt kimutatható. Az interspecifikus hibrid esetek a rovarok bizonyos csoportjaiban jól ismertek, ezért mindenképpen szükségesnek tartom bemutatni a tesztelések ide vonatkozó részét is. A fent leírt tesztelési eljárást – vagyis az UPGMA, valamint az NMDS és a K-közép módszerrel végzett klaszterezés által kapott eredmények összehasonlítását – elvégeztük öt olyan adatsorral is, melyben a *Formica*, *Lasius*, *Temnothorax* és *Messor* génuszokba tartozó fajták közötti hibridizálódás jelensége ismert, dokumentált

és publikált volt. Ezeket az eredményeket a **3. táblázatban** mutatom be. A hibridizációt illetve a hibridek és szülői fajok közötti elkülönítést korábbi integratív vizsgálatok igazolták: morfológiai vizsgálatokkal, (mind az öt esetben), magi DNS szakaszok elemzésével (három esetben), mitokondriális DNS szekvenciákkal (két esetben) és allozimekkel (két esetben). Eredményként az mondható el, hogy az NC-klaszterezés által felismert mintázat és az integratív taxonómia által megállapított fajhipotézis között átlagban 8.3% és 32.3%-os eltéréseket tapasztaltunk. Ennek megfelelően, a módszer nem ajánlható hibrid minták meghatározására. A magas hibaszázalék a hibrid esetek morfológiai karaktereinek magas fluktuációjával magyarázható. Az egyes egyedek nem a két szülői fenotípus egyes jellegeinek középértékeként írhatók le, hanem bizonyos karakterek véletlenszerűen egyik vagy másik szülői fenotípushoz nagyobb mértékben közelítenek. Ezzel az irodalomból egyébként ismert jelenséggel (Seifert 1984, 1999, 2006b, Kulmuni és mtsai. 2010, Seifert és mtsai. 2010, Steiner és mtsai. 2011, Bagherian és mtsai. 2012) morfológiai adatokkal dolgozó többváltozós algoritmusok nehezen tudnak megbirkózni.

Példák	Nk	Nm	UPGMA	Ward	NMDS	K-means-NC	K-Means	Hivatkozások
<i>Formica aquilonia</i> / <i>aquilonia</i> x <i>polycтена</i> / <i>poly</i>	16	145	7,59	3,45	35,86	17,24	34,48	[24]
<i>Formica polycтена</i> / <i>polycтена</i> x <i>rufa</i> / <i>rufa</i>	13	140	7,14	7,14	13,54	1,43	14,28	[25]
<i>Messor minor</i> / <i>minor</i> x <i>wasmanni</i> / <i>wasmanni</i>	17	30	6,67	6,67	6,67	6,67	20	[26]
<i>Temnothorax crassispinus</i> / <i>crassispinus</i> x <i>nyland</i>	18	143	6,99	6,99	30,07	32,16	42,65	[27]
<i>Lasius jensi</i> / <i>jensi</i> x <i>umbratus</i> / <i>umbratus</i>	14	199	18,59	18,59	26,13	24,12	50	[28]
Az 5 adatsorból számított átlag eltérés			9,39	8,57	22,45	16,32	32,28	

3. táblázat. Különböző módszerekkel (NC-klaszterezés+UPGMA, NC-klaszterezés+Ward módszer, NMDS, és K-közép módszer) elért osztályozás az ún. progresszív fajhipotézistől való százalékbeli eltérései hibrideket tartalmazó fajok esetén. Az ún. progresszív fajhipotézis leírását ld. a 5.2.4.2. Fejezetben. Magyarázat: „K-közép-NC” = a K-közép módszer bemeneti adata az NC-klaszterezés során transzformált adatsor volt, a „K-közép” = bemeneti adata megegyezett fészeksorozatok alapadatával. „Nm” = a minták száma. „Nk” = a karakterek száma az elemzésben (egy nyíl jelzi, ahol az elemzést megismételtük csökkentett karakter számmal). A „Hivatkozások” oszlopban található referenciák megtalálhatóak a referencialistában, és a következő cikkekre mutatnak: [24] Kulmuni és mtsai. (2010), [25] Seifert és mtsai. (2010), [26] Steiner és mtsai. (2011), [27] Pusch és mtsai. (2006), [28] Seifert (2006b). A táblázat Seifert, Ritz, **Csász** (2014) *Myrmecological News*, 19: 1–15. nyomán készült.

5.2.6 Konklúzió és ajánlások

Az általunk bevezetett NC-klaszterezés nagy segítséget jelent bármilyen euszociális szervezet – azaz hártácsszárnyúak (hangyák, méhek, darazsak), termeszek kolóniái, gubacs-képző tetvek, tripszek, ormányos bogarak és vakond patkányok – morfológiai alapú taxonómiai hipotézisalkotásban. Alkalmas a fent említett adatsorokban rejlő nem várt morfológiai klaszterek felismerésére. A meglepően jó eredményeket az NC-klaszterezés legalapvetőbb lépésének, a fészeksorozatok szerint csoportosított LDA-nak tulajdoníthatjuk, amelynek elsődleges funkciója esetünkben nem a fészeksorozatok elkülönítése, hanem a közöttük levő különbségek számszerűsítése. Az LDA által megadott lineáris diszkriminánsokból Euklideszi távolsággal meghatározzuk a minták közötti távolságot, majd klaszterezve ábrázoljuk. Ez a munkafolyamat – a tesztjeinkből úgy látszik – előfeltétele a jó klaszterezésnek, mivel a K-közép módszert mind diszkriminánsokkal, mind pedig alapadatokkal teszteltük és drámaian nagy különbséget mértünk az diszkriminánsokkal végzett elemzés (2.09%) és az alapadatokkal történt elemzés (21.5%) eredményessége között. Még a hibrid populációkkal terhelt adatsornál is meglátszik ez a különbség (16.32% és 32.28%), bár az ilyen forgatókönyvekre az NC-klaszteresést nem találtuk alkalmazhatónak.

Az új eljárás egyik nagy előnye, hogy alkalmazásával az eddig bonyolult, hosszadalmas és sok buktatóval terhelt morfológiai alapú taxonómiai döntéshozási folyamat nagymértékben felgyorsítható. A dendrogramból képet kaphatunk intraspecifikus morfológiai mintázatokról is, amik regionális populációs különbséget jeleznek, így elmondható, hogy a kifejezett taxonómiai alkalmazásokon túl a módszer – óvatos következtetések levonása mellett – alkalmazható intraspecifikus állatföldrajzi mintázatok vizsgálatára is.

5.2.6.1 Módszerek kombinált használata

Amint az 5.2.5. alfejezetben részletesen bemutatásra került, az egyes NC-klaszterezés által transzformált adatokkal működő módszerek által adott eredmények minimálisan bár, de eltérnek a *progresszív fajhipotézistől* (**2. táblázat**). Ezért javasolt több módszer szimultán alkalmazása ugyanazon adatsor elemzésére, ezzel is tovább növelve az osztályozás sikerességét. Kombinációként javasolható az NC-UPGMA, az NC-Ward, K-means, és az NMDS közül legalább kettő használata, illetve, amennyiben ezt szükségesnek találjuk, akár az összes itt felsorolt módszert is használhatjuk egymás mellett. Így a több, független elemzés típus metszéspontjában adott döntések alkotják a *végleges morfológiai fajhipotézist* (ld. **6. ábra**). Az egyes mintákra vonatkozó fajhipotézist csak akkor fogadjuk el, ha mind a három feltáró típusú elemzés ugyanarra a következtetésre jutott. Abban az eset-

ben, ha bármelyik módszer alternatív döntést hoz, a minta wild-card-ként került a megerősítő elemzés (LDA) adatmátrixába. Ezzel az elgondolással az elérhető legmegbízhatóbb morfológiai fajhipotézist vagyunk képesek megalkotni.

Nagyobb adatsoroknál azonban, melyekben több száz mintát kell elemezzünk (pl. **Csősz és mtsai.** 2015 adatsorában 525 mintát elemeztem), az iteratív beállításokkal partícionáló NMDS és K-közép módszer nagy számítási kapacitása miatt ezek helyett az NC-UPGMA és NC-Ward módszerek kombinációját tudom ajánlani, hogy bár több lépésben, de egyszerűbben lehet döntést hozni.

5.2.6.2 A minták mérete az NC-klaszterezés sikerességét befolyásoló tényező

Az NC-klaszterezés eredményeinek javítása egy ideálisan definiált, ismételhető morfometrikus karakterek kijelölése után adott vizsgálati anyag esetén két további fő paraméteren múlhat:

- a) A mért morfometrikus karakterek számán, valamint
- b) A minták méretén, azaz a fészeksorozatokba tartozó átlagos egyedszámon.

Az előbbi (a) pont lehetőségeit a vizsgált csoport alaktani jellemzői nagyban befolyásolják. Nagy mennyiségű morfometrikus karakter kellő definiálására nincs lehetőség kevésbé differenciált morfológiai jellemzőkkel rendelkező csoportok esetében. Ilyenek pl. az ún. Poneromorf fajok, melyeken sokkal kevesebb megfelelően értelmezhető struktúrát találhatunk morfometrikus vizsgálat céljára, mint egy izgalmasabb morfológiával rendelkező Myrmicinae alcsaládba tartozó hangyán, például *Nesomyrmex* fajokon (**11. ábra**).

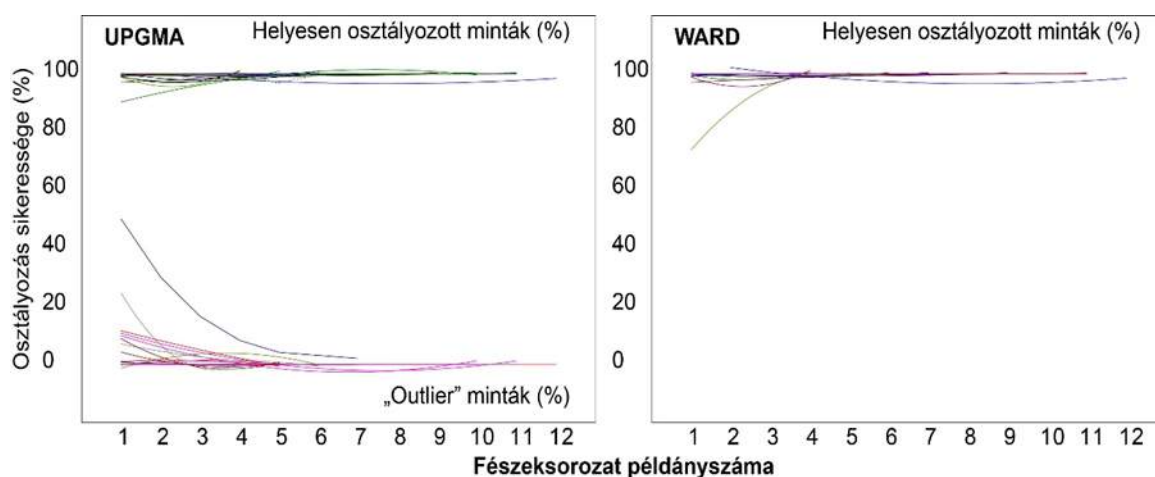


11. ábra. Két, eltérő mértékben karakter-gazdag külső morfológiával rendelkező hangyafaj összehasonlítása. A *Ponera* fajok (balról) sokkal egyszerűbb felépítése nem kedvez a gazdag morfometrikus adatsor felvételének. A sokkal izgalmasabb struktúrát viselő *Nesomyrmex spinosus* (jobbról) nagy és könnyen mérhető összetett szemei, kellően definiálható dorzális propodeális és petioláris tüskéi, a petiolustól elkülönülő postpetiolus megjelenése jelentősen megnövelik a mérhető karakterek számát.

Elmondható tehát, hogy a mért morfometrikus karakterek számát a vizsgált csoport jellemzői alapvetően meghatározzák. Emellett a felvett morfometrikus adatok számának igazodnia kell a rendelkezésünkre álló egyedszámhoz, a morfometrikus karaktereket nem szabad a vizsgált egyedszámtól függetlenül emelni.

Az irodalomban az LDA elemzések számára megfogalmazott javaslatok a karakterszámnál nagyobb mintaelemszámot tartanak megfelelőnek (Lachenbruch & Mickey 1968, Lesaffre és mtsai. 1989). Nagyobb mintaelemszám jobban biztosítja az osztályozás pontosságát. Amennyiben nem áll rendelkezésünkre megfelelő mennyiségű vizsgálható példány, a karakterszám növelése önmagában nem segít a taxonómiai problémák megoldásában.

A másik lehetőség a mintázatfelismerés pontosságának javítására a fészeksorozatok méretében, azaz a fészeksorozatok átlagos egyedszámában rejlik (*b* pont). Amennyiben lehetőségünk van rá, illetve a vizsgált anyag nagysága ezt megengedi, ez a kiterjesztés az *a* pontban leírt elgondolásokkal szemben valódi megoldást hozhat.



12. ábra. A 48 vizsgált fajpár 4844 fészekmintájának NC klaszterezéssel kapott osztályozási sikerességének változása a fészekminták egyedszámának növekedésével. Mind az NC-UPGMA, mind pedig a NC-Ward módszer átlagosan 95%-nál nagyobb sikerrel osztályozza a fészekmintákat kis egyedszám esetén is. Az outlierek aránya NC-UPGMA elemzésinél szintén négyenél több egyedet tartalmazó sorozatok esetén közelít a nullához. (Csősz, nem publikált)

Az általunk vizsgált 48 fajpár 4844 fészekmintájának NC-klasztterezéssel kapott osztályozási sikeressége a fészekminták egyedszámának növekedésével. Mind az NC-UPGMA (12. ábra, bal oldalon), mind pedig a NC-Ward (12. ábra, jobb oldalon) módszer átlagosan 95%-nál nagyobb sikerrel osztályozza a fészekmintákat kis egyedszám esetén is. Alapvetően jó

eredményeket várhatunk egy vagy két példánnyal rendelkező fészeksorozatoknál is, azonban három dolgozónál nagyobb fészekminták esetében a hibás csoportosítás lehetősége minimálisra csökken. Figyelemre méltó, hogy az UPGMA esetében jelen lévő *outlierek* előfordulása szintén a négy példányt tartalmazó vagy annál nagyobb fészekmintáktól kezdve a nullához közelít. Elmondható tehát, hogy bár az NC-klaszterezés az egy vagy két példányt tartalmazó mintákat is képes elfogadható sikerrel (>95%) osztályozni, a legjobb eredmény eléréséhez három vagy afölötti átlagos példányszámú fészeksorozatokat tartalmazó adat-sort érdemes használnunk, és a fészeksorozatok példányszáma valódi megoldást ad problémás osztályozása terén.

5.2.6.3 A négy vizsgált módszer nem javasolt hibrid esetek vizsgálatára

Az UPGMA, a Ward módszer, az NMDS és a K-Means teljesítményét öt különböző, interspecifikus hibridizációban vizsgáltuk (**3. táblázat**). Az esetek többségében a négy klaszterezési módszer olyan zavaros képeket mutatott, amelyek nem tettek lehetővé ésszerű hipotéziseket. Ennek ismeretében a négy klaszterezési módszer nem ajánlható hibrid esetek vizsgálatára. A kudarcot a hibridek fenotípusos karaktereinek ingadozó, „oda-vissza” jellegével magyarázhatjuk. Ugyanabban a hibridben egy karakter megközelítheti az A szülői faj helyzetét, a második karakter közel lehet a B szülői fajhoz, és egy harmadik karakter közbenső lehet (Kulmuni és mtsai. 2010, Seifert és mtsai 2010, Steiner és mtsai 2011, Bagherian és mtsai 2012). Ez a csomópontok kialakításában vagy a klaszterek elosztásában nagyfokú instabilitást okoz.

Hibrid rendszerek jelenlétére pusztán morfológiai jellegű vizsgálatokból történő következtetést amúgy sem javasolhatom. Mindenképpen szükséges valamilyen nukleáris gén (nuDNS) bevonása a vizsgálatokba, amellyel a hibridizáció jelenlétét kétségtelenül igazolni lehet. Amennyiben ez megtörtént, morfológiai oldalról a tradicionális morfometria helyett inkább geometriai morfometrikus módszerek javasolhatók, amelyet már sikerrel alkalmaztak *Myrmica* hangyákon (Bagherian és mtsai 2012). A főkomponens analízist szintén hasznosnak találták hibridizáció elemzésekor, abban az esetben, ha a vizsgált karakterkészletet a legindikatívabb szelekcióra csökkentették (Seifert, Ritz & Csősz 2014).

5.3 Az új módszertani elemek gyakorlati alkalmazása a biodiverzitás kutatásában – NC-klaszterezés alkalmazása Nyugat Palearktikus hangya fauna taxonómiai feltárásában

Az NC-klaszterezés a gyakorlati alkalmazását a dolgozatban két, saját magam által elvégzett (Csősz és mtsai. (2014) *Org. Divers. Evol.*, 14, 75-88. <http://link.springer.com/article/10.1007/s13127-013-0153-3>; Csősz, Mikó & Heinze (2015) *PLoS ONE* 10: e0140000. 62 p., <http://dx.doi.org/10.1371/journal.pone.0140000/>, etc.) integratív taxonómiai revízió mutat be, melyekben a Nyugat-Palearktikus *Temnothorax* fajok (13. ábra) morfológiai mintázatát – így az NC-klaszterezés által felismert mintázatot – mitokondriális, COI szekvenciák elemzésével tudom megerősíteni¹⁴.



13. ábra. Fáradt társát hazaszállító *Temnothorax affinis* dolgozó. A *Temnothorax* fajok Európában minden felé elterjedtek, többnyire néhány száz fős kolóniákban élnek és dolgozóik ritkán haladják meg a 3 mm-es testhosszúságot. (Fotó: Rahmé Nikola)

Az NC-klaszterezés elsődlegesen taxonómiai kérdések jobb és gyorsabb tisztázására kialakított eljárás. A módszert megjelenése óta rangos folyóiratokban publikált (Derkarabetian & Hedin (2014) *PLoS ONE*, 9: e104982.; Gratiashvili (2014) *Org. Divers. Evol.*, 14: 259-267.; Wachter és mtsai. (2015) *Mol. Ecol.* 24: 863-889., etc.), integratív taxonómiai munkákban használják morfológiai adatok elemzésére. A morfológiai alkalmazásokon túl ismeretes

¹⁴ Megjegyzésként szót kell ejteni arról a tényről, hogy a mitokondriális gének kizárólagos használata ellen elég sok érv szól (ld. McFadden és mtsai. 2011), ezért önálló taxonómiai hipotézis alkotásra kevésbé alkalmasak. Más forrásokból gyűjtött adatsorokkal például morfológiai adatokkal együtt, illetve kereszt ellenőrzésekre, azonban a COI szekvenciák biztonsággal alkalmazhatók.

kutikuláris hidrokarbonát vizsgálatoknak elemzésére történő alkalmazása is (Guillem és mtsai. (2014) *Current Zool*, 60: 62-75.; (2016) *J. Chem. Ecol.*, 42: 1052-1062.). Én magam a módszer bevezetése óta elvégzett revízióimban 4 európai és afrikai genusz 73 valid fajtát kezeltem faji szinten, melyekből 50 bizonyult tudományra új taxonnak. Ezek közül becsléseim alapján 25-35 taxon tartozhat a tipikusan *kriptikus fajok* kategóriájába, ezen taxonokat az NC-klaszterezés alkalmazása nélkül nagy valószínűséggel pusztán külső morfológiai vizsgálatok alapján nem tudtam volna felismerni.

A példaként szolgáló két revíziós munka a morfológiai karakterek alapján alkotott fajhipotézis támogatására rendelkezésre álló mtDNS szekvenciák mellett több okból is ideálisnak mondhatók a módszer bemutatására:

- a) *Temnothorax* genusz fajai monomorfok, azaz egy fajt vagy populációt alkotó példányok méret vagy alaki szórása nagyon kicsi. Ez esetben a populáció legnagyobb példánya kevesebb mint másfélszer akkora, mint az ugyanazon populáció legkisebb egyede.
- b) Sok az ún. *kriptikus faj*. A fajok közötti morfológiai különbségek nagyon kicsik, a különböző populációk morfológiai karakterei egymással széles skálán átfednek, amely rendkívül megnehezíti a hagyományos taxonómiai módszerekkel való munkát. A *kriptikus fajok* nagy aránya segít kiemelni a módszer nagyfokú megbízhatóságát.
- c) A genusz európai populációiból megfelelő mennyiségű mintát gyűjtöttünk morфомetrikus vizsgálatok céljára, amely elegendő volt egy robusztus mintázat felismeréséhez és teszteléséhez.
- d) A régió biogeográfiai viszonyairól és ökológiai paramétereiről rendelkezésre álló információk segítik a felismert biodiverzitási mintázatok megértését és ellenőrzését.

5.3.1 A munka aktualitása

A Myrmicinae alcsaládba tartozó *Temnothorax* hangyagenusz (korábban a *Leptothorax* genuszba tartozott szub-genuszként), a ma ismert 380 valid fajjal, a legdiverzebb hangyagenuszok közé tartozik világszerte (<http://www.antcat.org>), és viszonylag sok fajjal képviselteti magát Európában is. A genusz fajaival nagyon sokféle környezetben találkozhatunk, gazdag szociális magatartás repertoárral rendelkeznek (Fokuhl és mtsai. 2012), és egyéb fajokkal való kapcsolatuk (hangyavendégek, szociálpazita fajok) is széles körű (Hölldobler & Wilson 1990). A kis kolóniaméret (100-200 dolgozó alkot egy kolóniát), a

könnyű laboratóriumi gondozás, valamint a széles viselkedési repertoár miatt a *Temnothorax* fajokat (**14. ábra**) gyakran alkalmazzák modell organizmusokként viselkedés ökológiai és szociobiológiai kutatások számára.

Az itt példaként bemutatott munkák további aktualitását az a tény adta, hogy bár az elmúlt évtizedek intenzív taxonómiai munkáinak (Seifert 2007, Czechowski et al 2012, Boro-wiec 2014) köszönhetően az európai hangyafaunáról szóló ismereteink sokat gyarapodtak, a *Temnothorax* faunáról azonban jellemzően korábbi (Arnol'di 1977, Bernard 1956, Radchenko 1995, stb.) források álltak rendelkezésre. Az eddigi munkák alapján azonban elmondható, – amelyet saját kutatásaim is megerősítenek – hogy a *T. nylanderi* fajcsoport Ponto-Mediterrán régióban¹⁵ tapasztalt fajgazdagsága átlagon felüli. A fajcsoport nagy fajgazdagsága társul az ún. *kriptikus* fajok nagy arányával, melyek szélesen átfedő morfológiai karakterei nagymértékben megnehezítik a fajhatárok hagyományos technikákkal történő felismerését.



14. ábra. Az európai *Temnothorax* fajok kolóniái kedvelt modell szervezetek laboratóriumi vizsgálatokhoz. (Fotó: University of Regensburg)

A bemutatott munkákban ezért a faj határokat morfometrikus valamint molekuláris filogenetikai módszerek kombinációjára épülő integratív módon határoztuk meg. A morfológiai mintázat felismerésére az újonnan bevezetett feltáró adatelemzés típust, az NC-klaszterezést alkalmaztuk, amely előzetes fajhipotézis megalkotása nélkül segít felismerni az adatokban rejlő struktúrát. Az 5.2.6. fejezetben megfogalmazott ajánlásokkal egybehangzóan a morfológiai fajhipotézis megbízhatóságának növelése céljából több módszer (NC-Ward, NC-NMDS és NC-K-közép) egyesített eredményéből következtettünk fajhatárok helyzetére.

¹⁵ A Földközi tengert és a Fekete tengert környező térséget értve ez alatt.

5.3.2 Anyag és módszerek

A jelen dolgozat legfőbb témája a morfológiai adatok taxonómiai célú mintázatfelismerése, ezért e fejezetben a morfológiai adatelemző módszerekre fókuszálok, emellett csak az adatfelvétel legfontosabb sajátosságait közlöm. A morfológiai karakterek részletes leírása, valamint a molekuláris módszertani lépései a mellékletekben (9.3. és 9.4. melléklet-ekben) vannak megadva.

5.3.2.1 A morfometrikus karakterek mérésének protokollja

A legnagyobb konzisztencia elérése érdekében a vizsgálat alatt nem változtattam sem a beállításokon, sem pedig az eszközökön (mikroszkóp objektív, tárgytartó, stb.). Minden mérést Olympus SZX9 sztereó mikroszkóppal, x150 nagyításon, mikrométeres skálán végeztem. A mérések átlagos pontosságát (*error rate*) $\pm 2 \mu\text{m}$ -re becsültem. A minták előre-hátra és oldalirányú finom mozgathatóságú tárgyasztalra rögzítését forgatható tárgytartó állvány biztosította. Minden mérést magam végeztem. A hangyák ivaros egyedeinek (királynők és hímek) viszonylagos kis száma miatt a morfometrikus vizsgálatokat dolgozókon végeztem. Összesen 22 morfometrikus jelleget vettem fel, és elemeztem. Ezek pontos definícióját a 9.4 mellékletben adom meg. A morfológiai elemzéshez 627 fészeksorozat 2052 dolgozójának (Csősz és mtsai. 2014: 101 fészek / 359 egyed; Csősz és mtsai. 2015: 526 fészek / 1693 egyed) 22 morfometrikus jellegét mértem le, azaz összességében egy 40,000 cella adatait tartalmazó mátrix segítségével sikerült a Nyugat-Palearktikus *Temnothorax* fajok morfológiai mintázatára következtetnem.

Az itt bemutatott példában szereplő munkák mellett az NC-klaszterezési eljárást több Palearktikus hangya revízióban is sikerrel alkalmaztam. Az említett munkák szintetizálása a molekuláris filogenetikai támogatás hiánya (Seifert, Csősz & Schulz (2014) *Beitr. Entomol.*, 61: 47-57.; Seifert & Csősz (2015) *ZooKeys*, 479: 37-64.), vagy a revideált (*Tetramorium*) hangya genusz (Csősz és mtsai. (2014) *Zool. Anz.* 253: 469-481.) a *Temnothorax*-októl való rendszertani távolsága – egyben az eltérő morfológiai karakter-készlete – miatt nehézkes és bonyolult lenne. A módszert hét további alkalommal alkalmaztam trópusi ökoszisztémákon is.

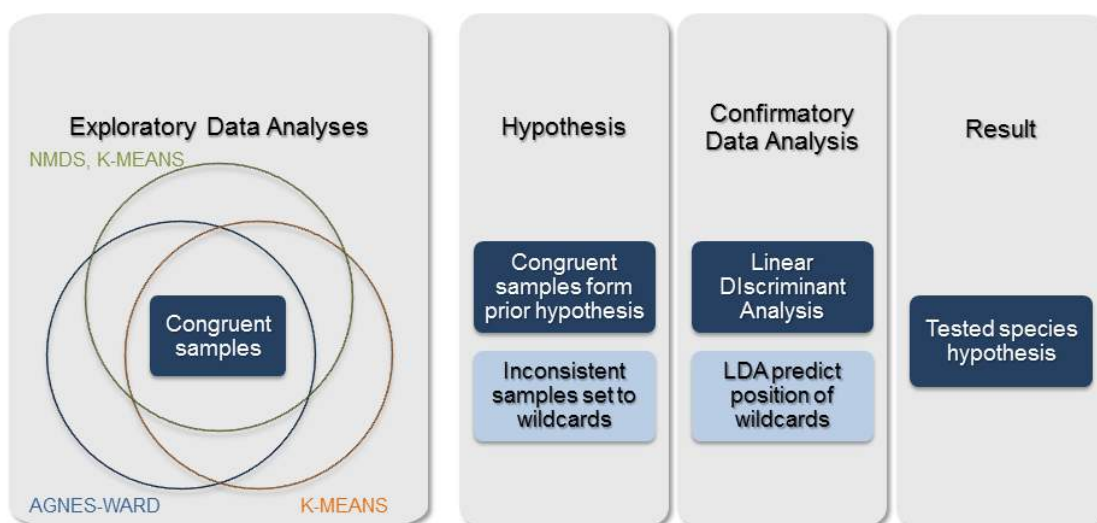
5.3.2.2 A statisztikai feldolgozás

A morfometrikus elemzések által kialakított fajhipotéziseket (azaz a felismert morfológiai mintázat érvényességét) megerősítő LDA-val teszteltük. Az így kialakított – és kvalitatív morfológiai jegyekkel megerősített – csoportosítás alkotta a *végleges morfológiai fajhipotézist* (6. ábra). A *végleges morfológiai fajhipotézis* megerősítésére, valamint kereszt-tesz-

telésére a COI mitokondriális gén szakasz alapján számolt törzsfát használtuk mindkét példaként említett munka esetében. A felismert klasztereket csak abban az esetben tekintettük biológiai tartalommal rendelkező egységeknek (azaz fajoknak, de egyaránt ideértve a korábban ismert taxonokat, valamint az új fajokat is), ha minden általunk alkalmazott diszciplína eredményei egyaránt erre utaltak. Eltérő hipotézisalkotás esetén, amikor egyik vagy másik diszciplína nem mutatott elkülönülést, a *stopping rulet* (Schlick-Steiner és mtsai. 2010) alkalmaztam, amely kimondja, hogy a klasztereket addig a pontig érdemes fajoknak tekinteni, ameddig a megközelítések között egyetértés van a klaszterek határait illetően.

5.3.2.3 Morfometrikus adatok feltáró elemzése

A nyugat-palearktikus *Temnothorax* fajok folytonos morfológiai változókból számított mintázatának értékelésére az általam bevezetett feltáró NC-klaszterezést alkalmaztam kombinálva az ordináló NMDS algoritmussal és a K-közép módszerrel. A három független elemzés típus metszéspontjában adott döntések alkották a *végleges morfológiai fajhipotézist* (bővebben ld. **6. ábra**), az egyes mintákra vonatkozó fajhipotézis csak akkor fogalmazódik meg, ha mind a három feltáró típusú elemzés ugyanarra a következtetésre jutott.



15. ábra. A végleges morfológiai fajhipotézis kombinált elemzési stratégiával történő megerősítése. Többszörös feltáró jellegű adatelemzés (Exploratory Data Analyses) és a megerősítő elemzésekkel (Confirmatory Data Analyses) végrehajtott taxonómiai folyamat az ábrázolt lépésekkel elvezet a végleges morfológiai fajhipotézishez. A három független elemzéstípus metszéspontjában található minták (congruent samples) alkotják a végleges morfológiai fajhipotézis elemeit. A metszésponton kívül eső elemek (inconsistent samples) csoportosítás nélkül (wild-card) kerülnek a megerősítő elemzés bemenetébe. (Csősz unpubl.)

Abban az esetben, ha bármelyik módszer alternatív döntést hozott, a minta wild-card-ként került a megerősítő elemzés adatmátrixába (**15. ábra**). Ezzel az elgondolással igyekeztem az elérhető legmegbízhatóbb morfológiai fajhipotézist megalkotni, amely megérett a további tesztelésekre. Ez az eljárás nagyfokú stabilitást eredményez a tökéletesített morfológiai fajhipotézis megalkotása terén, és működik közepes és kis adatmennyiségből való következtetések során (pl. **Csősz és mtsai. 2014**). Nagyobb adatsoroknál, melyekben száz feletti mintát kell elemezzünk (pl. **Csősz és mtsai. 2015** adatsorában 525 mintát elemeztünk) az iteratív beállításokkal partíciónálós NMDS és K-közép módszer olyan nagy számítási képességet igényel, hogy szintén több lépésben, de egyszerűbben kell döntést hozzunk. A **Csősz és mtsai. (2015)** munkafolyamata annyiban módosult, hogy a legkevesebb számítási képességet igénylő NC-klaszterezés egymagában alkotta a *végző morfológiai fajhipotézist*, viszont a teljes adatsorra elvégzett megerősítő elemzést a fajcsoportonként, azaz alegységenként megismételttem, minimalizálva a hibás osztályozás lehetőségét.

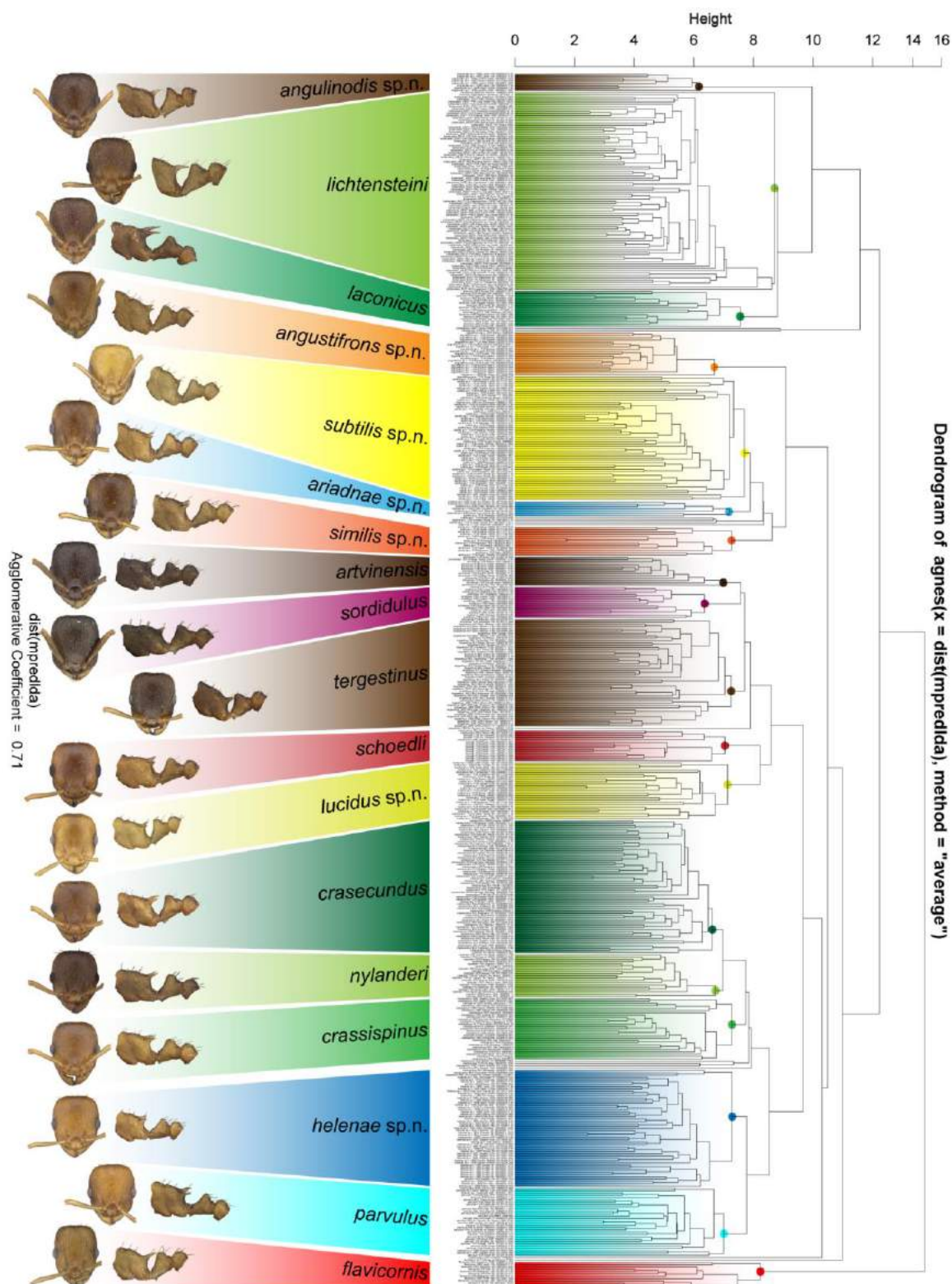
5.3.3 Eredmények

Az eredmények¹⁶ a *T. nylanderi* fajcsoport Ponto-Mediterrán elemeinek nagyfokú diverzitását mutatják. Az újonnan bevezetett morfometrikus elemzések nagyban hozzájárultak a régió fajgazdagságának morfológiai feltárásához, a sok ún. *kriptikus faj* felismerése hagyományos módszertannal nehezebb, vagy megoldhatatlan lett volna.

5.3.3.1 Morfológiai eredmények

Az NC-klaszterezés összesen 18 morfológiai klasztert ismert fel (**16. ábra**). Ezen klaszterek több mint fele külső (kvalitatív) morfológiai módszerekkel is igazolható módon elkülönül egymástól. Az ilyen taxonok más fajcsoportok fajainál megszokott morfológiai különbségeket mutatnak. Más klaszterek (az összes klaszter kb 40%-a) ún. *kriptikus fajok*, amelyek pusztán hagyományos morfológiai eljárásokkal nem különíthető el, de – mint lejjebb látjuk – morfometrikus változók statisztikai elemzése, valamint molekuláris vizsgálatok egymást támogatva igazolják a reprodukzív barrierek jelenlétét, azaz a faji szint ezek számára is indokolt. A kapott mintázat Lineáris Diszkriminációs Analízissel (LDA) is jól megerősíthető. Az osztályozás sikeressége összességében 95% volt, amely jó eredménynek mondható, és faji szintre lebontva is megfelelően megbízható képet kapunk (**4. táblázat**).

¹⁶ E fejezetben a két cikk eredményeit (**Csősz és mtsai. (2014) Org. Divers. Evol., 14: 75.; Csősz, Mikó & Heinze (2015) PLoS ONE 10: e0140000. 62 p., <http://dx.doi.org/10.1371/journal.pone.0140000>**) együtt mutatom be.



16. ábra. A *Temnothorax nylanderi* fajcsoport Ponto-Mediterrán fajainak UPGMA agglomerációs módszerrel kombinált NC klaszterezés dendrogramja. A cédulák információs sorrendje a következő: progresszív fajhipotézis – Ország_lelőhelyhez legközelebbi város_gyűjtés dátuma_minta kód (amennyiben rendelkezésre állt). Csősz, Mikó & Heinze (2015) PLoS ONE 10: e0140000. alapján.

observed \ predicted		classification success (%)																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			
		predicted																			
		observed																			

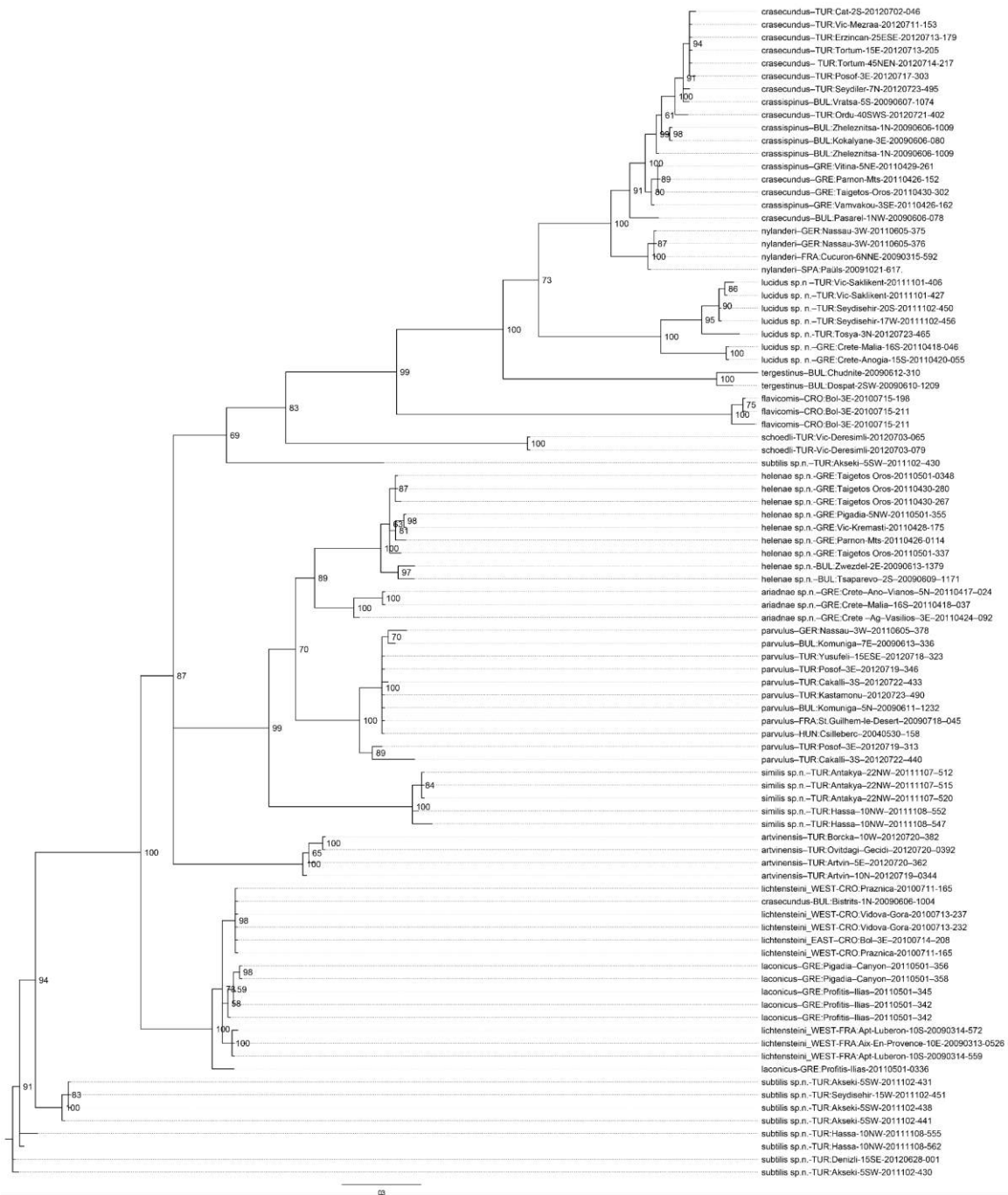
4. táblázat. A Temnothorax nylander fajcsoport, LDA-val végzett megerősítő elemzésének eredményei. A bal oldali táblázat a hagyományos, a jobb oldali tábla a LOOCV-LDA eredményeit mutatják.

Több faj osztályozása teljesen, vagy majdnem teljesen sikeres¹⁷ volt: *T. angulinodis* sp.n. (100%), *T. angustifrons* sp.n. (98.5%), *T. ariadnae* sp.n. (100%), *T. artvinensis* (95.8%), *T. flavicornis* (100%), *T. laconicus* (94.9%), *T. lichtensteini* (97.3%), *T. lucidus* sp.n. (94.3%), *T. parvulus* (97.1%), *T. schoedli* (93.8%), *T. similis* sp.n. (94.7%), *T. sordidulus* (95.6%) és *T. subtilis* sp.n. (95.5%). Néhány másik faj gyengébb eredményt mutatott: *T. crasecundus* (89.6%), *T. crassispinus* (90.4%), *T. helenae* sp.n. (91.3%), *T. nylanderi* (90%), and *T. tergustinus* (89.9%). A 90% körüli klasszifikációs siker már az a szint, ahol a fajhipotézisek újragondolása indokolt. Ezért ezekre az esetekre egy ismételt elemzést futtattam, melyekben a fajkomplexek fajainak újra elemzése megtörtént. Ezek eredményeit azután összevetve a mtDNS elemzések eredményeivel *progresszív fajhipotézissé* (5.2.4.2. alfejezet) alakítottam. Az ún. *kriptikus fajok* száma a *Temnothorax* génuszban is viszonylag magas, megközelíti az 50%-ot. Ilyenek például a *T. crasecundus* Seifert & Csősz, 2015 és a *T. crassispinus* (Karawaiev, 1926) dolgozói, melyek morfológiai értelemben annyira hasonlóak, hogy a biztonságos meghatározás eléréséhez a fészeksorozatok vizsgálata szükséges, vagyis egy fészekből 2-3 dolgozó vizsgálatának alapján vonjunk le faji következtetéseket. A fészeksorozatok meghatározása ugyanis lényegesen jobb (98.5%-100%) pontosságot eredményez ezekben az esetekben is.

5.3.3.2 Molekuláris eredmények

Bár a munkám és célkitűzéseim alapvetően a morfológiai eredményekre illetve a morfológiai döntéshozási protokollok jobbítására irányultak, az eredmények igazolására molekuláris filogenetikai megközelítést is alkalmaztunk. Ennek segítségével teszteltem a morfometrikus módszerek használhatóságát. A molekuláris COI szekvenciák elemzéséből született törzsfa (*Bayesian tree*, **17. ábra**) teljes mértékben alátámasztotta a morfometrikus alapon felállított fajhipotézist, melyet a fajok nagy részének (*T. ariadnae* sp.n., *T. artvinensis*, *T. flavicornis*, *T. helenae* sp.n., *T. lucidus* sp.n., *T. nylanderi*, *T. parvulus*, *T. schoedli*, *T. similis* sp.n.) magas posterior probability értékei is megerősítene. A molekuláris módszer nem tudta teljes mértékben alátámasztani a *T. laconicus* - *T. lichtensteini*, valamint a *T. crasecundus* - *T. crassispinus* fajpárokat. Ez tulajdonképpen megegyezik a morfometrikus módszerekkel kapott eredmények megállapításaival, vagyis, hogy ezek a fajpárok *kriptikus fajokból* állak, morfológiailag nehezen megkülönböztethetők, többváltozós morfometrikus eljárások szükségesek a megbízható elkülönítéshez.

¹⁷ A morfometrikus vizsgálatok a legritkább esetben adnak tökéletes elkülönülést. Valamilyen okból kifolyólag, amely lehet mérési hiba, fejlődési rendellenesség, hibridizáció vagy bármi egyéb ok, a minták egy része nem mindig esik a „megfelelő” klaszterbe. Amennyiben ez a hiba arány 5% alatt marad, akkor ezt egyszerű statisztikai hibaként kezelhetjük, ezt a határt jelentősen meghaladó eltérés esetén indokolt lehet a faji hipotézis újragondolása vagy elvetése.



17. ábra. Bayes-féle konszenzusfa a *Temnothorax nylanderii* fajcsoportba tartozó minták mitokondriális CO I génjének 658 bázispárból álló elemzéséről. A bayes-i posterior valószínűségeket (százalékosan) a csomópontokon adom meg. A molekuláris fa 97%-ban megerősíti a morfológiai alapon elkülönített minták csoportosítását. Csősz, Mikó & Heinze (2015) PLoS ONE 10: e0140000. alapján. Nagyobb felbontású verzió az alábbi linken <http://dx.doi.org/10.1371/journal.pone.0140000> elérhető.

5.3.4 Konklúzió

A *Temnothorax* genusz, ezen belül a *T. nylanderi* fajcsoport Ponto-Mediterrán elemeinek diverzitására irányuló integratív taxonómiai vizsgálat sorozat kombinált bizonyítékaiból összesen 18 faj jelenlétét sikerült kimutatni. Ezen fajokból kilenc, tehát több mint a teljes fajszaám fele, bizonyult tudományra új fajnak. A csoport általunk leírt fajgazdagsága alátámasztja a korábban feltételezett – de megfelelő módszerek hiányában ki nem mutatott – faji szintű sokféleséget. A fajcsoport elemeit faj-komplexekre osztottam, melyekben 1-5 faj található, és melyekről morfológiai karakterek vagy karakter kombinációk értékeléséből valószínűsíthetően monofiletikus¹⁸ egységet alkotnak.

A munkáim során feldolgozott taxonok listája a következő:

***Temnothorax angulinodis* complex.**

Temnothorax angulinodis Csősz, Mikó & Heinze, 2015

***Temnothorax angustifrons* complex.**

Temnothorax angustifrons Csősz, Mikó & Heinze, 2015

Temnothorax lucidus Csősz, Mikó & Heinze, 2015

Temnothorax similis Csősz, Mikó & Heinze, 2015

Temnothorax subtilis Csősz, Mikó & Heinze, 2015

***Temnothorax flavicornis* complex.**

Temnothorax flavicornis (Emery, 1870)

***Temnothorax lichtensteini* complex.**

¹⁸ Szükséges megjegyezni, hogy a monofília pusztán morfológiai és mitokondriális szakaszok adataiból teljes mértékben nem igazolható. Ehhez magi szakaszok vizsgálata lenne szükséges.

Temnothorax lichtensteini (Bondroit, 1918)

Temnothorax laconicus **Csősz** & al., 2014

***Temnothorax nylanderi* complex.**

Temnothorax nylanderi (Foerster, 1850)

Temnothorax crassispinus (Karavaiev, 1926)

Temnothorax crasecundus Seifert & **Csősz**, 2015

***Temnothorax parvulus* complex.**

Temnothorax parvulus (Schenck, 1852)

Temnothorax ariadnae **Csősz**, Mikó & Heinze, 2015

Temnothorax helenae **Csősz**, Mikó & Heinze, 2015

***Temnothorax sordidulus* complex.**

Temnothorax artvinensis Seifert, 2006

Temnothorax schoedli Seifert, 2006

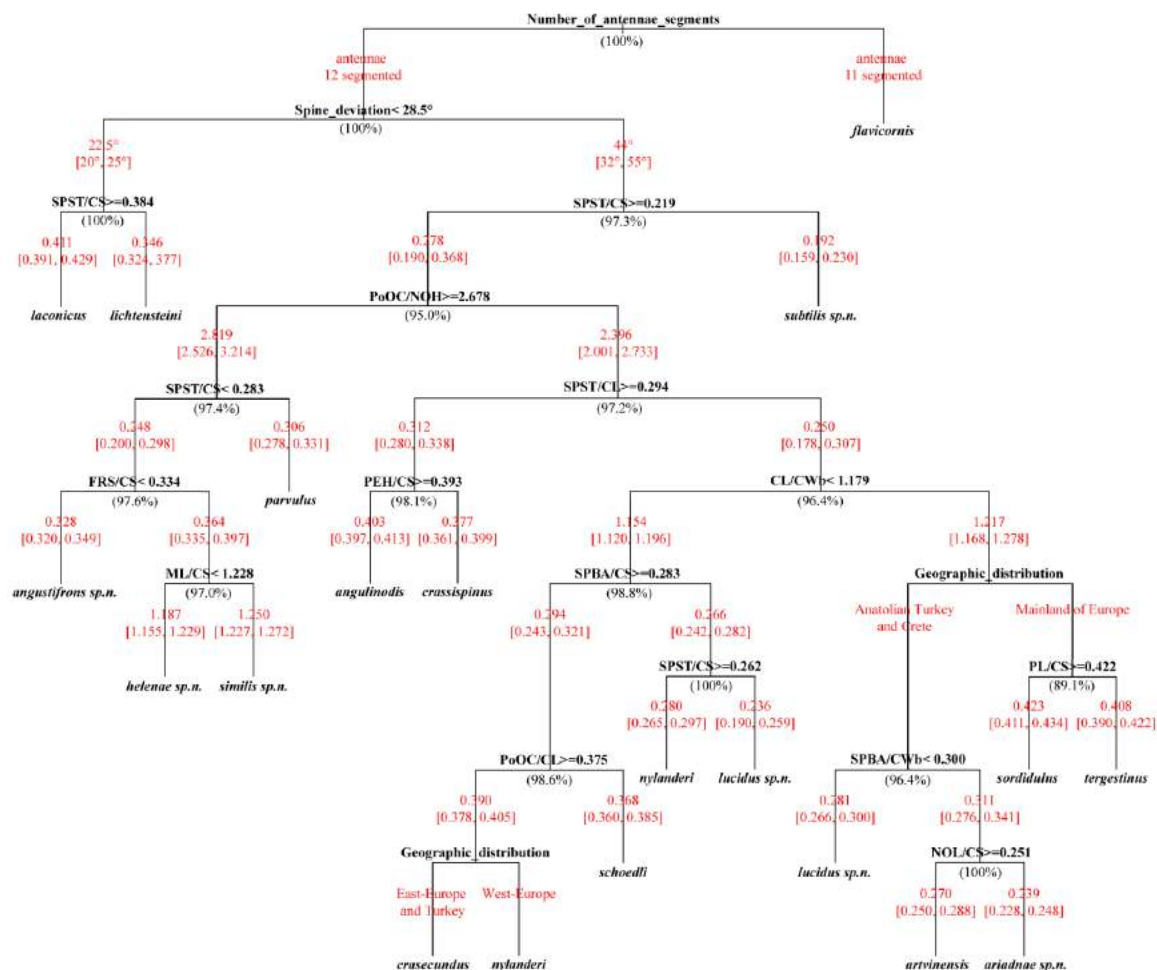
Temnothorax sordidulus (Müller, 1923)

Temnothorax carinthiacus (Bernard, 1957)

Temnothorax tergestinus (Finzi, 1928) stat.n. (lásd **Csősz**, Mikó & Heinze 2015)

Temnothorax saxonicus (Seifert, 1995) syn.n. (lásd **Csősz**, Mikó & Heinze 2015)

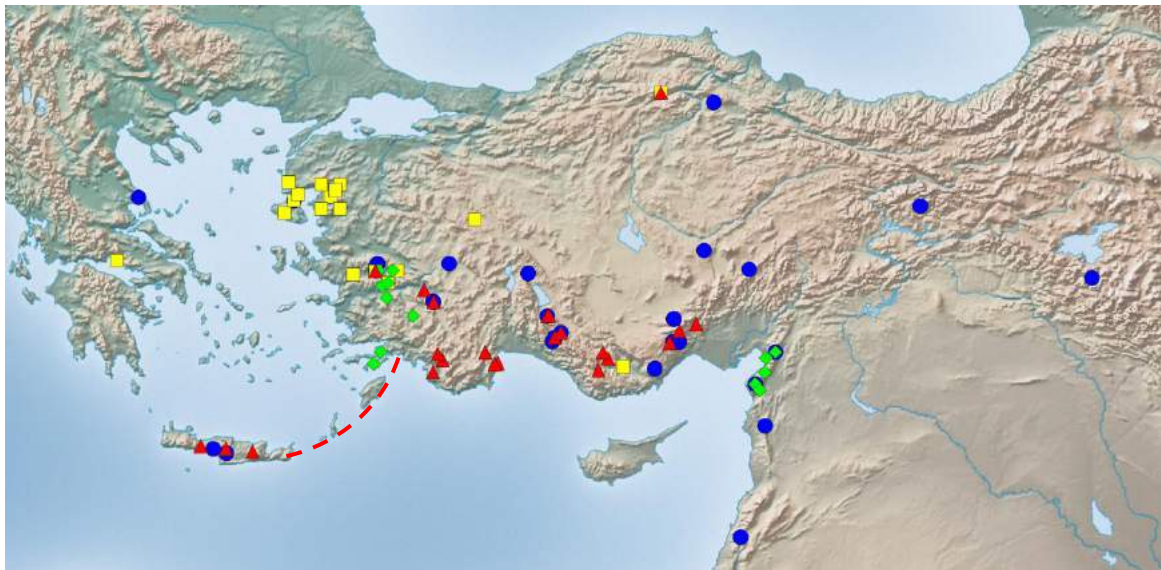
A fenti listában szereplő taxonok morfometrikus eljárásokkal valamint molekuláris vizsgálatokkal segített elkülönítése után erőfeszítéseket tettem a fajok hagyományos, külső morfológia alapján történő elkülönítésére.



18. ábra. A *Temnothorax nylander* fajcsoport dolgozóinak elágazásos formában történő azonosítására alkalmas döntési fa (Cross-Validated Decision Tree). Az ábra a dolgozók határozásához könnyen használható és gyorsan átlátható grafikus megjelenítést mutat be. A százalékos osztályozási sikert az egyes csomópontok alatt zárójelben adom meg. A csomópontok felett vastagon szedve a döntésben szerepet játszó egyszerű aránypárt mutatom. A morfometrikus arányok átlagai, minimális, maximális értékei szögletes zárójelben minden ágon vörös színnel láthatók. Két pozícióban (csomópont) az eredetileg javasolt folytonos morfometriai jellemzőket felváltottam megbízhatóbb geográfiai adatokkal. Csősz, Mikó & Heinze (2015) PLoS ONE 10: e0140000. alapján. Nagyobb felbontású verzió az alábbi linken <http://dx.doi.org/10.1371/journal.pone.0140000> elérhető.

Eredményül, fentebb felsorolt taxonok közel 50%-a bizonyult bár ún. *kriptikus faj*nak, ezek egymástól viszonylag egyszerűen és megfelelő osztályozási sikerrel valamelyik egyszerű morfológiai aránypár segítségével elkülöníthetők, ahogy azt a döntési fa (18. ábra) segítségével ábrázoltam. A fajcsoport fajai egymással kevés kivételtől eltekintve parapatrikus elterjedési mintázatot adnak, azaz csak egy nagyon szűk kontaktzónában mutatnak átfedést.

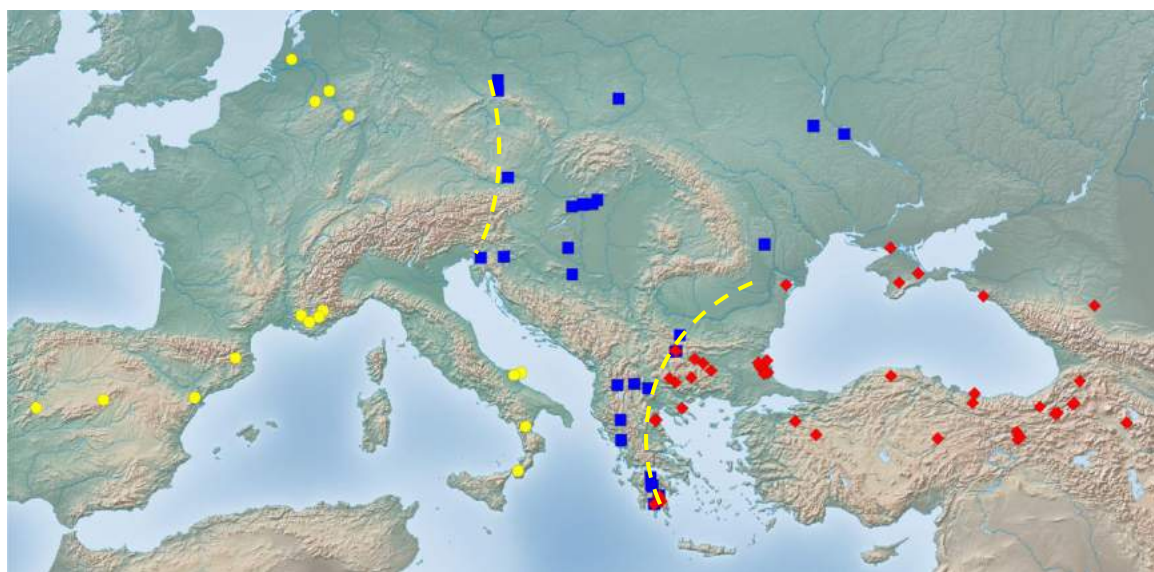
A felderített nagy fajgazdagság nem egyenletesen oszlik el a régióban, hanem különösen fajgazdag fajképződési gócok láthatóak sok endémizmussal a Peloponészoszi-félszigeten, valamint Anatólia területén. Az *angustifrons* fajkomplex szinte összes ismert lelőhelye (19. ábra) Anatólia területén és Kréta szigetén található, a két görögországi lelőhely lehet ember általi behurcolás eredménye is. Ez utóbbi elméletet támasztja alá, hogy mindkét ismert európai lelőhely kikötővároshoz nagyon közel fekszik. A fajcsoport fajainak Kréta-szigeti megjelenése magyarázható az elmúlt glaciális időszak tengereringadozásaival összefüggő folyosók létrejöttével. A Kréta-Rodosz tengeri hátság és a rajta fekvő szigetek (Karpathosz, Rodosz) az Anatóliai Marmarisz-öbölbe fut be.



19. ábra. A *Temnothorax angustifrons* komplex fajainak elterjedése ismert lelőhely adatok alapján. Színkódok: *T. angustifrons* sp.n. (sárga négyszög), *T. lucidus* (vörös háromszög), *T. similis* (zöld rombusz) és *T. subtilis* (kék kör). Vörös szaggatott segédvonal jelzi a Kréta-Rodosz tengeri hátságot és a rajta fekvő szigetekkel (Karpathosz, Rodosz). Ez a geológiai formáció köti össze Kréta szigetét az anatóliai Marmarisz-öböllel. Csősz, Mikó & Heinze (2015) PLOS ONE 10: e0140000 nyomán.

Ez az útvonal a jelenlegi, „magas” tengerszint mellett is lehetséges betelepülési útvonalat kínál a közepes röpképességű *Temnothorax* fajok számára, a múltban viszont, például a Würm glaciális idején uralkodó alacsonyabb tengerszint mellett, az anatóliai faunaelemek sokkal kevesebb akadály leküzdése árán juthattak a szigetre.

Érdekesebb kérdés a *T. angustifrons* fajkomplex fajainak európai hiánya. Bár a *T. subtilis* szórványos előfordulását Görögország északi részéről már jelezték (Lech Borowiec szóbeli közlése alapján), az *angustifrons* fajkomplex európai előfordulásának hiánya így is szembeötlő. Egyértelmű magyarázattal nem tudok szolgálni, de valószínűleg a már Európában jelen lévő rokon fajok hasonló ökológiai igényei, illetve ennek okán az ökológiai kizárás lehet a jelenség egyik oka. A *T. angustifrons* fajkomplex, eltérően a többi faj feletti egység fajaitól Anatóliában sok szinpatikus fajjal van jelen. Ennek lehet magyarázata a fajok kora, vagyis a fajképződés óta eltelt idő, melynek során mind az ökológiai igényekben mind pedig a reprodukív rendszerekben olyan szintű változások történhettek, amely lehetővé teszi a fajok együttes előfordulását.



20. ábra. A *Temnothorax nylanderi* fajkomplex fajainak elterjedése az ismert lelőhely adatok alapján. Színekódok: *T. crasecundus* (vörös rombusz), *T. crassispinus* (kék négyzet), *T. nylanderi* (sárga kör). A sárga szaggatott vonalak jelzik a tipikus elterjedési határokat. A „nyugati barrier” a Balkáni LGM refúgium és az Appennini-félsziget refúgiumaiból előretörő populációk mai elterjedésének közös határán jött létre. A „keleti” barrier döntően a nagyjából 20.000 éve létezett Balkáni LGM refúgium, valamint az ún. Ukrán LGM refúgium populációinak expanziója útján alakult ki. Csász, Mikó & Heinze (2015) PLOS ONE 10: e0140000 nyomán.

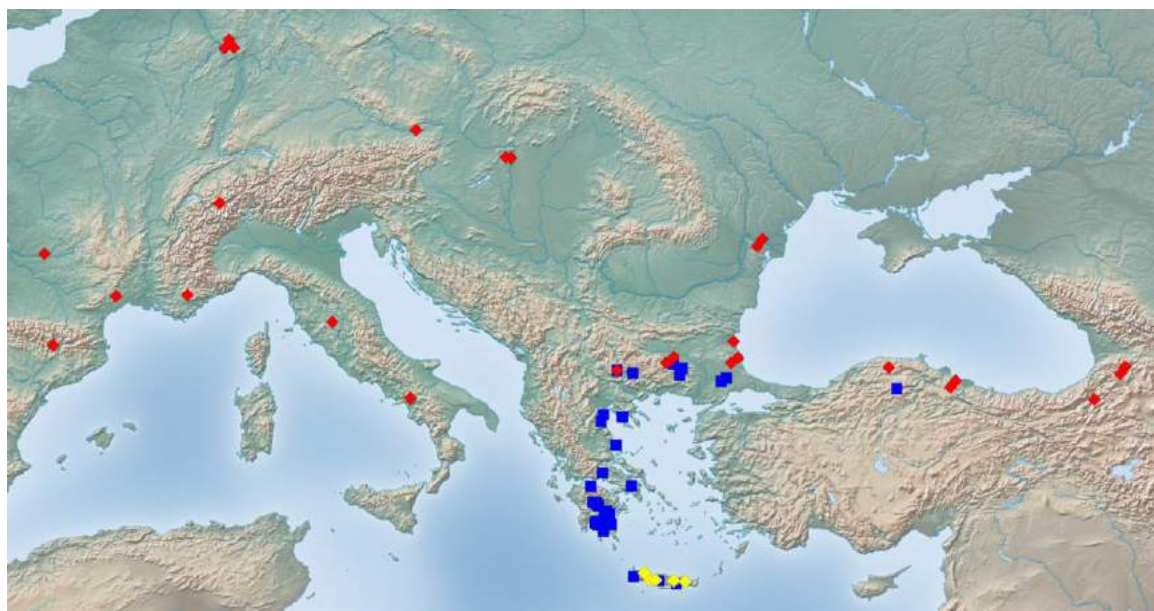
Érdemes megemlíteni még a *T. similis* széttagolt elterjedését (**19. ábra**). Az egymással nem érintkező két területen élő populációi morfológiailag nem különböznek egymástól, ezért ezeket a széttagolt áréában élő populációkat ugyanazon taxonba soroltam. Sajnálatos módon az Iskenderun térségében honos populációból nem sikerült működő COI szekvenciákat kinyerni, ezért a faji rangú azonosságot ebben az esetben morfometrikus elemzéseken túl más módszerrel nem tudtam alátámasztani.

A *T. nylanderi* fajcsoport Ponto-Mediterrán fajai – az *angustifrons* fajkomplex elemein kívül – döntően Európa déli részén honosak. A fajkomplexeken belüli mintázatokban sok hasonlóságot fedezhetünk fel. Az egyik ilyen elem a parapatikus fajpárok jelenléte (**20. ábra**), amely minden fajcsoportban – egy vagy több esetben – megjelenik.

A másik fontos közös jellemző az elterjedési határok (esetünkben kontakt zónák) elhelyezkedésének nagyfokú egyezése. Ezen európai fajok döntően két barrier mentén találkoznak: a keleti barrier Nyugat-Anatólia és a Görög, Balkán faunaterületek között húzódik, a „keleti” barrier elhelyezkedése a Rodope-hegységhez kötődik; a nyugati barrier a Balkán-félsziget északi részének populációit választja el az Appennini-félsziget elemeitől, a „nyugati” barrier Szlovéniától az Alpokálján keresztül húzódik északra (**20. ábra**). Ezeket a recens elterjedési mintázatokat a legutóbbi glaciális idején kialakult ún. refúgium területeknek, illetve az interglaciális időszakban innen szétterjedő populációk által kialakított mintázatoknak köszönhetjük (Hewitt 1999; Goropashnaya és mtsai. 2007; Litvinchuk és mtsai. 2013). A „keleti” barrier döntően a nagyjából 20.000 éve létezett ún. Balkáni LGM¹⁹ refúgiumokból, valamint a Nyugat-Anatólia vagy az ún. Ukrán LGM refúgium területéről szétterjedő populációk találkozási frontján alakult ki. A „nyugati barrier” a Balkáni LGM refúgiumokból és az Appennini-félsziget refúgiumaiból előretörő populációk mai elterjedésének közös határán jött létre. A felvázolt kép ismert más hangya csoportok esetében is. Hasonló „nyugati” parapatikus elterjedési mintázatot írtak le *Lasius piliferus* Seifert, 1992 és *Lasius psammophilus* Seifert, 1992 fajpár esetében (Seifert 1992), *Myrmica spinosior* Santschi, 1931 és *M. sabuleti* Meinert, 1861 populációi között (Seifert 2005) és a *Tetramorium moravicum* Kratochvil, 1941 két jól elkülönülő leszármazási vonala is hasonló mintázatot követ (Schlick-Steiner és mtsai. 2007b).

¹⁹ LGM refúgium = a Last Glacial Maxima, azaz az utolsó glaciális periódus (Würm) maximális eljegesedési időszak alatt (~20.000 évvel ezelőtt) refúgiumként szolgáló területek angolszász irodalomban meghonosított elnevezése.

A másik gyakran visszatérő elterjedési mintázati elem az ún. ugrásszerű szétterjedés („jump”-dispersal), amely egyes szatellit populációk gyors, nagy távolságra jutása útján jön létre. A jelenség több *Temnothorax* fajnál is megfigyelhető, esetünkben egyrészt Nyugat-anatóliai fajok (**19. ábra**), másrészt jellemzően Európai (Balkán-félszigeti) fajok enklávéi jelenhetnek meg Kréta szigetén (**21. ábra**), viszont az Égei-tenger többi, kisebb szigetéről e fajok populációi nem ismertek. Természetesen nem zárható ki az alulgyűjtöttség, mint egy lehetséges ok. Valószínűbbnek tartom azonban, hogy az LGM periódus idején az Anatólia és a Peloponnészosz *Temnothorax* faunája között az Égei-tenger – az alacsonyabb tengerszint miatt jóval nagyobb kiterjedésű – szigetein folyamatos kapcsolatot biztosíthattak. Mára azonban a közbülső területek állatvilágából ezek a fajok hiányoznak, eltűnhettek. Ha ebből a feltételezésből indulunk ki, akkor a recens kép egy rég letűnt időszak maradványait mutatja.



21. ábra. A *Temnothorax parvulus* fajkomplex fajainak elterjedése az ismert lelőhely adatok alapján. Szín kódok: *T. ariadnae* (sárga rombusz), *T. helenae* (kék négyszög), *T. parvulus* (vörös rombusz). A Balkáni LGM refúgiumainak *T. helenae* populációi (kék négyszög) megjelennek Kréta szigetén is. Csósz, Mikó & Heinze (2015) PLOS ONE 10: e0140000 nyomán.

A *T. nylanderi* fajcsoport általam feltárt Ponto-Mediterrán diverzitási mintázata bár döntően morfológiai eszközökkel történt, ellenőrzési céllal mitokondriális COI elemzést alkalmaztunk. A nagyszámú *kriptikus faj*, valamint a nagy fajgazdagság miatt, a *T. nylanderi* fajcsoport feltárása egy másik kutatócsoport több mint egy évtizedes munkája után sem hozott eredményt (Andreas Schulz személyes közlése), majd fel is hagytak vele. A feldolgozás lehetőségét az általam bevezetett eljárás, az NC-klaszterezés hozta el számunkra. A biodiverzitás kutatásának visszatérő eleme, hogy a felismert mintázat, azaz a leírt fajok mennyire feleltethetők meg biológiai fajoknak, vagyis mennyi biológiai információval tudjuk összekapcsolni a diverzitási adatokat.

A Nyugat-palearktikus *Temnothorax* fajok vizsgálatának a dolgozatban vázolt szeletét 2015-ben fejeztem be, hogy egy másik, hasonlóan izgalmas kérdéssel foglalkozhassak, amit a következő, 5.4. fejezetben fogok tárgyalni. A publikációk közlésének időpontjában korológiai, ökológiai információk, életmenet tulajdonság adatok korlátozottan álltak rendelkezésre²⁰. Az elmúlt két évben azonban több kolléga is jelezte, hogy egyes fajok, pl. *Temnothorax ariadnae* Csősz, Mikó & Heinze, 2015 kizárólag Kréta szigetének tűlevelű erdeiben fordul elő (Lech Borowiec személyes közlése), vagy 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. Az újonnan lefektetett Nyugat-palearktikus *Temnothorax* taxonómia tehát új információk társításának, rendezésének adott terepet.

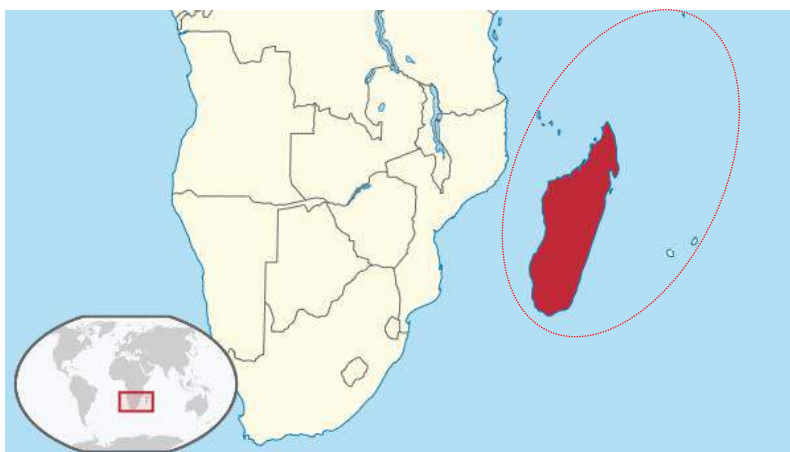
²⁰ A taxonómiai kutatás általános problémája, hogy akkor vagyunk kénytelenek leírni egy taxont, mikor az felfedezésre kerül, azaz egyéb jellemzői a legkevésbé ismertek.

5.4 A Malagasi *Nesomyrmex* fauna taxonómiai feltárása

Madagaszkár Földünk biodiverzitásának egyik kiemelt pontja (Myers et al. 2000), tele egyedülálló, endemikus elemekkel. Az elmúlt évtizedekben komoly erőfeszítések történtek a Malagasi hangyadiverzitás feltárására (Udvardy 1983), melyek alátámasztják a régió nagy fajgazdagságról alkotott képünket (Fisher 2005). A régió geográfaiilag is változatos képet mutat, és kimondottan szétszabdalt, magában foglalja Madagaszkár és a környező szigetvilág (Mauritius, Comores, Mayotte, Seychelles stb.) területeit is (**22. ábra**).

A helyi élővilág feltárásában nagy szerep jut a *California Academy of Sciences* (<http://www.calacademy.org/explore-science/madagascar>) és a *Madagascar Biodiversity Centre* (<http://www.madagascarbio.org/>) közös kezdeményezésének. Ebből az együttműködésből – több mint egy évtized eredményeként – hatalmas mennyiségű, reprezentatív gyűjtemény halmozódott fel a *California Academy of Sciences* gyűjteményében.

22. ábra. A Malagasi régió (piros ellipszissel jelölve) Afrika keleti partjaitól néhány száz kilométerre kerül el. A régiót a legnagyobb sziget, Madagaszkár (piros terület) mellett sok kisebb sziget alkotja. Forrás: <https://upload.wikimedia.org/>.



A Malagasi hangyafauna jobb feltárásának nagyszabású projektjéhez csatlakoztam 2014-ben egy Schlinger alapítvány²¹ által támogatott lehetőség keretében. Ebben a két évet felölelő projektben a Malagasi *Nesomyrmex* (**23. ábra**) fauna teljes feltárását tűztem ki célul. Később, további feladatként morфомetrikus módszertani fejlesztésekre látszott nagy szükség, valamint a Malagasi kollégák morфомetrikus módszerekre történő képzésén keresztül (amelyet a következő alfejezetben fogok részletesen tárgyalni) még inkább belemélyedtem a helyi fauna feltárásába. A *Nesomyrmex* hangyagénusz korábban négy ismert taxonja mellett további 29 tudományra új fajt írtam le (Csősz & Fisher (2015) *ZooKeys* 526: 19-59, <https://doi.org/10.3897/zookeys.526.6037>; Csősz & Fisher (2016a) *PeerJ* 4:e1796, <https://doi.org/10.7717/peerj.1796>; Csősz & Fisher (2016b) *PLoS ONE* 11(4):

²¹ <https://www.tgci.com/funding-sources/funders/schlinger-foundation>

e0152454, <http://dx.doi.org/10.1371/journal.pone.0152454>; Csősz & Fisher (2016c) *ZooKeys* 603: 105–130, <https://doi.org/10.3897/zookeys.603.8271>; Csősz & Fisher (2016d) *ZooKeys* 616: 125–159. <https://doi.org/10.3897/zookeys.616.9532>), ezzel a Malagasi *Nesomyrmex* fajszám 33-ra emelkedett.

23. ábra. *Nesomyrmex angulatus* Mayr, 1862 dolgozója profilból. (A kép forrása: California Academy of Sciences, antweb.org: CASENT0134948.)



A Malagasi *Nesomyrmex* genusz, az általam elvégzett elővizsgálatok eredményeinek alapján is sokkal fajgazdagabbnak tűnt, mint azt a korabeli ismereteink mutatták. A projekt kezdetén ismeretes négy faj helyett

becsléseim ennek közel tízszeresét mutatták, amely később be is igazolódott. A nagy fajszám megfelelő taxonómiai kezeléséhez külső kvalitatív morfológiai jellegek alapján előbb négy fajcsoportot (*N. angulatus*, *hafahafa*, *madecassus* és *sikorai* fajcsoportok, ld. Csősz & Fisher 2015) állítottam fel, majd végül szükség volt egy ötödik fajcsoport (*brevicornis* fajcsoport, ld. Csősz & Fisher 2016d) elkülönítésére is, melyek egyenként 4-9 fajt tartalmaztak. Ezek a fajcsoportok egymástól kvalitatív jellegek vizsgálatával is jól elkülöníthetőek, és vélhetően monofiletikus egységet alkotnak²².

Madagaszkár és az egész Malagasi régió egyedülállóan gazdag faunája napjaink környezetromboló tevékenysége által erősen veszélyeztetetté vált, így fennáll a veszélye annak, hogy fajok tömegei tűnnek el a színről még mielőtt felfedeznénk őket. Madagaszkár kivételesen nagy fajgazdagsága, a *kriptikus* fajok nagy száma valamint a diverzitás gyors ütemű²³ pusztulása (Ceballos és mtsai. 2015) együttesen komoly kihívásokat támaszt az

²² Szükségesnek tartom megjegyezni, hogy a fajcsoportok taxonómiai munkák során jól bevált módon, morfológiai jelleg-kombinációk alapján történő elkülönítése. Bár a fajcsoportok létrehozása bevett taxonómiai gyakorlat, a taxonómiai munkát nagyban megkönnyíti és ideális esetben származásbeli kapcsolatokról árulkodik, a döntően morfológiai alapon létrehozott fajcsoportok nem szükségszerűen alkotnak monofiletikus egységet.

²³ <http://www.commondreams.org/news/2013/05/28/un-accelerating-biodiversity-loss-fundamental-threat-survival-humankind>

élővilág helyi szeletének teljes megismerése elé. Ezért első sorban olyan taxonómiai eljárásokra van szükség, melyek a nagy pontosság mellett jelentősen meggyorsítják a döntéshozás mechanizmusát, és ezzel jelentős időt és energiát takarítanak meg napjaink taxonómiai számára.

5.4.1 Automatizált taxonómiai döntéshozási eljárások bevezetése a Madagaszkár hangya faunájának taxonómiai feltárásának meggyorsítása céljából

Az elmúlt években számos ígéretes algoritmus került publikálásra, melyek sikerrel alkalmazhatók rovar taxonómiában, komplex morfológiai mintázatok felismerésére (Ezard és mtsai. 2010, Baur & Leuenberger 2011, Derkarabetian & Hedin 2014). Ehhez az irányvonalhoz csatlakozva fektettem le saját morfológiai mintázatfelismerésben jól alkalmazható eljárásomat (Siefert, Ritz & Csősz 2014). Ezen algoritmusok képesek megkönnyíteni a taxonómiai munkát, a morfológiai diverzitás többváltozós stratégiák alapján történő feltárásán keresztül hasznos eszközökként tekinthetünk rájuk, melyekkel a taxonómiai döntéshozást meggyorsíthatjuk. A módszerek nyilvánvaló előnyei mellett azonban van egy olyan hiányosságuk, melynek kiküszöböléséhez a helyi diverzitási mintázatok sajátosságait jól ismerő szakember közreműködését igénylik: szükséges ugyanis, a klaszterek ideális számának definiálása.

Ezek a módszerek, bár nagy segítséget jelentenek az adatsorban rejlő mintázatok felismerésében, a klaszterek ideális számának eldöntését – azaz hány faj van az adatsorban – teljes mértékben a felhasználóra bízák. A fajok számára való következtetés éppen ezért olyan szakember beavatkozását igényli, akinek komoly tapasztalata van a területen. A Palearktikus hangya taxonómiában rendelkezésre állnak komoly referencia munkák, melyek segítségével (széles körben elfogadott taxonok morfológiai variabilitását alapul véve) „kalibrálni” lehet a mintázatban rejlő fajsza-
mót. a trópusi régiók faunájának kalibrálását is megkísérelhetjük hasonló stratégia mentén, azonban ez kockázatos vállalkozás egy olyan fauna esetében, ahol a fajgazdagságnak csak kis szelete ismert, az is többnyire múlt századi munkák alapján.

A saját projektem témájaként választott Malagasi *Nesomyrmex* fauna 2015 előtt mindössze 4 fajt számlált. A feladatom végeztével a génusz teljes régióbeli fajsza-
ma 29 általam leírt fajjal összesen 33-ra emelkedett. Ilyen körülmények között nehéz a meglévő taxonok ismert adataihoz „kalibrálni” a jövőbeni döntésünket. Megoldást jelenthet jól feltárt területek rokon génuszainak morfológiai diverzitására alapozni, de ez is sok éves vagy évtizedes szaktudást feltételez. Madagaszkár faunájának belátható időn belüli feltárása azonban csak helyi, feltörekvő szakemberek bevonásával végezhető el, így számolnunk kell ke-

véssé tapasztalt, fiatal generációhoz tartozó kollégák nagy számának bevonásával és közreműködésével. A szakemberek bevonásának kiterjesztése – a nyilvánvaló előnyökön túl – kockázatokat is magában hordoz.

Ilyen esetek által támasztott kihívásokra adhat megfelelő választ egy olyan, könnyen használható algoritmikus protokoll bevezetése, amivel a *szakértelem iránti igény (a)* és a *szubjektív interpretáció (b)* lehetősége egyaránt minimalizálható.

- a) A szakértelem iránti igény minimalizálása a zömében trópusi régiók helyi erők bevonásával történő taxonómiai feltárásának meggyorsítása miatt lényeges. A trópusi területek feltárásához ugyanis a nagyfokú környezetrombolás okozta gyors lokális kihalás, valamint a még leíratlan taxonok nagy száma miatt kénytelenek vagyunk kompromisszumokat kötni.
- b) A szubjektivitás esélyének csökkentése pedig a vizsgálatok ismételhetősége, tesztelhetősége és a következtetések széles körű összehasonlítása miatt esszenciális. Ezen kritériumok mindegyikét a tudományosság alapelveiként tartjuk számon.

Erre a problémára keresve megoldást – a Malagasi *Nesomyrmex* fauna feltárásán messze túlmutatóan – további módszertani elemek bevezetését tűztem ki célul. Korai elképzeléseim szerint, amennyiben ez a törekvésem sikerrel jár, kevésbé tapasztalt helyi kollégák nagy része válik képessé a helyi fauna mintázataira vonatkozó gyors és megbízható következtetések levonására.

Fő követelményként merült fel, hogy a módszer – a morfológiai adatsorban rejlő mintázat feltárása mellett – képes legyen becslést adni az adatsorban rejlő klaszterek számáról is. Ezzel nem csak az automatizálás javul, hanem a döntéshozás egy szeletének algoritmusos alapokra helyezésével egy további lépést teszünk az ismételhetőség és reprodukálhatóság felé. Egy olyan rendszer, amely a vizsgálótól függetlenül ugyanazt az eredményt hozza, egyúttal összehasonlíthatóvá is teszi a következtetéseket. Itt muszáj megjegyezni, hogy a taxonómiai döntéshozás algoritmusos megoldása, bár gyors és hathatós segítséget kínál a fajok számának eldöntéséhez, egy kiegészítés marad az eszköztárunkban. A kapott mintázat illetve a klaszterek száma szakember által, egyéb adatokból származó bizonyítékok figyelembe vétele után, felülbírálnak.

A feladat tehát adva volt: egy széleskörűen alkalmazható módszertani eljárás bevezetése a Malagasi *Nesomyrmex* fauna feldolgozásán bemutatva. Egy olyan az NC-klaszterezéssel kombinálható kiegészítő statisztikai eljárást kerestem, amely egyrészt képes becsléssel szolgálni a klaszterek számát illetően, másrészt az eseteket partícionálja, azaz csoportokba válogatja. A mintázat struktúrájának megjelenítése itt már nem követelmény, hiszen azt

az NC-klaszterezés által megjelenített dendrogram megfelelően képes értékelhető formában elénk tárni.

A feladat újdonságát jelzi, hogy – legjobb tudomásom szerint – ez volt az első olyan rovar-taxonómiában a gyakorlatban is alkalmazott eljárás, amely képes morfológiai adatsorok alapján gap statisztikai algoritmus segítségével becsléssel szolgálni a klaszterek, azaz fajok számát illetően. A *gap* statisztika megbecsüli azon részhalmazok számát, melyek a csoportokat legjobban jellemzik (**24. ábra**). Az algoritmus előzetes hipotézisalkotást nem igényel, könnyen értelmezhető megoldást kínál a „megfelelő” klaszterszám (K) megtalálására normál eloszlású adatsorban, mint amilyenek a folytonos morfometrikus adatsorok is. A *gap* statisztikát 2001-ben a Stanford egyetem munkatársai (Tibshirani, Walther & Hastie, 2001) publikálták. Azt az alapelvet követték, hogy összehasonlították az adatsorban található diszperziót (W_k) egy Monte Carlo újramintázással nyert ún. referencia eloszlásban tapasztalt diszperzióval, melyben nyilvánvalóan nem találunk elkülönült klasztereket ($K=1$).

Az optimális klaszterszám az eljárásuk szerint ott található, ahol a kapott W_k a legtávolabb esik a referencia görbétől:

$$\text{Gap}_n(k) = E_n^*\{\log W_k\} - \log W_k$$

A referencia adatsor az eredeti adatsorból generálható (ld. zöld keretben, **24. ábra**). A Monte Carlo ismétléseiből generált $\log W_k^*$ tapasztalati szórása ($\text{sd}(k)$) kvantifálható:

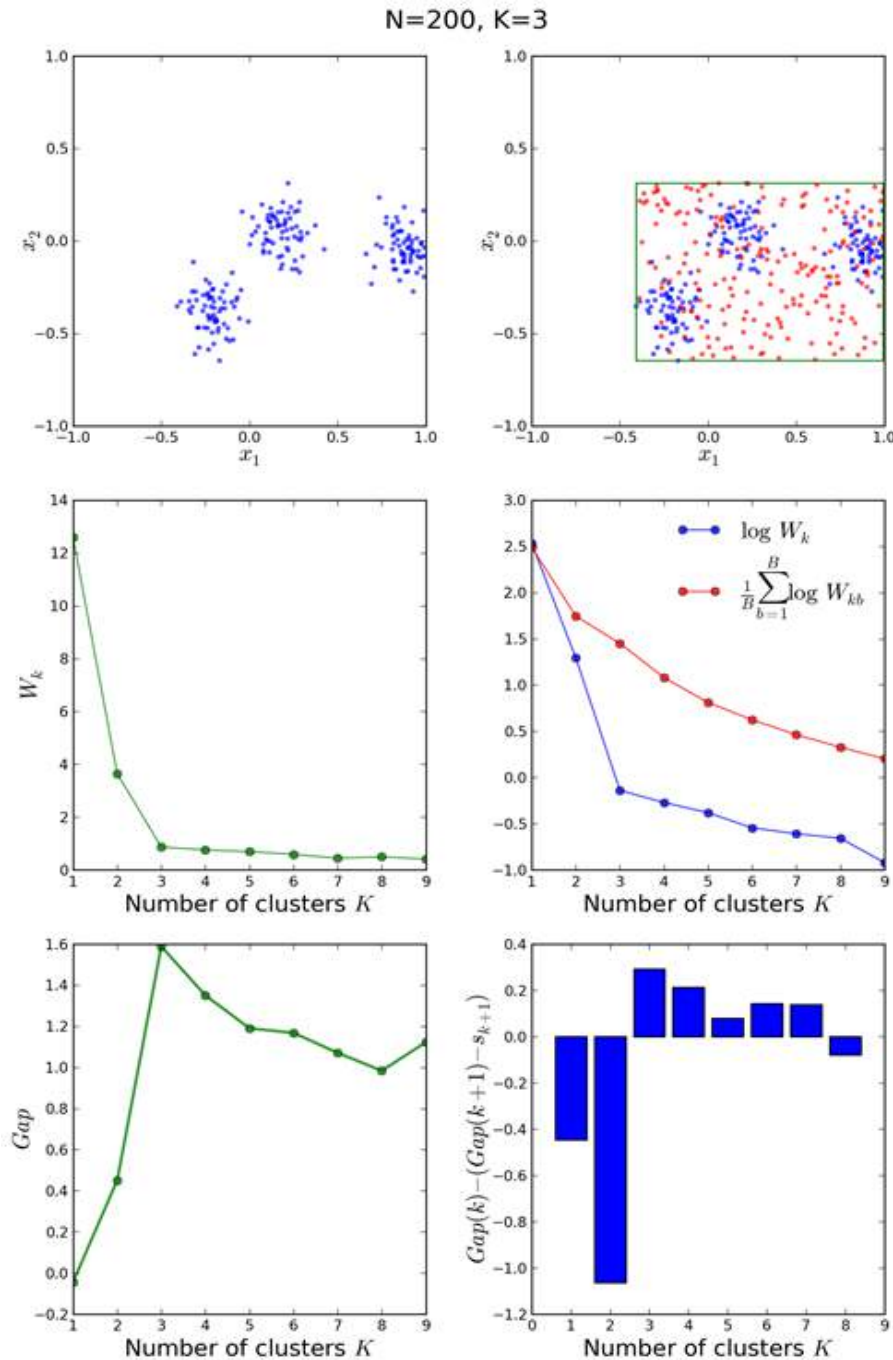
$$s_k = \sqrt{1 + 1/B} \text{sd}(k)$$

Végezetül az optimális klaszterszám (K) az az érték, ahol a kapott W_k a legtávolabb esik a referencia görbétől, és a szórás maximuma nem haladja meg az előtte lévő érték átlagát.

$$\text{Gap}(k) \geq \text{Gap}(k+1) - s_{k+1}$$

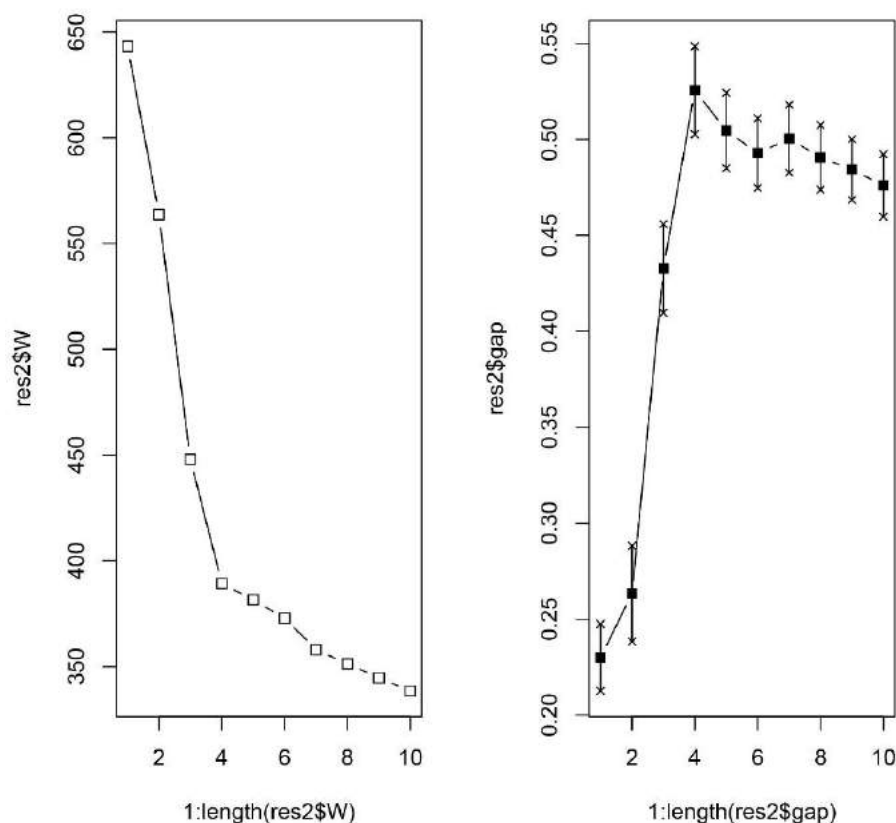
A **24. ábra** egy három klasztert tartalmazó adatsor *gap* statisztikai elemzésének lépéseit mutatja be sematikus²⁴.

²⁴ A *gap* statisztika rövid leírása, a képletek és a **24. ábra** forrása: „The Data Science Lab” (<https://datasci-encelab.wordpress.com/2013/12/27/finding-the-k-in-k-means-clustering/>)



24. ábra. A gap statisztika működésének vázlata. A bal felső doboz a 3 klaszteres valós állapotot mutatja. A jobb oldali doboz (zölddel körbekerítve) mutatja a Monte Carlo újramintázással készített ún. referencia eloszlást. Balra középen az eredeti adatsor diszperziójáról (W_k) készített hagyományos grafikon látható, amelynek segítségével meg tudjuk becsülni a klaszterek számát. Esetünkben a “könyök” hajlása 3 klaszter meglétére utal ($K = 3$). A gap statisztika azonban haladóbb a könyök-módszernél. Középen, jobb oldalon a gap statisztika módszerével készült csoportszám-becslést láthatjuk. Ahol a legnagyobb különbséget mérhetjük az eredeti (kék görbe) és a referencia eloszlás görbéje (piros görbe) között, ott határozzuk meg a klaszterek ideális számát. Az alsó két doboz a különbségek mértékét mutatja. Az optimális klaszter szám esetünkben $K = 3$.

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.



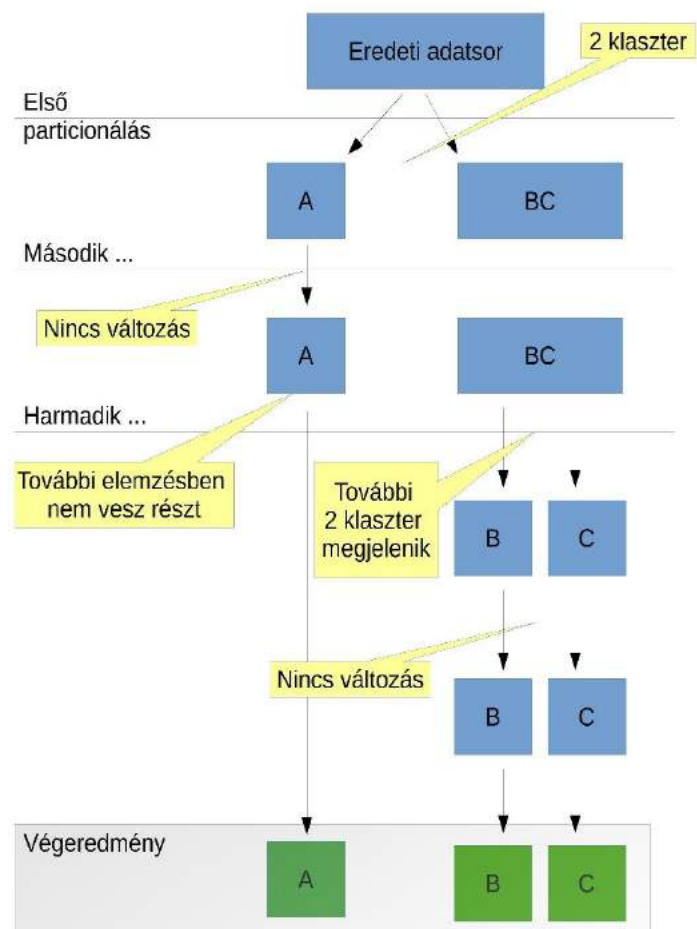
25. ábra. A *Nesomyrmex hafahafa* fajcsoport adatsorán végzett *gap* statisztika. A diszperziós görbe (bal oldali ábra) négy összetevőnél való megtörése (könyök), valamint a *gap*-görbe (jobb oldali ábra) csúcsa együttesen támogatják a 4 klaszteres megoldást. Az adatokban lévő klaszterek száma az X tengelyen látható.

A *gap* statisztikai algoritmus egy adatsorra történő egyszeri futtatása kis klaszterszám mellett sikeresen határozza meg az adatsorban rejlő részalmazok számát. A **25. ábra** a *Nesomyrmex hafahafa* fajcsoport négy fajt tartalmazó adatsorban rejlő optimális klaszterszámot mutatja.

Nagyobb adatsorok vagy fajgazdagabb fajcsoportok elemzésénél azonban a meghatározott részalmazokban további, rejtett klaszterek maradhatnak vissza, melyek a specifikusan a jelzett klaszterekre irányuló, ismételt *gap* statisztikai elemzésekkel bonthatók ki. Amennyiben teljes partícionálásra, valamint optimálisan meghatározott klaszter számra

van szükségünk, akkor a *gap* statisztika alapelveivel operáló, de lépésenként újraelemző algoritmusra van szükség. A *clusterGenomics* csomag (Nilsen & Lingjaerde 2013) tartalmaz egy pontosan erre a célra megalkotott algoritmust *Partitioning Based on Recursive Thresholding – PART* néven, amely nagy adatsorok elemzésére (például genetikai minták elemzésére) lett kidolgozva. Munkám során arra összpontosítottam, hogy a PART algoritmus működését összehangoljam a már széles körben kipróbált NC-klaszterezéssel, valamint a meglévő palearktikus és az újonnan gyűjtött Malagasi génuszok adatsorain a két módszer együttes működését részletesen leteszteljem.

Amennyiben a PART módszer a *gap* statisztika kritériumai szerint egynél több klasztert talál az adatsorban, akkor minden egyes részhalmozra optimalizálva egy új elemzés fut le (26. ábra) egészen addig, amíg el nem érünk egy határértéket, vagy a klaszter méret a megadott »minimum méret« (*minSize*) kritérium alatt marad, melyet mi a scriptben magunk is beállíthatunk. A »minimum méret« kritérium alapbeállítása *minSize* = 5, de a saját eredményeim alapján szükség szerint *minSize* = 3-ra leszállítható.



26. ábra. A PART módszer működésének sematikus ábrázolása. Első lépés: az eredeti adatsorban két klasztert ismert fel a *gap* statisztika („A” és „BC”, kék színnel). A második lépés: amennyiben az „A” klaszter nem bontható további elemeire, a későbbi lépésekben egységes klaszterként kezelendő. A harmadik lépésben a „BC” klaszter további két összetevőre bontható. Amennyiben az így kapott 3 klaszter egyike sem bontható tovább, nem történik további lépés. Végeredményben összesen 3 klasztert találunk („A”, „B” és „C” klaszterek zöld színnel) az adatsorban a rekurzív particionálás segítségével.

A *clusterGenomics* csomag továbbá lehetőséget ad két tetszőleges *bootstrap* ismétlésszámmal futtatott klaszter-módszer, a »kmeans« és a »hclust« partícionálásának kinyerésére majd összehasonlítására. A *bootstrap* ismétlések beállítását a felhasználó végzi. A több ismétléses beállítások robusztusabb eredményt adnak, de nagyobb adatsoroknál (> 500 minta) figyelembe kell venni a drámaian megnövekedett számítási időt. Saját tapasztalataimra alapozva úgy gondolom, az 500 vagy 1000 ismétléses beállítás megfelelő kompromiszsumnak tekinthető. Nagyobb adatsorok elemzése így is órákat vesz igénybe.

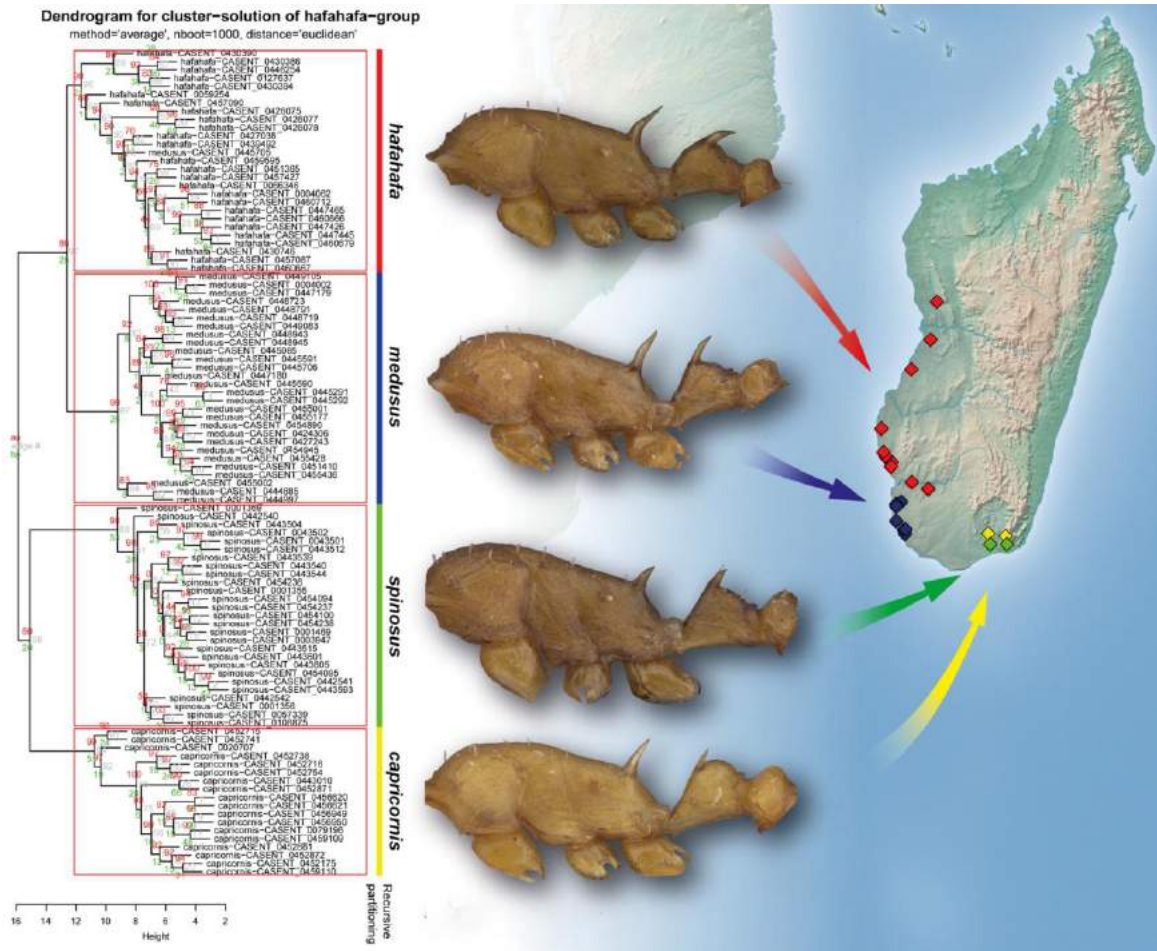
A partícionálási eredmények ezután a könnyű értékelhetőség érdekében a dendrogramra illesztve ábrázolhatók »mark.dendrogram« parancssorral (Beleites & Sergio 2015). A script R-ben íródott, és rangos folyóiratokban (Csósz & Fisher (2016a) PeerJ, 4, e1796. <https://peerj.com/articles/1796/>; Csósz & Fisher (2016b) PLoS ONE 11(4): e0152454. <http://dx.doi.org/10.1371/journal.pone.0152454>) publikáltam, valamint a jelen dolgozat **9.5. mellékletében** is megtalálható.

Ez az eljárás a NC-klaszterezés eredményét, a lineáris diszkrimináns mátrixot használja fel bemeneti adatként, ezért az eljárásnak NC-PART klaszterezés nevet adtam. Az NC-klaszterezés *gap* statisztikai algoritmussal való kombinációja egy nagymértékben automatizált feltáró jellegű adatelemzést eredményez, amely a fajok számára valamint az esetek csoportokba rendezésére statisztikai alapú döntéssel szolgál.

5.4.2 A munka aktualitása

Már esett szó a Madagaszkár kiemelkedően változatos és endémikus elemektől gazdag élővilágáról, amely a bolygónk biodiverzitásának egyik kiemelt pontja (Myers et al., 2000). Madagaszkár (és az egész Malagasi régió, beleértve Madagaszkárt és a környező szigeteket) egyedülálló és gazdag faunája azonban napjaink környezetromboló tevékenysége – ideértve az élőhelyek elpusztításától kezdve, a környezetszennyezésen át az invazív fajok behurcolását – által erősen veszélyeztetetté vált, és gyors ütemű pusztulásnak indult. Ezért fennáll a veszélye annak, hogy fajok tömegei tűnnek el a színről még azelőtt, mielőtt felfedeznénk őket.

A munka legfőbb aktualitását a fent vázolt, gyorsan hanyatló bióta megismerésének igénye adja. A természet védelmének egyik kulcseleme, hogy tisztában legyünk azzal, hogy mit is védünk, mivel állunk szemben. Ezen információk birtokában vagyunk csak képesek egy hathatós stratégiát kidolgozni, amellyel képesek lehetünk megállítani, vagy ideális esetben visszafordítani a negatív trendeket. Az első és legfontosabb kérdés tehát az, hogy »*hány fajunk van a kérdéses területen?*«. Ebben az adatgyűjtési munkában elsődleges szerep hárul a terület taxonómiai feltárására.



27. ábra. A *Nesomyrmex hafahafa* fajcsoport NC-PART-klaszterezéssel kapott dendrogramja. A fajcsoport négy újonnan leírt fajának mesosoma profilja ábrázolva a Madagasz-kári elterjedési adatokkal. A dendrogram jobb oldalán a fészekminták PART algoritmus 'hclust' klaszter módszerrel végzett partícionálásának eredménye látható különböző színekkel jelölve. Színkódok: *Nesomyrmex capricornis* (sárga), *N. hafahafa* (piros), *N. medusus* (kék), *N. spinosus* (zöld). A dendrogramon feltüntetett AU/BP (%) értékek magyarázata: AU=P-érték, BP=statisztikai igazítás előtti bootstrap valószínűség. Piros keretes négyszögekbe foglalva a végső fajhipotézis során egy fajba tartozó esetek láthatók. Csász & Fisher (2015) nyomán.

A *Nesomyrmex* genusz kiválasztásában a következő szempontok játszottak döntő szerepet:

- A génusz még soha nem volt modern eszközökkel feltárva.
- Előzetes becsléseim alapján kellően fajgazdagnak látszott ahhoz, hogy az ökoszisztémában jelentős nyomot hagyjon.
- A génusz populációiból megfelelő mennyiségű minta áll rendelkezésre morfológiai vizsgálatok céljára, amely elegendő egy robusztus mintázat felismeréséhez és teszteléséhez.

- d) A biodiverzitás kutatási törekvéseken túl, az 5.2.3.1. fejezetben leírt szempontrendszer alapján, a *Nesomyrmex* génusz módszertani alkalmazások tesztelésére és bemutatására is tökéletes tesztelési területként jöhet számításba.
- e) A *Nesomyrmex* génusz fajai monomorfok, azaz egy fajt vagy populációt alkotó példányok méret vagy alaki szórása nagyon kicsi. A saját meghatározásom szerint egy faj monomorf, ha a populáció legnagyobb példány testhossza kevesebb mint másfélszer akkora, mint ugyanazon populáció legkisebb egyedének hossza.
- f) Sok az ún. *kriptikus faj*. A fajok közötti morfológiai különbségek nagyon kicsik, a különböző populációk morfológiai karakterei egymással széles skálán átfednek, ami drasztikusan megnehezíti a hagyományos taxonómiai módszerekkel való munkát. A *kriptikus fajok* nagy aránya segít kiemelni a módszer nagyfokú megbízhatóságát.

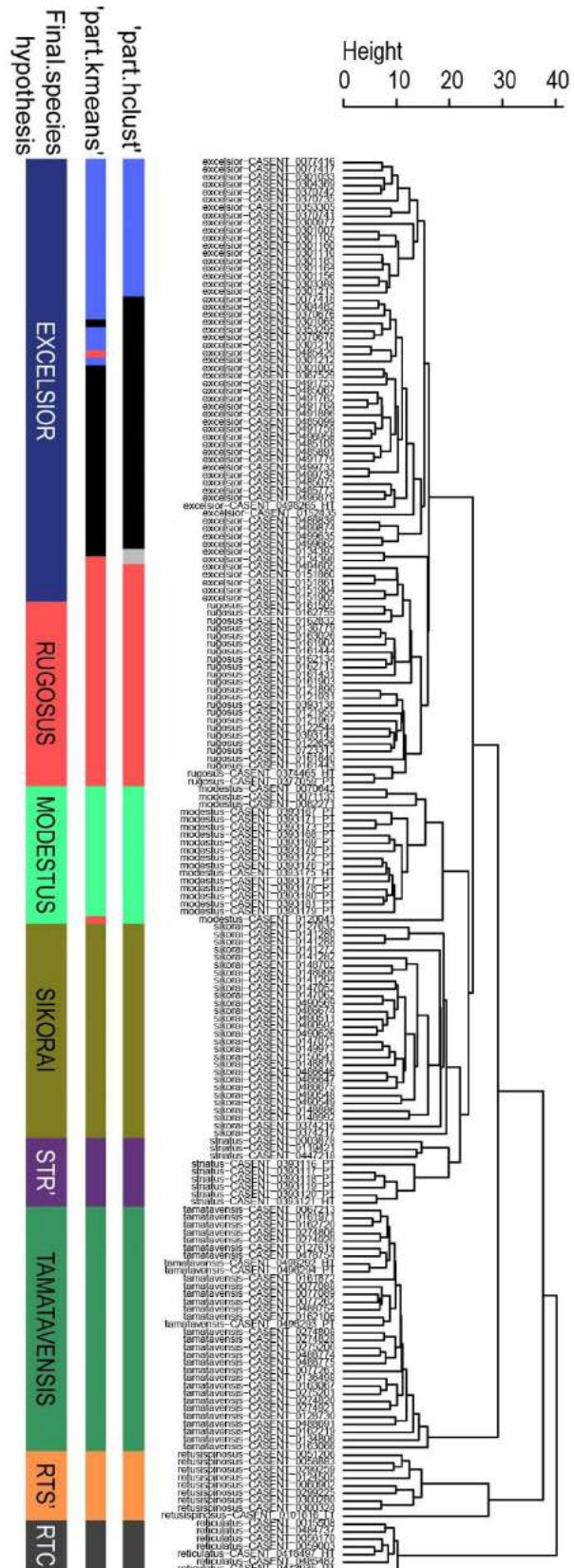
A munkának további aktualitást kölcsönöz – és egyúttal kiemeli a helyi fauna gyors ütemű pusztulása miatt a taxonómiai munkák kiemelt fontosságát – az a tény, hogy két általam leírt faj (*N. capricornis* és *N. spinosus*, ld. Csősz & Fisher 2015) populációit, melyeket 2002-ben nagyon szűk elterjedési területről gyűjtöttek (**27. ábra**), a 2015-ös expedíció során már célzott keresés ellenére sem találták meg. Nem elképzelhetetlen, hogy a két említett faj időközben a nagyszabású környezetpusztító tevékenységnek szinte szemünk láttára esett áldozatul, és leírásukra az utolsó utáni pillanatban került sor. Ezt a feltevést látszik alátámasztani az említett populációkban tapasztalt nagyfokú morfológiai aszimmetria, amely súlyos beltenyészetre és a genetikai állomány elszegényedésére utalhat.

5.4.3 Anyag és módszer

A Madagaszkár és környező szigetek faunájának feltárásában nagy szerep jut a *California Academy of Sciences* (<http://www.calacademy.org/explore-science/madagascar/>) és a *Madagascar Biodiversity Centre* (<http://www.madagascarbio.org/>) közös kezdeményezésének. Ebből az együttműködésből – több mint egy évtized munkájának eredményeként – hatalmas mennyiségű, reprezentatív anyag halmozódott fel a *California Academy of Sciences* gyűjteményében, amely alapvető fontosságú egy ekkora terv végrehajtásához. A vizsgált anyag döntő része a *California Academy of Sciences* (San Francisco, USA) rovargyűjteményében található. Kisebb részben vizsgáltam a *Madagascar Biodiversity Centre* (Antananarivo, Madagascar) anyagát, a *Museum of Comparative Zoology* (Cambridge, USA) gyűjteményét, valamint sok kisebb, de nem kevésbé fontos múzeum (*Naturhistorisches Museum*, Bázeli, Svájc, *Naturhistorisches Museum* Bécsi, Ausztria, *Muséum d'Histoire Naturelle*, Genfi, Svájc) típus anyagát kölcsönöztem összehasonlító vizsgálat céljából.

28. ábra. A *Nesomyrmex sikorai* fajcsoport NC-PART klaszterezésének dendrogramja. A két keskenyebb oszlop a PART algoritmus »kmeans« és »hclust« klaszter módszere által megadott faji osztályozást és fajhatárokat mutatja. Az eltérő módon besorolt minták wild-cardként kerültek a megerősítő szakaszba (bővebben a szövegben). A végleges morfológiai fajhipotézis oszlop mutatja a megerősítő elemzési fázis utáni taxonómiai döntést.

Színkódok: Nesomyrmex excelsior: sötét-kék, N. modestus: világoszöld, N. reticulatus (RTC') szürke, N. retusispinosus (RTS'): narancs, N. rugosus: piros, N. sikorai: keki, N. striatus (STR'): lila, N. tamatavensis: sötétzöld. A N. excelsior (a végleges morfológiai fajhipotézis oszlopban sötétkék) az exploratory elemzések egyaránt két részre bontották (világoskék és fekete) azonban a fajhatárookban jelentkező nagyfokú eltérés, valamint a faji szintű elkülönülést támogató egyéb biológiai információ hiánya miatt ezeket összevettem. Csősz & Fisher (2016b) PLoS ONE 11(4): e0152454 nyomán.



5.4.3.1 A morfometrikus karakterek mérésének protokollja

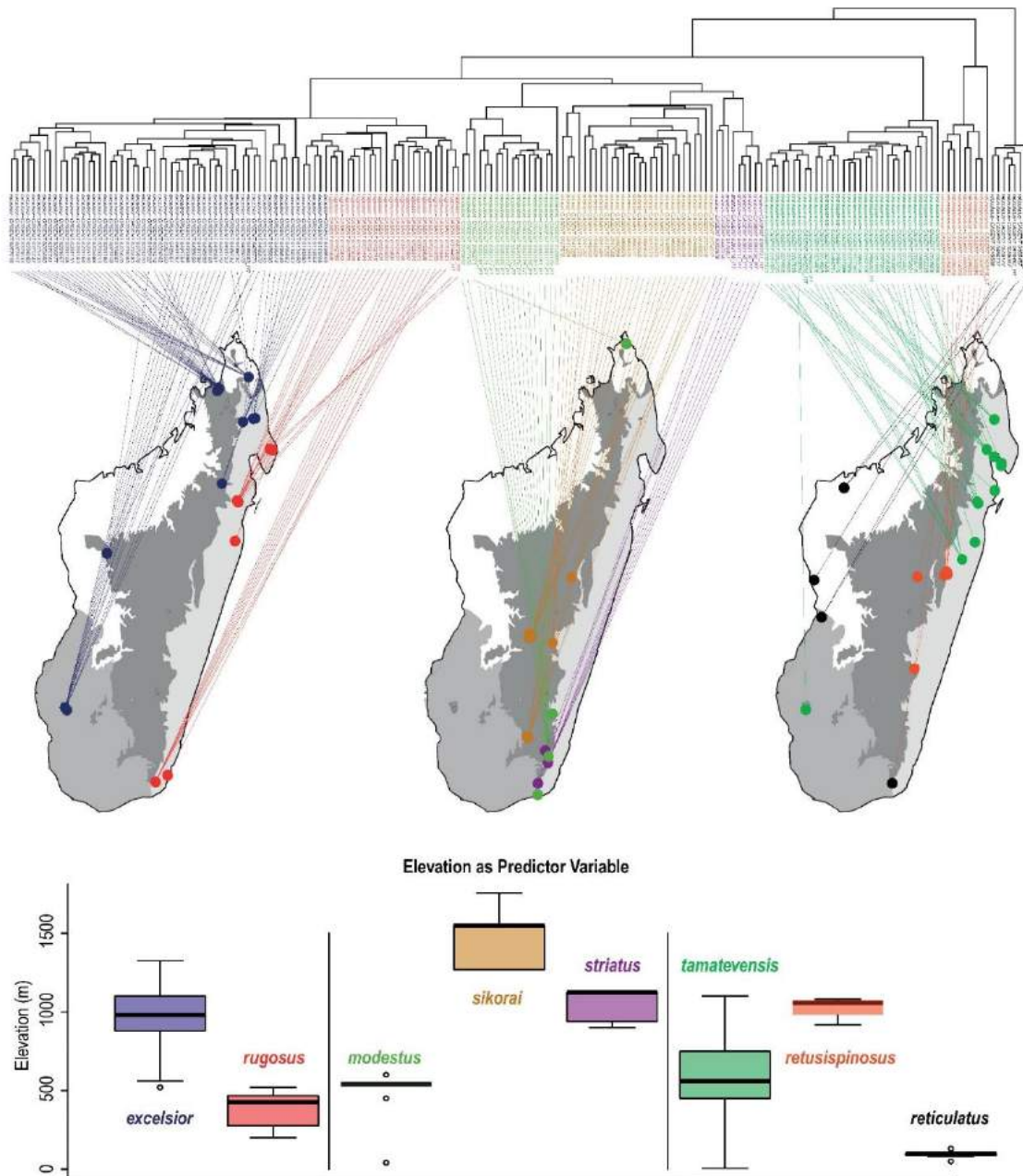
A morfológiai mintázat elemzéséhez 463 fészekminta 1160 dolgozóját mértem le. Fajcsoportonként változó, 19-22 morfometrikus jelleget vettem fel és elemeztem, melynek oka a fajcsoportonként némileg eltérő morfológia, amely egyes esetekben nem tette lehetővé bizonyos karakterek szabatos mérését minden fajcsoportban. A karakterek pontos definícióját lásd a **9.6. mellékletben**. Minden mérést magam végeztem. A hangyák ivaros egyedeinek (királynők és hímek) viszonylagos kis száma miatt a morfometrikus vizsgálatokat csakis dolgozókon végeztem. Minden mérést Leica MZ 12.5 sztereomikroszkóppal, x100 nagyításon, mikrométeres skálán végeztem. A mérések átlagos pontosságát (error rate) $\pm 2 \mu\text{m}$ -re becsültem. A minták előre-hátra és oldalirányú finom mozgatható tárgyasztalon rögzítését forgatható tárgytartó állvány biztosította.

5.4.3.2 Statisztikai feldolgozás

Ennek az egyedülálló és nagyrészt elhanyagolt faunának a feltárását az NC-klaszterezésre (Seifert, Ritz & **Csősz**, 2014) épülő, de PART (*Partitioning Algorithm based on Recursive Thresholding*) algoritmussal (Nilsen & Lingjaerde, 2013) kiegészített módszerrel végeztem. Az eljárást azzal a céllal állítottam össze, hogy a morfológiai adatsorban rejlő mintázat dendrogramon történő ábrázoláson túl képes legyen megbecsülni a klaszterek számát, másrészt a minták csoportokba válogatásával segítse a taxonómiai döntéshozást és javítsa az értékelés objektivitását.

Az NC-klaszterezés bizonyítottan jól használható nagy és összetett adatsorokban rejlő mintázatok felismerésére (**Csősz** et al., 2014; Guillem, Drijfhout & Martin, 2014; Wachter et al., 2015) a PART algoritmus pedig statisztikai küszöbértékek alapján objektíven képes kijelölni a klasztereket (Nilsen et al., 2013; Tibshirani, Walther & Hastie, 2001). A partícionálás eredményét »mark.dendrogram« parancssorral (Beleites & Sergio 2015 alapján) színes sávokkal jelöltem meg a dendrogram bal oldalán (ld. **28. ábra**).

Két klaszter módszert (»kmeans« és »hclust«) használtam a PART algoritmus klaszter elhatárolásaihoz, illetve az optimális klaszter szám becsléséhez. Ezeket külön ábrázolom a kombinált dendrogramon (**28. ábra**). A felismert klaszterek csak abban az esetben tekinthetők biológiai tartalommal rendelkező egységeknek (azaz fajoknak, egyaránt ideértve a korábban ismert taxonokat, valamint az új fajokat is) ha a klaszter módszerek egyezést mutatnak. Eltérő hipotézis alkotás esetén, azaz mikor egyik vagy másik klaszter módszer alternatív eredményt mutat, az adott minták nem lesznek részesei az elsődleges csoportosításnak, wild-card-ként kerülnek a megerősítő tesztelési lépés bemenő adatsorába (részleteket lásd még az **5.2.4.** és az **5.2.6. alfejezetekben**).

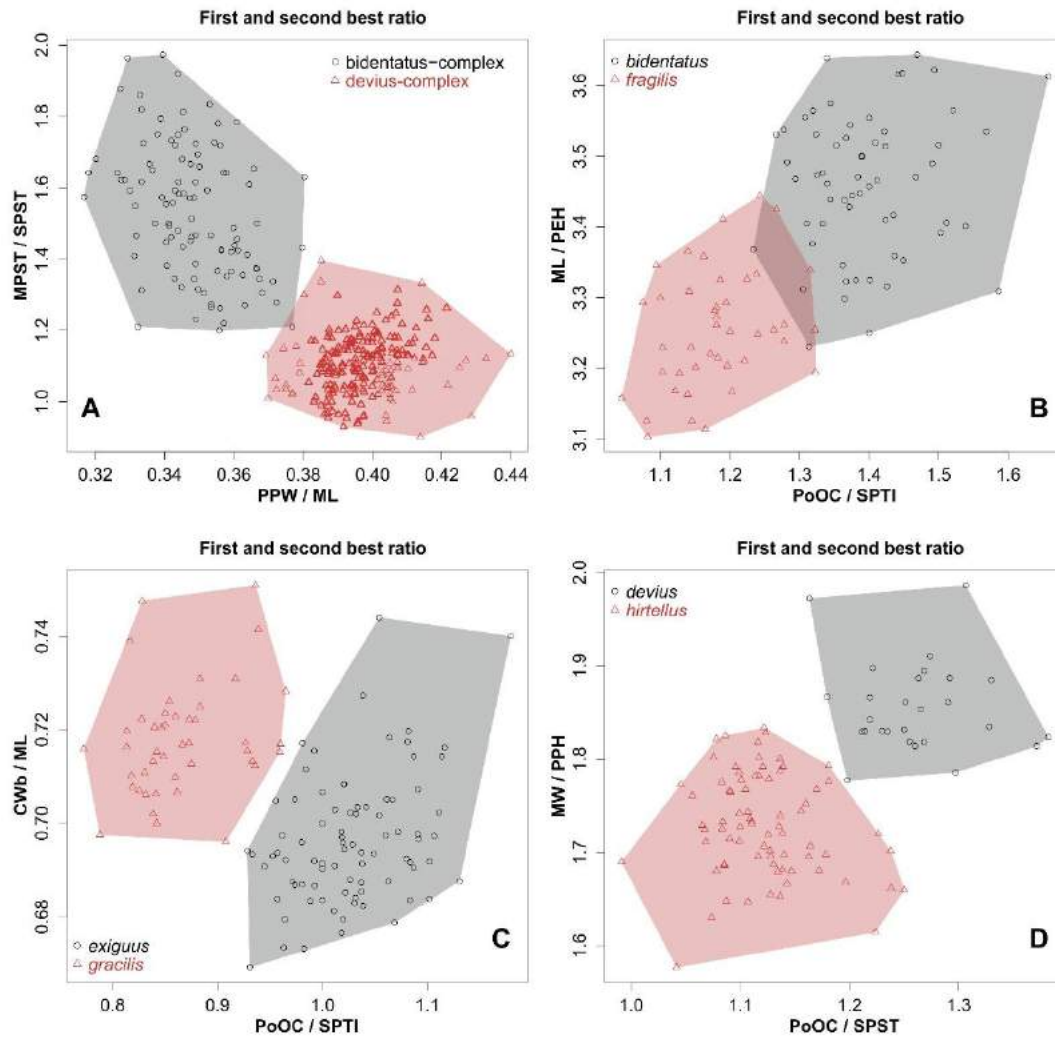


29. ábra. Az NC-klaszterezés eredményeinek megjelenítése Madagaszkár geográfiai térképén. A fajok elterjedése követi az ökológiai régiók kiterjedését (Schatz 2000, Cornet 1974): keleti esőerdő (világosszürke), centrális hegyi erdőségek (sötét szürke), nyugati száraz erdők (fehér), és dél-nyugati sivatagi cserjés (közép szürke). Színkódok: *Nesomyrmex excelsior* sp. n.: sötétkék, *N. modestus* sp. n.: világoszöld, *N. reticulatus* (RTC'): szürke, *N. retusispinosus* (RTS'): narancs, *N. rugosus* sp. n.: vörös, *N. sikorai*: keki, *N. striatus* (STR'): lilla, *N. tamatevensis* sp. n.: sötétzöld. Az ábra színezését utólag végeztem el. Csősz & Fisher (2016b) PLoS ONE 11(4): e0152454 nyomán.

A fauna általunk vizsgált szegmensének NC-klaszterezéssel dendrogramon megjelenített morfológiai mintázata megmutatja nekünk, hol vannak a nagyobb rések a homogénebb klaszterek között, azaz képet alkotunk arról, hogy vannak reprodukív akadályok által okozott morfológiai diszkontinuitások az adatsorban. Természetesen nem csak reprodukív akadályok okozhatnak fenotípusbeli változásokat. Azok sok más okból is kialakulhatnak, például környezeti hatásokra adott válaszként polimorfizmus (polifénizmus) okán. Esetleg paraziták szintűgy felelősek lehetnek stabilan eltérő fenotípusok kialakulásáért. 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. Éppen ezért lehet nagyon fontos a morfológiai mintázatot rávetíteni az elterjedési térképre (**29. ábra**), például az általam alkalmazott »phylo.to.map« parancssor segítségével, amely a »phytools« csomagban R környezetben elérhető (Revell, 2012).

Ezzel a módszerrel könnyen lehet következtetni a faji szintű mintázatok geográfiai eredetére, valamint egyes minták hibás osztályozására is gyorsan fény derülhet, amennyiben a fajok elterjedési területein lényegesen kívül eső mintával van dolgunk. Eme utóbbi esetekben a hibás morfológiai klasszifikáción túl megoldásként szóba jöhet még a hibás cédulázás, vagy az ember általi terjesztés lehetősége is. A módszer nagyszerű lehetőséget ad az új ismeretek már meglévő biogeográfiai modellekbe való beillesztésébe, azaz, hogy az újonnan felfedezett fajok elterjedési határai valóban egybeesnek-e a már ismert, több faj adataiból rekonstruált biogeográfiai mintázatokkal.

A morfometrikus elemzések által kialakított fajhipotéziseket (a felismert morfológiai mintázat érvényességét) megerősítő LDA-val teszteltem. Az így kialakított – és kvalitatív morfológiai jegyekkel megerősített – csoportosítás alkotta a *végleges morfológiai fajhipotézist*. A végső morfológiai fajhipotézis elemeinek, a fajok minél stabilabb elkülönítéséhez *Multivariate Ratio Extractor*-t, rövidítve *MRA*-t (Baur & Leuenberger, 2011) használtam, amely képes megtalálni két faj (vagy taxonok általunk definiált csoportjai) elkülönítéséhez legmegbízhatóbban használható aránypárt (**30. ábra**). Ezzel az eljárással a felhasználók kezébe olyan morfometrikus kulcsot vagyunk képesek adni, amely nagy valószínűséggel vezet sikeres meghatározáshoz, mindössze kettő vagy négy morfometrikus karakter alkalmazásával. Ez nagy segítséget jelenthet olyan felhasználók számára, akik a kérdéses csoport taxonómiájával, vagy a csoportra jellemző kvalitatív morfológiai jellegek változatosságával nincsenek kellően tisztában. Számukra egy kvantitatív aránypárral történő határozás – bár legtöbbször időigényesebb, mint a fajok hagyományos módszerrel történő elkülönítése – nagyobb biztonságot nyújt és az így a továbbiakban, az testarányok segítségével meghatározott anyag megfelelő alapot ad a csoport morfológiájának pontosabb megismeréséhez.



30. ábra. Az első és második legjobb morfológiai aránypárok szórásdiagrammjai a *Nesomyrmex angulatus* fajcsoport kriptikus fajtái és fajkomplexei között.

A: A két legjobb morfológiai karakter kombináció a *N. bidentatus* fajkomplex és a *N. devius* fajkomplexei között. B: A *N. bidentatus* fajkomplex két faja, a *N. bidentatus* és a *fragilis* elkülönítése két aránypár segítségével 97%-os sikert eredményez. A *devius* fajkomplex két kriptikus fajpárjának elkülönítése, a *N. exiguus* és a *N. gracilis* (C), valamint a *N. devius* és a *N. hirtellus* (D) esetében az elkülönítés teljes mértékű. Csősz & Fisher (2016a) PeerJ 4:e1796 nyomán.

5.4.4 Eredmények és konklúzió

5.4.4.1 A Malagasi *Nesomyrmex* fauna diverzitása

A Malagasi *Nesomyrmex* fauna feltárására irányuló projektem során elért eredmények aláhúzzák a Madagaszkár és a környező szigetvilág nagyfokú fajgazdagságáról alkotott ismereteinket. Döntően a morfometrikus adatokban rejlő mintázatok felismerésére építő – saját magam által összeállított – eljárás komoly segítséget jelentett a madagaszkári *Nesomyrmex* fauna jobb megismerésében: a korábban ismert 4 faj további 29 tudományra új fajjal gazdagodott, így 33-ra növelve a régióban ismert *Nesomyrmex* fajok számát. A fajsza ilyen nagymértékű hirtelen emelkedése nem egyedi a forró égővi területek kutatásának történetében. A korábbi munkák egy-egy kiragadott gyűjtési esemény jellegzetes fajait dolgozzák fel, így értelemszerűen nem adhatnak átfogó képet a regionális fajgazdagságról. Az újabban egyre erősödő gyűjtési aktivitás, valamint az újonnan bevezetett morfometrikus eljárás együttesen járultak hozzá a régió fajgazdagságának feltárásához.

A nagy mennyiségű, modern protokollok alapján gyűjtött minta kellő mértékben lefedi a régió területeit, az új, kvantitatív morfológiai megközelítéssel működő módszertan pedig a sok ún. *kriptikus faj* gyors és pontos felismeréséhez nyújt hathatós segítséget. A régió *Nesomyrmex* faunáját külső morfológiai jellegek kombinációi alapján először négy (Csősz & Fisher 2015), majd később öt (Csősz & Fisher 2016d) csoportra bontottam. Az ezeket alkotó összesen a már említett 29 tudományra új fajból 14 bizonyult ún. *kriptikus faj*nak, azaz a fajhatárok felismeréséhez többváltozós morfometrikus módszertant kellett használnom (31. ábra). A Malagasi *Nesomyrmex* fauna feltárására irányuló projektem (2014-2016) keretében feldolgozott fajok száma részletezve (teljes fajsza, új fajok száma és a *kriptikus faj*ok száma) az 5. táblázatban látható.

5.4.4.2 *Kriptikus Nesomyrmex* fajok aránya a régióban

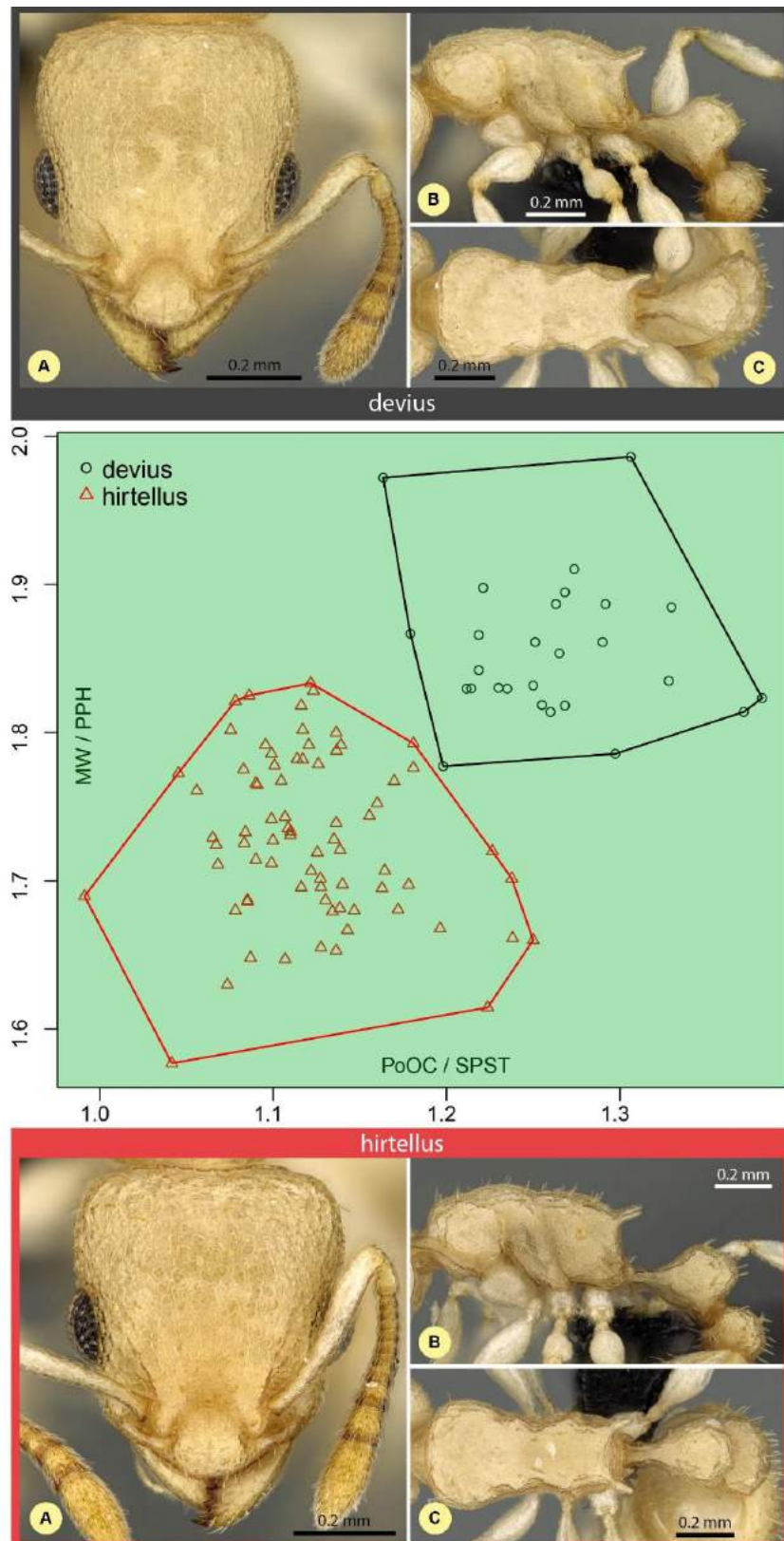
A jelen ismereteink szerint a *kriptikus* hangyafajok száma világviszonylatban megközelíti az 50%-ot, példaként a palearktikus szinten is jól kutatott *Lasius* fajok 46%-a (összesen 97 faj), a palearktikus *Formica* fajok 43%-a (67 faj) és a világon ismert 77 *Cardiocondyla* faj 52%-a bizonyult *kriptikus faj*nak (Seifert 2009). A régióban kimutatott *kriptikus Nesomyrmex* fajok aránya (5. táblázat) nem tér el a világ legjobban kutatott régiójában, a Palearktiszban vizsgált hangya genuszoknál tapasztalható aránytól. Mindez arra enged következtetni, hogy – ha elfogadjuk a fenti állítást, miszerint a forró égővi és mérsékelt övi hangya társulások *kriptikus faj*ainak aránya egymástól nem tér el lényeges mértékben²⁵ – a terület *Nesomyrmex* faunája a jelen ismereteink szerint közel teljesen fel lett tárva.

²⁵ A jelenlegi ismereteink alapján nincs okunk azt gondolni, hogy a trópusi és mérsékelt övi hangyák radiációja eltérő számú *kriptikus faj* képződéséhez vezetne.

Fajcsoport	Fajszám	Tudományra új fajok száma	Kriptikus fajok	Referencia
<i>angulatus</i>	8	7	6	[1]
<i>brevicornis</i> *	9	9	0	[2]
<i>hafahafa</i>	4	4	0	[3]
<i>madecassus</i>	4	2	3	[4]
<i>sikorai</i>	8	7	5	[5]
Összesen	33	29	14	

5. táblázat. A 2014 és 2016 között lebonyolított Malagasi Nesomyrmex taxonómiai projekt keretében feldolgozott fajok száma. A kriptikus Malagasi Nesomyrmex fajok aránya 42%, ami nem tér el lényegesen a világ legjobban kutatott régiójában, a Palearktiszban vizsgált hangya genuszoknál tapasztalható aránytól. A „Fajcsoport” oszlopban a publikációimban körülhatárolt aktuális fajcsoport megnevezése, a „Fajszám” oszlopban a teljes kezelt faj mennyiség (ideértve a már ismert és újonnan leírt fajok számát), a „Tudományra új fajok száma” oszlopban az aktuális munkákban leírt fajok száma, a „Kriptikus fajok” oszlopban az általam csak morfológiai módszerekkel elkülöníthető fajszámot mutatja. A „Referencia” oszlop a cikkek rövid referenciáját tartalmazza a DOI linkekkel. A referenciák sorszáma a következő munkákra utal: [1]: Csősz & Fisher (2016a) PeerJ 4:e1796, <https://doi.org/10.7717/peerj.1796>; [2]: Csősz & Fisher (2016d) ZooKeys 616: 125–159. <https://doi.org/10.3897/zookeys.616.9532>; [3]: Csősz & Fisher (2015) ZooKeys 526: 19–59, <https://doi.org/10.3897/zookeys.526.6037>; [4]: Csősz & Fisher (2016c) ZooKeys 603: 105–130, <https://doi.org/10.3897/zookeys.603.8271>; [5]: Csősz & Fisher (2016b) PLoS ONE 11(4): e0152454, <http://dx.doi.org/10.1371/journal.pone.0152454>. A csillaggal jelzett *N. brevicornis* fajcsoport taxonómiai feldolgozásához, a kis mintaszám miatt, csak hagyományos módszertant használtam, ezért itt kriptikus fajok definíció szerint nem bukkantak fel.

Külön megjegyzést érdemel a *N. brevicornis* fajcsoport feldolgozása (Csősz & Fisher (2016d) ZooKeys 616: 125–159. <https://doi.org/10.3897/zookeys.616.9532>), amelynek populációiból a sok éves, koncentrált gyűjtési aktivitás ellenére kevés minta állt rendelkezésre. A kevés gyűjtött anyag mögötti okokat csak találgatni lehet, de a csoport fajainak a valódi ritkasága mellett a fajok életvitele (éjszakai aktivitás, rejtett életmód stb.) is szóba jöhet, mint egy lehetséges magyarázat.



31. ábra. Kriptikus fajok. Két, egymástól nehezen elkülöníthető faj, a *N. devius* (felül) és *N. hirtellus* (alul) látható. A dolgozókon nem találunk hagyományos morfológiai karaktert a két faj elkülönítéséhez. Mégis, két morfometrikus aránypár kombinációjával (középső zöld tábla) a két faj könnyedén és teljes sikerrel szétválasztható.

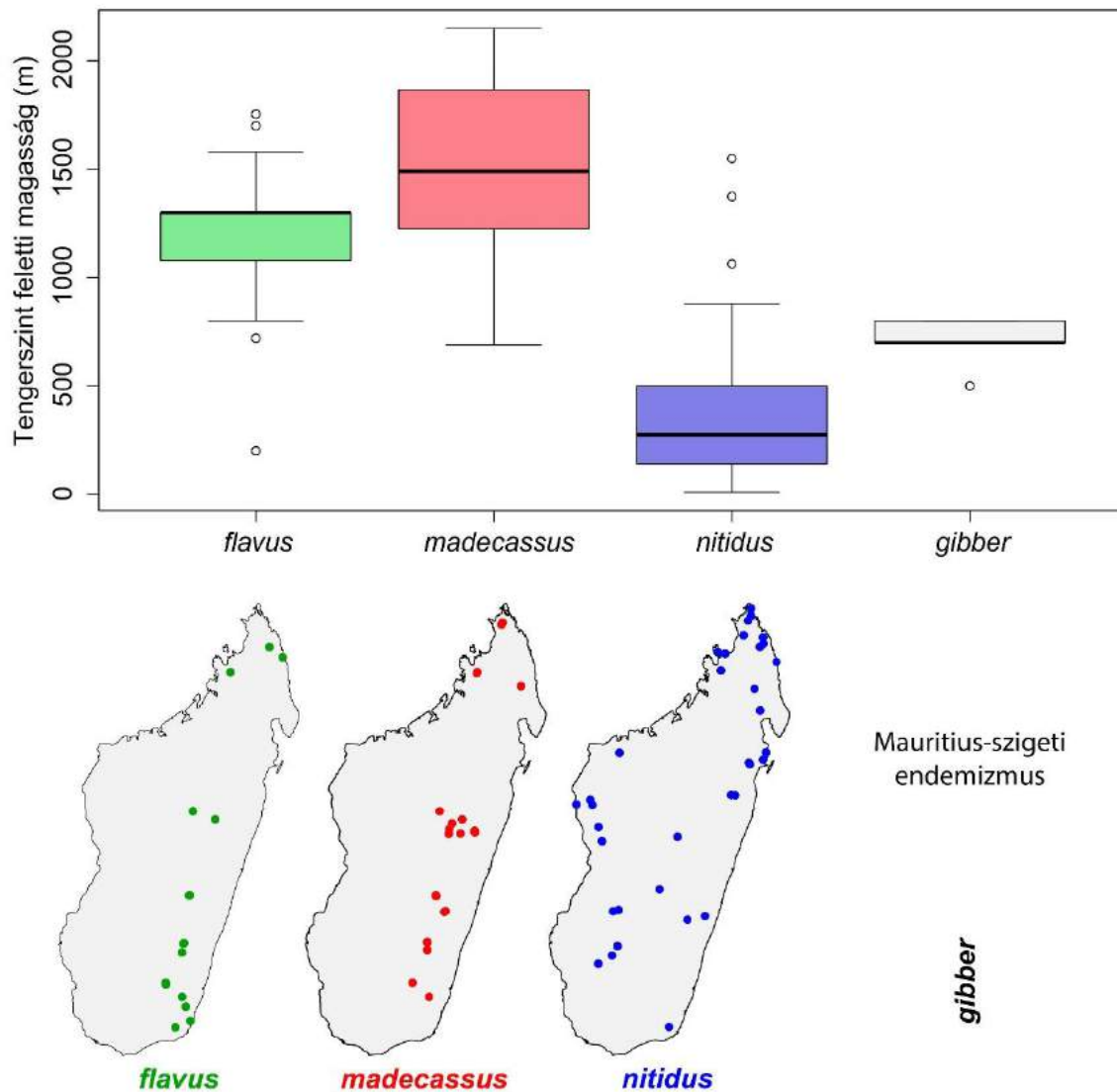
5.4.4.3 A *Nesomyrmex* fajok környezeti és geográfiai szegregációja

Madagaszkár – a Malagasi régió legnagyobb szigete – akkora, mint a Brit-szigetek, és ennek megfelelően nagyon változatos tájakat találunk rajta. A természetrombolás során az erdőségek legnagyobb részét bár kiirtották, még ma is lenyűgöző az élőhelyek változatosága. Részben ennek is tudható be a Malagasi állatvilág gazdagsága. A hangyák gyakorlatilag minden benépesíthető szárazföldi élőhelyen jelen vannak. Ugyanúgy megtalálhatóak a magashegységekben és a sivatagokban is. Madagaszkáron a *Nesomyrmex* fajok minden nagyobb ökológiai régióban (Schatz 2000, Cornet 1974), a sziget egész területén megtalálhatók. A fajcsoportok nem mutatnak különleges, csak rájuk jellemző geográfiai eloszlást, döntő többségük elterjedése jól illeszkedik a madagaszkári ökológiai régiók kiterjedéséről alkotott ismereteinkbe (**29. ábra**). Egyes fajcsoportokon belül viszont a fajok egymástól lehetőség szerint elkülönülnek. Ez az elkülönülés történhet vízszintes vagy függőleges mintázatok mentén.

- a) A *N. hafahafa* fajcsoport a délnyugati sivatagi cserjés régiókban elterjedt. A fajai (4 faj) ökológiai igényeiben – a rendelkezésre álló információk alapján – nem találunk lényeges különbséget, azonban a fajok egymással elterjedésükben nem fednek át, allopatrikus- vagy parapatrikus elterjedési mintázatot mutatnak (**27. ábra**).
- b) A *N. angulatus* és *N. sikorai* fajcsoport fajainak elterjedési mintázatai (fajcsoportonként egyaránt 8-8 fajjal) jól követik a madagaszkári ökológiai régiók határait (**29. ábra**). Ezek a fajok többnyire allopatrikus vagy parapatrikus fajpárokat tartalmaznak, azaz egymással át nem fedő elterjedési területtel bírnak, vagy csak egy keskeny kontaktzónában találkoznak.
- c) A *N. madecassus* fajcsoport négy faja viszont elterjedésükben nem különülnek el élesen egymástól, és nem követik szigorúan az ökológiai régiók (Schatz 2000, Cornet 1974) határait. Felismerhető azonban a fajok tengerszint feletti magasság szerinti szegregációja (**32. ábra**).

5.4.4.4 Malagasi endemizmusok a fauna vizsgált szegmensében

A Malagasi *Nesomyrmex* fajok egy faj, a *N. angulatus* kivételével, endemikusnak mutatkoznak a régióra, ezekről a fajokról nincs jelenleg adatunk más területekről. A fajcsoportok között mindössze egyetlen endemikus egységet találunk, ez a *N. hafahafa* fajcsoport, amely más régiókból, így például az Afrikai kontinensről sem ismert, a többi csoportba sorolható elemek azonban fellelhetők Afrikában is. Ez arra utal, hogy a Malagasi *Nesomyrmex* fauna döntően az afrikai fauna elemeinek a Mozambiki-csatorna (**33. ábra**) óceáni áramlataival történő egyirányú beáramlással jöhetett létre (Samonds és mtsai. 2013) az elmúlt néhány tíz millió évben.

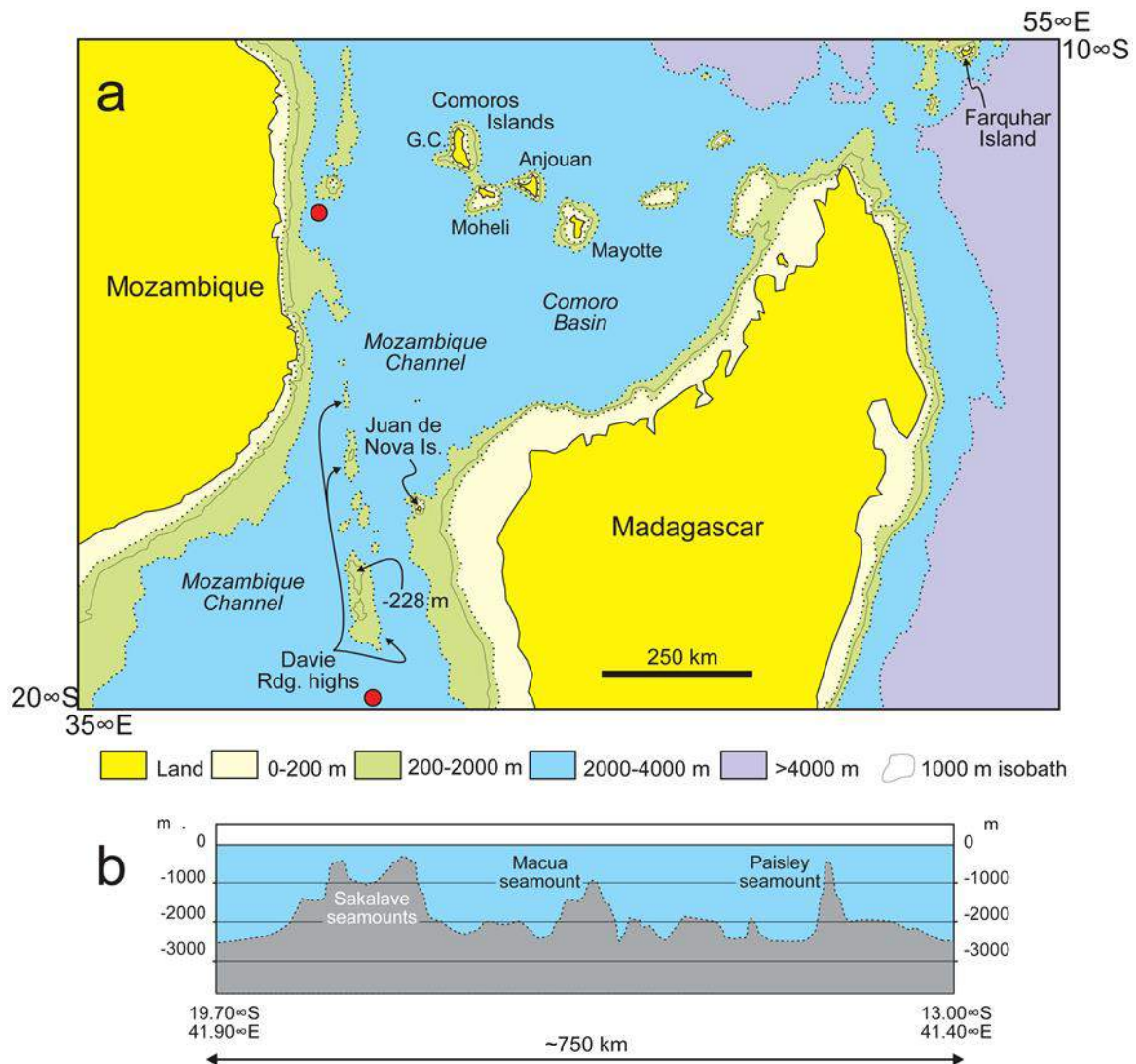


32. ábra. A *Nesomyrmex madecassus* fajcsoport fajainak vertikális és horizontális elterjedési mintázata a Malagasi régióban.

A Malagasi régióban található fajcsoportok 4/5-ödének afrikai jelenléte, valamint a fajcsoportok közepesen nagy fajsza (4-9 faj fajcsoportonként) a többszörös kolonizáció elméletét erősíti, azaz Madagaszkár *Nesomyrmex* faunájának beáramlása több, egymástól független eseményhez köthető. Egyedül a *N. hafahafa* fajcsoport elterjedési jellemzői mondhatók átlagtól eltérőnek. Maga a fajcsoport morfológiája is a három, dorzális (pronotális, propodeális és petioláris) helyzetű kutikuláris tüskepár miatt, rendkívül egyedi. Hasonló alaktani jellemzőkkel rendelkező fajokat az Afrikai kontinensről nem ismerünk²⁶. Lehetőségként szóba jöhet az a magyarázat, hogy a *N. hafahafa* fajcsoport Malagasi endemizmu-

²⁶ Ez nem jelenti azt, hogy a földtörténeti múltban nem lett volna jelen a kontinensen

sát autogén kialakulásnak köszönheti egy más, bazálisabb fajcsoportból történt leágazással, azonban valószínűbb, hogy a fajcsoport nem Madagaszkáron alakul ki, hanem – a többi leszármazási vonalhoz hasonlóan – afrikai eredetű. Ez utóbbi esetben az afrikai kontinensen kihalt fajcsoportról beszélhetünk. A *hafahafa* fajcsoport egésze Madagaszkáron is egy reliktnak tekinthető. Valószínűleg régebben szélesebb elterjedési területtel rendelkezhetett, a jelenben viszont visszaszorulóban van. Két ide sorolható faj, a *N. capricornis* és *N. spinosus*, populációi nagy valószínűséggel a kihalás szélén vannak (27. ábra). Az sem elképzelhetetlen, hogy a két említett faj időközben a nagyszabású környezetpusztító tevékenységnek köszönhetően mára kihalt, és leírásukra a kihalás után került sor (ld. még 5.4.2. alfejezetben írottakat).



33. ábra. A Mozambiki-csatorna területének egyszerűsített mélység-térképe. A képen (fölső kép, a) Madagaszkár, Mozambik, és a közeli szigetek viszonylagos elhelyezkedése láthatók, míg a (alsó kép, b) mélységmérő a szoros keresztmetszetét mutatja. Sammonds és mtsai. (2013) nyomán.

5.5 Malagasi kollégák kvantitatív taxonómiai munkára való felkészítése és a módszertan meghonosítása helyi szinten

Az imént bemutatott revíziós munkák modellként szolgálhatnak további Malagasi hangyagenuszok feldolgozásához is. A módszer szabadon elérhető, és letölthető valamely cikk mellékletéből. A Malagasi fauna feltárására sok nemzetközi projekt jött létre, azonban a helyi erőforrások bevonása nélkül ez nehezen valósítható meg. A 2002-ben alapított a *Madagascar Biodiversity Centre* (Antananarivo, Madagascar) a *California Academy of Sciences*-el szoros együttműködésben igyekszik feltárni a helyi faunát.

A taxonómiai képzettség lokális erősítése alapvető fontosságú a helyi fauna feltárásának szempontjából, ezért a helyi szakemberek számára a két intézet égisze alatt futó módszertani képzési programot indítottunk a Schlinger pályázatomban kezdeti szakaszától (2014–). A programban a *Madagascar Biodiversity Centre* (<http://www.madagascarbio.org/our-team.html>) két helyi munkatársa, a labor vezető taxonómusa, Jean Claude Rakotonirina (PhD) és egy fiatal, doktori fokozat előtt álló kolléga, Nicole Rasoamanana vesz részt. Azzal a konkrét elképzeléssel vágtam bele ebbe a munkába, hogy a *Madagascar Biodiversity Centre* (Antananarivo, Madagascar) munkatársainak morфомetrikus és analitikai tréningezésével hozzájáruljak a helyi taxonómiai tudás erősítéséhez, s ezen keresztül a helyi fauna feltárásának meggyorsításához.

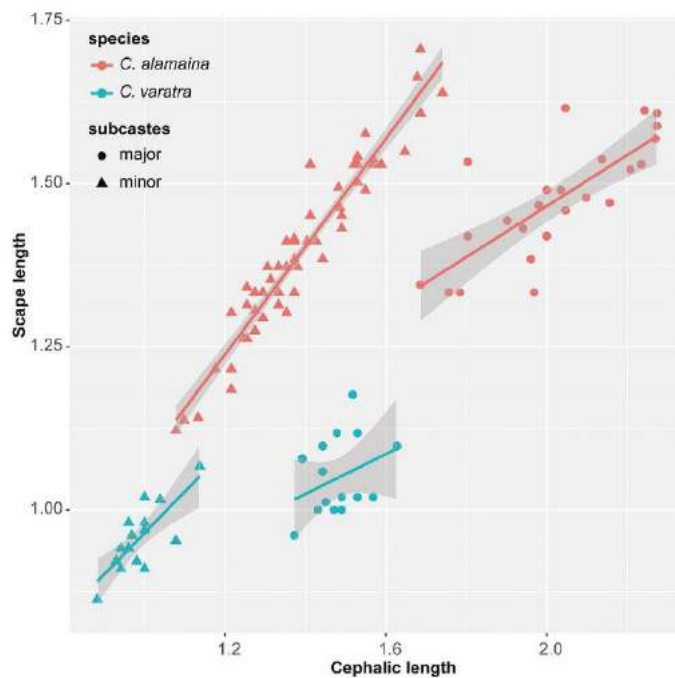
34. ábra. A *Madagascar Biodiversity Centre* morфомetrikus továbbképzési programjába bevont kollégák: Nicole Rasoamanana és Jean Claude Rakotonirina



Az említett madagaszkári kollégák a szélsőségesen fajgazdag *Camponotus* genusz revíziójához használják a megszerzett morфомetrikus mérési ismereteket és az elemzésükhöz szükséges ismereteket, többek között az általam összeállított NC-PART klaszterezést. A munkájuk folyamatos figyelemmel kísérésével nyújtok segítséget, nem csak a régió faunájának feltárásában, hanem hosszú távú stratégiai építkezésben is.

A feladat nehézségét jelzi, hogy a *Camponotus* világszerte a leginkább fajgazdag genuszok között van számon tartva. A több mint 1500 ismert fajból és alfajból (www.antcat.org) a madagaszkári régióban eddig összesen 78 *Camponotus* faj ismeretes, valamint becslések szerint további 100 taxon (www.antweb.org) előfordulása várható.

Mindemellett, a *Camponotus* fauna modern feltárásának számos, ismert nehézség szabhat gátat: elsőként a taxonok nagyfokú morfológiai változatossága, különösen az *intraspecifikus* és *intranidális* (kolónián belüli) polimorfizmus (**35. ábra**) magas szintje; másodsorban a három morfológiai kaszt (minor, média és major dolgozói kasztok) jelenléte; harmadszor pedig a morfológiai jellemzők diagnosztikai értékének – konkrét ismeretek és a tesztek hiányából adódó – bizonytalansága. Ezen három, együttesen jelentkező nehézség rendkívül nehézé teszi a génusz körüli hagyományos megközelítésekkel operáló taxonómiai munkát.



35. ábra. Két *Camponotus* faj, *C. alamaina* (piros) és *C. varatra* (kék) jellegeinek allometrikus összefüggései. A csápnyél (Scape length, y tengely) hossza mindkét faj esetében tört allometriát mutat a fej hosszán (Cephalic length, x tengely) ábrázolva. Az allometriában bekövetkező törés és a közöttük jelentkező rés két különböző fenotípusra osztja a populációkat: a háromszögek a kis dolgozókat (minor workers), a körök a nagy dolgozókat (major) jelzik. A sötétszürke területek a 95%-os konfidencia-intervallumot mutatják, a regressziós vonalat lineáris modell alapján számoltam.

Kvantitatív módszerekkel viszont ezek az akadályok könnyebben elháríthatók, illetve a revíziós munkák diverzitási mintázatokra vonatkozó része viszonylag rövid idő alatt elvégezhető. Az én feladatom ebben a munkában sokrétű: a) a morfometrikus megközelítés alapelveinek ismertetése, b) a kvantitatív taxonómia filozófiájának elsajátítása, c) protokollok felállítása, d) statisztikai alaptudás átadása, elemzések elvégzésének segítése, e) adatelemzés és interpretálás, f) kézirat elkészítése és javítása.

A teljes program lefutását öt évben határoztuk meg. A tervek szerint ennyi időre van ugyanis szükség a Malagasi *Camponotus* fauna teljes feltárásához. Az első év döntően képzési jellege után következő második évtől már kézzelfogható eredményekről is beszámolhatunk. Az értekezés leadásának idején három *Camponotus* fajcsoport kvantitatív taxonómiai feldolgozását bemutató cikk már elérhető (Rakotonirina, Csősz & Fisher (2016) 81–154, ZooKeys <https://doi.org/10.3897/zookeys.572.7177>, Rakotonirina, Csősz & Fisher (2017) Zootaxa 4238: 203 – 245 <http://www.mapress.com/j/zt/article/view/zootaxa.4238.2.2/> és Rasoamanana, Csősz & Fisher (2017) ZooKeys, 681: 119. <https://zookeys.pensoft.net/articles.php?id=13187>, egy következő munka pedig leadási vagy elbírálási szakaszban van.

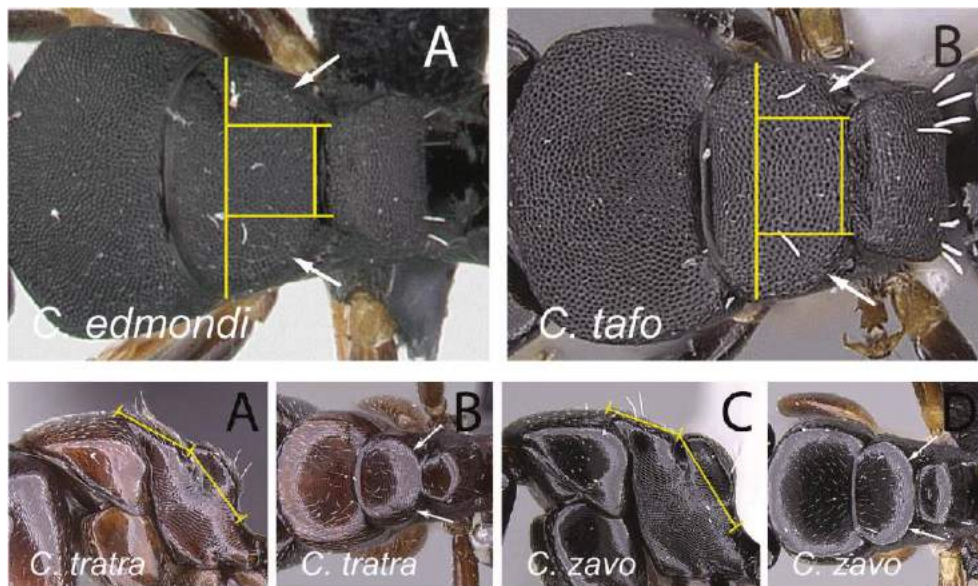
	teljes kezelt fajszám	tudományra új fajok	feldolgozott fajcsoport
<i>Rakotonirina</i> , Csősz & Fisher (2016)	15	11	1
<i>Rakotonirina</i> , Csősz & Fisher (2017)	8	4	2
<i>Rasoamanana</i> , Csősz & Fisher (2017)	4	3	1
Összesen	27	18	4

6. táblázat. A kvantitatív módszertannal elvégzett revíziók során feldolgozott Malagasi *Camponotus* fajok száma, tudományra új fajok száma és a revideált fajcsoportok száma publikációnkénti felbontásban.

A modern módszertannal revideált 27 fajból 18 bizonyult tudományra újnak (**6. táblázat**), amely a teljes fajszám több mint fele. A három megjelentetett cikk összesen négy, morfológiai jellegek alapján definiált fajcsoportot dolgoz fel.

A kriptikus fajok számát utólag nem egyszerű meghatározni²⁷, de becsléseim szerint a teljes fajszám 15-25%-át tekinthetjük a legszigorúbb meghatározás szerinti kriptikus fajnak. Ez a szám némileg alulmarad a Malagasi *Nesomryrmex* faunánál tapasztaltaktól, de a Palearktikus *Camponotus* fajokra sem jellemző a kriptikus fajok nagy száma. A morfológiai karakterek kis- és nagy dolgozóinak (egyes terminológiák szerint szokás ezeket *katonák*-nak nevezni) allometrikus összefüggéseinek letisztázása (**35. ábra**) sokat javított a morfológiai jellegek alkalmazhatóságán. Miután ezt az akadályt elhárítottuk, a revíziók elkészítéséhez kvantitatív karakterek (**36. ábra**), aránypárok biztonsággal használhatók.

²⁷ A fajhatárok kvantitatív alapon történő meghatározása után ugyanis, az esetek jelentős részében utólag sikerül valamilyen külső morfológiai jelleget találni, amellyel a fajok elkülönítése kvantitatív módszerek nélkül is elvégezhető.



36. ábra. Számszerűsített morfológiai jelegek alkalmazása megkönnyíti a Malagasi *Camponotus* fajok meghatározását. Rakotonirina, Csősz & Fisher (2016) nyomán.

6 Az eredmények összefoglalása

A karrieremnek a doktori (PhD) fokozata megszerzése utáni időszakát (2007-től napjainkig) az a célkitűzés motiválta, hogy a taxonómia modern, XXI. századi információ alapú tudománnyá alakításához tevékenyen hozzájáruljak. A taxonómia kvantitatív irányzatának alkalmazása munkásságomat alapjaiban meghatározta. A céljaim elérését olyan modern módszertani keretek között igyekeztem megvalósítani, amelyek teret adnak a taxonómiai eredmények széleskörű interpretálhatóságának és az ismeretek könnyebb átadásának.

6.1 A szemantikus fenotípus megközelítés alkalmazása hangyákon

A szemantikai kijelentések hangyákra történő kialakítása, az ontológiák fejlesztése mellett a világon elsőként a saját munkáimhoz kötődik. A megközelítés gyakorlati felhasználásának alapjait rangos folyóiratokban közöltem (Csősz, Mikó & Heinze 2015 és Csősz & Fisher 2016a). Munkám során folyamatosan igyekszem finomítani és tisztázni a korábbi taxonómiai revíziók során használt fenotípus megközelítéseket, és megvizsgálom a meglévő adatokkal való újítások lehetőségeit. A leírásokban és kulcsokban használt morfológiai állítások hártáásszárnyú-specifikus terminológiája, valamint a fenotípuson alapuló diagnózisok ontológiás leírásai *Hymenoptera anatómia ontológia* (HAO) URI táblázatban kerülnek csoportosításra (1. táblázat).

6.2 NC-klaszterezés és NC-PART morfológiai mintázat felismerő rendszerek

A taxonómiai döntéshozás gyakran hosszadalmas és fáradtságos folyamatát algoritmikus eljárások bevezetésével segítetttem felgyorsítani. Saját elgondolás és kezdeményezés alapján egy új eljárást vezettem be „NC-klaszterezés” néven (Seifert, Ritz & Csősz 2014). A módszer extenzív tesztelési folyamatainak eredményei alapján kijelenthető, hogy elméletileg bármely euszociális szervezet populációira – így mint hártáásszárnyúak (hangyák, méhek, darazsak), termeszek kolóniái, gubacsképző tetvek, stb. – biztonsággal alkalmazható.

Az NC-klaszterezés továbbfejlesztésével később kiküszöböltem az NC-klaszterezésnek azt a hiányosságát is, hogy bár látjuk a mintázatot, az adatsorban rejlő klaszterek számára nem kapunk becsléseket. Ezt az NC-klaszterezés és egy ún. rekurzív partíciónáló alkalmazás (PART, Nilsen & Lingjaerde 2013) kombinálásával sikerült elérnem. Az NC-klaszterezéssel kombinált PART algoritmus (NC-PART-klaszterezés) statisztikai küszöbértékek alapján objektíven képes kijelölni a klasztereket és a mintákat automatikusan partíciónálja a megfelelő klaszterekbe. Ezt a kombinációt szintén extenzív tesztelési folyamat után publikáltam (Csősz & Fisher 2016a, Csősz & Fisher 2016b). Az eljárás nagy segítséget jelent

sok bizonytalansági tényezővel bíró – például ismeretlen, trópusi – csoportok esetén, valamint kevésbé tapasztalt kollégák számára.



37. ábra. A karrierem során elvégzett taxonómiai revíziók során faji rangon kezelt taxonok száma éves lebontásban. Kék színnel a hagyományos módszerrel elvégzett revíziók taxon-száma, narancs színnel az NC-klaszterezéssel felismert összes fajszám látható.

Az új módszer hatékonyságának növekedése két fő tényezővel jellemezhető: A taxonómiai döntéshozás meggyorsításának képességével (a), valamint a felismert mintázat pontosságával (b) az új módszer alkalmazhatóságának értékelése szempontjából alapvető fontosságú.

- a) **A taxonómiai döntéshozás meggyorsításának képessége.** A bevezetett NC-PART-klaszterezés bizonyítottan képes jelentősen meggyorsítani a hangyafauna feltárását terén azokban az esetekben, ahol a fenotípusos jellegek jól számszerűsíthetők, és megfelelő mennyiségű anyag áll rendelkezésre. Ez a módszer mind tapasztalt taxonómusok, mind pedig kevésbé gyakorlott szakemberek számára is hathatós segítséget jelenthet. Az új eljárás használatának taxonómia meggyorsításában játszott szerepe a saját munkásságomban is szebeötlő: az NC-klaszterezéssel revideált fajok száma évenkénti bontásban jelentősen megnövekedett (**37. ábra**).

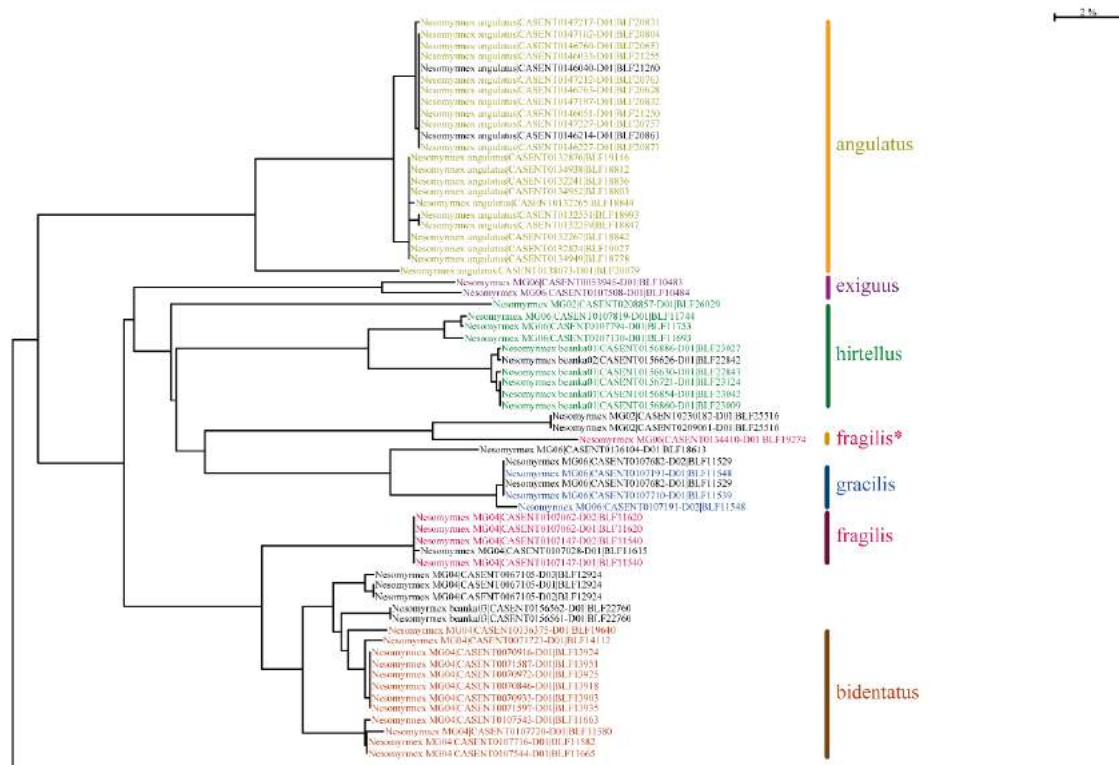
- b) **A felismert mintázat pontossága**, azaz a taxonómiai megoldóerő (*taxonomic resolving power*) és a megbízhatóság (*reliability*) együttes erősítése. A mintázat helyénvalóságát a legjobban független adatsorból történő kereszt-tesztelésekkel lehet igazolni. A rendelkezésre álló molekuláris (többnyire COI) adatsorok az NC-klaszterezéssel nagyon hasonló mintázatot adnak. Az összesen 18 fajt számláló, európai *Temnothorax* genusz *nylanderi* fajcsoportjának integratív vizsgálatai a COI és az NC-klaszterezés 97%-os egyezését²⁸ adták eredményül, amely átlagosnak mondható. Más csoportok, mint a 8 fajt magában foglaló Malagasy *Nesomyrmex angulatus* fajcsoport esetén, (**38. ábra**) még kisebb eltérést találunk a molekuláris (COI) és a morfológiás (NC-PART-klaszterezés) elemzések által kapott mintázatok között.

6.3 Taxonómiai munkásságom

A módszertani fejlesztések mellett komoly hangsúlyt helyeztem magára a – szűkebb értelemben vett – taxonómiai munkára is. A karrierem során ezen munkákban összesen 125 taxont kezeltem faji rangon, ebből 14 taxon státuszát változtattam meg, és 62 tudományra új fajt írtam le, ezen felül közel 20 új szinonímiát javasoltam.

- a) **A nyugat-palearktikus hangyafauna** jobb megismerése alapvető célkitűzésem közé tartozik. Ezen belül is kitüntetett helyen szerepel a Kárpát-medence faunájának a Ponto-Kaspi régióval való kapcsolatának kutatása (Csősz és mtsai. 2014a, 2014b, 2015). A nyugat-palearktikus taxonok közül jellemzően – de nem kizárólag – Myrmicinae hangyák (*Leptothorax*, *Messor*, *Myrmica*, *Temnothorax*, *Tetramorium* genuszok) taxonómiai feltárásával foglalkoztam, mely során összesen 69 faj taxonómiáját rendeztem, melyből 17 bizonyult tudományra újnak.
- b) **A madagaszkári hangyafauna**, azaz a biodiverzitás egy trópusi szegmensének feltárásához a karrierem későbbi szakaszában nyílt lehetőségem komolyan hozzájárulni (Csősz & Fisher 2015, 2016a, 2016b, 2016c, 2016d, Rakotonirina, Csősz & Fisher 2016, 2017). A Malagasi fauna feltárására irányuló projektem során elért eredmények aláhúzzák a Madagaszkár és a környező szigetvilág nagyfokú fajgazdagságáról alkotott ismereteinket. Döntően a morfometrikus adatokban rejlő mintázatok felismerésére építő – saját magam által összeállított – eljárás komoly segítséget jelentett a madagaszkári *Nesomyrmex* fauna jobb megismerésében: a korábban ismert 4 faj mellé 29 tudományra új faj került, így 33-ra növeltem a régióban ismert *Nesomyrmex* fajok számát.

²⁸ Az egyezés alatt az eltérő megközelítések által egyes klaszterekbe sorolt minták értendők. A morfológiai és molekuláris adatok elemzésével kapott dendrogramok topológiája nem volt összehasonlítás tárgya.



38. ábra. A *Nesomyrmex angulatus* fajcsoport Malagasy képviselőinek molekuláris (COI) törzsfája. Színkódokkal a morfológiai fajhipotézisek láthatók. A szekvenált, de morfológiailag nem vizsgált egyedek feketével jelennek meg az ábrán. A párosított 52 egyedből 51 (>98%) molekuláris és a morfológiai eljárások alapján történt osztályozása megegyezett. Az egyetlen eltérően osztályozott példány csillaggal jelölve.

6.4 Taxonómus-képzés Madagaszkáron – módszertani tréning helyi kollégák számára

A taxonómiai képzettség lokális erősítése alapvető fontosságú a helyi fauna feltárásának szempontjából, ezért az általam revideált csoportok taxonómiai tisztázásán túl a *Madagascar Biodiversity Centre* (<http://www.madagascarbio.org/>) munkatársainak morфомetrikus és többváltozós statisztikai képzését tekinthetem a magam után hagyott egyik fontos eredményemnek. Ezzel is hozzájárultam a taxonómia célkitűzésének mihamarabbi megvalósulásához ott, ahol arra a legnagyobb szükség mutatkozik. A Malagasi kollégák munkájának folyamatos figyelemmel kíséréseivel igyekszem segítséget nyújtani, nem csak a régió faunájának feltárásában, hanem hosszú távú stratégiai építkezésben is. Az eredmények nem maradnak elszigeteltek, jelenleg a madagaszkári ökológiai és biogeográfiai egységek határainak a hangyafauna elterjedési mintázataival történő tesztelését tervezzük, amelyben a modern módszerekkel revideált *Nesomyrmex* génusznak, valamint a helyi kollégák által – közreműködéssel – feldolgozás alatt álló *Camponotus* genusznak kitüntetett szerepe lesz.

7 Köszönetnyilvánítás

Első sorban nagyon köszönöm a családomnak, feleségemnek, Katának és lányomnak, Boginak a támogatást és a sok türelmet! Köszönöm barátaimnak és kollégáimnak, Buczkó Krisztnak, Kassai Ferencnek, Makranczy Györgynek, Mikó Istvánnak és Rózsa Lajosnak a támogatást, biztatást és a szakmai tanácsokat.

Köszönet illeti korábbi tanáraimat, Prof. Varga Zoltánt, Dr. Gallé Lászlót, Dr. Papp Lászlót, akik alapvetően meghatározták a pályafutásomat.

Kollégáim között kiemelt köszönettel tartozom Dr. Bernhard Seifertnek, aki egyfajta mentorként állt mellettem, és akinek a kvantitatív módszertan területén nyújtott iránymutatásából rengeteget merítettem.

Korábbi munkahelyeim közül a kiemelkedően támogató és stabil környezetért külön köszönet jár Prof. Podani Jánosnak (MTA-ELTE-MTA Ökológiai kutatócsoport) és Dr. Brian L. Fishernek (California Academy of Sciences).

Köszönet jár még azon kollégáimnak, Markó Bálintnak, Tartally Andrásnak, Sasha Radchenkonak és Andreas Schulznak, akik komoly hatással voltak szakmai gondolkodásom alakulására.

Köszönöm továbbá minden eddigi hallgatómnak szakdolgozómnak, Cserhalmi Katalinnak, Erős Katalinnak, Kurali Erzsébetnek, , Vajthó Szendrának, Vályi Nagy Mariannak, valamint Nicole Rasoamanananak és Jean Claude Rakotonirinanak a közös munkát és a kérdéseiket, melyek engem is fejlesztettek, előrelendítettek.

A fenti eredmények több kutatási projekt keretében, összehangolt és megtervezett munka eredményeként jöhettek létre: Schlinger Postdoctoral Fellowship (California Academy of Sciences, San Francisco, USA, „*Phylogeny and systematics of the Malagasy *Nesomyrmex* ants*”), Magyar Állami Eötvös ösztöndíj (SMNK, Karlsruhe, Németország, „*A *Temnothorax parvulus* faj komplex Balkán-félsziget, Törökország és a Közel-Kelet fajainak taxonómiai revíziója*”), National R&D Programme (No: 3B023-04: „*The origin and genesis of the fauna of the Carpathian Basin: diversity, biogeographical hotspots and nature conservation significance*”). Végül köszönöm a Magyar Tudományos Akadémia, MTA-NEI osztályának az egyéni tanulmányutakkal (Törökországba, Oroszországba és Romániába) való támogatását, melyek nélkül a munkáim bizonyosan nem érthették volna el ezt a szakmai színvonalat.

8 Irodalomjegyzék

Az irodalmi hivatkozások között a szövegben félkövér betűtípussal kiemelt saját publikációim listáját külön felsorolásban adom meg (lehetőség szerint DOI linkkel ellátva).

8.1 Az értekezéshez felhasznált publikációim listája

a) A doktori fokozat megszerzése utáni (2007-2017) publikációk listája

- Csősz S** & Fisher BL (2015). Diagnostic survey of Malagasy *Nesomyrmex* species-groups and revision of *hafahafa* group species via morphology based cluster delimitation protocol. *ZooKeys*, 526: 19-59. <https://doi.org/10.3897/zookeys.526.6037>. [IF: 0,938]
- Csősz S** & Fisher BL (2016a) Taxonomic revision of the Malagasy members of the *Nesomyrmex angulatus* species group using the automated morphological species delineation protocol NC-PART-clustering. *PeerJ*, 4, e1796. <https://doi.org/10.7717/peerj.1796>. [IF: 2,2]
- Csősz S** & Fisher BL (2016b) Toward objective, morphology-based taxonomy: a case study on the Malagasy *Nesomyrmex sikorai* species group (Hymenoptera: Formicidae). *PloS one*, 11(4), e0152454. <http://dx.doi.org/10.1371/journal.pone.0152454>. [IF: 2,806]
- Csősz S** & Fisher BL (2016c) Taxonomic revision of the Malagasy *Nesomyrmex madecassus* species-group using a quantitative morphometric approach. *ZooKeys*, 603: 105-130. <https://doi.org/10.3897/zookeys.603.8271>. [IF: 1,031]
- Csősz S** & Fisher BL (2016d) Taxonomy of Malagasy *Nesomyrmex brevicornis* species-group using conventional morphology-based approach. *ZooKeys*, 616: 125–159. <https://doi.org/10.3897/zookeys.616.9532>. [IF: 1,031]
- Csősz S** & Majoros G (2009) Ontogenetic origin of mermithogenic *Myrmica* phenotypes (Hymenoptera, Formicidae). *Insectes sociaux*, 56(1), 70-76. [IF: 1,480]
- Csősz S** & Schulz A (2010) A taxonomic review of the Palaearctic *Tetramorium ferox* species-complex (Hymenoptera, Formicidae). *Zootaxa*, 2401, 1-29. [IF: 0,853]
- Csősz S** (2012) Nematode infection as significant source of unjustified taxonomic descriptions in ants (Hymenoptera: Formicidae). *Myrmecological News*, 17, 27-31. [IF: 2,157]
- Csősz S**, Heinze J & Mikó I (2015) Taxonomic Synopsis of the Ponto-Mediterranean ants of *Temnothorax nylander* species-group. *PloS one*, 10(11), e0140000. <http://dx.doi.org/10.1371/journal.pone.0140000>. [IF: 3,057]
- 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. [IF: 2,888]

- Csősz S**, Wagner H C, Bozsó M, Seifert B, Arthofer W, Schlick-Steiner B C, Steiner F M, Péntes Z (2014) Tetramorium indocile Santschi, 1927 stat. rev. is the proposed scientific name for Tetramorium sp. C sensu based on combined molecular and morphological evidence (Hymenoptera: Formicidae). *Zoologischer Anzeiger-A Journal of Comparative Zoology*, 253(6), 469-481. [IF: 1,483]
- Deans AR, Lewis SE, Huala E, Anzaldo SS, Ashburner M, Balhoff JP, Blackburn DC, Blake JA, Burleigh JG, Chanet B, Cooper LD, Courtot M, **Csősz S**, Cui H, Dahdul W, Das S, Dececchi TA, Dettai A, Diogo R, Druzinsky RE, Dumontier M, Franz NM, Friedrich F, Gkoutos GV, Haendel M, Harmon LJ, Hayamizu TF, He Y, Hines HM, Ibrahim N, Jackson LM, Jaiswal P, James-Zorn C, Kohler S, Lecointre G, Lapp H, Lawrence CJ, Le Novere N, Lundberg JG, Macklin J, Mast AR, Midford PE, Miko I, Mungall CJ, Oellrich A, Osumi-Sutherland D, Parkinson H, Ramirez MJ, Richter S, Robinson PN, Ruttenberg A, Schulz KS, Segerdell E, Seltmann KC, Sharkey MJ, Smith AD, Smith B, Specht CD, Squires RB, Thacker RW, Thessen A, Fernandez-Triana J, Vihinen M, Vize PD, Vogt L, Wall CE, Walls RL, Westerfeld M, Wharton RA, Wirkner CS, Woolley JB, Yoder MJ, Zorn AM & Mabee P (2015) Finding Our Way through Phenotypes. *PLoS Biology* 13(1): e1002033. doi:10.1371/journal.pbio.1002033. <http://dx.doi.org/10.1371/journal.pbio.1002033>. [IF: 9,343]
- Rakotonirina JC, **Csősz S** & Fisher BL (2016) Revision of the Malagasy *Camponotus edmondi* species group (Hymenoptera, Formicidae, Formicinae): integrating qualitative morphology and multivariate morphometric analysis. *ZooKeys*, 572: 81–154. <https://doi.org/10.3897/zookeys.572.7177>. [IF: 1,031]
- Rakotonirina JC, **Csősz S** & Fisher BL (2017) Taxonomic revision of the Malagasy *Camponotus grandidieri* and *niveosetosus* species groups (Hymenoptera, Formicidae) using qualitative and quantitative morphology. *Zootaxa*, 4238(2), 203-245. <http://www.mapress.com/j/zt/>. [IF: 0,994]
- Rasoamanana N, **Csősz S** & Fisher BL (2017) Taxonomic revision of imitating carpenter ants, *Camponotus* subgenus *Myrmopytia* (Hymenoptera, Formicidae) of Madagascar, using morphometry and qualitative traits. *ZooKeys*, 681: 119. <https://zookeys.pensoft.net/articles.php?id=13187>. [IF: 1,031]
- Seifert B & **Csősz S** (2015) *Temnothorax crasecundus* sp. n.—a cryptic Eurocaucasian ant species (Hymenoptera, Formicidae) discovered by Nest Centroid Clustering. *ZooKeys*, 479: 37. <http://zookeys.pensoft.net/articles.php?id=4610>. [IF: 0,938]
- Seifert B, Ritz M & **Csősz S** (2014) Application of exploratory data analyses opens a new perspective in morphology-based alpha-taxonomy of eusocial organisms. *Myrmecological News*, 19: 1–15. [IF: 2,898]
- Seifert B, **Csősz S** & Schulz A (2014) NC-Clustering demonstrates heterospecificity of the cryptic ant species *Temnothorax luteus* (Forel 1874) and *T. racovitzai* (Bondroit 1918)(Hymenoptera: Formicidae). *Contribution to Entomology*, 64(1): 47-57. [IF: -]

b) A doktori fokozat megszerzése előtti (1998-2007) publikációk listája

- Csősz S**, Radchenko AG & Schulz A (2007) Taxonomic revision of the Palaearctic *Tetramorium chefketi* species complex (Hymenoptera: Formicidae). *Zootaxa*, 1405(1): 1-38. [IF: 0,691]
- Schlick-Steiner BC, Steiner FM, Konrad H, Markó B, **Csősz S**, Heller G, Ferencz B, Sipos B, Christian E, Stauffer C (2006). More than one species of *Messor* harvester ants (Hymenoptera: Formicidae) in Central Europe. *European Journal of Entomology*, 103: 469-476. [IF: 0,782]
- Csősz S** & Markó B (2004) Redescription of *Tetramorium hungaricum* Rösler, 1935, a related species of *T. caespitum* (Linnaeus, 1758)(Hymenoptera: Formicidae). *Myrmecologische Nachrichten*, 6: 49-59. [IF: -]
- Csősz S** & Seifert B (2003) *Ponera testacea* Emery, 1895 stat. n.-a sister species of *P. coarctata* (Latreille, 1802)(Hymenoptera, Formicidae). *Acta Zoologica Academiae Scientiarum Hungaricae*, 49(3): 201-214. [IF: 0,5]

8.2 Az értekezés elkészítéséhez fel nem használt publikációim listája

- Tartally A, Antonova V, Espadaler X, **Csősz S** & Czechowski W (2016) Collapse of the invasive garden ant, *Lasius neglectus*, populations in four European countries. *Biological invasions*, 18(11): 3127-3131.
- Csata E, Czekes Z, Erős K, Német E, Hughes M, **Csősz S** & Markó B (2013) Comprehensive survey of Romanian myrmecoparasitic fungi: new species, biology and distribution. *North-Western Journal of Zoology*, 9 (1): 23-29.
- Czekes Zs, Radchenko AG., **Csősz S**, Szász-Len A, Tăușan I, Benedek K& Markó B (2012) The genus *Myrmica* Latreille, 1804 (Hymenoptera: Formicidae) in Romania: distribution of species and key for their identification. *Entomologica romanica*, (17): 29-50.
- Csősz S**, Lapeva-Gjonova A & Markó B (2011) New data on the geographical distribution and host utilization of the entomopathogenic fungus *Myrmicinosporidium durum*. *Journal of insect science*, 12: 129-129.
- Csősz S**, Markó B & Gallé L (2011) The myrmecofauna (Hymenoptera: Formicidae) of Hungary: an updated checklist. *North-Western Journal of Zoology*, 7(1): 55-62.
- Arnaldo PS, Wynhoff I, Soares P, Rodrigues MC, Aranha J, **Csősz S**, Maravalhas E, Tartally A (2011). *Maculinea alcon* exploits *Myrmica aloba* in Portugal: unusual host ant species of a myrmecophilous butterfly in a peripheral region. *Journal of Insect Conservation*, 15(3), 465-467.
- De Jesús-Bonilla V S, Nunes J F, Penteado-Dias A M, Csősz S, Zaldívar-Riverón A (2011). A new synonym of the Neotropical parasitoid wasp genus *Notiospathius* (Braconidae, Doryctinae), with redescription of two species and description of five new species from Brazil. *ZooKeys*, (122), 71.
- Nagy C, Tartally A, Vilisics F, Merkl O, Szita É, Szél G, Podlussány A, Rédei D, Csősz S, Pozsgai G, Orosz A, Szövényi G, Markó V (2009) Effects of the invasive garden ant, *Lasius neglectus* VAN LOON, BOOMSMA & ANDRÁSFALVY, 1990 (Hymenoptera: Formicidae), on arthropod assemblages: pattern analyses in the type supercolony. *Myrmecological News*, 12, 171-181.
- Borowiec ML, Borowiec L, Csősz S, Radchenko AG (2009). Ants collected during 2006 Polish expedition to Kyrgyzstan (Hymenoptera: Formicidae). *Genus*, 20(2), 367-379.
- Ionescu-Hirsch, A., Markó, B., & Csősz, S. (2009). *Camponotus tergestinus* Müller, 1921 (Hymenoptera: Formicidae): first records of a rare species for Romania and Hungary. *Entomologica romanica*, 14, 19-22.
- Markó, B., Sipos, B., Csősz, S., Kiss, K., Boros, I., & Gallé, L. (2006). A comprehensive list of the ants of Romania (Hymenoptera: Formicidae). *Myrmecologische Nachrichten*, 9, 65-76.

- Csősz S & Markó B (2005) European ant species (Hymenoptera: Formicidae) in the ant collection of the Natural History Museum of Sibiu (Hermannstadt/Nagyszeben), Romania II. Subfamily Formicinae. In *Annales Historico-Naturales Musei Nationalis Hungarici* (Vol. 97, pp. 225-240).
- Gallé L, Markó B, Kiss K, Kovács É, Dürgeő H, Kőváry K, Csősz S (2005) Ant fauna of Tisza River Basin (Hymenoptera: Formicidae) In: Gallé L. *Vegetation and Fauna of Tisza River Basin*. 197 p. Szeged: Szegedi Tudományegyetem, 2005. pp. 149-197. (Tiscia Monograph Series 7)
- Tartally A & Csősz S (2004) Adatok a *Maculinea* boglárkalepkék (Lepidoptera: Lycaenidae) kárpát-medencei hangyagazdájairól.[Data on the ant hosts of the *Maculinea* butterflies (Lepidoptera: Lycaenidae) of Hungary.]. *Természetvédelmi közlemények*, 11:309-17.
- Csősz S (2003) A key to the Ponerinae species of the Carpathian Basin (Hymenoptera: Formicidae). In *Annales historico-naturales Musei nationalis hungarici* (Vol. 95, pp. 147-160).
- Markó B & Csősz S (2002) Die europäischen Ameisenarten (Hymenoptera: Formicidae) des Hermannstädter (Sibiu, Rumänien) Naturkundemuseums I.: Unterfamilien Ponerinae, Myrmicinae und Dolichoderinae. In *Annales historico-naturales Musei nationalis hungarici* (Vol. 94, pp. 109-121).
- Csősz S, Markó B, Kiss K, Tartally A, Gallé L (2002) The ant fauna of the Fertő-Hanság National Park (Hymenoptera: Formicoidea). In: Mahunka S (szerk.) *The fauna of the Fertő-Hanság National Park*. 423 p. Budapest: Magyar Természettudományi Múzeum, 2002. pp. 617-629. (Natural History of the National Parks of Hungary; vol. 1-2.)
- Csősz S (2001) Taxonomical and distributional notes on two new and a rare *Leptothorax* Mayr, 1855 species for the Hungarian ant fauna (Hymenoptera, Formicidae). In *Annales historico-naturales Musei nationalis hungarici* (Vol. 93, pp. 99-106).
- Csősz, S., Markó, B., & Galle, L. (2001). Ants (Hymenoptera: Formicidae) of Stana Valley (Romania): Evaluation of the effectiveness of a myrmecological survey. *Entomologica romanica*, 6(2002), 121-126.
- Markó B & Csősz S (2001) Nine new ant species in the Romanian fauna (Hymenoptera: Formicidae): morphology, biology, and distribution. *Entomologica romanica*, 6, 127-132.
- Gallé L, Csősz S, Tartally A, Kovács É (1998) A check-list of Hungarian ants. *Folia Entomologica Hungarica*, 59: pp. 213-220.

8.3 Az értekezéshez felhasznált egyéb publikációk listája

- Arnol'di KV (1977) Novyi i maloizvestnye vidy murav'ev roda *Leptothorax* MAYR (Hymenoptera, Formicidae) evropejskoj chasti SSSR i Kavkaza. *Entomol Oboz*, 56: 198–204.
- Bagherian A, Münch W & Seifert B (2012) First demonstration of interspecific hybridisation in *Myrmica* ants by geometric morphometrics (Hymenoptera: Formicidae). *Myrmecological News*, 17: 121–131.
- Balhoff JP, Mikó I, Yoder MJ, Mullins PL & Deans AR (2013) A semantic model for species description applied to the ensign wasps (Hymenoptera: Evaniidae) of New Caledonia. *Systematic Biology*, 62(5): 639–659.
- Baur H & Leuenberger C (2011) Analysis of ratios in multivariate morphometry. *Systematic Biology*, 60: 813–825. doi: 10.1093/sysbio/syr061 PMID: 21828084
- Beleites C, Sergo V (2015) 'hyperSpec: a package to handle hyperspectral data sets in R', R package version 0.98-20150304. <http://hyperspec.r-forge.r-project.org>
- Bernard F (1956) Révision des *Leptothorax* (Hyménoptères Formicidae) d'Europe occidentale, basée sur la biométrie et les genitalia males. *Bull Soc Zool France*, 81: 151–165.
- Borowiec L. (2014) Catalogue of ants of Europe, the Mediterranean Basin and adjacent regions (Hymenoptera: Formicidae). *Genus*, 25: 1–340.
- Brock G, Pihur V, Datta S & Datta S (2008) clValid, an R package for cluster validation. *Journal of Statistical Software*, 25: 1–22.
- Burleigh JG, Alphonse K, Alverson AJ, Bik HM, Blank C, Cirranello AL, Cui H, Daly M, Dietterich TG, Gasparich G, Irvine J, Julius M, Kaufman S, Law E, Liu J, Moore L, O'Leary MA, Passarotti M, Ranade S, Simmons NB, Stevenson DW, Thacker RW, Theriot EC, Todorovic S, Velazco PM, Walls RL, Wolfe JM & Yu M (2013) Next-generation phenomics for the Tree of Life. *PLOS Currents Tree of Life. Edition 1*. doi: 10.1371/currents.tol.085c713acafc8711b2ff7010a4b03733.
- Ceballos G, Ehrlich PR, Barnosky AD, García A, Pringle RM, Palmer TM. (2015) Accelerated modern human induced species losses: Entering the sixth mass extinction. *Science Advances*, 1: e1400253. doi: 10.1126/sciadv.1400253 PMID: 26601195.
- Cornet A (1974) Essai de cartographie bioclimatique à Madagascar. Notice Explicative. Paris: Editions ORSTOM.
- Cremer S, Ugelvig LV, Drijfhout FP, Schlick-Steiner BC, Steiner FM, Seifert B, Hughes DP, Schulz A, Petersen KS, Konrad H, Stauffer C, Kiran K, Espadaler X, D'Ettorre P, Aktaş N., Eilenberg J, Jones GR, Nash DR, Pedersen JS & Boomsma JJ (2008) The evolution of invasiveness in garden ants. *PLoS ONE*, 3:e3838.
- Czechowski W, Radchenko A, Czechowska W & Vepsäläinen K (2012) *The ants of Poland with reference to the myrmecofauna of Europe*. Fauna Poloniae 4. Warsaw: Natura Optima Dux Foundation.

- Csereháti Z (2004) Outlierek meghatározása és kezelése gazdaságstatisztikai felvételekben. *Statisztikai Szemle*, 82: 728–746.
- Day WHE & Edelsbrunner H (1984) Efficient algorithms for agglomerative hierarchical clustering method. *Journal of Classification*, 1: 7–24.
- Dayrat B (2005) Towards integrative taxonomy. *Biological journal of the Linnean society*, 85(3): 407–415.
- Derkarabetian S & Hedin M (2014) Integrative taxonomy and species delimitation in harvestmen: A revision of the western North American genus *Sclerobunus* (Opiliones: Laniatores: Tra-vunioidea) *PLoS ONE*, 9: e104982. doi: 10.1371/journal.pone.0104982 PMID: 25144370
- Ezard TH, Pearson PN & Purvis A (2010) Algorithmic approaches to aid species' delimitation in multidimensional morphospace. *BMC Evolutionary Biology*, 10: 1–11.
- Feltens R, Görner R, Kalkhof S, Gröger-Arndt H & von Bergen M (2010) Discrimination of different species from the genus *Drosophila* by intact protein profiling using matrix-assisted laser desorption ionization mass spectrometry. *BMC Evolutionary Biology*, 10: 95.
- Ferreira RS, Poteaux C, Delabie JHC, Fresneau D & Rybak F (2010) Stridulations Reveal Cryptic Speciation in Neotropical Sympatric Ants. *PLoS ONE*, 5(12): e15363. doi:10.1371/journal.pone.0015363.
- Fisher BL (2005) A model for a global inventory of ants: a case study in Madagascar. *Proceedings-California Academy of Sciences*, 56:86.
- Fokuhl G, Heinze J & Poschlod P (2012) Myrmecochory by small ants—Beneficial effects through elaiosome nutrition and seed dispersal. *Acta Oecologica*, 38: 71–76.
- Galimberti A, Spada M, Russo D, Mucedda M, Agnelli P, Crottini A, Ferri E, Martinoli A & Casiraghi M (2012) Integrated Operational Taxonomic Units (IOTUs) in Echolocating Bats: A Bridge between Molecular and Traditional Taxonomy. *PLoS ONE*, 7(6): e40122. doi:10.1371/journal.pone.0040122.
- Godfray HCJ (2002) Challenges for taxonomy. *Nature*, 417(6884): 17–19.
- Goloboff PA, Mattoni CI & Quinteros AS (2006) Continuous characters analyzed as such. *Cladistics*, 22: 589–601.
- Goropashnaya AV, Fedorov VB, Seifert B & Pamilo P (2007) Phylogeography and population structure in the ant *Formica exsecta* (Hymenoptera, Formicidae) across Eurasia as reflected by mitochondrial DNA variation and microsatellites. *Annales Zoologici Fennici*, 44: 462–474.
- Gratiashvili N, Bernadou A, Suefuji M, Seifert B & Heinze J (2014) The Caucaso-Anatolian slave-making ant *Myrmoxenus tamarae* (Arnoldi, 1968) and its more widely distributed

- congener *Myrmoxenus ravouxi* (André, 1896): a multidisciplinary comparison (Hymenoptera: Formicidae). *Organisms Diversity & Evolution*, 14(3): 259-267. doi:10.1007/s13127-014-0174-6
- Guillem RM, Drijfhout FP & Martin SJ (2016) Species-specific cuticular hydrocarbon stability within European *Myrmica* ants. *Journal of chemical ecology*, 42(10): 1052-1062. DOI: 10.1007/s10886-016-0784-x
- Hewitt GM (1999) Post-glacial re-colonization of European biota. *Biological Journal of the Linnean Society*, 68: 87–112.
- Hölldobler B & Wilson EO. The Ants. Cambridge, Mass.: Harvard University Press; 1990.
- Kaufman L & Rousseeuw PJ (1990) *Finding groups in data. An introduction to cluster analysis*. Wiley, New York
- Klingenberg CP (2011) MorphoJ: an integrated software package for geometric morphometrics. *Molecular Ecology Resources*, 11: 353–357.
- Kruskal JB (1964) Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, 29(1): 1-27.
- Kulmuni J, Seifert B & Pamilo P (2010) Segregation distortion causes large-scale differences between male and female genomes in hybrid ants. *Proceedings of the National Academy of Sciences of the United States of America*, 107(16): 7371–7376.
- Lachenbruch P. & Mickey ,M. (1968) Estimation of error rates in discriminant analysis. *Technometrics*, 10: 1-11.
- Legendre P & Legendre L (1998) *Numerical ecology*. – Elsevier, Amsterdam, 853 pp.
- Lesaffre E, Willems JL & Albert A (1989) Estimation of error rate in multiple group lo-gistic discrimination. The approximate leaving-one-out method. *Communication in Statistics: Theory and Methods*, 18: 2989-3007.
- Litvinchuk SN, Crottini A, Federici S, de Pous P, Donaire D, Andreone F, Kalezić ML, Džukić G, Lada GA, Borkin LJ & Rosanov JM (2013) Phylogeographic patterns of genetic diversity in the common spadefoot toad, *Pelobates fuscus*, reveals evolutionary history, postglacial range expansion and secondary contact. *Organisms Diversity & Evolution*, doi:10.1007/s13127-013-0127-5.
- Lloyd S. (1982) Least squares quantization in PCM. *IEEE Transactions on Information Theory*, 28: 129-137.
- Maechler M., Rousseeuw P., Struyf A., Hubert M. & Hornik K. (2012) *Cluster: cluster analysis basics and extensions*. R package version 1.14.3. <<http://CRAN.R-project.org/package=cluster>>, retrieved in October 2012.
- McFadden CS, Benayahu Y, Pante E, Thoma JN, Nevarez PA & France SC (2011) Limitations of mitochondrial gene barcoding in Octocorallia. *Molecular ecology resources*, 11(1): 19-31.

- Mikó I, Copeland R, Balhoff J, Yoder M & Deans A (2014) Folding wings like a cockroach: a review of transverse wing folding ensign wasps (Hymenoptera: Evaniidae: Afrevania and Trissevania). *PLoS ONE*, 9: e94056.
- Molet M, Wheeler DE & Peeters C (2012) Evolution of novel mosaic castes in ants: modularity, phenotypic plasticity, and colonial buffering. *The American Naturalist*, 180(3), 328-341.
- Myers N, Mittermeier RA, Mittermeier CG, Da Foneseca GA & Kent J (2000) Biodiversity hotspots and conservation priorities. *Nature* 403:853–858 DOI 10.1038/35002501.
- Nascimento DG, Rates B, Santos DM, Verano-Braga T, Barbosa-Silva A, Dutra AA, Biondi I, Martin-Eauclaire MF, De Lima ME & Pimenta AM (2006) Moving pieces in a taxonomic puzzle: venom 2D-LC/MS and data clustering analyses to infer phylogenetic relationships in some scorpions from the Buthidae family (Scorpiones). *Toxicon*, 47(6): 628–39.
- Ni Z, Kim ED, Ha M, Lackey E, Liu J, Zhang Y & Chen ZJ (2009) Altered circadian rhythms regulate growth vigour in hybrids and allopolyploids. *Nature*, 457(7227): 327-331.
- Nilsen G & Lingjaerde OC (2013) clusterGenomics: Identifying clusters in genomics data by recursive partitioning. *R package version*, 1(0).
- O'Brien SJ & MacIntyre RJ (1969) An analysis of gene-enzyme variability in natural populations of *Drosophila melanogaster* and *Drosophila simulans*. *American Naturalist*, 103: 97–114.
- Odong TL, van Heerwaarden J, Jansen J, van Hintum TJL & van Eeuwijk FA (2011) Determination of genetic structure of germplasm collections: are traditional hierarchical clustering methods appropriate for molecular marker data? *Theoretical and Applied Genetics*, 123: 195–205.
- Padial JM, Miralles A, De la Riva I & Vences M (2010) The integrative future of taxonomy. *Frontiers in zoology*, 7(1): 16.
- Pecsénye K & Meglész E (1995) Enzyme polymorphism in *Drosophila melanogaster* populations collected in two different habitats in Hungary. *Genetica*, 96(3): 257-268. doi:10.1007/BF01439580
- Pecsénye K (1989) A comparison of the level of enzyme polymorphism in cosmopolitan *Drosophila* species between populations collected in distilleries and in their surroundings in Hungary. *Genetics Selection Evolution*, 21: 147–157.
- Pogodin S, Hasan J, Baulin VA, Webb HK, Truong VK, Nguyen THP, Boshkovikj V, Fluke CJ, Watson CS, Watson JA, Crawford RJ & Ivanova EP (2013) Biophysical model of bacterial cell interactions with nanopatterned cicada wing surfaces. *Biophys J* 104: 835–840. <http://dx.doi.org/10.1016/j.bpj.2012.12.046>

- Pusch K, Heinze J & Foitzik S (2006) The influence of hybridization on colony structure in the ant species *Temnothorax nylanderi* and *T. crassispinus*. *Insectes sociaux*, 53(4): 439–445.
- R Development Core Team (2011) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.
- Radchenko AG (1995) A Review of the Ant Genus *Leptothorax* (Hymenoptera, Formicidae) of Central and East Palaearctics. Communication 3. Group *nylanderi*, *korbi*, *nassonovi* and *susamyri*. *Vestnik Zoologici*, 4: 3–11.
- Rousseeuw P, Struyf A, Hubert M & Maechler M (2012). *cluster: Cluster Analysis Extended*. R package version 1.14.2, URL <http://CRAN.R-project.org/package=cluster>.
- Rousseeuw PJ (1986). *A visual display for hierarchical classification*, in *Data Analysis and Informatics 4*; edited by E. Diday, Y. Escoufier, L. Lebart, J. Pages, Y. Schektman, and R. Tomassone. North-Holland, Amsterdam, 743–748. http://www.unesco.org/webworld/idams/advguide/Chapt7_1_4.htm.
- Salmon AB, Leonard S, Masamsetti V, Pierce A, Podlutzky AJ, Podlutzkaya N, Richardson A, Austad SN & Chaudhuri AR (2009) The long lifespan of two bat species is correlated with resistance to protein oxidation and enhanced protein homeostasis. *FASEB J*, 23: 2317–2326. <http://doi.org/10.1096/fj.08-122523>.
- Samonds KE, Godfrey LR, Ali JR, Goodman SM, Vences M, Sutherland MR, és mtsai. (2013) Imperfect Isolation: Factors and Filters Shaping Madagascar’s Extant Vertebrate Fauna. *PLoS ONE*, 8(4): e62086. <http://doi:10.1371/journal.pone.0062086>.
- Schatz GE (2000) Endemism in the Malagasy tree flora. In *Biogeography of Madagascar* (eds W.R. Lourenço and S.M. Goodman) pp. 1–9. Mémoires de la Société de Biogéographie, Paris
- Schlick-Steiner BC, Seifert B, Stauffer C, Christian E, Crozier RH, Steiner FM (2007a) Without morphology, cryptic species stay in taxonomic crypsis following discovery. *Trends in Ecology and Evolution*, 22: 391–392. <http://dx.doi.org/10.1016/j.tree.2007.05.004>.
- Schlick-Steiner BC, Steiner FM, Sanetra M, Seifert B, Christian E & Stauffer C (2007b) Lineage specific evolution of an alternative social strategy in *Tetramorium* ants (Hymenoptera: Formicidae). *Biological Journal of the Linnean Society*, 91: 247–255.
- Schlick-Steiner BC, Steiner FM, Seifert B, Stauffer C, Christian & Crozier RH (2010) Integrative Taxonomy: a multi-source approach to exploring biodiversity. *Annual Review of Entomology*, 55: 421–438. <http://10.1146/annurev-ento-112408-085432>.
- Seifert B & Goropashnaya A (2004) Ideal phenotypes and mismatching haplotypes errors of mtDNA treeing in ants (Hymenoptera: Formicidae) detected by standardized morphometry. *Organisms, Diversity & Evolution*, 4(4):295–305.

- Seifert B & Schultz R (2009a) A taxonomic revision of the *Formica rufibarbis* Fabricius, 1793 group (Hymenoptera: Formicidae). *Myrmecological News*, 12: 255–272.
- Seifert B & Schultz R (2009b) A taxonomic revision of the *Formica subpilosa* Ruzsky, 1902 group (Hymenoptera: Formicidae). *Myrmecological News*, 12: 67–83.
- Seifert B (1984) Nachweis einer im Freiland aufgetretenen Bastardierung von *Leptothorax nigriceps* Mayr und *Leptothorax unifasciatus* (Latr.) mittels einer multiplen Diskriminanzanalyse. *Abhandlungen und Berichte des Naturkundemuseums Görlitz*, 58(7): 1–8.
- Seifert B (1988) A revision of the European species of the ant subgenus *Chthonolasius*. *Entomologische Abhandlungen Museum Tierkunde Dresden*, 51: 143–180.
- Seifert B (1991) *Lasius platythorax* n.sp., a widespread sibling species of *Lasius niger*. *Entomologia Generalis*, 16: 69–81.
- Seifert B (1992) A taxonomic revision of the Palaearctic members of the ant subgenus *Lasius* s.str. (Hymenoptera: Formicidae). *Abhandlungen und Berichte des Naturkundemuseums Görlitz*, 66: 1–67.
- Seifert B (1992) A taxonomic revision of the Palaearctic members of the ant subgenus *Lasius* s.str. (Hymenoptera: Formicidae). *Abhandlungen und Berichte des Naturkundemuseums Görlitz*, 66: 1–67.
- Seifert B (1999) Interspecific hybridisations in natural populations of ants by example of a regional fauna (Hymenoptera:Formicidae). *Insectes Sociaux*, 46: 45–52.
- Seifert B (2000) A taxonomic revision of the ant subgenus *Coptoformica* Mueller, 1923. *Zoosystema* (Paris), 22(3): 517–568.
- Seifert B (2003a) The ant genus *Cardiocondyla* (Hymenoptera: Formicidae) - a taxonomic revision of the *elegans*, *bulgarica*, *batesii*, *nuda*, *shuckardi*, *stambuloffii*, *wroughtonii*, *emeryi*, and *minutior* species groups. *Annalen des Naturhistorischen Museums Wien, Serie B*, 104: 203–338.
- Seifert B (2003b) A taxonomic revision of the *Formica cinerea* group (Hymenoptera: Formicidae). *Abhandlungen und Berichte des Naturkundemuseums Görlitz*, 74(2): 245–272.
- Seifert B (2004) *Hypoponera punctatissima* (Roger) and *H. schauinslandi* (Emery) - Two morphologically and biologically distinct species (Hymenoptera: Formicidae). *Abhandlungen und Berichte des Naturkundemuseums Görlitz*, 75(1): 61–81
- Seifert B (2005) Rank elevation in two European ant species: *Myrmica lobulicornis* Nylander, 1857, stat.n. and *Myrmica spinosior* Santschi, 1931, stat.n. (Hymenoptera: Formicidae). *Myrmecologische Nachrichten*, 7: 1–7.

- Seifert B (2005) Rank elevation in two European ant species: *Myrmica lobulicornis* Nylander, 1857, stat.n. and *Myrmica spinosior* Santschi, 1931, stat.n. (Hymenoptera: Formicidae). *Myrmecologische Nachrichten*, 7: 1–7.
- Seifert B (2006a) *Temnothorax saxonicus* (Seifert 1995) stat. n., comb.n. – a parapatric, closely-related species of *T. sordidulus* (Müller 1923) comb.n. and description of two new closely-related species, *T. schoedli* sp.n. and *T. artvinense* sp.n., from Turkey (Hymenoptera: Formicidae). *Myrmecologische Nachrichten*, 8: 1–12.
- Seifert B (2006b) Social cleptogamy in the ant subgenus *Chthonolasius* – survival as a minority. *Abhandlungen und Berichte des Naturkundemuseums Görlitz*, 77: 251–276.
- Seifert B (2007) *Die Ameisen Mittel- und Nordeuropas*. Iutra, Tauer: Iutra Verlags- und Vertriebsgesellschaft, 368 pp.
- Seifert B (2008) Removal of allometric variance improves species separation in multi-character discriminant functions when species are strongly allometric and exposes diagnostic characters. *Myrmecological News*, 11: 91–105.
- Seifert B (2009) Cryptic species in ants (Hymenoptera:Formicidae) revisited: we need a change in the alpha-taxonomic approach. *Myrmecological News*, 12: 149–166.
- Seifert B (2011) A taxonomic revision of the Eurasian *Myrmica salina* species complex (Hymenoptera: Formicidae). *Soil Organisms*, 83: 169–186.
- Seifert B (2012a) Clarifying naming and identification of the outdoor species of the ant genus *Tapinoma* Förster, 1850 (Hymenoptera: Formicidae) in Europe north of the Mediterranean region with description of a new species. *Myrmecological News*, 16: 139–147.
- Seifert B (2012b) A review of the West Palaearctic species of the ant genus *Bothriomyrmex* Emery, 1869 (Hymenoptera: Formicidae). *Myrmecological News*, 17: 91–104.
- Seifert B, Kulmuni J & Pamilo P (2010) Independent hybrid populations of *Formica polyctena* X *rufa* wood ants (Hymenoptera: Formicidae) abound under conditions of forest fragmentation. *Evolutionary Ecology*, 24(5):1219–1237.
- Seifert B, Schlick-Steiner BC & Steiner FM (2009) *Myrmica constricta* Karavajev, 1934 – a cryptic sister species of *Myrmica hellenica* Finzi, 1926 (Hymenoptera: Formicidae). *Soil Organisms*, 81: 53–76.
- Sokal R & Michener C (1958) A statistical method for evaluating systematic relationships. *University of Kansas Science Bulletin*, 38: 1409–1438.
- Steiner FM, Schlick-Steiner BC, Seifert B (2009) Morphology-based taxonomy is essential to link molecular research to nomenclature. *Contributions to Natural History*, 12: 1295–1315.
- Steiner FM, Seifert B, Grasso DA, Le Moli F, Arthofer W, Stauffer C, Crozier RH & Schlick-Steiner BC (2011) Mixed colonies and hybridisation of Messor harvester ant species

- (Hymenoptera: Formicidae). *Organisms Diversity & Evolution*, 11(2): 107–134. <http://dx.doi.org/10.1007/s13127-011-0045-3>.
- Steiner FM, Seifert B, Moder K & Schlick-Steiner BC (2010) A multisource solution for a complex problem in biodiversity research: Description of the cryptic ant species *Tetramorium alpestre* sp.n. (Hymenoptera: Formicidae). *Zoologischer Anzeiger*, 249:223–254.
- Struyf A, Mia H & Rousseeuw PJ (1997) Integrating robust clustering techniques in S-PLUS. *Computational Statistics & Data Analysis*, 26: 17–37.
- Tabacchi M, Mantegazza R, Spada A & Ferrero A (2006) Morphological traits and molecular markers for classification of *Echinochloa* species from Italian rice fields. *Weed Science*, 54:1086–1093.
- Tibshirani R, Walther G, & Hastie T (2001) Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(2): 411–423.
- Udvardy M (1983) *Dinamikus állatföldrajz*. Tankönyvkiadó, Budapest, 496.
- Wachter GA, Muster C, Arthofer W, Raspotnig G, Föttinger P, Komposch C, Steiner FM & Schlick-Steiner BC (2015) Taking the discovery approach in integrative taxonomy: decrypting a complex of narrow-endemic Alpine harvestmen (Opiliones: Phalangidae: Megabunus). *Molecular ecology*, 24(4): 863–889. DOI: 10.1111/mec.13077.
- Ward JH (1963) Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association*, 48: 236–244.
- Ward PS & Sumnicht TP (2012) Molecular and morphological evidence for three sympatric species of *Leptanilla* (Hymenoptera: Formicidae) on the Greek island of Rhodes. *Myrmecological News*, 17: 5–11.
- Wiens JJ (2000) Reconstructing phylogenies from allozyme data: comparing method performance with congruence. *Biological Journal of the Linnean Society*, 70: 613–632.
- Will KW, Mishler BD & Wheeler QD (2005) The perils of DNA barcoding and the need for integrative taxonomy. *Systematic Biology*, 54: 844–851.
- Yeates DK, Seago A, Nelson L, Cameron SL, Joseph LEO & Trueman JW (2011) Integrative taxonomy, or iterative taxonomy?. *Systematic Entomology*, 36(2): 209–217.
- Yoder MJ, Miko I, Seltmann KC, Bertone MA & Deans AR (2010) A gross anatomy ontology for Hymenoptera. *PLoS One*, 5: e15991. <https://doi.org/10.1371/journal.pone.0015991>.
- Younens-Clark K, Buckler E, Casstevens T, Chen C, Declerck G, Derwent P, Dharmawardhana P, Jaiswal P, Kersey P, Karthikeyan AS, Lu J, McCouch SR, Ren L, Spooner W, Stein

JC, Thomason J, Wei S & Ware D (2011) Gramene database in 2010: updates and extensions. *Nucleic Acids Research*, 39: D1085–1094.
<https://doi.org/10.1093/nar/gkq1148>.

Zhang ZQ (2008) Contributing to the progress of descriptive taxonomy. *Zootaxa*, 1968: 65-68.

9 Mellékletek

9.1 Melléklet. A leírásokban és a kulcsokban használt morfológiai állítások terminológiája, valamint a fenotípuson alapuló diagnózisok, ontológiák URI táblázatban rögzített reprezentációja.

Character	State	Semantic Phenotype
Dorsal region of mesosoma sculpture	rugose with areolate ground sculpture	has part some (mesosoma and (has part some (dorsal region and (bearer of some rugose) and (bearer of some areolate))))
Scape length vs. absolute cephalic size (SL/CS)	0.797-0.816 (m=0.808)	has part some (scape length and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.797f$, $\leq 0.816f$])))))
Absolute cephalic size	510-627 μm (m=566, nlot=53)	(has part some absolute cephalic size) and (bearer of some (length and (is quality measured as some ((has measurement unit label value micrometer) and (has measurement value some (float [$\geq 510.0f$] and float [$\leq 566.0f$])))))
Dorsal profile of petiolar node contour line angle value to frontal profile of petiole contour line in lateral view	105-110°	has part some (dorsal petiolar profile contour line and (bearer of some (angle and (is quality measured as some ((has measurement unit label value degree) and (has measurement value some (float [$\geq 105.0f$] and float [$\leq 110.0f$]))))) and (towards some anterior profile of petiolar node contour line)))
Absolute cephalic size	1069 [958, 1189] μm (n=56)	(has part some absolute cephalic size) and (bearer of some (length and (is quality measured as some ((has measurement unit label value micrometer) and (has measurement value some (float [$\geq 958.0f$] and float [$\leq 1189.0f$])))))
Cephalic length vs. Maximum width of head capsule (CL/CWb)	1.139-1.221 (m=1.173)	has part some (maximum cephalic length in median line and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some dorsal head width))) and (has measurement value some float [$\geq 1.139f$, $\leq 1.221f$])))))
Eye length vs. absolute cephalic size (EL/CS)	0.244-0.271 (m=0.262)	has part some (maximum diameter of the compound eye and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.244f$, $\leq 0.271f$])))))
Apical spine distance vs. absolute cephalic size (SPTI/CS)	0.373-0.482 (m=0.432)	has part some (apical spine distance and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.373f$, $\leq 0.482f$])))))
Dorsal profile of petiolar node contour	72-120°	has part some (dorsal petiolar profile contour line and (bearer of some (angle and (is quality measured as

line angle value to frontal profile of petiole contour line in lateral view		some ((has measurement unit label value degree) and (has measurement value some (float [$\geq 72.0f$] and float [$\leq 120.0f$])))) and (towards some anterior profile of petiolar node contour line)))
Body color pattern	mesosoma, antenna and legs, waist and anterior region of 1st gastral tergite lighter than head and posterior region of gaster	(has part some (cranium and (bearer of some (color brightness and (similar in magnitude relative to some (color brightness and (inheres in some (posterior region and (part of some gaster))))))))) and (has part some (leg and (bearer of some (color brightness and (similar in magnitude relative to some (color brightness and (inheres in some mesosoma))))))) and (has part some (mesosoma and (bearer of some (color brightness and (increased in magnitude relative to some (color brightness and (inheres in some cranium))))))))) and (has part some (mesosoma and (bearer of some (color brightness and (similar in magnitude relative to some (color brightness and (inheres in some abdominal segment 1))))))))) and (has part some (mesosoma and (bearer of some (color brightness and (similar in magnitude relative to some (color brightness and (inheres in some antenna))))))))) and (has part some (mesosoma and (bearer of some (color brightness and (similar in magnitude relative to some (color brightness and (inheres in some (anterior region and (part of some gaster)))))))))))))))
Body color pattern	Body concolorous	(has part some body) and (bearer of some mono-colored)
Antennomere count	12	has part some (Error3 and antenna)
Scape length vs. absolute cephalic size (SL/CS)	0.764-0.805 (m=0.784)	has part some (scape length and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size)))) and (has measurement value some float [$\geq 0.764f$, $\leq 0.805f$])))))
Cephalic length vs. Maximum width of head capsule (CL/CWb)	1.145-1.233 (m=1.183)	has part some (maximum cephalic length in median line and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some dorsal head width)))) and (has measurement value some float [$\geq 1.145f$, $\leq 1.233f$])))))
Eye lenght vs. absolute cephalic size (EL/CS)	0.236-0.280 (m=0.255)	has part some (maximum diameter of the compound eye and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size)))) and (has measurement value some float [$\geq 0.236f$, $\leq 0.28f$])))))
Facial area of the scape absolute setal angle	0-15°	has part some (facial area of the scape and (has part some (seta and (bearer of some (setal angle and (is quality measured as some ((has measurement unit label value degree) and (has measurement value some (float [$\geq 0.0f$] and float [$\leq 15.0f$]))))))))
Eye lenght vs. absolute cephalic size (EL/CS)	0.237-0.262 (m=0.250)	has part some (maximum diameter of the compound eye and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size)))) and (has measurement value some float [$\geq 0.237f$, $\leq 0.262f$])))))

Dorso-caudal petiolar profile contour line in lateral view shape	convex	(has part some caudal petiolar profile) and (bearer of some convex)
Eye length vs. absolute cephalic size (EL/CS)	0.232 [0.219, 0.249]	has part some (maximum diameter of the compound eye and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.219f$, $\leq 0.249f$]))))))
Body color	brown	(has part some body) and (bearer of some brown)
Maximum spine distance vs. absolute cephalic size (SPWI/CS)	0.299-0.347 (m=0.321)	has part some (apical spine distance and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.299f$, $\leq 0.347f$]))))))
Smooth median region on frons count	absent	has part some (frons and (not (has part some (medial region and (bearer of some smooth))))
Dorsal region of petiole sculpture	ground sculpture areolate, main sculpture rugose-reticulate	has part some (abdominal segment 3 and (has part some (dorsal region and (bearer of some rugose) and (bearer of some areolate)))
Metanotal depression shape	shallow	(has part some metanotal depression) and (bearer of some increased depth)
Frontal carina distance vs. absolute cephalic size (FRS/CS)	0.313 [0.295, 0.331]	has part some (frontal carina line and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.295f$, $\leq 0.331f$]))))))
Scape length vs. absolute cephalic size (SL/CS)	0.756-0.811 (m=0.784)	has part some (scape length and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.756f$, $\leq 0.811f$]))))))
Metanotal depression count	absent	not (has part some metanotal depression)
Anterodorsal rim of petiole count	absent medially	(has part some anterodorsal rim of petiole) and (bearer of some split medially)
Spine length vs. absolute cephalic size (SPST/CS)	0.220-0.335 (m=0.276)	has part some (spine length and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.22f$, $\leq 0.335f$]))))))
Scape length vs. absolute cephalic size (SL/CS)	0.880 [0.844, 0.919]	has part some (scape length and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.844f$, $\leq 0.919f$]))))))
Eye length vs. absolute cephalic size (EL/CS)	0.237-0.280 (m=0.258)	has part some (maximum diameter of the compound eye and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.237f$, $\leq 0.28f$]))))))
Spine length vs. absolute cephalic size	0.288-0.356 (m=0.329)	has part some (spine length and (bearer of some (

(SPST/CS)		length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.288f$, $\leq 0.356f$]))))
Cephalic length vs. Maximum width of head capsule (CL/CWb)	1.079 [1.037, 1.111]	has part some (maximum cephalic length in median line and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some dorsal head width))) and (has measurement value some float [$\geq 1.037f$, $\leq 1.111f$]))))
Cephalic length vs. Maximum width of head capsule (CL/CWb)	1.147-1.214 (m=1.184)	has part some (maximum cephalic length in median line and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some dorsal head width))) and (has measurement value some float [$\geq 1.147f$, $\leq 1.214f$]))))
External area of the scape absolute setal angle	30-45°	has part some (external area of the scape and (has part some (seta and (bearer of some (setal angle and (is quality measured as some ((has measurement unit label value degree) and (has measurement value some (float [$\geq 30.0f$] and float [$\leq 45.0f$])))))))
Postocular distance vs. cephalic length (PoOc/CL)	0.391 [0.371, 0.413]	has part some (minimum spine distance and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.371f$, $\leq 0.413f$]))))
Vertex contour line in frontal view shape	straight	has part some (cranium and (has part some (dorsal side and (has part some (posterior margin and (bearer of some straight))))
Frontal carina distance vs. absolute cephalic size (FRS/CS)	0.320-0.349 (m=0.328)	has part some (frontal carina line and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.32f$, $\leq 0.349f$]))))
Postocular distance vs. cephalic length (PoOc/CL)	0.379-0.403 (m=0.390)	has part some (minimum spine distance and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.379f$, $\leq 0.403f$]))))
Maximum spine distance vs. absolute cephalic size (SPWI/CS)	0.282-0.319 (m=0.298)	has part some (apical spine distance and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.282f$, $\leq 0.319f$]))))
Postocular distance vs. cephalic length (PoOc/CL)	0.356-0.378 (m=0.368)	has part some (minimum spine distance and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.356f$, $\leq 0.378f$]))))
Mesopleuron sculpture	areolate ground sculpture superimposed by dispersed rugulae	has part some (mesopleural area of the mesonoto-metanoto-mesopecto-metapecto-propodeal complex and (bearer of some rugose) and (bearer of some areolate))
Metanotal depression count	present	has part some metanotal depression
Scape length vs. absolute cephalic size	0.776-0.813 (m=0.794)	has part some (scape length and (bearer of some (

(SL/CS)		length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.776f$, $\leq 0.813f$]))))))
Dorso-caudal petiolar profile contour line in lateral view shape	straight	(has part some caudal petiolar profile) and (bearer of some flat)
Postocular distance vs. cephalic length (PoOc/CL)	0.371-0.398 (m=0.386)	has part some (minimum spine distance and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.371f$, $\leq 0.398f$]))))))
Frontal carina distance vs. absolute cephalic size (FRS/CS)	0.315 [0.291, 0.335]	has part some (frontal carina line and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.291f$, $\leq 0.335f$]))))))
Apical spine distance vs. absolute cephalic size (SPTI/CS)	0.455 [0.386, 0.569]	has part some (apical spine distance and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.386f$, $\leq 0.569f$]))))))
Spine length vs. absolute cephalic size (SPST/CS)	0.190-0.259 (m=0.236)	has part some (spine length and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.19f$, $\leq 0.259f$]))))))
Anterodorsal rim of petiole count	present medially	(has part some anterodorsal rim of petiole) and (bearer of some undivided)
Postocular distance vs. cephalic length (PoOc/CL)	0.386-0.418 (m=0.401)	has part some (minimum spine distance and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.386f$, $\leq 0.418f$]))))))
Maximum spine distance vs. absolute cephalic size (SPWI/CS)	0.303-0.336 (m=0.314)	has part some (apical spine distance and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.303f$, $\leq 0.336f$]))))))
Maximum spine distance vs. absolute cephalic size (SPWI/CS)	0.362-0.421 (m=0.389)	has part some (apical spine distance and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.362f$, $\leq 0.421f$]))))))
Postocular sides of cranium contour frontal view shape	feebly convex	(has part some postocular contour line) and (bearer of some convex)
Spine length vs. absolute cephalic size (SPST/CS)	0.265-0.297 (m=0.280)	has part some (spine length and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.265f$, $\leq 0.297f$]))))))
Dorsal region of petiole sculpture	ground sculpture areolate, main sculpture dispersed rugose	has part some (abdominal segment 3 and (has part some (dorsal region and (bearer of some rugose) and (bearer of some areolate)))

Mellékletek

Vertex sculpture	main sculpture rugose, ground sculpture areolate	has part some (vertex and (bearer of some rugose) and (bearer of some areolate))
Absolute cephalic size	544-688 μ m (m=623, nlot=42)	(has part some absolute cephalic size) and (bearer of some (length and (is quality measured as some ((has measurement unit label value micrometer) and (has measurement value some (float [$\geq$ 544.0f] and float [$\leq$ 688.0f])))))))
Median carina of clypeus count	present	has part some median carina of clypeus
Apical spine distance vs. absolute cephalic size (SPTI/CS)	0.342-0.397 (m=0.366)	has part some (apical spine distance and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq$ 0.342f , $\leq$ 0.397f]))))))
Frontal carina distance vs. absolute cephalic size (FRS/CS)	0.329-0.360 (m=0.342)	has part some (frontal carina line and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq$ 0.329f , $\leq$ 0.36f]))))))
Maximum spine distance vs. absolute cephalic size (SPWI/CS)	0.266-0.336 (m=0.282)	has part some (apical spine distance and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq$ 0.266f , $\leq$ 0.336f]))))))
Maximum spine distance vs. absolute cephalic size (SPWI/CS)	0.305-0.342 (m=0.327)	has part some (apical spine distance and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq$ 0.305f , $\leq$ 0.342f]))))))
Apical spine distance vs. absolute cephalic size (SPTI/CS)	0.277-0.370 (m=0.320)	has part some (apical spine distance and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq$ 0.277f , $\leq$ 0.37f]))))))
Maximum spine distance vs. absolute cephalic size (SPWI/CS)	0.280-0.331 (m=0.306)	has part some (apical spine distance and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq$ 0.28f , $\leq$ 0.331f]))))))
Postocular distance vs. cephalic length (PoOc/CL)	0.363-0.390 (m=0.374)	has part some (minimum spine distance and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq$ 0.363f , $\leq$ 0.39f]))))))
Dorsal profile of petiolar node contour line angle value to frontal profile of petiole contour line in lateral view	95-100°	has part some (dorsal petiolar profile contour line and (bearer of some (angle and (is quality measured as some ((has measurement unit label value degree) and (has measurement value some (float [$\geq$ 95.0f] and float [$\leq$ 100.0f]))))) and (towards some anterior profile of petiolar node contour line)))
Dorsal profile of petiolar node contour line angle value to frontal profile of petiole contour line in lateral view	88-104°	has part some (dorsal petiolar profile contour line and (bearer of some (angle and (is quality measured as some ((has measurement unit label value degree) and (has measurement value some (float [$\geq$ 88.0f] and float [$\leq$ 104.0f]))))) and (towards some anterior profile of petiolar node contour line)))

Postocular sides of cranium contour frontal view shape	straight	(has part some postocular contour line) and (bearer of some straight)
Apical spine distance vs. absolute cephalic size (SPTI/CS)	0.291-0.326 (m=0.321)	has part some (apical spine distance and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.291f$, $\leq 0.326f$]))))))
Frontal carina distance vs. absolute cephalic size (FRS/CS)	0.335-0.375 (m=0.360)	has part some (frontal carina line and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.335f$, $\leq 0.375f$]))))))
Maximum spine distance vs. absolute cephalic size (SPWI/CS)	0.401-0.485 (m=0.438)	has part some (apical spine distance and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.401f$, $\leq 0.485f$]))))))
Genae contour from anterior view orientation	converging	has part some (torular head width and (bearer of some length) and (decreased in magnitude relative to some (length and (inheres in some ventral head width)))
Dorsal profile of petiolar node contour line angle value to frontal profile of petiole contour line in lateral view	100-110°	has part some (dorsal petiolar profile contour line and (bearer of some (angle and (is quality measured as some ((has measurement unit label value degree) and (has measurement value some (float [$\geq 100.0f$] and float [$\leq 110.0f$])))))) and (towards some anterior profile of petiolar node contour line)))
Gena sculpture	rugoso-reticulate with feeble areolate ground sculpture	has part some (gena and (bearer of some rugose) and (bearer of some areolate))
Cephalic length vs. Maximum width of head capsule (CL/CWb)	1.181-1.261 (m=1.225)	has part some (maximum cephalic length in median line and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some dorsal head width))) and (has measurement value some float [$\geq 1.181f$, $\leq 1.261f$]))))))
Frontal carina distance vs. absolute cephalic size (FRS/CS)	0.346-0.369 (m=0.356)	has part some (frontal carina line and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.346f$, $\leq 0.369f$]))))))
Absolute cephalic size	587-678 μm (m=625, nlot=20)	(has part some absolute cephalic size) and (bearer of some (length and (is quality measured as some ((has measurement unit label value micrometer) and (has measurement value some (float [$\geq 587.0f$] and float [$\leq 678.0f$]))))))
Absolute cephalic size	540-634 μm (m=582, nlot=15)	(has part some absolute cephalic size) and (bearer of some (length and (is quality measured as some ((has measurement unit label value micrometer) and (has measurement value some (float [$\geq 540.0f$] and float [$\leq 634.0f$]))))))
Smooth median region on frons count	present	has part some (frons and (has part some (medial region and (bearer of some smooth)))
Concentric carinae laterally surrounding	absent	not (phenotype_27768)

Dorsal profile of petiolar node contour line angle value to frontal profile of petiole contour line in lateral view	110-120°	has part some (dorsal petiolar profile contour line and (bearer of some (angle and (is quality measured as some ((has measurement unit label value degree) and (has measurement value some (float [$\geq 110.0f$] and float [$\leq 120.0f$]]))) and (towards some anterior profile of petiolar node contour line)))
Postocular distance vs. cephalic length (PoOc/CL)	0.374-0.408 (m=0.388)	has part some (postocular distance and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some maximum cephalic length in median line))) and (has measurement value some float [$\geq 0.374f$, $\leq 0.408f$]])))
Frontal carina distance vs. absolute cephalic size (FRS/CS)	0.363-0.379 (m=0.372)	has part some (frontal carina line and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.363f$, $\leq 0.379f$]])))
Postocular distance vs. cephalic length (PoOc/CL)	0.377-0.401 (m=0.389)	has part some (minimum spine distance and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.377f$, $\leq 0.401f$]])))
Absolute cephalic size	500-590 μm (m=546, nlot=16)	(has part some absolute cephalic size) and (bearer of some (length and (is quality measured as some ((has measurement unit label value micrometer) and (has measurement value some (float [$\geq 500.0f$] and float [$\leq 5465.0f$]])))
Lateral region of pronotum sculpture	areolate ground sculpture, main sculpture dispersed costate	has part some (pronotum and (has part some (lateral side and (bearer of some rugose) and (bearer of some areolate)))
Absolute cephalic size	488-586 μm (m=550, nlot=29)	(has part some absolute cephalic size) and (bearer of some (length and (is quality measured as some ((has measurement unit label value micrometer) and (has measurement value some (float [$\geq 488.0f$] and float [$\leq 586.0f$]])))
Maximum spine distance vs. absolute cephalic size (SPWI/CS)	0.283-0.372 (m=0.333)	has part some (apical spine distance and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.283f$, $\leq 0.372f$]])))
Eye length vs. absolute cephalic size (EL/CS)	0.260-0.266 (m=0.263)	has part some (maximum diameter of the compound eye and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.26f$, $\leq 0.266f$]])))
Vertex sculpture	main sculpture homogeneously forked costate, ground sculpture areolate	has part some (vertex and (bearer of some areolate) and (bearer of some (rugose and (has part some (rugae and (bearer of some branched))))
Cephalic length vs. Maximum width of head capsule (CL/CWb)	1.171-1.222 (m=1.203)	has part some (maximum cephalic length in median line and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some dorsal head width))) and (has measurement value some float [$\geq 1.171f$, $\leq 1.222f$]])))
Apical spine distance	0.543 [0.463, 0.607]	has part some (apical spine distance and (bearer of

Mellékletek

vs. absolute cephalic size (SPTI/CS)		some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.463f$, $\leq 0.607f$]))))
Spine length vs. absolute cephalic size (SPST/CS)	0.398 [0.355, 0.427]	has part some (spine length and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.355f$, $\leq 0.427f$]))))
Apical spine distance vs. absolute cephalic size (SPTI/CS)	0.443 [0.354, 0.504]	has part some (apical spine distance and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.354f$, $\leq 0.504f$]))))
Maximum spine distance vs. absolute cephalic size (SPWI/CS)	0.290-0.387 (m=0.334)	has part some (apical spine distance and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.29f$, $\leq 0.387f$]))))
Cephalic length vs. Maximum width of head capsule (CL/CWb)	1.121-1.196 (m=1.155)	has part some (maximum cephalic length in median line and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some dorsal head width))) and (has measurement value some float [$\geq 1.121f$, $\leq 1.196f$]))))
Postocular distance vs. cephalic length (PoOc/CL)	0.368-0.400 (m=0.380)	has part some (minimum spine distance and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.368f$, $\leq 0.4f$]))))
Frontal carina distance vs. absolute cephalic size (FRS/CS)	0.359-0.397 (m=0.374)	has part some (frontal carina line and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.359f$, $\leq 0.397f$]))))
Apical spine distance vs. absolute cephalic size (SPTI/CS)	0.298-0.339 (m=0.320)	has part some (apical spine distance and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.298f$, $\leq 0.339f$]))))
Eye length vs. absolute cephalic size (EL/CS)	0.229-0.265 (m=0.244)	has part some (maximum diameter of the compound eye and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.229f$, $\leq 0.265f$]))))
Absolute cephalic size	499-628 μm (m=556, nlot=55)	(has part some absolute cephalic size) and (bearer of some (length and (is quality measured as some ((has measurement unit label value micrometer) and (has measurement value some (float [$\geq 499.0f$] and float [$\leq 628.0f$]))))
Longitudinal carinae on medial region of frons shape	not forked	has part some (frons and (has part some (medial region and (has part some (carina and (bearer of some unbranched))))
Absolute cephalic size	474-585 μm (m=535, nlot=85)	(has part some absolute cephalic size) and (bearer of some (length and (is quality measured as some ((has

		measurement unit label value micrometer) and (has measurement value some (float [$\geq 474.0f$] and float [$\leq 585.0f$]))))
Metapleuron sculpture	fine areolate ground sculpture, superimposed by dispersed rugulae	has part some (metapleural area of the mesonoto-metanoto-mesoplecto-metaplecto-propodeal complex and (bearer of some rugose) and (bearer of some areolate))
Apical spine distance vs. absolute cephalic size (SPTI/CS)	0.300-0.369 (m=0.331)	has part some (apical spine distance and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.3f$, $\leq 0.369f$]))))
Gena contour line in frontal view shape	feebly convex	(has part some gena contour line) and (bearer of some convex)
Lateral region of pronotum sculpture	inconspicuous areolate ground sculpture, main sculpture interspersed costate	has part some (pronotum and (has part some (lateral side and (bearer of some rugose) and (bearer of some areolate)))
Eye length vs. absolute cephalic size (EL/CS)	0.245-0.268 (m=0.254)	has part some (maximum diameter of the compound eye and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.245f$, $\leq 0.268f$]))))
Cephalic length vs. Maximum width of head capsule (CL/CWb)	1.056 [0.980, 1.113]	has part some (maximum cephalic length in median line and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some dorsal head width))) and (has measurement value some float [$\geq 0.98f$, $\leq 1.113f$]))))
Apical spine distance vs. absolute cephalic size (SPTI/CS)	0.318-0.411 (m=0.371)	has part some (apical spine distance and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.318f$, $\leq 0.411f$]))))
Absolute cephalic size	533-557 μm (m=543, nlot=5)	(has part some absolute cephalic size) and (bearer of some (length and (is quality measured as some ((has measurement unit label value micrometer) and (has measurement value some (float [$\geq 533.0f$] and float [$\leq 557.0f$]))))
Cephalic length vs. Maximum width of head capsule (CL/CWb)	1.199-1.258 (m=1.228)	has part some (maximum cephalic length in median line and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some dorsal head width))) and (has measurement value some float [$\geq 1.199f$, $\leq 1.258f$]))))
Apical spine distance vs. absolute cephalic size (SPTI/CS)	0.366-0.398 (m=0.386)	has part some (apical spine distance and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.366f$, $\leq 0.398f$]))))
Frontal carina distance vs. absolute cephalic size (FRS/CS)	0.352-0.369 (m=0.361)	has part some (frontal carina line and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.352f$, $\leq 0.369f$]))))
Frontal carina distance vs. absolute	0.366-0.388 (m=0.374)	has part some (frontal carina line and (bearer of some (

Mellékletek

cephalic size (FRS/CS)		length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.366f$, $\leq 0.388f$]))))))
Postocular sides of cranium contour frontal view shape	broadly convex	(has part some postocular contour line) and (bearer of some convex)
Cephalic length vs. Maximum width of head capsule (CL/CWb)	1.196-1.293 (m=1.223)	has part some (maximum cephalic length in median line and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some dorsal head width))) and (has measurement value some float [$\geq 1.196f$, $\leq 1.293f$])))))))
Concentric carinae laterally surrounding antennal foramen count	present	has part some circumantennal carinae
Absolute cephalic size	539-719 μm (m=614, nlot=57)	(has part some absolute cephalic size) and (bearer of some (length and (is quality measured as some ((has measurement unit label value micrometer) and (has measurement value some (float [$\geq 539.0f$] and float [$\leq 719.0f$])))))))
Postocular distance vs. cephalic length (PoOc/CL)	0.390 [0.381, 0.403]	has part some (minimum spine distance and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.381f$, $\leq 0.403f$])))))))
Dorsal region of petiole sculpture	ground sculpture areolate, main sculpture absent	has part some (abdominal segment 3 and (has part some (dorsal region and (bearer of some areolate))))
Absolute cephalic size	465-516 μm (m=496, nlot=12)	(has part some absolute cephalic size) and (bearer of some (length and (is quality measured as some ((has measurement unit label value micrometer) and (has measurement value some (float [$\geq 465.0f$] and float [$\leq 516.0f$])))))))
Spine length vs. absolute cephalic size (SPST/CS)	0.232-0.282 (m=0.261)	has part some (spine length and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.232f$, $\leq 0.282f$])))))))
Ground sculpture of submedian area of clypeus	smooth	has part some (submedian area of clypeus and (bearer of some smooth))
Absolute cephalic size	1062 [974, 1142] μm (n = 48)	(has part some absolute cephalic size) and (bearer of some (length and (is quality measured as some ((has measurement unit label value micrometer) and (has measurement value some (float [$\geq 974.0f$] and float [$\leq 1142.0f$])))))))
Spine length vs. absolute cephalic size (SPST/CS)	0.227-0.288 (m=0.265)	has part some (spine length and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.227f$, $\leq 0.288f$])))))))
Dorsal region of mesosoma sculpture	fine areolate ground sculpture, superimposed by dispersed rugae	has part some (mesosoma and (has part some (dorsal region and (bearer of some rugose) and (bearer of some areolate))))

Body color pattern	head, mesosoma, waist and anterior region of 1st gastral tergite lighter than antenna and legs and posterior region of gaster	(has part some (cranium and (bearer of some (color brightness and (increased in magnitude relative to some (color brightness and (inheres in some leg)))))) and (has part some (cranium and (bearer of some (color brightness and (similar in magnitude relative to some (color brightness and (inheres in some abdominal segment 1)))))) and (has part some (cranium and (bearer of some (color brightness and (similar in magnitude relative to some (color brightness and (inheres in some mesosoma)))))) and (has part some (cranium and (bearer of some (color brightness and (similar in magnitude relative to some (color brightness and (inheres in some (anterior region and (part of some gaster)))))))) and (has part some (leg and (bearer of some (color brightness and (similar in magnitude relative to some (color brightness and (inheres in some antenna)))))) and (has part some (leg and (bearer of some (color brightness and (similar in magnitude relative to some (color brightness and (inheres in some (posterior region and (part of some gaster))))))))
Apical spine distance vs. absolute cephalic size (SPTI/CS)	0.265-0.298 (m=0.284)	has part some (apical spine distance and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.265f$, $\leq 0.298f$]))))))
Frontal carina distance vs. absolute cephalic size (FRS/CS)	0.316 [0.289, 0.333]	has part some (frontal carina line and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.289f$, $\leq 0.333f$]))))))
Spine length vs. absolute cephalic size (SPST/CS)	0.303-0.420 (m= 0.358)	has part some (spine length and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.303f$, $\leq 0.42f$]))))))
Apical spine distance vs. absolute cephalic size (SPTI/CS)	0.270-0.346 (m=0.312)	has part some (apical spine distance and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.27f$, $\leq 0.346f$]))))))
Maximum spine distance vs. absolute cephalic size (SPWI/CS)	0.296-0.350 (m=0.331)	has part some (apical spine distance and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.296f$, $\leq 0.35f$]))))))
Eye length vs. absolute cephalic size (EL/CS)	0.247-0.268 (m=0.256)	has part some (maximum diameter of the compound eye and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.247f$, $\leq 0.268f$]))))))
External area of the scape absolute setal angle	30°	has part some (external area of the scape and (has part some (seta and (bearer of some (setal angle and (is quality measured as some ((has measurement unit label value degree) and (has measurement value value

Mellékletek

		30))))))
Frontal carina distance vs. absolute cephalic size (FRS/CS)	0.362-0.399 (m=0.377)	has part some (frontal carina line and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.362f$, $\leq 0.399f$]))))))
Frontal carina distance vs. absolute cephalic size (FRS/CS)	0.347-0.380 (m=0.361)	has part some (frontal carina line and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.347f$, $\leq 0.38f$]))))))
Absolute cephalic size	523-665 μm (m=592, nlot=49)	(has part some absolute cephalic size) and (bearer of some (length and (is quality measured as some ((has measurement unit label value micrometer) and (has measurement value some (float [$\geq 523.0f$] and float [$\leq 665.0f$])))))))
Lateral region of pronotum sculpture	areolate ground sculpture, main sculpture forked costate	has part some (pronotum and (has part some (lateral side and (bearer of some rugose) and (bearer of some areolate)))
Cephalic length vs. Maximum width of head capsule (CL/CWb)	1.143-1.196 (m=1.172)	has part some (maximum cephalic length in median line and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some dorsal head width))) and (has measurement value some float [$\geq 1.143f$, $\leq 1.196f$]))))))
Scape length vs. absolute cephalic size (SL/CS)	0.807-0.832 (m=0.821)	has part some (scape length and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.807f$, $\leq 0.832f$]))))))
Maximum spine distance vs. absolute cephalic size (SPWI/CS)	0.321-0.362 (m=0.343)	has part some (apical spine distance and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.321f$, $\leq 0.362f$]))))))
Postocular sides of cranium contour frontal view shape	convex	(has part some postocular contour line) and (bearer of some convex)
Dorsal profile of petiolar node contour line in lateral view shape	widely rounded or slightly angulate area	(has part some dorsal petiolar profile) and (bearer of some convex)
Absolute cephalic size	560-670 μm (m=618, nlot=24)	(has part some absolute cephalic size) and (bearer of some (length and (is quality measured as some ((has measurement unit label value micrometer) and (has measurement value some (float [$\geq 560.0f$] and float [$\leq 670.0f$])))))))
Cephalic length vs. Maximum width of head capsule (CL/CWb)	1.152-1.242 (m=1.197)	has part some (maximum cephalic length in median line and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some dorsal head width))) and (has measurement value some float [$\geq 1.152f$, $\leq 1.242f$]))))))
Spine length vs. absolute cephalic size (SPST/CS)	0.220-0.267 (m=0.245)	has part some (spine length and (bearer of some (length and (is quality measured as some ((has measure-

		ment unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.22f$, $\leq 0.267f$]))))
Absolute cephalic size	572-696 mm (m=652, nlot=11)	(has part some absolute cephalic size) and (bearer of some (length and (is quality measured as some ((has measurement unit label value micrometer) and (has measurement value some (float [$\geq 572.0f$] and float [$\leq 696.0f$]))))
Postocular distance vs. cephalic length (PoOc/CL)	0.384-0.424 (m=0.400)	has part some (minimum spine distance and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.384f$, $\leq 0.424f$]))))
Vertex sculpture	main sculpture dispersed forked costate sculpture, ground sculpture areolate	has part some (vertex and (bearer of some areolate) and (bearer of some (rugose and (has part some (ruga and (bearer of some branched))))
Apical spine distance vs. absolute cephalic size (SPTI/CS)	0.266-0.312 (m=0.288)	has part some (apical spine distance and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.266f$, $\leq 0.312f$]))))
Postocular distance vs. cephalic length (PoOc/CL)	0.378-0.405 (m=0.374)	has part some (minimum spine distance and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.378f$, $\leq 0.405f$]))))
Ground sculpture of submedian area of clypeus	carinate	has part some (submedian area of clypeus and (has part some carina))
Postocular distance vs. cephalic length (PoOc/CL)	0.374 [0.342, 0.393]	has part some (minimum spine distance and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.342f$, $\leq 0.393f$]))))
Dorsal region of postpetiole sculpture	ground sculpture areolate, main sculpture dispersed rugose	has part some (abdominal segment 3 and (has part some (dorsal region and (bearer of some rugose) and (bearer of some areolate))))
Scape length vs. absolute cephalic size (SL/CS)	0.763-0.836 (m=0.791)	has part some (scape length and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.763f$, $\leq 0.836f$]))))
Postocular distance vs. cephalic length (PoOc/CL)	0.373-0.393 (m= 0.386)	has part some (minimum spine distance and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.373f$, $\leq 0.393f$]))))
Eye length vs. absolute cephalic size (EL/CS)	0.247-0.263 (m=0.254)	has part some (maximum diameter of the compound eye and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.247f$, $\leq 0.263f$]))))
Frontal profile of petiolar node contour line in lateral view	straight	(has part some anterior profile of petiolar node contour line) and (bearer of some straight)

Mellékletek

shape		
Postocular distance vs. cephalic length (PoOc/CL)	0.388 [0.361, 0.406]	has part some (minimum spine distance and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.361f$, $\leq 0.406f$]))))
Apical spine distance vs. absolute cephalic size (SPTI/CS)	0.332-0.395 (m=0.364)	has part some (apical spine distance and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.332f$, $\leq 0.395f$]))))
Apical spine distance vs. absolute cephalic size (SPTI/CS)	0.280-0.337 (m=0.313)	has part some (apical spine distance and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.28f$, $\leq 0.337f$]))))
Dorsal region of postpetiole sculpture	ground sculpture areolate, main sculpture absent	has part some (abdominal segment 3 and (has part some (dorsal region and (bearer of some areolate)))
Dorso-caudal petio-lar profile contour line in lateral view shape	concave	(has part some caudal petiolar profile) and (bearer of some concave)
Eye length vs. absolute cephalic size (EL/CS)	0.241 [0.225, 0.267]	has part some (maximum diameter of the compound eye and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.225f$, $\leq 0.267f$]))))
Maximum spine distance vs. absolute cephalic size (SPWI/CS)	0.311-0.392 (m=0.350)	has part some (apical spine distance and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.311f$, $\leq 0.392f$]))))
Cephalic length vs. Maximum width of head capsule (CL/CWb)	1.135-1.238 (m=1.189)	has part some (maximum cephalic length in median line and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some dorsal head width))) and (has measurement value some float [$\geq 1.135f$, $\leq 1.238f$]))))
Maximum spine distance vs. absolute cephalic size (SPWI/CS)	0.335-0.428 (m=0.391)	has part some (apical spine distance and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.335f$, $\leq 0.428f$]))))
Body color	black	(has part some body) and (bearer of some black)
Maximum spine distance vs. absolute cephalic size (SPWI/CS)	0.402-0.506 (m=0.458)	has part some (apical spine distance and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.402f$, $\leq 0.506f$]))))
Cephalic length vs. Maximum width of head capsule (CL/CWb)	1.121-1.160 (m=1.140)	has part some (maximum cephalic length in median line and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some dorsal head width))) and (has measurement value some float [$\geq 1.121f$, $\leq$

		1.16f]])))))
Frontal carina distance vs. absolute cephalic size (FRS/CS)	0.336-0.380 (m=0.356)	has part some (frontal carina line and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.336f$, $\leq 0.38f$]])))))
Scape length vs. absolute cephalic size (SL/CS)	0.775-0.830 (m=0.802)	has part some (scape length and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.775f$, $\leq 0.83f$]])))))
Vertex contour line in frontal view shape	slightly concave	has part some (cranium and (has part some (dorsal side and (has part some (posterior margin and (bearer of some concave)))))
Maximum spine distance vs. absolute cephalic size (SPWI/CS)	0.390-0.419 (m=0.409)	has part some (apical spine distance and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.39f$, $\leq 0.419f$]])))))
Cephalic length vs. Maximum width of head capsule (CL/CWb)	1.226-1.299 (m=1.266)	has part some (maximum cephalic length in median line and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some dorsal head width))) and (has measurement value some float [$\geq 1.226f$, $\leq 1.299f$]])))))
Apical spine distance vs. absolute cephalic size (SPTI/CS)	0.256-0.322 (m=0.282)	has part some (apical spine distance and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.256f$, $\leq 0.322f$]])))))
Eye length vs. absolute cephalic size (EL/CS)	0.239-0.260 (m=0.247)	has part some (maximum diameter of the compound eye and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.239f$, $\leq 0.26f$]])))))
Spine length vs. absolute cephalic size (SPST/CS)	0.205-0.299 (m=0.255)	has part some (spine length and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.205f$, $\leq 0.299f$]])))))
Dorsal profile of petiolar node contour line angle value to frontal profile of petiole contour line in lateral view	72-82°	has part some (dorsal petiolar profile contour line and (bearer of some (angle and (is quality measured as some ((has measurement unit label value degree) and (has measurement value some (float [$\geq 72.0f$] and float [$\leq 82.0f$]])) and (towards some anterior profile of petiolar node contour line))))
Frontal carina distance vs. absolute cephalic size (FRS/CS)	0.356-0.389 (m=0.370)	has part some (frontal carina line and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.356f$, $\leq 0.389f$]])))))
Postocular sides of cranium contour frontal view orientation	parallel	has part some (dorsal head width and (bearer of some (length and (similar in magnitude relative to some (length and (inheres in some postocular head width)))))

Mellékletek

Frontal carina distance vs. absolute cephalic size (FRS/CS)	0.351-0.373 (m=0.364)	has part some (frontal carina line and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.351f$, $\leq 0.373f$]))))))))
Apical spine distance vs. absolute cephalic size (SPTI/CS)	0.293-0.328 (m=0.310)	has part some (apical spine distance and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.293f$, $\leq 0.328f$]))))))))
Scape length vs. absolute cephalic size (SL/CS)	0.766-0.804 (m=0.785)	has part some (scape length and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.766f$, $\leq 0.804f$]))))))))
Postocular sides of cranium contour frontal view orientation	converging anteriorly	has part some (dorsal head width and (bearer of some (length and (decreased in magnitude relative to some (length and (inheres in some postocular head width)))))))
Postocular distance vs. cephalic length (PoOc/CL)	0.360-0.385 (m=0.368)	has part some (minimum spine distance and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.36f$, $\leq 0.385f$]))))))))
Body color pattern	mesosoma, antenna and legs, waist and anterior region of 1st gastral tergite lighter than head dorsum and posterior region of gaster	(has part some (cranium and (has part some (dorsal region and (bearer of some (color brightness and (similar in magnitude relative to some (color brightness and (inheres in some (posterior region and (part of some gaster))))))))))))) and (has part some (leg and (bearer of some (color brightness and (similar in magnitude relative to some (color brightness and (inheres in some mesosoma))))))))))) and (has part some (mesosoma and (bearer of some (color brightness and (increased in magnitude relative to some (color brightness and (inheres in some (dorsal region and (part of some head))))))))))) and (has part some (mesosoma and (bearer of some (color brightness and (similar in magnitude relative to some (color brightness and (inheres in some abdominal segment 1))))))))))) and (has part some (mesosoma and (bearer of some (color brightness and (similar in magnitude relative to some (color brightness and (inheres in some antenna))))))))))) and (has part some (mesosoma and (bearer of some (color brightness and (similar in magnitude relative to some (color brightness and (inheres in some (anterior region and (part of some gaster)))))))))))))
Vertex contour line in frontal view shape	feebly convex	has part some (cranium and (has part some (dorsal side and (has part some (posterior margin and (bearer of some convex)))))))
Dorsal profile of petiolar node contour line angle value to frontal profile of petiole contour line in lateral view	95-105°	has part some (dorsal petiolar profile contour line and (bearer of some (angle and (is quality measured as some ((has measurement unit label value degree) and (has measurement value some (float [$\geq 95.0f$] and float [$\leq 105.0f$])))))) and (towards some anterior profile of petiolar node contour line)))

Vertex sculpture	main sculpture absent, ground sculpture areolate	(has part some vertex) and (bearer of some areolate)
Spine length vs. absolute cephalic size (SPST/CS)	0.385 [0.333, 0.437]	has part some (spine length and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.333f$, $\leq 0.437f$]))))))
Scape length vs. absolute cephalic size (SL/CS)	0.895 [0.861, 0.927]	has part some (scape length and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.861f$, $\leq 0.927f$]))))))
Eye length vs. absolute cephalic size (EL/CS)	0.257-0.274 (m=0.265)	has part some (maximum diameter of the compound eye and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.257f$, $\leq 0.274f$]))))))
Cephalic length vs. Maximum width of head capsule (CL/CWb)	1.193-1.254 (m=1.224)	has part some (maximum cephalic length in median line and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some dorsal head width))) and (has measurement value some float [$\geq 1.193f$, $\leq 1.254f$]))))))
Body color pattern	mesosoma, antenna and legs excluding femora, waist and anterior region of 1st gastral tergite lighter than head, femora and posterior region of gaster	(has part some (cranium and (bearer of some (color brightness and (similar in magnitude relative to some (color brightness and (inheres in some femur)))))) and (has part some (cranium and (bearer of some (color brightness and (similar in magnitude relative to some (color brightness and (inheres in some (posterior region and (part of some gaster))))))))) and (has part some (mesosoma and (bearer of some (color brightness and (increased in magnitude relative to some (color brightness and (inheres in some cranium))))))) and (has part some (mesosoma and (bearer of some (color brightness and (similar in magnitude relative to some (color brightness and (inheres in some abdominal segment 1))))))) and (has part some (mesosoma and (bearer of some (color brightness and (similar in magnitude relative to some (color brightness and (inheres in some antenna))))))) and (has part some (mesosoma and (bearer of some (color brightness and (similar in magnitude relative to some (color brightness and (inheres in some (anterior region and (part of some gaster))))))))) and (has part some ((not (femur)) and (part of some (leg and (bearer of some (color brightness and (similar in magnitude relative to some (color brightness and (inheres in some mesosoma)))))))))))
Scape length vs. absolute cephalic size (SL/CS)	0.763-0.809 (m=0.785)	has part some (scape length and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.763f$, $\leq 0.809f$]))))))
Body color	yellow	(has part some body) and (bearer of some yellow)
Body color pattern	head, mesosoma, antenna	has part some (propodeal spine and (has part some (

Mellékletek

	and legs excluding femora, waist and anterior region of 1st gastral tergite lighter than femora and posterior region of gaster	median anatomical line and (bearer of some (angle and (is quality measured as some ((has measurement unit label value degree) and (has measurement value some (float [$\geq 20.0f$] and float [$\leq 25.0f$])))) and (towards some Weber length))))
Scape length vs. absolute cephalic size (SL/CS)	0.757-0.798 (m=0.777)	has part some (scape length and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.757f$, $\leq 0.798f$]))))
Eye length vs. absolute cephalic size (EL/CS)	0.238-0.270 (m=0.251)	has part some (maximum diameter of the compound eye and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.238f$, $\leq 0.27f$]))))
Postocular distance vs. cephalic length (PoOc/CL)	0.373-0.400 (m=0.384)	has part some (minimum spine distance and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.373f$, $\leq 0.4f$]))))
Dorsal profile of petiolar node contour line in lateral view shape	slightly convex	(has part some dorsal petiolar profile) and (bearer of some convex)
Absolute cephalic size	594-657 μm (m=626, nlot=8)	(has part some absolute cephalic size) and (bearer of some (length and (is quality measured as some ((has measurement unit label value micrometer) and (has measurement value some (float [$\geq 594.0f$] and float [$\leq 657.0f$]))))
Longitudinal carinae on median region of frons count	absent	has part some (frons and (has part some (medial region and (not (has part some carina))))
Scape length vs. absolute cephalic size (SL/CS)	0.781-0.824 (m=0.799)	has part some (scape length and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.781f$, $\leq 0.824f$]))))
Median region of antennal rim vs. frontal carina in frontal view structure	not fully overlapped by frontal carina	(has part some torulus) and (has part some (medial region and (not (adjacent to some frontal carina))))
Spine length vs. absolute cephalic size (SPST/CS)	0.332-0.369 (m=355)	has part some (spine length and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.332f$, $\leq 0.369f$]))))
Spine length vs. absolute cephalic size (SPST/CS)	0.324-0.377 (m=0.346)	has part some (spine length and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.324f$, $\leq 0.377f$]))))
Dorsal profile of petiolar node contour line angle value to	90-100°	has part some (dorsal petiolar profile contour line and (bearer of some (angle and (is quality measured as some ((has measurement unit label value degree) and (

frontal profile of petiole contour line in lateral view		has measurement value some (float [$\geq 90.0f$] and float [$\leq 100.0f$])))) and (towards some anterior profile of petiolar node contour line))))
Eye length vs. absolute cephalic size (EL/CS)	0.228-0.268 (m=0.249)	has part some (maximum diameter of the compound eye and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.228f$, $\leq 0.268f$])))))))
Postocular sides of cranium contour frontal view shape	strongly convex	(has part some postocular contour line) and (bearer of some convex)
Eye length vs. absolute cephalic size (EL/CS)	0.247-0.276 (m=0.262)	has part some (maximum diameter of the compound eye and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.247f$, $\leq 0.276f$])))))))
Antenna color pattern	clava concolorous funicle	(has part some antenna) and (bearer of some mono-colored)
Apical spine distance vs. absolute cephalic size (SPTI/CS)	0.381-0.462 (m=0.413)	has part some (apical spine distance and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.381f$, $\leq 0.462f$])))))))
Body color pattern	head, mesosoma, waist and anterior region of 1st gastral tergite lighter antenna and legs except femora lighter than femora and posterior region of gaster	((has part some cranium) and (has part some ((not (femur)) and (part of some (leg and (bearer of some (color brightness and (similar in magnitude relative to some (color brightness and (inheres in some (posterior region and (part of some gaster))))))))))) and (bearer of some (color brightness and (similar in magnitude relative to some (color brightness and (inheres in some ((not (femur)) and (part of some leg)))))))) and ((has part some cranium) and (bearer of some (color brightness and (increased in magnitude relative to some (color brightness and (inheres in some femur)))))) and ((has part some cranium) and (bearer of some (color brightness and (similar in magnitude relative to some (color brightness and (inheres in some abdominal segment 1)))))) and ((has part some cranium) and (bearer of some (color brightness and (similar in magnitude relative to some (color brightness and (inheres in some antenna)))))) and ((has part some cranium) and (bearer of some (color brightness and (similar in magnitude relative to some (color brightness and (inheres in some mesosoma)))))) and ((has part some cranium) and (bearer of some (color brightness and (similar in magnitude relative to some (color brightness and (inheres in some (anterior region and (part of some gaster))))))))
Scape length vs. absolute cephalic size (SL/CS)	0.758-0.795 (m=0.775)	has part some (scape length and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.758f$, $\leq 0.795f$])))))))

Mellékletek

Spine length vs. absolute cephalic size (SPST/CS)	0.300 [0.258, 0.330]	has part some (spine length and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.258f$, $\leq 0.33f$]))))))))
Apical spine distance vs. absolute cephalic size (SPTI/CS)	0.261-0.318 (m=0.290)	has part some (apical spine distance and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.261f$, $\leq 0.318f$])))))))
Dorso-caudal petiolar profile contour line in lateral view shape	strongly convex	(has part some caudal petiolar profile) and (bearer of some convex)
Gena contour line in frontal view shape	convex	(has part some gena contour line) and (bearer of some convex)
Scape length vs. absolute cephalic size (SL/CS)	0.907 [0.849, 0.997]	has part some (scape length and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.849f$, $\leq 0.997f$])))))))
Gena sculpture	rugoso-reticulate with areolate ground sculpture	has part some (gena and (bearer of some rugose) and (bearer of some areolate))
Eye length vs. absolute cephalic size (EL/CS)	0.245-0.270 (m=0.263)	has part some (maximum diameter of the compound eye and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.245f$, $\leq 0.27f$])))))))
Spine length vs. absolute cephalic size (SPST/CS)	0.391-0.429 (m=0.411)	has part some (spine length and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.391f$, $\leq 0.429f$])))))))
Cephalic length vs. Maximum width of head capsule (CL/CWb)	1.169-1.253 (m=1.206)	has part some (maximum cephalic length in median line and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some dorsal head width))) and (has measurement value some float [$\geq 1.169f$, $\leq 1.253f$])))))))
Lateral carinae of clypeus count	present	has part some lateral carina of clypeus
Absolute cephalic size	505-635 μm (m=571, nlot=13)	(has part some absolute cephalic size) and (bearer of some (length and (is quality measured as some ((has measurement unit label value micrometer) and (has measurement value some (float [$\geq 505.0f$] and float [$\leq 635.0f$]))))))
Dorsal profile of petiolar node contour line in lateral view shape	slightly convex to conspicuously rounded	(has part some dorsal petiolar profile) and (bearer of some convex)
Scape length vs. absolute cephalic size (SL/CS)	0.763-0.809 (m=0.787)	has part some (scape length and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.763f$, $\leq 0.809f$])))))))

Mesopleuron sculpture	fine areolate ground sculpture, superimposed by dispersed rugulae	has part some (mesopleural area of the mesonoto-metanoto-mesopecto-metapecto-propodeal complex and (bearer of some rugose) and (bearer of some areolate))
Longitudinal carinae on median region of frons count	present	has part some (frons and (has part some (medial region and (has part some carina))))
Cephalic length vs. Maximum width of head capsule (CL/CWb)	1.100-1.180 (m=1.140)	has part some (maximum cephalic length in median line and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some dorsal head width))) and (has measurement value some float [$\geq 1.1f$, $\leq 1.18f$])))))
Scape length vs. absolute cephalic size (SL/CS)	0.784-0.817 (m= 0.803)	has part some (scape length and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.784f$, $\leq 0.817f$])))))
Spine length vs. absolute cephalic size (SPST/CS)	0.278-0.331 (m=0.306)	has part some (spine length and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.278f$, $\leq 0.331f$])))))
Dorsal region of mesosoma sculpture	rugulose with areolate ground sculpture	has part some (mesosoma and (has part some (dorsal region and (bearer of some rugose) and (bearer of some areolate)))
Spine length vs. absolute cephalic size (SPST/CS)	0.220-0.249 (m=0.237)	has part some (spine length and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.22f$, $\leq 0.249f$])))))
Spine length vs. absolute cephalic size (SPST/CS)	0.227-0.281 (m=0.258)	has part some (spine length and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.227f$, $\leq 0.281f$])))))
Spine length vs. absolute cephalic size (SPST/CS)	0.159-0.230 (m=0.192)	has part some (spine length and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.159f$, $\leq 0.23f$])))))
Maximum spine distance vs. absolute cephalic size (SPWI/CS)	0.353-0.415 (m=0.384)	has part some (apical spine distance and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.353f$, $\leq 0.415f$])))))
Antennomere count	11	has part some (Error4 and antenna)
Dorsal profile of petiolar node contour line angle value to frontal profile of petiole contour line in lateral view	100-115°	has part some (dorsal petiolar profile contour line and (bearer of some (angle and (is quality measured as some ((has measurement unit label value degree) and (has measurement value some (float [$\geq 100.0f$] and float [$\leq 115.0f$])))) and (towards some anterior profile of petiolar node contour line)))
Spine length vs. absolute cephalic size (SPST/CS)	0.200-0.264 (m= 0.231)	has part some (spine length and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.2f$, $\leq 0.264f$])))))

Mellékletek

Apical spine distance vs. absolute cephalic size (SPTI/CS)	0.307 [0.221, 0.361]	has part some (apical spine distance and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.221f$, $\leq 0.361f$])))))))
Eye length vs. absolute cephalic size (EL/CS)	0.239 [0.220, 0.265]	has part some (maximum diameter of the compound eye and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.22f$, $\leq 0.265f$])))))))
Scape length vs. absolute cephalic size (SL/CS)	0.763-0.796 (m=0.778)	has part some (scape length and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.763f$, $\leq 0.796f$])))))))
Postocular sides of cranium contour frontal view orientation	converging posteriorly	has part some (dorsal head width and (bearer of some (length and (increased in magnitude relative to some (length and (inheres in some postocular head width)))))))
Apical spine distance vs. absolute cephalic size (SPTI/CS)	0.289-0.322 (m=0.300)	has part some (apical spine distance and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.289f$, $\leq 0.322f$])))))))
Vertex sculpture	main sculpture dispersed forked costate, ground sculpture inconspicuous areolate	has part some (vertex and (bearer of some areolate) and (bearer of some (rugose and (has part some (ruga and (bearer of some branched)))))))
Frontal carina distance vs. absolute cephalic size (FRS/CS)	0.353-0.376 (m=0.361)	has part some (frontal carina line and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.353f$, $\leq 0.376f$])))))))
Scape length vs. absolute cephalic size (SL/CS)	0.758-0.808 (m=0.783)	has part some (scape length and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.758f$, $\leq 0.808f$])))))))
Cephalic length vs. Maximum width of head capsule (CL/CWb)	1.206-1.263 (m=1.240)	has part some (maximum cephalic length in median line and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some dorsal head width))) and (has measurement value some float [$\geq 1.206f$, $\leq 1.263f$])))))))
Postocular distance vs. cephalic length (PoOc/CL)	0.392-0.413 (m=0.405)	has part some (minimum spine distance and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.392f$, $\leq 0.413f$])))))))
Cephalic length vs. Maximum width of head capsule (CL/CWb)	1.192-1.278 (m=1.238)	has part some (maximum cephalic length in median line and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some dorsal head width))) and (has measurement value some float [$\geq 1.192f$, $\leq 1.278f$])))))))

Scape length vs. absolute cephalic size (SL/CS)	0.927 [0.907, 0.948]	has part some (scape length and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.907f$, $\leq 0.948f$]))))))
Dorsal profile of petiolar node contour line in lateral view shape	strongly convex	(has part some dorsal petiolar profile) and (bearer of some convex)
Cephalic length vs. Maximum width of head capsule (CL/CWb)	1.038 [0.993, 1.075]	has part some (maximum cephalic length in median line and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some dorsal head width))) and (has measurement value some float [$\geq 0.993f$, $\leq 1.075f$]))))))
Postocular distance vs. cephalic length (PoOc/CL)	0.383-0.403 (m=0.396)	has part some (minimum spine distance and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.383f$, $\leq 0.403f$]))))))
Vertex sculpture	main sculpture parallelly costate, ground sculpture areolate	has part some (vertex and (bearer of some areolate) and (bearer of some (rugose and (has part some (ruga and (bearer of some (not (branched)))))))))
Postocular distance vs. cephalic length (PoOc/CL)	0.375-0.389 (0.384)	has part some (minimum spine distance and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.375f$, $\leq 0.389f$]))))))
Dorsal region of mesosoma sculpture	areolate ground sculpture, superimposed by dispersed rugae	has part some (mesosoma and (has part some (dorsal region and (bearer of some rugose) and (bearer of some areolate)))
Vertex sculpture	main sculpture rugoso-reticulate, ground sculpture areolate	has part some (vertex and (bearer of some rugose) and (bearer of some areolate))
Eye length vs. absolute cephalic size (EL/CS)	0.258-0.280 (m=0.266)	has part some (maximum diameter of the compound eye and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.258f$, $\leq 0.28f$]))))))
Spine length vs. absolute cephalic size (SPST/CS)	0.397 [0.367, 0.432]	has part some (spine length and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.367f$, $\leq 0.432f$]))))))
Frontal profile of petiolar node contour line in lateral view shape	concave	(has part some anterior profile of petiolar node contour line) and (bearer of some concave)
Scape length vs. absolute cephalic size (SL/CS)	0.735-0.810 (m=0.782)	has part some (scape length and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.735f$, $\leq 0.81f$]))))))
Dorsal profile of petiolar node contour	concave	(has part some dorsal petiolar profile) and (bearer of some concave)

Mellékletek

line in lateral view shape		
Maximum spine distance vs. absolute cephalic size (SPWI/CS)	0.272-0.339 (m=0.304)	has part some (apical spine distance and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.272f$, $\leq 0.339f$]))))))
Eye length vs. absolute cephalic size (EL/CS)	0.230 [0.212, 0.248]	has part some (maximum diameter of the compound eye and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.212f$, $\leq 0.248f$]))))))
Frontal profile of petiolar node contour line in lateral view shape	convex	(has part some anterior profile of petiolar node contour line) and (bearer of some convex)
Dorsal profile of petiolar node contour line angle value to frontal profile of petiole contour line in lateral view	105-115°	has part some (dorsal petiolar profile contour line and (bearer of some (angle and (is quality measured as some ((has measurement unit label value degree) and (has measurement value some (float [$\geq 105.0f$] and float [$\leq 115.0f$])))) and (towards some anterior profile of petiolar node contour line)))
Metapleuron sculpture	areolate ground sculpture superimposed by dispersed rugulae	has part some (metapleural area of the mesonoto-metanoto-mesopecto-metapecto-propodeal complex and (bearer of some rugose) and (bearer of some areolate))
Body color pattern	Body concolorous, only clava darker	(has part some ((not (clava)) and (part of some (body and (bearer of some color brightness) and (increased in magnitude relative to some (color brightness and (inheres in some clava)))))) and (has part some ((not (clava)) and (part of some (body and (bearer of some mono-colored)))))
Longitudinal carinae on medial region of frons shape	forked	has part some (frons and (has part some (medial region and (has part some (carina and (bearer of some branched))))))
Spine length vs. absolute cephalic size (SPST/CS)	0.253-0.322 (m=0.289)	has part some (spine length and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.253f$, $\leq 0.322f$]))))))
Mesopleuron sculpture	areolate ground sculpture, superimposed by dispersed rugae	has part some (mesopleural area of the mesonoto-metanoto-mesopecto-metapecto-propodeal complex and (bearer of some rugose) and (bearer of some areolate))
Scape length vs. absolute cephalic size (SL/CS)	0.767-0.832 (m=0.796)	has part some (scape length and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.767f$, $\leq 0.832f$]))))))
Metapleuron sculpture	areolate ground sculpture, superimposed by dispersed rugae	has part some (metapleural area of the mesonoto-metanoto-mesopecto-metapecto-propodeal complex and (bearer of some rugose) and (bearer of some areolate))
Frontal carina distance vs. absolute cephalic size (FRS/CS)	0.357-0.375 (m=0.367)	has part some (frontal carina line and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.357f$, $\leq 0.375f$]))))))

External area of the scape absolute setal angle	10-40	has part some (external area of the scape and (has part some (seta and (bearer of some (setal angle and (is quality measured as some ((has measurement unit label value degree) and (has measurement value some (float [$\geq 10.0f$] and float [$\leq 40.0f$])))))))
Frontal carina distance vs. absolute cephalic size (FRS/CS)	0.335-0.373 (m=0.357)	has part some (frontal carina line and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.335f$, $\leq 0.373f$])))))))
Frontal carina distance vs. absolute cephalic size (FRS/CS)	0.354-0.383 (m=0.373)	has part some (frontal carina line and (bearer of some (length and (is quality measured as some ((has measurement unit label some (length and (inheres in some absolute cephalic size))) and (has measurement value some float [$\geq 0.354f$, $\leq 0.383f$])))))))

9.2 Melléklet. A Semantic phenotype approach gyakorlati alkalmazásának bemutatása.

A melléklet a szemantikus fenotípus megközelítés alkalmazását és szerkezetét mutatja be, amit cikkeimben alkalmaztam. A bemutatott részlet a Csősz és mtsai (2015) cikk alapján. A szemantikus algoritmus nyelve angol, ezért mellékletben szereplő rész itt is fordítás nélkül szerepel.

We represent natural language phenotypes in an EQ format: Entity attribute: value (S6 Table). Semantic statements of natural language phenotypes (S6 Table) were composed in Protégé 5.0 (<http://protege.stanford.edu/>) using the OWL Manchester syntax (<http://www.w3.org/TR/owl2-manchester-syntax/>) following [49] and [50]. The full data set, represented in OWL (Web Ontology Language; <http://www.w3.org/TR/owl2-overview/> last accessed February 4, 2014), was deposited as a Resource Description Framework (RDF)-XML file (<http://www.w3.org/TR/REC-rdf-syntax/> in the Dryad repository (<http://dx.doi.org/10.5061/dryad.xxxx>).

Hereby we propose two new template EQ expression as an expansion of the semantic statement types described by Balhoff et al. [49] and Mikó et al. [50]:

Angle between anatomical lines

Angles between anatomical lines are often used in taxonomic treatments of Formicidae:

NL: Dorsal profile of petiolar node contour line angle value to frontal profile of petiole contour line in lateral view: 72–82°

We expressed angle value between anatomical lines using PATO's [angle](#) and UO's [degree](#) with the combination of our previously proposed relative and absolute measurement phenotypes:

SP: 'has part' some ('dorsal petiolar profile contour line' and (bearer_of some (angle and ('is quality measured as' some (('has measurement unit label' value degree) and ('has measurement value' some (float[>= 72.0f] and float[<= 82.0f])))) and (towards some 'anterior profile of petiolar node contour line'))))

Coloration

We present color hue and intensity as separated entity attributes:

NL: Body color: brown

SP: 'has part' some (body and ('bearer of' some brown))

and

NL: Body color pattern: mesosoma, antenna and legs excluding femora, waist and anterior region of 1st gastral tergite lighter than head, femora and posterior region of gaster.

SP: ('has part' some (cranium and ('bearer of' some ('color brightness' and (similar_in_magnitude_relative_to some ('color brightness' and ('inheres in' some femur)))))) and ('has part' some (cranium and ('bearer of' some ('color brightness' and (similar_in_magnitude_relative_to some ('color brightness' and ('inheres in' some ('posterior region' and ('part of' some gaster)))))))) and ('has part' some (mesosoma and ('bearer of' some ('color brightness' and (increased_in_magnitude_relative_to some ('color brightness' and ('inheres in' some cranium)))))) and ('has part' some (mesosoma and ('bearer of' some ('color brightness' and (similar_in_magnitude_relative_to some ('color brightness' and ('inheres in' some 'abdominal segment 1')))))) and ('has part' some (mesosoma and ('bearer of' some ('color brightness' and (similar_in_magnitude_relative_to some ('color brightness' and ('inheres in' some antenna)))))) and ('has part' some (mesosoma and ('bearer of' some ('color brightness' and (similar_in_magnitude_relative_to some ('color brightness' and ('inheres in' some ('anterior region' and ('part of' some gaster)))))))) and ('has part' some ((not (femur)) and ('part of' some (leg and ('bearer of' some ('color brightness' and (similar_in_magnitude_relative_to some ('color brightness' and ('inheres in' some mesosoma))))))))

9.3 Melléklet. Az 5.3. fejezetben bemutatott integratív taxonómiai munkák molekuláris filogenetikai módszertanának kivonata.

A bemutatott részlet a (Csősz és mtsai. (2014) *Organisms Diversity & Evolution*, 14, 75-88.) munkából származik.

Mitochondrial DNA We isolated DNA from workers of colonies of *Temnothorax lichtensteini* from France, Croatia, and Greece and several other species of *Temnothorax* using the CTAB method including digestion with proteinase K (Sambrook and Russell 2001) and amplified a 658 bp fragment of the CO I gene using the primers LCO1490 and HCO2198 (Folmer és mtsai. 1994). Other workers from the same colonies were investigated by morphometry (see below).

PCR was carried out with GO-Taq Hot Start Master Mix from Promega (Madison, WI). The final primer concentration was 0.7 μ M and 1 μ l DNA was added. PCR conditions consisted of an initial denaturation at 94 °C for 240 s, 38 cycles of denaturation at 94 °C, 45 s; annealing at 50 °C, 45 s; elongation at 72 °C, 1 min; and a final step of 72 °C, 300 s. For sequencing we used an ABI 3100 capillary sequencer (Applied Biosystems, Foster City, CA). Sequences were aligned using Bioedit (Hall 1999) and ambiguous stretches at the 5' and 3' ends of the fragments were removed, resulting in fragments of 652 bp for further analysis. The models underlying the Bayesian analysis of the data were estimated using MrModeltest GUI (available under <http://genedrift.org>) in conjuncture with PAUP 4.0b10 (Swofford 1998), using a sequence from *T. cf. korbi* from Nur Dağları, Turkey, as an out-group. The analysis was performed with MrBayes 3.2.1 (Ronquist és mtsai. 2012) using HKY + I + G as a model and with the default number of four Markov chains (three heated, one cold) and the default heating parameter set at 0.2 with an MCMC length of 3,000,000 generations. The first 500 generations (burn-in) were discarded and every 100th generation was sampled. The phylogenetic tree was drawn using FigTree version 1.2.2 (available at <http://tree.bio.ed.ac.uk/software>). GenBank accession numbers are given in Table 2. We checked the most divergent sequence (PELO-GRE:Profitis-Ilias-20110501-336) visually for inversions, deletions, or stop codons to exclude that it represents a pseudogene. Its amino acid sequence showed 95 % homology to published sequences of CO I from ants in GenBank.

9.4 Melléklet. Az 5.3. fejezetben bemutatott integratív taxonómiai munkákban használt morfológiai karakter felvételek pontos protokolljának kivonata.

A bemutatott részlet a (Csósz, Mikó & Heinze (2015) munkából származik.

Protocol of morphometric character recording

Morphometric characters are defined as in [48]. All measurements were made in μm using a pin-holding stage, permitting rotations around X, Y, and Z axes. An Olympus SZX9 stereomicroscope was used at a magnification of x150 for each character, allowing an estimated precision of $\pm 2 \mu\text{m}$. Morphometric data (in μm) are provided in S3 Table . All measurements were made by the first author. Measured characters are defined and abbreviated as follows:

CL: maximum cephalic length in median line; the head must be carefully tilted to the position with the true maximum. Excavations of hind vertex and/or clypeus reduce CL (Fig. 1A).

CS: cephalic size; the arithmetic mean of CL and CWb.

CWb: maximum width of head capsule, measured just posterior to the eyes (Fig. 1A).

EL: maximum diameter of the eye.

FRS: distance of the frontal carinae immediately caudal of the posterior intersection points between frontal carinae and the lamellae dorsal to the torulus. If these dorsal lamellae do not laterally surpass the frontal carinae, the deepest point of scape corner pits may be taken as reference line. These pits take up the inner corner of scape base when the scape is fully switched caudad and produce a dark triangular shadow in the lateral frontal lobes immediately posterior to the dorsal lamellae of scape joint capsule (Fig. 1B).

ML: mesosoma length from caudalmost point of propodeal lobe to transition point between anterior pronotal slope and anterior propodeal shield (preferentially measured in lateral view; if the transition point is not well defined, use dorsal view and take the center of the dark-shaded borderline between pronotal slope and pronotal shield as anterior reference point) (Fig. 1D).

MW: maximum mesosoma width; this is in workers pronotal width (Fig. 1C).

NOdL: anterior length of petiole measured in dorsal view. Distance from the centre of anteriormost seta pit on the petiolar node to the level of the constriction of articulation condyle with propodeum. Measuring requires a change of focus from above (seta pit) to below (constriction). Dorsal view is achieved when the dorsalmost point of anterior petiolar peduncle at the level of its strongest constriction and the dorsalmost point of caudal petiolar margin are in the same focal level (Fig. 1F).

NOH: maximum height of the petiolar node, measured in lateral view from the uppermost point of the petiolar node perpendicular to a reference line set from the petiolar spiracle to the imaginary midpoint of the transition between dorso-caudal slope and dorsal profile of caudal cylinder of the petiole (Fig. 1E).

NOL: length of the petiolar node, measured in lateral view from petiolar spiracle to dorso-caudal corner of caudal cylinder. Do not erroneously take as reference point the dorso-caudal corner of the helcium, which is sometimes visible (Fig. 1E).

PEH: maximum petiole height. The chord of ventral petiolar profile at node level is the reference line perpendicular to which the maximum height of petiole is measured (Fig. 1D).

PEW: maximum width of petiole (Fig. 1C).

PL: total petiole length measured in dorsal view; distance between the dorsalmost point of caudal petiolar margin and the dorsalmost point of anterior petiolar peduncle at the transversal level of its strongest constriction. Positioning of petiole as in NOdL (Fig. 1F).

PoOC: postocular distance. Use a cross-scaled ocular micrometer and adjust the head to the measuring position of CL. Caudal measuring point: median occipital margin; frontal measuring point: median head at the level of the posterior eye margin (Fig. 1A).

PPH: maximum height of the postpetiole in lateral view measured perpendicularly to a line defined by the linear section of the segment border between dorsal and ventral petiolar sclerite (Fig. 1D).

PPL: maximum length of the postpetiole measured in lateral view perpendicular to the straight section of lateral postpetiolar margin (Fig. 1E).

PPW: maximum width of postpetiole (Fig. 1C).

SL: maximum straight line scape length excluding the articular condyle (Fig. 1A).

SPL: minimum distance between the center of propodeal spiracle and the subspinal excavation measured in lateral view (i.e. the same view that is applied to measure ML). Note: in lateral view propodeal spiracle and the caudal margin of propodeal declivity might not be in the same focal level, hence slight adjust might be necessary while measuring SPL between the two endpoints (Fig. 1E).

SPBA: the smallest distance of the lateral margins of the spines at their base. This should be measured in dorsofrontal view, since the wider parts of the ventral propodeum do not interfere with the measurement in this position. If the lateral margins of spines diverge continuously from the tip to the base, a smallest distance at base is not defined. In this case, SPBA is measured at the level of the bottom of the interspinal meniscus (Fig. 1C).

SPST: distance between the center of propodeal stigma and spine tip. The stigma centre refers to the midpoint defined by the outer cuticular ring but not to the centre of real stigma opening that may be positioned eccentrically (Fig. 1E).

SPTI: the distance of spine tips in dorsal view; if spine tips are rounded or truncated, the centers of spine tips are taken as reference points (Fig. 1C).

SPWI: maximum distance between outer margins of spines; measured in same position as SPBA (Fig. 1C).

The **deviation of propodeal spines from longitudinal mesosomal axis** (i.e. angle between the spines and the mesosomal axis) is often of high diagnostic value within the *Temnothorax nylanderi* species-group. Longitudinal mesosomal axis is defined by Seifert és mtsai. [39] as follows: *in lateral view as straight line from the center of propodeal lobe to the border point between anterior pronotal shield and propleuron.*

9.5 Melléklet. A NC-PART klaszterezés és a 'mark.dendrogram' alkalmazás R scriptje.

A script a (Csósz & Fisher (2016a) PeerJ 4:e1796. 35 p., <https://doi.org/10.7717/peerj.1796>; Csósz & Fisher (2016b) PLoS ONE 11(4): e0152454. 31 p., <http://dx.doi.org/10.1371/journal.pone.0152454>) tudományos cikkek supplemetjeként szabadon letölthető.

```
rm(list=ls()); ls()                # tidying up the work space
library(MASS)                      # setup environment
library(cluster)

d <- read.csv("~angulatus-group_rawdata.csv", header=TRUE, na.strings="") # loading the input file of the
data

vari<-c(4:24)                      # define columns with (morphometric) variables
grouping<-2                        # define grouping variable (column)
sname<-3                           # define species name (column)

summary(d)
dim(d); n<-dim(d)[1]              # identify dimensions
ddrop<-rep(FALSE,n)               # make empty vector ddrop (n = number of measurement vari-
ables)
ddrop<-is.na(d[,grouping]); colnames(d)[grouping]
sum(ddrop)
dm<-d[!ddrop,vari]; summary(dm)    # dm is data matrix
sum(is.na(dm))                   # check for missing values
n<-dim(dm)[1]                    # number of observations (rows)
gid<-paste(d[!ddrop,sname],"-",as.character(d[!ddrop,grouping]),sep="")
#gid                             # show (new) sample names
table(gid);length(table(gid));sum(is.na(gid))
ddrop2<-NULL
tmp<-apply(apply(dm,1,is.na),2,sum)>0 # search for zeros
tmp3<-is.na(gid)                  # check for missing values
ddrop2<- tmp | tmp3;
rm(tmp,tmp3)
sum(ddrop2)
dm<-dm[!ddrop2,]; summary(dm)
sum(is.na(dm))                   # check for missing values
n<-dim(dm)[1]
gid<-gid[!ddrop2]
table(gid);length(table(gid));sum(is.na(gid))
#LDA using nest samples as groups
formlda<-lda(dm,as.factor(gid)); formlda # gid=group ID (site)
predlda<-predict(formlda,dm)$class
sum(predlda==gid)/length(gid)
x<-table(predlda==gid,gid); x
mpredlda<-predict(formlda,formlda$means)$x
plot(agnes(mpredlda,method="average"),which.plots=2,hang=-1,cex=0.7)

#####
### PART Partitioning based on recursive thresholding
#####
```

```

library(clusterGenomics)
groups1 = dm$gid
#Run PART.hclust
res <- part(mpredlda,Kmax=14,minSize=5,Kmax.rec=8,B=1000, cl.method = "hclust",
  linkage="average")          # run PART with cluster method 'hclust' using default setup
cbind(res$lab.hatK, groups1)   #extract hypothesis
summary.hclust<-cbind(res$lab.hatK, groups1)
hyp.part.hclust <- summary.hclust
thyp.part.hclust <- t(hyp.part.hclust)      # transpose partitioning results

#Run PART.kmeans
resa <- part(mpredlda, cl.method = "kmeans", minSize=5,Kmax.rec=8,nstart = 10,B=1000)
cbind(resa$lab.hatK, groups1)   #extract hypothesis
summary.kmeans<-cbind(resa$lab.hatK, groups1)
hyp.part.kmeans <- summary.kmeans
thyp.part.kmeans <- t(hyp.part.kmeans) # transpose partitioning results

#####
### Let's map the results of method PART on the NC clustering dendrogram
#####
#Source of the following fragment of the script: https://r-forge.r-project.org/scm/viewvc.php/src/hyperSpec/R/mark.dendrogram.R?root=hyperSpec&r1=451&r2=450&pathrev=451&diff\_format=s, see R package hyperSpec

dend <- agnes(mpredlda,method="average")
par(xpd = TRUE, mar = c (10, 4, 4, 2))      # allows plotting into the margin
par(mfrow=c(1,1))
plot(dend,which.plots=2,hang=-0.1,cex=0.65)  # set letter size

G2 <- c(1:12)                                # optionally type final species hypothesis, separated by
comma
dend$Group <- G2
clusters <- as.factor (dend$Group)
Z<-0.05                                     # set the bars' height (dendrogram$height in
mark.dendrogram function)

mark.dendrogram <- function (dendrogram, groups, col = seq_along (unique (dend$Group)),
  pos.marker = pos.marker,
  height = Z * max (dendrogram$height),
  pos.text = pos.text,
  border = NA, text.col = "white", label, label.right = FALSE,
  ...){
  if (! is.factor (groups))
    groups <- as.factor (dend$Group)

  groups.x <- groups [dendrogram$order]      # clusters in order on x axis

  rle.groups <- rle (as.integer (groups.x))   # run-length encoding gives borders

  end <- cumsum (rle.groups$lengths) + 0.5
  start <- c (0.5, (head (end, -1)))
  text <- (start + end) / 2

  text.col <- rep (text.col, length.out = length (text))

  for (g in seq_along (rle.groups$lengths)){

```

```

rect (xleft = start [g], ybottom = pos.marker - height,
      xright = end [g], ytop = pos.marker,
      col = col [rle.groups$values[g]], border = border, ...)
if (! is.na (text.col [g]))
text (x = text [g], y = pos.text,
      levels (groups) [rle.groups$values [g]],
      col = text.col [rle.groups$values [g]], ...)
}

if (! missing (label))
text (x = label.right * tail (end, 1) * 1.01, y = pos.marker - height/2,
      label, adj = c(1 - label.right, .3))
}

mark.dendrogram (dend, as.factor (thyp.part.hclust), pos.marker = -32, pos.text = -33, cex=0.7, label =
"part.hclust") # display partitioning results of cl.method hclust on the dendrogram
mark.dendrogram (dend, as.factor (thyp.part.kmeans), pos.marker = -36, pos.text = -37, cex=0.7, label =
"part.kmeans") # display partitioning results of cl.method kmeans on the dendrogram

mark.dendrogram (dend, as.factor (dend$Group), pos.marker = -300, pos.text = -300, cex=0.6, label = "fi-
nal.sp.hyp") # optional: read final species hypothesis

```

9.6 Melléklet. Az 5.4. fejezetben bemutatott taxonómiai munkák morfológiai karaktereinek és mérési pontos protokoljának kivonata.

A bemutatott részlet a (Csősz & Fisher (2016a) PeerJ 4:e1796, <https://doi.org/10.7717/peerj.1796>) munkából származik.

Protocol of morphometric character recording

CL: Maximum cephalic length in median line. The head must be carefully tilted to the position providing the true maximum. Excavations of hind vertex and/or clypeus reduce CL (Fig. 1A).

CW: Maximum width of the head. Includes compound eyes (Fig. 1A).

CWb: Maximum width of head capsule without the compound eyes. Measured just posterior of the eyes (Fig. 1A).

CS: Absolute cephalic size. The arithmetic mean of CL and CWb.

Cdep: Antero-median clypeal depression. Maximum depth of the median clypeal depression on its anterior contour line as it appears in fronto-dorsal view (Fig. 1B).

EL: Maximum diameter of the compound eye (not shown).

FRS: Frontal carina distance. Distance between the frontal carinae immediately caudal of the posterior intersection points between the frontal carinae and the torular lamellae. If these dorsal lamellae do not laterally surpass the frontal carinae, the deepest point of the scape corner pits may be taken as the reference line. These pits take up the inner corner of the scape base when the scape is directed caudally and produce a dark triangular shadow in the lateral frontal lobes immediately posterior to the dorsal lamellae of the scape joint capsule (Fig. 1B).

ML (Weber length): Mesosoma length from caudalmost point of propodeal lobe to transition point between anterior pronotal slope and anterior pronotal shield. Preferentially measured in lateral view; if the transition point is not well defined, use dorsal view and take the centre of the dark-shaded borderline between pronotal slope and pronotal shield as anterior reference point. In gynes: length from caudalmost point of propodeal lobe to the most distant point of steep anterior pronotal face (Fig. 1C).

MW: Mesosoma width. In workers MW is defined as the longest width of the pronotum in dorsal view excluding the pronotal spines (Fig. 1E).

MPST: Maximum distance from the center of the propodeal stigma to the anteroventral corner of the ventrolateral margin of the metapleuron (Fig. 1D).

NOH: Maximum height of the petiolar node. Measured in lateral view from the uppermost point of the petiolar node perpendicular to a reference line from the petiolar spiracle to the imaginary midpoint of the transition between dorso-caudal slope and dorsal profile of caudal cylinder of the petiole (Fig. 1D).

NOL: Length of the petiolar node. Measured in lateral view from the center of the petiolar spiracle to dorso-caudal corner of caudal cylinder. Do not take as the reference point the dorso-caudal corner of the helcium, which is sometimes visible (Fig. 1D).

PEH: Maximum petiole height. The chord of the ventral petiolar profile at node level is the reference line perpendicular to which the maximum height of petiole is measured (Fig. 1F).

PEL: Diagonal petiolar length in lateral view; measured from anterior corner of subpetiolar process to dorso-caudal corner of caudal cylinder (Fig. 1F).

PEW: Maximum width of petiole in dorsal view. Nodal spines are not considered (Fig. 1G).