

Markers and phylogenetics
oooooooooooooooooooo

Phylogenomics intro
oooooooooooo

Single copy orthologs
oooooooooooo

Reduced representation
oooooooooooo

Alignment-free methods
oooo

Outro
o

Молекулярная филогенетика и филогеномика

Полина Дроздова

9–10 октября 2023

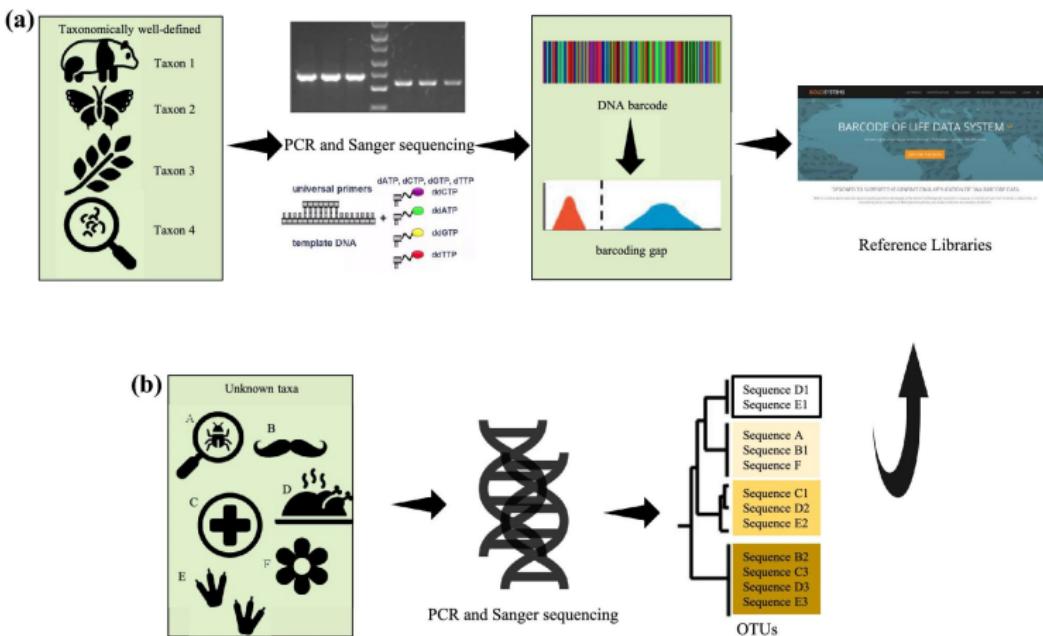


https://github.com/drozdovapb/ZIN_school_2023

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;



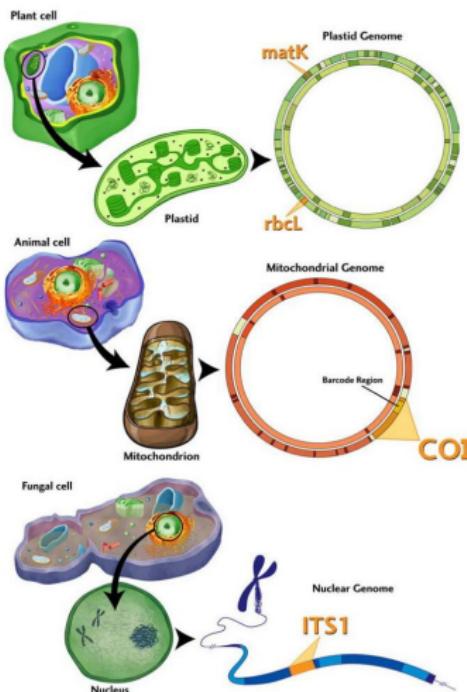
EPIC = exon primed intron crossing

3. желательна селективная нейтральность.

Примеры часто используемых маркеров

Table: key of tree of life

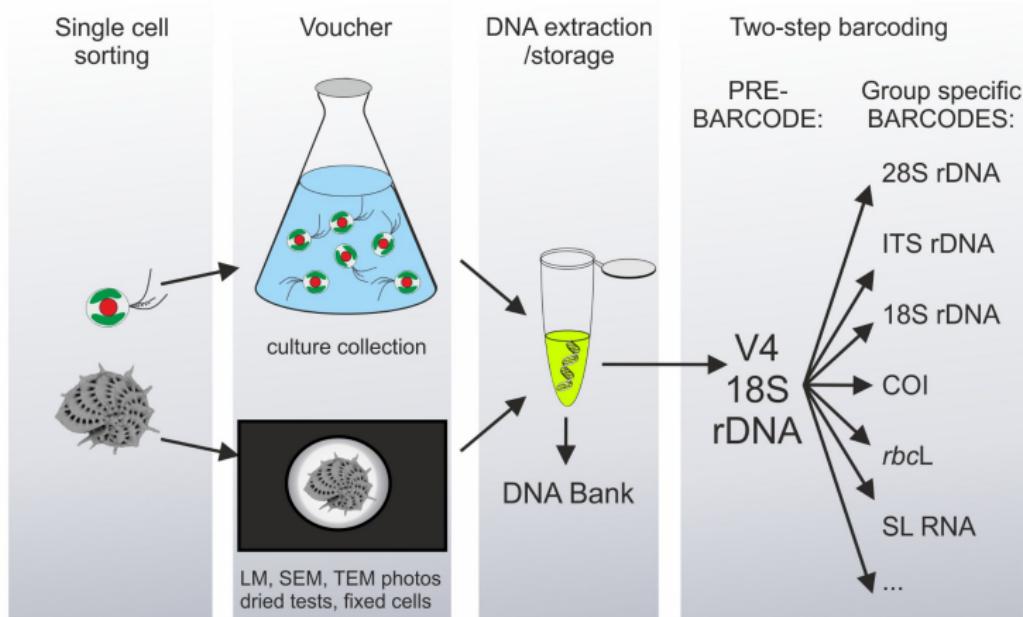
Color	Clade	Primary barcode	Secondary barcode
	Animals	COI	COI, 16S
	Fungi	ITS	LSU D1/D2
	Green algae	tufA	LSU D2/D3
	Land plants	rbcL/matK	psbA-trnH/ITS
	Algae	COI-5P	LSU D2/D3



<https://doi.org/10.1155/2022/1846485>

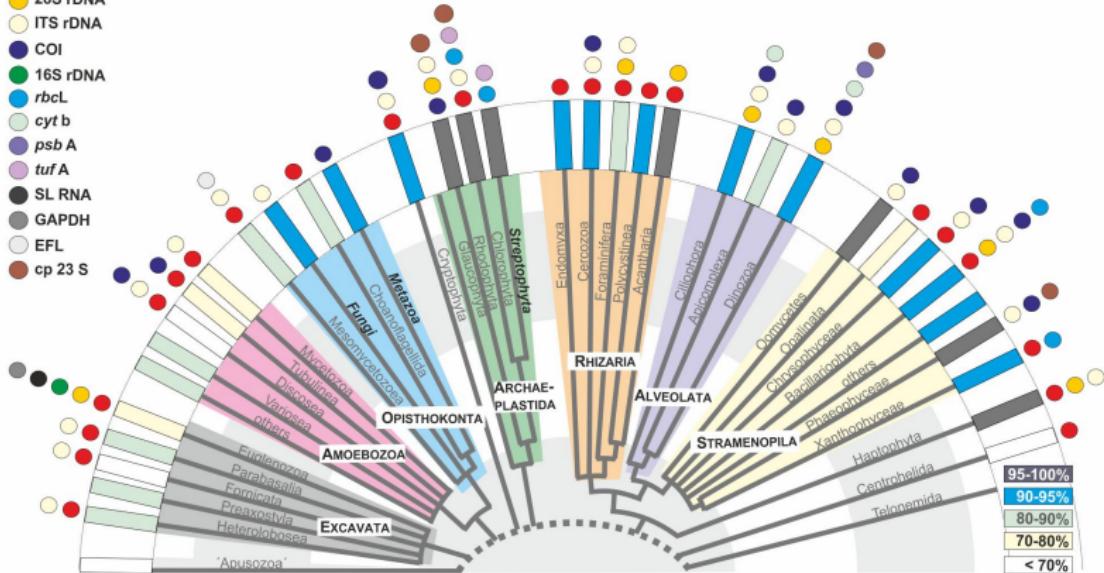
<https://www.cbd.int/doc/publications/cbd-ts-94-en.pdf>

Примеры часто используемых маркеров: протисты



Примеры часто используемых маркеров: протисты

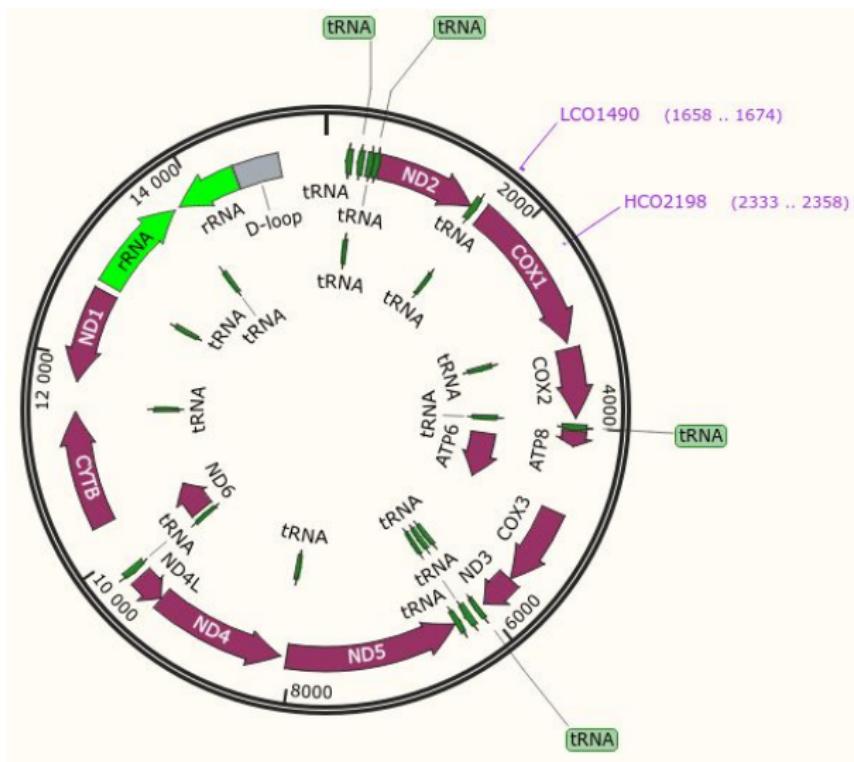
- 18S rDNA
- 28S rDNA
- ITS rDNA
- COI
- 16S rDNA
- *rbcL*
- *cyt b*
- *psb A*
- *tufA*
- SL RNA
- GAPDH
- EFL
- cp 23 S



A.

B. C.

Примеры часто используемых маркеров: животные



Структура локуса рДНК

Bacteria / Archaea:



Eukaryota:

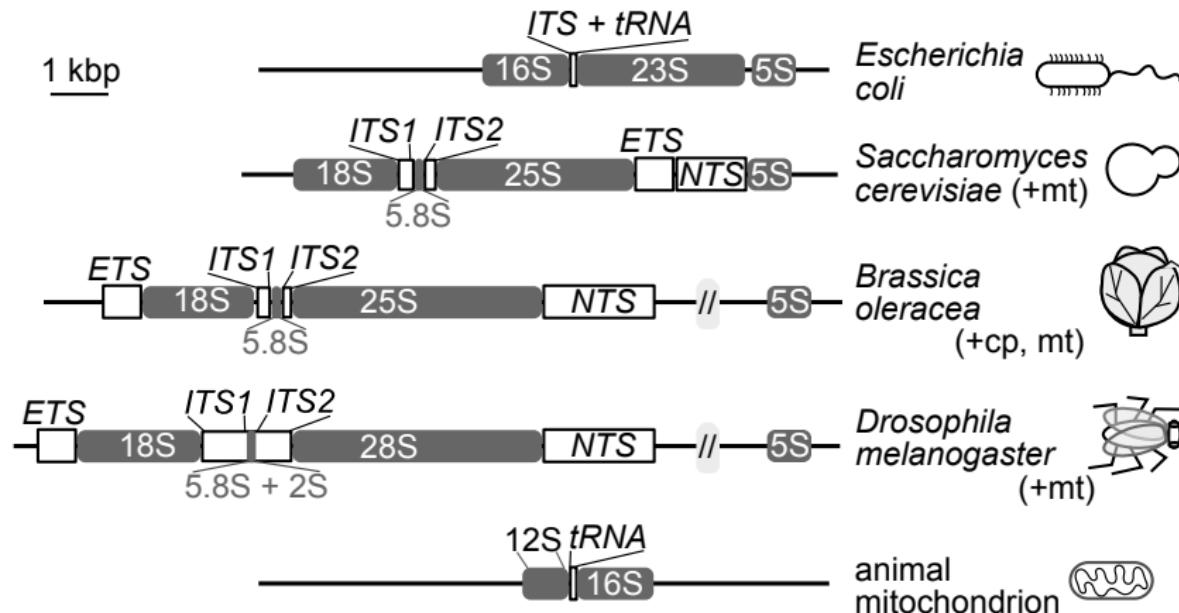


ITS = internal transcribed spacer

ETS = external transcribed spacer

IGS / NTS = intergenic / non-transcribed spacer

Структура локуса рДНК у разных организмов



Штрихкодирование ДНК (DNA barcoding)

Albacore Tuna



Esoclar

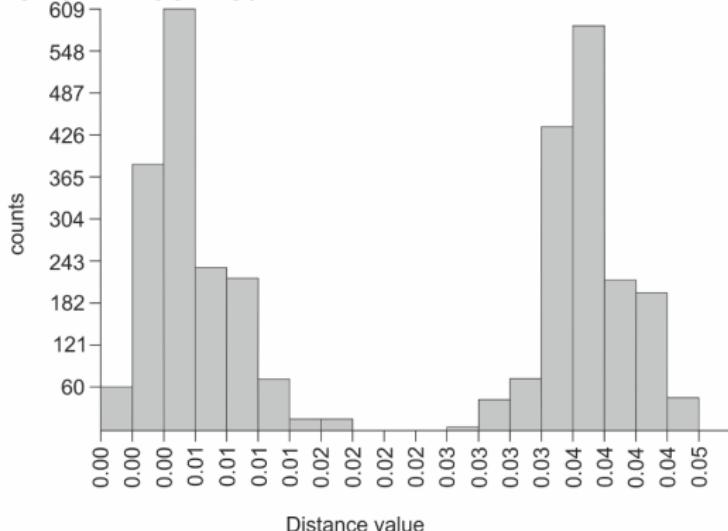


Atlantic Cod



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

Barcode gap — разрыв между внутривидовой и межвидовой изменчивостью.



Markers and phylogenetics

oooooooooooo●oooooooo

Phylogenomics intro

oooooooooooo

Single copy orthologs

oooooooooooo

Reduced representation

oooooooooooo

Alignment-free methods

oooo

Outro

o

Проект Barcode of Life



The image shows the homepage of the Barcode of Life Data System (BOLD). At the top left, it says "BOLD SYSTEMS". On the right side, there is a search icon and a menu icon. The main title "BARCODE OF LIFE DATA SYSTEM" is displayed prominently in large white text, with "v4" in smaller orange text next to "SYSTEM". Below the title, a subtitle reads "Advancing biodiversity science through DNA-based species identification." At the bottom center, there is an orange button with the text "EXPLORE THE DATA". The background of the page features a world map with various animal silhouettes.

<https://boldsystems.org/>

Markers and phylogenetics
oooooooooooo●oooo

Phylogenomics intro
oooooooooooo

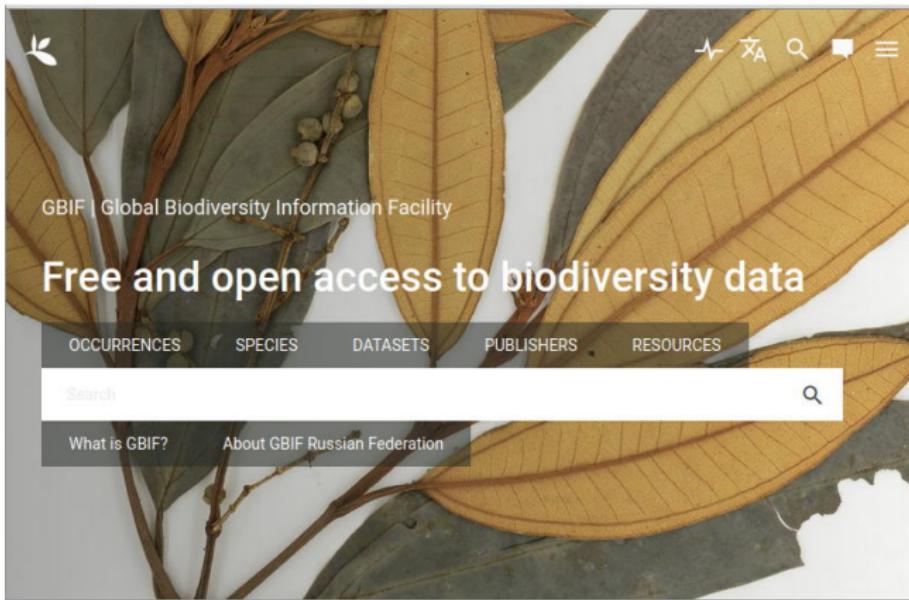
Single copy orthologs
oooooooooooo

Reduced representation
oooooooooooo

Alignment-free methods
oooo

Outro
o

База GBIF

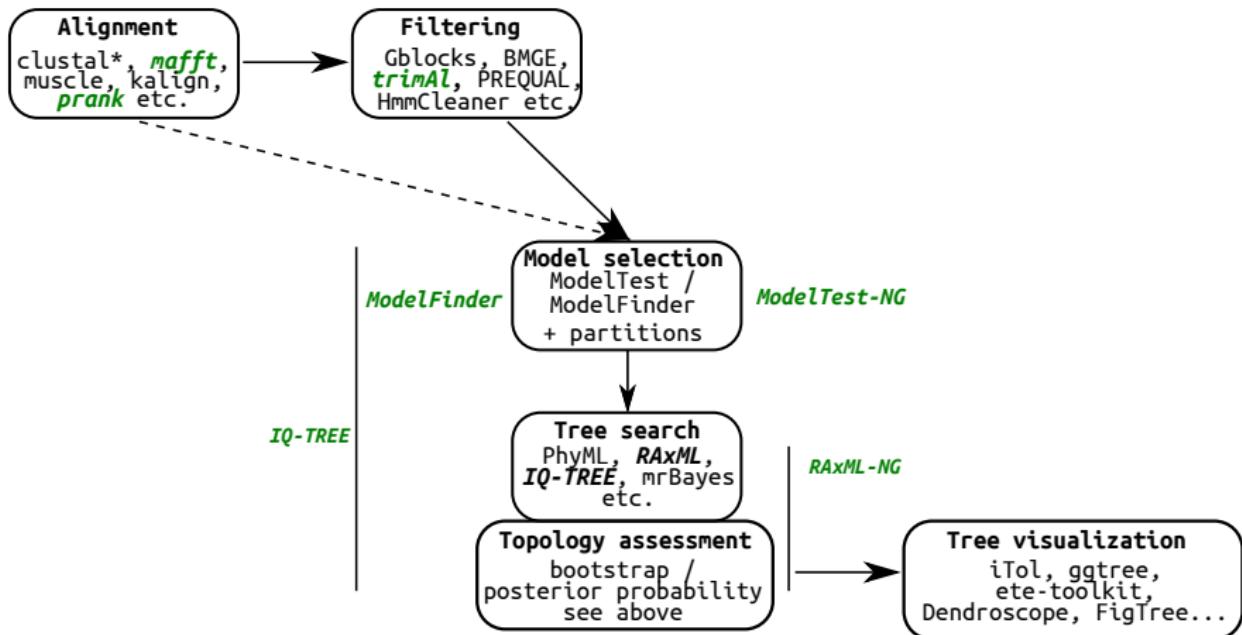


iNaturalist

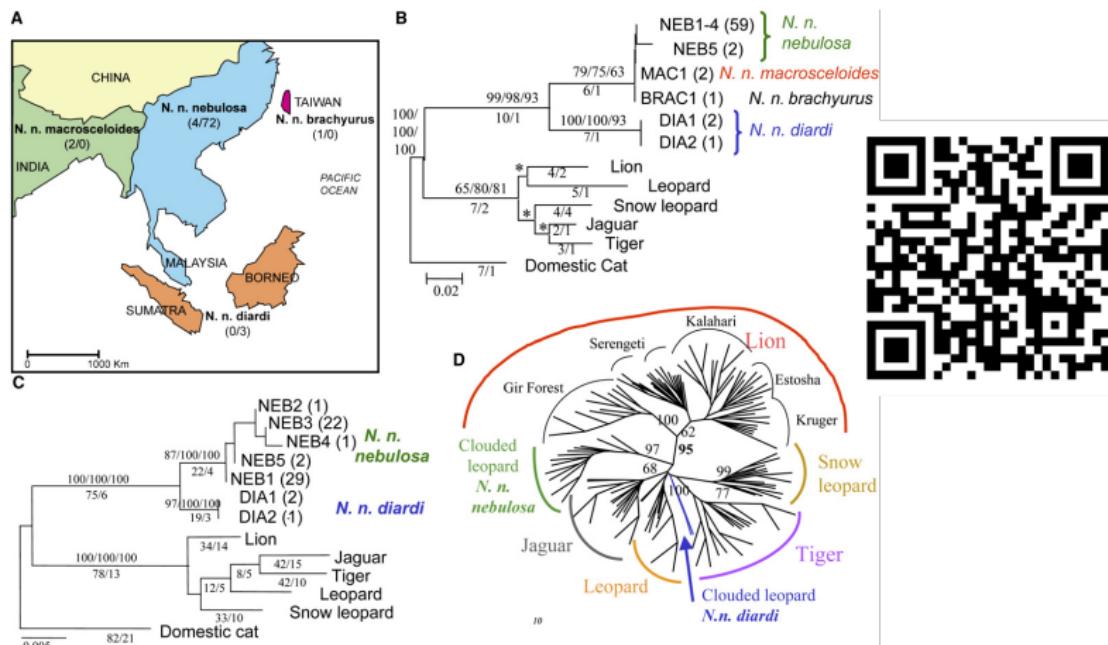
BOLD SYSTEMS

<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: филогенетия 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*

Markers and phylogenetics

Phylogenomics intro

Single copy orthologs

Reduced representation

Alignment-free methods

Outro

oooooooooooo●

oooooooooo

oooooooooo

oooooooooo

oooo

o

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

Markers and phylogenetics
oooooooooooooooooooo

Phylogenomics intro
●oooooooooooo

Single copy orthologs
oooooooooooo

Reduced representation
oooooooooooo

Alignment-free methods
oooo

Outro
o

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...

Markers and phylogenetics
oooooooooooooooo

Phylogenomics intro
○●○○○○○○○○

Single copy orthologs
○○○○○○○○○○

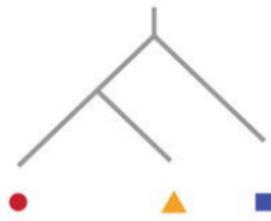
Reduced representation
○○○○○○○○○○

Alignment-free methods
○○○○

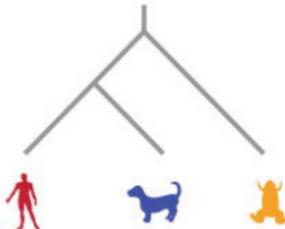
Outro
○

Потеря гомологов

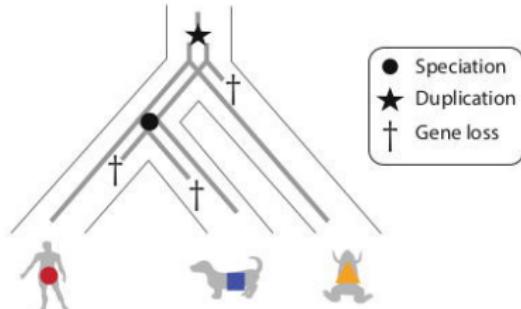
Gene Tree



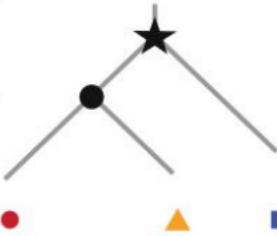
Species Tree



Reconciled Tree
(Full Representation)

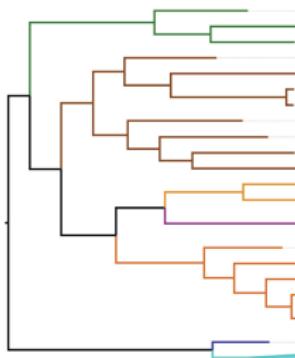


Reconciled Tree
(Simple Representation)

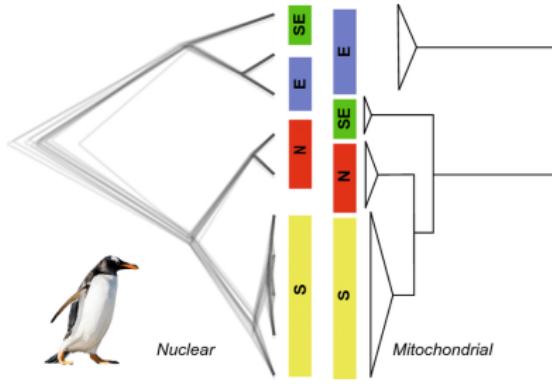
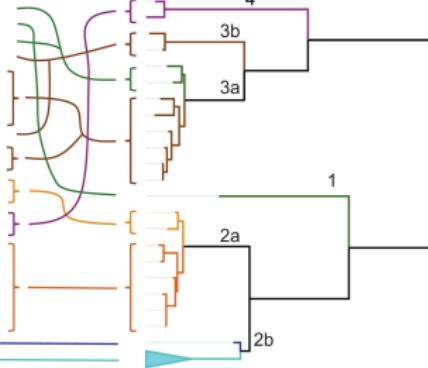


Mito-nuclear discordance

A Autosomal SNPs

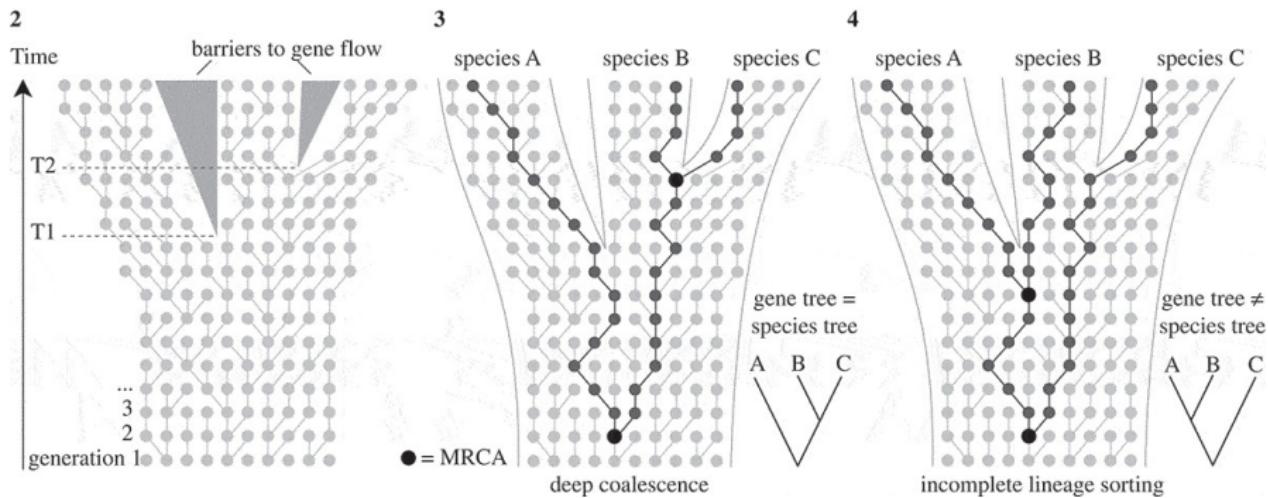
BGI_RF01
BGI_OFS01
BGI_SJS01
BB037
BB034
BB049
BB059
BB020
EB027
WB039
GRZ
ABC1
BGI_ABC06
BGI_GP01
CON001
BGI_ABC02
ABC2
BGI_ABC04
BGI_ABC01
BGI_ABC03
BGI_ABC05

Mitochondrial genomes





Incomplete lineage sorting

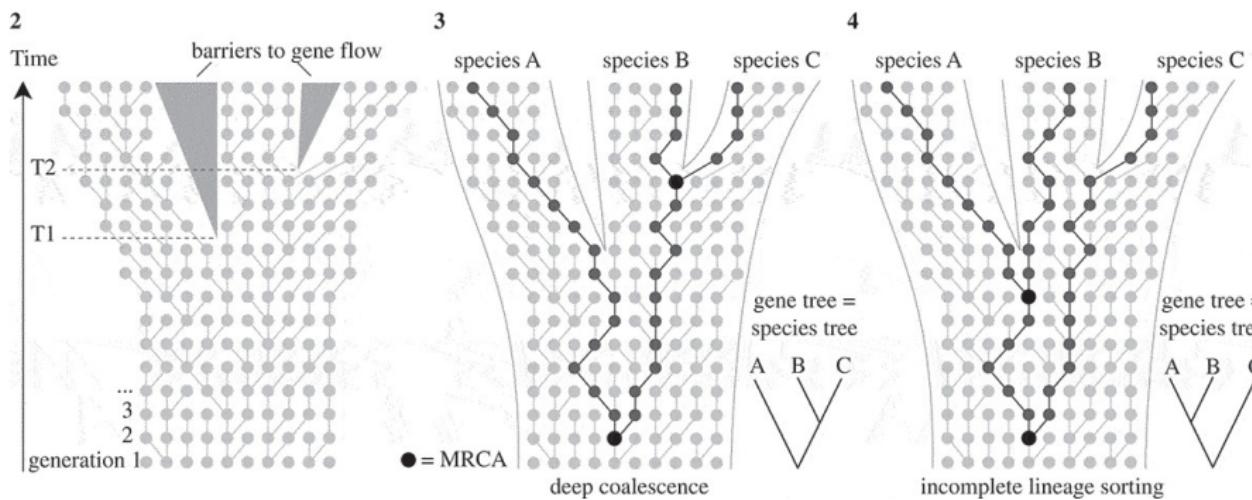


Leliaert et al., 2014,

<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

Markers and phylogenetics

Phylogenomics intro
○○○○●○○○○

Single copy orthologs



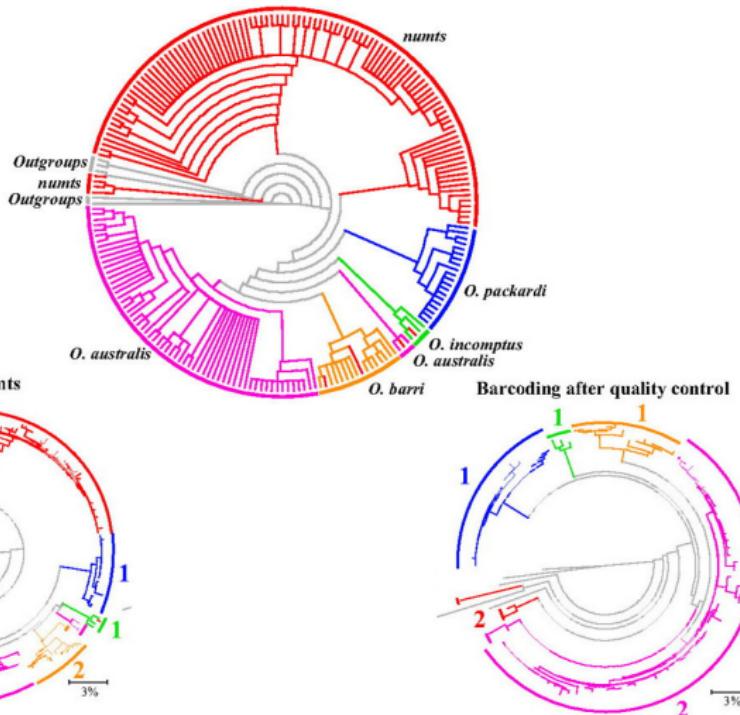
Reduced representation
oooooooooooo

Alignment-free methods

Outro

NUMTs = nuclear mitochondrial DNA segment

[new might]



Markers and phylogenetics
oooooooooooooooooooo

Phylogenomics intro
ooooo●oooo

Single copy orthologs
oooooooooooo

Reduced representation
oooooooooooo

Alignment-free methods
oooo

Outro
o

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...)

Markers and phylogenetics
oooooooooooooooooooo

Phylogenomics intro
oooooooo●○

Single copy orthologs
oooooooooooo

Reduced representation
oooooooooooo

Alignment-free methods
oooo

Outro
o

Упражнение №2: восстановление филогении по 13 митохондриальным белкам

Markers and phylogenetics
oooooooooooooooo

Phylogenomics intro
oooooooo●

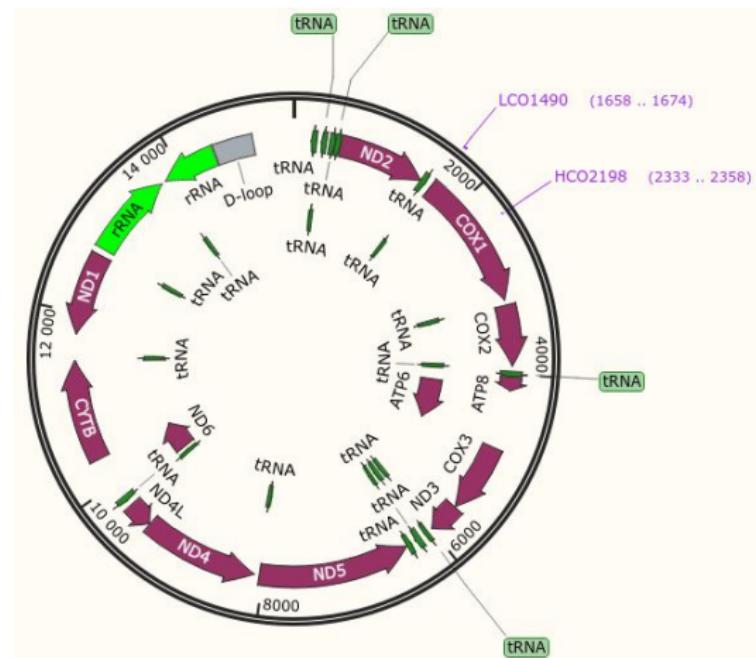
Single copy orthologs
oooooooo

Reduced representation
oooooooo

Alignment-free methods
oooo

Outro
○

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*

Markers and phylogenetics
ooooooooooooooo

Phylogenomics intro
oooooooooooo

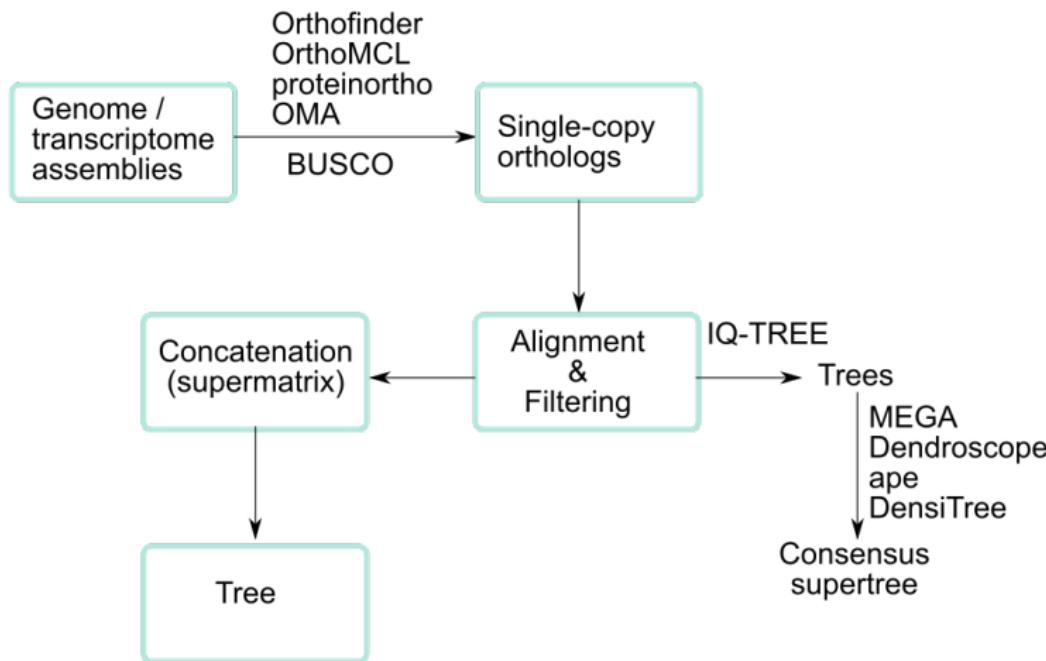
Single copy orthologs
●oooooooooooo

Reduced representation
oooooooooooo

Alignment-free methods
oooo

Outro
o

Single copy orthologs



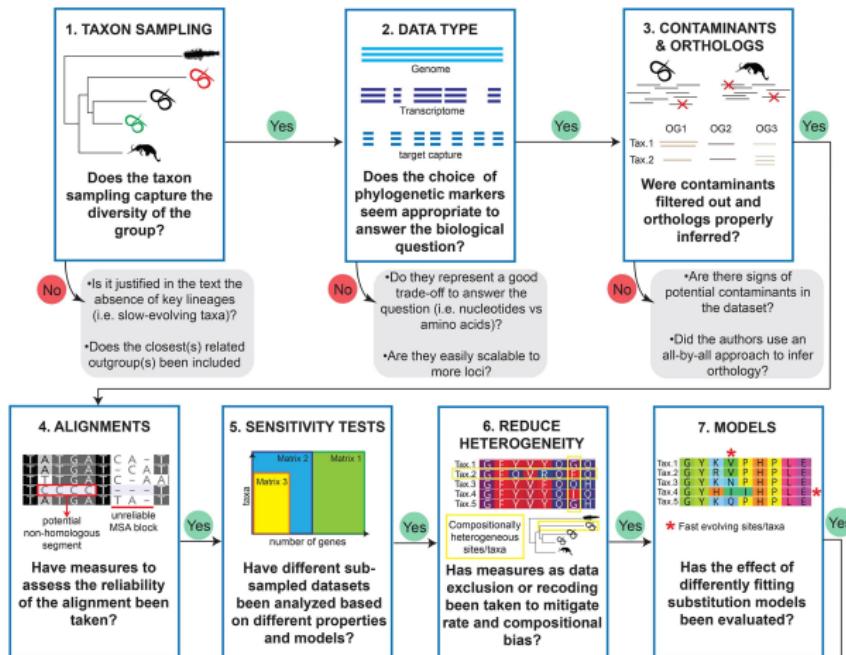
Работает как с геномами, так и с транскриптомами

Recommended!

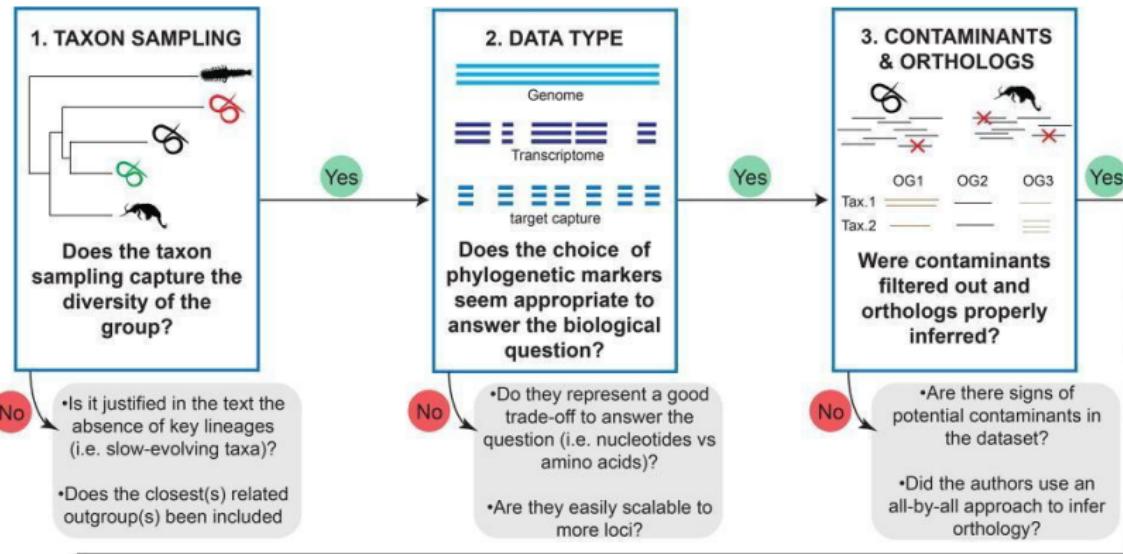
A Practical Guide to Design and Assess a Phylogenomic Study

Jesus Lozano-Fernandez ①,2,*

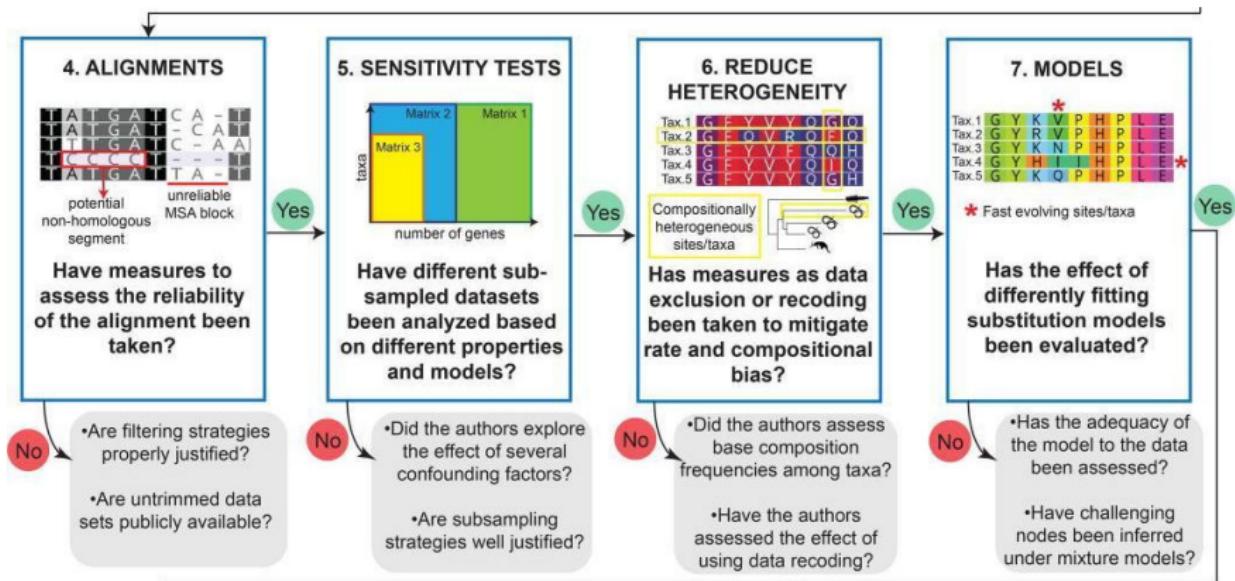
GBE



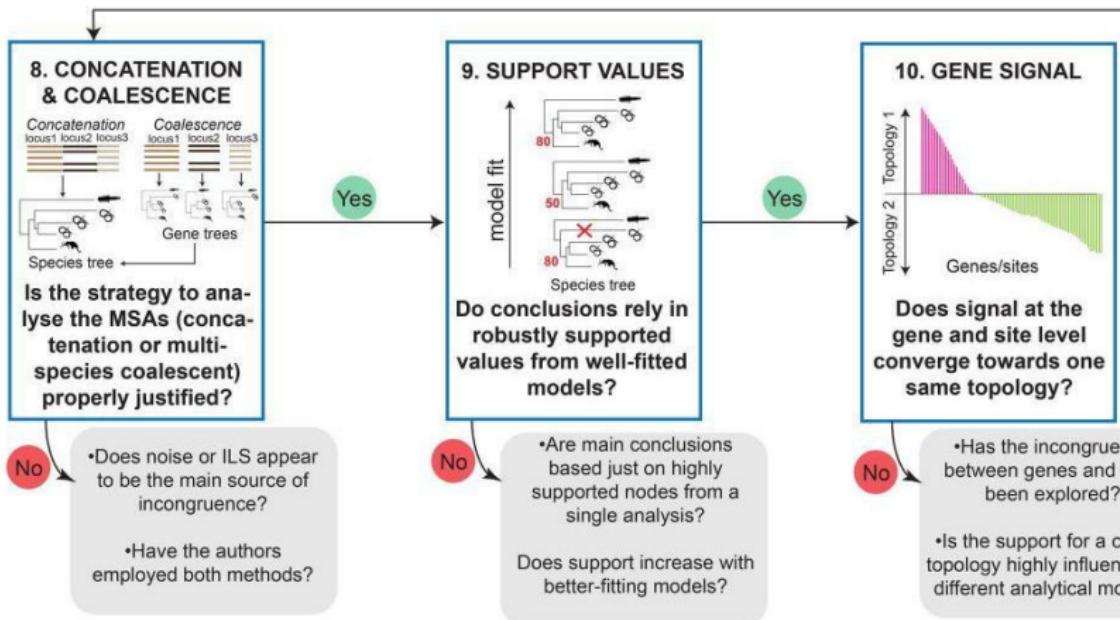
Recommended!



Recommended!

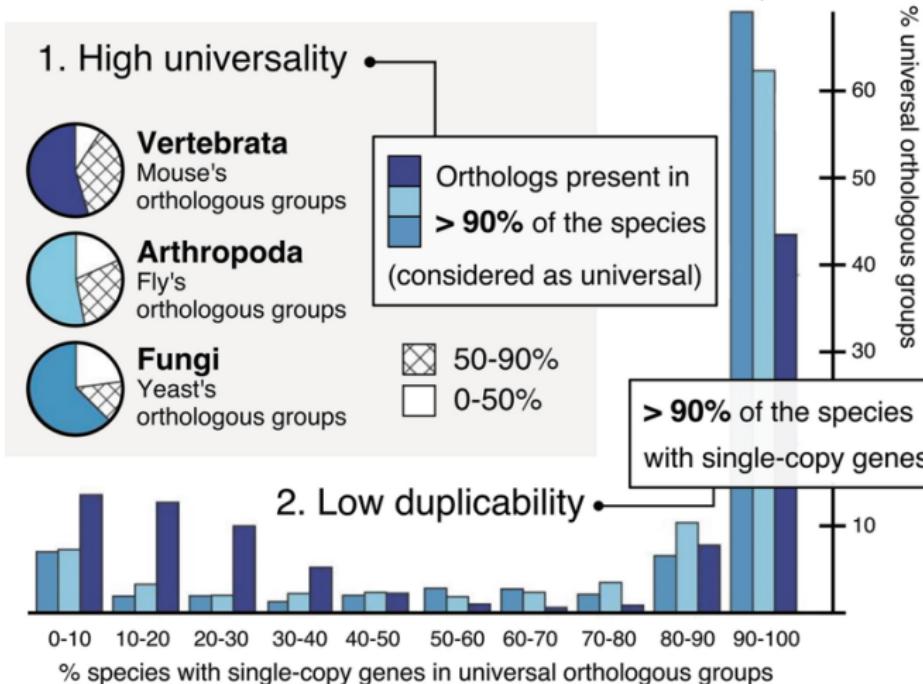


Recommended!



BUSCO

BUSCO sampling space



Markers and phylogenetics
oooooooooooooooooooo

Phylogenomics intro
oooooooooooo

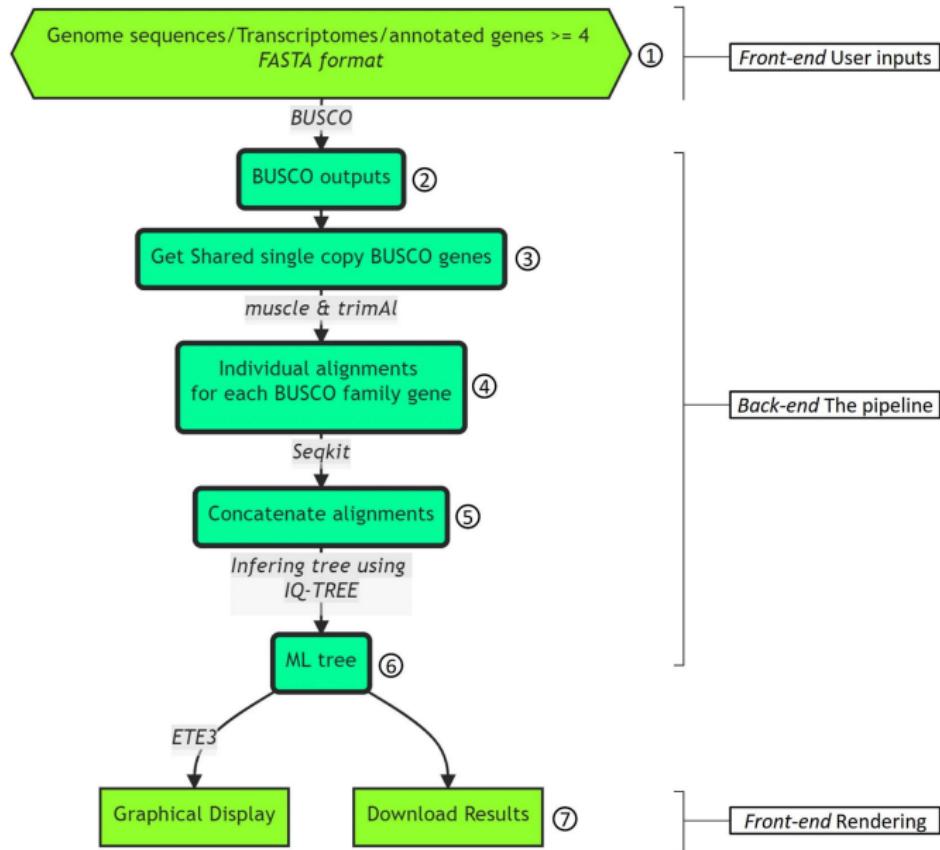
Single copy orthologs
ooo●oooooo

Reduced representation
oooooooooo

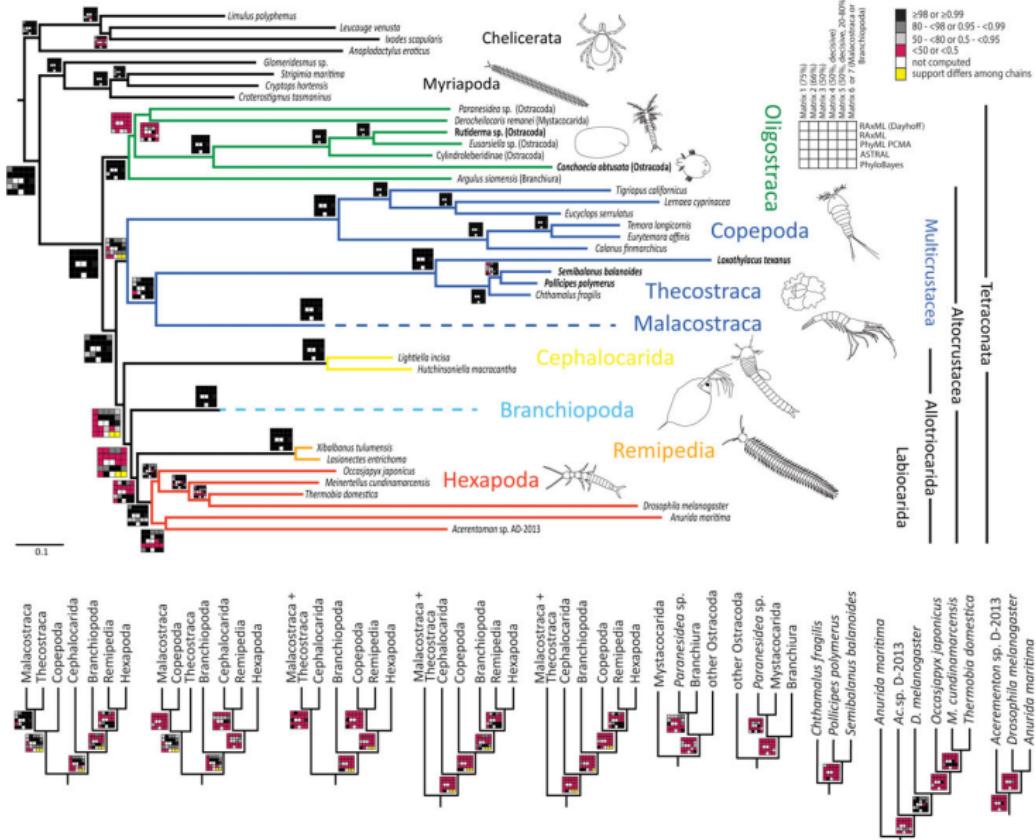
Alignment-free methods
oooo

Outro
o

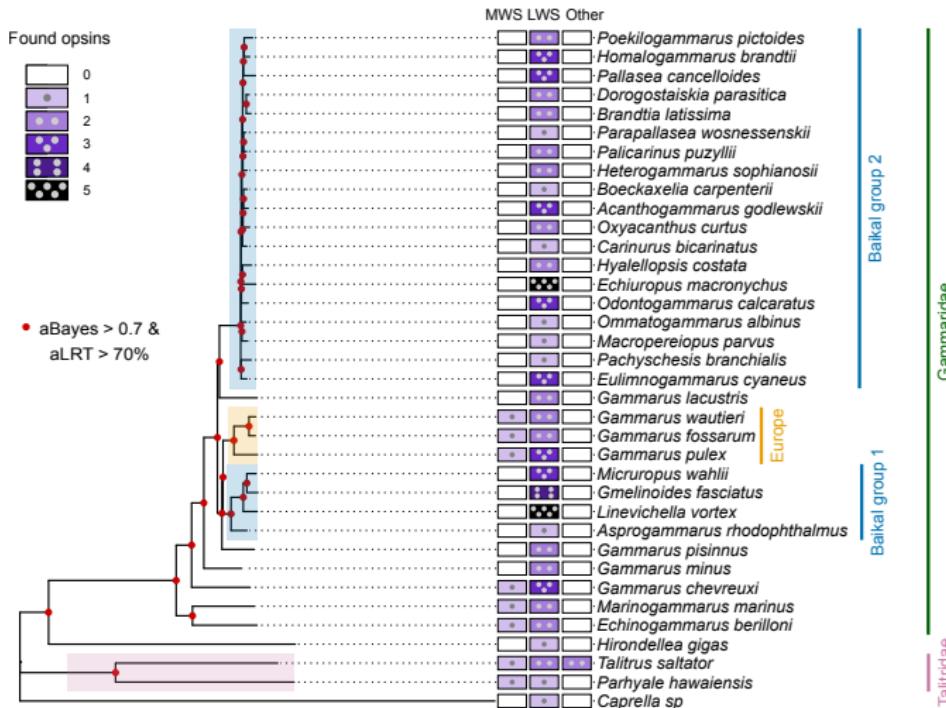
BuscoPhylo



Single copy orthologs: examples



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

Источник собранных геномов: Ensembl

www.ensembl.org/Felis_catus/Info/Index

Ensembl

BLAST/BLAT | VEP | Tools | BioMart | Downloads | Help & Docs | Blog

Search Cat...

Cat (Felis_catus_9.0) ▾

Search Cat (Felis catus)

Search all categories ▾ Search

e.g. PPP2R2A or B1:31889947-31974534 or rs43782442 or RDX

Go

Genome assembly: Felis_catus_9.0 (GCA_000181335.4)

- More Information and statistics
- Download DNA sequence (FASTA)
- Display your data in Ensembl

Other assemblies

Felis_catus_6.2 (Ensembl release 80) ▾ Go

Comparative genomics

What can I find? Homologues, gene trees, and whole genome alignments across multiple species.

More about comparative analysis

Example transcript

Pax6 INS FOXP2 DMD ssh

Example gene

Example region

Gene annotation

What can I find? Protein-coding and non-coding genes, splice variants, cDNA and protein sequences, non-coding RNAs.

- More about this genebuild
- Download FASTA files for genes, cDNAs, ncRNA, proteins
- Download GTF or GFF3 files for genes, cDNAs, ncRNA, proteins
- Update your old Ensembl IDs

Variation

What can I find? Short sequence variants; disease and other phenotypes

More about variation in Ensembl

Example variant

ATCGAGCT
ATCCAGCT
ATCGAGAT

Markers and phylogenetics
oooooooooooooooooooo

Phylogenomics intro
oooooooooooo

Single copy orthologs
oooooooo●

Reduced representation
oooooooooo

Alignment-free methods
oooo

Outro
o

Упражнение 3: филогения 12 видов млекопитающих по однокопийным ортологам

Markers and phylogenetics
oooooooooooooooooooo

Phylogenomics intro
oooooooooooo

Single copy orthologs
oooooooooooo

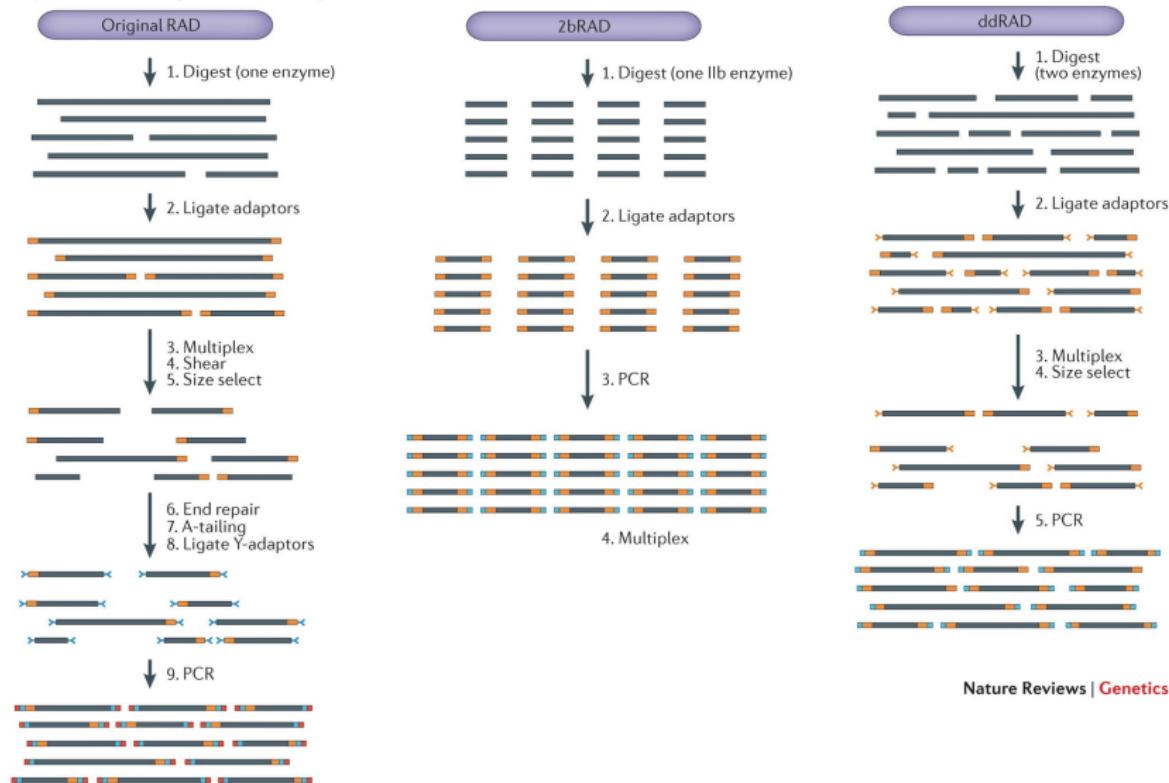
Reduced representation
●oooooooooooo

Alignment-free methods
oooo

Outro
○

RADseq

Sequence next to single restriction enzyme cut sites



Nature Reviews | Genetics

ddRADseq

OPEN  ACCESS Freely available online

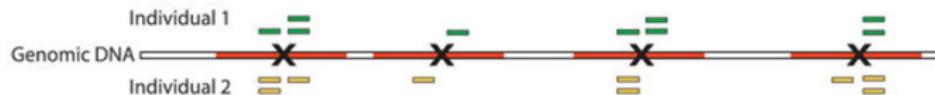
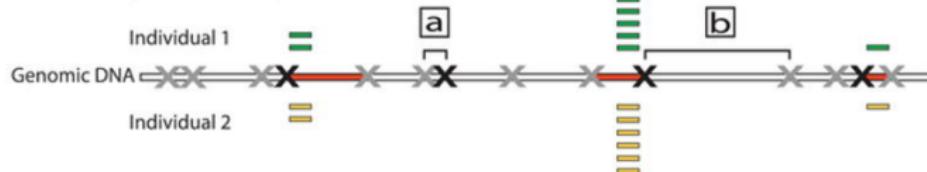
Double Digest RADseq: An Inexpensive Method for *De Novo* SNP Discovery and Genotyping in Model and Non-Model Species

Brant K. Peterson*, Jesse N. Weber, Emily H. Kay, Heidi S. Fisher, Hopi E. Hoekstra

Department of Organismic & Evolutionary Biology, Department of Molecular & Cellular Biology, Museum of Comparative Zoology, Harvard University, Cambridge, Massachusetts, United States of America

A**RAD sequencing**

X Rare cut site — Genomic interval present in library
X Common cut site — Sequence reads

**B****double digest RADseq**

Markers and phylogenetics
oooooooooooooooooooo

Phylogenomics intro
oooooooooooo

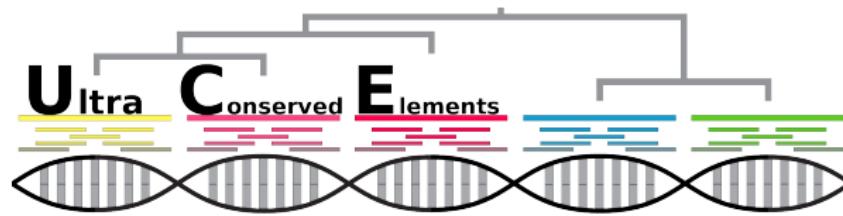
Single copy orthologs
oooooooooooo

Reduced representation
oo●oooooooo

Alignment-free methods
oooo

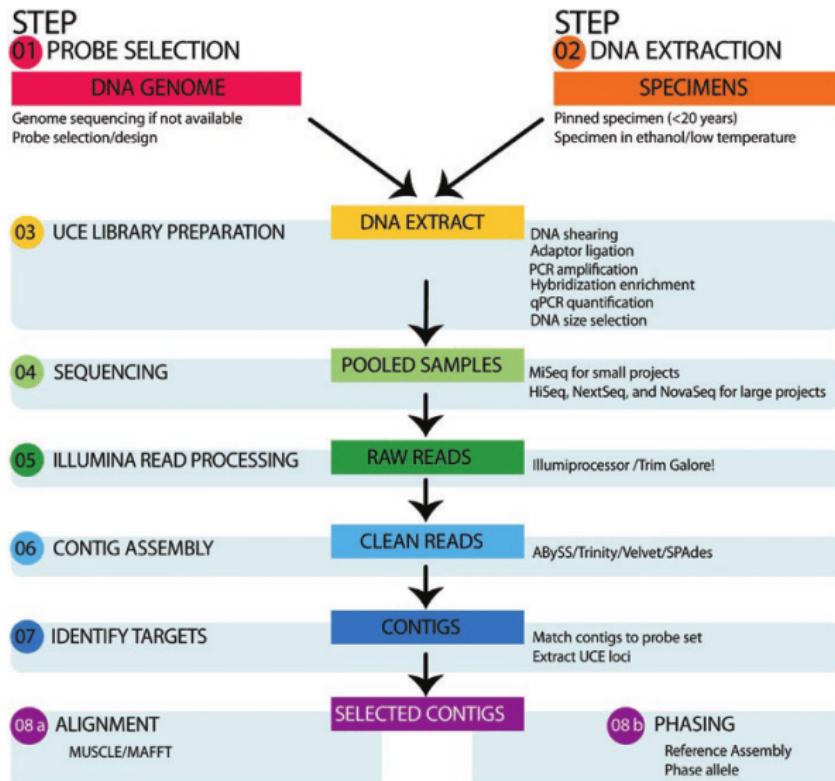
Outro
o

Ultraconserved elements = UCEs



<https://www.ultraconserved.org/>

Ultraconserved elements = UCEs



Markers and phylogenetics
oooooooooooooooo

Phylogenomics intro
oooooooooooo

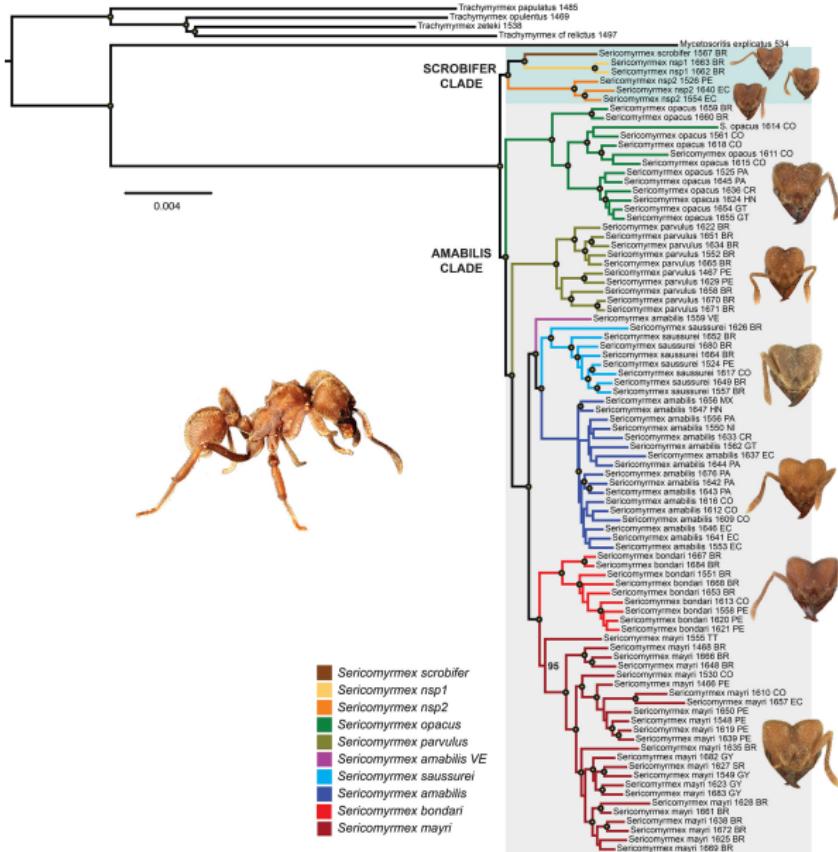
Single copy orthologs
oooooooooooo

Reduced representation
oooo●oooo

Alignment-free methods
oooo

Outgroups
○

Ultraconserved elements = UCEs (example)



Markers and phylogenetics
oooooooooooooooooooo

Phylogenomics intro
oooooooooooo

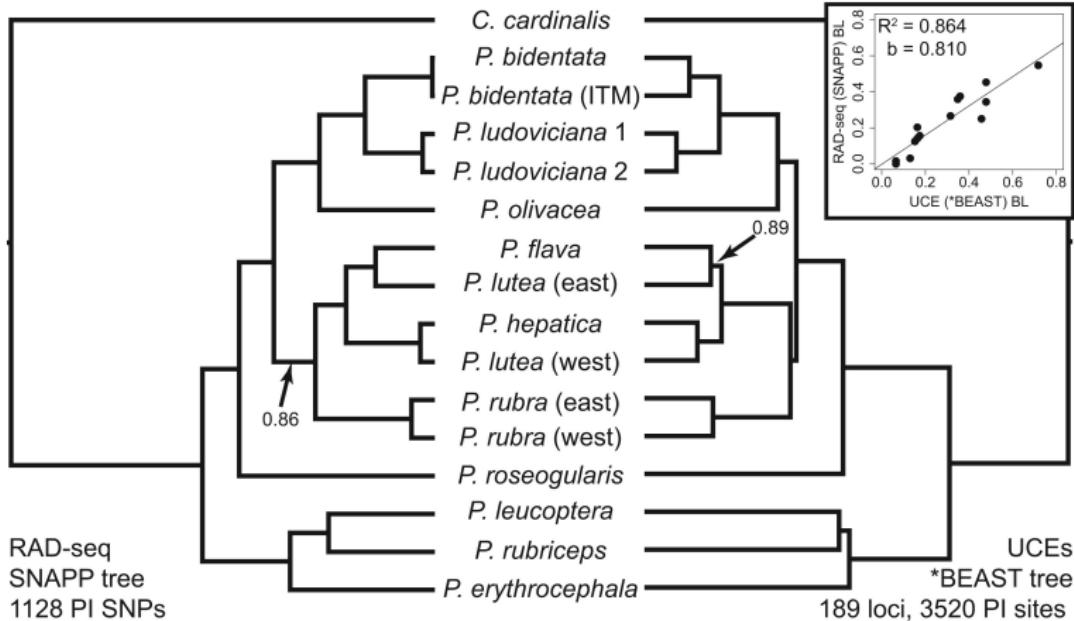
Single copy orthologs
oooooooooooo

Reduced representation
ooooo●oooo

Alignment-free methods
oooo

Outro
○

RAD-seq и UCE: пример



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

Markers and phylogenetics
oooooooooooooooo

Phylogenomics intro
oooooooooooo

Single copy orthologs
oooooooooooo

Reduced representation
oooooooo●○○

Alignment-free methods
oooo

Outo
○

Recommended!

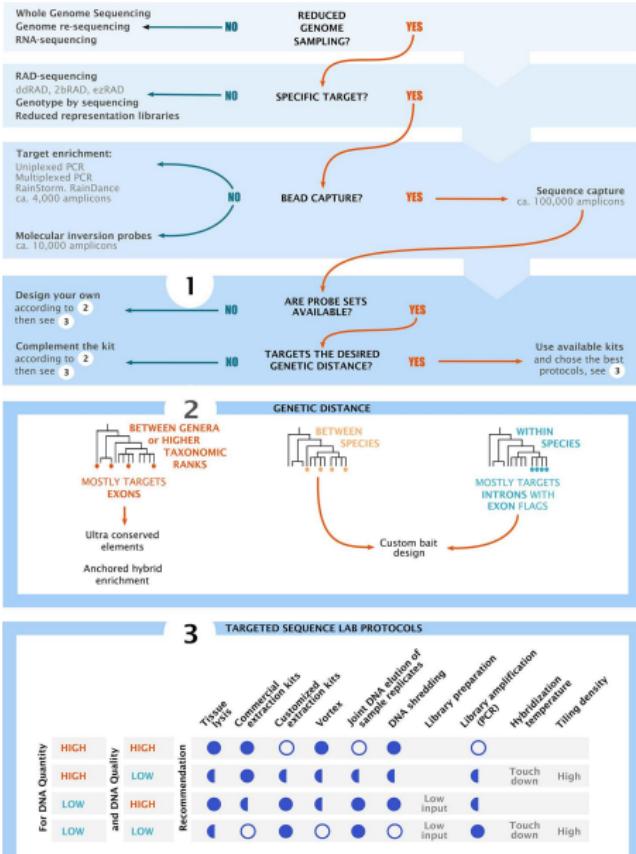


REVIEW
published: 21 February 2020
doi: 10.3389/fgene.2019.01407

A Guide to Carrying Out a Phylogenomic Target Sequence Capture Project

Tobias Andermann^{1,2*}, María Fernanda Torres Jiménez^{1,2*}, Pável Matos-Maraví^{1,3}, Romina Batista^{2,4,5}, José L. Blanco-Pastor^{1,6}, A. Lovisa S. Gustafsson⁷, Logan Kistler⁸, Isabel M. Liberal¹, Bengt Oxelman^{1,2}, Christine D. Bacon^{1,2} and Alexandre Antonelli^{1,2,9}

[https://doi.org/10.3389/fgene.](https://doi.org/10.3389/fgene.2019.01407)
2019.01407



Markers and phylogenetics
oooooooooooooooooooo

Phylogenomics intro
oooooooooooo

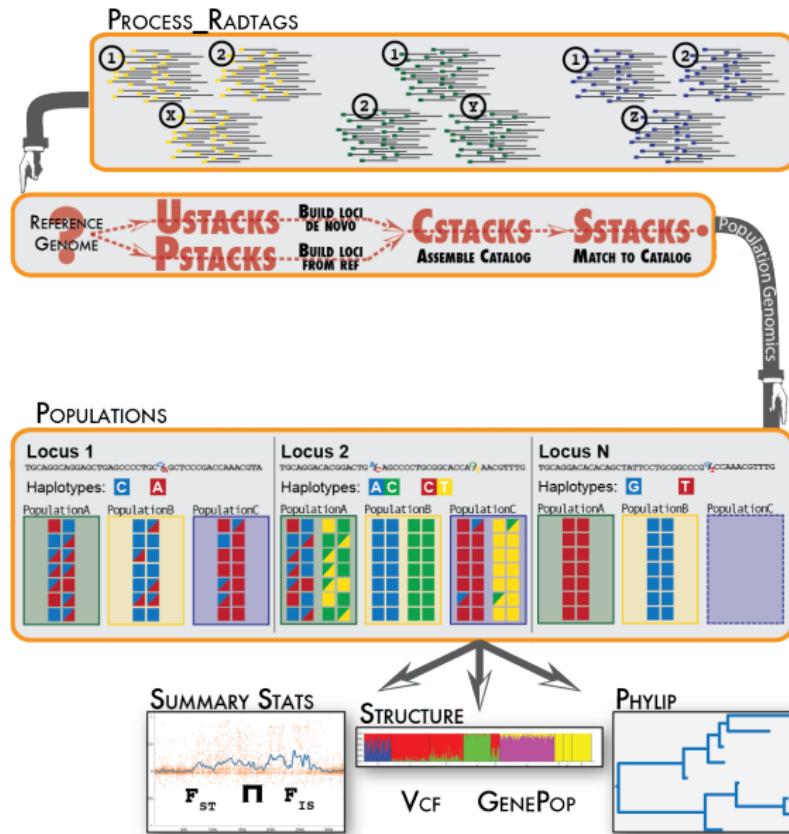
Single copy orthologs
oooooooooooo

Reduced representation
oooooooo●○

Alignment-free methods
oooo

Outro
○

RADseq: анализ



Markers and phylogenetics
oooooooooooooooooooo

Phylogenomics intro
oooooooooooo

Single copy orthologs
oooooooooooo

Reduced representation
oooooooo●●

Alignment-free methods
oooo

Outro
o

Упражнение №5: обработка данных RADseq для трёхиглой колюшки

Источники данных:

- ✓ [https://journals.plos.org/plosgenetics/article?id=10.1371/](https://journals.plos.org/plosgenetics/article?id=10.1371)
- ✓ <https://training.galaxyproject.org/training-material/topics/ecology/tutorials/de-novo-rad-seq/tutorial.html>



КОЛЮШКА И КОРЮШКА
ПЛЫВУТ В ЯПОНИЮ

Markers and phylogenetics
oooooooooooooooo

Phylogenomics intro
oooooooooooo

Single copy orthologs
oooooooooooo

Reduced representation
oooooooooooo

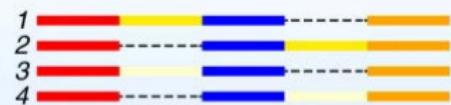
Alignment-free methods
●○○○

Outro
○

Alignment-free methods

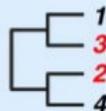


A classical approach



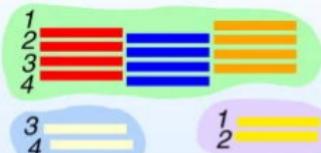
multiple sequence
alignment

phylogenetic
inference



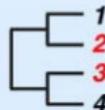
phylogenetic
tree

B alternative approach



alignment-free method

phylogenetic
inference



Markers and phylogenetics
○○○○○○○○○○○○○○○○

Phylogenomics intro
○○○○○○○○○○○○○○○○

Single copy orthologs
○○○○○○○○○○○○○○○○

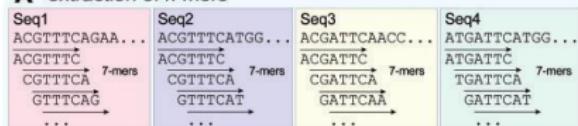
Reduced representation
○○○○○○○○○○○○○○○○

Alignment-free methods
○●○○

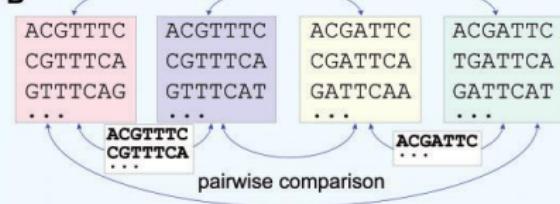
Outro
○

Alignment-free methods

A extraction of k -mers



B



pairwise comparison

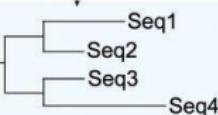
C

	Seq1	Seq2	Seq3	Seq4
Seq1	0.0			
Seq2	0.3	0.0		
Seq3	0.5	0.4	0.0	
Seq4	0.7	0.6	0.4	0.0

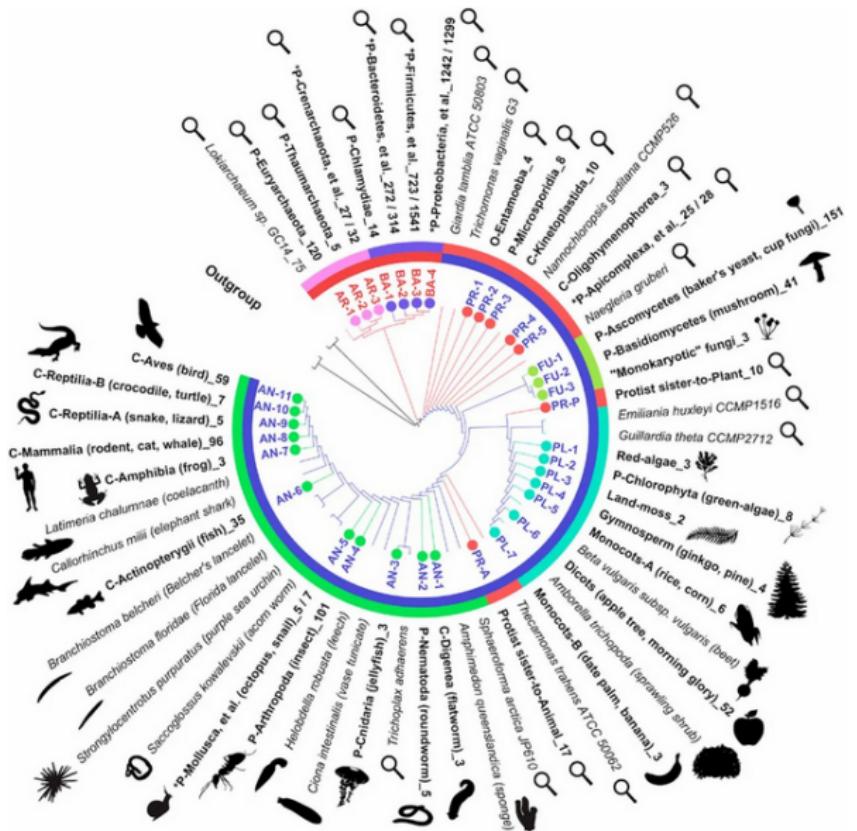
pairwise
distance
matrix

D

phylogenetic
tree



Alignment-free methods



Markers and phylogenetics
oooooooooooooooooooo

Phylogenomics intro
oooooooooooo

Single copy orthologs
oooooooooooo

Reduced representation
oooooooooooo

Alignment-free methods
ooo●

Outro
o

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

FFP <https://github.com/jaejinchoi/FFP/tree/master>

Интегративная таксономия

A

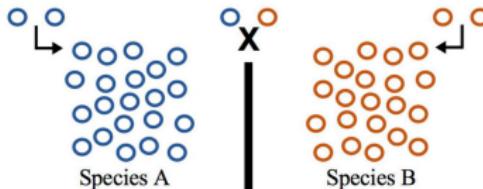
Phylogenetic



- Species are distinguishable based on shared, derived traits
- Individuals of the same species can be grouped based on derived traits from a common ancestor

B

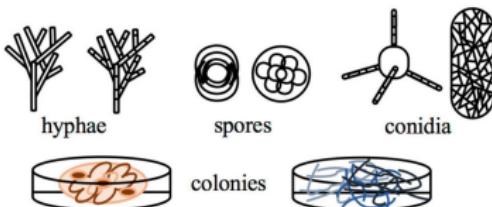
Biological



- Species are distinguishable via reproductive isolation
- Interbreeding of individuals in the same species produces viable offspring

C

Morphological



- Species are distinguishable based on features and appearance
- Individuals within a species have a shared suite of distinct morphologies and characteristics

D

Ecological



- Species are distinguishable via niche partitioning
- Individuals of the same species share preferences and tolerances for biotic and abiotic conditions