

# **Bioinformatics**

**CS300**

**Horizontal Gene Transfer  
Database, Tools, Multiseq Alignment**

**Fall 2019**

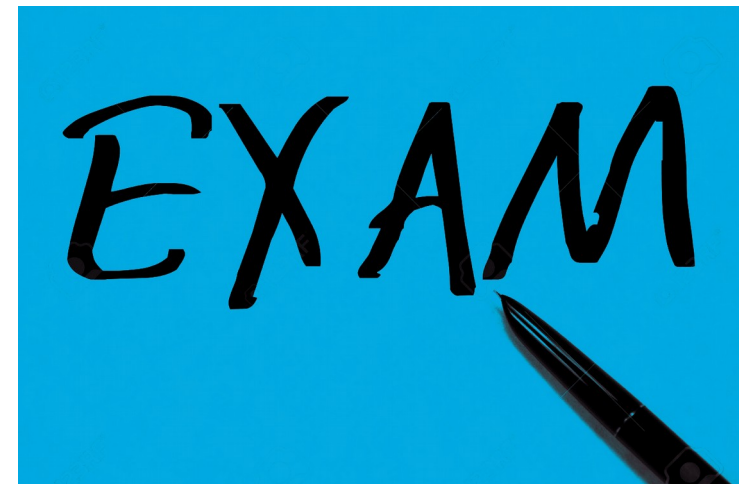
**Oliver BONHAM-CARTER**



# On Exam 1

Thursday, 10 October

- Algorithms: design and implementation in Python
- Basic Python programming: syntax, keywords and definitions
- Tracking influenza by sequence study
- Sequence alignment
- Topics from recent lessons

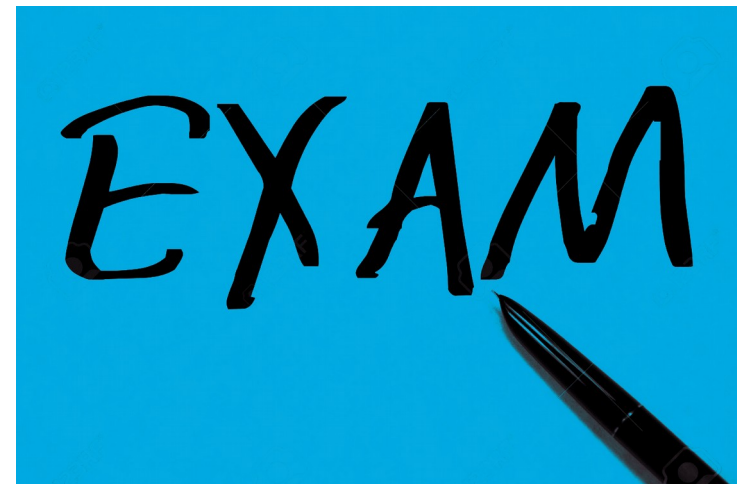




# On Exam 1

Thursday, 10 October

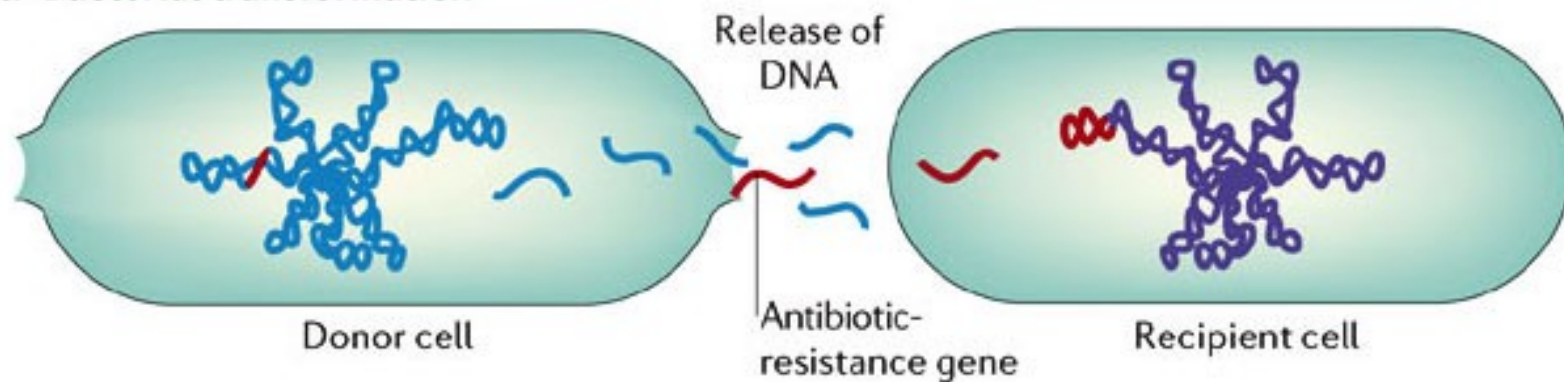
- Questions similar to those on worksheets
- Central Dogma of Biology
  - Transcription, Translation
- Mutations: types, causes and effects
- Genetic disorders and their inheritance
- STUDY YOUR SLIDES!



# Horizontal Gene Transfer

**Horizontal gene transfer (HGT)** refers to the **transfer of genes** between organisms in a manner other than traditional reproduction.

**a Bacterial transformation**





# What is HGT? (Horizontal Gene Transfer)

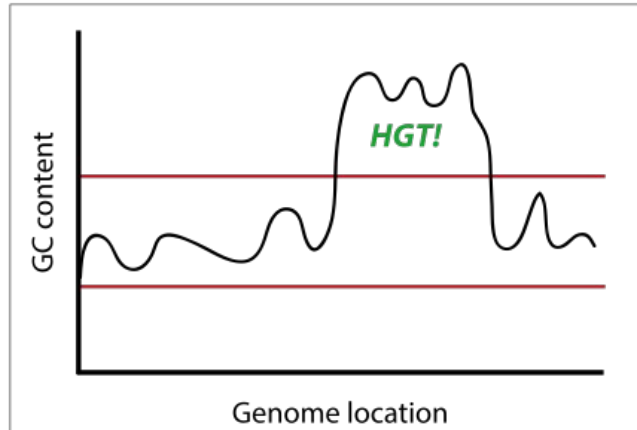
- The transmission of portions of genomic DNA between *distant* organisms
- A process decoupled from vertical inheritance (no mating).
- Various fragments of the genome are the result of different evolutionary histories and come from unrelated organisms.
- This can therefore complicate the investigations of evolutionary relatedness of lineages and species
- Bacteria can acquire new traits such as antibiotic resistance or pathogenic toxins.

# Detection of HGT: Fragments that do not belong

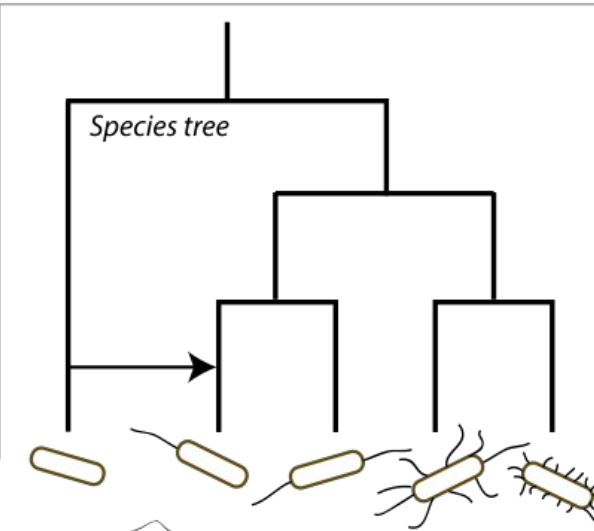
Statistical  
tests:  
Elevated  
GC content  
in DNA  
regions

Species  
that are  
too closely  
related  
to each  
other

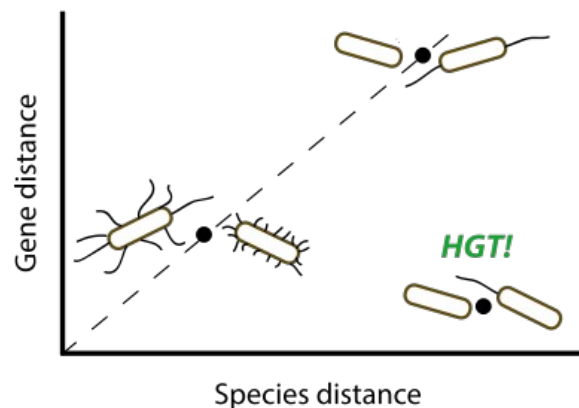
## 1. Parametric methods



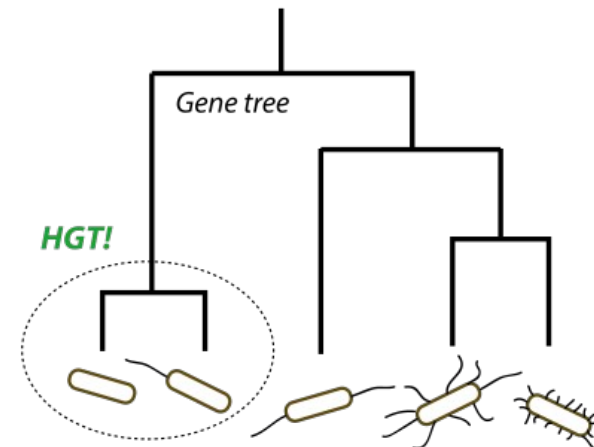
## 2. Phylogenetic methods



### 2a. Implicit phylogenetic methods



### 2b. Explicit phylogenetic methods

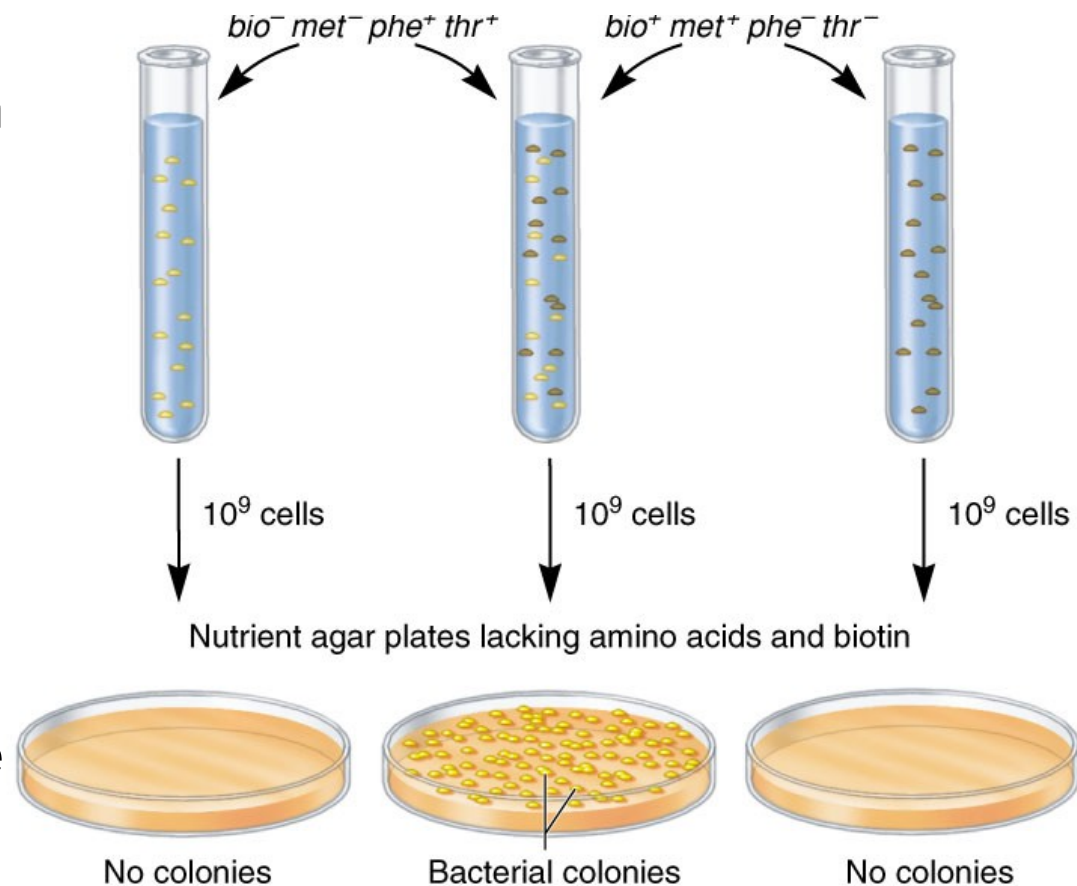


Phylogenetic  
tests using  
relationship  
trees:  
Organisms  
living  
physically near  
each other  
show close  
relations

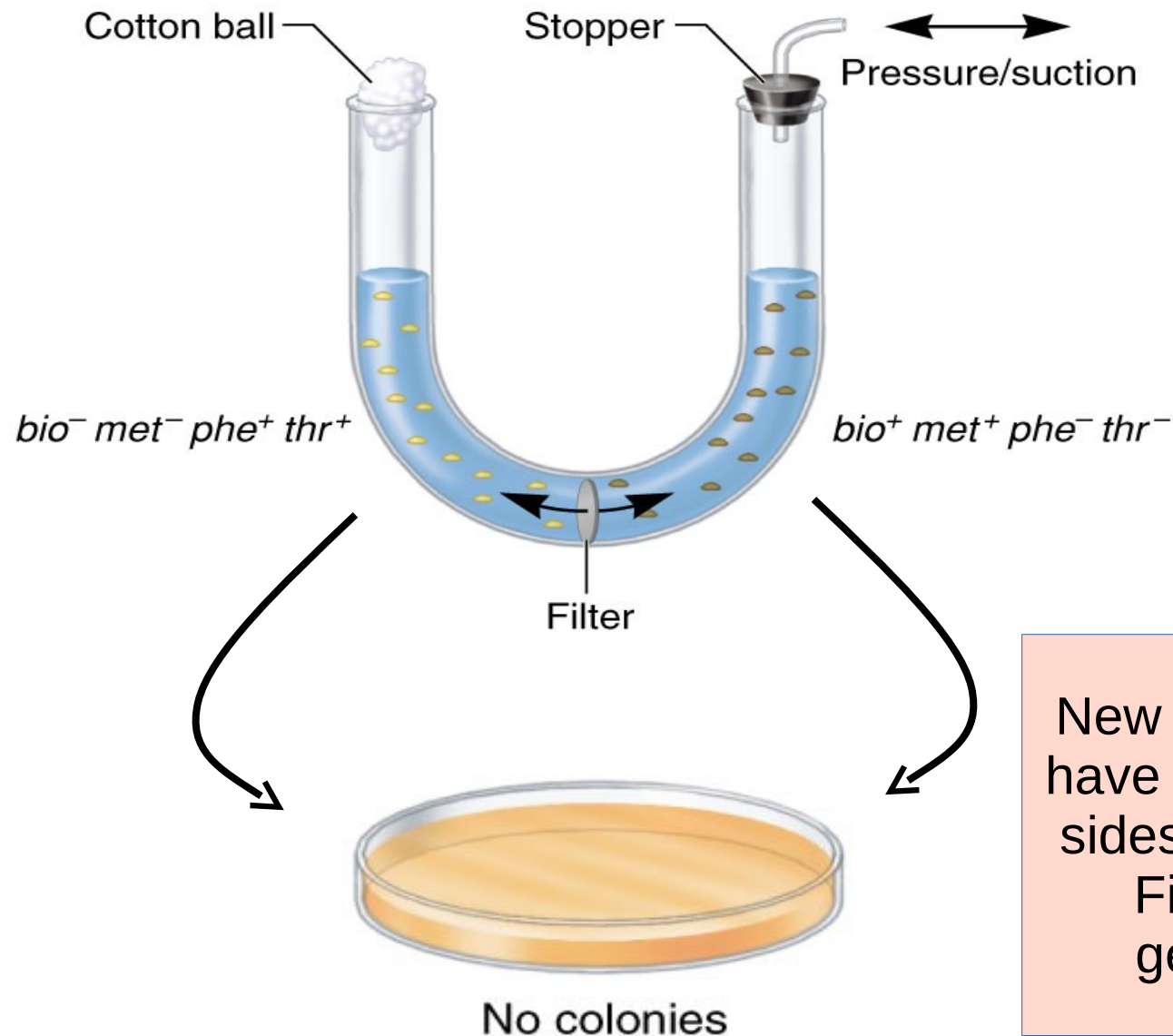
# Genetic Sharing

	<i>bio</i>	<i>met</i>	<i>phe</i>	<i>thr</i>
Strain 1	+	+	-	-
Strain 2	-	-	+	+

- Colonies have opposing growth requirements: each colony requires two extra genes to survive under conditions
- Str 1 needs *phe* and *thr*, Str 2 needs *bio*, *met*.
- In the experiment, colonies randomly share missing genes with each other.
- Cells with HGT survive to make a new colony.



# Transfer of Genes From Physical Contact



New colony needs to have genes from both sides of filter to grow. Filter prevents gene passage



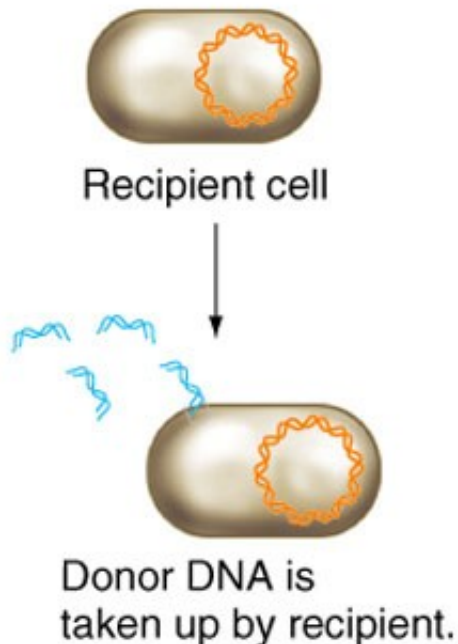
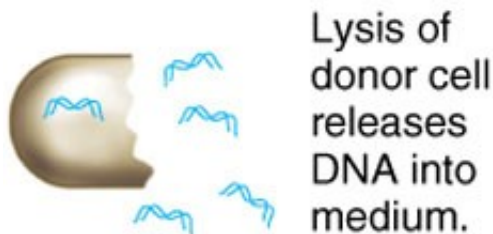


# Mechanisms of DNA Transfer

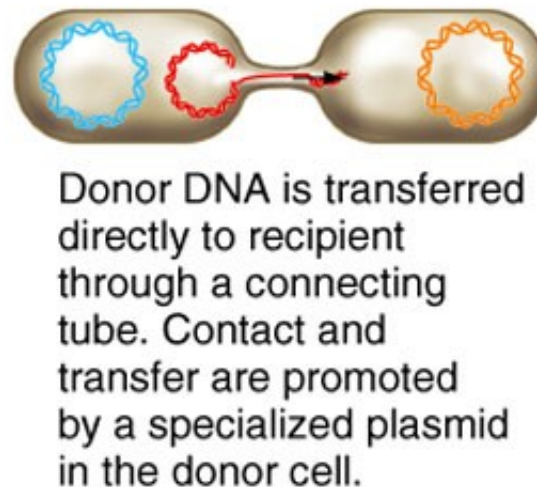
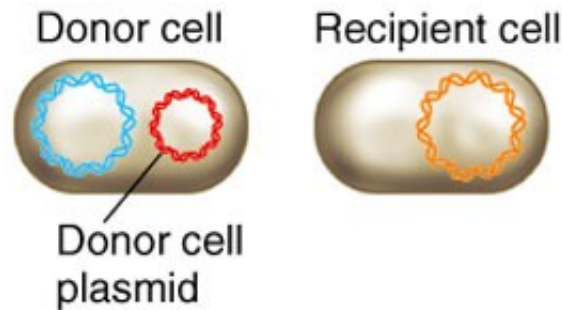
- Conjugation
  - Conjugation is the process by which one bacterium transfers genetic material to another through direct contact.
- Transduction
  - Virus mediated transfer of DNA between bacteria
- Transformation
  - Requires release of DNA into the environment and then the manual take-up of the DNA by bacteria

# Mechanisms of DNA Transfer

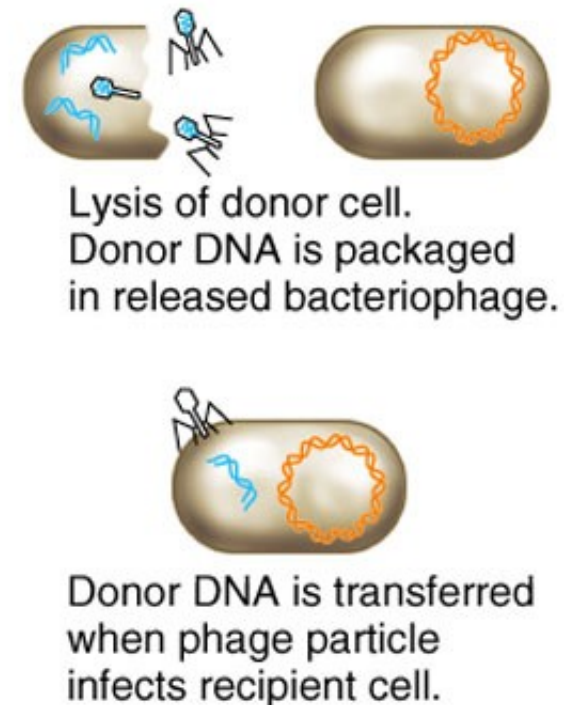
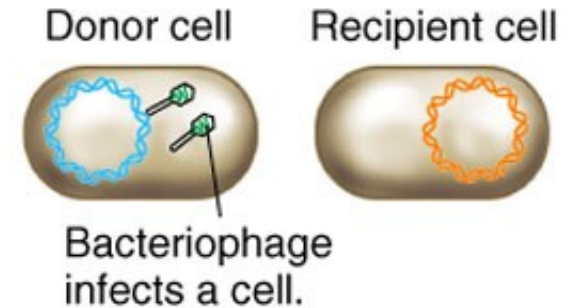
## Transformation



## Conjugation

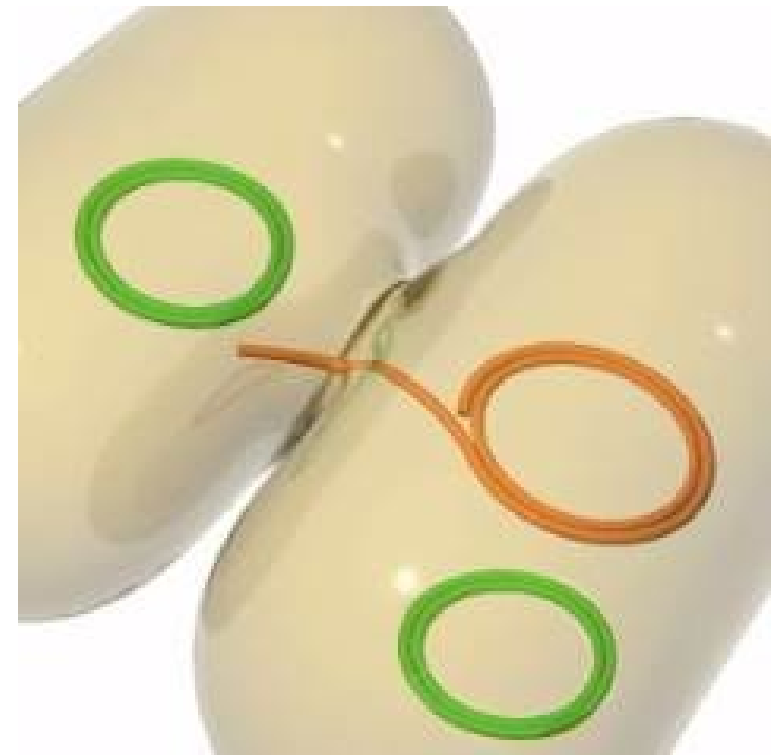


## Transduction



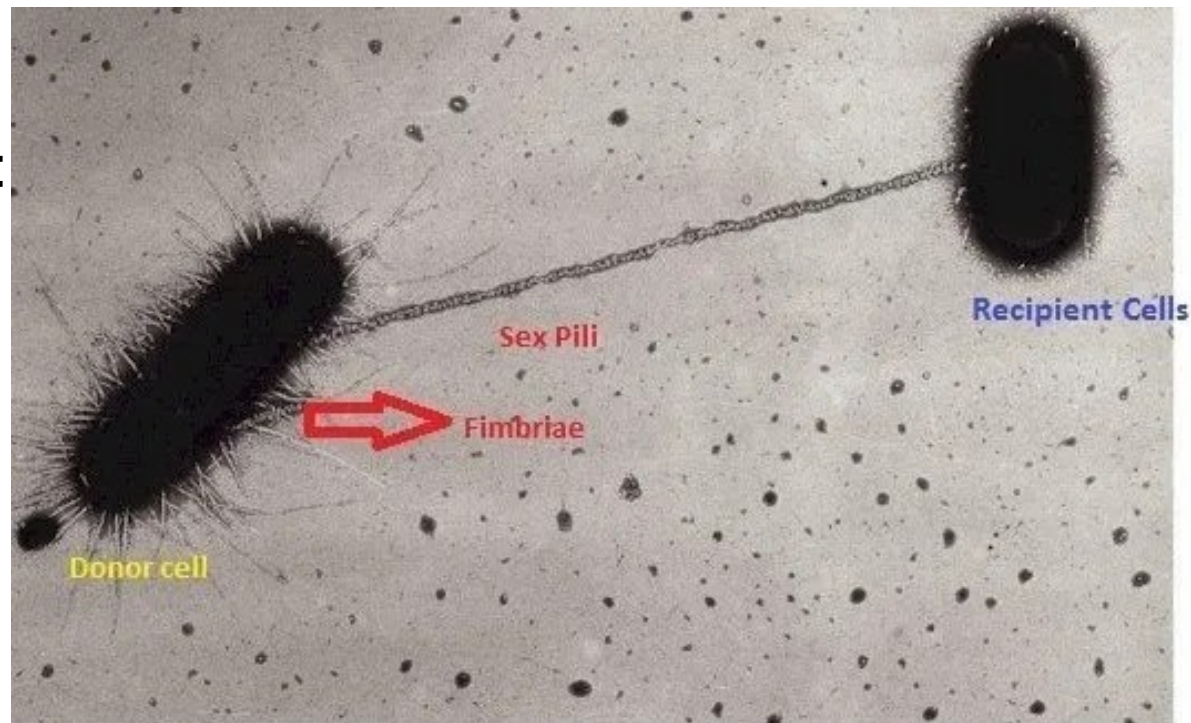
# Bacterial Conjugation

- Only specific bacteria can serve as donors
  - Arber, Werner. "*Horizontal gene transfer among bacteria and its role in biological evolution.*" *Life* 4.2 (2014): 217-224.
- The donor bacterium carries a DNA sequence called the fertility factor, or F-factor.
- The F-factor allows the donor to produce a thin, tubelike structure called a *pilus*, which the donor uses to contact the recipient and transfer genetic content to recipient.



# A Pilus : An Exchange Bridge

- Gene sharing possible if F-Factor present
- Fimbriae or pili (singular: pilus): hair like filaments (tiny hollow projections)
- Extend from the cell membrane into the external environment
- A pilus is composed of subunits of the protein pilin



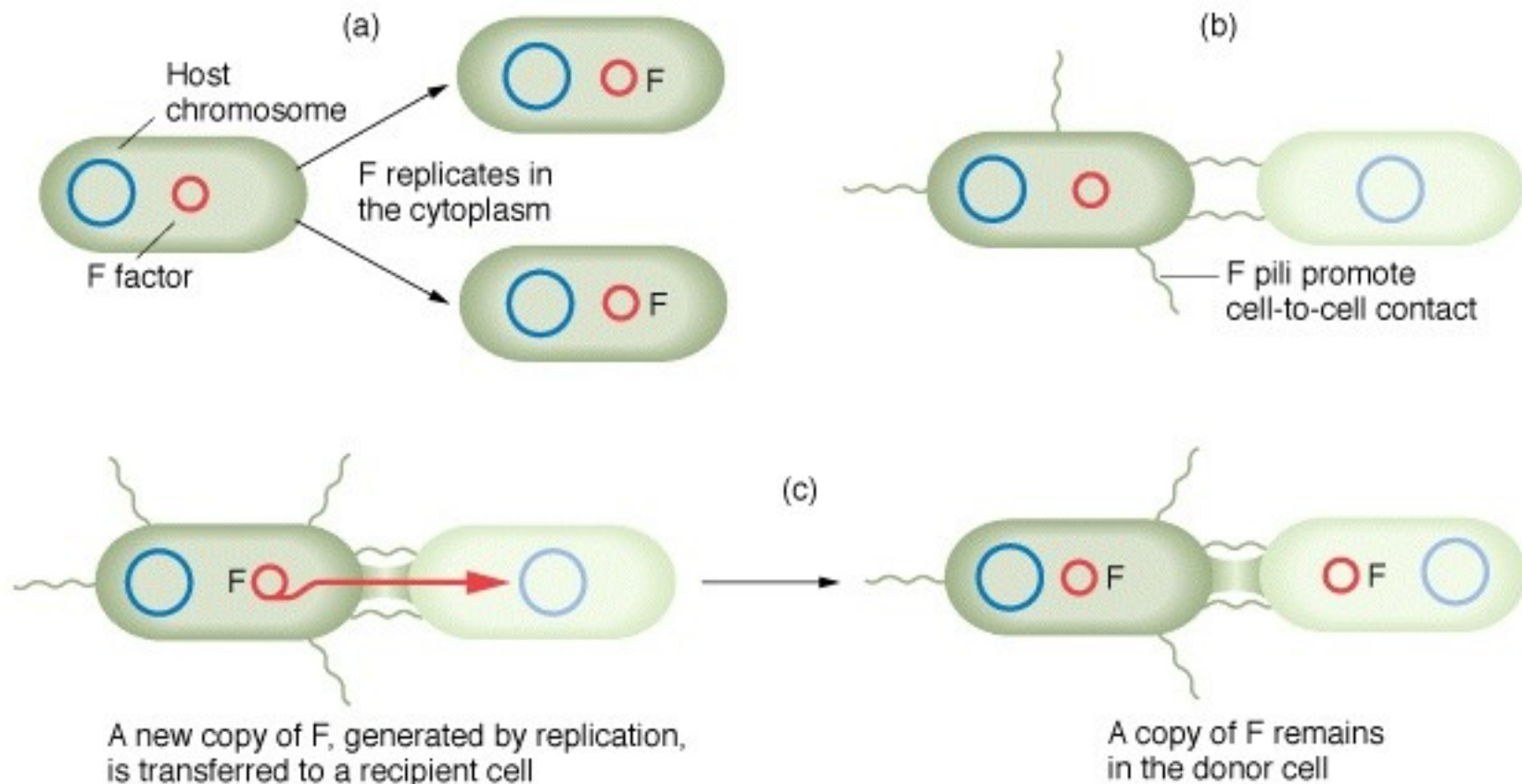
microbeonline.com

TEM

1  $\mu$ m

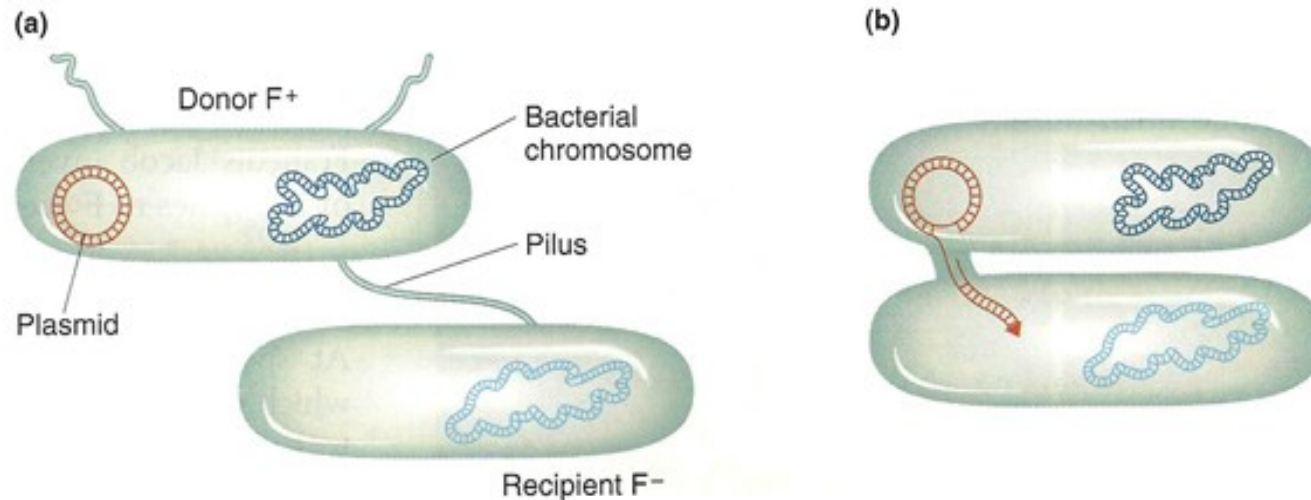
# The F-Factor Material

- **F-Factor:** a material (i.e., the *fertility factor*) that is encoded on the plasmid (bacterial DNA). F<sup>+</sup> (has factor) F<sup>-</sup> (does not have factor)
- Enables the creation of the pilus bridge to transfer circular piece of DNA





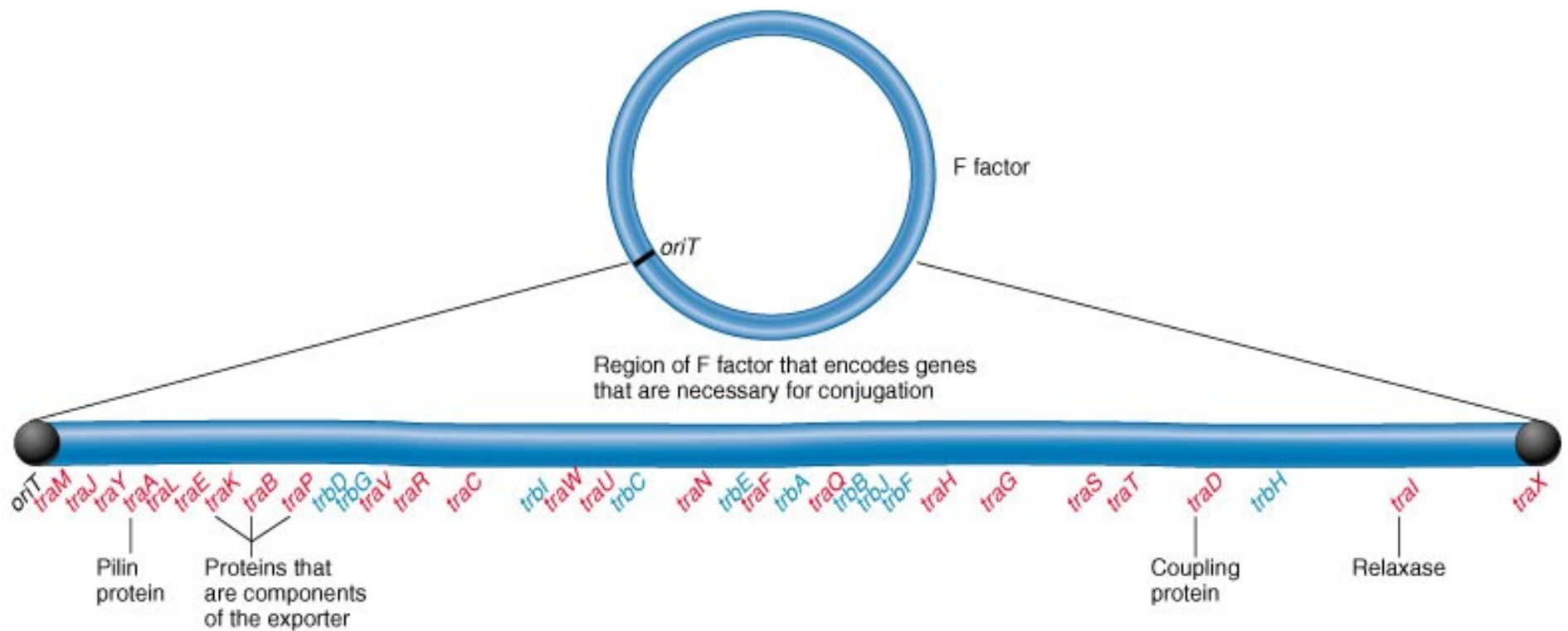
# Transfer of Genetic Material



Plasmids are transferred containing genes of donor to recipient through the pilus

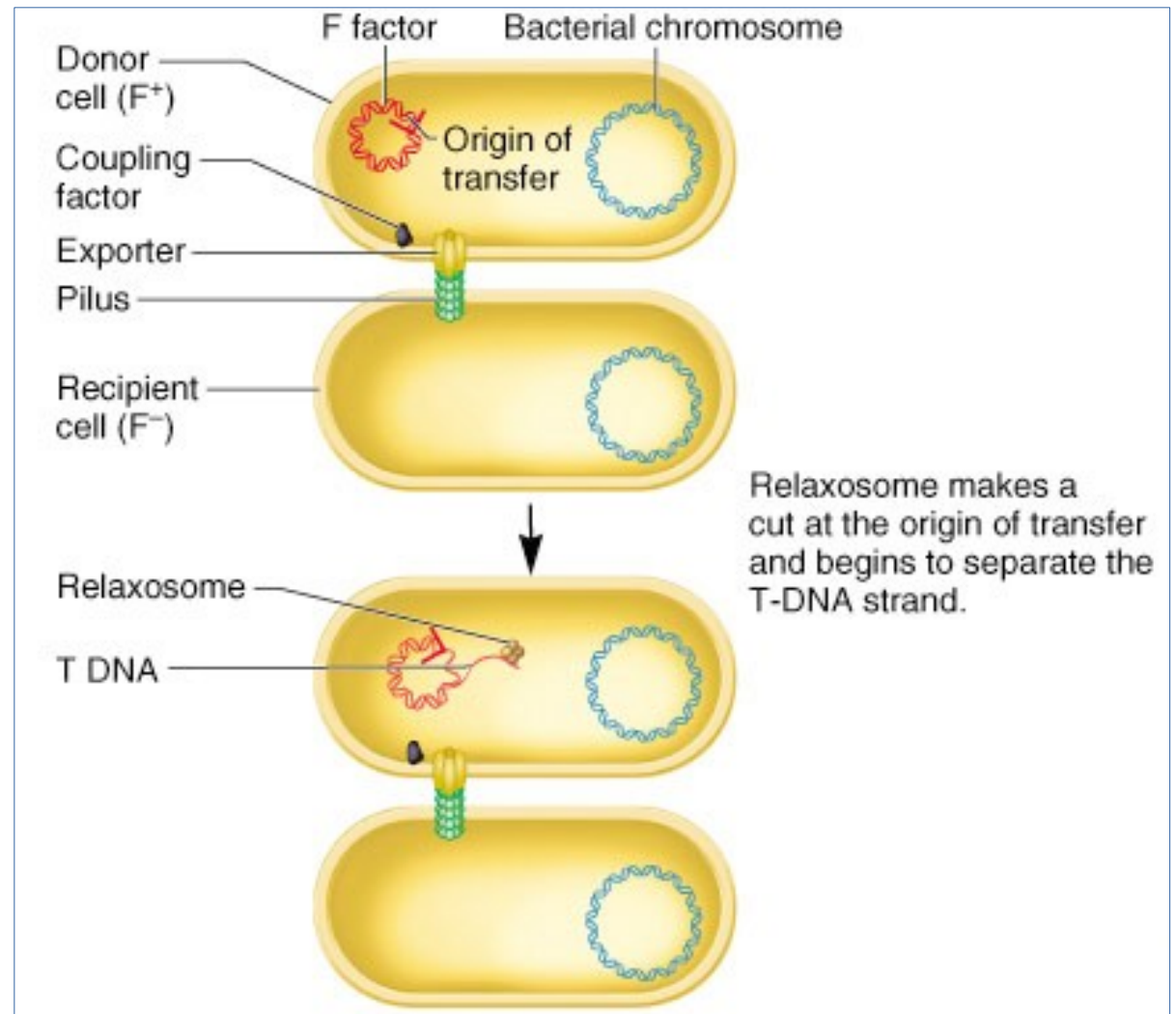


# Plasmids Contain Genes



# Mechanism of Transfer

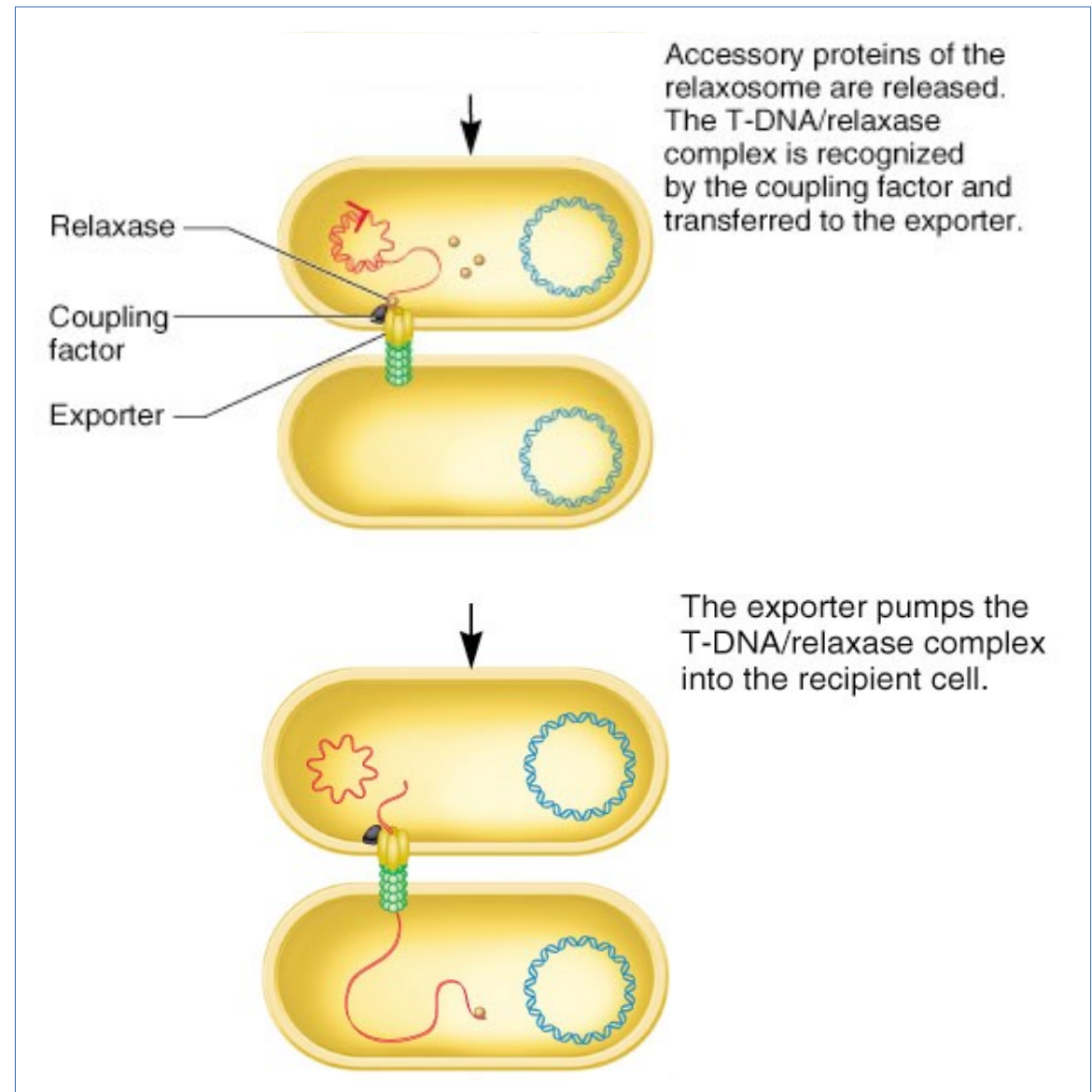
- The relaxosome: a complex of proteins facilitating plasmid transfer during bacterial conjugation
- Relaxosome begins the transfer of T-DNA (transfer DNA) from donor to recipient





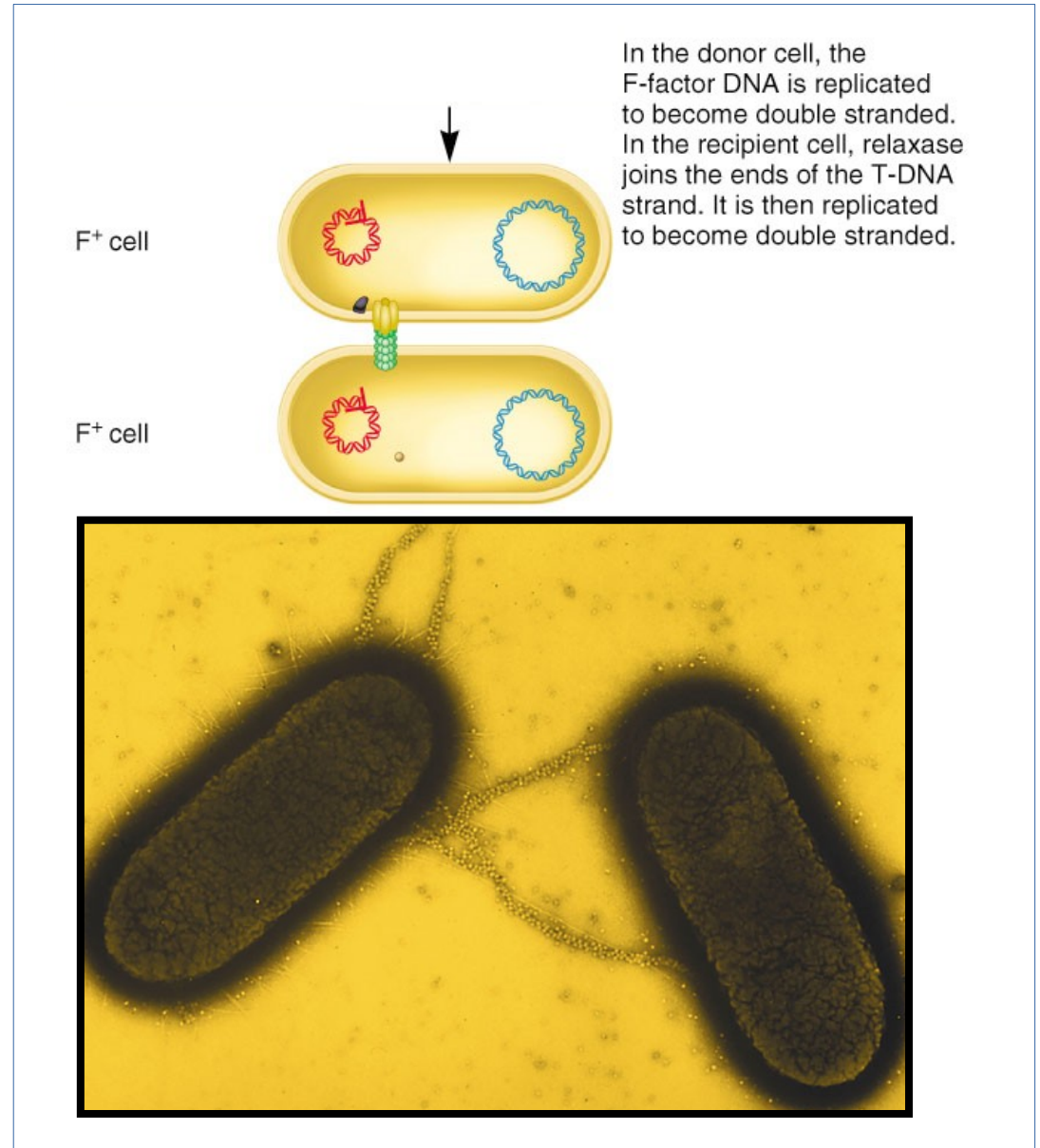
# Mechanism of Transfer

- T-DNA is separated but bound to relaxase protein.
- Complex called nucleoprotein
- Complex is recognized by a coupling factor, fed through exporter (conjugation tube)



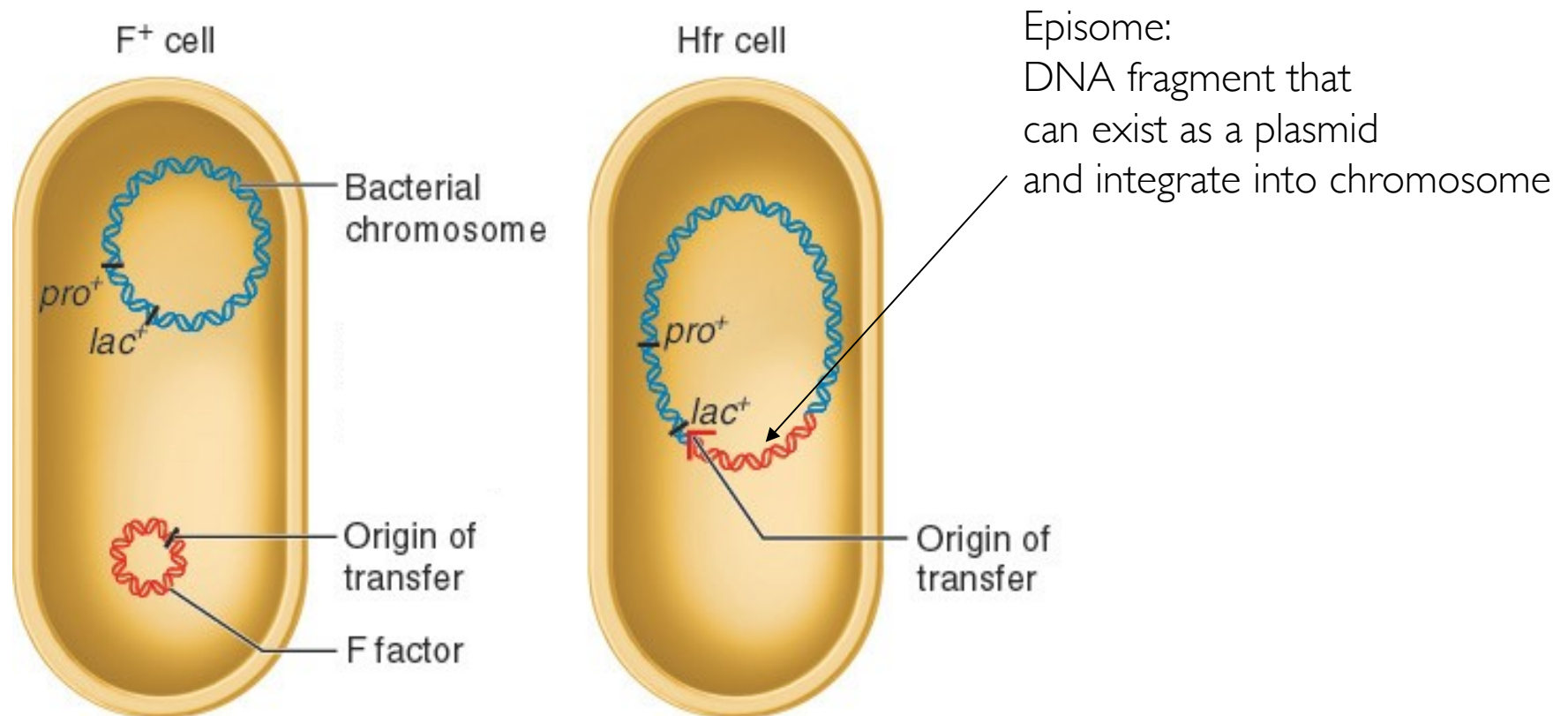
# F Factor Transfer

- Relaxase joins ends to produce circular molecule
- Single strand of F-Factor are in both cells (DNA replication)



# Integration of DNA into Chromosome

- Genes encoded on F-Factor can integrate into host DNA and alter its genotype (genetics) and phenotype (look-like)
- An *Hfr* strain was derived from an *F*<sup>+</sup> strain

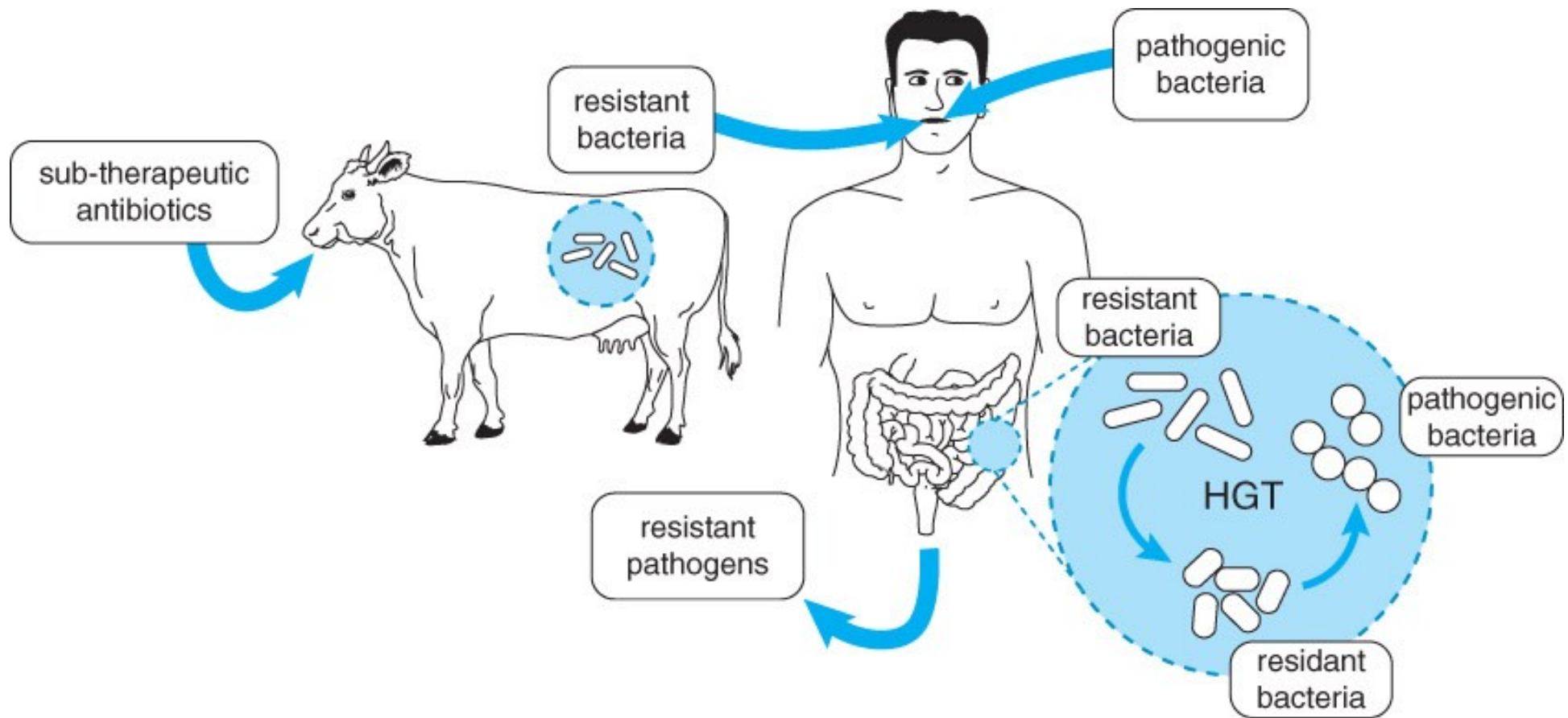




# Known Instances of HGT

- Antibiotic resistance genes on plasmids
- Insertion sequences
- Pathogenicity islands
- Toxin resistance genes on plasmids
- Viruses and viroids
- Organelle to nucleus transfers

# Horizontal Gene Transfer of Antibiotic Resistance

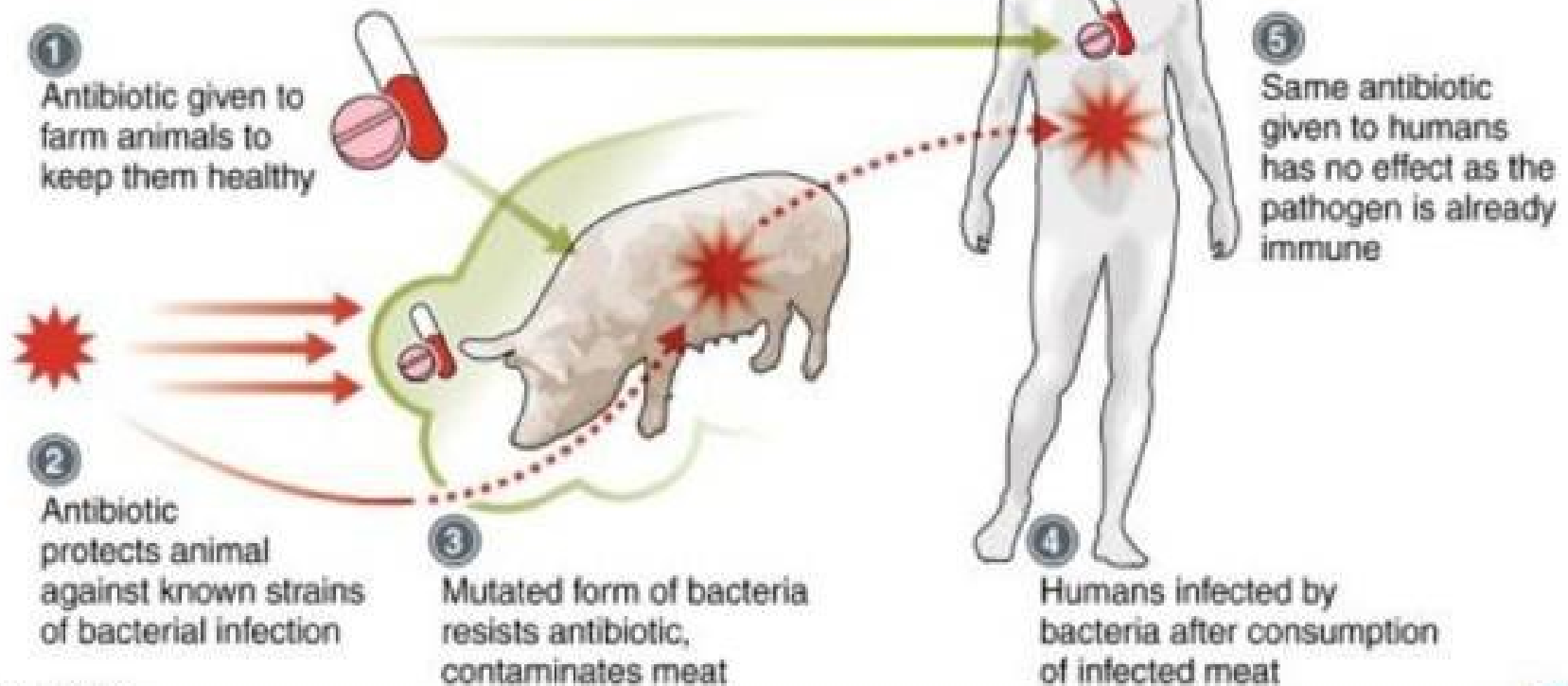


# From Farm-to-Fork Spread of Antibiotic Resistance

## Antibiotic drug abuse

Overuse of antibiotics in agriculture is widening the spread of superbugs that are immune to common drugs

### Path of resistance

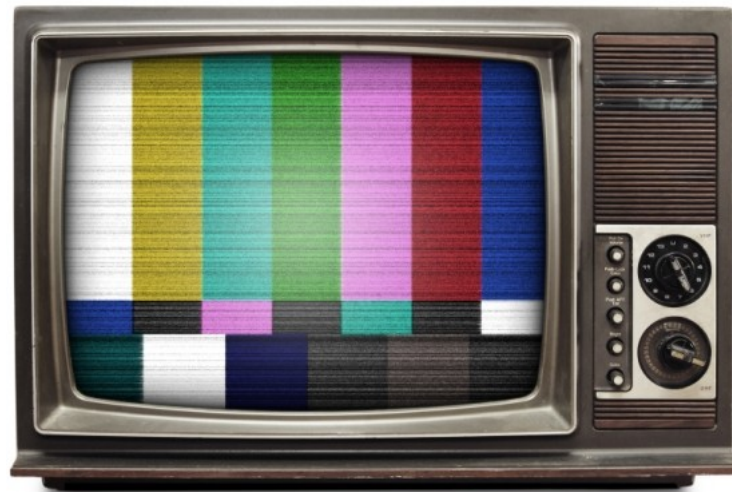


Source: WHO



# Supporting Videos

- Horizontal Gene Transfer animation (1.5mins)
  - <https://www.youtube.com/watch?v=tCZhkxpzxDY>
- Antibiotic Resistance (4mins)
  - <https://ed.ted.com/lessons/how-antibiotics-become-resistant-over-time-kevin-wu>





ALLEGHENY  
COLLEGE



*How do we  
find common  
regions in  
genetic  
material?*



# Multi-Sequence Alignment

- Clustal Omega: MultiSequence Alignment
- <https://www.ebi.ac.uk/Tools/msa/clustalo/>

Where are the  
chunks of  
common  
genes?

Are these  
chunks from  
horizontal gene  
transfers?

Input form | Web services | Help & Documentation

Tools > Multiple Sequence Alignment > Clustal Omega

EMBL-EBI to be HTTPS by default from 1st October

On the 1st October the majority of services hosted on [www.ebi.ac.uk](https://www.ebi.ac.uk) will be served over HTTPS by default. automatically redirect users accessing the site on insecure HTTP URLs to secure HTTPS URLs. Users of EMBL-EBI services may wish to update links, bookmarks or API clients to use the HTTPS URLs.

Results for job clustalo-l20171005-050706-0613-19128639-oy

Alignments | Result Summary | Phylogenetic Tree | Submission Details

Download Alignment File | Send to Simple\_Phylogeny

CLUSTAL O(1.2.4) multiple sequence alignment

```

AB011005.1  -----
X95927.1    -----GGGAGG-CAAGAATCAGGC-CTCAAAACCCCTCAAACCTCTA
X95928.1    TATTATTTTCATGAAACTAATGTGAGGAGAAAAATTAGGCAGACGTATAAGTGACAGTAGA
KU325497.1  -----
KU325498.1  -----

AB011005.1  -----
X95927.1    ACCTTGGAACAAACAGCCCTAGTGGGATAGCCCTATCAAACCCCTTCCTTC--AGGGATT
X95928.1    -----

```



# Clustal: Load Input Sequences

## Multiple Sequence Alignment

Clustal Omega is a new multiple sequence alignment program that uses seeded guide trees and HMM profile-profile techniques to generate alignments between **three or more** sequences. For the alignment of two sequences please instead use our [pairwise sequence alignment tools](#).

**Important note:** This tool can align up to 4000 sequences or a maximum file size of 4 MB.

### STEP 1 - Enter your input sequences

Enter or paste a set of

DNA

sequences in any supported format:

Or, upload a file:  No file chosen



# Using Example DNA Sequences

Results for job clustalo-E20191003-040355-0978-26932726-p1m

[Alignments](#)[Result Summary](#)[Guide Tree](#)[Phylogenetic Tree](#)[Results Viewers](#)[Submission Details](#)[Download Alignment File](#)

CLUSTAL O(1.2.4) multiple sequence alignment

```
test1      ATGAGTCTCTCTGATAAGGACAAGGCTGCTGTGAAAGCCCTATGG----- 45
test2      -----CTGTCTCCTGCCGACAAGACCAACGTCAAGGCCGCCTGGGGTAAG----- 45
test3      -----ACAAAAGCAACATCAAGGCTGCCTGGGGGAAGATTGGTGGC 41
              *****          *  *  *  *          ***
```

```
test1      ---- 45
test2      ---- 45
test3      CATG 45
```



# Using Example Protein Sequences

Results for job clustalo-E20191003-040019-0643-26585412-p1m

Alignments

Result Summary

Guide Tree

Phylogenetic Tree

Results Viewers

Submission Details

Download Alignment File

Show Colors

CLUSTAL O(1.2.4) multiple sequence alignment

```
sp|P69905|HBA_HUMAN      MVLSPADKTNVKAAWGKVGAGHAGEYGAEALERMFLSFPTTKTYFPHFDLSHGSAQVKGHG  60
sp|P01942|HBA_MOUSE      MVLSGEDKSNIAAWGKIGGHGAEYGAEALERMFASFPPTTKTYFPHFDVSHGSAQVKGHG  60
sp|P13786|HBAZ_CAPHI     MSLTRTERTIILSLWSKISTQADVIGTETLERLFSCYPQAKTYFPHFDLHSGSAQLRAHG  60
* *:  ::: : : *.*:.. ..  *:*:***:* .:* :*****:  ****:..**

sp|P69905|HBA_HUMAN      KKVADALTNAVAHVDDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPAEFTP  120
sp|P01942|HBA_MOUSE      KKVADALASAAGHLDDLPGALSALSDLHAHKLRVDPVNFKLLSHCLLVTLASHHPADFTP  120
sp|P13786|HBAZ_CAPHI     SKVVAAVGDAVKSIDNVTSALSKLSELHAYVLRVDPVNFKFLSHCLLVTLASHFPADFTA  120
.***. *: .*.  :::.*** **:***: *****:*****:* **:**

sp|P69905|HBA_HUMAN      AVHASLDKFLASVSTVLTSKYR      142
sp|P01942|HBA_MOUSE      AVHASLDKFLASVSTVLTSKYR      142
sp|P13786|HBAZ_CAPHI     DAHAAWDKFLSIVSGVLTEKYR      142
.***: *****: ** ***.***
```

Results took about five minutes to get! :-/  
Is there a way to get faster results?

# BLAST: A Heuristic Approach to Database Searching

- <https://blast.ncbi.nlm.nih.gov/Blast.cgi>
- Basic Local Alignment Search Tool
  - NCBI – DNA and protein sequence
  - Compares one sequence to database of > 100 million
  - Finds best hits (optimal alignments) in a matter of seconds
  - (would take >3 years using Needleman-Wunsch algorithm)





# BLAST: A Heuristic Approach to Database Searching

**NIH** U.S. National Library of Medicine **NCBI** National Center for Biotechnology Information

**BLAST**<sup>®</sup> >> blastn suite [Home](#) [Recent Results](#)

*Escherichia coli* O157:H7 str. Sakai Nucleotide BLAST

[blastn](#) [blastp](#) [blastx](#) [tblastn](#)

Enter Query Sequence

BLASTN programs search nucleotide databases using a nucleotide query. [more...](#)

Enter accession number(s), gi(s), or FASTA sequence(s) [?](#) [Clear](#)

atatatatatatatatatat

Query subrange [?](#)

From

To

Or, upload file  No file chosen [?](#)

Job Title

Enter a descriptive title for your BLAST search [?](#)

☐ Align two or more sequences [?](#)

Where does my sequence show up in *other organisms*?



# BLAST: Output

[Download](#) ▼

[GenBank](#) [Graphics](#)

Sort by: E value ▼

▼ [Next](#) ▲ [Previous](#)

## **Tepidimicrobium xylanilyticum strain DSM 23310, whole genome shotgun sequence**

Sequence ID: [NZ\\_FNNG01000007.1](#) Length: 162671 Number of Matches: 4

**Range 1: 102686 to 102707** [GenBank](#) [Graphics](#)

▼ [Next Match](#) ▲ [Previous Match](#)

Score	Expect	Identities	Gaps	Strand
44.1 bits(22)	0.005	22/22(100%)	0/22(0%)	Plus/Plus
Query 1	ATATATATATATATATATATATAT			22
Sbjct 102686	ATATATATATATATATATATATAT			102707

**Range 2: 102686 to 102707** [GenBank](#) [Graphics](#)

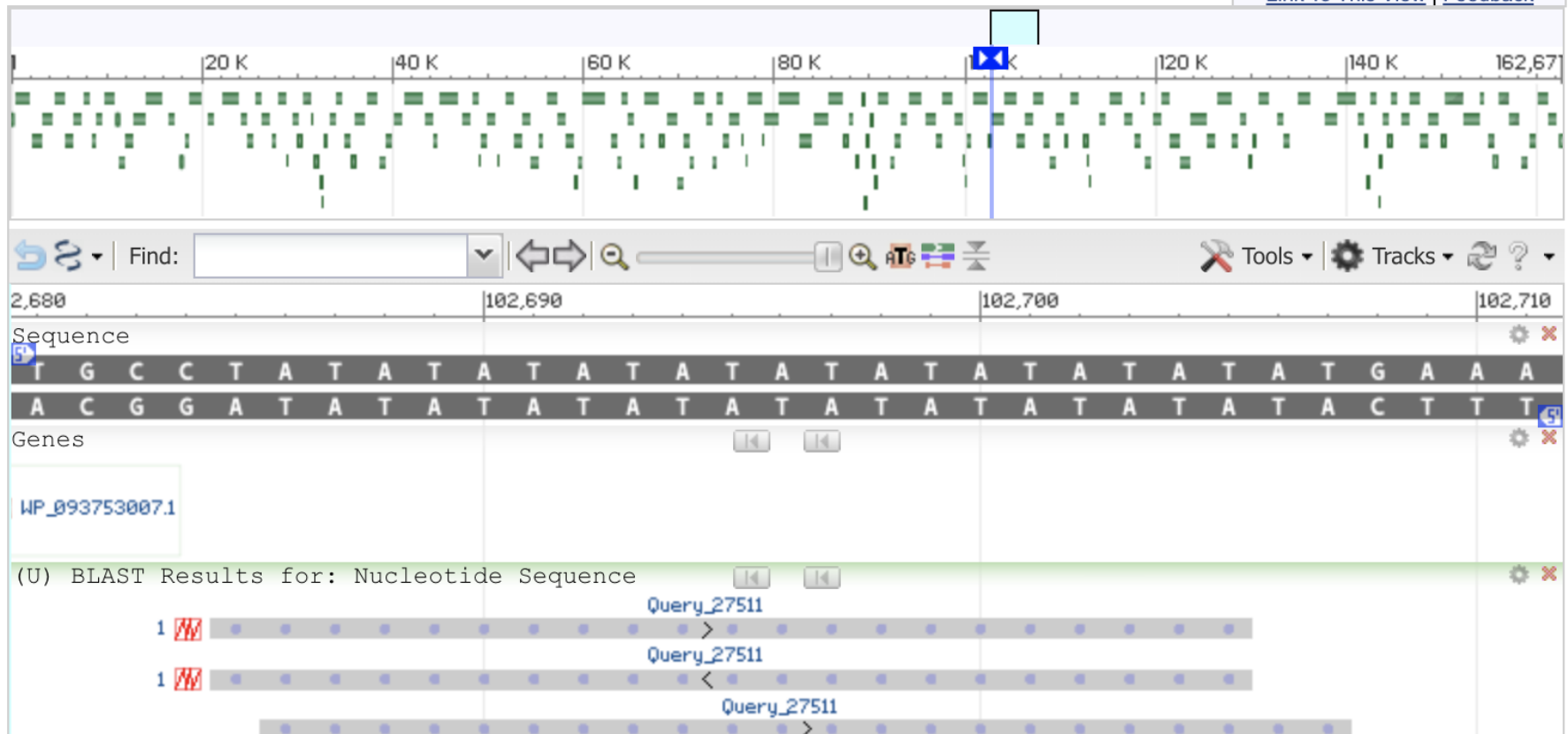
▼ [Next Match](#) ▲ [Previous Match](#) ▲ [First Match](#)

Score	Expect	Identities	Gaps	Strand
44.1 bits(22)	0.005	22/22(100%)	0/22(0%)	Plus/Minus
Query 1	ATATATATATATATATATATATAT			22
Sbjct 102707	ATATATATATATATATATATATAT			102686

**Range 3: 102685 to 102705** [GenBank](#) [Graphics](#)

▼ [Next Match](#) ▲ [Previous Match](#) ▲ [First Match](#)

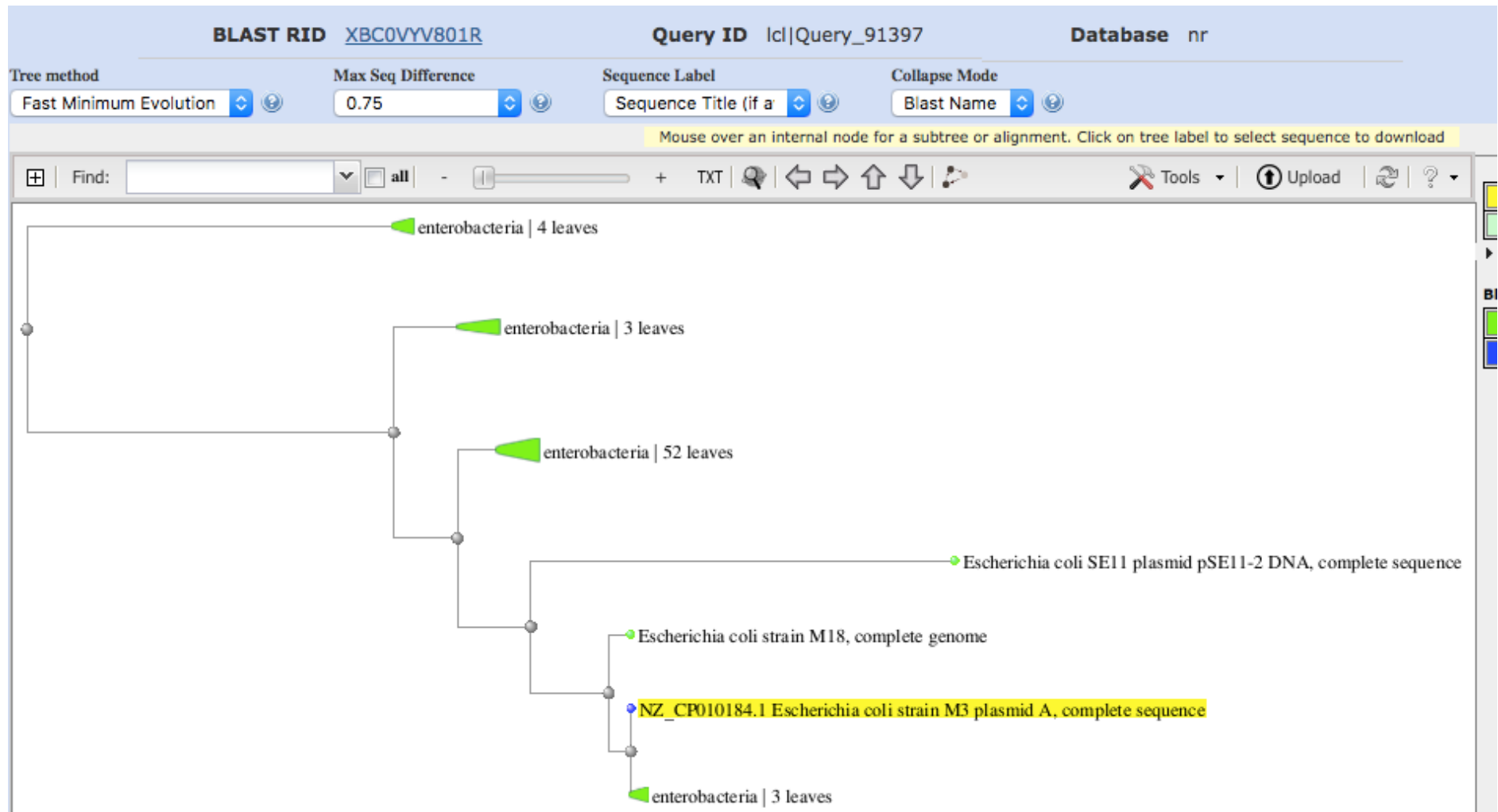
Score	Expect	Identities	Gaps	Strand
42.1 bits(21)	0.019	21/21(100%)	0/21(0%)	Plus/Plus
Query 2	TATATATATATATATATATATAT			22
Sbjct 102685	TATATATATATATATATATATAT			102705







# BLAST: Tree of Relations





# BLAST Demo

## Cyprinus carpio SIRT1 mRNA, partial cds

GenBank: KF881970.1

[FASTA](#) [Graphics](#)

Go to: ☐

LOCUS KF881970 375 bp mRNA linear VRT 06-NOV-2014  
DEFINITION Cyprinus carpio SIRT1 mRNA, partial cds.  
ACCESSION KF881970  
VERSION KF881970.1  
KEYWORDS .  
SOURCE Cyprinus carpio (common carp)  
ORGANISM [Cyprinus carpio](#)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;  
Cypriniformes; Cyprinidae; Cyprinus.  
REFERENCE 1 (bases 1 to 375)  
AUTHORS Fuentes,E.N., Zuloaga,R., Nardocci,G., Fernandez de la Reguera,C.,  
Simonet,N., Fumeron,R., Valdes,J.A., Molina,A. and Alvarez,M.  
TITLE Skeletal muscle plasticity induced by seasonal acclimatization in  
carp involves differential expression of rRNA and molecules that  
epigenetically regulate its synthesis  
JOURNAL Comp. Biochem. Physiol. B, Biochem. Mol. Biol. 172-173, 57-66  
(2014)  
PUBMED [24769445](#)  
REFERENCE 2 (bases 1 to 375)  
AUTHORS Fernandez de la Reguera,C. and Alvarez,M.  
TITLE Isolation of the Sirtuin 1 (Sirt1) in the common carp

Customize view

Analyze this se  
Run BLAST

Pick Primers

Highlight Sequenc

Find in this Seque

Articles about t  
LOC109064745

Skeletal muscle p  
se [Comp Biocher

Reference sequ  
information

RefSeq mRNA  
See reference mF  
LOC109064745 g

**Sirt1** (member of the sirtuin family) is a nicotinamide adenosine dinucleotide (NAD)-dependent deacetylase that removes acetyl groups from various proteins. Sirt1 performs a wide variety of functions in biological systems.



# BLAST Demo

NIH U.S. National Library of Medicine NCBI Sign in to NCBI

**BLAST**® >> blastn suite Home Recent Results Saved Strategies Help

Standard Nucleotide BLAST

blastn blastp blastx tblastn tblastx

BLASTN programs search nucleotide databases using a nucleotide query. [more...](#) [Reset page](#) [Bookmark](#)

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) [Clear](#)

KF881970.1

Query subrange [?](#)

From

To

Or, upload file

Choose File No file chosen [?](#)

Job Title



Enter a descriptive title for your BLAST search [?](#)

☐ Align two or more sequences [?](#)

We are checking the occurrence of this RNA sequence, **Cyprinus carpio SIRT1 mRNA**, partial cds GenBank: **KF881970.1**) in many other sequences to determine common parts of code.



# BLAST Demo

Job Title	gb KF881970.1		
RID	<a href="#">TRT19U51014</a>	Search expires on 10-09 09:19 am	<a href="#">Download All</a> ▼
Program	BLASTN 	<a href="#">Citation</a> ▼	
Database	refseq_rna	<a href="#">See details</a> ▼	
Query ID	<a href="#">KF881970.1</a>		
Description	Cyprinus carpio SIRT1 mRNA, partial cds.		
Molecule type	rna		
Query Length	375		
Other reports	<a href="#">Distance tree of results</a>	<a href="#">MSA viewer</a>	

We get basic information from the search...



# BLAST Demo

<input checked="" type="checkbox"/> select all 28 sequences selected		<a href="#">GenBank</a>	<a href="#">Graphics</a>	<a href="#">Distance tree of results</a>			
	Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
<input checked="" type="checkbox"/>	<a href="#">Cyprinus carpio SIRT1 mRNA, partial cds</a>	693	693	100%	0.0	100.00%	<a href="#">KF881970.1</a>
<input checked="" type="checkbox"/>	<a href="#">PREDICTED: Cyprinus carpio NAD-dependent protein deacetylase sirtuin-1-like (LOC1</a>	654	654	100%	0.0	98.13%	<a href="#">XM_019081783.1</a>
<input checked="" type="checkbox"/>	<a href="#">PREDICTED: Sinocyclocheilus rhinoceros NAD-dependent protein deacetylase sirtuin</a>	621	621	100%	1e-174	96.53%	<a href="#">XM_016538446.1</a>
<input checked="" type="checkbox"/>	<a href="#">PREDICTED: Sinocyclocheilus grahami NAD-dependent protein deacetylase sirtuin-1-li</a>	616	616	100%	5e-173	96.27%	<a href="#">XM_016247026.1</a>
<input checked="" type="checkbox"/>	<a href="#">PREDICTED: Carassius auratus NAD-dependent protein deacetylase sirtuin-1-like (LO</a>	604	604	100%	1e-169	95.73%	<a href="#">XM_026277791.1</a>
<input checked="" type="checkbox"/>	<a href="#">PREDICTED: Cyprinus carpio NAD-dependent protein deacetylase sirtuin-1-like (LOC1</a>	604	604	100%	1e-169	95.73%	<a href="#">XM_019107331.1</a>
<input checked="" type="checkbox"/>	<a href="#">PREDICTED: Sinocyclocheilus anshuiensis NAD-dependent protein deacetylase sirtuin</a>	604	604	100%	1e-169	95.73%	<a href="#">XM_016487894.1</a>
<input checked="" type="checkbox"/>	<a href="#">PREDICTED: Sinocyclocheilus anshuiensis NAD-dependent protein deacetylase sirtuin</a>	593	593	100%	2e-168	95.20%	<a href="#">XM_016487300.1</a>

Cases (sequences) in which the subsequence is found in other sequences



# BLAST Demo

**PREDICTED: Anabas testudineus sirtuin 1 (sirt1), mRNA**

Sequence ID: [XM\\_026356561.1](#) Length: 2767 Number of Matches: 1

Range 1: 1074 to 1442 [GenBank](#) [Graphics](#)

▼ [Next Match](#) ▲ [Previous Match](#)

Score	Expect	Identities	Gaps	Strand
252 bits(136)	2e-63	295/373(79%)	5/373(1%)	Plus/Plus
Query 1	TGCGACTGCCTCCTGTCTTATCTGTAAACATAAGGTTGACTGTGAGGCCGTAA			
Sbjct 1074	TGCTACTGCATCATGTCTTGTCTGTAAACACAAAGTGGATTGTGAGGCTATAA			
Query 61	TATATTCAACCAGGTTGTTCTCCTCACTGTCCCAGGTGTCCGGCAGACGTCCCGT			
Sbjct 1134	CATCTTTAACCAGATTGTCCCTCATTGTCCACGATGT-C--CAGATATTCC-T			
Query 120	TCATGAAACCAGACATCGTCTTCTTTGGCGAGAACCTTCCAGAGTTTTTCCAC			
Sbjct 1190	TCATGAAACCAGACATTGTCTTCTTTGGAGAGAATCTTCCAGAAATGTTTCAC			
Query 180	TGAAGCAGGATAAAGATGAGGTGGATCTTCTCATCGTGATCGGCTCCTCGCTG			
Sbjct 1250	TGAAGCAGGATAAAGATGAGGTGGACCTCTTGATTGTCATTGGCTCTTCACTT			
Query 240	GGCCAGTGGCTCTCATACCCAGCTCTATACCTCATGACGTGCCTCAAGTCCTG			
Sbjct 1310	GACCAGTTGCCCTGATCCCTAACTCCATTCCTCATGAAGTGCCTCAGGTCCTG			
Query 300	GCGAGCCGCTGCCGCATCTGAACTTCGACGTGGAGCTGCTCGGAGACTGTGAC			
Sbjct 1370	GGGAGCAGCTGCCTCACCTCAACTTTGATGTGGAGCTACTTGGTGATTGTGAT			
Query 360	TGAACGAACTCTG 372			
Sbjct 1430	TAAACGAGCTCTG 1442			





ALLEGHENY  
COLLEGE

# BLAST Tutorial: DigitalWorldBiology



<https://digitalworldbiology.com/tutorial/blast-for-beginners>