

- Presumed common progenitor of all extant organisms
- Presumed common progenitor of archaeobacteria and eukaryotes

Гомологичные последовательности – последовательности, имеющие общее происхождение (общего предка).

Признаки гомологичности белков

- **сходная 3D-структура**
- **в той или иной степени похожая аминокислотная последовательность**
- **разные другие соображения...**

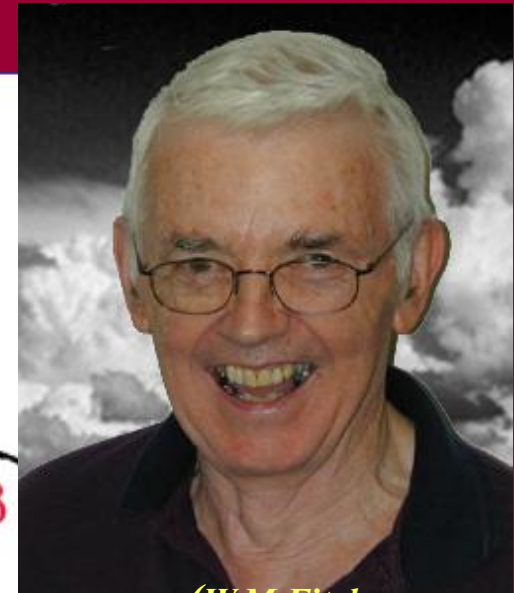
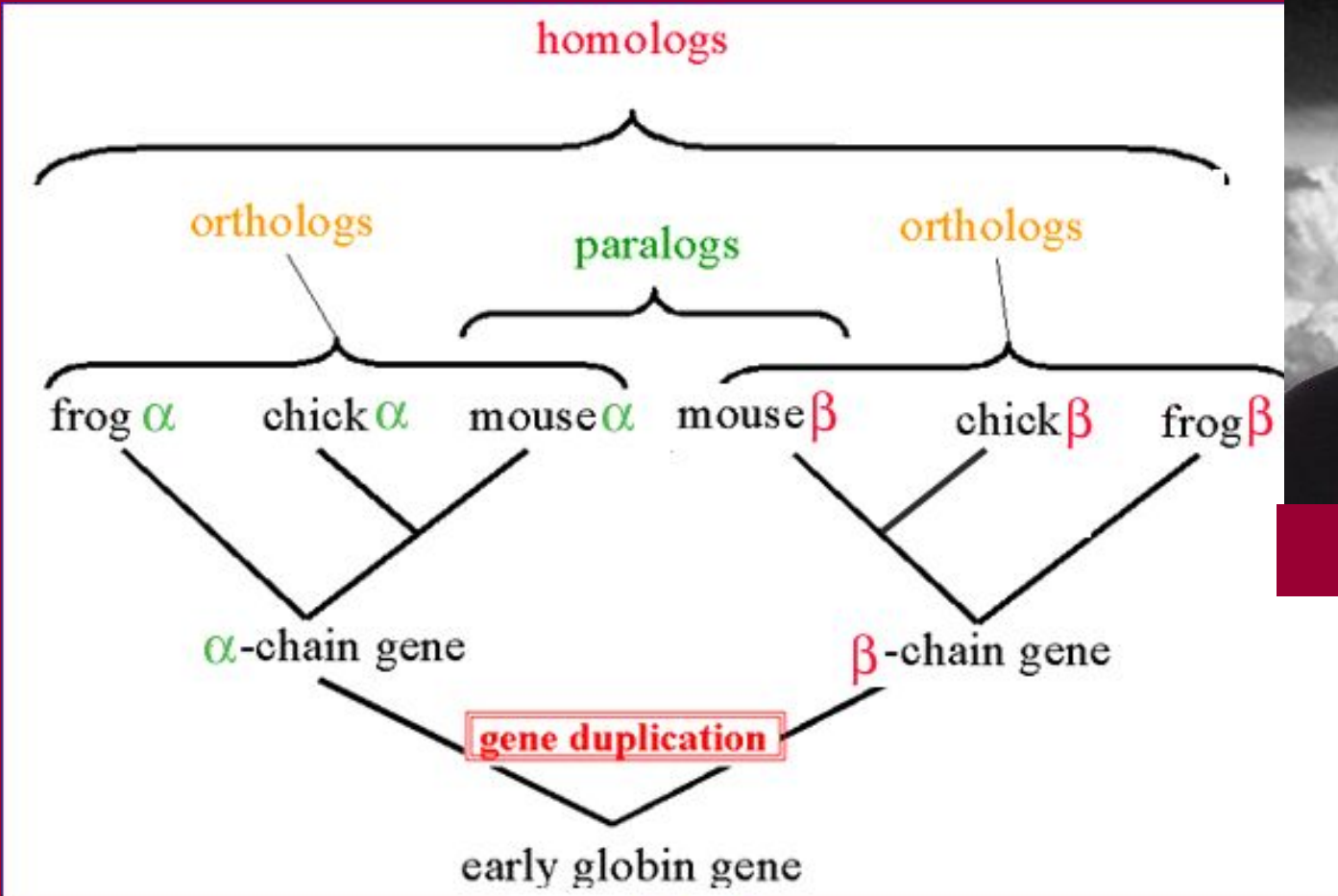
Гомологи

Ортологи

Паралоги

Ксенологи

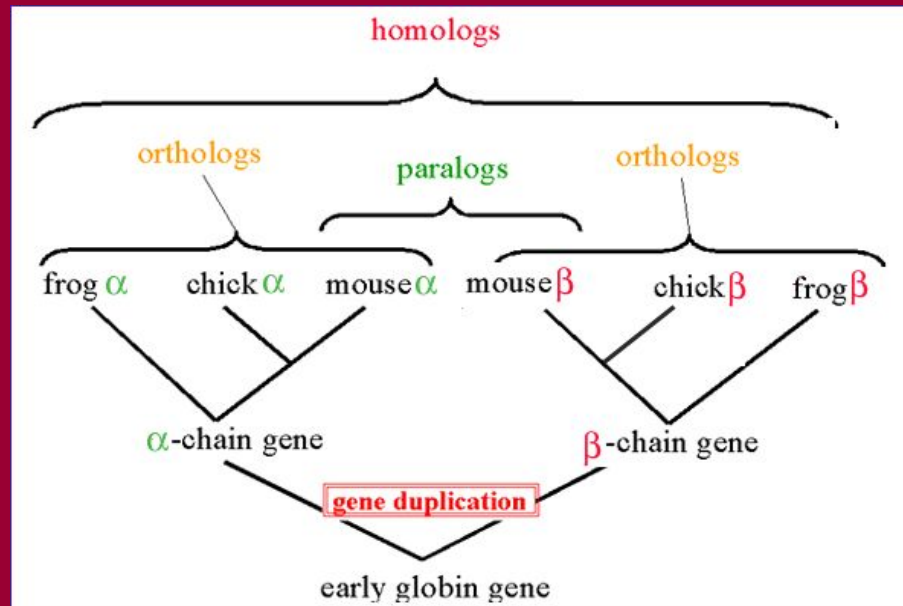
?



(W.M. Fitch,
Syst. Zool. 19, 99 (1970))

Ортологи – последовательности, возникшие из одного общего предшественника в процессе видообразования. Ортологи, как правило, имеют одну и ту же функцию

Паралоги – последовательности, возникшие из одного общего предшественника в результате дупликации одного гена в одном организме. Паралоги, как правило, имеют разные функции.



Задача выравнивания = задача поиска сходства. Сходство \neq гомология

«Идеальное» выравнивание – запись последовательностей одна под другой так, чтобы гомологичные фрагменты оказались друг под другом.

Дом

Скупидом

Водомерка ?

Домов---ой

Домо-строй

?

Домов-----ой

Домо-строй

Гэп – пропуск в последовательности

Как получить «оптимальное» выравнивание?

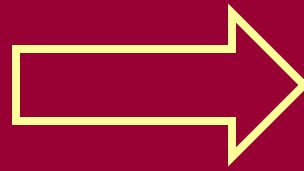
Один из способов:

```
>EC_Tr : MQNRLTI KDI ARLSGVVGKSTVSRVLNNEYR  
>EC_Fr : MKLDEI ARLAGVSR TTASYVI NGKAKQYR
```

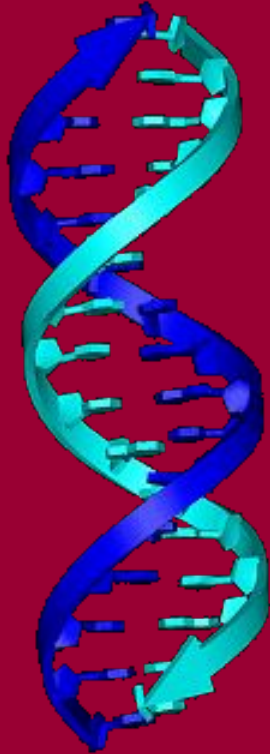
**Гэп – пропуск в
последовательности**

```
>EC_Tr : MQNRLTI KDI ARLSGVVGKSTVSRVLNNE---YR : 30  
>EC_Fr : ----MKLDEI ARLAGVSR TTASYVI NGKAKQYR : 29  
          | ARL GV T S V N YR
```

Мутации,
рекомбинация,



?



Example: point mutations

(a) Ile Cys Ile Lys Ala Leu Val Leu Leu Thr
 ATA TGT ATA AAG GCA CTG GTC CTG TTA ACA
 ATA TGT ATA AAG GCA CTG GTA CTG TTA ACA
 Ile Cys Ile Lys Ala Leu Val Leu Leu Thr

Silent mutation

(b) Ile Cys Ile Lys Ala Asn Val Leu Leu Thr
 ATA TGT ATA AAG GCA AAC GTC CTG TTA ACA
 ATA TGT ATA AAG GCA AAC TTC CTG TTA ACA
 Ile Cys Ile Lys Ala Asn Phe Leu Leu Thr

Amino acid change

(c) Ile Cys Ile Lys Ala Asn Val Leu Leu Thr
 ATA TGT ATA AAG GCA AAC GTC CTG TTA ACA
 ATA TGT ATA TAG GCAAACGTCCTGTTAACA
 Ile Cys Ile Ter

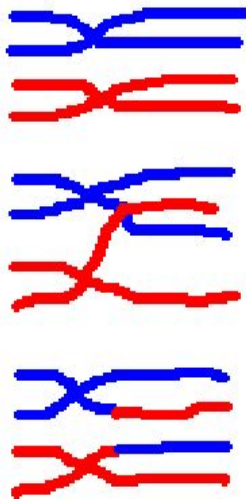
Termination

Ile Cys Ile Lys Ala Asn Val Leu Leu Thr
 ATA TGT ATA AAG GCA AAC ~~X~~GTC CTG TTA ACA
 ATA TGT ATA AAG GCA AAC TCC TGT TAA AC
 Ile Cys Ile Lys Ala Asn Ser Cys Stop

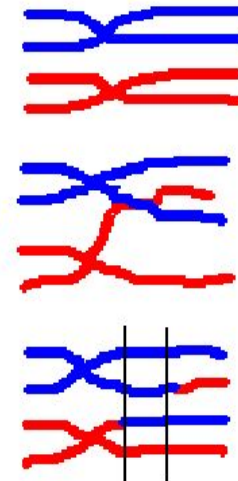
Frame shift

Recombination

Crossing over:
reciprocal recombination








Gene conversion:
Non-reciprocal recombination

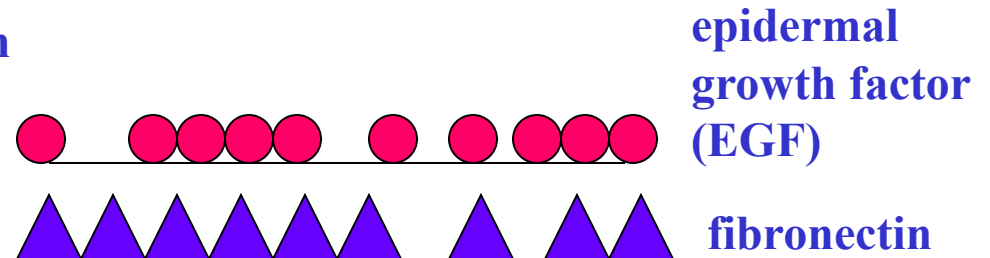


Domain insertion: "Mosaic proteins"

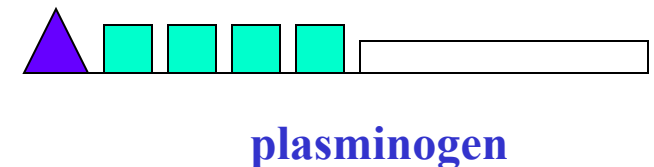
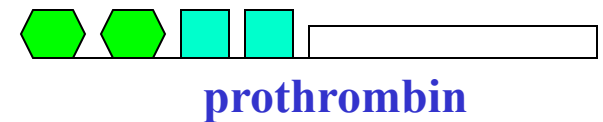
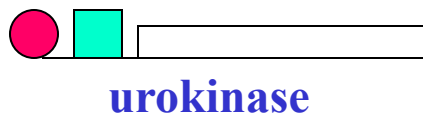
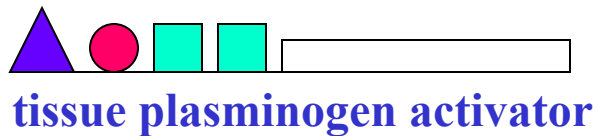
Structural modules:

-  plasminogen kringle domain
-  EGF domain
-  fibronectin finger domain
-  vit. K-dependent calcium-binding domain (osteocalcin)
-  trypsin-like serine protease

Domain origins:



Mosaic proteins



Домен – единица

**эволюции, структуры и функции
белков.**

**Домен – компактная,
относительно
независимо сворачивающаяся
структура,
относительно консервативная в
процессе
эволюции.**

**Белки могут состоять из одного или
многих доменов.**

Бывает и такое....

(пример структурного выравнивания)

```
.kqrrsrttFSASQLDELERAFert...qYPDIYTREELAQ
mdskrt rqtYTRYOTLELEKEFhfn...rYITRRRRRIDIAN
.mrrkrrvlFSQAQVYELERRREkqq...kYLSAPEREHLAS
.tkpyrghrFTKENVRILESWFakni enpYLDTKGLENLMK
.rkrkr rttISIAAKDALERHFgeh...sKPSSQEIMRMAE
.kspkgkssISPQARAFLEEVFrk...qSLNSKEKEEVAK
```

AAAAAAAAAAAA



AACDEAAAAA

Что лучше?

AAAAAAAAAAAA

AA---AAAAAAAA

AACDE-AAAA

AACDE-----AAAA