

## IRODALOM

- Beklemishev, V. N. (1969): *Principles of Comparative Anatomy of Invertebrates*, Vol. 1: *Promorphology*. Oliver and Boyd, Edinburgh
- Clarke, Shoa L. – VanderMeer, Julia E. – Wenger, Aaron M. et al. (2012): Human Developmental Enhancers Conserved between Deuterostomes and Protostomes. *PLOS Genetics*. 8(8), e1002852 DOI: 10.1371/journal.pgen.1002852 journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1002852
- Davidson, Eric H. (2010): Emerging Properties of Animal Gene Regulatory Networks. *Nature*. 468, 911–920. DOI:10.1038/nature09645
- Davidson, Eric H. – Erwin, Douglas H. (2006): Gene Regulatory Networks and the Evolution of Animal Body Plans. *Science*. 311, 796–800. DOI:10.1126/science.1113832

- Hoerner, Sighard F. (1965): *Fluid-dynamic Drag*. Hoerner Fluid Dynamics, Brick Town, NJ <http://dl.kashti.ir/ENBOOKS/NEW/FDD.pdf>
- Holló Gábor (2014): Animals Are Both Radially and Bilaterally Symmetrical: Accommodating Seemingly Mutually Exclusive Paradigms. *BioEssays*. 36, 901–902. DOI:10.1002/bies.201400089
- Holló Gábor (2015): A New Paradigm for Animal Symmetry. *Interface Focus*. 5, 20150032 DOI: 10.1098/rsfs.2015.0032 • <http://rsfs.royalsocietypublishing.org/content/5/6/20150032>
- Holló Gábor – Novák Mihály (2012): The Manoeuvrability Hypothesis to Explain the Maintenance of Bilateral Symmetry in Animal Evolution. *Biology Direct*. 7, 22. • <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3438024/>
- Vogel, Steven (1994): *Life in Moving Fluids*. 2<sup>nd</sup> edition. Princeton University Press, Princeton, NJ



# FAJOK ÉS GÉNEK TÖRTÉNETÉNEK NYOMÁBAN

Gertheis Zsófia N. Szöllösi Gergely J.

doktorandusz, PhD, tudományos főmunkatárs,  
ELTE–MTA „Lendület” Evolúciós Genomika Kutatócsoport  
ELTE Biológiai Fizika Tanszék szzolo@gmail.com

## Bevezetés

A biológia történeti tudomány. A biológiai rendszerek szerkezete, a fehérjéktől egészen az ökoszisztémáig mind egy hosszú, több mint hárommilliárd évvel ezelőtt kezdődött evolúciós folyamat eredménye. Bár az evolúció folyamatainak és múltbeli lefolyásának tanulmányozását hagyományosan a többi biológiai diszciplínától különállóként kezelték, az a gondolat, hogy az élet története kulcsfontosságú az élő rendszerek megértésében, mára kezd széles körben elfogadottá válni.

A DNS-láncok nukleotidsorozatának megfejtésére kidolgozott technológiák rohamos fejlődése alapjaiban változtatta meg a biológia forrásait és kérdéseit. Különösen igaz ez az élővilág történetének kutatására. A DNS-szekvenciák hiányában néhány évtizede még a fajok származását, őseik testfelépítését és életmódját a ma élő egyedek közös és eltérő tulajdonságai, valamint fennmaradt fosszíliaik alapján próbáltuk meg felderíteni. Ma már a fajok teljes genetikai állománya és ezáltal az őseiktől örökölt gének közötti hasonlóságok és különbségek sokasága áll rendelkezésünkre. Ezek a DNS-szekvenciák a múlt dokumentumai (Zuckerlandl – Pau-

ling, 1965; Boussau et al., 2010), amelyek segítségével minden eddiginél részletesebben rekonstruálhatjuk az élet történetét és érthetjük meg ma fellelhető diverzitását.

## A szekvenálástól a géncsaládokig

Az örökítőanyag minden egyes szaporodási ciklusban megkettőződik, így adódik tovább (kisebb-nagyobb másolási hibákkal, mutációkkal) az utódoknak, akik azt saját utódaiknak örökítik, újabb módosulásokkal. Ily módon kerül a DNS-láncok nukleotid sorozatába, a DNS-szekvenciákba az élőlények leszármazási történetének lenyomata. A molekuláris filogenetika feladata, hogy csupán egyetlen jelenbeli pillanatképből, a laboratóriumban megfejtett szekvenciák közötti különbségekből kikövetkeztesse ennek a folyamatnak a lépéseit, rekonstruálja a hozzájuk vezető osztódások sorozatát.

Ezekből a pillanatképekből azonban nagyon sok áll rendelkezésünkre, mára több tízezer élőlény teljes genetikai állománya (genomja) ismert. Az evolúciós múlt DNS-szekvenciák alapján történő rekonstrukciójában az első feladatunk, hogy a genom-szekvenciákat, amelyek önmagukban csak a négy nukleotidból (A, T, G, C) álló hosszú sorozatok, valamilyen módon értelmezzük. Ezt a

genomok automatizált számítógépes elemzése segítségével (annotálásával) tesszük. Ennek a folyamatnak kulcsfontosságú lépése a gének predikciója.

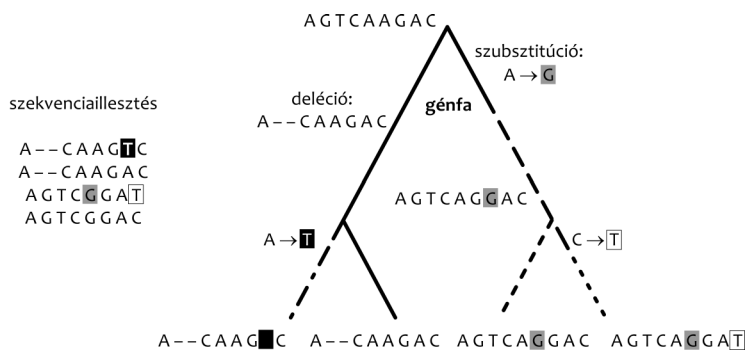
A filogenetikai rekonstrukció elemi egységének ugyanis a fehérjét (és más funkcionális makromolekulákat, például riboszómális vagy transzfer-RNS-eket) kódoló géneket tekintjük. Ennek az oka az, hogy ezekre a szekvenciákra olyan funkcionális egységekként gondolunk, amelyekről jó közelítéssel feltételezhetjük, hogy részei (végső soron az egyes nukleotidok) közös evolúciós történettel rendelkeznek.

A rekonstrukció következő lépése, hogy a génszekvenciák tömegében megkeressük a gének olyan *homológ* csoportjait, melyek minden tagja egyértelműen közös eredetű a csoport többi tagjával (Boussau et al., 2010). Egy-egy ilyen homológ csoport, egy gén család általában több különböző faj hasonló funkcióval rendelkező génjeit tömöríti. Ilyen például a citokróm-c fehérjét kódoló gének családja. Előfordul, hogy ugyanazon faj két vagy több rokon génje is megtalálható egy „többpéldányos” családban: erre példa az emberi hemoglobin oldalláncok családja.

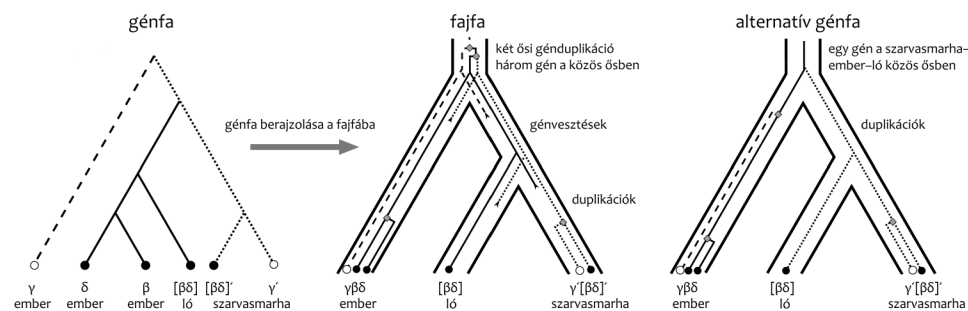
Hogy az egy gén családba tartozó szekvenciák közötti különbségeket megtaláljuk, meg

kell tudnunk mondani, hogy egy adott génben melyik nukleotid melyik másik nukleotidnak feleltethető meg a család többi génjében. Ehhez egy ún. szekvenciaillesztési (i. ábra) táblázatban a szekvenciákat egymás alá írjuk úgy, hogy a táblázat egyes oszlopaiba kerüljenek azok a szekvenciapozíciók, melyek a gén család közös ősi szekvenciájában ugyanazon szekvenciapozícióra vezetnek vissza történetüket. A szekvenciaillesztés nehézsége abban rejlik, hogy a szekvenciák evolúciója során kisebb-nagyobb szekvenciabeszúrások és -kivágások történhettek. Ennek eredményeképpen a gén család szekvenciái általában különböző hosszúságúak, és mivel a szekvenciabeszúrások és -kivágások történetét *a priori* nem ismerjük, nagyon sok szekvenciaillesztési táblázat lehetséges. Ezt a problémát a gyakorlatban úgy oldjuk meg, hogy azt a táblázatot választjuk, amely a lehető legkevesebb törés beiktatásával illeszti a lehető leghasonlóbb karaktereket egy oszlopba.

A szekvenciaillesztés birtokában már felvázolhatjuk, milyen események sorozata vezetett a szekvenciák közti különbségek kialakulásához. Ehhez azonban ismernünk és modelleznünk kell a szekvenciát módosító folyamatokat, azaz a nukleotidok cseréjét (szubsztitúcióját), beszúrását és kivágását



1. ábra • Egy gén család evolúciójának története a génfa mentén.



2. ábra • A Zuckerkandl és Pauling által rekonstruált első génfa és annak értelmezése a fajfa kontextusában

(inszercióját és delécióját). Az 1. ábrán bemutatott négy rokonszekvencia létrejöttét például a jobb oldali eseménysor magyarázza meg. Itt a fa elágazásai mind egy-egy DNS-replikációs eseménynek felelnek meg, ahonnan a két új szekvencia más-más úton fejlődött tovább, és különböző mutációk létrejöttével eltávolodtak egymástól. Nagyon sok lehetséges evolúciós történet és ezeknek megfelelő *génfa* létezik, melyek ugyanarra a szekvenciaillesztésben összefoglalt megfigyelésre vezetnek. Ezek közül a legkevesebb eseményt igénylő (szofisztikáltabb matematikai módszereket alkalmazva a legvalószínűbb) génfát a célunk megtalálni.

Az utóbbi évtizedekben a génfa-rekonstrukciós eljárások rohamos tempóban fejlődtek. Mára számos matematikailag kifinomult és biokémiailag jól informált módszer létezik génfa rekonstrukciójára az 1. ábrán illusztrált szekvenciaillesztések alapján (Felsenstein, 2004). A rekonstrukció számára azonban alapvető korlátot szab a rendelkezésre álló szekvenciák, illetve a közöttük lévő különbségek végessége, valamint távolabbi rokonszekvenciák esetén az evolúciós múlt évmilliárdos távlatai. Ennek eredményeként a génfa-rekonstrukció az egyedi gén családok szintjén

jellemzően több, részben hasonló génfával is hasonlóan jól megoldható, ezért csak egy valamelyest elmosódott képet ad a gén család evolúciós történetéről.

#### Minden gén család története egyedi

Emil Zuckerkandl és Linus Pauling ötven éve publikálta az első molekuláris szekvenciák alapján készült génfa-rekonstrukciót a béta-típusú globinok családjáról. Ennek a gén családnak az emberi genomban több tagja van, amelyek különböző életszakaszokban aktivizálódnak, hogy legyártódhasson belőlük az adott életkorban a megfelelő hemoglobin alegység. Zuckerkandl és Pauling három emberi hemoglobin gén – a magzati korban jelenlévő  $\gamma$ -globin, valamint a felnőttekre jellemző  $\beta$ - és  $\delta$ -globin – szekvenciáját hasonlította össze a ló, és két szarvasmarha megfelelő génjeivel. Zuckerkandl és Pauling rekonstrukciójának eredménye a 2. ábra bal oldalán látható.

Ha ezt alaposabban szemügyre vesszük, feltűnik, hogy a család humán, szarvasmarha és ló genomokból származó tagjai nem olyan rokoni kapcsolatban állnak egymással, mint amire számíthatnánk – a  $\beta$ - és  $\delta$ -globinokat tekintve az ember a ló közelebbi rokonának

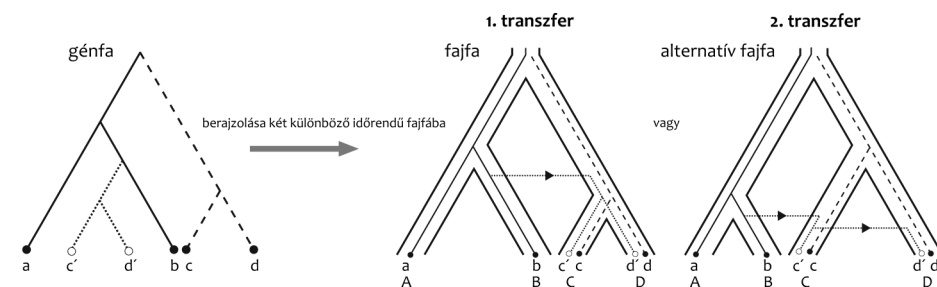
tűnik, mint a szarvasmarha. Ha a  $\gamma$  alegységet is megfigyeljük, még bonyolultabbnak tűnik a helyzet. Téves lenne tehát Zuckerkandl és Pauling híres munkája? Nos, a válasz korántsem ilyen egyszerű: ahhoz, hogy megértsük, hogyan lehet pontos egyszerre az itt látott  $\beta$ -globin fa, és a számításba vett három faj jól ismert rokonsági kapcsolata is, a szemléletünkön kell változtatnunk.

A 2. ábra jobb oldalán található, vastag körvonalú „csőrendszer” a kérdéses három faj leszármazási kapcsolatait ábrázolja: ez fajfa, amelynek elágazásai ősi fajképződési eseményeket jelölnek. A génfák általában a fajfát követik, hiszen a fajképződések során a keletkező két új faj a közös ős génjeit öröklí. A gének sorsa viszont ennél bonyolultabban is alakulhat: végbemehet például a *duplikáció*, egy ritka esemény, amelyet követően a faj genomjában a duplikálódott gén két, eleinte azonos szekvenciájú példánya lesz jelen. Duplikációs eseményeket követően a génfa elágazik, két új ágban fejlődik tovább. Hasonlóan megtörténhet, hogy a faj elveszi az adott gén egy példányát, ebben az esetben a génfa érintett vonala nem ér el a jelenig (Maddison, 1997). Duplikációk és génvesztések sorozatát feltételezve Zuckerkandl és Pauling génfáját is bele tudjuk rajzolni a fajfába, mint ahogy az ábrán látható.

Ma már ismert, hogy főként prokarióta, de néha eukarióta fajok között is előfordul a *horizontális géntranszfer* – erre kiváló példa a mitokondrium és a kloroplasztisz génjeinek bekerülése az eukarióta sejtanyagba (endoszimbiotikus géntranszfer). Egy ilyen esemény azt eredményezi, hogy a génfa vonalai a fajfa ágaira merőlegesen haladnak, és átlépik a távolabbi rokon fajok közötti határokat is (3. ábra), ezzel tovább gazdagítva a lehetséges génfa-berajzolások halmazát.

#### Úton a pontosabb faj- és géntörténetek felé

Mint azt fentebb is említettük, a géncsaládok szekvenciája alapján rekonstruált gének gyakran elmosódtak. Ha a fajfát ismerjük, vagy ismertnek tételezzük fel, nagy mértékben csökkenteni tudjuk ezt az elmosódottságot. Ekkor ugyanis a lehetséges génfák közül nem csak aszerint tudunk választani, hogy melyeket feltételezve van szükségünk a legkevesebb szubsztitúcióra (vö. 1. ábra), hanem az alapján is, melyik génfa magyarázható kisebb számú duplikáció, horizontális géntranszfer és génvesztés segítségével. A 2. ábra példája esetén is ez a helyzet, kiderül ugyanis, hogy a fajfa ismeretében egy alternatív génfa lesz a legvalószínűbb, amelynek a fajfába való berajzoláshoz kevesebb duplikáció- és génvesztés-eseményre van szükség.



3. ábra • Horizontális géntranszfer események mint molekuláris fosszília

De honnan tudjuk, hogyan néz ki a fajfa? A  $\beta$ -globinokon kívül géncsaládok millióit ismerjük. Ezek közül azokban a géncsaládokban, amelyekben duplikáció-, génvesztés- vagy transzfereseemények mentek végbe, azt várjuk, hogy a génfa különböző lesz a fajfától. Valójában a géncsaládok szinte mindegyike ebbe a kategóriába sorolható, olyan eszenciális géncsaládok ritka kivételével, amelyek hiánya drasztikus következményekkel jár az élőlények számára – ilyen például az információkezeléssel (transzkripció, transláció) foglalkozó gének egy része. Amennyiben egy ilyen géncsalád tagjai pontosan egy példányban szerepelnek minden élőlény genomjában, és feltételezzük, hogy nem történt egyetlen génduplikáció, horizontális géntranszfer vagy génvesztés-esemény sem, akkor a géncsalád története, pontosabban az azt leíró génfa megegyezik az őket hordozó fajok történelmével, vagyis az azt leíró fajfával. 1977-ben Carl Woese és társai (Fox et al., 1977) is egy ilyen „egypéldányos” géncsalád, a riboszómák alkotórészét adó kis RNS-alegység, az ún. 16S rRNS gének szekvenciáit vizsgálva rekonstruálták az élővilág fajfáját, és fedezték fel, hogy az addig prokariótaként ismert csoport valójában két doménre oszlik: a baktériumokra és archeákra.

Célravezetőnek tűnhet tehát, hogy a kis számú esszenciális géncsaládot, a gének „1%-át felhasználva kövessük vissza az élet történetét. Azonban ha így járunk el, a géncsaládok maradék 99%-a által hordozott információt nem hasznosítjuk. Ez azt eredményezi, hogy az egyedi géncsaládok történetének elmosódottsága miatt sok esetben nem rajzolható fel egyértelműen a fajfa sem.

Erre a problémára keresték a megoldást Bastien Boussau és munkatársai, akik 2013-ban harminchat emlősfajban 6966 géncsalád

történetét rekonstruálták a fajfával közösen. A lehetséges fajfák terét egy számítógépes algoritmus segítségével járták be: egy adott fajfára elkészítették az összes géncsaládhoz tartozó, az adott fajfához és a szekvenciákhoz együttesen legjobban illő, mindkét folyamat szerint legvalószínűbb génfát. Az így kapott „közös” valószínűségek szorzata adta a fajfa valószínűségét. Ezt követően minden szomszédos fajfára is elvégezték ugyanezt a számítást, és ha ezek között találtak olyan szomszédot, melynek valószínűsége nagyobb volt, mint a kezdeti fajfa, akkor az eljárást ettől a fajfától elindulva megismételték. Az algoritmus akkor ért véget, amikor olyan fajfát találtak, amelynek nem volt szomszédja, amely nála valószínűbb lett volna. Ez az eljárás egyszerre használja a fajfa adta keretrendszert arra, hogy pontosabb génfákat kapjunk, miközben a fajfát is nagy pontossággal, több ezer géncsalád alapján képes meghatározni — ez az ún. *közös rekonstrukció*. Az eredmények biztatóak: a kapott fajfa nem mond ellent semmilyen ismert rendszertani ténynek, és a géntörténetek is valószínűek, ugyanis nem mutatnak a mainál jóval nagyobb ősi genomokat (nem megfelelő rekonstrukciós módszerek gyakran vezetnek „felduzzasztott” ősi genomokhoz).

#### Géntranszfer: zaj helyett információ

Az emlősök történetének kutatásában elegendő csupán a duplikációkat és a génvesztéseket figyelembe vennünk, más élőlényeknél viszont nem hanyagolható el a gének történetét bonyolító harmadik folyamat sem: a horizontális géntranszfer. A transzfereseeményekre sokan a fajfa vonalait elmosó „zajként” tekintenek, hiszen ilyenkor a génfák vonalai átlépik a fajok szabta határokat, és a fajfa csöveinek összekötésével egy bonyolult háló-

zatot hoznak létre. Ha azonban mégis arra vállalkozunk, hogy megvizsgáljuk a transzferen keresztülment gencsaládok történetét is, értékes információkhoz juthatunk.

Először is figyelembe kell vennünk, hogy minden transzfer két, térben és időben egymás mellett élő faj között történhet meg – fordított szemzőből nézve, ha ismerünk egy transzferet is tartalmazó géntörténetet, megtudhatjuk, a fajfa mely ágai fedtek át egymással időben (3. ábra). Több transzferet felhasználva akár a fajfa összes elágazásának idejét is meghatározhatjuk (Szöllösi et al., 2012).

Ugyanígy jelentőséggel bírhat a tény, hogy az ősi horizontális transzferek során mára már kihalt fajok génjei is bekerülhettek a mostani élőlények őseinek genomjába. Ha ezek a gének beépültek az új genomba, és ott egészen napjainkig fennmaradtak, akkor az élővilág számunkra eddig ismeretlen részeinek lenyomatát adhatják (Szöllösi et al., 2013).

#### *A múlt életre kel a laboratóriumban*

A filogenetikai rekonstrukció eredménye nemcsak egy génfa, hanem, mint az az 1. ábrán is látható, a szekvenciaváltozásokat a génfa mentén visszafelé követve az ősi szekvenciákat is megkapjuk. Ugyanazon szekvenciaillesztés esetén különböző alternatív génfák különböző ősi szekvenciákhoz vezetnek. Ezek az ősi szekvenciák az evolúciós múltról tett alternatív predikciók, amelyeket a laboratóriumban egymással összehasonlíthatunk. Rendelkezésünkre állnak ugyanis molekuláris biológiai módszerek, melyek segítségével az

ősi szekvenciákat mesterségesen szintetizálva, majd baktériumokba bejuttatva az általuk kódolt ősi fehérjéket is legyárthatjuk.

Mathieu Groussin és társai (Groussin et al., 2015) ilyen módon „feltámasztott,” több száz millió éves LeuB-enzimek biokémiai tulajdonságait hasonlították össze. Arra az eredményre jutottak, hogy mind a fajfa figyelembe vételével, mind pedig az anélkül rekonstruált génfáknak megfelelő enzim nagy vonalakban hasonló enzimeket eredményezett: mindkettő a mai LeuB-enzimeknél magasabb hőmérsékleten működött optimálisan, megerősítve azt, hogy az ősi baktérium, amelyből származtak, magas hőmérsékleten élő, ún. termofil organizmus volt. A részletesebb enzimaktivitásra irányuló vizsgálatok azonban egyértelműen kimutatták, hogy csak a fajfa felhasználásával rekonstruált enzim mutatott a jelenbeli enzimekkel összemérhető, valószínű biokémiai paramétereket.

A LeuB-enzimekkel végzett kísérletek közvetlen bizonyítékát adják annak, hogy a fajfa és génfák közös kezelésével pontosabb rekonstrukciókat kapunk. Megmutatják továbbá, hogy nemcsak a fajok és gének lezármazásának történetéről lehet élesebb fogalmunk, hanem az ősi enzimekről és rajtuk keresztül az ősi élőlények tulajdonságairól is részletes információkhoz juthatunk.

**Kulcsszavak:** DNS-szekvenálás, génduplikáció, horizontális géntranszfer, szekvenciaillesztés, génfák, fajfák

#### IRODALOM

Boussau, Bastien – Szöllösi Gergely J. – Duret, Laurent et al. (2013): Genome-scale Coestimation of Species and Gene Trees. *Genome Research*. 23, 2, 323–330. DOI: 10.1101/gr.141978.112 • <http://genome.cshlp.org/content/23/2/323.full>

Boussau, Bastien – Daubin, Vincent (2010): Genomes as Documents of Evolutionary History. *Trends in Ecology and Evolution*. 25, 4, 224–232. DOI: 10.1016/j.tree.2009.09.007

Felsenstein, Joseph (2004): *Inferring Phylogenies*. Sinauer Associates, Sunderland, USA

Fox, George E. – Magrum, Linda J. – Balch, William E. et al. (1977): Classification of Methanogenic Bacteria by 16S Ribosomal RNA Characterization. *Proceedings of the National Academy of Sciences of the USA*. 74, 10, 4537–4541. DOI: 10.1073/pnas.74.10.4537 • <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC431980/>

Groussin, Mathieu – Hobbs, Joanne K. – Szöllösi Gergely J. et al. (2015): Toward More Accurate Ancestral Protein Genotype–Phenotype Reconstructions with the Use of Species Tree-aware Gene Trees. *Molecular Biology and Evolution*. 32, 1, 13–22. DOI: 10.1093/molbev/msu305 • <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4271536/>

Maddison, Wayne P. (1997): Gene Trees in Species Trees. *Systematic Biology*. 46, 3, 523–536. DOI: 10.1093/sysbio/46.3.523 • <http://sysbio.oxfordjournals.org/content/46/3/523.full.pdf+html>

Szöllösi Gergely J. – Tannier, Eric – Lartillot, Nicolas – Daubin, Vincent (2013): Lateral Gene Transfer from the Dead. *Systematic Biology*. 62, 3, 386–397. DOI:

10.1093/sysbio/syto03 • <http://sysbio.oxfordjournals.org/content/early/2013/01/25/sysbio.syto03.full.pdf+html>

Szöllösi Gergely J. – Boussau, Bastien – Abby, Sophie S. et al. (2012): Phylogenetic Modeling of Lateral Gene Transfer Reconstructs the Pattern and Relative Timing of Speciations. *Proceedings of the National Academy of Sciences of the USA*. 109, 43, 17513–17518. DOI: 10.1073/pnas.1202997109 • <http://www.pnas.org/content/109/43/17513.full>

Zuckermandl, Emile – Pauling, Linus (1965): Evolutionary Divergence and Convergence in Proteins. *Evolving Genes and Proteins*. 97, 97–165. • <http://authors.library.caltech.edu/5456/1/hrst.mit.edu/hrs/evolution/public/papers/zuckermandlpauling1965/zuckermandlpauling1965.pdf>

Zuckermandl, Emile – Pauling, Linus (1965): Molecules as Documents of Evolutionary History. *Journal of Theoretical Biology*. 8, 2, 357–366. DOI: 10.1016/0022-5193(65)90083-4 • [http://lectures.molgen.mpg.de/phylogeny\\_ws05/papers/zuckermandl\\_pauling.pdf](http://lectures.molgen.mpg.de/phylogeny_ws05/papers/zuckermandl_pauling.pdf)

