

ВВЕДЕНИЕ В ИНФОРМАЦИОННУЮ БИОЛОГИЮ

АЛЬТЕРНАТИВНЫЙ КУРС

Н. А. Колчанов

Факультет Естественных Наук НГУ

kol@bionet.nsc.ru

РЕВОЛЮЦИЯ В СОВРЕМЕННОЙ БИОЛОГИИ

Расшифровка структуры двойной спирали ДНК

- Рентгено-структурные методы, расшифровка пространственной структуры белков
Методы расшифровки аминокислотных последовательностей
- Методы расшифровки нуклеотидных последовательностей
- Генетическая инженерия
- Генодиагностика
- Трансгенез
- Клонирование
- Молекулярные биотехнологии
- ДНК-микрочипы
- Массовое секвенирование геномов

1953

2001

Расшифровка генома человека

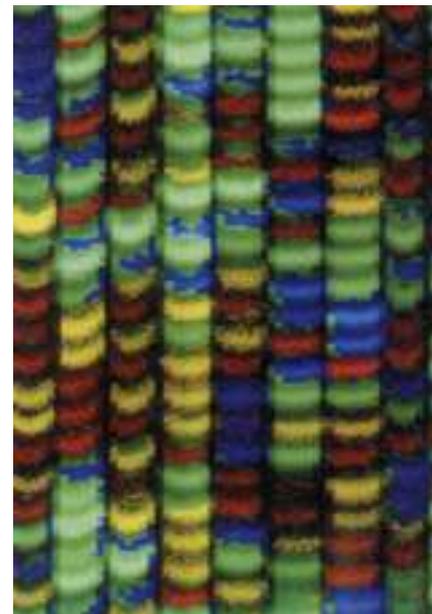
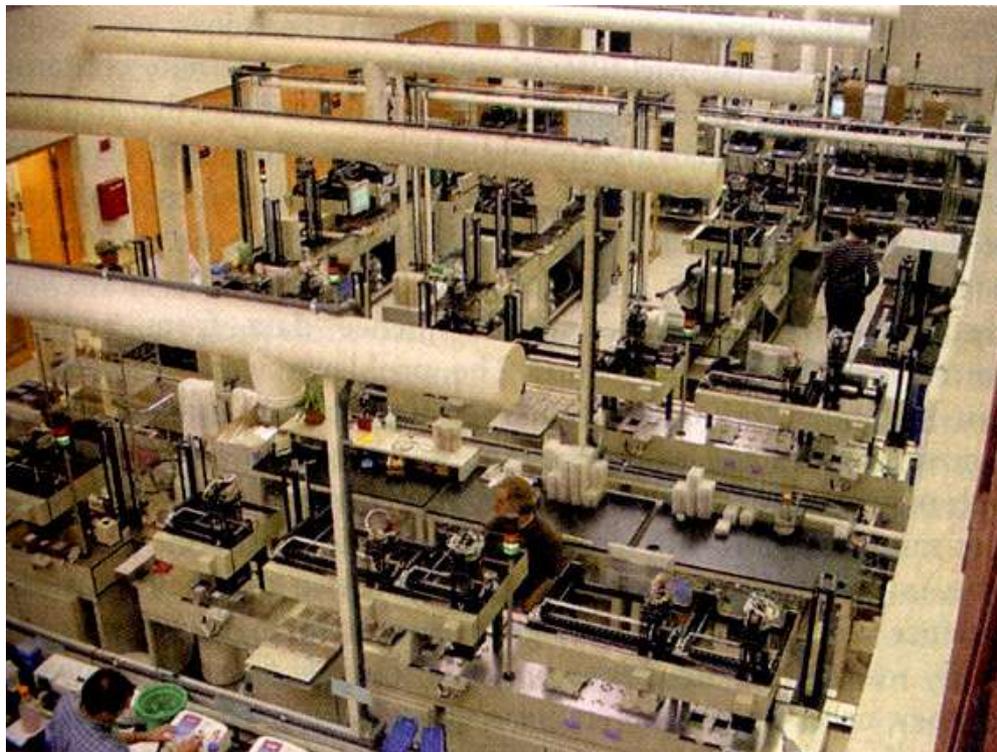
Протеомика

Транскриптомика

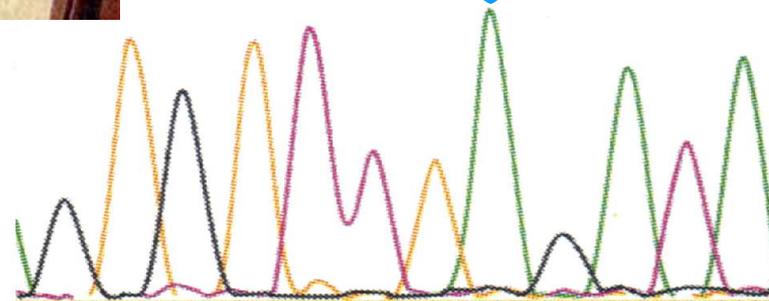
Молекулярная медицина, генотерапия

Конструирование молекулярно-генетических систем с заданными свойствами

ИНФОРМАЦИОННЫЙ ВЗРЫВ В МОЛЕКУЛЯРНОЙ БИОЛОГИИ И ГЕНЕТИКЕ: автоматическая расшифровка нуклеотидных последовательностей



Фабрика автоматического
секвенирования (расшифровки) ДНК



... . **GGTTCATAAATTTGGATCGTACTTATGGTTAATGC...**

ИНФОРМАЦИОННЫЙ ВЗРЫВ В МОЛЕКУЛЯРНОЙ БИОЛОГИИ И ГЕНЕТИКЕ

**В РЕЗУЛЬТАТЕ ОСУЩЕСТВЛЕНИЯ
КРУПНОМАСШТАБНЫХ ГЕНОМНЫХ ПРОЕКТОВ
РАСШИФРОВАНЫ СТРУКТУРЫ ГЕНОМНОЙ ДНК
ТЫСЯЧ ВИРУСОВ, СОТЕН БАКТЕРИЙ, ГЕНОМЫ
ДРОЖЖЕЙ, S.ELEGANS, ДРОЗОФИЛЫ, ГЕНОМЫ
РЯДА РАСТЕНИЙ, ЗАКОНЧЕНА ЧЕРНОВАЯ
РАСШИФРОВКА ГЕНОМА ЧЕЛОВЕКА**

ИНФОРМАЦИОННЫЙ ВЗРЫВ В МОЛЕКУЛЯРНОЙ БИОЛОГИИ И ГЕНЕТИКЕ: РАЗМЕРЫ ГЕНОМОВ

5 ТОМОВ



E. Coli
 5×10^6 п.о.

E. Coli

10 ТОМОВ



Дрожжи
 $\sim 10^7$ п.о.

100 ТОМОВ



Дрозофила
 $\sim 10^8$ п.о.

3000 ТОМОВ



Человек
 3.3×10^9 п.о.

ИНФОРМАЦИОННЫЙ ВЗРЫВ В МОЛЕКУЛЯРНОЙ БИОЛОГИИ И ГЕНЕТИКЕ

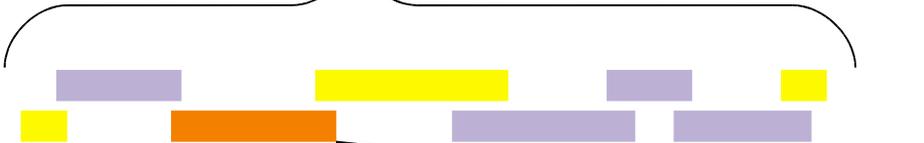
Характеристики генома человека:

размер - 3.3 миллиарда нуклеотидов **ATGCCCGGGTTTAATGCGTCAGTGAСТGCACA...**

количество генов - 35000:
цитогенетическая карта:

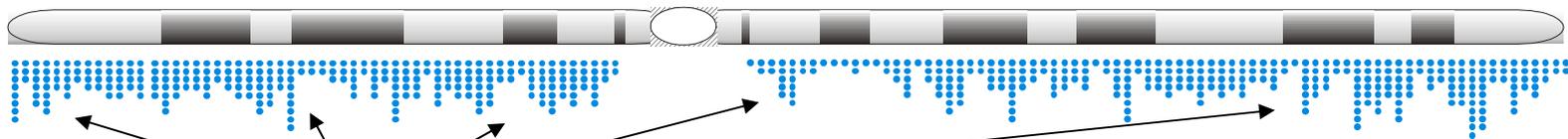


физическая карта:



Локализация генов

— FABP2B
— CNKL, ARSA, ACR
— SC02, EOGF1, OPT1B
— BVL, MARK12, MARK11
— PKDREJ, B99, CELSR1
— FBLN1, E46L, PRARA
— ANKBAR, NUP50, UPR3
— SULT3, CG151, CG156
— PACSIN2, BIK, BZFR
— **DIA1**, A14GALT, ANFGAR1
— PRO2455, CYP2D6, CG196
— SREBF2, NGA, NDUFA6
— PML1, G22P1, NHRP2L1
— TEF70B2, ACO2
— RBX1, EP300, RANBP1
— GRP24, RMR24, ST13
— CASNA11, GVAR2, AOSL
— TAB1, MGAT3, ATF4
— **ROGFB**, RPL3, SYNGR1
— GTRP1, HSP018, DNAL4
— P72, DMC1, C22ORF2
— CSNK1E, KCNJ4, KDELR3
— PLAZG6, MAFF, C22ORF5
— **SOX10**, PRKCA, PRKCB, SLC16A8
— GCAT, GALR3, HSP025
— OGAT1, GALS1, HIF0
— MFG, M5E5L, GALS2
— HSP022, RAC2, PSCD4
— TSTL2RB, SSTR3
— RAU1, NCF4, **OSF2B**
— TXN2, EIF3S7, CACNG2
— RBM9, CG12-1, APOE
— NMOX1, MCM5, MB
— **TMR3L**, LARGE, TOM1
— RPR13, FBX07, SYN3
— YWHAH, SLC5A1, RFR12
— ZNF278, DRG1, AET
— **SMTN**, AGT12, LIMK2
— CS1, PEST1, **TGN2**
— SF3A1, SEC14L2, HSP0242
— HSP051, MIMR3, LIF
— PK13, **NP2**
— AP1B1, NEFN
— GARP2, RRP22
— PTPN8, RADD3
— STYB4, MNI
— PPS12, SNYB1
— **SNYB2**, ADPAK2



Локализация мутаций: 1.5 миллиона полиморфизмов

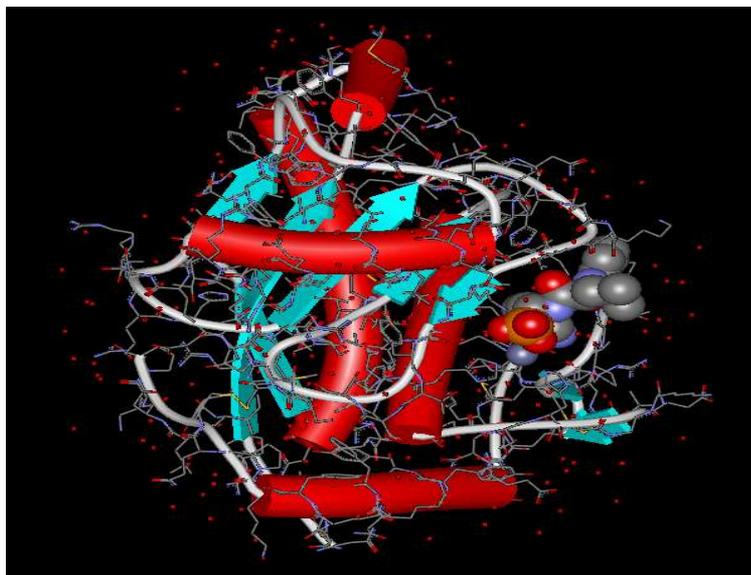
ИНФОРМАЦИОННЫЙ ВЗРЫВ В МОЛЕКУЛЯРНОЙ БИОЛОГИИ И ГЕНЕТИКЕ

Расшифрованы
аминокислотные
последовательности
сотен тысяч белков

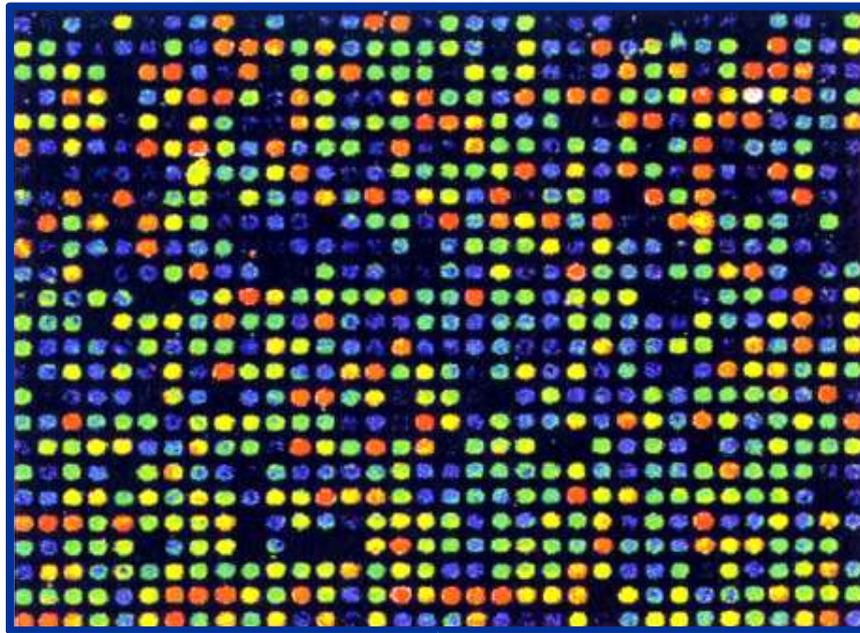
b

	330	340	350	360	370	380					
1ATI_A	◆ LGTMCDP	◆ kLSIGIVQDHS	◆ PINLLMGV	◆ TMAHEIGHNIGME	◆ HDg~~~k	◆ DCLRg~~~~a	◆ SLCIMRPGL				
query	◆ VEGMCRA	◆ eSSGGVST	◆ DHS	◆ ELPIGAAATMAHEIGH	◆ SLGISHD	◆ p~~~d	◆ GCCV	◆ eaaaaes	◆ GGCV	◆ MAAAT	
462302	◆ MGGMCNA	◆ kNSVGIVK	◆ DSSNVFM	◆ VAVTMTHEIGHN	◆ LGMEHD	◆ d~kd	◆ KCKC	◆ ~~~~~	◆ EAC	◆ IMSAVI	
462095	◆ VGGMCQL	◆ kHSTGVIQ	◆ DHSAINLL	◆ VALTMAHEIGHN	◆ LGMNH	◆ Dg~~~n	◆ QCHC	◆ g~~~~	◆ aNSCV	◆ MAAML	
123525	◆ TGGMCYP	◆ kNSVGIVQ	◆ DHPKTL	◆ LIAVTMAHEIGHN	◆ LGMNH	◆ Dg~~~n	◆ HCHC	◆ s~~~~	◆ aSFC	◆ IMPPSI	
2231613	◆ MASMCDP	◆ kRSVGIIQ	◆ DSTINLM	◆ MAVTMAHEM	◆ GHNIGMD	◆ HDe~~~k	◆ YCTC	◆ g~~~~	◆ aKSCV	◆ MAKAL	
1916617	◆ LEGMCTA	◆ eNSGGVSM	◆ DHS	◆ ENAIGAAATMAHEIGH	◆ NFGMSH	◆ Dd~~~	◆ GCCV	◆ eatpeq	◆ GGC	◆ IMAAAT	
7993727	◆ IMSMCTA	◆ dQSGGIVM	◆ DSDNPL	◆ GAAVTLAHEL	◆ GHNFGMNH	◆ Dtdldr	◆ GCSC	◆ qmavek	◆ GGC	◆ IMNAST	
1709103	◆ VSALCSR	◆ ~HSGAVN	◆ QDHSKNS	◆ IGVASTMAHEL	◆ GHNIGMSH	◆ Dd	◆ edip	◆ GCYC	◆ pepreg	◆ GGC	◆ IMTESI
2499914	◆ VSAMCSH	◆ ~SSGAVN	◆ QDHSKNP	◆ VGVACTMAHEM	◆ GHNIGMD	◆ HDe	◆ envq	◆ GCRC	◆ qerfea	◆ GRC	◆ IMAGSI

Расшифрованы
пространственные
структуры десятков
тысяч белков



ТЕХНОЛОГИЯ ДНК-ЧИПОВ ОТКРЫВАЕТ ВОЗМОЖНОСТЬ ПЕРЕХОДА НА КАЧЕСТВЕННО НОВЫЙ УРОВЕНЬ В ГЕНОДИАГНОСТИКЕ, ИЗУЧЕНИИ ЭКСПРЕССИИ ГЕНОВ И ДРУГИХ ЗАДАЧАХ МОЛЕКУЛЯРНОЙ БИОЛОГИИ И ГЕНЕТИКИ



ДНК-чип

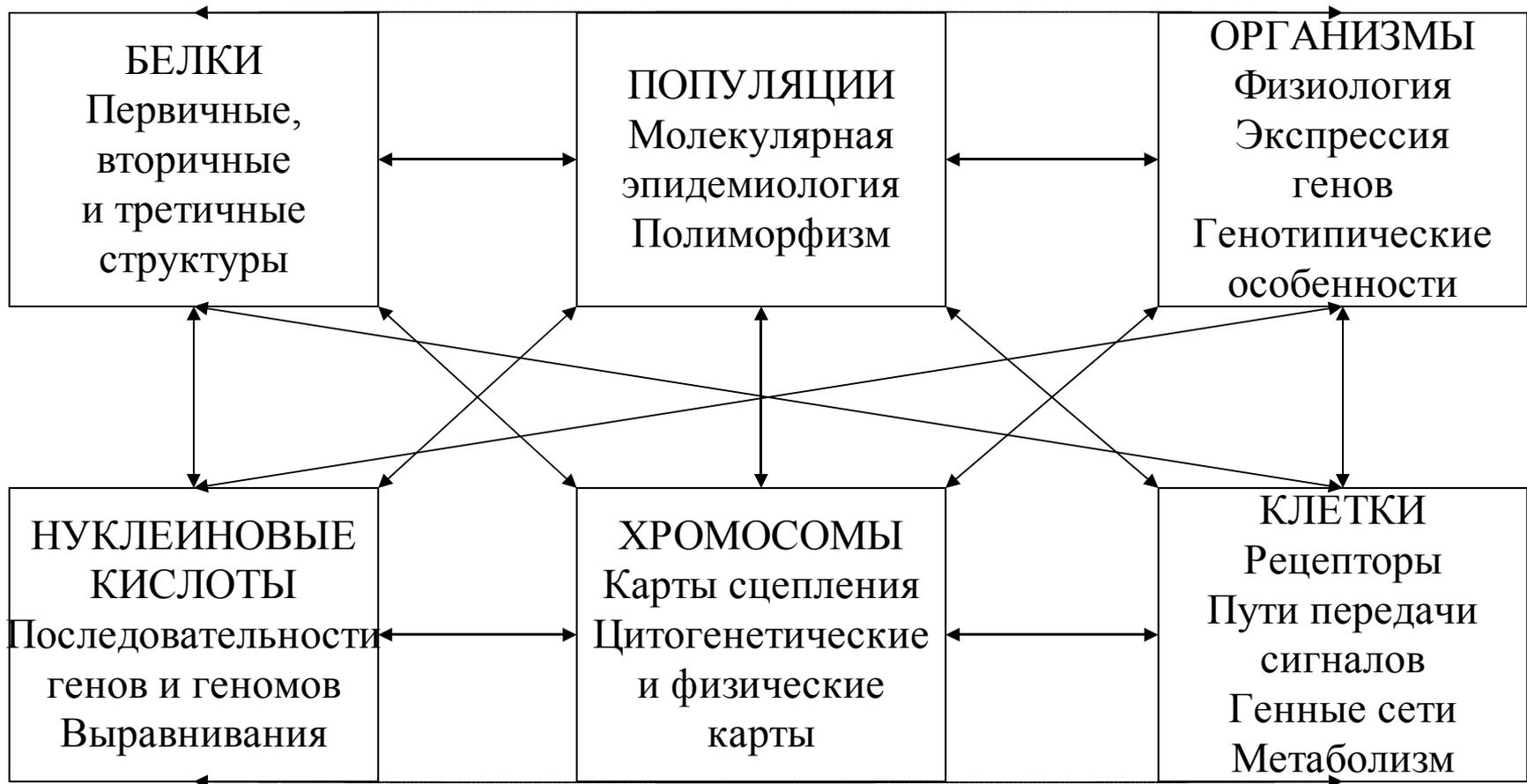


Современная биология стала производителем беспрецедентно огромных объемов экспериментальных данных. Их осмысление невозможно без привлечения современных информационных технологий, эффективных методов анализа данных и моделирования биологических систем и процессов на различных уровнях организации живой материи: от молекулярно-генетического, включая организменный и заканчивая экосистемным и биосферным.

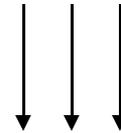
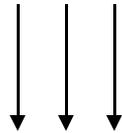
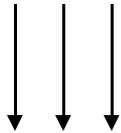
В ответ на этот вызов возникает *НОВАЯ наука, называемая *информационной биологией.**

ЧЕЛОВЕК: ИНФОРМАЦИОННЫЕ РЕСУРСЫ СОВРЕМЕННОЙ БИОЛОГИИ

для решения проблем здоровья



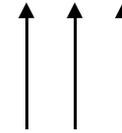
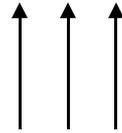
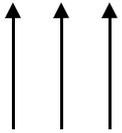
Экспериментальные технологии



*Изучение механизмов
фенотипического
проявления
генетически
обусловленных
патологий*

*Поиск и создание
фармакологических
препаратов для
генотип-
специфической
коррекции патологий*

*Поиск генотип-
специфических
стратегий коррекции
патологий
с использованием
созданных
препаратов*



Информационно-компьютерные технологии

**ДВЕ БАЗОВЫХ КОМПОНЕНТЫ СОВРЕМЕННОЙ
БИОЛОГИИ**

ИНФОРМАЦИОННАЯ БИОЛОГИЯ

О значимости и масштабах исследований в области информационной биологии свидетельствует тот факт, что только в ближайшие 4 года в индустриально развитых странах мира (США, АНГЛИЯ, ЯПОНИЯ, ГЕРМАНИЯ, ФРАНЦИЯ) в информационные технологии, ориентированные на молекулярную биологию, генетику, биотехнологию, медицину, агробиологию, будет инвестировано более 30 миллиардов долларов.

БАЗЫ ДАННЫХ ПО БИОЛОГИЧЕСКОМУ РАЗНООБРАЗИЮ ЖИВОТНОГО И РАСТИТЕЛЬНОГО МИРА

Зеленая книга Сибири - Microsoft Internet Explorer

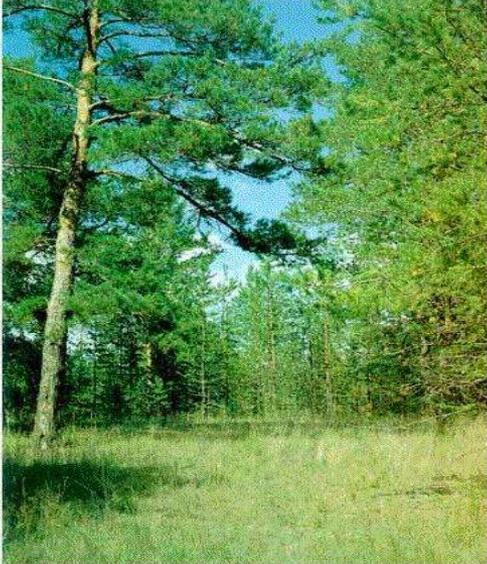
File Edit View Go Favorites Help

Back Forward Stop Refresh Home Search Favorites History Channels Fullscreen Mail Print

Address <http://www-sbras.nsc.ru/win/elbib/bio/green/62.html>

Редкие и нуждающиеся в охране растительные сообщества Сибири

Сосновый осоково-злаковый ксерофитно-псаммофитный лес (*Pinus sylvestris* - *Koeleria glauca* + *Carex supina*)



Место конкретного описания. Алтайский край, Волчихинский район, 17км восточнее с. Усть-Волчиха, боровая лента.

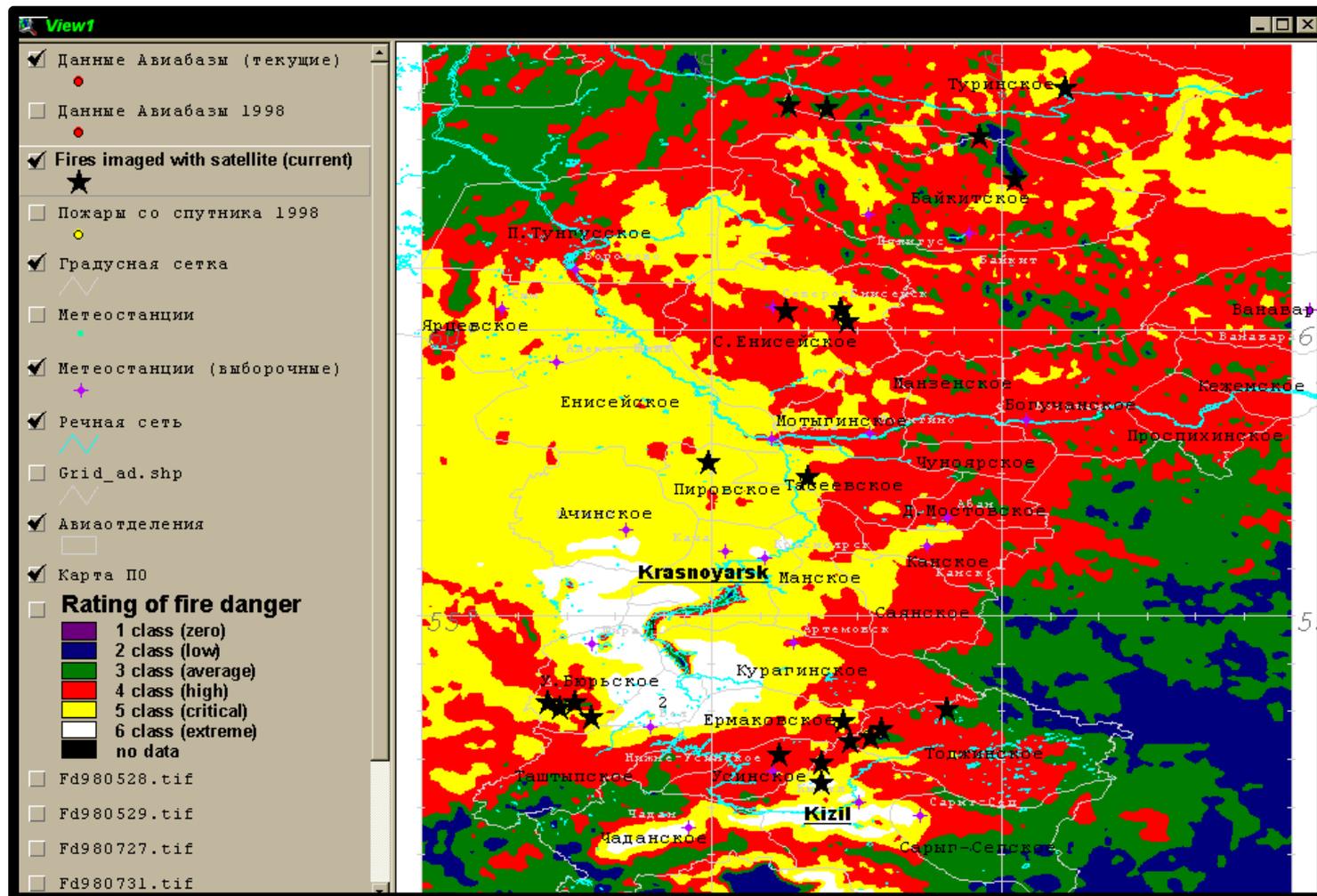
Географический ареал. Сообщества имеют дизъюнктивный ареал, в пределах зональных полос настоящих и сухих степей северной Евразии. В Сибири и Северном Казахстане сообщества приурочены к южным частям ленточных боров.

Экологический ареал. Сообщества представляют крайне южные форпосты равнинной лесной растительности в семиаридных зональных климатических условиях. Они целиком связаны с сухими песчаными местообитаниями, занимают выпуклые части дюнных всхолмлений широких долин древнего стока, непосредственно контактируют с участками псаммофитных степей. Почвы песчаные боровые олиготрофные.

Основные дестабилизирующие факторы. Сообщества формируются в составе ленточных боров, входящих в I группу лесов, подлежащих охране. Основной

Internet zone

ОПИСАНИЕ И МОДЕЛИРОВАНИЕ ДИНАМИКИ ЭКОСИСТЕМ: ГЕОГРАФИЧЕСКИЕ ИНФОРМАЦИОННЫЕ СИСТЕМЫ

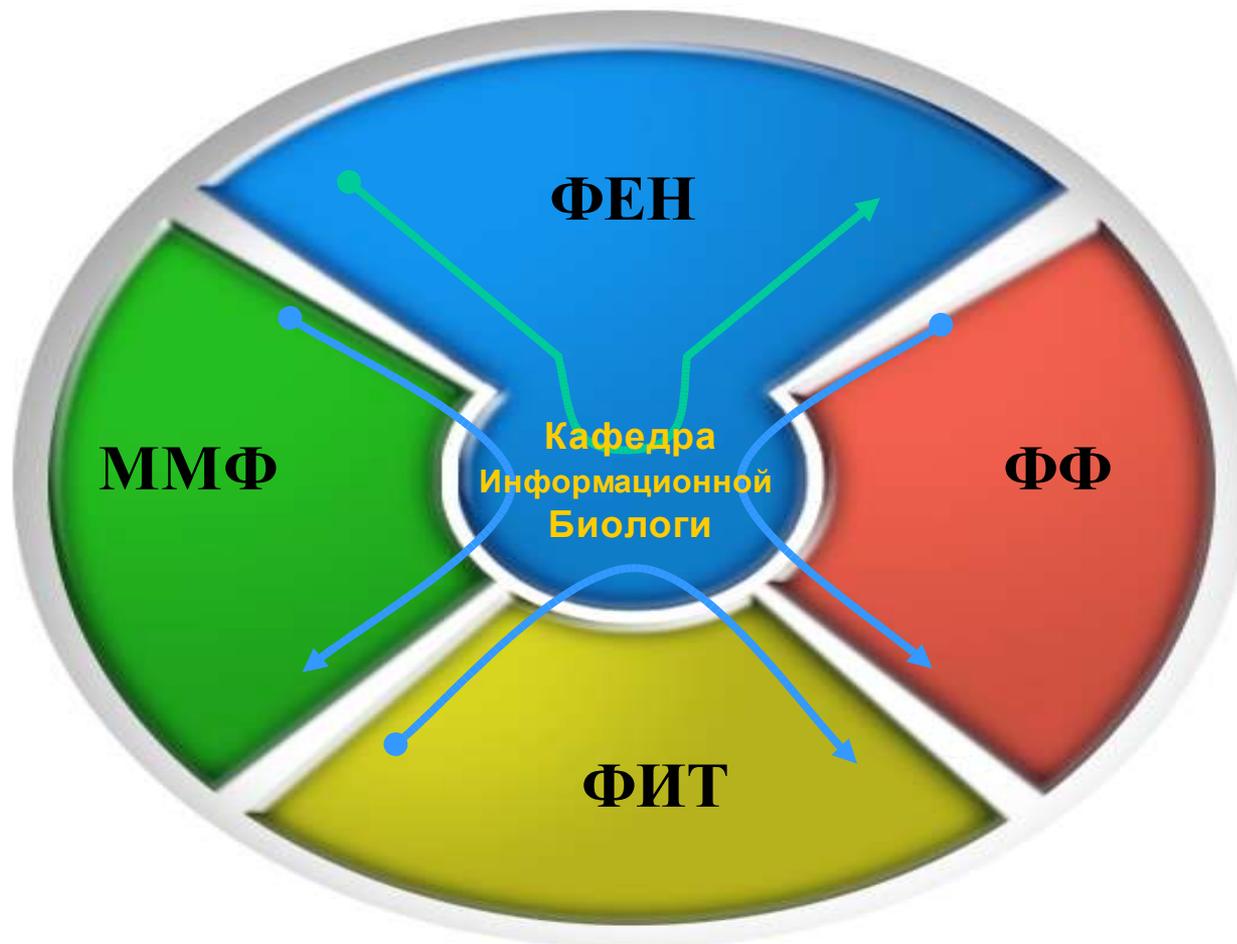


КАФЕДРА ИНФОРМАЦИОННОЙ БИОЛОГИИ

Задача - подготовка специалистов, владеющих методами биоинформатики, теоретического и компьютерного анализа и моделирования, необходимыми для решения широкого круга фундаментальных и прикладных проблем молекулярной биологии, молекулярной генетики, клеточной биологии, физиологии, биофизики, биомедицины, фармакологии, общей биологии, экологии, возникающих на стыке с математикой, информатикой и физикой.



КАФЕДРА ИНФОРМАЦИОННОЙ БИОЛОГИИ: ВЗАИМОДЕЙСТВИЯ С ФАКУЛЬТЕТАМИ



**ИЦИГ СО РАН – базовый
институт кафедры
информационной биологии.
Лаборатория теоретической
генетики ИЦИГ СО РАН –
базовая лаборатория
кафедры информационной
биологии**

КУРСЫ КАФЕДРЫ ИНФОРМАЦИОННОЙ БИОЛОГИИ ДЛЯ СТУДЕНТОВ 4-5 ГОДА ОБУЧЕНИЯ

Введение в информационную биологию - 45 час.

Цикл «Теоретическая биология»

Теория предбиологической эволюции - 10 час.

Математическая теория молекулярной эволюции – 20 час.

Теория популяций – 16 час.

Цикл «Основы компьютерной геномики»

Комбинаторика символьных последовательностей – 16 час.

Теория кодирования информации – 10 час.

Компьютерные методы анализа генетических текстов – 20 час.

Большой практикум по компьютерной геномике - 40 час.

Цикл «Физика биополимеров»

Физическая теория биополимеров – 20 час.

Конформационный анализ и молекулярная динамика – 15 час.

Компьютерные методы предсказания структуры биополимеров – 10 час.

Большой практикум по структуре биополимеров – 25 час.

КУРСЫ КАФЕДРЫ ИНФОРМАЦИОННОЙ БИОЛОГИИ ДЛЯ СТУДЕНТОВ 4-5 ГОДА ОБУЧЕНИЯ

Цикл "Управление в биологических системах"

Механизмы биологической регуляции - 45 час.

*Управление в молекулярно-генетических системах (генных сетях)
(20 час.)*

Управление в физиологических системах - 15 час.

Управление в популяциях и экосистемах - 15 час.

Математическая теория управления - 20 час.

Цикл "Моделирование биологических систем"

*Моделирование и анализ молекулярно-генетических систем
(курс лекций и практикум) – 36 час.*

*Моделирование и анализ экологических систем
(курс лекций и практикум) – 36 час*

**Синергетика - теория самоорганизации неравновесных систем –
20 час.**

КУРСЫ КАФЕДРЫ ИНФОРМАЦИОННОЙ БИОЛОГИИ ДЛЯ СТУДЕНТОВ 4-5 ГОДА ОБУЧЕНИЯ

Цикл «Информационные ресурсы в биологии»

Информационные ресурсы по молекулярной биологии и генетике (курс лекций и большой Интернет-практикум) – 36 час.

Информационные ресурсы по экологии и биоразнообразию (курс лекций и большой Интернет-практикум) – 36 час.

Географические информационные системы (курс лекций и практикум) – 36 час.

Электронные библиотеки в биологии (курс лекций и практикум) – 36 час.

КУРСЫ КАФЕДРЫ ИНФОРМАЦИОННОЙ БИОЛОГИИ ДЛЯ СТУДЕНТОВ 4-5 ГОДА ОБУЧЕНИЯ

Цикл "Методы обнаружения закономерностей"

Методы продукции знаний и поиска закономерностей – 10 час.

Методы анализа данных и знаний – 10 час.

Методы классификации – 20 час.

Семинар и практикум (освоение стандартных и специализированных пакетов программ для решения задач анализа в различных областях биологии) – 15 час.

Цикл «Спецразделы математики»

Теория графов –16 час. (Ф)

Методы оптимизации (решения экстремальных задач) - 20 час.

Методы идентификации систем (решения обратных задач) –20 час.

Дифференциальные уравнения - 30 час. (Ф)

Дискретный анализ и комбинаторика 18 час. (Ф)

Цикл «Основы информатики»

Информационные технологии и языки программирования – 40 час.

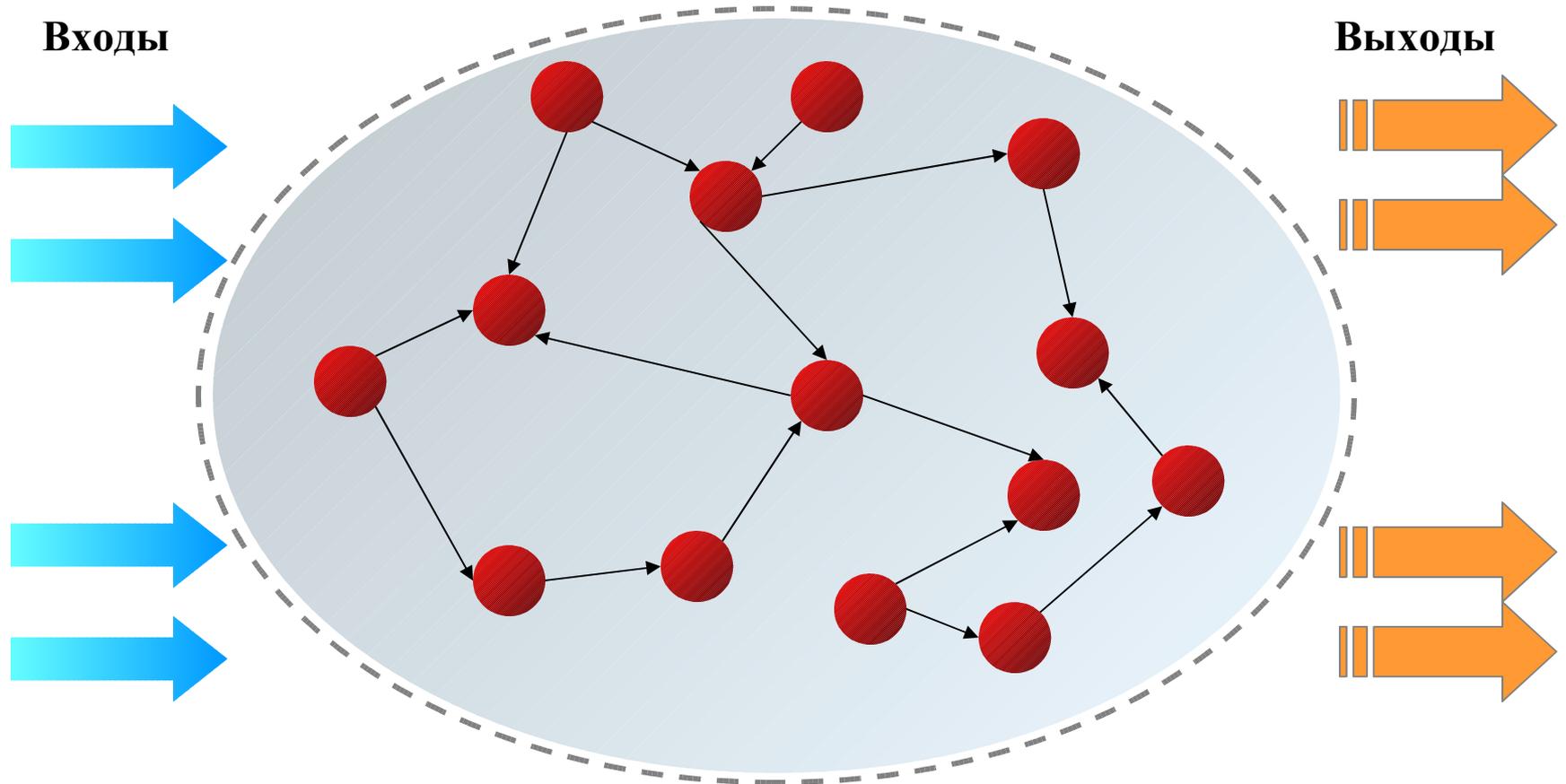
Большой вычислительный практикум (методы вычислений, методы статистического анализа) - 30 час.

Лекция 1. Часть 2

**Биологические
самовоспроизводящиеся
системы: основные понятия
и определения**

Системы бывают двух типов:

- (а) замкнутые, когда отсутствует взаимодействие между ними и окружающей средой и
- (б) открытые, когда через определенные элементы (входы и выходы) осуществляется обмен потоками вещества, энергии и информации с окружающей средой.





САМОВОСПРОИЗВОДЯЩИЕСЯ СИСТЕМЫ



Самовоспроизводящейся называется любая система, способная к воспроизведению своих копий на основе содержащейся в ней наследственной информации.

Самовоспроизводящиеся системы являются открытыми, и воспроизведение подобных себе копий является одним из результатов их взаимодействия со средой.

Известно два типа самовоспроизводящихся систем: (а) биологические и (б) социальные.

Характерная особенность биологических СВ систем - запись подавляющей части наследственной информации с использованием двух типов кодирующих биополимеров: ДНК или/и РНК



Примеры реальных биологических самовоспроизводящихся систем (СВ систем)

- ВИРУС ИЛИ ФАГ - ПРОСТЕЙШАЯ СВ СИСТЕМА, ВОСПРОИЗВОДЯЩАЯ СВОИ КОПИИ НА ОСНОВЕ ИНФОРМАЦИИ, ЗАПИСАННОЙ В ИХ ГЕНОМАХ, НО С ИСПОЛЬЗОВАНИЕМ МОЛЕКУЛЯРНО- ГЕНЕТИЧЕСКОГО АППАРАТА КЛЕТКИ-ХОЗЯИНА
- ОДНОКЛЕТОЧНЫЙ ПРО- ИЛИ ЭУКАРИОТИЧЕСКИЙ ОРГАНИЗМ - СВ СИСТЕМА МИНИМАЛЬНОГО УРОВНЯ СЛОЖНОСТИ, СПОСОБНАЯ К ПОЛНОСТЬЮ АВТОНОМНОМУ ВОСПРОИЗВЕДЕНИЮ СВОИХ КОПИЙ
- ПАРА РАЗНОПОЛЫХ ОСОБЕЙ - МИНИМАЛЬНАЯ ЕДИНИЦА САМОВОСПРОИЗВЕДЕНИЯ ДЛЯ ЖИВОТНЫХ И РАСТЕНИЙ С ПОЛОВЫМ ПРОЦЕССОМ РАЗМНОЖЕНИЯ
- ПОПУЛЯЦИЯ - СВ СИСТЕМА, СПОСОБНАЯ К УСТОЙЧИВОМУ (ПО ОТНОШЕНИЮ К ИЗМЕНЕНИЯМ ОКРУЖАЮЩЕЙ СРЕДЫ) ВОСПРОИЗВЕДЕНИЮ
- БИОГЕОЦЕНОЗ - СОВОКУПНОСТЬ ПОПУЛЯЦИЙ, СУЩЕСТВУЮЩИХ НА ОПРЕДЕЛЕННОЙ ТЕРРИТОРИИ, ПРИНАДЛЕЖАЩИХ РАЗЛИЧНЫМ БИОЛОГИЧЕСКИМ ВИДАМ И ОБЪЕДИНЕННЫХ ИЕРАРХИЕЙ ТРОФИЧЕСКИХ СВЯЗЕЙ
- БИОСФЕРА ЗЕМЛИ - ГЛОБАЛЬНАЯ СВ СИСТЕМА ИЕРАРХИЧЕСКИ НАИБОЛЕЕ ВЫСОКОГО УРОВНЯ



Принципиальная схема самовоспроизводящей системы

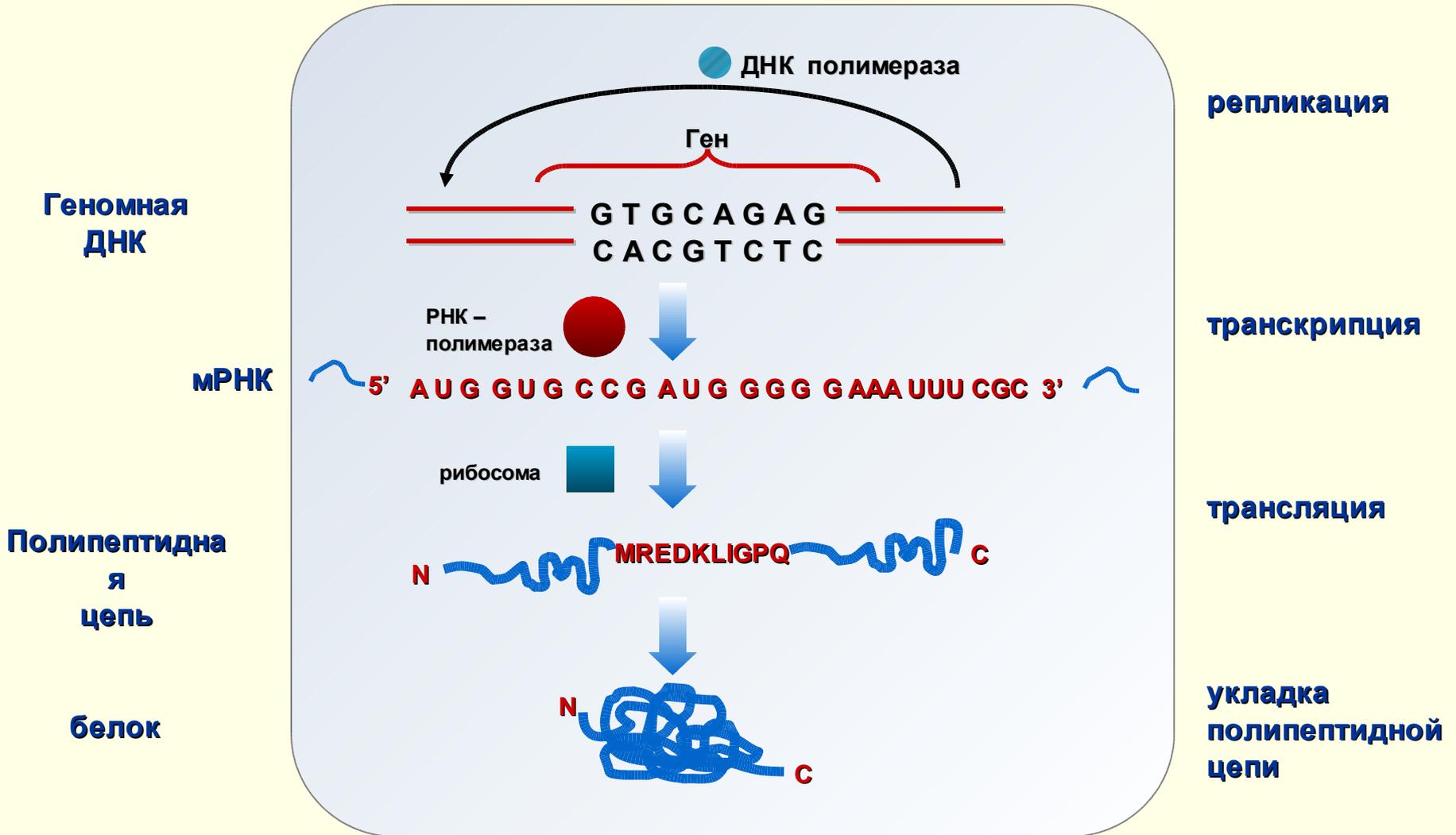


СВ система получает два типа возмущающих воздействий со стороны окружающей среды:

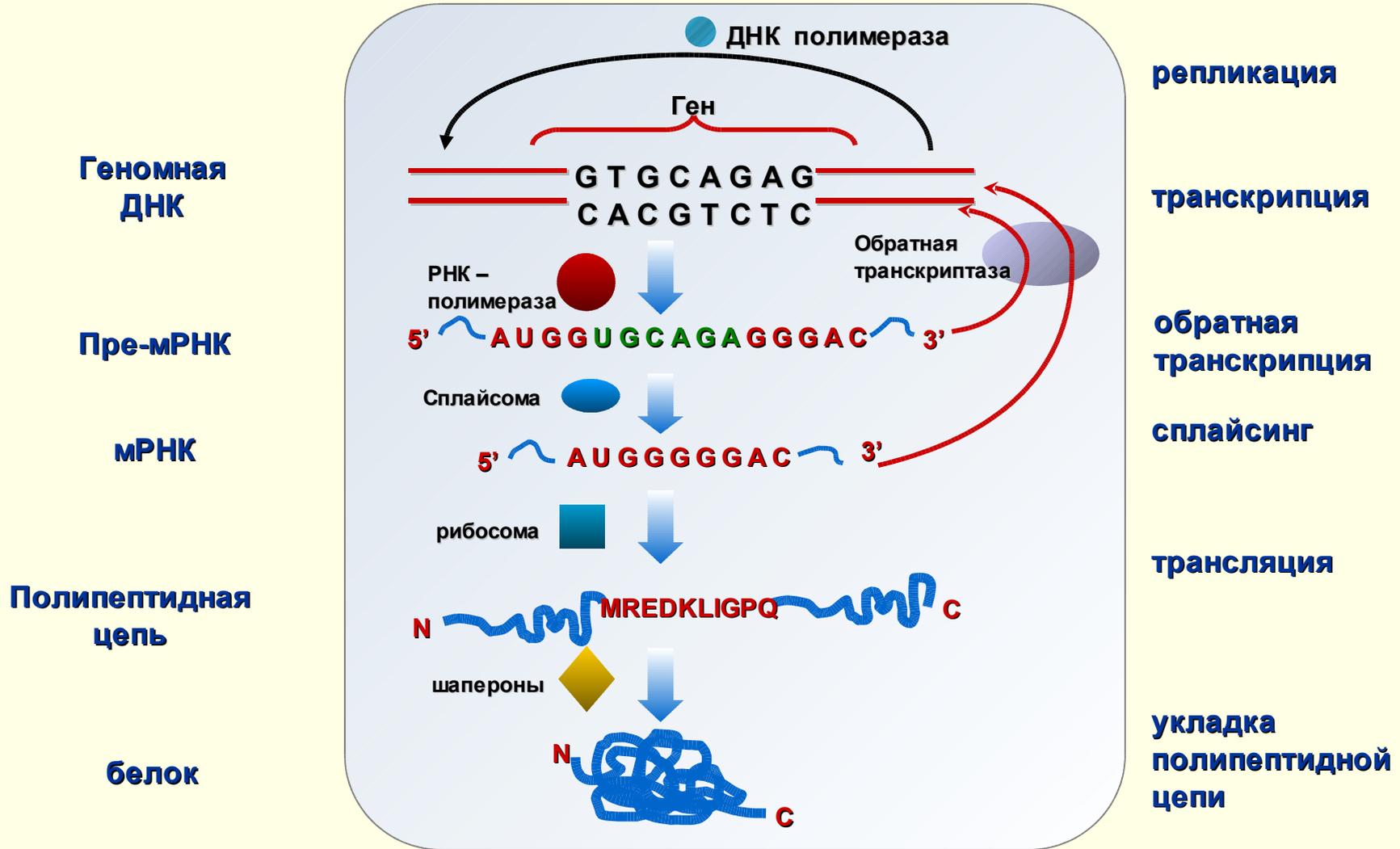
- (1) возмущения через входные каналы (качественные и количественные изменения потоков вещества, энергии и информации);
- (2) мутации наследственной памяти системы.



Центральная догма молекулярной биологии: поток информации генетической информации (Френсис Крик, 1963)



Генетические макромолекулы и фундаментальные генетические процессы (современный взгляд на проблему)





ПРИНЦИПИАЛЬНЫЕ КОМПОНЕНТЫ САМОВОСПРОИЗВОДЯЩИХСЯ СИСТЕМ



Наследственная память содержит информацию о механизмах функционирования и воспроизведения СВ системы



ГЕНЕТИЧЕСКАЯ ПАМЯТЬ

многоклеточной СВ системы

```
graph TD; A[ГЕНЕТИЧЕСКАЯ ПАМЯТЬ многоклеточной СВ системы] --> B[память генеративных элементов]; A --> C[память соматических элементов];
```

память генеративных элементов

(хранение и реализация наследственной информации, необходимой для воспроизведения копии СВ системы)

память соматических элементов

(функционирование и поддержание структурной организации СВ систем)



Химическая структура двойной спирали ДНК – материальной основы наследственной памяти самовоспроизводящихся систем

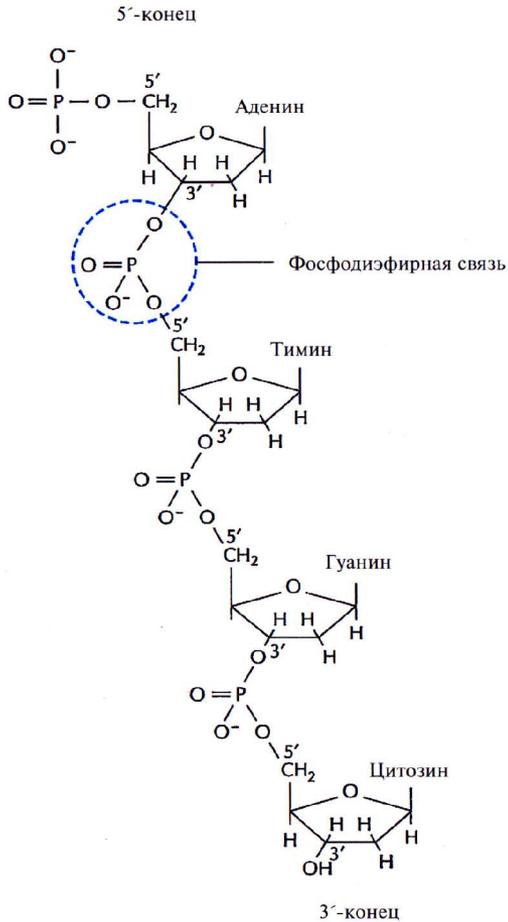


Рис. 3.2. Одна из цепей молекулы ДНК.

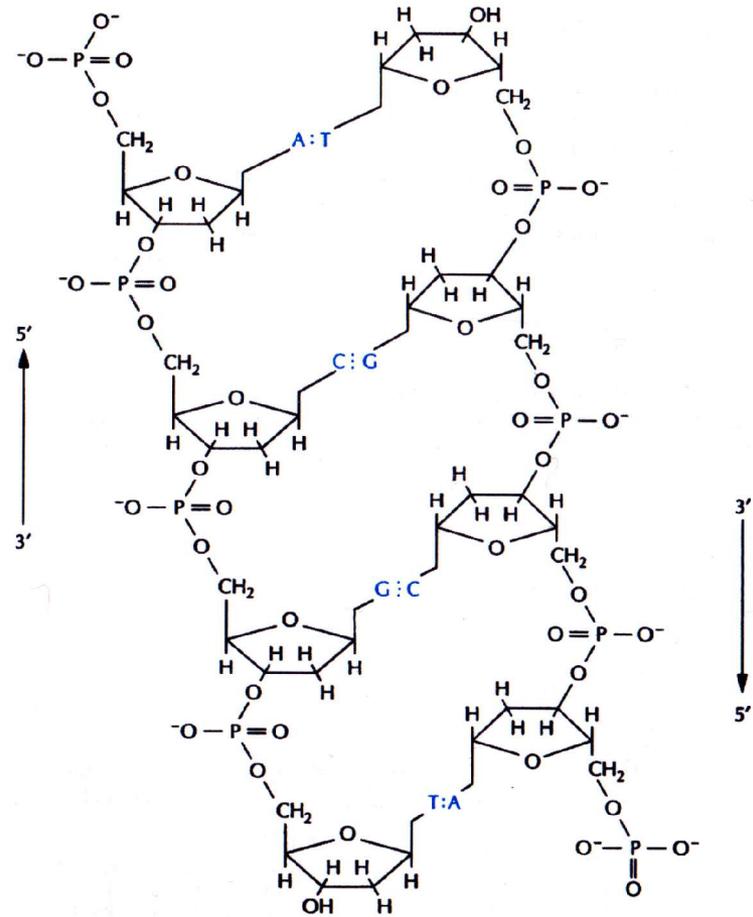


Рис. 3.4. Фрагмент двухцепочечной ДНК.



Пространственное строение двойной спирали ДНК

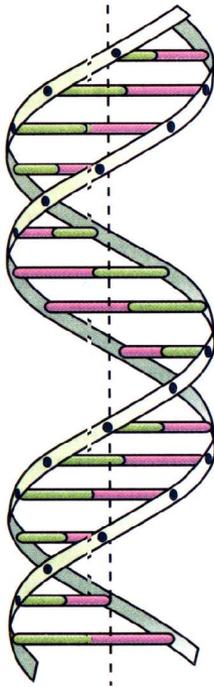
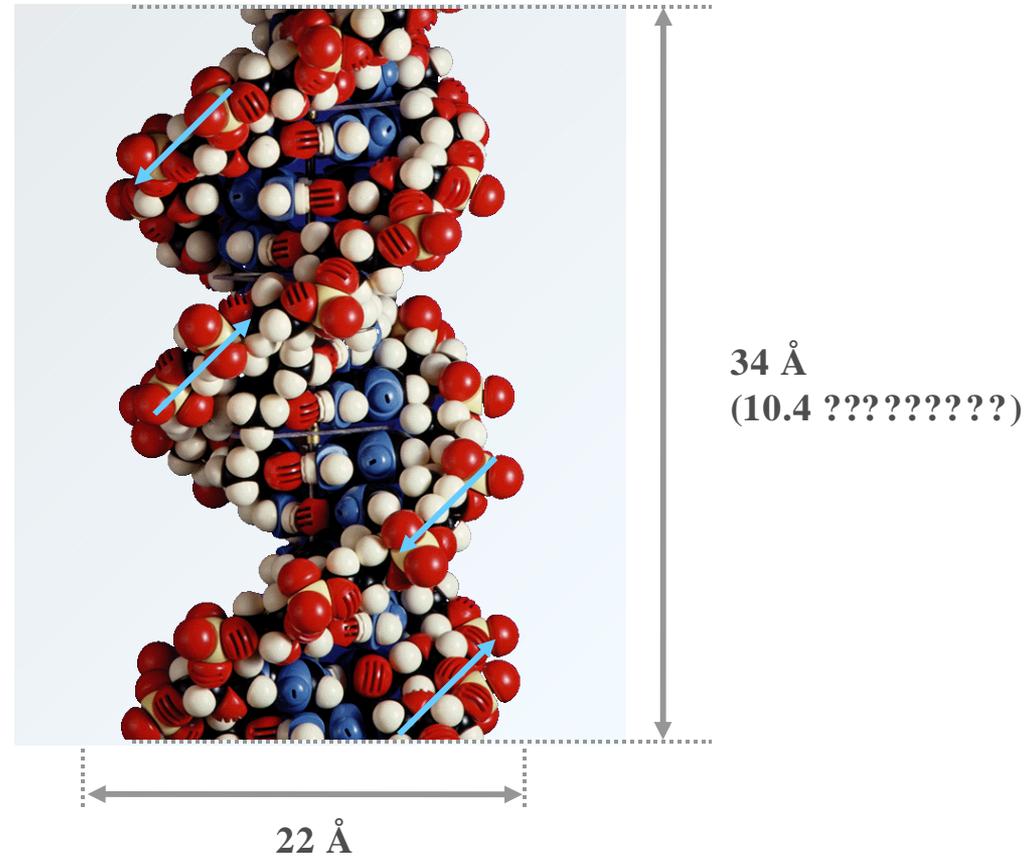
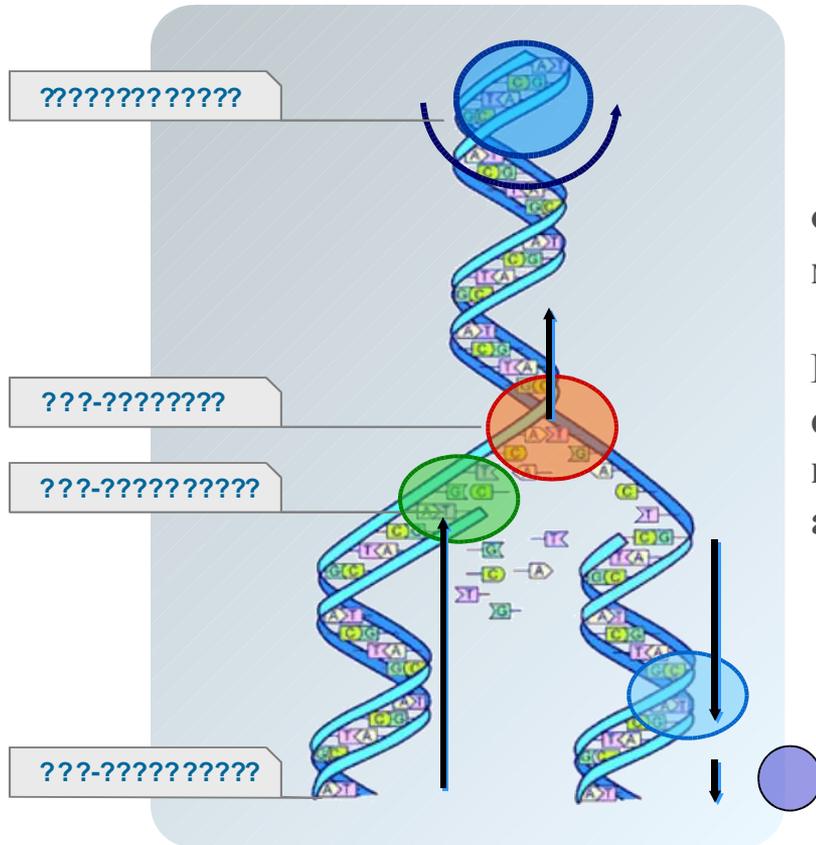


Рис. 3.3. Модель двойной спирали ДНК. Поперечные перекладины – комплементарные пары оснований, «бювины» – сахарофатный остов.





ФУНДАМЕНТАЛЬНЫЕ ГЕНЕТИЧЕСКИЕ ПРОЦЕССЫ: РЕПЛИКАЦИЯ.



Физико-химическое содержание:
матричный синтез ДНК по ДНК.

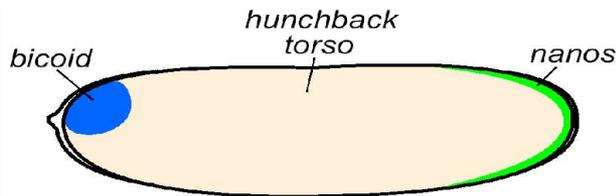
Информационное содержание:
самовоспроизведение (копирование)
генетической ДНК-памяти в исходном
алфавите {А, Т, G, С }



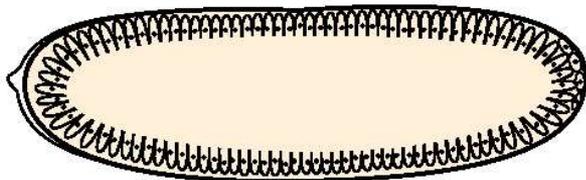
Эпигенетическая информация (КОНЦЕНТРАЦИЯ ОПРЕДЕЛЕННОГО ВЕЩЕСТВА, ОПРЕДЕЛЯЮЩАЯ НАЧАЛЬНЫЕ УСЛОВИЯ ФУНКЦИОНИРОВАНИЯ МОЛЕКУЛЯРНО-ГЕНЕТИЧЕСКИХ СИСТЕМ КЛЕТКИ)

Позиционированная

(концентрация вещества важна в определенных местах клетки)

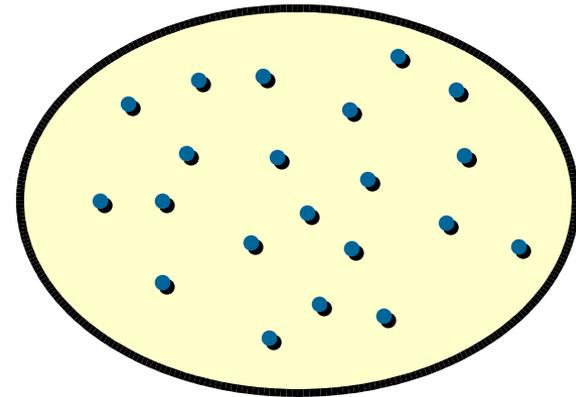


Распределение морфогенов РНК по продольной оси яйца дрозофилы (Из: Lawrence, 1992, p. 42).



Непозиционированная

(концентрация вещества важна вне зависимости от его положения в клетке)





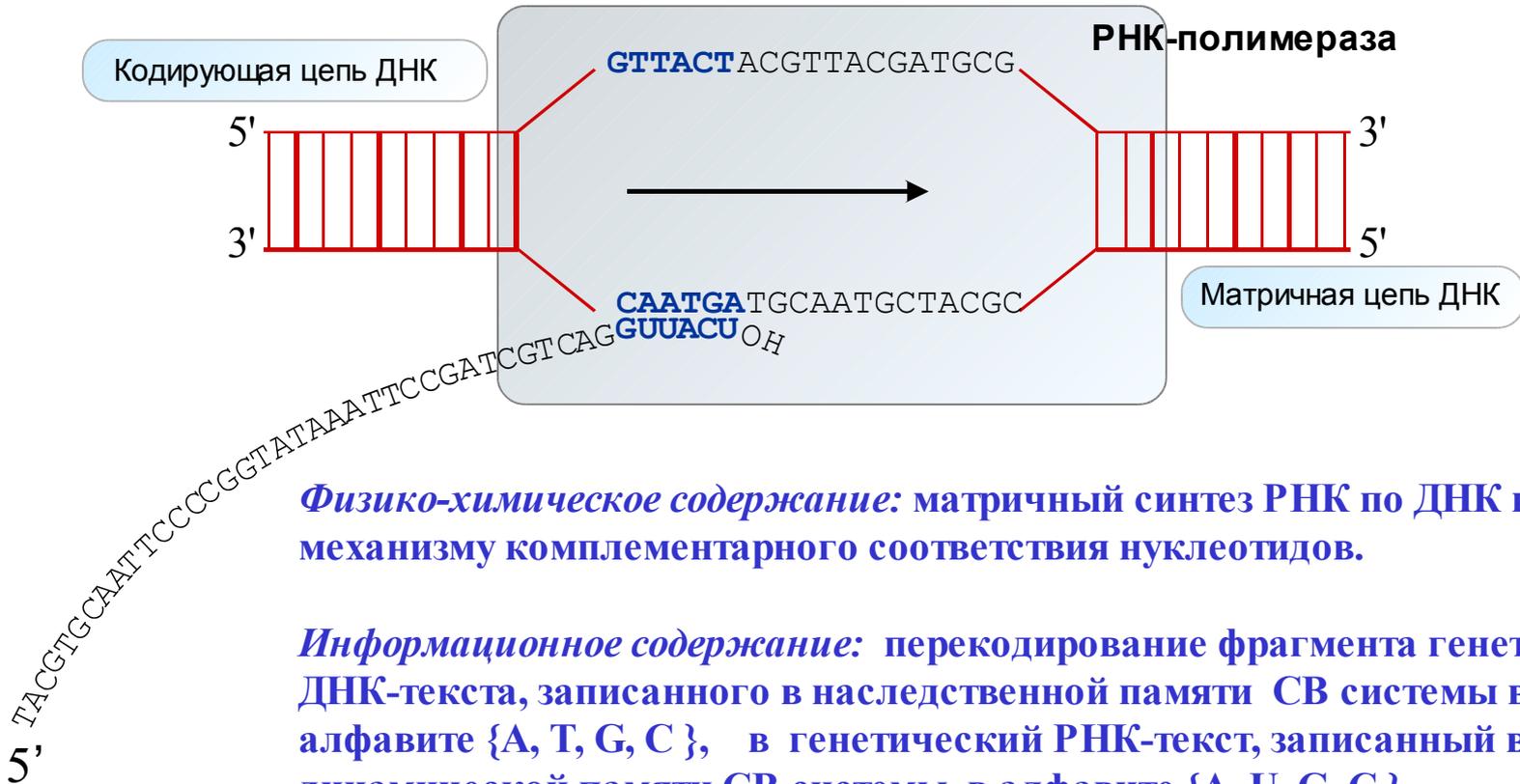
ПРИНЦИПИАЛЬНЫЕ КОМПОНЕНТЫ САМОВОСПРОИЗВОДЯЩИХСЯ СИСТЕМ



Считывающе-перекодирующее устройство - совокупность механизмов, структур и процессов, ответственных за считывание информации из наследственной памяти и ее перекодирование в различные типы генетических сообщений, необходимых для функционирования СВ системы.



Фундаментальные генетические процессы: транскрипция



Физико-химическое содержание: матричный синтез РНК по ДНК по механизму комплементарного соответствия нуклеотидов.

Информационное содержание: перекодирование фрагмента генетического ДНК-текста, записанного в наследственной памяти СВ системы в алфавите {А, Т, G, С}, в генетический РНК-текст, записанный в динамической памяти СВ системы в алфавите {А, U, G, С}

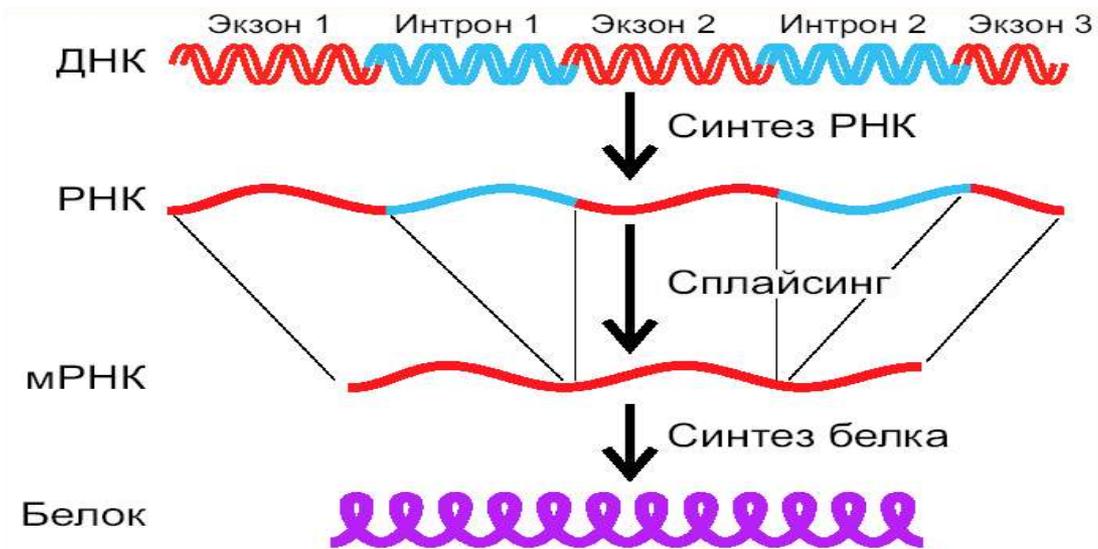


ФУНДАМЕНТАЛЬНЫЕ ГЕНЕТИЧЕСКИЕ ПРОЦЕССЫ: СПЛАЙСИНГ



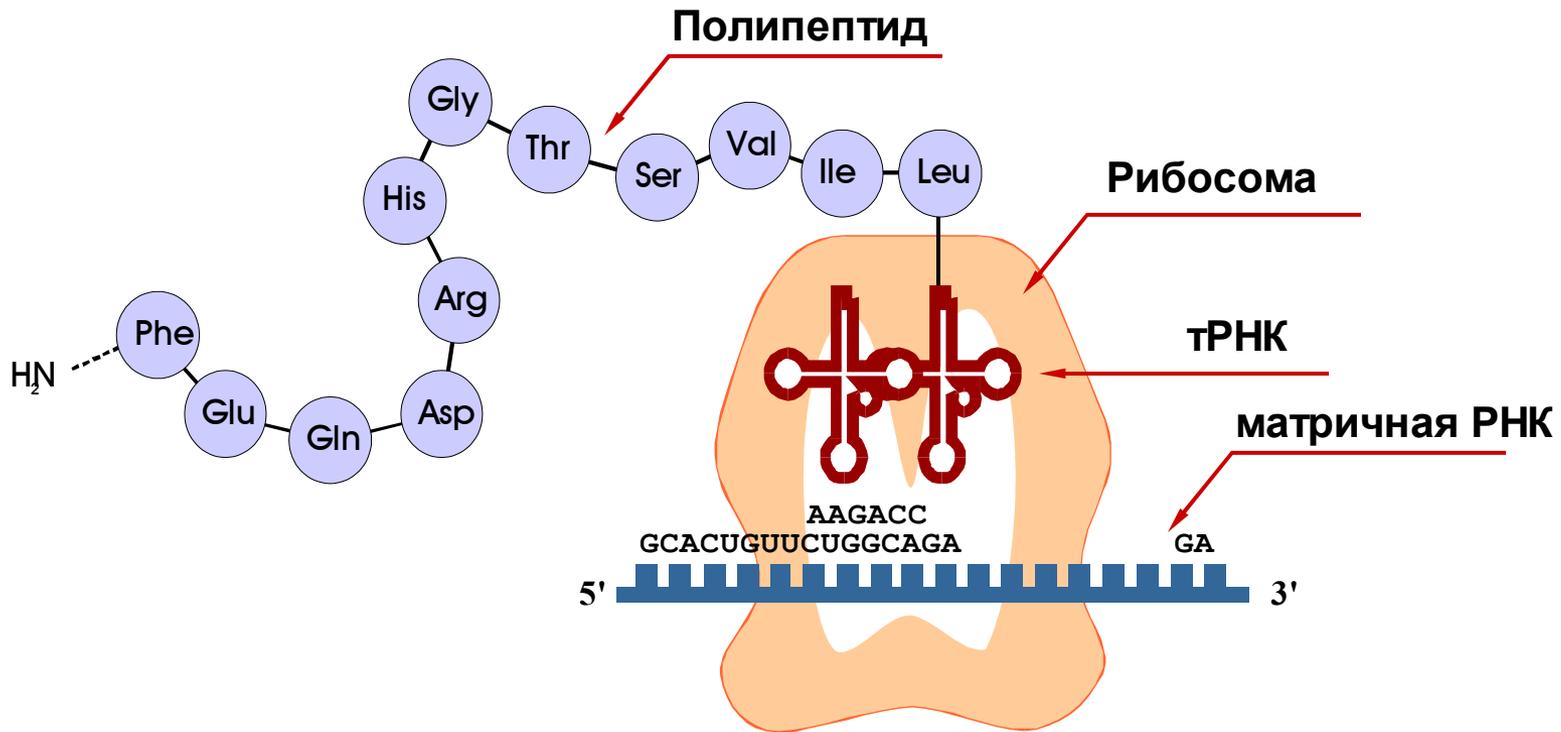
Физико-химическое содержание: вырезание фрагментов из пре-мРНК

Информационное содержание: удаление фрагментов, не несущих генетической информации из генетического РНК-текста





Фундаментальные генетические процессы: трансляция





ПРИНЦИПИАЛЬНЫЕ КОМПОНЕНТЫ САМОВОСПРОИЗВОДЯЩИХСЯ СИСТЕМ

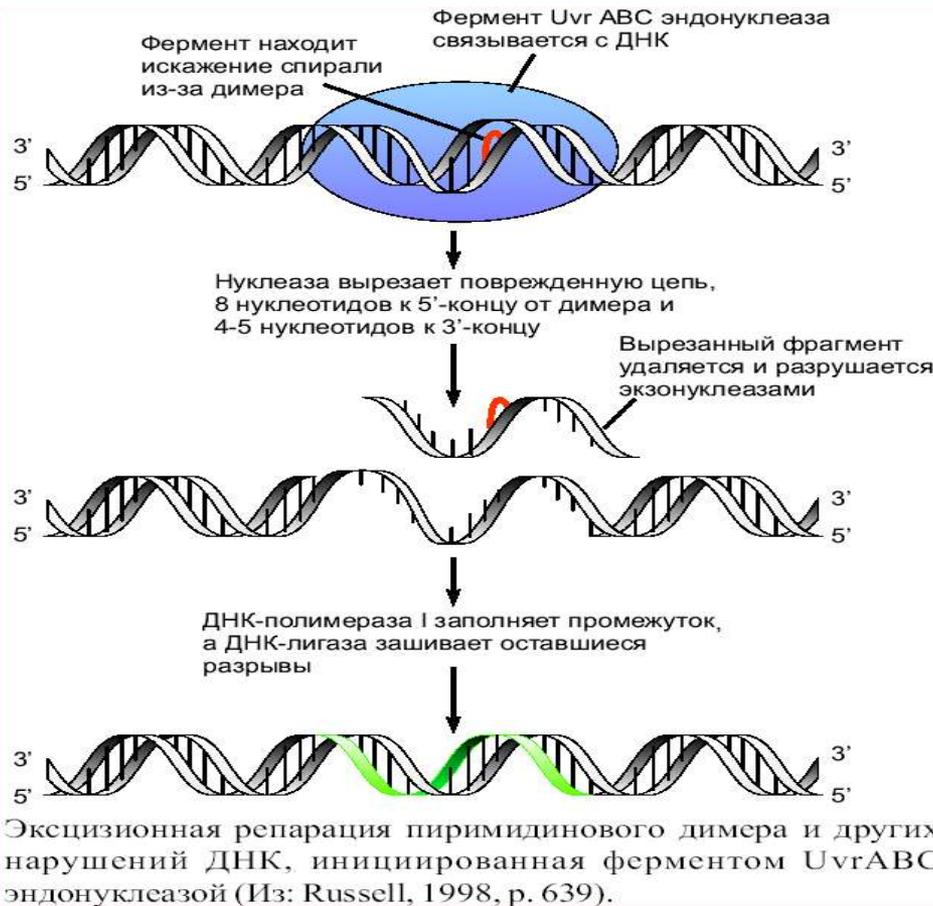


Репарирующая подсистема - совокупность структур и процессов, обеспечивающих восстановление поврежденных блоков наследственной памяти

ФУНДАМЕНТАЛЬНЫЕ ГЕНЕТИЧЕСКИЕ ПРОЦЕССЫ: ЭКСЦИЗИОННАЯ РЕПАРАЦИЯ ДНК

Физико-химическое содержание: восстановление комплементарности ДНК

Информационное содержание: восстановление поврежденных фрагментов ДНК-памяти



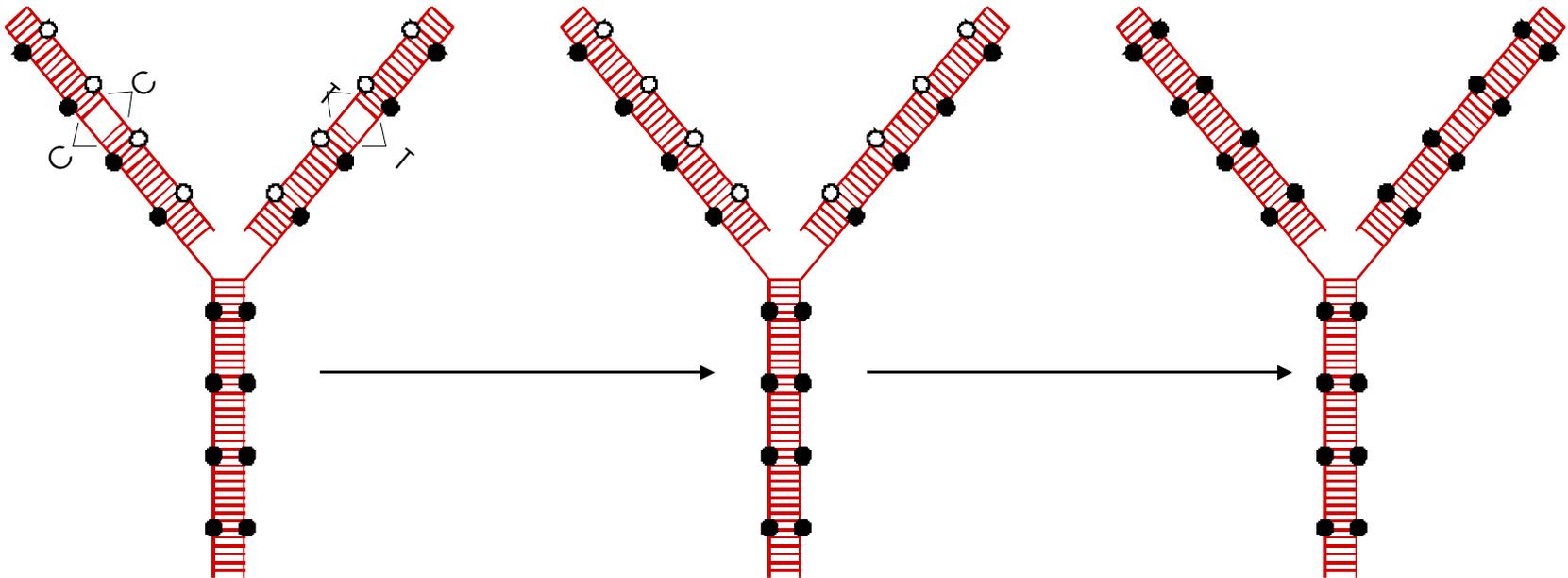


ФУНДАМЕНТАЛЬНЫЕ ГЕНЕТИЧЕСКИЕ ПРОЦЕССЫ: ПОСТРЕПЛИКАТИВНАЯ РЕПАРАЦИЯ ДНК



Физико-химическое содержание:
восстановление
комплементарности ДНК

Информационное содержание:
восстановление правильных
символов в поврежденной ДНК-памяти



Сигнал родительской нити ДНК - $GA^{Me}TC$ ●
В дочерней нити ДНК GATC первоначально не метилирован ○



ПРИНЦИПИАЛЬНЫЕ КОМПОНЕНТЫ САМОВОСПРОИЗВОДЯЩИХСЯ СИСТЕМ



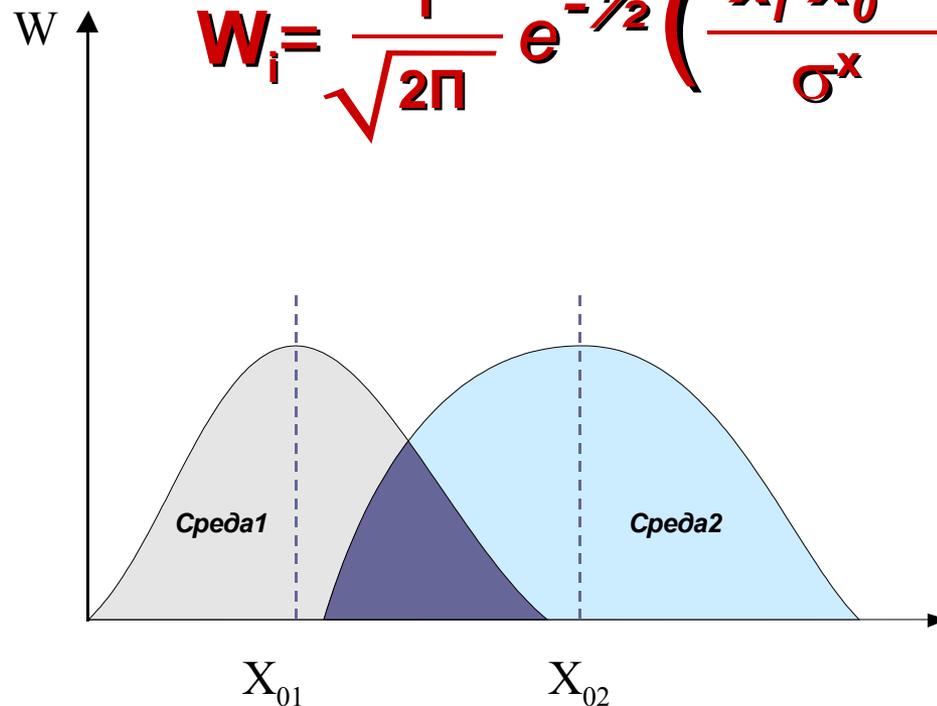
Подсистема генерации генетической изменчивости. *Наличие необходимого уровня изменчивости генетической памяти - критическое условие для выживания СВ систем при резких изменениях условий окружающей среды.*



НЕВЫРОЖДЕНИЕ ПОПУЛЯЦИЙ В ИЗМЕНЯЮЩИХСЯ УСЛОВИЯХ ВНЕШНЕЙ СРЕДЫ:

в популяции имеется резерв генетической изменчивости, обеспечивающей ее выживание в новых условиях среды.

$$W_i = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{X_i - X_0}{\sigma^x} \right)^2}$$

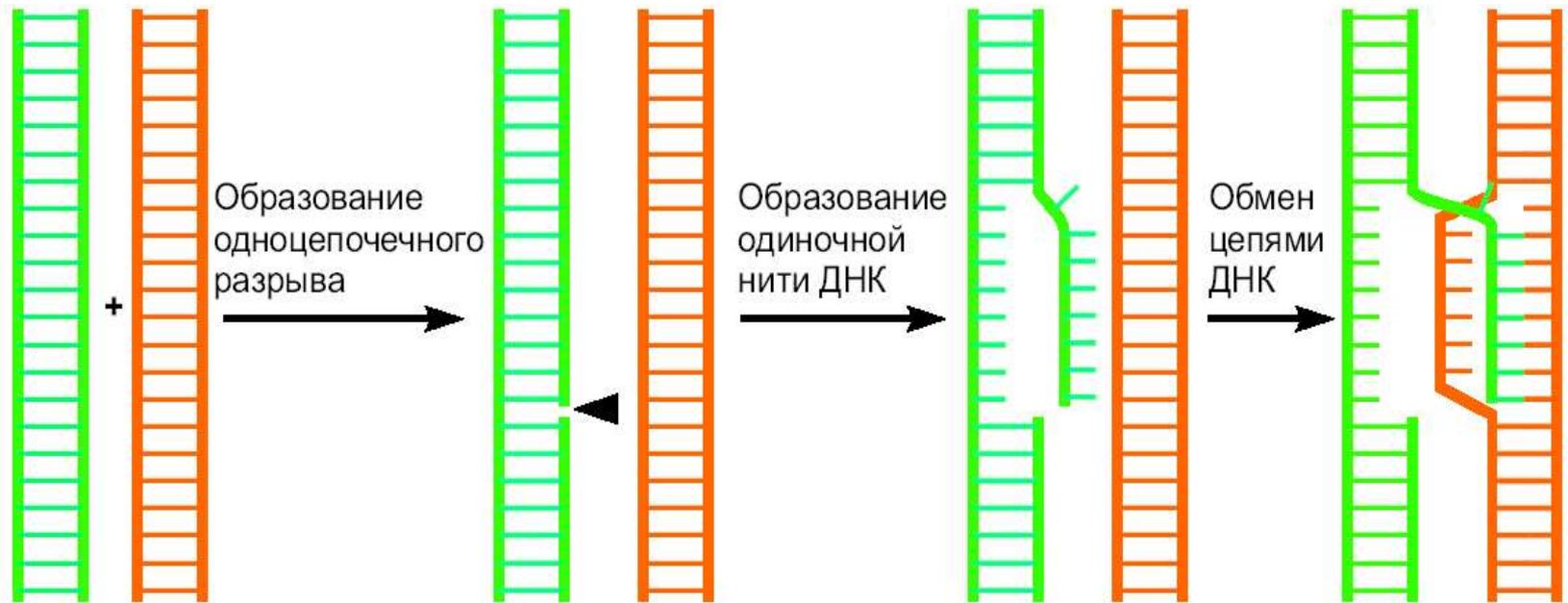


Распределение особей по приспособленности W_i для двух состояний внешней среды.

X_i - критический параметр организма, определяющий его приспособленность

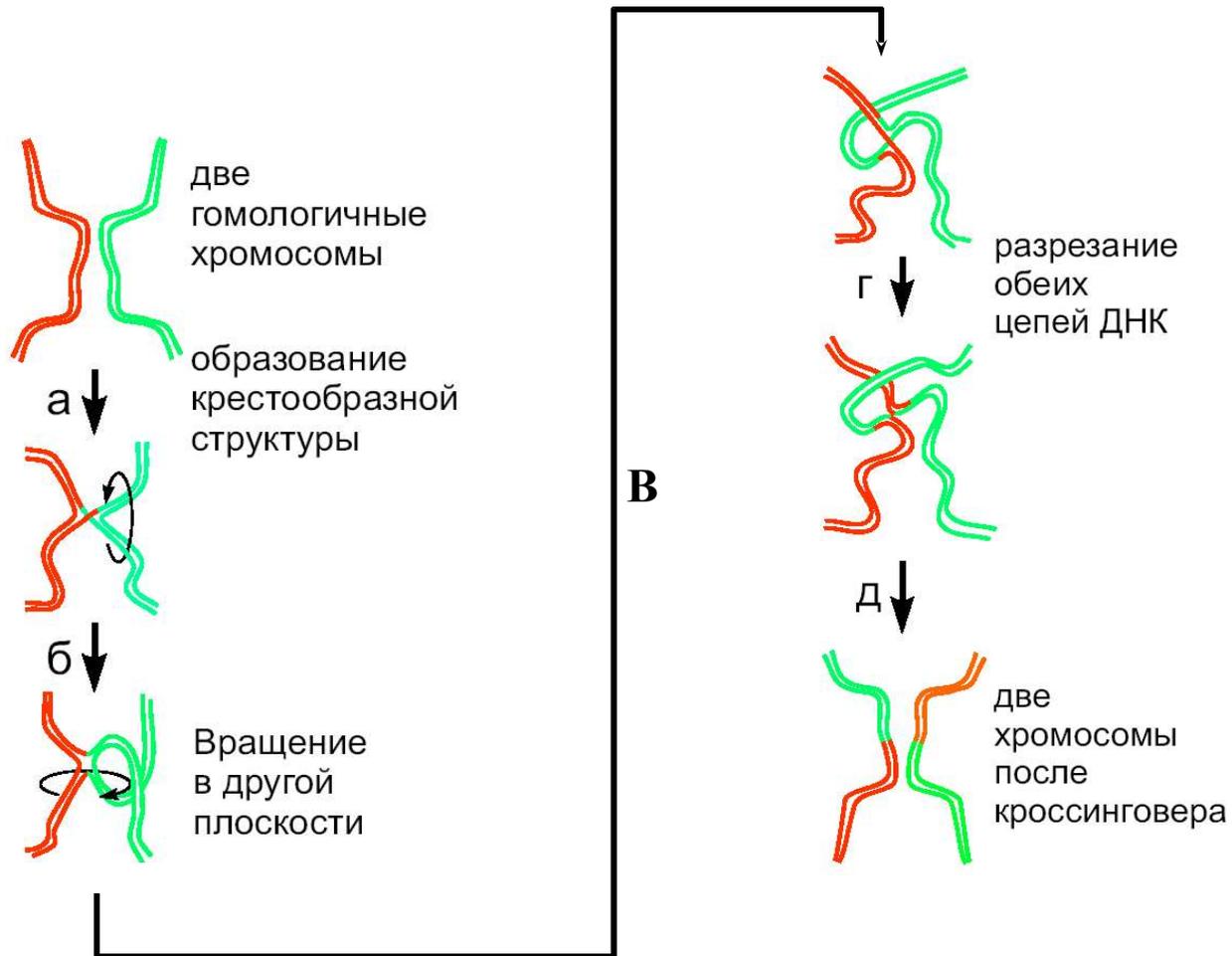
X_{01} , X_{02} - оптимальные значения критического параметра для среды 1 и среды 2

ПРОЦЕССЫ ГЕНЕРАЦИИ ГЕНЕТИЧЕСКОЙ ИЗМЕНЧИВОСТИ: Гомологичная рекомбинация ДНК



Первоначальный обмен нитями в схеме общей рекомбинации (Из: Alberts et al., 1994, 265).

ПРОЦЕССЫ ГЕНЕРАЦИИ ГЕНЕТИЧЕСКОЙ ИЗМЕНЧИВОСТИ: ГОМОЛОГИЧНАЯ РЕКОМБИНАЦИЯ ДНК





ПРИНЦИПИАЛЬНЫЕ КОМПОНЕНТЫ САМОВОСПРОИЗВОДЯЩИХСЯ СИСТЕМ



Подсистема безопасности - совокупность структур и процессов, обеспечивающих защиту СВ системы от проникновения в нее других СВ систем и чужеродной генетической информации



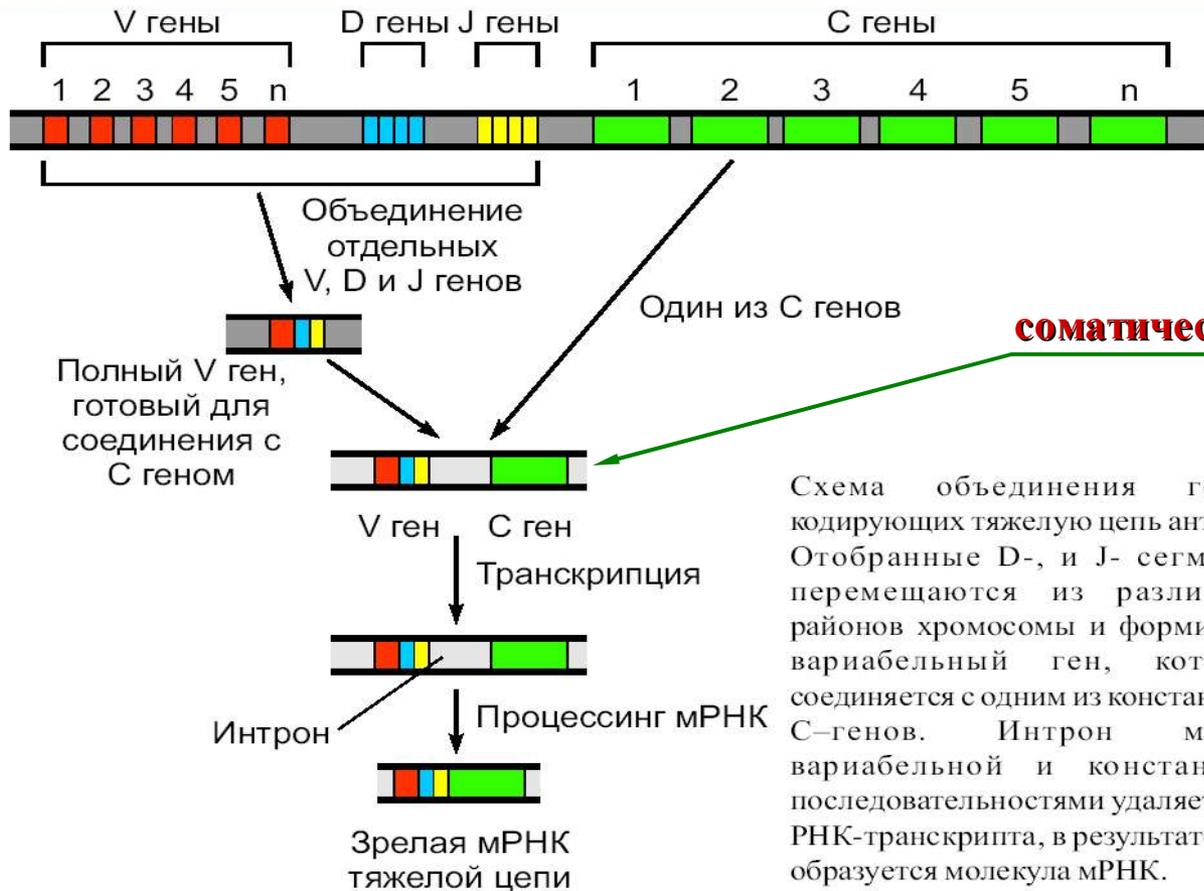
ФУНКЦИИ ПОДСИСТЕМ БЕЗОПАСНОСТИ СВ СИСТЕМ



- Защита СВ системы от проникновения в нее чужеродной генетической информации
- Подавление развития паразитических подсистем в пределах данной СВ системы
- Предотвращение «смешивания» наследственной памяти данной СВ системы с наследственной памятью других СВ систем
- Распознавание и уничтожение «чужеродной» информации, проникшей в СВ систему



Подсистемы безопасности самовоспроизводящихся систем: иммунная система



соматический мутагенез

Схема объединения генов, кодирующих тяжелую цепь антител. Отобранные D-, и J- сегменты перемещаются из различных районов хромосомы и формируют переменный ген, который соединяется с одним из константных C-генов. Интрон между переменной и константной последовательностями удаляется из РНК-транскрипта, в результате чего образуется молекула мРНК.



ПРИНЦИПИАЛЬНЫЕ КОМПОНЕНТЫ САМОВОСПРОИЗВОДЯЩИХСЯ СИСТЕМ



Исполняющая подсистема - инструктируемая наследственной памятью совокупность структур и процессов, обеспечивающих выполнение всех жизненно-важных функций СВ системы

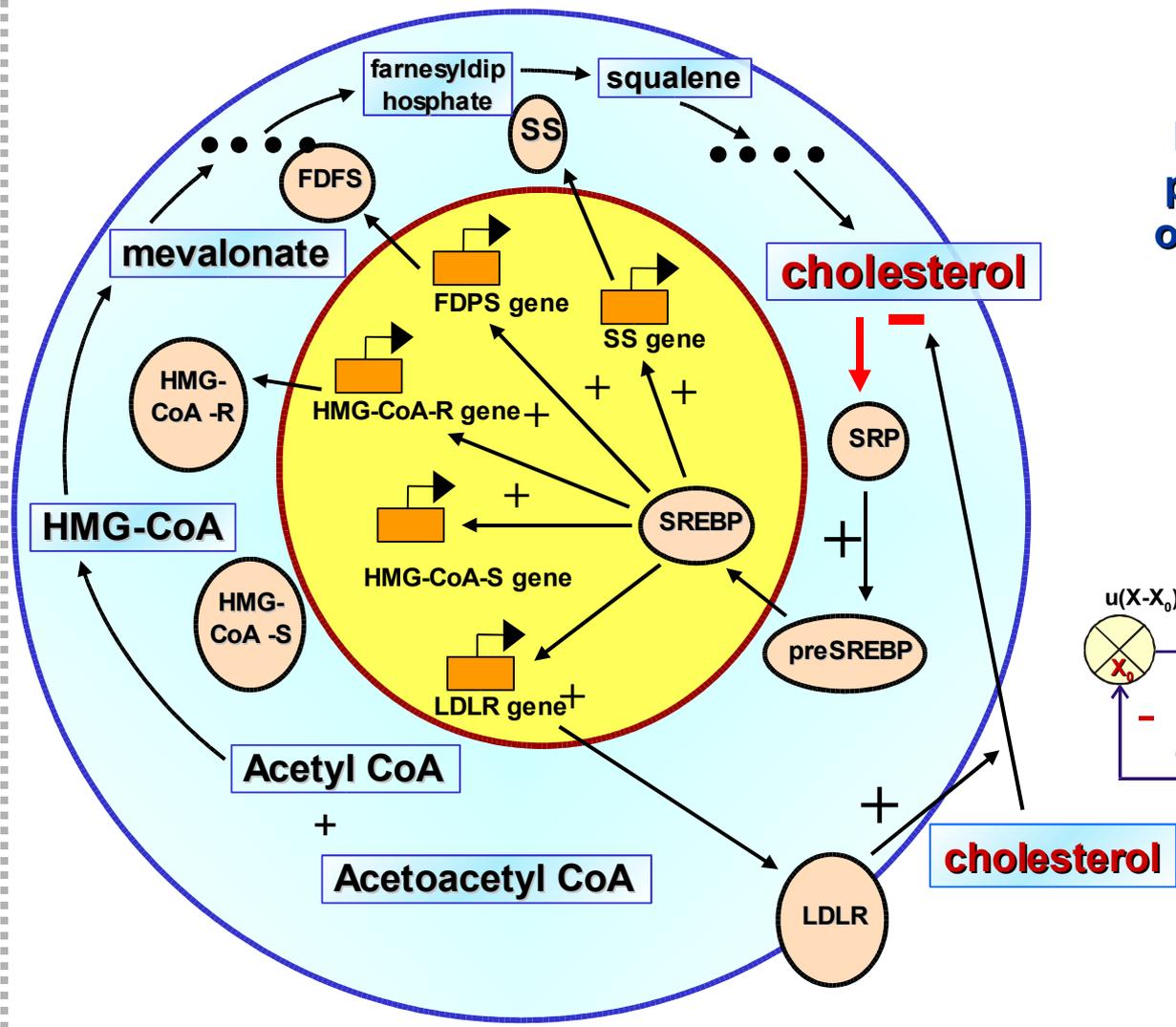


ИСПОЛНЯЮЩИЕ ПОДСИСТЕМЫ СВ СИСТЕМ - ГЕННЫЕ СЕТИ

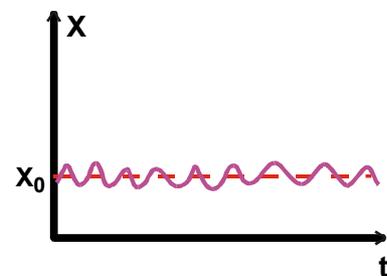


Генная сеть - группа
координированно функционирующих
генов, контролирующих
физиологические, биохимические,
молекулярные
и др. функции организмов

БАЗА ДАННЫХ GENENET: ЦЕНТРАЛЬНЫЙ ФРАГМЕНТ ГЕННОЙ СЕТИ БИОСИНТЕЗА ХОЛЕСТЕРИНА В КЛЕТКЕ (РЕГУЛЯЦИЯ ПО МЕХАНИЗМУ ОТРИЦАТЕЛЬНОЙ ОБРАТНОЙ СВЯЗИ)



Принципиальная схема регуляторного контура с отрицательной обратной связью



**ИНФОРМАЦИОННОЕ
СОДЕРЖАНИЕ ГЕНОМОВ И
КОДИРОВАНИЕ
ГЕНЕТИЧЕСКОЙ
ИНФОРМАЦИИ**



ИНФОРМАЦИОННОЕ СОДЕРЖАНИЕ ГЕНОМОВ И КОДИРОВАНИЕ ГЕНЕТИЧЕСКОЙ ИНФОРМАЦИИ



Под генетическим кодом будем понимать любой тип нуклеотидного контекста, значимый для выполнения определенной биомолекулярной функции. Изучение природы кодов функционирования генетических систем - одна из центральных задач информационной биологии.



ИНФОРМАЦИОННОЕ СОДЕРЖАНИЕ ГЕНОМОВ И КОДИРОВАНИЕ ГЕНЕТИЧЕСКОЙ ИНФОРМАЦИИ: ТРИПЛЕТНЫЙ ГЕНЕТИЧЕСКИЙ КОД



Триплетный генетический код обеспечивает отображение последовательности мРНК в аминокислотную последовательность.



Свойства генетического кода



		Second Position of Codon					
		T	C	A	G		
F i r s t P o s i t i o n	T	TTT Phe [F]	TCT Ser [S]	TAT Tyr [Y]	TGT Cys [C]	T	T h i r d P o s i t i o n
		TTC Phe [F]	TCC Ser [S]	TAC Tyr [Y]	TGC Cys [C]	C	
		TTA Leu [L]	TCA Ser [S]	TAA <i>Ter</i> [end]	TGA <i>Ter</i> [end]	A	
		TTG Leu [L]	TCG Ser [S]	TAG <i>Ter</i> [end]	TGG Trp [W]	G	
	C	CTT Leu [L]	CCT Pro [P]	CAT His [H]	CGT Arg [R]	T	
		CTC Leu [L]	CCC Pro [P]	CAC His [H]	CGC Arg [R]	C	
		CTA Leu [L]	CCA Pro [P]	CAA Gln [Q]	CGA Arg [R]	A	
		CTG Leu [L]	CCG Pro [P]	CAG Gln [Q]	CGG Arg [R]	G	
	A	ATT Ile [I]	ACT Thr [T]	AAT Asn [N]	AGT Ser [S]	T	
		ATC Ile [I]	ACC Thr [T]	AAC Asn [N]	AGC Ser [S]	C	
		ATA Ile [I]	ACA Thr [T]	AAA Lys [K]	AGA Arg [R]	A	
		ATG Met [M]	ACG Thr [T]	AAG Lys [K]	AGG Arg [R]	G	
	G	GTT Val [V]	GCT Ala [A]	GAT Asp [D]	GGT Gly [G]	T	
		GTC Val [V]	GCC Ala [A]	GAC Asp [D]	GGC Gly [G]	C	
		GTA Val [V]	GCA Ala [A]	GAA Glu [E]	GGA Gly [G]	A	
		GTG Val [V]	GCG Ala [A]	GAG Glu [E]	GGG Gly [G]	G	

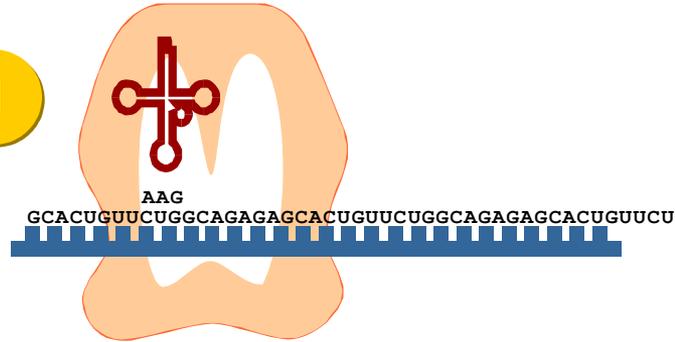
Генетический код универсален.



МОЛЕКУЛЯРНЫЕ КОМПОНЕНТЫ ТРАНСЛЯЦИОННОЙ МАШИНЫ

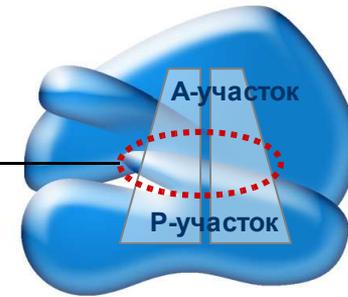


1

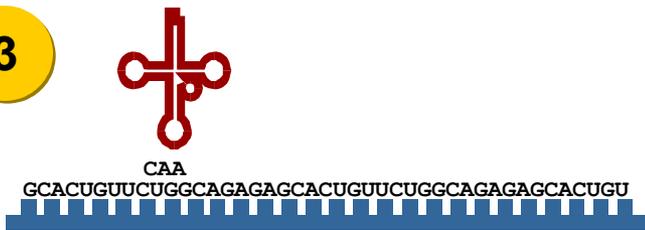


2

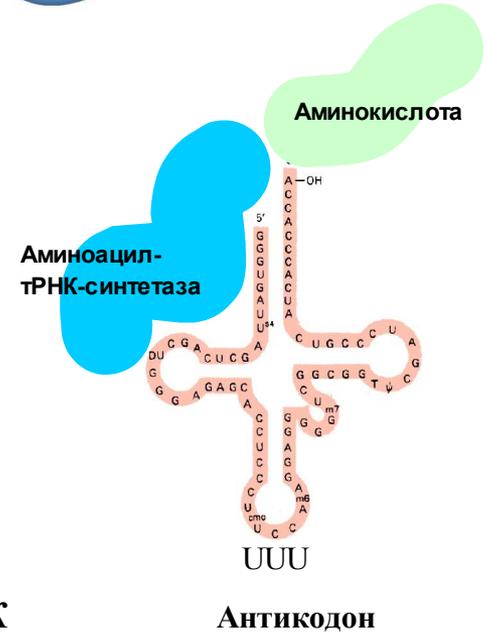
Участок связывания мРНК



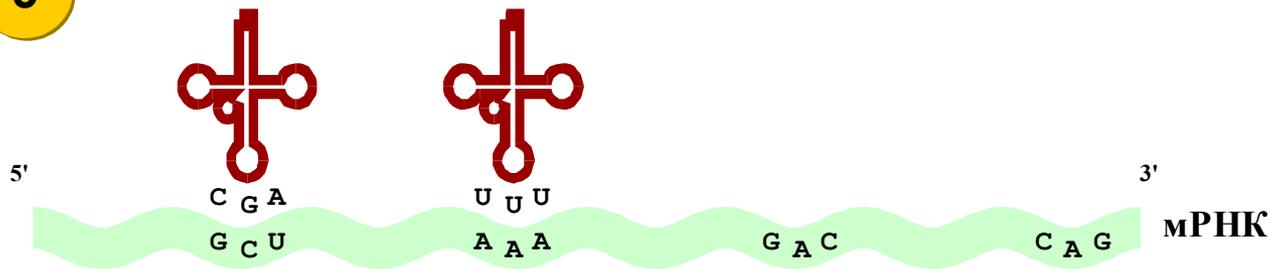
3



4

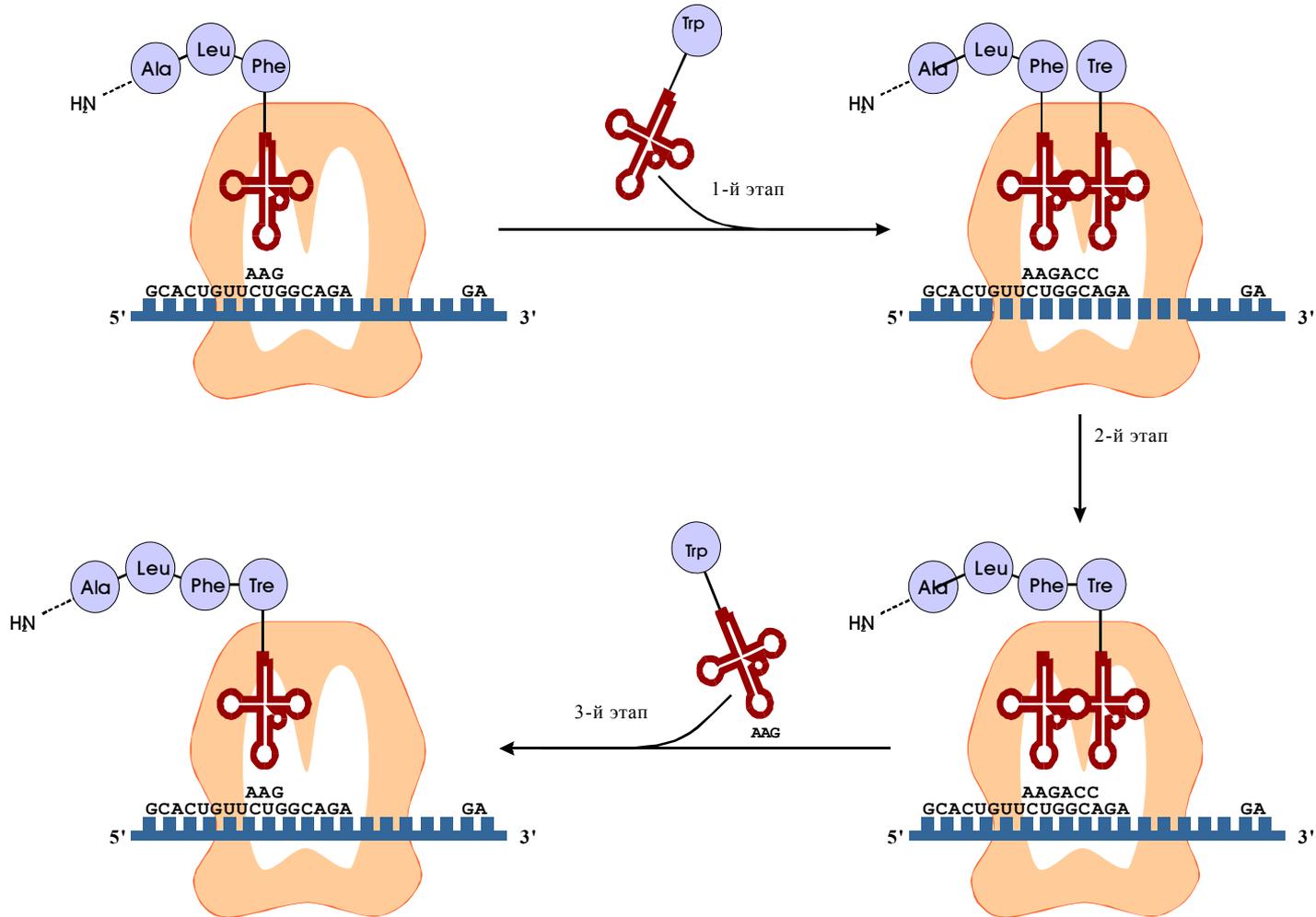


5



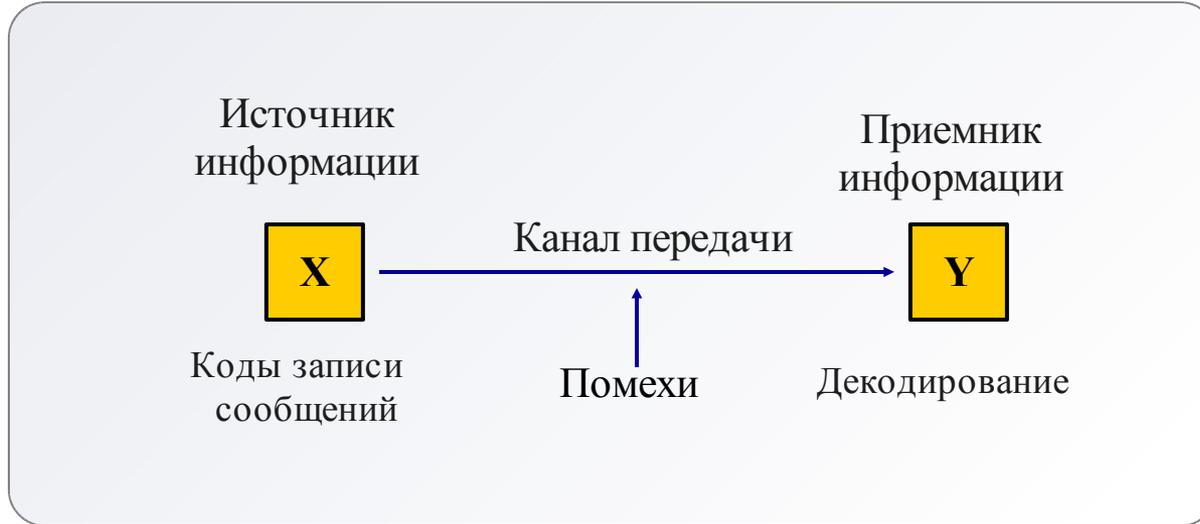


Принципиальная схема процесса трансляции



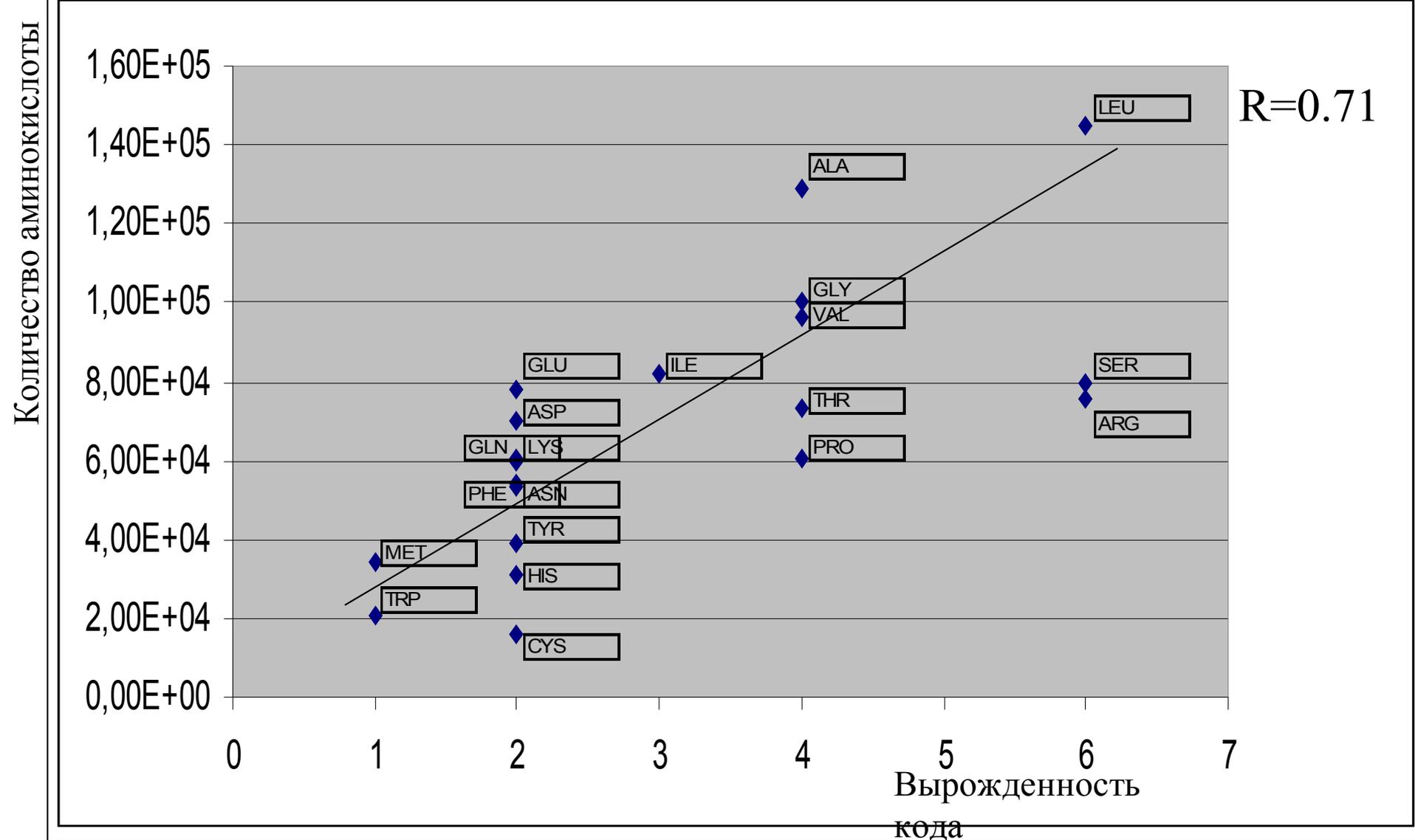


ЭЛЕМЕНТАРНЫЙ ОБЪЕКТ ТЕОРИИ ИНФОРМАЦИИ - КАНАЛ ПЕРЕДАЧИ ИНФОРМАЦИИ С ШУМОМ



Примеры передачи информации в биологических СВ системах

- (1) Передача информации клетке-потомку от родительской клетки;
- (2) Передача информации от родителя (родителей) к потомку в процессе самовоспроизведения биологической системы
- (3) Передача информации по внутренним коммуникациям СВ системы в ходе ее функционирования.



Помехоустойчивость генетического кода: положительная зависимость между частотами использования аминокислот и вырожденностью их кодовых серий



ПОМЕХОУСТОЙЧИВОСТЬ ГЕНЕТИЧЕСКОГО КОДА ПРИ ОДИНОЧНЫХ НУКЛЕОТИДНЫХ ЗАМЕНАХ



ГЕНЕТИЧЕСКИЙ КОД ОБЕСПЕЧИВАЕТ ПОВЫШЕННУЮ ЧАСТОТУ
АМИНОКИСЛОТНЫХ ЗАМЕН
С СОХРАНИЕМ ФИЗИКО-ХИМИЧЕСКИХ СВОЙСТВ АМИНОКИСЛОТ

СРЕДНЕЕ ИЗМЕНЕНИЕ ГИДРОФОБНОСТИ АМИНОКИСЛОТ ПРИ ЗАМЕНАХ

870 ккал/моль

1280 ккал/моль

Примеры замен полярных аминокислот(п) на гидрофобные (г) и наоборот при одиночных нуклеотидных заменах

АМИНОКИСЛОТА

LEU (Г) → ARG (П) → ILE (Г) → ARG (П) → VAL (Г) → GLU (П)
CUU CGU AUA AGA GUU GAU

КОДОН

Примеры замен с сохранением гидрофобности (Г) или полярности(П) при одиночных нуклеотидных заменах

LEU (Г) → PHE (Г) → LEU (Г) → ILE (Г) → GLN (П) → LYS (П)
UUA UUU CUU AUU GAA AAA



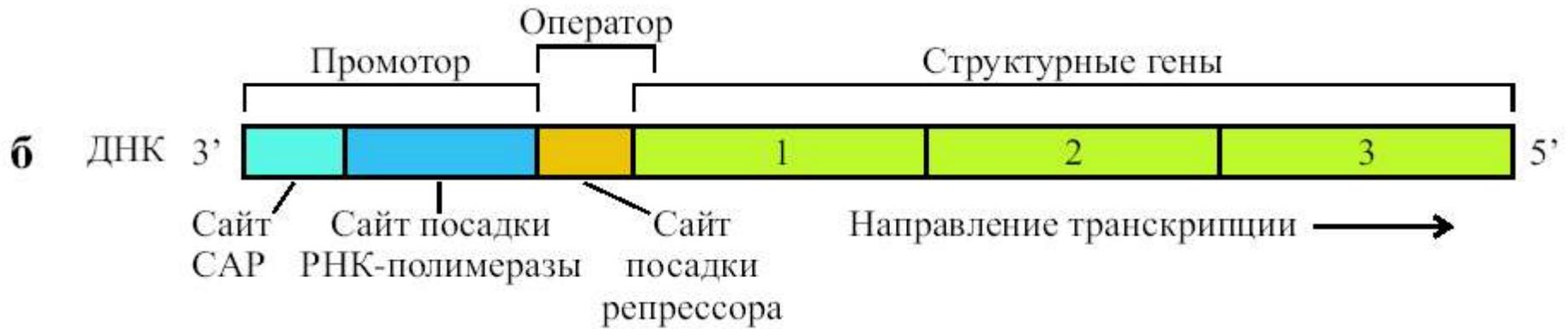
***Исходная мотивация
рассмотрения проблемы –
парадокс некодирующей ДНК:
для геномов эукариот характерна
низкая плотность белок-кодирующих
областей (до нескольких процентов
геномной ДНК).***



Характерные особенности геномов прокариот

- 1. Небольшие размеры – от 500 000 до 3-4-х миллионов п. о.**
- 2. Небольшое количество генов - от 400 до 4000;**
- 3. Высокая плотность кодирования генетической информации (выше 90%);**
- 4. Малое расстояние между генами (до нескольких сот п.о.);**
- 5. Простая структурная организация генов;**
- 6. Малые размеры регуляторных районов (до 60 – 100 п.о.).**
- 7. Объединение групп генов в опероны;**
- 8. Небольшое количество факторов транскрипции;**
- 9. Обедненность повторяющимися последовательностями;**
- 10. Отсутствие изофункциональных семейств генов.**

СТРОЕНИЕ ГЕНА (ОПЕРОНА) ПРОКАРИОТ





ИНФОРМАЦИОННОЕ СОДЕРЖАНИЕ ГЕНОМОВ И КОДИРОВАНИЕ ГЕНЕТИЧЕСКОЙ ИНФОРМАЦИИ



РАЗМЕРЫ ГЕНОМОВ ПРО- И ЭУКАРИОТ

	?.?.	n
<i>E.coli</i>	5×10^6	1
<i>S. cerevisiae</i>	12×10^6	16
<i>D. melanogaster</i>	1×10^8	4
<i>M. musculus</i>	3×10^9	20
<i>H. sapiens</i>	3×10^9	23



Характерные особенности геномов многоклеточных эукариот

- 1. Большие размеры – до 10^{12} п.о.;**
- 2. Большое количество генов – до 30000 - 40000;**
- 3. Низкая плотность белок-кодирующих районов (2 – 4%);**
- 4. Большие расстояния между генами (до сотен тысяч п.о.);**
- 5. Мозаичная (экзон-интронная) структура генов;**
- 6. Сплайсинг (вырезание интронов);**
- 7. Протяженные регуляторные районы (до десятков тысяч п.о.);**
- 8. Большое количество факторов транскрипции (более 1000);**
- 9. Высокая насыщенность повторяющимися последовательностями;**
- 10. Наличие кластеров изофункциональных семейств генов и механизмов их регуляции (Locus control regions);**
- 11. Нуклеосомная упаковка ДНК и сложноорганизованный хроматин.**

СТРОЕНИЕ И СХЕМА ЭКСПРЕССИИ ГЕНА ЭУКАРИОТ

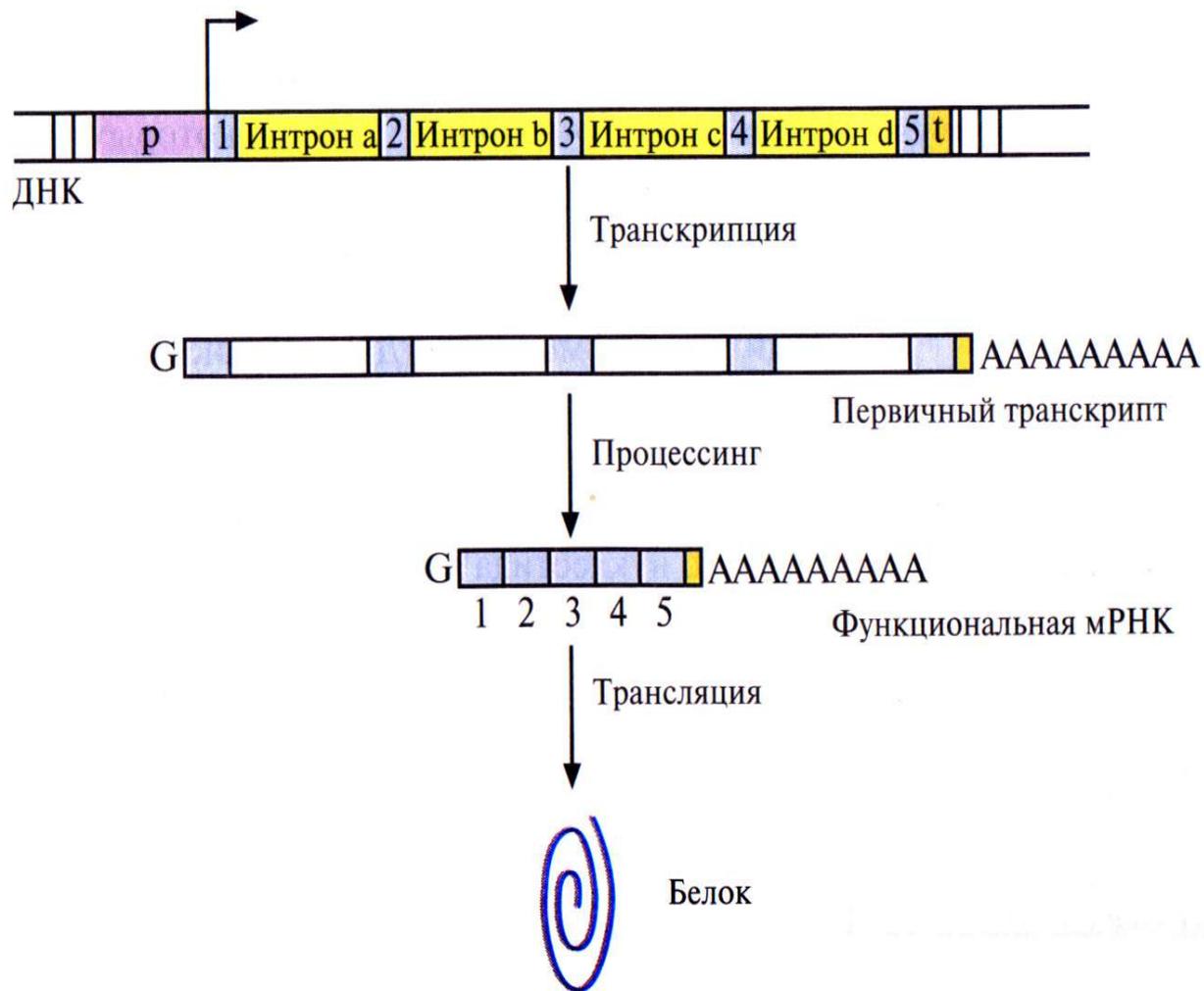


Рис. 3.11. Схематическое изображение эукариотического структурного гена. Указаны промотор (p), сайт инициации транскрипции и ее направление (горизонтальная стрелка), область терминации транскрипции, узнаваемая РНК-полимеразой (t). 1–5 – экзоны, а–d – интроны. Первичный транскрипт содержит poly(A)-«хвост» на 3'-конце и метилированный нуклеотид G («кэп») на 5'-конце. После транскрипции интроны из первичного транскрипта вырезаются (процессинг), и на образовавшейся функциональной РНК синтезируется белковая молекула (трансляция).



ИНФОРМАЦИОННОЕ СОДЕРЖАНИЕ ГЕНОМОВ И КОДИРОВАНИЕ ГЕНЕТИЧЕСКОЙ ИНФОРМАЦИИ



ХАРАКТЕРИСТИКИ ГЕНОМА ЧЕЛОВЕКА

Размер генома, включая разрывы	2.91 млрд. пар осн.
Количество А/Т оснований	54%
Количество G/C оснований	38%
Количество повторов	35%
Количество генов	26 000 -39 000
Ген с наибольшим количеством экзонов	Титин (234 экзона)
Средний размер гена	27 тыс. пар осн.
Хромосома, несущая наибольшее количество генов	№ 19 (23 гена/млн. пар осн.)
Хромосома, несущая наименьшее количество генов	№ 13 (5 генов/млн. пар осн.), Y -хромосома (5 генов/млн. пар осн.)
Количество пар оснований, составляющих гены	25,5-37,8%
Количество пар оснований, составляющих экзоны	1,1-1,4%
Количество пар оснований, составляющих интроны	24,4-36,4%
Количество пар оснований в промежутках между генами	63,6-74,5%
Самый протяженный промежуток между генами	3 038 416 пар оснований (Хромосома № 13)
Частота SNP -вариаций	1/1250 пар оснований

**ТРИПЛЕТНЫЙ КОД ЗАНИМАЕТ НЕ БОЛЕЕ 3 %
ГЕНОМА ЭУКАРИОТ.
КАКАЯ ГЕНЕТИЧЕСКАЯ ИНФОРМАЦИЯ
ХРАНИТСЯ В ОСТАЛЬНЫХ 97% ГЕНОМА?**

“Преобладающие в геномах эукариот некодирующие последовательности кодируют нечто иное, что не требует привлечения традиционного триплетного кода. Иными словами, кроме триплетного кода и трансляционной машины, в клетке имеются другие коды и средства их чтения. При этом под кодом понимается любой тип нуклеотидного контекста, значимый для выполнения определенной биомолекулярной функции”.

Существует множество нетриплетных кодов для записи генетической информации

УРОВЕНЬ ДНК

- ☀ Код конформации ДНК
- ☀ Код активности функциональных сайтов
- ☀ Код формирования хроматина
- ☀ Код эволюционной модуляции (быстрой адаптации) генома
- ☀ Код регуляции транскрипции

УРОВЕНЬ РНК

- ☀ Код сплайсинга РНК
- ☀ Код укладки РНК
- ☀ Код стабильности РНК
- ☀ Код регуляции трансляции мРНК

УРОВЕНЬ БЕЛКА

- ☀ Код укладки белка
- ☀ Код адресации белка
- ☀ Код стабильности белка