Молекулярная филогенетика и филогеномика

Полина Дроздова

9–10 октября 2023

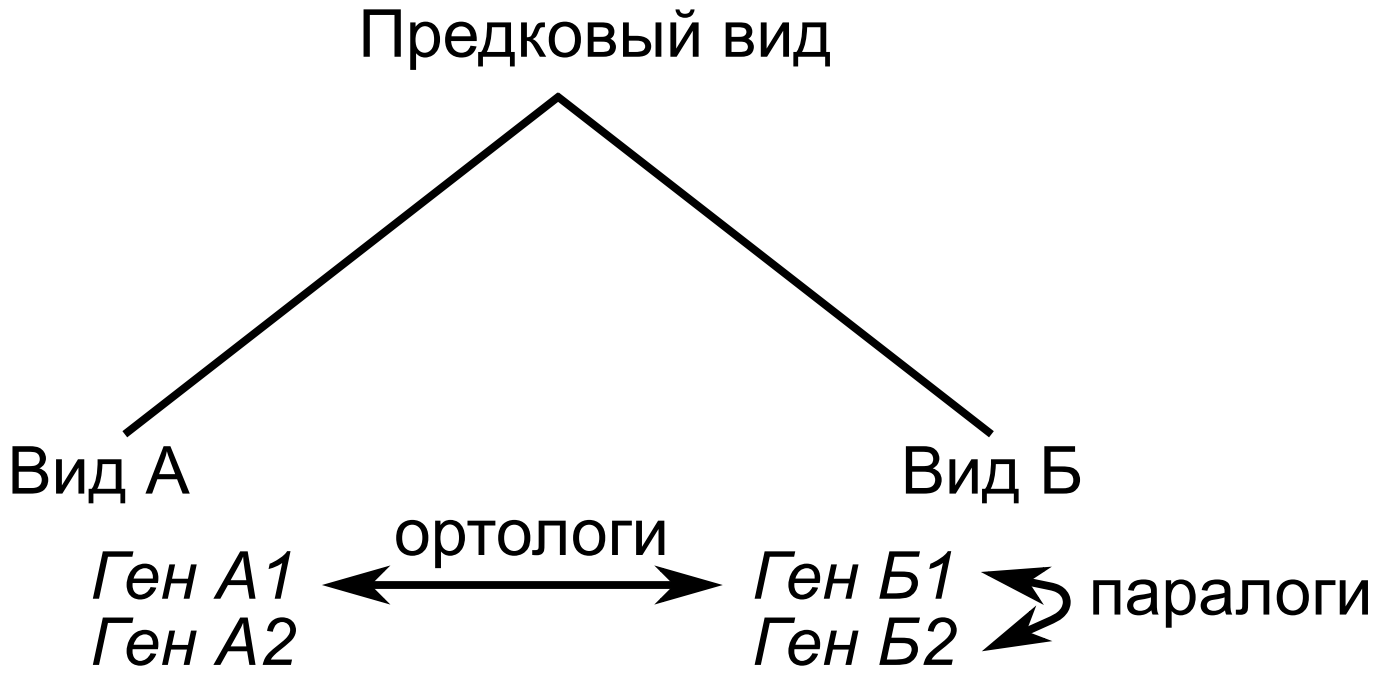
https://github.com/drozdovapb/ZIN_school_2023

Терминология

Гомология (homology) — сходство признаков, обусловленное общим происхождением.
Дивергенция (divergence) — расхождение признаков в ходе эволюции.
Гомоплазия (homoplasy) — сходство признаков, *не* обусловленное общим происхождением. Часто обусловлена **параллельной (конвергентной)** эволюцией.

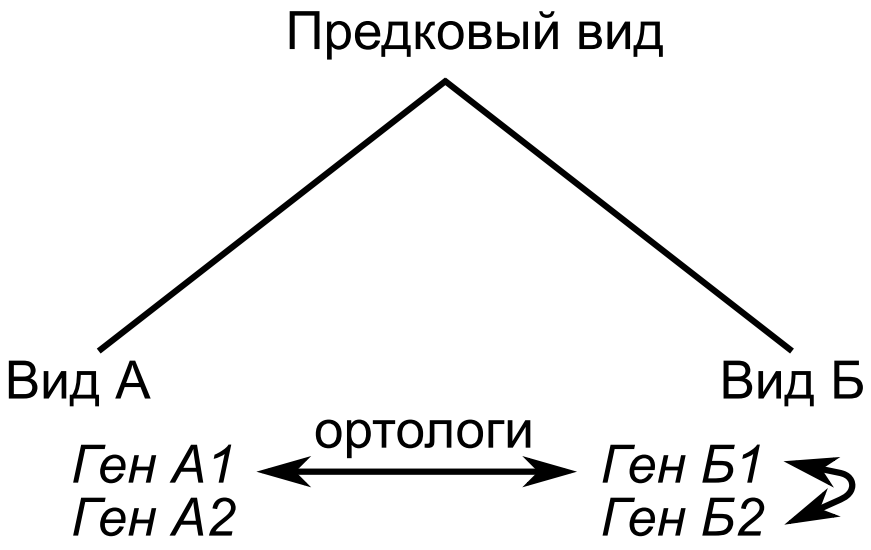
Гомологи = ортологи (orthologs/orthologues) + паралоги (paralogs/paralogues).

Варианты гомологии



Термины paralogous / orthologous:
Fitch WM. Distinguishing homologous from analogous proteins.
Systematic Biology. 1970 Jun 1;19(2):99-113.

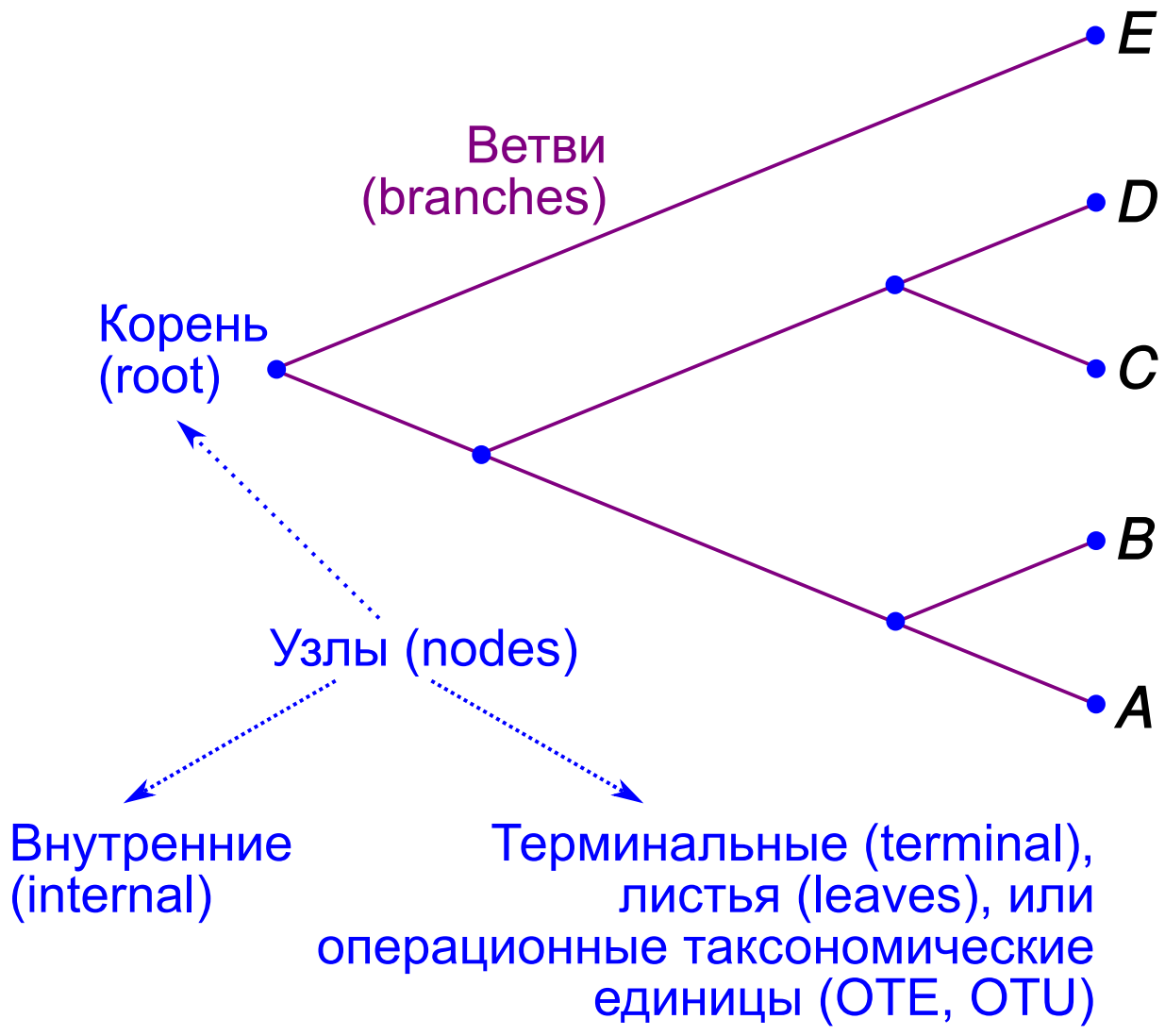
Варианты гомологии



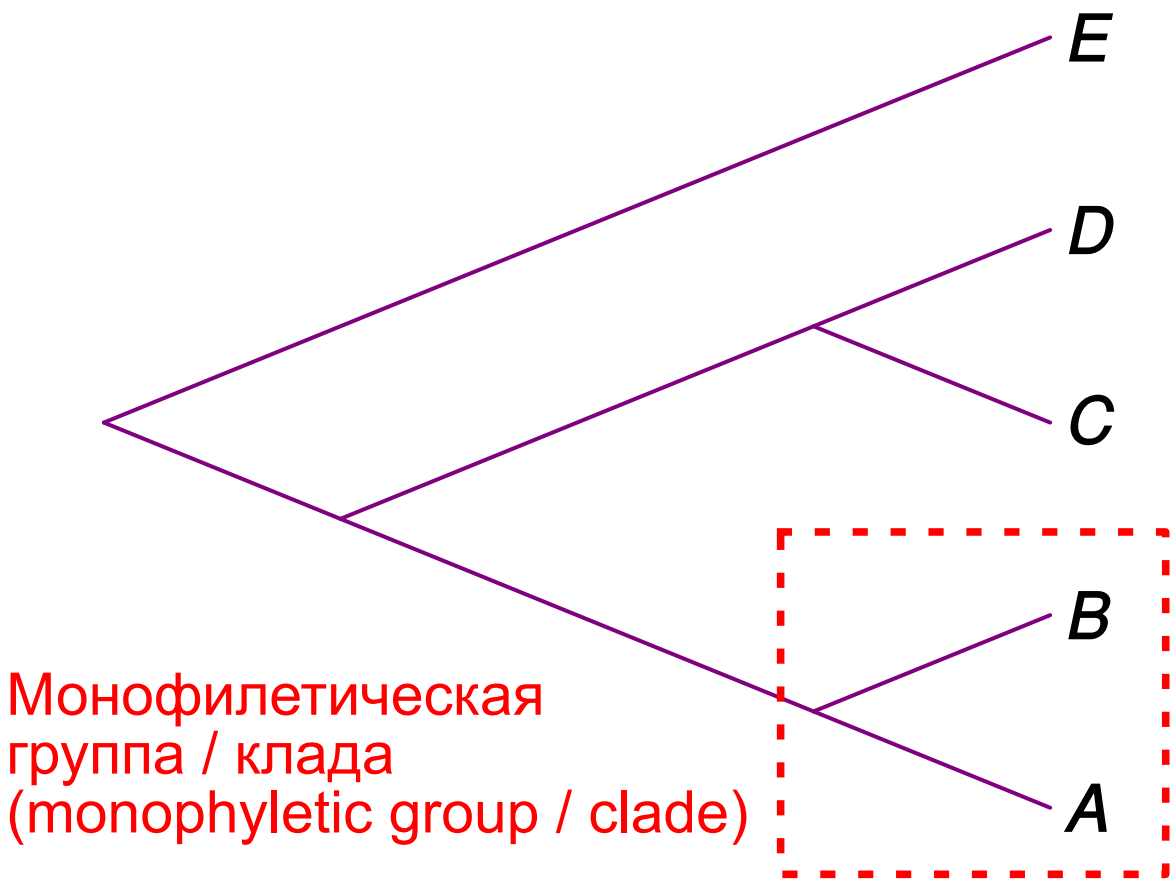
Inparalogs (synparalogs) /
Outparalogs (alloparalogs)
Co-orthologs
Pseudoothologs
Xenologs
Ohnologs
Pseudoparalogs

см. Koonin EV. Orthologs, paralogs, and evolutionary genomics 1. Annu. Rev. Genet. 2005;39:309-38.

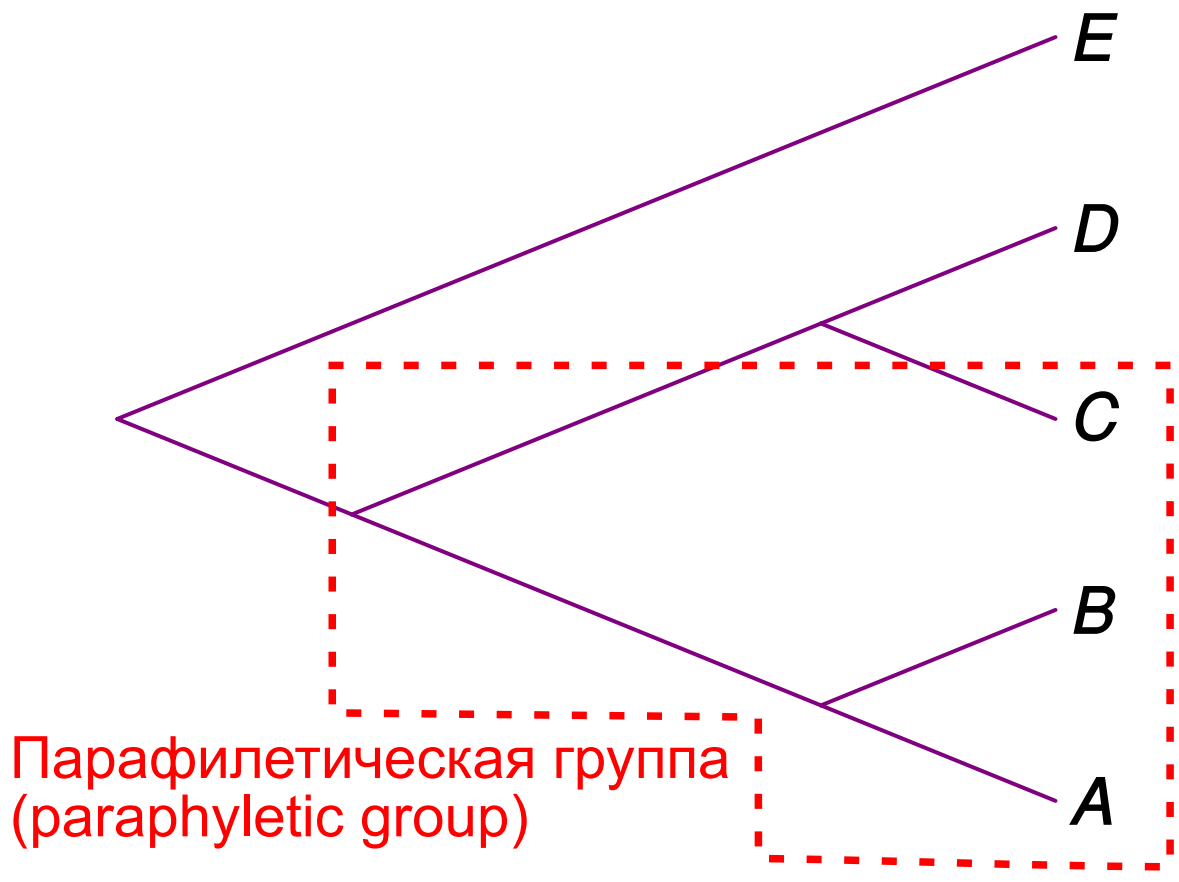
Названия частей дерева



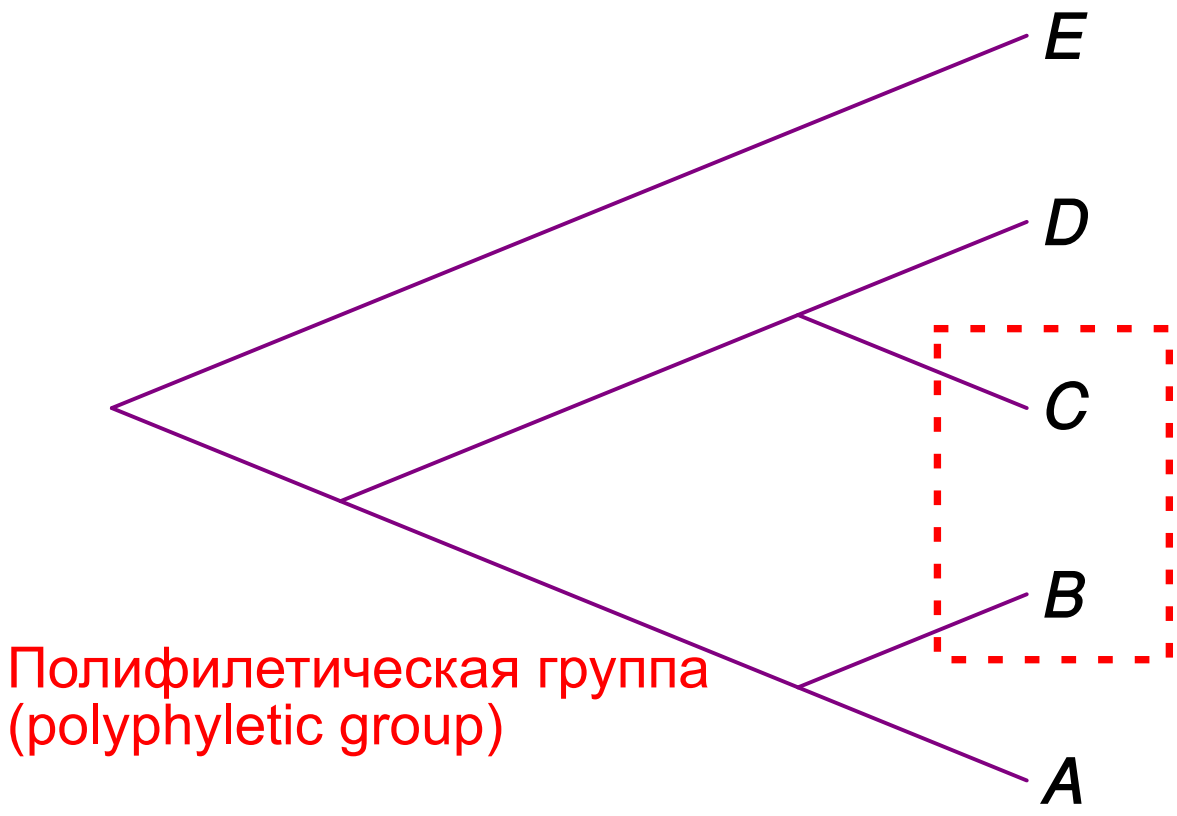
Монофилия, полифилия и парафилия



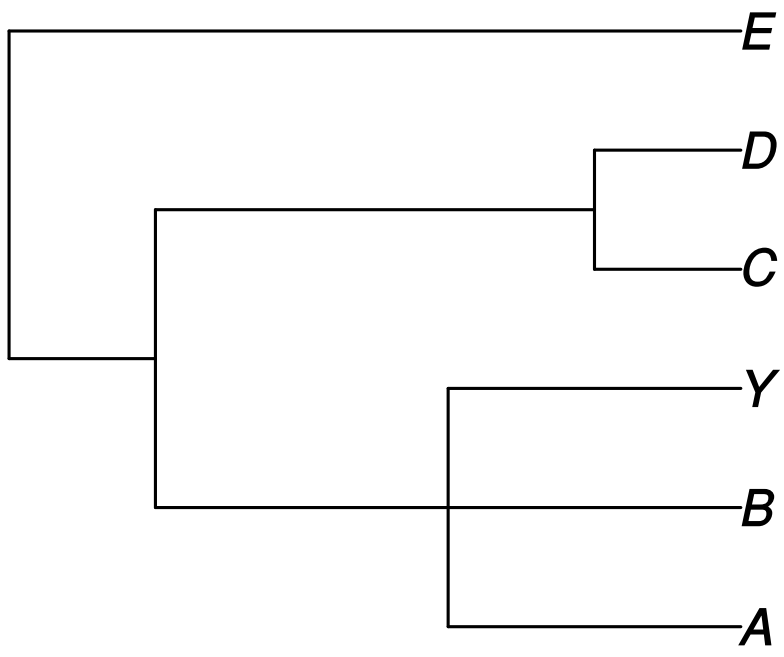
Монофилия, полифилия и парафилия



Монофилия, полифилия и парафилия



Политомия



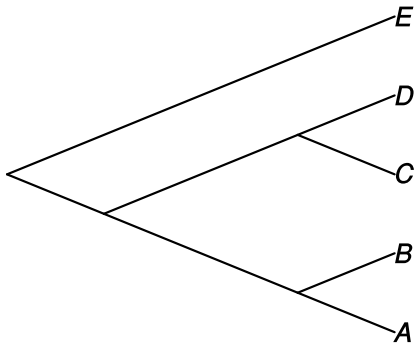
Soft polytomy (из-за недостатка данных)

vs.

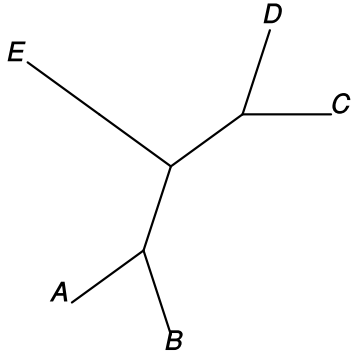
Hard polytomy (очень быстрое видообразование)

Типы дендрограмм

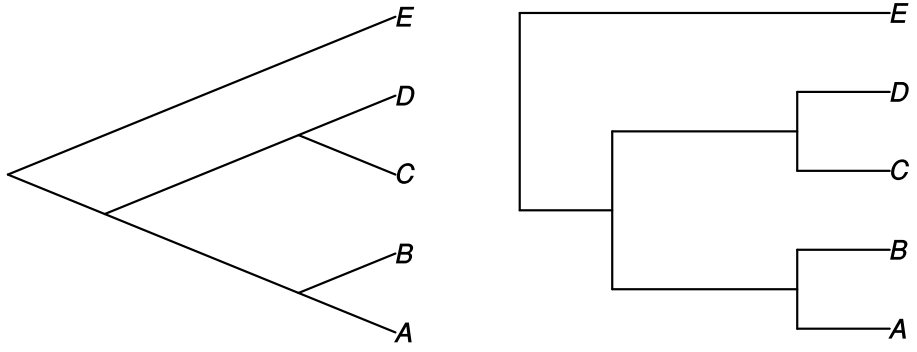
Укоренённое (rooted) дерево



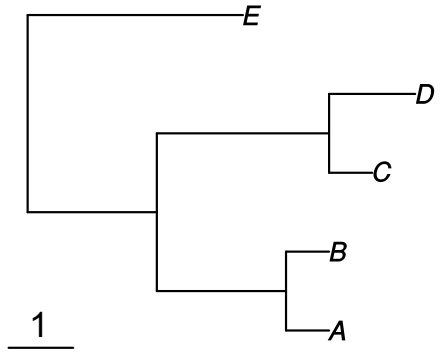
Неукоренённое (unrooted)



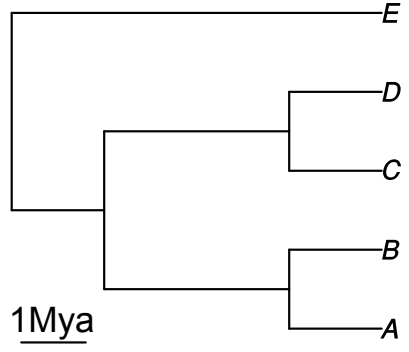
Кладограммы (длина ветвей не означает ничего)



Филограммы (длина ветвей что-то означает)



Аддитивное (additive):
длина ветви — число замен

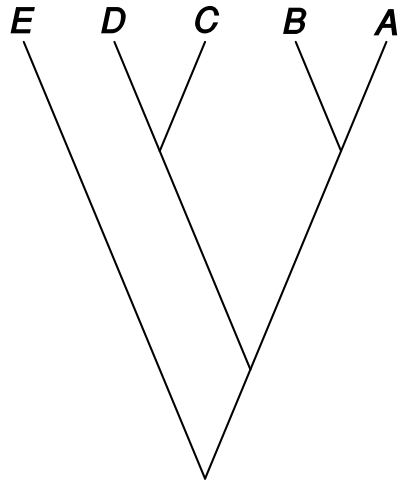


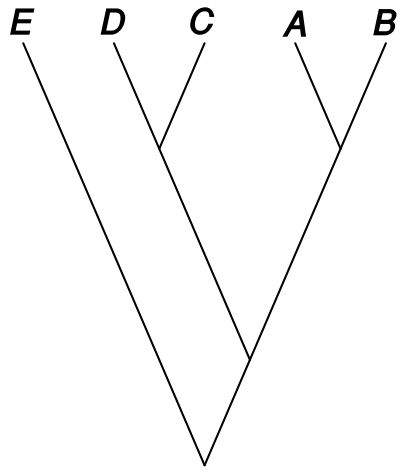
Ультраметрическое (ultrametric):
равная длина ветвей
(напр., длина ветви — время)

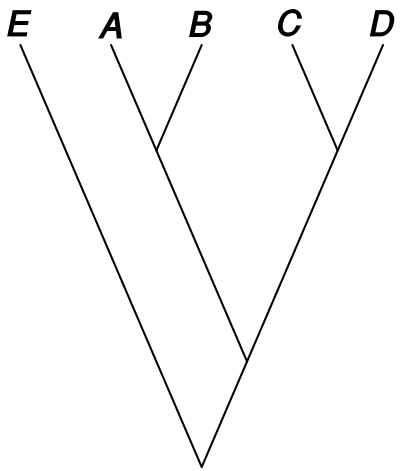
Число возможных топологий для дерева

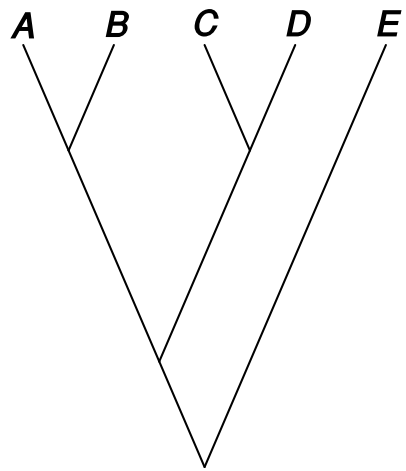
Число OTE	Возможных неукоренённых деревьев	Возможных укоренённых деревьев
3	1	3
4	3	15
...	...	...
22	0,5 моля	
...	...	...
n	$(2n-5)!!$	$(2n-3)!!$
n	$\frac{(2n-5)!}{2^{n-3}\cdot(n-3)!}$	$\frac{(2n-3)!}{2^{n-2}\cdot(n-2)!}$

Разные варианты отображения топологии







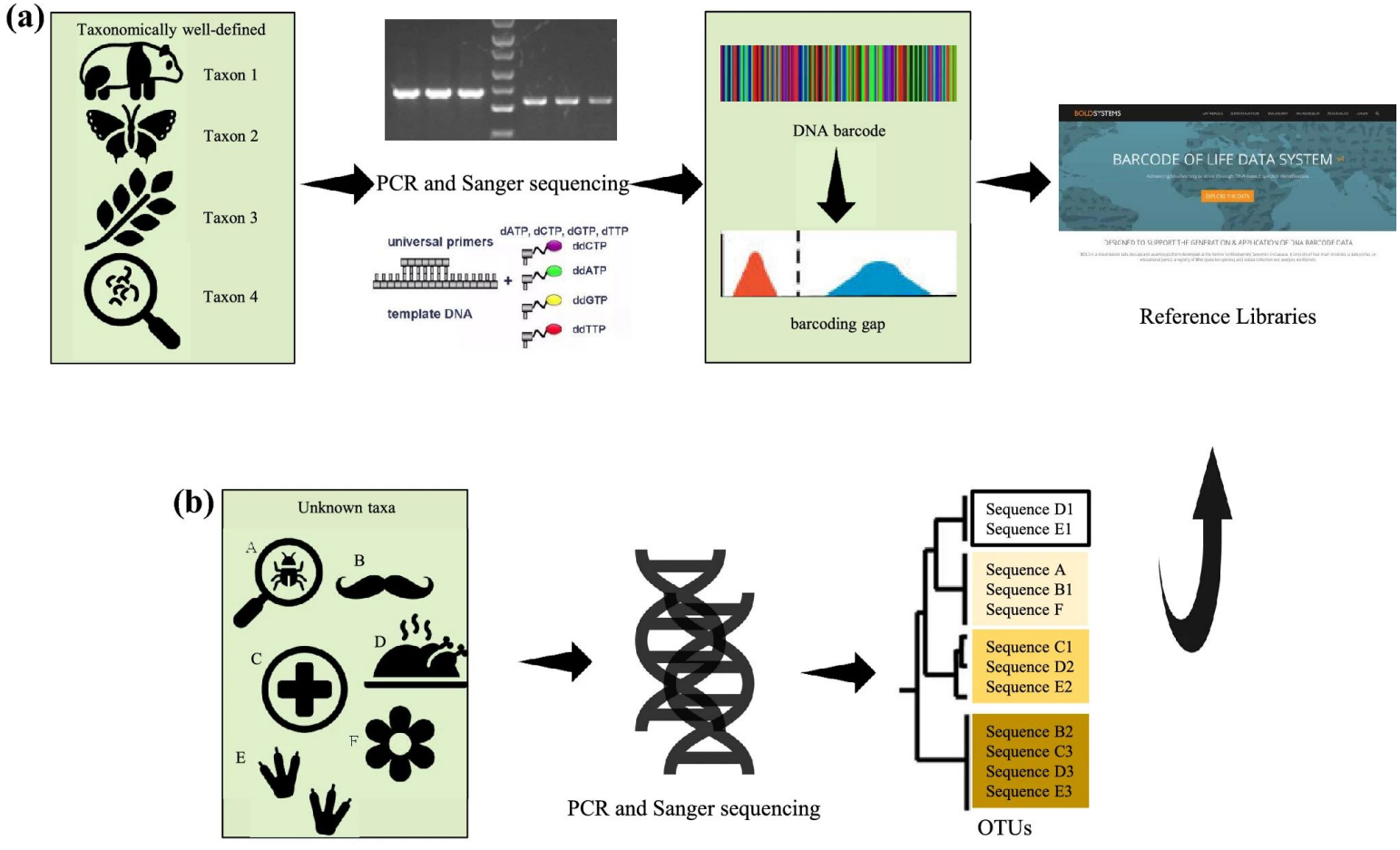


Это одно и то же дерево!

Почему молекулярные данные?

- ✓ Биополимеры есть у всех живых организмов и состоят из одинаковых элементов.
- ✓ Общая методология для разных организмов => проще обучить специалистов;
- ✓ Иногда последовательность ДНК — это единственное, что мы знаем об организме.
- ✓ Нужно мало материала / подходит повреждённый материал.
- ✓ Можно комбинировать с морфологическими данными.

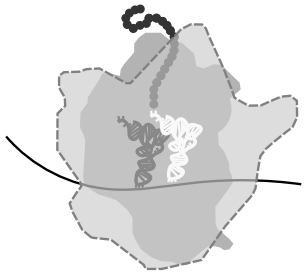
Баркодинг



https://link.springer.com/article/10.1007/s12686-022-01291-2/

Необходимые свойства маркера:

1. **универсальность** => наличие ортологов, желательно без проблем с паралогами;



(Например, компоненты рибосомы)

2. подходящий уровень **разнообразия** (например, межвидовые различия больше внутривидовых), но при этом редкие множественные замены;

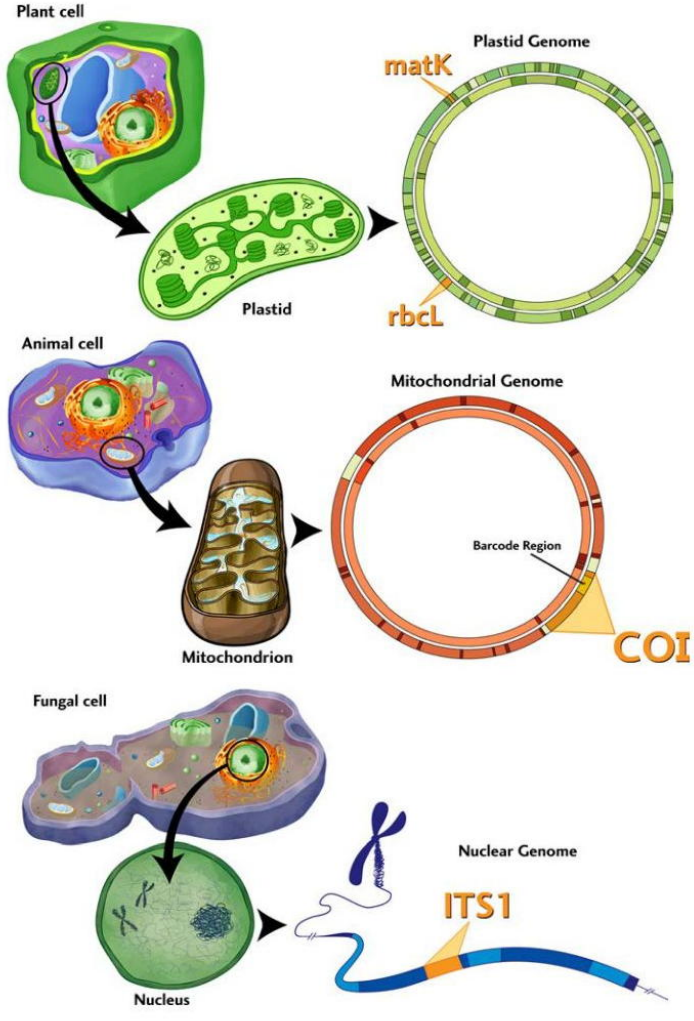
EPIC = exon primed intron crossing

3. желательна селективная нейтральность.

Примеры часто используемых маркеров

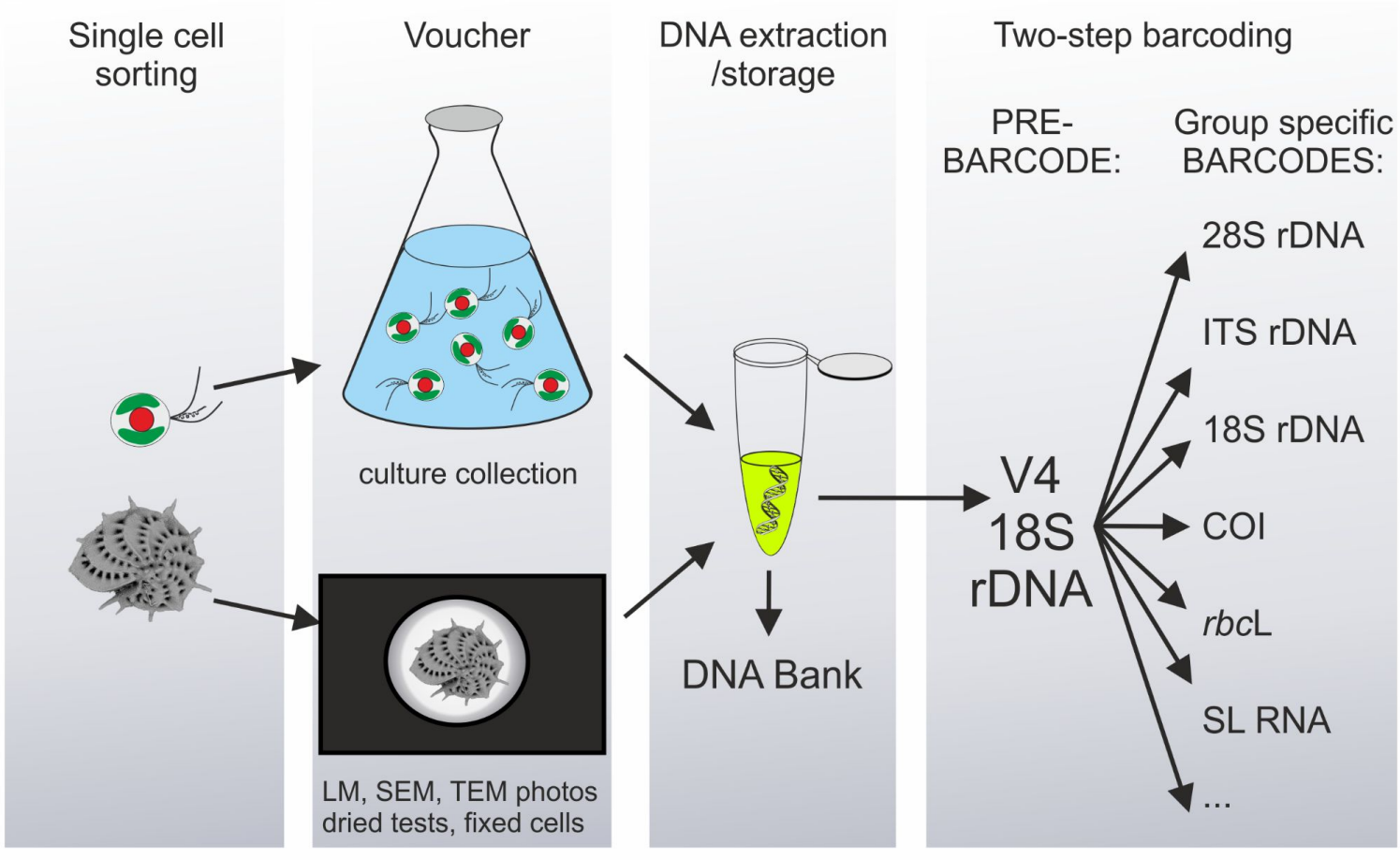
Table: key of tree of life

Color	Clade	Primary barcode	Secondary barcode
	Animals	COI	COI, 16S
	Fungi	ITS	LSU D1/D2
	Green algae	tufA	LSU D2/D3
	Land plants	rbcL/matK	psbA-trnH/ITS
	Algae	COI-5P	LSU D2/D3



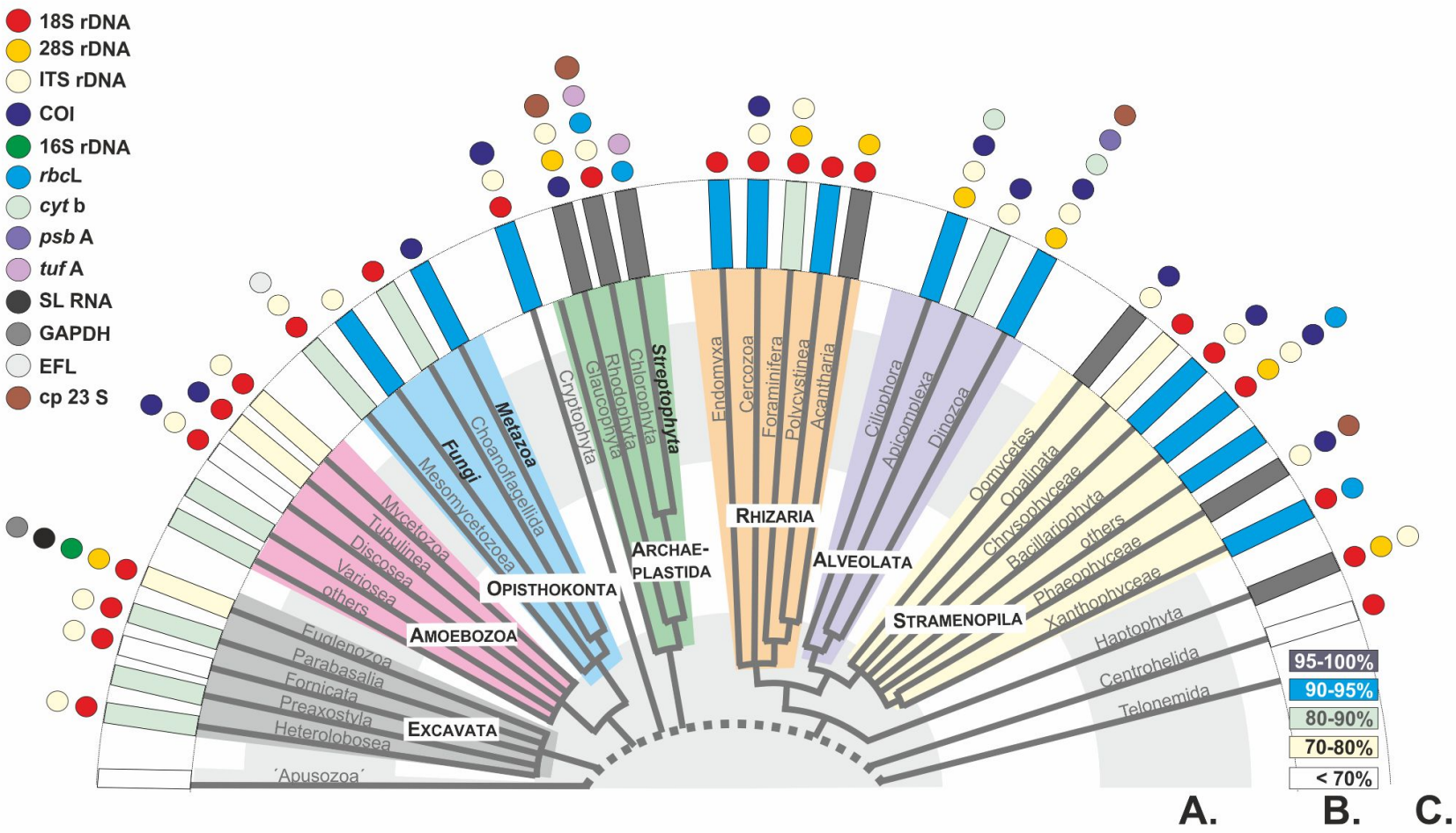
https://doi.org/10.1155/2022/1846485
https://www.cbd.int/doc/publications/cbd-ts-94-en.pdf

Примеры часто используемых маркеров: протисты



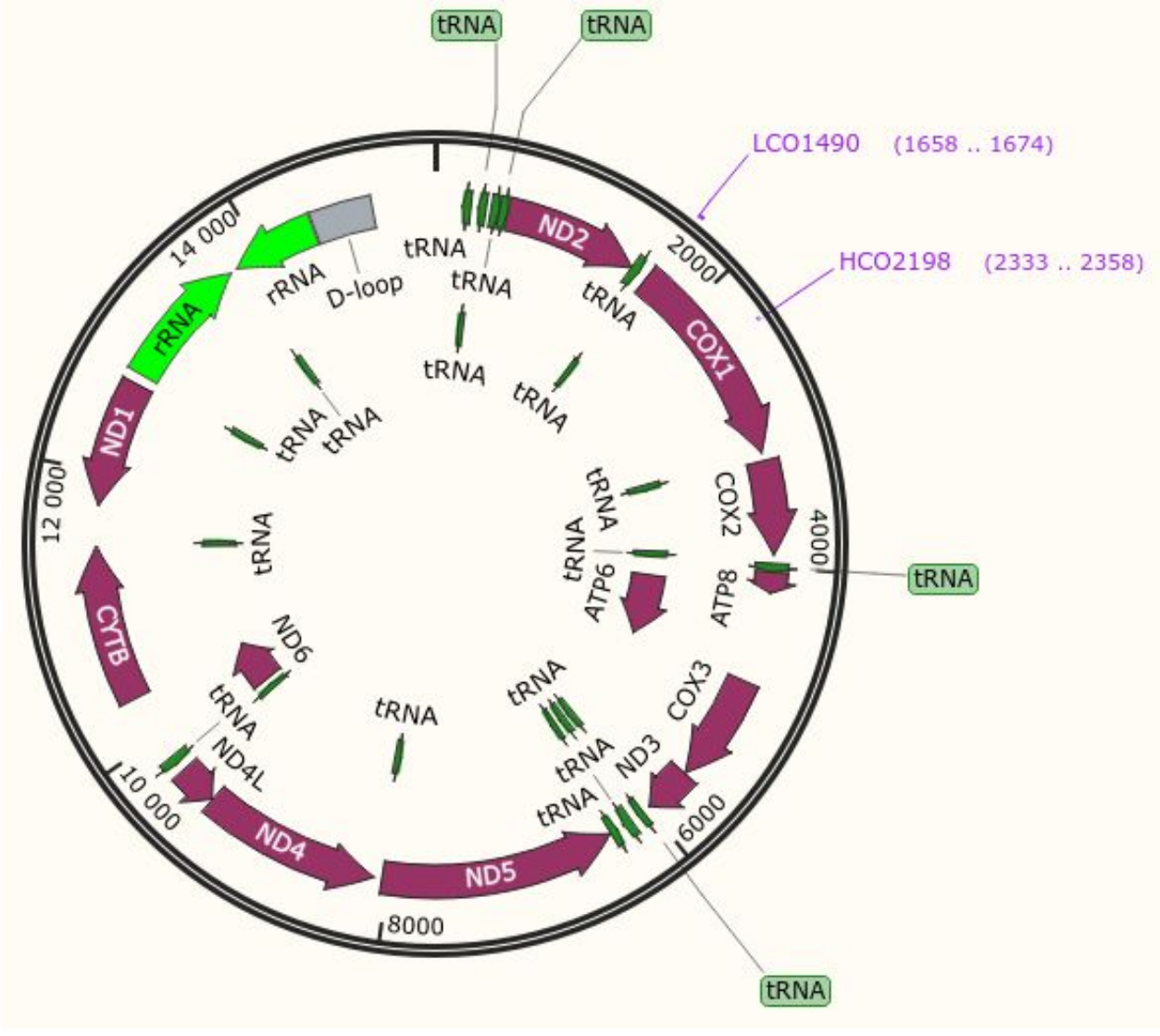
https://doi.org/10.1371/journal.pbio.1001419

Примеры часто используемых маркеров: протисты



https://doi.org/10.1371/journal.pbio.1001419

Примеры часто используемых маркеров: животные



Структура локуса рДНК

Bacteria / Archaea:

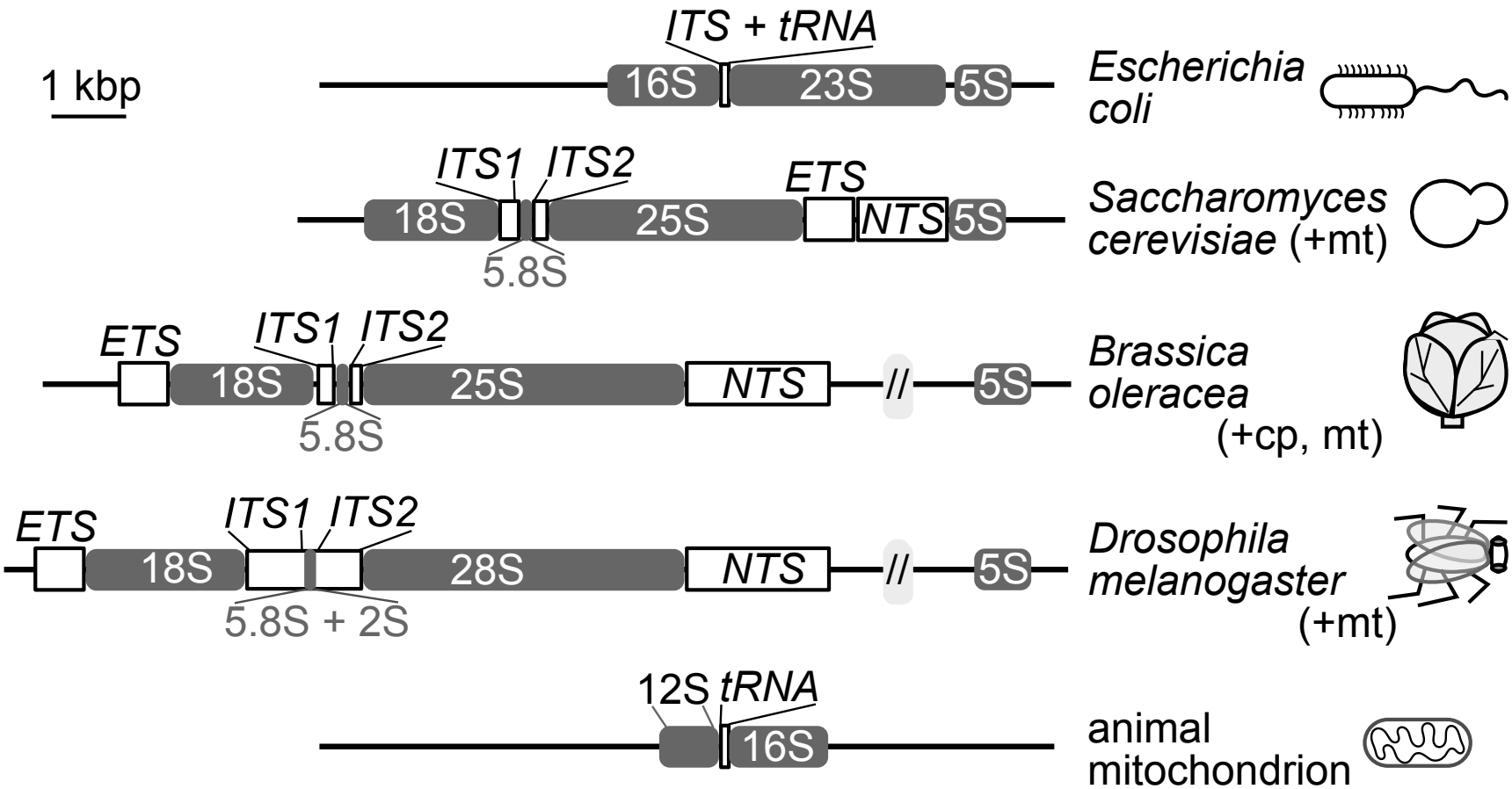
Eukaryota:

ITS = internal transcribed spacer

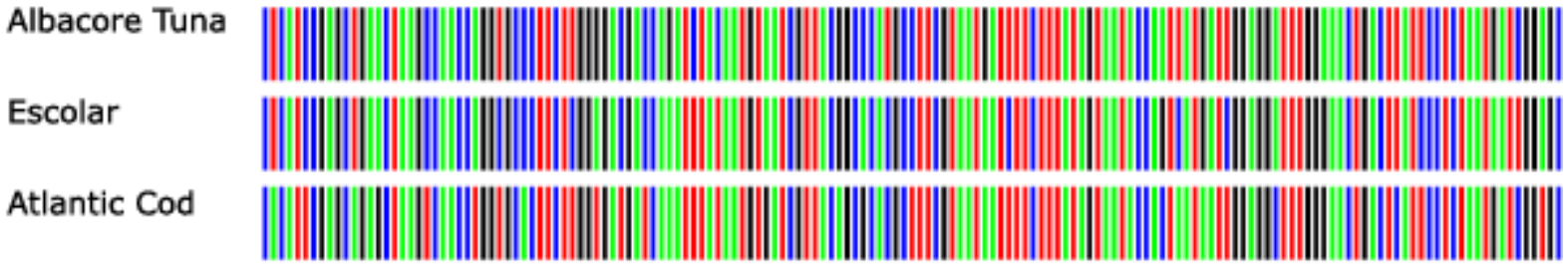
ETS = external transcribed spacer

IGS / NTS = intergenic / non-transcribed spacer

Структура локуса рДНК у разных организмов

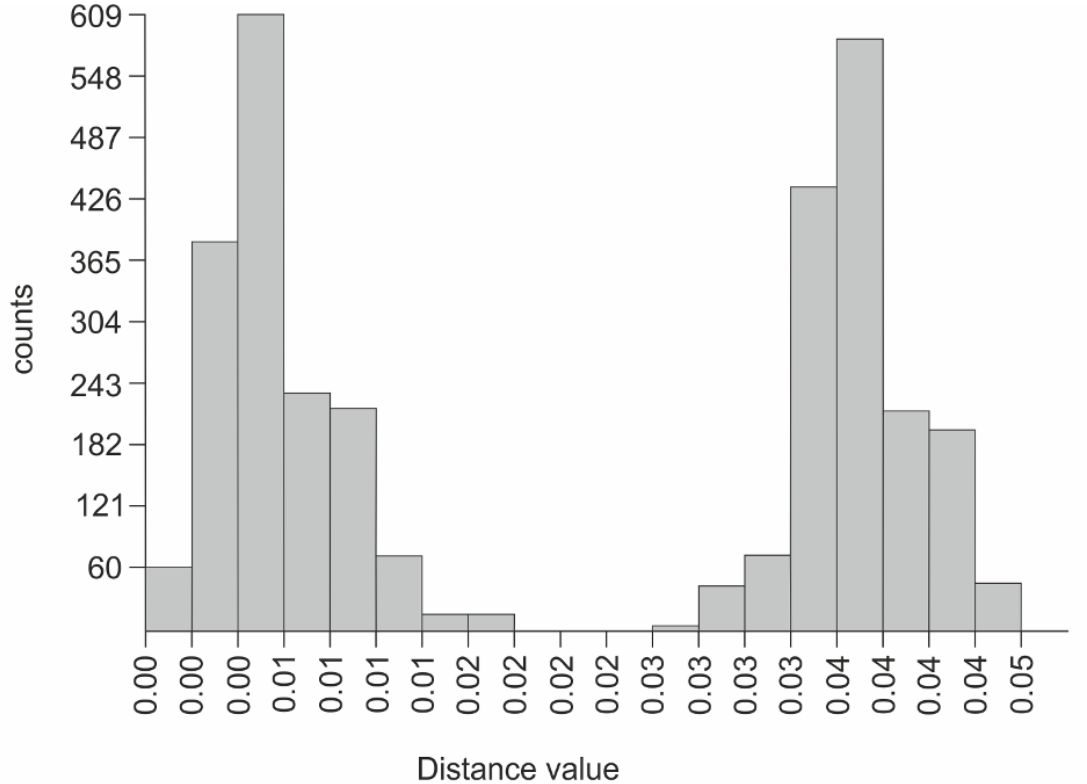


Штрихкодирование ДНК (DNA barcoding)



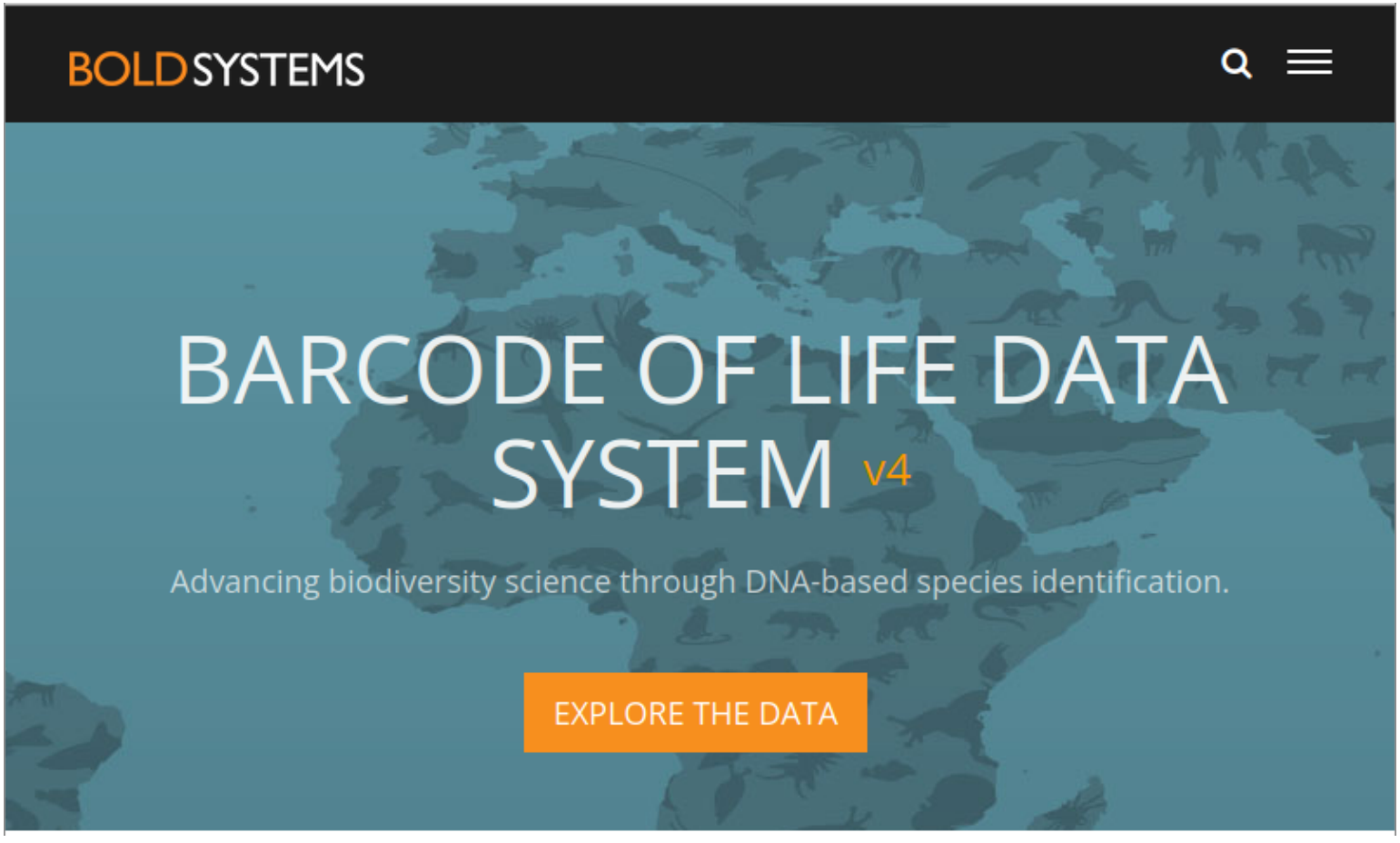
http://www.ibolproject.org/research.php

Barcode gap — разрыв между внутривидовой и межвидовой изменчивостью.



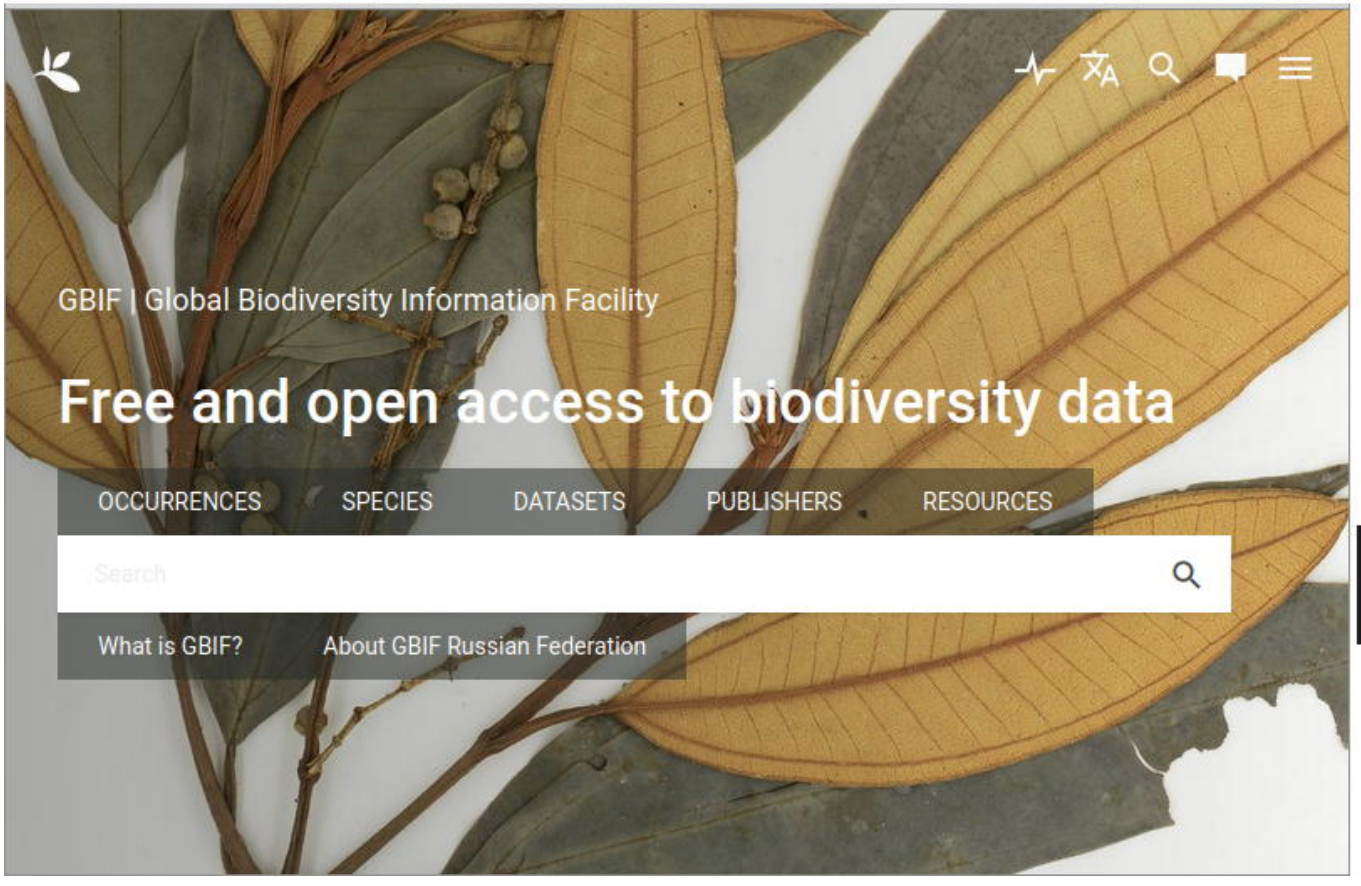
https://peerj.com/articles/2128/

Проект Barcode of Life



https://boldsystems.org/

База GBIF

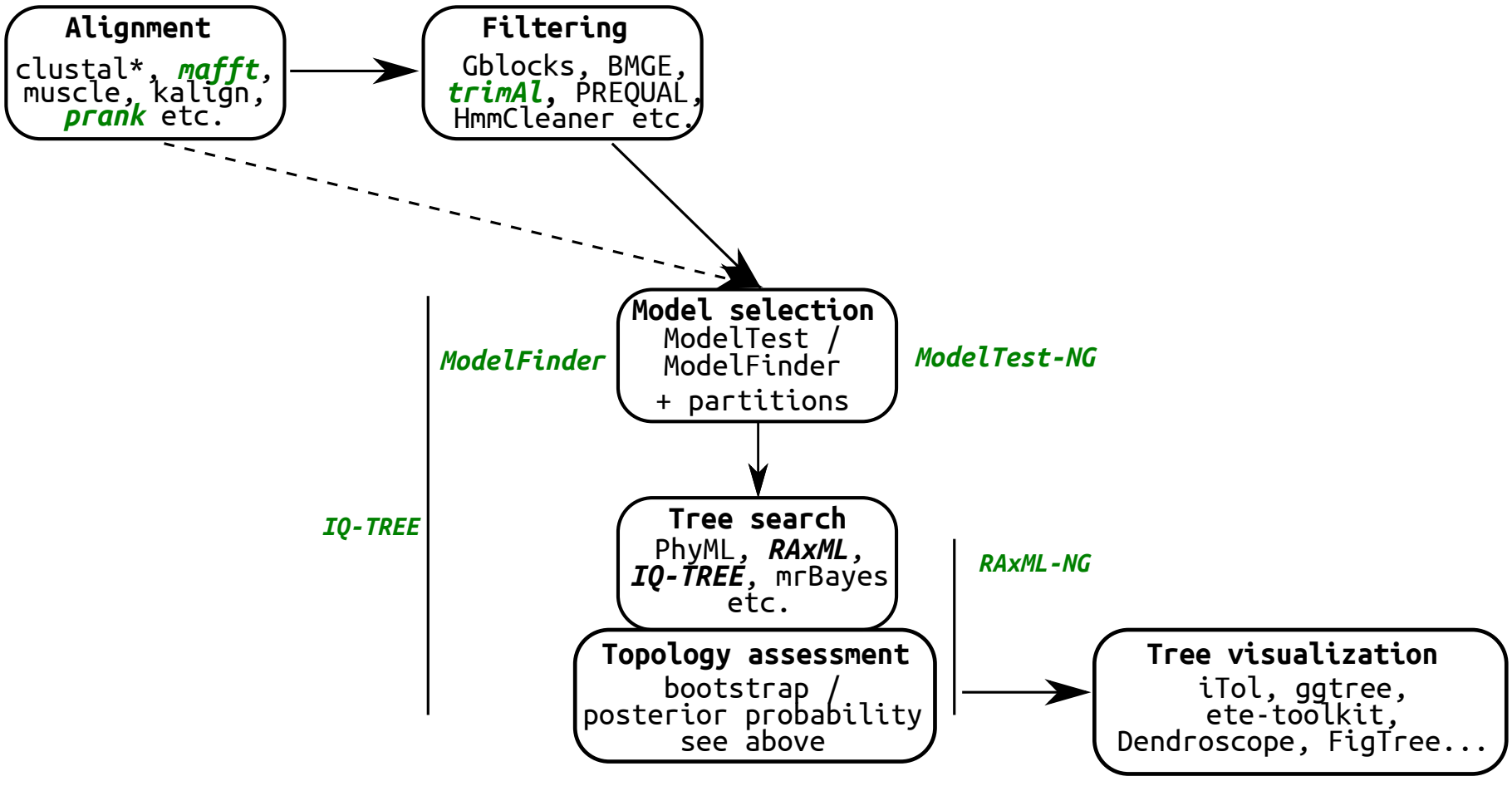


iNaturalist

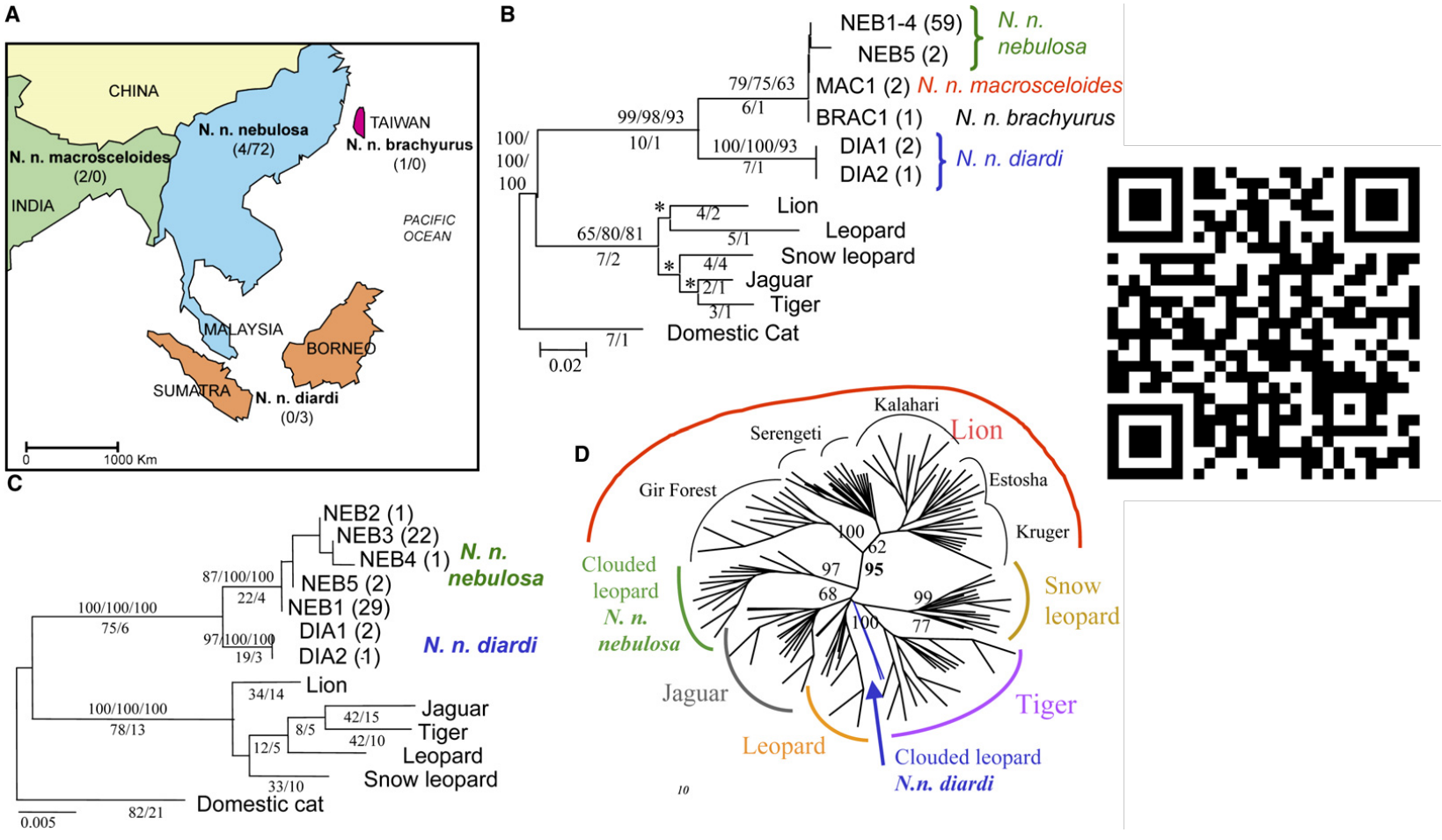
BOLD SYSTEMS

https://www.gbif.org/

Общая схема филогенетического анализа



Упражнение №1: восстановление филогении по фрагменту гена (*cytb*)



https://doi.org/10.1016/j.cub.2006.08.066

Упражнения: филогения 12 видов млекопитающих

- ✓ *Physeter catodon*
- ✓ *Ovis aries*
- ✓ *Equus caballus*
- ✓ *Felis catus*
- ✓ *Canis lupus familiaris*
- ✓ *Myotis lucifugus*
- ✓ *Mus spretus*
- ✓ *Mus musculus*
- ✓ *Marmota marmota*
- ✓ *Macaca mulatta*
- ✓ *Ornithorhynchus anatinus*
- ✓ *Vombatus ursinus*

Упражнения: филогения 12 видов млекопитающих

Упражнение №2: филогения 12 видов млекопитающих по одному гену (*cytb*)

Филогенетика и филогеномика

Филогенетика: один или несколько генов.

Филогеномика: много генов (полный геном / транскриптом или его существенная часть)

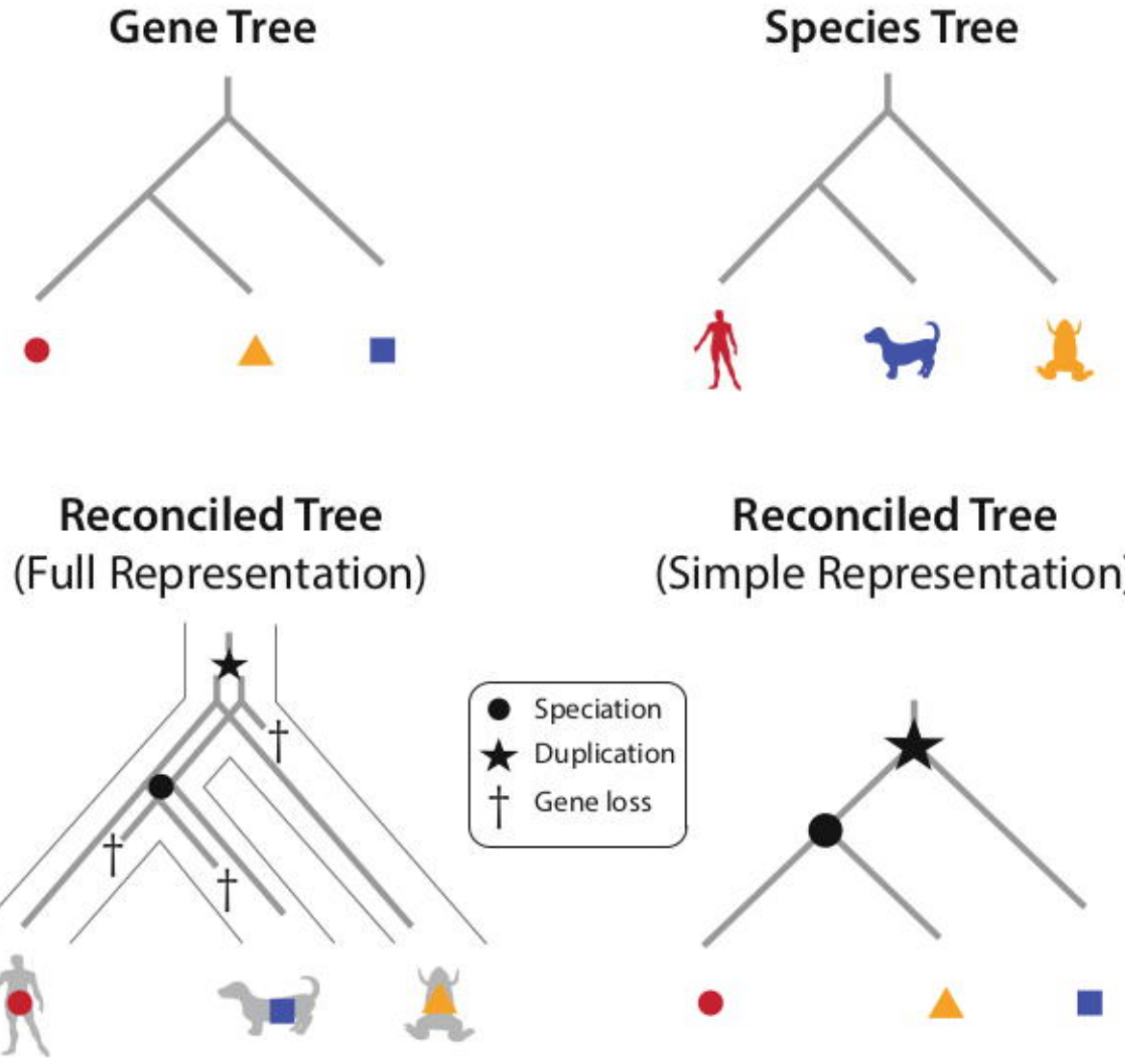
Филогенетика и филогеномика

Филогенетика: один или несколько генов.
Филогеномика: много генов (полный геном / транскриптом или его существенная часть)

Зачем использовать много генов?

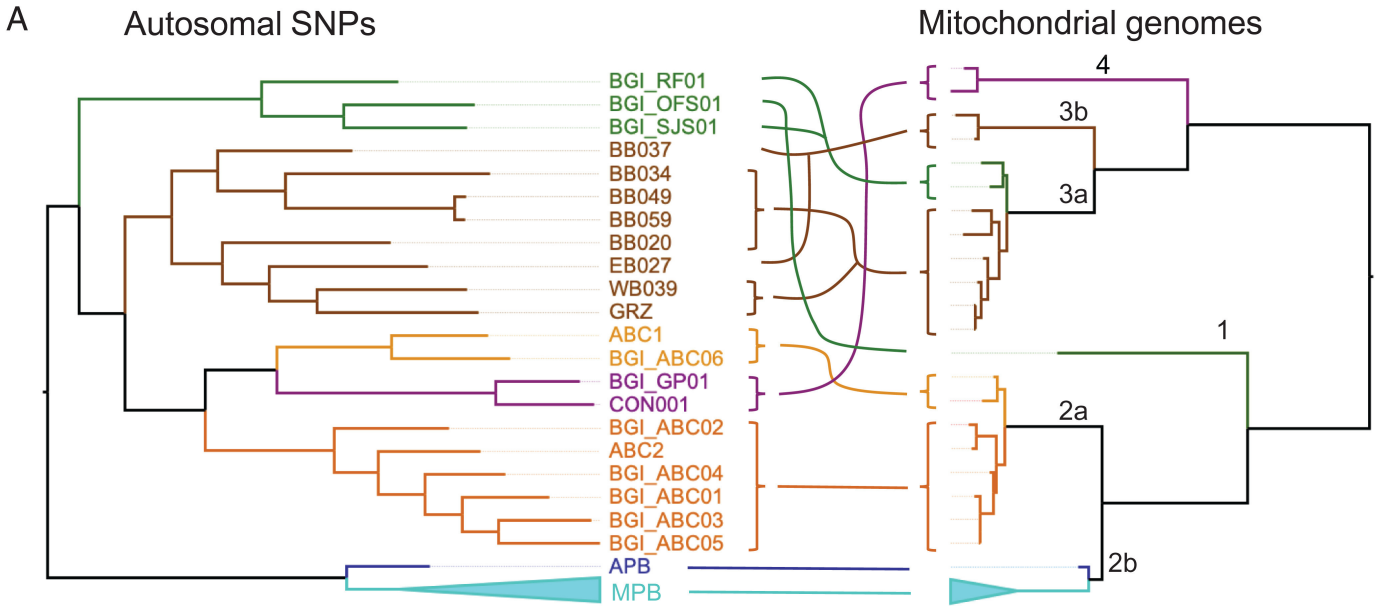
- ✓ Иногда дерево генов не соответствует дереву видов...

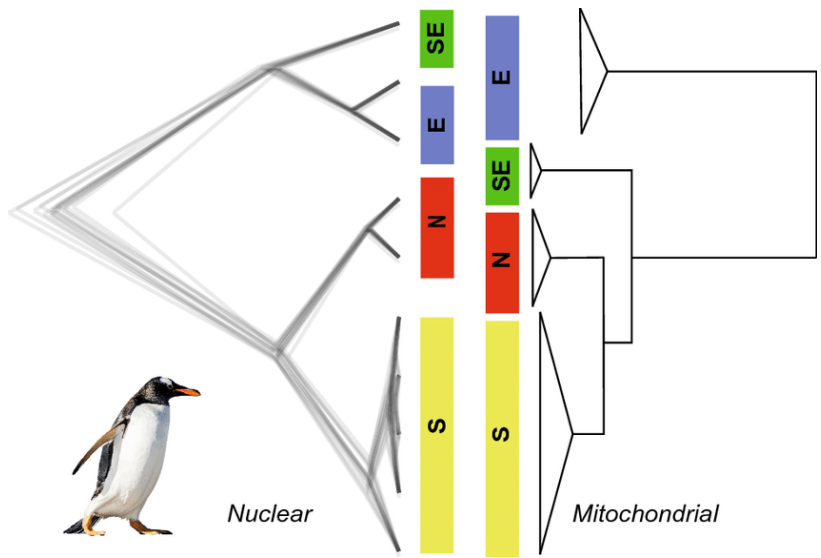
Потеря гомологов



https://www.researchgate.net/publication/334265032_Inferring_Orthology_and_Paralogy

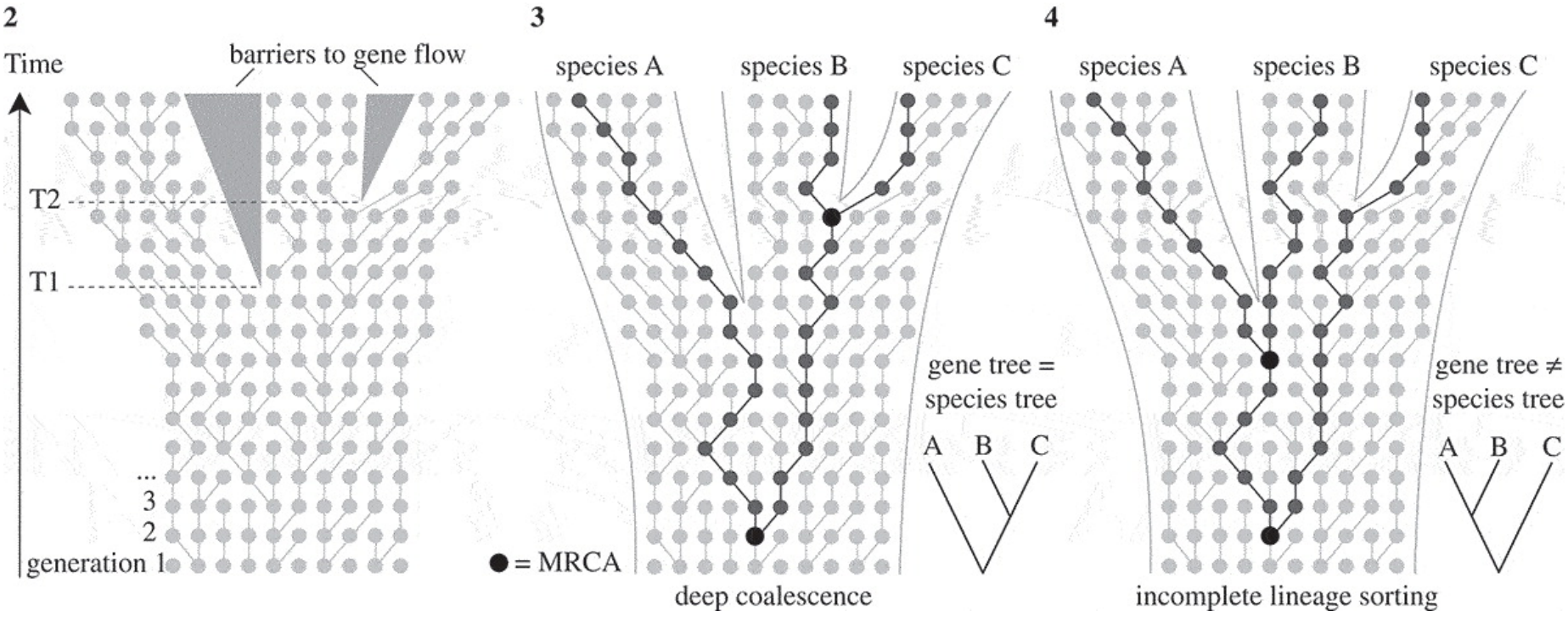
Mito-nuclear discordance





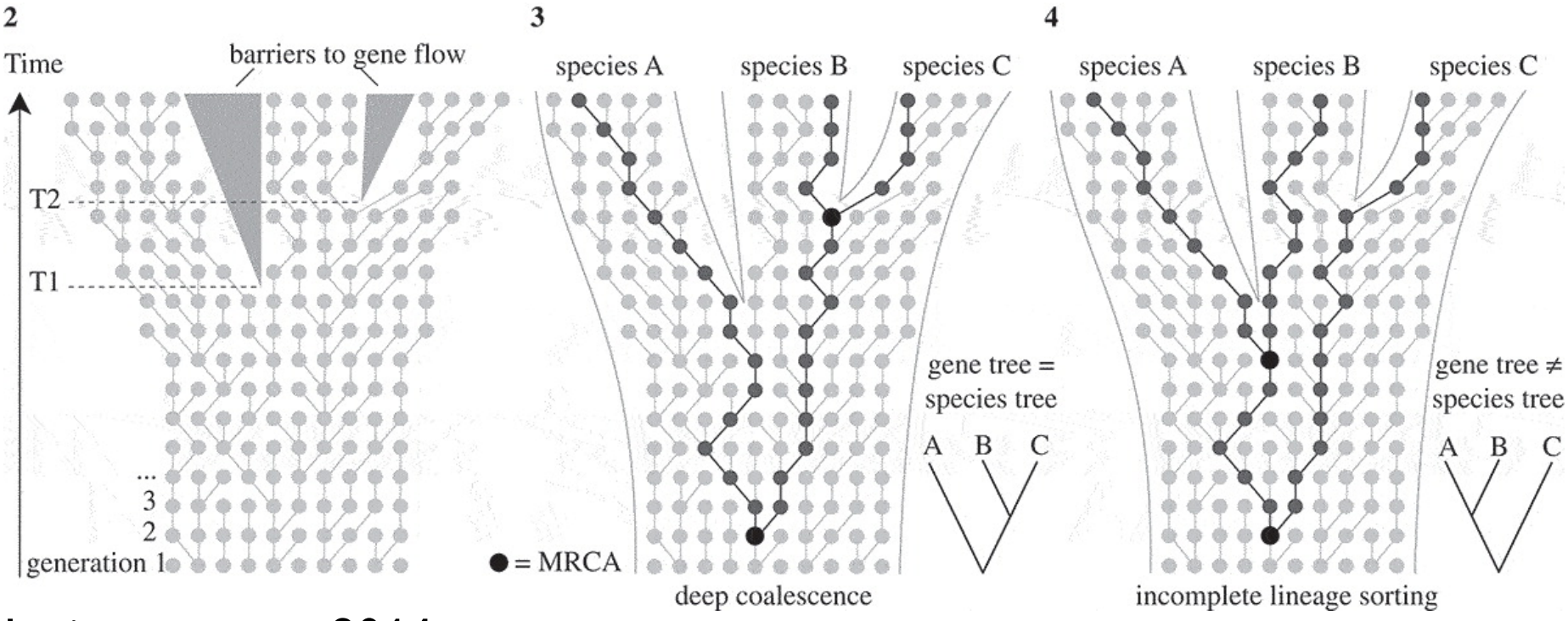
https://doi.org/10.1073/pnas.2200016119
http://dx.doi.org/10.1038/s41598-022-07562-0

Incomplete lineage sorting



Leliaert et al., 2014,
https://dx.doi.org/10.1080/09670262.2014.904524

Incomplete lineage sorting

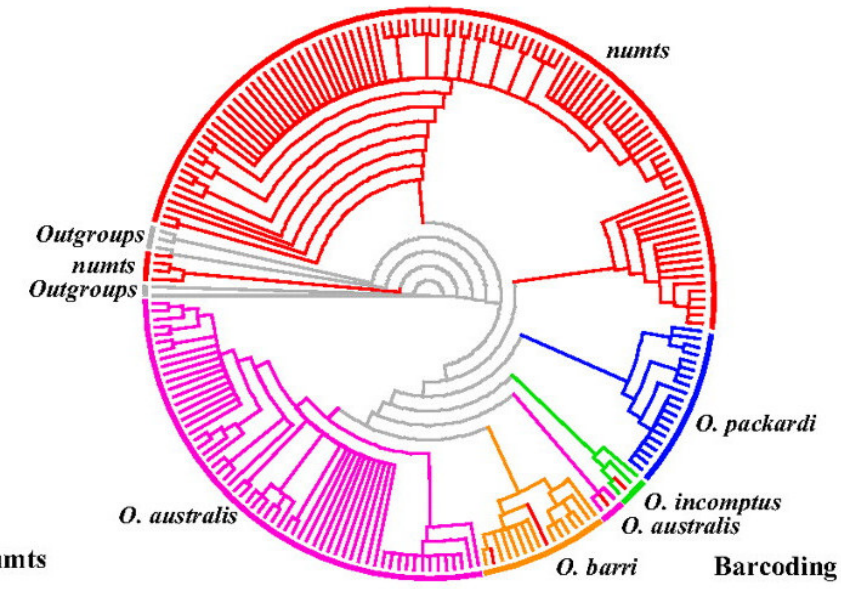


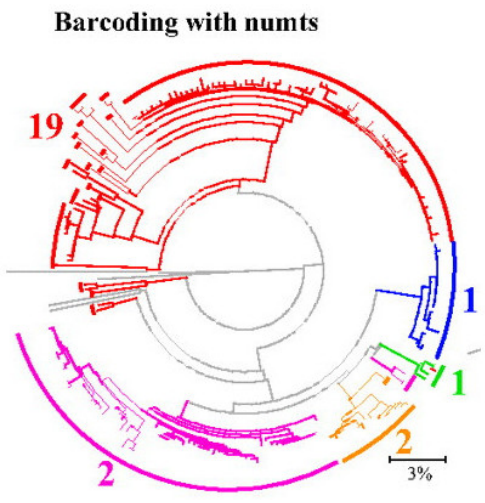
Leliaert et al., 2014,
https://dx.doi.org/10.1080/09670262.2014.904524

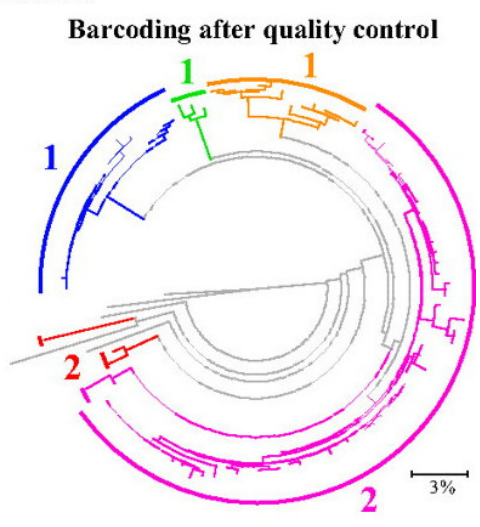
Mito-nuclear discordance часто именно ILS

NUMTs = nuclear mitochondrial DNA segment

[new might]







https://doi.org/10.1073/pnas.0803076105

Филогенетика и филогеномика

Филогенетика: один или несколько генов.

Филогеномика: много генов (полный геном / транскриптом или его существенная часть)

Филогенетика и филогеномика

Филогенетика: один или несколько генов.
Филогеномика: много генов (полный геном / транскриптом или его существенная часть)

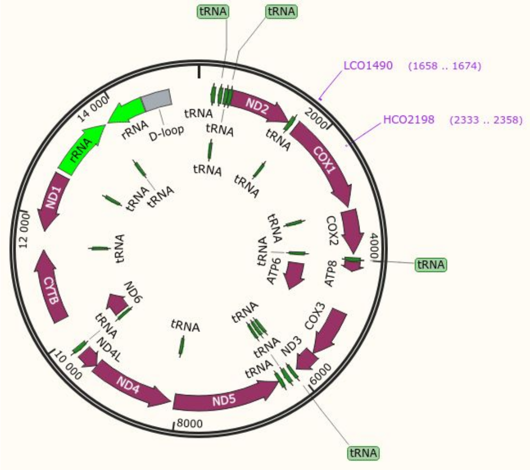
Зачем использовать много генов?

- ✓ Иногда дерево генов не соответствует дереву видов...
- ✓ Лучшее разрешение в близких таксонах.
- ✓ Лучшее разрешение в макротаксономии
- ✓ (Эти данные у нас уже есть...)

Упражнения: филогения 12 видов млекопитающих

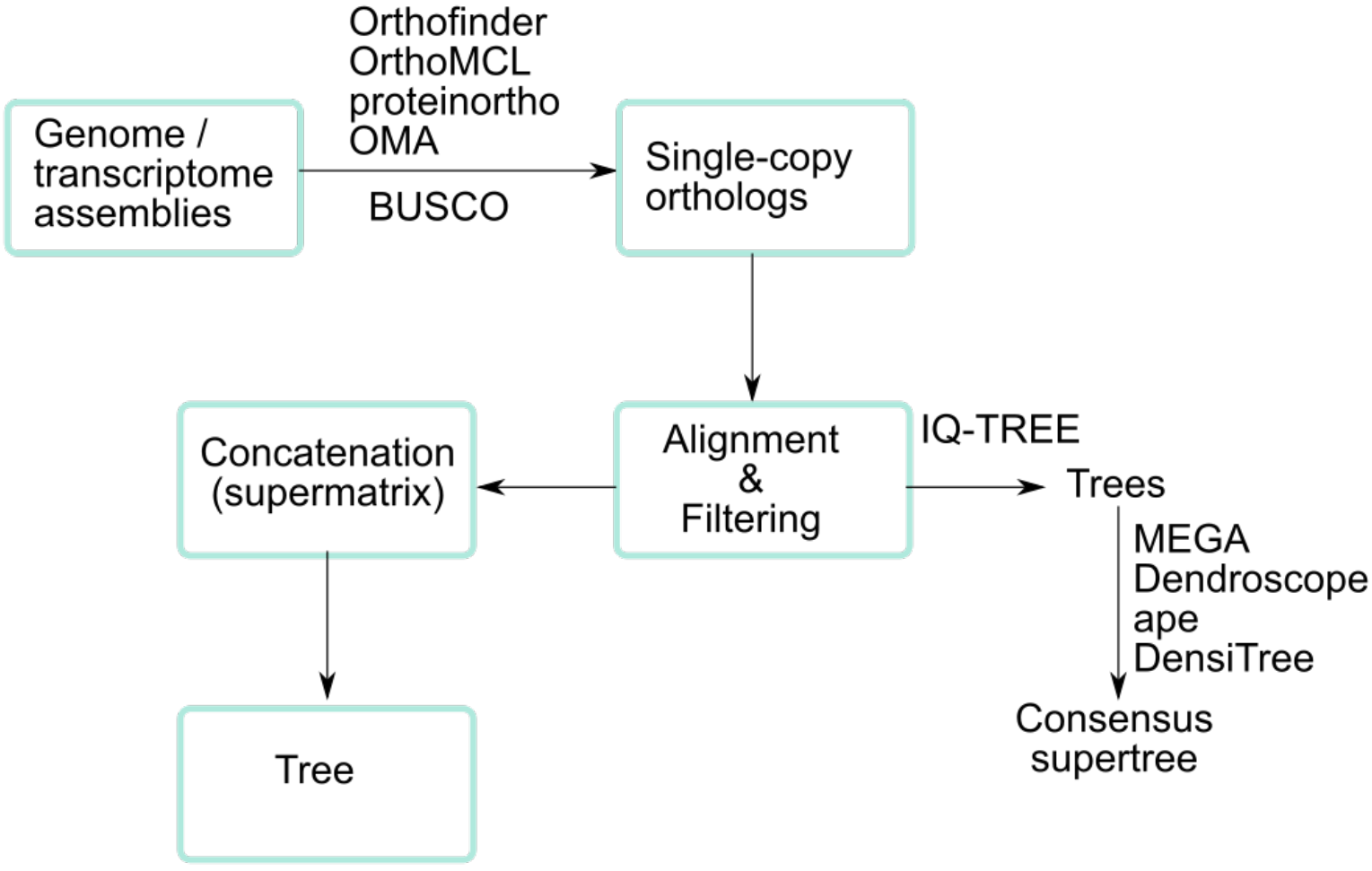
Упражнение №2: восстановление филогении по 13 митохондриальным белкам

Митохондриальный геном животных



1. *ATP6*
2. *ATP8*
3. *COX1*
4. *COX2*
5. *COX3*
6. *CYTB*
7. *ND1*
8. *ND2*
9. *ND3*
10. *ND4*
11. *ND4L*
12. *ND5*
13. *ND6*

Single copy orthologs



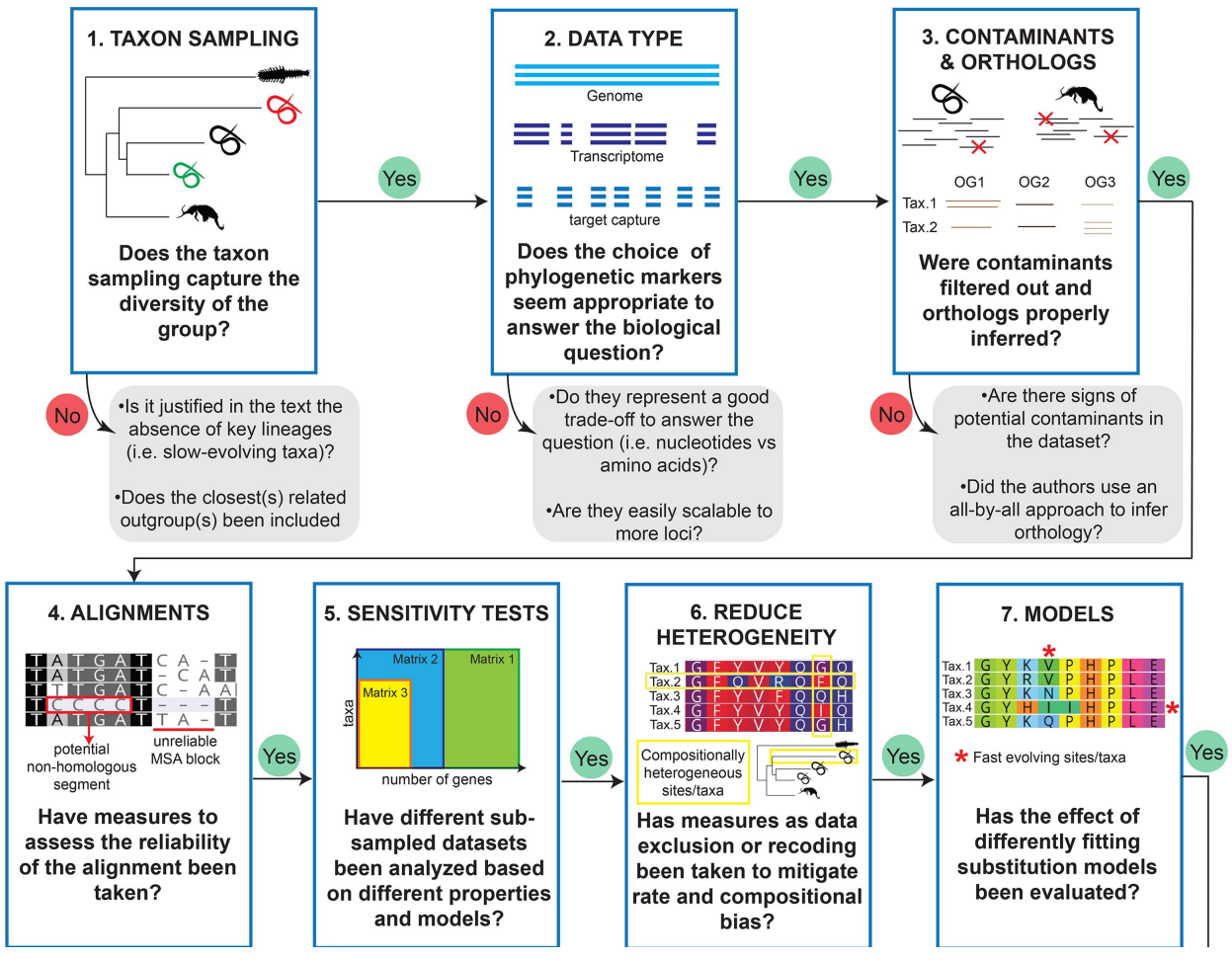
Работает как с геномами, так и с транскриптомами

Recommended!

A Practical Guide to Design and Assess a Phylogenomic Study

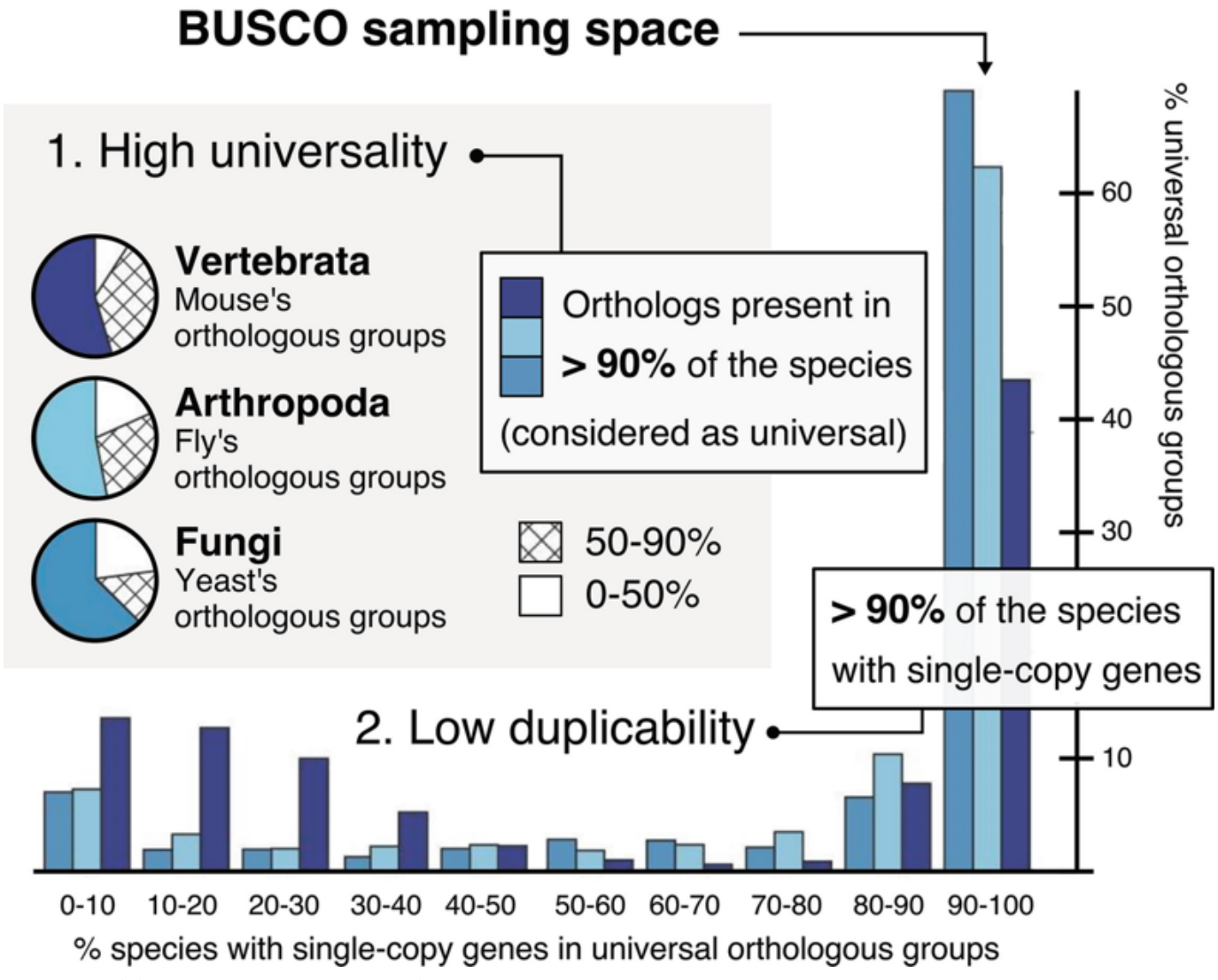
Jesus Lozano-Fernandez [1,2,*]

GBE



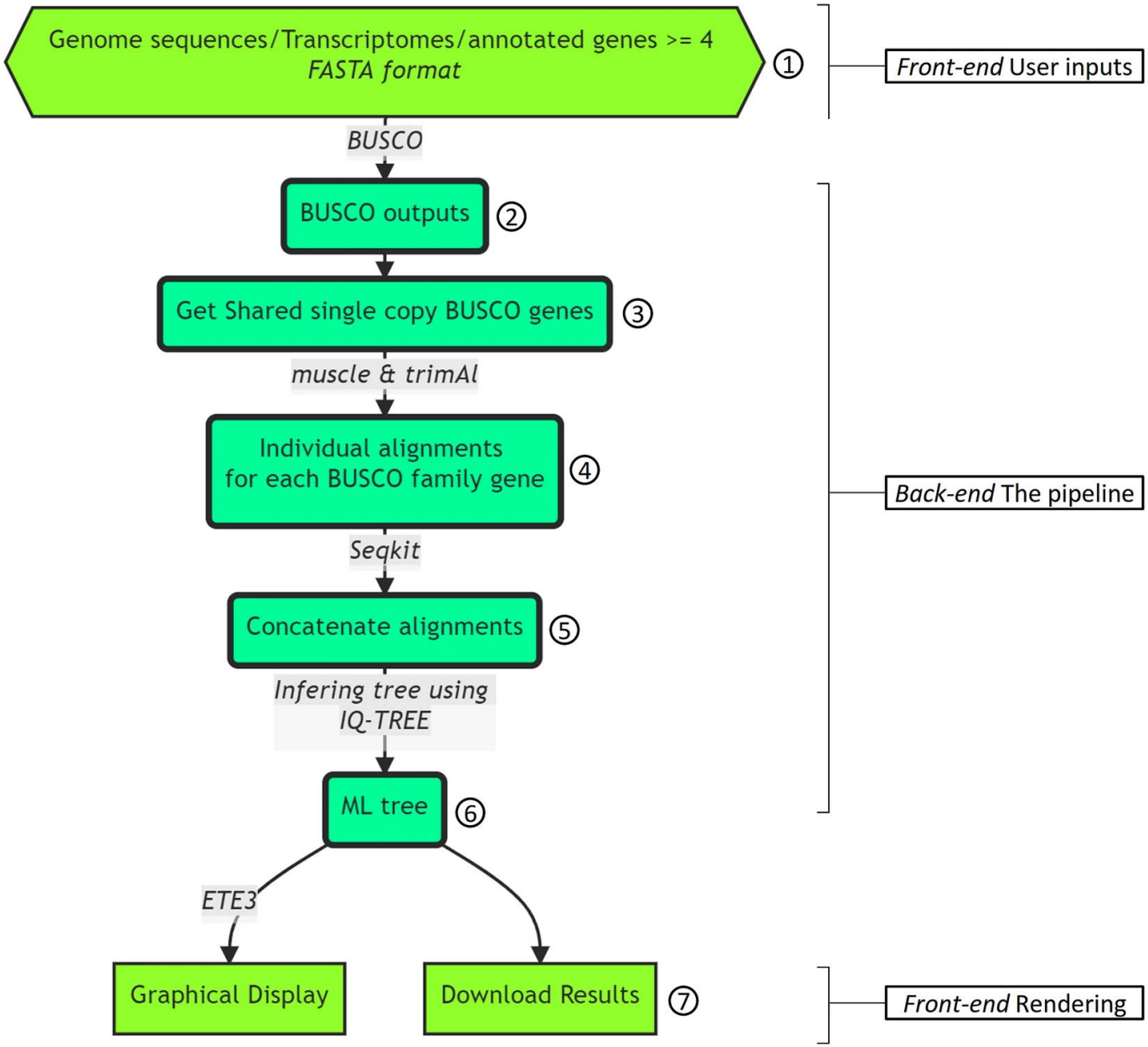
https://doi.org/10.1093/gbe/evac129

BUSCO



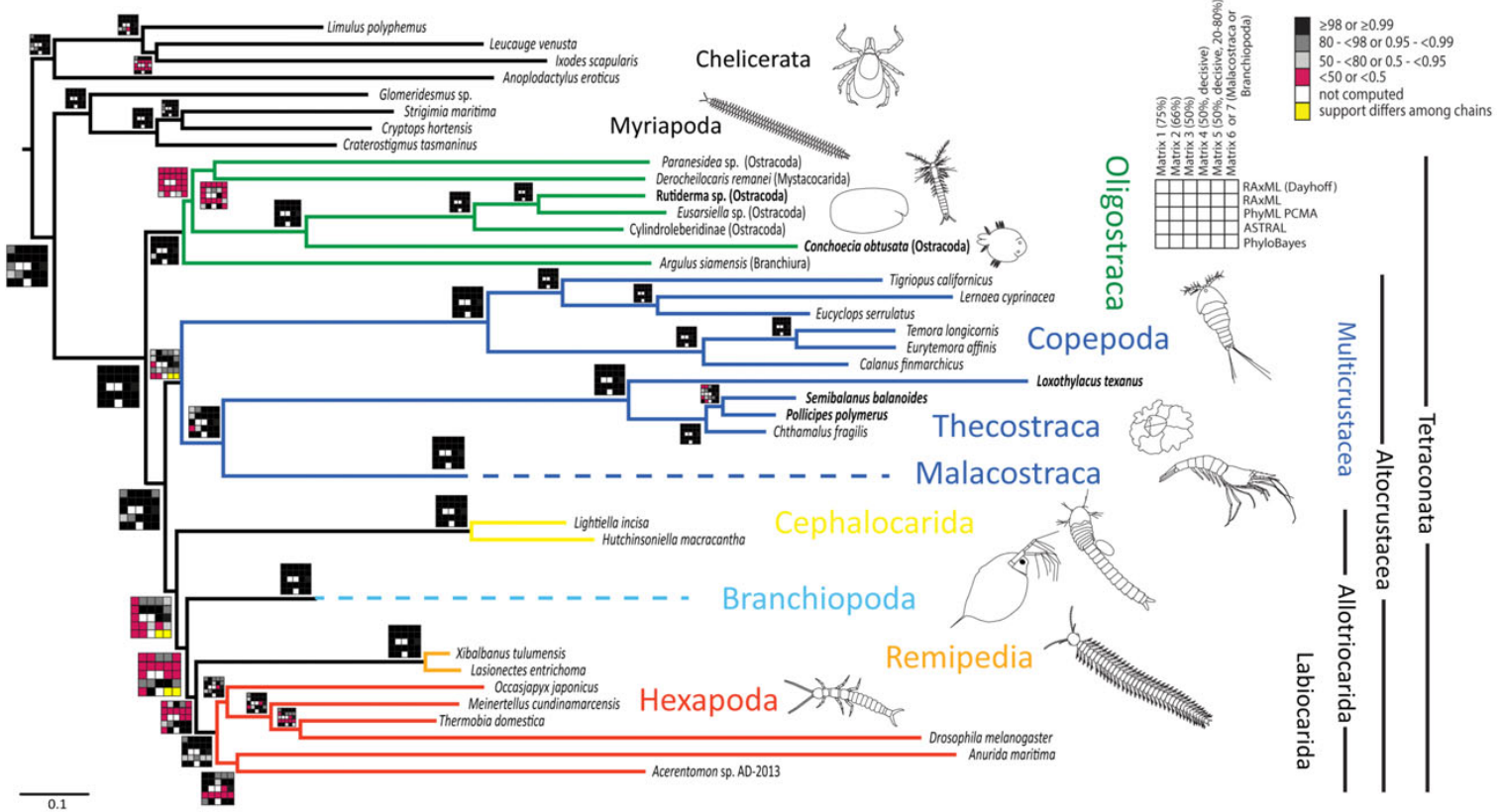
https://busco.ezlab.org/

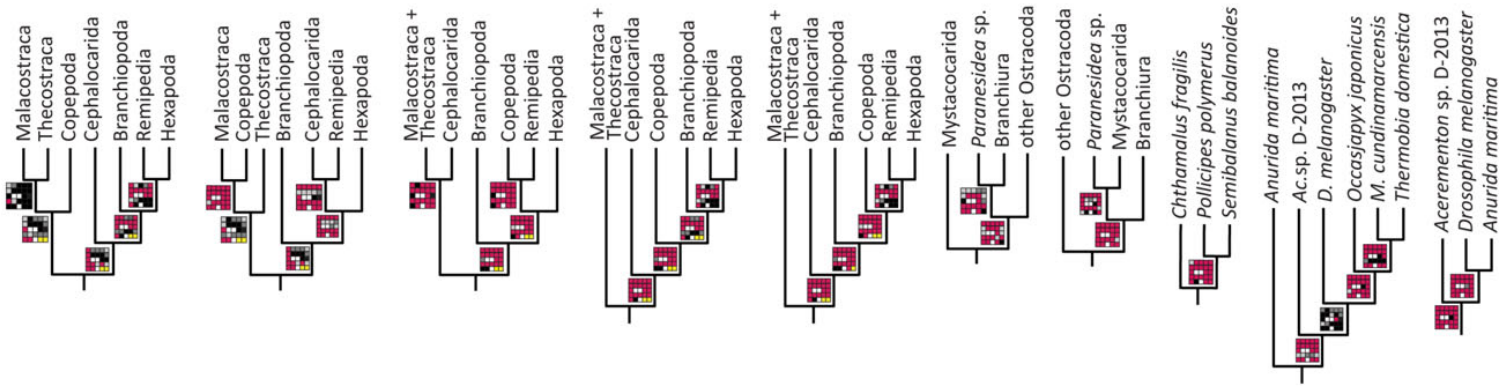
BuscoPhylo



https://www.nature.com/articles/s41598-022-22461-0

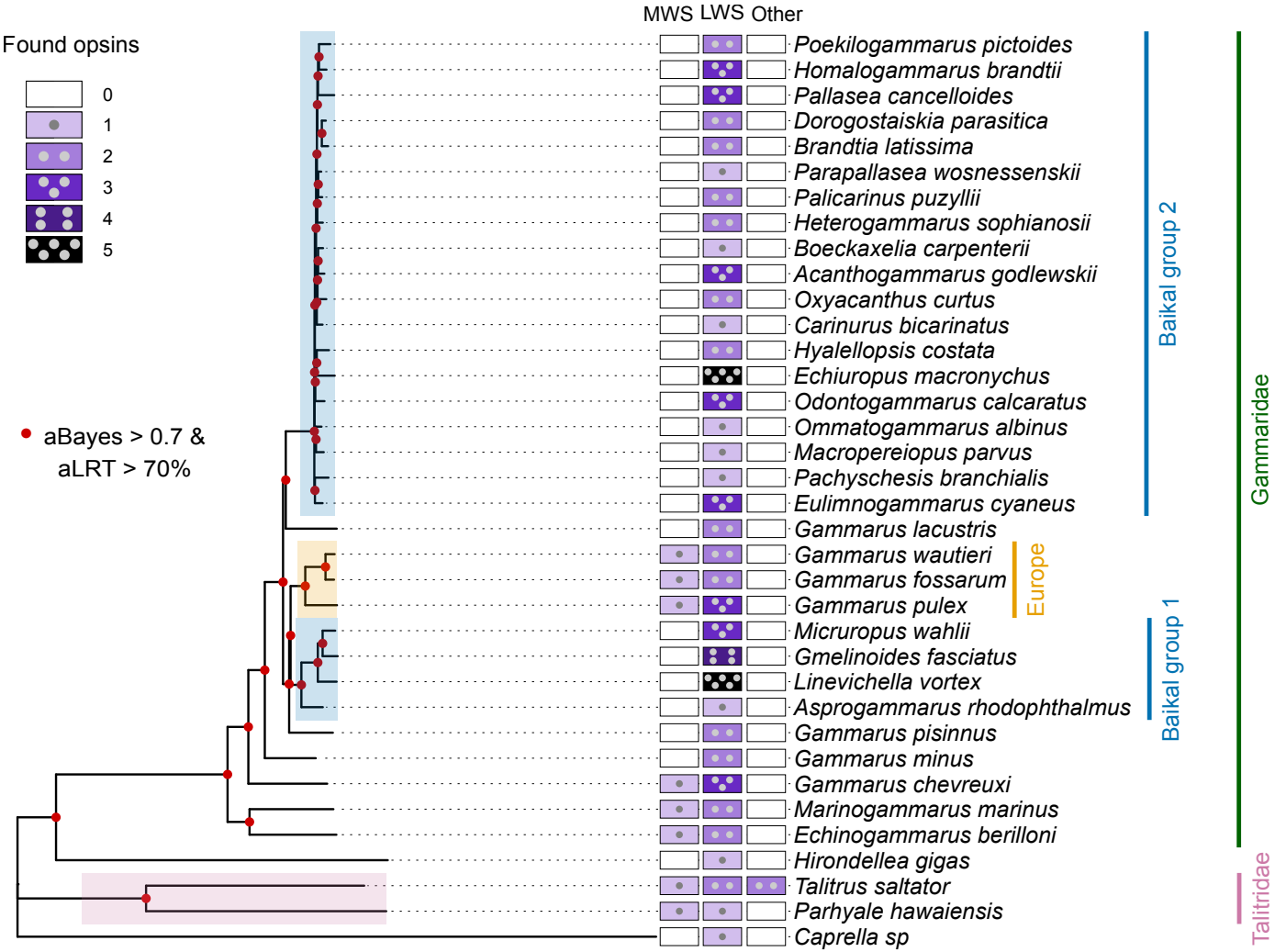
Single copy orthologs: examples





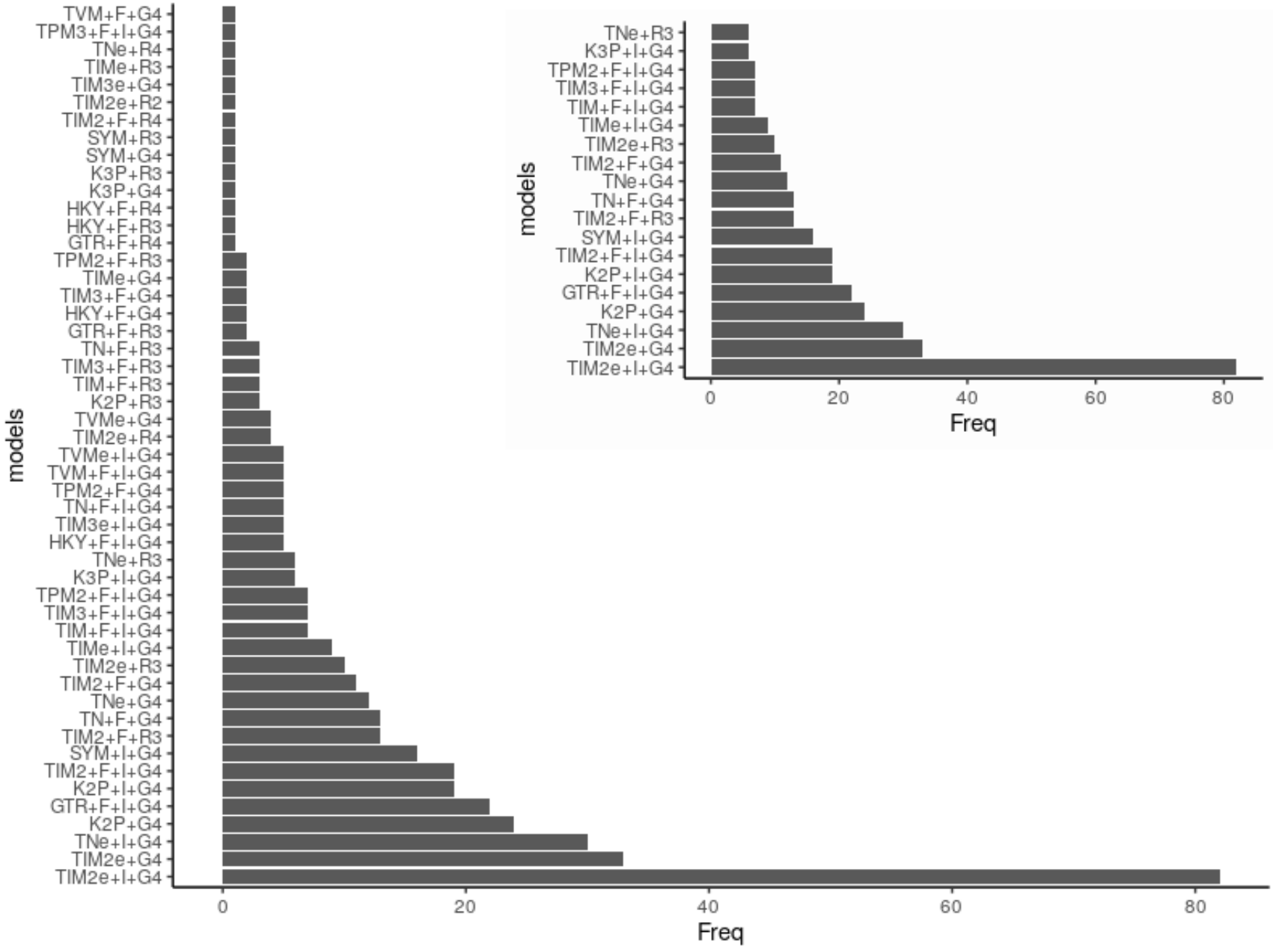
Schwentner *et al.*, 2018, http://dx.doi.org/10.1098/rspb.2018.1524

Single copy orthologs: examples



Drozdova *et al.*, 2021, https://doi.org/10.1186/s12862-021-01806-9
https://github.com/AlenaKizenko/pia3_amphipod_opsins/blob/master/other_scripts/2.3_species_tree.sh, 2.4_Fig2_S2B_species_tree.R

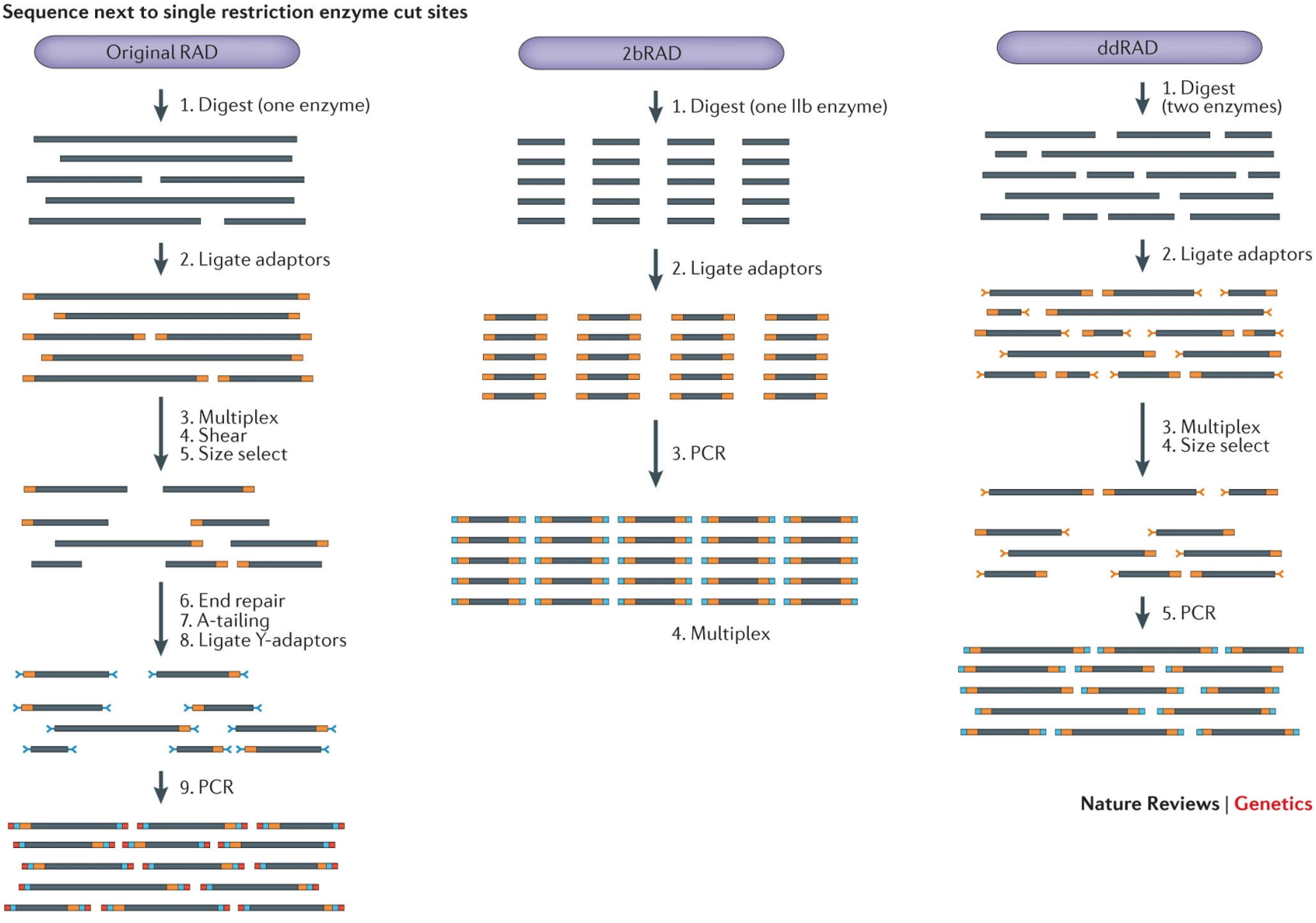
Statistics on best models



Упражнения: филогения 12 видов млекопитающих

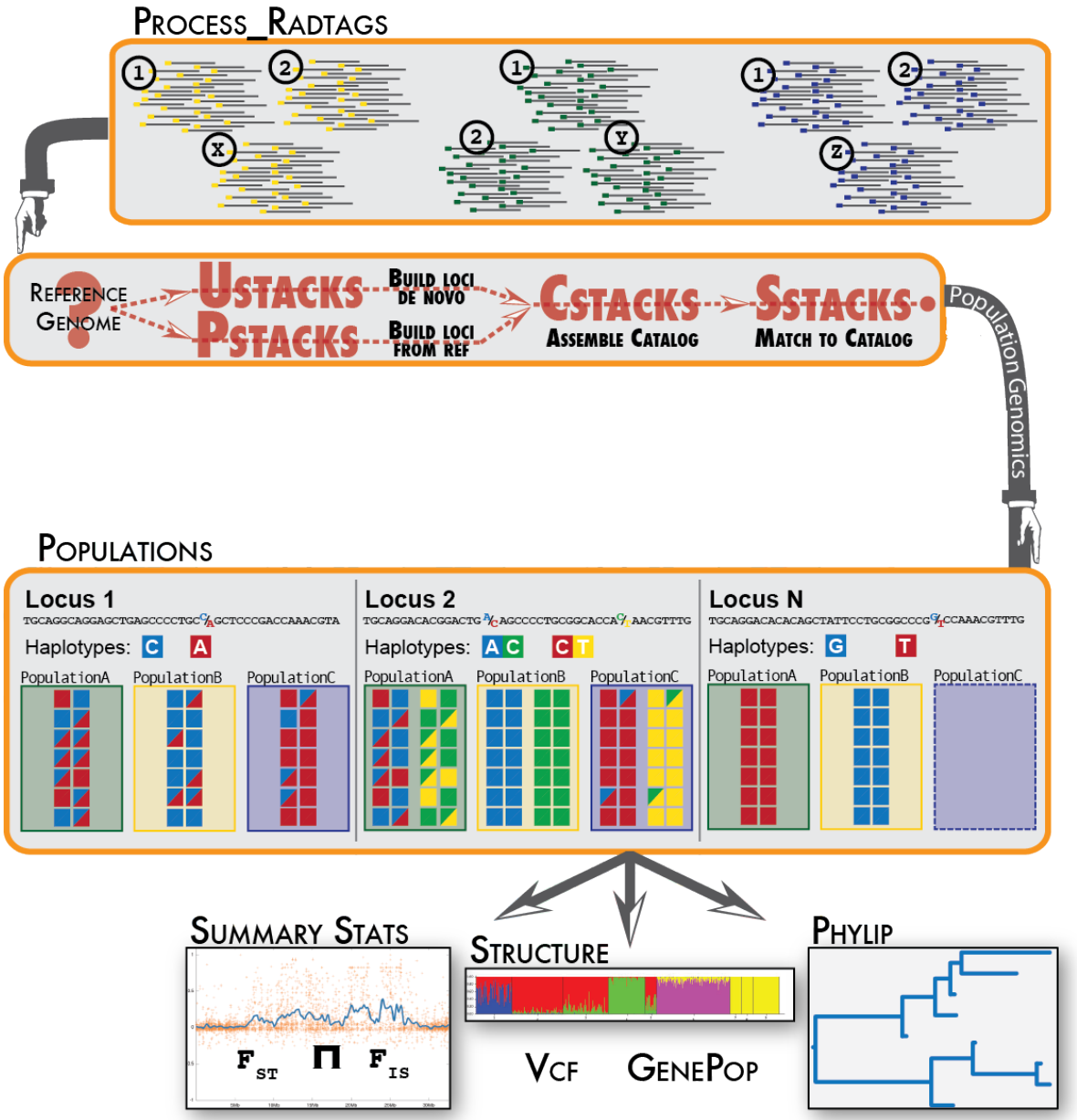
Упражнение 3: филогения 12 видов млекопитающих по однокопийным ортологам

RADseq

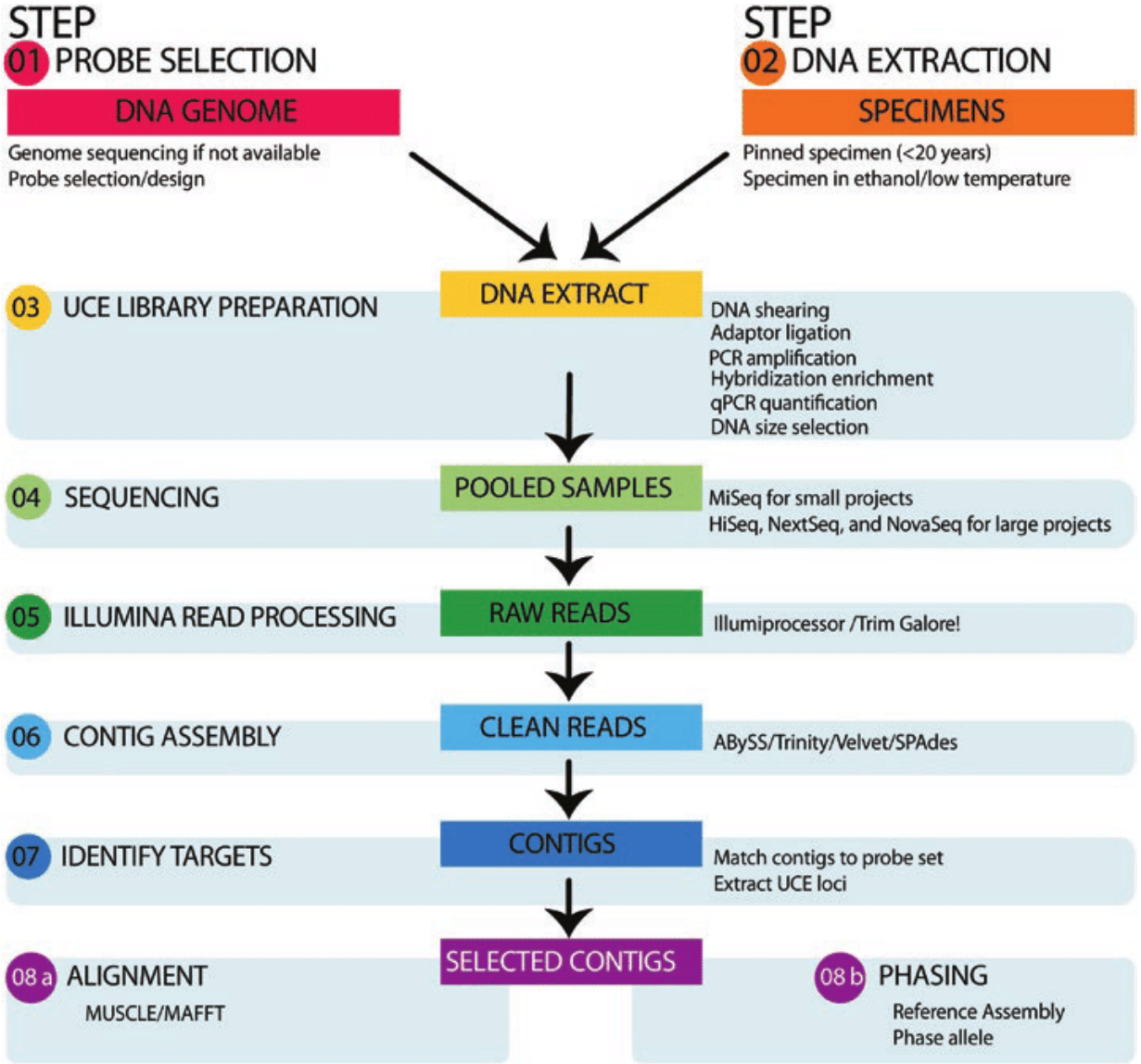


https://www.nature.com/articles/nrg.2015.28

RADseq: analysis

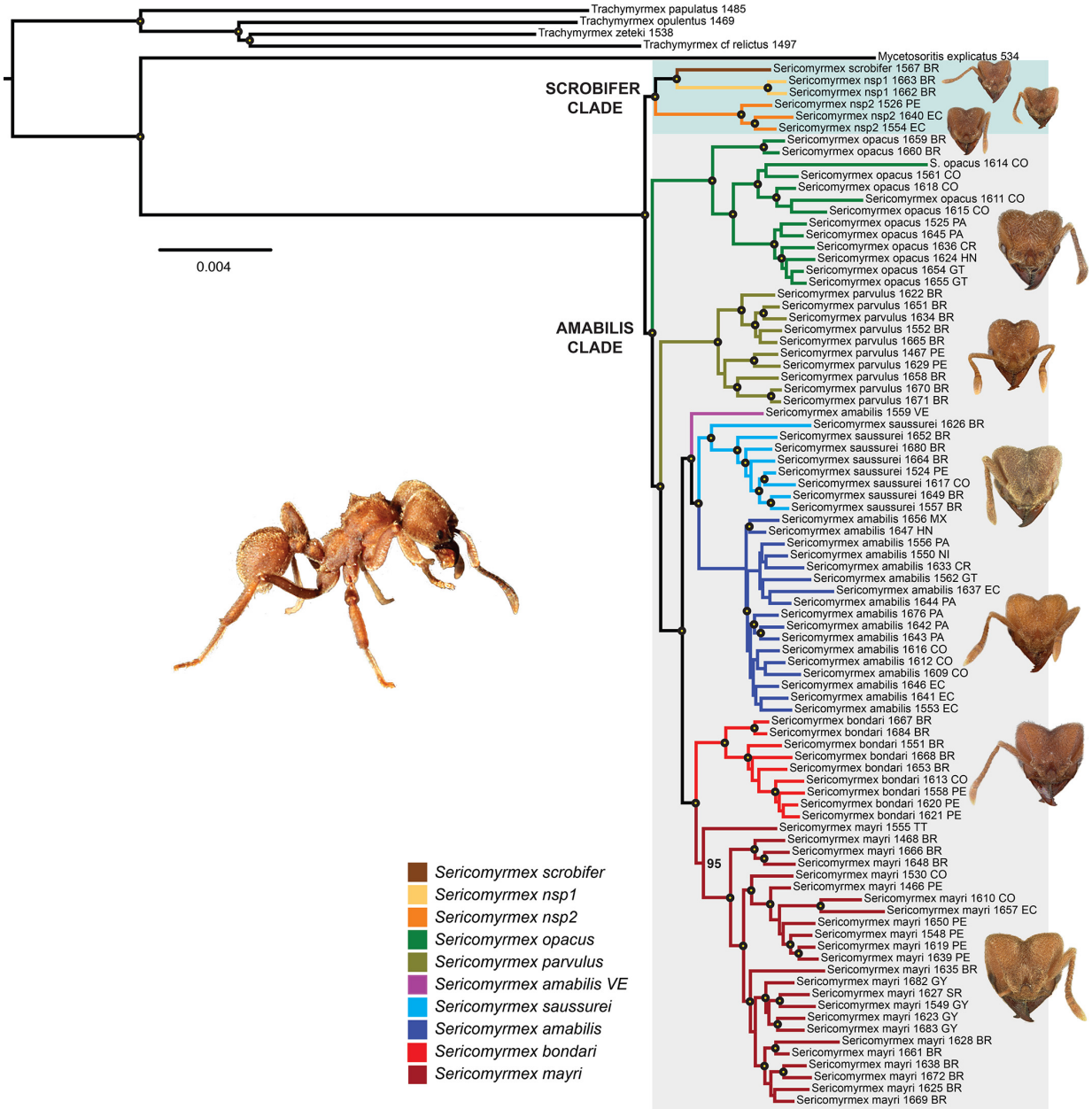


https://catchenlab.life.illinois.edu/stacks/manual-v1/



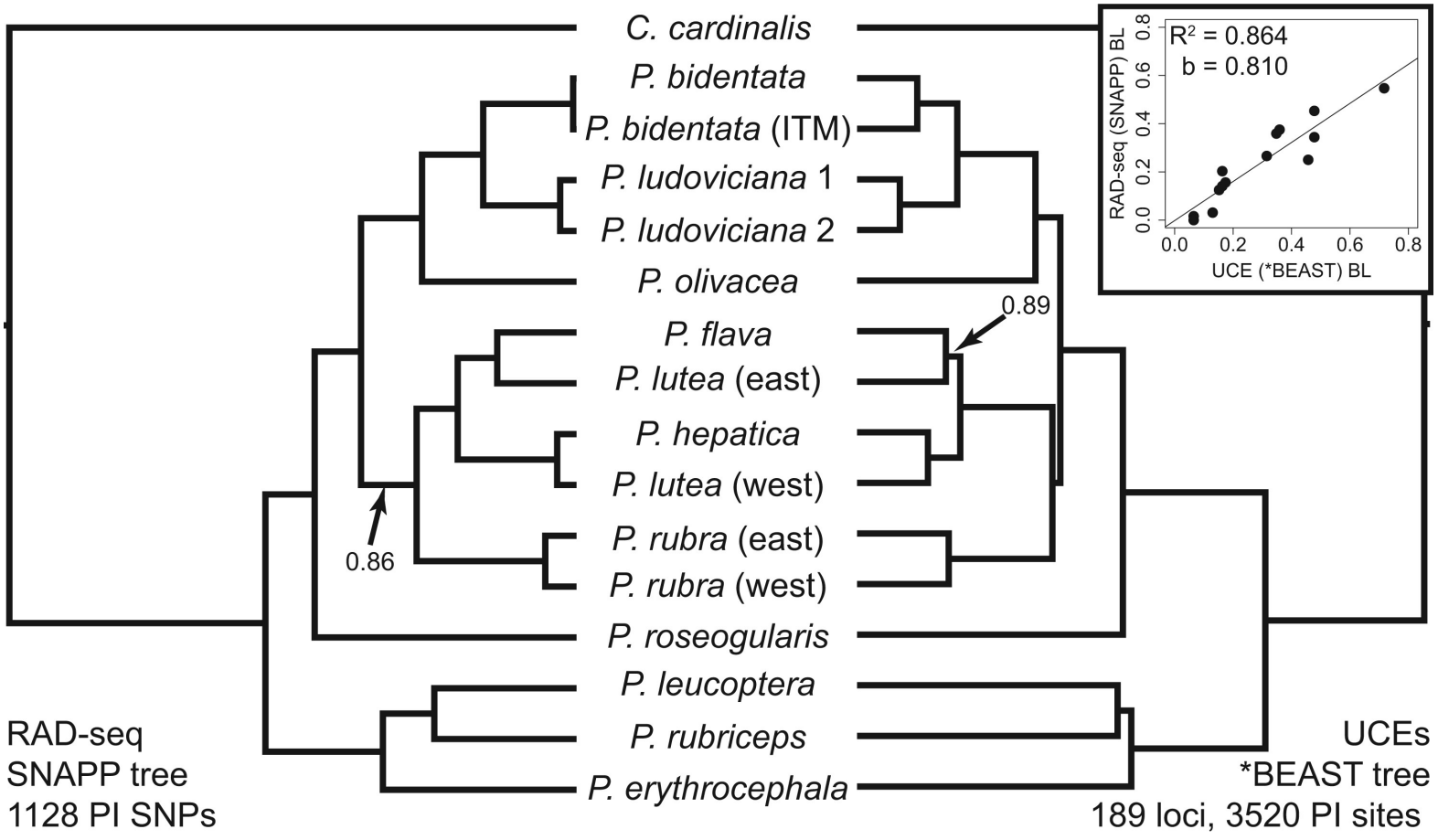
Miles Zhang *et al.*, 2019, http://dx.doi.org/10.1093/isd/ixz016

Ultraconserved elements = UCEs (example)



Jesovnik *et al.*, 2017, http://dx.doi.org/10.1111/syen.12228

RAD-seq and UCE: example



https://academic.oup.com/sysbio/article/65/4/640/1753369

Recommended!

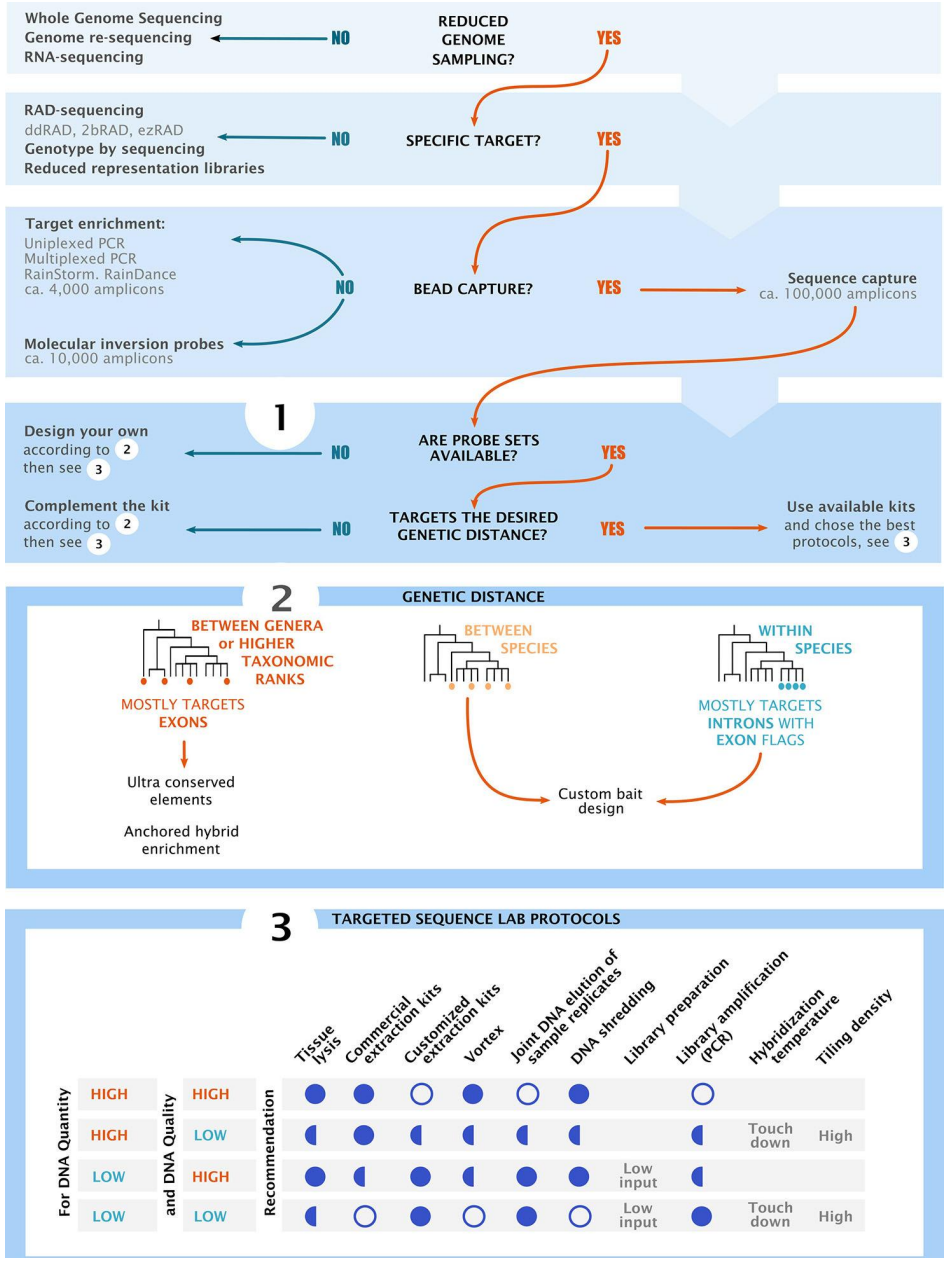
frontiers in Genetics

REVIEW
published: 21 February 2020
doi: 10.3389/fgene.2019.01407

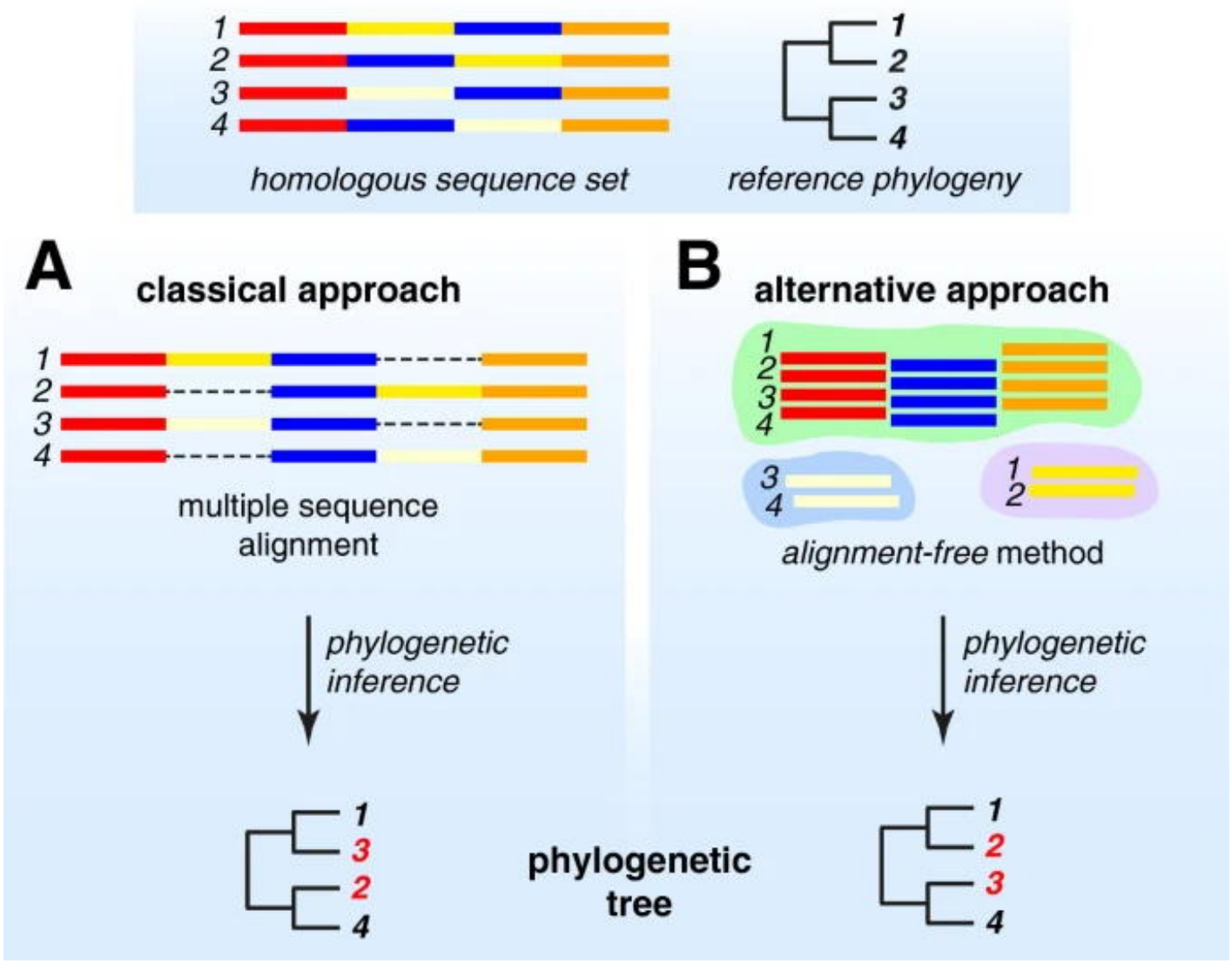
A Guide to Carrying Out a Phylogenomic Target Sequence Capture Project

Tobias Andermann[1,2†], Maria Fernanda Torres Jiménez[1,2*†], Pável Matos-Maraví[1,2,3], Romina Batista[2,4,5], José L. Blanco-Pastor[1,6], A. Lovisa S. Gustafsson[7], Logan Kistler[8], Isabel M. Liberal[1], Bengt Oxelman[1,2], Christine D. Bacon[1,2] and Alexandre Antonelli[1,2,9]*

https://doi.org/10.3389/fgene.2019.01407

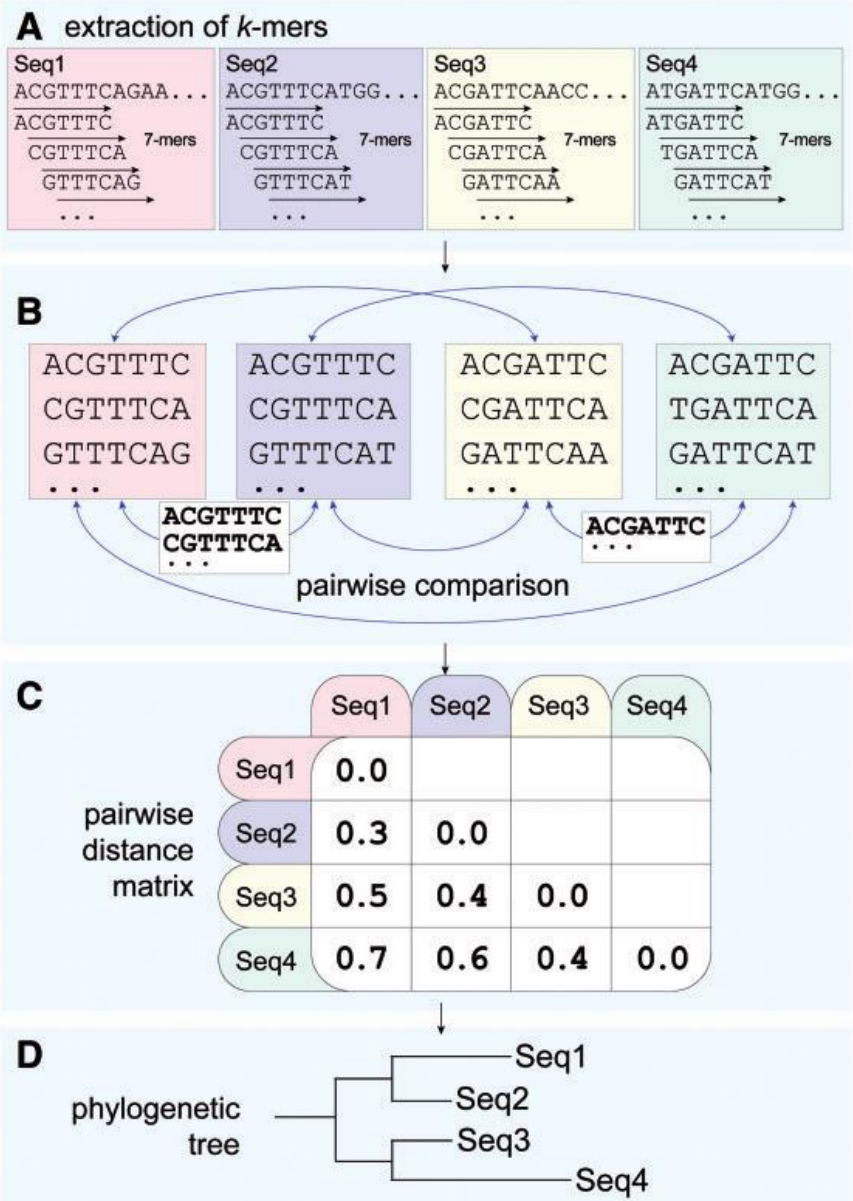


Alignment-free methods



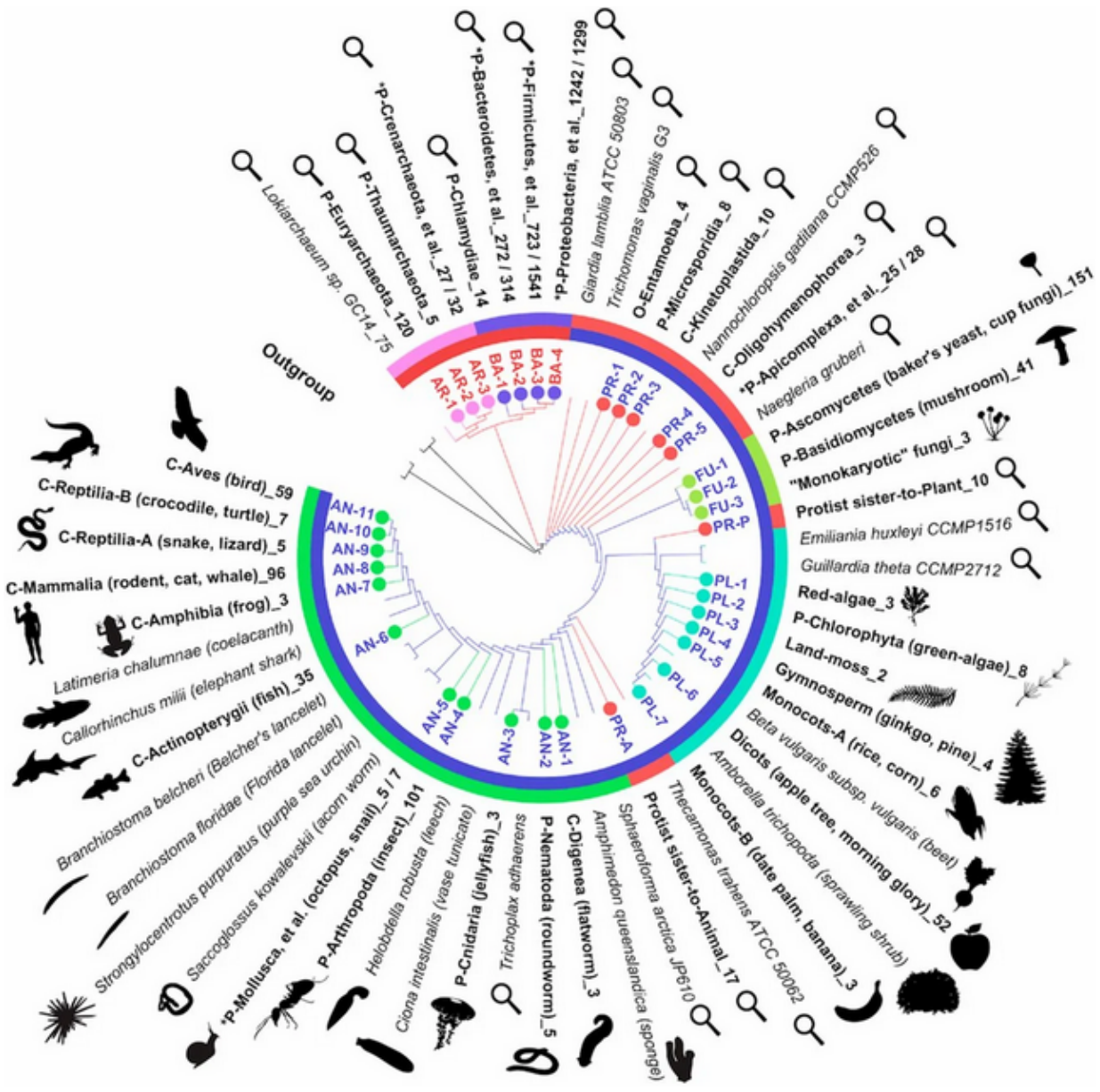
https://doi.org/10.1186/1745-6150-8-3

Alignment-free methods



https://doi.org/10.1093/bib/bbx067

Alignment-free methods



Choi & Kim *et al.*, 2020, https://doi.org/10.1073/pnas.1915766117

Упражнение №5: восстановление филогении 12 видов млекопитающих по аминокислотному составу протеомов

Интегративная таксономия