

*Федеральное государственное бюджетное образовательное учреждение высшего образования
"Красноярский государственный медицинский университет имени профессора В.Ф. Войно-Ясенецкого"
Министерства здравоохранения Российской Федерации*

Кафедра медицинской кибернетики и информатики

Дисциплина:

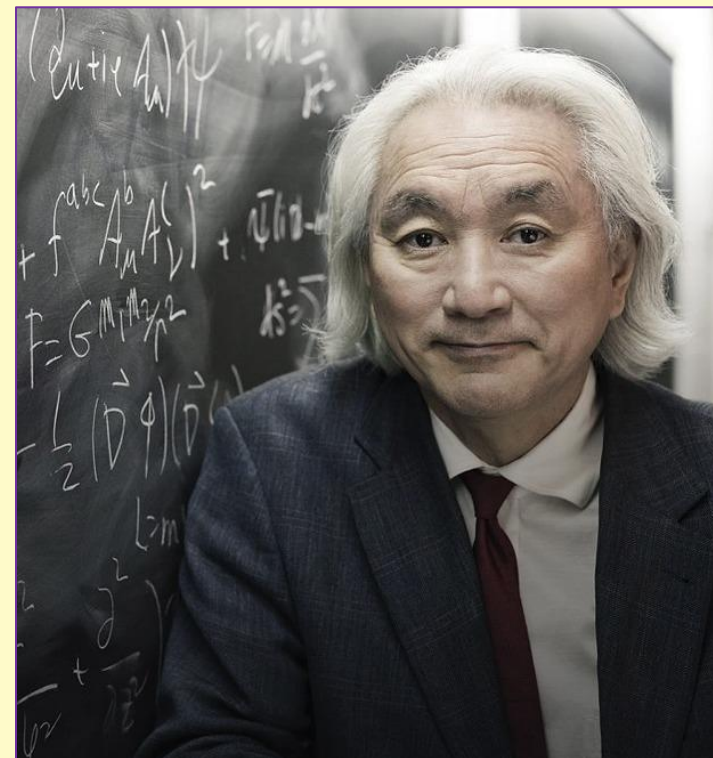
КОМПЬЮТЕРНОЕ МОДЕЛИРОВАНИЕ В МЕДИЦИНЕ

Лекция 9

Введение в биоинформатику

Шадрин Константин Викторович

Будущее медицины



Митио Каку
(род. 1947)



Пла

1. Понятие биоинформатики.
2. Задачи и методы биоинформатики.
3. Молекула ДНК.
4. Основы анализа нуклеиновых последовательностей.
5. Выравнивание нуклеиновых последовательностей.
6. Анализ последовательностей с использованием компьютера.

Молекулярная

эволюция
□ Молекулярная эволюция — это наука, изучающая изменения генетических макромолекул (ДНК, РНК, белков) в процессе эволюции, закономерности и механизмы этих изменений, а также реконструирующая эволюционную историю генов и организмов.

Объекты исследования:

1. Последовательности нуклеиновых кислот (ДНК и РНК) как носителей генетической информации.
2. Последовательности белков.
3. Структура белков.
4. Геномы организмов.

Основные задачи:

- 1. Выявление закономерностей эволюции генетических макромолекул.
- 2. Реконструкция эволюционной истории генов и организмов. 4

Биоинформатика

□ *Биоинформатика* — совокупность методов и подходов, включающих в себя:

- **математические методы компьютерного анализа** в сравнительной геномике (геномная биоинформатика);
- **разработку алгоритмов и программ для предсказания** пространственной **структуры** биополимеров (структурная биоинформатика);
- **исследование стратегий, соответствующих** вычислительных методологий, а также общее управление информационной сложностью биологических систем.

Молекула

1868 г. - Фридрих Мишер открыл ДНК



1953 г. – структура ДНК расшифрована Francis Crick, James Watson, Maurice Wilkins and Rosalind Franklin.



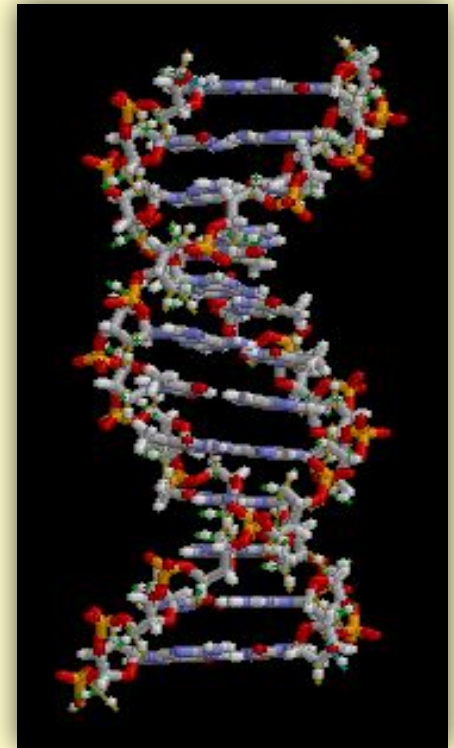
Francis Harry
Compton Crick



James Dewey
Watson



Maurice Hugh
Frederick Wilkins



Молекула ДНК обеспечивает хранение и передачу информации о развитии и функционировании живых организмов из поколения в поколение.

Молекула ДНК

Структура:

ДНК состоит из повторяющихся единиц – **нуклеотидов**:

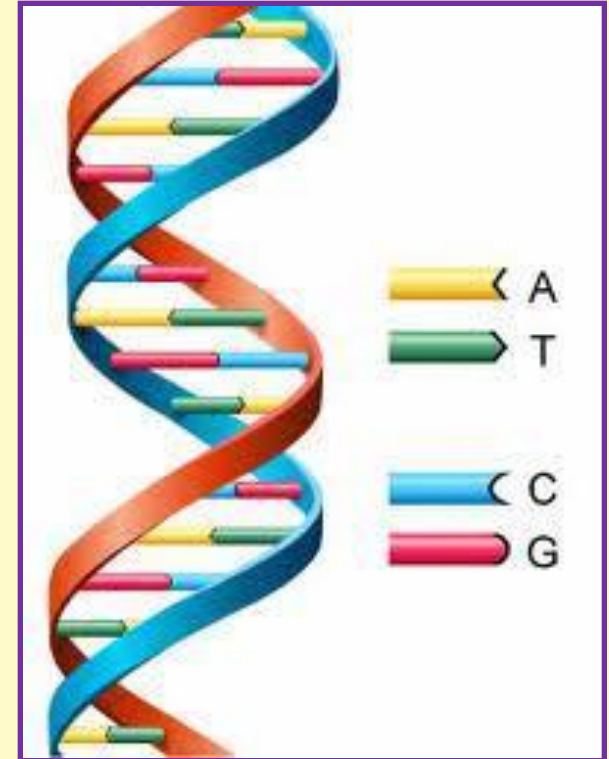
A – Adenine;
T – Thymine;
C – Cytosine;
G – Guanine.

Молекула ДНК состоит из двух цепей, которые соединены через нуклеотиды по принципу комплиментарности:

A – T and C – G.

Пурины – аденин, гуанин.

Пиримидины – тимин, цитозин.



✓ При воспроизведении генетической информации возникают ошибки.



- ✓ Преимущество белого медведя перед бурым при охоте «в снегу».



✓ Один медведь не эволюционирует,
эволюционирует **ПОПУЛЯЦИЯ!!!**



Кто родственник человека???

(участок ДНК, отвечающий за усвоение сахаров)



Все живые объекты –

результат

молекулярной эволюции

Выравнивание

Выравнивание последовательностей — процедура сопоставления остатков в последовательностях, направленная на поиск соответствия между ними.

- Выравнивание последовательностей

```
MALWMRLLPLLALLALWGPDPAAAFVNQHLGSHLVEALYLVCGERGFFYTPKTRREAED
LQVGQVELGGGPGAGSLQPLALEGSLQKRGIVEQCCTSIQSLYQLENYCN


MALWTRLRPLLALLALWPPPARAFVNQHLGSHLVEALYLVCGERGFFYTPKARREVEG
PQVGALELAGGPGAGGL-----EGPPQKRGIVEQCASVCSLYQLENYCN

1 MALWMRLLPLLALLALWGPDPAAAFVNQHLGSHLVEALYLVCGERGFFYTPKTRREAED 60 P01308 INS_HUMAN
1 MALWTRLRPLLALLALWPPPARAFVNQHLGSHLVEALYLVCGERGFFYTPKARREVEG 60 P01317 INS_BOVIN
*** ** ***** * ** *****

61 LQVGQVELGGGPGAGSLQPLALEGSLQKRGIVEQCCTSIQSLYQLENYCN 110 P01308 INS_HUMAN
61 PQVGALELAGGPGAGGL-----EGPPQKRGIVEQCASVCSLYQLENYCN 105 P01317 INS_BOVIN
*** ** ***** * ** *****

1 MALWMRLLPLLALLALWGPDPAAAFVNQHLGSHLVEALYLVCGERGFFYTPKTRREAED 60 P01308 INS_HUMAN
1 MALWTRLRPLLALLALWPPPARAFVNQHLGSHLVEALYLVCGERGFFYTPKARREVEG 60 P01317 INS_BOVIN
*** ** ***** * ** *****

61 LQVGQVELGGGPGAGSLQPLALEGSLQKRGIVEQCCTSIQSLYQLENYCN 110 P01308 INS_HUMAN
61 PQVGALELAGGPGAGGL-----EGPPQKRGIVEQCASVCSLYQLENYCN 105 P01317 INS_BOVIN
*** ** ***** * ** *****
```



сравнить

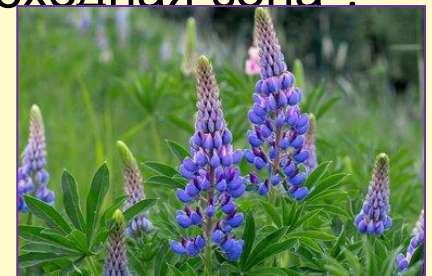
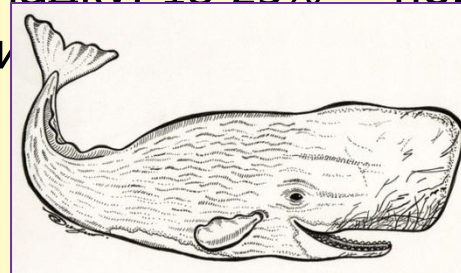
- Считается, что белки с более чем 45% идентичных остатков в выравнивании, это **ОЧЕНЬ** похожие белки, более 25% - имеют сходную пространственную укладку. 18-25% - "переходная зона".

Менее 18% сходства, как и

кашалота и

леггемоглобин люпина

(15%) действительно



Анализ последовательностей

с использованием

1. Задание последовательности.
компьютера

2. Статистический анализ.

3. Выравнивание последовательностей.

Задание

последовательностей

randseq

Seq=randseq(SeqLength)

Seq=randseq(SeqLength, ...'Alphabet', AlphabetValue,...)

```
randseq(20)
```

```
ans =
```

```
TAGCTGGCCAAGCGAGCTTG
```

Задание

последовательностей

Преобразование аминокислотной последовательности в нуклеотидную.

```
nucleo_seq= aa2nt(amino_seq, 'GeneticCode', code_number,  
... 'alphabet', 'alphabet_name'),
```

Пример:

```
>> nucl_seq1=aa2nt('LifeSciense','alphabet','rna')
```

```
nucl_seq1 =
```

```
CUCAUUUUUGAGAGUUGCAUUGAGAAUUCUGAA
```

Статистический анализ **aaccount**

```
amino_s=aaccount(amino_seq, 'chart', chart_value),
```

```
>> seq=randseq(25,'alphabet','amino')
```

```
>> amino_acid=aaccount(seq)
```

```
seq =
```

```
RCYDTLVRHNVASTWRGQTHYDQNN
```

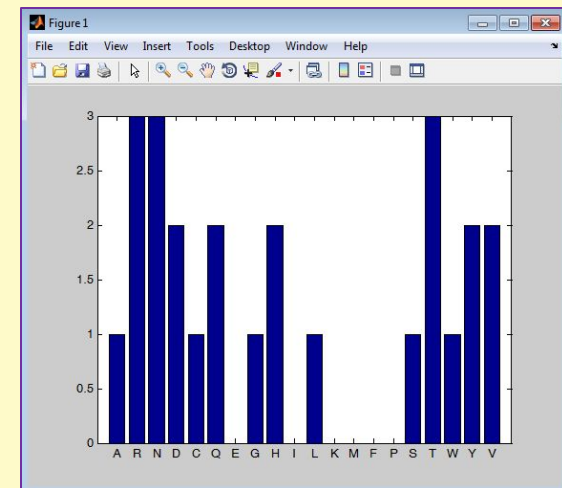
```
amino_acid =
```

```
A: 1
```

```
R: 3
```

```
N: 3
```

```
D: 2
```



Статистический анализ **ntdensity**

```
nucl_dens=ntdensity(nucleotide_seq),
```

Пример:

```
>> S=randseq(20)
```

```
S =
```

```
TTATGACGTTATTCTACTTT
```

```
>> nucl_dens =ntdensity(S)
```

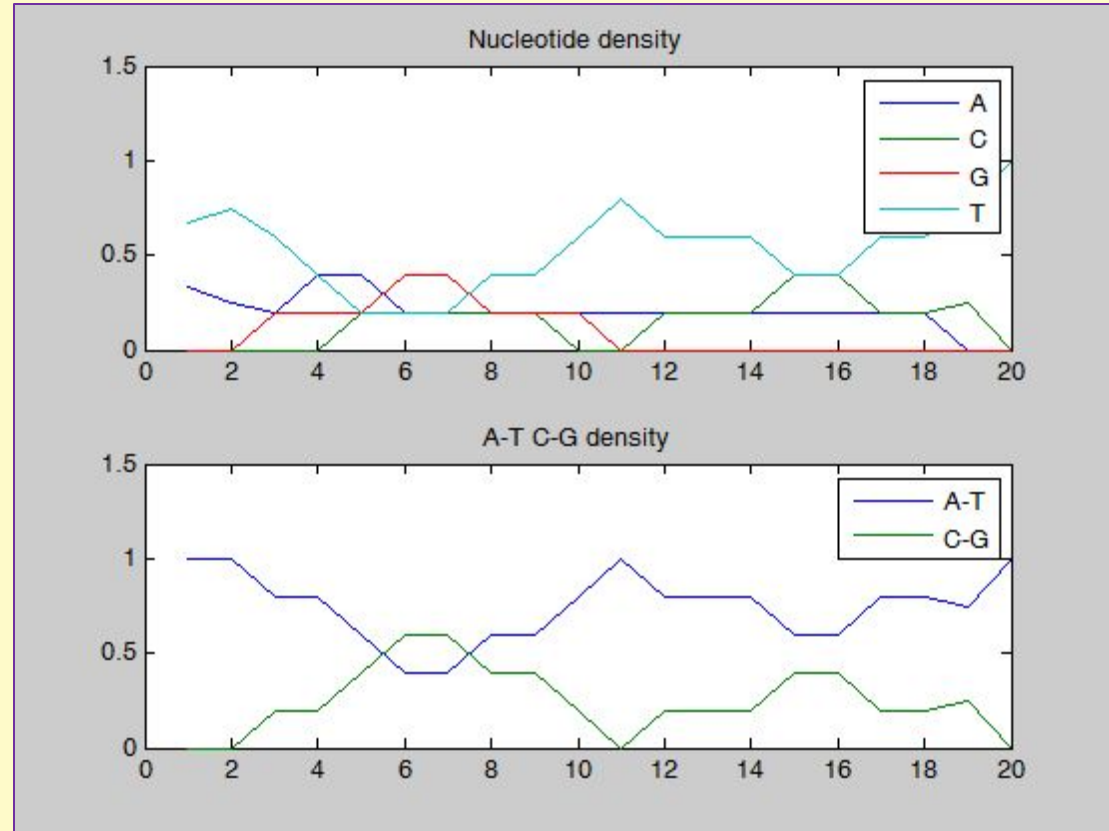
```
nucl_dens =
```

```
A: [1x20 double]
```

```
C: [1x20 double]
```

```
G: [1x20 double]
```

```
T: [1x20 double]
```



Выравнивание

ие

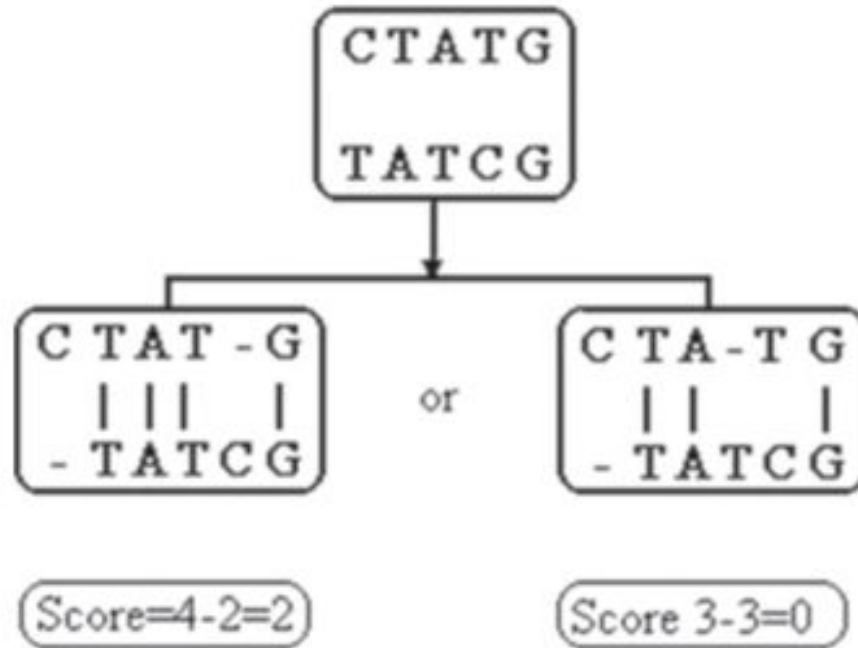


Рисунок 12.3. Пример определения качества выравнивания

Способы выравнивания

1. Глобальное выравнивание
(последовательности одинаковой
длины)
2. Локальное выравнивание

Точечная матрица сходства

Точечная матрица сходства – это матрица, в которой строки соответствуют одной последовательности, а столбцы другой. В простейшем варианте на пересечении совпадающих элементов выставляют единицы, а в остальных ячейках нули.

	A	A	G	C	T	C	T	C	A	T	G
A	1	2							1		
G			1								1
C				1		1		1			
T					1		1			1	
C				1		1		1			
A	1	1							1		
C				1		1		1			
C				1		1		2			
A	1	1							1		
T					1		1			1	

Позволяет

- установить участки локального совпадения.
- установить наличие повторов и палиндромов.
- сопоставить по крайней мере фрагменты последовательностей.

AAGCTCT - CATG

- AGCTCACCAT -

AAGCTCTCATG - -

- AG - - CTCACCAT

Выравнивание Глобальное

Выравнивание *Алгоритм Нидлмана-Вунша*

```
[score, alignment]=nwalign(seq_1, seq_2, 'showscore', true)
```

```
s_1='lifescience';  
s_2='sefesici';
```

```
[score, alignment]=nwalign(s_1, s_2, 'showscore', true)
```

Результат:

score =

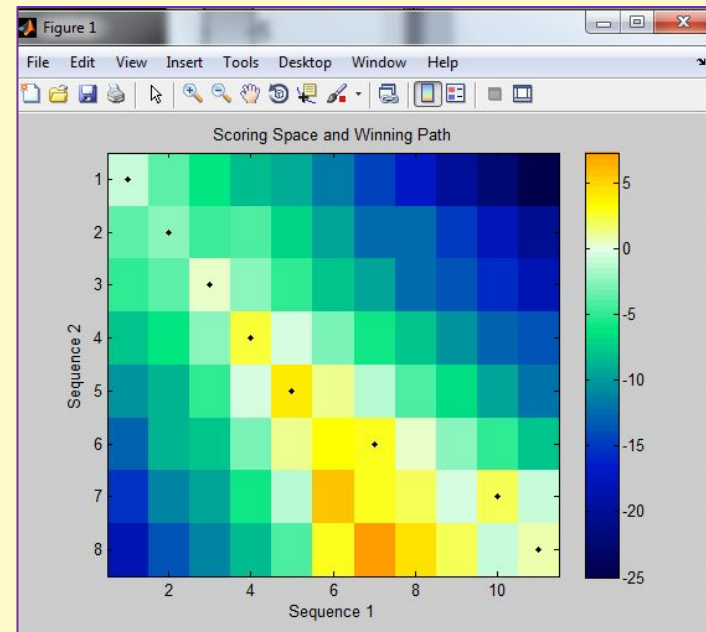
0.6667

alignment =

L I F E S C I E N C E

| | | | |

S E F E S - I - - C I



Выравнивание Множественное выравнивание

Более двух последовательностей

```
aligned_seqs = multialign(sequence_set),
```

Пример:

```
>> seqs = strvcat ('CGTTAT', 'TCGTTAC', 'TAGTTGTGC',  
'GAGTTAATG');
```

```
>> ma = multialign(seqs)
```

```
ma =
```

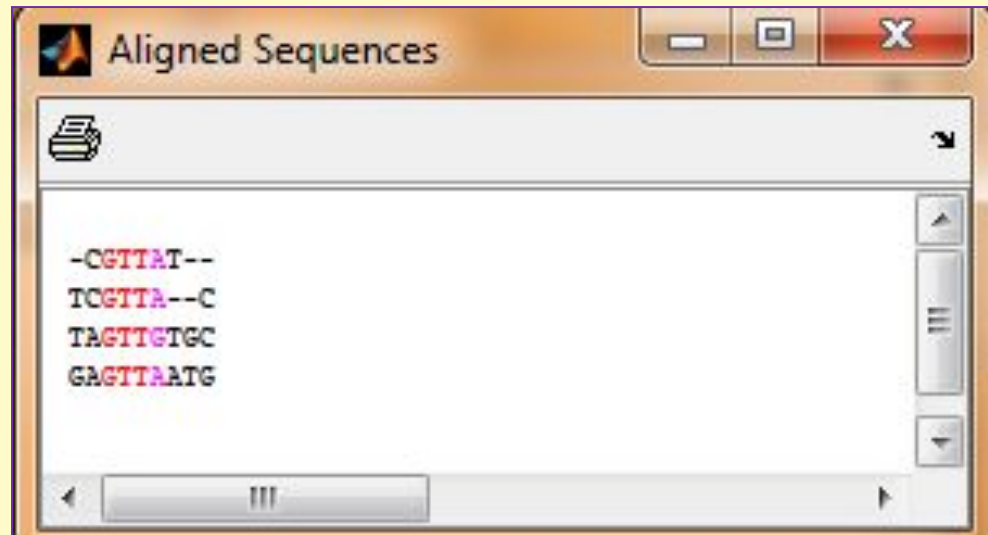
```
CGTTAT - -  
TCGTTA - - C  
TAGTTGTGC  
GAGTTAATG
```


Выравнивание Множественное выравнивание

Более двух последовательностей
`aligned_seqs = multialign(sequence_set),`

Пример:

```
>> showalignment(ma)
```



Выравнивание Множественное

Для длинных последовательностей (длиннее 64 символов) нужно использовать другую форму команды **showalignment**:

```
showalignment(Alignment, 'Columns', Columns_Value)
```

где Columns_Value свойства «Columns» является максимальным целым числом символов, отображаемых в одной строке; например, `showalignment(Alignment, 'Columns', 64)` отображает 64 буквы в одной строке.

Выравнивание Множественное

выравнивание
multialignviewer (Alignment)

The screenshot displays the Multiple Sequence Alignment Viewer (ma) software interface. The window title is "Multiple Sequence Alignment Viewer - ma". The menu bar includes "File", "Edit", "Display", "Window", and "Help". The toolbar contains various icons for navigation and editing. The main display area shows a multiple sequence alignment of four DNA sequences. The consensus sequence is "T A G T T A T - C". The four sequences are:

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

The alignment is visualized with colored blocks representing matches and gaps. A secondary visualization at the bottom shows a horizontal bar chart of the alignment, with a red box highlighting a specific region from position 1 to 8. The status bar at the bottom indicates "4 sequences", "Sequence 4", "Seq 4", and "Aln 9".

Литерату ра

Обязательная

Бутиловский, А.В. Основные методы молекулярной эволюции: Монография / А.В. Бутиловский, Е.В. Барковский, В.Э. Бутиловский, В.В. Давыдов, Е.А. Черноус, В.В. Хрусталеv; под общ. ред. проф. Е.В. Бирковского. – Минск, 2009. – 210 с.

Лазарев, Ю. Ф. Начала программирования в среде Matlab: Учебное пособие. / Ю. Ф. Лазарев. - Киев: НТУУ "КПИ", 2003. - 424 с.

Дащенко, О. Ф. Matlab в инженерных и научных расчетах: Монография / О. Ф. Дащенко, В. Х. Кириллов, Л. В. Коломиец, В. Ф. Оробей. – Одесса: Астропринт, 2003. – 214 с.

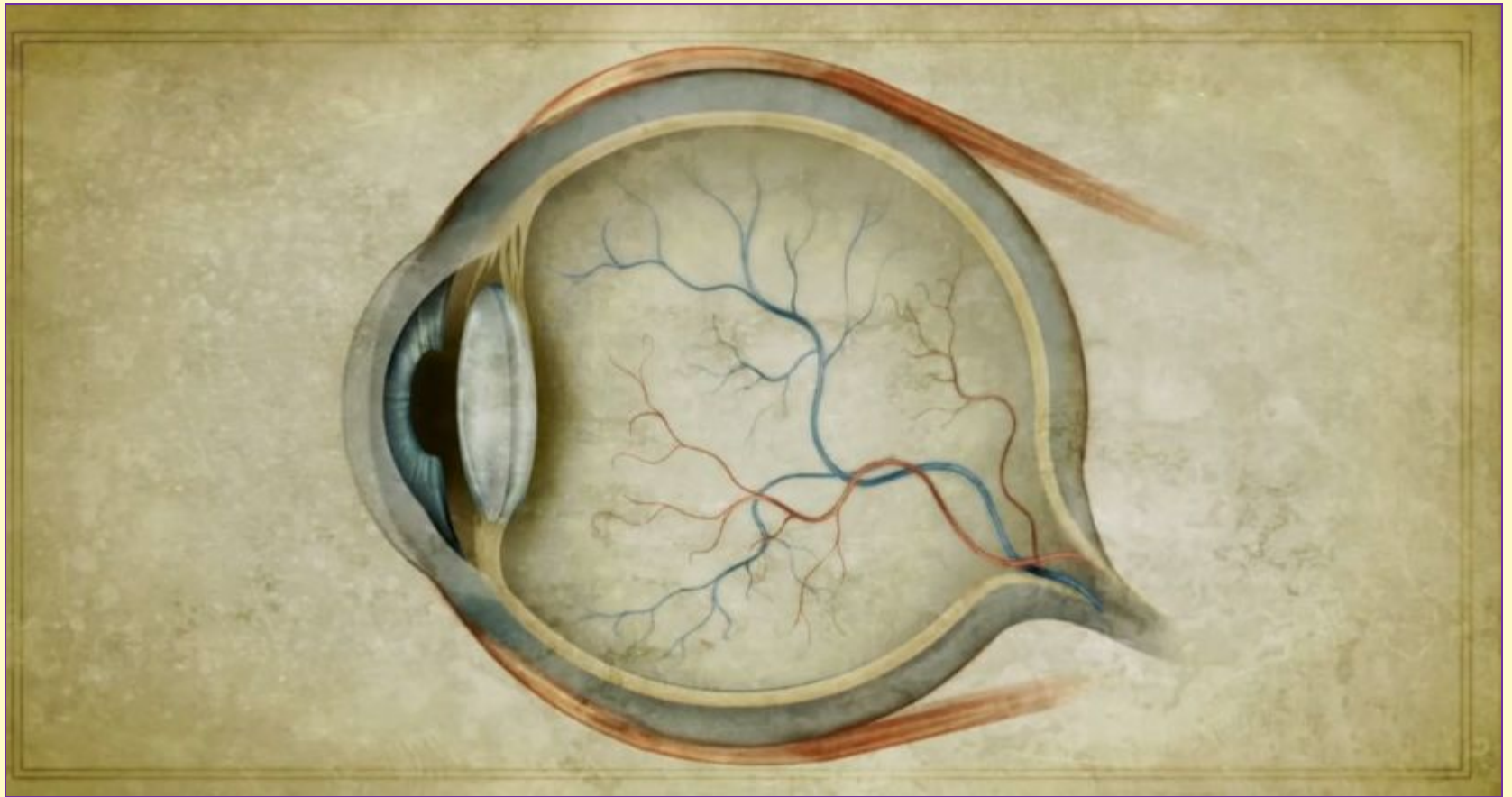
Дополнительная

Burstein, L. MATLAB® in bioscience and biotechnology. / L. Burstein. - Woodhead Publishing, 2011. – 248 p.

Мироновский, Л.А. Введение в Matlab: Учебное пособие / Л.А. Мироновский, К.Ю. Петрова. СПб: ГУАП., 2005. – 122 с.

Дьяконов, В. Matlab. Обработка сигналов и изображений. Специальный справочник. – СПб.: Питер, 2002. – 608 с.

Глаз – продукт слепой эволюции???



«Слепой мир»



Так видит светочувствительная бактерия



Бактерии научились избегать солнечного света



Появление светочувствительного пятна

«Отыскивали» солнечный свет



Плоские черви (чешуйчатое углубление)

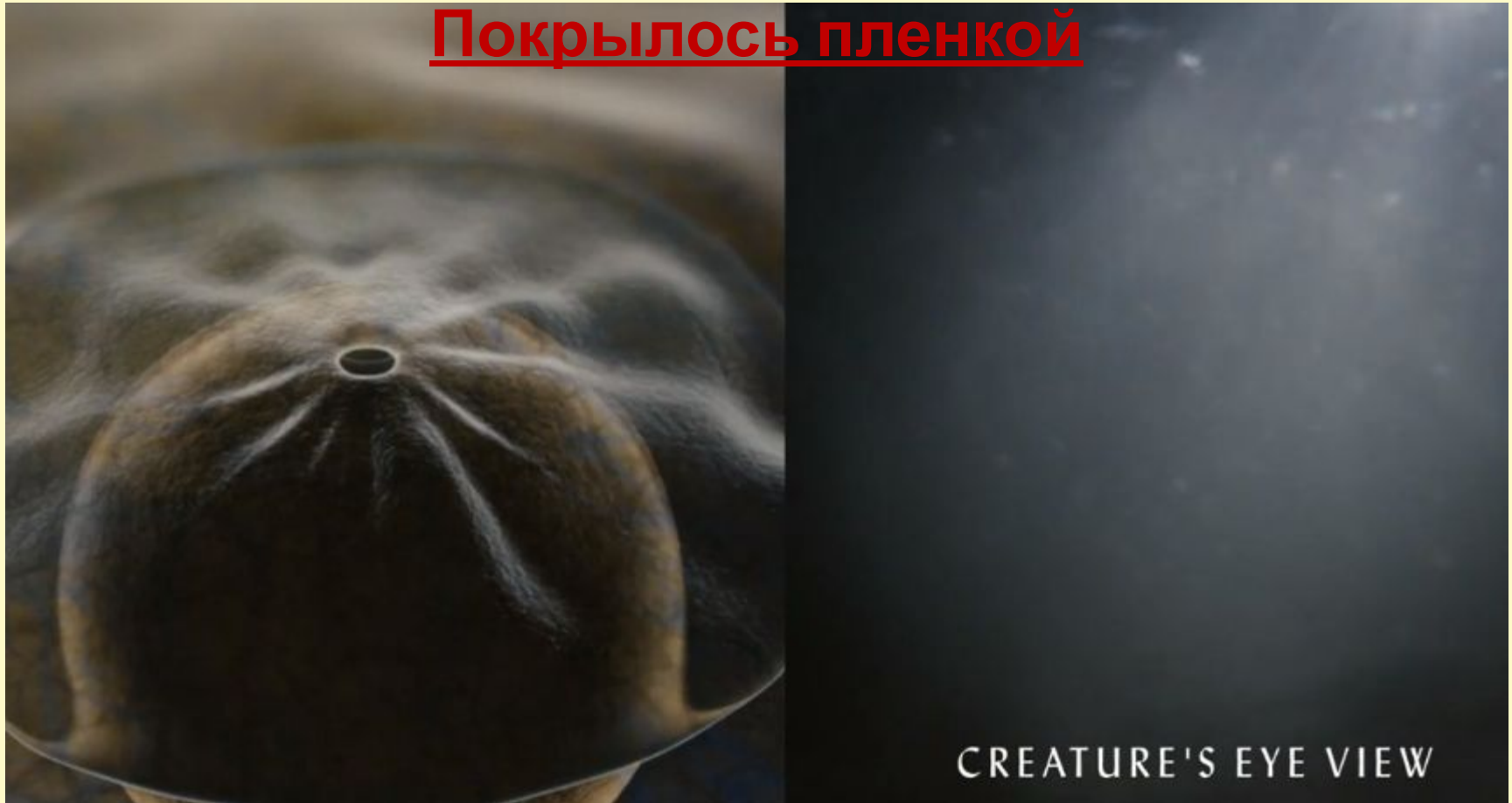
Видны очертания



Появилось углубление



Сужение отверстия до размеров ТОЧКИ



Сужение отверстия до размеров точки

Узкое отверстие пропускает меньше света, но изображение более четкое



Появление хрусталика

Обеспечивал и яркость и резкость

