Bioinformatics

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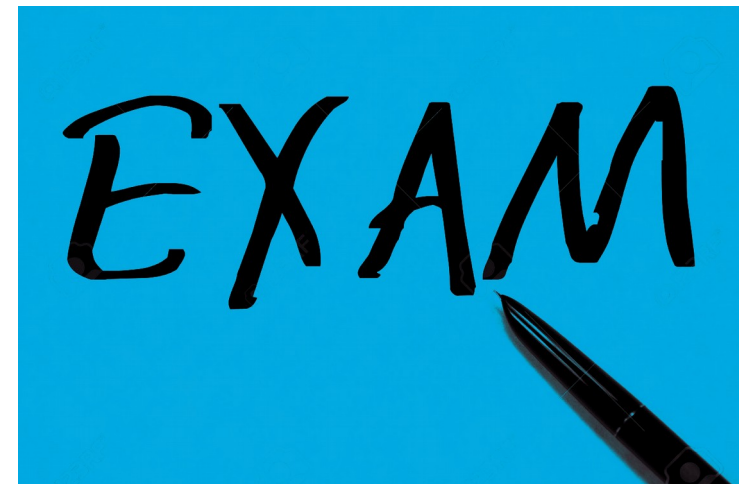## Database, Tools, Multiseq Alignment

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

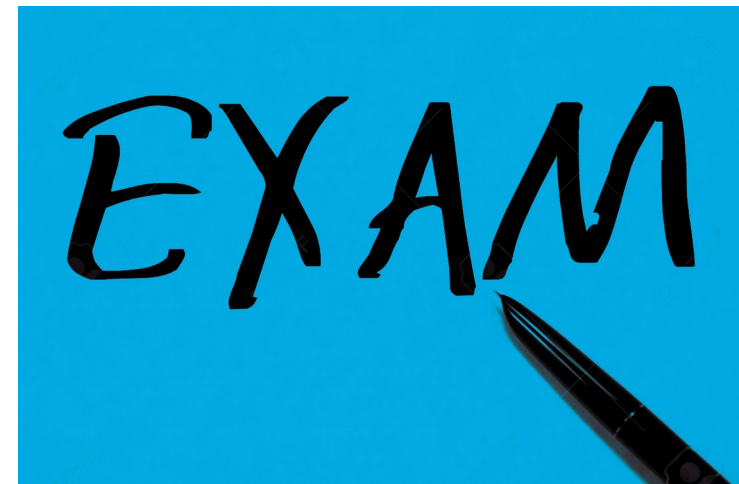- Sequence alignment

- 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

Release of DNA

Donor cell

Antibiotic-resistance gene

Recipient cell
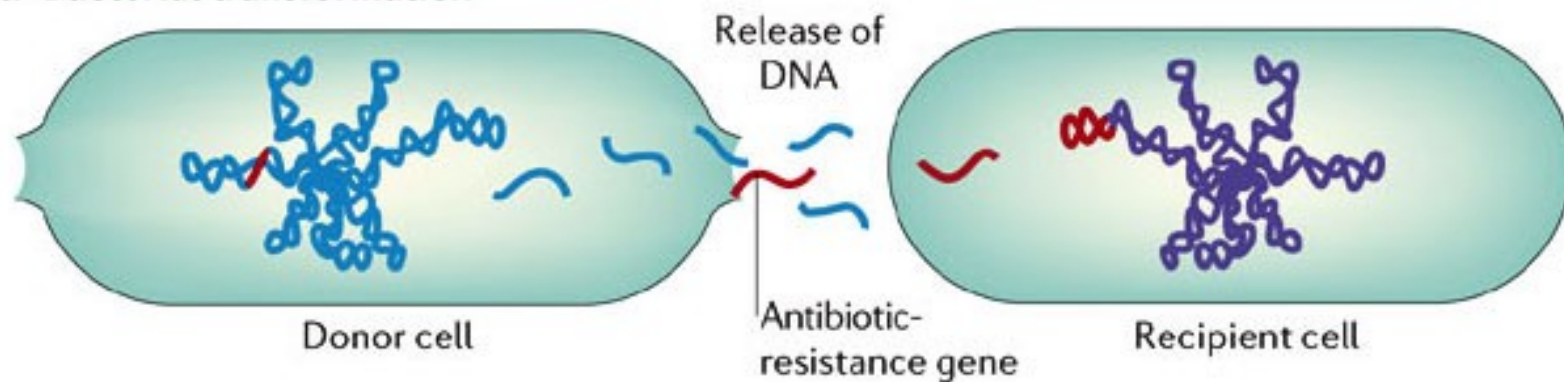
What is HGT?
(Horizontal Gene Transfer)

- 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.
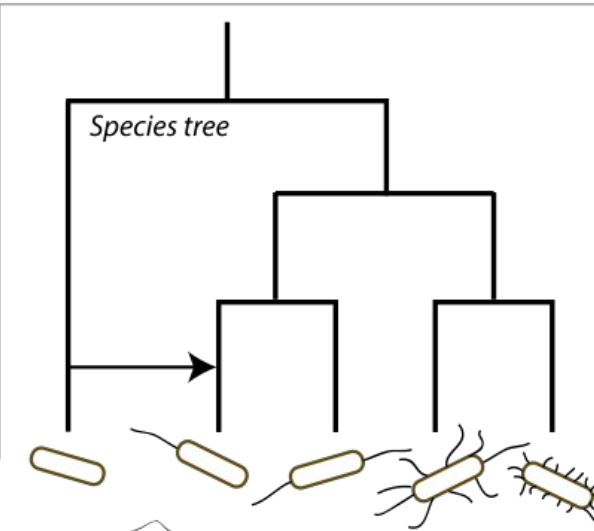
Detection of HGT:
Fragments that do not belong

Statistical tests: Elevated GC content in DNA regions

Species that are too closely related to each other
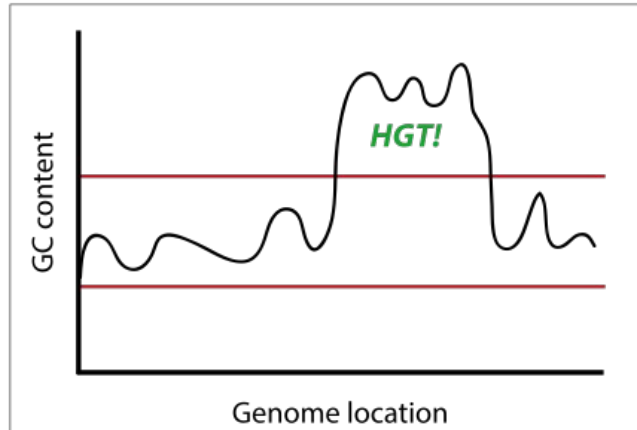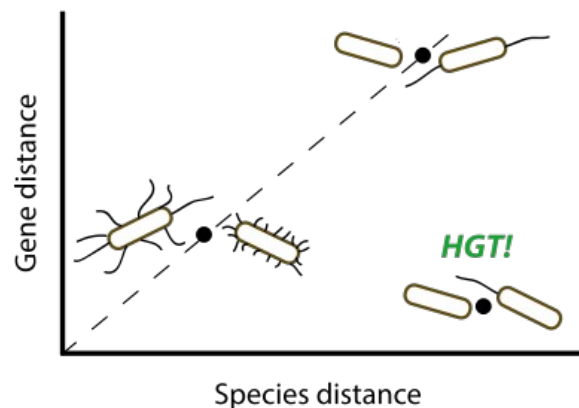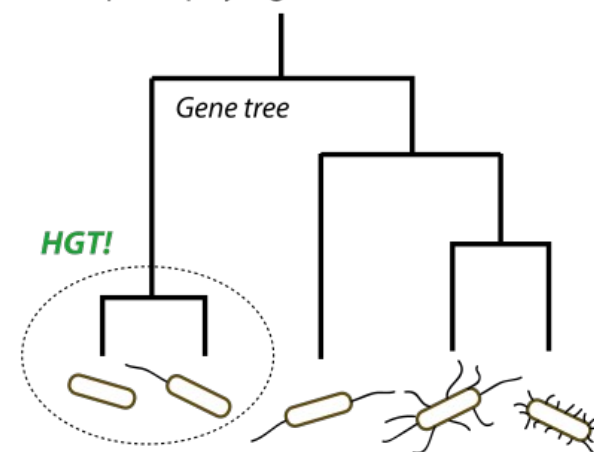
Phylogenetic tests using relationship trees: Organisms living physically near each other show close relations

1. Parametric methods

GC content

HGT!

Genome location

2. Phylogenetic methods

Species tree

2a. Implicit phylogenetic methods

Gene distance

HGT!

Species distance

2b. Explicit phylogenetic methods

Gene tree

HGT!

Genetic Sharing

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

- Colonies have opposing growth requirements: each colony requires two extra genes to survive under conditions
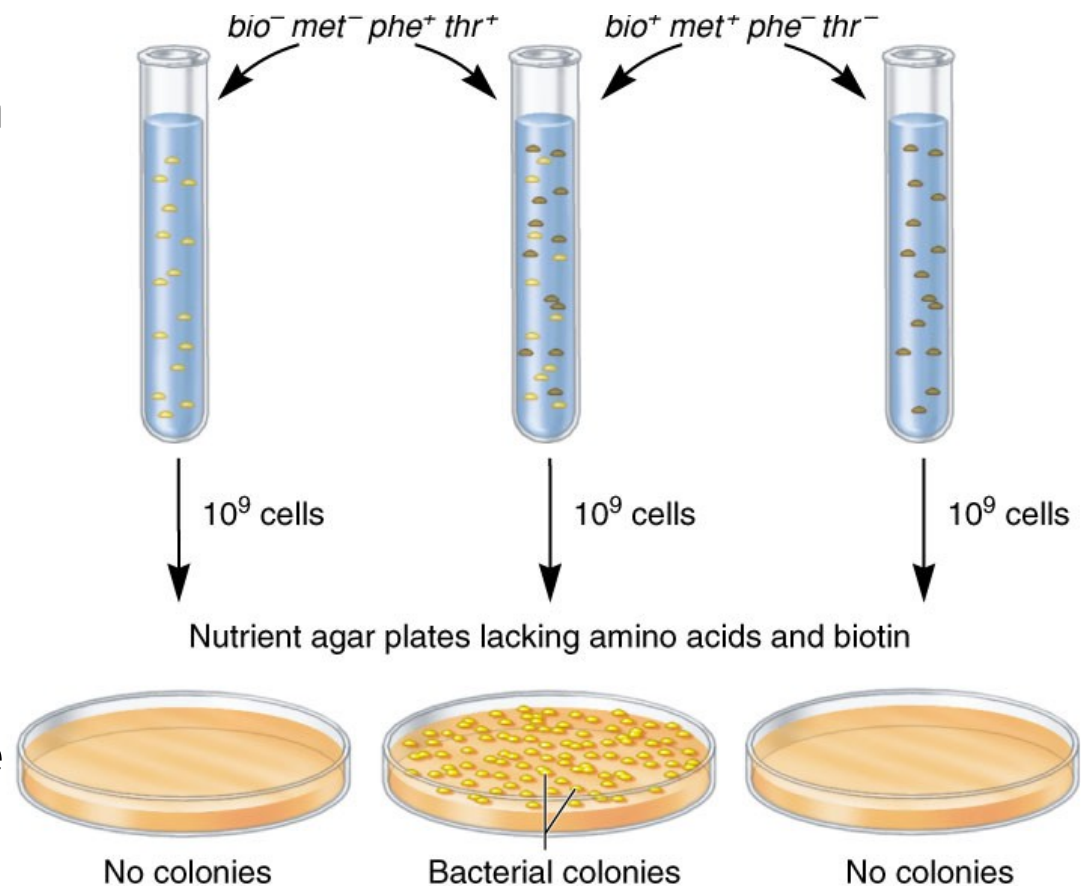
- Str 1 needs *phe* and *thr*, Str 2 needs *bio*, *met*.

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

- Cells with HGT survive to make a new colony.

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

10^9 cells 10^9 cells 10^9 cells

Nutrient agar plates lacking amino acids and biotin

No colonies   Bacterial colonies   No colonies

ALLEGHENY COLLEGE

Transfer of Genes From Physical Contact

Cotton ball

Stopper

Pressure/suction

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

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

Filter

No colonies
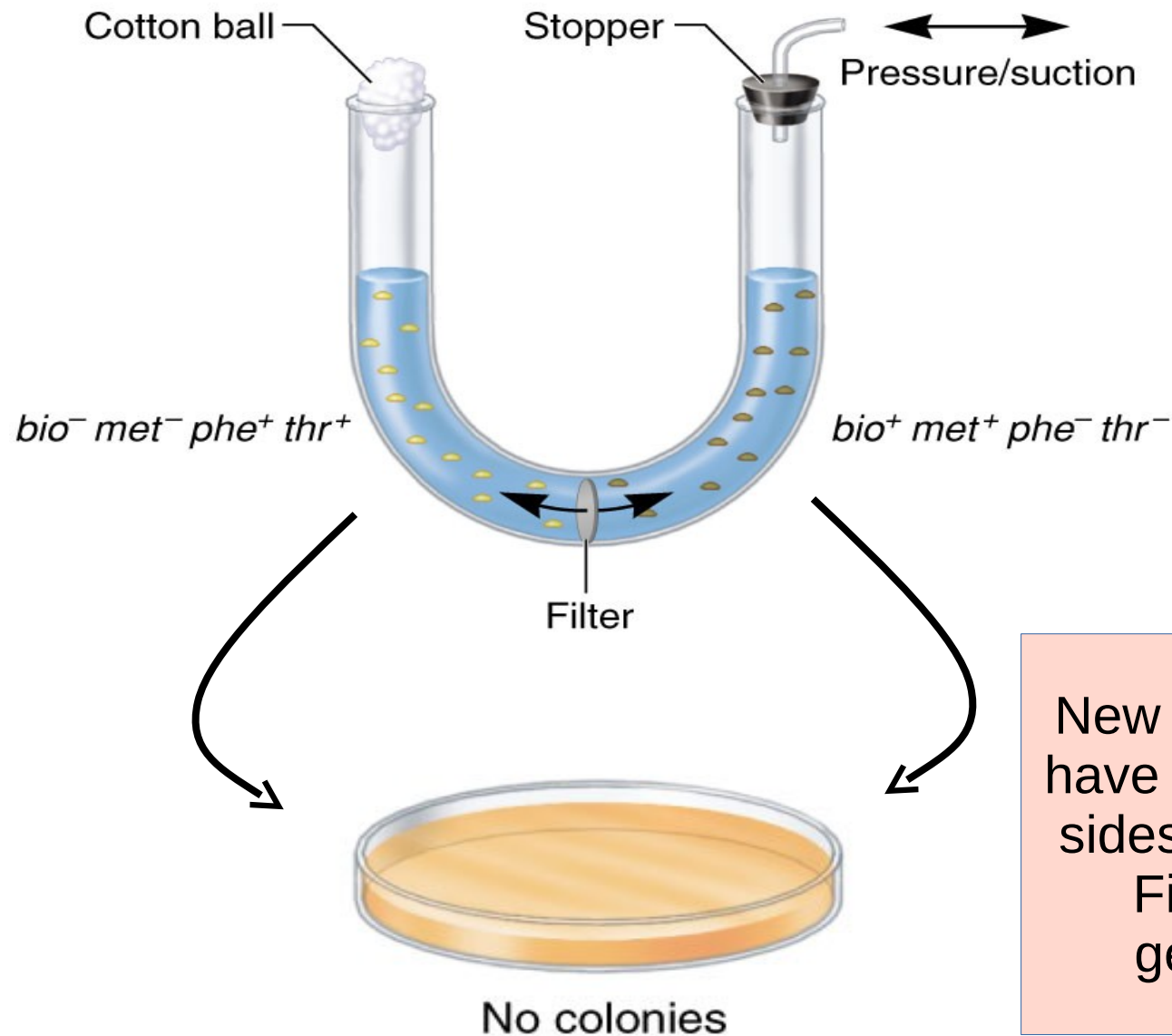
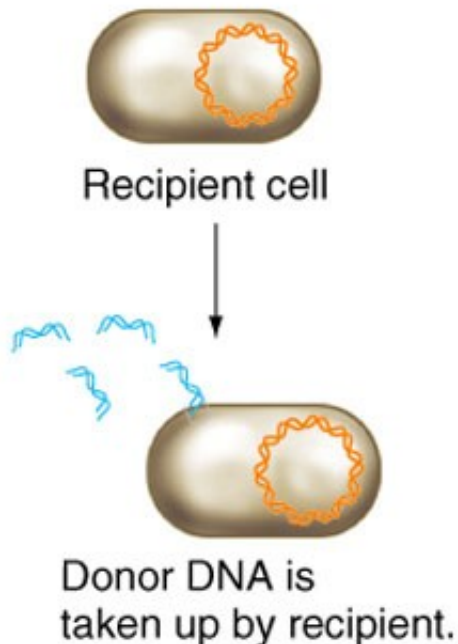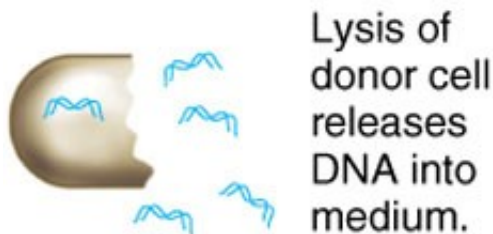New colony needs to have genes from both sides of filter to grow. Filter prevents gene passage
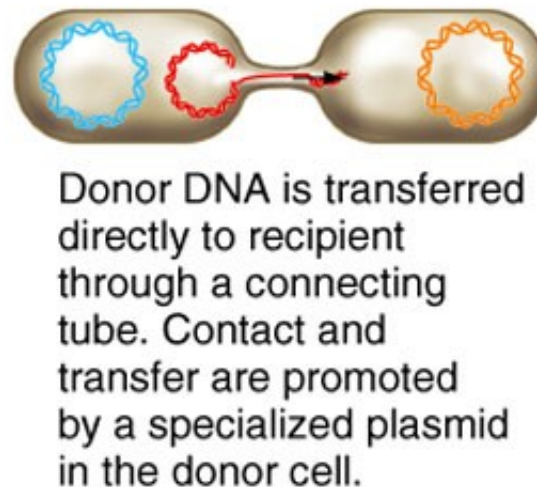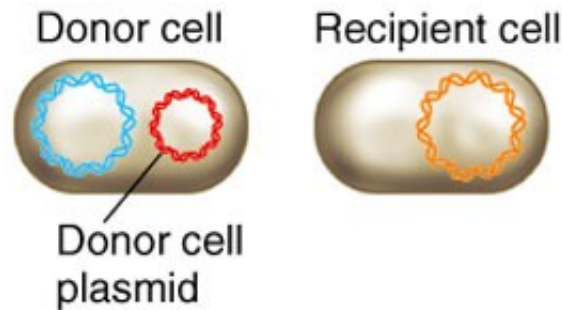
Mechanisms of DNA Transfer

- 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.

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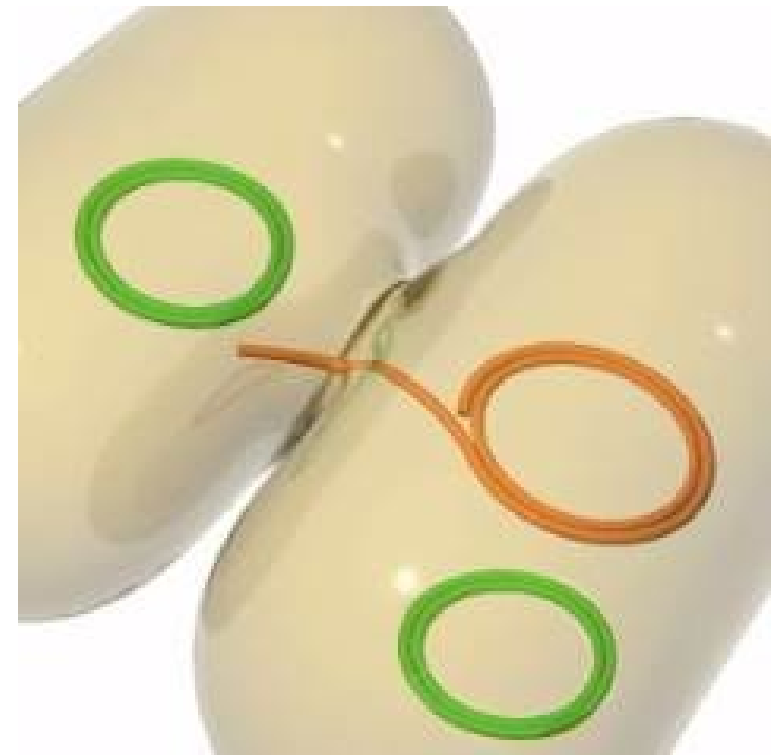
A Pilus : An Exchange Bridge

- Gene sharing possible if F-Factor present

- Fimbriae or pili (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

Recipient Cells

Sex Pili

Fimbriae

Donor cell
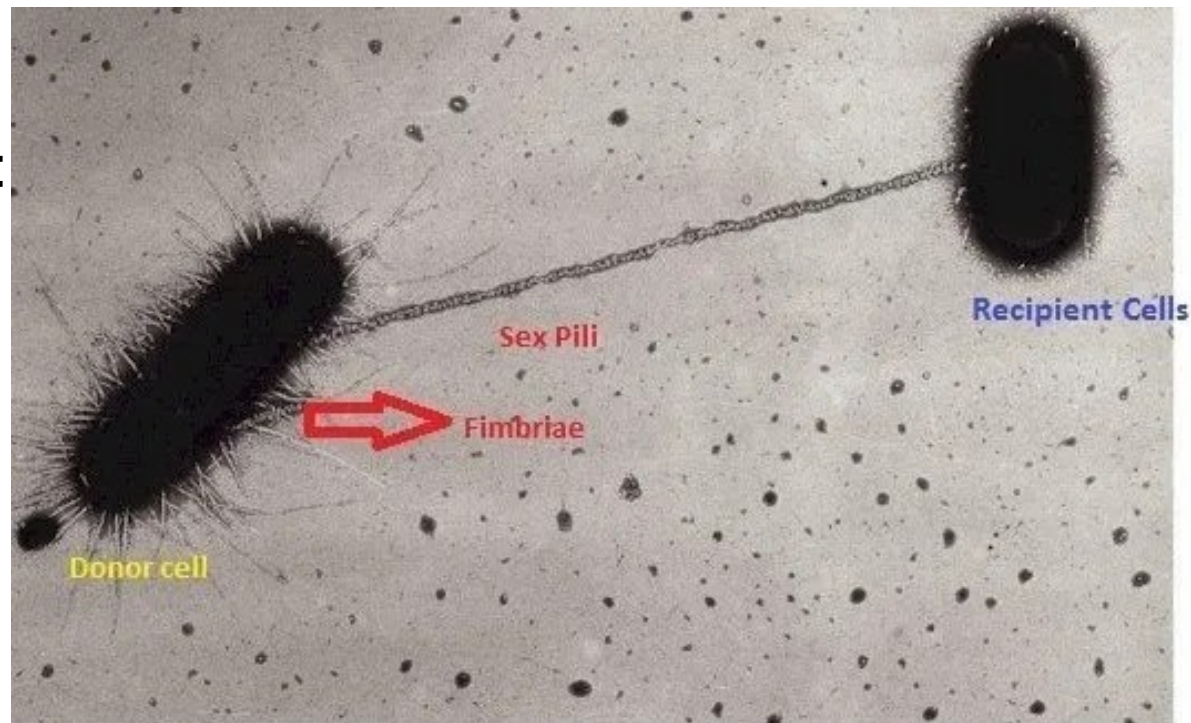
microbeonline.com

TEM

1 μm

The F-Factor Material

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

- Enables the creation of the pilus bridge to transfer circular piece of DNA

(a) Host chromosome / F factor / F replicates in the cytoplasm → F, F

(b) F pili promote cell-to-cell contact

(c) A new copy of F, generated by replication, is transferred to a recipient cell → A copy of F remains in the donor cell
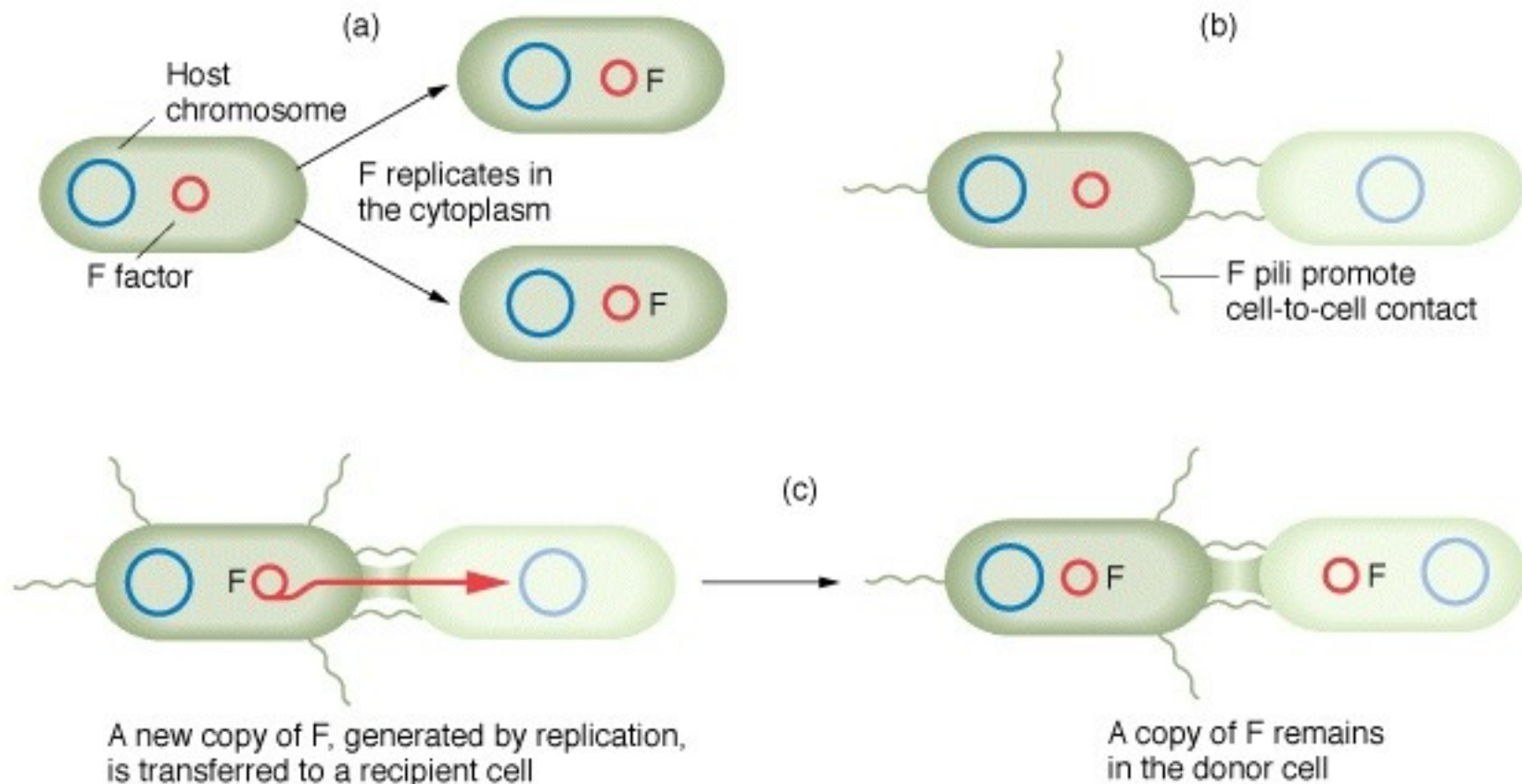
Transfer of Genetic Material
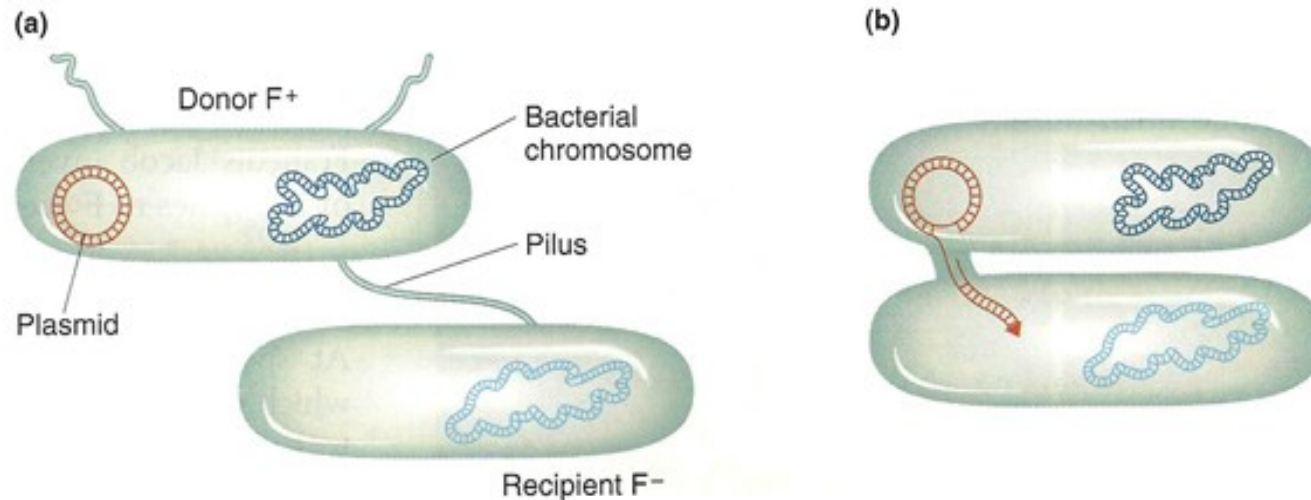
Plasmids are transferred containing genes of donor to recipient through the pilus

Plasmids Contain Genes
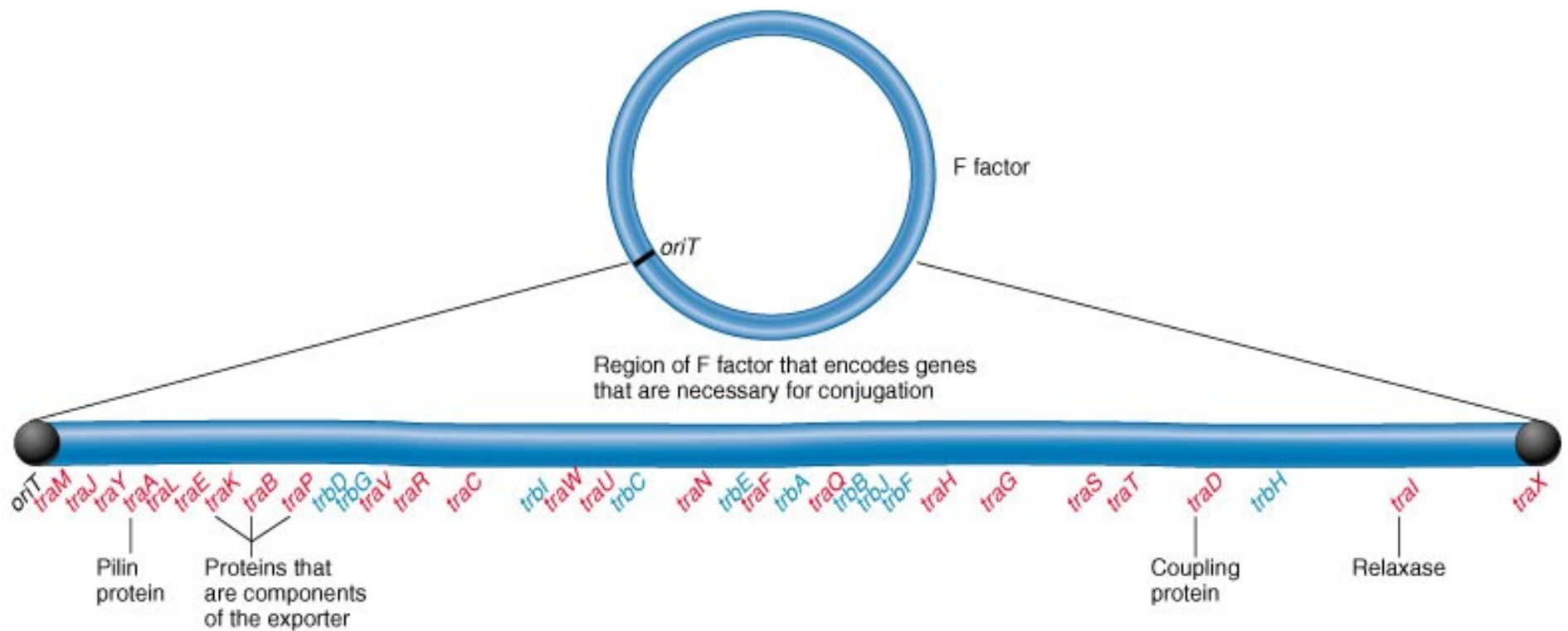
Mechanism of Transfer

- 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

F factor   Bacterial chromosome

Donor cell (F⁺)

Origin of transfer

Coupling factor

Exporter

Pilus

Recipient cell (F⁻)

Relaxosome

T DNA

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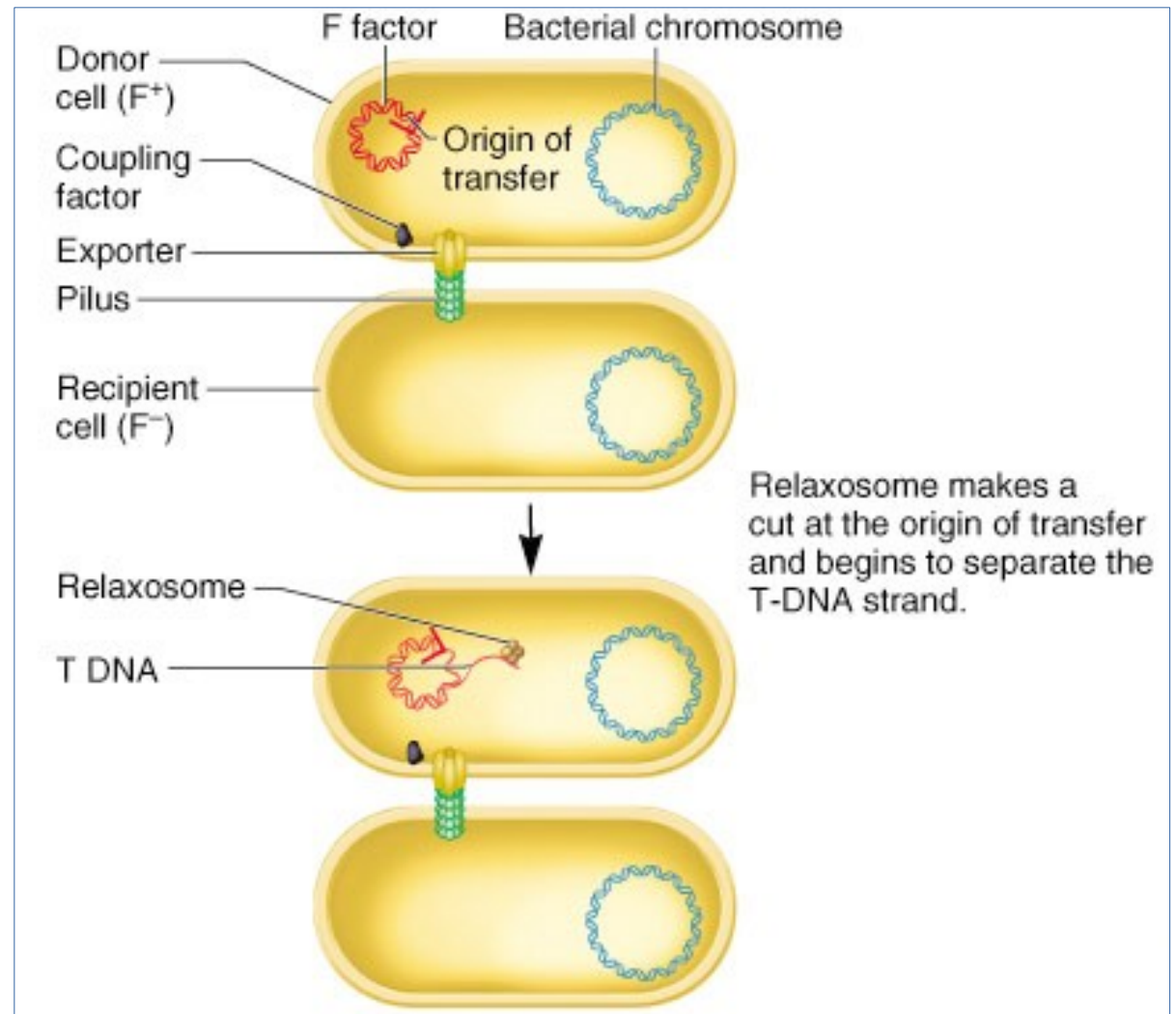
Mechanism of Transfer

- 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.
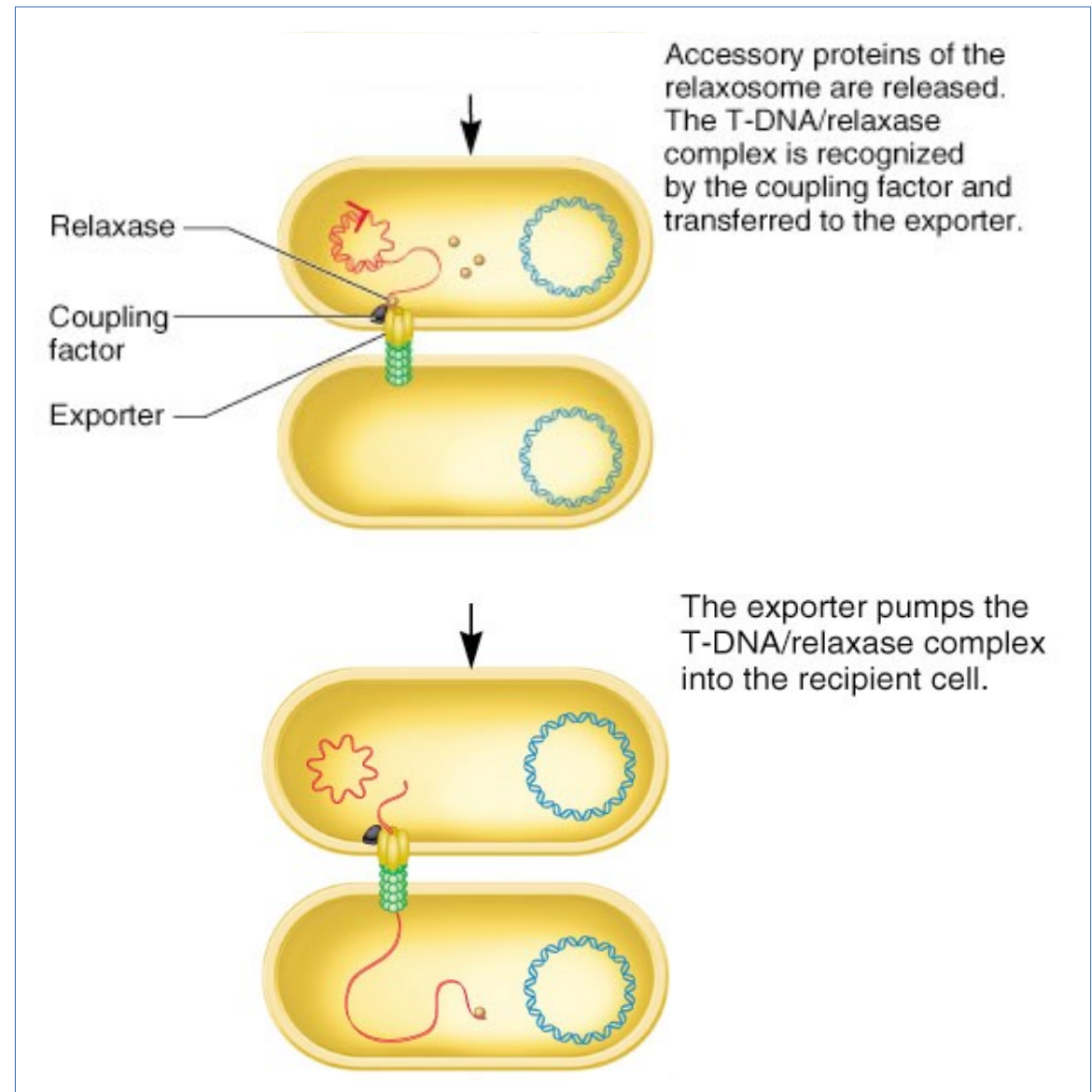
F Factor Transfer

- 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
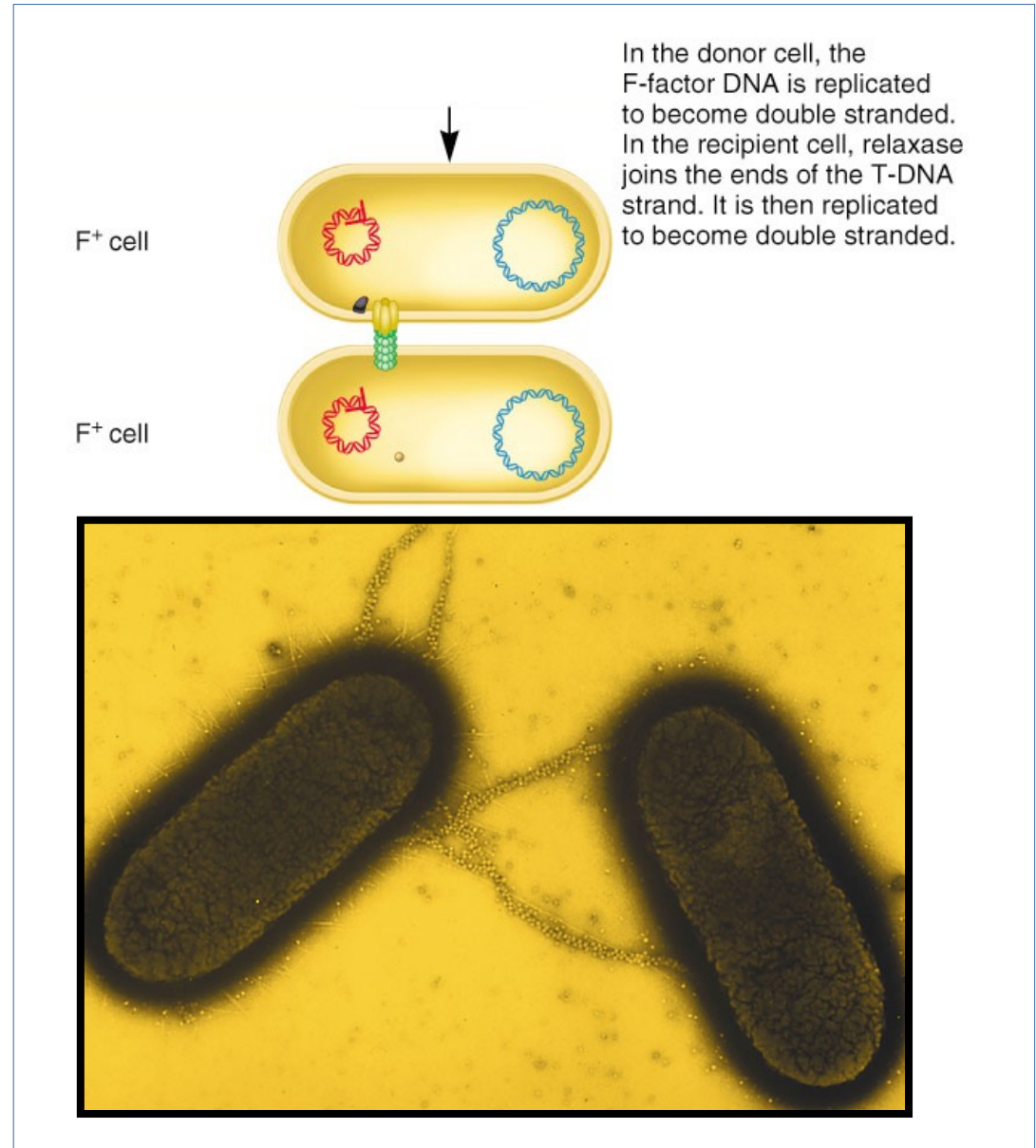
Integration of DNA into Chromosome

- 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

F⁺ cell

Hfr cell

Bacterial chromosome

Origin of transfer

F factor

pro⁺
lac⁺

pro⁺
lac⁺

Origin of transfer

Episome:
DNA fragment that can exist as a plasmid and integrate into chromosome
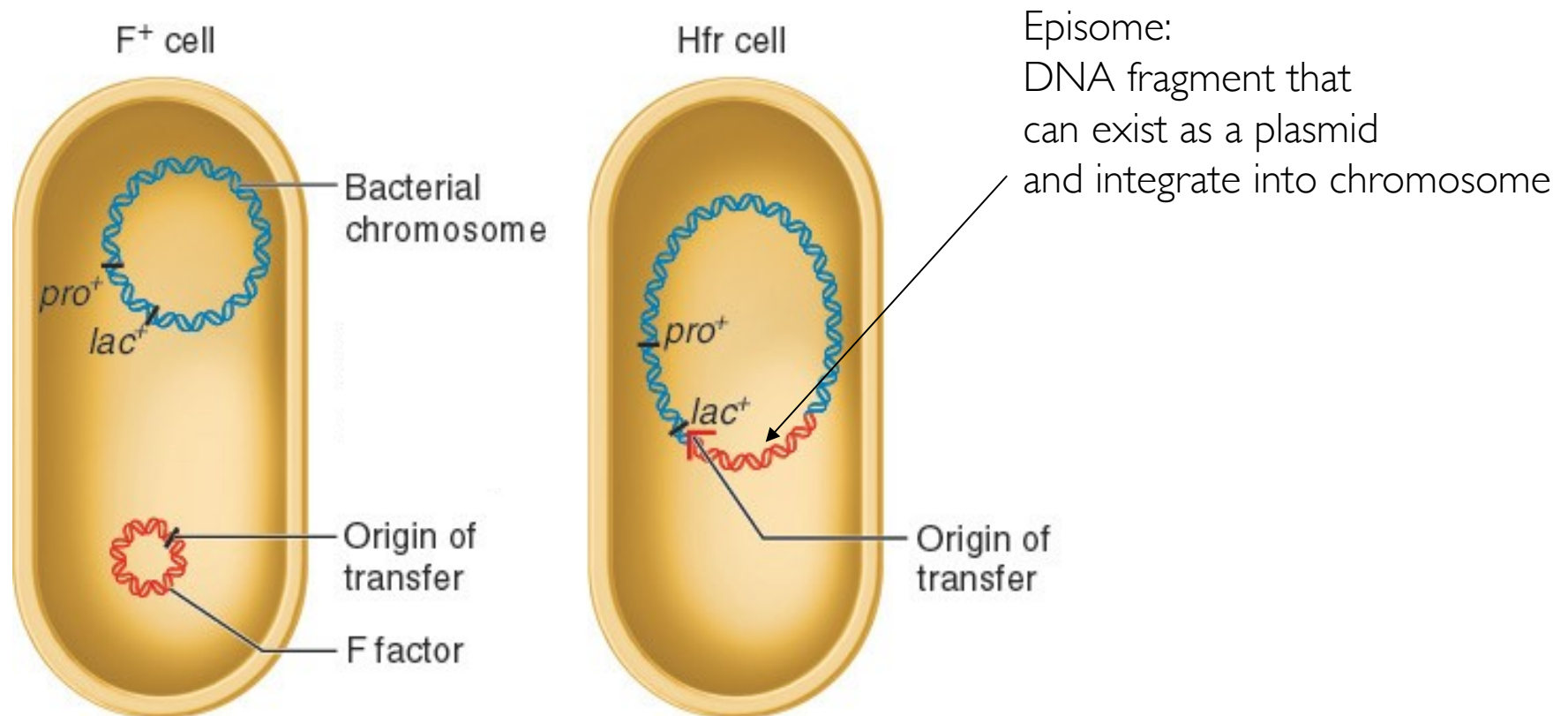
Known Instances of HGT

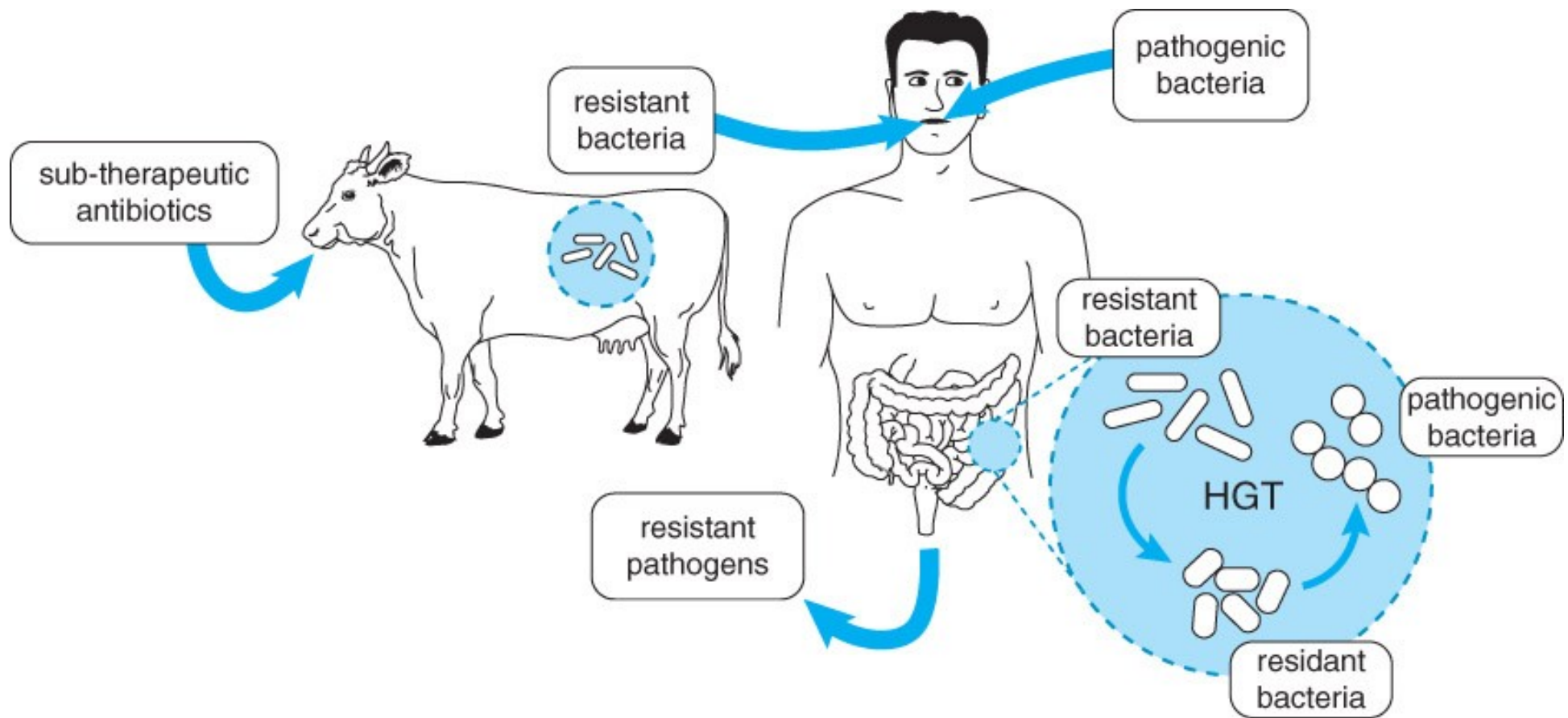- Antibiotic resistance genes on plasmids

- Insertion sequences

- Pathogenicity islands

- Toxin resistance genes on plasmids

- Viruses and viroids

- Organelle to nucleus transfers

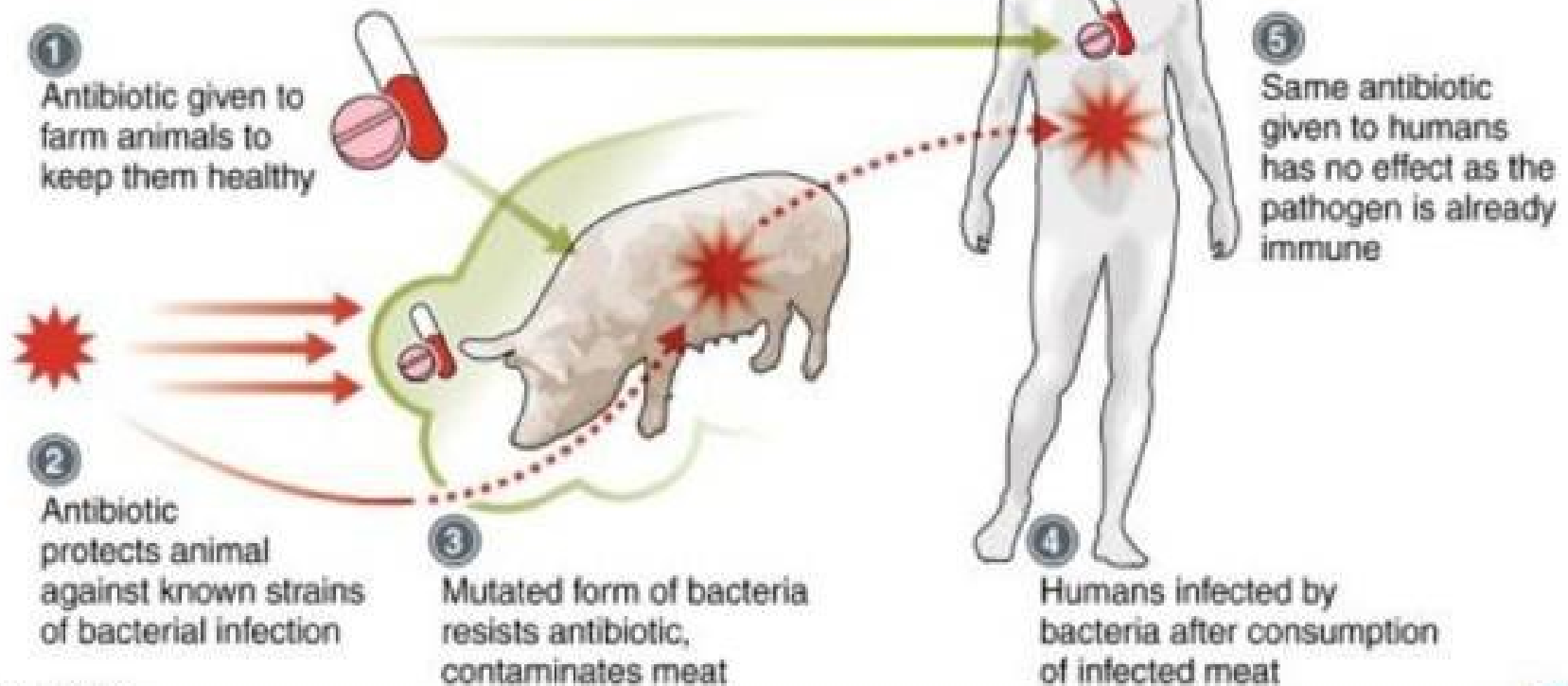# Horizontal Gene Transfer of Antibiotic Resistance

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

1. Antibiotic given to farm animals to keep them healthy
2. Antibiotic protects animal against known strains of bacterial infection
3. Mutated form of bacteria resists antibiotic, contaminates meat
4. Humans infected by bacteria after consumption of infected meat
5. Same antibiotic given to humans has no effect as the pathogen is already immune

Source: WHO

AFP

Supporting Videos

- Horizontal Gene Transfer animation (1.5mins)
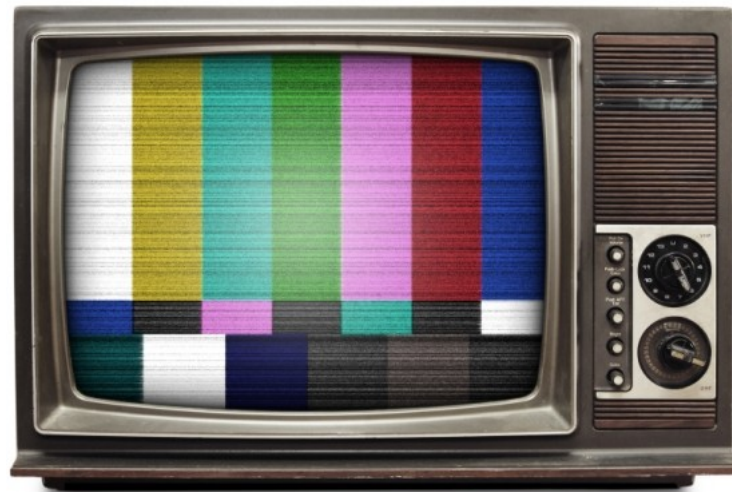
  – https://www.youtube.com/watch?v=tCZhkxpzxDY

- Antibiotic Resistance (4mins)

  – https://ed.ted.com/lessons/how-antibiotics-become-resistant-over-time-kevin-wu

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
                         .**: ****: ** ***.***
```

This 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 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