Publications

	Total	as lead author
Publications with impact factor:	52	25
of these Publications with IF>10	24	13

Cumulative IF: 533.135 (mean IF: 10.25) of this **273.79 as lead author** (mean IF: 10.952)
H-index: 30 (google scholar)
Citations: 3935 (google scholar)

Leading interdisciplinary publications

Journal	Total	as lead author	impact factor
Nature	1	0	69.5
Science	1	1	63.7
PNAS	3	2	12.8
Nature Communications	3	2	17.7

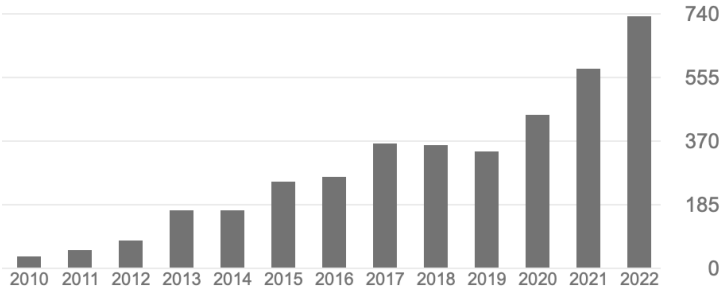
Leading discipline specific publications

Systematic Biology	6	5	15.7
Molecular Biology and Evolution	8	4	16.2
Nature Ecology and Evolution	4	1	15.5
Genetics	3	2	4.4
Phys. Rev. Lett.	2	1	9.2

Citations

	All	Since 2017
Citations	3935	2826
h-index	30	24
i10-index	46	39

Citations per year



ORCID: 0000-0002-8556-845X
https://orcid.org/0000-0002-8556-845X

Google Scholar: sPrYT-oAAAAJ
https://scholar.google.hu/citations?hl=en&user=sPrYT-oAAAAJ&view_op=list_works&sortby=pubdate

Publications in peer-reviewed journals

supervision: ✍ undergraduate co-author; ☕ PhD student co-author; ⚛ postdoc co-author;
top five publications: ☆ — see also summaries starting **page 9**

2022

as lead author

☆**#2** Demeter MC ☕, Derényi I, **Szöllősi GJ**
Trade-off between reducing mutational load and increasing commitment to differentiation determines tissue organization
Nature Communications 13 (1) 1666 https://doi.org/10.1038/s41467-022-29004-1

Tibély G ⚛, Schrempf D ⚛, Derényi I, **Szöllősi GJ**
Simultaneous estimation of per cell division mutation rate and turnover rate from bulk tumour sequence data
PLoS Computational Biology 18 (4): e1010048 https://doi.org/10.1371/journal.pcbi.1010048

Szánthó L ☕, Lartilott N, **Szöllősi GJ*****, Schrempf D* ⚛
**co-corresponding and equally contributing authors*
Compositionally constrained sites drive long branch attraction
Systematic Biology *to appear* https://doi.org/10.1101/2022.03.03.482715

Szöllősi GJ, Höhna S, Williams TA, Schrempf D ⚛, Daubin V, Boussau B
Relative time constraints improve molecular dating
Systematic Biology *syab084* https://doi.org/10.1093/sysbio/syab084

as contributing author

Ocaña-Pallarès E ⚛, Williams T, López-Escardó D, Arroyo A, Pathmanathan J, Bapteste E, Tikhonenkov D, Keeling P, **Szöllősi GJ**, Ruiz-Trillo I
Divergent genomic trajectories predate the origin of animals and fungi
Nature 609, 747 https://doi.org/10.1038/s41586-022-05110-4

Harris B, Clarck J, Schrempf D, **Szöllősi GJ**, Donoghue P, Hetherington A, Williams TA
Divergent evolutionary trajectories of bryophytes and tracheophytes from a complex common ancestor of land plants
Nature Ecology and Evolution published online https://doi.org/10.1038/s41559-022-01885-x

Foster PG, Schrempf D ⚛, **Szöllősi GJ**, Williams TA, Cox CJ, Embley MT
Recoding amino acids to a reduced alphabet may increase or decrease phylogenetic accuracy
Systematic Biology syac042 https://doi.org/10.1093/sysbio/syac042

Moody ERR, Mahendrarajah TA, Dombrowski N, Clark JW, Petitjean C, Offre P, **Szöllősi GJ**, Spang A, Williams TA
Universal markers support a long inter-domain branch between Archaea and Bacteria
eLIFE 11:e66695 https://doi.org/10.7554/eLife.66695

Morel B, Schade S, Lutteropp S, **Szöllősi GJ**, Stamatakis A
SpeciesRax: A tool for maximum likelihood species tree inference from gene family trees under duplication, transfer, and loss
Molecular Biology and Evolution 39(2) msab365 https://doi.org/10.1093/molbev/msab365

Borges R, Boussau B, **Szöllősi GJ,** Kosiol C
Nucleotide usage biases distort inferences of the species tree
Genome Biology and Evolution 14(1) *evab290* https://doi.org/10.1093/gbe/evab290

2021

as lead author

☆**#1** Coleman G^, Davín A^, Mahendrarajah T, Szánthó L ✍, Spang A, Hugenholtz P*, **Szöllősi GJ***, Williams TA* ^*equal contribution *co-corresponding and equally contributing authors*
A rooted phylogeny resolves early bacterial evolution
Science 372 (6542) 574. https://doi.org/10.1126/science.abe0511

as contributing author

Blanquart S, Groussin M, Le Roy A, **Szöllősi GJ**, Girard E, Franzetti B, Gouy M, Madern D
Resurrection of Ancestral Malate Dehydrogenases Reveals the Evolutionary History of Halobacterial Proteins: Deciphering Gene Trajectories and Changes in Biochemical Properties.
Molecular Biology and Evolution 38 (9) 3754–3774. https://doi.org/10.1093/molbev/msab146

Williams TA, Schrempf D ⚛, **Szöllősi GJ,** Cox CJ, Foster PG, Embley TM
Inferring the deep past from molecular data
Genome Biology and Evolution 13 (5) evab067 https://doi.org/10.1093/gbe/evab067

2020

as lead author

Grajzel D ☕, Derényi I, **Szöllősi GJ**
A compartment size dependent selective threshold limits mutation accumulation in hierarchical tissues
Proceedings of the National Academy of Sciences 117 (3), 1606-1611
https://doi.org/10.1073/pnas.1913104117

D Schrempf D ⚛, Lartillot N, **Szöllősi GJ**
Scalable empirical mixture models that account for across-site compositional heterogeneity
Molecular Biology and Evolution 37 (12), 3616–3631. https://doi.org/10.1093/molbev/msaa145

Morel B, Kozlov AM, Stamatakis A, **Szöllősi GJ**
GeneRax: A Tool for Species-Tree-Aware Maximum Likelihood-Based Gene Family Tree Inference under Gene Duplication, Transfer, and Loss
Molecular Biology and Evolution 37 (9), 2763-2774. https://doi.org/10.1093/molbev/msaa141

Davín AA ⚛, Tricou T, Tannier E, de Vienne DM, **Szöllősi GJ**
Zombi: A phylogenetic simulator of trees, genomes and sequences that accounts for dead lineages
Bioinformatics 36 (4), 1286–1288. https://doi.org/10.1093/bioinformatics/btz710

as contributing author

Willimas TA, Cox C, Foster P, **Szöllősi GJ**, Embley T
Phylogenomics provides robust support for a two-domains tree of life
Nature Ecology and Evolution 4 (1), 138-147. https://doi.org/10.1038/s41559-019-1040-x

Miyauchi S, .. , Gertheis ZN ✍, ..,**Szöllősi GJ**, .., Martin FM
Large-scale genome sequencing of mycorrhizal fungi provides insights into the early evolution of symbiotic traits
Nature Communications 11 (1), 1-17. https://doi.org/10.1038/s41467-020-18795-w

2019

as contributing author

Varga T, .., **Szöllősi GJ**, .., Nagy LG
Megaphylogeny resolves global patterns of mushroom evolution
Nature Ecology an Evolution, 3, 668. https://doi.org/10.1038/s41559-019-0834-1

Borges R, **Szöllősi GJ**, Kosiol C
Quantifying GC-biased gene conversion in great ape genomes using polymorphism-aware models
Genetics 212, 1321-1336. https://doi.org/10.1534/genetics.119.302074

2018

as lead author

☆**#3** Davín AA, Tannier E, Williams TA, Boussau B, Daubin B, **Szöllősi GJ**
Gene transfers can date the tree of life
Nature Ecology and Evolution 2 (5), 904. https://doi.org/10.1038/s41559-018-0525-3

Nagy LG, **Szöllősi GJ**
Fungal phylogeny in the age of genomics: insights into phylogenetic inference from genome-scale datasets
Advances in Genetics 100, 49-72. https://doi.org/10.1016/bs.adgen.2017.09.008

as contributing author

Kellner S, Spang A, Offre P, **Szöllősi GJ**, Petitjean C, Williams TA
Genome size evolution in the Archaea
Emerging Topics in Life Sciences 2 (4), 595-605. https://doi.org/10.1042/ETLS20180021

Chauve V, Rafiey A, Davin AA, Scornavacca C, Veber P, Boussau B, **Szöllősi GJ**, Daubin V, Tannier E
MaxTiC: Fast ranking of a phylogenetic tree by Maximum Time Consistency with lateral gene Transfers
bioRxiv, 127548 (2017) Reviewed and Recommended by ***Peer Community In Evolutionary Biology***
https://doi.org/10.24072/pci.evolbiol.100037

Duchemin W, .., **Szöllosi GJ,** Zhang L, Tannier E, Daubin V.
RecPhyloXML: a format for reconciled gene trees
Bioinformatics 34 (21), 3646-3652. https://doi.org/10.1093/bioinformatics/bty389

2017

as lead author

Derényi I, **Szöllősi GJ**
Hierarchical tissue organization as a general mechanism to limit the accumulation of somatic mutations
Nature Communications 8, 14545. https://doi.org/10.1038/ncomms14545

as contributing author

Williams TA, **Szöllősi GJ**, Spang A, Foster PG, Heaps SE, Boussau B, Ettema TJG, Embley TM
Integrative modeling of gene and genome evolution roots the archaeal tree of life
Proceedings of the National Academy of Sciences 114 (23), E4602-E4611
https://doi.org/10.1073/pnas.1618463114

Bailly-Bechet M, Martins-Simões P, **Szöllősi GJ**, Mialdea G, Sagot MF, Charlat S
How Long Does Wolbachia Remain on Board?
Molecular Biology and Evolution 34 (5), 1183-1193. https://doi.org/10.1093/molbev/msx073

2016

as contributing author

Groussin M, Boussau B, **Szöllősi GJ**, Eme L, Gouy M, Brochier-Armanet C, Daubin V
Gene acquisitions from bacteria at the origins of major archaeal clades are vastly overestimated.
Molecular Biology and Evolution 33, 305-310. doi: 10.1093/molbev/msv249

Jacox E, Chauve C, **Szöllősi GJ**, Ponty J, Scornavacca C
ecceTERA: Comprehensive gene tree-species tree reconciliation using parsimony
Bioinformatics 32 2056-2058. https://doi.org/10.1093/bioinformatics/btw105

2015

as lead author

Szöllősi GJ, Davin AA,Tannier E, Daubin V, Boussau B
Genome-scale phylogenetic analysis finds extensive gene transfer among Fungi.
Philosophical Transactions of the Royal Society B 370 20140335
https://doi.org/10.1098/rstb.2014.0335

Scornavacca C, Jacox E, **Szöllősi GJ**
Joint Amalgamation of Most Parsimonious Reconciled Gene Trees
Bioinformatics, 31**,** *841-8*. https://doi.org/10.1093/bioinformatics/btu728

Derényi I, **Szöllősi GJ**
The effective temperature of mutations
Physical Review Letters 114, 058101. https://doi.org/10.1103/PhysRevLett.114.058101

Szöllősi GJ, Tannier Eric, Daubin V, Boussau B
The inference of gene trees with species trees
Systematic Biology 64, 42-62. https://doi.org/10.1093/sysbio/syu048

as contributing author

Groussin M, Hobbs JK, **Szöllősi GJ**, Gribaldo S, Arcus VL, Gouy M
Towards accurate ancestral protein genotype–phenotype reconstructions with species tree-aware gene trees.
Molecular Biology and Evolution 32, 13-22. https://doi.org/10.1093/molbev/msu305

2013

as lead author

☆**#4 Szöllősi GJ**, Tannier E, Lartillot N, Daubin V
Lateral Gene Transfer from the Dead.
Systematic Biology 62 ,386-397. https://doi.org/10.1093/sysbio/syt003

(follow up to above #4☆) **Szöllősi GJ,** Rosikiewicz W, Boussau B, Tannier E, Daubin V
Efficient Exploration of the Space of Reconciled Gene Trees
Systematic Biology 62**,** 901-912. https://doi.org/10.1093/sysbio/syt054

as contributing author

Patterson M, **Szöllősi GJ**, Daubin V, Tannier E
Lateral gene transfer, rearrangement, reconciliation
BMC Bioinformatics 14, *S47.* https://doi.org/10.1186/1471-2105-14-S15-S4

Boussau B, **Szöllősi GJ**, Duret L, Gouy M, Tannier E, Daubin V
Genome-scale coestimation of species and gene trees
Genome Research 23, *323-330.* https://doi.org/10.1101/gr.141978.112

2012

as lead author

#5☆ **Szöllősi GJ**, Boussau B, Abby SS, Tannier E, Daubin V
Phylogenetic modeling of lateral gene transfer reconstructs the pattern and relative timing of speciations
Proceedings of the National Academy of Sciences 109, 17513-17518.
https://doi.org/10.1073/pnas.1202997109

as contributing author

Berard S, Gallien C, Boussau B, **Szöllősi GJ**, Daubin V, Tannier E
Evolution of gene neighbourhoods within reconciled phylogenies
Bioinformatics, 28 I382-I388*.* https://doi.org/10.1093/bioinformatics/bts374

2011

as lead author

Schiffels S*, **Szöllősi GJ***, Mustonen V, Laessig M ***equal contrib.**
Emergent Neutrality in Adaptive Asexual Evolution
Genetics 189, 1361-1375. https://doi.org/10.1534/genetics.111.132027

as contributing author

Rauscher AA, Simon Z, **Szöllősi GJ,** Graf L, Derényi I, Malnasi-Csizmadia A
Temperature dependence of internal friction in enzyme reactions
FASEB Journal 25, 2804-2813. https://doi.org/10.1096/fj.11-180794

Czovek A, **Szöllősi GJ**, Derényi I
Neck-Linker Docking Coordinates the Kinetics of Kinesin's Heads
Biophysical Journal 100, 1729-1736. https://doi.org/10.1016/j.bpj.2011.01.039

2010

as contributing author

Katifori E, **Szöllősi GJ**, Magnasco MO
Damage and Fluctuations Induce Loops in Optimal Transport Networks
Physical Review Letters 104, 048704. https://doi.org/10.1103/PhysRevLett.104.048704

2009

as lead author

Szöllősi GJ, Derényi I
Congruent Evolution of Genetic and Environmental Robustness in Micro-RNA
Molecular Biology and Evolution 26, 867-874. https://doi.org/10.1093/molbev/msp008

2008

as lead author

Szöllősi GJ, Derényi I
Evolutionary games on minimally structured populations
Physical Review E **78:**(3) *031919.* https://doi.org/10.1103/PhysRevE.78.031919

Szöllősi GJ, Derényi I
The effect of recombination on the neutral evolution of genetic robustness
Mathematical Biosciences 214, 58-62. https://doi.org/10.1016/j.mbs.2008.03.010

as contributing author

Czovek A, **Szöllősi GJ**, Derényi I
The relevance of neck linker docking in the motility of kinesin
Biosystems 93, 29-33. https://doi.org/10.1016/j.biosystems.2008.04.006

2006

as lead author

Szöllősi GJ, Derényi I, Vellai T
The maintenance of sex in bacteria is ensured by its potential to reload genes
Genetics **174:**(4) *pp. 2173-2180.* https://doi.org/10.1534/genetics.106.063412

as contributing author

Szabo B, **Szöllősi GJ**, Gonci B, Juranyi Z, Selmeczi D, Vicsek T
Phase Transition in The Collective Migration of Tissue Cells: Experiment And Model
Physical Review E 74 061908. https://doi.org/10.1103/PhysRevE.74.061908

2004

as lead author

Szöllősi GJ, Derényi I, Vörös J
Reversible mesoscopic model of protein adsorption: From equilibrium to dynamics
Physica A 343, *359-375.* https://doi.org/10.1016/j.physa.2004.06.062

Further publications:

Conference Proceedings

Doyon JP, Scornavacca C, Gorbunov KY, **Szöllősi GJ**, Ranwez V, Berry V
An Efficient Algorithm for Gene/Species Trees Parsimonious Reconciliation with Losses, Duplications and Transfers
In: Tannier Eric (ed.) Lecture Notes in Computer Science. 6398 pp 93-108 (2010)
https://doi.org/10.1007/978-3-642-16181-0_9

Book chapters

Davín AA*,* Schrempf D ⚛*,* Williams TA, Hugenholtz P, **Szöllősi GJ**
Dating microbial evolution with horizontal gene transfers
Methods in Molecular Biology Book Chapter (2022)
https://doi.org/10.1007/978-1-0716-2691-7

Harris B**,** Gubry-Rangin C, **Szöllősi GJ**, Williams TA
Rooting species trees using gene tree-species tree reconciliation
Methods in Molecular Biology Book Chapter (2022)
https://doi.org/10.1007/978-1-0716-2691-7

Schrempf D ⚛, **Szöllősi GJ**
The sources of phylogenetic conflicts
Chapter 3.1 in Phylogenetics in the genomic era. Open Access Book
Editors: Celine Scornavacca, Frédéric Delsuc and Nicolas Galtier;
https://hal.archives-ouvertes.fr/hal-02535070/document

Daubin V, **Szöllősi GJ**
Horizontal gene transfer and the history of life. In: Howard Ochman (ed.)
Microbial Evolution. New York: Cold Spring Harbour Press, Woodbury, New York,
https://doi.org/10.1101/cshperspect.a018036

Szöllősi GJ, Daubin V
Modeling gene family evolution and reconciling phylogenetic discord. In: Maria Anisimova (ed.)
Evolutionary Genomics: Statistical and Computational Methods, Volume 2. New York: Humana Press *pp. 29-51.*
https://doi.org/10.1007/978-1-61779-585-5_2

Publications in Hungarian:

Gertheis ZN ✍, **GJ Szöllősi**
Fajok és gének történetének nyomában
Magyar Tudomány 178(2), 140-143. (2017)
http://www.epa.hu/00600/00691/00161/pdf/EPA00691_mtud_2017_02_141-147.pdf

Derényi I, **Szöllősi GJ**
A hierarchikus sejtdifferenciálódás szerepe a mutációk felhalmozódásának és a rák kockázatának minimalizálásában
Magyar Tudomány 177(1), 23-25. (2016)
http://epa.oszk.hu/00600/00691/00148/pdf/EPA00691_mtud_2016_01_080-083.pdf
Hungarian language journal

Open-source software with my involvement

ALE — https://github.com/ssolo/ALE

GeneRax — *Resolving phylogenetic conflicts* — https://github.com/BenoitMorel/GeneRax

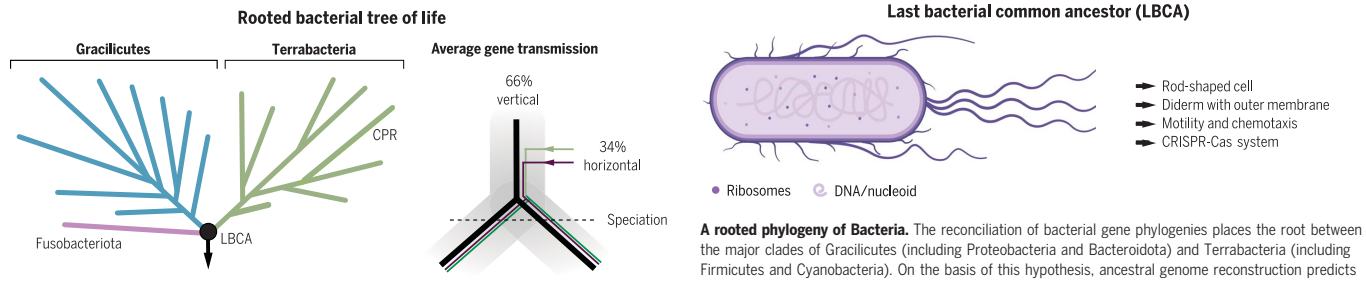
Zombi — https://github.com/AADavin/Zombi

recPhyloXML — https://github.com/WandrilleD/recPhyloXML

EDCLUSTER — https://github.com/dschrempf/EDCluster

Top five publications

#1 *A rooted phylogeny resolves early bacterial evolution*
Coleman G^, Davín A^, Mahendrarajah T, Szánthó L, Spang A , Hugenholtz P*, **Szöllősi GJ***, Williams TA*
Science 372 (6542) 574 (2021) ^eq. contrib. *co-corr. & eq. contrib.
https://doi.org/10.1126/science.abe0511
author's copy of pdf: http://ssolo.web.elte.hu/Coleman_2021.pdf



A rooted phylogeny of Bacteria. The reconciliation of bacterial gene phylogenies places the root between the major clades of Gracilicutes (including Proteobacteria and Bacteroidota) and Terrabacteria (including Firmicutes and Cyanobacteria). On the basis of this hypothesis, ancestral genome reconstruction predicts that the last bacterial common ancestor (LBCA) was a complex, double-membraned cell and that, on average, two-thirds of gene transmissions have been vertically inherited along this rooted tree.

In a collaboration between the groups of Tom Williams, Philip Hugenholtz and my own group we reconstructed and rooted the bacterial tree by applying a hierarchical probabilistic model of genome evolution along the specie tree that I have been developing for the last ten years (cf. publication #3). This approach, which explicitly uses information from gene duplications and losses within a genome as well as gene transfers between genomes, allowed us to root the tree without including an archaeal outgroup. This is important, both because the position of the universal root is uncertain and because the long branch separating Bacteria from Archaea has the potential to distort the reconstruction of within-Bacteria relationships. It also allowed us to quantitatively model both the vertical and horizontal components of bacterial evolution and integrate information from 11,272 gene families to resolve the root of the bacterial tree.

Gene families inferred to have been present at the root indicate that the last bacterial common ancestor was already a complex double- membraned cell capable of motility and chemotaxis that possessed a CRISPR-Cas system. We also found that, although ~92% of gene families have experienced horizontal transfers during their history, tracing their evolution along the most likely rooted tree revealed that about two-thirds of gene transmissions have been vertical. Thus, bacterial evolution has a major vertical component, despite a profound impact of horizontal gene transfer through time.

This paper was also covered in a Perspective in Science: Illuminating the first bacteria.

Related publications:

A series of papers demonstrating the ability of the probabilistic model of genome evolution we developed (cf. publication #3) to access *bona fide* phylogenetic signal was instrumental both in achieving methodological maturity and credibility in the field:

Gene transfers can date the tree of life (#3 below)
Davin A, Tannier E, Williams T, Boussau B, Daubin V, **Szöllősi GJ**
Nature Ecology and Evolution 2(5), 904-909. (2018)
https://doi.org/10.1038/s41559-018-0525-3
This paper which demonstrates that timing information can be extracted from transfers was also covered in Quanta magazine: Chronological Clues to Life's Early History Lurk in Gene Transfers

Integrative modeling of gene and genome evolution roots the archaeal tree of life
Williams TA, **Szöllősi GJ**, Spang A, Foster PG, Heaps SE, Boussau B, Ettema TJG, Embley TM
Proceedings of the National Academy of Sciences 114 (23), E4602-E4611 (2017)
https://doi.org/10.1073/pnas.1618463114
In this work we demonstrated the utility of reconciliation methods for out-group free rooting and the reconstruction of ancestral phenotypes at deep evolutionary time scales.

Phylogenetic modeling of lateral gene transfer reconstructs the pattern and relative timing of speciations (#5 below)
Szöllősi GJ, Boussau B, Abby SS, Tannier E, Daubin V
Proceedings of the National Academy of Sciences 109, 17513-17518. (2012)
https://doi.org/10.1073/pnas.1202997109
A proof of concept study that illustrated the potential of reconciliation methods for reconstructing genome scale phylogenies.

Top five publications

#2 *Trade-off between reducing mutational accumulation and increasing commitment to differentiation determines tissue organization*
Demeter M, Derényi I, **Szöllősi GJ**
Nature Communications 13, 1666 (2022)
https://doi.org/10.1038/s41467-022-29004-1 (open access)

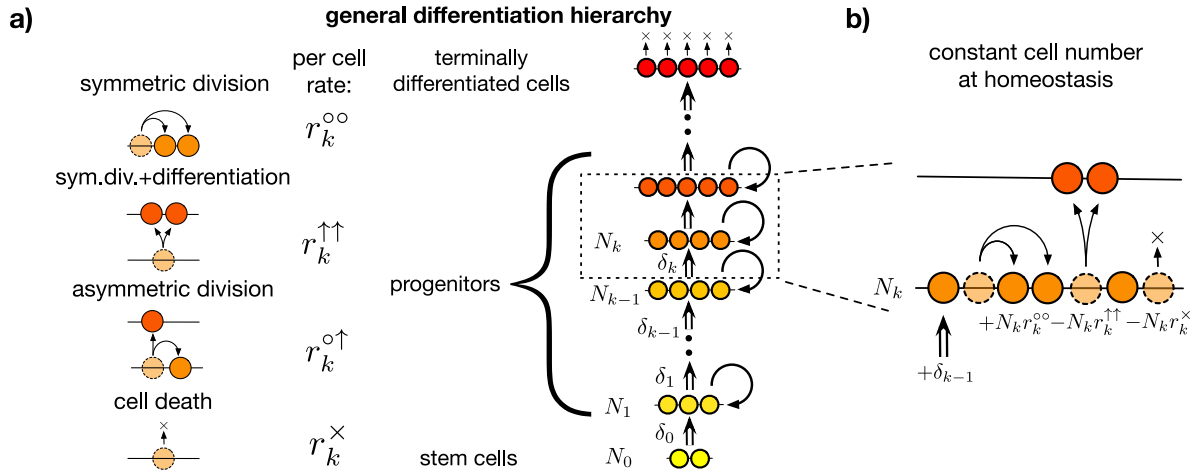


Fig. 1 Minimal generic model of hierarchically differentiating tissues and corresponding cell lineage trees. a cells are organized into $n+1$ hierarchical levels based on their differentiation state. The bottom level (level 0) corresponds to tissue-specific stem cells, higher levels represent progressively differentiated progenitor cells, and the top level (level n) is comprised of terminally differentiated cells. Four microscopic events can occur with a cell: (i) asymmetric cell division, (ii) symmetric cell division with differentiation, (iv) symmetric cell division without differentiation and (iv) cell death. The number of cells on level k under normal homeostatic conditions is denoted by N_k. Under homeostatic conditions each level (except for the terminally differentiated one at the top) provides the next level with newly differentiated cells at a rate δ_k. Terminally differentiated cells at the top of the hierarchy cannot divide and are destined to wear away (i.e., leave the tissue). **b** At progenitor levels $k>0$ in fully developed tissues under homeostatic conditions self-renewal replenishes only a fraction of the cells lost as cells arrive by differentiation from lower levels and, as a result, progenitor cells always have an inherent proliferative disadvantage. **c** Cell lineage trees in "wild type" tissues are shown for two different values of a uniform amplification factor $\gamma_k=\gamma$. Given the

Renewed tissues of multicellular organism accumulate mutations that lead to ageing and cancer. To mitigate these effects, self-renewing tissues produce cells along differentiation hierarchies, which have been shown to suppress somatic evolution both by limiting the number of cell divisions, and thus reducing mutational load, and by differentiation "washing out" mutations. In this paper we use the mathematical model of homeostatic tissue with multiple levels of differentiation developed in (Derényi and Szöllősi 2017, see below) to explore the organisation of hierarchical tissues evolved to limit the incidence of cancer. Developing an accurate approximation of the probability of accumulating a given number of critical mutations with selective effects that can change the proliferation dynamics of individual cells we show that under general conditions there exists a trade-off between minimising mutation accumulation and maximising the proliferative disadvantage of cells. This trade-off provides an explanation for the observed higher division rate of stem cells than what would be expected solely from the minimisation of the accumulation of mutations. Our results explain differences in the organisation of widely different hierarchical tissues, such as colon and blood.

Related publications:

A compartment size dependent selective threshold limits mutation accumulation in hierarchical tissues
Grajzel D, Derényi I, **Szöllősi GJ**
Proceedings of the National Academy of Sciences 117 (3), 1606-1611 (2020)
https://doi.org/10.1073/pnas.1913104117 (open access)

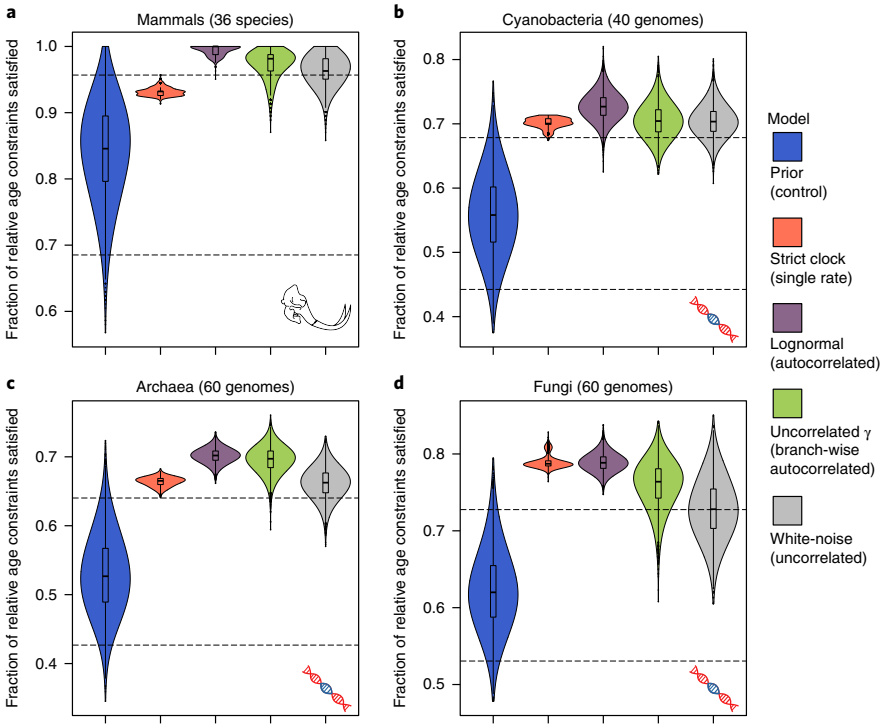
In this paper we analytically demonstrate the existence of a third mechanism: a compartment size-dependent threshold in proliferative advantage, below which mutations cannot persist, but are rapidly expelled from the tissue by differentiation.

Hierarchical tissue organization as a general mechanism to limit the accumulation of somatic mutations
Derényi I, **Szöllősi GJ**
Nature Communications 8, 14545 (2017)
https://doi.org/10.1038/ncomms14545 (open access)

To limit the accumulation of somatic mutations, renewing tissues must minimise the number of times each cell divides during differentiation. Here, we analytically derive the lower limit of lifetime divisional load of a tissue, show that hierarchically differentiating tissues can approach this limit, and that this depends on uneven divisional rates across the hierarchy.

Top five publications

#3 *Gene transfers can date the tree of life.*
Davin A, Tannier E, Williams T, Boussau B, Daubin V, **Szöllősi GJ**
Nature Ecology and Evolution 2(5), 904-909. (2018)
https://doi.org/10.1038/s41559-018-0525-3



Agreement between transfer-based relative ages and molecular clocks. a, Relative ages derived from 12 fossil calibrations from a phylogeny of 36 extant mammals were compared with node ages sampled from four different relaxed molecular clock models implemented in Phylobayes and with node ages derived from random chronograms, keeping the species phylogeny fixed. **b**–**d**, Relative ages derived from gene transfers in Cyanobacteria (**b**), Archaea (**c**) and Fungi (**d**) using the MaxTiC algorithm were compared with estimates from the same five models as in **a**. For each model and each sampled chronogram, we calculated the fraction of relative age constraints that are satisfied. Each violin plot shows the distribution of the fraction of relative age constraints satisfied by 5,000 sampled chronograms. The blue distribution corresponds to random chronograms drawn from the prior with the 95% confidence interval denoted by broken lines.

Biodiversity has always been predominantly microbial, and the scarcity of fossils from bacteria, archaea and microbial eukaryotes has prevented a comprehensive dating of the tree of life. Here, we show that patterns of lateral gene transfer deduced from an analysis of modern genomes encode a novel and abundant source of information about the temporal coexistence of lineages throughout the history of life.

This paper was also covered in Quanta magazine: Chronological Clues to Life's Early History Lurk in Gene Transfers

#4 *Lateral Gene Transfer from the Dead.*
Szöllősi GJ, Tannier E, Lartillot N, Daubin V
Systematic Biology 62, 386-397. (2013) https://doi.org/10.1093/sysbio/syt003 (open access)
and *Efficient Exploration of the Space of Reconciled Gene Trees*
Szöllősi GJ, Rosikiewicz W, Boussau B, Tannier E, Daubin V
Systematic Biology 62, 901-912. (2013) https://doi.org/10.1093/sysbio/syt054 (open access)

In this series of two publications published in the same year in Systematic Biology we demonstrate that if the number of sampled species is small compared with the total number of existing species, the overwhelming majority of gene transfers involve speciation to and evolution along extinct or unsampled lineages. We show that the evolution of genes along extinct or unsampled lineages can to good approximation be treated as those of independently evolving lineages described by a few global parameters.
Using this result, in the second publication we derive an algorithm to efficiently calculate the likelihood of an alignment under a joint model of sequence evolution and gene tree–species tree reconciliation. We demonstrate that gene trees reconstructed using the joint likelihood are substantially more accurate than those reconstructed using sequence alone and exhibit a striking reduction in apparent phylogenetic discord, with respectively 24%, 59%, and 46% reductions in the mean numbers of duplications, transfers, and losses per gene family.
The open source implementation of ALE is available since 2013 at https://github.com/ssolo/ALE.git.

Top five publications

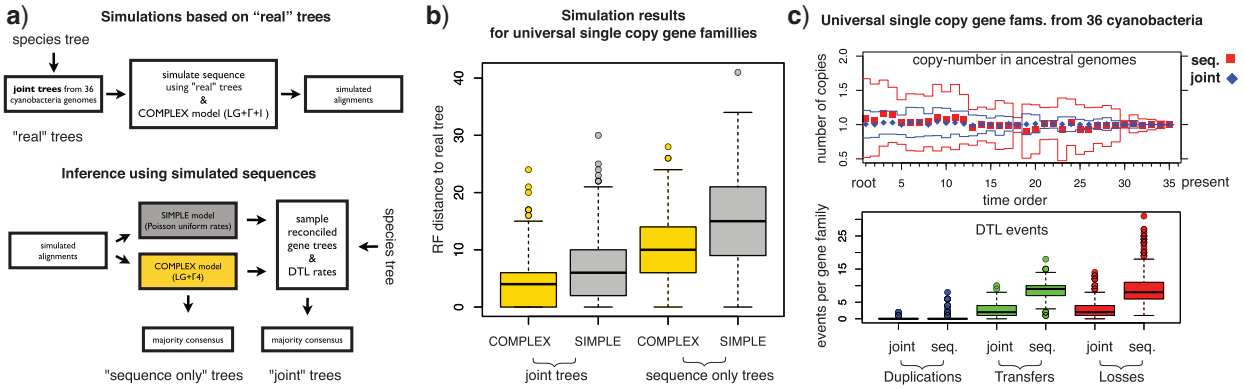


FIGURE 2. Validating joint likelihood-based inference. a) We (i) reconstructed reconciled gene trees that maximise the joint likelihood using homologous gene families from 36 cyanobacterial genomes together with the species tree show in Figure A.4; (ii) simulated sequences using the reconstructed "real" trees and a COMPLEX model of sequence evolution; (iii) sampled gene tree topologies using both a SIMPLE model and the COMPLEX model; (iv) attempted to reconstruct the "real" trees from the simulated sequences using only the sequence alone, and using the joint likelihood together with the species tree for samples from both the SIMPLE and the COMPLEX models. b) The Robinson-Foulds distance to the real trees demonstrates that trees reconstructed from simulated sequences using the joint likelihood are more accurate than those reconstructed based on the sequence alone regardless of the model of sequence evolution used. c) In the top panel, we compare the distribution of the number of genes in ancestral genomes based on reconciliations of gene trees reconstructed from 342 universal single-copy cyanobacterial gene families. The mean number of copies for joint (diamonds, blue online) and sequence trees (squares, red online) is plotted together with the standard deviation (dark and light gray lines, blue and red online). The time order of the speciations corresponds to Figure 3 of Szöllősi et al. (2012). In the lower panel, we compare the number of Duplication, Transfer, and Loss events needed to reconcile joint and sequence trees. For details of the inferences presented see Appendix 1.

Related publications:

Groussin M, Hobbs JK, **Szöllősi GJ**, Gribaldo S, Arcus VL, Gouy M
Towards accurate ancestral protein genotype–phenotype reconstructions with species tree-aware gene trees.
Molecular Biology and Evolution 32, 13-22. (2014) https://doi.org/10.1093/molbev/msu305 (open access)

In this paper we show with simulations that gene trees reconstructed using joint likelihood (using ALE) significantly improve ancestral sequence reconstruction (ASR) accuracy. This underscores the importance of the tree topology in the inference of putative ancestors. We validate our *in silico* predictions using *in vitro* resurrection of the LeuB enzyme for the ancestor of the Firmicutes, a major and ancient bacterial phylum.

Morel B, Kozlov AM, Stamatakis A, **Szöllősi GJ**
GeneRax: A Tool for Species-Tree-Aware Maximum Likelihood-Based Gene Family Tree Inference under Gene Duplication, Transfer, and Loss
Molecular Biology and Evolution 37, 138. (2020) https://doi.org/10.1093/molbev/msaa141 (open access)

We provide an efficient maximum likelihood implementation of the ALE algorithm in collaboration with the Exelixis Lab, developers of RAxML.

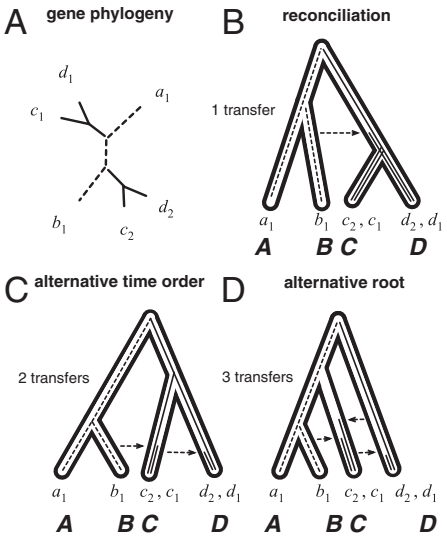


Fig. 1. A gene tree–species tree reconciliation invoking gene transfer and loss. (*A*) A gene tree topology. (*B*) A possible reconciliation of the gene tree in *A* with the species tree invoking one event of transfer. Note that the transfer from the branch leading to ***B*** to above the ancestor of ***C*** and ***D*** implies that the ancestor of ***A*** and ***B*** is older than the ancestor of ***C*** and ***D***. (*C* and *D*) Alternative time orders (*C*) or rootings (*D*) of the species tree violate this condition and only allow reconciliations with a larger number of events.

#5 Phylogenetic modeling of lateral gene transfer reconstructs the pattern and relative timing of speciations **Szöllősi GJ**, Boussau B, Abby SS, Tannier E, Daubin V
Proceedings of the National Academy of Sciences 109, 17513-17518. (2012)
https://doi.org/10.1073/pnas.1202997109

This proof of concept study is ,y first significant publication as a postdoc. Using a combination of simulations and several genome scale datasets we illustrated the potential of reconciliation methods for reconstructing genome scale phylogenies and demonstrated that lateral gene transfers, detected by probabilistic models of genome evolution, can be used as a source of information on the timing of evolution, providing a valuable complement to the limited prokaryotic fossil record.