Bioinformatics

CS300
Horizontal Gene Transfer
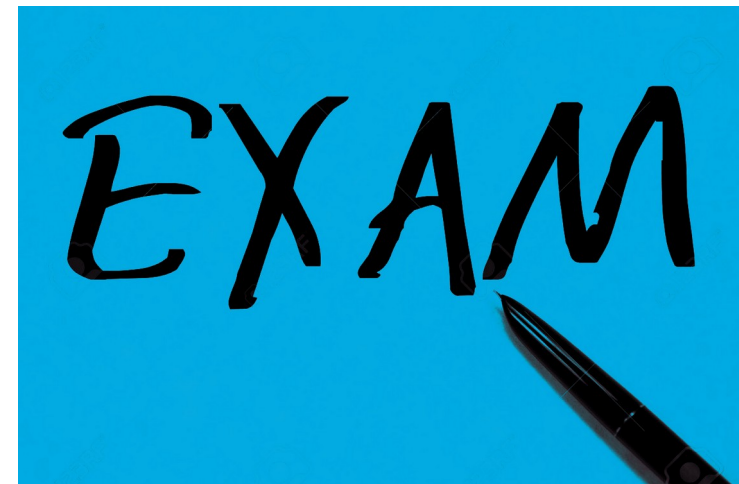## Database, Tools, Multiseq Alignment

Spring 2021
Oliver BONHAM-CARTER

Exam 1
During lab Monday, 12 April

- Differences between DNA and RNA
- Basic Python programming: syntax, keywords and definitions (covered in class)
- Global and local alignment
- Terms and definitions

Exam 1
During lab Monday, 12 April

- Some questions are similar to those on activities
- Central Dogma of Biology
  - Transcription, Translation
- Detecting genetic disorders
- STUDY YOUR SLIDES!

Exam 1
During lab Monday, 12 April

- Multiple choice, Short answer, Short essay,

- Given online on Sakai's "*Tests and Quizzes"* Tab at the beginning of lab.

- You will have 24 hours to take the exam starting from Monday's lab (2:50pm)

- The exam will be due on the following day at 2:50pm and will likely take about two hours to complete during that time.
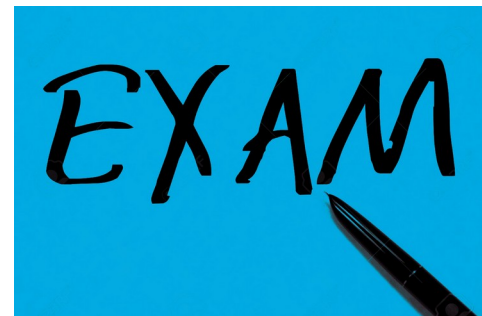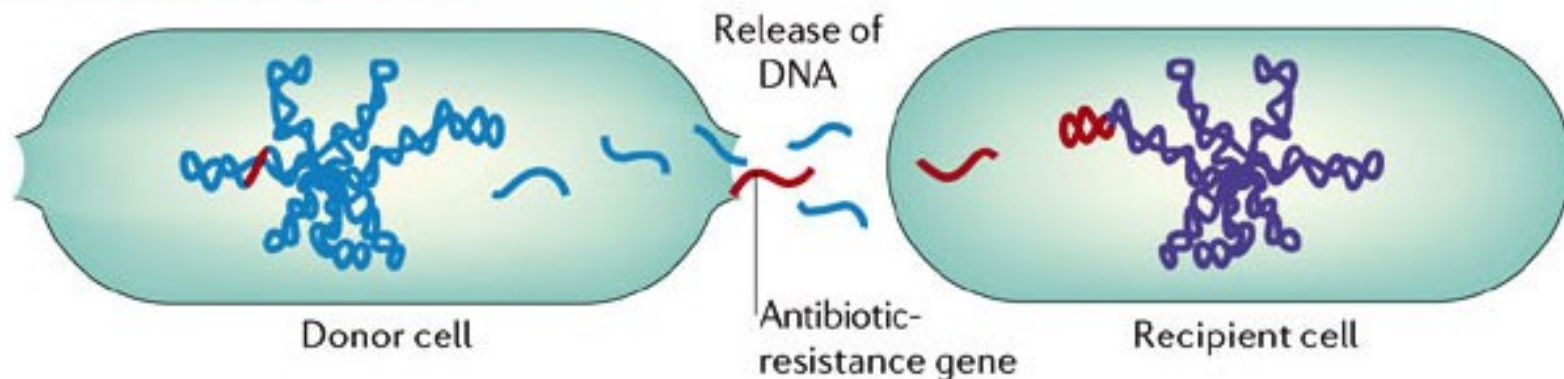
Horizontal Gene Transfer

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

Generally stated: the movement of genetic information between organisms, a process that includes the spread of antibiotic resistance genes among bacteria (except for those from parent to offspring), fueling pathogen evolution.

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 un-related 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.

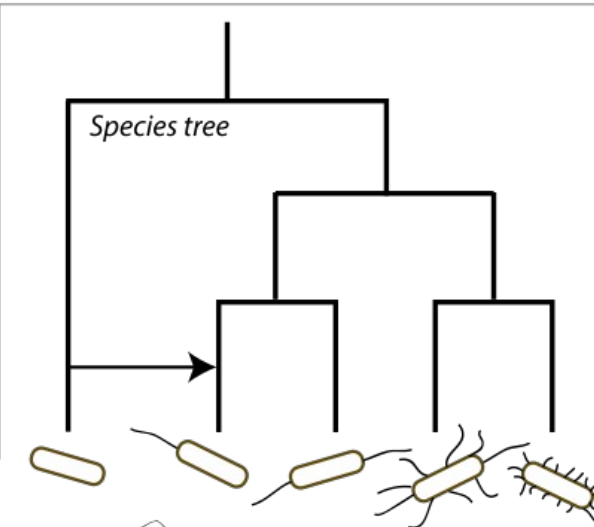# Detection of HGT:
Fragments that do not belong
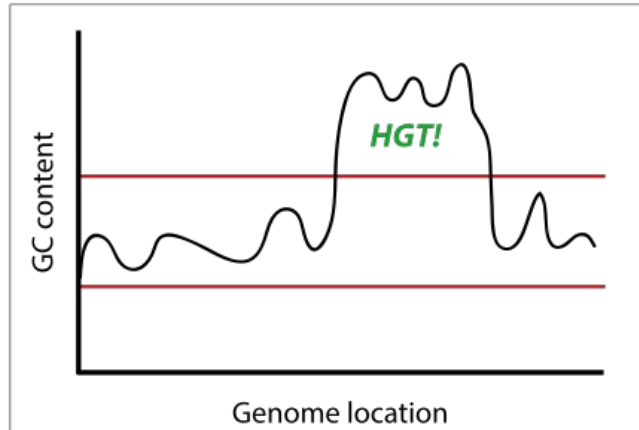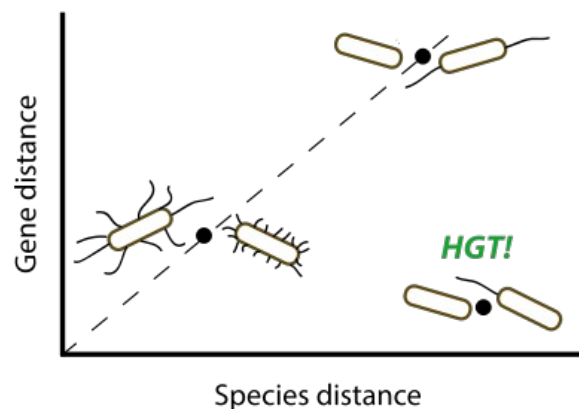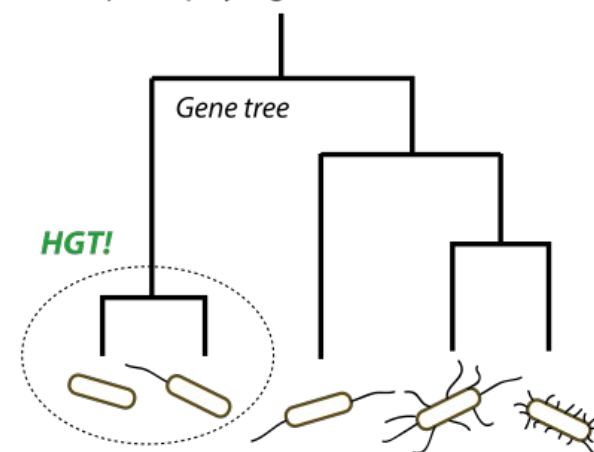
Genetic Sharing: Experiment

	bio	met	phe	thr
Strain 1	+	+	-	-
Strain 2	-	-	+	+

- Colonies have opposing growth requirements: each colony requires two extra genes to survive under conditions

- Strain 1 needs *phe* and *thr*, Strain 2 needs *bio and met*.

- In the experiment, colonies randomly share missing genes with each other.

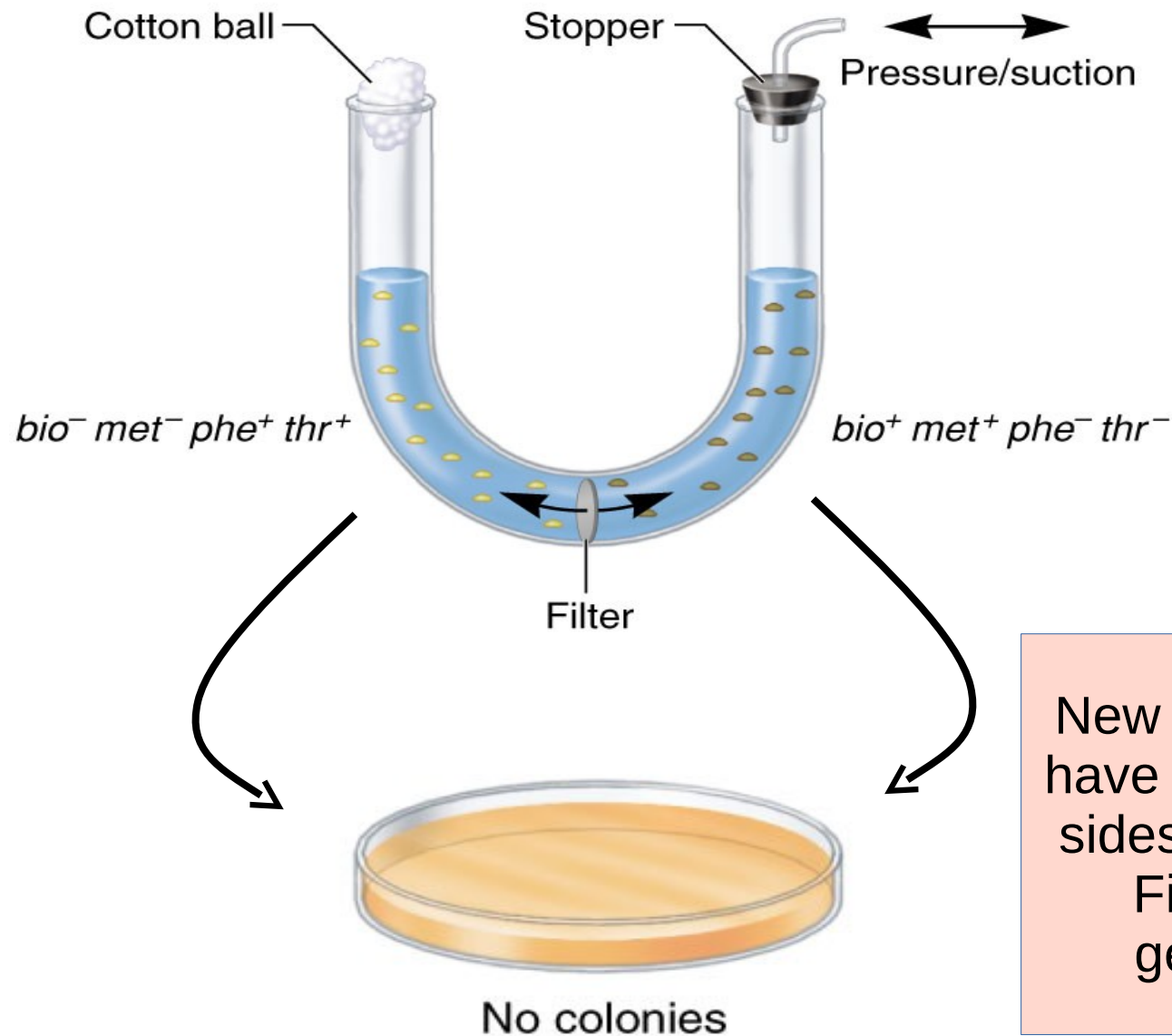- Surviving cells make a new colony and have all four necessary genes from HTG.

bio⁻ met⁻ phe⁺ thr⁺   bio⁺ met⁺ phe⁻ thr⁻

10⁹ cells   10⁹ cells   10⁹ cells

Nutrient agar plates lacking amino acids and biotin

No colonies   Bacterial colonies   No colonies
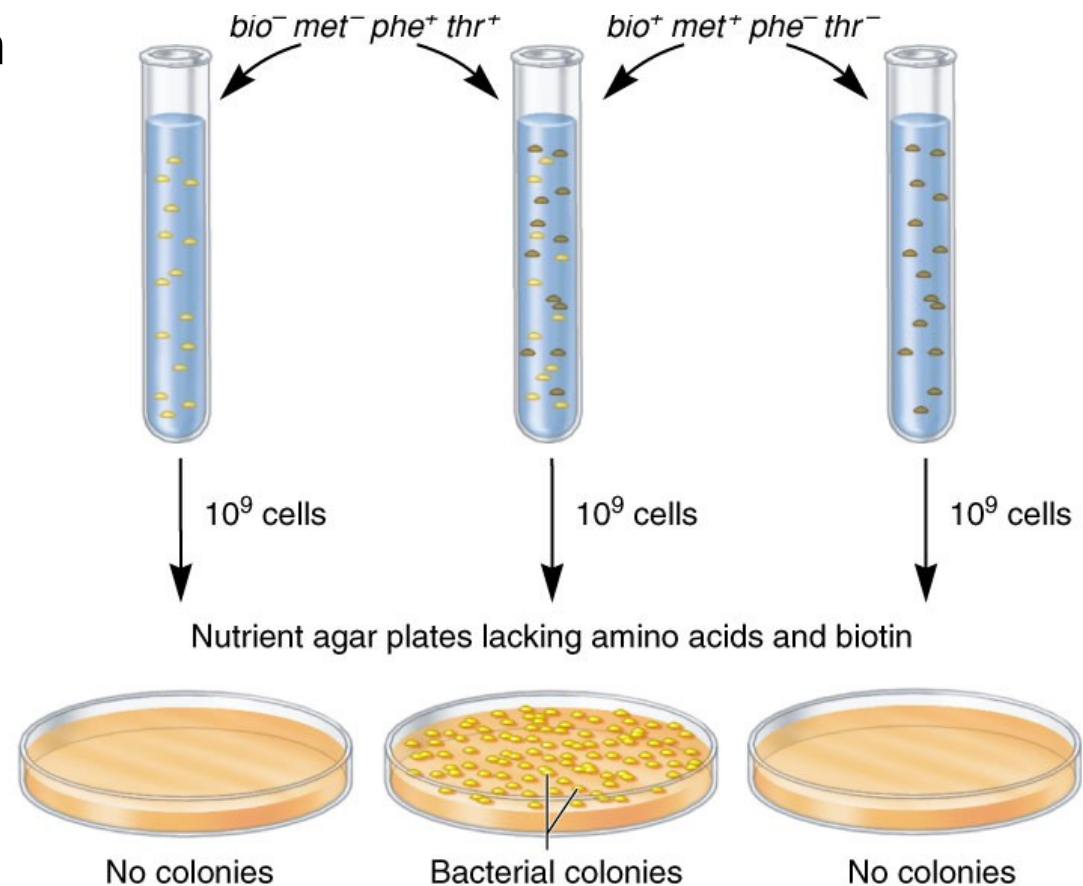
Transfer of Genes From Physical Contact

Cotton ball

Stopper

Pressure/suction

$bio^- met^- phe^+ thr^+$

$bio^+ met^+ phe^- thr^-$

Filter

No colonies

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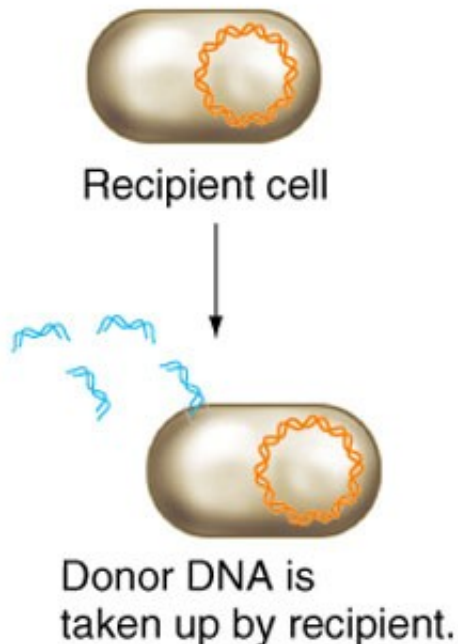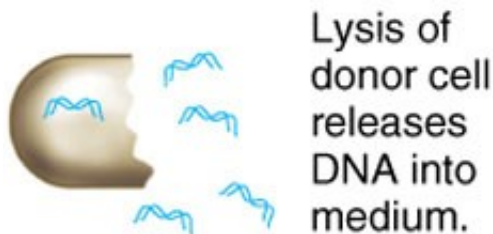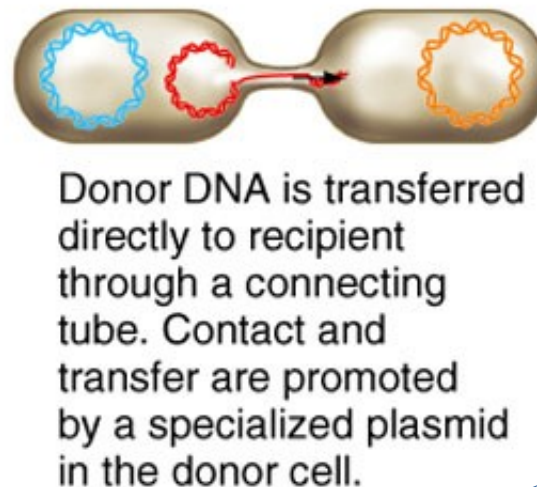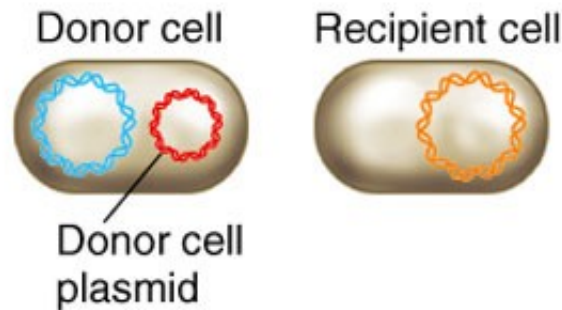
Mechanisms of DNA Transfer

- 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
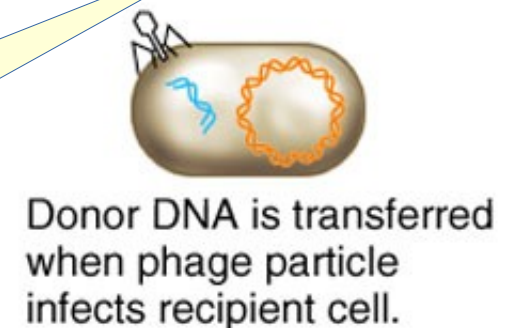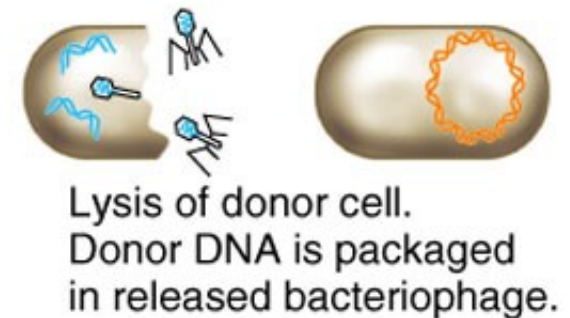
Mechanisms of DNA Transfer

Transformation

Lysis of donor cell releases DNA into medium.

Recipient cell

Donor DNA is taken up by recipient.

Conjugation

Donor cell

Recipient cell

Donor cell plasmid

Donor DNA is transferred directly to recipient through a connecting tube. Contact and transfer are promoted by a specialized plasmid in the donor cell.

Transduction

Donor cell

Recipient cell

Bacteriophage infects a cell.

Lysis of donor cell. Donor DNA is packaged in released bacteriophage.

Bacteriophage: a virus that parasitizes a bacterium by infecting it and reproducing inside it.

Donor DNA is transferred when phage particle infects recipient cell.

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, tube-like structure called a *pilus*, which the donor uses to contact the recipient and transfer genetic content to recipient.
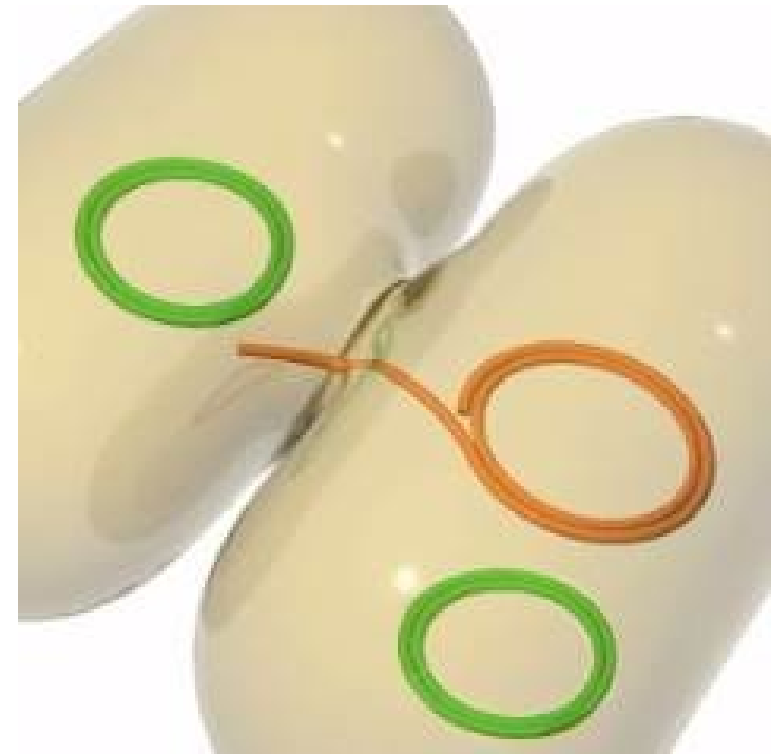
A Pilus : An Exchange Bridge

- Gene sharing possible if F-Factor present

- Fimbriae or pili (singlular: 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

Donor cell • Sex Pili • Fimbriae • Recipient Cells • microbeonline.com • TEM • 1 μm

The F-Factor
Fertility Material

- **F-Factor**: a material (i.e., the *fertility factor*) that is encoded on the plasmid (bacterial DNA). F+ (means has factor) F- (means does not have factor)

- Enables the creation of the pilus bridge to transfer circular piece of DNA
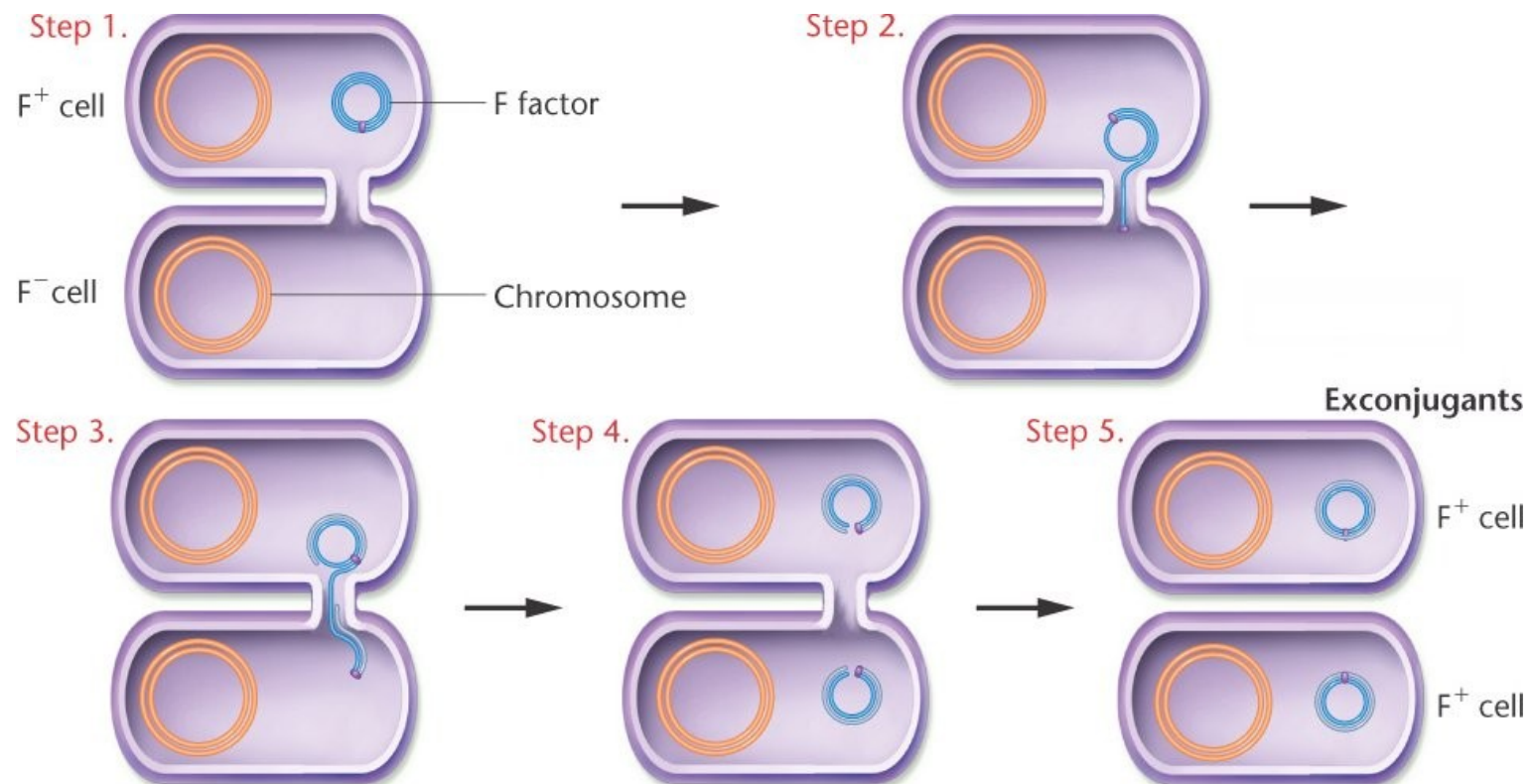
Transfer of Genetic Material
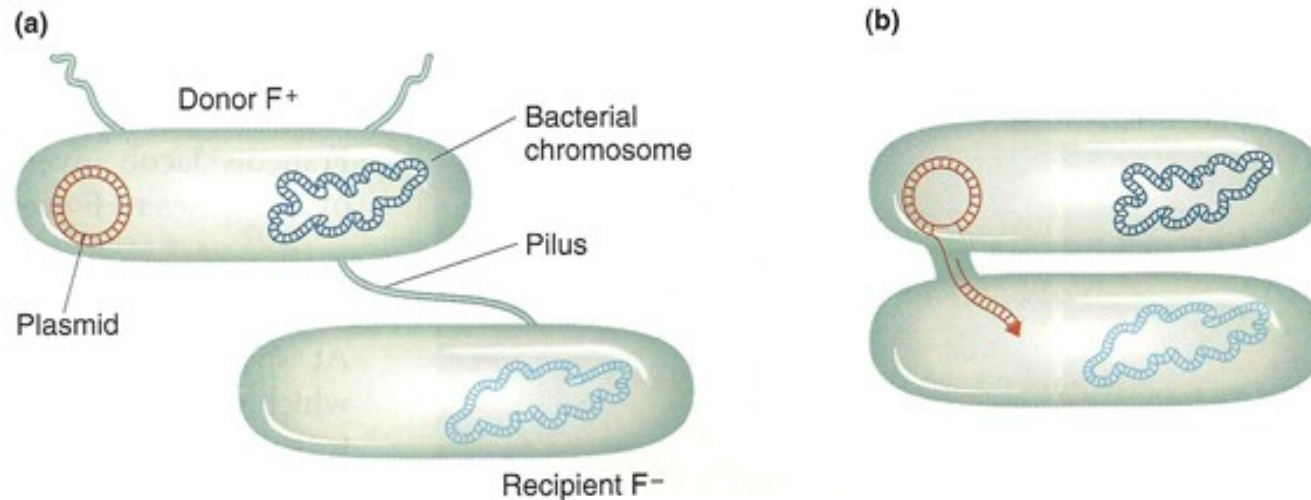
Plasmids are transferred containing genes of donor to recipient through the pilus
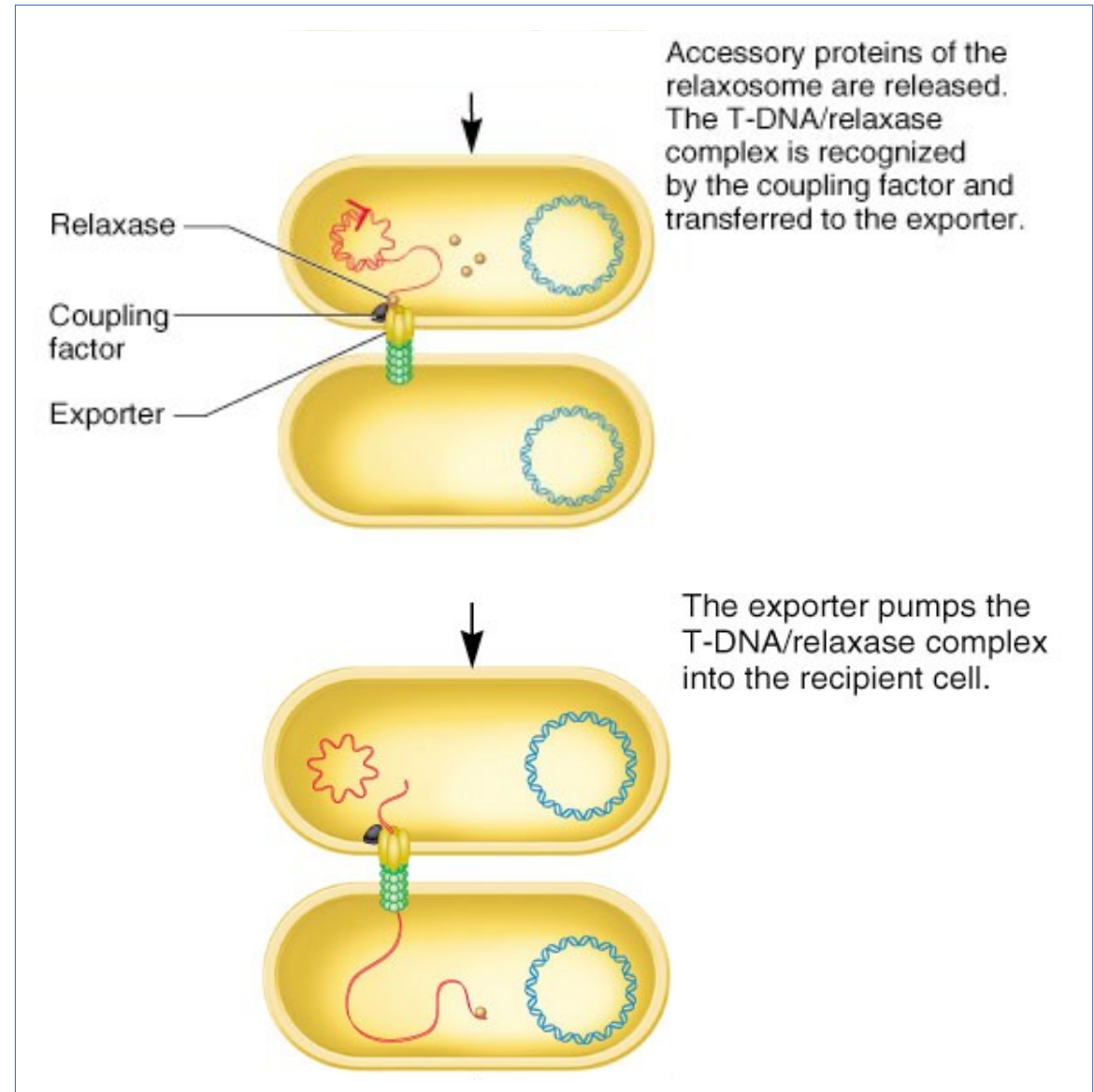
Mechanism of Transfer

- 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

Relaxosome makes a cut at the origin of transfer and begins to separate the T-DNA strand.
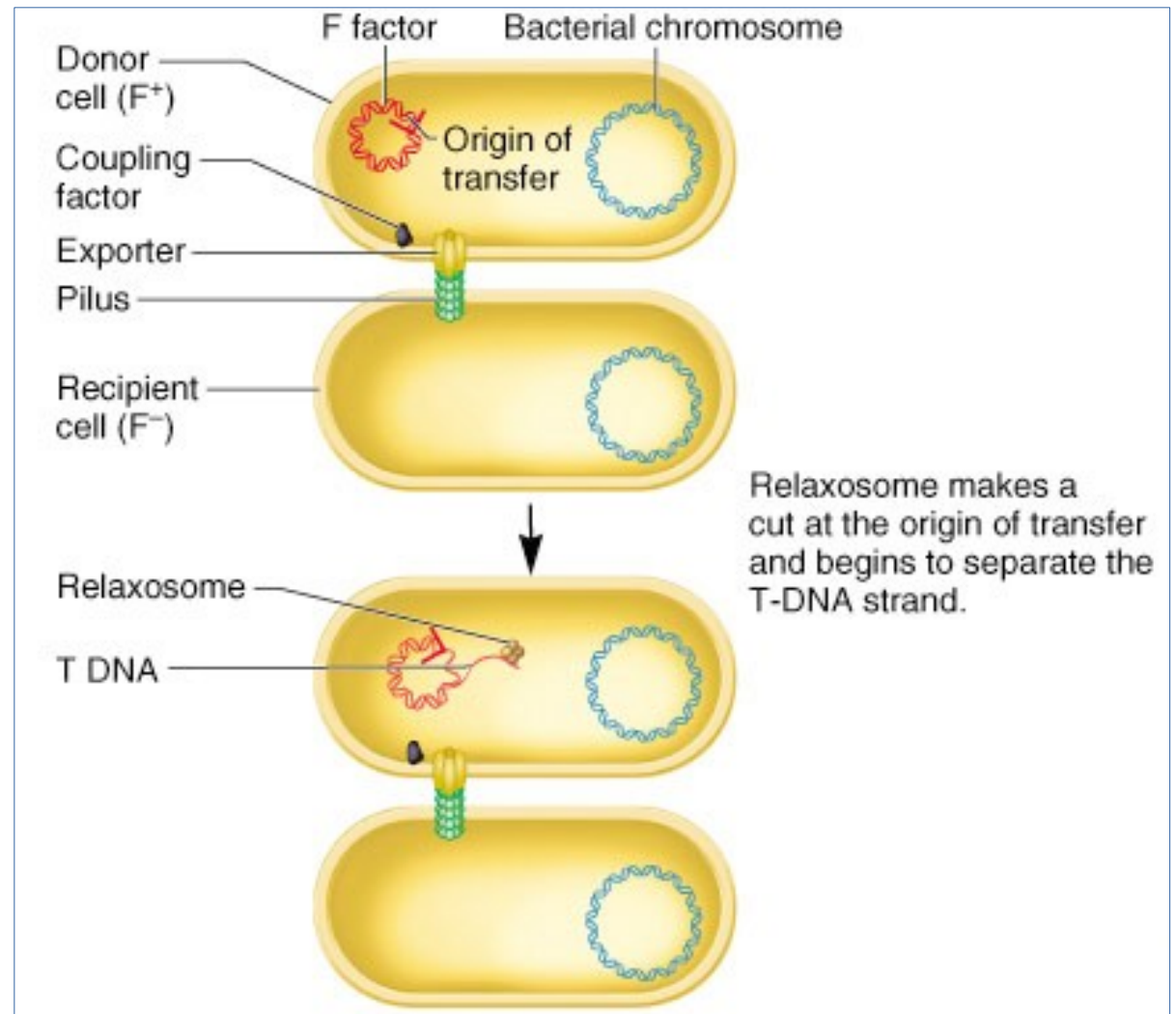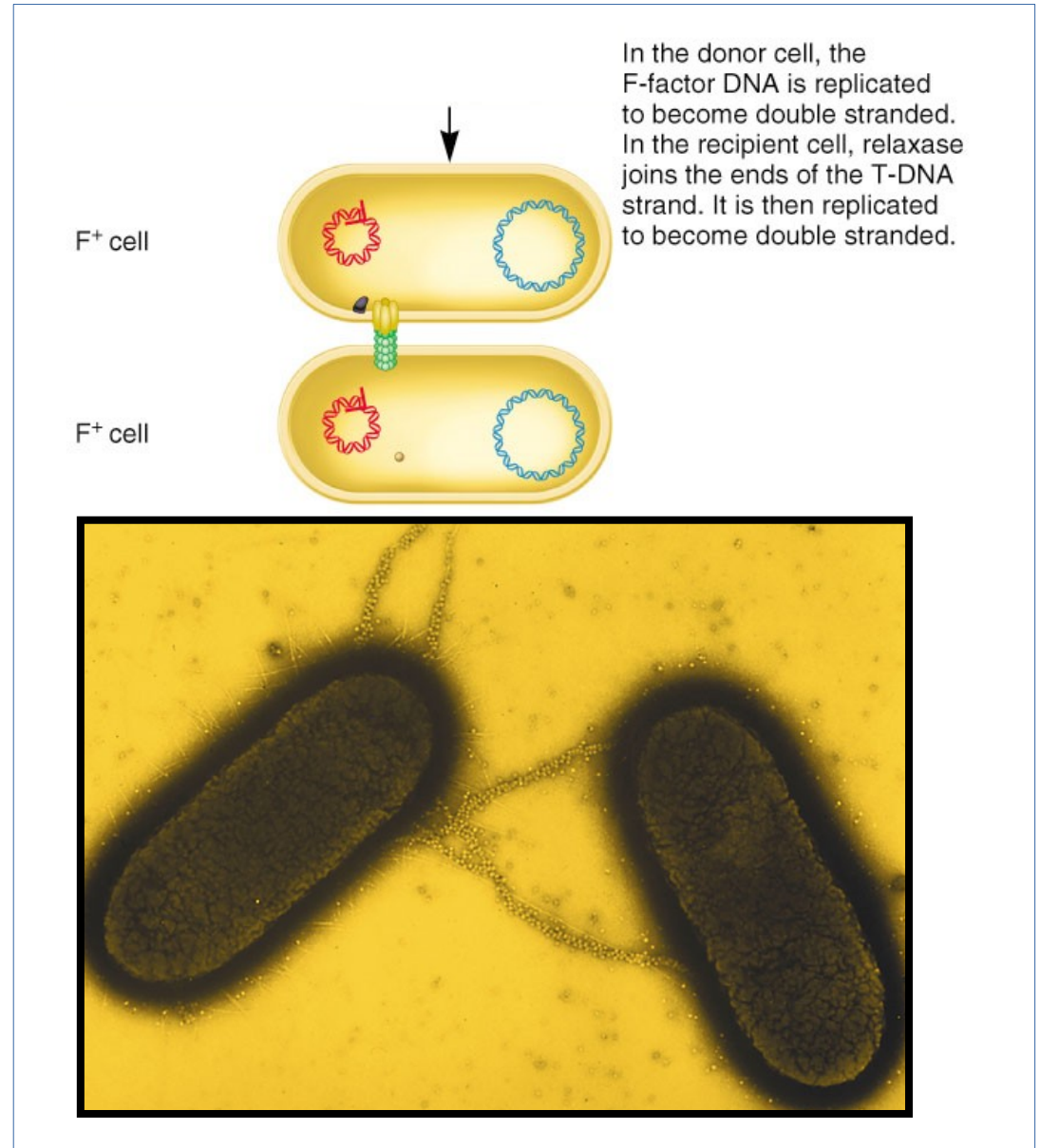
Mechanism of Transfer

- T-DNA is separated but bound to relaxase protein.

- Complex called nucleoprotein

- Complex is recognized by a coupling factor, fed through exporter (conjugation tube)

Accessory proteins of the relaxosome are released. The T-DNA/relaxase complex is recognized by the coupling factor and transferred to the exporter.

Relaxase

Coupling factor

Exporter

The exporter pumps the T-DNA/relaxase complex into the recipient cell.

F Factor Transfer

- Relaxase joins ends to produce circular molecule

- Single strand of F-Factor are in both cells (DNA replication)

In the donor cell, the F-factor DNA is replicated to become double stranded. In the recipient cell, relaxase joins the ends of the T-DNA strand. It is then replicated to become double stranded.
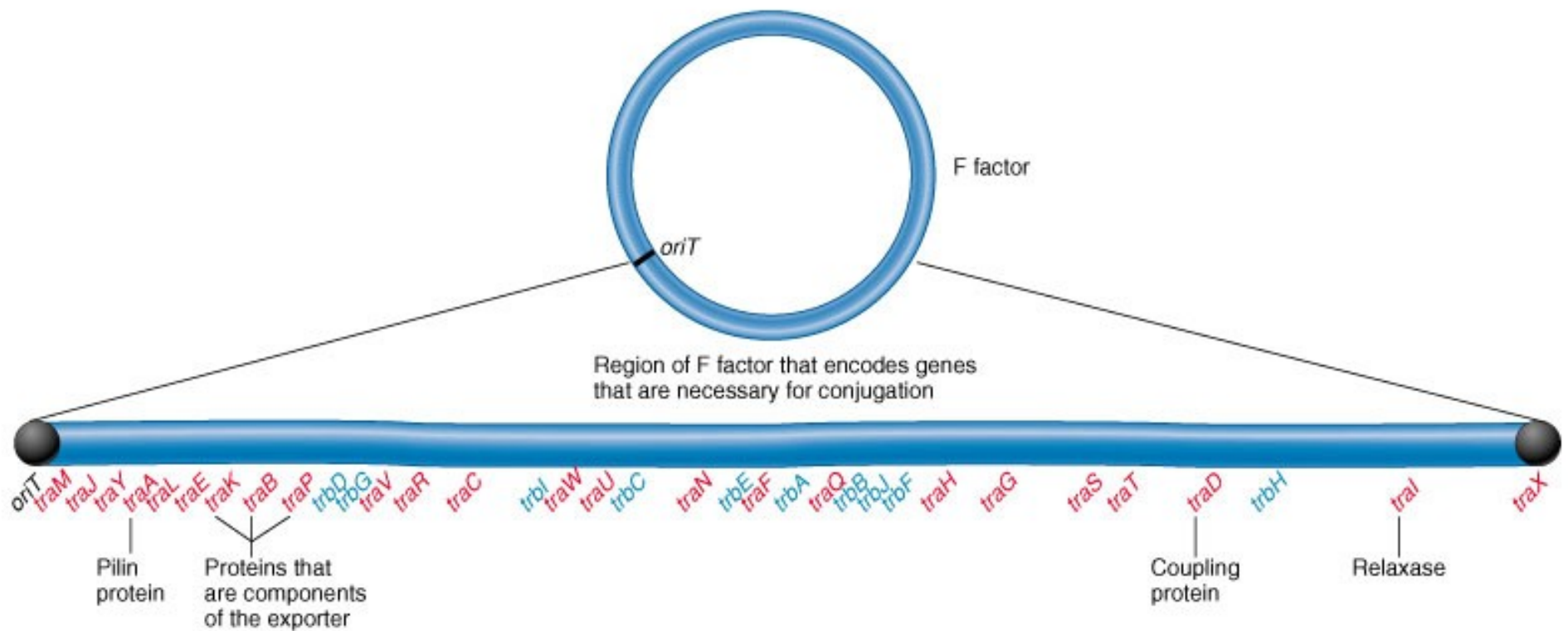
F⁺ cell

F⁺ cell

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+ strain

Episome:
DNA fragment that can exist as a plasmid and integrate into chromosome
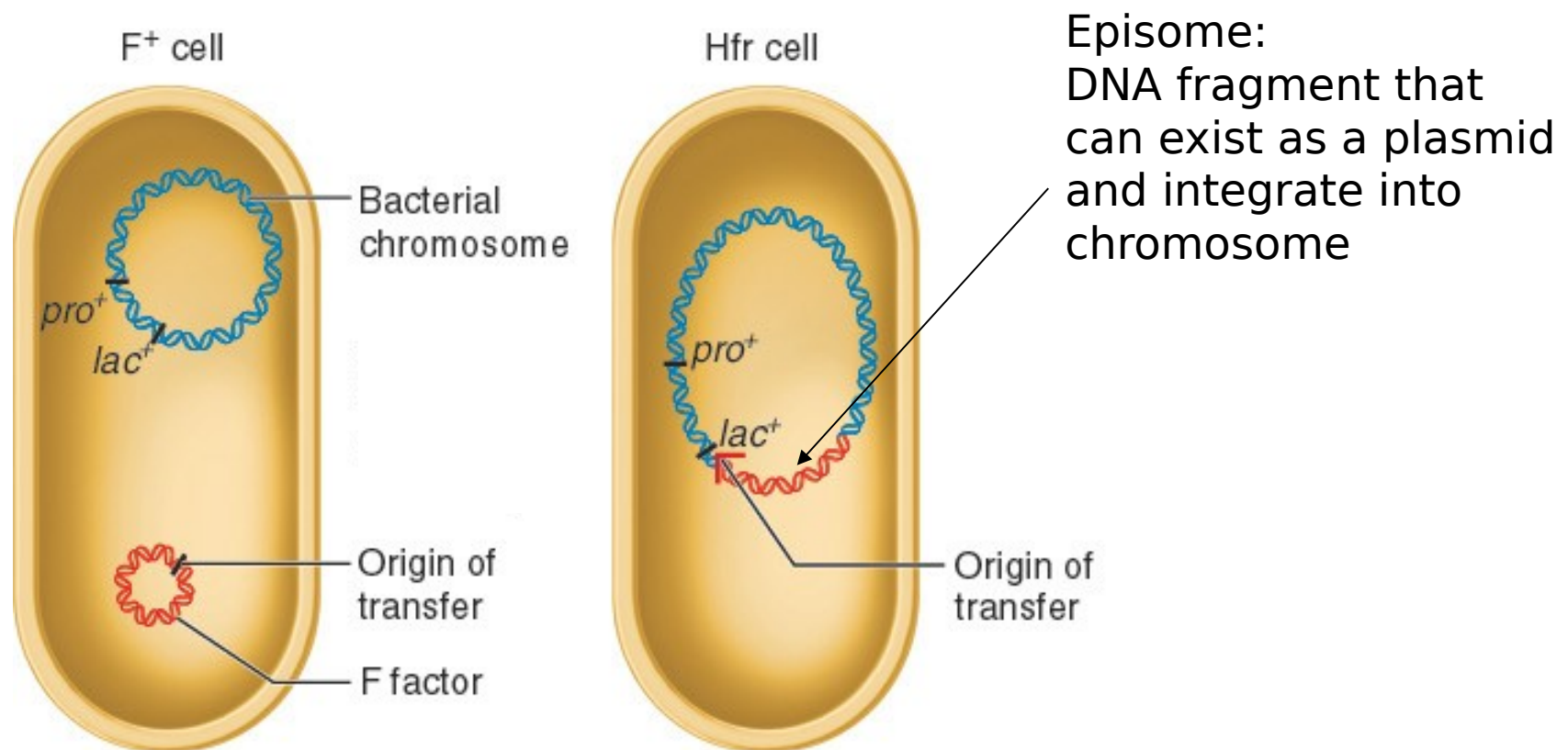
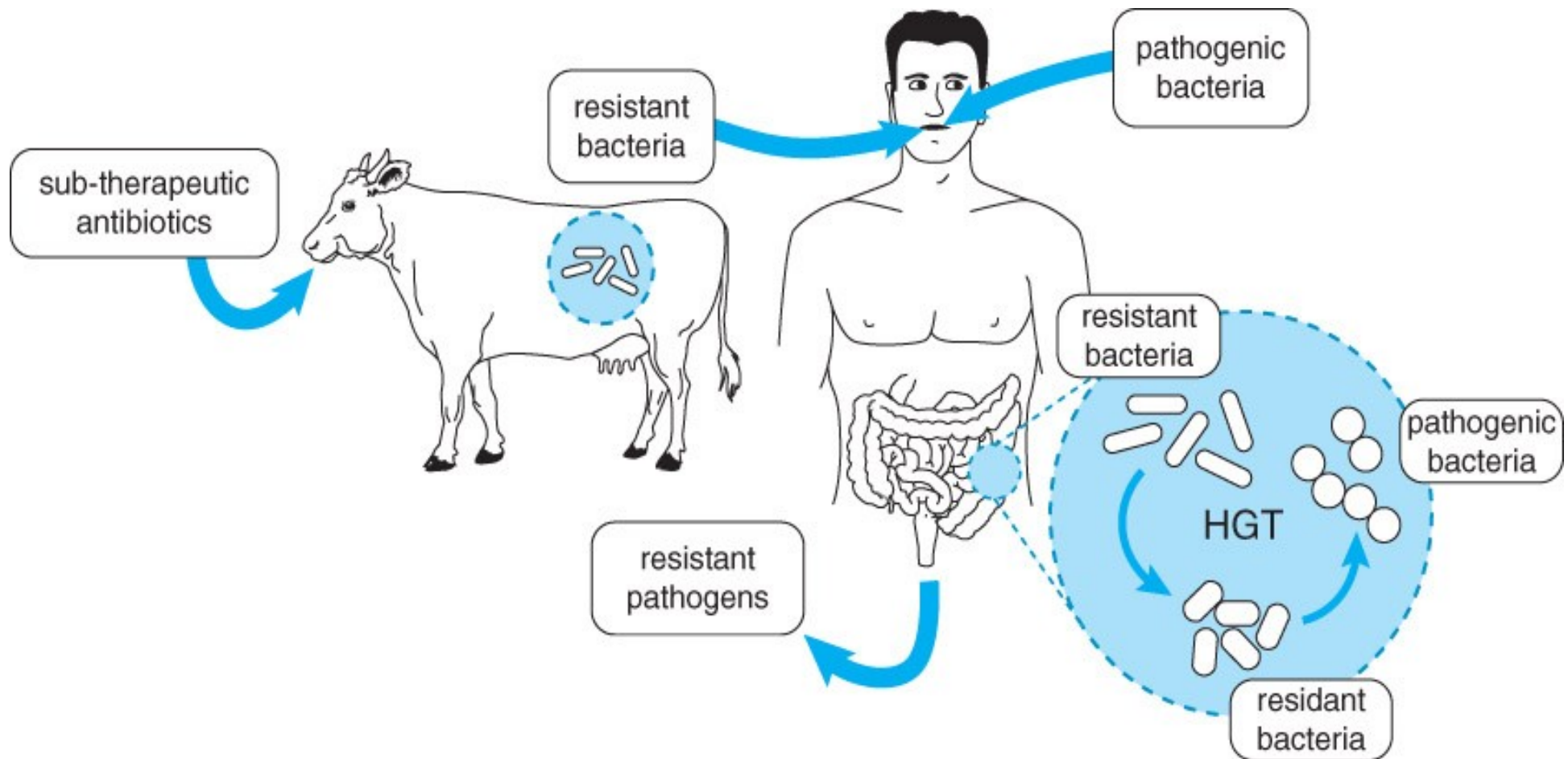# Plasmids Contain (lots of) Genes

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

Genetic code that did not
come from common ancestry.
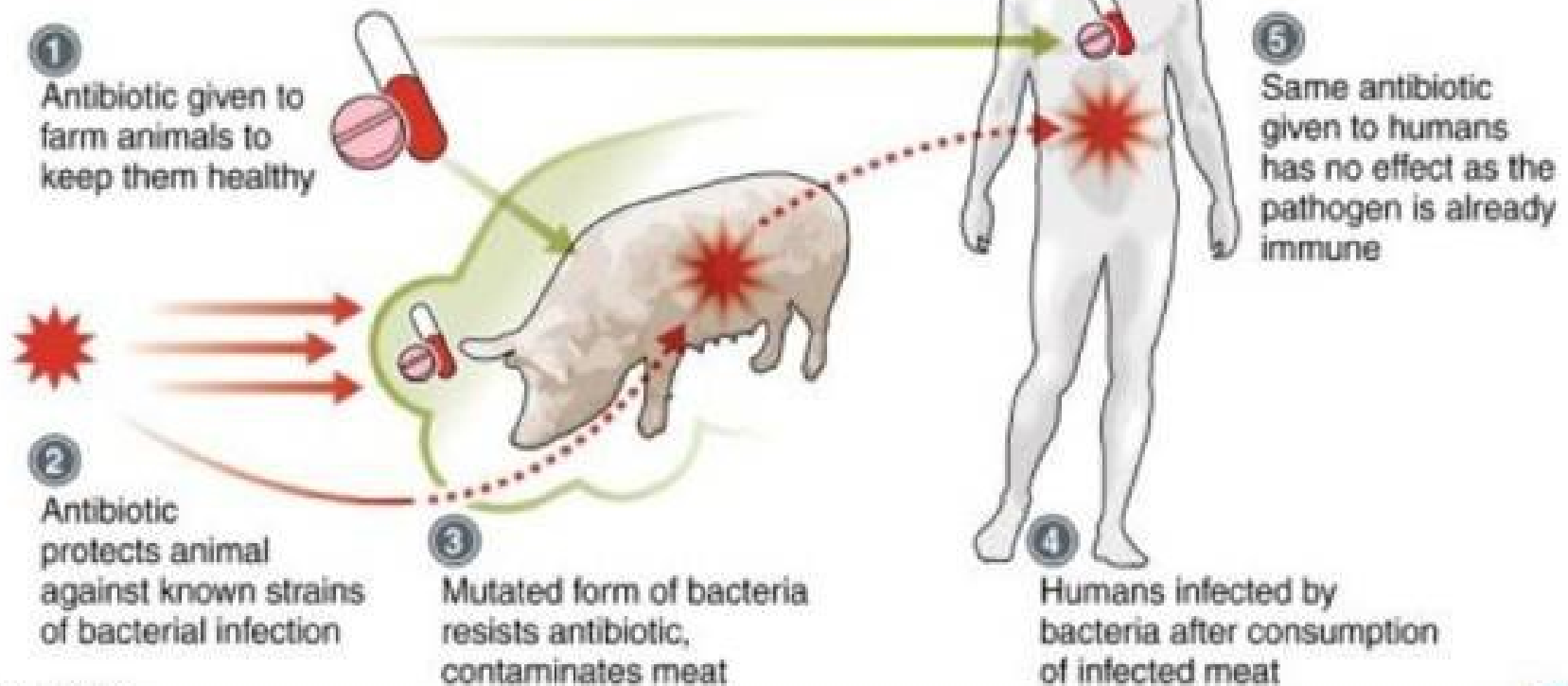
Horizontal Gene Transfer of Antibiotic Resistance

From Farm-to-Fork Spread
of Antibiotic Resistance

Supporting Videos

- Why the sex pilus is so dangerous - horizontal gene transfer
    - https://www.youtube.com/watch?v=GzCLp1KBf4Q
- Antibiotic Resistance (4mins)
    - https://ed.ted.com/lessons/how-antibiotics-become-resistant-over-time-kevin-wu

Multi-Sequence Alignment

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

Where are the chucks 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 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-I20171005-050706-0613-19128639-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-CTCAAAACCCTCAAACTCTA
X95928.1        TATTATTTCATGAAACTAATGTGAGGAGAAAAATTAGGCAGACGTATAAGTGACAGTAGA
KU325497.1      ------------------------------------------------------------
KU325498.1      ------------------------------------------------------------


AB011005.1      ------------------------------------------------------------
X95927.1        ACCTTGGAACAAACAGCCCTAGTGGGATAGCCCTATCAAACCCTTCCCTTC--AGGGATT
```

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: [Choose 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      MVLSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKTYFPHFDLSHGSAQVKGHG    60
sp|P01942|HBA_MOUSE      MVLSGEDKSNIKAAWGKIGGHGAEYGAEALERMFASFPTTKTYFPHFDVSHGSAQVKGHG    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
- <u>B</u>asic <u>L</u>ocal <u>A</u>lignment <u>S</u>earch <u>T</u>ool
  - 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 Needlman-Wunsch algorithm)

BLAST: A Heuristic Approach
to Database Searching

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

BLAST: Output

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       ATATATATATATATATATATAT  22
               ||||||||||||||||||||||
Sbjct  102686  ATATATATATATATATATATAT  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       ATATATATATATATATATATAT  22
               ||||||||||||||||||||||
Sbjct  102707  ATATATATATATATATATATAT  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       TATATATATATATATATATAT  22
               |||||||||||||||||||||
Sbjct  102685  TATATATATATATATATATAT  102705
```

BLAST: Output

BLAST: Tree of Relations

BLAST Demo

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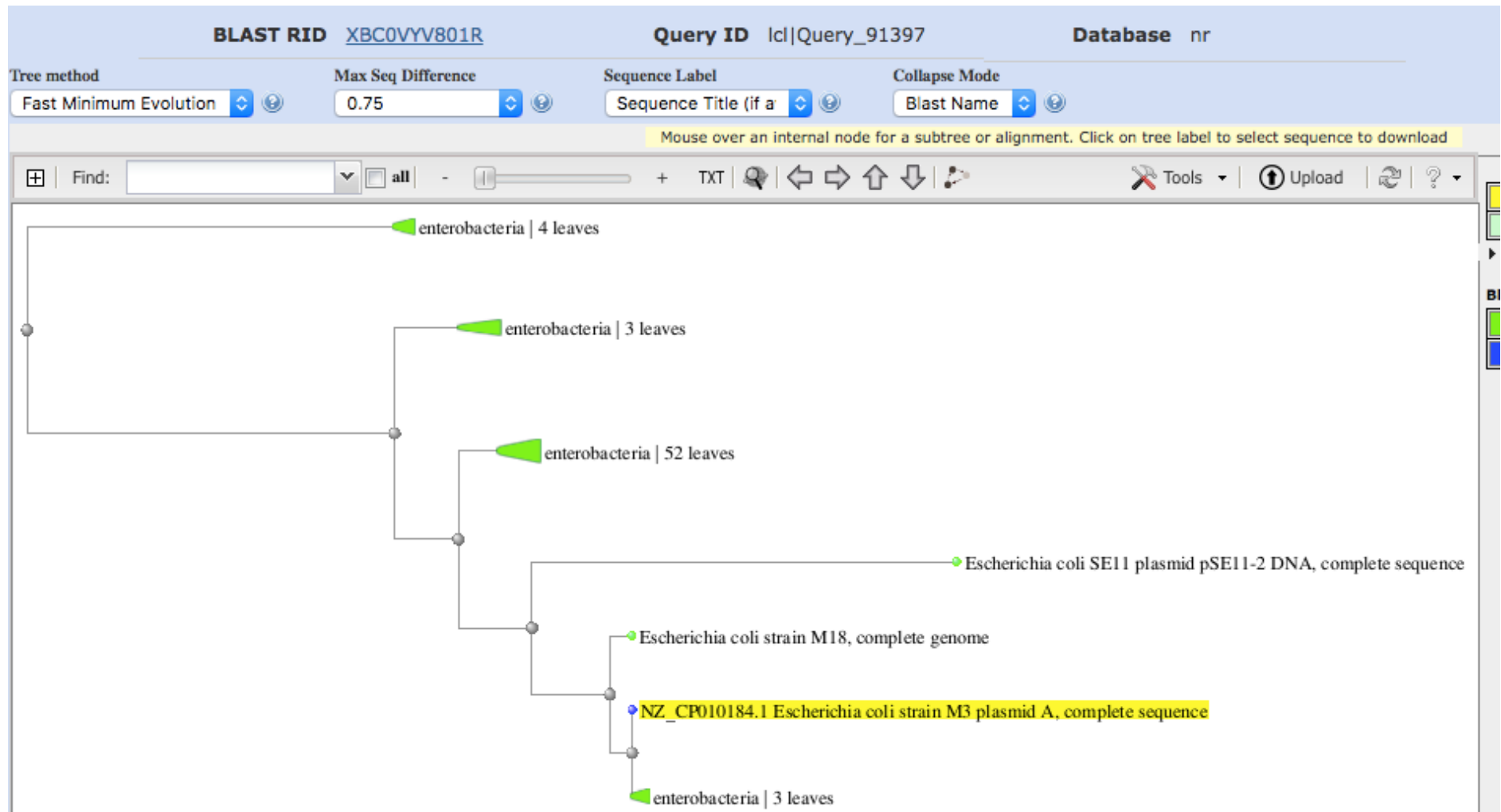
We are checking the occurrence of this RNA sequence, **Cyprinus carpio SIRT1 mRNA**, partial cds GenBank: **KF881970.1**) in many other sequences to determine common parts of code.

BLAST Demo

Job Title	**gb|KF881970.1|**
RID	TRT19U51014 *Search expires on 10-09 09:19 am* Download All ˅
Program	BLASTN ❓ Citation ˅
Database	refseq_rna See details ˅
Query ID	KF881970.1
Description	Cyprinus carpio SIRT1 mRNA, partial cds.
Molecule type	rna
Query Length	375
Other reports	Distance tree of results MSA viewer ❓

We get basic information from the search...

BLAST Demo

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    TATATTCAACCAGGTTGTTCCTCACTGTCCCAGGTGTCCGGCAGACGTCCCGT
             ||  || |||||| |||| ||||| ||||  |||||  | ||||| |  || |
Sbjct  1134  CATCTTTAACCAGATTGTCCCTCATTGTCCACGATGT-C--CAGATATTCC-T

Query  120   TCATGAAACCAGACATCGTCTTCTTTGGCGAGAACCTTCCAGAGTTTTTCCAC
             |||||||||||||||| |||||||||||| ||||| |||||||| ||| |||
Sbjct  1190  TCATGAAACCAGACATTGTCTTCTTTGGAGAGAATCTTCCAGAAATGTTTCAC

Query  180   TGAAGCAGGATAAAGATGAGGTGGATCTTCTCATCGTGATCGGCTCCTCGCTG
             ||||||||||||||||||||||||||| |||| ||| || || || ||||| |
Sbjct  1250  TGAAGCAGGATAAAGATGAGGTGGACCTCTTGATTGTCATTGGCTCTTCACTT

Query  240   GGCCAGTGGCTCTCATACCCAGCTCTATACCTCATGACGTGCCTCAAGTCCTG
             | ||||| ||| || || || | ||| || |||||||||||||||||||||||
Sbjct  1310  GACCAGTTGCCCTGATCCCTAACTCCATTCCTCATGAAGTGCCTCAGGTCCTG
```

BLAST Tutorial: DigitalWorldBiology

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