

Лекция 4

Модели молекулярной эволюции, кладистика по Хеннигу и метод максимальной парсимонии

ДНК:

1 5 10
tagcaaaatg

Модели молекулярной эволюции

ДНК:

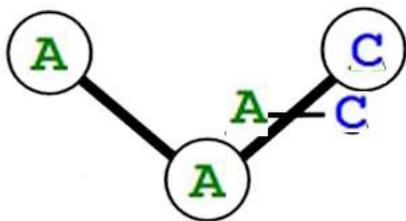
1 5 10
tagcaaaatg

Соотношения между нуклеотидными заменами и нуклеотидными различиями

Единичная замена

(a) Single substitution

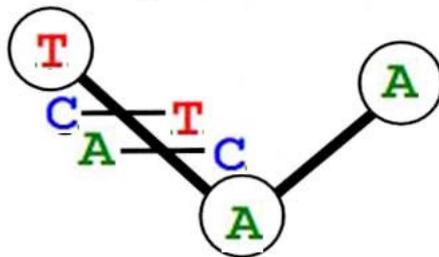
1 change, 1 difference



Множественные замены

(b) Multiple substitution

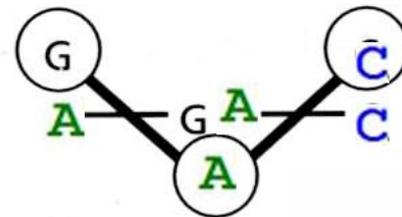
2 changes, 1 difference



Одновременные замены
в разных линиях

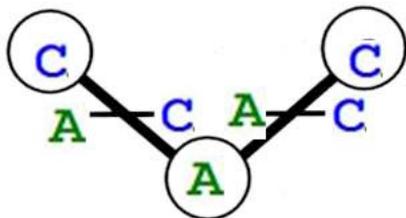
(c) Coincidental substitution

2 changes, 1 difference



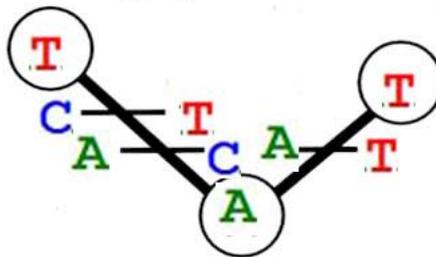
Параллельные замены
(d) Parallel substitution

2 changes, no difference



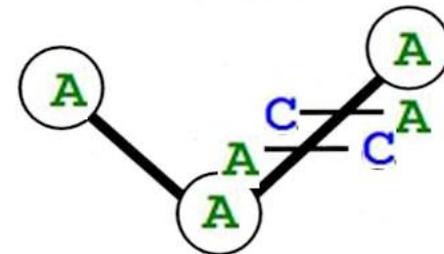
Конвергентные замены
(e) Convergent substitution

3 changes, no difference



Обратная замена
(f) Back substitution

2 changes, no difference

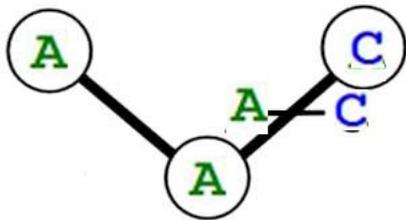


Число нуклеотидных замен \geq числа наблюдаемых нуклеотидных различий

Единичная замена

(a) Single substitution

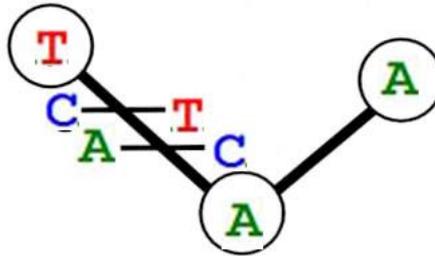
1 change, 1 difference



Множественные замены

(b) Multiple substitution

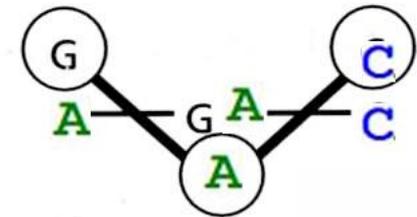
2 changes, 1 difference



Одновременные замены в разных линиях

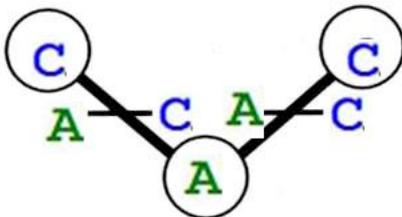
(c) Coincidental substitution

2 changes, 1 difference



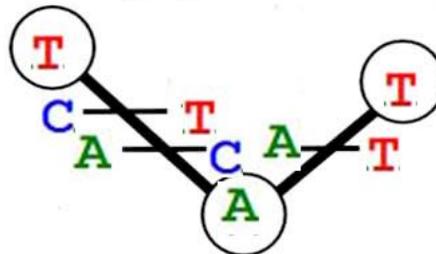
Параллельные замены (d) Parallel substitution

2 changes, no difference



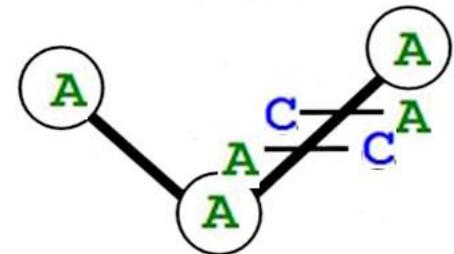
Конвергентные замены (e) Convergent substitution

3 changes, no difference



Обратная замена (f) Back substitution

2 changes, no difference



явные и скрытые генетические дистанции

	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

Что такое генетическая дистанция?

$d = p$, где p - доля различающихся сайтов

d - это "сырая" дистанция

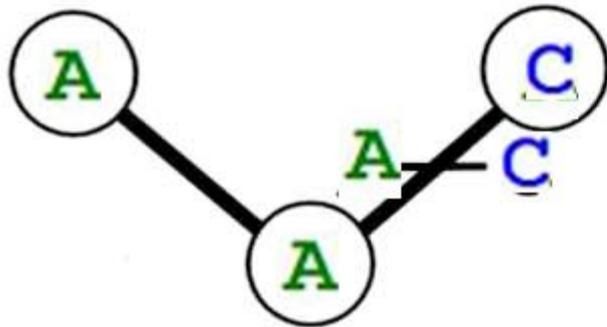
$$d = 3/32 = 9.375\%$$

Число нуклеотидных замен \geq числа наблюдаемых нуклеотидных отличий

Единичная замена

(a) Single substitution

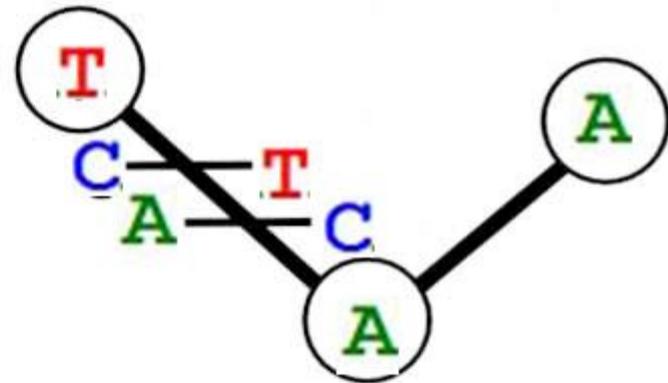
1 change, 1 difference



Множественные замены

(b) Multiple substitution

2 changes, 1 difference



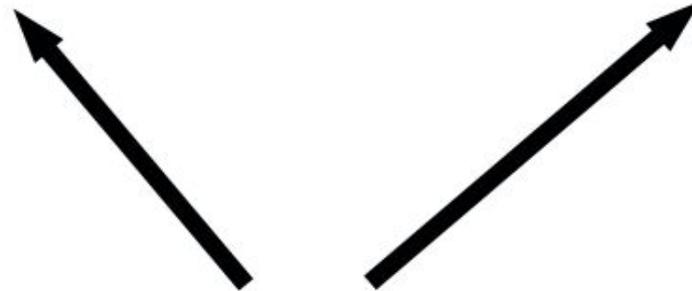
Проблема дистанций состоит в том, что наблюдаемые дистанции могут быть меньше, чем реальные дистанции, так как не все замены видны при сравнении сиквенсов

- Наблюдаемые генетические дистанции как правило меньше реальных эволюционных дистанций, так как есть скрытые замены
- Но как выявить эти реальные эволюционные дистанции?
- Нужно знать возраст таксонов (время дивергенции) и скорость замен

Закономерности накопления замен

ACGTACGTAC

ACGTACGTAC



$d=0$

ACGTACGTAC

Первая замена - в сайте 1. $d=0.1$

Наблюдаемая дистанция = реальной дистанции

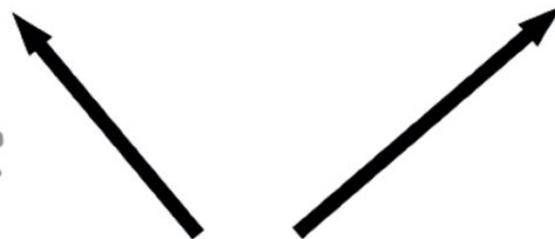
CCGTACGTAC

ACGTACGTAC

A → C

$d=0, 1$

ACGTACGTAC



Вторая замена -

Имеется вероятность 0.1, что она будет повторной (т.е. тоже в сайте 1) и вероятность 0.9, что она будет неповторной

Если она все же будет в первой позиции, то
Наблюдаемая дистанция = 0.1 (или даже 0),
а истинная дистанция = 0.2

1	2	3	4	5	6	7	8	9	10
C	C	G	A	A	C	G	T	A	C
A	C	G	T	A	C	G	T	A	C

Но скорее всего (с вероятностью 0.9), вторая замена не будет в сайте 1

Третья замена имеет большую вероятность быть повторной, четвертая – еще большую, и. т.д.

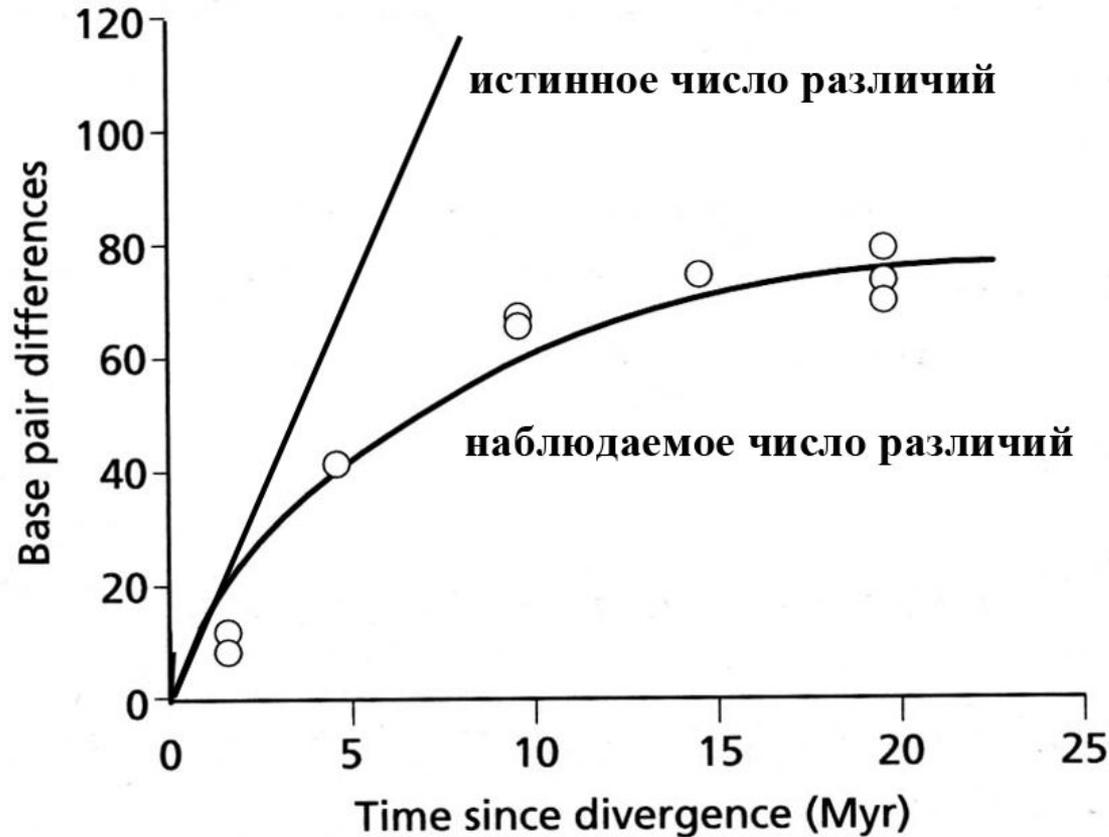
Т.е. чем больше замен, тем больше вероятность повторных замен.

Если все 10 позиций испытали замены, то любая следующая замена будет повторной.

CGTACGTACG
ACGTACGTAC

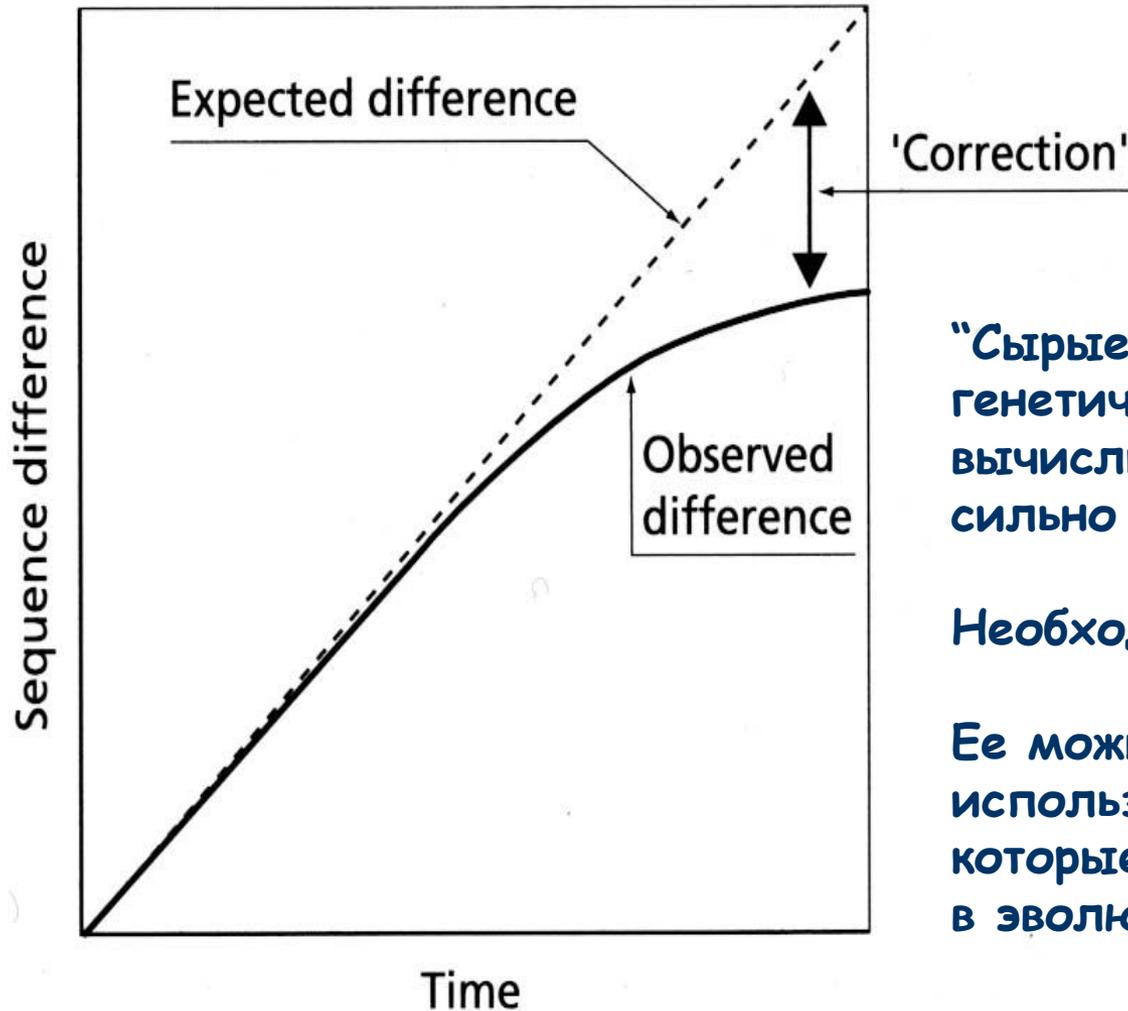
После этого замены продолжают накапливаться, а наблюдаемые различия не растут

Зависимость между временем дивергенции и числом наблюдаемых нуклеотидных отличий в гене *CytB* у жвачных копытных животных



Происходит насыщение нуклеотидными заменами – число замен растет, но уровень отличий выходит на плато и не меняется

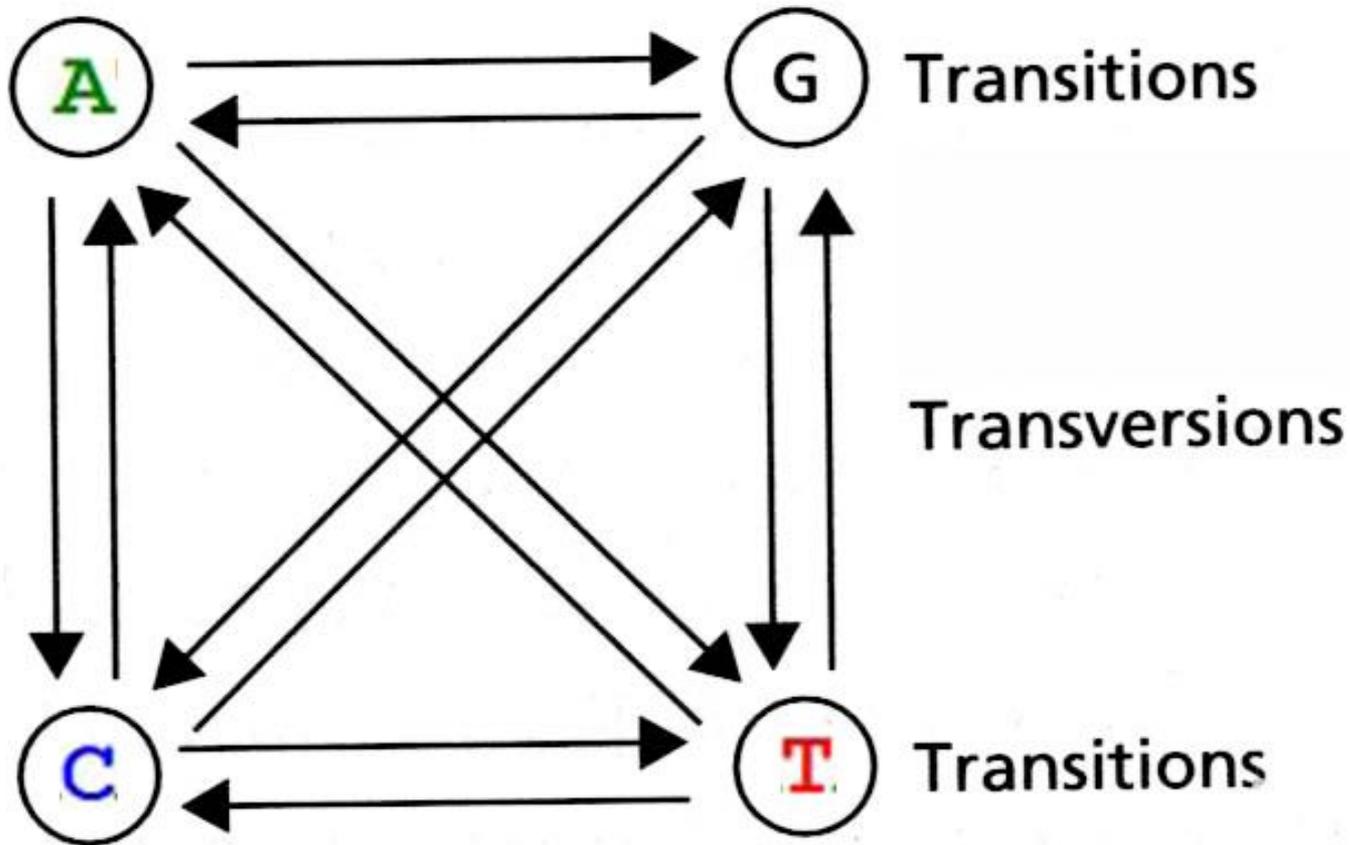
NB: каждая точка – это пара особей разных линий



“Сырые” (нескорректированные) генетические дистанции легко вычислить, но они могут быть сильно занижены.

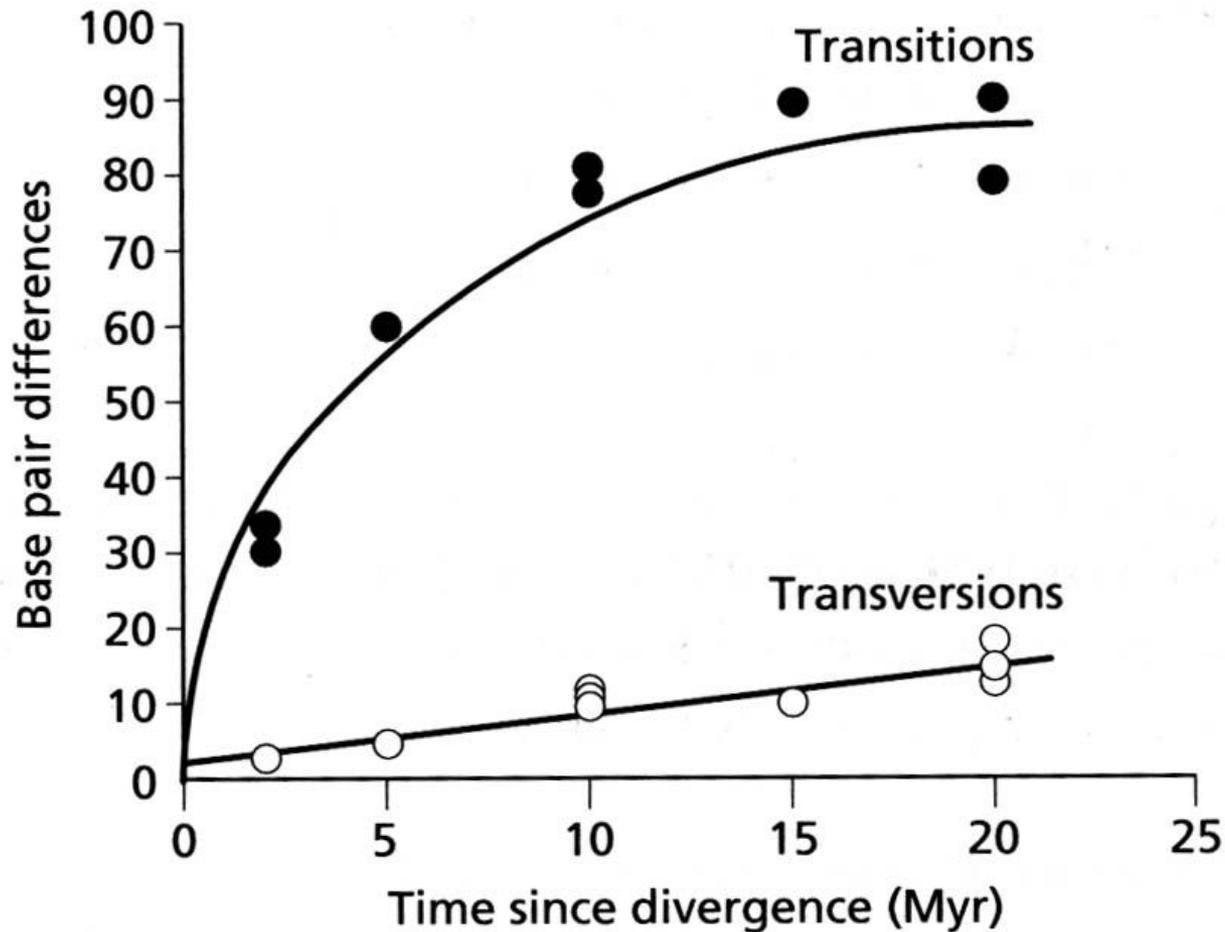
Необходима коррекция

Ее можно сделать с использованием моделей, которые учитывают разницу в эволюции разных признаков



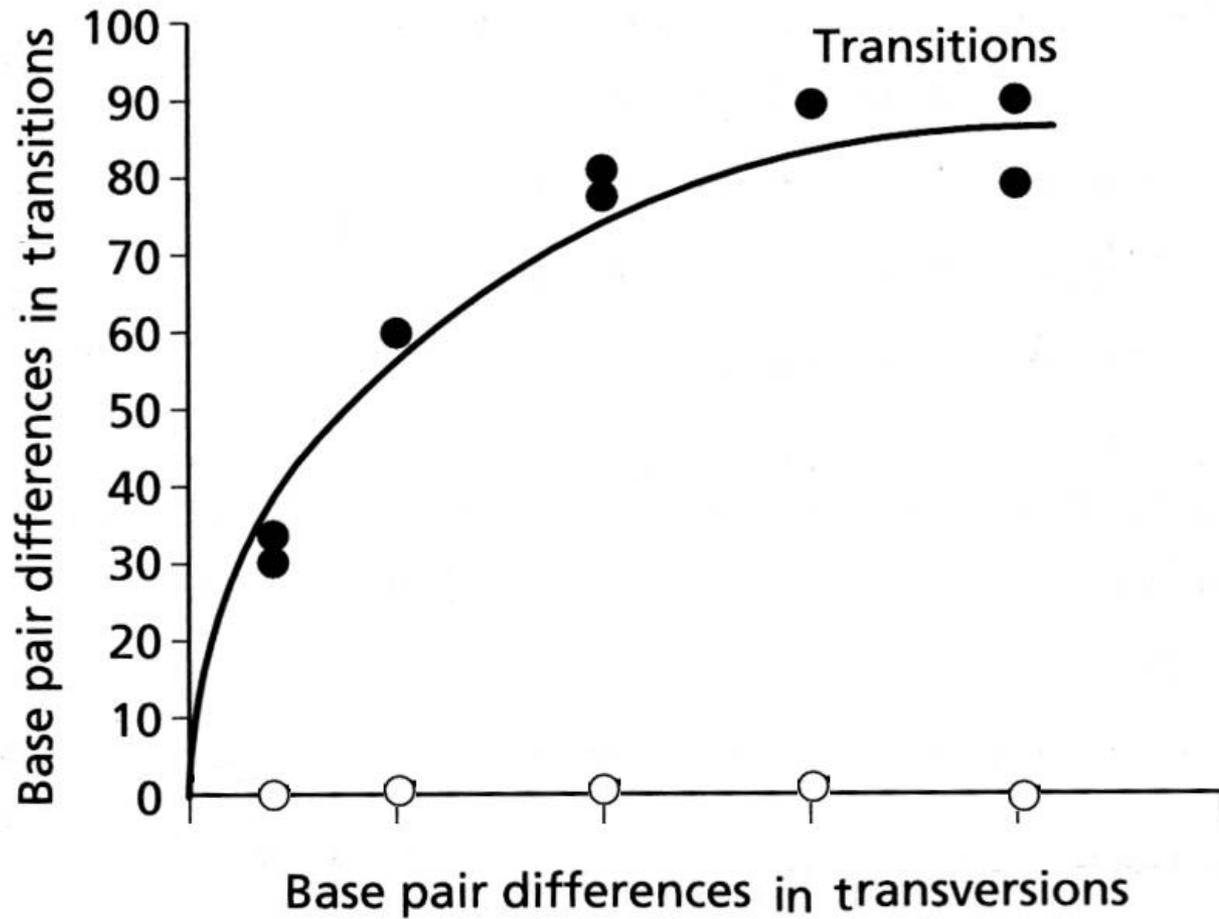
Purines = adenin and guanine
Pirimidines = cytosine and thymine

Кривые накопления повторных замен для транзиций и трансверсий



Каждая точка – это сравнение, т.е. пара видов

Кривая накопления транзиций по отношению к трансверсиям



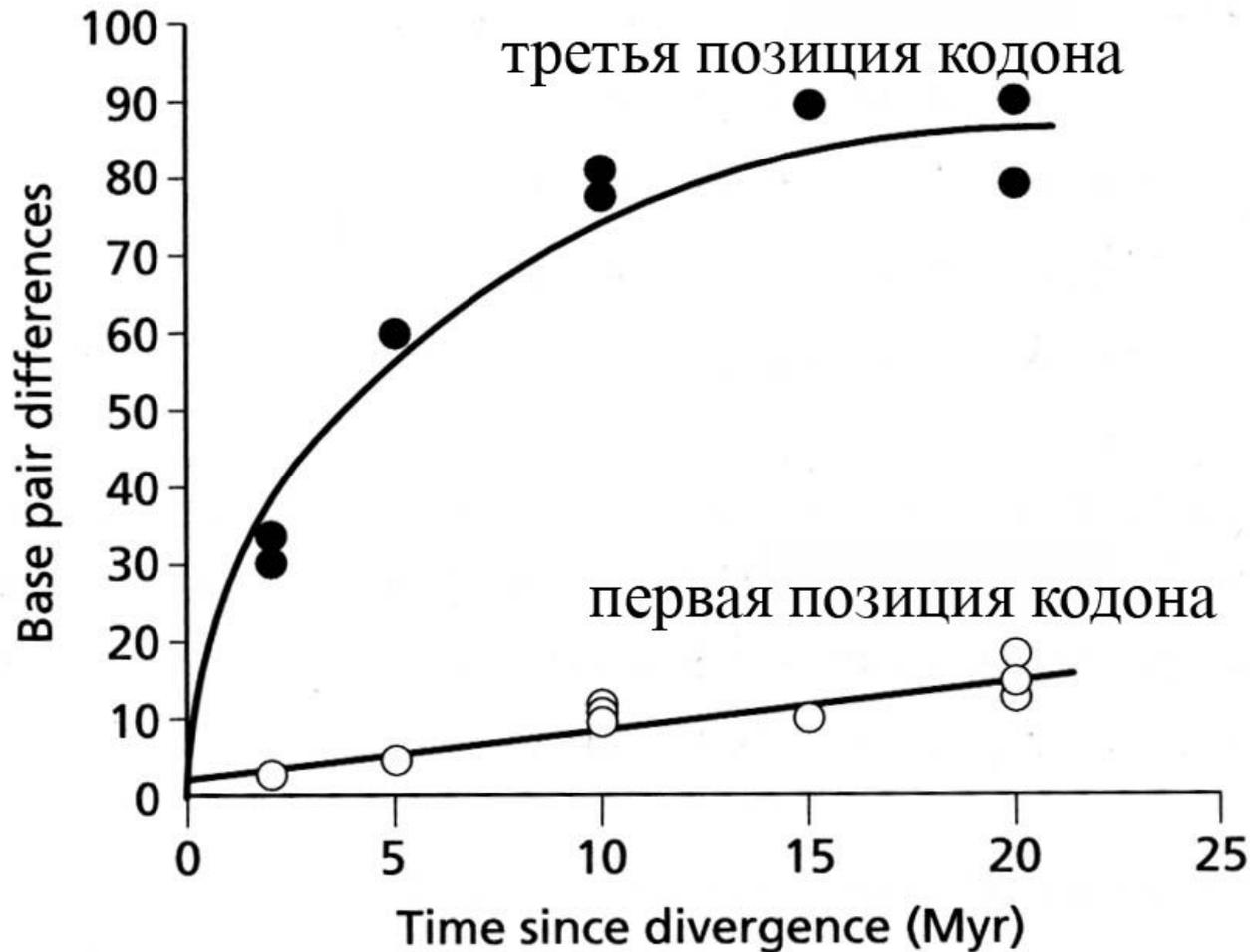
Генетический код

- Замена в первой позиции кодона ведет к замене аминокислоты
- Замена в третьей позиции кодона как правило синонимична
- → нуклеотиды в третьей позиции эволюционируют быстрее

The genetic code.

Codon	Amino acid						
UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys
UUC	Phe	UCC	Ser	UAC	Tyr	UGC	Cys
UUA	Leu	UCA	Ser	UAA	Stop	UGA	Stop
UUG	Leu	UCG	Ser	UAG	Stop	UGG	Trp
CUU	Leu	CCU	Pro	CAU	His	CGU	Arg
CUC	Leu	CCC	Pro	CAC	His	CGC	Arg
CUA	Leu	CCA	Pro	CAA	Gln	CGA	Arg
CUG	Leu	CCG	Pro	CAG	Gln	CGG	Arg
AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser
AUC	Ile	ACC	Thr	AAC	Asn	AGC	Ser
AUA	Ile	ACA	Thr	AAA	Lys	AGA	Arg
AUG	Met	ACG	Thr	AAG	Lys	AGG	Arg
GUU	Val	GCU	Ala	GAU	Asp	GGU	Gly
GUC	Val	GCC	Ala	GAC	Asp	GGC	Gly
GUA	Val	GCA	Ala	GAA	Glu	GGA	Gly
GUG	Val	GCG	Ala	GAG	Glu	GGG	Gly

Кривые накопления повторных замен для третьей и первой позиций кодона



	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

Какие параметры можно извлечь из нуклеотидного выравнивания?

Какие параметры можно извлечь из нуклеотидного выравнивания?

- (1) Длина
- (2) Доля изменчивых сайтов
- (3) Доля инвариантных сайтов
- (4) Соотношение нуклеотидов разных типов
- (5) Доля транзиций и трансверсий
- (6) Доля нуклеотидных замен разных типов
- (7) Все это (2-6) отдельно для каждой позиции кодона
- (8) Доля синонимичных и несинонимичных замен
- (9) Доля синонимичных и несинонимичных замен отдельно для каждой позиции кодона
- (10) Распределение замен по длине нуклеотидной последовательности

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

Какие параметры можно извлечь из нуклеотидного выравнивания?

(1) Длина выравнивания

Зависит от задач и технических возможностей

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

Какие параметры можно извлечь из нуклеотидного выравнивания?

- (2) Доля изменчивых сайтов
- (3) Доля инвариантных сайтов

```
AY557140 A achaemenes voucher  
EF104615 A achaemenes voucher  
EF104621 A actinides voucher V  
LOWA513-06|2005-LOWA-513|Agrod  
LOWA503-06|2005-LOWA-503|Agrod  
LOWA502-06|2005-LOWA-502|Agrod  
LOWA501-06|2005-LOWA-501|Agrod  
LOWA512-06|2005-LOWA-512|Agrod  
LOWA511-06|2005-LOWA-511|Agrod
```

```
1           10           20           30  
ATAACATAAGATTCTGATTACTACCCCATCA  
ATAATATAAGATTGTGA-TACTACCCCATCA  
ATAACATAAGATTCTGATTATTACCACCATCA  
ATAACGTAAGATTCTGATTATTACCACCATCA  
ATAACATAAGATTCTGATTATTACCACCATCA  
ATAACAAAAGATTCTGATTATTACCACCATCA  
ATAACATAAGATTCTGATTATTACCACCATCA  
ATAACATAAGATTCTGATTATTACCACCATCA  
ATAACATAAGATTCTGATTATTACCACCATCA
```

Какие параметры можно извлечь из нуклеотидного выравнивания?

(4) Соотношение нуклеотидов разных типов (A C G T)

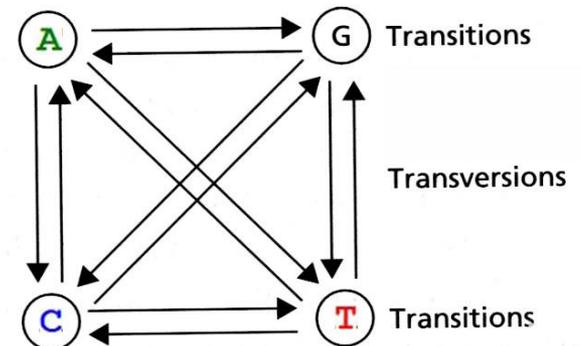
	1	10	20	30
AY557140 A achaemenes voucher	A	T	A	A
EF104615 A achaemenes voucher	A	T	A	A
EF104621 A actinides voucher V	A	T	A	A
LOWA513-06 2005-LOWA-513 Agrod	A	T	A	A
LOWA503-06 2005-LOWA-503 Agrod	A	T	A	A
LOWA502-06 2005-LOWA-502 Agrod	A	T	A	A
LOWA501-06 2005-LOWA-501 Agrod	A	T	A	A
LOWA512-06 2005-LOWA-512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

Какие параметры можно извлечь из нуклеотидного выравнивания?

(5) Доля транзиций и трансверсий

	1	10	20	30																											
AY557140 A achaemenes voucher	A	T	A	A	G	A	T	T	C	T	G	A	T	T	A	C	T	A	C	C	C	A	T	C	A						
EF104615 A achaemenes voucher	A	T	A	A	T	A	A	G	A	T	T	G	T	G	A	-	T	A	C	T	A	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	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	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	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	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	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	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	A	T	C	A

Type of sequences	Transition/transversion ratio (κ)
mtDNA	9.0
12S rRNA	1.75
α - and β -globins	0.66
Pseudo η -globin	2.70



Какие параметры можно извлечь из нуклеотидного выравнивания?
 (6) Доля нуклеотидных замен разных типов

	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

p_{AA} p_{AC} p_{AG} p_{AT}
 p_{CA} p_{CC} p_{CG} p_{CT}
 p_{GA} p_{GC} p_{GG} p_{GT}
 p_{TA} p_{TC} p_{TG} p_{TT}

Какие параметры можно извлечь из нуклеотидного выравнивания?
(7) Все это (2-6) отдельно для каждой позиции кодона

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

Какие параметры можно извлечь из нуклеотидного выравнивания?

(8) Доля синонимичных и несинонимичных замен

(9) Доля синонимичных и несинонимичных замен

отдельно для каждой позиции кодона

(10) Распределение замен по длине нуклеотидной последовательности

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

Модели нуклеотидных замен

- Предпосылки
- 1) нуклеотидные замены одного типа равновероятны в разных частях одного гена
- 2) нуклеотидные замены обратимы (здесь не работает принцип необратимости эволюции) (A <-> T)

Если вероятности нуклеотидных замен (p) и частоты нуклеотидов (f) константны во времени, то суммарная эволюционная дистанция (доля измененных нуклеотидов) =

$$\mathbf{P}_t = \begin{bmatrix} p_{AA} & p_{AC} & p_{AG} & p_{AT} \\ p_{CA} & p_{CC} & p_{CG} & p_{CT} \\ p_{GA} & p_{GC} & p_{GG} & p_{GT} \\ p_{TA} & p_{TC} & p_{TG} & p_{TT} \end{bmatrix} \quad \mathbf{f} = [f_A \ f_C \ f_G \ f_T]$$

Где t это время, P_{AC} -

$$P_{AC} = P_{CA}$$

Если вероятности нуклеотидных замен (p) и частоты нуклеотидов (f) константны во времени, то суммарная эволюционная дистанция (доля измененных нуклеотидов) =

General reversible model (REV)

$$\mathbf{P}_t = \begin{bmatrix} \cdot & \pi_C a & \pi_G b & \pi_T c \\ \pi_A a & \cdot & \pi_G d & \pi_T e \\ \pi_A b & \pi_C d & \cdot & \pi_T f \\ \pi_A c & \pi_C e & \pi_G f & \cdot \end{bmatrix}, \quad \mathbf{f} = [\pi_A \ \pi_C \ \pi_G \ \pi_T]$$

- ЧАСТОТЫ НУКЛЕОТИДОВ И ДОЛИ ЗАМЕН
разного типа берутся
непосредственно из выравнивания

AY557140 A achaemenes voucher	ATAACATAAGATTCTGATTACTACCCCATCA
EF104615 A achaemenes voucher	ATAATATAAGATTGTGA-TACTACCCCATCA
EF104621 A actinides voucher V	ATAACATAAGATTCTGATTATTACCACCATCA
LOWA513-06 2005-LOWA-513 Agrod	ATAACGTAAGATTCTGATTATTACCACCATCA
LOWA503-06 2005-LOWA-503 Agrod	ATAACATAAGATTCTGATTATTACCACCATCA
LOWA502-06 2005-LOWA-502 Agrod	ATAACAAAAGATTCTGATTATTACCACCATCA
LOWA501-06 2005-LOWA-501 Agrod	ATAACATAAGATTCTGATTATTACCACCATCA
LOWA512-06 2005-LOWA-512 Agrod	ATAACATAAGATTCTGATTATTACCACCATCA
LOWA511-06 2005-LOWA-511 Agrod	ATAACATAAGATTCTGATTATTACCACCATCA

ЖС

Вероятности всех замен одинаковы, частоты нуклеотидов равны

Jukes–Cantor (JC)

$$\mathbf{P}_t = \begin{bmatrix} . & \alpha & \alpha & \alpha \\ \alpha & . & \alpha & \alpha \\ \alpha & \alpha & . & \alpha \\ \alpha & \alpha & \alpha & . \end{bmatrix}, \quad \mathbf{f} = \left[\frac{1}{4} \quad \frac{1}{4} \quad \frac{1}{4} \quad \frac{1}{4} \right]$$

$$d = -\frac{3}{4} \ln \left(1 - \frac{4}{3} p \right)$$

где p - это сырая дистанция

Двухпараметрическая модель Кимуры K_2P
 Вероятности транзиций и трансверсий разные,
 частоты нуклеотидов равны

Kimura's 2 parameter model (K2P)

α - транзиция

β - трансверсия

$$P_t = \begin{bmatrix} . & \beta & \alpha & \beta \\ \beta & . & \beta & \alpha \\ \alpha & \beta & . & \beta \\ \beta & \alpha & \beta & . \end{bmatrix}, \quad \mathbf{f} = \left[\frac{1}{4} \frac{1}{4} \frac{1}{4} \frac{1}{4} \right].$$

Type of sequences	Transition/transversion ratio (κ)
mtDNA	9.0
12S rRNA	1.75
α - and β -globins	0.66
Pseudo η -globin	2.70

F81

Вероятности всех замен одинаковы, но частоты нуклеотидов разные

Felsenstein (1981)

$$\mathbf{P}_t = \begin{bmatrix} . & \pi_C \alpha & \pi_G \alpha & \pi_T \alpha \\ \pi_A \alpha & . & \pi_G \alpha & \pi_T \alpha \\ \pi_A \alpha & \pi_C \alpha & . & \pi_T \alpha \\ \pi_A \alpha & \pi_C \alpha & \pi_G \alpha & . \end{bmatrix}, \quad \mathbf{f} = [\pi_A \ \pi_C \ \pi_G \ \pi_T]$$

HKY model

Вероятности транзиций и трансверсий разные,
частоты нуклеотидов разные

Hasegawa, Kishino and Yano (1985)

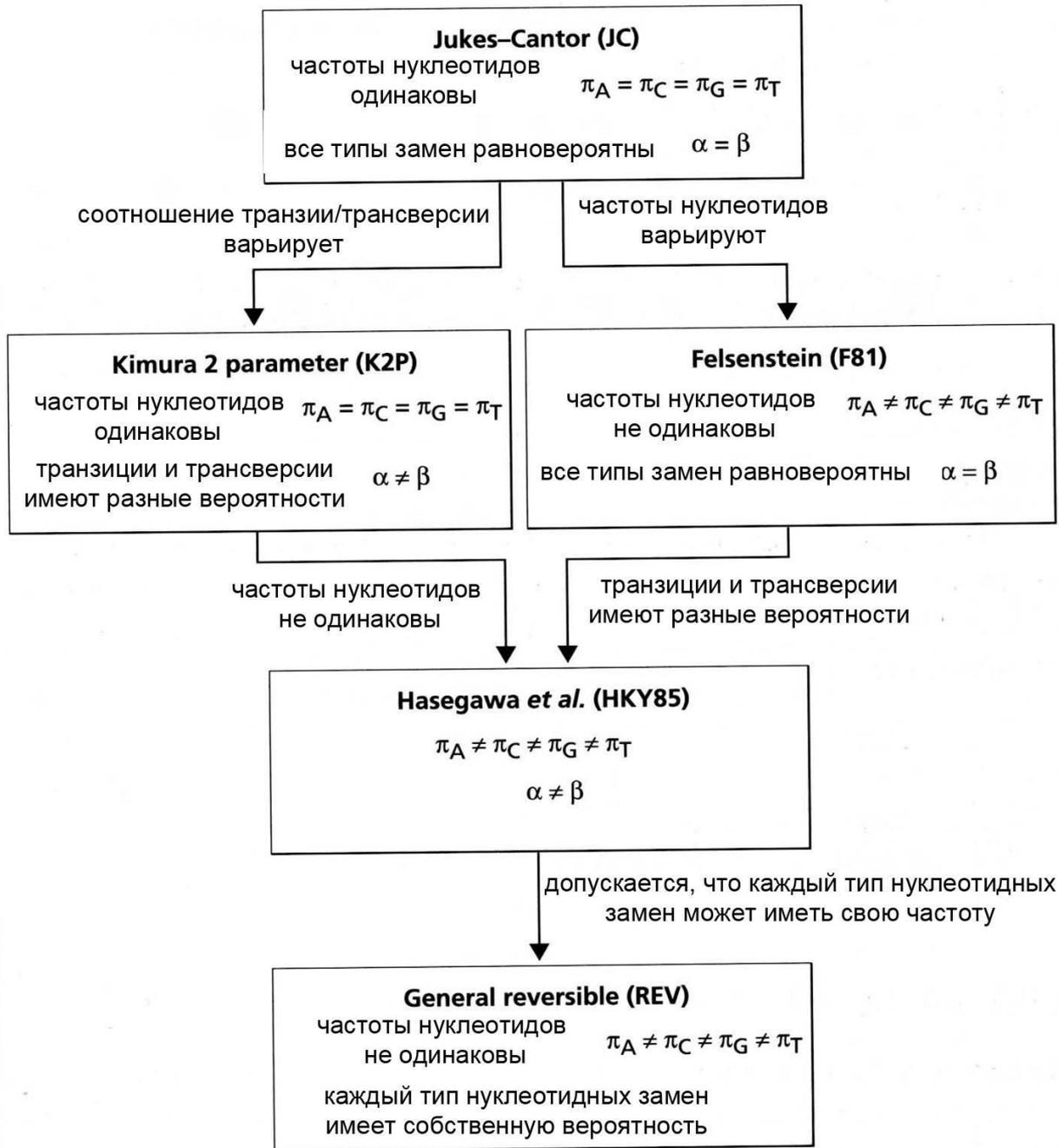
$$\mathbf{P}_t = \begin{bmatrix} . & \pi_C \beta & \pi_G \alpha & \pi_T \beta \\ \pi_A \beta & . & \pi_G \beta & \pi_T \alpha \\ \pi_A \alpha & \pi_C \beta & . & \pi_T \beta \\ \pi_A \beta & \pi_C \alpha & \pi_G \beta & . \end{bmatrix}, \quad \mathbf{f} = [\pi_A \ \pi_C \ \pi_G \ \pi_T]$$

REV

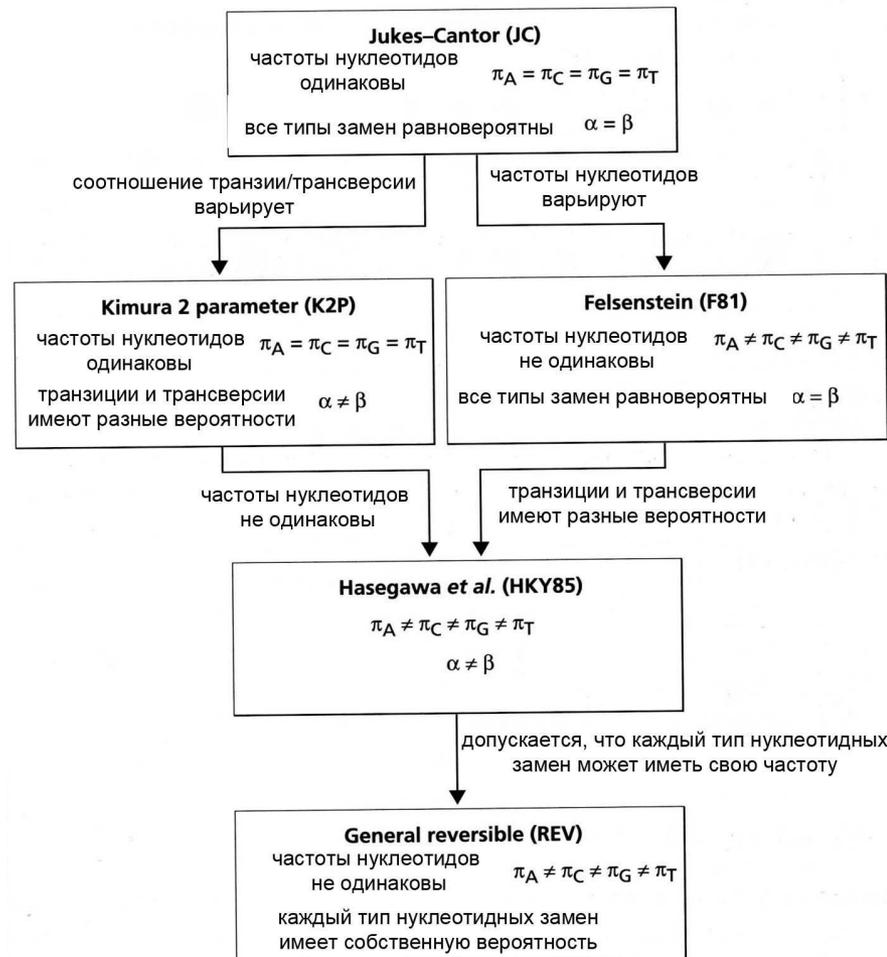
Вероятности ВСЕХ ЗАМЕН разные,
частоты нуклеотидов разные

General reversible model (REV)

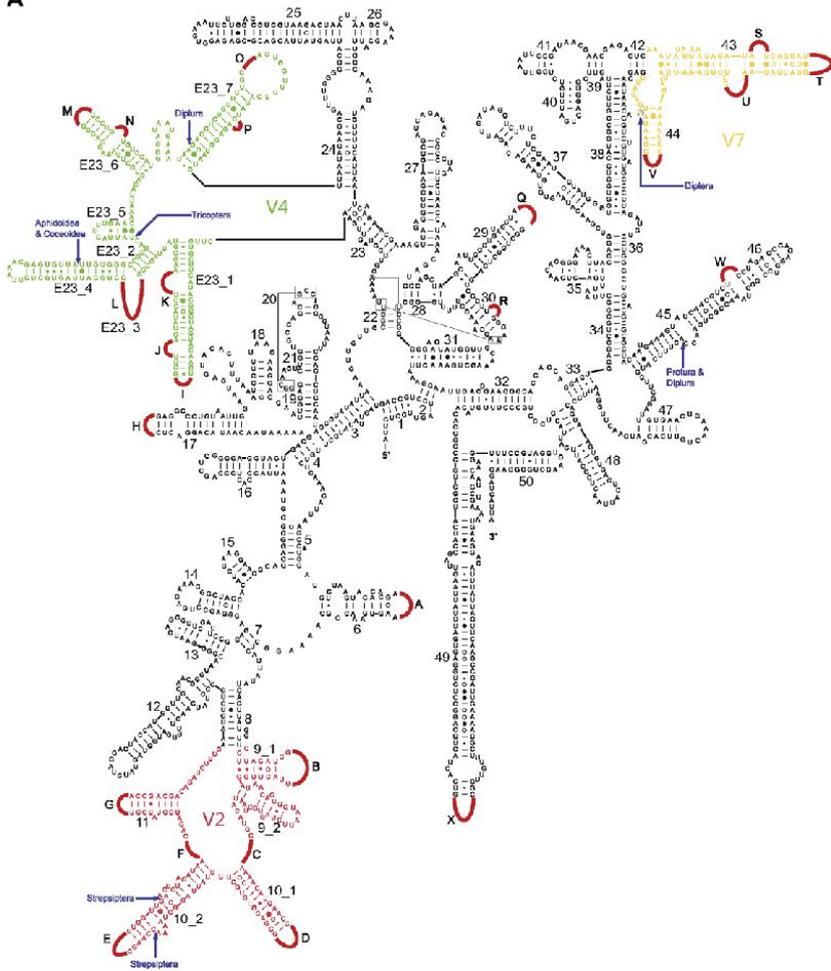
$$\mathbf{P}_t = \begin{bmatrix} \cdot & \pi_C a & \pi_G b & \pi_T c \\ \pi_A a & \cdot & \pi_G d & \pi_T e \\ \pi_A b & \pi_C d & \cdot & \pi_T f \\ \pi_A c & \pi_C e & \pi_G f & \cdot \end{bmatrix}, \quad \mathbf{f} = [\pi_A \ \pi_C \ \pi_G \ \pi_T]$$



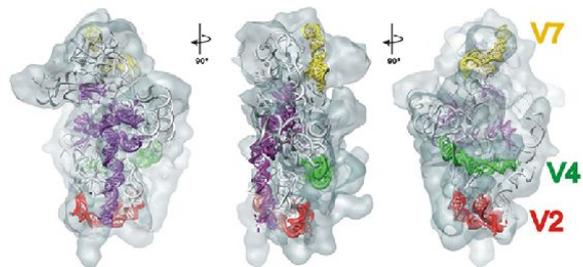
Чем хороши и чем плохи сложные и простые модели?

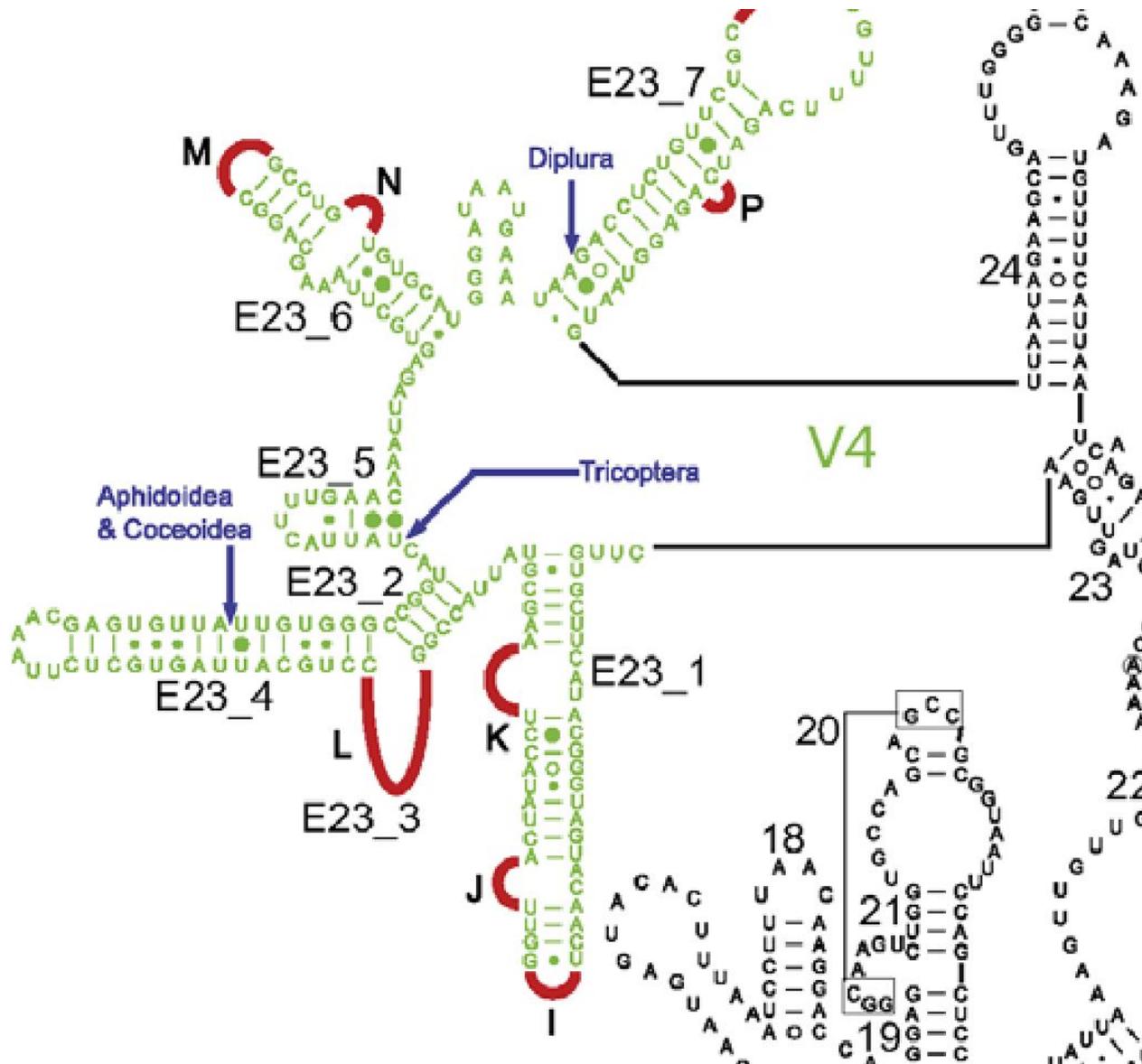


- Это стохастические модели, которые предполагают, что все замены случайны и независимы друг от друга
- А если нет ...

A

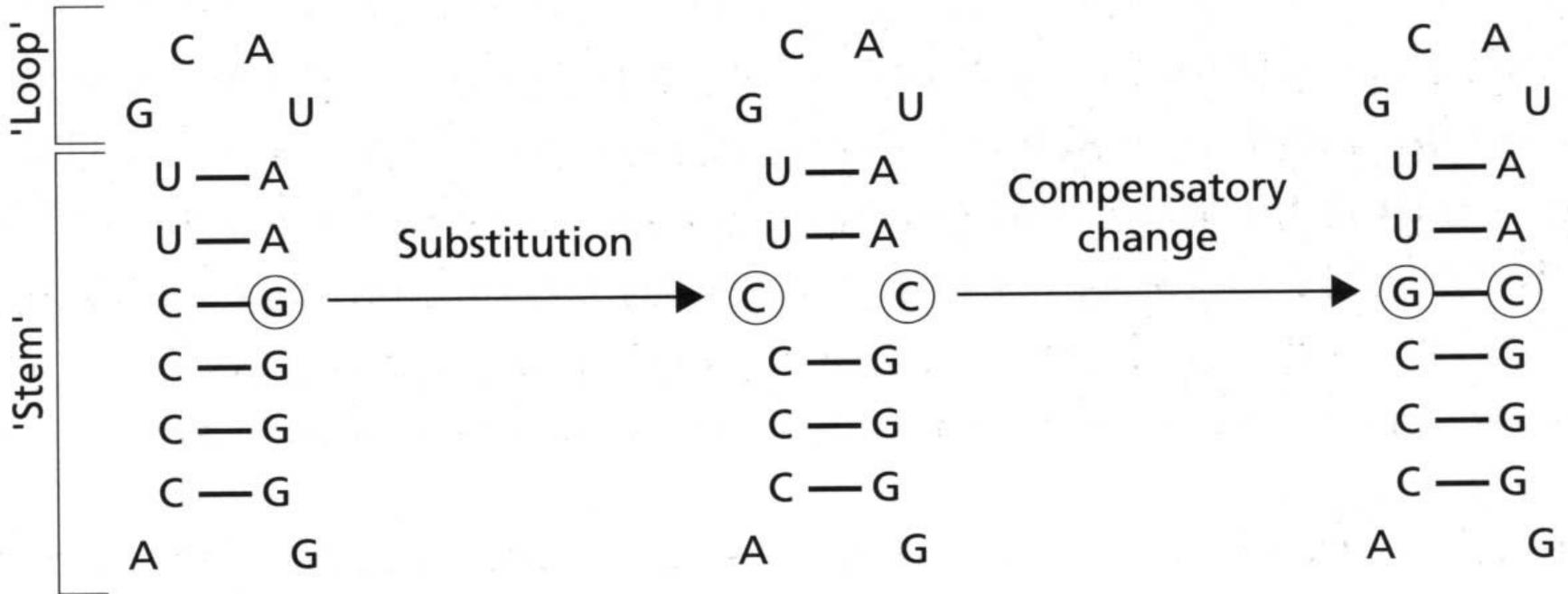
Структура 18S rDNA

B



18S rDNA
(фрагмент)

- Все замены ^{Условия, при которых работают эти модели}случайны и независимы друг от друга
- А если нет ...
- Зависимые компенсаторные замены



Общие принципы построения филогений

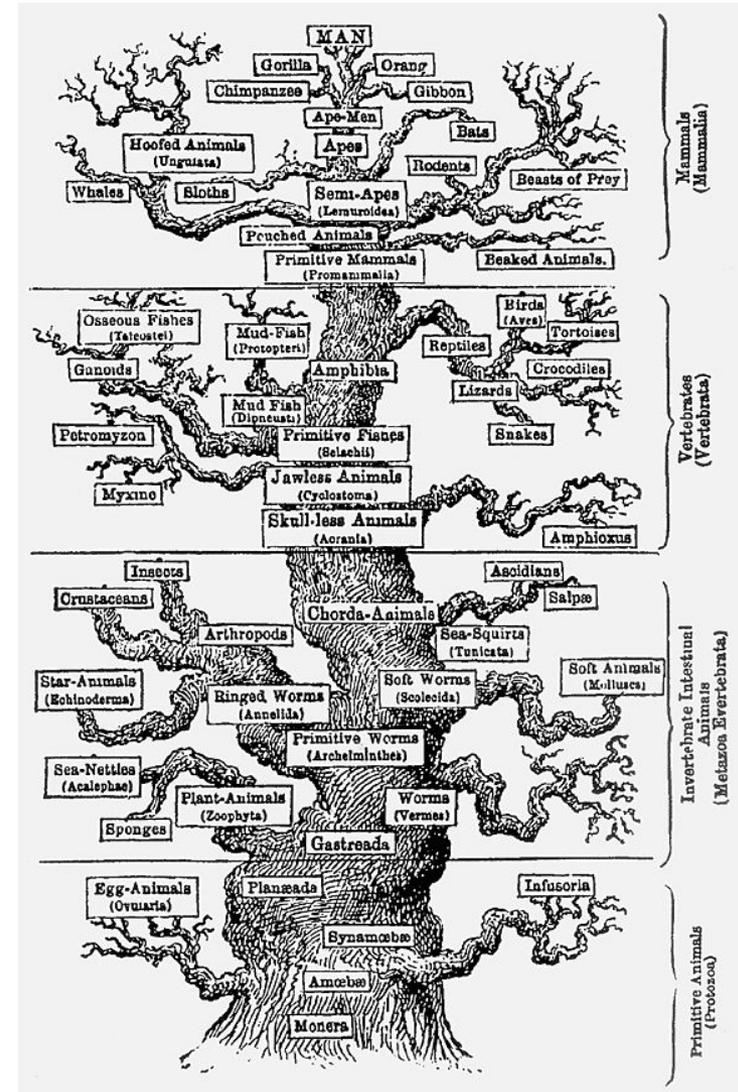
- 1) Анализ признаков,
- 2) выбор оптимальной модели эволюции признака,
- 3) выбор методов и алгоритмов для построения дерева**

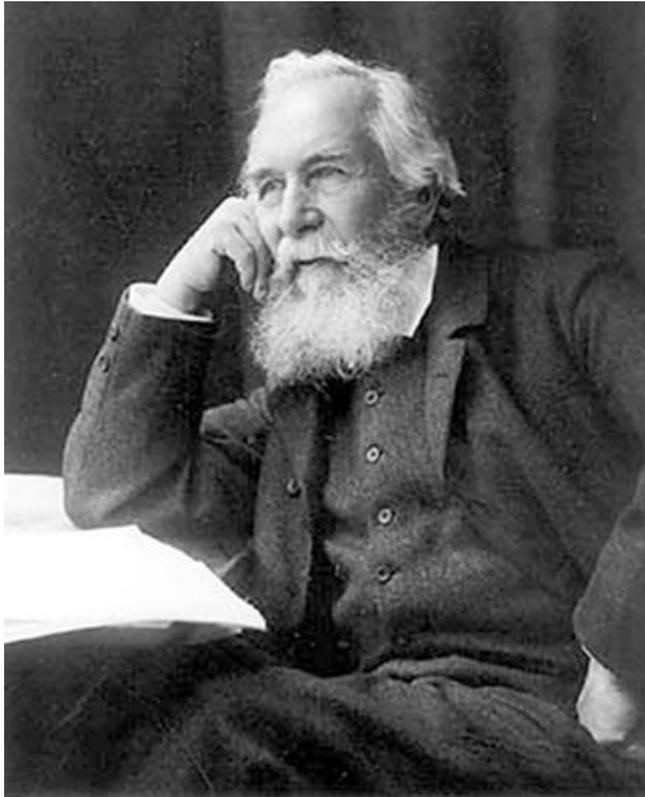
Подходы к выявлению филогений

- **традиционный** (Геккелевский, эмпирико-интуитивный)
- **традиционная кладистика** (Hennig, 1950, 1966)
- **фенетика**
- **метод максимальной парсимонии**
- **метод максимального правдоподобия**
- **метод Байеса**
- **методы, основанные на анализе генетических дистанций**

Традиционный (эмпирико-интуитивный) метод выведения филогений

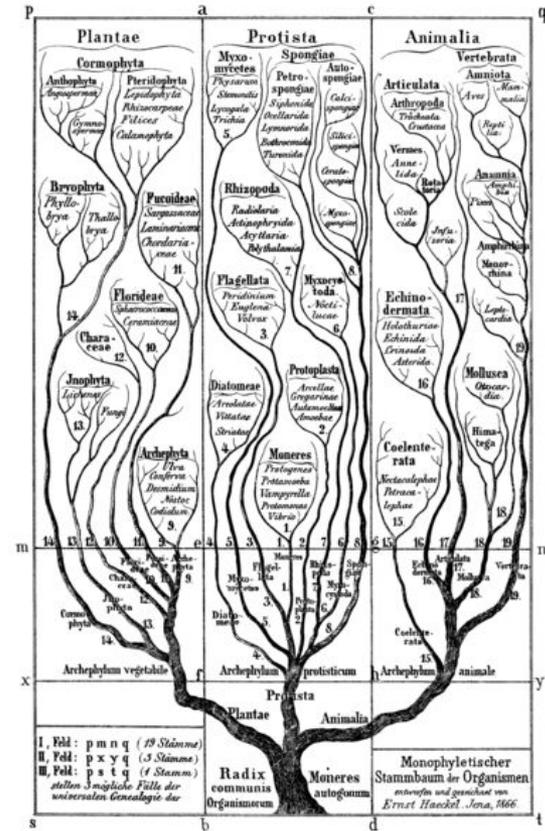
- Строго **научный** и, как правило, очень качественный **анализ признаков** сочетается с частично или полностью **интуитивным методом** их филогенетического **обобщения**,
 - то есть с отсутствием универсальных, четких и формализованных алгоритмов филогенетического анализа
 - Модели эволюции примитивны и не формализованы



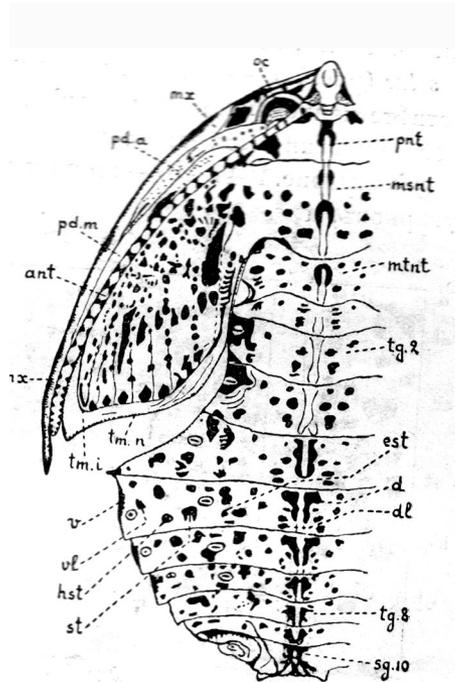


Ernst Haeckel (1834-1919)

По Геккелю филогенетика - наука о путях, закономерностях и причинах исторического развития организмов



Н.Я.Кузнецов. Насекомые чешуекрылые. Т. 1. Фауна России.
Петроград, 1915



великолепный анализ
морфологии

Выявление гомологий

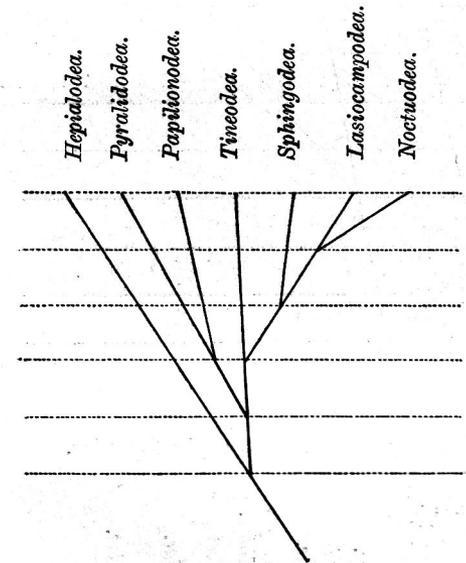


Рис. 1.—Схема взаимоотношений между подотрядами и сериями *Lepidoptera.*

Однако обоснование филогений
ограничивается словами: «Я
предлагаю принять
филогенетические отношения,
представленные на рисунках»

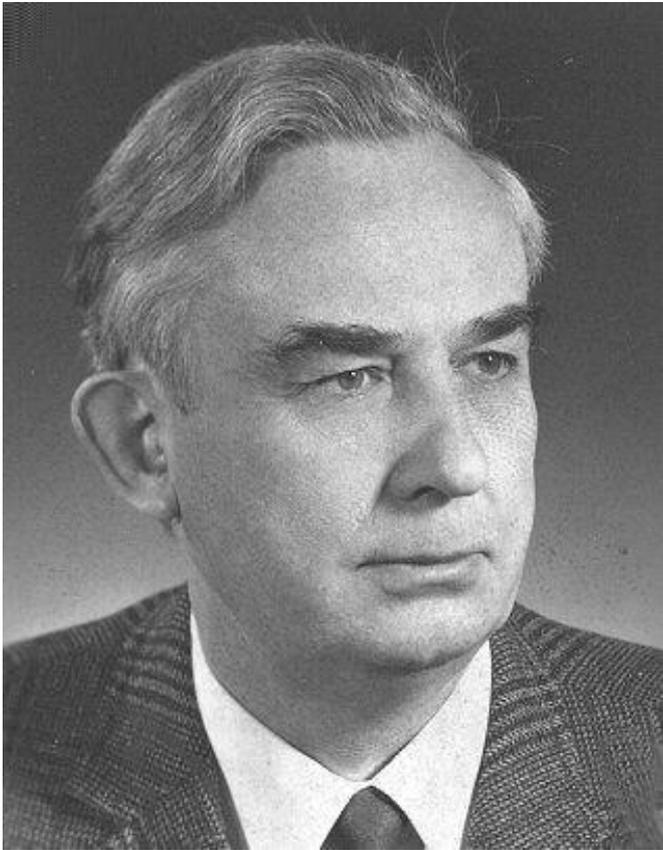
“Недавно в лабораторию [Моргана] пришла почта с произведениями Северцова с многочисленными филогенетическими деревьями, на которые я указал Моргану. Его реплика была такова: “Я думал, что такие идиоты могут существовать только в *Museum of Natural History*”. После этого я со сладострастием наблюдал, как все это пошло на свалку”

Ф.Г. Добржанский (из письма к
Филиппенко, 23 июля 1928)

Ю.А.



Ф.Г.Добржанский
фото 1935 г.



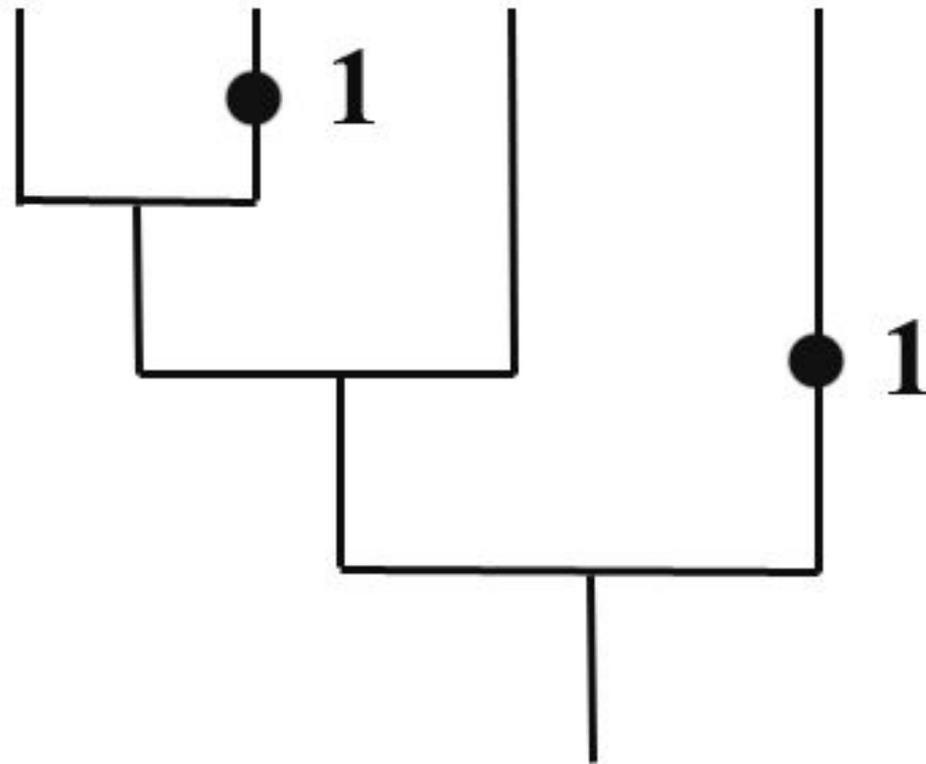
**Willi Hennig
(1913-1976)**

**Традиционная кладистика
(Hennig, 1950, 1966)**

**Хенниг предложил строго
научные принципы
перехода от анализа
признаков к реконструкции
филогений**

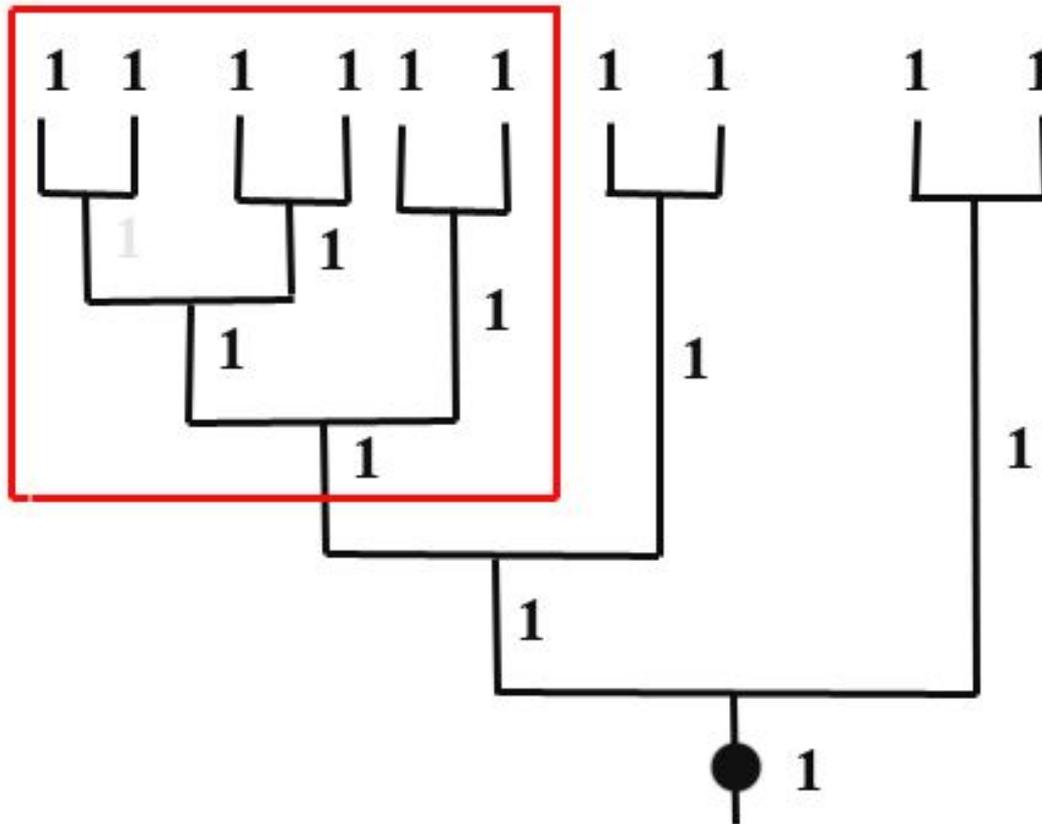
- Признаки
 - Негомологичные (гомоплазии)
 - Гомологичные
 - Плезiomорфии
 - Апоморфии
 - Синапоморфии

Гомоплазии - независимо возникшие признаки. Они не несут никакой информации о филогении

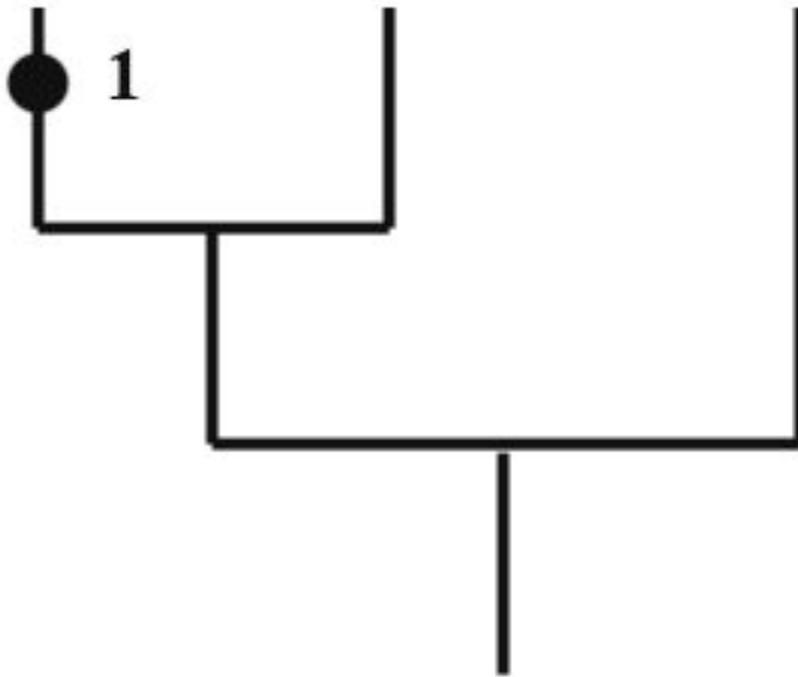


1 - гомоплазия

Плезиоморфии - древние (=исходные; =примитивные) гомологичные признаки. Они не несут никакой информации о топологии поздних ветвлений.

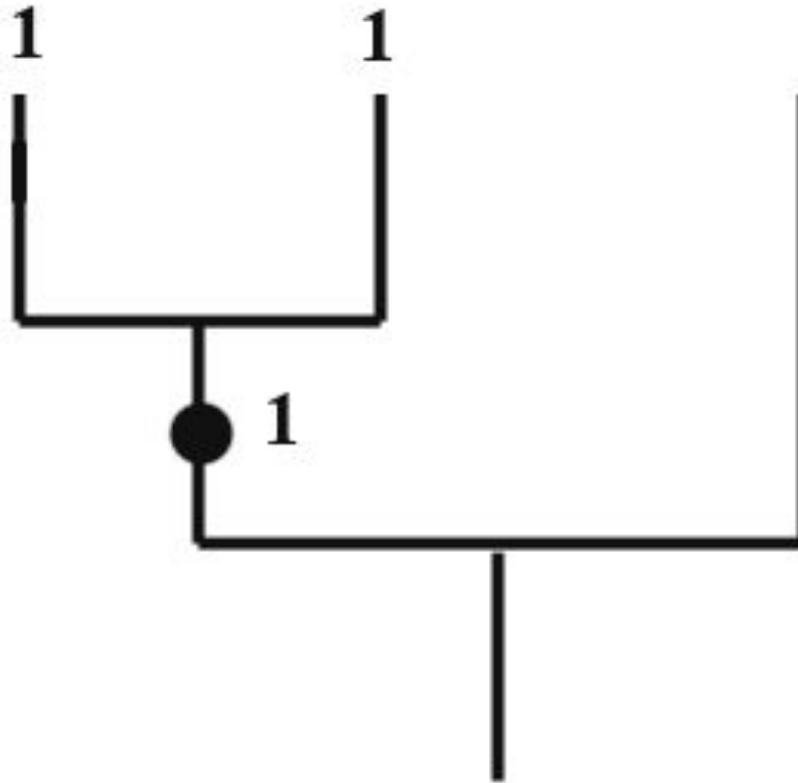


Апоморфия - новый
(= продвинутый; = производный;
= прогрессивный)
гомологичный признак.



Апоморфия является специфическим маркером эволюционной линии

Единичная апоморфия, возникшая в концевой ветви, метит только эту ветвь и не несет никакой информации о топологии

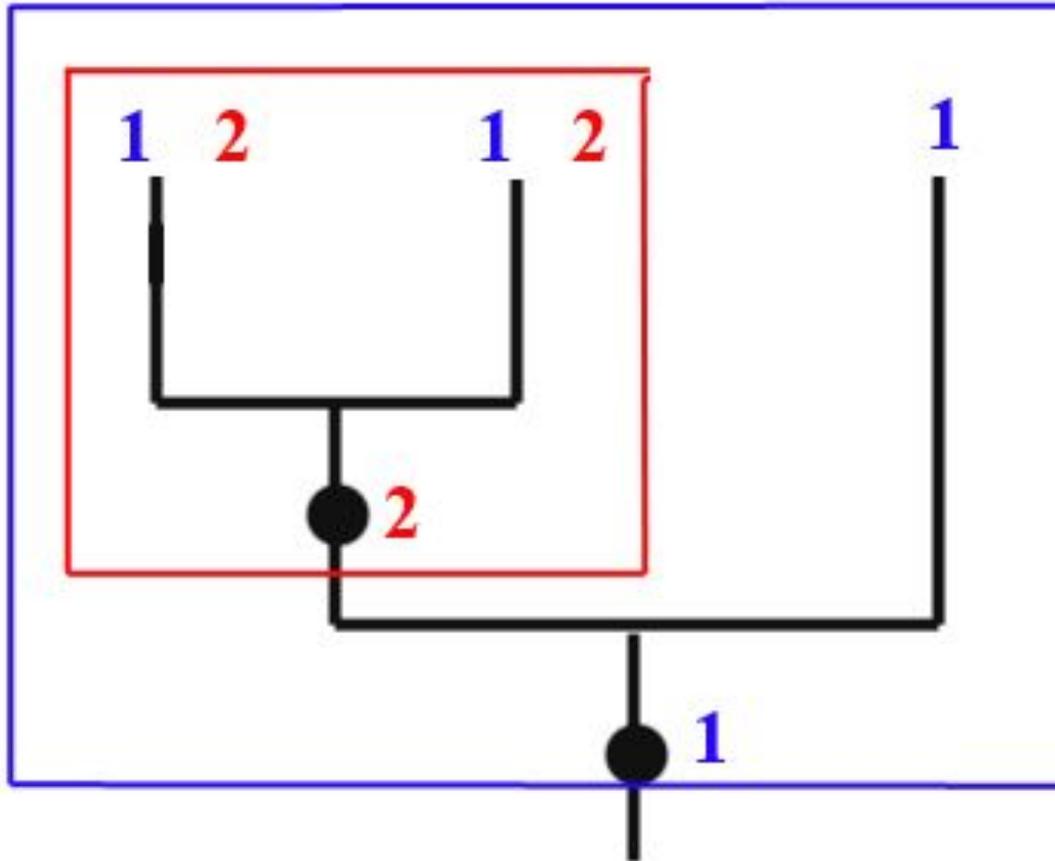


Но если апоморфия возникла до разделения ветвей и передалась в обе ветки, то наличие такой апоморфии указывает на существование клады, состоящей из двух таксонов.

*Такая апоморфия называется **синапоморфией**.*

Синапоморфия несет информацию о филогении!!!

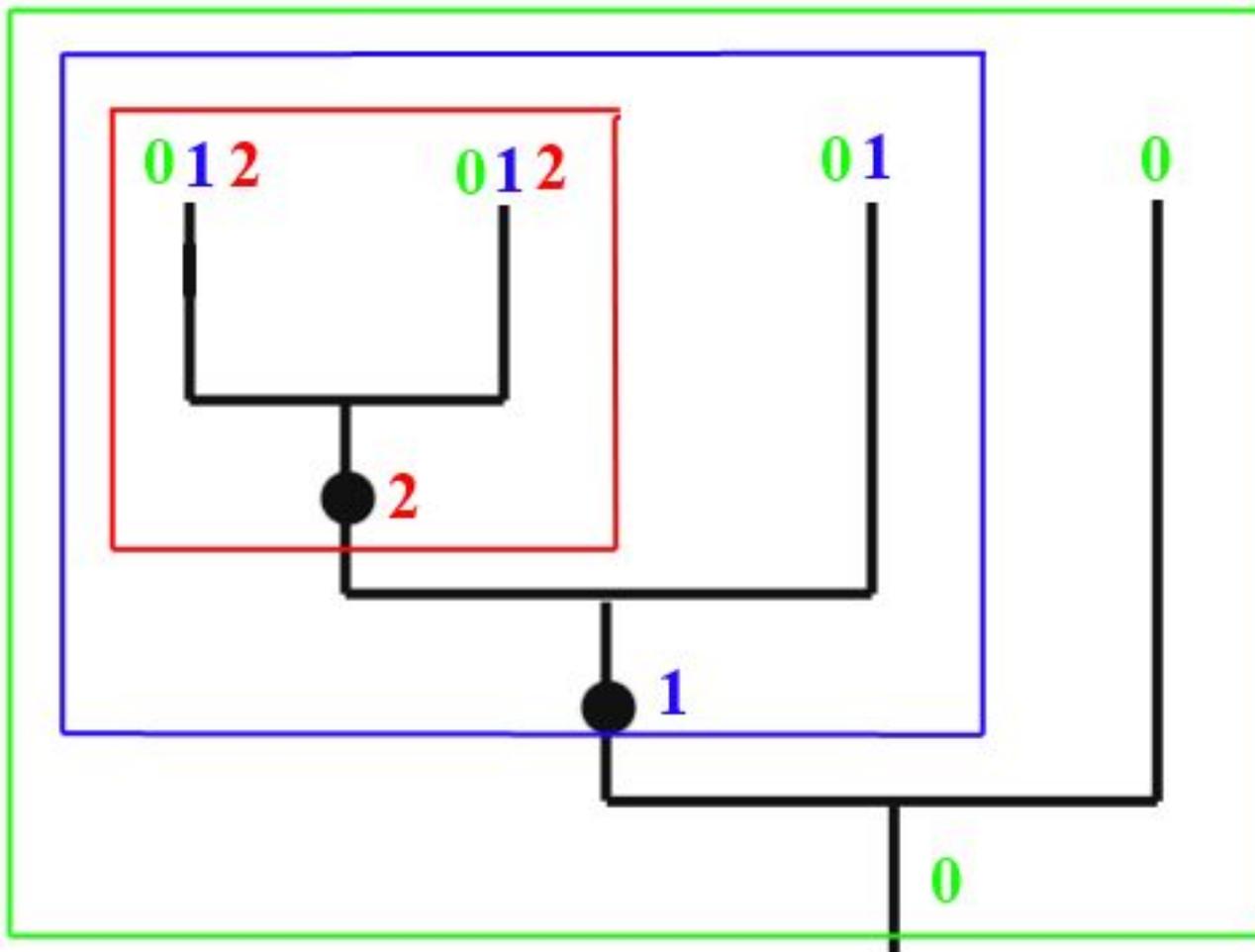
Для построения
необходимо



я)

Для построения филогении трех таксонов (два ветвления)
необходимо наличие одной синапоморфии

В общем виде
ветвления
на каждом



ещей n
одной

В общем виде для полного разрешения филогении, включающей n ветвлений, необходимо и достаточно $n-1$ синапоморфий (по одной на каждый узел, кроме базального)

- Филогения строится как система соподчиненных (вложенных одна в другую) клад (монофилетических групп), каждая из которых выявляется по наличию синапоморфий

- Модель эволюции в кладистике по Геннигу
 - Топология - строгая дихотомия
 - Процесс - накопление синапоморфий.

Алгоритм анализа

- Одна истинная синапоморфия может разрешить узел ветвления филогенетического дерева
- Выявление филогении – многоступенчатый процесс выдвижения и тестирования филогенетических гипотез, в ходе которого представление о филогенезе постепенно уточняется и конкретизируется

Построение молекулярного дерева с использованием кладистики по Хеннигу

- 1 AAGT
- 2 AAGT
- 3 ACGT

Построение молекулярного дерева таксонов 1-4 с использованием клатистики по Хеннигу

- 1 AAGTT
- 2 AAGTT
- 3 ACGTT
- 4 ACGTA
- 5 ACGTA
- 6 ACGTA
- 7 ACGTA

Состояние АСГТА плезиморфно

- 1 ААГТТ
- 2 ААГТТ
- 3 АСГТТ
- 4 АСГТА
- 5 АСГТА
- 6 АСГТА
- 7 АСГТА

А во второй позиции - синапоморфия 1 + 2

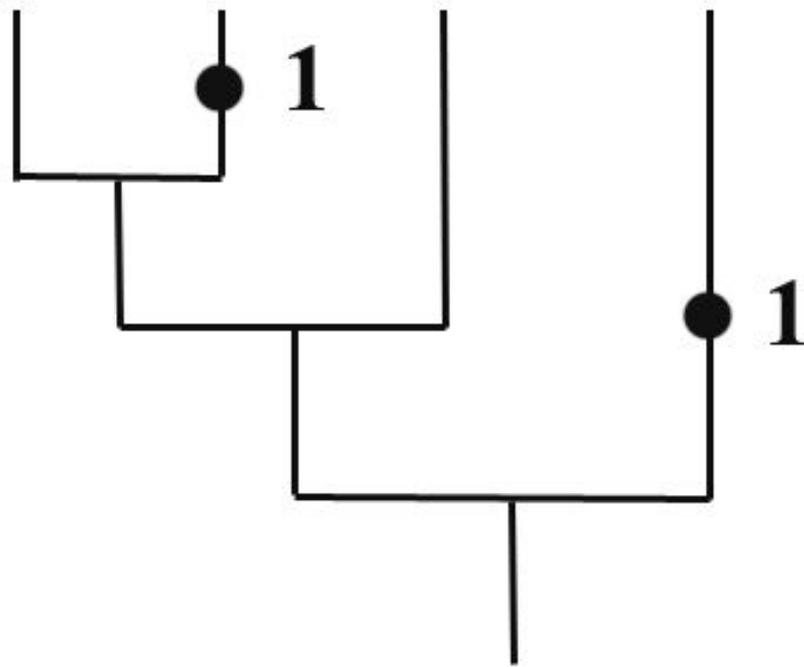
- 1 ААGТТ
- 2 ААGТТ
- 3 АСGТТ
- 4 АСGТА
- 5 АСGТА
- 6 АСGТА
- 7 АСGТА

Т в пятой позиции -
синапоморфия 1 + 2 + 3

- 1 AAGTT
- 2 AAGTT
- 3 ACGTT
- 4 ACGTA
- 5 ACGTA
- 6 ACGTA
- 7 ACGTA

Проблема гомоплазий

- Презумпция: Синапоморфии встречаются чаще, чем гомоплазии



Конфликт между потенциальными синапоморфиями

- 1 AAGTT
- 2 AACTT
- 3 ACCTT
- 4 ACGTT

Принципы традиционной кладистики

- Если возникает конфликт между потенциальными синапоморфиями, то основной путь его решения – переисследование материала, поиск и изучение дополнительных признаков и таксонов

Другие проблемы генниговской кладистики:

- “Надежных” синапоморфий может быть мало, недостаточно для того, что разрешить все узлы ветвления разрабатываемой филогении

Проблемы традиционной кладистики

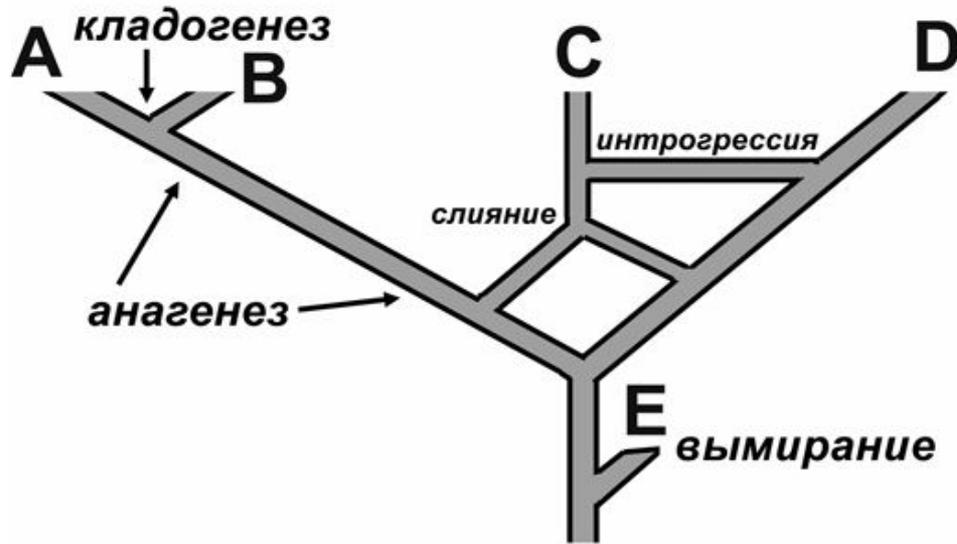
- “Надежных” синапоморфий может быть мало, недостаточно для того, что разрешить все узлы ветвления разрабатываемой филогении
- отбрасывая «ненадежные» признаки, мы теряем филогенетическую информацию, так как «ненадежные» признаки также могут содержать филогенетический сигнал

Проблемы традиционной кладистики

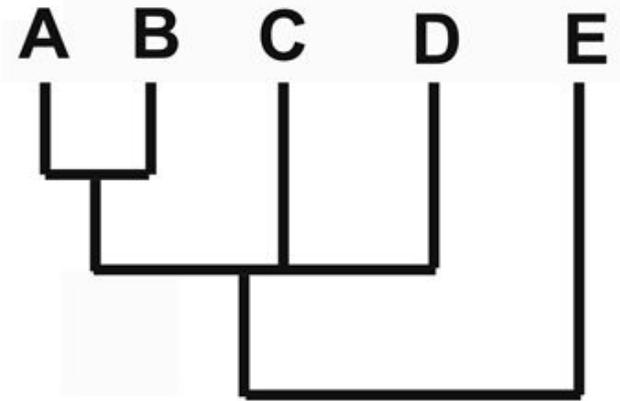
- “Надежных” синапоморфий может быть мало, недостаточно для того, что разрешить все узлы ветвления разрабатываемой филогении
- отбрасывая «ненадежные» признаки, мы теряем филогенетическую информацию, так как «ненадежные» признаки также могут содержать филогенетический сигнал
- Реконструируется только топология!

Картинь филогенезов, которе создает кладистический (по Геннигу и парсимониальньй) анализ, неполнь и однобоки:

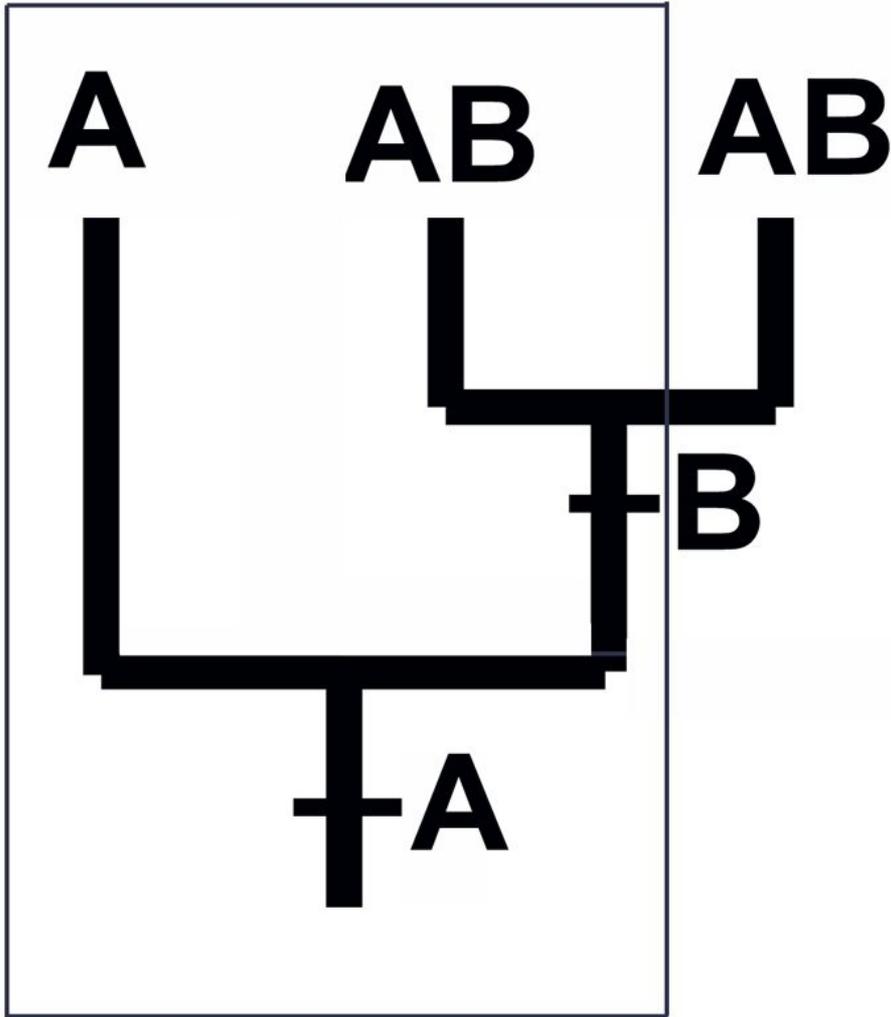
- Анагенез не учитывается
- Ретикулогенез (слияния+интрогрессии) не выявляется
- Некоторые узлы принципиально не могут быть выявлены



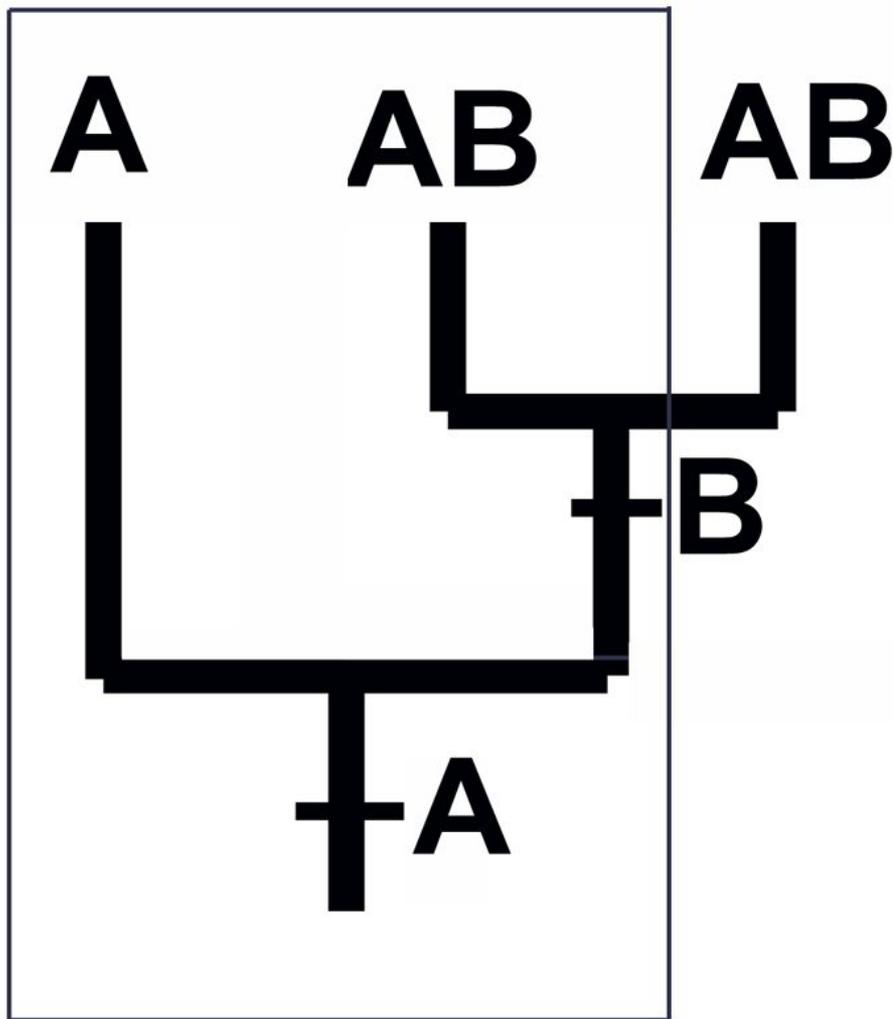
филогения



кладаграмма



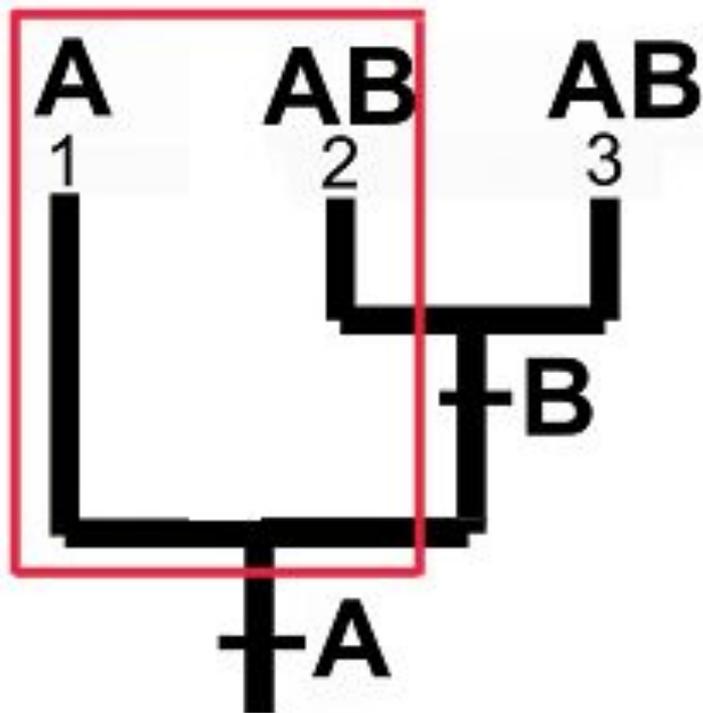
Принцип монофилии лежит в самой основе алгоритма построения дерева в хенниговской кладистике. Сипапоморфии однозначно определяют только монофилетические линии, а немонофилетические группы, например, парафилетические группировки не могут быть определены однозначно.



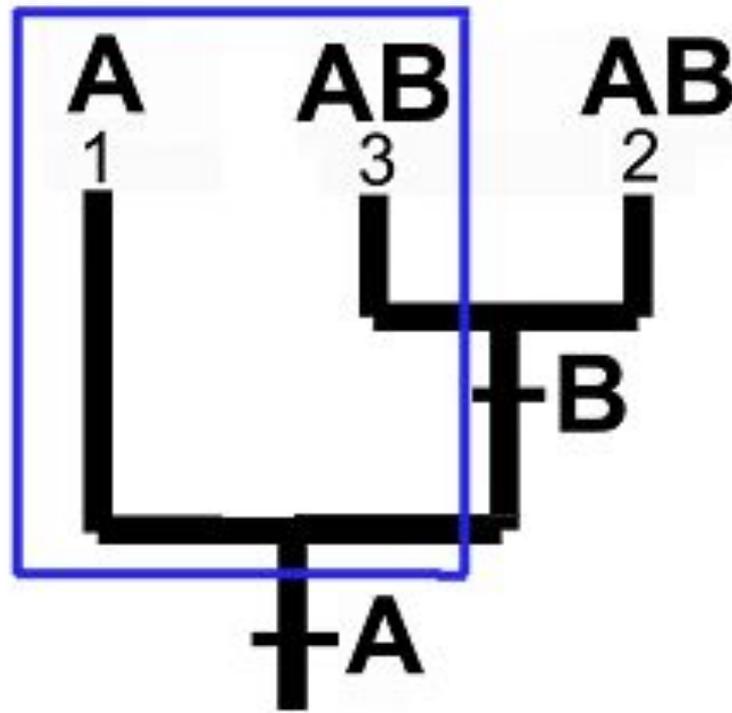
Кладизм объявляет парафилетические группы вне закона просто по той причине, что он не умеет их выявлять

(поскольку парафилетические группы не имеют синапоморфий)

Проблемы парафилетических таксонов



1+2 = парафилетический таксон.
Признак A не уникален, признак B
характеризует лишь часть таксона
1+2 и тоже не уникален

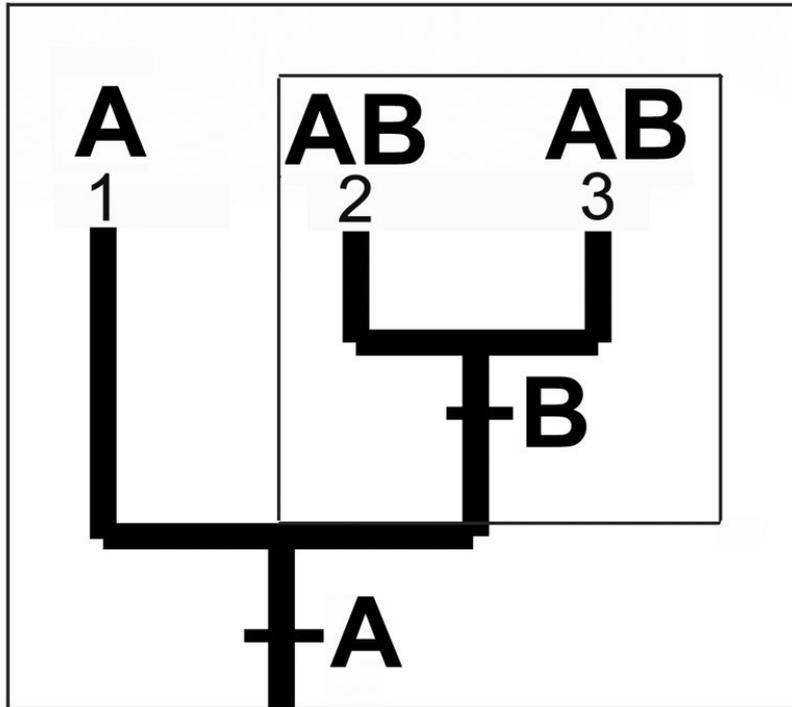


1+3 = парафилетический таксон.
Признак A не уникален, признак B
характеризует лишь часть таксона
1+3 и тоже не уникален

Существует несколько вариантов частично
пересекающихся парафилетических таксонов

Монофилетический таксон - группа, которая включает предка и всех его потомков

Монофилетические группы могут иметь синапоморфии

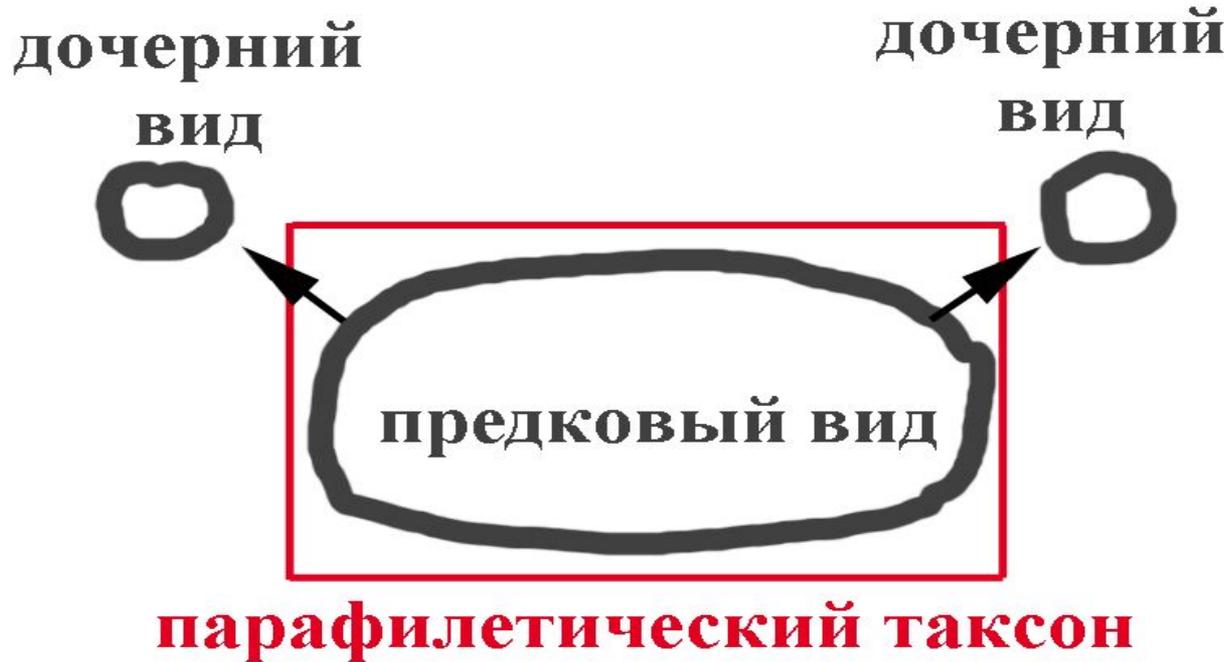


A – это синапоморфия таксона $1+(2+3)$
→ A однозначно характеризует таксон $1+(2+3)$

B, синапоморфия таксона $2+3$
→ B однозначно характеризует таксон $2+3$

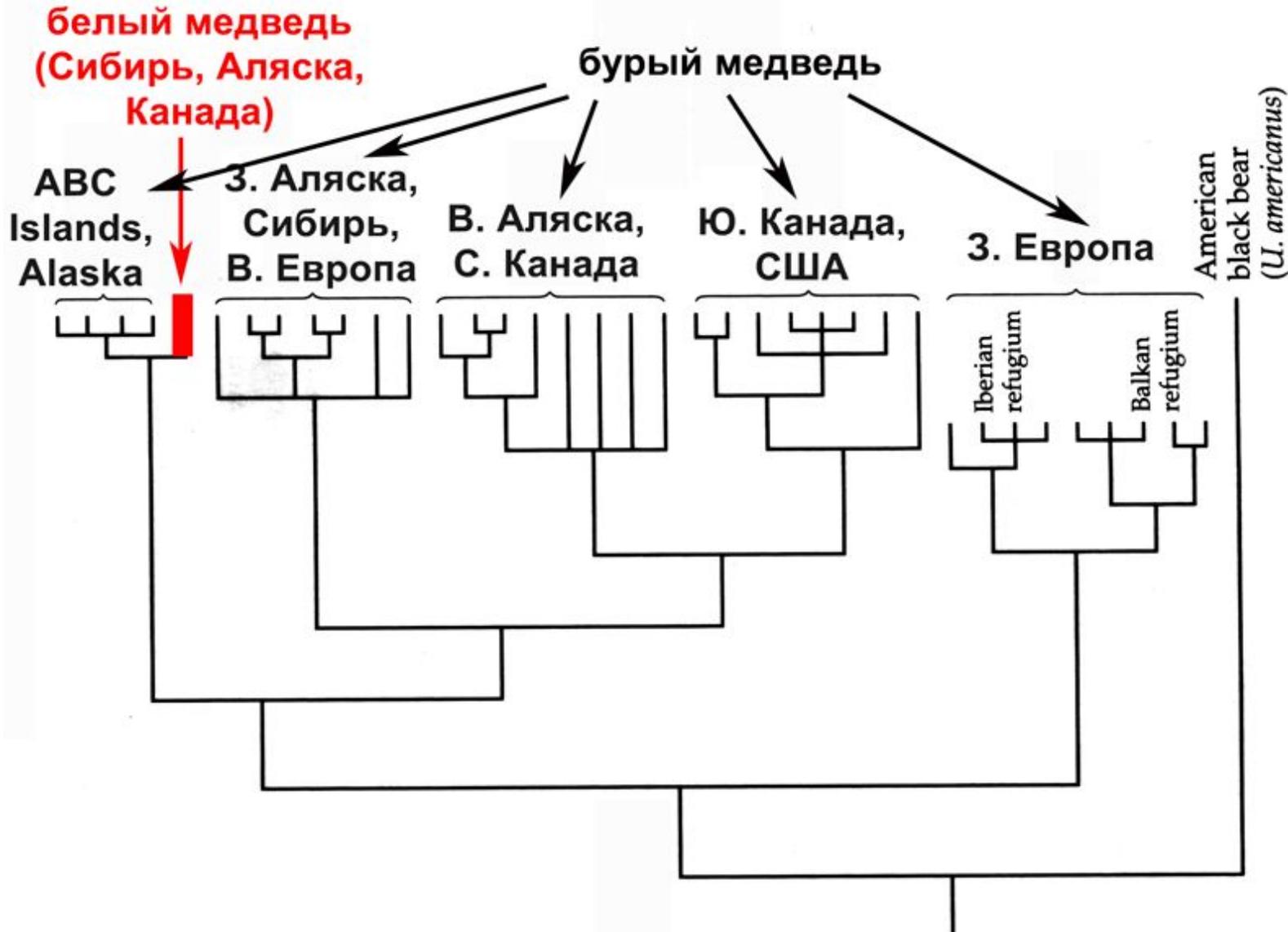
Другие варианты монофилетических таксонов не существуют

Перипатрическое видообразование: предковый таксон при этом не исчезает, но он становится парафилетическим.



Несмотря на парафилию, такой вид представляет собой единое репродуктивное сообщество, изолированное от дочерних видов

Филогеография медведей, основанная на
кладистическом анализе (MP) нуклеотидных замен в
митохондриальном геноме (Avice, 2004)



- Кладистика по Геннигу остается рабочим инструментом филогенетики!

Фенетика

В кладистике процедура выявления гомологичных признаков (дифференциация от гомоплазий) не формализована. Это может быть причиной субъективизма

Фенетика

- 📌 Отказ от доминирования принципа гомологии (в фенетике все признаки имеют равный вес)
- 📌 Степень родства = степени сходства
- + попытка ввести объективность в систематику и филогенетику
- + широкое внедрение методов статистики в систематику

Фенетика

- Кластерный анализ (выявление группировок по степени их сходства).
- Иерархии таких группировок можно интерпретировать в качестве филогении.

Фенетика

- Пример научной, но неправильной (неадекватной) методологии
- Научность - строгое следование принципам научной логики, избегание субъективизма
- Неправильность - основана на неадекватной аксиоматике (на ложных предпосылках)

Традиционная и нумерическая кладистика

- Увеличение числа признаков приводит к противоречиям между предполагаемыми синапоморфиями, которые свидетельствуют о наличии гомоплазий
- При наличии противоречий между "синапоморфиями" возможны разные варианты филогении
- Как выбрать правильный вариант?
-

- Если возникает конфликт между потенциальными синапоморфиями, то есть два пути его решения:
- 1) переисследование материала, поиск и изучение дополнительных признаков и таксонов с целью выявления "истинных" синапоморфий -
Традиционная кладистика

- Если возникает конфликт между потенциальными синапоморфиями, то есть два пути его решения:
- 1) переисследование материала, поиск и изучение дополнительных признаков и таксонов с целью выявления "истинных" синапоморфий
- 2) наоборот - использование большого числа признаков, получение нескольких (многих) деревьев и выбор "лучшего" из них с использованием определенного критерия - нумерическая кладистика

Нумерическая кладистика и метод максимальной парсимонии

- При наличии противоречий между “синапоморфиями” возможны разные варианты филогении
- Как выбрать “правильное” дерево?
 - - критерий максимальной парсимонии

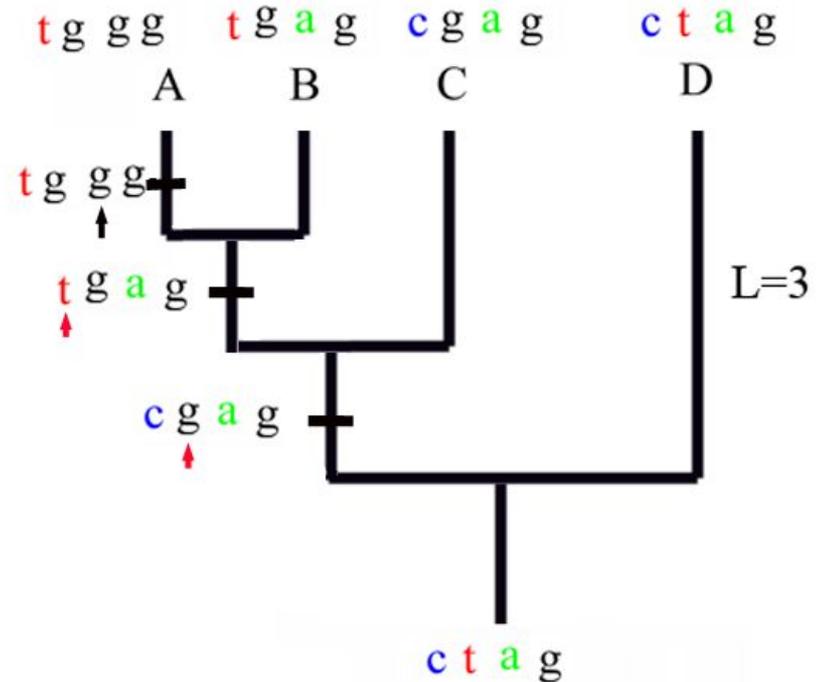
Метод максимальной парсимонии
(наибольшей экономии)

Нет гомоплазий - одно возможное дерево

Таксон\сайт	1	2	3	4
A	t	g	g	g
B	t	g	a	g
C	c	g	a	g
D	c	t	a	g

- c(1) симплезиоморфия
- t(1) – синапоморфия
- t(2) плезиоморфия
- g(2) – синапоморфия
- a(3) симплезиоморфия
- g(3) – аутапоморфия
- g(4) симплезиоморфия

Сайты 1 и 2 – парсимониально информативные
 Сайты 3 и 4 – парсимониально неинформативные

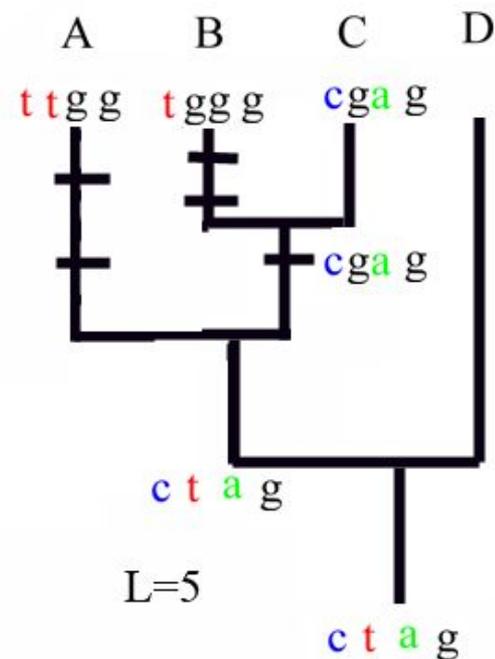
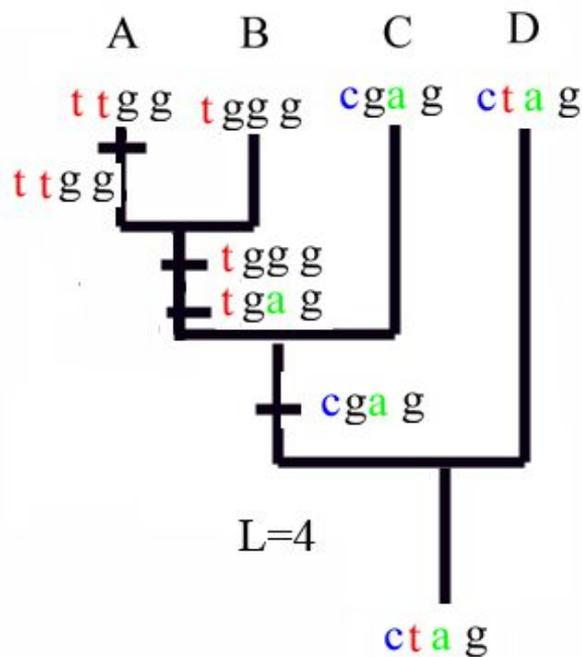


Число шагов (L) = 3

Сайт 4 – инвариантный, сайт 3 - переменный

Первое дерево более парсимониальное, оно короче
 Происходит голосование "синапоморфиями"

Таксон\сайт	1	2	3	4
A	t	t	g	g
B	t	g	g	g
C	c	g	a	g
D	c	t	a	g



- в реальности у нас исходно нет ни топологии дерева, ни распределения признаков по нему, ни анцестрального состояния.
- Как все это найти?

AY557140 A achaemenes voucher	ATAACATAAGATTCTGATTACTACCCCATCA
EF104615 A achaemenes voucher	ATAATATAAGATTGTGA-TACTACCCCATCA
EF104621 A actinides voucher V	ATAACATAAGATTCTGATTATTACCACCATCA
LOWA513-06 2005-LOWA-513 Agrod	ATAACGTAAGATTCTGATTATTACCACCATCA
LOWA503-06 2005-LOWA-503 Agrod	ATAACATAAGATTCTGATTATTACCACCATCA
LOWA502-06 2005-LOWA-502 Agrod	ATAACAAAAGATTCTGATTATTACCACCATCA
LOWA501-06 2005-LOWA-501 Agrod	ATAACATAAGATTCTGATTATTACCACCATCA
LOWA512-06 2005-LOWA-512 Agrod	ATAACATAAGATTCTGATTATTACCACCATCA
LOWA511-06 2005-LOWA-511 Agrod	ATAACATAAGATTCTGATTATTACCACCATCA

Шаг 1: выявление признаков и их состояний

Признак – цвет глаз

Состояния – коричневый, голубой, зеленый

Признак – группа крови

Состояния – первая, вторая, третья, четвертая

Шаг 2: кодирование признаков и их состояний

Признак – цвет глаз

Состояния – коричневый (0), голубой (1), зеленый (2)

Признак – группа крови

Состояния – первая (0), вторая (1), третья (2), четвертая (3)

0 – обычно анцестральное состояние

Шаг 3: Составление матрицы признаков

Table 1.1: A simple data set with 0/1 characters.

Species	Characters					
	1	2	3	4	5	6
Alpha	1	0	0	1	1	0
Beta	0	0	1	0	0	0
Gamma	1	1	0	0	0	0
Delta	1	1	0	1	1	1
Epsilon	0	0	1	1	1	0

Бинарная матрица

Table 1.1: A simple data set with 0/1 characters.

Species	Characters					
	1	2	3	4	5	6
Alpha	1	0	0	1	1	0
Beta	0	0	1	0	0	0
Gamma	1	1	0	0	0	0
Delta	1	1	0	1	1	1
Epsilon	0	0	1	1	1	0

Матрица множественных состояний

Species	Characters					
	1	2	3	4	5	6
Alpha	1	0	0	1	1	0
Beta	0	0	1	0	0	3
Gamma	1	1	2	0	0	0
Delta	2	1	3	1	1	1
Epsilon	0	2	1	1	1	0

- Нуклеотидное (или аминокислотное) выравнивание - это уже готовая матрица признаков
- 4 состояния - **A C G T**

AY557140 A achaemenes voucher	ATAACATAAGATTCTGATTACTACCCCATCA
EF104615 A achaemenes voucher	ATAATATAAGATTGTGA-TACTACCCCATCA
EF104621 A actinides voucher V	ATAACATAAGATTCTGATTATTACCACCATCA
LOWA513-06 2005-LOWA-513 Agrod	ATAACGTAAGATTCTGATTATTACCACCATCA
LOWA503-06 2005-LOWA-503 Agrod	ATAACATAAGATTCTGATTATTACCACCATCA
LOWA502-06 2005-LOWA-502 Agrod	ATAACAAAAGATTCTGATTATTACCACCATCA
LOWA501-06 2005-LOWA-501 Agrod	ATAACATAAGATTCTGATTATTACCACCATCA
LOWA512-06 2005-LOWA-512 Agrod	ATAACATAAGATTCTGATTATTACCACCATCA
LOWA511-06 2005-LOWA-511 Agrod	ATAACATAAGATTCTGATTATTACCACCATCA

Шаг 4: выбор модели эволюции

- Модель Камина-Сокола (Camin- Sokal parsimony):
анцестральное состояние известно, тогда $0 \rightarrow 1$
- Всегда дает укорененное дерево

- Модель Долло (Dollo parsimony) (основана на принципе необратимости эволюции) - допускаются изменения признака в любую сторону, но только один раз (вернее повторные изменения менее вероятны)

Модель Фитча-Вагнера (Fitch-Wagner parsimony) -
симметричная модель



дерево неукорененное!!!

Модель Фитча-Вагнера (Fitch-Wagner parsimony)
для множественных состояний признака

0 \longleftrightarrow 1

0 \longleftrightarrow 2

0 \longleftrightarrow 3

1 \longleftrightarrow 2

1 \longleftrightarrow 3

2 \longleftrightarrow 3

дерево неукорененное!!!

Модель Фитча-Вагнера (Fitch-Wagner parsimony) для нуклеотидных замен

$A \leftrightarrow C$ $A \leftrightarrow G$ $A \leftrightarrow T$

$C \leftrightarrow G$ $C \leftrightarrow T$

$G \leftrightarrow T$