

Лекция 3

Общие принципы построения филогений

•

ДНК:

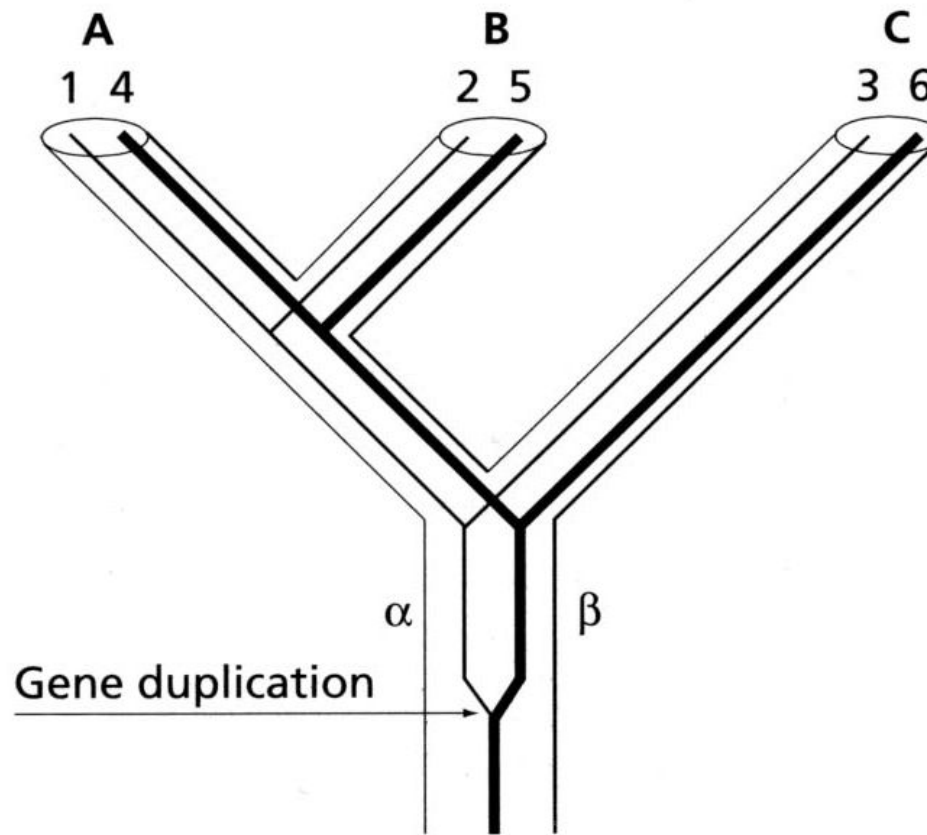
1 5 10
tagcaaaatg

Какие гены и какие нуклеотиды гомологичны?

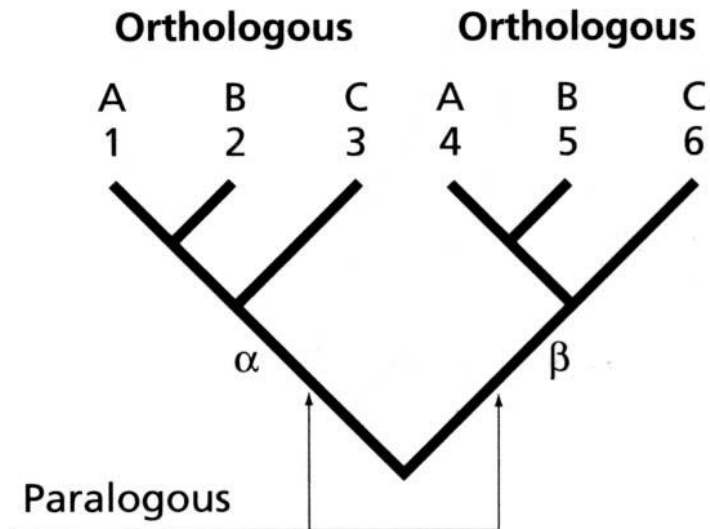
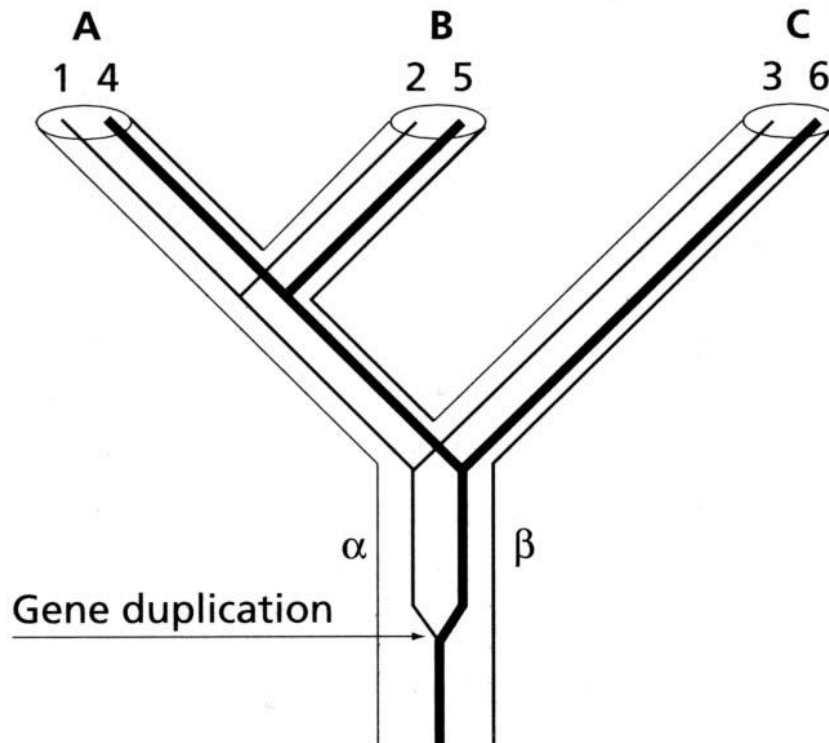
	1	10	20	30																												
AY557140 A achaemenes voucher	A	T	A	A	C	A	T	A	A	G	A	T	T	C	T	G	A	T	T	A	C	T	A	C	C	C	C	A	T	C	A	
EF104615 A achaemenes voucher	A	T	A	A	T	A	T	A	A	G	A	T	T	G	T	G	A	-	T	A	C	T	A	C	C	C	C	A	T	C	A	
EF104621 A actinides voucher V	A	T	A	A	C	A	T	A	A	G	A	T	T	C	T	G	A	T	T	A	T	T	A	C	C	A	C	C	A	T	C	A
LOWA513-06 2005-LOWA-513 Agrod	A	T	A	A	C	G	T	A	A	G	A	T	T	C	T	G	A	T	T	A	T	T	A	C	C	A	C	C	A	T	C	A
LOWA503-06 2005-LOWA-503 Agrod	A	T	A	A	C	A	T	A	A	G	A	T	T	C	T	G	A	T	T	A	T	T	A	C	C	A	C	C	A	T	C	A
LOWA502-06 2005-LOWA-502 Agrod	A	T	A	A	C	A	A	A	A	G	A	T	T	C	T	G	A	T	T	A	T	T	A	C	C	A	C	C	A	T	C	A
LOWA501-06 2005-LOWA-501 Agrod	A	T	A	A	C	A	T	A	A	G	A	T	T	C	T	G	A	T	T	A	T	T	A	C	C	A	C	C	A	T	C	A
LOWA512-06 2005-LOWA-512 Agrod	A	T	A	A	C	A	T	A	A	G	A	T	T	C	T	G	A	T	T	A	T	T	A	C	C	A	C	C	A	T	C	A
LOWA511-06 2005-LOWA-511 Agrod	A	T	A	A	C	A	T	A	A	G	A	T	T	C	T	G	A	T	T	A	T	T	A	C	C	A	C	C	A	T	C	A

Дублицированные гены негомологичны!

Они не получены в ходе генеалогической передачи признаков! Они занимают разные локусы и каждый имеет свою эволюционную судьбу. Это РАЗНЫЕ признаки!



В результате дупликаций возникают пары похожих, но НЕ ГОМОЛОГИЧНЫХ генов, которые обретают собственную судьбу и эволюционируют независимо. Смешение настоящих гомологичных (=ортологичных) и негомологичных (паралогичных) ведет к ошибочной реконструкции филогенеза



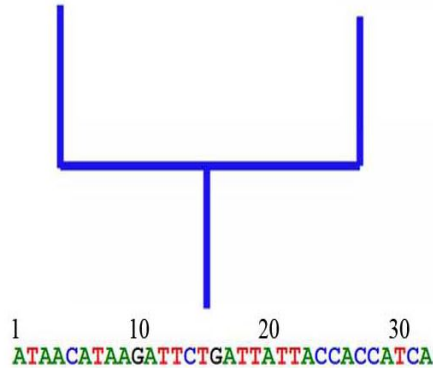
Гомология – происхождение от общего генеалогического предка.
 Признак, возникший вне генеалогического ряда, не является гомологией, несмотря даже на генетическую идентичность

ген A

ген A¹

1 10 20 30
ATAACATAAGATTCTGATTATTACCACCATCA

1 10 20 30
ATAACATAAGATTCTGATTATTACCACCATCA



ген A

Гомология генов

Гены A и A¹ гомологичны
В данном случае они идентичны по структуре и положению в геноме, и, самое главное, происходят от общего предка

Гомология нуклеотидов – это позиционная гомология

1 10 20 30
ATAACATAAGATTCTGATTATTACCACCATCA
ATAACATAAGATTCTGATTATTACCACCATCA
1 10 20 30

ген A

ген A¹

Сайт 1 гена A гомологичен сайту 1 гена A¹;
Сайт 2 гена A гомологичен сайту 2 гена A¹

.....

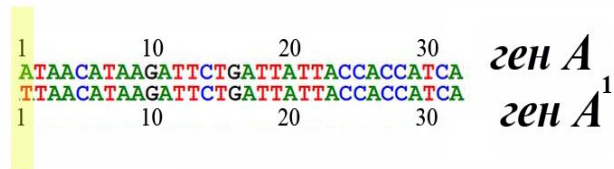
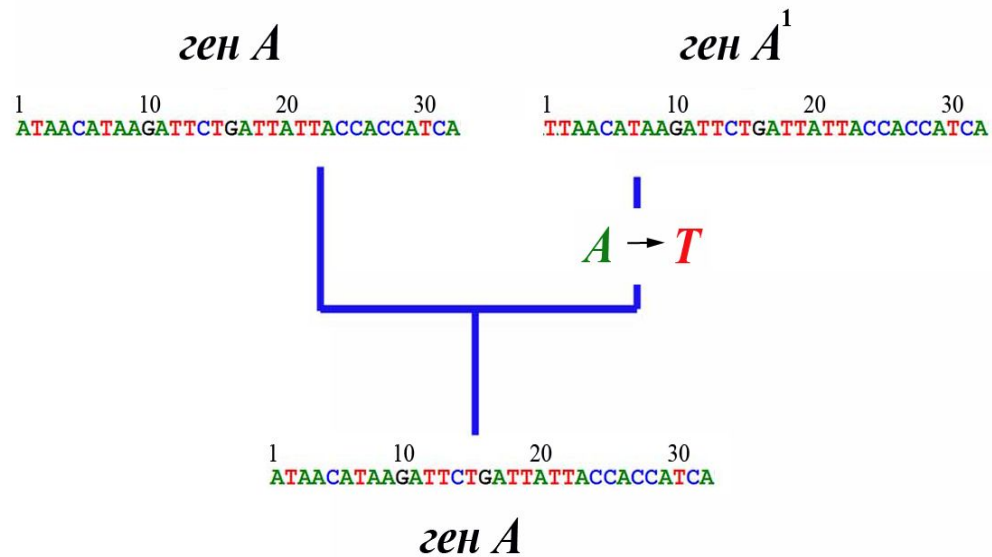
Сайт 10 гена A гомологичен сайту 10 гена A¹

Гомология – это не то же самое, что идентичность или сходство генов

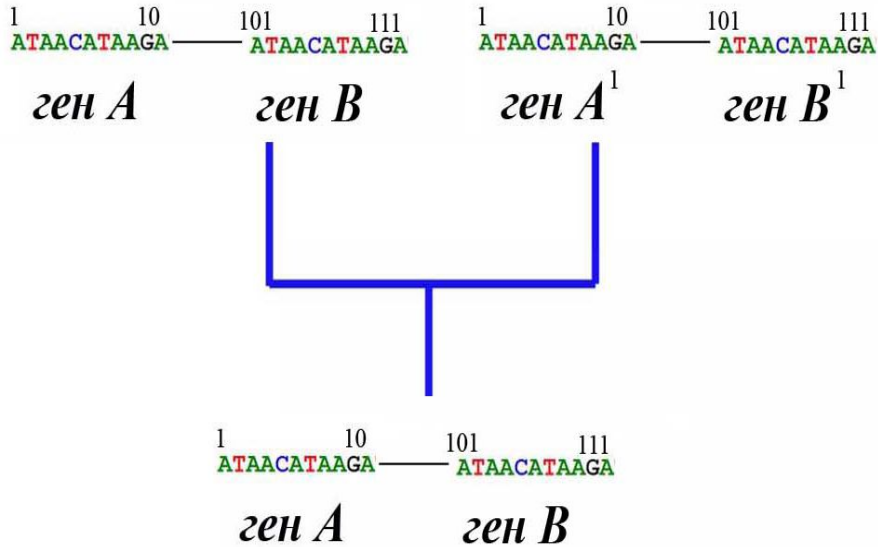
Мутации создают новые состояния признаков, не меняя их гомологичность

Сайт 1 гена A гомологичен сайту 1 гена A^1 .

но они не идентичны



Гомология нуклеотидных последовательностей ("генов")



Гены А и А¹ гомологичны
 В данном случае они идентичны по структуре и положению в геноме

Гены А и В¹ не гомологичны,
 Они имеют разное положение в геноме, но при этом они идентичны по структуре

Гомология нуклеотидов – это позиционная гомология

Сайт 1 гена А гомологичен сайту 1 гена А¹:

НО

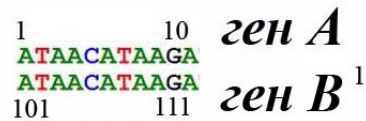
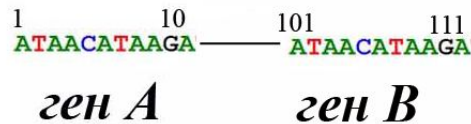
Сайт 1 гена А не гомологичен сайту 101 гена В¹

1 10 ген А
 АТААСАТААГА
 АТААСАТААГА ген А¹
 1 10

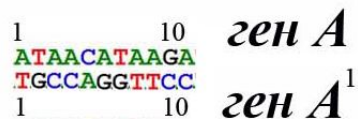
1 10 ген А
 АТААСАТААГА
 АТААСАТААГА ген В¹
 101 111



Бывают ли “проценты гомологии”?



Гены А и В¹:
 Гомология отсутствует,
 Сходство = 100%



Гены А и А¹:
 Гомология присутствует,
 Сходство = 0%

	1	10	20	30																												
AY557140 A achaemenes voucher	A	T	A	A	C	A	T	A	A	G	A	T	T	C	T	G	A	T	T	A	C	T	A	C	C	C	C	C	A	T	C	A
EF104615 A achaemenes voucher	A	T	A	A	T	A	T	A	A	G	A	T	T	G	T	G	A	-	T	A	C	T	A	C	C	C	C	C	A	T	C	A
EF104621 A actinides voucher V	A	T	A	A	C	A	T	A	A	G	A	T	T	C	T	G	A	T	T	A	T	T	A	C	C	A	C	C	A	T	C	A
LOWA513-06 2005-LOWA-513 Agrod	A	T	A	A	C	G	T	A	A	G	A	T	T	C	T	G	A	T	T	A	T	T	A	C	C	A	C	C	A	T	C	A
LOWA503-06 2005-LOWA-503 Agrod	A	T	A	A	C	A	T	A	A	G	A	T	T	C	T	G	A	T	T	A	T	T	A	C	C	A	C	C	A	T	C	A
LOWA502-06 2005-LOWA-502 Agrod	A	T	A	A	C	A	A	A	A	G	A	T	T	C	T	G	A	T	T	A	T	T	A	C	C	A	C	C	A	T	C	A
LOWA501-06 2005-LOWA-501 Agrod	A	T	A	A	C	A	T	A	A	G	A	T	T	C	T	G	A	T	T	A	T	T	A	C	C	A	C	C	A	T	C	A
LOWA512-06 2005-LOWA-512 Agrod	A	T	A	A	C	A	T	A	A	G	A	T	T	C	T	G	A	T	T	A	T	T	A	C	C	A	C	C	A	T	C	A
LOWA511-06 2005-LOWA-511 Agrod	A	T	A	A	C	A	T	A	A	G	A	T	T	C	T	G	A	T	T	A	T	T	A	C	C	A	C	C	A	T	C	A

Сравнение правомочно, если

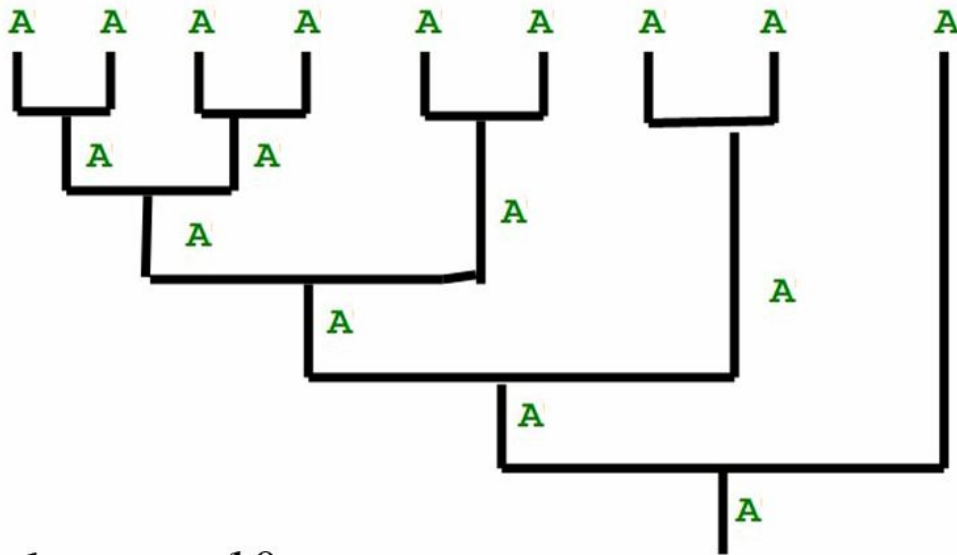
а) гены ортологичны (=гомологичны, не паралогичны),

б) сравниваются нуклеотиды, занимающие один и тот же сайт

ТО ЕСТЬ, ЕСЛИ ГОМОЛОГИЗАЦИЯ ПРОВЕДЕНА ПРАВИЛЬНО

Какие признаки пригодны для филогенетического анализа?

- **Только наследуемые:** молекулярные признаки - **да!**
- Только гомологичные! Но как выявить гомологию молекулярных признаков?
- Не гомоплазии (гомоплазии, несмотря на сходство, не несут информации о филогенетическом родстве)

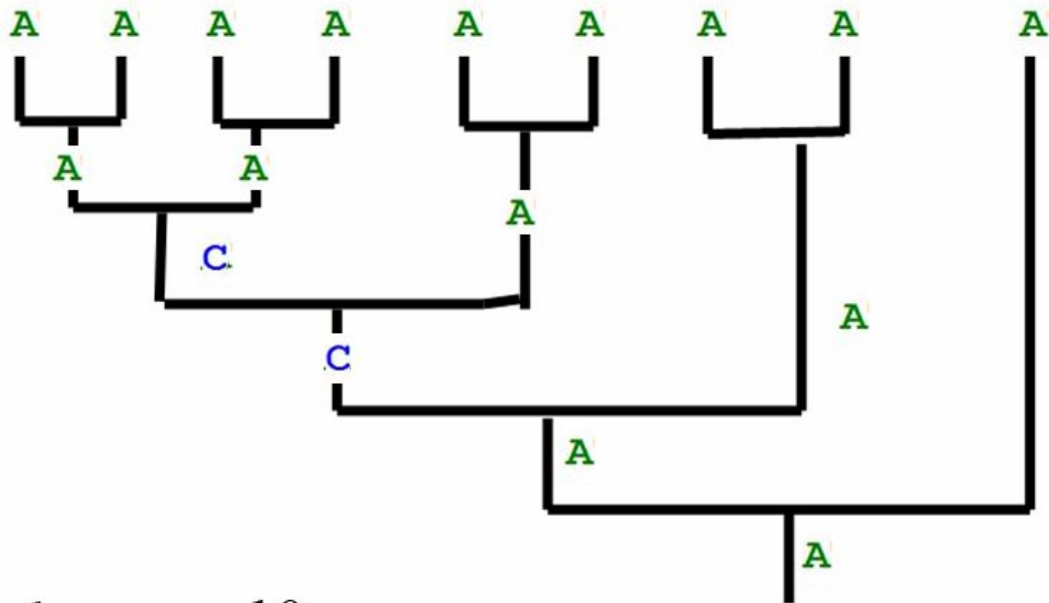


Признак А в первом сайте является гомологией (не является Гомоплазией!!!), так как он непосредственно унаследован от общего предка

1 10

ATAACATAAGATTCTGATTACTACCCCATCA
ATAATATAAGATTGTGA-TACTACCCCATCA
ATAACATAAGATTCTGATTATTACCACCATCA
ATAACGTAAGATTCTGATTATTACCACCATCA
ATAACATAAGATTCTGATTATTACCACCATCA
ATAACAAAAGATTCTGATTATTACCACCATCA
ATAACATAAGATTCTGATTATTACCACCATCA
ATAACATAAGATTCTGATTATTACCACCATCA
ATAACATAAGATTCTGATTATTACCACCATCA

Эволюция нуклеотидов в позиции № 1



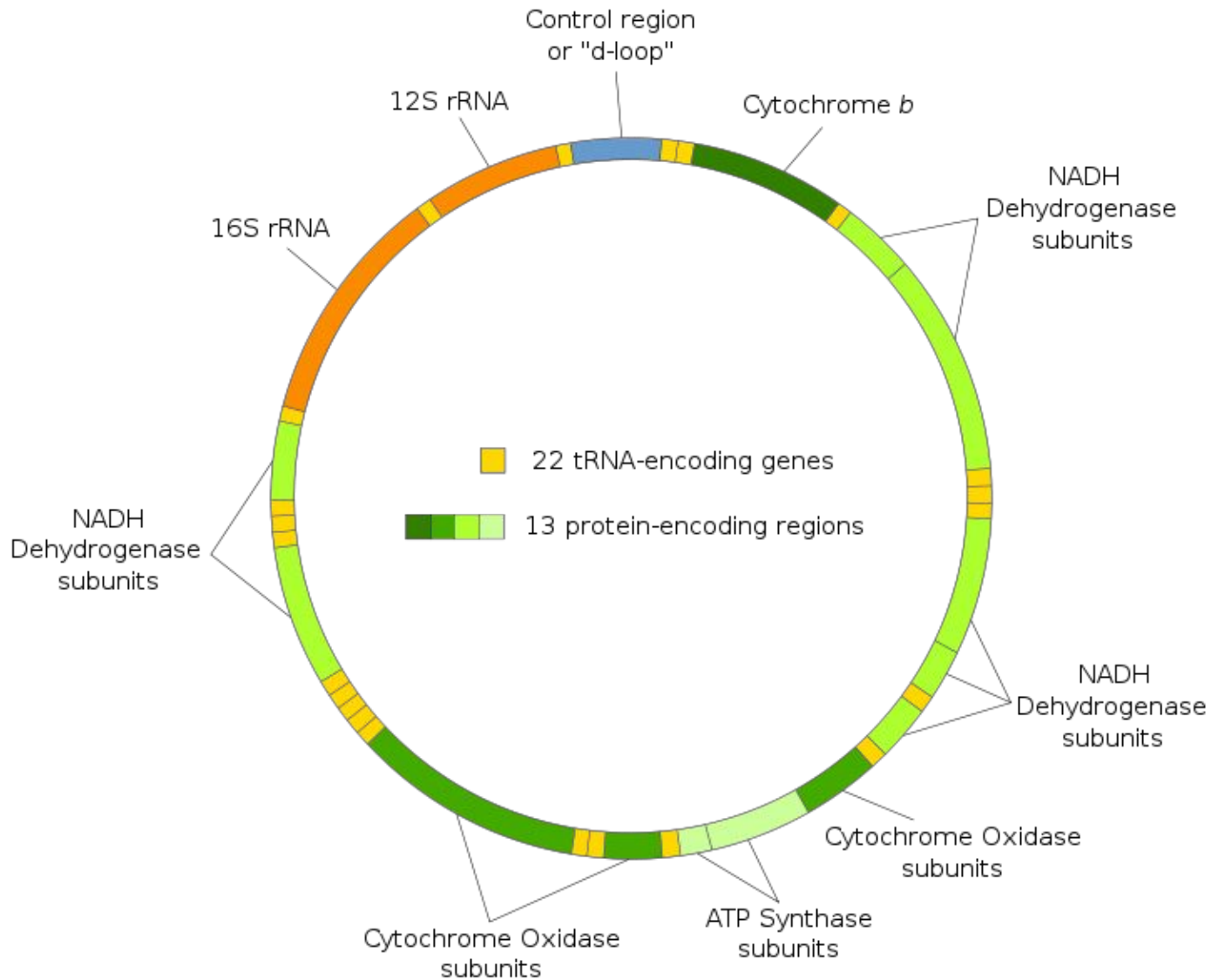
Совпадения буквы А в первой позиции являются следствием реверсии (=гомоплазии)

1 10

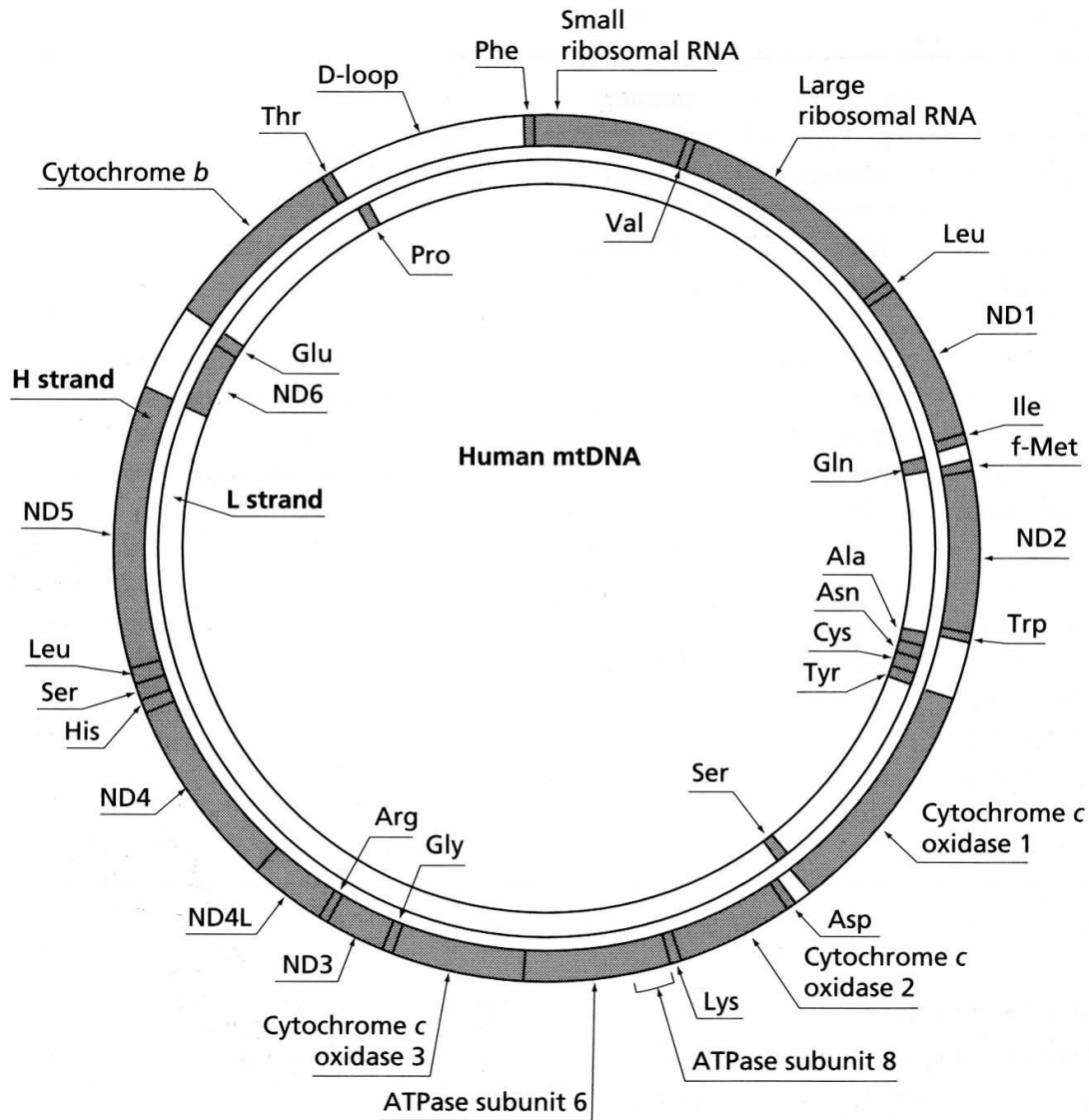
ATAACATAAGATTCTGATTACTACCCCATCA
 ATAATATAAGATTGTGA-TACTACCCCATCA
 ATAACATAAGATTCTGATTATTACCACCATCA
 ATAACGTAAGATTCTGATTATTACCACCATCA
 ATAACATAAGATTCTGATTATTACCACCATCA
 ATAACAAAAGATTCTGATTATTACCACCATCA
 ATAACATAAGATTCTGATTATTACCACCATCA
 ATAACATAAGATTCTGATTATTACCACCATCA
 ATAACATAAGATTCTGATTATTACCACCATCA
 ATAACATAAGATTCTGATTATTACCACCATCA

Гомология признаков и гомология состояний признаков: изучаемый признак (позиция 1) гомологичен для всех сиквенсов, в состояние А гомологично не для всех

Митохондриальная хромосома



Митохондриальная ДНК



Гаплоидность

По материнской линии

Нет рекомбинации

Высокая скорость нуклеотидных замен

Малое время коалесценции

Свой генетический код

Множество копий в клетке

Транскрипция в две стороны

Key transitions in animal evolution: a mitochondrial DNA perspective

Integrative and Comparative Biology, volume 47, number 5, pp. 734–743
doi:10.1093/icb/icm045

Dennis V. Lavrov

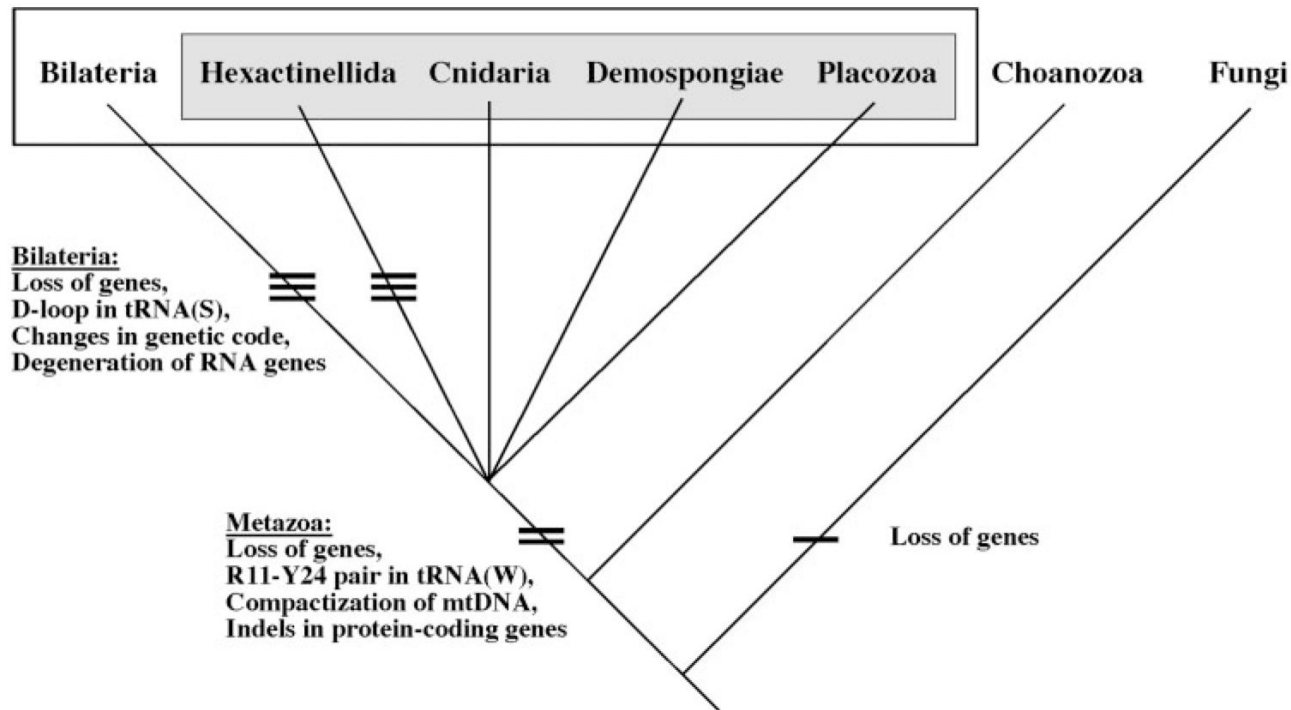
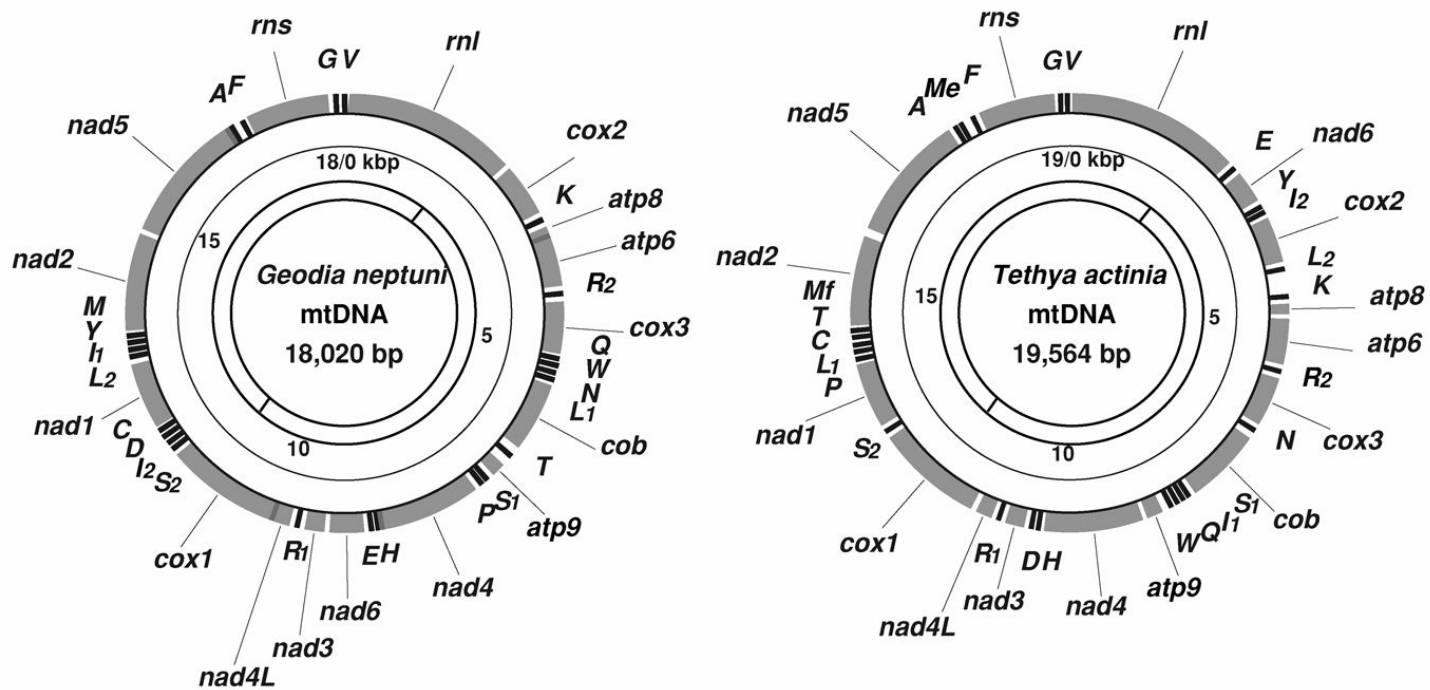


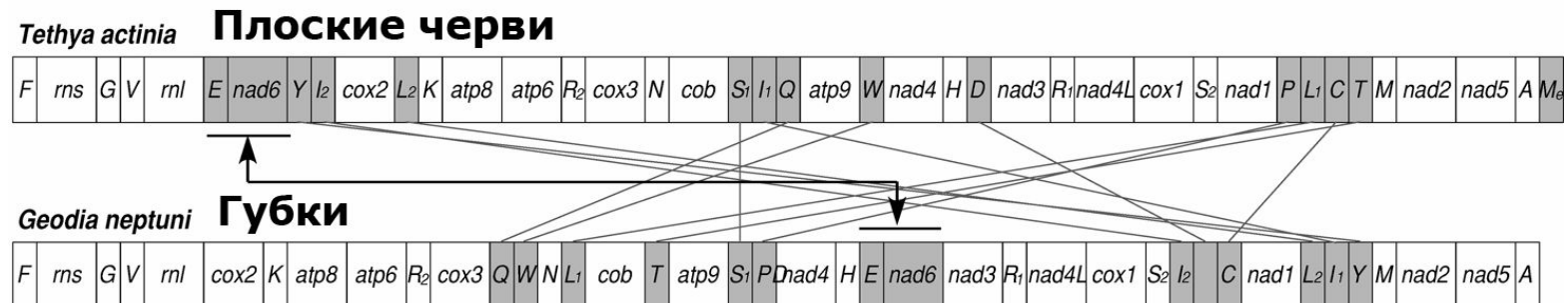
Fig. 2 mtDNA evolution in the Metazoa. Phylogenetic relationships among bilaterian and nonbilaterian animals are currently unresolved and are represented here by a polytomy (empty box). The present review is mostly limited to mtDNA in “nonbilaterian” animals (light-gray box). Only the taxa for which mtDNA data are available are shown.

Генные карты (A) и схема реорганизаций положения генов (B) для митохондриальных геномов *Geodia neptuni* и *Tethya actinia* (по: Lavrov, Lang, Syst. Biol. 54:551-559 (2005))

A

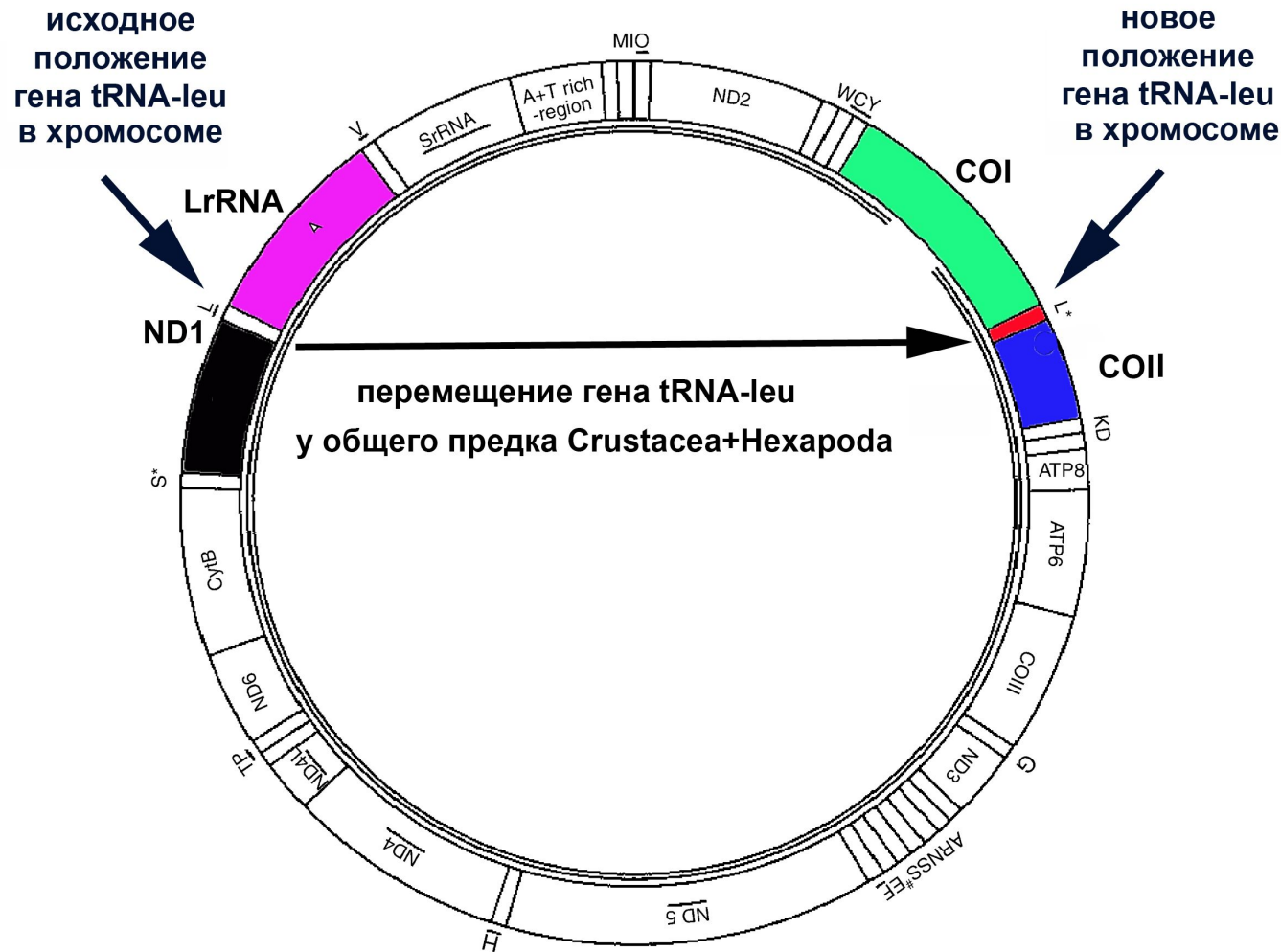


B



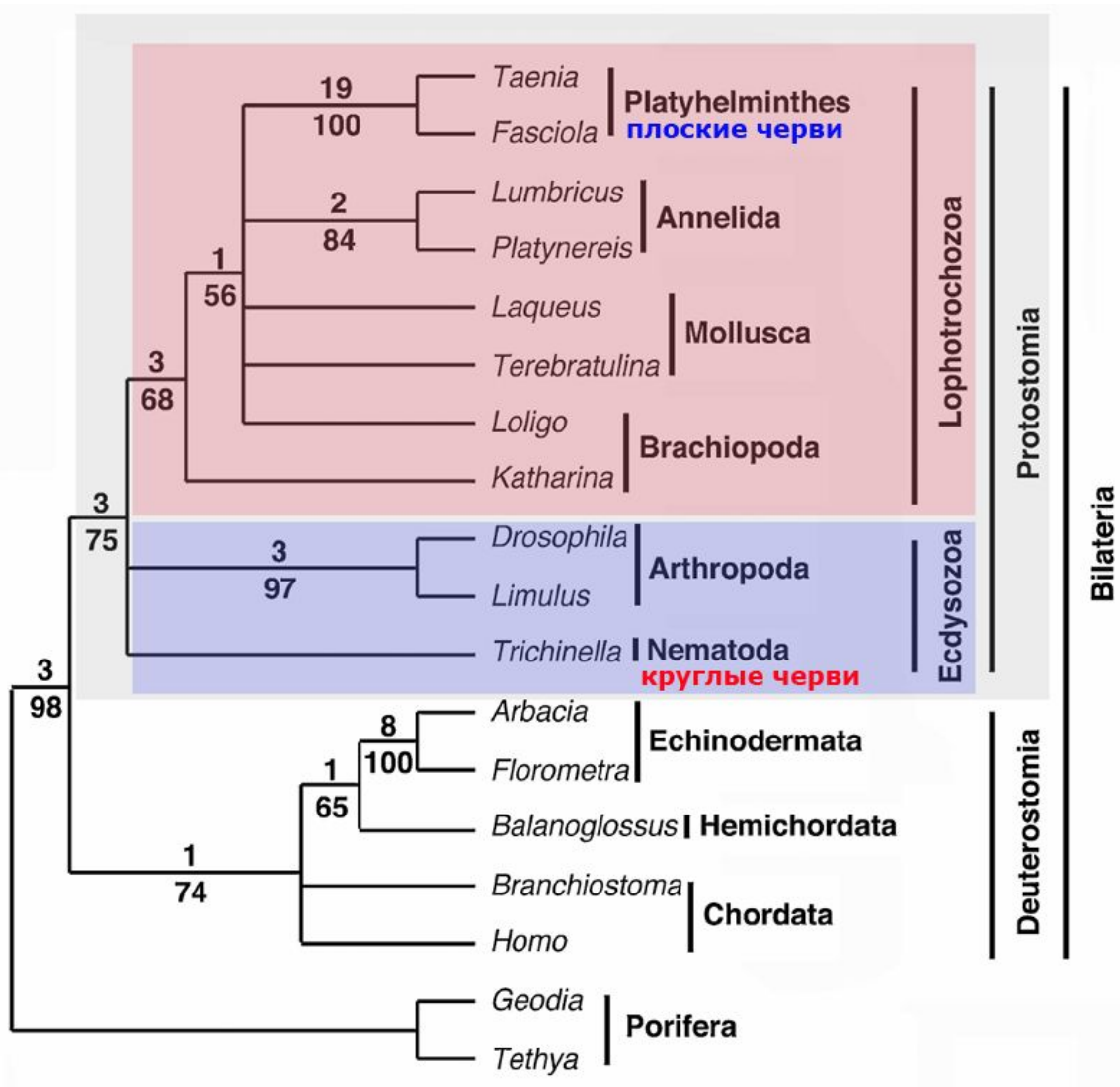
У общего предка Hexapoda и Crustacea произошла перестройка митохондриальной хромосомы

(Boore et al. 1998. Gene translocation links insects and crustaceans. *Nature* 392:667-668)



Генная карта
кольцевой
митохондриальной
хромосомы
насекомых
и ракообразных
(по: Kim et al., 2006.
*Insect Molecular
Biology*, с изменениями)

Филогения основных групп билатеральных животных, основанная на анализе хромосомных перестроек митохондриального генома (Lavrov, Lang, 2005, Systematic Biology)

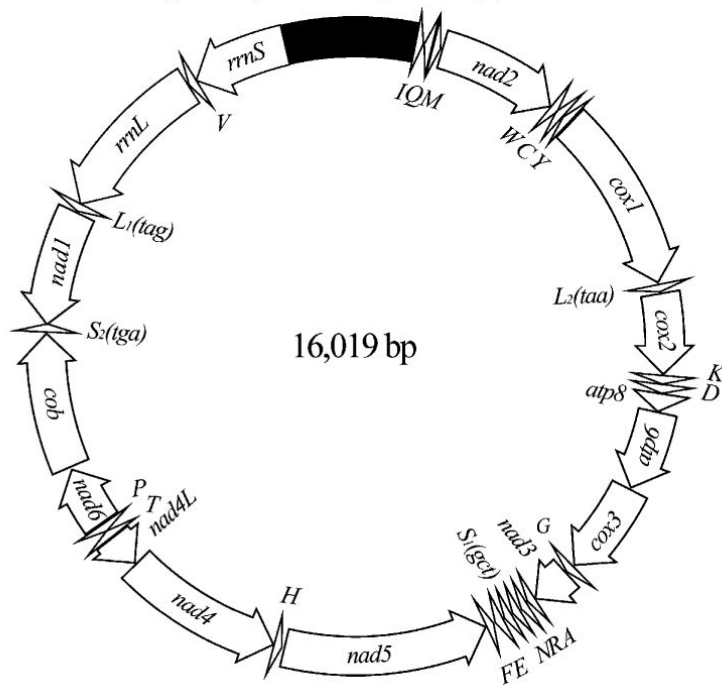


The single mitochondrial chromosome typical of animals has evolved into 18 minichromosomes in the human body louse, *Pediculus humanus*

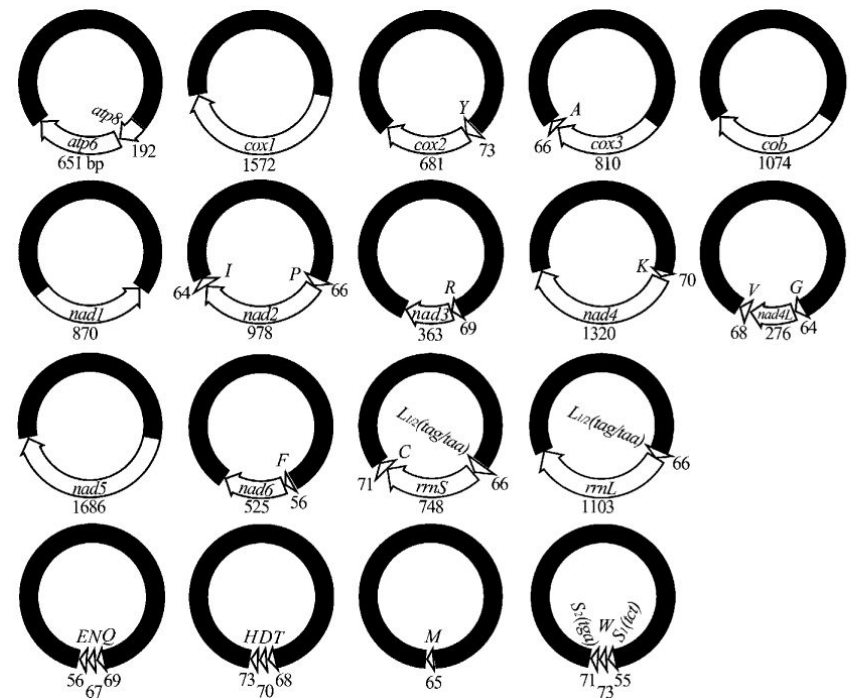
Renfu Shao, Ewen F. Kirkness and Stephen C. Barker

Genome Res. 2009 19: 904-912 originally published online March 31, 2009

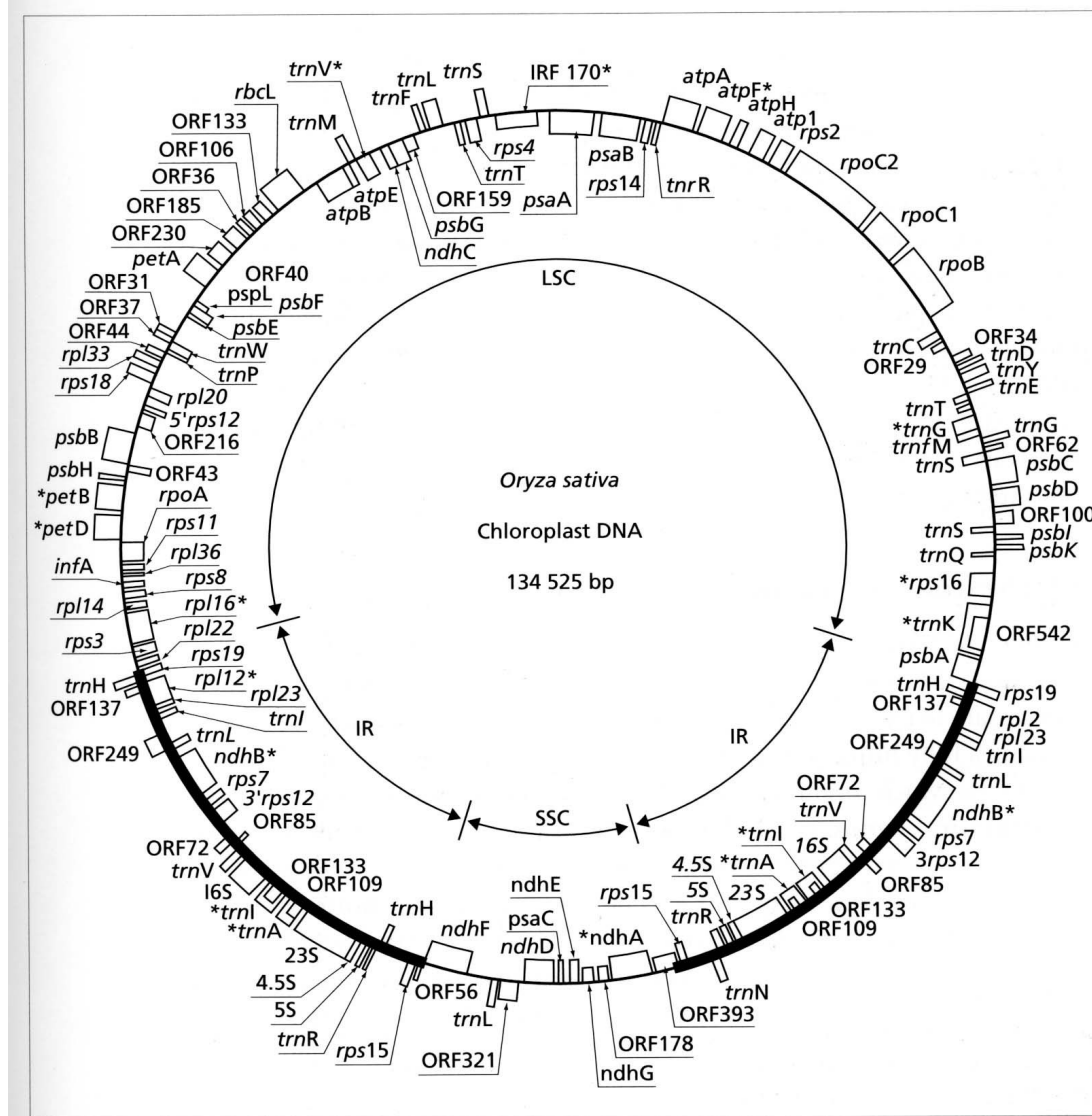
A *Drosophila yakuba* (Fruitfly)



B *Pediculus humanus* (Human body louse)



Хлоропластная ДНК



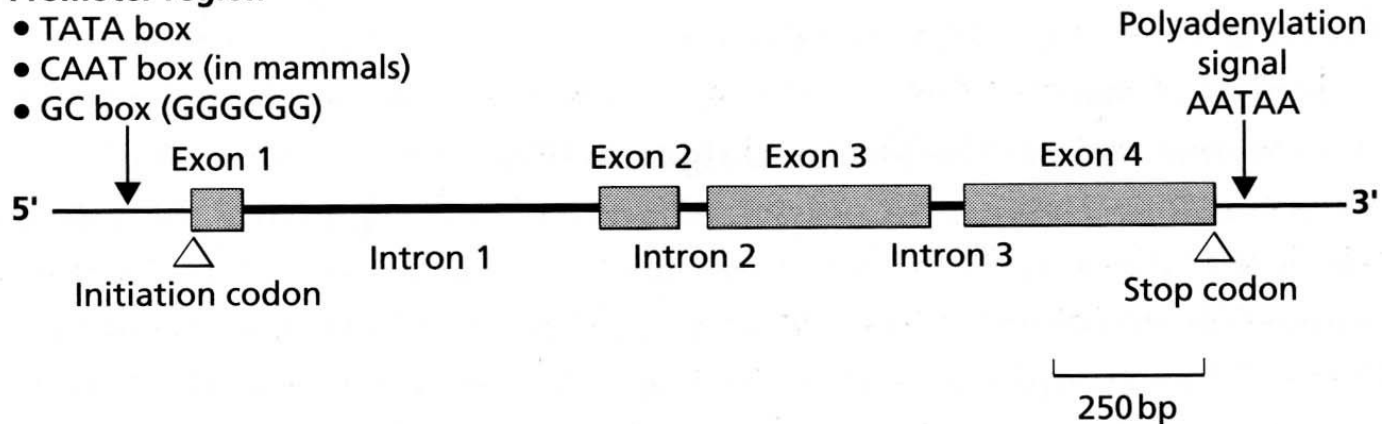
The chloroplast genome (cpDNA) of rice (*Oryza sativa*).

Ядерная ДНК эукариот и ДНК прокариот

Eukaryote (*Drosophila melanogaster*)

Promoter region

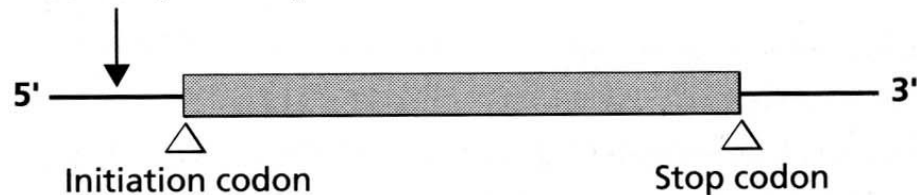
- TATA box
- CAAT box (in mammals)
- GC box (GGGCGG)



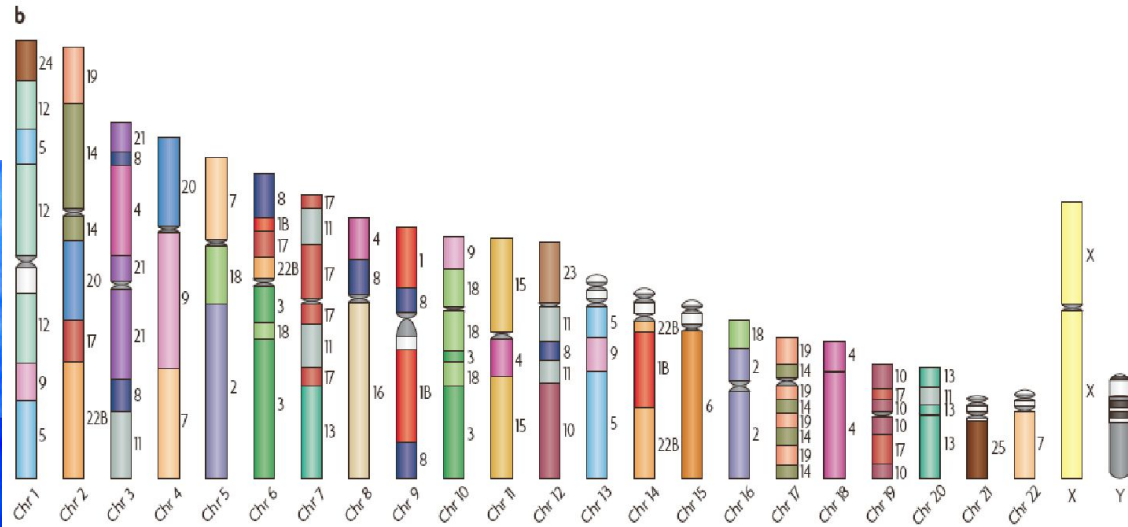
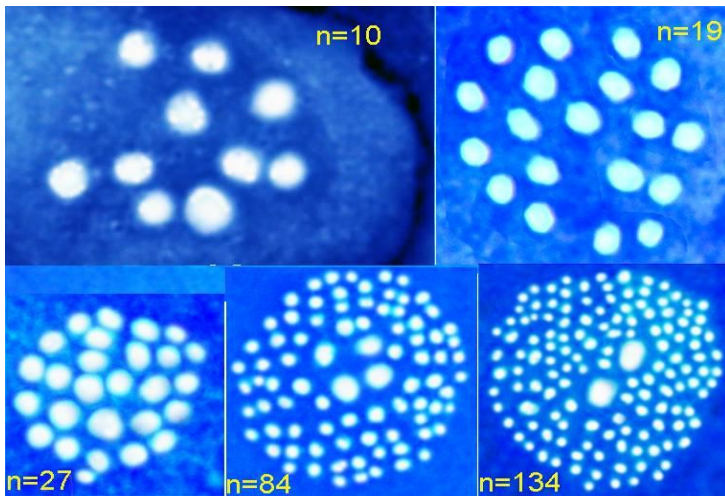
Bacteria (*Bacillus stearothermophilus*)

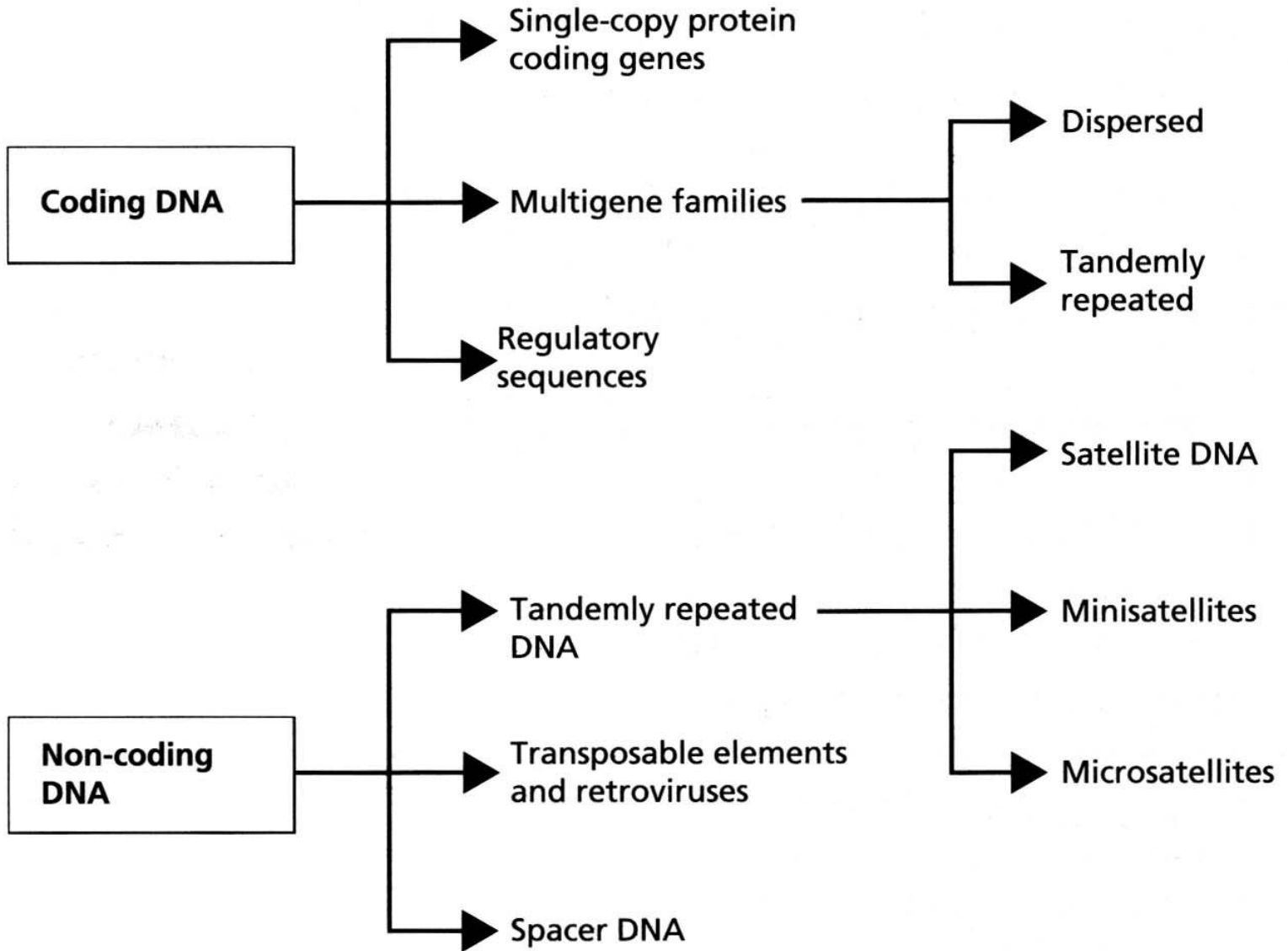
Promoter region

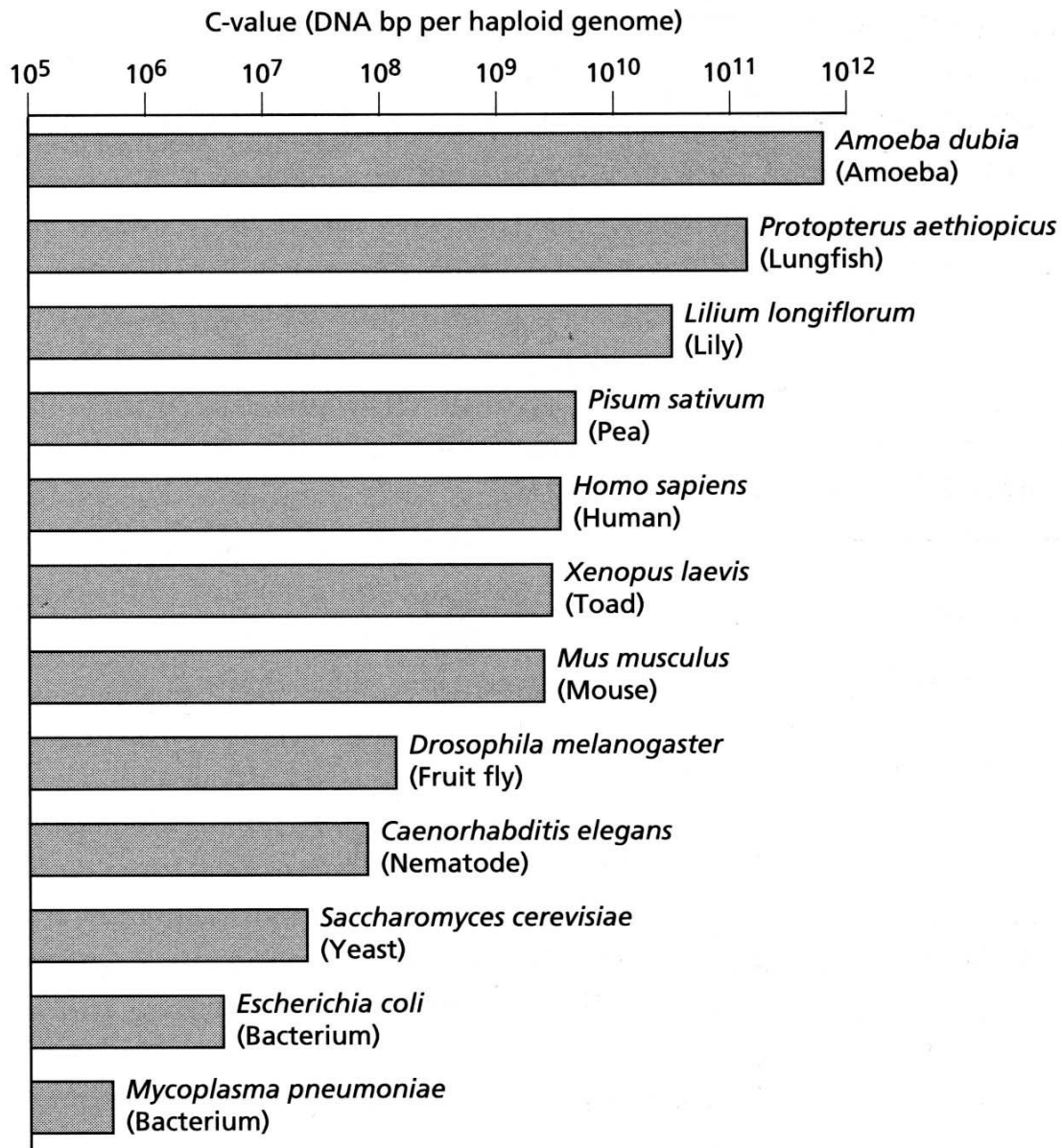
- Shine-Dalgarno box (AGGAGG)
- -10 site (TATAAT)
- -35 site (TTGACA)



- Хромосомы, геномы
- Эухроматин (содержит большинство генов)
- Гетерохроматин
- Центромеры, теломеры, хромосомные плечи, хромосомный бэндинг





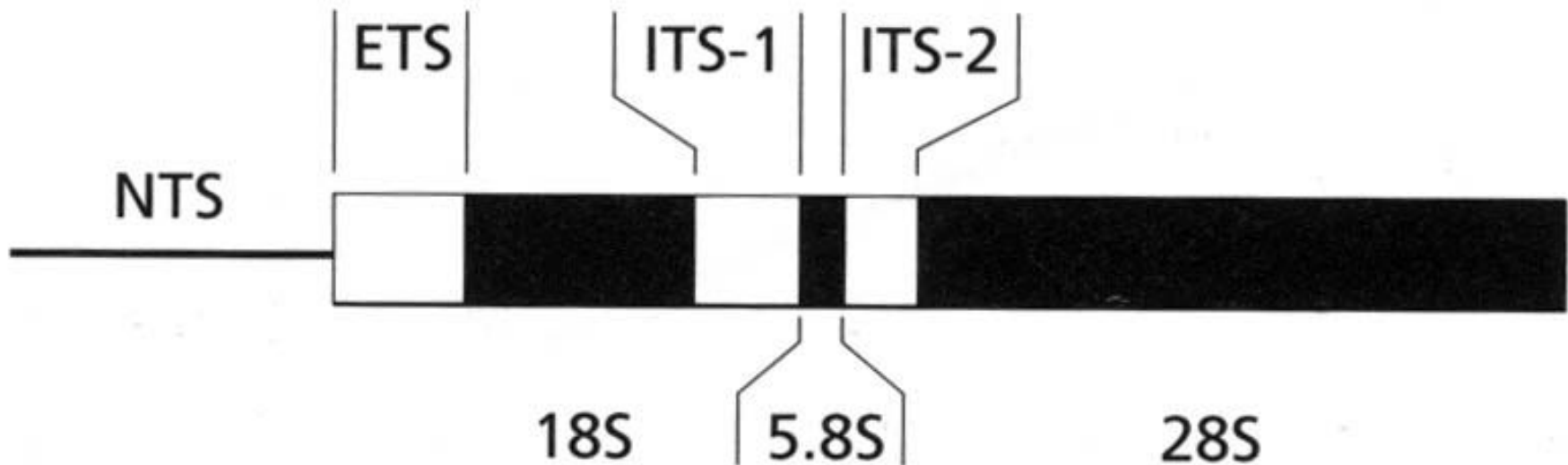


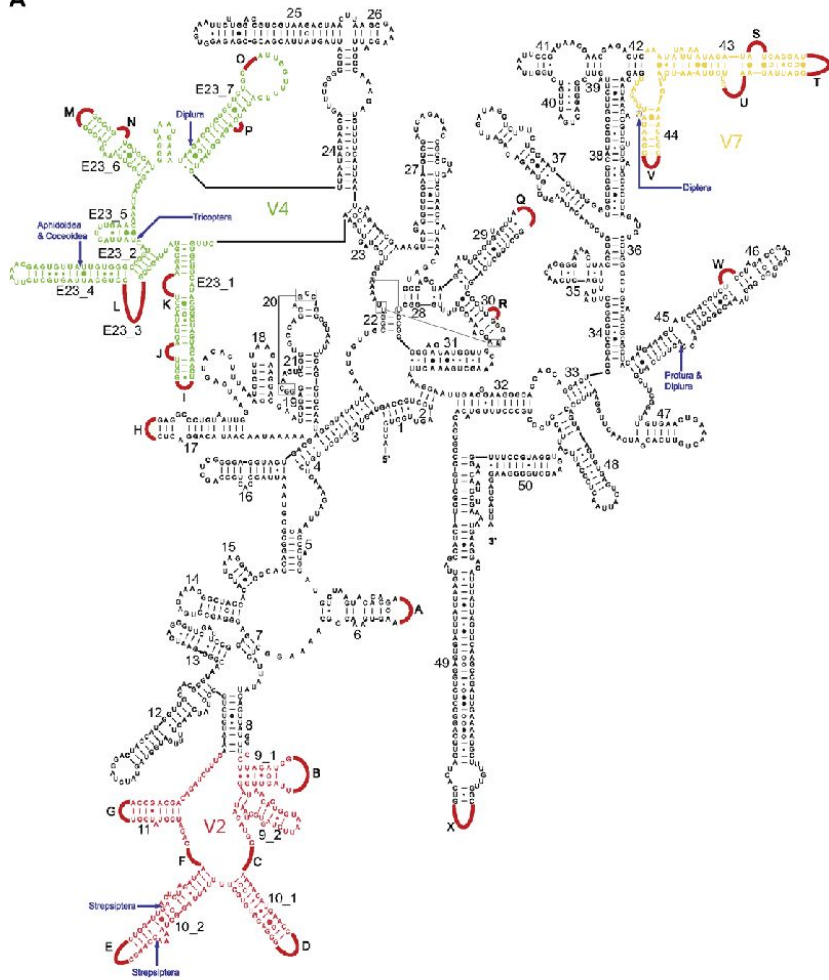
Некодирующая
ДНК и
C- парадокс

Кластеры рибосомальной ДНК

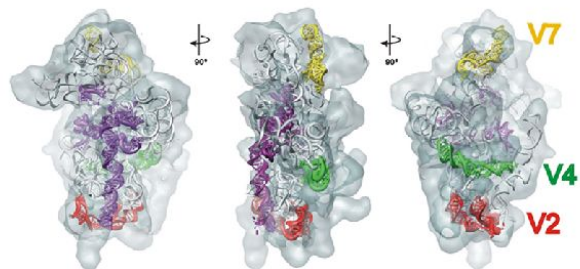
Консервативные, переменные и гиперпеременные участки

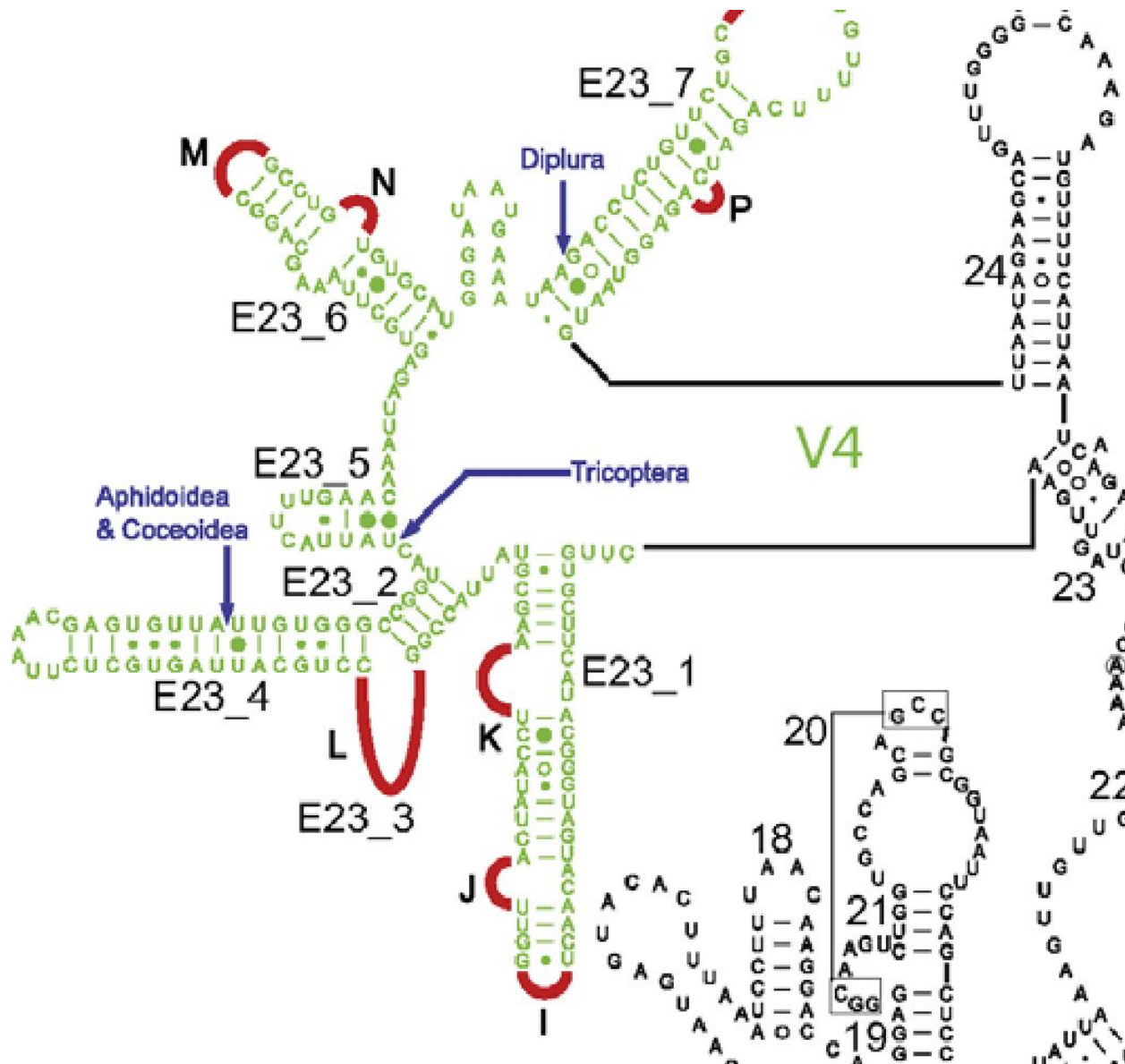
Множество копий генов. Понятие согласованной (концертной) эволюции.



A

Структура 18S rDNA

B

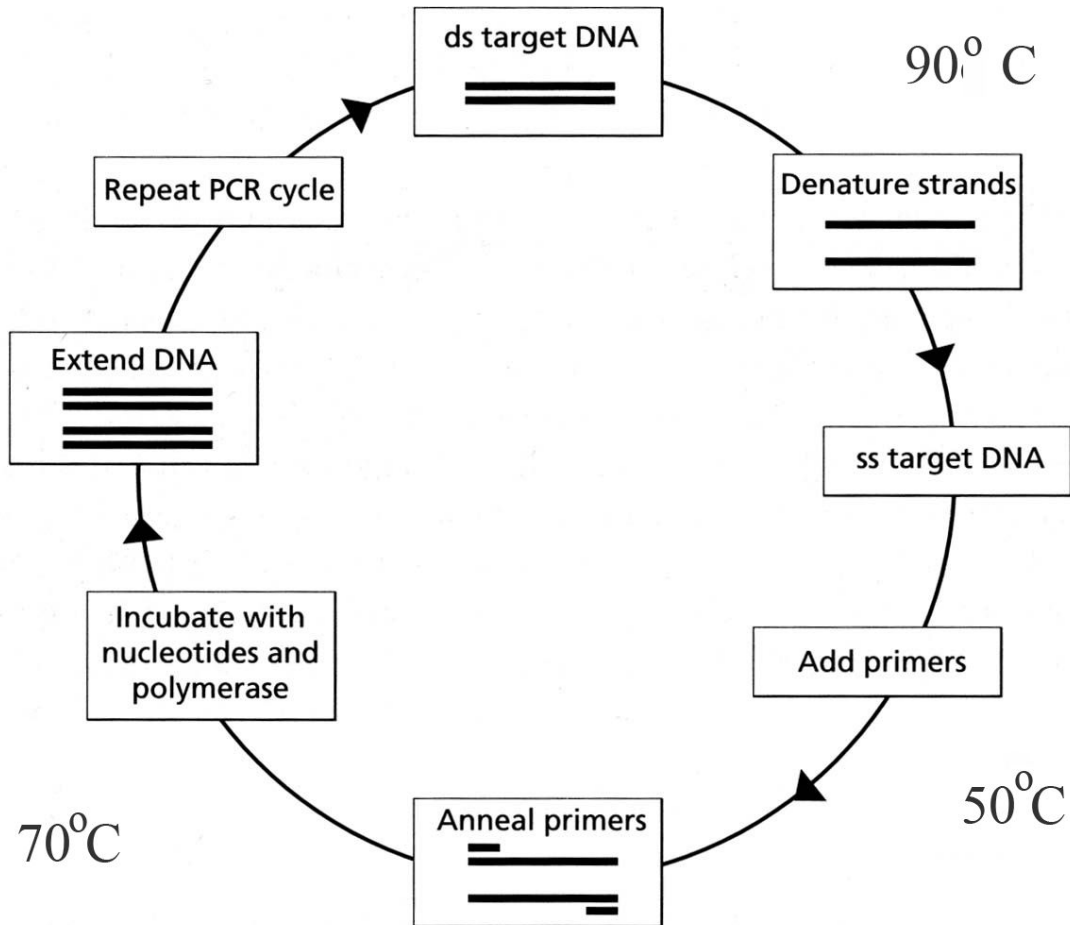


18S rDNA
(фрагмент)

Техники ДНК-анализа

- Клонирование ДНК
- ПЦР и секвенирование
- Техники, основанные на электрофорезе
 - RAPD (randomly amplified polymorphic DNA)
 - RFLP (restriction fragment length polymorphism)
 - AFLP (amplified fragment length polymorphism)

ПЦР



- как на практике перейти от анализа распределения признаков к филогениям?
- Для этого нужны:
 - (1) сами признаки,
 - (2) модели эволюции этих признаков и
 - (3) методы филогенетического анализа, т.е. обоснованные и систематизированные совокупности шагов и действий, которые необходимо предпринять, чтобы на основании изучения признаков и с учетом модели эволюции этих признаков решить поставленную задачу.

Общие принципы построения филогений

- 1) Анализ признаков,**
- 2) выбор оптимальной модели
эволюции признака,**
- 3) выбор методов и алгоритмов для
построения дерева**

Общие принципы построения филогений

- 1) **Анализ признаков,**
- 2) **Выбор признаков –необходимые и достаточные условия**
- 2) **выбор оптимальной модели эволюции признака,**
- 3) **выбор методов и алгоритмов для построения дерева**

- Необходимое условие - гомологичность
- Достаточное условие - соответствие поставленной задаче
 - Число генов (локусов)
 - Уровень изменчивости генов (локусов) (=скорость молекулярной эволюции)

- Сколько генов необходимо?
- Зависит от решаемой проблемы.

Число генов (локусов)

1- проблемы идентификации видов и популяций

Идентификация объектов при помощи ДНК-баркодинга

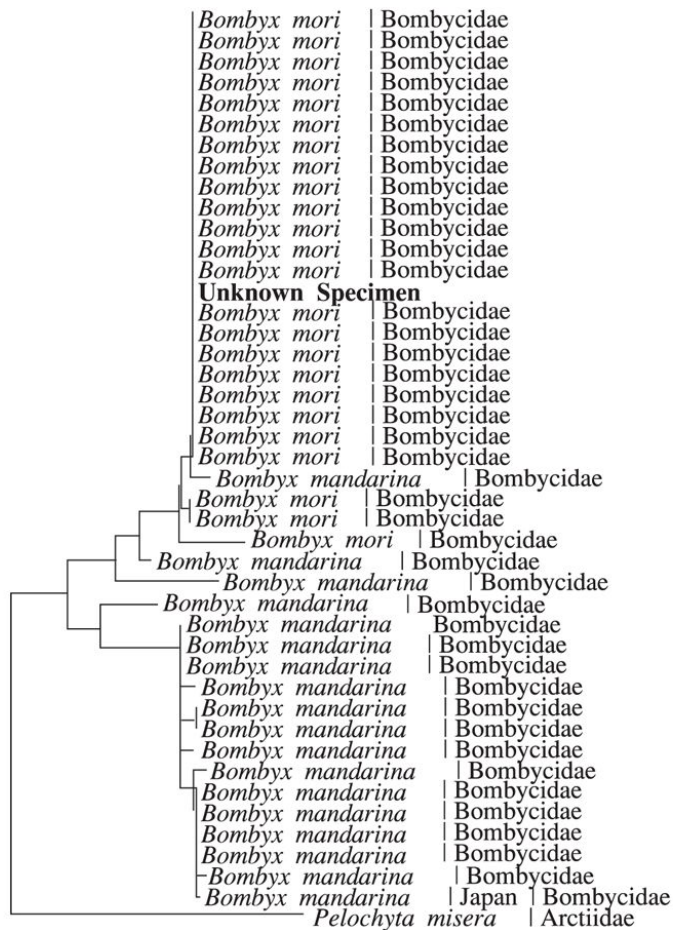


Рис. 1. Видовая идентификация в системе ДНК-баркодинга. Цифровые значения соответствуют частоте гомо-СОВ

материал (у насекомых лучше избегать использования высушенных и вторично размоченных экземпляров, поскольку фрагментация молекул ДНК, происходящая при гниении, сильно затрудняет анализ).

Более серьезной проблемой для широкого распространения баркодинга является второй этап – выделение ДНК и наработка ее путем амплификации. Методически он не сложен и не дорогостоящ, но, к сожалению, пока не удалось создать такие технологии, которые бы полностью исключали ручной труд. По-видимому, эта проблема также будет решена довольно скоро.

Третий этап – секвенирование – до недавнего времени считался самым сложным. Однако в настоящее время уже организована сеть промышленных лабораторий, которые секвенируют любые участки ДНК по заказу клиента (как это делается, например, при печати фотографий в фотолабораториях). Эта сеть быстро расширяется, и ее услуги становятся общедоступными. Так что массовое внедрение методов секвенирования – дело совсем недалекого будущего.

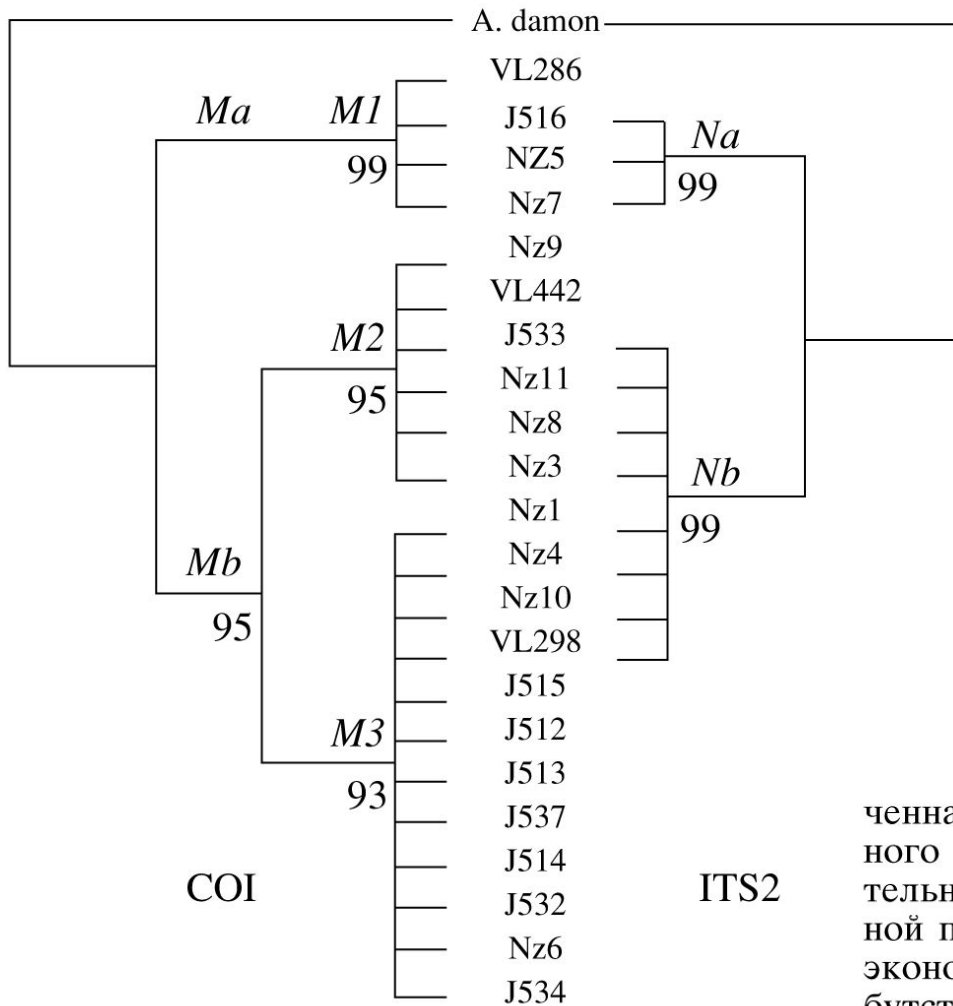
Остается четвертая проблема – создание базы данных баркодов, которая будет использоваться для определения. Такая база данных будет работать наиболее эффективно, если хранящиеся в ней ДНК-баркоды будут обработаны таксономически, т.е. будет проанализировано их соответствие ре-

- Сколько генов необходимо?
- Зависит от решаемой проблемы.

Число генов (локусов)

1- проблемы идентификации видов и популяций

2 - анализ равновесия по сцеплению



Дендрограмма гаплотипов голубянок, полученная на основе анализа фрагмента митохондриального гена COI и ядерной нуклеотидной последовательности ITS2 с использованием метода максимальной парсимонии (50%-ный консенсус 1498 наиболее экономных деревьев). В узлах показаны значения бутстреп-поддержки.

Число генов (локусов)

1- проблемы идентификации

2 - анализ равновесия по сцеплению

3 и более - филогенетические
реконструкции

Какие гены ?

Уровень изменчивости генов (локусов) (=скорость молекулярной эволюции)

1) консервативные кодирующие районы ДНК

2) переменные кодирующие районы ДНК

3) некодирующая ДНК

интроны

спейсеры

повторы (например, микросателлиты)

Общие принципы построения филогений

- 1) Анализ признаков,
- 2) выбор оптимальной модели эволюции признака,
- 3) выбор методов и алгоритмов для построения дерева

- Модели – это или словесные, или имеющие вид математических формул, описания закономерностей эволюционных преобразований признаков.
- На ранних этапах развития филогенетики в качестве моделей часто использовались нечетко сформулированные (а иногда не сформулированные вообще) интуитивные представления о том, как могла идти эволюция изучаемых признаков.

- Параметрические модели
 - Включают параметры с известными свойствами (распределениями)
- Непараметрические модели
 - Тип распределения неизвестен

Принципы моделирования

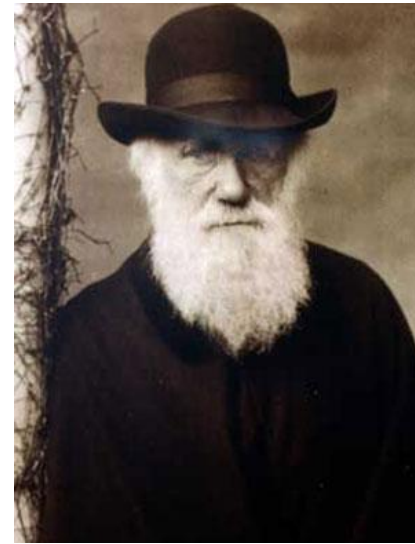
- Любая модель использует две группы данных
 - параметры, которые выявляют при разработке модели в ходе изучения процесса (
 - данные, которые выявляют при обработке конкретных измерений

2) выбор оптимальной модели эволюции признака

В основе идея биологической эволюции вообще



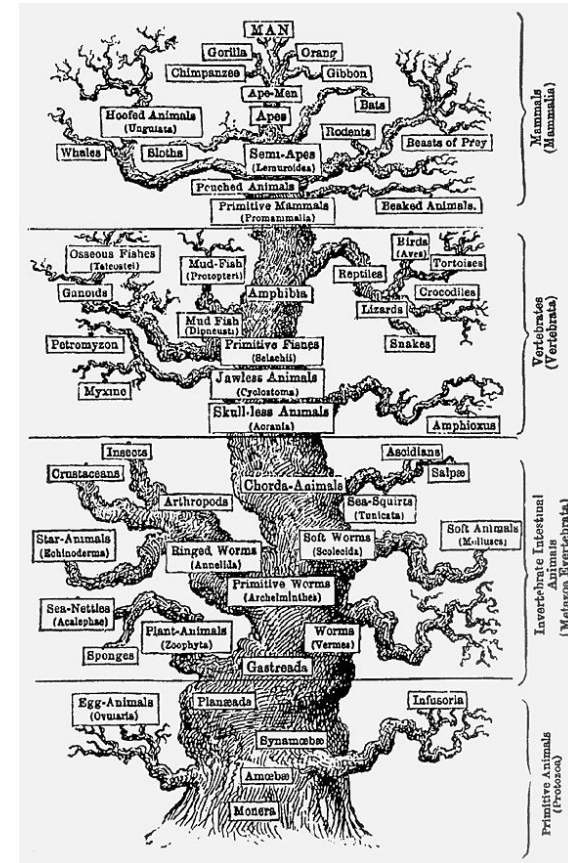
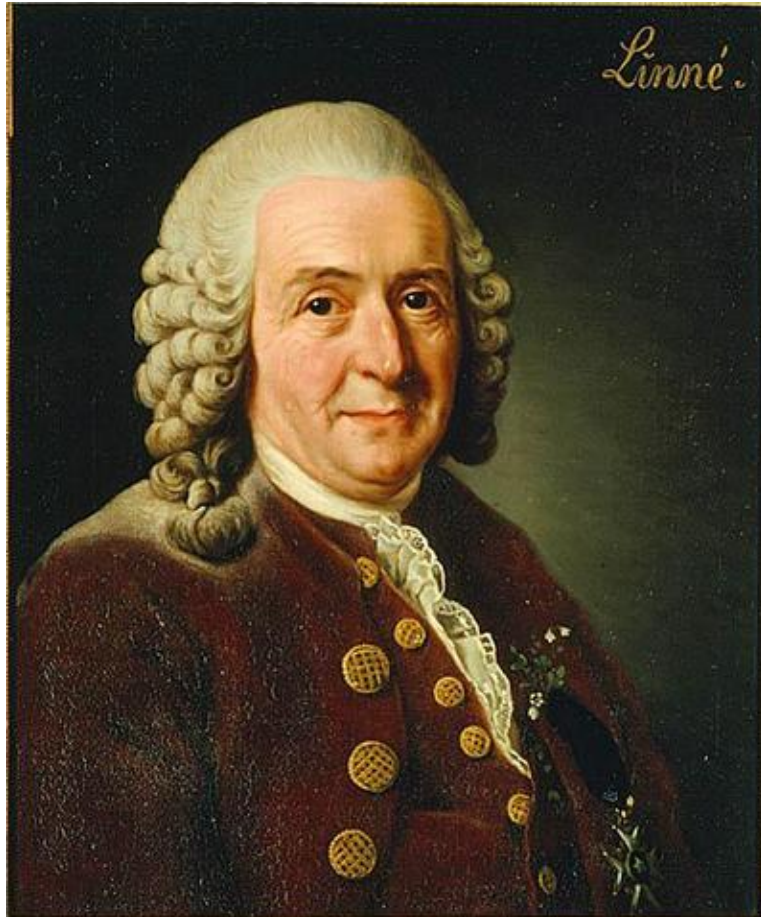
Жан Батист Пьер Антуан
де Моне шевалье де Ламарк



ТОПОЛОГИЯ

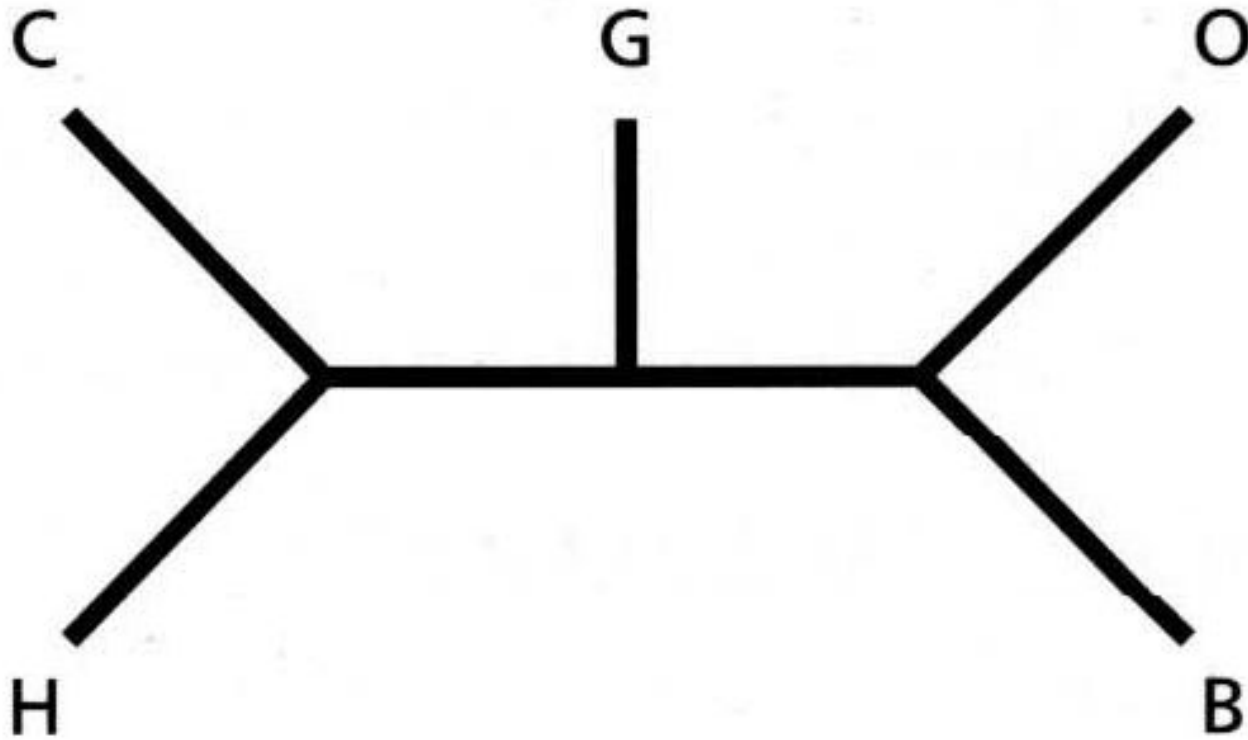
- Обязательный компонент любой филогенетической модели - это топология, то есть геометрическая, обычно двухмерная схема, показывающая генеалогические связи между единицами филогенетического анализа.

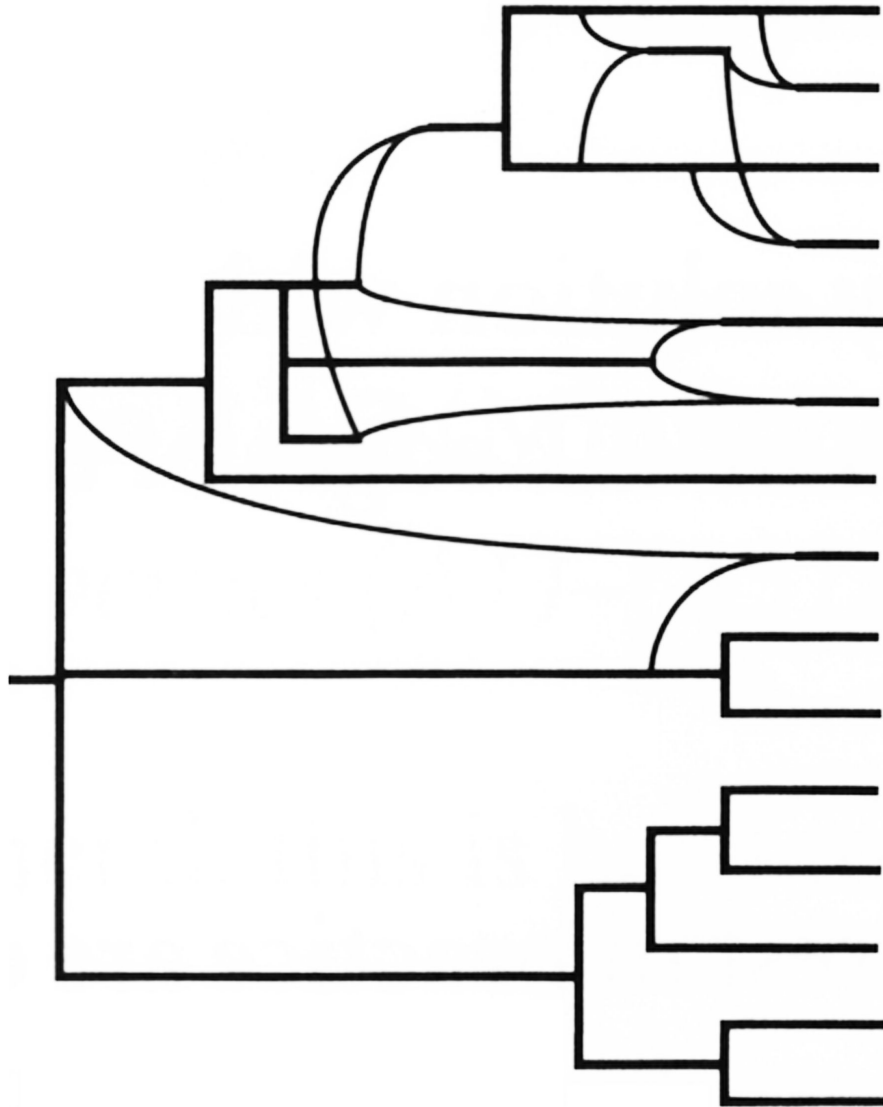
Выбор эволюционной модели



Иерархическая система Линнея – принципиальная конкретная основа любой эволюционной модели

Если направление передачи признака неизвестно, можно использовать модель неукорененного дерева





Для представления
филогении в случаях
ретикулярной эволюции
удобно использовать
модель
филогенетической сети

Здесь: укорененная сеть

- Кроме того, у филогенетических моделей могут быть различные качественные и количественные параметры, выраженные словами, числами, соотношениями и вероятностями.
- Примеры таких параметров:
 - признак, который был потерян организмом в ходе эволюции, не может снова появиться в своем исходном виде (модель Долло)
 - эволюционные изменения признака полностью обратимы (модель Фитча-Вагнера)