

Bacteriologia para Ciências Biomédicas - BMM0584

Genômica de procariotos

Robson Francisco de Souza
LEEP: Laboratório de Estrutura e Evolução de Proteínas
ICB/USP – 2018

Tópicos

- Conceitos básicos de genômica
 - Diversidade de procariotos
 - O que é um genoma?
 - Estrutura de genomas em procariotos
 - Como sequenciar um genoma?
- Evolução de genomas
 - Perda de gene / redução de genomas
 - Ganho de genes
 - Duplicação
 - Transferência lateral
 - Rearranjos
- Anotação de genomas
- Desdobramentos
 - Metagenômica
 - Biologia de sistemas

Diversidade de procariotos

A árvore da vida

A divisão dos organismos em três domínios ou super-reinos (abaixo) foi proposta por Carl Woese com base na análise do RNA ribossomal 16S e depois corroborada pela análise de proteínas e outros dados

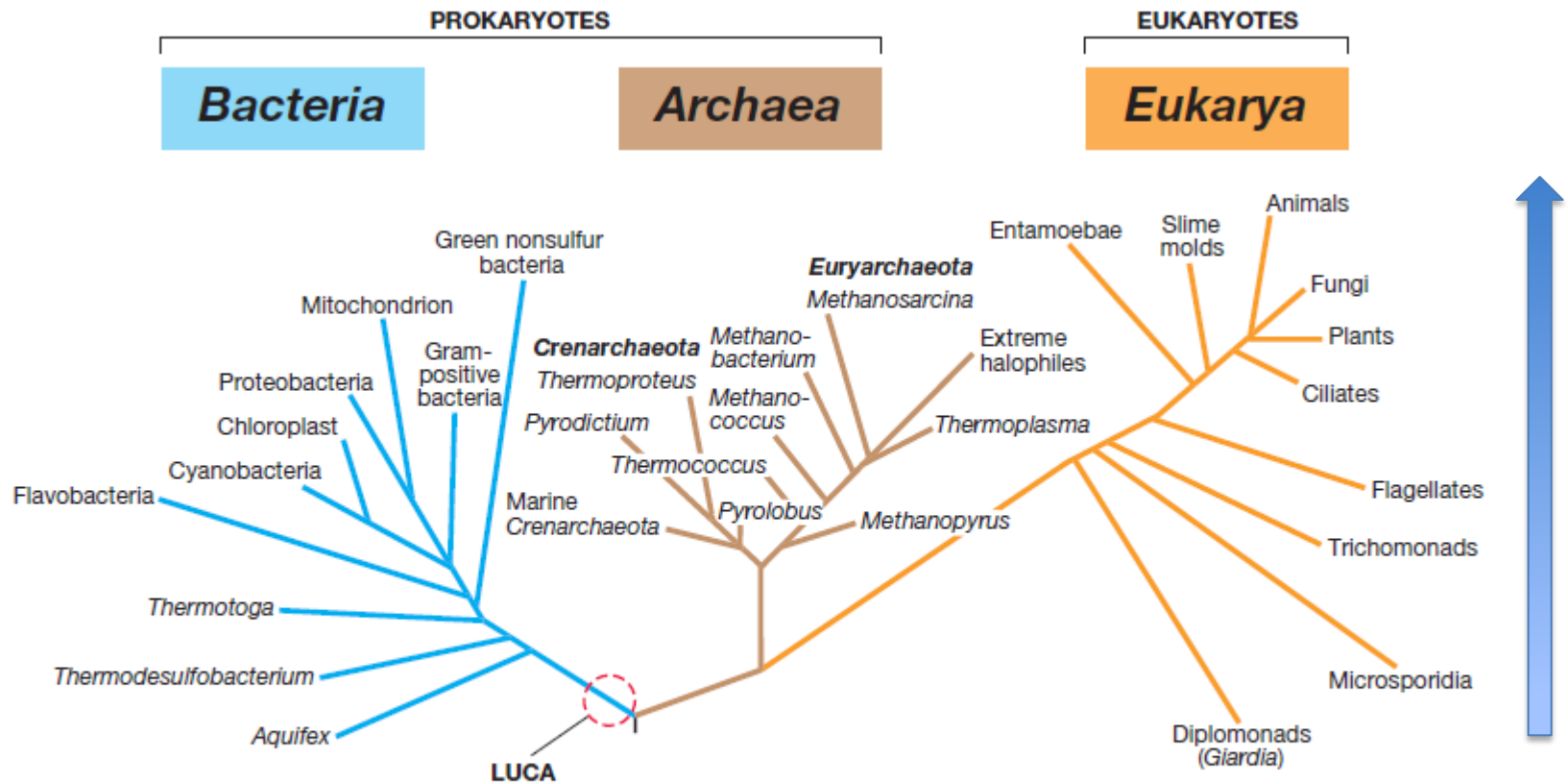


Figure 16.16 Universal phylogenetic tree as determined from comparative SSU rRNA gene sequence analysis. Only a few key organisms or lineages are shown in each domain. At least 80 lineages of *Bacteria* have now been identified although many of these have not yet been cultured. LUCA, last universal common ancestor.

<http://itol.embl.de>
<http://tolweb.org/tree/>

Diversidade de procariotos

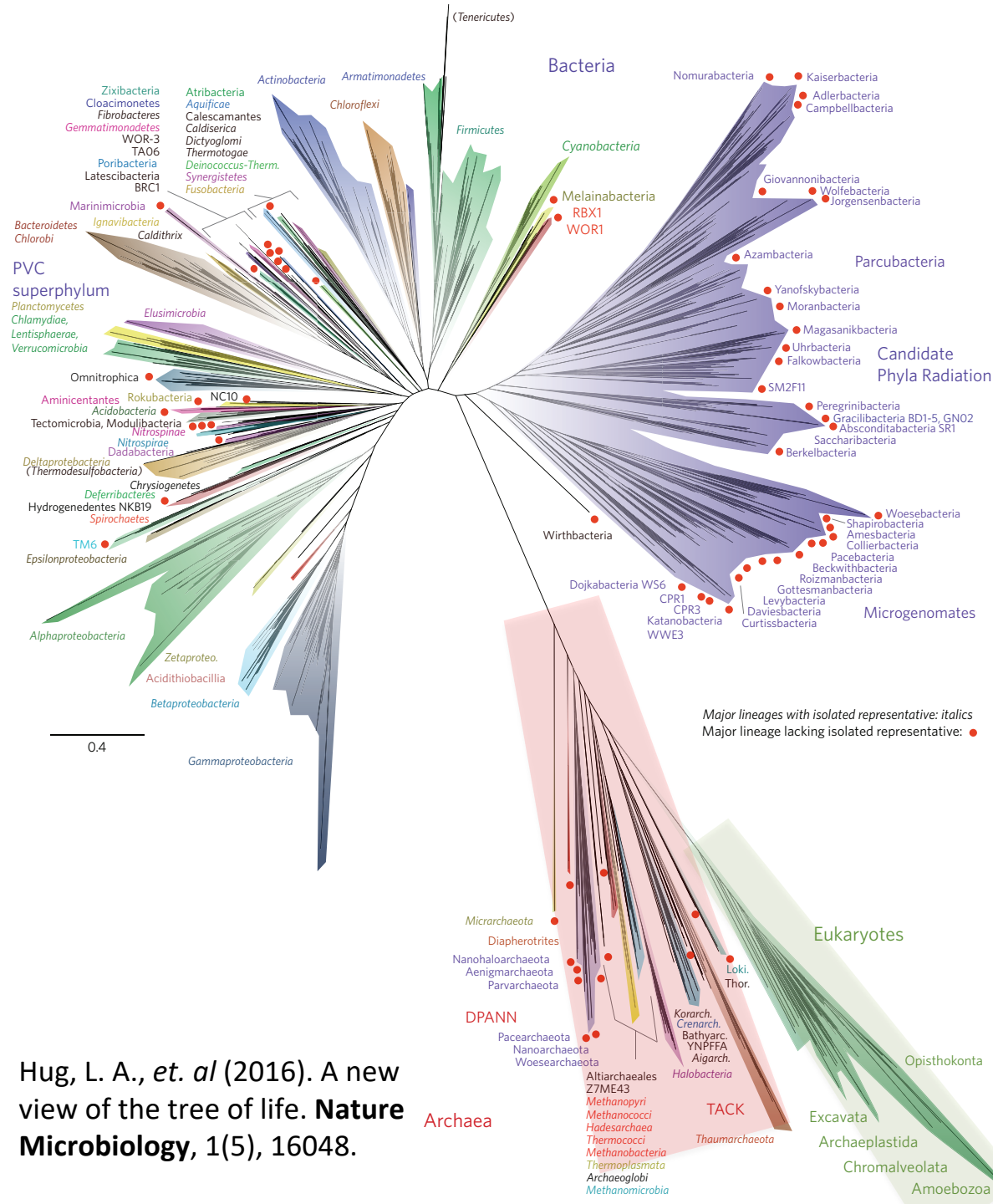
A árvore da vida

Much like the physical sciences community has coordinated to define and understand the universe's dark matter, microbiologists have embarked on a similar voyage using DNA sequencing to discover the hidden diversity and genetic potential of Earth's microbiomes.

As a result, we are rapidly and continually growing new branches on the tree of life

Blaser, M. J., *et al.* (2016). Toward a Predictive Understanding of Earth's Microbiomes to Address 21st Century Challenges. *mBio*, 7(3), e00714–16.

Hug, L. A., *et al.* (2016). A new view of the tree of life. **Nature Microbiology**, 1(5), 16048.



Genoma

Definição

Corresponde a todo o material genético em uma célula/organismo ou vírus, incluindo tanto regiões codificantes como regiões não-codificantes

Estrutura

- Formado por um ou mais **replicons** característicos da espécie, incluindo cromossomos, plasmídeos e outro elementos extra-cromossomais (*e.g.* operon ribossomal das mitocôndrias em *T. cruzi*).
- Os replicons podem ser lineares ou circulares e feitos de DNA ou RNA.

Genomas: histórico

- Primeiro genoma completo (**1974**)
 - Fago Φ X174 (RNA)
- Primeiro genoma de uma bactéria (**1995**)
 - *Haemophilus influenzae*: foi também o primeiro feito com “Whole-genome random sequencing”
- Primeiro genoma de eucarioto (**1996**)
 - *Saccharomyces cerevisiae*
- Primeiro genoma de um animal (**1998**)
 - *Caenorhabditis elegans* Strain: Bristol N2
- **Genoma humano (2001)**
 - Craig Venter vs. International consortium
- 1000 Genomes (**2012**)
- Genomas Online Database
 - Quase 40000 projetos de sequenciamento de genomas em Feb/2014

Genoma: tipos de moléculas

Organismo	Elemento	Ácido nucléico	Descrição
Procarioto	Cromosomo	DNA dupla fita	A maioria é circular, muito longo
Eucarioto	Cromosomo	DNA dupla fita	Maioria linear, extremamente longo
Todos	Plasmídeo*	DNA dupla fita	Relativamente curto, linear ou circular
Mitocondria ou cloroplasto	Genoma	DNA dupla fita	Pequeno ou médio, geralmente circular
Vírus	Genoma	DNA ou RNA, fita dupla ou simples	Relativamente curto, circular ou linear

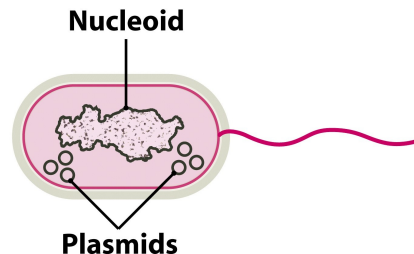
* Plasmídeos são muito raros em eucariotos

Cromossomos

- Codificam genes essenciais para o organismo
- Codificam os genes necessários para replicação e segregação

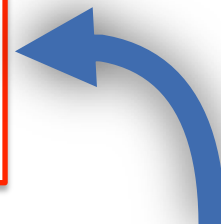
Plasmídeos

- Usam as polimerases do cromosomo
- Controlam seu número na célula
- Codificam genes para segregação





Definição: o conjunto de toda a informação genética de um organismo, incluindo tanto regiões codificantes como regiões não-codificantes



The diagram illustrates the structure of DNA. On the left, a 3D model of the DNA double helix is shown. The two strands are twisted around each other. The 'Base pair' is indicated at the top, showing complementary base pairing (A with T, G with C). The 'Minor groove' and 'Major groove' are labeled on the left side of the helix. The 'Sugar-phosphate backbone' is labeled on the right side of the helix. The 'Hydrogen' bonds between base pairs are indicated at the bottom. On the right, a detailed chemical structure of a DNA segment is shown. It includes the 5' end and 3' end of the strands. The sugar-phosphate backbone is shown with phosphate groups (P) and deoxyribose sugars (C). The nitrogenous bases (A, T, G, C) are shown in the center, connected by hydrogen bonds (indicated by red lines). The 5' end is labeled '5' end' and the 3' end is labeled '3' end'.

Figure 1-46. *Continued* © Garland Science 2007



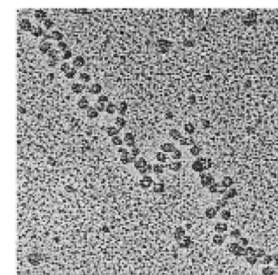
Metaphase chromosome, diameter 1 μm

0.4 μm

30-nm fiber

Nucleosome

B-DNA, diameter 20 Å

*con-*

DNA: procariotos

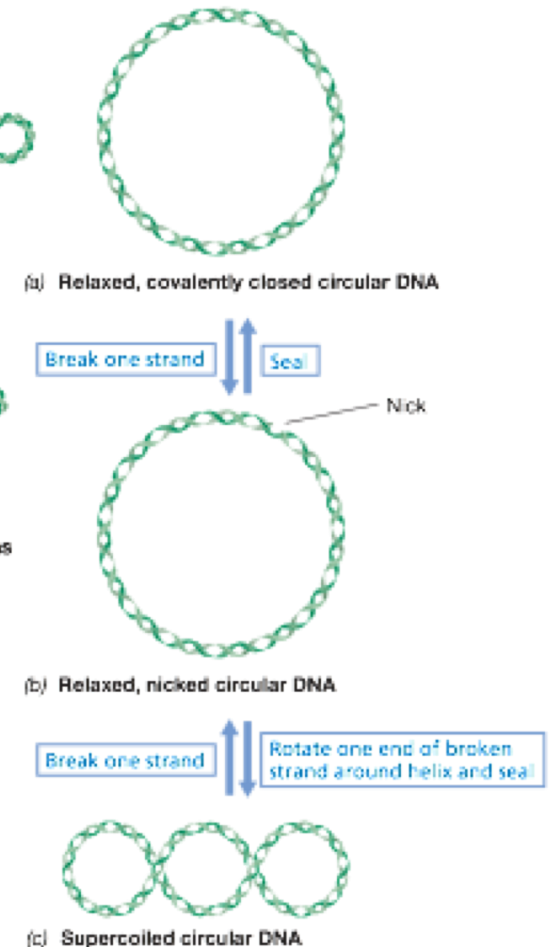
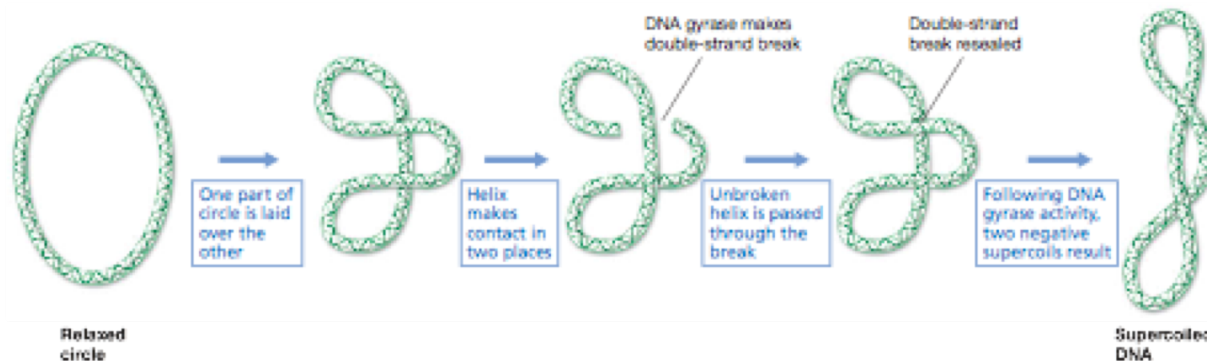
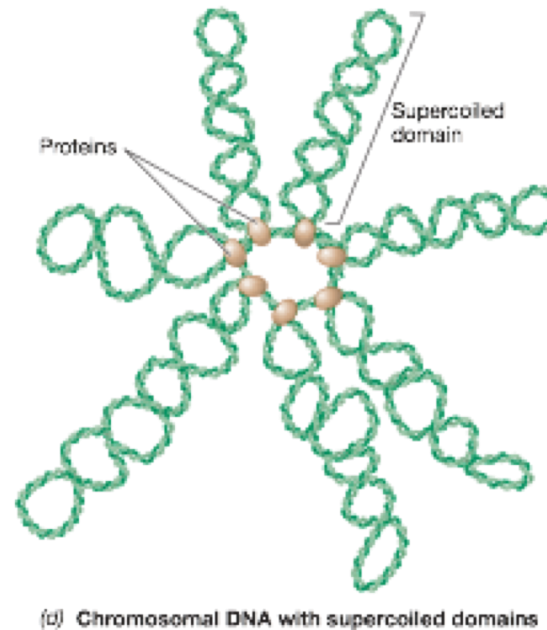
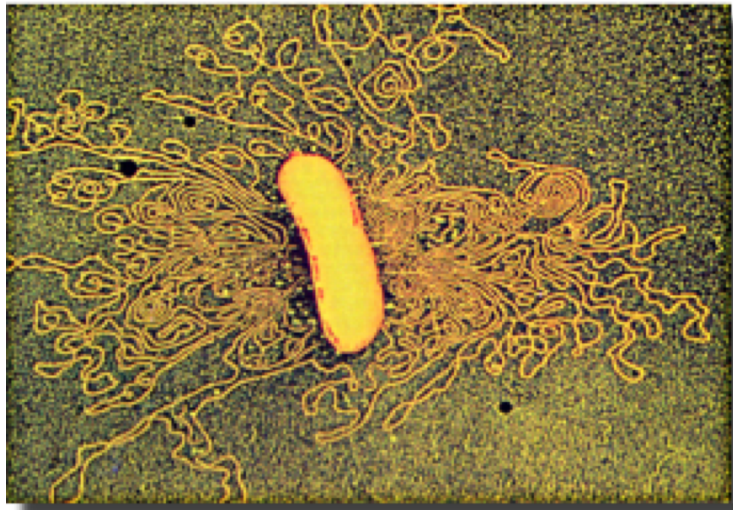
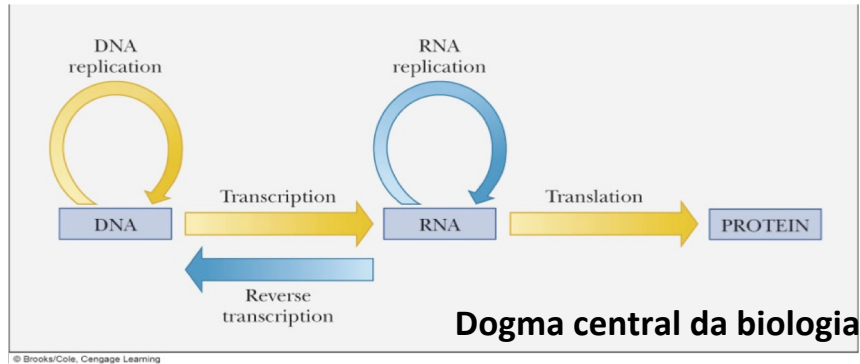


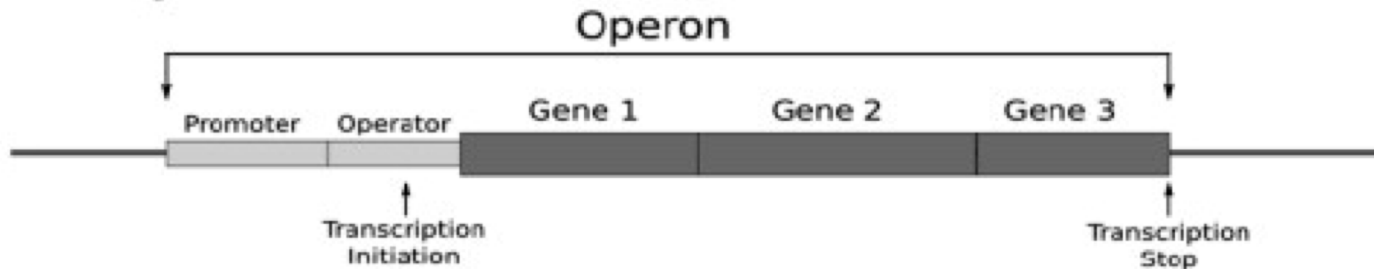
Figure 6.9 DNA gyrase. Introduction of negative supercoiling into circular DNA by the activity of DNA gyrase (topoisomerase II), which makes double-strand breaks.

Genes e genomas



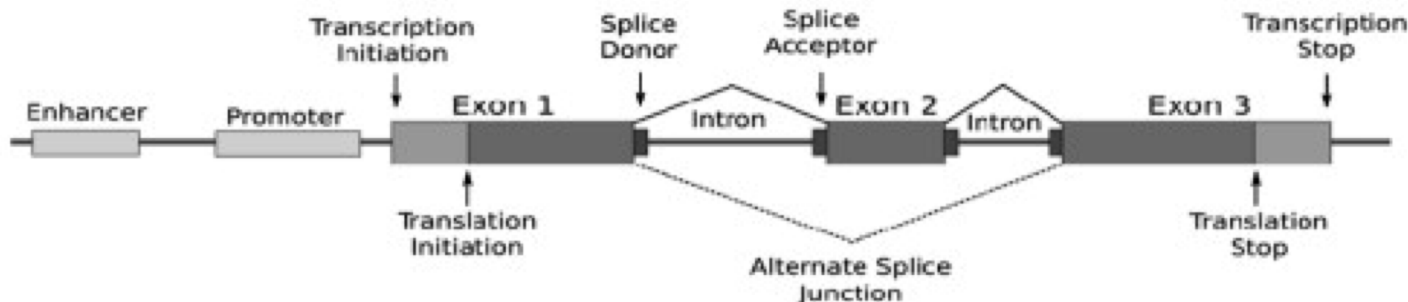
Gene: região de uma sequência genômica que corresponde a uma unidade de herança e contém a informação necessária para produzir uma ou mais proteína e/ou RNA, podendo incluir também as regiões regulatórias e/ou outras sequências que regulam sua atividade

Prokaryotic Genes



ORF == gene

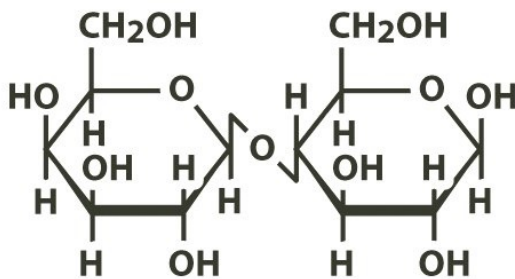
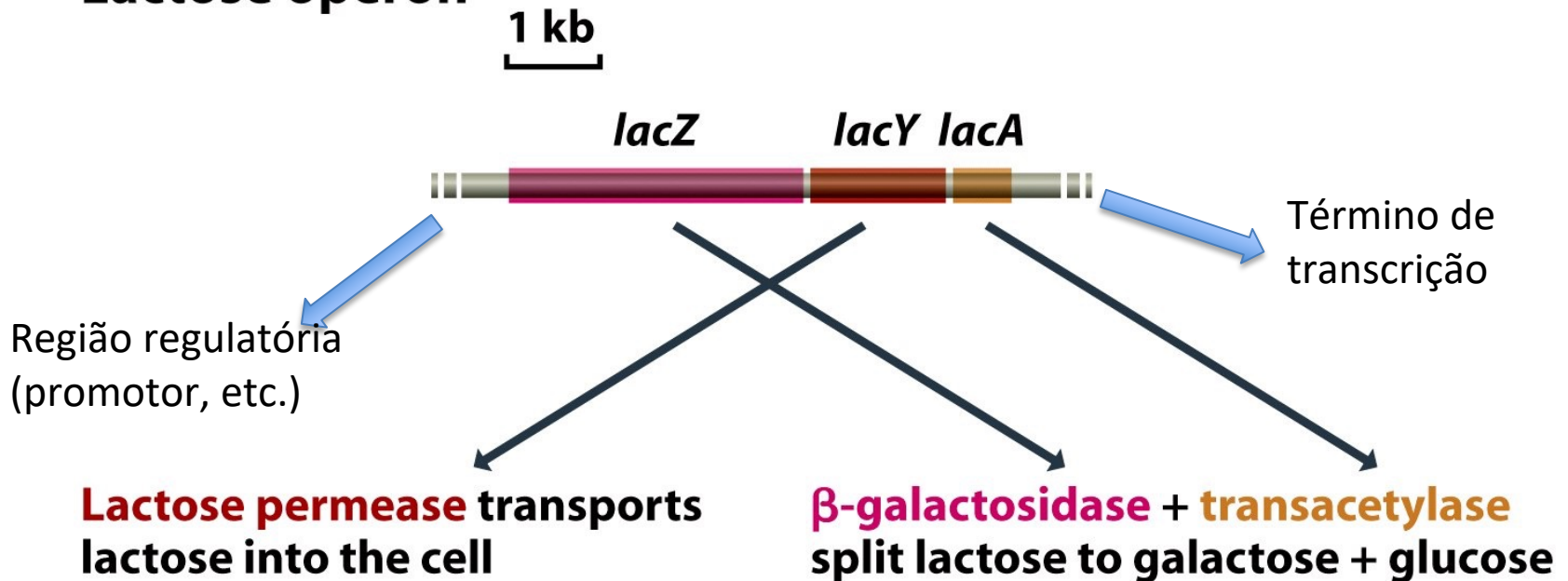
Eukaryotic Gene



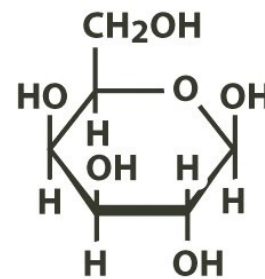
1 ou mais ORFs

Genomas de procariotos: operons

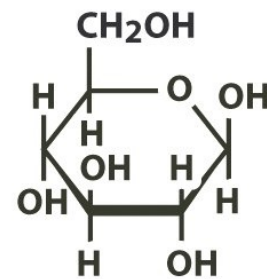
Lactose operon



Lactose



Galactose

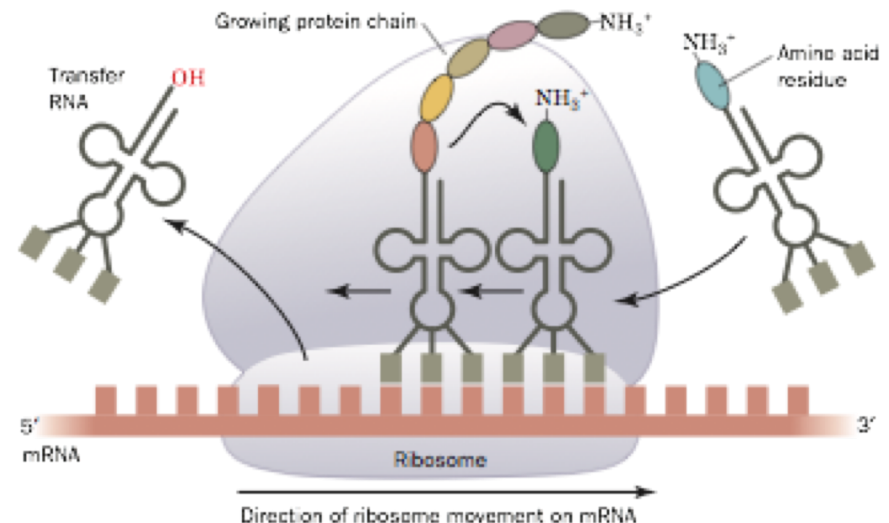
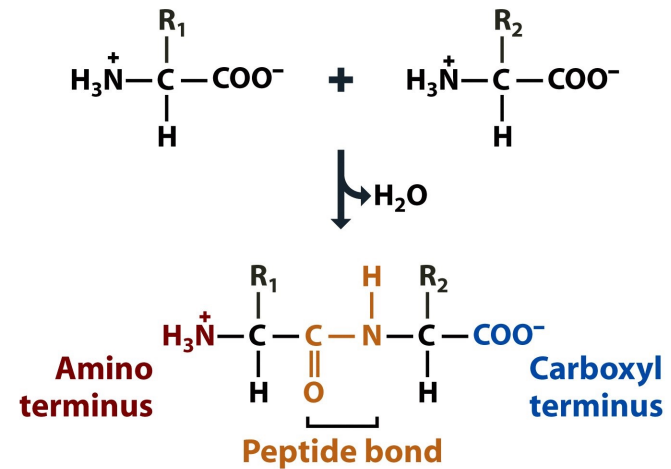
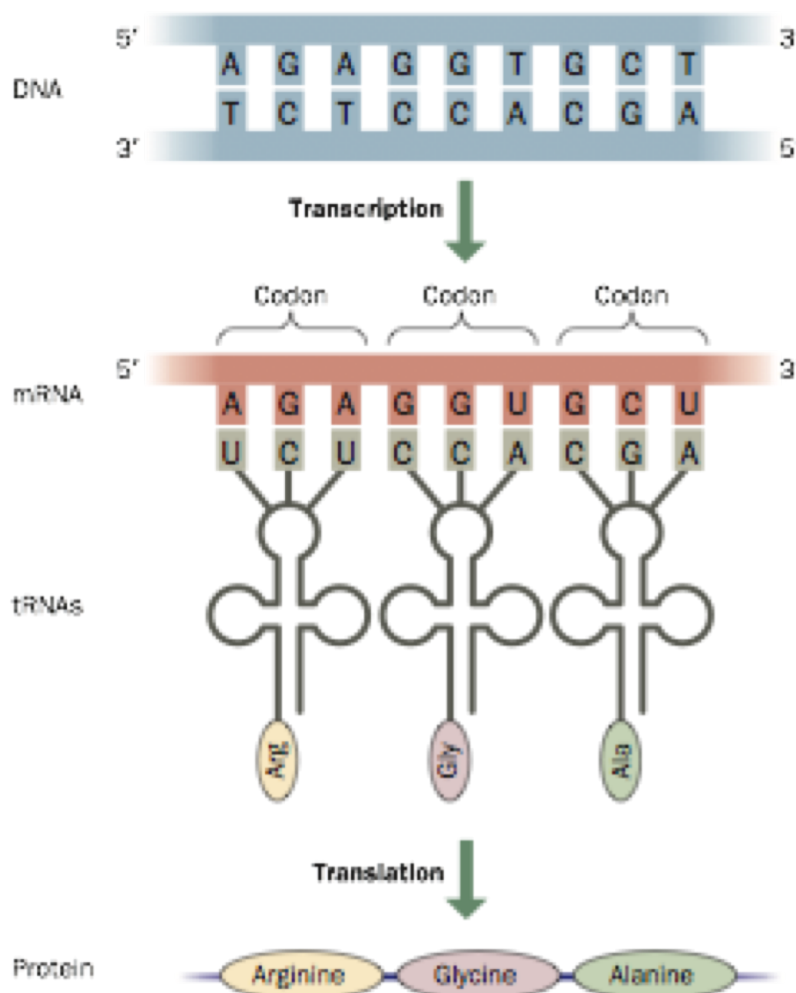


Glucose

Operons

- Genes de uma mesma via metabólica muitas vezes formam operons no genoma de bactérias
- Agregar genes com funções relacionadas em operons permite coordenar a expressão desses genes usando um único fator de transcrição
- Como não têm núcleo, as bactérias executam transcrição e tradução no mesmo compartimento, o que permite aos genes em operons acoplar os processos de transcrição, tradução e formação de complexos, resultando em maior eficiência (hipótese de “genomic channeling”)

Síntese de Proteínas: tradução



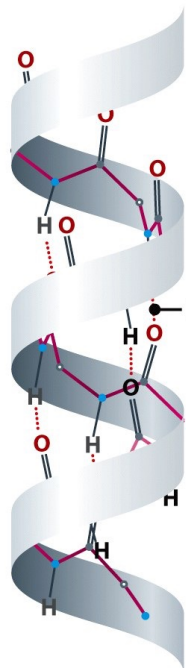
Proteínas: estrutura



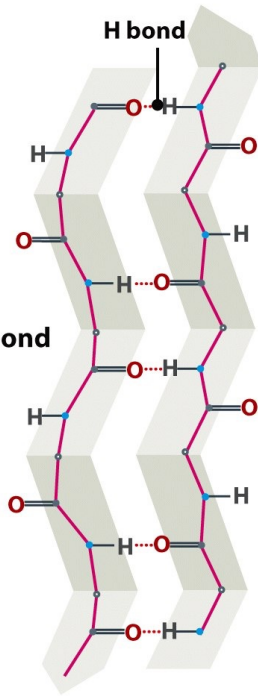
Como os nucleotídeos, proteínas são moléculas informacionais

A diversidade das cadeias laterais dos aminoácidos garante a versatilidade funcional das proteínas

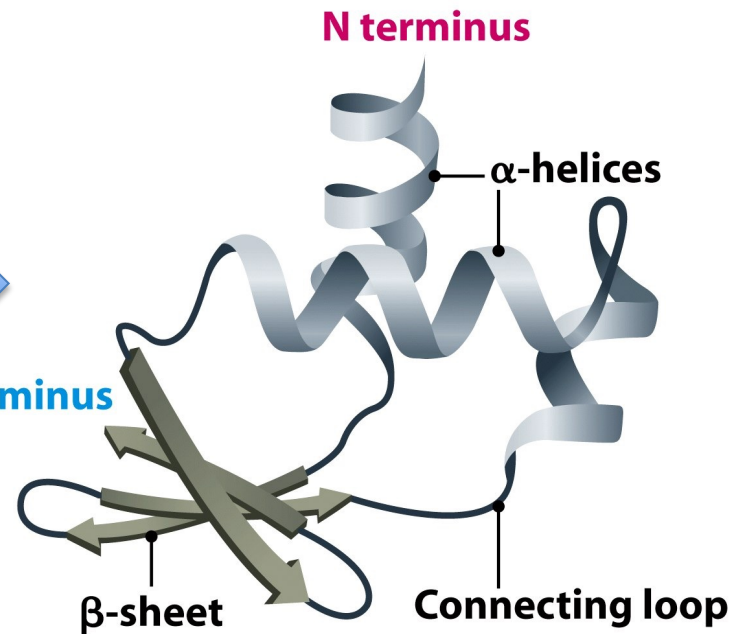
(A) α -helix



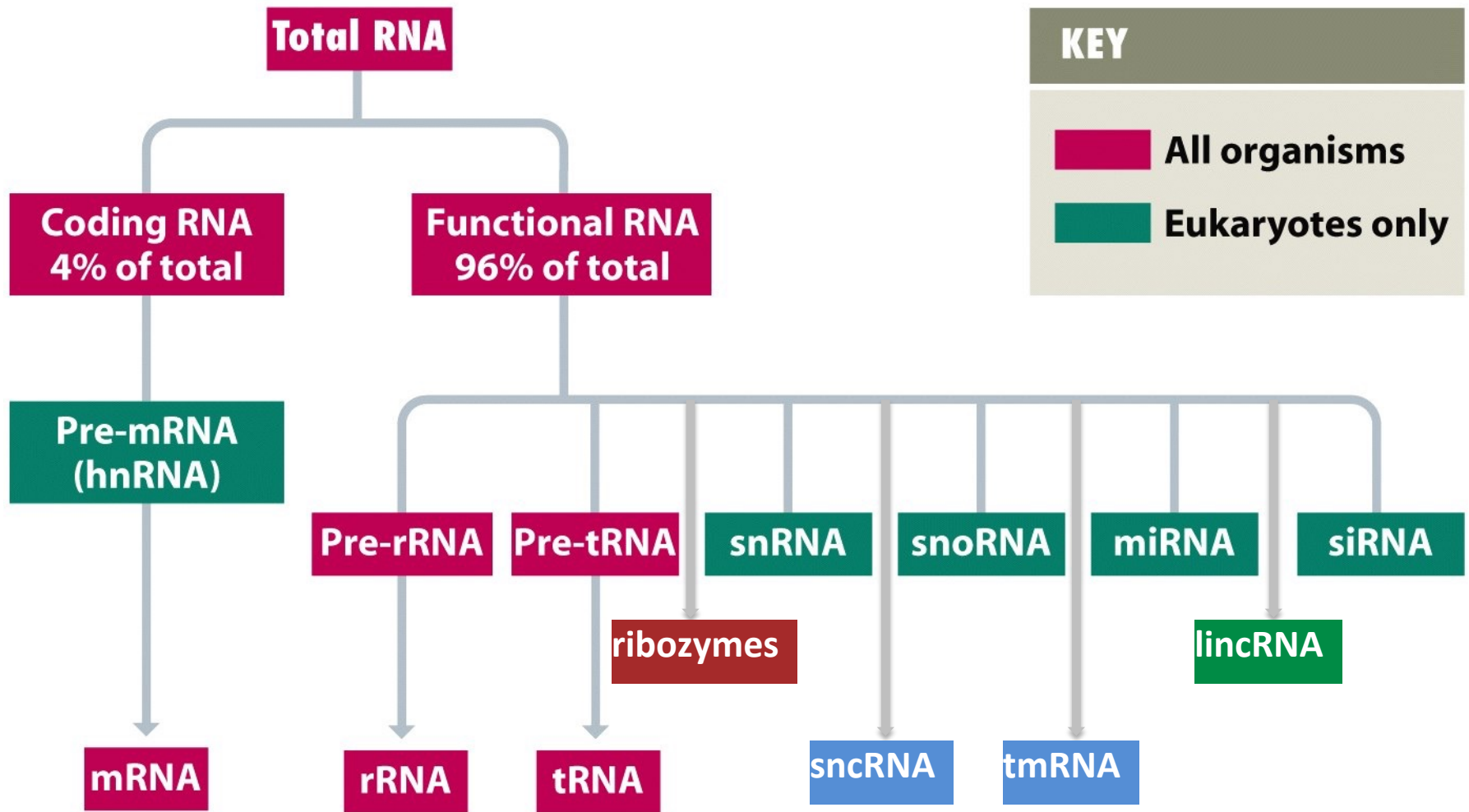
(B) β -sheet



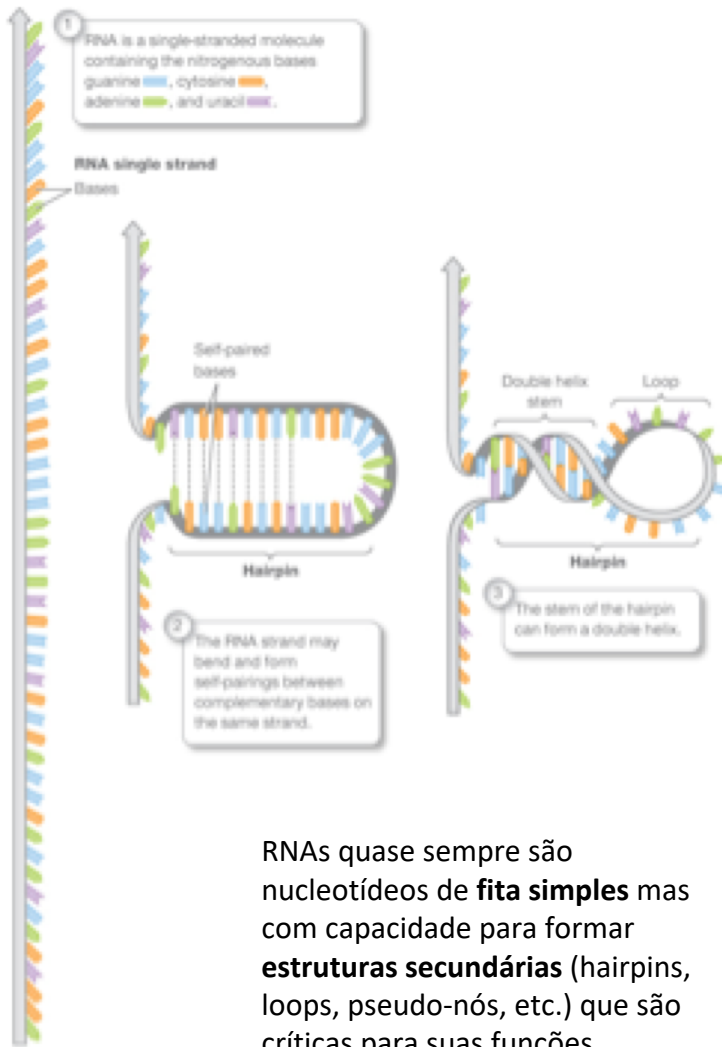
C terminus



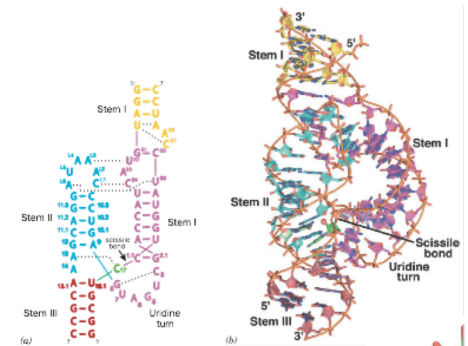
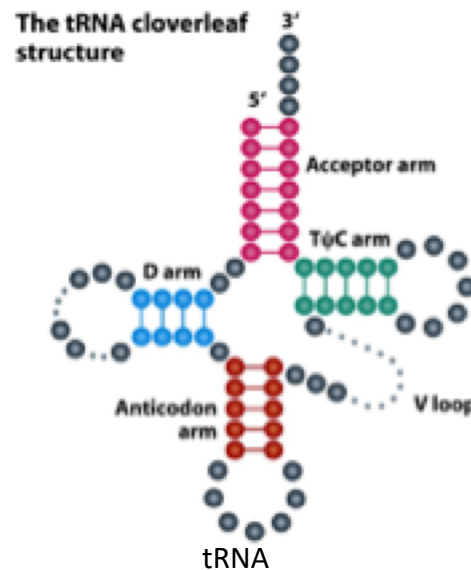
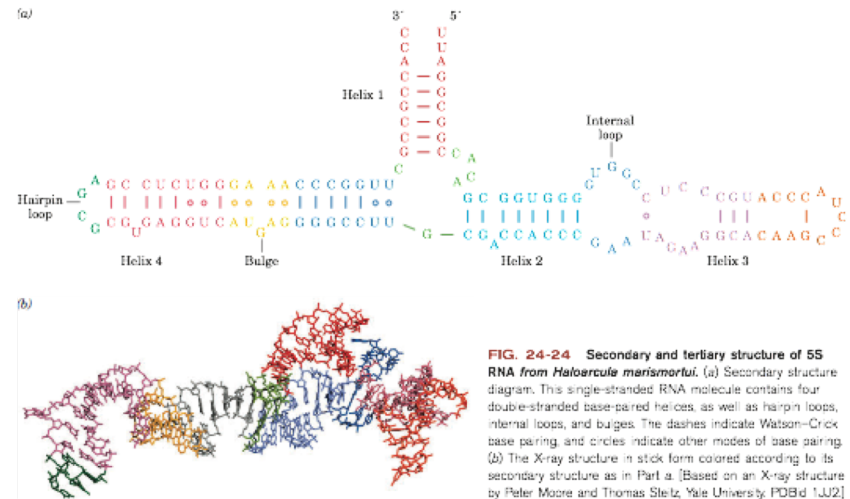
RNA: tipos



RNA: estrutura



RNAs quase sempre são nucleotídeos de **fita simples** mas com capacidade para formar **estruturas secundárias** (hairpins, loops, pseudo-nós, etc.) que são críticas para suas funções



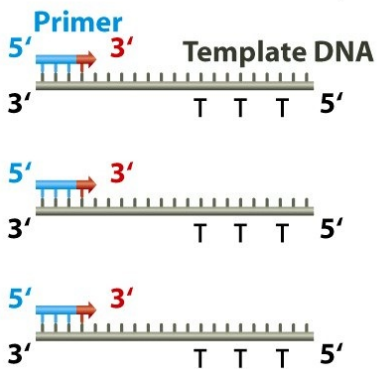
Ribozima “hammerhead” catalisa a própria clivagem

Genomas: como conseguir um?

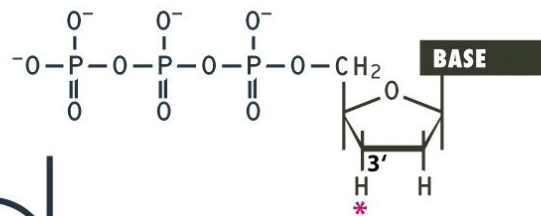
Sequenciamento de DNA Método Sanger



(A) Initiation of strand synthesis

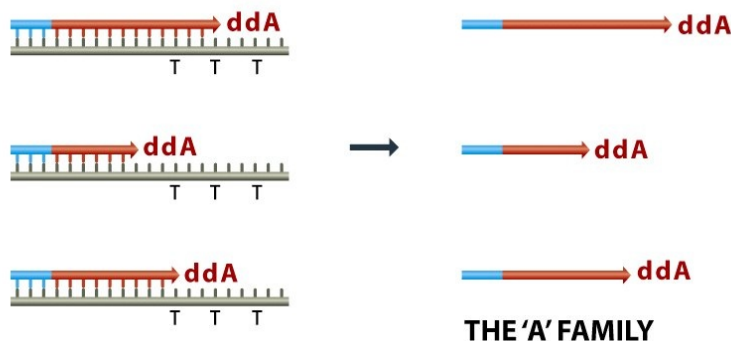


(B) A dideoxynucleotide



* Position where the -OH of a dNTP is replaced by -H

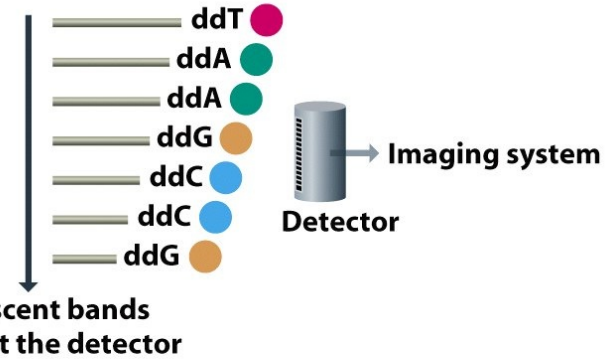
(C) Strand synthesis terminates when a ddNTP is added



(A)

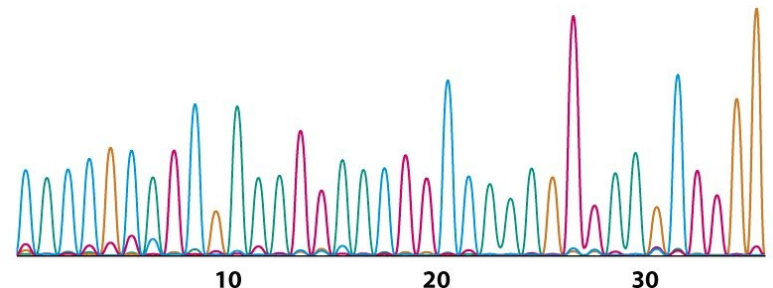
ddA ● ddC ● ddT ● ddG ● ddNTPs – each with a different fluorescent label

↓ Sequencing reactions, fractionation of products



(B)

C A C C G C A T C G A A A T T A A C T T C C A A A G T T A A G C T T G G



Genomas: como conseguir um?

Shotgun sequencing (1995)

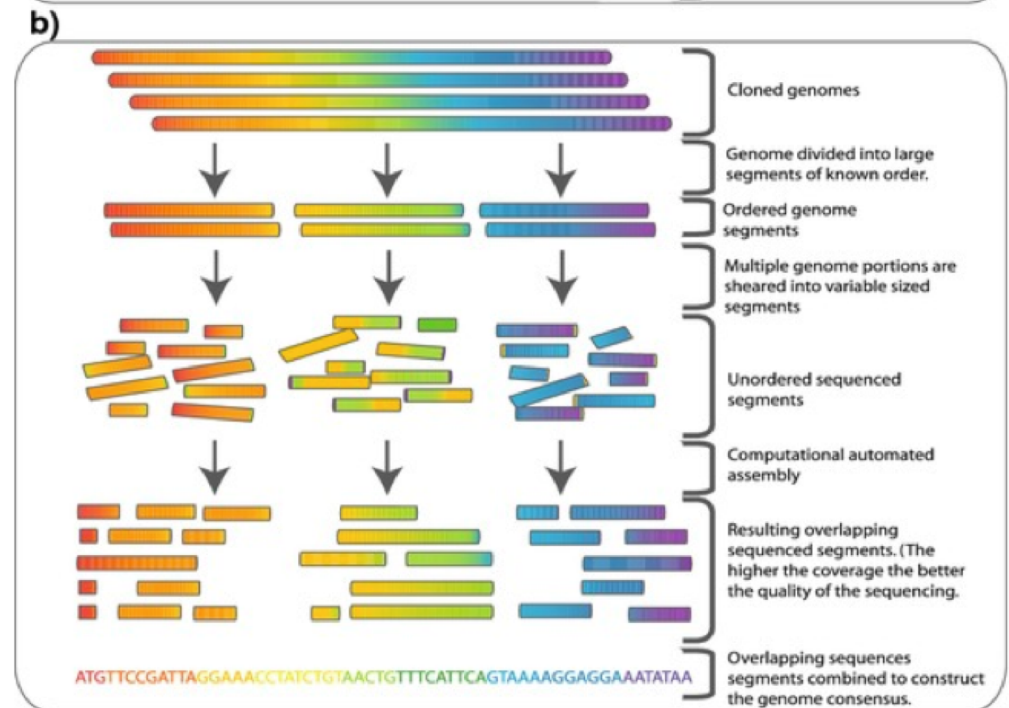
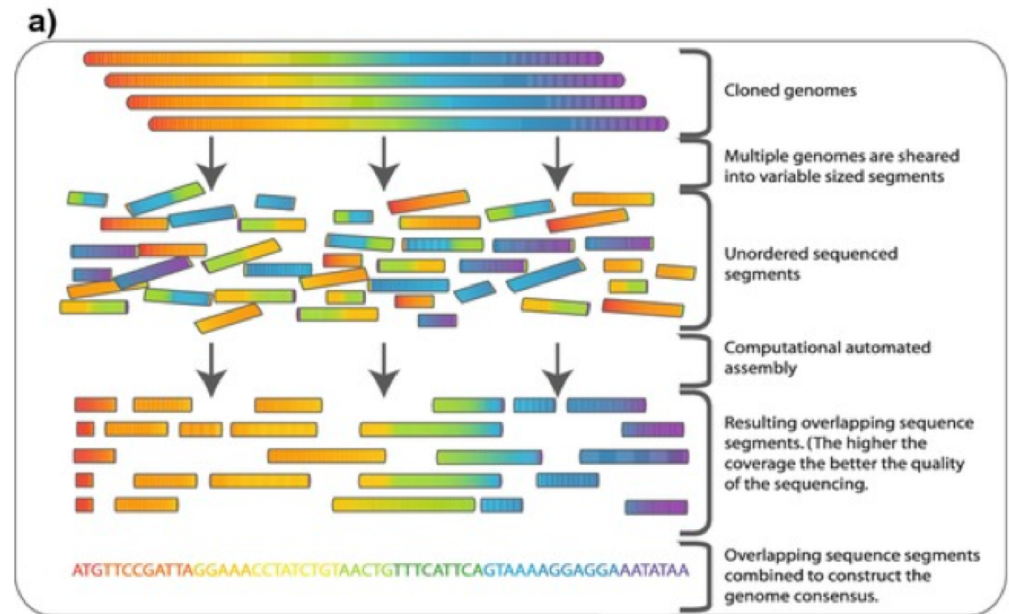
Começa com a (laboriosa) construção de uma **biblioteca de vetores** contendo fragmentos do genoma

Dependendo do tamanho e complexidade do genoma, duas estratégias poderiam ser utilizadas:

a) Genoma total (“*whole genome*”)

b) Hierárquicos (BACs)

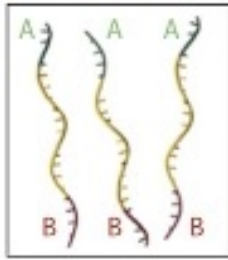
O método de sequenciamento Sanger precisa de grandes quantidades de DNA e cada um de seus produtos tem que ser processado em um único capilar



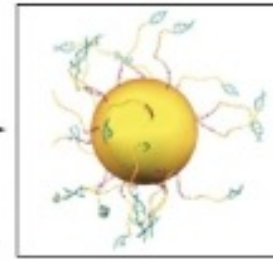
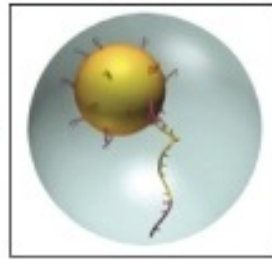
Sequenciadores de 2ª geração (2007)

Plataforma 454

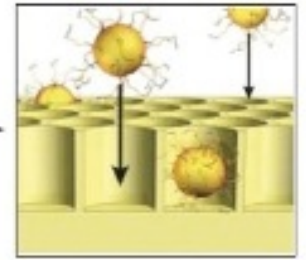
Construção da biblioteca



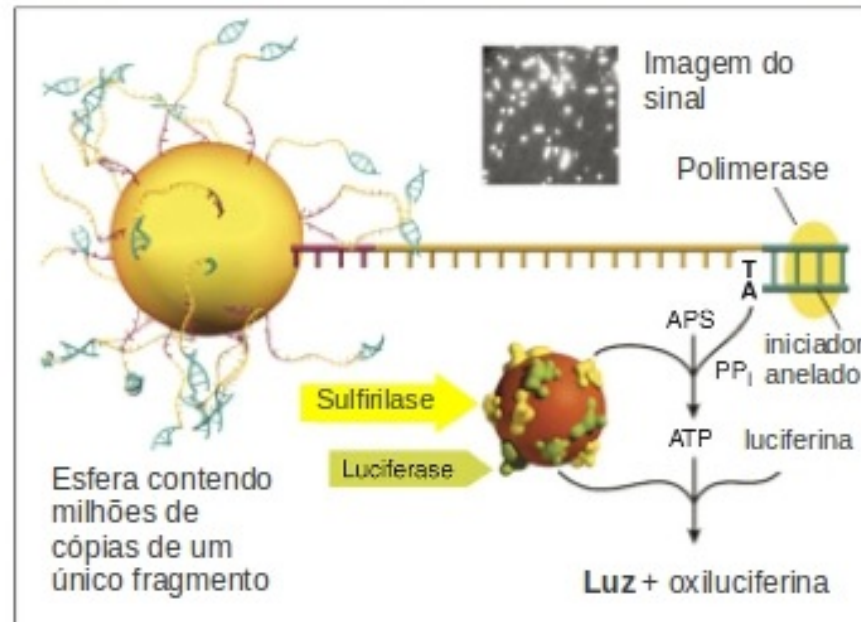
PCR em emulsão



Deposição na placa



Não precisa clonagem!!!



Reação de Pirosequenciamento

A miniaturização e a automação permitiram o aumento da escala: genomas em dias

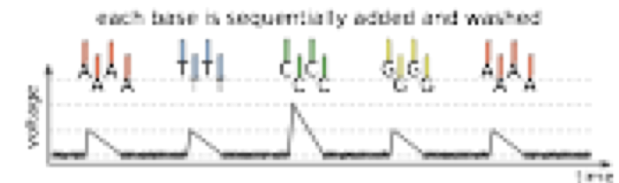
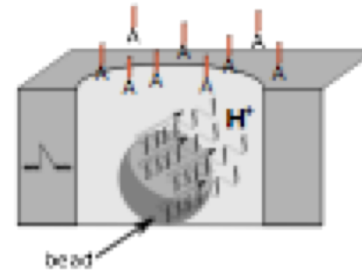
Novas tecnologias de sequenciamento

Da mesma forma que o 454, as novas tecnologias de sequenciamento aplicam avanços nas áreas de semicondutores, **microfluídica**, processamento e captura de sinais em tempo real.

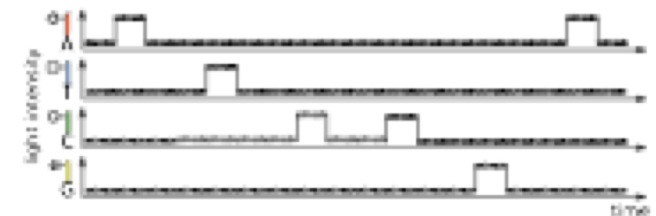
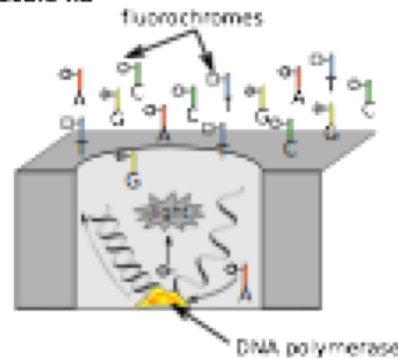
Essas tecnologias são exemplos reais de aplicações de **nanotecnologia** em biomedicina

template: TAGGCT

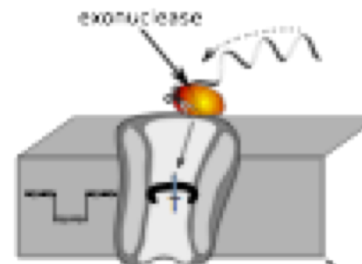
(A) Ion Torrent PGH



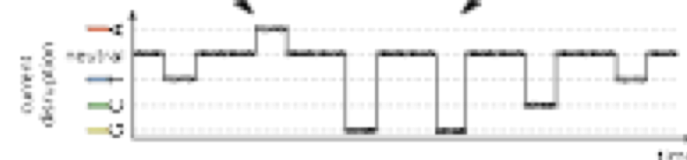
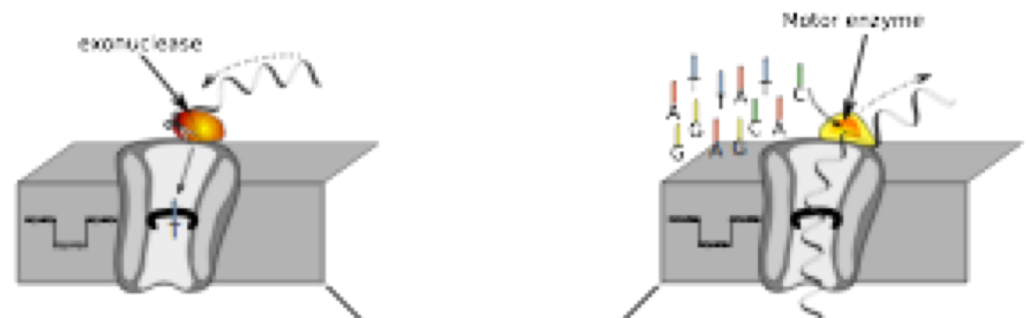
(B) PacBio RS



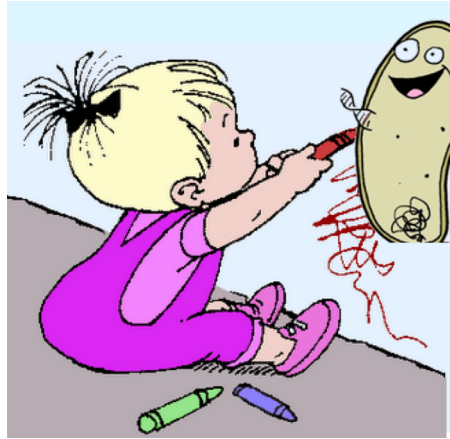
(C) GridION - exonuclease sequencing



(D) GridION - strand sequencing



Sequenciadores de bancada



Minion: secuenciador de bolso



454 GS Junior
(Roche)



**Ion Torrent PGM - Personal
Genome Machine**
(Life Technologies)



MiSeq
(Illumina)

A melhor tecnologia de sequenciamento de DNA?

Depende da aplicação...

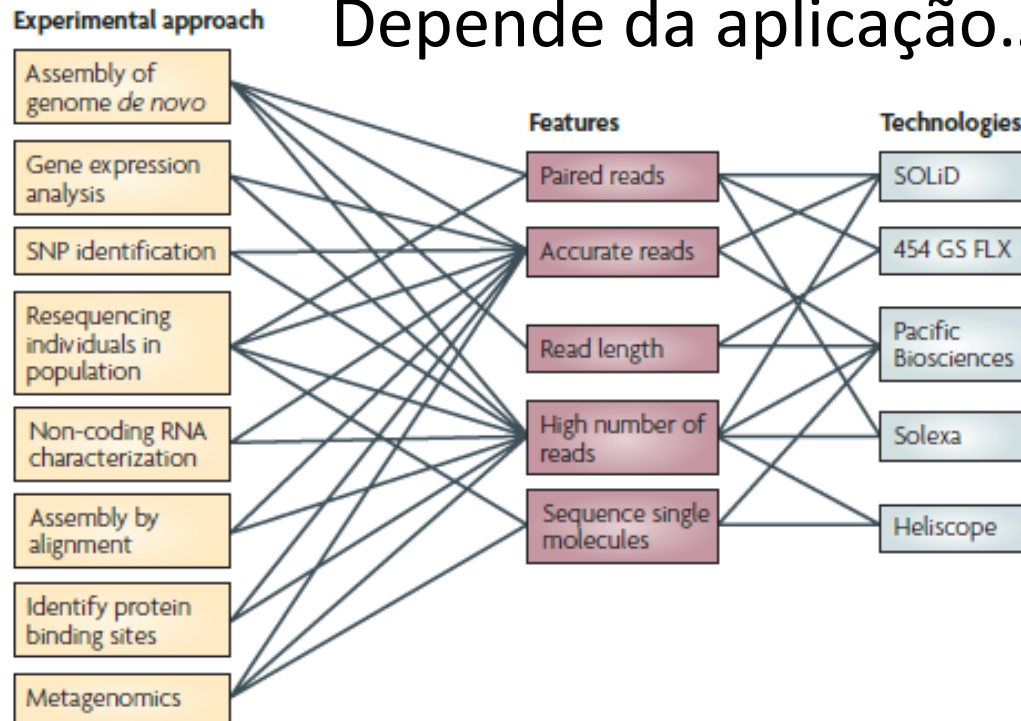
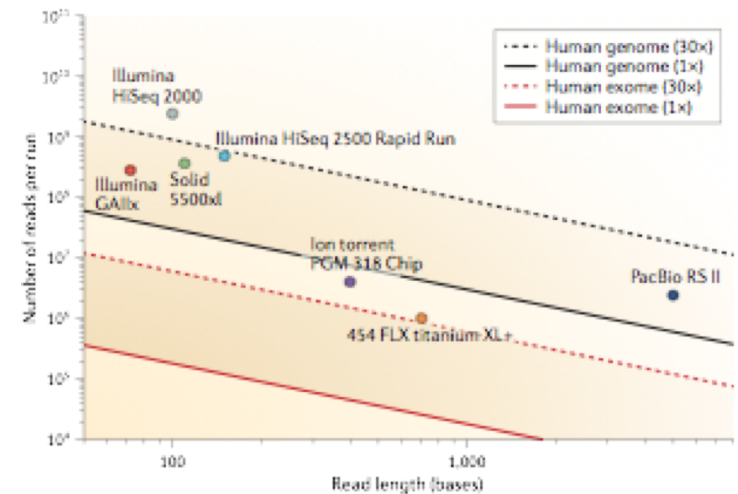


Figure 2 | Selecting a technology for an experiment. The yellow boxes represent different experimental approaches. The lines from each experimental approach lead to a 'wish list' of features (brown boxes) that are especially useful for the successful execution of an experiment or analysis. The features are then linked to the technologies that provide them. SNP, single-nucleotide polymorphism.

MacLean D, Jones JD, Studholme DJ. Application of 'next-generation' sequencing technologies to microbial genetics. *Nat Rev Microbiol.* 2009 Apr;7(4):287-96. doi: 10.1038/nrmicro2122.

Sims, D. et al. (2014) Sequencing depth and coverage: key considerations in genomic analyses. *Nat. Rev. Genet.*, 15, 121–32.



GAIIx, Genome Analyzer Iix; PacBio, Pacific Biosciences; PGM, personal genome machine.

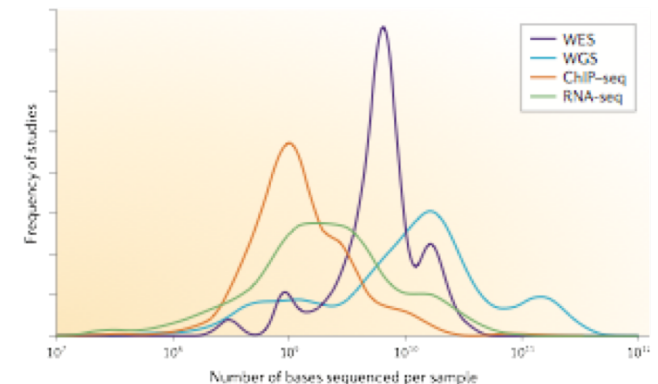


Figure 1 | Sequencing depths for different applications. The frequency of studies that use read counts of all runs (which are typically flow-cell lanes) and that were deposited from 2012 to June 2013 for the Illumina platform in the European Nucleotide Archive (ENA) is shown. The plot provides an overview of sequencing depths that are usually chosen for the four most common experimental strategies. Densities have been smoothed and normalized to provide an area under the curve that is equal to one. The depth and therefore the cost of an experiment increase in the order of chromatin immunoprecipitation followed by sequencing (ChIP-seq), RNA sequencing (RNA-seq), whole-exome sequencing (WES) to whole-genome sequencing (WGS). Although ChIP-seq, WES and WGS have typical applications and thus standardized read depths, the sequencing depth of RNA-seq data sets varies over several orders of magnitude. Multimodal distributions of WES and WGS reflect different target coverage. To generate this figure, runs were summed by experiment and, for each study, one experiment was chosen at random to avoid counting large studies more than once. Note that the ENA archive only contains published data sets and excludes medically relevant data sets. The plot was created from 771 studies.

Comparação da plataformas de sequenciamento

Plataforma	Instrumento	Reads	Comprimento	Erro	Notas
Illumina	HiSeq X Ten	375,000,000	300	substitution	Whole human genome sequencing only.
Illumina	NextSeq 500 HO	400,000,000	300	substitution	Exoma, transcriptomas, genoma completo e resequenciamento.
Illumina	HiSeq HO v4	250,000,000	250	substitution	Sequenciamento <i>de novo</i> e resequenciamento de genomas.
Illumina	HiSeq Rapid Run	150,696,000	300	substitution	Same as HiSeq run, except option available for faster run time, slightly less reads
Illumina	MiSeq v3	25,000,000	600	substitution	Opção mais rápida da Illumina. Custo por base > HiSeq.
Illumina	MiSeq v2	16,000,000	250	substitution	
Ion	Proton I	60,000,000	200	indel	https://genohub.com/ngs-instrument-guide/
PacBio	PacBio RS II	47,000	8500	indel	Reads mais longos dentre os sequenciadores em uso
Roche 454	GS FLX+	700,000	700	indel	Good for sequencing of small genomes.
SOLiD	5500xl W	266,666,667	100	A/T Bias	High throughput, good for re-sequencing.
SOLiD	5500	81,500,000	100	A/T Bias	

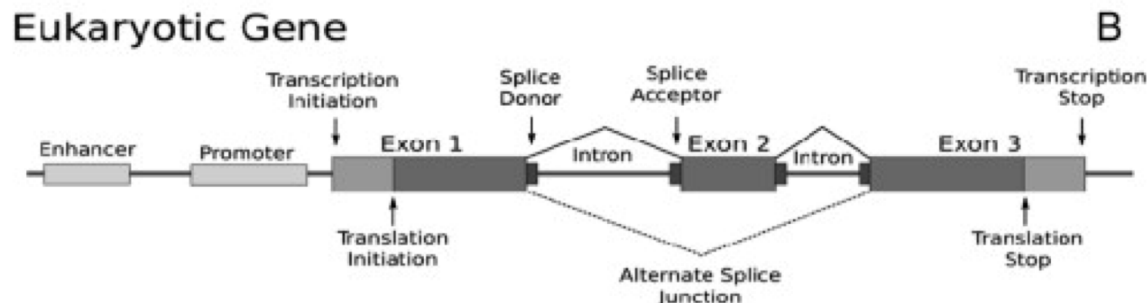
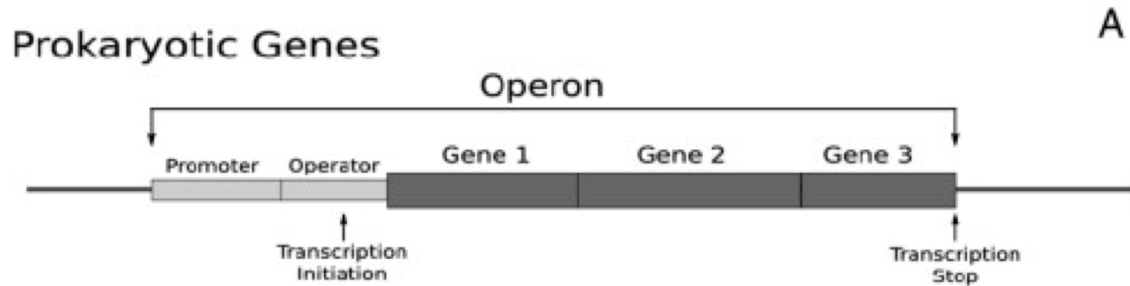
Anotação de genomas

Consiste na tarefa de identificar cada região ou regiões do genoma que possuem alguma **atividade** (função biológica) e/ou **história** independentes

- Anotação estrutural
 - Achar genes e identificar as ORFs
 - Identificar elementos reguladores, como regiões promotoras e terminadoras de transcrição
 - Achar as origens de replicação
- Anotação funcional
 - Atribuir aos genes uma função com base na sequência (IAE)
 - Identificar elementos móveis

Anotação estrutural: como achar um gene?

GGGGTTCGGGGAGCAATGGAACAGGGAAGTCAGTTAAGCCGGTATCAATGAAATGTCGGAGCGCTGGTAGCACAGCAATGCTTGTACT
TCCGCCCCACTGCCACATGGGCACTGATCATTTCTCCCAATCTTAGGACCTTGTCTGCGTAAAGGCCCTTGGGTTGCCAATTGGCGATG
TGAGGCAAAGTAGCGATAGATGTGCGTCAGTCCGGCAATCAGGGACTGAAGCAACTCATCGCGTTTGTCCGGCGCAATCTCTGGCGGG
CGCATGGCGGGATCAGGATCATGTTTCATGGGTCAAGATCATGATTGGCAGCATGGGCCCCGCCAAATTCTTCGCTGCCGATCAACTCTT
GCCAACTATTGGGCCGTAATTGAATGCCGCGCATGAAACCACGTGCCCAGTCGTTACCGTGTACGGCCCCATCCGCATCTTCGAGTAG
CACGGGAAGGTAGACATCGTCTTTCTCCAGAGTGCAGGAACAATTCTGATGCAATCGTGTTCCAGTGCCGCATGACCAAGCCAAGAATT
TCGGCGGCTTGGTCATTGCTGTGCGAAGCAATGGTCTTCCCCGAATACTTGTGGGCAGTGTTTCGCTGGGCAAAACCGTTTCTGGACCAC
AAATGAGTGCAGCAAGAACCATCGAGCATTTTCGAGATTCATTGCAGGTGCGCCGACTCCTTCGAGAAAGTCGCCAGCCGATCAAG
CTGGGCATCGTTCAACGGTTGTTGCGTCATTTGTTGCATGGTGAATTTCCAAGGTTTTTAAGGATCATTCGGTTTGGTCAACGGTCTC
AGCTCGCCGCGCGACAGCTTCCGGCACCGAGAACGAATAGCGTCCGAGCATGTTGATGTGCTCGTGGCCGAACGGCGACAGCCGTG



Gene

ATG GGC AGT GCA

TTT TAG

GAC →
 TGA →
 ATG →
 5' = ATGACGAGAGAGCAGCCATTTTAG = 3'
 3' = TACTGCTCTCTCGTCGGTAAAATC = 5'
 ← ATC
 ← AAT
 ← AAA

[illegible]

Anotação funcional

Associação dos elementos estruturais e/ou seus produtos (RNA e proteínas) com atividades e processos biológicos

Muitas vezes limitada a descrições textuais de cada proteína/gene

```
>gi|113373879|dbj|BAF03560.1| clam ADP-ribosylating protein CARP-1  
MPAGKYLYRVLRDDEDPSKGLSARNPGATTAKVSSHVNGLKRSPISTASKHAAQAFYWLSVKKWA  
KKNKKEREDPQKTFKIVEIDKDTLIKENKSVEVIDLSDKSVRDEYLGKNKKLQNYAAKYEEVLVKGF  
IPANCVKEIHLAALTDTESSEGYKENDYDVDTDDLAVAGAAGMTLNKP
```

```
>gi|108759255|ref|YP_634578.1| hypothetical protein MXAN_6455  
MATLLALLLLTLPVKPENPGSAGAPPASQADSTRYVFAWRGVPVGTVTCLKLEPGRFTYVSQHLHTRA  
GKAGARRREQSLAVDAQGRVRGSPAVPQALWLWKGVPVAVGCVVGREELTGREGPHCVTRVEDRRVEG  
TLLGTAFQARYGADGQMEALEVGDSRFTVAAPGERLRAPPELFADGLPVEGARGALSLEPPLDVPAR  
PAGLTPWNAAAARALAAEVHAAFPKTPGGADWSEGEAGEAGGCLAHALRFAAGARARGVKVALVHG  
LLVVDGGPARPHAWVRVALSKGGTLDLDPTSLDEVSPQTHLPIALAEAQGPALAGARWLSLLRGH  
RVVRRP
```


Anotação de genomas: exemplo de anotação funcional

ORF

BlastP

códon inicial
nome
produto

PFAM

domínio
nome
produto

COG

grupos ortólogos
nome
produto

KEGG

EC #
via metabólica
nome
produto

Vizinhos

clusters
operons

Psi-blast

alinhamento
por
posição

Blast vs.

Transp.
Número TC

BlastX

frameshift
baixa
qualidade

PSORT

localização

BlastN

confirmação

Completa

Dúvida

nome
produto
tamanho
número EC
cat. primária
cat. secundária
observações

Incompleta

Registrar

Dúvida

Completa

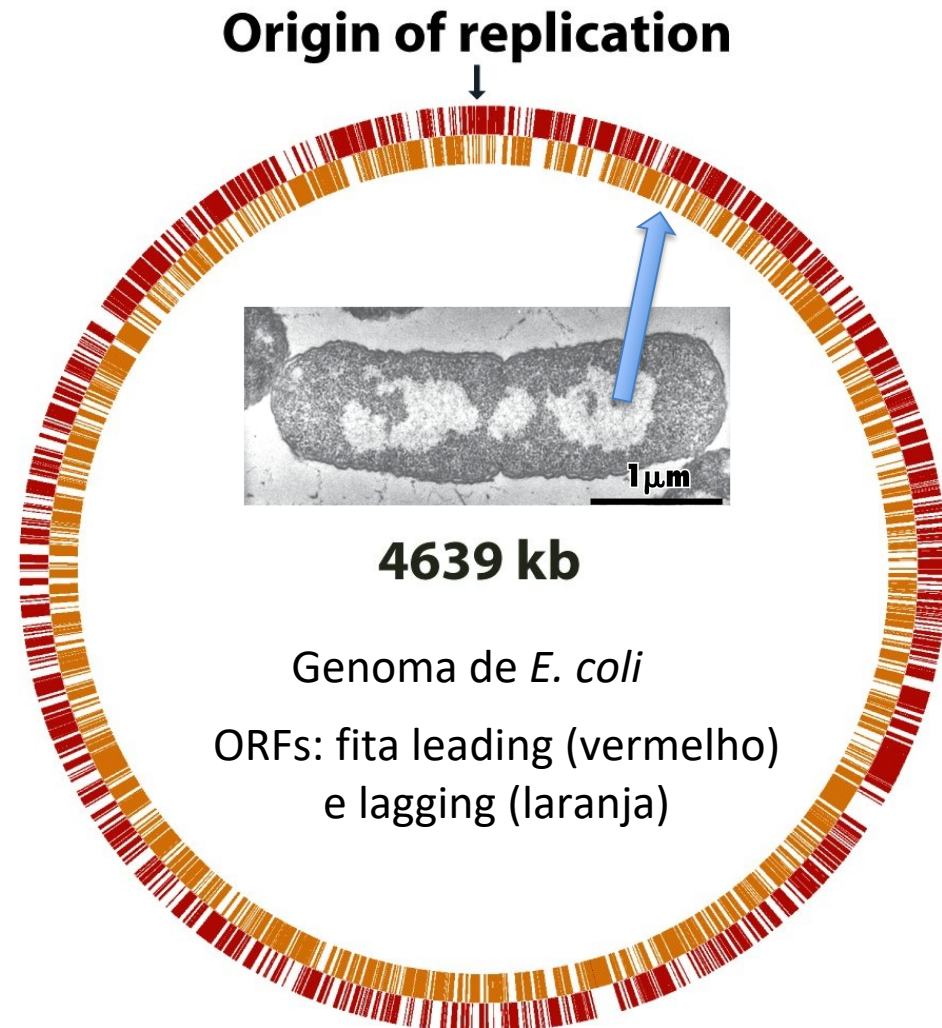
Genomas: histórico

- Primeiro genoma completo (**1974**)
 - Fago Φ X174 (RNA)
- Primeiro genoma de uma bactéria (**1995**)
 - *Haemophilus influenzae*: foi também o primeiro feito com “Whole-genome random sequencing”
- Primeiro genoma de eucarioto (**1996**)
 - *Saccharomyces cerevisiae*
- Primeiro genoma de um animal (**1998**)
 - *Caenorhabditis elegans* Strain: Bristol N2
- **Genoma humano (2001)**
 - Craig Venter vs. International consortium
- 1000 Genomes (**2012**)
- Genomas Online Database
 - Quase 40000 projetos de sequenciamento de genomas em Feb/2014

Genomas de procariotos

>gi|49175990|ref|NC_000913.2| Escherichia coli str. K-12 substr. MG1655, complete

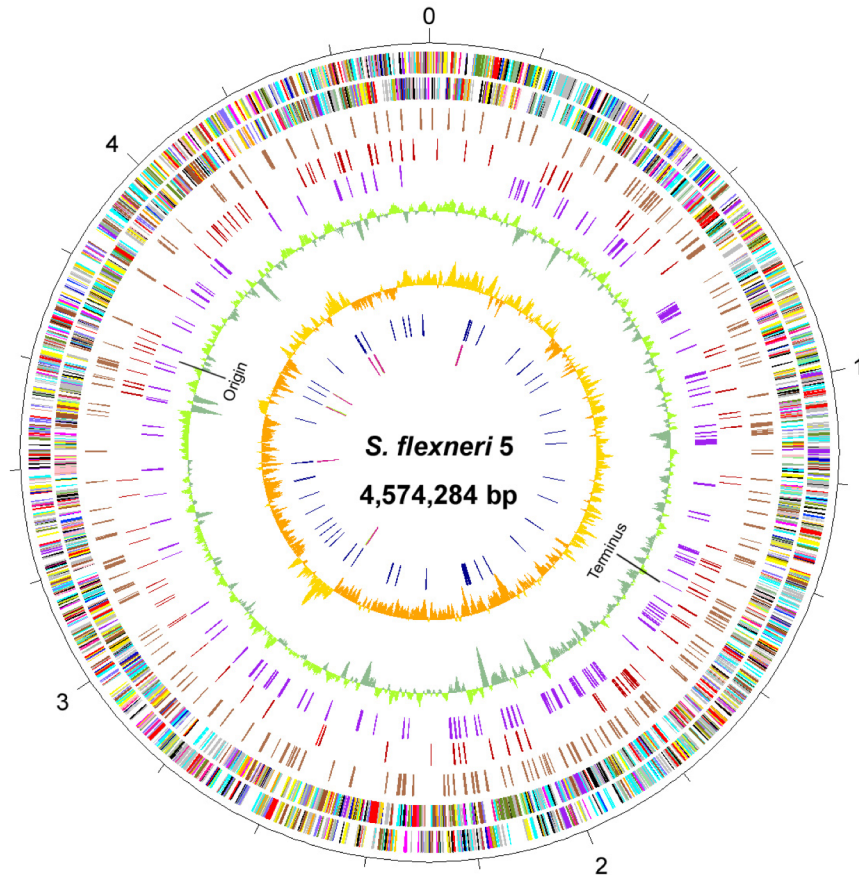
AGCTTTTCATTCTGACTGCAACGGGCAATATGTCTCTGTGTGGATTAAAAAAGAGTGTCTG
 ATAGCAGCTTCTGAAGTGGTTACCTGCGGTGAGTAAATTTAAATTTTATTGACTTAGGTCAC
 TAAATACTTTAACCAATATAGGCATAGCGCACAGACAGATAAAAAATTACAGAGTACACAACA
 TCCATGAAACGCATTAGCACCACCATTACCACCACCATCACCATTACCACAGGTAACGGTGC
 GGGCTGACGCGTACAGGAAACACAGAAAAAGCCCGCACCTGACAGTGC GGCTTTTTTTTT
 CGACCAAAGGTAACGAGGTAACAACCATGCGAGTGTGAAGTTCGGCGGTACATCAGTGGCA
 AATGCAGAACGTTTTCTGCGTGTGCGCATATTCTGGAAAGCAATGCCAGGCAGGGGCAGGT
 GGCCACCGTCCTCTCTGCCCCGCCAAAATCACCACCACTGGTGGCGATGATTGAAAAA
 CCATTAGCGGCCAGGATGCTTTACCAATATCAGCGATGCCGAACGTATTTTTGCCGAACCT
 TTGACGGGACTCGCCGCCGCCAGCGGGGTTCCCGCTGGCGCAATTGAAAACTTTTCGTCGA
 TCAGGAATTTGCCCAAATAAAACATGTCTGCATGGCATTAGTTTGTGGGCGAGTGCCTGG
 ATAGCATCAACGCTGCGCTGATTTGCCGTGGCGAGAAAATGTCGATCGCCATTATGGCCGCG
 GTATTAGAAGCGCGCGGTCAACGTTACTGTTATCGATCCGGTCGAAAACTGCTGGCAGT
 GGGGCATTACCTCGAATCTACCGTCGATATTGCTGAGTCCACCCGCCGTATTGCGGCAAGCC
 GCATTCCGGCTGATCAGTGGTCTGATGGCAGGTTTACC CGCCGTAATGAAAAAGGCGAA
 CTGGTGGTGCTTGACGCAACGTTCCGACTACTCTGCTGCGGTGCTGGCTGCCTGTTTACG
 CGCCGATTGTTGCGAGATTTGGACGGACGTTGACGGGGTCTATACCTGCGACCCGCGTCAGG
 TGCCCGATGCGAGGTTGTTGAAGTCGATGTCCTACCAGGAAGCGATGGAGCTTTCCTACTTC
 GGCCTAAAGTTCTTACCCCCGCACCATTACCCCCATCGCCAGTTCAGATCCCTTGCT
 GATTAAAAATACCGGAAATCCTCAAGCACCAGGTACGCTCATTGGTGCCAGCCGTGATGAAG
 ACGAATTACCGGTCAAGGGCATTTCGAATCTGAATAACATGGCAATGTTACGCTTTCTGGT
 CCGGGGATGAAAGGGATGGTCCGCATGGCGGCGCGCTTTTGACGCGATGTCACGCGCCCCG
 TATTTCCGTGGTGCTGATTACGCAATCATCTCCGAATACAGCATCAGTTTCTGCGTTCCAC
 AAAGCGACTGTGTGCGAGCTGAACGGGCAATGCAGGAAGAGTTCTACCTGGAAGTGAAGAA
 GGCTTACTGGAGCCGCTGGCAGTGACGGAACGGCTGGCCATTATCTCGGTGGTAGGTGATGG
 TATGCGCACCTTGCGTGGGATCTCGCGGAAATCTTTGCCGCACTGGCCCGCGCCAATATCA
 ACATTGTCGCCATTGCTCAGGGATCTTCTGAACGCTCAATCTCTGTCGTGGTAAATAACGAT
 GATGCGACCACTGGCGTGCGGTTACTCATCAGATGCTGTTCAATACCGATCAGGTTATCGA
 AGTGTGTTGTGATTGGCGTCGGTGGCGTTGGCGGTGCGCTGCTGGAGCAACTGAAGCGTCAGC
 AAAGCTGGCTGAAGAATAAACATATCGACTTACGTGTCTGCGGTGTTGCCAACTCGAAGGCT
 CTGCTACCAATGTACATGGCCTTAATCTGGAAAAGTGGCAGGAAGAACTGGCGCAAGCCAA
 AGAGCCGTTTAATCTCGGGCGCTTAATTG...



Gene Map of the *Mycoplasma pneumoniae* Genome

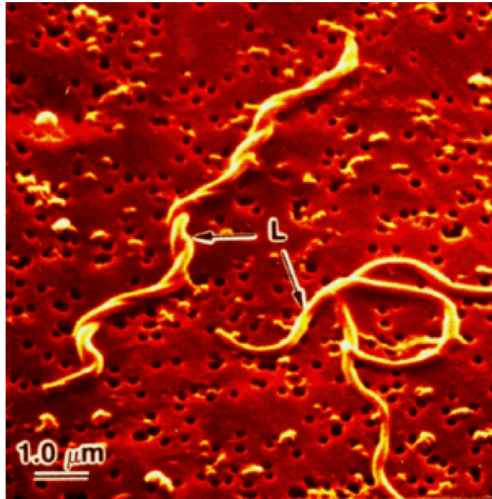


Genomas completos: exemplos



Complete genome sequence of *Shigella flexneri* 5b and comparison with *Shigella flexneri* 2a. BMC Genomics (2006) 7:173

Genomas completos



Borrelia burgdorferi



21 plasmídeos, alguns lineares
Cromossomo de ~930Kb

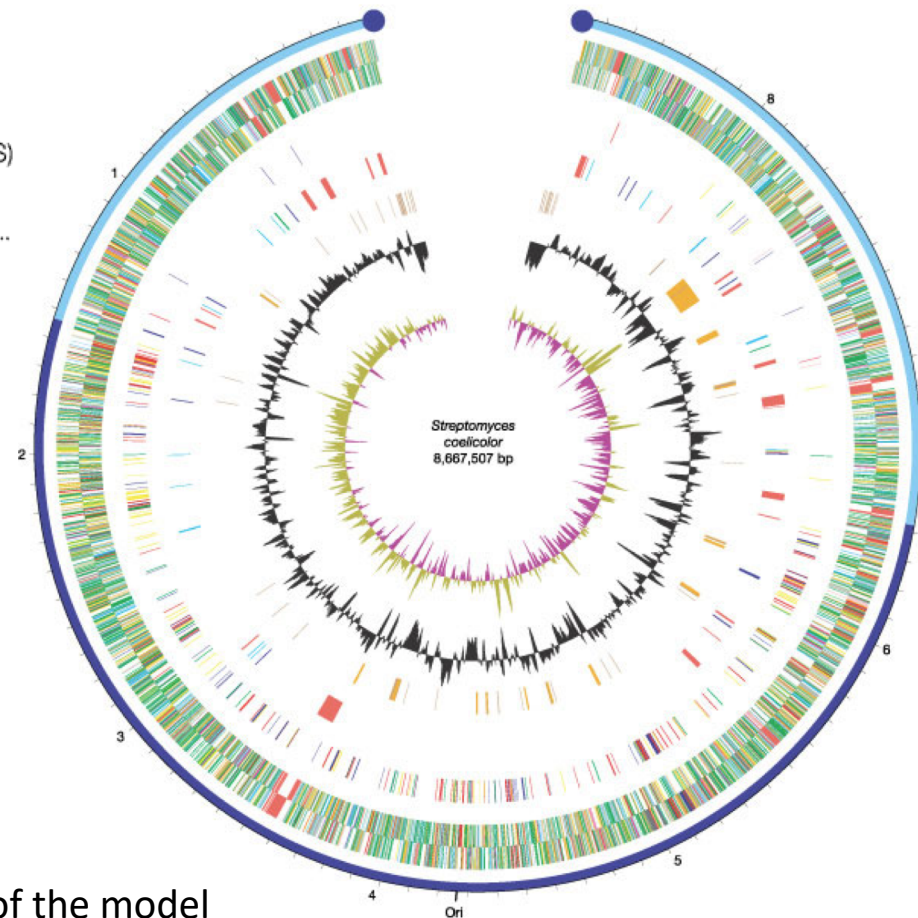


Streptomyces

Table 1 General features of the chromosome

Component of chromosome	Property
Total size	8,667,507 bp
Terminal inverted repeat	21,653 bp
G + C content	72.12%
Coding sequences	7,825
...of which pseudogenes	55
Coding density	88.9%
Average gene length	991 bp
Ribosomal RNAs	6 × (16S–23S–5S)
Transfer RNAs	63
Other stable RNAs	3

The outer scale is numbered anticlockwise (to correspond with a previously published map) in megabases and indicates the core (dark blue) and arm (light blue) regions of the chromosome. Circles 1 and 2 (from the outside in), all genes (reverse and forward strand, respectively) colour-coded by function (black, energy metabolism; red, information transfer and secondary metabolism; dark green, surface associated; cyan, degradation of large molecules; magenta, degradation of small molecules; yellow, central or intermediary metabolism; pale blue, regulators; orange, conserved hypothetical; brown, pseudogenes; pale green, unknown; grey, miscellaneous); circle 3, selected 'essential' genes (for cell division, DNA replication, transcription, translation and amino-acid biosynthesis, colour coding as for circles 1 and 2); circle 4, selected 'contingency' genes (red, secondary metabolism; pale blue, exoenzymes; dark blue, conservon; green, gas vesicle proteins); circle 5, mobile elements (brown, transposases; orange, putative laterally acquired genes); circle 6, G + C content; circle 7, GC bias ((G - C/G + C), khaki indicates values >1, purple <1). The origin of replication (Ori) and terminal protein (blue circles) are also indicated.



Bentley, S.D. et al. (2002) Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature*, 417, 141–7.

Evolução de genomas

À medida que duas linhagens de organismos divergem, seus genomas tendem a mudar pela fixação de certos eventos que produzem mudanças em escalas que vão de um gene a cromossomos inteiros.

Esses eventos podem ser separados nas seguintes classes:

- Ganho de genes
 - Duplicação gênica (parálogos e famílias gênicas)
 - Duplicação de cromossomos inteiros (poliploidia)
 - Transferência lateral (xenólogos)
- Perda de genes
- Rearranjos cromossômicos
 - Mediada por elementos móveis
 - Mediada pela recombinação entre cromossomos irmãos durante a replicação
 - Incluem fenômenos como inversões e translocação de regiões do genoma

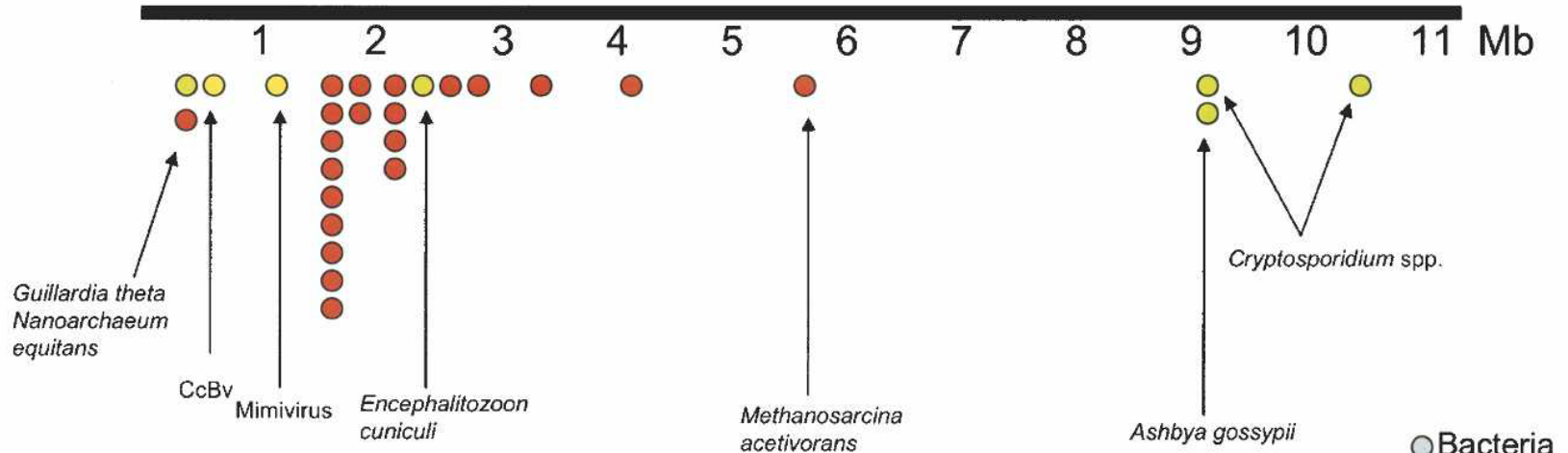
Genomas: variação de tamanho em microorganismos

Bacterial genomes <1Mb:

Mycoplasma spp.
Ureaplasma urealyticum
Phytoplasma asteris
Mesoplasma florum
Wigglesworthia glossinidia
Buchnera aphidicola
Blochmannia floridanus
Tropheryma whipplei
Borrelia garinii

Bacterial genomes >5.5Mb:

Escherichia coli (some), *Ralstonia solanacearum*
Photobacterium luminescens, *Bacillus anthracis*
Burkholderia mallei, *pseudomallei*
Nocardia farcinica
Pseudomonas aeruginosa, *putida*, *syringae*
Nostoc sp., *Bacteroides thetaiotaomicron*
Photobacterium profundum
Sinorhizobium meliloti
Rhodopirellula baltica, *Mesorhizobium loti*
Streptomyces coelicolor, *avermitilis*
Bradyrhizobium japonicum



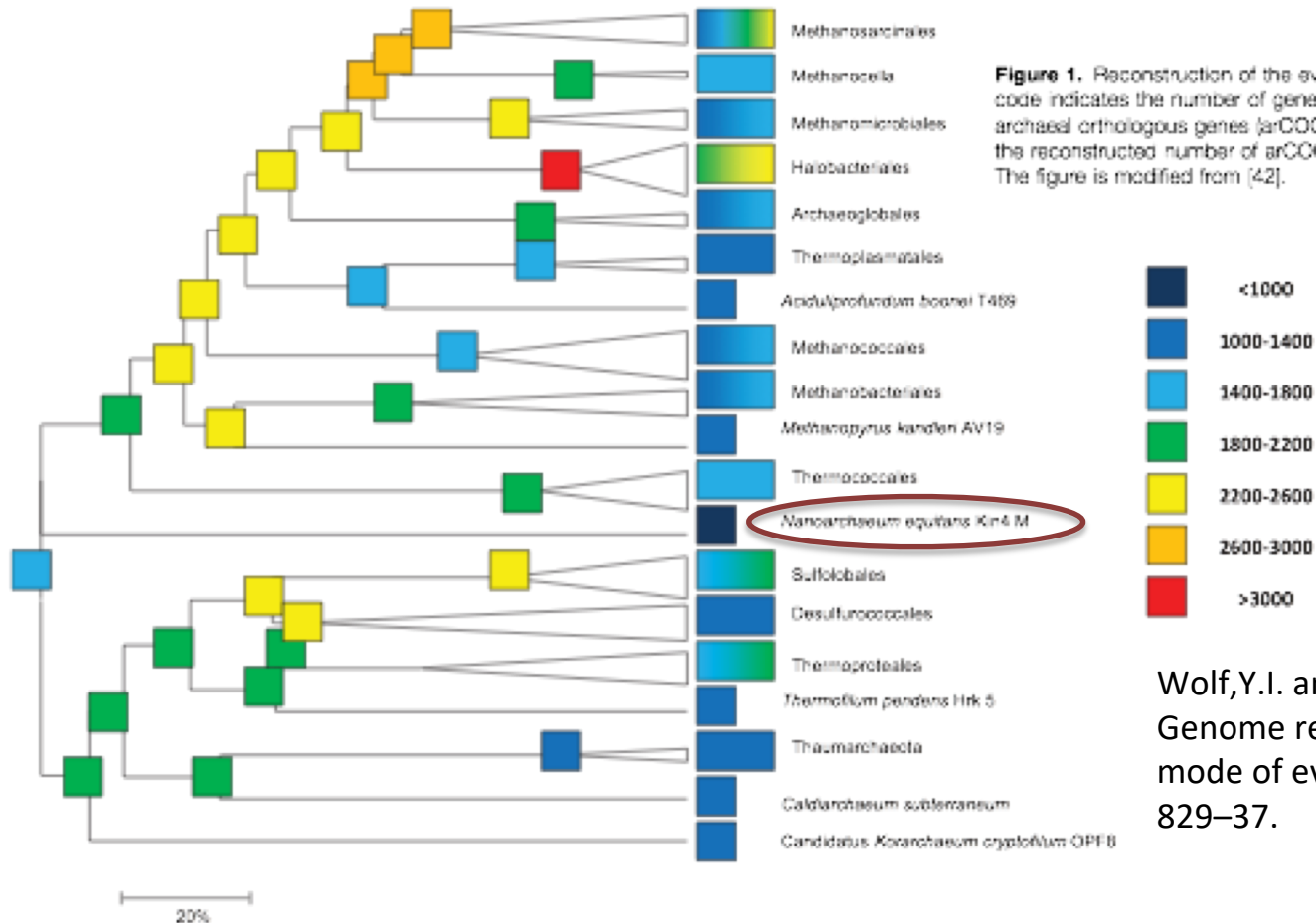
Insights on biology and evolution from microbial genome sequencing

Claire M. Fraser-Liggett

Genome Res. 2005 15: 1603-1610

Access the most recent version at doi:[10.1101/gr.3724205](https://doi.org/10.1101/gr.3724205)

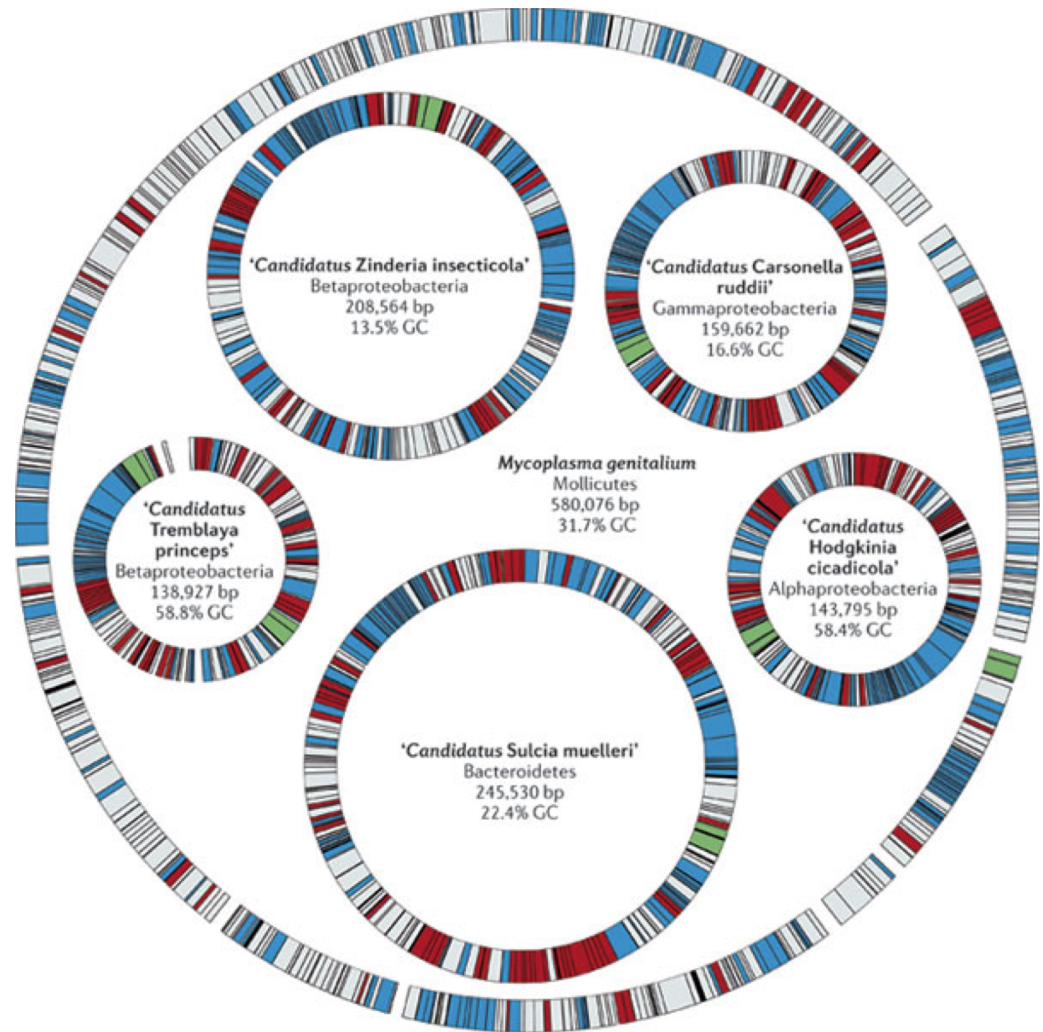
Perda de genes: arquea



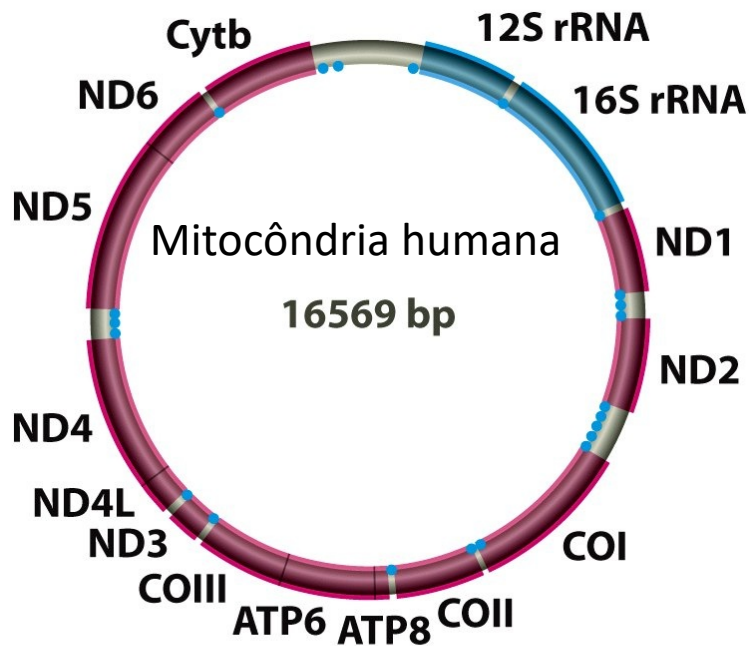
Wolf, Y.I. and Koonin, E. V (2013)
Genome reduction as the dominant
mode of evolution. Bioessays, 35,
829–37.

Perda de genes: bactérias simbiotes

The genome of **Mycoplasma genitalium**, the free-living organism with the **smallest** genome, is two to four times as large as the genomes of five symbionts recently shown to have tiny genomes (that is, smaller than 300 kb): 'Candidatus Sulcia muelleri', 'Candidatus Zinderia insecticola', 'Candidatus Carsonella ruddii', 'Candidatus Hodgkinia cicadicola' and 'Candidatus Tremblaya princeps'. Genes involved in informational processes are in blue, those involved in vitamin or amino acid biosynthesis are in maroon, ribosomal RNA genes are in green, other genes are light grey and breaks are non-coding regions.

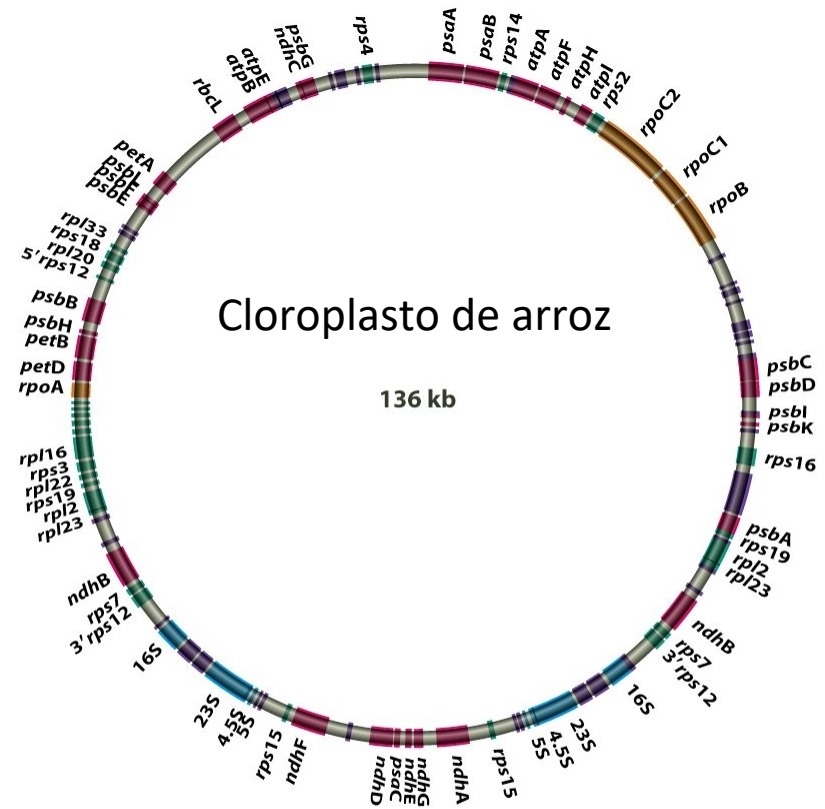


Redução de genoma: organelas de eucariotos



KEY

- Respiratory complex gene
- Ribosomal RNA gene
- Transfer RNA gene



KEY

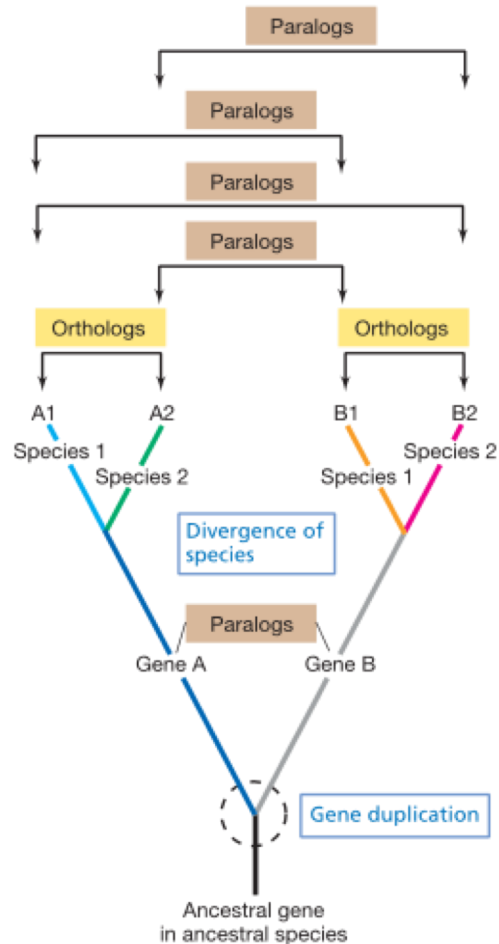
- Photosynthesis gene
- Ribosomal protein gene
- Transfer RNA gene
- Ribosomal RNA gene
- RNA polymerase gene

Endosimbiose de proteobactérias (mitocôndria) e cianobactérias (cloroplasto) seguida de redução e transferência de genes para o núcleo

Ganho de genes: duplicação gênica

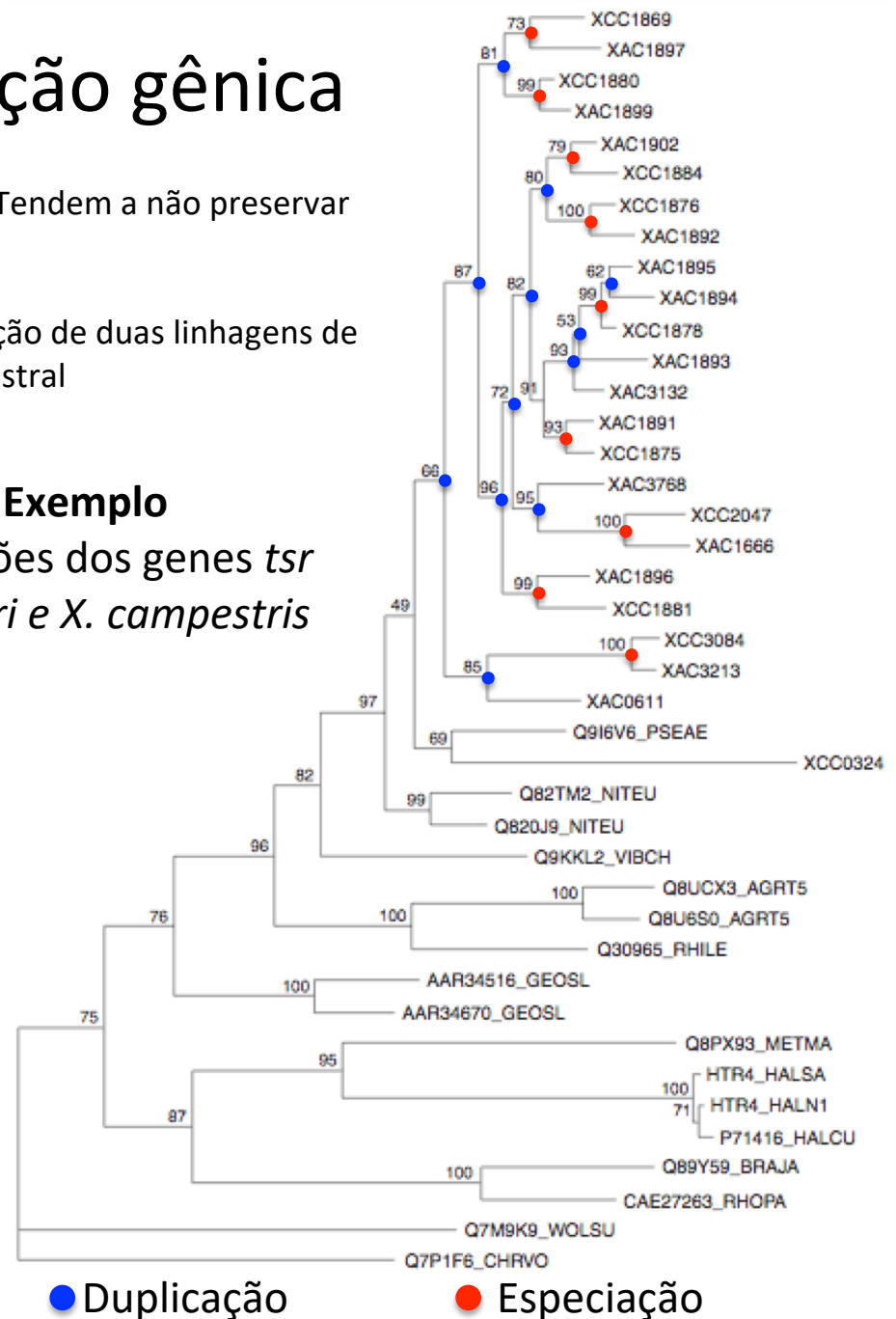
Parálogos: divergem por **duplicação** do gene codificante. Tendem a não preservar a função original

Ortólogos: divergem por **especiação**, ou seja, pela separação de duas linhagens de organismos. Tendem a preservar a mesma função do ancestral



Exemplo

Duplicações dos genes *tsr* em *X. citri* e *X. campestris*



Ganho de genes: transferência lateral

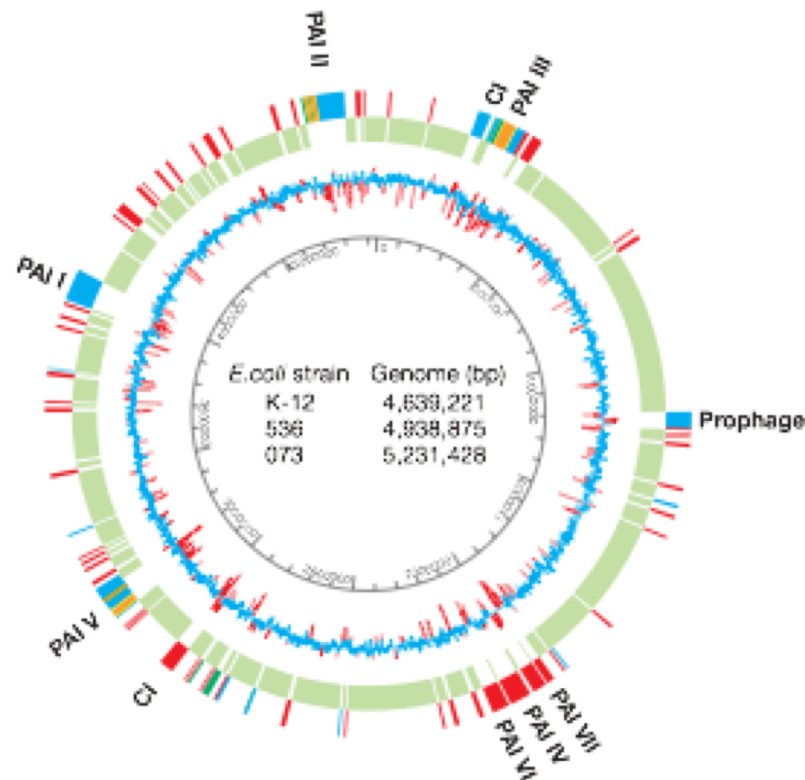


Figure 12.17 Pathogenicity islands in *Escherichia coli*. Genetic map of *E. coli* strain 536, a urinary tract pathogen, compared with a second pathogenic strain (073) and the wild-type strain K-12. The pathogenic strains contain pathogenicity islands, and thus their chromosomes are larger than that of K-12. Inner circle, nucleotide base pairs. Jagged circle, DNA GC distribution; regions where GC content varies dramatically from the genome average are in red. Outermost circle, three-way genomic comparison: green, genes common to all strains; red, genes present in the pathogenic strains only; blue, genes found only in strain 536; orange, genes of strain 536 present in a different location in strain 073. Some very small inserts deleted for clarity. PAI, pathogenicity islands; CI, chromosomal island. Prophage, DNA from a temperate bacteriophage. Note the correlation between genomic islands and skewed GC content. Data adapted from *Proc. Natl. Acad. Sci. (USA)* 103: 12879–12884 (2006).

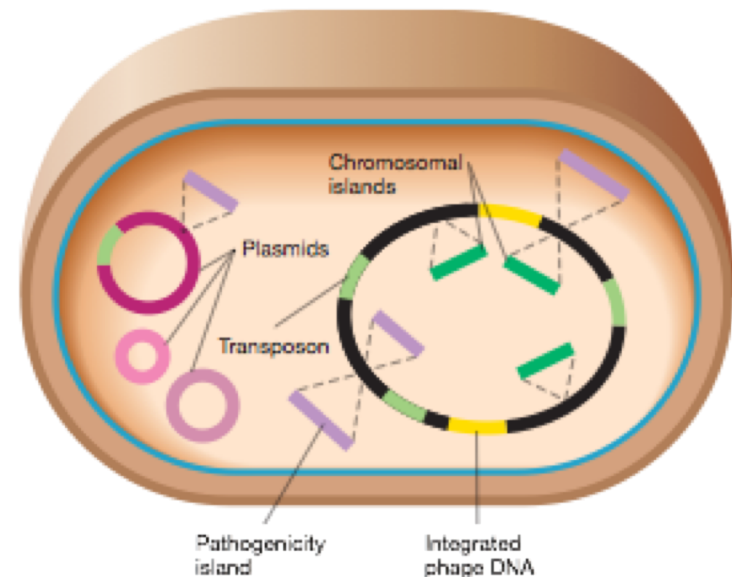
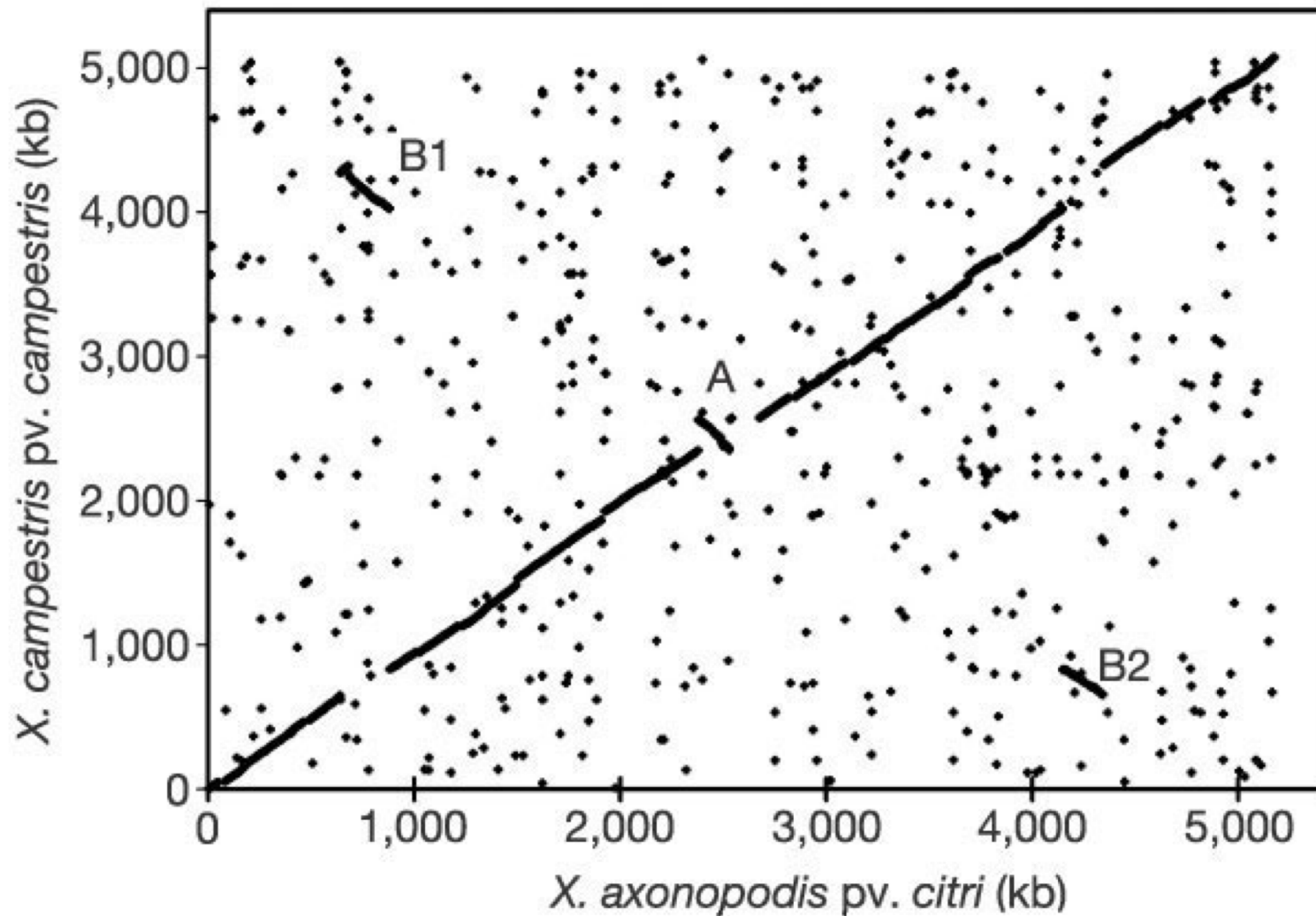


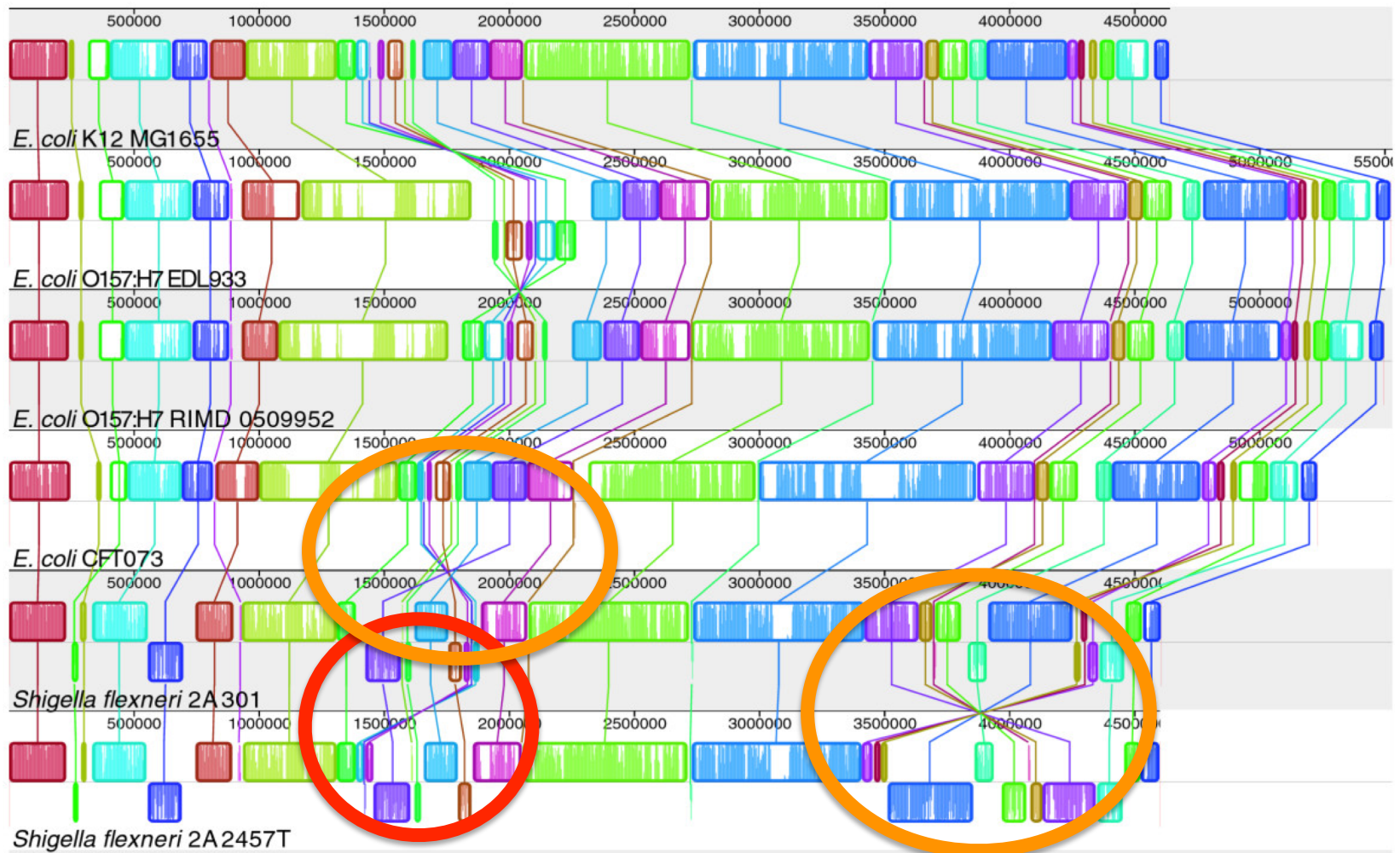
Figure 12.18 Pan genome versus core genome. The core genome is represented by the black regions of the chromosome and is present in all strains of a species. The pan genome includes elements that are present in one or more strains but not in all strains. Each colored wedge indicates a single insertion. Where two wedges emerge from the same location, they represent alternative islands that can insert at that site. However, only one insertion can be present at a given location. Plasmids, like the chromosome, may have insertions that are not present in all strains.

Evolução de genomas: rearranjos genômicos



Rearranjos genômicos

Comparação do genoma de diferentes linhagens de *E. coli* e *Shigella*
São visíveis regiões homólogas com mais de 1000 pares de bases (1Kbp)

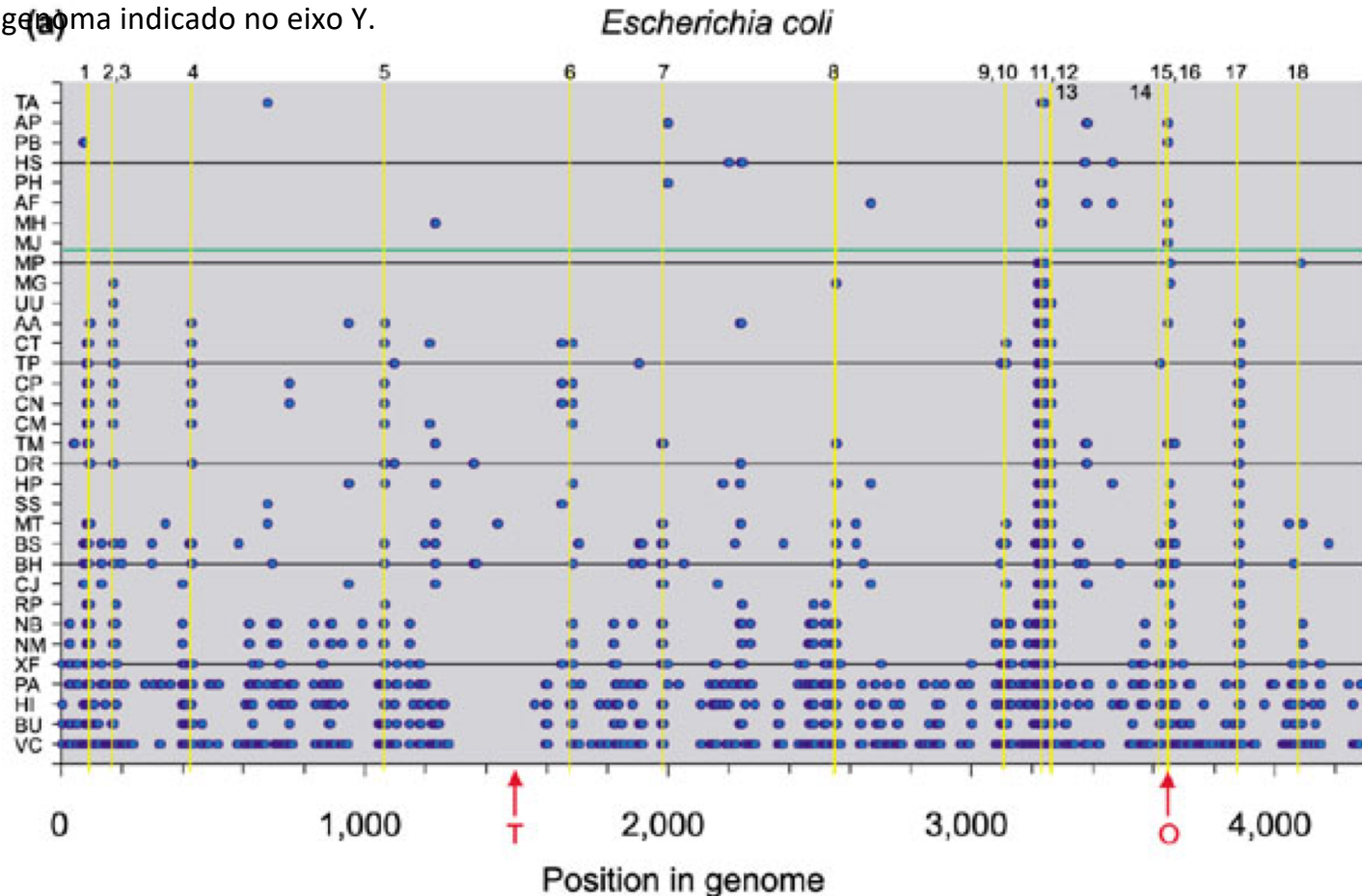


Rearranjos: efeito a longo prazo

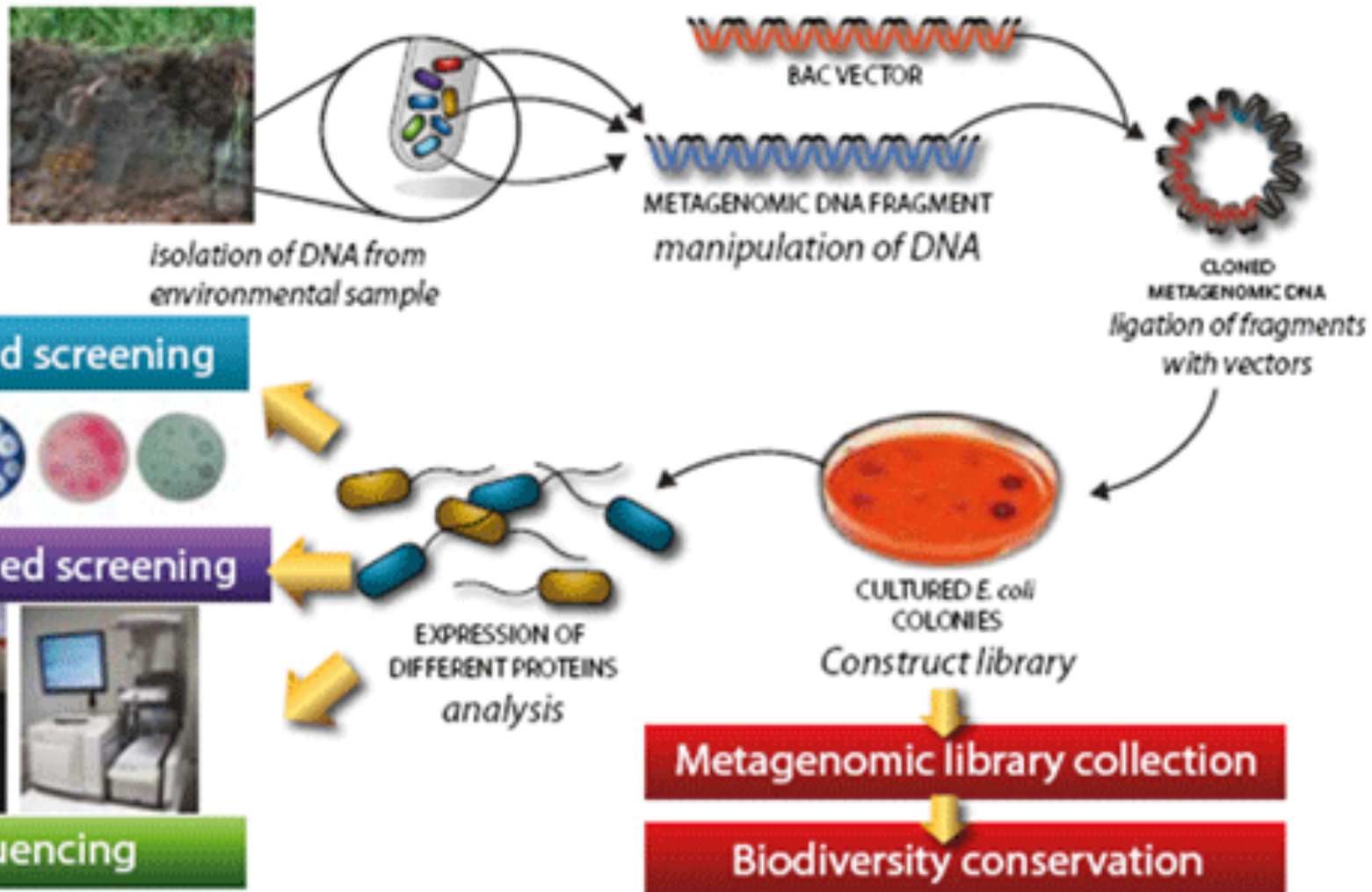
Quanto maior o tempo de divergência entre dois organismos, menor o grau de conservação da ordem de seus genes.

A figura abaixo mostra a comparação da ordem dos genes em *Escherichia coli* (eixo X) com a observada no genoma de várias espécies de procariotos (identificadas no eixo Y), incluindo algumas arqueas (no alto).

Cada ponto no gráfico representa três genes de *E. coli* cujos ortólogos são encontrados na mesma ordem no genoma indicado no eixo Y.



Metagenômica



Bactérias não cultiváveis e Metagenômica

- Apenas 1% das bactérias existentes no ambiente são cultiváveis
- Metagenômica
 - Identificação de microrganismos não cultiváveis
 - Genes com algum interesse específico presentes em diferentes amostras ambientais.

Metagenômica oral

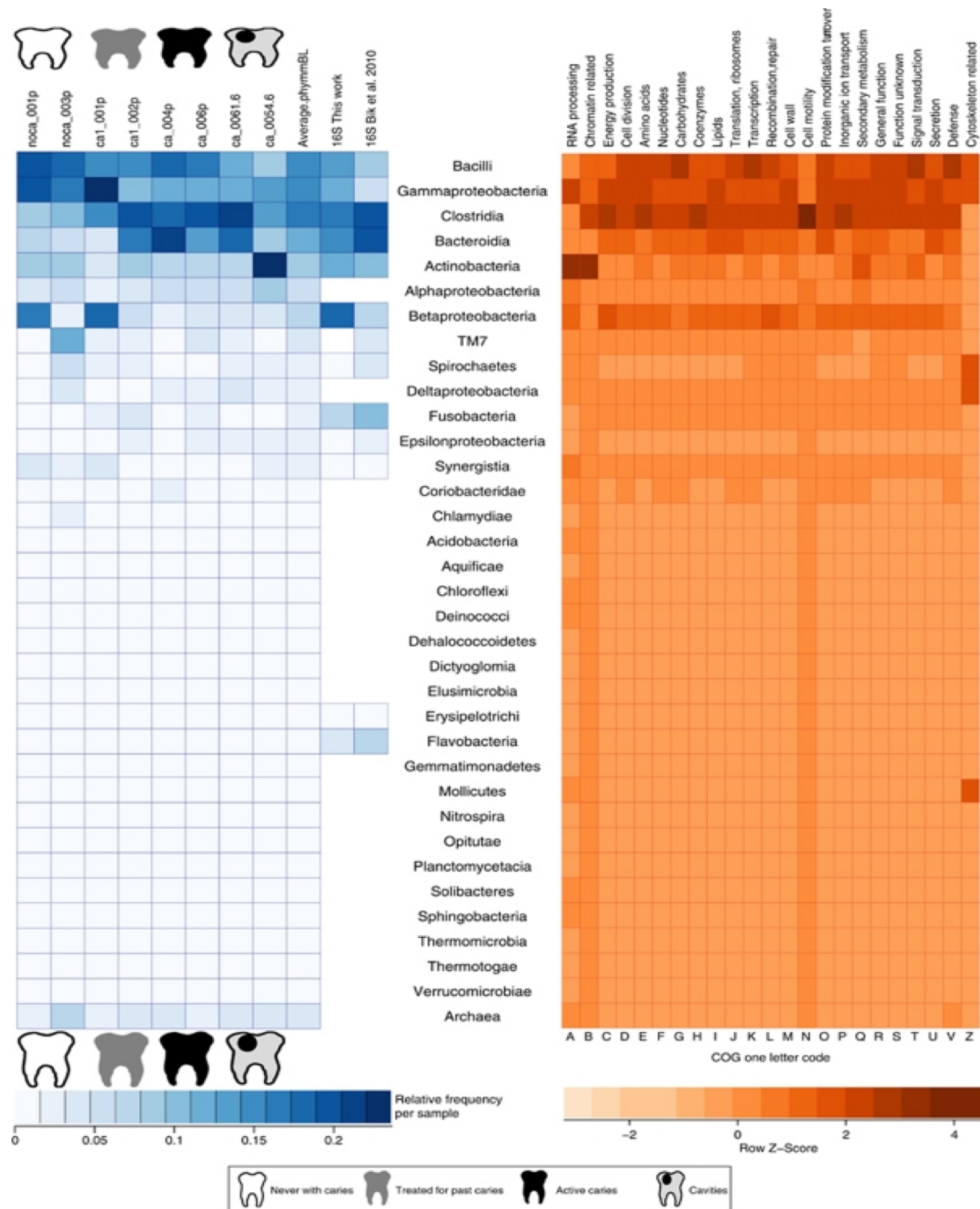
Diversidade de bactérias da cavidade oral (1Gbp sequenciados).

À esquerda: diversidade taxonômica na amostra (frequência relativa dos taxons)

À direita: contribuição relativa de cada grupo para o repertório de genes codificantes do ecossistema bucal

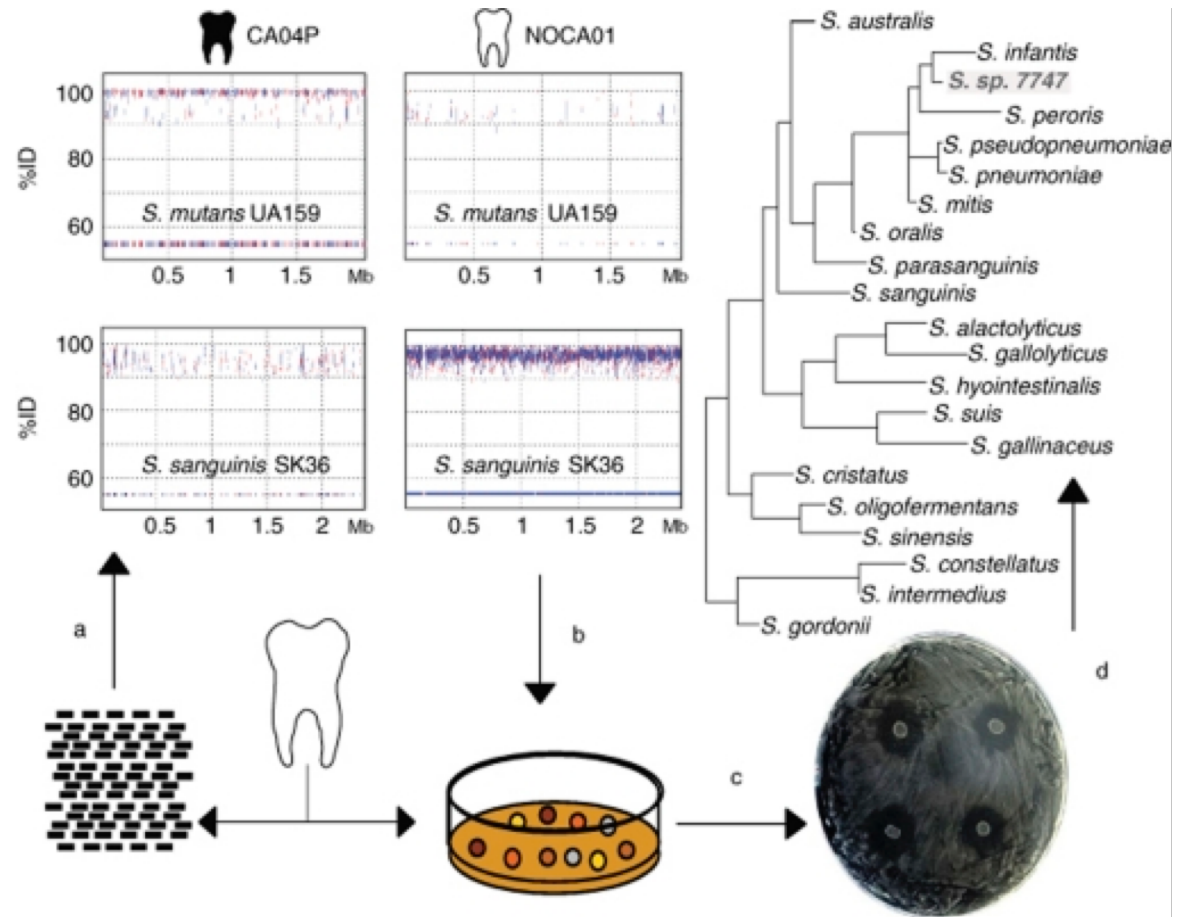
A análise revelou que indivíduos saudáveis possuem flora bucal com excesso de genes para percepção da população (“quorum sensing”) e peptídeos antimicrobianos

Belda-Ferre *et al* (2012) **The oral metagenome in health and disease**. ISME J. 2012 January; 6(1): 46–56.
doi: 10.1038/ismej.2011.85



Metagenômica oral

Os autores procuraram bactérias com atividade anti-cárie entre as bactérias menos frequentes na boca de indivíduos doentes mas abundantes na boca dos saudáveis.



Belda-Ferre *et al* (2012) **The oral metagenome in health and disease.**
 ISME J. 2012 January; 6(1): 46–56. doi: 10.1038/ismej.2011.85

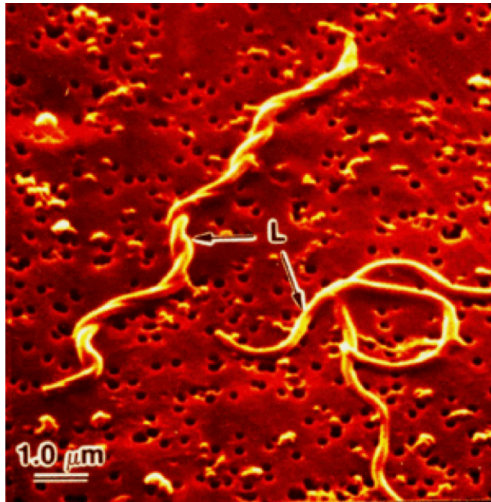
Biologia de sistemas

- A introdução da genômica possibilitou o acesso ao repertório completo de genes (genomas) para organismos modelos
- Os avanços da genômica foram acompanhado pela introdução de tecnologias e procedimentos que permitem a caracterização do estado de muitos, quando não todos, esses genes e seus produtos em um ou mais experimentos.
- Os métodos e procedimentos usados na **obtenção simultânea de dados biológicos em larga escala** recebem o nome um tanto infeliz de “ômicas”
- A disponibilidade de dados em larga escala inspirou a construção de modelos globais do funcionamento da célula, dando origem à área hoje conhecida como biologia de sistemas

Biologia de sistemas: precursores e/ou prérequisitos

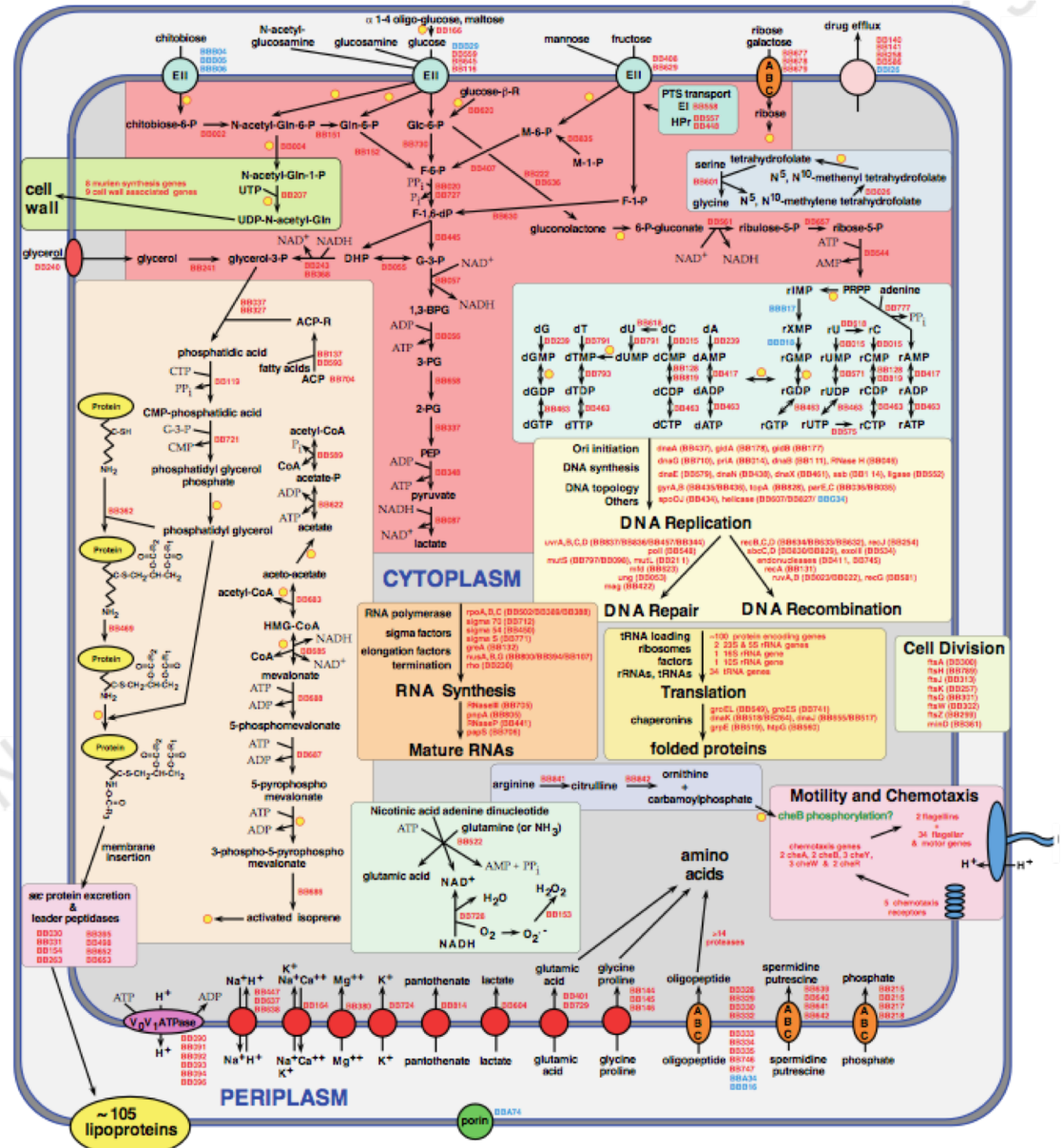
- Transcriptômica: níveis de expressão gênica, ou seja, a quantidade (relativa) de moléculas de RNA para cada gene
- Proteômica: a concentração de todas as proteínas numa dada situação
- Metabolômica: a concentração de todas as pequenas moléculas dentro da célula

Biologia de sistemas modelos globais de vias metabólicas



Borrelia burgdorferi

A sequência do genoma de *B. burgdorferi* possibilitou a Inferência e análise das vias metabólicas e mecanismos de transporte e patogenicidade



Banco de dados KEGG

<http://www.genome.jp/kegg/kegg2.html>

- Um dos melhores repositórios que disponibiliza vias metabólicas para organismos com genomas completos
- Contém mapas, tabelas com gene ortólogos, modelos de reações bioquímicas e outros
- Os dados podem ser baixados em formato compatível com ferramentas de análise usadas em enzimologia e biologia de sistemas

Perguntas

- Que estilo de vida é o mais comum entre bactérias com genoma reduzido?
- Que fenômenos podem alterar a ordem dos genes ao longo da evolução dos organismos?

Demonstração

- Plasmídeo pXAC33 no NCBI
- BLAST de uma proteína
- Breve olhada no KEGG

Referências

- Microbiologia de Brock (13ª edição)
 - Unidade 3: Biologia Molecular e Expressão Gênica
 - **Capítulo 12: Genômica de Microorganismos**
 - Unidade 6: Evolução e diversidade de microorganismos
- Fundamentals of Biochemistry (Voet, 4ª edição)
 - Capítulos 24 a 28
- Genomes 3
 - Texto mais avançado
- Artigos