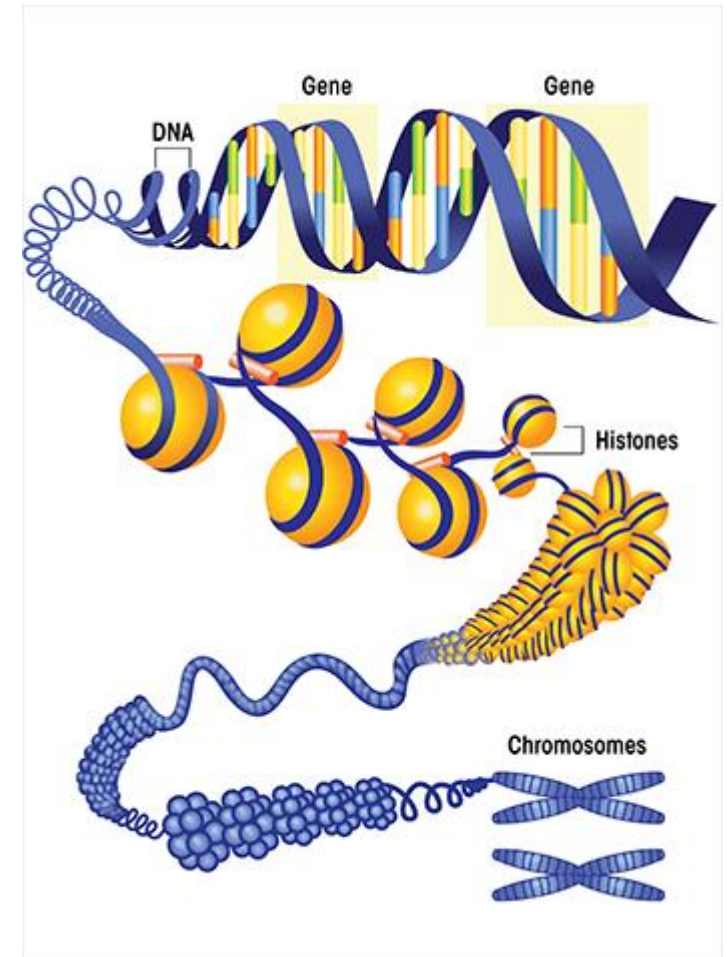
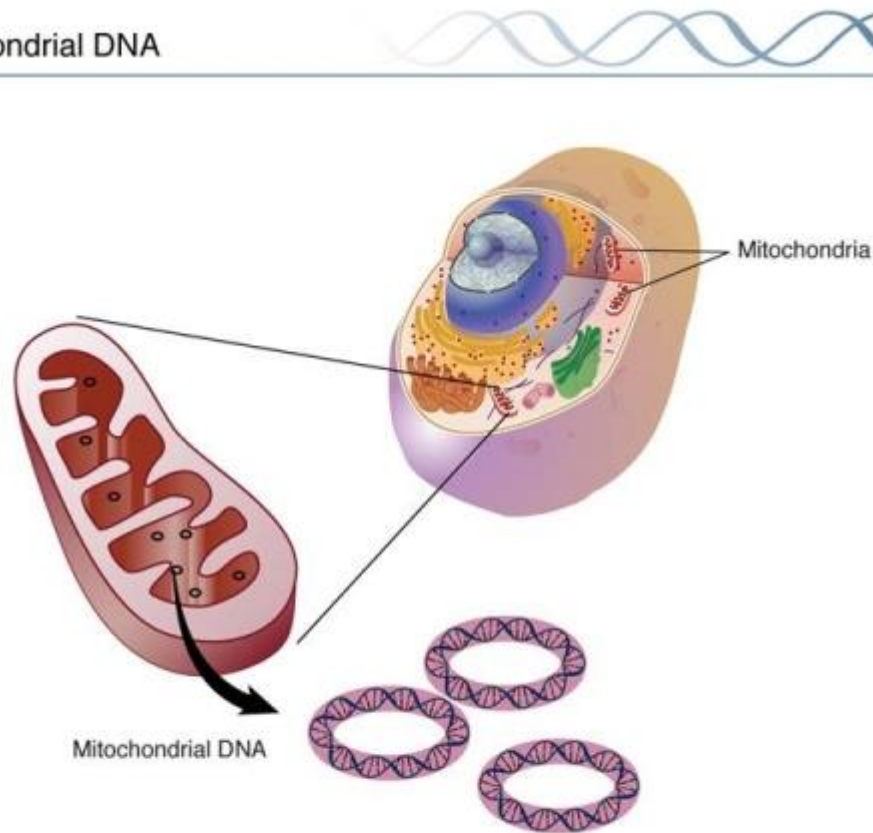


Аннотация ядерной МИТОХОНДРИАЛЬНОЙ ДНК

Mitochondrial DNA



Сравнение митохондриальной и ядерной ДНК

Table 1 | **Comparison between the human nuclear and mitochondrial genomes***

Characteristic	Nuclear genome	Mitochondrial genome
Size	~3.3 x 10 ⁹ bp	16,569 bp
Number of DNA molecules per cell	23 in haploid cells; 46 in diploid cells	Several thousand copies per cell (polyploidy)
Number of genes encoded	~20,000–30,000	37 (13 polypeptides, 22 tRNAs and 2 rRNAs)
Gene density	~1 per 40,000 bp	1 per 450 bp
Introns	Frequently found in most genes	Absent
Percentage of coding DNA	~3%	~93%
Codon usage	The universal genetic code	AUA codes for methionine; TGA codes for tryptophan; AGA and AGG specify stop codons
Associated proteins	Nucleosome-associated histone proteins and non-histone proteins	No histones; but associated with several proteins (for example, TFAM) that form nucleoids
Mode of inheritance	Mendelian inheritance for autosomes and the X chromosome; paternal inheritance for the Y chromosome	Exclusively maternal
Replication	Strand-coupled mechanism that uses DNA polymerases α and δ	Strand-coupled and strand-displacement models; only uses DNA polymerase γ
Transcription	Most genes are transcribed individually	All genes on both strands are transcribed as large polycistrons
Recombination	Each pair of homologues recombines during the prophase of meiosis	There is evidence that recombination occurs at a cellular level but little evidence that it occurs at a population level

*Table modified from REF. 171 © (1999) John Wiley and Sons. TFAM, mitochondrial transcription factor A; rRNA, ribosomal RNA.

Задачи:

- Найти митохондриальную ДНК, встроенную в ДНК ядра (numt) организмов *Candida albicans*, *Arabidopsis thaliana*, *Apis melliformes*
- Найти гены, встроившиеся в полном размере
- Определить функциональную активность встроенных генов

Объекты исследования



Arabidópsis thaliána

Размер генома: 30,43 Mb

Размер митохондрина: 366924 bp



Ápis melliféra

Размер генома: 453 Mb

Размер митохондрина: 16343 bp

Использованные сторонние данные и софт

- Геномные и хромосомные последовательности были скачаны с сервера NCBI, ftp был не доступен
Файлы с хромосомами конкатенировались поорганизменно
- Последовательности скачивались в формате fasta (ядерная ДНК) или в формате genbank (для митохондриальной ДНК)

```
>Unknown sequence #1  
ACTACGCTATCAATATACTCCCACAAATATCAAGAGCCTTCCCAGTATTA AATTGCTA  
AATTCAATACGAACTTCACACTCCACAGCCTCACGCGAAATTAATAAATCGTATTTAAAT  
ATACCATGAACSTATCGTTTAGTACATGAATTTACACACGTCAGCCOGATCAAATGTTTAT  
CATTATATATGTACATTTTCAGTTTGTGTATATAGACATAACATTAATGTAATAAAGACAT  
TAGTACATTAATTGATTGTCCTCAAGCATATAAGCAAGTACTAGACATTCACTAGCGGTA
```

Использованный сторонний софт

- BLAST+
- Igv- browser

Поиск последовательностей numtDNA

- Локальный бласт с параметрами:

```
makeblastdb.exe -in A.thaliana_all_chr.fasta -dbtype nucl
```

```
blastn -task blastn -db A.thaliana_all_chr.fasta -query  
A.thaliana_mitoch.fasta -outfmt 7 -evaluate 1e-06 > a.out
```

- Numt1.py - скрипт на Python с использованием Biopython (на входе – blast_output и mit_dna.gb) – преобразование вывода blast и поиск numt
- Numt2.py - скрипт на Python для поиска функциональных генов в numt
- Анализ данных в виде таблицы Excel

Blast-output fragment

BLASTN 2.2.28+

Query: gi|26556996|ref|NC_001284.2| Arabidopsis thaliana mitochondrion, complete genome

Database: A.thaliana_all_chr.fasta

Fields: query id, subject id, % identity, alignment length, mismatches, gap opens, q. start, q. end, s. start, s. end, evalue, bit score

825 hits found

gi 26556996 ref NC_001284.2	gi 240254678 ref NC_003071.7	99.66	74085	95	87	44698	1187363435331	3509305	0.0	1.322e+
gi 26556996 ref NC_001284.2	gi 240254678 ref NC_003071.7	99.67	66106	84	49	2176882837653313244		3247243	0.0	1.181e+
gi 26556996 ref NC_001284.2	gi 240254678 ref NC_003071.7	99.44	44829	63	70	1115711562873367972		3412723	0.0	7.956e+
gi 26556996 ref NC_001284.2	gi 240254678 ref NC_003071.7	99.71	37960	60	39	3056703436073360738		3322806	0.0	6.783e+

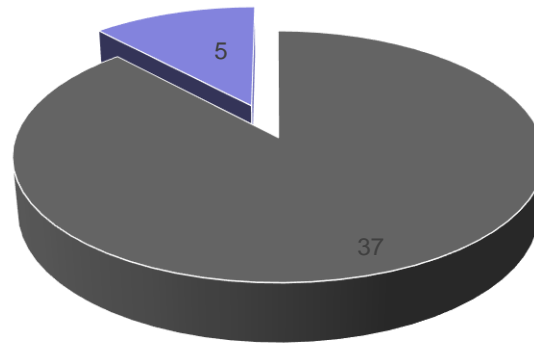
Фрагмент вывода для скрипта

gi|323388985|ref|NC_007072.3|,LG3,9646351,9647650,1337,84.44,ND2gi|323388985|ref|NC_007072.3|,L
G3,12673440,12672786,735,78.64,ND4Lgi|323388985|ref|NC_007072.3|,LG3,11520453,11521003,59
0,80.17,ND3gi|323388985|ref|NC_007072.3|,LG3,13134469,13134749,282,96.10,ATP8gi|323388985|r
ef|NC_007072.3|,LG3,12659723,12660660,988,64.57,ND4Lgi|323388985|ref|NC_007072.3|,LG3,1266
0388,12660656,298,66.44,ATP8gi|323388985|ref|NC_007072.3|,LG3,12659655,12660168,545,64.95,
ATP8gi|323388982|ref|NC_007075.3|,LG6,12559955,12559370,635,84.09,ATP8gi|323388986|ref|NC_
007071.3|,LG2,12060424,12060808,393,89.31,ND4Lgi|323388981|ref|NC_007076.3|,LG7,11973595,1
1972861,822,81.39,ATP6gi|323388981|ref|

Количественные характеристики найденных numt

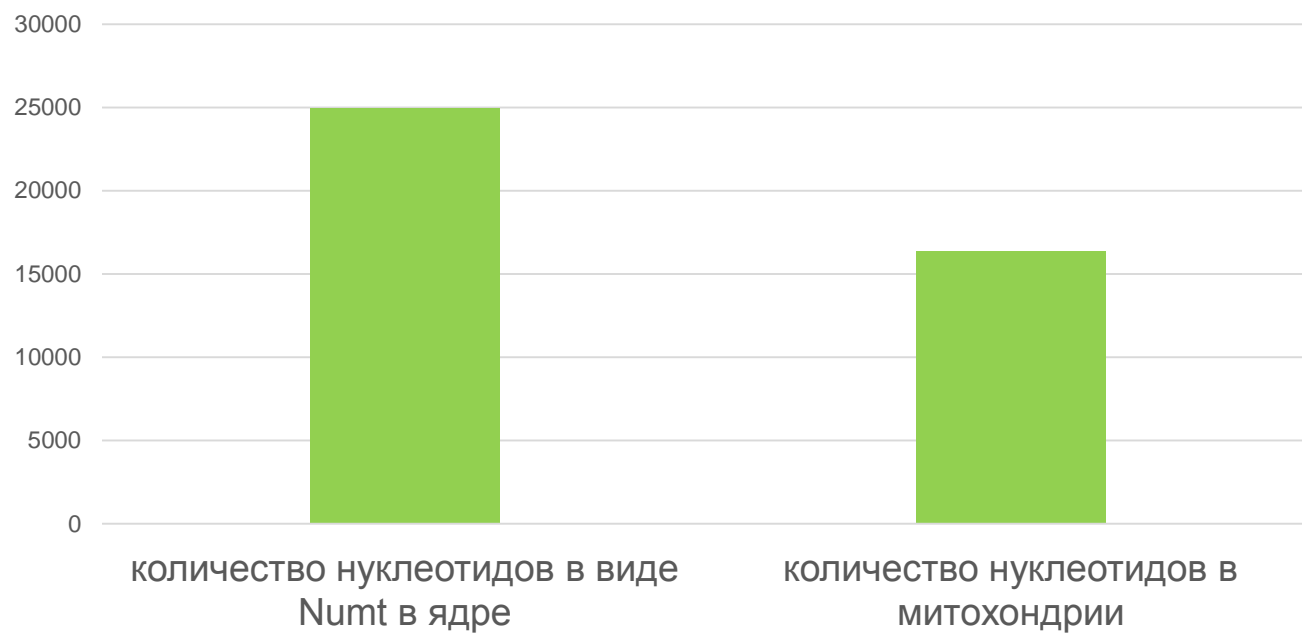
	<i>A.thaliana</i>	<i>A.melifera</i>
Количество хитов, содержащих митохондриальную ДНК = количество numt	2572	6915
Количество numt, содержащих полные гены митохондриального происхождения	123	44
Количество numt, содержащих гены, попавших в открытые рамки считывания	120	40

Отношение перенесенных генов к общему количеству генов в митохондрии *Apis mellifera*



- количество генов в митохондрии
- количество Numt с полным митохондриальным геном в ядерном геноме

Количество нуклеотидов в виде Numt и в митохондриальном геноме *Apis mellifera*



Что это за гены?

Для *A.thaliana* найдено 68 генов:

- большая часть из них не имеют известных белковых продуктов *ArthMp*
- Функционально это: субъединицы АТФазы, tRNA, matR матураза, белки, участвующие в биогенезе цитохрома С.
- Практически все гены встроились несколько раз
- Для *A.melifera* найдено 5 генов
 - ND2, NADH dehydrogenase subunit 2, protein coding
 - ND3, NADH dehydrogenase subunit 3, protein coding
 - ND4L, NADH dehydrogenase subunit 4L, protein coding
 - ATP8, ATP synthase F0 subunit 8, protein coding,
 - ATP6, ATP synthase F0 subunit 6, protein coding

Ген	Количество
ND4L	9
ND2	1
ND3	8
АТР6	2
АТР8	24
ArthMpo67	80
ArthMpo66	48
tRNA-Ser	9
tRNA-Tyr	7
ArthMpo71	6
ArthMpo75	8
ArthMpo21	11