

Ősi írásaink II.

A Magyarságkutató Intézet Kiadványai 31.

ŐSI ÍRÁSAINK II.

**Tanulmánykötet a Magyarságkutató Intézet
által 2020. december 10–11-én rendezett
konferencián elhangzott előadásokból**

SZERKESZTETTE:

FEHÉR BENCE - FERENCZI GÁBOR



Magyarságkutató Intézet

Budapest, 2021

Szakmai lektorok: Fehér Bence, Szabó Ádám
Nyelvi lektorok: Ferenczi Gábor, Nagy Dóra, Rási Szilvia

Az MKI szerkesztőbizottsága: Vizi László Tamás (elnök), Fehér Bence,
Katona József Álmos, Kovács Attila, Pomozi Péter, Virág István

A kötet megjelenését az EMMI támogatta.



EMBERI ERŐFORRÁSOK
MINISZTERIUMA

© Szerzők, Szerkesztők, 2021

ISBN

ISSN 2677-0261

TARTALOM

Előszó	7
Fehér Bence: Székelydália 1–4.	9
Katona-Kiss Attila – Bíró Gyöngyvér: Orgondaszentmiklóstól Komlósig. Újabb középkori rovásfeliratos csaták a Maros–Körös közéről.	31
Novotnik Ádám – Fehér Bence: Írásos emlékek a Tolna-Mözs-Fehérvize- dűlői avar temetőben.	55
Puskás Timea – Hosszú Gábor: Az írásevolúció kladisztikai megközelítése	85
Rumi Tamás – Sípos László – Somfai Tamás: Jelenkori rovás emlékek – digitális tartalomfejlesztés	117
Szabó Ádám: Az apahidai Mauric (Mauricus ≈ Mauricius)	139
Szabó Géza: Kaukázusi régészeti és néprajzi megfigyelések a rovásfeliratos Nagyszentmiklósi kincs értékeléséhez az újabb archaeometriai vizsgálatok tükrében	175
Tóth Anna Judit: Komáromi Csipkés György és a nyelvek eredete	237
Visy Zsolt: Rovásírásos és latin nyelvű feliratok Énlakán	265
Zelliger Erzsébet: Bonyhai Moga Mihály írásbelisége nyelvész szemmel ...	307

AZ ÍRÁSEVOLÚCIÓ KLADISZTIKAI MEGKÖZELÍTÉSE

PUSKÁS TIMEA – HOSSZÚ GÁBOR

ABSZTRAKT: A tanulmány olyan írások evolúciójának kladisztikai elemzésével foglalkozik, amelyben a vizsgált írások nem csak egy elődtől származnak. A kidolgozott eljárás a fenetikai leírást evolúciós információkkal kibővítő, ún. kiterjesztett fenetikai modellből indul ki, amelynek segítségével a vizsgált írások elődírásai meghatározhatók. A jelen kladisztikai vizsgálat célja, hogy felderítse a tanulmányozott írásoknak egymáshoz és a korábbi kutatásokból meghatározott elődírásaikhoz való evolúciós leszármazási kapcsolatait. A kladisztikai elemzésben a Wagner-módszer segítségével hoztuk létre a kladogramokat, és az eljárást a steppei népek írásainak vizsgálatára használtuk fel. Az elemzések megvalósítása Python programozási nyelven történt.

KULCSSZAVAK: fenetika, írásinformatika, kladisztika, konzisztenciaindex, összetartási index, rovás írások, Wagner-módszer

ABSTRACT: A cladistic approach to script evolution. The study deals with the cladistic analysis of the evolution of scripts that originate from more than one predecessor script. The starting point of the elaborated procedure is the 'extended' phenetic model that expands the phenetic description with evolutionary information with the help of which the predecessor scripts of the scripts under examination can be defined. The objective of the present cladistic analysis is to reveal the evolutionary lineage relations of the examined scripts

with one another and with predecessor scripts identified in previous researches. In the cladistic analysis, cladograms are created using the Wagner method and the described procedure is used to examine the scripts of the Steppe peoples. The analyses were implemented in Python programming language.

KEY WORDS: cladistics, consistency index, phenetics, retention index, Rorash scripts, scriptinformatics, Wagner method

Bevezetés

A világban számtalan megfejtetlen írásemlék van. Az írásinformatika hosszútávú célja, hogy ezeknek az értelmezésében, megfejtésében segítsen az írások evolúciós kapcsolatainak feltárásával (Hosszú 2010, 2014a, 2014b, 2015, 2019). Az írásinformatikai kutatások hátteréhez érdemes figyelembe venni, hogy az írások fejlődése gyakran több évszázadon keresztül zajló folyamat, azonban ebből sokszor csak egy-egy írásemlék maradt fenn, amely nem elég az írás teljes fejlődésének nyomon követéséhez. Tegyük fel például, hogy biztosan tudjuk, hogy egy írást egy évezreden keresztül használtak, de összesen csak két írásemlék maradt fenn róla, egy a hatodik századból, egy a nyolcadikból. Ez önmagában nem elég arra, hogy az írás fejlődését végig követhessük, és megmondhassuk, mikor, hogyan és milyen behatás alatt fejlődött az írás, és milyen formákban jelent meg.

További probléma, hogy nem minden megtalált írásemléken használt írást lehet azonosítani és a leírt szöveget megfejteni. Például tekintsük a Kr. e. 196-ban, Egyiptomban készült Rosette-i követ, amelyet 1799-ben találtak meg (Parkinson 1999). Esetében azért volt lehetőség a rajta található szövegek elolvasására, mert a három részből álló felirat elég hosszú ahhoz, hogy fel lehessen ismerni a felirat egyes részeihez használt írásokat. Ezek az egyiptomi hieroglif, a démotikus és a görög írás. Az előbbi két írással készült felirat óegyiptomi, az utóbbi ógörög nyelvű szöveget tartalmaz, azonban ezek lényegében azonos tartalmúak. Ezzel szemben például egy egyszerű egy-két szá-

vas, ismeretlen írással készült sírfelirat megfejtése sokkal nehezebb feladat. A régészek ezerszámra találnak olyan rövid írásemlékeket, amelyek esetén még a felirathoz használt írás meghatározása is bizonytalan, így a felirat megfejtése is gyakran lehetetlen. Ennek az egyik oka az, hogy a feliraton található jelalakok olyan írással vagy írásváltozattal készültek, amely addig ismeretlen volt. Ezen a gondon segítené, ha az írások fejlődését a mainál részletesebben ismernénk.

Az írásfejlődés modellezéséhez a biológiai evolúcióban alkalmazott módszerek csak módosítva alkalmazhatók, s ez a két tudományterület tárgya közötti eltérésekből adódik. Az ezzel kapcsolatos kutatások járulékos haszna lehet, hogy az eredetileg csak a biológiai evolúcióban alkalmazott módszerek átdolgozva további evolúciós folyamatok leírására is alkalmassá válhatnak, ilyen terület például a szoftverevolúció.

A filogenetikus rendszertan eredetileg az élőlények rendszerezésével foglalkozik; az élőlények törzsfjlődése során kialakult rokonsági kapcsolatok alapján rendszerez, és az ezen kapcsolatokat kifejező filogenetikus családfa segítségével osztályoz. A filogenetika feltételezése, hogy a hasonló dolgoknak közös ősük van. Két fő része van: a nevezéktan és az osztályozási rendszer. A neveket a leszármazás alapján adják, egy osztályba pedig a genetikailag hasonló dolgok tartoznak. A közös csoportba tartozás és a közös ősrel rendelkezés ekvivalens. A *kladisztika* a filogenetika egy, leszármazási mintákat feltáró módszertana. Ezzel szemben a *fenetika* a morfológiai vagy más megfigyelhető jellegbeli hasonlóságokon alapuló elemzés, amely nem veszi figyelembe az élőlények közötti leszármazásbeli vagy evolúciós tulajdonságokat (Podani 2007, 114–115., 197–203.).

Az evolúciós folyamatok nem csak a biológiában találhatók, ezért az evolúció tanulmányozására kidolgozott módszereket más tudományterületeken is lehet alkalmazni. Ennek egyik példája a szoftverevolúció, amely a szoftverek fejlődésének vizsgálatával foglalkozik (Sampaio 2007). Szintén az evolúciós módszerek alkalmazásának területe az írásinformatika (Hosszú 2019, 120), amelyben az írás az osztályozási alapegység (taxon), a taxonokat leíró jellegek pedig a grafémák és az írásszabályok. A kutatás során vizsgált írások egy részéhez csak igen kevés írásemlék ismert, ráadásul azok nagy részének a kora

ismeretlen vagy csak durva közelítéssel határozható meg, így ezen írások teljeskörű kladisztikai vagy filogenetikai elemzése túlmutat a kutatás eddigi keretén. Ezért elsődlegesen a fenetikai elemzésre kellett korlátozni a vizsgálatokat, s jóllehet ez elvileg nem nyújt ismeretet evolúciós összefüggésekre, de az írások egymás közötti hasonlósági viszonyait leírja. A vizsgálatok során kapott fenetikai eredmények, vagyis az, hogy egyes írások hasonlóbba egymáshoz, mint mások, utalnak az evolúciós hátterükre, de nem jelentenek közvetlen evolúciós ismereteket.

Az írások sajátossága, hogy a biológiai fajokhoz képest sokkal kevesebb jellelleggel lehet azokat leírni, ezért nagy a véletlen egybeesések, ún. homopláziák (véletlen hasonlóság) esélye. Ennek következtében a fenetikai analízis könnyen téves összefüggéseket mutathat ki. Ennek elkerülésére evolúciós jellegű megfontolásokat is figyelembe vettünk a Budapesti Műszaki és Gazdaságtudományi Egyetem Elektronikus Eszközök Tanszékének ezirányú kutatásait felhasználva (Hosszú 2017, 2019, 2021). Így egy kiterjesztett fenetikai analízist lehetett elvégezni, s ez azzal a járulékos előnnyel járt, hogy az ebből kapott eredmények alkalmassá váltak közelítő evolúciós következtetések levonására is. Ezen vizsgálatok alapján meghatároztuk a steppei népek által használt írások egy csoportjának, a rovás írásoknak a valószínű elődeit (Hosszú 2021a, 229–231.). Ezek szerint a rovás írások elsődleges elődje az arám-középiráni írások egy csoportja (hivatalos arám, párthus, középperzsa és szogd), s a másodlagos elődje a bráhmí írás.

Az írások egy részének nem csak egy elődje van, így a fejlődésüket nem lehet leírni egyetlen filogenetikai fával, vagyis egyszerű előd-utód kapcsolatként nem egyszerű törzsfá mentén haladt, hanem az eltérő ágak esetenként összekapcsolódtak, s ezt *retikulációs evolúciónak* (*hálózatos evolúció, anasztomózis*) nevezünk (Podani 1997: 175). A retikuláris evolúció egyik mechanizmusa a *hibridizáció*. Ha több elődje van egy írásnak, akkor ezek hibridizációját nem fával, hanem hálózatos gráffal ábrázolhatjuk. A hibridizáció két fajtája az *írás-szintű hibridizáció* és a *grafémaszintű hibridizáció* (Hosszú 2021a, 21–22.). Az írásszintű hibridizáció írások keveredésével, azonban a grafémaszintű hibridizáció különböző írásokból származó grafémák összeolvadásával keletkezik. Mivel a kladisztikai elemzés előfeltétele az egyetlen elődtől való leszármazás,

ezért a hibridizáció esetét általában nem lehet kladisztikailag vizsgálni. Azonban a vizsgált utódírások jellegeit leíró, kiterjesztett fenetikai modell felbontható olyan részmodellekre, amelyekben már csak egy elődírás szerepel. Természetesen egy ilyen részmodell nem írja le pontosan a vizsgált utódírások teljes fejlődését, azonban közelítőleg megmutatja, hogy egy adott elődírás hogyan hatott a vizsgált utódírásokra. Különösen akkor lehet hatékony ez a módszer, ha a vizsgált utódírásokra nem egy időben hatottak az elődírások, hanem eltérő időszakokban. Ugyanis az elődírások hatásának elkülönítésével a vizsgált utódírások fejlődési korszakai leírhatókká lehetnek.

A jelen vizsgálat az egyes elődírások hatásainak elkülönítése céljából részekre bontott kiterjesztett fenetikai modellen alapuló kladisztikai vizsgálatokkal foglalkozik. A kladisztikai vizsgálat célja, hogy felderítse a tanulmányozott írásoknak egymáshoz és a korábbi kutatásokból meghatározott elődírásaihoz való evolúciós leszármazási kapcsolatokat. A kladisztikai analízisben általában a legegyszerűbb és ezért a legvalószínűbbnek tartott, ún. legparszimónikusabb kladogramok létrehozatala a cél (Podani 1997, 177.). Megjegyzendő, hogy a *parszimónia*, vagyis a lehető legkevesebb jellegállapot-változás feltételezése, más szóval a *takarékossági elv* egy általános tudományos alapelv. Ennek a filozófiában *Occam* [Ockham] *borotvája* felel meg, amely szerint a legegyszerűbb magyarázat a helyes. A kladisztikában az optimális fa minimális számú jellegállapot-változással magyarázza a leszármazási viszonyokat.

Az írásinformatikai adatbázisra alkalmazott kladisztikai vizsgálatok elméleti hátterét Hosszú Gábor dolgozta ki, a szükséges szoftvert Pythonban Puskás Timea készítette el. A vizsgálatok eredményét a valamikori steppei népek által használt négy írás példáján mutatjuk be, nevezetesen a türk rovás (TR), székely-magyar rovás (SZMR), a Kárpát-medencei rovás (KMR) és a steppei rovás (SR) (Hosszú 2019, 138–140.); azonban a kidolgozott módszer az elődök hatásának elkülönítésével általánosan is felhasználható taxonok retikulációs evolúciójának kladisztikai elemzésére és előd-utód kapcsolatok meghatározására. A cikk először a módszer elméleti hátterével, majd a kidolgozott eljárás leírásával, a vizsgálati eredmények ismertetésével, végül a következtetések levonásával foglalkozik.

Elméleti háttér és irodalmi áttekintés

Evolúciós és írásinformatikai alapfogalmak

A *taxon* (taxonómiai egység) a vizsgált objektumok (egyedek, statisztikai egyedek, adatpontok, objektumok) azonos kategóriába sorolt csoportja, amelyet a rendszertanban egységként kezelnek. Általában a modellezés elvonatkoztatási szintjétől függ, mit tekintünk taxonnak. A taxon azon egyedek egy csoportja, amelyeknek a leszármazási viszonyait vizsgáljuk. A biológiai evolúció modellezésében a taxon lehet egy faj vagy egy fajta. Az írásinformatikában a taxon általában egy írás, de lehet egy írásváltozat is (Hosszú 2019, 125.). A biológiai egyedek írásinformatikai megfelelői az adott írással készült feliratok (Hosszú 2017, 183.). A *jelleg* az a tulajdonság, amelynek alapján vizsgáljuk a taxont, pl. egy írás jellegei közé tartoznak, milyen grafémákat (betűk, ligatúrák, számjegyek vagy írásjelek) tartalmaz. Az írásinformatikában a jelleg lehet graféma vagy írásszabály. A *jellegállapot* egy adott jellegnek egy taxonbeli megléte (1) vagy hiánya (0). A *homoplázia* olyan tulajdonság, amely több nem közös őstől származó taxon között megegyezik (Torres-Montúfar és mtsai. 2018, 543–551.). Ezzel szemben a *szünapomorfia* két vagy több taxon megfelelő jellegének közös leszármazott jellegállapota, amely a közös törzsvonalon evolúciós újdonságként jött létre, és a legközelebbi közös elődjüktől ered.

Bár a kladisztika alkalmazása manapság általánosan elterjedt, azonban egyes esetekben a fenetika használata kedvezőbb. Ennek egyik példája, ha nincs nyilvánvaló előd-utód kapcsolat a vizsgált taxonok között, mivel ekkor a kladisztikai módszerek nem elég hatásosak és helyettük fenetikai eljárásokat kell alkalmazni (Podani–Morrison 2017, 191.). A fenetika alkalmazásának másik példája, ha nem csak egyszerű elágazásos fejlődés van, hanem valószínűleg hálózatos, akkor a kladisztika alapfeltételei nem teljesülnek. Így annak ellenére, hogy a fenetika kevesebbet ígér, mint a kladisztika, mégis érdemes végrehajtani a fenetikai elemzést, mert a vizsgált taxonokat leíró elég sok jelleg ismerete esetén a fenetikai elemzés biztosan végrehajtható.

A kladisztika csak azokat a jellegeket (fenotípusokat) használja, amelyek evolúciós szempontból lényegesek. Ezzel szemben a fenetikai osztályozás minden észlelhető jelet figyelembe vesz, így nemcsak azokat, amelyeknek evolúciós jelentőségük van. A kladisztikai eredményeket kladogramon, a fenetikai eredményeket fenogramon ábrázolják. Mind a kladogram, mind a fenogram olyan gráf, amelynek terminális csomópontjai az osztályozott taxonoknak felelnek meg. A dendrogram a taxonok egymáshoz tartozását szemlélteti.

Egyszerű fenetikai modell

A fenetikai modellezés során az írások mint taxonok és a jellegeik megléte vagy hiánya alapján egy kétdimenziós, bináris taxon–jelleg-adatmátrix hozható létre (Hosszú 2017, 200–201.). A jellegtervezés során dönteni kell minden vizsgált írás esetén az egyes jellegek meglétéről vagy hiányáról. Egy-egy jelleghez tartozó taxonok (esetünkben írások) kiválasztása egy szöveges adatbázisban történik, amelyben az egy jelleghez tartozó adatok együttes elnevezése a *hasonlósági jellegcsoport* (*Similarity Feature Group*, SFG). Az SFG-k meghatározása a gépi tanulást előkészítő jellegtervezésbe tartozik (Hosszú 2017, 182). A hivatkozott közleményben szereplő fenetikai modell egy részlete látható az 1. ábrán, ahol néhány SFG szerepel.

SFG-8: *NE-Iberian, Celtiberian* ▷ <a>; *SHR* ◁ <a>

SFG-9: *Carian* Ⓐ, *SE-Iberian* Ⓐ, *Elymian* Ⓐ, *Latin (epigraphic cursive)* Ⓐ, (*cursive majuscule, Pompeii*) ⓧ, *Raetic* Ⓐ, Ⓐ, *Lepontic* Ⓐ, Ⓐ, *Gallo-Etruscan* Ⓐ, Ⓝ, *Camunic* Ⓐ, *Runic (older fuþark)* †; † <a>. (i) The Runic † could be an autapomorphy.

SFG-10: *Elymian* Ⓐ, *Raetic* Ⓐ, Ⓐ <a>

SFG-11: *Parthian* Ⓝ <'> /a, ā/; *Sogdian* Ⓝ, Ⓝ <'> /a, ā, ə/; *Syriac* Ⓝ <'>

1. ábra: A 2017-ben közölt 66 írás 186 jellegét tartalmazó fenetikai modell egy részlete (Hosszú 2017, 207)

Az 1. ábrán látható SFG-k közül az SFG-9, SFG-10 és az SFG-11 problémamentes, az ezekben szereplő betűalak SFG-nként ugyanahhoz a grafémához tartozik, így az azonos SFG-ben szereplő írások biztosan ugyanazt a jelet

tartalmazzák. Az SFG-9 különlegessége, hogy kiegészítésképpen tartalmaz egy olyan jelleget is, amely valószínűleg a rúna (*Runic*) íráson belüli fejlődés eredménye, és más írásban nem jelenik meg, ezt autapomorfiának (*autapomorphy*) nevezik. Mivel az autapomorf jelleg csak egyetlen írásban (taxonban) jelenik meg, ezért a filogenetikai elemzésben általában nincs szerepe, továbbá nem képez összehasonlítási alapot, s az 1. ábrán látható SFG-9-ben csak a teljesség igénye miatt jelenik meg.

Az 1. ábra SFG-8 mélyebb vizsgálatot igényel. Az északkelet-ibér (*NE-Iberian*), a keltibér (*Celtiberian*) valamint a SZMR (*SHR*) egyaránt tartalmazza az SFG-8-beli szár nélküli, háromszög alakú betűalakot. Az előbbi kettő egymással kétségtelenül szoros rokonságban lévő óhiszán írás, ezzel szemben a SZMR-nak biztosan nincs közvetlen köze az óhiszán írásokhoz, legfeljebb azokkal közös elődei lehetnek, s az SFG-8-ban szereplő betűalak talán ilyen közös elődtől való leszármazásból adódik. De az is lehet, hogy a homoplázia (jellegállapotok nem evolúciós eredetű azonossága) egyik esetével, a párhuzamos evolúcióval állunk szembe, amikor egymástól függetlenül kifejlődött mind az óhiszán írásokban, mind a SZMR-ban az <a> betűnek az SFG-8-ban szereplő háromszögszerű alakja. A kifejlődés oka is hasonló lehetett: a szár nélküli betűalakot könnyebb vésni.

Kiterjesztett fenetikai modell

Amint azt az SFG-8 példáján láthatjuk, a rendelkezésre álló ismeretek alapján gyakran nem lehet meghatározni, hogy két graféma hasonlósága evolúciós eredetű-e, vagy csak a véletlen műve (homoplázia). Ilyenkor az egyértelmű döntés információvesztéssel járhat. Ezért célszerű egy jellegnek egy vizsgált írás szempontjából való elemzésébe evolúciós megfontolásokat is bevonni, bár ez általában nem a fenetika, hanem a kladisztika jellemzője (Hosszú 2020, 52.). A biológiai evolúcióval szemben az írásevolúcióban sokkal gyakoribb a homoplázia, s ez a taxont meghatározó jellegek kisebb változatosságából adódik. Ezért az egyszerű fenetikai modellezés tovább lett fejlesztve az ún. kiterjesztett fenetikai modellé (Hosszú 2019, 2021), amelyben egy SFG-ben a bizonytalan esetekben több, alternatív lehetőség is figyelembe vehető anél-

kül, hogy hiányos információ alapján dönteni kellene közöttük. Ezek az alternatív lehetőségek valamilyen (történelmi, földrajzi stb.) indokra épülnek, és a vizsgálat időpontjában a jellegtervezés során nincs elég ismeret annak eldöntésére, hogy az alternatív lehetőségek közül melyik szerint zajlott le egy adott jelleg fejlődése. Ezen alternatív lehetőségek elnevezése az *eredetmodell*. Egy adott SFG-beli eredetmodellek csak arra az egy SFG-re vonatkoznak, és nem feltétlenül terjednek ki a taxon – vagyis a vizsgált írás – többi jellegére. A korábbi kutatások során létrehozott kiterjesztett fenetikai modellben (Hosszú 2019, 2021) számos olyan jelleg van, amelyben csak egy eredetmodell jöhet szóba, de a jellegek többségénél egynél több alternatíva lehetséges, s közülük kizárólag csak az adott jelleg vizsgálata alapján nem lehet biztosan eldönteni, melyik a valóságban tényleg megtörtént evolúciós lépés leírása. Más szavakkal, a rendelkezésre álló adatok alapján gyakran nem dönthető el, hogy az adott jelleghez tartozó utódgraféma melyik számba vehető elődírásból származik (erre példa az 1. ábrán szereplő SFG-8). Egy jelleghez (utódírásbeli grafémához vagy írásszabályhoz) több alternatív elődírásbeli graféma vagy írásszabály is szóba jöhet, amelyek mind ugyanabba az SFG-be tartoznak (Hosszú 2020, 53.). Az SFG egy olyan adatstruktúra, amely tartalmazza egy adott jelleghez (graféma vagy írásszabály) az összes információt. Vagyis egy SFG magában foglalja egy jelleghez kapcsolódóan az összes taxont (írást), amelyben az adott jelleg előfordul, valamint a hozzájuk tartozó egyéb információkat is. Ha nem tudható biztosan, hogy két hasonló jelleg evolúciósan megegyezik-e, akkor inkább külön SFG-be kerülnek (Hosszú 2014a, 2014b, 2015, 2017).

Gyakran nem lehet megmondani, hogy ha egy tulajdonság több taxonban is jelen van, akkor az csak a véletlen műve (homoplázia), vagy az érintett taxonok evolúciós kapcsolatban állnak. Erre egy megoldás, ha különböző háttérinformációk felhasználásával eldöntjük, melyik igaz a két lehetőség közül. Ahhoz, hogy leírható legyen az összes számba vehető elődírásjelölt, a hagyományos fenetikai modell ki lett egészítve az adott jellegre vonatkozó, egyes származási alternatívák figyelembevételével. Az így létrejött kiterjesztett fenetikai modell (P_c) lehetővé teszi az abból adódó probléma elkerülését, hogy az egyszerű fenetikai modell (Hosszú 2017) nem veszi figyelembe a vizsgált

taxonok egyes tulajdonságainak evolúciósan távolabbi, elődként valószínűsíthető taxonokból származó eredetét, hanem csak a vizsgált taxonok egymáshoz viszonyított jellemzőit vizsgálja. A vizsgált utódírások (eredményírások, utódtaxonok) leírására, különböző evolúciós megfontolások (eredetmodellek) alapján elkészített kiterjesztett fenetikai modell (P_e) azonban nem egyértelmű, mivel a modellt alkotó hasonlósági jellegcsoportok (SFG-k) egy része több elődíráshoz tartozó elődgrafémát tartalmaz (Hosszú 2019, 182.). Ennek egyik oka az, hogy az SFG-kben több, alternatív eredetmodell szerepel, amelyek eltérő evolúciós hátteret (földrajzi, történelmi, nyelvészeti adatokat) írnak le, és az ugyanazon SFG-ben szereplő eredetmodellek általában eltérő elődíráshoz tartozó elődgrafémát tartalmaznak. Az SFG-k egy részében található elődírások többféleségének másik oka az, hogy még egyetlen eredetmodellből is gyakran egynél több lehetséges elődgraféma következik. Ugyanis a különböző írások egyes grafémái nagyon hasonlóak lehetnek, így az alakzati és hangértékbeli tulajdonságok alapján sokszor nem lehet eldönteni, hogy egy utódgraféma melyik – egymással rokon – elődírásból származik. A kiterjesztett fenetikai modellben szereplő alternatív lehetőségekről nem tudható, melyik írja le a ténylegesen végbement evolúciós folyamatot, s ez csökkenti a modell egyértelműségét.

A fenetikai analízis és az elődanalízis eredménye

A P_e kiterjesztett fenetikai modellből azon hasonlósági jellegcsoportok (SFG-k) értékelhetők ki legegyszerűbben, amelyek valamilyen szempontból egyértelműek. Ha csak az egyetlen elődírást tartalmazó SFG-ket vesszük figyelembe a modell kiértékelésénél, akkor a P_e -beli SFG-knek csak a töredékét vehetjük figyelembe, ez pedig csökkenti a modell *felbontását*, vagyis a vizsgálat pontosságát. A kiértékeléskor felhasználható SFG-k számának növelése érdekében figyelembe vehető, hogy az SFG-k egy részénél szereplő elődírásjelöltek vagy közeli rokonai egymásnak, vagy egy korban, egy földrajzi térben használt, összefüggő kulturális közegből származnak. Így a bennük található elődgrafémák nagyon hasonlóak lehetnek, és ezért bizonytalan, hogy ezen rokon elődírásjelöltek közül pontosan melyik a tényleges előd. Ráadásul,

figyelembe véve a retikulációs evolúció lehetőségét, számításba lehetne venni a grafémaszintű hibridizációt is (Hosszú 2017, 185.), vagyis egy graféma tényleges evolúciója során akár több elődje is lehetett. Azonban a jelen vizsgálatban a grafémaszintű hibridizáció lehetőségével nem számolunk. Ha nem elődírásokat, hanem hasonló jellegekkel rendelkező elődírások csoportjait keressük a modell kiértékelésekor, akkor az egyértelmű SFG-k száma növekedik azon SFG-kkel, amelyekben ugyan több elődírás van, de ezek rokonok. Ezért érdemes elődíráscsoportokat alkotni (Hosszú 2020, 54–56.; Hosszú 2021a, 210–211.). A P_e kiterjesztett fenetikai modell kiértékelése szukcesszív eliminációs eljárással történik (Hosszú 2019, 383–388.; Hosszú 2020, 56–70.; Hosszú 2021a; Hosszú 2021b). Ennek során az elődíráscsoportok átkonfigurálhatók, ha ezzel a szukcesszív eliminációs eljárással kapott fenetikai modellben az elődíráscsoportokra nézve egyértelmű SFG-k száma, vagyis a modell felbontása növekszik (Hosszú 2021a, 222–223.).

A jelen vizsgálatokban a P_e -beli utódírások az eurázsiai steppe népeinek egy része által használt írások (TR, SZMR, KMR és SR); az elődírások pedig a P_e -beli valamelyik SFG-ben elődgrafémát adó írások (Hosszú 2019: 188–370). A hivatkozott közleményben (Hosszú 2019) szereplő modell – felépítési elveinek változatlansága mellett – az azóta végzett kutatások alapján tovább lett fejlesztve (Hosszú 2021a).

Az elődíráscsoportok használatának járulékos előnye, hogy ezzel kezelhetővé válik az az eset is, amikor egy hajdani grafémaátadás nem pontosan azokból az írásokból történt, amelyeket ma önálló elődírásaként veszünk figyelembe, hanem ezek valamilyen jelenleg ismeretlen átmeneti formájából. Ilyenkor az elődíráscsoportban lévő írások grafémái csak közelítik a feledésbe merült hajdani tényleges elődgrafémákat (Hosszú 2020, 54.).

A P_e kiterjesztett fenetikai modell szukcesszív eliminációs eljárással történt elemzése alapján létrejött fenetikai modell részhalmaza a négy vizsgált írásnak (TR, SZMR, KMR és SR) egymáshoz képesti hasonlósági viszonyait leíró egyszerű fenetikai modell. Az ebből származó fenogramot (Hosszú, 2021a, 228.), vagyis a fenetikai elemzés eredményeképpen kapott dendrogramot bemutatja a 2. ábra.



2. ábra: A négy vizsgált írás fenogramja

Az alternatív eredetmodellel kibővített kiterjesztett fenetikai modell (P_e) az elődírásnak jelölt írásokból számos közel-keleti, kis-ázsiai, indiai, belső-ázsiai és közép-európai írást tartalmazott. A kiterjesztett fenetikai modell szukcesszív eliminációs eljárással történt elemzése (Hosszú 2021b, 207–231) azt mutatta, hogy a vizsgált írások két nagy elődírás csoport hibridizációjából származnak. Ezek egyike az arám-középiráni írás csoport (hivatalos arám, párthus, középperzsa és szogd írás), a másik csoport egyetlen tagja az indiai eredetű, de egy időben Belső-Ázsiában is használt bráhmii írás (Hosszú 2021a, 222–223.). Megjegyzendő, hogy a bráhmii írásnak is voltak egymástól eltérő változatai Belső-Ázsiában (Hosszú 2021a, 257–258.).

Az adatbázisban megjelenő 119 SFG-ből 33 van, amelyről biztosan tudjuk, hogy az arám-középiráni írás csoportból származik, azaz 33 olyan tulajdonságról beszélhetünk, amely az arám-középiráni írás csoport behatása alapján fejlődött ki. Az 1. táblázatban olyan SFG-k egy-egy részlete látható, amelyben egy vagy több rovás grafémának (jellegnek) megfelel az arám-középiráni írás csoport valamelyik írásának megfelelő grafémája (jellege). Az 1. táblázatban és a később következő 2. táblázatban csak egy-két alakváltozat látható. A további alakváltozatok, átbetűzési és átírási értékek, irodalmi hivatkozások, valamint a közölt grafémaalakok forrása megtalálható a vizsgálatot részletező könyvben (Hosszú 2021a).

SFG	Elődjellel	Utódjellel
SFG-3	hivatalos arám 𐭠 'ālap <'>; szogd 𐭡 <'>	SZMR (Konstantinápoly) 𐭠 <e>; TR 𐭠 <A> /ä/
SFG-4	szogd 𐭢 <'>	TR 𐭣 <A>
SFG-8	párthus 𐭤 <'>; szogd 𐭥 <'>	KMR 𐭤 <A>; SR 𐭥 <A>
SFG-17	hivatalos arám 𐭦 bēl ; párthus 𐭧 ; szogd 𐭨 	TR 𐭩 <b ¹ >; SR 𐭪 <b ¹ >
SFG-21	szogd 𐭫 gāmal <γ>	SZMR (Vargyas) 𐭫 <γ>; TR 𐭬 <g ¹ >, KMR N <γ>; SR 𐭭 <g ¹ >
SFG-25	hivatalos arám 𐭮, 𐭯 dālet <d>; párthus 𐭰 <d>; szogd 𐭱 <d>	KMR 𐭱 <d>; SR 𐭲 <d>
SFG-31	hivatalos arám 𐭳, 𐭴 wāw <w>	TR 𐭵, 𐭶 <W>; KMR 𐭷 <W>
SFG-32	párthus 𐭸 wāw <w>; szogd 𐭹 <w>	SZMR 𐭺 <o>
SFG-42	hivatalos arám 𐭼, 𐭽 hē <h>; párthus 𐭾 <h>	SR 𐭿, 𐮀 <χ>
SFG-50	hivatalos arám 𐭿, 𐮀 yōd <y> /y, ī/; párthus 𐮁 <y>	SZMR 𐮁, 𐮂 <i, j>; TR 𐮁, 𐮂 <I>; KMR 𐮁 <I>; SR 𐮁, 𐮂 <I>
SFG-51	középpirzsa 𐮃 <y>	KMR 𐮃 <I>; SR 𐮃 <I>
SFG-92	hivatalos arám 𐮄 rēš <r>; párthus 𐮅 <r>; szogd 𐮆 <r>	SZMR 𐮇 <r>; TR 𐮈, 𐮉 <r ¹ >; SR 𐮈 <r>
SFG-97	hivatalos arám 𐮊, 𐮋 śin/šin <š>; párthus 𐮌 <š> /š, ž/	SZMR 𐮍 <ž> /š, ž/; TR 𐮍, 𐮎 <š>; SR 𐮍 <š>
SFG-103	hivatalos arám 𐮏, 𐮐 tāw <t>; párthus 𐮑 <t>; szogd 𐮒 <t>	TR 𐮓 <t ² >

1. táblázat: Az arám-középiráni íráscsoport néhány grafémájának megfeleltetése a kiterjesztett fenetikai modellben (Hosszú 2021, 53–206.)

Az adatbázisban megjelenő 119 SFG-ből 29 van, amelyről biztosan tudjuk, hogy a bráhmii írásból származik, azaz 29 olyan tulajdonságról beszélhetünk, amely a bráhmii íráscsoport hatása alapján fejlődött ki. A 2. táblázatban olyan SFG-k egy-egy részlete látható, amelyben egy vagy több rovás grafémának megfelel a bráhmii íráscsoport valamelyik írásának megfelelő grafémája.

SFG	Elődjellel	Utódjellel
SFG-48	bráhmii 𑀓 <tā>	TR 𑀓 <t ¹ >
SFG-61	bráhmii 𑀕 <kha>; bráhmii 𑀖 <khi>	SZMR 𑀕, 𑀖 <k>; TR 𑀗 <q ¹ >
SFG-62	bráhmii 𑀙, 𑀚 <ka>	TR 𑀛, 𑀜 <w ^q >
SFG-68	bráhmii 𑀞 <ma>; 𑀟 <ma>	SZMR 𑀞 <m>
SFG-74	bráhmii 𑀡 <ña>	TR 𑀢 <n ¹ >
SFG-75	bráhmii 𑀣 <ne>	TR 𑀤, 𑀥 <n ² >; SR 𑀦 <n ² >
SFG-82	bráhmii 𑀨 <tṣa>	SZMR 𑀨 <č>; SR 𑀩, 𑀪 /s/; 𑀫 /tʃ/ <č>
SFG-88	bráhmii 𑀬, 𑀭 <kā>	KMR 𑀬 <q>; SR 𑀬 <q>; SR 𑀮 <w ^q >
SFG-98	bráhmii 𑀱 <śa>	SZMR 𑀱 <š>; TR 𑀲, 𑀳 <š>; SR 𑀴 <š>
SFG-99	bráhmii 𑀶, 𑀷 <śa>; bráhmii 𑀸 <śa>	SZMR 𑀸 us <š>; KMR 𑀸, 𑀹 <š>
SFG-100	bráhmii 𑀺 <śa>	TR 𑀻, 𑀼 <š>
SFG-102	bráhmii 𑀽 <śa>	SZMR 𑀽, 𑀾 <z>
SFG-108	bráhmii 𑀿, 𑁀 <tu>	KMR 𑁁 <t>

2. táblázat: A bráhmii írás néhány grafémájának megfeleltetése a kiterjesztett fenetikai modellben (Hosszú 2021, 53–206.)

A vizsgált utódírásoknak a szukcesszív eliminációból kapott két fő elődje közül nagyon valószínű, hogy az arám-középiráni íráscsoport hatása megelőzte a bráhmét (Hosszú 2021a, 229.). Ez azt jelenti, hogy a vizsgált taxonok fejlődése az arám-középiráni íráscsoportból indult, és ezzel alakult ki egy feltételezett írás, a *protorovás (Hosszú 2019, 390–392.). Ezek után a *protorovás írásba horizontális átadás történt a bráhmí írásból, ez pedig a hálózatos evolúció egyik mechanizmusa. Így viszont az utódírás csoport fejlődése nem vizsgálható kladisztikai módszerekkel, mert az rendszerint csak elágazásos fejlődést vesz figyelembe, hálózatos fejlődést általában nem.

A szukcesszív eliminációval kapott fenetikai modellben szereplő olyan SFG-k különválaszthatók, amelyek egyértelműen vagy az arám-középiráni, vagy a bráhmí elődírás csoportból származnak. Így az eredetileg egy fenetikai modell három partícióra bontható, abból az egyik tartalmazza az egyértelműen arám-középiráni eredetű jellegeket (SFG-eket), a másik az egyértelműen bráhmí eredetűeket, a harmadik partíció pedig az összes többi SFG-t. Az arám-középiráni eredetű jellegeket tartalmazó partíciónak a csak a vizsgált utódírásokra vonatkozó része az (1) képletben látható $P_{arám-középiráni}$ azonban a bráhmí eredetű jellegeket tartalmazó partíciónak a csak a vizsgált utódírásokra vonatkozó része a (2) képletben látható $P_{bráhmí}$. Összesen 33 db SFG volt a szukcesszív elimináció eredményeképpen kapott fenetikai modellben, tehát egyértelműen az arám-középiráni íráscsoportból származnak a jellegek, ennek megfelelően a $P_{arám-középiráni}$ oszlopainak száma 33. Hasonlóan a $P_{bráhmí}$ oszlopainak száma 29. Mindkét mátrixban a sorok rendre azt jelzik, hogy az oszlopoknak megfelelő jellegek szerepelnek-e a vizsgált taxonokban (TR, SZMR, KMR és SR).

$$P_{arám-középiráni} = \begin{pmatrix} 110101001101000100111110111011100 \\ 100001001011000100100100111010000 \\ 001011100100101110100111110000001 \\ 001101110000111111101011111110110 \end{pmatrix} \quad (1)$$

$$P_{bráhmí} = \begin{pmatrix} 10110010110111010011010001010 \\ 01001000101000100101101101100 \\ 11000100000000001000100110100 \\ 11000001000010101101000100101 \end{pmatrix} \quad (2)$$

A kladisztikai analízis előkészítése

Mivel mind a négy vizsgált írás (TR, SZMR, KMR és SR) jellegei között szerepel mind az arám-középiráni, mind a bráhmi íráscsoport hatása, ezért kézenfekvő feltételezések, hogy a négy írás teljes szétválása előtt hatott rájuk mind a két íráscsoport. Így lehetőség nyílik arra, hogy az arám-középiráni és a bráhmi hatást egyesítsük, s egy kladisztikai elemzést úgy végezzünk el, hogy az arám-középiráni és bráhmi hatásokat megtestesítő jellegek összességét hordozó *protorovást mint kulcsoportot tekintjük, és ehhez képest nézzük az egyes vizsgált taxonok fejlődését, s ez az arám-középiráni és bráhmi jellegek részbeni, fokozatos elvesztését jelenti.

A közös arám-középiráni és bráhmi íráscsoport bevezetésének járulékos haszna, hogy bizonyos SFG-ket, amelyeknél nem volt egyértelmű, hogy a bennük szereplő jellegek az arám-középiráni íráscsoportból vagy a bráhmi írásból származnak, és ezért ki kellett volna zárni az elemzésből, az arám-középiráni és bráhmi hatás egyesítésével egyértelműnek tekinthetünk. Az olyan SFG-k száma 72, amelyek egyértelműen az arám-középiráni-bráhmi íráscsoportból származnak. Ez jelentősen több, mint az (1)-ből és (2)-ből adódó $33 + 29 = 62$; így a felhasználható SFG-k száma, vagyis az elemzés felbontása nőtt.

Az arám-középiráni és bráhmi jellegeket hordozó *protorovás segítségével vizsgált taxonokra (TR, SZMR, KMR és SR) lehetőség nyílt egy kladisztikai elemzés elvégzésére. A kladisztikai analízis célja a tanulmányozott írások törzsfelődésének evolúciós elágazási sorozatára vonatkozó következtetések levonása. A kladisztikai analízist egy Python nyelven készült szoftver végzi el és eredményét kladogram alakjában jeleníti meg. A szoftver bemenete egy olyan adatmátrix, amelyben a sorok a taxonoknak (esetünkben írások), az oszlopok pedig az SFG-k által meghatározott jellegeknek felelnek meg. Az adatmátrix bináris, vagyis az elemeinek értéke 0 vagy 1, mivel a jellegek egy-egy taxonbeli megléte vagy hiánya (jellegállapot) tartalmazza az információt. Így egy taxon jellegállapotainak sorozata egy bitsorozatot jelent.

Módszer

Az eredeti Wagner-módszer

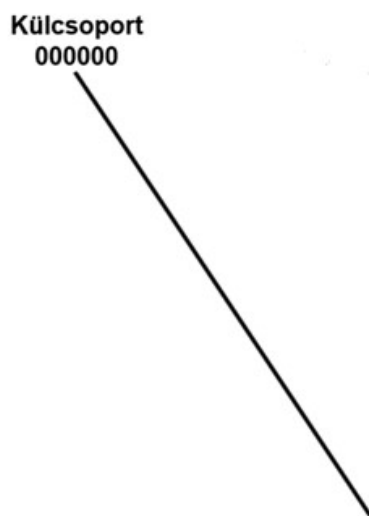
A kladogram, vagyis az a dendrogram, amelyet egy kladisztikai elemzés eredményeként kapunk, többféleképpen létrehozható. Az egyik megoldást Hennig dolgozta ki (Hennig 1966). Eljárásának lényege, hogy ismertek a jellegek ősi állapotai, és a jellegállapotok változásai visszafordíthatatlanok. Ezért egy kladogramon belül a jellegek csak egyszer változhatnak. Ha ezek a feltételek együtt nem teljesülnek, akkor enyhíteni kell Hennig feltételezésein, és valamilyen stratégia, pl. a parszimónia alapján lehet meghatározni egy elfogadható kladogramot (Podani 1997, 177.). A parszimónia alkalmazásának egyik eszköze lehet a Wagner-módszer, amelyet Kluge és Farris alkotott meg 1969-ben (Farris 1969; Farris 1970). Ekkor a kladogramot a taxonok fokozatos hozzácsatolásával hozzuk létre úgy, hogy mindig a lehető legkevesebb jellegállapot-változást próbáljuk elérni, amikor kiválasztjuk, melyik taxont hova kapcsoljuk. Ennél a módszernél nem kell az ősi jellegállapotokat meghatározni. A Wagner-módszer azt feltételezi, hogy az evolúció minimális számú állapotváltozással halad (parszimónia elve). A kidolgozott szoftver a Wagner-módszer egy, az igények szerint módosított változatát alkalmazza. Ugyanis az eredeti módszer szerint a Wagner-fa készítésének első lépése megkeresni a külcsoportot (*outgroup*): ez azon taxon (írás), amelynek minden jellegállapota nulla (Farris 1982). Esetünkben azonban a külcsoport a *protorovás (Hosszú 2019, 390–392.), amelynek minden arám-középiráni-bráhmii eredetű jellegállapota 1, és a taxonok fejlődése ezen jellegállapotok egy részének fokozatos 0-vá alakulása.

A Wagner-fa szerkesztésekor egy új belső csomópont egyik jellegállapota akkor lesz 1, ha azon a helyen minden levelén 1 van, és akkor lesz 0, ha minden levelének megfelelő jellegállapota 0. Ha a leveleinek érintett jellegállapotai nem egyeznek meg (1 és 0 vagy 0 és 1), akkor a fában előtte található csomópont fog dönteni, vagyis ha az előtte álló csomóponton 1 van, akkor 1 lesz a fához most hozzáadandó belső csomópont megfelelő jellegállapota, ha pedig az előtte lévő csomópont megfelelő jellegállapota 0, akkor az új belső csomópont megfelelő

jellegállapota is 0 lesz. Az eredeti Wagner-fa szerkesztésének bemutatásához tekintsük a 3. táblázattal megadott példát, amelyben az A, B, C és D vizsgált taxonokat jelöl. Az eredeti Wagner-módszer 1. lépésének eredményét a 3. ábra szemlélteti.

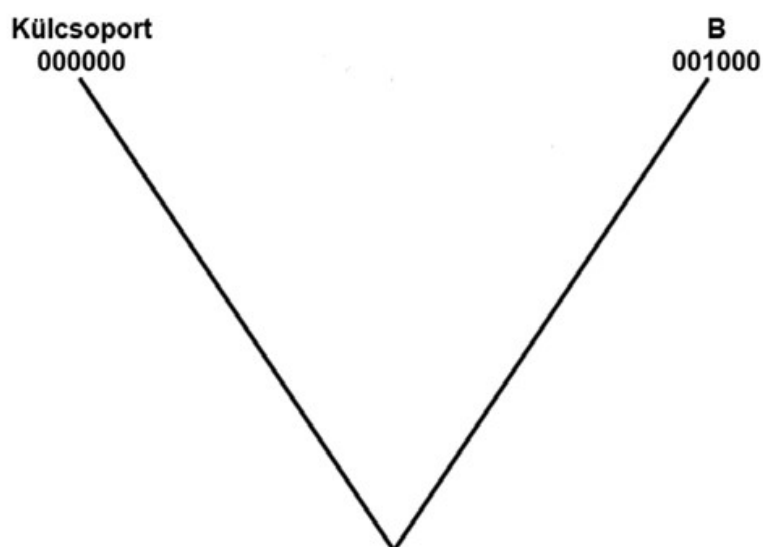
Taxon	Jelleg1	Jelleg2	Jelleg3	Jelleg4	Jelleg5	Jelleg6
Külcsoport	0	0	0	0	0	0
A	1	1	1	0	0	1
B	0	0	1	0	0	0
C	1	0	1	0	0	0
D	1	0	1	0	0	1

3. táblázat: Példamátrix



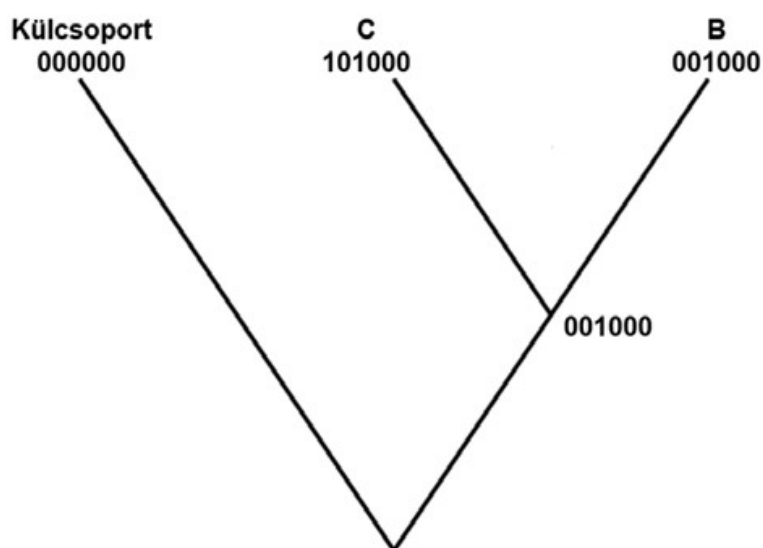
3. ábra: Az eredeti Wagner-módszer 1. lépése

Az ezt követő lépés annak a kikeresése, melyik taxonban van a legkevesebb származtatott állapot, azaz olyan tulajdonság, amely a külcsoporthoz képest megkülönbözteti a taxont. Ebben az esetben ez azt jelenti, hogy a mátrixban a tulajdonság alatt '1' található, azaz a tulajdonság megjelenik a taxonban, lásd 4. ábra.



4. ábra: A Wagner-módszer 2. lépése

A Wagner-módszer következő lépése megnézi a megmaradt, mátrixbeli taxonok közül a legkevesebb leszármazott állapotot tartalmazó taxont, és hozzárajzolja a kladogramhoz, majd a kettő találkozási pontjánál megjelöli, hogy melyek a közös leszármazott állapotaik. Tehát például, ha van egy C taxonunk, amelynek a jellegállapotai: (101000), és egy B taxonunk (001000) jellegállapottokkal, akkor a találkozási pontjukra a (001000) írható, lásd 5. ábra.



5. ábra: A Wagner-módszer 3. lépése

A Wagner-módszer ezután ismét kikeresi a maradék taxonok közül a legkevesebb leszármazott állapottal rendelkező taxont, viszont most már kérdéses, hova tudja hozzáilleszteni a fához. Ehhez használatos a fa egyes ágaihoz rendelt hosszúság mérésére szolgáló Wagner-képlet, lásd (3).

$$d(A, B) = \sum_i |X(A_i) - X(B_i)| \quad (3)$$

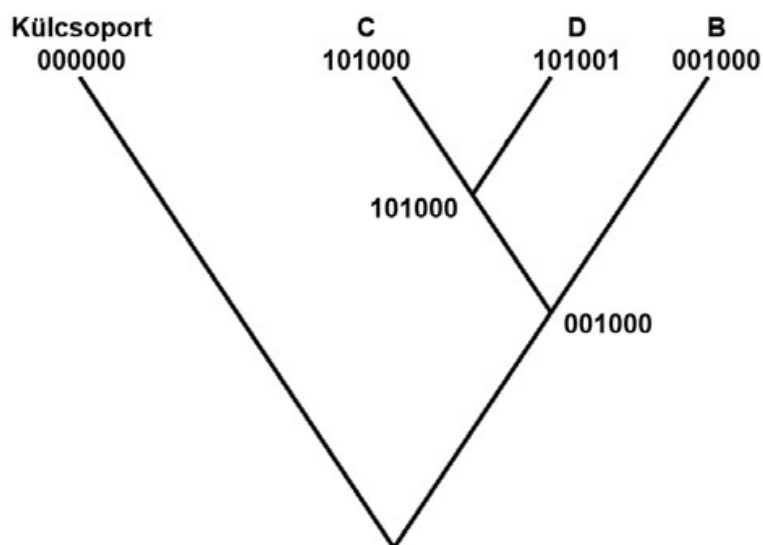
ahol $d(A, B)$ az A és B közötti jellegállapotok változásainak száma, $X(A_i)$ az A taxonhoz tartozó i -edik jelleg állapota és $X(B_i)$ az B taxonhoz tartozó i -edik jelleg állapota.

Ennek alapján összehasonlítjuk a soron következő vizsgált taxont az összes lehetséges fához való csatlakozási lehetőséggel. Vegyünk példaként két, B és C nevű taxont egy fában. A $B(001000)$ és a $C(101000)$ taxonok neve utáni zárójelben az egyes jellegek állapota látható. Ennek a két taxonnak a közös csatlakozási pontja a $BC(001000)$; a közös csatlakozási pont is leírható a jellegállapotok sorozatával. A Wagner-módszer következő lépésében az előző lépésben kapott fához megpróbálja hozzácsatlakoztatni a $D(101001)$ taxont. A D taxon a B -től két jellegállapotban, a C taxontól egy jellegállapotban, míg a BC -től szintén két jellegállapotban tér el (4. táblázat).

Taxon	Jellegállapotok	Különbség
D	101001	-
B	001000	2 db
C	101000	1 db
BC	001000	2 db

4. táblázat: A Wagner-módszer 3. lépésében kapott gráf tulajdonságai

A 4. táblázat alapján a D -t közvetlenül a C taxonhoz kell csatlakoztatni. Vagyis ebben a lépésben is a legkevesebb leszármazott állapottal rendelkező taxont kell kikeresni, és a Wagner-képlet alapján eldönteni, hogy hova kerüljön beillesztésre a gráfon belül. Ekkor létrejön a $DC(101000)$ csatlakozási pont (6. ábra).



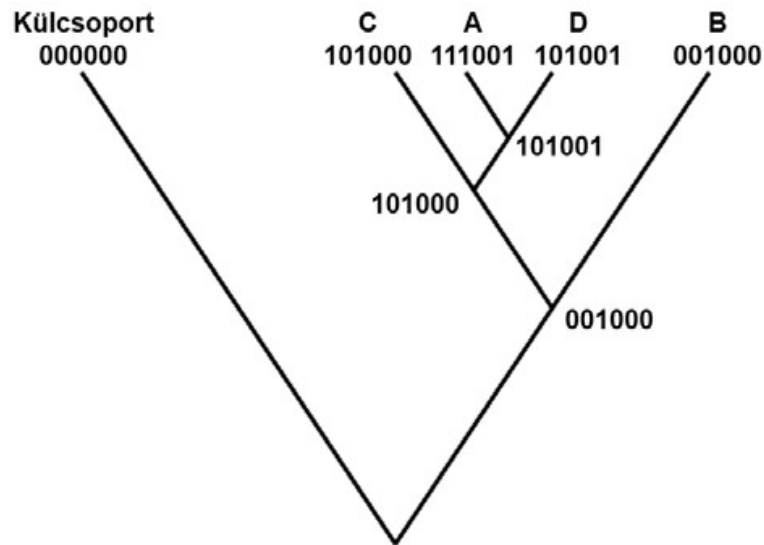
6. ábra: A Wagner-módszer 4. lépése

A Wagner-módszer 4. lépésében kapott gráfot (dendrogramot) az 5. táblázat írja le.

Taxon	Jellegállapotok	Különbség
A	111001	–
D	101001	1 db
B	001000	3 db
C	101000	2 db
CD	101000	2 db
BCD	001000	3 db

5. táblázat: A Wagner-módszer 4. lépésében kapott gráf tulajdonságai

Az 5. táblázat alapján a fentiekben leírtak szerint kerül csatlakoztatásra a következő (A) taxon (7. ábra).



7. ábra: Az algoritmus 5. lépése

A *B* taxon, amely először leválik, a harmadik elemében különbözik a kulcsoporttól, ez egy változásnak minősül (4. ábra). Ezután a *C* taxon válik le, amely az első elemében különbözik az előzőtől, ez plusz egy jellegállapot-változás (5. ábra). A *D* taxon a hatodik, az *A* taxon pedig a második elemében különbözik az előtte lévő jellegállapottól (6. és 7. ábra). A következő felsorolásban vastagítottan jelöljük azokat a jellegeket, amelyek az adott taxonban leszarmazott jellegállapotok a kladogram előző szintjéhez képest: *A*(11**1**001), *B*(00**1**000), *C*(**1**01000), *D*(10**1**001). Ez összesen négy jellegállapot-változás, így a korábban meghatározott Wagner-fa hossza $L=4$. A kulcsoport esetén minden jellegállapot (bit) zérus: (000000).

Megjegyzendő, hogy változásnak számít a 0-ról 1-re, és az 1-ről 0-ra váltás is. Ha lenne olyan jellegállapot-változás a példában, ahol a leszarmazott jellegállapot 0, pedig az elődjénél ugyanennek a jellegnek az állapota 1 (pl.: *A*(11**0**001), *D*(10**0**001)) akkor az szintén jellegállapot-változásnak számítana, az L hosszhoz ugyanúgy hozzáadódna.

A Wagner-módszer lépései addig folytatódnak, amíg van olyan taxon a kiindulási adatmátrixban, amely még nem lett elhelyezve a fába. Előfordulhat, hogy a különbség két opcionál megegyezik, ebben az esetben az elsőt választjuk.

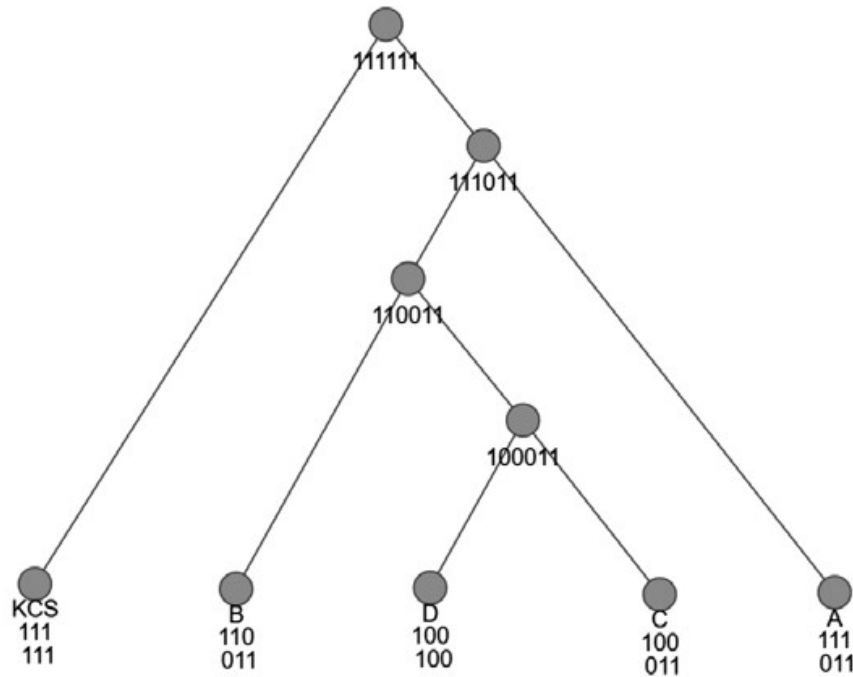
Ezt azért tehetjük meg, mert a Wagner-módszerrel kapott gráf egy kezdetinek tekinthető kladogram, amelyet csupán arra használunk, hogy élhosszúságainak összegével felső határt szabjunk arra, hogy milyen összes élhosszúság várható a majd meghatározandó optimális, ún. legparszimónikusabb kladogramtól. Ez csak felső határ, mivel a kladogramot készítő Wagner-módszer mohó algoritmus, vagyis mindig helyileg optimalizál (Hosszú 2005, 84–85). Ezért az általa kialakított gráf nem feltétlenül legparszimónikusabb megoldás, vagyis nem biztos, hogy a lehető legkevesebb jellegállapot-változást tartalmazó kladogram. Így szükség van egy további másik algoritmusra, amely a Wagner-módszerrel kapott kezdeti kladogramot optimalizálja a legparszimónikusabb kladogram megtalálása érdekében.

A módosított Wagner-módszer

A kifejlesztett szoftver ugyan a Wagner-fa alkotásának elvét veszi alapul, de ettől eltér annyiban, hogy a külcsoport csupa 0 helyett csupa 1-ből áll, ezért megfordul a hozzáadás sorrendje: minden lépésben a legkevesebb leszármazási állapot helyett a legtöbb leszármazási állapotot tartalmazó taxon kerül bele a kladogramba. A 3. táblázatbeli példamátrix esetén a sorrend úgy nézne ki, hogy: A, D, C, B; továbbá a belső csomópontok is megváltoznak. Egy egyszerű példa: $KCS(111111)$, $A(111011)$, $B(110011)$, $C(100011)$, $D(100100)$, lásd 8. ábra.

1. Felrajzoljuk a külcsoportot: $KCS(111111)$.
2. Hozzákapcsoljuk az A-t (5 db 1-es).
3. Az A-hoz hozzákapcsoljuk a B-t (4 db 1-es). A harmadik jellegállapotuk eltér, ezért a fában előző csomópont (amely maga a külcsoport) megfelelő jellegének állapota dönt (amely adott, hogy csupa 1), így ide is 1 kerül. A negyedik bitjük azonosan 0, ezért ide is 0 kerül. Közös pontjuk: (111011) .
4. A következő taxon a C (3 db 1-es) ez a B-hez áll legközelebb, ezért oda kötjük be. Közös pontjuk: (110011) , eltérés a második bitnél, egyel korábban (A és B közös belső csomópontja) 1 van, ezért ide is 1 kerül, harmadik bitnél minkettő 0, ezért ide is 0 kerül.

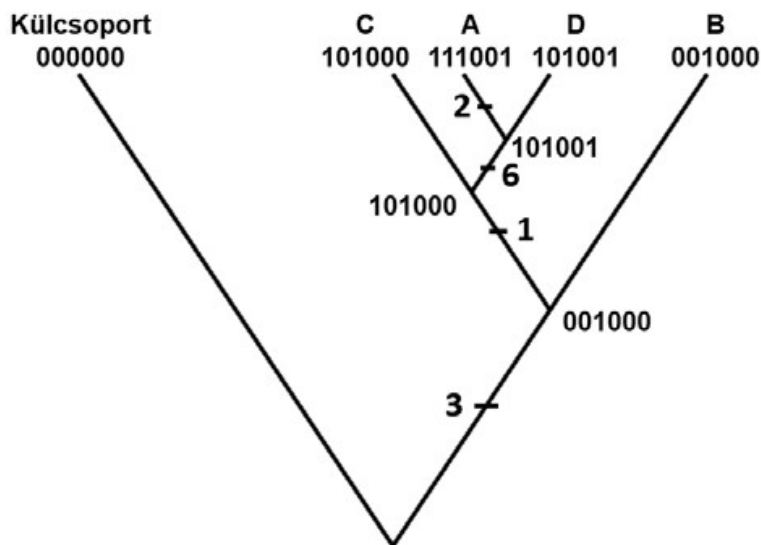
5. Már csak a D (2 db 1-es) van hátra, ez C-hez áll a legközelebb. Közös pontjuk: (100011). Eltérések: negyedik bit: előzőnél (C-B) itt 0 van, ezért ide is 0 kerül, ötödik bit: előzőben itt 1 van ezért ide is 1 kerül, hatodik bit: előzőben itt 1 van, ezért ide is 1 kerül.



8. ábra: Egyszerű példa a módosított Wagner-módszerre

A kladogram optimalizálása

Ahhoz, hogy eldönthessük, melyik fa a legparszimónikusabb, az eredeti vagy módosított Wagner-módszerrel kapott fa (Wagner-fa) L hosszát kell megnéznünk. A Wagner-fa L hossza az a skaláris mennyiség, amellyel azt jellemezzük, hány jellegállapot-változás ment végbe a fán. Ez azon jellegállapot-változások összegzéséeként kapható, amelyek ahhoz szükségesek, hogy eljussunk az összes csomóponttól a gyökérig. Minél kisebb az L hossz, azaz minél kevesebb jellegállapot-változásra van szükség, annál parszimónikusabb a fa. A fentiekben kiszámított Wagner-fában a 9. ábrán látható jellegek változnak.



9. ábra: A Wagner-fa hosszának számítása

Az ág és kötés (*branch and bound*) módszeren (Hendy–Penny 1982) alapuló kladogram-optimalizáló algoritmus úgy működik, hogy kiszámít minden olyan lehetséges faelrendezést, amelynek hossza kisebb a korábban megkapott Wagner-fa hosszánál. Vagyis egy-egy faelrendezés kialakítása közben a kladogram-optimalizáló algoritmus folyamatosan számítja a még csak részben kialakított faelrendezés hosszát. Ha ez a hossz meghaladja a kezdeti Wagner-fa hosszát, akkor tudható, hogy az adott – félkész – faelrendezés biztosan nem optimális, így újabb taxonok figyelembevételével nem kell tovább folytatni a fa kiszámítását. Ezzel jelentősen gyorsul a kladogram-optimalizáló algoritmus működése.

Miután a kladogram-optimalizáló algoritmus létrehozta az összes, a kezdeti Wagner-fa hosszánál nem nagyobb hosszúságú faelrendezést, közülük kiválasztja a legkisebb hosszúságút, mert ez lesz a parszimónikusabb kladogram. Előfordulhat, hogy nem egyetlen legrövidebb fa létezik, ha két, különböző fa hossza megegyezik, ebben az esetben a szoftver az összes ilyen eltárolja és ki-rajzolja.

A kladogram ellenőrzése

A kladogramok minőségét nemcsak a korábban bevezetett hosszúságukkal lehet mérni, hanem azzal is, hogy a jellegállapotok között mekkora a homopláziák aránya. A konzisztenciaindex azt adja meg, hogy a jellegváltozások száma hogyan aránylik az elméletileg elképzelhető minimumhoz. Mértéke az állapotváltozások minimális száma (figyelembe véve az adatokban szereplő jellegek számát), osztva az állapotváltozások aktuális számával (Kluge–Farris 1969; Podani 1997, 197–198). Egyetlen jellegre is értelmezhető, de kiszámítható egy τ kladogram összes jellegére is. Az utóbbi esetben a neve: *átlagos konzisztenciaindex* (4).

$$CI(\tau) = \frac{\sum_{j=1}^n m_j}{\sum_{j=1}^n s_j}, \quad (4)$$

ahol a τ kladogram j -edik jellegállapotával ($j=1,2,\dots,n$) a fában s_j változás történik, pedig az adatokra felírható egy olyan fa is, amelyben a j -edik jelleg állapotában csak minimális számú, m_j változás áll be.

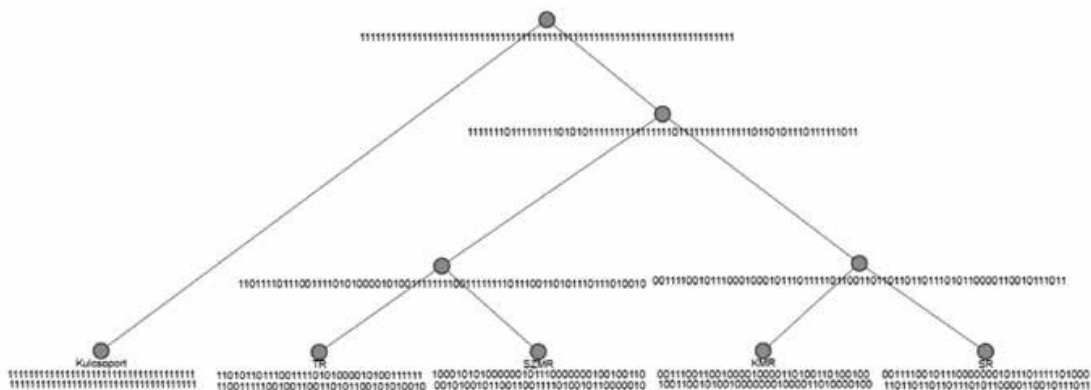
A τ kladogramban lévő szünapomorfiák mennyiségét méri az *összetartási index*, amelynek mértéke: a bármely fán mérhető állapotváltozások maximális száma (jellegenként az 1-as állapotú vagy a 0-s állapotú taxonok számának összege attól függően, melyik a kisebb, összegezve az összes jellegre) mínusz az állapotváltozások száma az aktuális fán, osztva a bármely fán mérhető lépések maximális száma és az állapotváltozások minimális számának különbségével (Farris 1989: 417), lásd (5).

$$RI(\tau) = \frac{\sum_{j=1}^n (M_j - s_j)}{\sum_{j=1}^n (M_j - m_j)}, \quad (5)$$

ahol M_j a j -edik jelleg lehetséges megváltozásainak maximális száma. Ennek értéke annál nagyobb, minél kisebb a homoplázia részesedése a szünapomorfiák kialakulásában. Ha nincs homoplázia, akkor $RI(\tau)=1$; ha pedig az összes szünapomorfiát homoplázia okozza, akkor $RI(\tau)=0$ (Podani 1997, 198.).

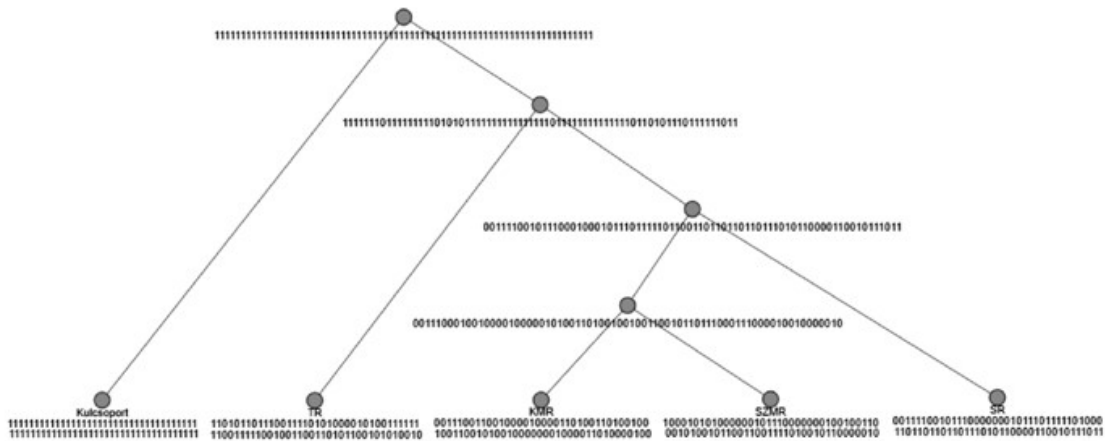
Vizsgálati eredmények és értékelésük

Az elvégzett vizsgálatok során a négy vizsgált írás (TR, SZMR, KMR és SR) korábban meghatározott két elődjét (bráhmi írás és az arám-középiráni elődírás-csoport) egyetlen íráscsoportban kezeltük és feltételeztük, hogy ez az egyesített íráscsoport a négy vizsgált írás közös elődjére, az ún. *protorovásra hatott. Az egyértelműen arám-középiráni-bráhmi eredetűnek minősíthető, és ezért az elemzésben felhasználható SFG-k száma 72, amelyből 33 arám-középiráni, 29 bráhmi és 10 vagy arám-középiráni vagy bráhmi eredetű. A módosított Wagner-módszert alkalmazó szoftverrel számított Wagner-fa (τ_0 , 10. ábra) jellemzői: hossza 111, 60, $CI(\tau_0)=60$, $RI(\tau_0)=21$.



10. ábra: A Wagner-fa (τ_0)

A kladogram-optimalizálás után kapott legparszimónikusabb fa (τ_1 , 11. ábra) jellemzői: hossza 101, $CI(\tau_1)=66$, $RI(\tau_1)=39$.



11. ábra: A legparszimónikusabb (τ_1) fa

Következtetések

A cikkben ismertetett eljárás az írások leszármazási kapcsolatainak modellezésére kidolgozott, evolúciós megfontolásokkal kiegészített, ún. kiterjesztett fenetikai modell kladsztikai elemzéséhez használható olyan esetekben, amelyekben a kiterjesztett fenetikai modell kiértékelése során végzett elődanalízisből az az eredmény adódott, hogy a vizsgált írások fejlődése nem egy elődtől, hanem többtől indult, és a fejlődés egy adott pontján írásszintű hibridizáció történt. A kladsztikai elemzést megelőző fenetikai és elődanalízis azért volt szükséges, mert igen hiányosak a vizsgált írásokhoz (mint taxonokhoz) rendelkezésre álló adatok: néhány vizsgált íráshoz csak kevés írásemlék ismert, ráadásul azok nagy részének a kora ismeretlen, vagy csak durva közelítéssel határozható meg. A vizsgált írásoknak a kiterjesztett fenetikai modellben szereplő jellegeinek alkalmas szétválogatásával lehetővé vált, hogy egy elődírás csoporttól való függésük önállóan legyen vizsgálható.

Az elvégzett vizsgálatok során egy olyan szoftver lett kifejlesztve, amely megfelelően strukturált bementi adatokat a gépi tanulásból ismert adatfeldolgozó módszerekkel elemzi, majd a kapott eredményeket kladogramokban megjeleníti. A jelenlegi adatbázisra épülő modell alapján számított kladogram

szerint a KMR és SZMR állnak egymáshoz a legközelebb, és evolúciósan az SR közelebb áll hozzájuk, mint a TR. A kidolgozott kladisztikai eljárás azt mutatja, hogy az eredetmodellekkel kibővített, kiterjesztett fenetikai modellre épülve kladisztikai eszközökkel is lehet vizsgálni a tanulmányozott taxonok törzsfelődése evolúciós elágazási sorozatát. A kapott eredmények alkalmasak lehetnek arra, hogy más tudományágak felhasználják pl. ismeretlen feliratokon használt írások meghatározása során.

IRODALOMJEGYZÉK

- Ashlock 1974.** Peter D. Ashlock: The Uses of Cladistics. *Annual Review of Ecology and Systematics*, 5. (1974) 81–99.
- Farris 1969.** James S. Farris: A successive approximations approach to character weighting. *Systematic Zoology*, 18. (1969) 4. sz. 374–385.
- Farris 1970.** James S. Farris: Methods for Computing Wagner trees. *Systematic Zoology*, 19. (1970) 1. sz. 83–92.
- Farris 1982.** James S. Farris: Outgroups and parsimony. *Systematic Zoology*, 31. (1982) 3. sz. 328–334.
- Farris 1989.** James S. Farris: The retention index and the rescaled consistency index. *Cladistics*, 5. (1989) 4. sz. 417–419.
- Hendy–Penny 1982.** M. D. Hendy – David Penny: Branch and Bound Algorithms to Determine Minimal Evolutionary Trees. *Mathematical Biosciences* 59. (1982) 2. sz. 277–290.
- Hennig 1966.** Willi Hennig: *Phylogenetic systematics*. University of Illinois, Urbana (IL), 1966.
- Hosszú 2005.** Hosszú Gábor: *Az internetes kommunikáció informatikai alapjai*. Novella Kiadó, Budapest, 2005.
- Hosszú 2010.** Hosszú Gábor: Az informatika írástörténeti alkalmazásai. (Meghívott plenáris előadás) In: Cserny László (szerk.): *Informatika Korszerű Technikái Konferencia*. (2010. március 5–6.). Dunaújvárosi Főiskola Informatikai Intézet, Dunaújváros, 2010. 5–21.
- Hosszú 2014a.** Hosszú Gábor: Topológiai eltérések minimalizálására visszavezetett graféma leszármazási vizsgálatok. In: Cserny László – Hadaricsné Dudás Nóra – Nagy Bálint (szerk.): *Informatika Korszerű Technikái Konferencia* (2012. november 16–17.). Dunaújvárosi Főiskola – Új Mandátum Könyvkiadó, Budapest, 2014. 60–71.
- Hosszú 2014b.** Gábor Hosszú: Mathematical Statistical Examinations on Script Relics. In: V. Bhatnagar (ed.): *Data Mining and Analysis in the Engineering Field*. Chapter 8. Information Science Reference, Hershey–New York, 2014. 142–158.
- Hosszú 2015.** Gábor Hosszú: A Novel Computerized Paleographical Method

- for Determining the Evolution of Graphemes. In: Mehdi Khosrow-Pour (ed.): *Encyclopedia of Information Science and Technology*. Third edition. Chapter 194. Information Science Reference, Hershey–New York, 2015. 2017–2031.
- Hosszú 2017.** Gábor Hosszú: Phenetic Approach to Script Evolution. In: Hannah Busch – Franz Fischer – Patrick Sahle (eds.): *Kodikologie und Paläographie im Digitalen Zeitalter 4*. Codicology and Palaeography in the Digital Age 4. Books on Demand, Norderstedt, 2017. 179–252.
- Hosszú 2019.** Hosszú Gábor: Íráslelékek grafémaalakjainak térstatisztikai és fenetikai elemzése. In: Zelliger Erzsébet (szerk.): *Rovás – magyar nyelvtörténet – művelődéstörténet*. Magyarságkutató Intézet, Budapest, 2019. 120–450.
- Hosszú 2020.** Hosszú Gábor: Evolúciós számítások írásinformatikai alkalmazása. In: Fehér Bence – Ferenczi Gábor (szerk.): *Ősi írásaink*. Tanulmánykötet a Magyarságkutató Intézet által 2019. december 12–13-án rendezett konferencián elhangzott előadásokból. Magyarságkutató Intézet, Budapest, 2020. 51–76.
- Hosszú 2021a.** Gábor Hosszú: *Scriptinformatics*. Extended Phenetic Approach to Script Evolution. Nap Kiadó, Budapest, 2021. http://napkiado.hu/media/Hosszu_Gabor-Scriptinformatics.pdf
- Hosszú 2021b.** Gábor Hosszú: Data-Driven Phenetic Modeling of Scripts’ Evolution. In: Shifeng Liu – Gábor Bohács – Xianliang Shi – Xiaopu Shang – Anqiang Huang (eds.): *LISS 2020*. Proceedings of the 10th International Conference on Logistics, Informatics and Service Sciences. Chapter 28. Springer, Singapore, 2021. megjelenés alatt.
- Kluge–Farris 1969.** Arnold G. Kluge – James S. Farris: Quantitative Phyletics and the Evolution of the Anurans. *Systematic Zoology*, 18. (1969) 1. sz. 1–32.
- Parkinson 1999.** Richard Parkinson: *Cracking Codes*. The Rosetta Stone and Decipherment. University of California Press, Berkeley–Los Angeles, 1999.
- Podani 1997.** Podani János: *Bevezetés a többváltozós biológiai adatfeltárás rejtelmeibe*. Scientia Kiadó, Budapest, 1997.
- Podani 2007.** Podani János: *A szárazföldi növények evolúciója és rendszertana*. ELTE Eötvös Kiadó, Budapest, 2007.

- Podani–Morrison 2017.** János Podani – David A. Morrison: Categorizing ideas about systematics: alternative trees of trees and related representations. *Rendiconti Lincei. Scienze Fisiche e Naturali*, 28. (2017) 1. sz. 191–202.
- Sampaio, 2007.** Alberto Sampaio: Software Phenetics, Phylogeny and Evolution. In: *Third International IEEE Workshop on Software Evolvability 2007*, Paris, France, 2007. 60–66.
- Torres-Montúfar és mtsai. 2018.** Alejandro Torres-Montúfar – Thomas Borsch – Helga Ochoterena: When Homoplasy Is Not Homoplasy: Dissecting Trait Evolution by Contrasting Composite and Reductive Coding. *Systematic Biology*, 67. (2018) 3. sz. 543–551.