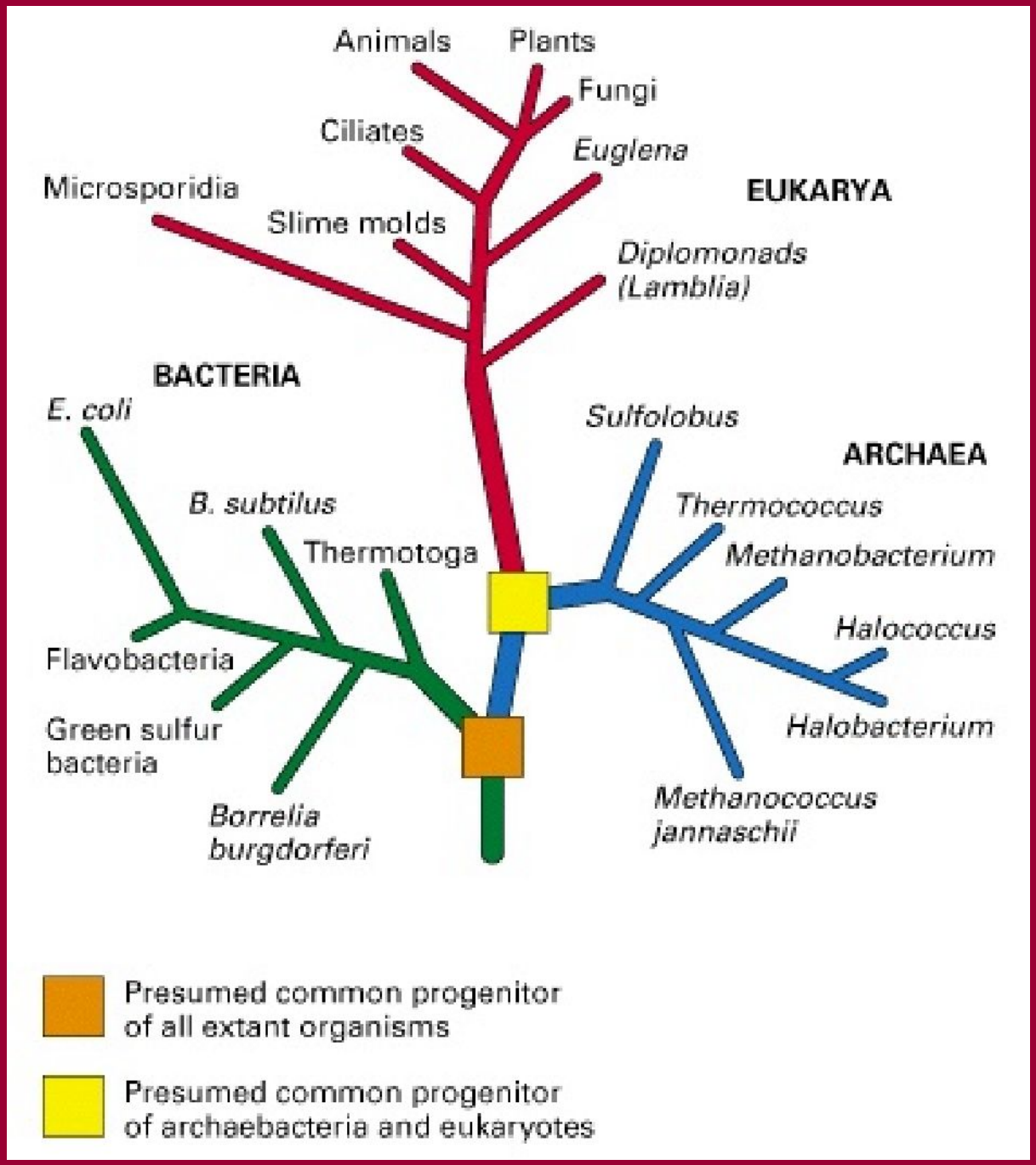

Animals
Plants
Fungi
Ciliates
Euglena
Microsporidia
EUKARYA
Slime molds
Diplomonads
(Lamblia)
BACTERIA
E. coli
Sulfolobus
ARCHAEA
B. subtilus
Thermococcus
Thermotoga
Methanobacterium
Halococcus
Flavobacteria
Halobacterium
Green sulfur
bacteria
Borrelia
burgdorferi
Methanococcus
jannaschii
Presumed common progenitor
of all extant organisms
Presumed common progenitor
of archaebacteria and eukaryotes

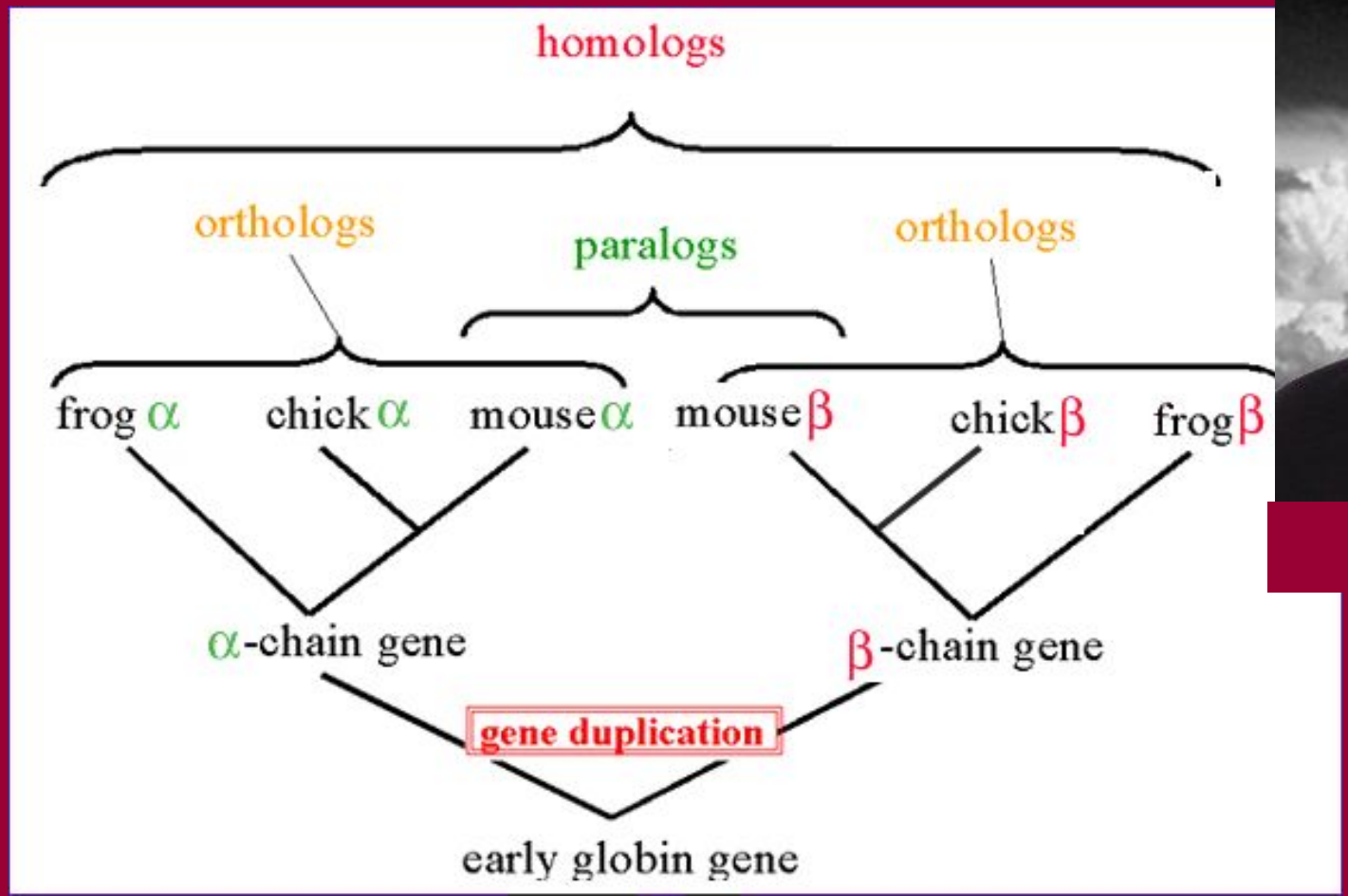
Гомологичные последовательности – последовательности, имеющие общее происхождение (общего предка).

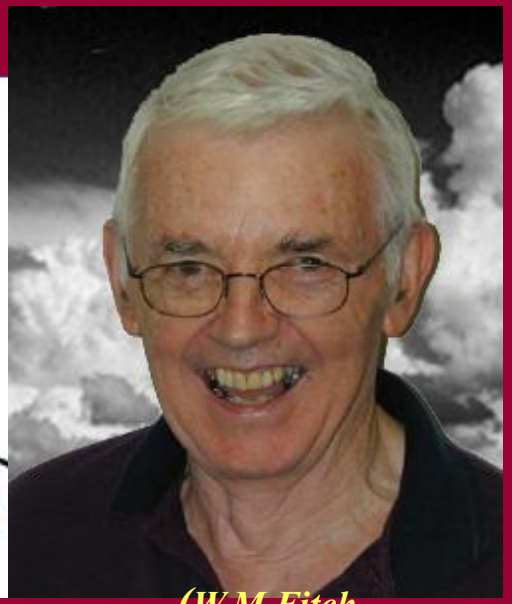
Признаки гомологичности белков

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

Гомологи

Ортологи **Паралоги** Ксенологи ?

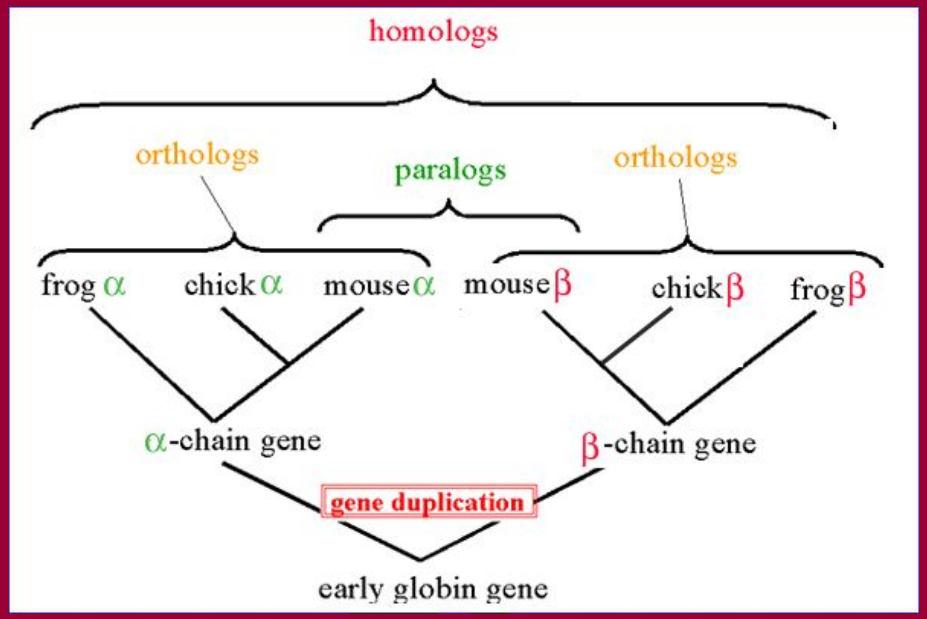




(W.M.Fitch, Syst.Zool.19,99(1970)

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 ≠ гомология

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

Дом

Скупидом

Водомерка ?

Гэп – пропуск в последовательности

Домов---ой

Домо-строй

?

Домов-----ой

Домо-строй

Как получить «оптимальное» выравнивание?

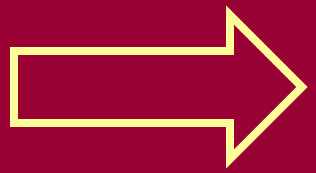
Один из способов:

```
>EC_Tr  :  MQNRLTI KDI ARLSGVGKSTVSRVLNNEYR
>EC_Fr  :  MKLDEI ARLAGVSRTTASYVI NGKAKQYR
```

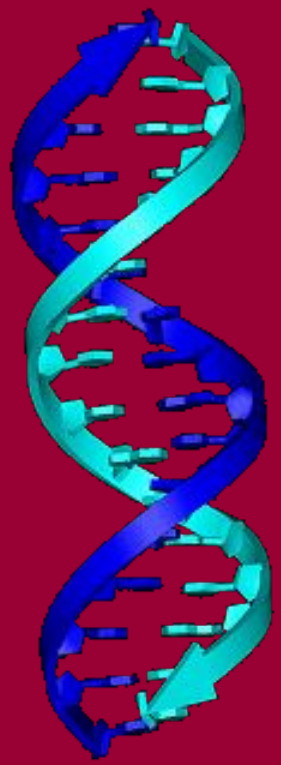
Гэп – пропуск в последовательности

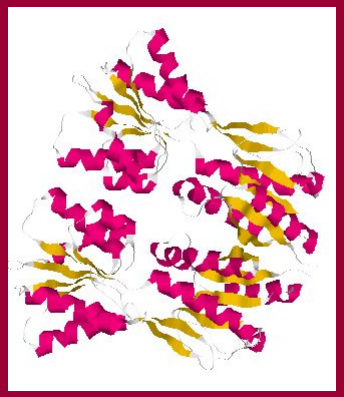
```
>EC_Tr  :  MQNRLTI KDI ARLSGVGKSTVSRVLNNE---YR  :  30
>EC_Fr  :  ----MKLDEI ARLAGVSRTTASYVI NGKAKQYR  :  29
                      I ARL  GV    T S V N     YR
```

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						

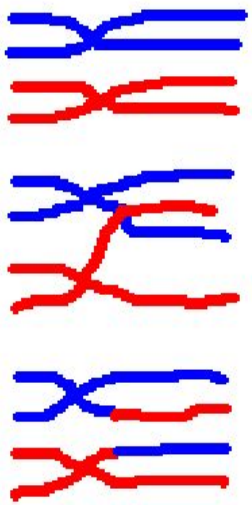
Termination

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

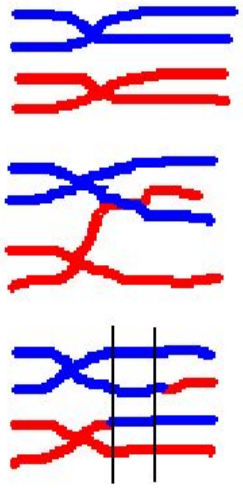
Frame shift

Recombination

Crossing over:
reciprocal recombination



Gene conversion:
Non-reciprocal recombination

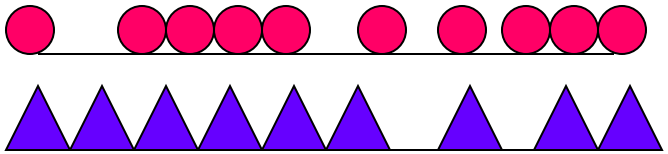


Domain insertion: “Mosaic proteins”

Structural modules:

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

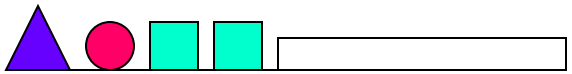
Domain origins:



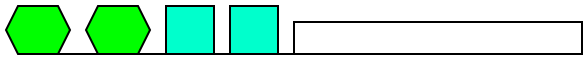
epidermal growth factor (EGF)

fibronectin

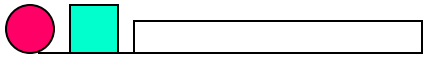
Mosaic proteins



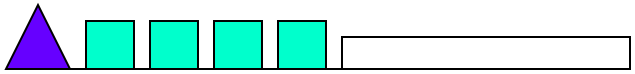
tissue plasminogen activator



prothrombin



urokinase



plasminogen

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
mdskrtrqtYTRYQTLELEKEFhfn...rYITRRRRIDIAN
.mrrkrrvlFSQAQVYELERRFkqq...kYLSAPEREHLAS
.tkpyrghrFTKENVRILESWFaknienpYLDTKGLENLMK
.rkrkrrttISIAAKDALERHFgeh...sKPSSQEIMRMAE
.kspkgkssISPQARAFLEEVFrrk...qSLNSKEKEEVAK
```

AAAAAAAAAA ⇨ AACDEAAAA

Что лучше?

```
AAAAAAAAAA        AA---AAAAAAAAA
AACDE-AAAA        AACDE----AAAA
```