apply!

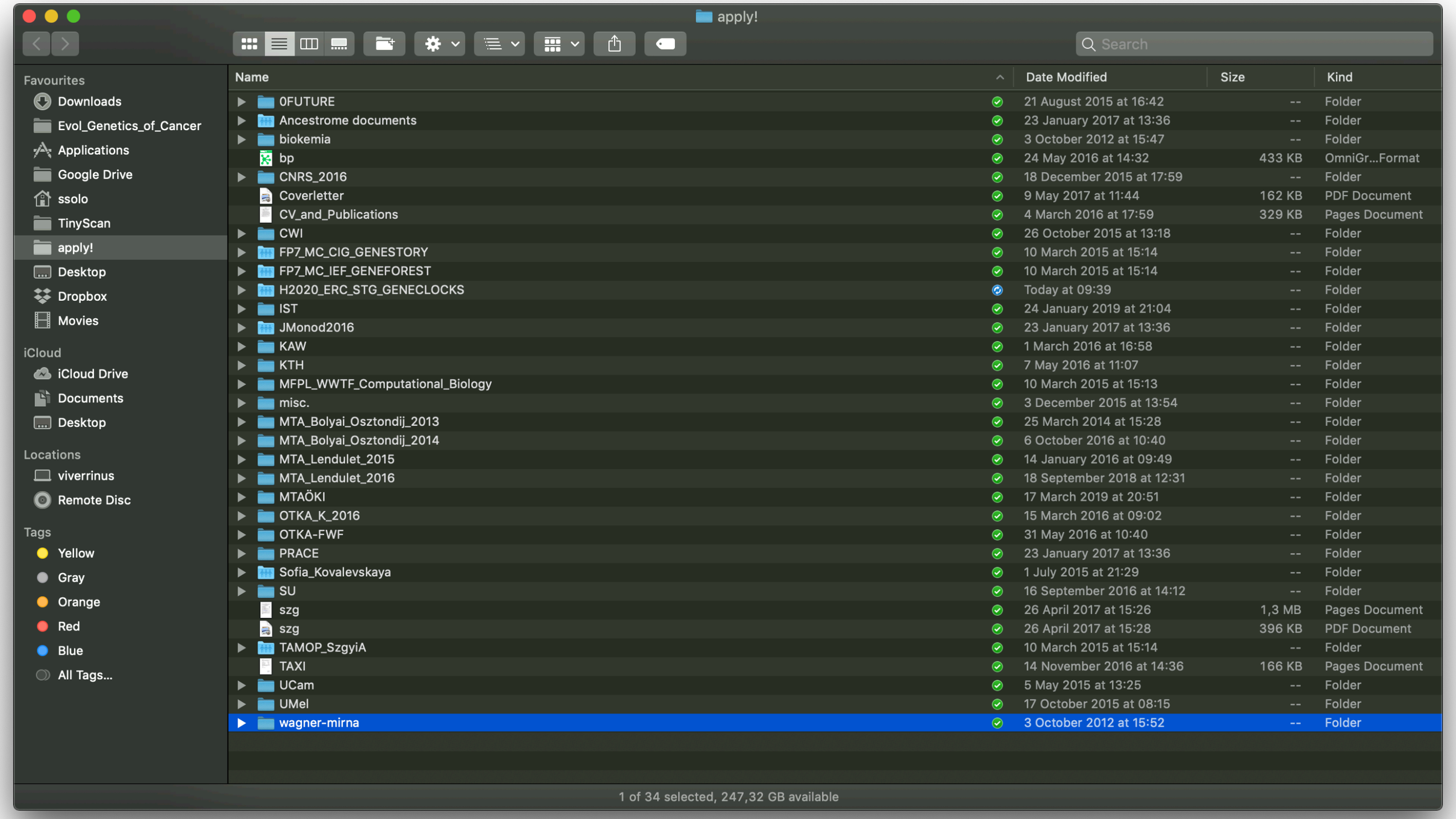
2009 PhD in Physics, Eötvös Uni. Budapest, Hungary — Theoretical Evolution

2010–2011 CNRS Post Doc LBBE, Lyon France

2011–2013 **“GENEFOREST” Marie Curie fellowship** Lyon France

2013– **“GENESTORY” Lead Researcher,** Eötvös Uni. Budapest — Evolutionary Genomics

apply!


apply!
Search
Favourites
Downloads
Evol_Genetics_of_Cancer
Applications
Google Drive
ssolo
TinyScan
apply!
Desktop
Dropbox
Movies
iCloud
iCloud Drive
Documents
Desktop
Locations
viverrinus
Remote Disc
Tags
Yellow
Gray
Orange
Red
Blue
All Tags...
Name
Date Modified
Size
Kind
0FUTURE 21 August 2015 at 16:42 -- Folder
Ancestrome documents 23 January 2017 at 13:36 -- Folder
biokemia 3 October 2012 at 15:47 -- Folder
bp 24 May 2016 at 14:32 433 KB OmniGr...Format
CNRS_2016 18 December 2015 at 17:59 -- Folder
Coverletter 9 May 2017 at 11:44 162 KB PDF Document
CV_and_Publications 4 March 2016 at 17:59 329 KB Pages Document
CWI 26 October 2015 at 13:18 -- Folder
FP7_MC_CIG_GENESTORY 10 March 2015 at 15:14 -- Folder
FP7_MC_IEF_GENEFOREST 10 March 2015 at 15:14 -- Folder
H2020_ERC_STG_GENECLOCKS Today at 09:39 -- Folder
IST 24 January 2019 at 21:04 -- Folder
JMonod2016 23 January 2017 at 13:36 -- Folder
KAW 1 March 2016 at 16:58 -- Folder
KTH 7 May 2016 at 11:07 -- Folder
MFPL_WWTF_Computational_Biology 10 March 2015 at 15:13 -- Folder
misc. 3 December 2015 at 13:54 -- Folder
MTA_Bolyai_Osztondij_2013 25 March 2014 at 15:28 -- Folder
MTA_Bolyai_Osztondij_2014 6 October 2016 at 10:40 -- Folder
MTA_Lendulet_2015 14 January 2016 at 09:49 -- Folder
MTA_Lendulet_2016 18 September 2018 at 12:31 -- Folder
MTAÖKI 17 March 2019 at 20:51 -- Folder
OTKA_K_2016 15 March 2016 at 09:02 -- Folder
OTKA-FWF 31 May 2016 at 10:40 -- Folder
PRACE 23 January 2017 at 13:36 -- Folder
Sofia_Kovalevskaya 1 July 2015 at 21:29 -- Folder
SU 16 September 2016 at 14:12 -- Folder
szg 26 April 2017 at 15:26 1,3 MB Pages Document
szg 26 April 2017 at 15:28 396 KB PDF Document
TAMOP_SzgyiA 10 March 2015 at 15:14 -- Folder
TAXI 14 November 2016 at 14:36 166 KB Pages Document
UCam 5 May 2015 at 13:25 -- Folder
UMel 17 October 2015 at 08:15 -- Folder
wagner-mirna 3 October 2012 at 15:52 -- Folder
1 of 34 selected, 247,32 GB available

127

5 Key Questions (ERC and B1 Excellence)

Educate the Evaluator with 'Facts' and 'Figures'

Why bother? (what new knowledge are you generating?)

Will this establish European International leader?

Is the knowledge already available (state-of-the art)?

Why now? (Why was this not done before now?)

Why you? (Are you the best people to do this work?)

Questions must be answered in the first 15 seconds of the proposal!

“If you want a grant, impress your peers, understand the evaluation process… [and] stop writing boring proposals! You are killing the evaluators. ”

Seán McCarthy

European Research Council

Established by the European Commission

5 Key Questions (ERC and B1 Excellence)	Educate the Evaluator with ‘Facts’ and ‘Figures’
Why bother? (what new knowledge are you generating?)	Are the key journals going to publish the results?
Will this establish Europe as International leader?	Are the key journals going to publish the results?
Is the knowledge already available (state-of-the art)?	Is it beyond state-of-the-art?
Why now? (Why was this not done before now?)	why the hell wasn’t this done before?
Why you? (Are you the best people to do this work?)	are you really the best scientist for the job?

Questions must be answered in the first 15 seconds of the proposal!

understand the evaluation process…

EVALUATION CRITERIA

Criterion 1 - RESEARCH PROJECT
Ground-breaking nature and potential impact of the research project.
To what extent does the proposed research address important challenges?
To what extent are the objectives ambitious and beyond the state of the art (e.g. novel concepts and approaches or development across disciplines)?
To what extent is the proposed research high risk/high gain?

Scientific Approach.
To what extent is the outlined scientific approach feasible bearing in mind the extent that the proposed research is high risk/high gain?
To what extent is the proposed research methodology appropriate to achieve the goals of the project?
To what extent does the proposal involve the development of novel methodology?
To what extent are the proposed timescales and resources necessary and properly justified?

Criterion 2 - PRINCIPAL INVESTIGATOR
Intellectual capacity, creativity and commitment
The questions below can have one of the following four responses: Outstanding/Excellent/Very good/Non-competitive
To what extent has the PI demonstrated the ability to propose and conduct ground-breaking research?
To what extent does the PI provide evidence of creative independent thinking?
To what extent have the achievements of the PI typically gone beyond the state of the art?
To what extent does the PI demonstrate the level of commitment to the project necessary for its execution and the willingness to devote a significant amount of time to the project (min 50% of the total working time on it and min 50% in an EU Member State or Associated Country)?

understand the evaluation process…

high risk / high gain

100% izgalmas kérdések

100% ambiciózus célok

50-80% megvalósíthatóság

The panel was impressed by the innovative work of the PI in developing new phylogenetic methods based on the idea of using horizontal transfer events as information, rather than as a problem.

The proposal, which centred on this idea and methodological breakthrough, was well received. In particular, it was clear to the panel that the application of these methods will offer a new way to evaluate the timing of ancient evolutionary events. It was also clear that this method has the potential to improve the estimates of the timings of evolutionary events. Some concerns were raised, however, as to whether the method could be usefully applied across such a wide timescales, and across so many major evolutionary events, as suggested in the proposal. Some reflection on these issues would be helpful

Proposal ID **714774** *Acronym* **GENECLOCKS** *Go to*

1 - General information

Topic	ERC-2016-STG	Type of action	ERC-STG
Call identifier	ERC-2016-STG	Acronym*	GENECLOCKS

Proposal title* Reconstructing a dated tree of life using phylogenetic incongruence

Note that for technical reasons, the following characters are not accepted in the Proposal Title and will be removed: < > " &

Duration in months* *60*

Primary ERC Review Panel* LS8 - Evolutionary, Population and Environmental Biology

Secondary ERC Review Panel LS2 - Genetics, Genomics, Bioinformatics and Systems Biology (if applicable)

Proposal ID **714774** Acronym **GENECLOCKS** Go to

Abstract*

With the advent of genome-scale sequencing, molecular phylogeny, which reconstructs gene trees from homologous sequences, has reached an impasse. Instead of answering open questions, new genomes have reignited old debates. The problem is clear, gene trees are not species trees, each is the unique result of series of evolutionary events. If, however, we model these differences in the context of a common species tree, we can access a wealth of information on genome evolution and the diversification of species that is not available to traditional methods. For example, as horizontal gene transfer (HGT) can only occur between coexisting species, HGTs provide information on the order of speciations. When HGT is rare, lineage sorting can generate incongruence between gene trees and the dating problem can be formulated in terms of biologically meaningful parameters (such as population size), that are informative on the rate of evolution and hence invaluable to molecular dating.
My first goal is to develop methods that systematically extract information on the pattern and timing of genomic evolution by explaining differences between gene trees. This will allow us to, for the first time, reconstruct a dated tree of life from genome-scale data. We will use parallel programming to maximise the number of genomes analysed.
My second goal is to apply these methods to open problems, e.g.: i) to resolve the timing of microbial evolution and its relationship to Earth history, where the extreme paucity of fossils limits the use of molecular dating methods, by using HGT events as "molecular fossils"; ii) to reconstruct rooted phylogenies from complete genomes and harness phylogenetic incongruence to answer long standing questions, such as the of diversification of animals or the position of eukaryotes among archaea; and iii) for eukaryotic groups such as Fungi, where evidence of significant amounts of HGT is emerging our methods will also allow the quantification of the extent of HGT.

Remaining characters 0

Why now? (Why was this not done before now?)

Why bother? (what new knowledge are you generating?)

why the hell wasn't this done before?

Are the key journals going to publish the results?

With the advent of genome-scale sequencing, molecular phylogeny, which reconstructs gene trees from homologous sequences, has reached an impasse. Instead of answering open questions, new genomes have reignited old debates. The problem is clear, gene trees are not species trees, each is the unique result of series of evolutionary events.

Why now? (Why was this not done before now?)

Why bother? (what new knowledge are you generating?)

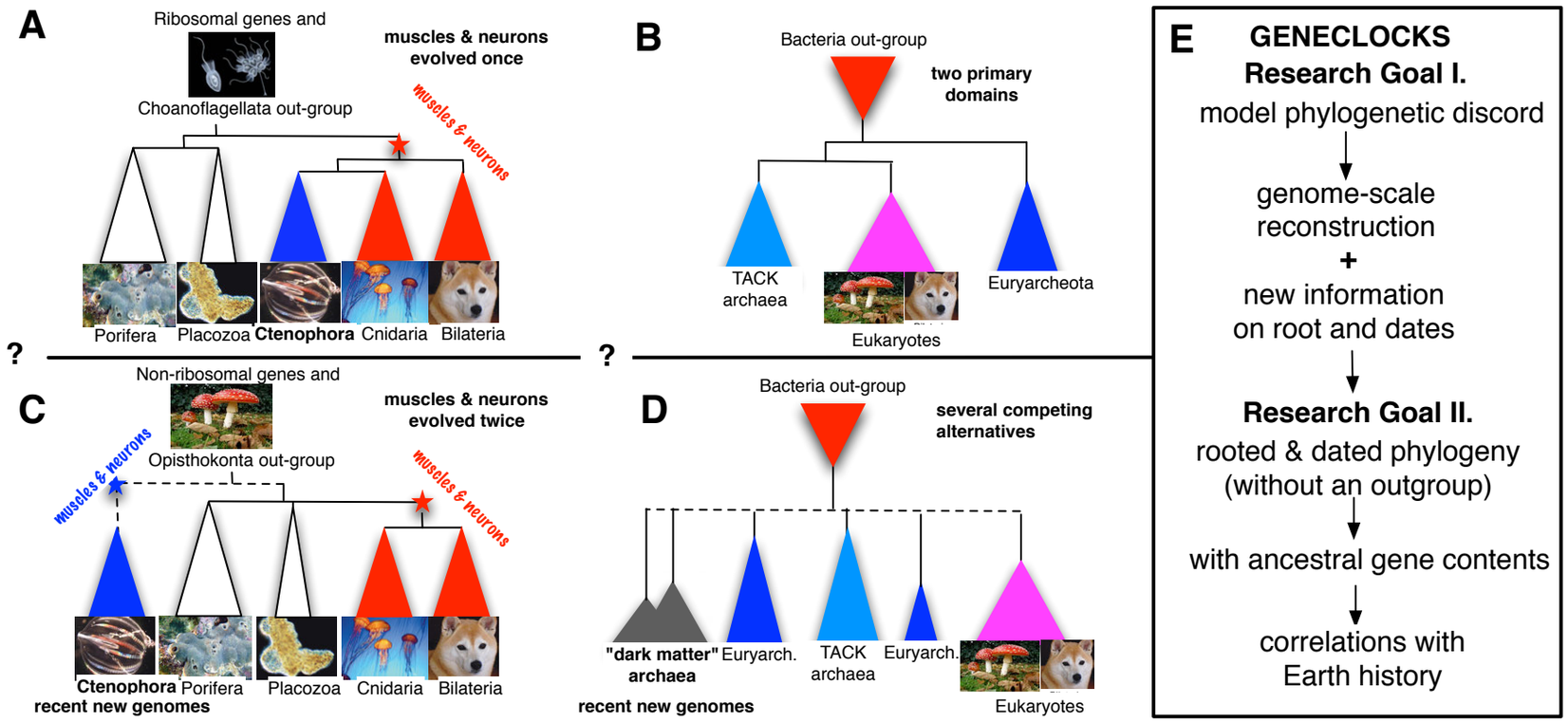


Fig 1. New genomes old questions. The first wave of phylogenetic studies attempting to use complete genomes have in practice relied on a small minority (1%-10%) of genes selected to minimise phylogenetic discord. Nonetheless, these studies produced important results, they (**A**) brought back traditional views on deep animal relationships (Philippe 2009), and (**B**) found support for (only) two primary domains of life (Williams 2012, 2013). In the last two years, however, inclusion of new Ctenophore genomes (Ryan 2013, Moroz 2014) has produced support for alternative branching orders implying multiple origins of the animal nervous system and muscles (**A** vs **C**), which later proved to be conditional on model details such as the choice of out-group and gene set (Noshenko 2013, Pisani 2015). Similarly, new sequences from uncultivated archaeal "dark matter" (**B** vs. **D**) recovered a traditional three domain phylogeny (Rinke 2013) and opened a debate about the branching order of major groups of Archaea (Raymann 2015). (**E**) GENECLOCKS proposes to resolve both questions by modelling phylogenetic discord, in order to i) use of the remaining (90-99%) of genes and more importantly ii) to extract novel information by modelling phylogenetic discord in biological terms.

Educate the Evaluator with 'Facts' and 'Figures'

why the hell wasn't this done before?

Are the key journals going to publish the results?

Is the knowledge already available (state-of-the art)?

Why you? (Are you the best people to do this work?)

Is it beyond state-of-the-art?

The solution is to model how gene trees are generated along the species tree

If, however, the evolution of genomes is modelled as a series of DTL (duplication, transfer and loss) and other genome evolutionary events generating a plurality of gene histories, gene and species phylogenies can be simultaneously reconstructed. Using parallel computing to efficiently consider dozens of complete genomes, we showed in two proof-of-concept studies that this allows genome-scale phylogenetic inference (Szöllősi 2012, Boussau 2013 & Fig.1 part B2) that results in gene trees and ancestral gene contents that are dramatically more accurate, and in general more similar to the species tree, e.g. we found that 2 out of 3 transfer events inferred by traditional species-tree unaware methods are the result of reconstruction errors (Szöllősi 2013a). ***I propose to develop and use novel methods that extract bona fide conflict among gene trees and interpret them in biological terms. These methods will allow us to exploit this novel source of information and hence offer a great hope to resolve issues that have been left pending by traditional methods.***

are you really the best scientist for the job?

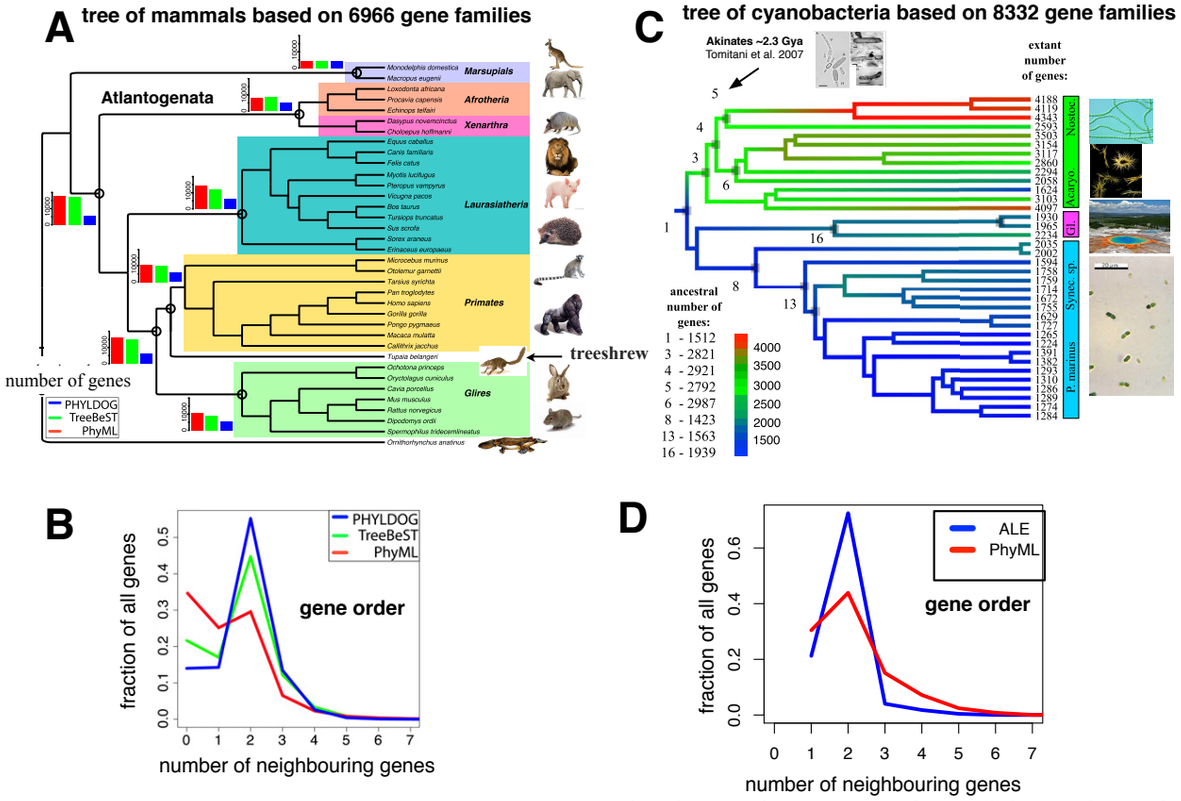


Fig 1. Joint reconstruction produces more accurate gene trees. Using the genetic sequences of several thousand gene families from complete genomes, we *jointly reconstructed the species tree and the forest of gene trees that evolved along it.* (**A**) Reconstruction based on 36 genomes representative of mammalian diversity (Boussau et al. 2013). In the absence of transfer, errors in gene trees result in an overestimation of the number ancestral gene copies, as extra duplications are required to explain spurious discord with the species tree. For select nodes we show the number of gene copies with red bars corresponding to the sequence only PhyML method, green bars to TreeBEST, the method used in the reference database EnsEMBL, and blue to our *joint reconstruction* method PHYLDOG. Our method exhibits the least error. (**B**) As independent validation, we also reconstructed

understand the evaluation process…

Dear Dr. SZOLLOSI,

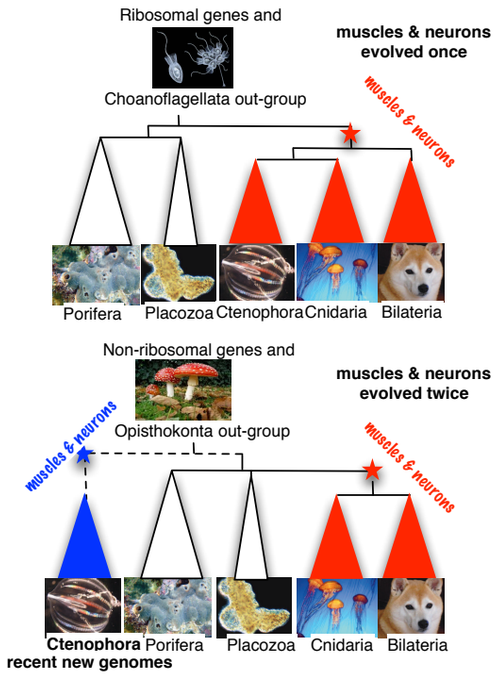
Subject: Initial information on the outcome of the evaluation of proposals submitted to the Call for Proposals ERC-STG-2016 - Proposal n° 714774 GENECLOCKS

I am pleased to inform you that the ERC evaluation panels, composed of independent experts, have favourably reviewed your proposal in Step 1 of the evaluation process. We cordially invite you to attend an interview with the evaluation panel.

Date:	The interviews for your panel will take place from 14 June to 17 June. Please make sure to keep these days free.
Place:	Brussels, Belgium
Interview content:	Interviews will last between 20 and 30 minutes. They will include a short presentation by the applicant and time for questions and answers.

Please also note that as a successful Step 1 applicant, you will not receive further feedback on the Step 1 review of your proposal.

A B2 is fontos! Itt kell elkölteni a pénzt..



?

?

E GENECLOCKS

Research Goal I.

model phylogenetic discord

↓

genome-scale reconstruction

+

new information on root and dates

↓

Research Goal II.

rooted & dated phylogeny (without an outgroup)

↓

with ancestral gene contents

↓

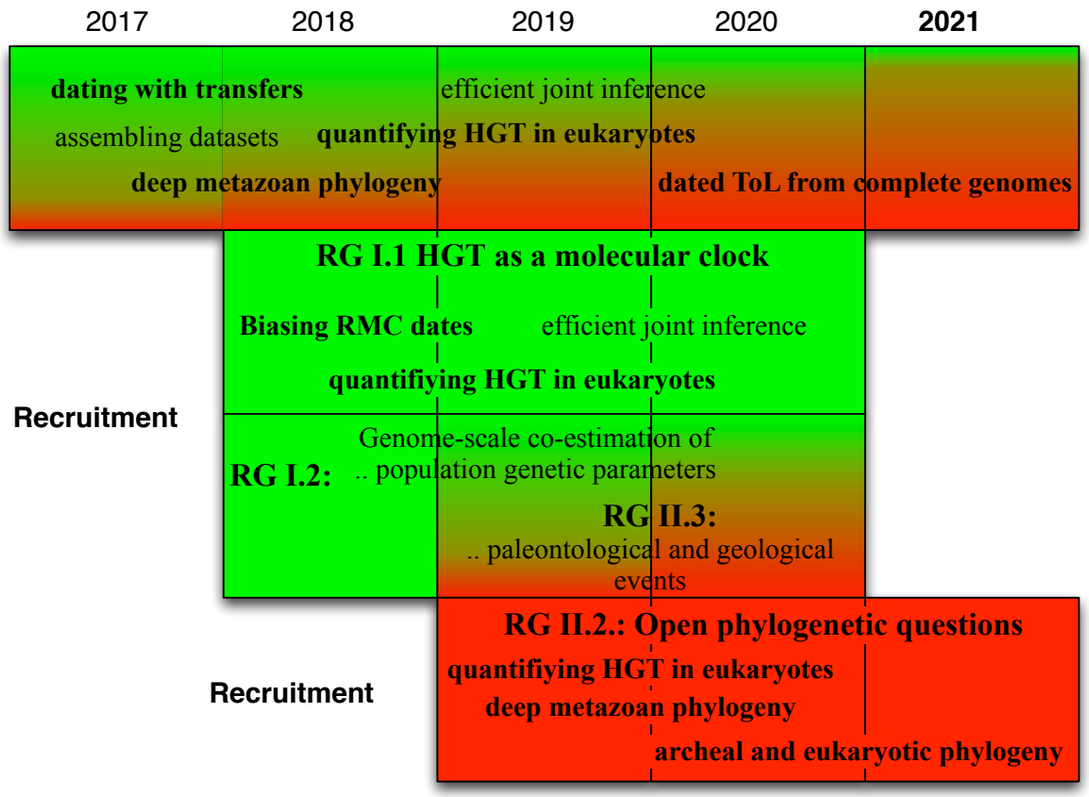
correlations with Earth history

PI

PostDoc 1

PhD 1

PostDoc 2



ERC Starting Grant 2016

RECONSTRUCTING A DATED TREE OF LIFE USING PHYLOGENETIC INCONGRUENCE

GENECLOCKS

PI:
Gergely J. Szöllősi

Why you? (Are you the best people to do this work?)

PI's Brief CV

Dr. Gergely J Szöllősi

2009 PhD in Physics, Eötvös Uni. Budapest, Hungary — Theoretical Evolution

2010–2011 CNRS Post Doc LBBE, Lyon France

2011–2013 **"GENEFOREST" Marie Curie fellowship** Lyon France

2013– **"GENESTORY" Lead Researcher,** Eötvös Uni. Budapest — Evolutionary Genomics

Publications

25 publications, **11 as first 3 as last author;**
> 800 citations, IF: 163 (95 as first or last author);
h-index of 15 (11 as first or last author).

Citations per year

Funding as PI ~ 300 000 EUR

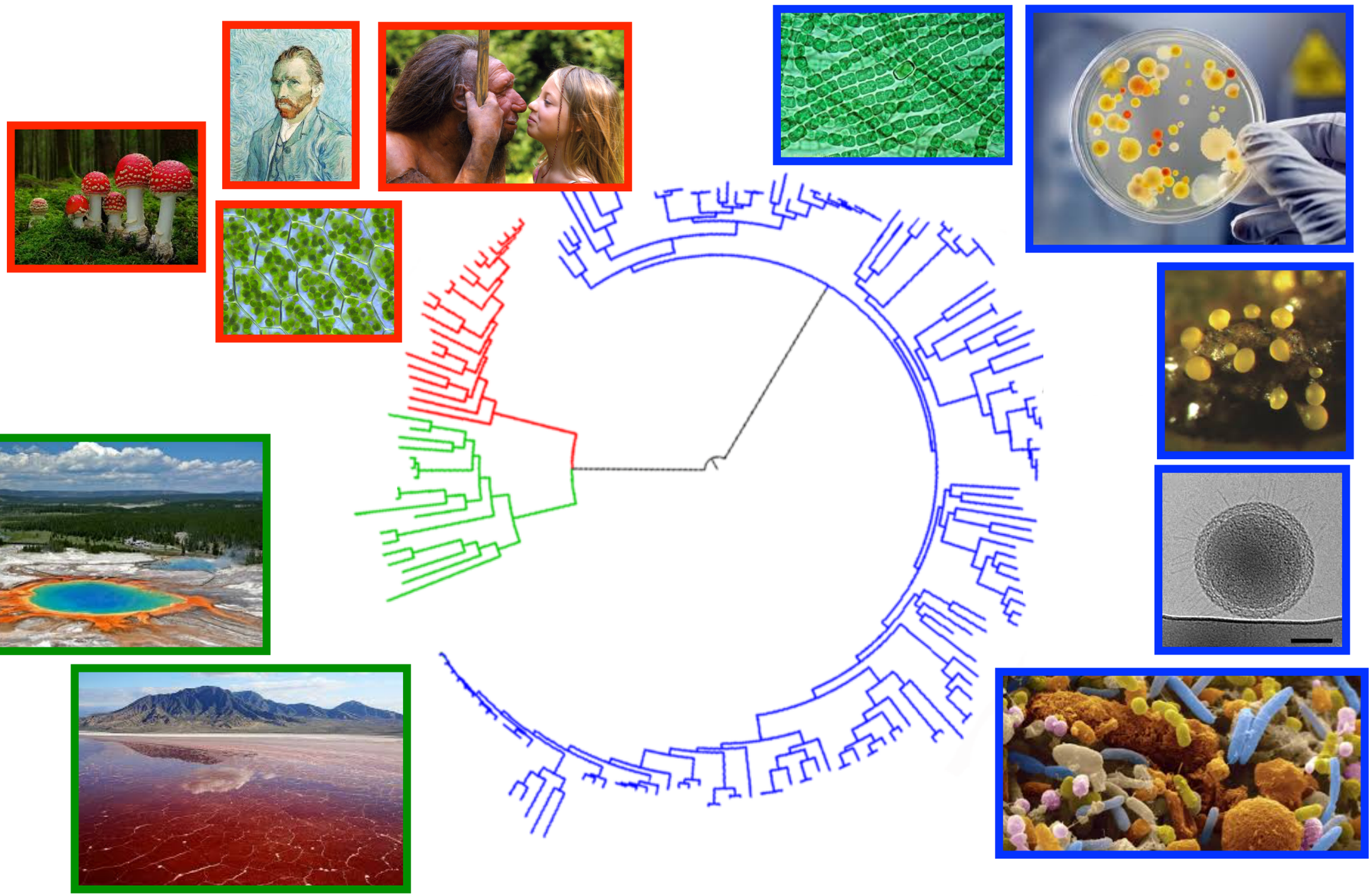
2014-2017	MTA Bolyai fellowship	15 000 EUR
2013-2016	FP7-PEOPLE-CIG "GENESTORY"	87 500 EUR
2013-2014	Alber Szentgyörgyi Excellence Fellowship	26 000 EUR
2011-2013	FP7-PEOPLE-IEF "GENEFOREST"	178 000 EUR

Student supervision: 1 PhD, 4 MSc, 5 BSc

2015- supervisor, PhD Biological physics Zsófia Kéri at Eötvös University, Budapest
2013-2015 supervisor, M.Sc. thesis in Biological Physics of **Zsófia Kéri** at Eötvös University, Budapest
2014 co-supervisor, M.Sc. thesis in Bioniformatics of **Benjamin Horvilleur,** at LBBE in Lyon
2013 supervisor, B.Sc. thesis in Physics of **Zsófia Kéri,** Eötvös University, Budapest
2010 co-supervisor, M.Sc. thesis in Bioniformatics of **Wojciech Roskiewicz,** LBBE in Lyon

Why now? (Why was this not done before now?)

The golden age of molecular evolution?



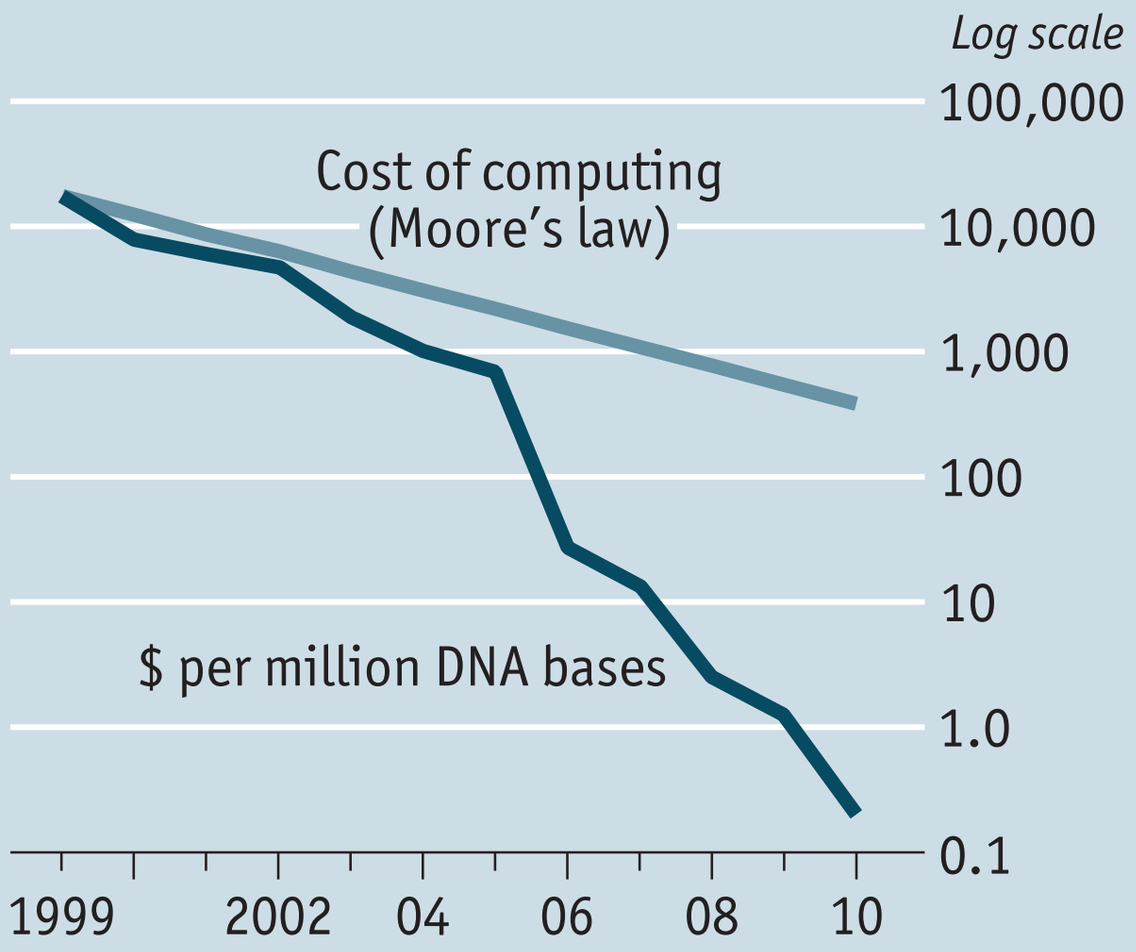
Why now? (Why was this not done before now?)

Biology 2.0

Baseline information

1

Cost of genome sequencing compared with Moore's law for computers



Source: Broad Institute

Educate the Evaluator with 'Facts' and 'Figures'

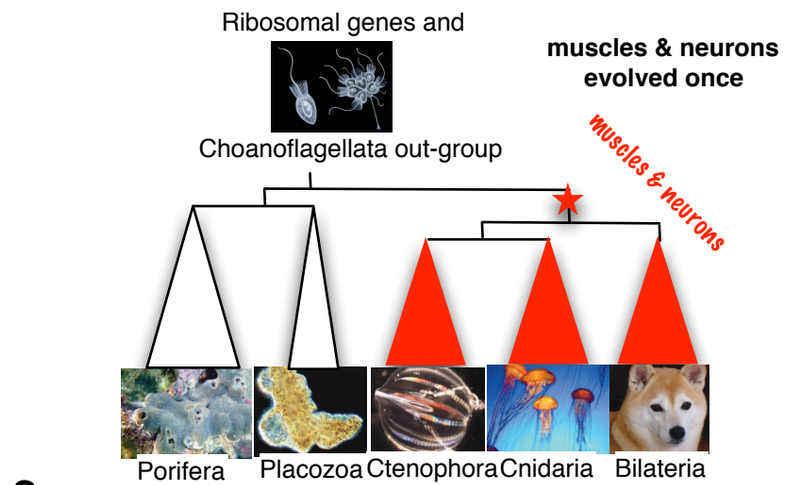
GENECLOCKS

Szöllősi GJ

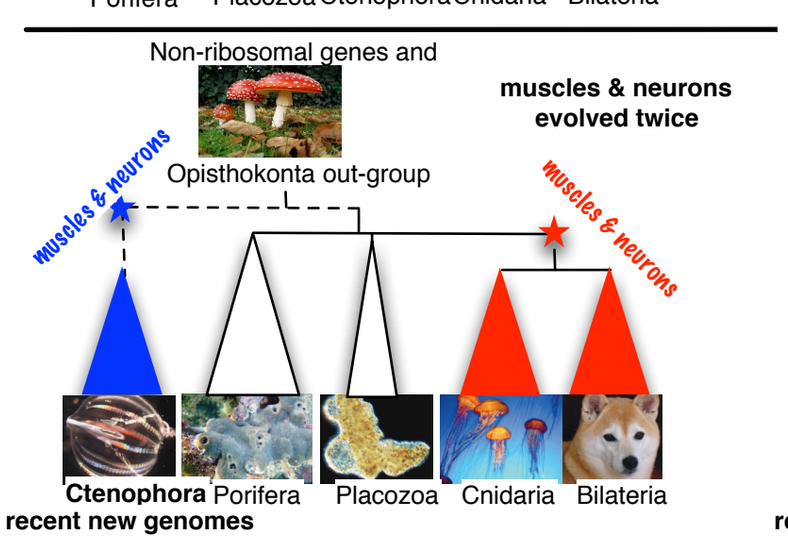
Why now? (Why was this not done before now?)

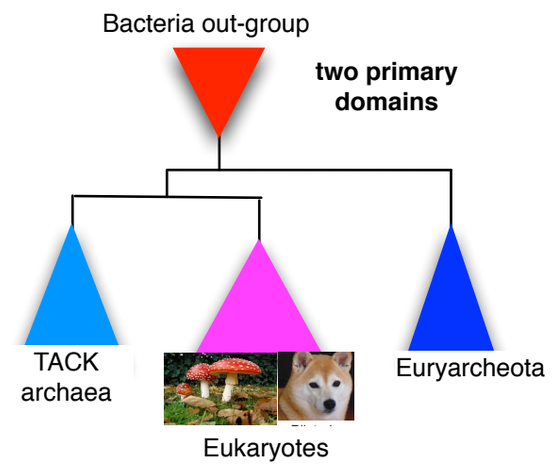
New genomes, old questions

New genomes instead of bringing into sharper focus major evolutionary events such as the origin of eukaryotes or the diversification of major animal lineages have instead reignited old debates.

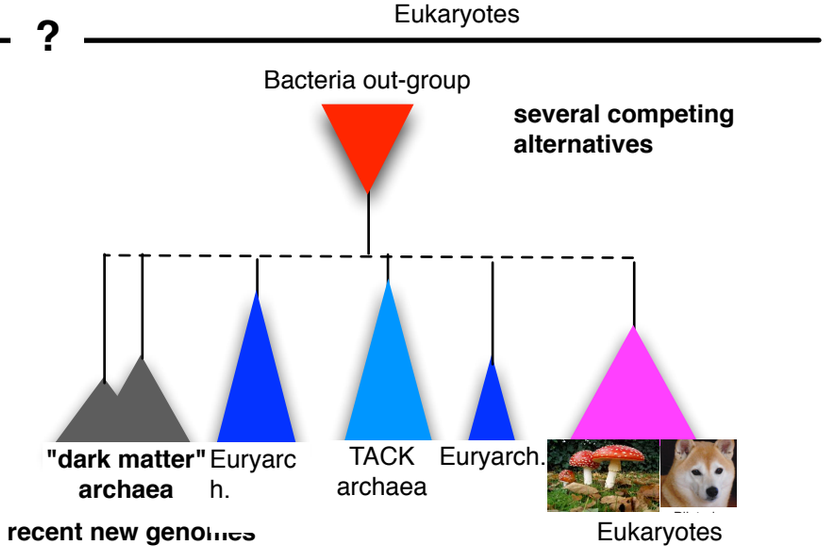


?





?



Why bother? (what new knowledge are you generating?)

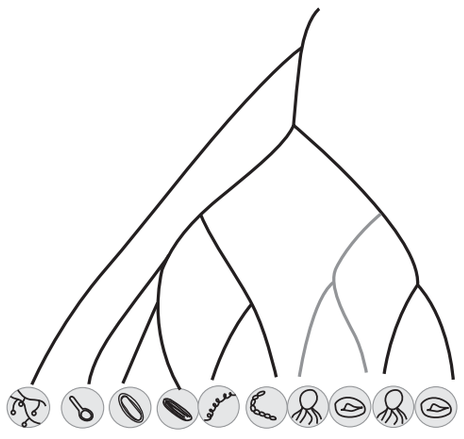
Are the key journals going to publish the results?

Educate the Evaluator with 'Facts' and 'Figures'

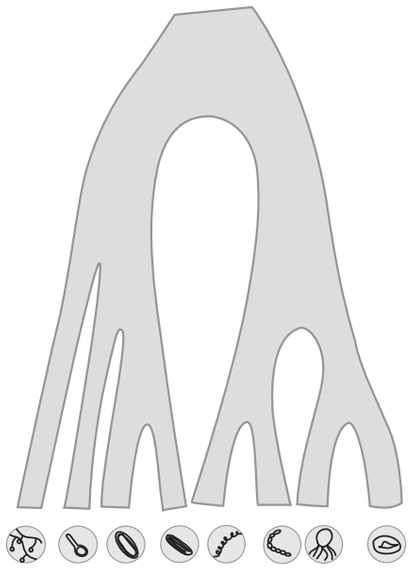
Is the knowledge already available (state-of-the art)?

The problem is gene trees are not species trees

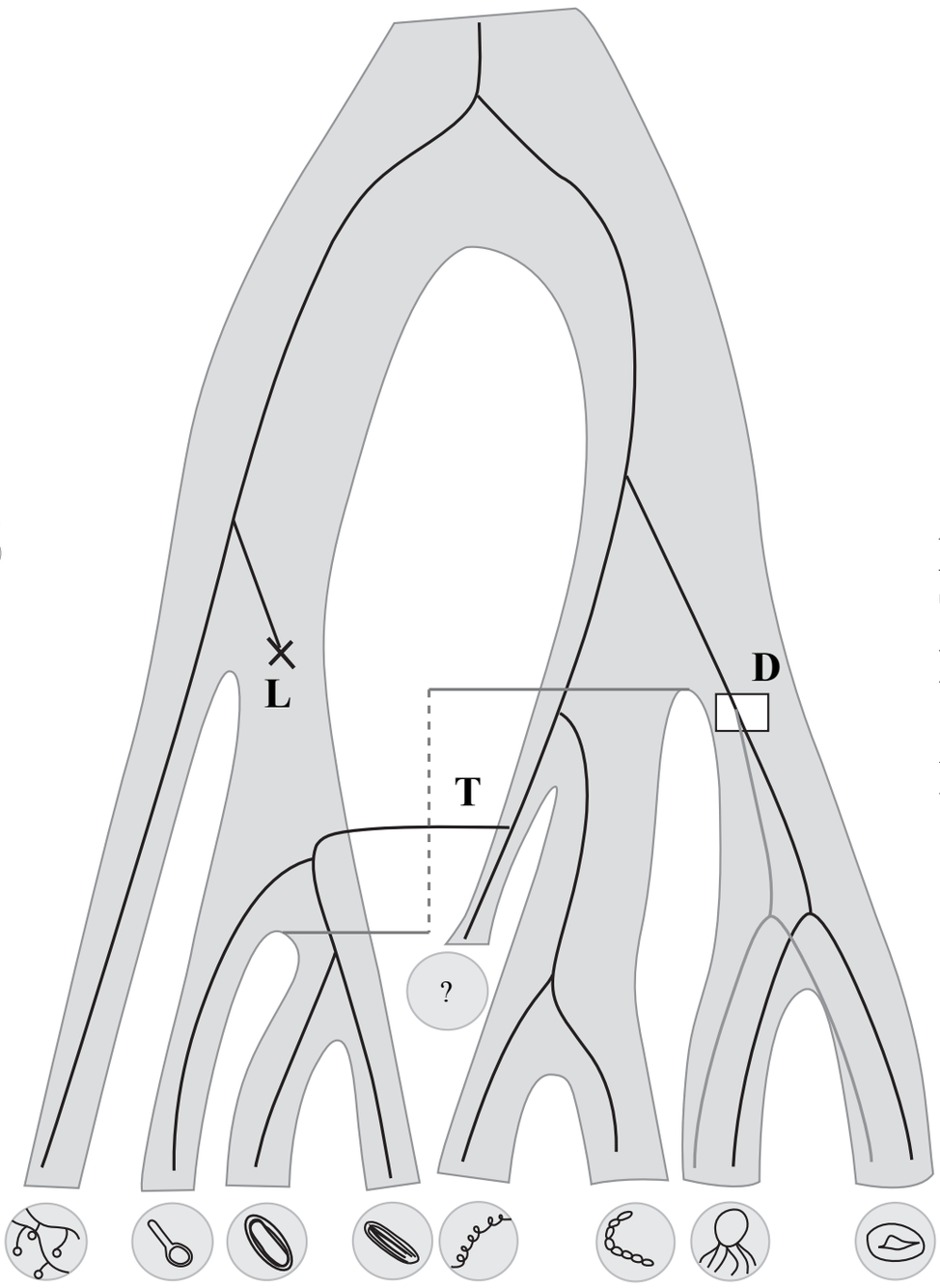
Gene tree



Species tree



Gene tree evolving along the species tree



Duplication
Transfer
Loss

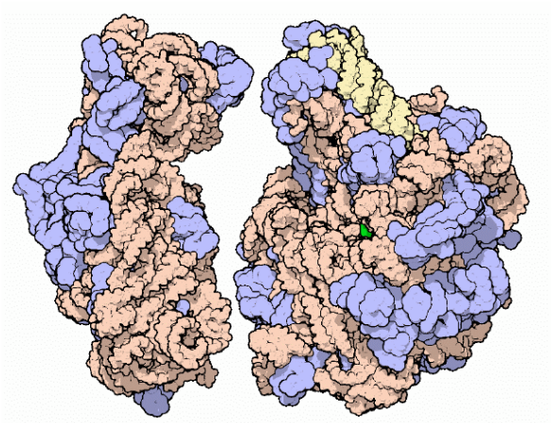
ILS

Educate the Evaluator with 'Facts' and 'Figures'

Daubin & Szöllősi 2016

Is the knowledge already available (state-of-the art)?

The history of the 1%

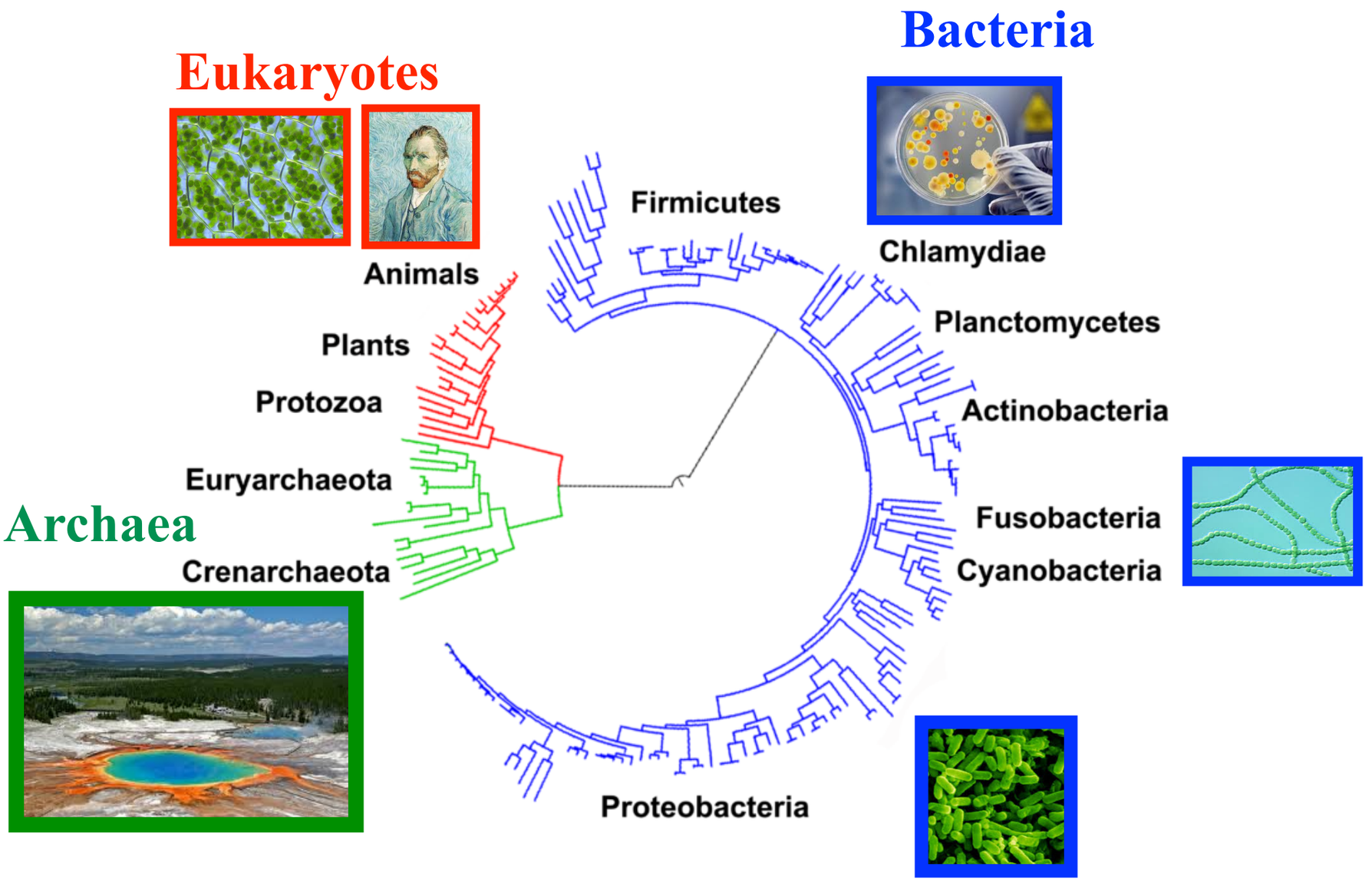


16S rRNA

Carl Woese, 1977

at most a few dozen

Ciccarelli, 2006



Educate the Evaluator with ‘Facts’ and ‘Figures’

.. but gene trees are generated along the species tree

The stories of individual gene families all take place along the same species tree. If we can model the process generating gene trees along the species tree, we can hope to infer better gene trees and species trees.

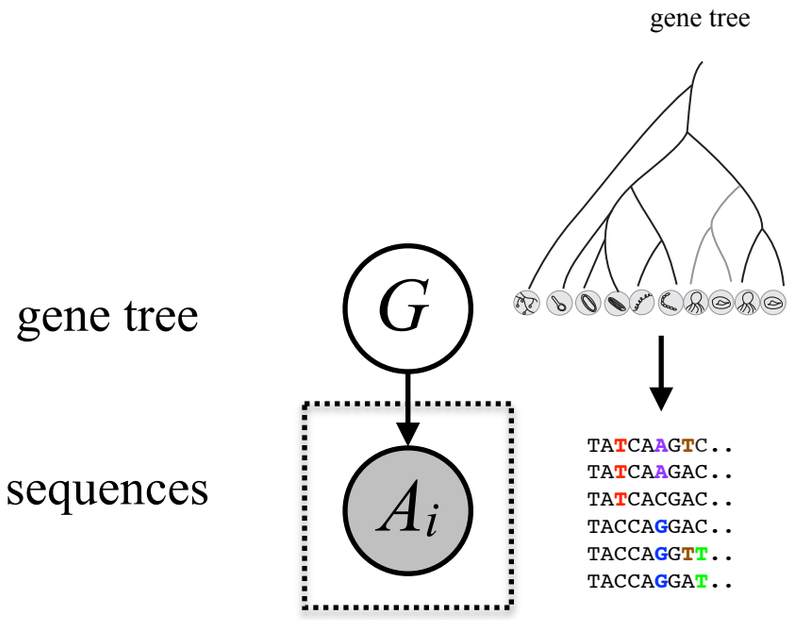
Daubin & Boussau 2011

Is the knowledge already available (state-of-the art)?

The solution is to model how gene trees are generated along the species tree

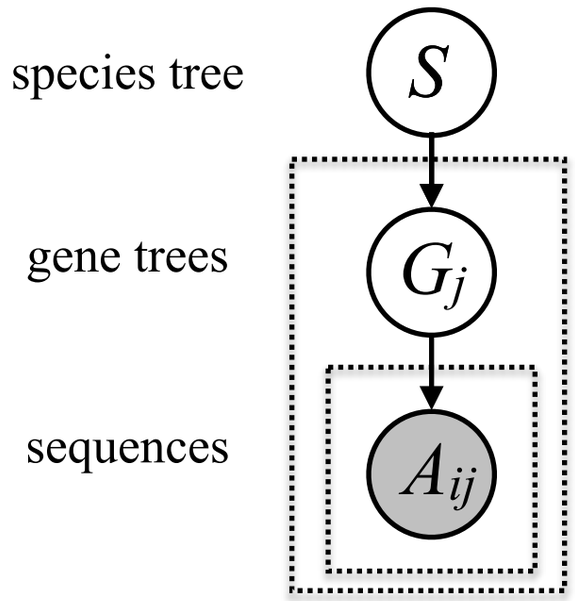
“sequence only” inference

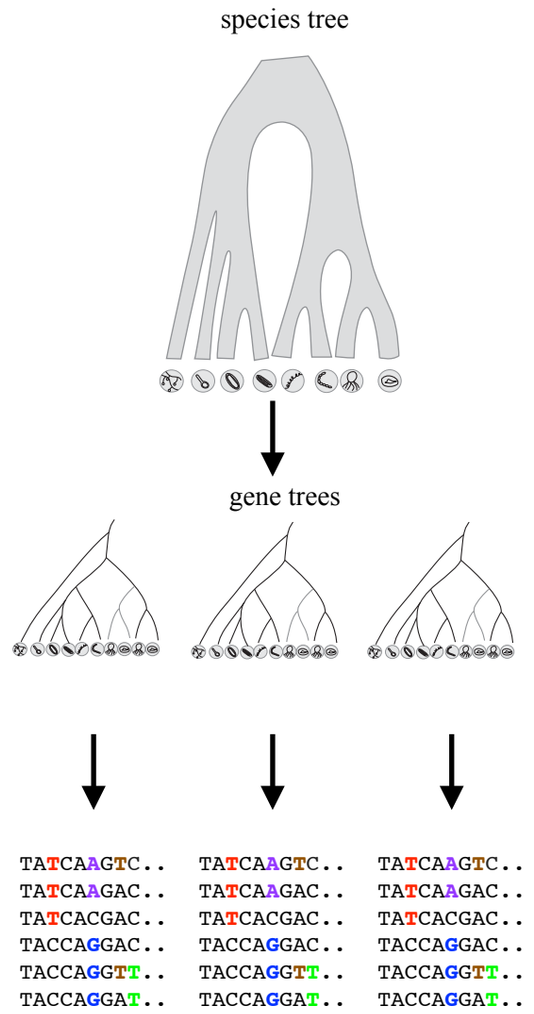
species tree-unaware



“joint” inference

species tree-aware





Szöllősi, Tannier, Lartillot & Daubin ***Systematic Biology*** (2013)
Lateral Gene Transfer from the Dead

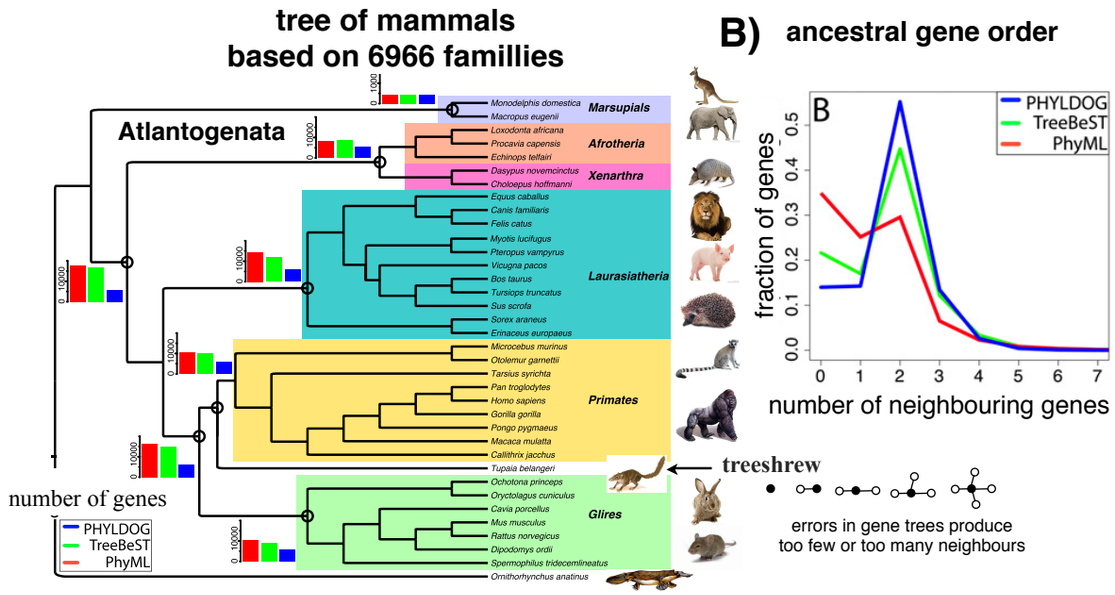
Szöllősi, Rosikiewicz, Boussau, Tannier & Daubin ***Systematic Biology*** (2013)
Efficient exploration of the space of reconciled gene trees

Szöllősi, Tannier, Daubin & Boussau ***Systematic Biology*** (2015)
The inference of gene trees with species trees

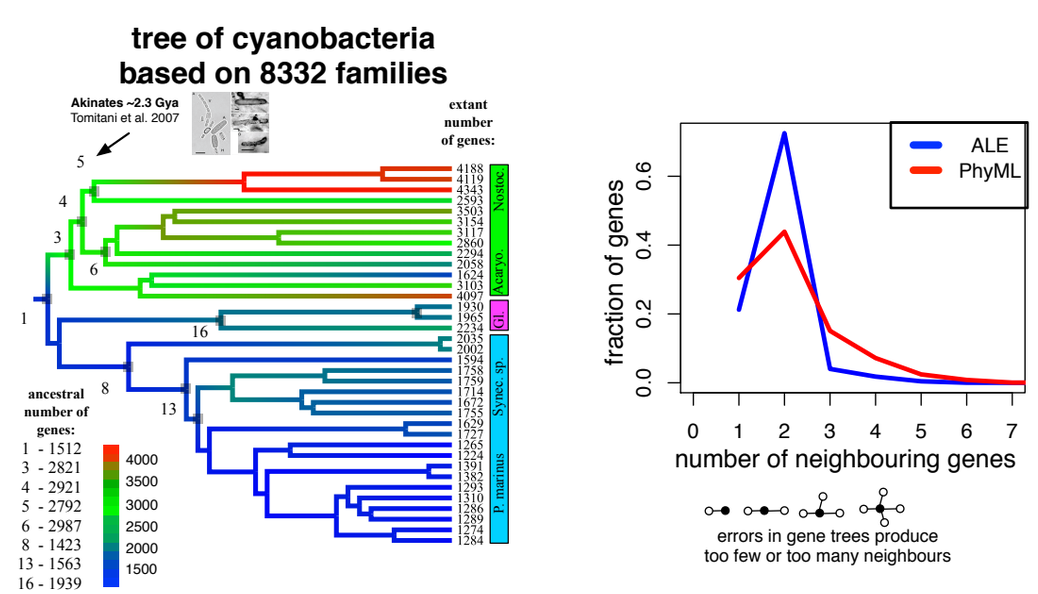
Genome-scale reconstruction of gene and species trees

In the last 5 years I developed methods that reconstruct gene and species phylogenies simultaneously. Using parallel computing, these methods have been able to efficiently consider datasets composed of a large number of complete genomes with unprecedented accuracy.

Aside of the quantitative advance of leveraging genome scale data, these **open the possibility of extract qualitatively new information from differences in gene histories (i.e. phylogenetic discord).**



Boussau, **Szöllősi**, Duret, Gouy, Tannier & Daubin ***Genome Res.*** (2013)
Genome-scale coestimation of species and gene trees



Szöllősi, Boussau, Abby, Tannier & Daubin ***PNAS*** (2012)
Phylogenetic modeling of lateral gene transfer reconstructs the pattern and relative timing of speciations

Why you? (Are you the best people to do this work?)

GENECLOCKS are you really the best scientist for the job? **Szöllősi GJ**

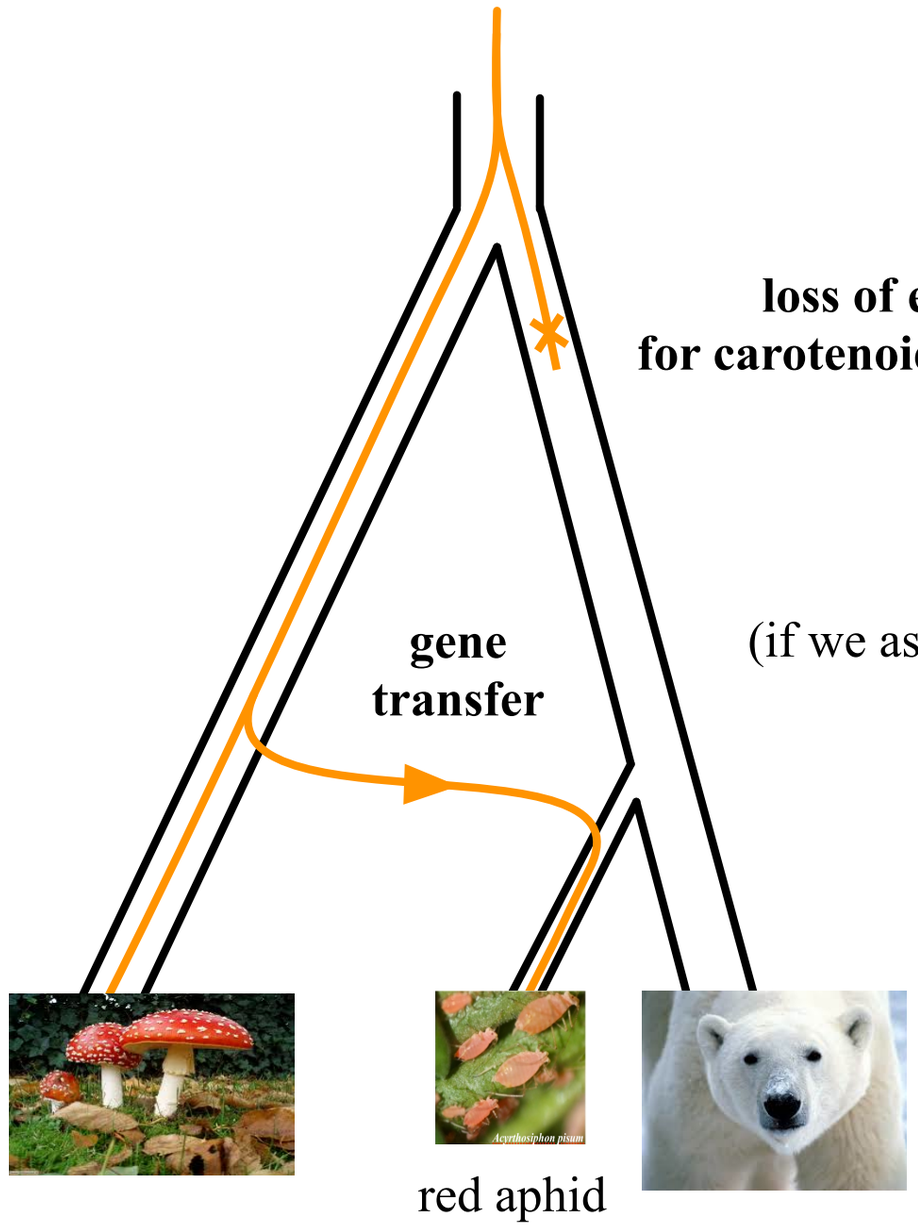
Is the knowledge already available (state-of-the art)?

E.g. horizontal gene transfer

Horizontal gene transfer is common among unicellular organisms, but examples are know even among animals.

pea aphids

Moran & Jarvik 2010

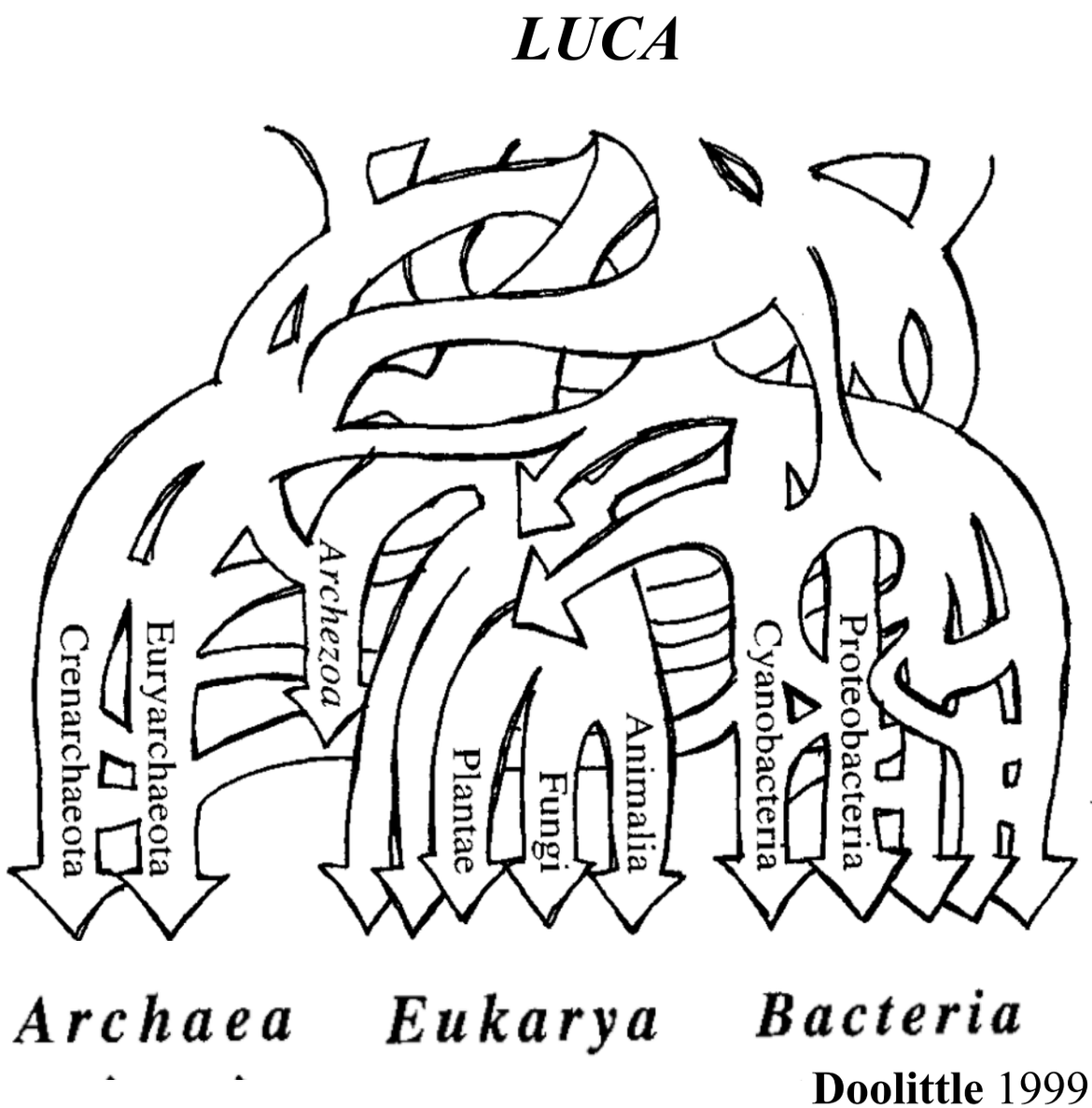


Educate the Evaluator with 'Facts' and 'Figures'

Is the knowledge already available (state-of-the art)?

Horizontal gene transfer as noise

Gene transfers result in apparently contradicting gene phylogenies, fungi can seem closely related to aphids. A potentially high rate of transfer esp. early in the evolution of life, suggests that the vertical signal may be drowned in noise.



Doolittle 1999

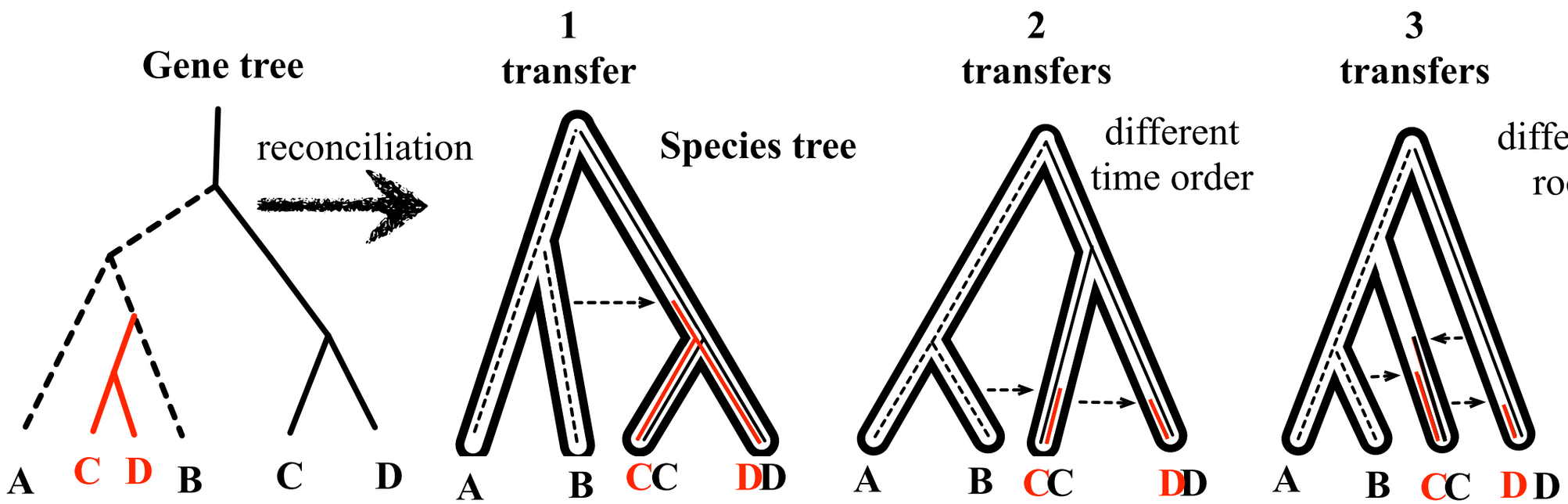
Educate the Evaluator with 'Facts' and 'Figures'

Why you? (Are you the best people to do this work?)

Horizontal gene transfer as information

Transfer events, encoded in the topologies of gene trees can be thought of as "*molecular fossils*" that record the order of speciation events.

Educate the Evaluator with 'Facts' and 'Figures'

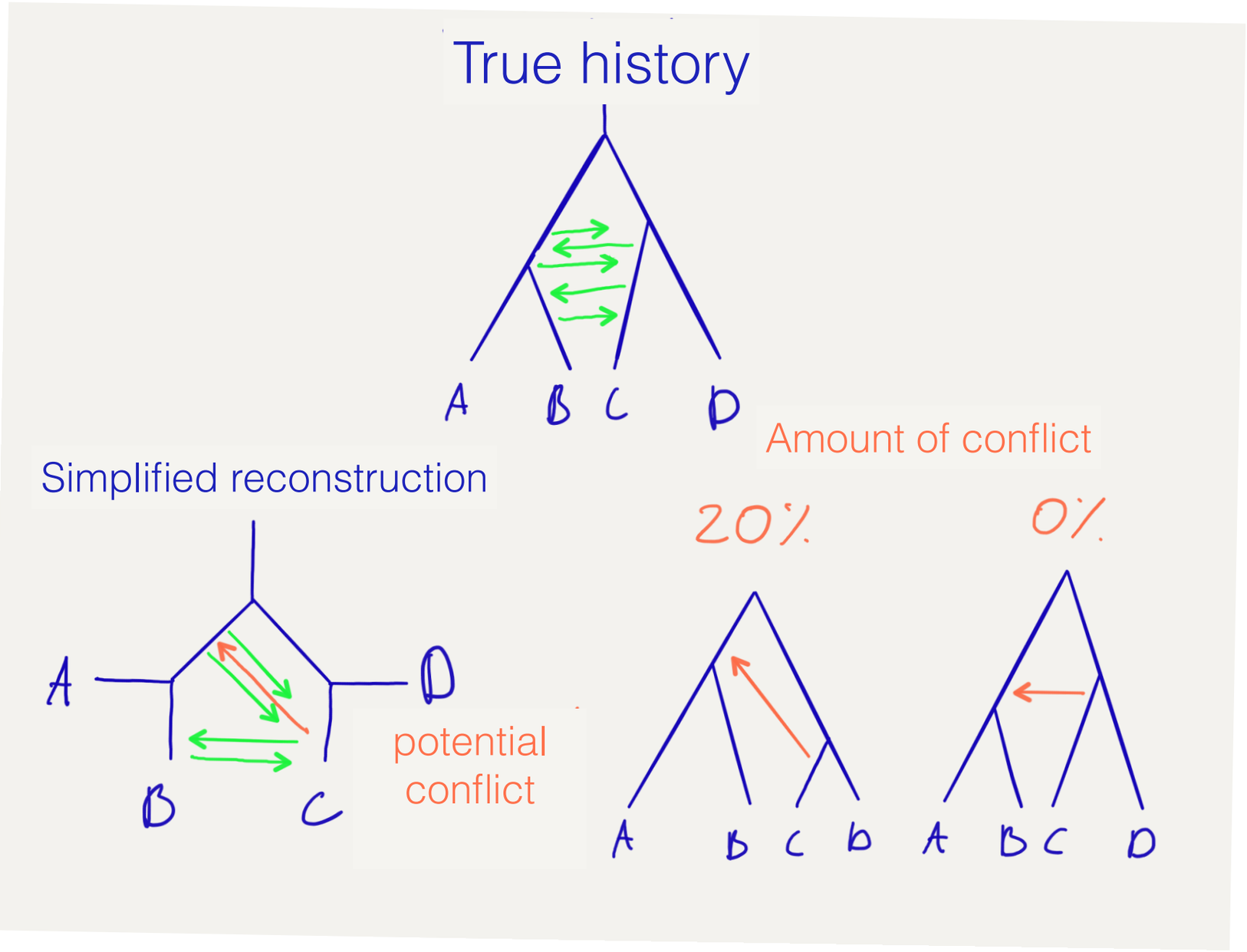


Szöllősi, Tannier, Lartillot & Daubin ***Systematic Biology*** (2013)
Lateral Gene Transfer from the Dead

Szöllősi, Boussau, Abby, Tannier & Daubin ***PNAS*** (2012)
Phylogenetic modeling of lateral gene transfer reconstructs the pattern and relative timing of speciations

Why you? (Are you the best people to do this work?)

Preliminary results:
Phylogenetic incongruence as molecular fossils



Why now? (Why was this not done before now?)

Is it beyond state-of-the-art?

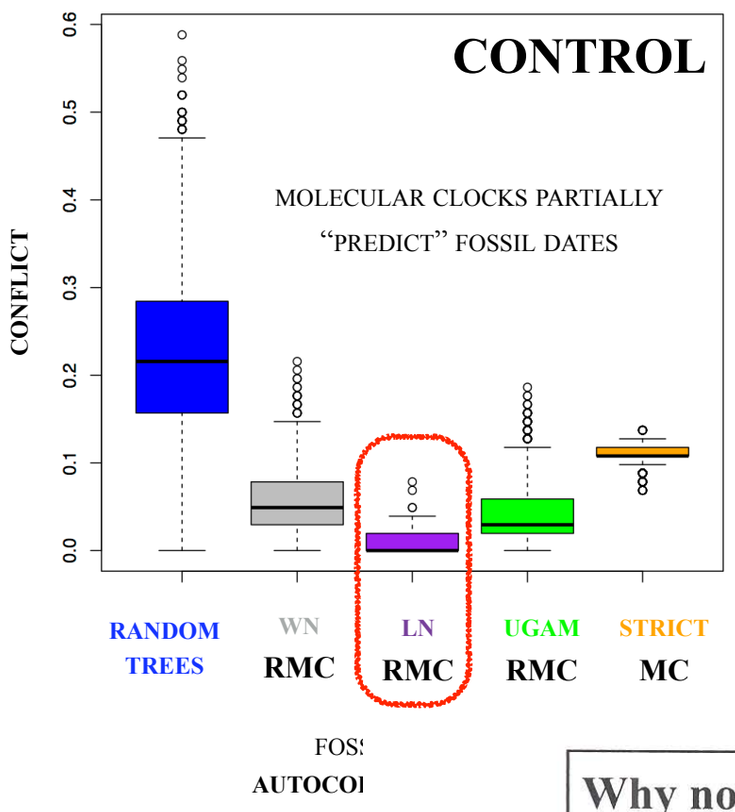
Why you? (Are you the best people to do this work?)

Preliminary results:
Phylogenetic incongruence as molecular fossils

Using a simplified model of to infer transfers from genome scale data we have been able to demonstrate that
i) **gene transfers and sequence based molecular clocks carry partially overlapping dating signal** and
ii) dating information conveyed by gene transfer events can distinguish between RMC methods.

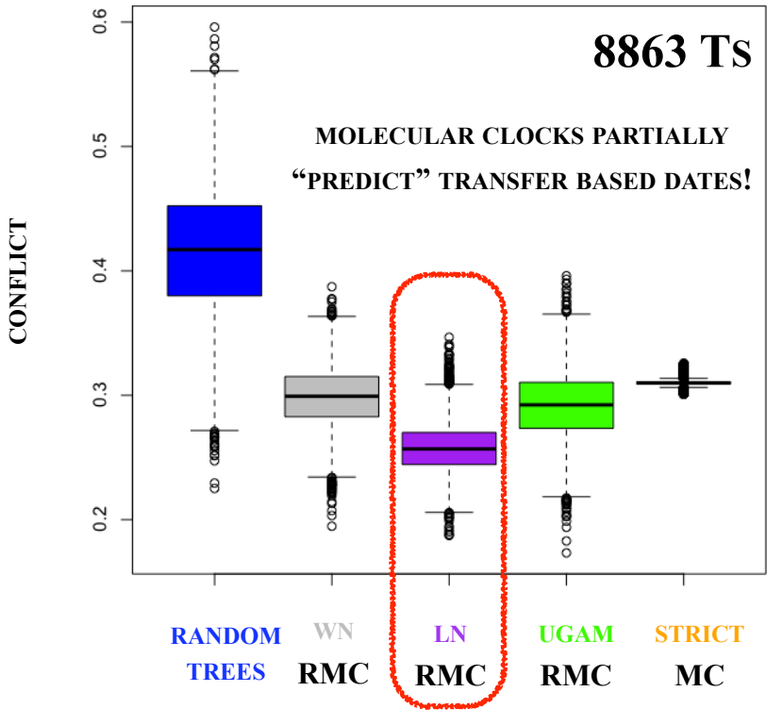
Dating methods combining information from transfers and sequence based molecular clocks have the potential to provide unprecedented resolution using genome-scale data.

CONFLICT WITH **FOSSIL BASED RELATIVE DATES** IN 36 MAMMALS

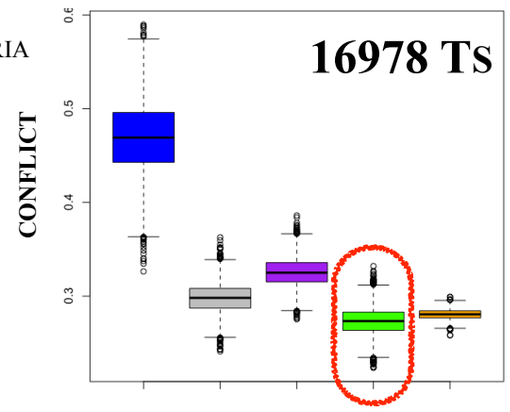


FOS
AUTOCOI

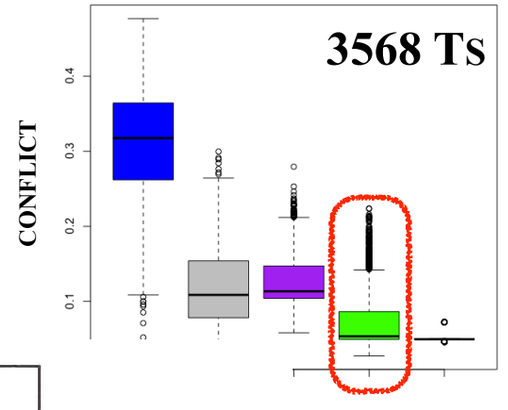
CONFLICT WITH **TRANSFER BASED RELATIVE DATES** IN 36 CYANOBACTERIA



60 Δ-PROTEOBACTERIA



28 FUNGI



NSFERS SUPPORT
UNCORRELATED MODEL

Why now? (Why was this not done before now?)

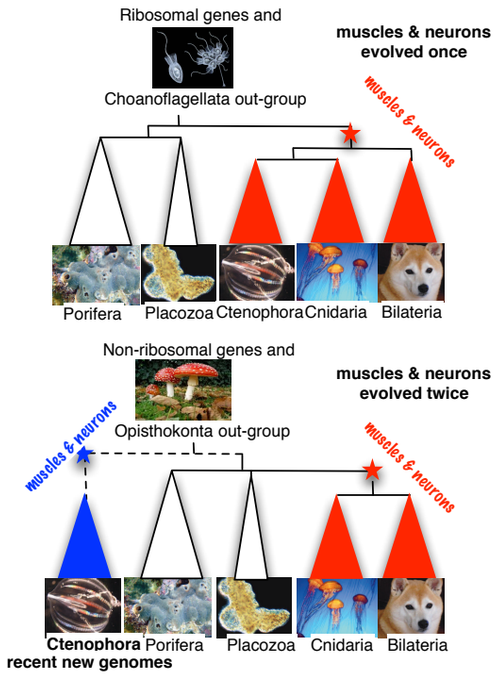
Is it beyond state-of-the-art?

GENECLOCKS

Szöllősi GJ

Goals & Resources

high risk / high gain



GENECLOCKS

Research Goal I.

model phylogenetic discord

↓

genome-scale reconstruction

+

new information on root and dates

↓

Research Goal II.

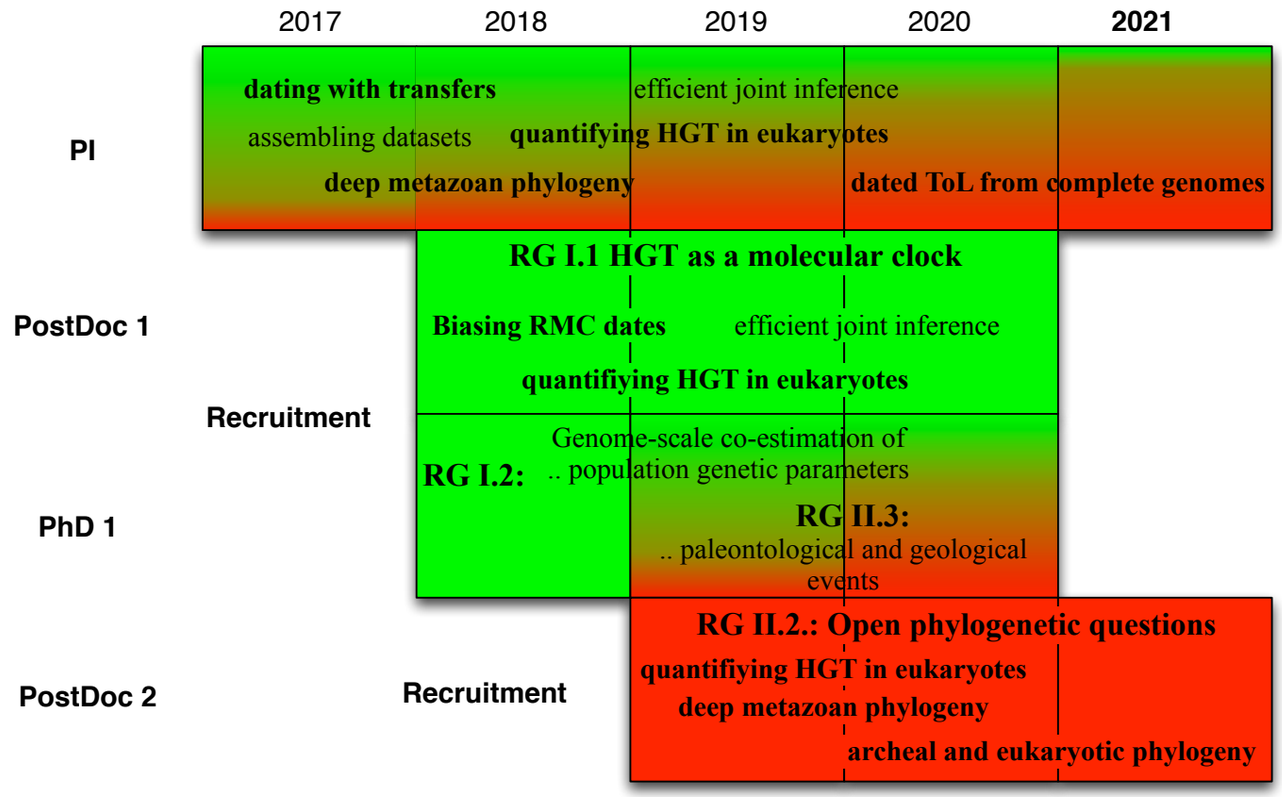
rooted & dated phylogeny (without an outgroup)

↓

with ancestral gene contents

↓

correlations with Earth history



Computing resources provided by host

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

100% izgalmas kérdések

100% ambiciózus célok

50-80% megvalósíthatóság

GENECLOCKS

Szöllősi GJ

EVALUATION CRITERIA

Criterion 1 - RESEARCH PROJECT
Ground-breaking nature and potential impact of the research project.
To what extent does the proposed research address important challenges?
To what extent are the objectives ambitious and beyond the state of the art (e.g. novel concepts and approaches or development across disciplines)?
To what extent is the proposed research high risk/high gain?

Scientific Approach.
To what extent is the outlined scientific approach feasible bearing in mind the extent that the proposed research is high risk/high gain?
To what extent is the proposed research methodology appropriate to achieve the goals of the project?
To what extent does the proposal involve the development of novel methodology?
To what extent are the proposed timescales and resources necessary and properly justified?

Criterion 2 - PRINCIPAL INVESTIGATOR
Intellectual capacity, creativity and commitment
The questions below can have one of the following four responses: Outstanding/Excellent/Very good/Non-competitive
To what extent has the PI demonstrated the ability to propose and conduct ground-breaking research?
To what extent does the PI provide evidence of creative independent thinking?
To what extent have the achievements of the PI typically gone beyond the state of the art?
To what extent does the PI demonstrate the level of commitment to the project necessary for its execution and the willingness to devote a significant amount of time to the project (min 50% of the total working time on it and min 50% in an EU Member State or Associated Country)?

PANEL SCORE AND RANKING RANGE

Final panel score :	A (fully meets the ERC's excellence criterion and is recommended for funding if sufficient funds are available)	Ranking range*: 44%-46%

* Ranking range of your proposal out of the proposals evaluated by the panel in Step 2, in percent, from 1% for the highest ranked proposals to 100% for the lowest ranked.

high risk / high gain

100% izgalmas kérdések

100% ambiciózus célok

50-80% megvalósíthatóság

The panel was impressed by the innovative work of the PI in developing new phylogenetic methods based on the idea of using horizontal transfer events as information, rather than as a problem.

The proposal, which centred on this idea and methodological breakthrough, was well received. In particular, it was clear to the panel that the application of these methods will offer a new way to evaluate the timing of ancient evolutionary events. It was also clear that this method has the potential to improve the estimates of the timings of evolutionary events. Some concerns were raised, however, as to whether the method could be usefully applied across such a wide timescales, and across so many major evolutionary events, as suggested in the proposal. Some reflection on these issues would be helpful

The panel therefore recommends the proposal to be retained for funding with a grant not exceeding 1 453 859.00 Euro, if additional budget becomes available.

ssolo@elte.hu