Microbiology of Fermented Foods & Beverages - MFFB

Notes for the course NFOK14019U, at the University of Copenhagen September-November 2024

Lucas Daniel Paz Zuleta, TZS159

MSc student at University of Copenhagen

Last compiled: 17-10-2024

Contents

Chapter 1
Course description

1.1 Course content

The course has major focus on the microorganisms involved in the processing of various fermented foods and beverages. The course includes the taxonomy of important microorganisms especially lactic acid bacteria (LAB) and yeasts covering both phenotypic characteristics and molecular typing techniques for their identification. Fermentation is introduced as a sustainable green technology and innovative technologies to improve fermented foods and beverages are considered.

Methodologies for isolation will be covered, including both culture dependent and culture independent techniques. Techniques such as high-throughput sequencing and digital programmes for bioinformatics are applied.

Various fermentation techniques is introduced covering the use of starter cultures and other fermentation techniques such as back-slopping. Further, the role of fermentation in sustainable food production and in prevention of food waste will be discussed.

An introduction to various fermented foods and beverages will be given including products such as cheese, bread, wine and beer as well as a number of traditional indigenous fermented food and beverages. Focus will additionally be on microbial interactions including topics such as quorum sensing, bacteriocin formation, etc.

1.1.1 Learning objectives

The objective of the course is to give the students a thorough knowledge on the microbiology behind production of fermented food and beverages and to give the students skills within isolation and identification of microorganisms occurring in these products. Additionally the students will be able to evaluate the functionalities and applications of microbial starter cultures.

Knowledge

- Show overview of fermented food and beverages in general and the microorganisms involved in their production
- Describe important groups of microorganisms identified from fermented food and beverages
- Comprehend microbial taxonomic systems
- Describe microbial interactions and their importance in food systems
- Reflect on microbial cytology and physiology
- Define molecular techniques for identification and typing to species and strain level

Skills

- Apply procedures for isolation and identification of the predominant microorganisms in fermented food and beverages
- Explain at the molecular level the behaviour and interactions between various groups of microorganisms
- Assess the most important parameters leading to optimal product quality and food safety
- Apply food fermentation to develop innovative food products

Competences

- Predict the composition of the microbiota of specific fermented food and beverages
- Discuss presumed functionalities of microorganisms in fermented food or beverages related to product quality and food safety
- Communicate and work independently on own data and discuss the results in relation to existing literature

1.1.2   Teaching and learning methods

Lectures, theoretical and laboratory practicals. The lectures will introduce issues of importance for the understanding of microbial behaviour during production of fermented food and beverages. The theoretical and laboratory practicals will give the students practice on how to identify various microorganisms from fermented food and beverages including skills within various molecular techniques and digital tools for bioinformatic purposes. Knowledge on food innovation will be obtained throughout the course.

1.1.3   Exam

At the time of the oral examination, 20 min, one theoretical question is drawn, and the examination proceeds without preparation time. Two weeks before the exam, the questions will be given to the students. At the oral exam, the drawn question and the curriculum will account for 75% of the grade. The discussion of the laboratory work and the experimental results will account for 25% of the grade.

Chapter 2
Lecture Notes

2.1 04.09.24 - Microscope

2.1.1 Parts of the microscope

The Microscope is put together by different components, as shown in Figure 2.1.
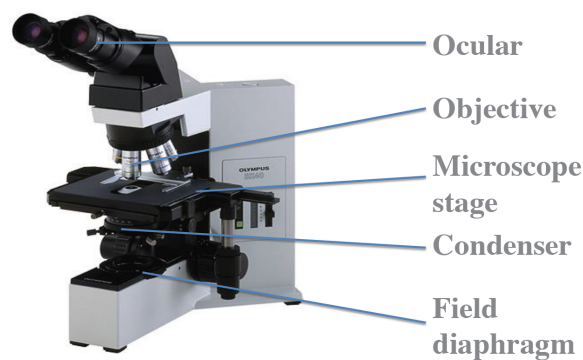
Figure 2.1: The different components of a microscope.

The main components are:

- Ocular

  This is the set of lenses that you peer into, when you use a microscope. The "only" purpose of this part is to relay the intermediate image created by the objective to the retina of our eyes, i.e. to make us see an image. The ocular also magnifies the intermediate image with the factor of the ocular (usually around 10x); however, increasing the magnification of the ocular does not increase the resolving power (finer details of the specimen will not be revealed)

- Objective

  Resolves the structure of the specimen, creating an intermediate image inside the microscope. The image is magnified with the factor depicted on the objective (often 10X, 40X and 100X). The objectives can be changed by rotating the objective turret, but this must be done cautiously to avoid mechanical damage. Objectives fitted for phase contrast will also be labelled with the size of the phase ring (usually this is depicted on the objective as either PH1, PH2 or PH3)

- Microscope stage

   The (translation) microscope stage holds the specimen firmly with a specimen holder, and allows the specimen to be moved in fine increments in two dimensions

---

- Condenser

   * The condenser is located underneath the microscope stage. The purpose of this optical part is to illuminate our specimen optimally

   * Our microscopes for microbiological work have a condenser with a rotating turret that enables the use of phase contrast imaging, as a specific annular ring in this turret must be matched with the size of the phase ring in the objective. There is a cut-out on the front of the turret that depict the actual selection. In this picture the Ph2 ring has been selected

   * Apart from the phase rings, the turret also includes a blank opening that allows for traditional brightfield observation depicted with a 0, and a setting for darkfield depicted with a DF

2.1.2 Contrast methods

Brightfield

This simply magnifies the specimen. However it is almost impossible to see cells, without a staining procedure, that selectively stains parts of the cells. Quite often, a stained specimen is also fixed, which means that the cells have been killed, so for live specimens (e.g. live bacteria), we need a different contrast method.

Darkfield

This is only useful for 10X objectives or smaller, so for observing virtually all living cells.

Phase contrast

This is the best method. The theory behind phase contrast is a little complicated (but hey - the guy won the Nobel Prize), but you can utilize phase contrast without knowing the theory. All you have to remember is:

- The number on the phase objective and the number on the condenser turret must match to obtain a proper image (e.g. Ph 2 for the 40 x objective)
- The Field diaphragm in the bottom of the microscope should be completely open

2.1.3 How to use the microscope

Preparing the microscope

Set the ocular distance to match YOUR interpupillary distance; The distance between the eyes of people differs a lot. Our microscopes are equipped with two eyepieces, in order for you to look down the microscope in the same way as you look through a binocular, and ultimately the same way as you look on the world as such. Therefore, the first thing to do when you approach a microscope is to adjust the ocular distance to your individual interpupillary distance. The ocular distance can be read on a mm-scale (commonly 50-75) between the oculars.

The interpupillary distance is extremely easy to obtain, because you simply ask a friend to measure the distance between your pupils with an ordinary ruler. Memorise this value (because it is the same for the rest of your life), and adjust the oculars to your distance before you do anything else. This increases the comfort of using a microscope tremendously.

2.2   04.09.24 - Yeast: Cytology, taxonomy and physiology

2.2.1   Background

Yeast are quite different from bacterias, first of all, they are bigger than bacterias. Yeast are also eukaryotic, which means that they have a nucleus, and they have a lot of organelles. The yeast *Saccharomyces cerevisiase* has a pair of 16 chromosomes, which is quite a lot for a yeast and thus making it a non-haploid. The many differences from bacteria will influence their identification, their growth characteristics as well as the interactions with the food matrix. Proper identfication of yeast is therefore important for the correct use of starter cultures as well as for spoilage yeasts.

2.2.2   Eukaryotes

Since yeast are Eukaryotes, here are some key features of Eukaryotes:

- Eukaryotes contain far more DNA than prokaryotes (x 100-1000). Yeast genome sizes range from 10-15 Mb and encode 5000-10000 genes
- The DNA is organised in a number of chromosomes
- The chromosomes are organised in a nucleus
- The presence of mitochondria
- The presence of different organelles such as: the Golgi apparatus and the endoplasmic reticulum
- Besides partly being involved in killer toxin production, the function of plasmids in yeasts is not really known

2.2.3   Organelles

- **CW: Cell wall (physical protection, adhesion etc.)**
- P: Periplasm
- **CM: Plasma membrane (transport in/out of the cytoplasma)**
- CMI: Invagination
- BS: Bud scar
- C: Cytoplasm
- **N: Nucleus (DNA in chromosomes)**
- **M: Mitochondrion (ATP synthesis, enzymes)**
- S: Secretory vesicles
- **V: Vacuole (protein trafficking, storage)**
- PER: Peroxisome
- ER: Endoplasmic reticulum
- G: Golgi apparatus

In this course, we will focus on the cell wall, the plasma membrane, the nucleus, the mitochondrion and the vacuole. Though the other listed organelles are important, they are not as important for the physiology of yeast.

Cell wall

The cell wall gives the yeast physical protection against the environment. It also maintains the shape of the yeast cell, and it is important for the osmotic stability of the yeast cell. The cell wall also has permeability barriers, which means that solutes larger than 620-760 Da can not pass the cell wall. The cell wall also has enzyme support, cation binding, cell-cell recognition and adhesion to other surfaces.

Plasma membrane

The primary functions of the yeast plasma membrane are to dictate what enters and what leaves the cytoplasma i.e. to be a primary barrier for passage of hydrophilic molecules and prevent cytoplasmic contents to mix with the environment. The plasma membrane is also important for the uptake of nutrients, the excretion of waste products, the maintenance of the cell shape and the cell wall, and the cell division.

Nucleus

2.2.4 Methodologies for species identification of yeast

There are various methods for species identification of yeast, including:

- Micro- and macromorphological identification added phenotypic characterisation
- Sequencing of the D1/D2 domain of the 26S rRNA gene (> 99% homology = belong to the same species)
- Sequencing of multiple genes
- Sequencing of the ITS region (mostly moulds and yeasts identified by metagenomics)
- Identification of protein patterns by MALDI-TOF MS
- Analysis of biomas by Fourier Transform Infrared Spectroscopy (FTIR)
- Others...

2.2.5 Yeast taxonomy

The taxonomy of yeast is quite complex, but it is important to know the different levels of taxonomy. The taxonomy of yeast is as follows:

- Kingdom: Fungi (Both yeast and moulds are fungi)
- **Phylum:** e.g Ascomycota // Basidiomycota
- **Class:** e.g *Saccharomycetes*
- **Order:** e.g *Saccharomycetales*
- Family: e.g *Saccharomycetaceae*
- Genus: e.g *Saccharomyces*
- Species: e.g *Saccharomyces cerevisiae*

The following is not universally recognized, but are used in some contexts:

- Subspecies: e.g *Saccharomyces cerevisiae* subsp. *cerevisiae*
- Variety: e.g *Saccharomyces cerevisiae* var. *cerevisiae*
- Strain: e.g *Saccharomyces cerevisiae* subsp. *cerevisiae* strain *S288C*

2.2.6 Substrates for yeasts

Following are some examples of some substrates which are greatly used for the propagation of yeasts:

- Malt yeast extract Glucose Peptone agar (MYGP)
- Yeast extract Peptone Glucose (YPG) agar
- Malt Agar (MA)
- Lysine agar a.o. (for non-Saccharomyces yeasts)
- WLN medium (allows separation by coloration)
- Others...

There are som stress tolerant yeasts where high concentrations of substrates are needed in the media for the growth of specific yeast. Some examples of medias for stress tolerant yeasts can be seen in 2.2.

Purpose	Media
Moderately xerotolerant yeasts	Yeast nitrogen base + sodium chloride (10% (w/v)) + glucose (5 (w/v)
Xerotolerant yeast	50% (w/v) glucose-yeast extract agar
Xerotolerant yeast from foods of high salt and sugar	Wort agar + sucrose (3.5% w/v) + glucose (1.0% w/v)
Zygosaccharomyces rouxii	Potato-dextrose agar + sucrose (60% w/v)
Sugar-tolerant yeast from concentrated orange juice	Glucose-citric acid-tryptone agar
Isolation and cultivation of xerotolerant yeast	Malt-extract agar + glucose (2, 20, 40, 50% w/v)

Figure 2.2: Examples of medias for stress tolerant yeasts.

2.2.7 Conventional methods for identification of yeast

There are various methods for identification of yeast, here are a list of some of the most common methods:

- By their cultural characteristics
  - Solid media: colony morphology, color, size, shape, surface, margin, elevation, texture
  - Liquid media: turbidity, sediment, pellicle, ring, flocculation
- By their cell morphology and arrangement
  - Cell form and size
  - Vegetative reproduction by budding or fission
  - Pseudomycelium or true mycelium
- By their sexual characteristics
  - Perfect stage (*teleomorph*) or imperfect stage (*anamorph*)
  - By their formation and arrangement of spores
- Biochemical characteristics
  - Assimilation of carbon or nitrate sources (e.g. API20c, ID32c)
  - fermentation of carbohydrates
- Physical and chemical tolerance
  - Growth in the presence of 100 or 1000 ppm cycloheximide
  - Growth in the presence of 1% acetic acid
  - Growth on 50% glucose-yeast extract agar
  - Etc.

2.2.8 Macro-morphological characteristics

There are many different macro-morphological characteristics of yeast, some of the ones we will be using in this course laboratory exercises can be seen in figure 2.3.
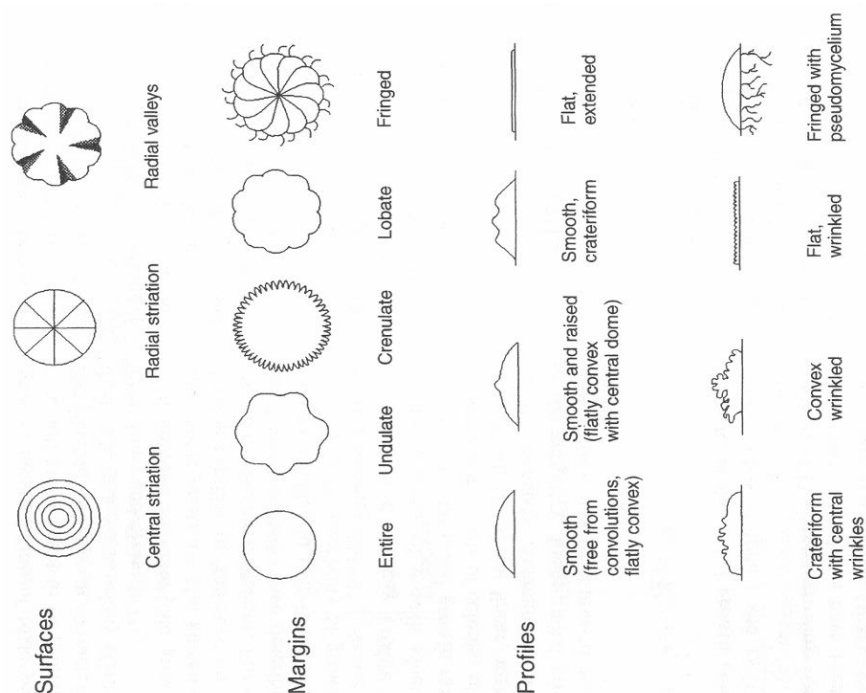
Figure 2.3: Examples of macro-morphological characteristics of yeast that will be evaluated in the laboratory exercises.

Furthermore, the lecture included a list of some key aspects to keep in mind while looking at the macro-morphological characteristics colonies. These are:

- Tecture (mucoid, viscous, matted, coherent etc.)
- Color
- Size
- Surface (glistening, smooth, verrucose, rough etc.)
- Elevation
- Margin (entire, rhizoid etc.)
- Mycelium (x100 magnification)

2.2.9   Micro-morphological characteristics

Micro-morphological characteristics are also important when identifying yeast. Some of the key aspects are listed below:

- Cell form (spheroidal, ellipsoidal, ovoid, lemon-shaped, elongated, triangular etc.)
- Cell size (length, width)
- Cell arrangement (pairs, aggregates)
- Vegetative reproduction
    Budding (monopolar, bipolar, multilateral)
    Fission
- Mycelium
    Pseudomycelium (not distint septa)
    True mycelium (distint septa)

2.2.10   Assimilation and fermentation

Assimilation refers to the process by which an organisms takes in and incorporates nutrients from the environment. The assimilation of carbon compounds is important for the growth of yeast. Yeast assimilates carbon compounds by various pathways, like Krebs cycle (TCA pathway) or the glycolysis.

Fermentation is the process by which yeast converts sugars into alcohol and carbon dioxide. The fermentation of carbohydrates typically happens under anaerobic conditions. The fermentation process is a less efficient energy source than respiration. Nevertheless, without it, wine and beer could not be made. Brewers therefore exploit the yeasts fermentative capability by introducing yeast into an anaerobic environment.

Assimilation of carbon compounds

To propagate yeast while having assimilation in mind, the growth medium Yeast Nitrogen Base from "Difco" can be used. This medium contains:

- Hexoses
- Trisaccharides
- Pentoses
- Organic acids
- Disaccharides
- Polysaccharides
- Alcohols
- Glycosides

Fermentation of carbon compounds

Here are a list of some of the carbon compounds that yeast can ferment:

- Glucose
- Maltose
- Trehalose
- Galactose
- Raffinose
- Melibiose
- Saccharose
- Lactose
- Inulin

Assimilation of nitrogen compounds

To propagate yeast while having assimilation in mind, the growth medium Yeast Carbon Base from "Difco" can be used. This medium contains:

- Nitrate
- Ethylamine hydrochloride
- L-lysine
- Creatine
- Nitrite
- Cadaverine dihydrochloride
- Creatinine

2.2.11   Vegetative reproduction (genera)

Vegetative reproduction is the process by which a plant or fungus reproduces asexually. Yeast can reproduce vegetatively by budding or fission. Some examples of yeast vegetative reproduction is:

Bipolar
- *Hanseniaspora (Kloeckera)*
- *Rhodotorula**
- *Saccharomycodes*
- *Trichosporon*

Fission
- *Schizosaccharomyces*

Arthroconidia
- *Galactomyces (Geotrichum)*
- *Saccharomycopsis**

Enterobalstic Budding
- *Phaffia*

Multilateral
- *Candida*
- *Debaryomyces*
- *Dekkera (Brettanomyces)*
- *Kazachstania*
- *Kluyveromyces*
- *Pichia*
- *Rhodotorula**
- *Saccharomyces*
- *Saccharomycopsis**
- *Torulaspora*
- *Yarrowia*
- *Zygosaccharomyces*

*Two types of vegetative reproduction my be present

2.2.12 Lecture conclusion

- Yeasts are predominantly unicellular eukaryotes
- Yeasts might have both imperfect and perfect names
- Yeast taxonomy is changing continuously especially due to new molecular techniques
- Yeasts are currently mostly identified by sequencing of the D1/D2 region of the 26S rRNA gene
  – however, never forget to verify by examining morphology and include phenotypic tests
- Remember to consult the newest scientific literature in order to use up dated taxonomic names

2.3 09.09.24 - Taxonomy and the species concept

2.3.1 Lecture content

During this lecture the following topics will be the focus point.

- Definitions of taxonomy and phylogeny
- Phylogeny
- Microbial taxonomy: from domain to species
      Species concept in Bacteria
      Species concept in Yeast and Moulds
- Subspecies, variant and strains

2.3.2 Taxonomy and phylogeny

From figure 2.4 we can see that the common ancestor of Bacteria, Archaea and Eukaryota is the root of the tree, and the branches of the three domains is what sums up the tree.

Figure 2.4: The comon ancestor of Bacteria, Archaea and Eukaryota and the som of their respective branches

A definition **Taxonomy**, this the naming and classification of organisms based on shared properties, which can be based on

- Genotypic properties
- phenotypic properties
- Phylogeny - evolutionary relationships

While working with taxonomy and phylogeny in a broder sense, some of the following terms are important to know:

- **Isolation:** The separation of a pure culture from a mixed culture
    Sampling: The collection of a representative part of a population
    Colony purification: The isolation of a single colony from a mixed culture
    DNA extraction: The isolation of DNA from a pure culture
- **Identification:** The determination of the taxonomic position of an organism
    Microscopy: The determination of the morphology of an organism through a microscope
    Sequencing: The determination of the sequence of a gene or a genome

While looking at the phylogeny we can look deeper into the classification, in table 2.1 we can see an example of the classification of a Bacteria and Eukarya where we stop at species. This can be further divided into subspecies, variant and strains.

Table 2.1: An example of the classification of a Bacteria and Eukarya

Domain	Bacteria	Eukarya
Kingdom	None assigned	Fungi
Phylym	Firmicutes	Ascomycota
Class	Bacilli	Saccharomycetes
Order	Lactobacillales	Saccharomycetales
Family	Lactobacillaceae	Saccharomycetaceae
Genus	*Lactobacillus*	*Saccharomyces*
Species	*L. acidophilus*	*S. cerevisiae*

In the following table, there will be listed some differences between prokaryotes and eukaryotes.

Table 2.2: Differences between prokaryotes and eukaryotes

	Prokaryotes	**Eukaryotes**
DNA	DNA is naked DNA is circular Usually no introns	DNA is bound to protein DNA is linear Usually has introns
Organelles	No nucleus No membrane-bound 70S robossoms	Has a nucleus Membrane-bound 80S ribosomes
Reproduction	Binary fission Single chromosomes (haploid)	Mitosis & meiosis Chromosomes paired (diploid or more)
Average size	Smaller (1-5 μm)	Larger (10-100 μm)

2.3.3   Gram-positive & -negative bacteria

(Almost) all bacteria can be divided into two groups, Gram-positive and Gram-negative bacteria. The difference between the two groups is the cell wall. The cell wall protects the bacteria e.g against osmotic stress. The cell wall differs where the Gram-positive bacteria have a thick cell wall, while the Gram-negative bacteria

have a thin cell wall. The difference in the cell wall is due to the peptidoglycan layer, which is thicker in the Gram-positive bacteria than in the Gram-negative bacteria. The peptidoglycan layer is also the reason why the Gram-positive bacteria are more resistant to antibiotics than the Gram-negative bacteria. Here are some bulletpoints of the two groups:

- **Gram-positive bacteria**
    Lactic Acid Bacteria (LAB)
    Bacillus
- **Gram-negative bacteria**
    Most Acetic Acid Bacteria (AAB)

2.3.4   Differentiating between bacterial species

There are several methods to differentiate between bacterial species, some of the most common methods are:

- **DNA:DNA hybridization (DDH)**:
    The DNA of two organisms is mixed and the degree of hybridization is measured.
    If DDH > 70% the organisms are considered to belong to the same species
- **16S rRNA gene sequencing**:
    The 16S rRNA gene is sequenced and compared to a database, if the sequence is > 97.5-98.5% identical to a known species, the organism is considered to belong to the same species
- **Whole genome sequencing**:
    Average nucleotide identity (ANI) > 95-96% is considered to belong to the same species
    Digital DNA-DNA hybridization (dDDH) > 70% is considered to belong to the same species

2.4   11.09.24 - Methodologies for identification and typing

2.4.1   Intended leaning outcomes

After this lecture, the students should be able to:

- Get to know how different microbial identification and typing methods works
- Apply this knowledge to comprehend microbial identification and taxonomy
- Reflect on the application possibilities of the different methods for microbial identification and typing

2.4.2 Background

- When you want to know what microorganisms are in a fermented food
    IDENTIFICATION
- Often identification is coupled with quantification (predominant/minor microorganisms)
- Many different methods are available
    Make an overview of the most used methods
    Elaborate the steps of the methods → insights into the structure/mechanistics
    Show how the methods are applied
- All methods have strengths and limitations

Note: typing is identification of isolates below species level

2.4.3 Methodology overview

Culture dependent methods

This is the most common method for identification of microorganisms. The method is based on the growth of the microorganisms on a solid or liquid medium. Here are an example of a method which is based on this:

- Plating
- Isolation
- Purification (The next 4 steps are methods for identification)
- Finger print based[1]
    rep-PCR, RAPD, RFLP or PFGE (All these are based on PCR)
- Sequence based[2]
    rRNA gene, MLST or MLSA
- Biochemical identification[3]
    MALDI-TOF or FTIR
- Classical[4]
    Macro/micro morphology, phenotypic tests (fermentation, assimilation or growth conditions)

Culture independent methods

- Direct extraction of DNA or RNA
- Molecular methods (the next two steps are both molecular methods methods)
- PCR[1]
    Species specific PCR
- qRT-PCR[2]
    Next generation sequencing

2.5 16.09 - From 1st to 3rd generation DNA sequencin strategies

DNA is quite small, 2 nm thick. Though, if folded out, the DNA would be around 1.8 m long.

2.5.1 Sanger Sequencing of DNA 1st generation sequencing

5'- TACAACTGAGCGACT -3'
3'- ATGTTGACTCGCTGA -5'
Here we have a blue sequence and a black complimenting sequence. If we want to sequence the blue sequence, we can use the black sequence as a primer. The primer will bind to the blue sequence, and the DNA polymerase will start to build a new strand of DNA. The DNA polymerase will build the new strand by adding nucleotides to the 3' end of the primer. The nucleotides are added in a 5' to 3' direction. The DNA polymerase will continue to add nucleotides until it reaches the end of the DNA strand. The DNA polymerase will then stop, and we will have a new strand of DNA.

Though with the Sanger sequencing, we will have a mix of DNA strands, where each strand will end with a different nucleotide. The DNA strands will then be separated by size, and the sequence can be read by the size of the DNA strands.

With Sanger sequencing, the -A, -C, -G and -T nucleotides will cut the DNA strand at different lengths, because it cuts at the respective letter. An example of how the Sanger sequencing works can be seen in figure 2.5.
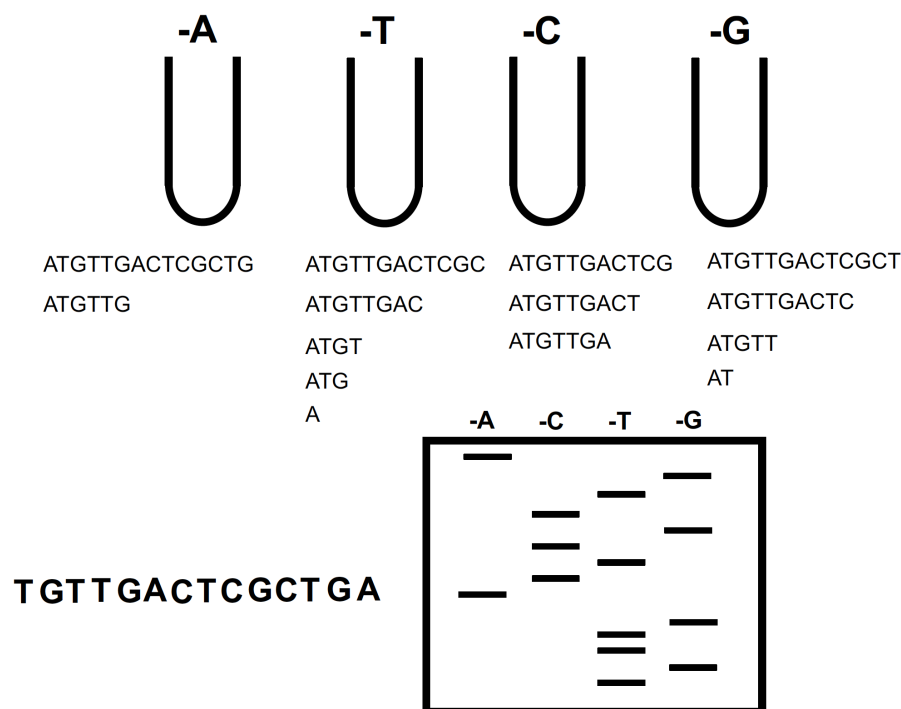
Figure 2.5: An example of how the method of Sanger sequencing of DNA works with the given blue sequence

2.5.2 Next generation sequencing

Next generation sequencing (NGS), has been evolving a lot since extensive research started in the 90's. Today we have a lot of different NGS methods, one of the most common methods is the Illumina sequencing. The

Illumina sequencing is based on the Sanger sequencing, but with a few modifications. Their HiSeqX can sequence around 50 human genomes in a day with a cost sub 1000 USD/genome.

Sequencing by synthesis

The Illumina sequencing is based on the sequencing by synthesis. It needs a primer, but these a fairly simple to make. After the primer has been added, the DNA polymerase will start to build a new strand of DNA. The DNA polymerase will add nucleotides from -A, -T, -G, anc -C, to the 3' end of the primer, and the nucleotides will be added in a 5' to 3' direction. The DNA polymerase will continue the flashing with nuclouetides until it reaches the end of the DNA strand. The DNA polymerase will then stop, and we will have a new strand of DNA.

The maximum read out length of the Illumina is between 75 bp to 350 bp, but the average read length is around 150 bp.

Bridge amplification

A piece of DNA is attached to a glass slide, and the DNA is then amplified. The DNA is then denatured, and the DNA is then amplified again. The DNA is then denatured again, and the DNA is then amplified again. This process is repeated until the DNA is amplified enough. The DNA is then sequenced, and the sequence is read by the computer. This method grows exponentially like PCR, though the DNA is amplified in a half circle, and not in a straight line.

An example of how the Illumina bridge amplification works can be seen in figure 2.6.

(a) CHere an example of how the DNA is attached to the glass slide on both ends forming the bridge.

(b) After heating the bridge opens up and the DNA is amplified.
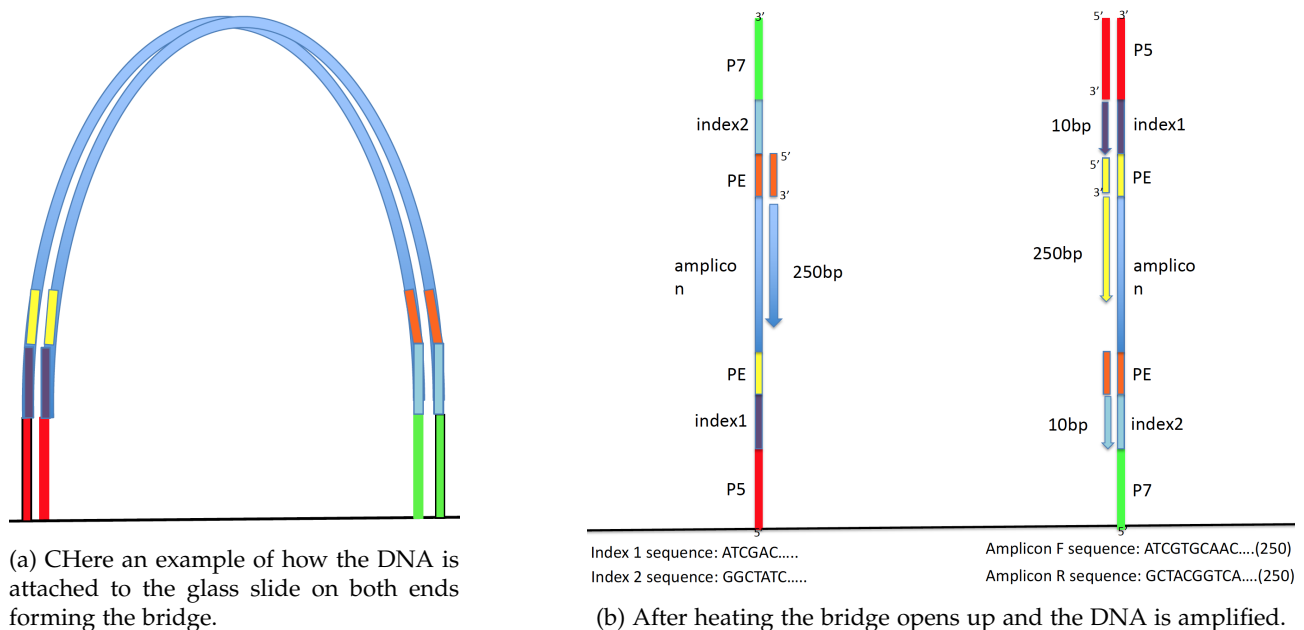
Figure 2.6: An example of how the Illumina bridge amplification works.

2.5.3 3rd generation sequencing

Chapter 3
Laboratory Exercises

Chapter 4

Lecture Exercises

4.1 11.09.24 - Exercise 1 - Replication, Transcription and Translation in prokaryotes

For an automated version of answering this exercise, please see the file w2e1.py in the Python folder. This file will automatically replicate, transcribe and translate the given DNA sequence. Use the file as a template for future exercises if needed.

Figure 4.1 shows a DNA sequence which will used in Exercise 1 and its sub tasks. The given DNA sequence can be seen in figure 4.1

5'-_____-3'

3'-G A T T A G G T A A C T G T G A T T C G T A C G T A A C G T G A C G A T A T T A G C A T C C A C C G C A T A C A G A C G A T A T G C A T A G C T G A T C A T A T C G C C-5'

Figure 4.1: Sequence of DNA which is used in Exercise 1

4.1.1 Replication in prokaryotes

Replicate the DNA sequence based on the anti-sense (template) strand written below using the following primer **5'- AUCCAUU -3'** (This RNA complementary to the DNA sequence and is the defined -35 region, which will be used in the transcription subsection).

- While replicating the DNA, identify the leading and lagging strands. Furthermore remember that G and C are the most stable base pairs, A and T are the least stable base pairs
- The DNA sequence is replicated from the 3' to 5' direction
- Use the primer as to define where to start from

5'- X X X **A T C C A** **T T G A C A** C T A A G C A T G C A T T G C A C T G C **T A T A A T** C G T A G G T G G C G T **A T G** T C T G C T A T A C G T A T C G A C T A G T A **T A G** C G G -3' (This is the replicated)

3'- *G A T* **T A G G T A A** C T G T G A T T C G T A C G T A A C G T G A C G A T A T T A G C A T C C A C C G C A T A C A G A C G A T A T G C A T A G C T G A T C A T A T C G C C-5' (This is the given)

So, to begin with, we will look at the given RNA primer, and convert it to a DNA primer. 5'-AUCCAUU-3' will be converted to 5'-ATCCATT-3' (since RNA has Uracil instead of Thymine).

Afterwards we will look at the given DNA sequence and find the complementary strand to it. The complementary strand will be the anti-sense strand. 5'-ATCCATT-3' will be converted to 3'-TAGGTA-5' (since A pairs with T, T pairs with A, C pairs with G, and G pairs with C).

Now we will highlight the given sequence in <mark>yellow</mark> and start translating the DNA sequence from the 3' to 5' direction. We will start from the primer and move towards the end of the sequence.

This concludes this part of the exercise.

4.1.2 Transcription in prokaryotes

Identify the Pribnow (-10 region) and (-35) regions in the above DNA double-helix (sense strand). Transcribe DNA into a Messenger-RNA (mRNA).

In this course the Pribnow box (-10 region) is defined as the region 5'-TATAAT-3' and the -35 region is defined as 5'-TTGACA-3'. Wil will therefore

We found the Pribnow box sequence and highlighted it in <mark>green</mark>. The mRNA strand will be transcribed from the anti-sense strand, so we will use the antis sense strand to transcribe the mRNA until the -35 region which we highlighted in <mark>turquoise</mark>.

We can now go to the right end (10-ish letters) of the pribnow box and start transcribing the mRNA. In this course, the start codon is defined as the first AUG codon after the Pribnow box. The mRNA strand will be transcribed from the 3' to 5' direction.

The start codon is highlighted in <mark>apricot</mark> starting from the 13th letter to the right. The stop codon is highlighted in <mark>bittersweet</mark>.

The Pribnow 10 region is **5'- ATCCATT -3'** and the -35 region is **5'- CGATAGG -3'**.

Now that we have identified both the start- and stop codon, we can start to transcribe the mRNA strand. The first codon will start from number 10 after the the Pribnow box. The mRNA strand will therefore be:

<mark>5'- GUA UGU CUG CUA UAC GUA UCG ACU AGU AUA GCG -3'</mark>

4.1.3 Translation in prokaryotes

Translate the mRNA strand from transcription step into amino acids.

The mRNA strand is written in the last subtask, and have already been translated into codons. It's now possible to translate the codons to amino acids.

H_2N- Val - Cys - Leu - Leu - Tyr - Val - Ser - Thr - Ser - Iso - Ala -COOH

4.2 11.09.24 - Exercise 2 - PCR and Restriction Enzymes

A part of DNA fragment shown below is to be amplified using PCR and the primers F and R. Subsequently the PCR product is cut with restriction enzyme HpyCH4V and the resultant products are run on a gel. The

```
              20              40              60              80
              |               |               |               |
5'- TGTCGGTGTG CCATAACCTA AGTCATAACG ATCCGTGAAT GCATACCCGT TTTGAATAAT TGAATCCAAG AAACTTGTAT CTGTACTTGA-3'
3'- ACAGCCACAC GGTATTGGAT TCAGTATTGC TAGGCACTTA CGTATGGGCA AAACTTATTA ACTTAGGTTC TTTGAACATA GACATGAACT-5'
```

Figure 4.2: Sequence of DNA which is used in Exercise 2

The two given primers are:

- Primer F: 5'- AACCT-3'
- Primer R: 5'-TGGAT-3' (3'-TAGGT-5')

4.2.1 What is the size (in base pairs) of the PCR-product?

The Primer F starts from the 3' end of the DNA sequence, and the Primer R starts from the 5' end of the DNA sequence. The PCR product will therefore be the DNA sequence between the two primers.
Primer F has to fit to the upper strand, and Primer R has to fit on the lower strand. The primners sequence are therefore:

- Primer F: 5'-AACCT-3'
- Primer R: 5'-TGGAT-3' -> 3'-TAGGT-5'

Primer F starts from the 15th to the right and ends at the 19th base pair. Primer R starts from the 64th from the right and ends at the 68th base pair. The PCR product will therefore be the DNA sequence between the 19th and 64th base pair.
The size of the PCR-product is therefore 68-14 = 54 base pairs.

4.2.2 How many copies of DNA do you have after 4 PCR cycles if you start with 3 copies of the DNA- fragment to be copied (considering ideal conditions)?

After each PCR cycle, the amount of DNA doubles. After 4 PCR cycles, the amount of DNA will be 2^4 = 16 times the original amount of DNA. Though, we are starting with 3 copies of the DNA fragment, so the amount of DNA after 4 PCR cycles will be $3 \cdot 2^4$ = 48 copies of the DNA fragment.

4.2.3 What is the size of 2 fragments after cutting with HpyCH4V?

The restriction enzyme HpyCH4V recognizes the sequence 3'-ACGT-5'. The DNA sequence is cut at the recognition site, and the two fragments will be the DNA sequence before the recognition site and the DNA sequence after the recognition site.
5'- TGTCGGTGTG CCATAACCTA AGTCATAACG ATCCGTGAAT GCATACCCGT TTTGAATAAT TGAATC-CAAG AAACTTGTAT CTGTACTTGA-3'
3'- ACAGCCACAC GGTATTGGAT TCAGTATTGC TAGGCACTTA CGTATGGGCA AAACTTATTA ACT-TAGGTTC TTTGAACATA GACATGAACT-5'

HpyCH4V will therefore cut the DNA sequence at number 41, and the two fragments will be:
5'- AACCTA AGTCATAACG ATCCGTGAA**T GCA**TACCCGT TTTGAATAAT TGAATCCA-3'
3'- TTGGAT TCAGTATTGC TAGGCACTT**A CGT**ATGGGCA AAACTTATTA ACTTAGGT-5'

There will therefore be 27 base pairs in the first fragment and 27 base pairs in the second fragment.

4.2.4 Place the bands that you would observe after running your product on the an agarose gel (path 1 before cutting with enzyme, path 2 after cutting)

When you run the PCR product on a gel electroforesis, you will see the DNA fragments separated based on their size. The smaller fragments will move faster through the gel than the larger fragments. This means that both the DNA fragment before and after cutting with the restriction enzyme will be visible on the gel since they doesn't have the same fragment size .

Therefore you will see two bands on the gel, one at 54 base pairs (before cutting) and one at 27 base pairs (after cutting). This can be seen in Figure 4.3.
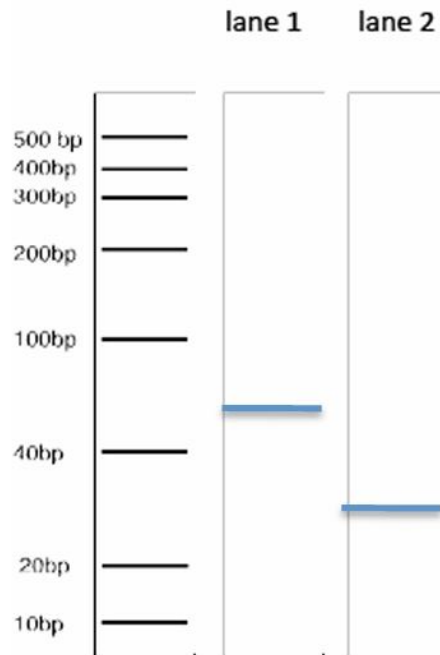
Figure 4.3: Gel electrophoresis of the DNA fragments and the PCR products' bands

4.3 11.09.24 - Exercise 3 - Primer3plus

Based on the sequence below design primers that fulfill following conditions:
The following information is given:

- size (22-24bp)
- Tm (55-57)
- GC% (35-70)
- amplicon size (300-400bp)

The following sequence is given:

```
TTGCAAGCCCTGATTTAAGCATTACACCAGCATCATCAAAGTAATACGCATTTCCATTTTGATCAGTTAT
ATATTGGTCAAAAACTTGATGACCTTTCTTATCGAAATAAACAATTGTACCATCTTCGTTTTGTAAGAAG
CTCTTGACCAGTTCGATACCATTAGGTAAGAAGAAGTAGGTATGGTTATTAATCTTTTGCAAACCTGTTA
CTAAATGACCTGTTTTATCAAAATAGTAATAGTTATTATCATCATCTTGAATAAATGCATTTTGGGCACG
ATAACCACTTAATGTATAATAGATGATTCCTTTATCATCGTGTGTAAAGCCAGTTTCTGACAGATCATTA
GTTAACTGTTTTGGTAAGTAGTCACCATCCTCAGTGTTCGAAACGACCTTAAAGTACTTATCAGAACCCA
TATCTTTCAATACGTATCCAGCACCTTTACCTTGGATGTTAGAGCCATTGAAGTACTTAGCCGCCCACTC
AGTAATCTTTACACTGCCATCAATTGGAACGCCAGTTGAGATTTGATTCACTTTAAATAGGGATGGATAG
AGTGCCTGTAACTCTTCTAAGAAGGCACCACCATACATCTCTTGATATTGACCACCCCCACGACTTTGTA
CAACATATAAGGCATTGTCAATATCAGAATCTGTATCGTCATCTCCAAATGAATTTGTTCTTGTGACAGT
AGCTAATTCTTGCTCTGGCAAATTATAAATTTGGTCCGGCACCCAATCGGCAATGGCTTGAATACCGCTA
GCATGTAAGGCTTTAATAGCATCGCGCAACTGATCAGCAGTTCCATATTTTGTCGGTGTGCCATAACCTA
AGTCATAACGATCCGTGAATGCATACCCGTTTTGAATAATTGAATCCAAGAAACTTGTATCTGTACTTGA
ACGATATTGTGGTGCCAATTGGAAGCTTGTCACACCCCATTGCTTAAATTGGTCCGCATTCTGAGCGATG
```

```
ACTACGTTTGTATATTCACTGCTGTCTGTAGCAAATGCTTGGAAGTTTGAGAAACCTTCGTAGATGACTT
GAGAATCAAGAGCAGCGTTTGAATGGAACACTTTATCACTCGTGTTTGTTGTTGTATCAGAGGCCGTTCG
TGCATCTTGATCTTGTTGCGCACCTACCGGAACCCAAACTGCCAAGTAACCAGAAACTTGTGGATTTTGT
ACACCATAAATTGATTCATTCGTAAAAATCAAATCGCCGTTAGCATCTGTGTACGCCACAGGTGCATTTT
CATCAGTATCATAATAAGCTAATCCATCTGCAGTTGTTGATAACAAAGCACGATAAGCTTGGTTCTTATG
AGCTGCCCCCATATGCAATGTGACAGTATGCCCATCCTCTAATTGTAGCTCCGCGTTATTGCTGACGATG
ACTCCAATACCTTCAGTACGCGTCTCAGATGTTCCAGTGTCAGAAGCCGTCATGGCATCTTTACCATAGC
GAACACTTGTTAACACGTCATTACTATCAACGGACATCGATTGGCCACCAGCAACATATTGGACTCTAGC
CTTCAGCAAAGTGTTAATCGCATCATAGTATGGTGACTTTGTTGCCATATATTGACCATCATCTGTATAT
AAATCGCCATAATAGACACGAGGAACAGTATCCTTGTTGGTTAGCAACATCGCATAAGCACTAGCCATAT
TATATTGTGTGTACTTTTTGTCTGCTAATTTTTCATCTTCATTGTATACTTTGAAAGCAGCTGCCAATTG
```

Use Primer3Plus (http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi/)

While setting the conditions, it's important to remember that the amplicon size is the size of the DNA fragment that will be amplified by the primers.This is named "Product range size" on the web-side under general settings.

Write answers in the table bellow:

Table 4.1: The given table which should be filled out with the information from the Primer3Plus web-side

Primer	Sequence (5'-3')	Primer size (bp)	Tm (°C)	GC content%
Forward	TCAGTAATCTTTACACTGCCATC	23	56.1	39.1
Reverse	TTATGACTTAGGTTATGGCACAC	23	56.1	39.1

The second part of this exercise was to use this information to deduct which bacteria this sequence belongs to. The sequence is from the bacteria *Leuconostoc mesenteroides*.

4.4   11.09.24 - Exercise 4 - Primers check and restriction enzymes

The following information is given:
- Primer F: 5'- GGAACCCAAACTGCCAAGT-3'
- Primer R: 5'-CCTCGTGTCTATTATGGCGATT-3'

Using **"In silico PCR amplification"** (http://insilico.ehu.es/PCR/) answer following questions:

4.4.1   What *Leuconostoc* strain will the primers align to?

We start by going to the web-side. Then we choose *Leuconostoc* and press "Next step". Afterwards we paste the primer sequences into the "Forward primer" and "Reverse primer" fields. We then choose "APPLY TO ALL Leuconostoc" and press "Amplify".
Then we can see 2 bands where the primers has traveled at 485 and 485 base pairs. The primers will therefore align to the 2 *Leuconostoc mesenteroides* strains.

4.4.2   What would be the PCR product size (bp) if we use those primers against that species?

Based on the bands that was produced the PCR product size will be 485 base pairs.

For the next subsections, the following information is given:
Using NEBcutter V2.O http://tools.neb.com/NEBcutter2/ retrieve the amplicon from "In silico PCR ampli-fication" and answer following questions:

4.4.3   What will be the size of DNA products after digesting the amplicon with Hind III restriction enzyme?

To use the web-side you need to copy the DNA sequence from the "In silico PCR amplification" web-side and paste it into the "Enter DNA sequence" field and press sumbit. Then you look at the sumary and find the restriction enzyme HendIII and hold the mouse over it to see where it cuts.

For the two *Leuconostoc* strains the HINDIII cuts at:
   • Cuts at bp 167 (Leuconostoc mesenteroides subsp. mesenteroides ATCC 8293)
   • Cuts at bp 167 (Leuconostoc mesenteroides subsp. mesenteroides J18)

This means that the DNA products will be 167- and 318 base pairs long after digesting the amplicon with HindIII restriction enzyme. This makes sense since the total sequence, for both, is 485 base pairs long. Thus 167+318 = 485 base pairs.

4.4.4   What are the reaction conditions requires for ApeKI?

To answer this, the following table has to be filled out. The information can be found on the NEBcutter web-side by pressing "Enzyme list" and then "ApeKI". Then in the bottom right corner of the first table, there is a button called "NEV Restriction Enzyme - Activity/Performance Chart". Then find the enzyme ApeKI and press on the buffer, there all the needed information is found.

Table 4.2: The given table which should be filled out with the information from the NEBcutter web-side

Buffer	NEBuffer^{TM} r3.1
Salt	100 mM NaCl
Main	50 mM Tris-HCl
pH	7.9 at 25°C
Mg	10 mM MgCl$_2$
rAlbumin	100 µg/ml Recombinant Albumin
Reaction temp. (°C)	75°C (Obtained from the solutions)

4.4.5   What is the digestion pattern of AciI?

The AciI enzyme cuts at the following base pairs:
Primer F 5'-...AA$^{\downarrow}$CGTT...-3'
Primer R 3'-...TTGC$_{\uparrow}$AA...-5'
This information can be found on the page before you press "NEV Restriction Enzyme - Activity/Performance Chart".

4.4.6   Using ENDMEMO

For this exercise we need to use this web-side (http://www.endmemo.com/bio/egel.php) In the graph on the right place bands you would receive after digesting the amplicon with HindIII, ApeKI and AciI in one reaction.
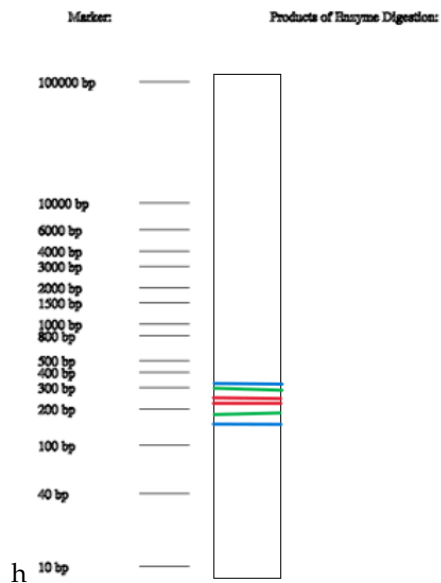
Figure 4.4: Gel electrophoresis of the DNA fragments and the PCR products' bands

Here we simply need to insert a sequence and choose the right restriction enzymes. The bands will then be shown on the gel.

- HindIII cuts at 167 bp; bands will be at 167- and 318 bp
    Blue color
- ApeKI cuts at 184 bp; bands will be at 184- and 301 bp
    Green color
- AciI cuts at 233 bp; bands will be at 233- and 252 bp
    Red color

Chapter 5

Literature résumés

5.1 1st Lecture

5.1.1 Article 1 - Fermented Foods as Experimentally Tractable Microbial Ecosystems

Introduction

The study of microbial communities faces significant challenges, largely due to their complexity and the difficulty in understanding how these multi-species communities are assembled, organized, and function. As a result, many aspects of microbial ecology remain unexplored.

The article highlights that the vast diversity and complexity of microbial communities are key factors contributing to our limited understanding. Additionally, the difficulty in isolating the majority of species from natural environments and cultivating them in vitro presents another notable challenge [5].

One approach to addressing these challenges is the development of simplified and controlled environments to study microbial communities. These controlled systems allow for reproducible results through measurable substrates, microbial growth, and community formation [5].

Several models with these properties have already been developed. These range from synthetic mixtures of microbial strains to more complex models composed of culture-able strains from free-living or host-associated communities, as well as naturally occurring communities that have been extensively sampled or perturbed in situ [5].

Fermented Foods as Experimentally Tractable Ecosystems

Fermented foods have been around for thousands of years and are an excellent example of how humans have optimized conditions to promote various properties of foods. Here are some of the main reasons why fermented foods have been produced for so long [5]:

- Good preserving effects
- Flavour can be improved
- Aroma can be enhanced
- Texture can be changed

Some factors for manipulating microbial growth are listed below [5]:

- Temperature
- Salinity
- Moisture

The article describes some micorbial communities of fermented foods (MCoFFs) and some examples of these, which will be listed bellow [5]:

- Multi-species biofilms associated with surfaces
    e.g. cheese rinds

- Suspended biofilms in liquid
    e.g. kombucha, kefir, and vinegar

- Dispersed growth in liquid
    e.g. lambic beers, natural wines, and yogurt

- Semi-solid substrates
    e.g. kimchi and miso

Many MCoFFs have been studied in detail, and the article highlights that some of these communities have been shown to have bacterial- and fungi species that have co-evolved over time. The microbial diversity within and across fermented foods is vast, and the article emphasizes that these communities are excellent models for studying microbial ecology. In the article, cheese rind (the MCoFFs on the surface of the cheese) was used as an example of the characterization and development of MCoFFs, since on the surface of cheese the microbial communities form as it ages [5].

Using Fermented Foods to Link Patterns, Processes, and Mechanisms of Microbial Community Assembly

In order to understand how species come together to form a community within a given environment microbiologists have developed several approaches for both plant- and animal communities. The ones named in the article is listed below [5]:

- **Dispersal**
    Contributions from the amount and timing of microbial propagules colonizing a habitat
- **Biotic selection**
    Interactions between species
- **Abiotic selection**
    Interactions between a species and the environment
- **Drift**
    Stochastic (random) changes in the relative abundances of species within communities
- **Diversification**
    Evolution of new species within communities

MCoFFs are ideal for linking microbial patterns with ecological processes. For instance, many fermented

foods exhibit clear microbial successions where early colonizers are replaced by other species. This process can be studied to understand how environmental changes and species interactions drive these transitions. New techniques like RNA sequencing and transposon mutagenesis have proven valuable in uncovering the molecular mechanisms behind these interactions. Additionally, tools like imaging mass spectrometry are enhancing our understanding of how microbial species communicate chemically [5].

Taste of Place? Microbial Biogeography of Fermented Foods

Though there are many different types of MCoFFs, many share some similarities. If a similar MCoFFs is used in two different cheese factories in two different locations, the microbial communities will be different, and with some distinct properties which can affect the texture, taste or aroma composition of the final product. This also enables the study of microbial biogeography, which is the study of how microbial communities are distributed across different locations. MCoFFs can therefore help linking patterns in microbial diversity from place to place. This is possible because each community has clear species abundance distributions [5].

Microbial Evolution in a Community Context

MCoFFs offer valuable systems for studying both ecological processes and microbial evolution. Microorganisms used in beer, wine, dairy, and sake production have been useful in understanding how microbes adapt to fermentation environments. Research suggests that these organisms often lose genes not needed in nutrient-rich settings, gain new traits through gene transfer, and adjust their metabolism to fit the fermentation niche [5].

Looking ahead, linking ecological and evolutionary processes could provide further insights. For example, studying long-term microbial communities could help explore how species interactions influence evolution. Experimentally testing coevolution within these communities could offer useful data on microbial adaptation over time [5].

Translation to Other Microbial Communities and Potential Applications

The study of microbial community assembly in traditional fermented foods (MCoFFs) can significantly enhance food quality and safety. Insights from these ecosystems can be applied to other microbial communities, such as the human microbiome. For example, the competitive mechanisms of *Lactobacillus reuteri*, found in both sourdough and human gut microbiomes, illustrate how patterns observed in MCoFFs can be relevant to other systems [5].

Research on cheese rind microbial diversity has revealed similarities with the human skin microbiome, suggesting that moisture is a major driver of surface biofilms in both environments. This indicates that ecological selection processes may be conserved across different microbial communities [5].

MCoFFs can directly impact human health as they are consumed and can interact with the human gut microbiome. While the probiotic effects of simplified MCoFFs like yogurt have been extensively studied, the broader microbial diversity in MCoFFs could introduce additional beneficial microbes and metabolites to the human digestive system. Studies have shown that MCoFFs can remain viable during passage through the human digestive tract [5].

Furthermore, MCoFFs offer valuable opportunities for designing synthetic microbial communities for applications in medicine, industry, and agriculture. Understanding the interactions within these communities can inform the construction of synthetic microbial ecosystems. By studying pre-existing MCoFFs, researchers can

learn about design principles and explore the potential for combining microbes from different MCoFFs into new, functional compositions [5].

Ultimately, research on food microbial communities could result in safer, more delicious foods and contribute to the development of essential principles for microbial community design [5].

5.1.2 Article 2 - Diversity of Microorganisms in Global Fermented Foods and Beverages

Introduction

The article describes how rice with fermented and non-fermented legumes is a staple diet in many counties in Asia. Wheat/barley-based breads/loaves followed bt milk and fermented milk products, meat and fermented meats are more common in the western world (West Asia, Europe and North America). The staple diet in Africa and South America comprise of sorghum and maize with wild legume seeds, meat and milk products [4]. Consortia of microorganisms are found naturally in uncooked plant and animal materials, utensils, food containers, earthen pots and in general in the environment. When producing fermented foods, it is normal to introduce a starter culture, which contains functional microorganisms [4]. The microorganisms in fermented food has converted the chemical composition of the raw food, which results in a change in the nutritional value of the food, making it more enriched [4].

Microorganisms in fermented foods

Lactic acid bacteria (LAB) are one of the most important groups of microorganisms in fermented foods. They are used in the production of fermented foods and beverages. Some of the major genra of the LAB are listed below [4]:

- *Alkalibacterium*
- *Carnobacterium*
- *Enterococcus*
- *Lactobacillus*
- *Lactococcus*
- *Leuconostoc*
- *Oenococcus*
- *Pediococcus*
- *Streptococcus*
- *Tetragenococcus*
- *Vagococcus*
- *Weissella*

Bacillus has ben found in alkaline-fermented foods in Asia anf Africa. The species of *Bacillus* that are found in abundance in legume-based fermented foods are listed bellow [4]:

- *Bacillus amyloliquefaciens*
- *Bacillus circulans*
- *Bacillus coagulans*
- *Bacillus firmus*
- *Bacillus licheniformis*
- *Bacillus megaterium*
- *Bacillus pumilus*
- *Bacillus subtilis*
- *Bacillus subtilis variety natto*
- *Bacillus thuringiensis*

There has been reported several species of *Kocuria, Micrococcus* and *Staphylococcus* in fermented foods. The species of *Kocuria* are found in abundance in fermented milk products, fermented sausages, meat-, and fish products [4].

Yeasts are also associated with fermentation of foods and alcoholic beverages. The yeasts named in the article are listed bellow [4]:

- *Brettanomyces*
- *Candida*
- *Cryptococcus*
- *Debaryomyces*
- *Dekkera*
- *Galactomyces*
- *Geotrichum*
- *Hansenula*
- *Hanseniaspora*
- *Hyphopichia*
- *Issatchenkia*
- *Metschnikowia*
- *Saccharomyces*
- *Pichia*
- *Kazachstania*
- *Rhodotorula*
- *Saccharomycodes*
- *Kluyveromyces*
- *Rhodosporidium*
- *Saccharomycopsis*
- *Schizosaccharomyces*
- *Sporobolomyces*
- *Torulaspora*
- *Torulopsis*
- *Trichosporon*
- *Yarrowia*
- *Zygosaccharomyces*

Filamentous molds are also found in fermented foods. They hold a major role in various fermented products in respect of enzyme production and in the degradation of anti-nutritive factors. The listed filamentous molds are listed bellow [4].

- *Actinomucor*
- *Amylomyces*
- *Aspergillus*
- *Monascus*
- *Mucor*
- *Neurospora*
- *Paracilomyces*
- *Penicillium*
- *Rhizopus*
- *Ustilago*

Taxonomic tools for identification of microorganisms from fermented foods

A commonly used method for profiling both culturable and non-culturable microbial populations from fermented foods is DNA extraction. The technique is based on the separation of PCR-amplified DNA fragments by denaturing gradient gel electrophoresis (DGGE). This method is both used for profiling bacterial populations and yeast populations in fermented foods. Also, in fermented foods a combination of Propidium MonoAzide (PMA) treatment, which is done before DNA-extraction, and molecular quantifying methods can be used to enumerate the viable microorganisms accurately [4]. Random amplification of polymorphic DNA (RAPD) is a typing method based on the genomic DNA fragment profiles amplified by rep-PCR, and is commonly used for differentiation of LAB strains from fermented foods and other bacteria or yeast in the consortia. This technique is able to differentiate between strains of the same species [4]. Amplified fragment length polymorphism (AFLP) is another technique which is used to identify and discriminate LAB strains. The technique is based on the selective amplification and separation of genomic restriction fragments [4].

DGGE techniques and temperature gradient gel electrophoresis (TGGE) are used to study the microbial diversity in fermented foods. The techniques are based on the separation of PCR-amplified 16s rDNA and 26s rDNA gene fragments by denaturing gradient gel electrophoresis. The techniques are used to study the microbial diversity in fermented foods. DGGE is quite time consuming and is not able to determine relative abundance of the dominant species, nor is it able to distinguish between viable and non-viable cells [4].

Next generation sequencing (NGS) are effective tools for studying the microbial diversity in fermented foods. The techniques are able to sequence all the genetic material present in a sample, providing a comprehensive view of the microbial community's diversity and functional potential. These techniques require extensive training and a well-equipped molecular laboratory [4].

In table 5.1 some of the NGS which is listed in the text are described. These NGS are used to study the microbial diversity in fermented foods [4].

Global fermented foods

A study reported approximately 3500 fermented food- and beverages, which were divided into 250 categories. Globally, fermented foods arae classified into 9 groups based of the substrates used from the plant- or animal

Table 5.1: Some NGS techniques with their respective key-points

Metagenomics	Phylobiomics	Metatranscriptomics
• Sequences all the genetic material present in a sample, providing a comprehensive view of the microbial community's diversity and functional potential. • Allows the study of microbial communities without the need to culture individual species, overcoming limitations of traditional microbiology methods. • By sequencing entire genomes, metagenomics can identify functional genes involved in processes like metabolism, antibiotic resistance, or nutrient cycling.	• Is primarily concerned with analyzing the evolutionary relationships between microbial species, often using 16S rRNA or other conserved genes. • It helps characterize the diversity and phylogenetic structure of microbial communities, determining how different species are related within an ecosystem. • This technique is useful for identifying species and classifying them taxonomically based on their evolutionary lineage.	• Focuses on sequencing the RNA transcripts present in a microbial community, providing a snapshot of active genes and gene expression levels. • Reveals what genes are actively being expressed in a community at a given time. • It helps assess how microbial communities respond to environmental changes, stress, or disease by analyzing shifts in gene expression.

source. These groups are listed in table 5.2 [4].

Table 5.2: The 9 groups of fermented foods and beverages

Number	Group
1	Fermented cereals
2	Fermented vegetables and bamboo shoots
3	Fermented legumes
4	Fermented roots/tubers
5	Fermented milk products
6	Fermented and preserved meat products
7	Fermented, dried and smoked fish products
8	Miscellaneous fermented products
9	Alcoholic beverages

Fermented milk products

Fermented milk products can be classified into two main groups based on the microorganisms involved; Lactic fermentation and Fungal-lactic fermentations. The lactic fermentation is dominated by Lactic Acid Bacteria (LAB), including thermophilic- (e.g., yogurt, Bulgarian buttermilk), probiotic- (e.g., acidophilus milk, bifidus milk), and mesophilic types (e.g., natural fermented milk, cultured milk, cultured cream, cultured buttermilk). The Fungal-lactic fermentations involves cooperation between LAB and yeasts to produce products like alcoholic milks and moldy milks [4].

Natural fermentation is an ancient method of milk processing that uses raw or boiled milk to ferment spontaneously. Starter cultures are used to initiate fermentation and can be either primary or secondary. The primary cultures mostly consist of *Lactococcus lactis*, *Lactobacillus delbrueckii*, and *Streptococcus thermophilus*, which participate in acidification. The secondary cultures are used in cheese-making and include *Brevibacterium linens*, *Propionibacterium freudenreichii*, and *Penicillium camemberti*, which develop flavor and texture during ripening [4].

Non-Starter Lactic Acid Bacteria (NSLAB) are present in high numbers in fermented milk and include *Enterococcus durans*, *Lactobacillus casei*, and *Staphylococcus* spp [4].

Fermented cereal foods

Fermented cereal foods are prepared differently in various regions of the world. In Asia, rice is fermented to produce alcoholic beverages or food beverages using mixed cultures. In Europe, America, and Australia, cereals like wheat, rye, barley, and maize are fermented using natural fermentation or commercial baker's yeast. In Africa, fermented cereal foods are used as staples, complementary, and weaning foods for infants and young children [4]. In Europe, traditional methods of bread preparation are still practiced without using commercial baker's yeast. Yeasts and LAB conduct dough fermentation, resulting in sourdough breads with higher lactic acid and acetic acid content [4]. Some LAB as *Enterococcus*, *Lactococcus*, *Lactobacillus*, *Leuconostoc*, *Pediococcus*, *Streptococcus*, and *Weissella* are deeply involved in cereal fermentation. Also some yeasts have a significant role such as the strains of *Saccharomyces cerevisiae* are the principal yeast, but other non-*Saccharomyces* yeasts like *Candida*, *Debaryomyces*, and *Hansenula* is also found in cereal fermentation [4].

Fermented vegetable foods

Many vegetables such as leafy greens, radish, cucumbers, and young edible bamboo shoots, are perishable and sesasonal, therefore they are traditionally fermented into edible products to ensure longer shelflife and safe consumption [4]. The main microorganisms involved in the fermentation of vegetables are LAB, which are responsible for the acidification of the product. The most common LAB in fermented vegetables are *Lactobacillus* and *Pediococcus*, followed by *Leuconostoc*, *Weissella*, *Tetragenococcus*, and *Lactococcus* [4]. Some examples of fermented vegetable products are listed below in table 5.3 [4]:

Table 5.3: Examples of fermented vegetable products

Food	Description
Kimchi	A Korean fermented vegetable product, with a characterized microbial profile of LAB
Sauerkraut	A German fermented cabbage product, with a reported species of LAB
Gundurk	have a native population of LAB. Made by fermenting leafy greans which then sun dried to create a sour and slightly bitter flavor

Fermented soybeans and other legumes

There are two types of fermented soybean foods, bacillus-fermented and mould-fermented. The bacillus-fermented foods are fermented by *Bacillus* spp. (mostly *B. subtilis*), and are characterized by stickiness. The mould-fermented foods are fermented by filamentous molds, mostly *Aspergillus*, *Mucor*, and *Rhizopus*.

These types of foods are concentrated in the "Kinema-Natto-Thua Nao (KNT)-triangle" region, which includes Japan, east Nepal, north-east India, and northern Thailand. Examples of bacillus-, mould-, and other fermented foods are listed in table 5.4 [4].

Table 5.4: Examples of Bacillus-, mould-, and other fermented foods

Bacillus-fermented		Mould-fermented		Fermented foods	
Foodstuff	**Origin**	**Foodstuff**	**Origin**	**Foodstuff**	**Origin**
Natto	Japan	Miso and shoyu	Japan	Locust beans	Africa
Kinema	India	Tempe	Indonesia	Black-grams	India
Chungkokjang	Korea	Douchi and sufu	China	Maseura	Nepal
Thua nao	Thailand	Doenjang	Korea		
Pekok	Myanmar	Pekok	Myanmar		
Sieng	Cambodia	Sieng	Cambodia		

Fermented meat products

Overall, fermented meat products are divided into two categories, whole meat pieces (e.g. jamón serrano) and chopped meat pieces (e.g. chorizo). The main microbial groups invovled in fermenting meats are LAB, coagulase-negative cocci, and yeasts. The LAB are responsible for the acidification of the meat, while the coagulase-negative cocci are responsible for the ripening of the meat. The yeasts are responsible for the development of the flavor and aroma of the meat [4].

Fermented fish products

Traditionally fish is preserved by either fermentation, sun/smoke drying and salting. Both bacteria and yeasts have been reported from these preservation methods.

Fermented miscellaneous products

The article discusses various fermented foods and the microorganisms involved in their production. These foods include [4]:

- Vinegar
    Produced from sugar or ethanol-containing substrates and hydrolyzed starchy materials by aerobic conversion to acetic acid.
    Dominant microorganisms include *Acetobacter* species and yeast species such as *Candida* and *Saccharomyces*.

- Fermented Teas
    Such as miang, puer tea, fuzhuan brick, and kombucha.
    Microorganisms involved include *Aspergillus*, *Penicillium*, and *Saccharomyces* species.

- Nata
    A delicacy from the Philippines, produced by *Acetobacter xylinum*.
    Two types of nata are known: nata de piña and nata de coco.

- Chocolate
    Produced from cocoa bean fermentation.
    The predominant microorganisms are *Acetobacter pasteurianus*, *Lactobacillus plantarum*, and *Saccharomyces cerevisiae*.

Continues on the next page.

> • Pidan
>   A preserved egg prepared from alkali-treated fresh duck eggs, with a strong hydrogen sulfide and ammonia smell.
>   Dominant microorganisms include *Bacillus cereus*, *Staphylococcus cohnii*, and *Staphylococcus epidermidis*.

Some common microorganisms found in miscellaneous fermented foods include [4]:

> • LAB; such as *Lb. fermentum*, *Lb. plantarum*, and *Weissella* species.
> • Acetobacter; such as *Acetobacter pasteurianus* and *Acetobacter xylinum*.
> • Yeast; such as *Saccharomyces*, *Candida*, and *Hanseniaspora* species.
> • Mold; such as *Aspergillus* and *Penicillium* species.

Amolytic starters

Amylolytic starters are traditional cultures of microorganisms used to produce alcoholic beverages and fermented foods from starchy materials. These starters are composed of consortia of filamentous molds, amylolytic, and alcohol-producing yeasts and LAB, and are typically made with rice or wheat as the base [4]. The article described some of the different types of amylolytic starters which are used in various countries, including [4]:

> • **Marcha** (India and Nepal)
> • **Hamei, humao, phab** (India)
> • **Men** (Vietnam)
> • **Ragi** (Indonesia)
> • **Bubod** (Philippines)
> • **Chiu/chu** (China and Taiwan)
> • **Loogpang** (Thailand)
> • **Mae/dombae/buh/puh** (Cambodia)
> • **Nuruk** (Korea)

Furthermore the article named some of the common microorganisms involved in amylolytic starters which included [4]:

> • Filamentous molds; such as *Mucor circinelloides*, *Rhizopus chinensis*, and *Aspergillus oryzae*.
> • Yeasts; such as *Saccharomyces cerevisiae*, *Pichia anomala*, and *Candida glabrata*.
> • LAB; such as *Pediococcus pentosaceus*, *Lactobacillus bifermentans*, and *Lactobacillus brevis*.
> • Amylase-producing bacilli; such as *Bacillus subtilis* and *Bacillus amyloliquefaciens*.
> • Acetic acid bacteria; such as *Acetobacter orientalis* and *Acetobacter pasteurianus*.

Alcoholic beverages

Alcoholic beverages can be classified into 10 types based on their production methods and ingredients. These types include [4]:

- **Non-distilled and unfiltered beverages produced by amylolytic starters**:, such as kodo ko jaanr (fermented finger millets) and bhaati jaanr (fermented rice) of India and Nepal.
- **Non-distilled and filtered beverages produced by amylolytic starters**:, such as saké of Japan.
- **Distilled beverages produced by amylolytic starters**:, such as shochu of Japan and soju of Korea.
- **Beverages produced by human saliva**: Such as chicha of Peru, which involves the use of amylase in human saliva.
- **Mono-fermentation beverages**: Produced by a single-strain fermentation, such as beer.
- **Honey-based beverages**: Such as tej of Ethiopia.
- **Plant-based beverages**: Such as pulque of Mexico, toddy of India, and kanji of India.
- **Malted beverages**: Produced by malting (germination), such as sorghum ("Bantu") beer of South Africa, pito of Nigeria and Ghana, and tchoukoutou of Benin.
- **Fruit-based beverages without distillation**: Such as wine and cider.
- **Distilled fruit and cereal beverages**: Such as whisky and brandy.

5.1.3   Article 3 - Molecular biology of the cell; chapter 6

In this chapter

The following topics will be further described in this chapter [1]:

The central dogma of molecular biology; The discovery of the structure of DNA in the 1950s has led to significant progress in cell and molecular biology. We now know the complete genome sequences for thousands of organisms, including humans and our extinct relatives, such as the Neanderthals. The flow of genetic information in cells is from DNA to RNA to protein, a principle known as the central dogma of molecular biology [1].

Variations in the central dogma; While the central dogma is universal, there are important variations between organisms in the way information flows from DNA to protein. In eukaryotic cells, RNA transcripts are subject to processing steps in the nucleus, including RNA splicing, before they are translated into protein [1].

The role of RNA; RNA plays a crucial role in the cell, not only as a template for protein synthesis but also as a structural and catalytic molecule. Some RNAs fold into precise three-dimensional structures, while others act as regulators of gene expression [1].

The complexity of genomes; The genomes of most multicellular organisms are surprisingly disorderly, reflecting their chaotic evolutionary histories. Genes are often arranged in a long string of alternating short exons and long introns, with small bits of DNA sequence that code for protein interspersed with large blocks of seemingly meaningless DNA [1].

Decoding genomes; Decoding genomes is a complex task, even with the aid of powerful computers. Cells in our body can automatically locate the beginning and end of genes and decipher when and where each gene is expressed, but researchers still face significant challenges in understanding how genomes are decoded and used [1].

The challenges of cell biology; Despite significant progress in cell and molecular biology, there is still much to be discovered about how the information stored in an organism's genome produces even the simplest unicellular bacterium, let alone how it directs the development of a human with approximately 30,000 genes.

The next generation of cell biologists will face many fascinating challenges in understanding the complexities of genomes and how they are decoded and used [1].

From DNA to RNA

Transcription and translation are the means by which cells read out, or express, the genetic instructions in their genes [1].

Regulation of gene expression is when the cells can synthesize a large amount of protein from a gene when necessary, and genes can be transcribed and translated with different efficiencies, allowing the cell to make vast quantities of some proteins and tiny amounts of others [1].

RNA molecules are single-stranded

Transcription is the first step a cell takes in reading out a needed part of its genetic instructions is to copy a particular portion of its DNA nucleotide sequence—a gene—into an RNA nucleotide sequence [1].

RNA structure is a linear polymer made of four different types of nucleotide subunits linked together by phosphodiester bonds. It differs from DNA chemically in two respects: it contains the sugar ribose and the base uracil (U) instead of thymine (T) [1].

RNA folding is single-stranded and can fold up into a particular shape, allowing some RNA molecules to have precise structural and catalytic functions [1].

Transcription produces RNA complementary to one strand of DNA

Transcription process begins with the opening and unwinding of a small portion of the DNA double helix to expose the bases on each DNA strand. One of the two strands of the DNA double helix then acts as a template for the synthesis of an RNA molecule [1].

RNA synthesis happens when nucleotide sequence of the RNA chain is determined by the complementary base-pairing between incoming nucleotides and the DNA template. The RNA chain is elongated one nucleotide at a time, and it has a nucleotide sequence that is exactly complementary to the strand of DNA used as the template [1].

Differs from DNA replication in several crucial ways. Unlike a newly formed DNA strand, the RNA strand does not remain hydrogen-bonded to the DNA template strand. Instead, the RNA chain is displaced and the DNA helix re-forms, releasing the RNA molecules as single strands [1].

RNA polymerase carry out transcription

RNA polymerases are enzymes that perform transcription. They catalyze the formation of the phosphodiester bonds that link the nucleotides together to form a linear chain [1].

Mechanism of RNA polymerase involves the RNA polymerase moving stepwise along the DNA, unwinding the DNA helix just ahead of the active site for polymerization to expose a new region of the template strand for complementary base-pairing, and extending the growing RNA chain by one nucleotide at a time in the 5′-to-3′ direction [1].

Characteristics of RNA polymerase include the ability to start an RNA chain without a primer, being absolutely processive, and having a modest proofreading mechanism to correct errors in the growing RNA chain [1].

Comparison with DNA polymerase reveals that RNA polymerase is less accurate, with an error rate of about one mistake for every 10^4 nucleotides copied into RNA, despite catalyzing essentially the same chemical reaction as DNA polymerase [1].

RNA polymerase binds to DNA at specific sites

Types of RNA include messenger RNA (mRNA), which directs the synthesis of proteins, and noncoding RNAs, which do not code for protein and serve as enzymatic, structural, and regulatory components [1].
Functions of noncoding RNAs include directing the splicing of pre-mRNA to form mRNA, forming the core of ribosomes, and serving as adaptors that select amino acids and hold them in place on a ribosome for incorporation into protein[1].
Transcription units are segments of DNA that are transcribed into RNA. In eukaryotes, a transcription unit typically carries the information of just one gene, while in bacteria, a set of adjacent genes is often transcribed as a unit [1].
RNA composition in cells reveals that RNA makes up a few percent of a cell's dry weight, with most of the RNA being rRNA, and mRNA comprising only 3–5% of the total RNA in a typical mammalian cell [1].

Signals encoded in DNA tell RNA poluymerase where to start and stop

Initiation of transcription is a crucial step in gene expression, where RNA polymerase recognizes where to start and stop transcribing a gene. The bacterial RNA polymerase core enzyme is a multisubunit complex that synthesizes RNA using the DNA template as a guide, with the help of the sigma (σ) factor [1].
During **transcription bubble formation**, the RNA polymerase holoenzyme opens up the double helix to expose a short stretch of nucleotides on each strand. The first ten or so nucleotides of RNA are synthesized using a "scrunching" mechanism, and the polymerase then moves down the DNA, synthesizing RNA in a stepwise fashion [1].
Chain elongation continues until the enzyme encounters the *Termination* signal, which consists of a string of A-T nucleotide pairs preceded by a twofold symmetric DNA sequence. The formation of a hairpin structure through Watson–Crick base-pairing helps to disengage the RNA transcript from the active site, marking the end of transcription [1].

Transcription Start and Stop Signals Are Heterogeneous in Nucleotide Sequence

Transcription initiation and termination involve a series of structural transitions in protein, DNA, and RNA molecules, and the signals encoded in DNA that specify these transitions are often difficult to recognize. Despite this, researchers have identified common features among bacterial promoters, which are often summarized in the form of a **consensus sequence**. A consensus sequence is derived by comparing many sequences with the same basic function and tallying up the most common nucleotides found at each position [1].
Promoter Strength and Sequence Variation The DNA sequences of individual bacterial promoters differ in ways that determine their strength, with promoters for genes that code for abundant proteins being much stronger than those associated with genes that encode rare proteins. The nucleotide sequences of their promoters are responsible for these differences [1].
Terminator Sequences Like bacterial promoters, transcription terminators also have a wide range of sequences, with the potential to form a simple hairpin RNA structure being the most important common feature. Terminator sequences are even more heterogeneous than promoter sequences [1].

Challenges in Locating Promoters and Terminators Despite our knowledge of bacterial promoters and terminators, their variation in nucleotide sequence makes it difficult to definitively locate them simply by analysis of the nucleotide sequence of a genome. Additional information, some of it from direct experimentation, is often needed to locate and accurately interpret the short DNA signals in genomes [1].

Promoter Orientation and Template Strand Promoter sequences are asymmetric, ensuring that RNA polymerase can bind in only one orientation. The promoter orientation specifies the strand to be used as a template, and genome sequences reveal that the DNA strand used as the template for RNA synthesis varies from gene to gene, depending on the orientation of the promoter [1].

Transcription Initiation in Eukaryotes Requires Many Proteins Eukaryotic nuclei have three types of RNA polymerase, with RNA polymerase II transcribing most genes, including those that encode proteins [1].

Key Differences from Bacterial RNA Polymerase

Eukaryotic RNA polymerase II differs from bacterial RNA polymerase in two main ways [1]:

- It requires many transcription-initiation factors, collectively called the general transcription factors.
- It must initiate transcription on DNA packaged into *nucleosomes* and higher-order chromatin structures.

RNA Polymerase II Requires a Set of General Transcription Factors

Eukaryotic RNA polymerase II requires a set of general transcription factors, denoted as TFIIA, TFIIB, TFIIC, TFIID, and so on, to position correctly at the promoter and initiate transcription [1].

Assembly of General Transcription Factors The assembly process begins when TFIID binds to the TATA sequence, a short double-helical DNA sequence primarily composed of T and A nucleotides. The binding of TFIID causes a large distortion in the DNA, which serves as a physical landmark for the location of an active promoter [1].

Role of TFIIH TFIIH, the most complicated of the general transcription factors, performs several enzymatic steps needed for the initiation of transcription, including unwinding the DNA and exposing the template strand [1].

Phosphorylation of RNA Polymerase II The phosphorylation of the C-terminal domain (CTD) of RNA polymerase II, specifically the serine located at the fifth position in the repeat sequence (Ser5), is a key step in the transition from initiation to elongation. This phosphorylation allows the polymerase to disengage from the cluster of general transcription factors and undergo conformational changes that tighten its interaction with DNA [1].

Release of General Transcription Factors Once the polymerase II has begun elongating the RNA transcript, most of the general transcription factors are released from the DNA, making them available to initiate another round of transcription with a new RNA polymerase molecule [1].

Polymerase II Also Requires Activator, Mediator, and Chromatin-Modifying Proteins

In addition to general transcription factors, RNA polymerase II requires several other proteins to initiate transcription in a eukaryotic cell, including [1]:

- **Transcriptional activators**, which bind to specific sequences in DNA (called enhancers) and help to attract RNA polymerase II to the start point of transcription.
- **Mediator**, a large protein complex that allows activator proteins to communicate properly with the polymerase II and with the general transcription factors.

Continues on the next page.

- **Chromatin-modifying enzymes**, including chromatin remodeling complexes and histone-modifying enzymes, which increase access to the DNA in chromatin and facilitate the assembly of the transcription initiation machinery onto DNA.

Assembly of Transcription Initiation Machinery The assembly of the transcription initiation machinery onto DNA involves the recruitment of many proteins (well over 100 individual subunits) and does not follow a prescribed pathway. The order of assembly differs from gene to gene, and some protein complexes may be brought to DNA as preformed subassemblies [1].

Release of RNA Polymerase II To begin transcribing, RNA polymerase II must be released from the large complex of proteins. This release often requires the in situ proteolysis of the activator protein [1].

Transcription Elongation in Eukaryotes Requires Accessory Proteins

Once RNA polymerase has initiated transcription, it requires accessory proteins to facilitate elongation. These proteins include [1]:

- **Elongation factors**, which decrease the likelihood that RNA polymerase will dissociate before it reaches the end of a gene.
- **ATP-dependent chromatin remodeling complexes**, which help RNA polymerase move through chromatin structure by either moving with the polymerase or rescuing stalled polymerases.
- **Histone chaperones**, which partially disassemble nucleosomes in front of a moving RNA polymerase and assemble them behind.

Modification of Histones During Transcription Elongation As RNA polymerase moves along a gene, enzymes bound to it modify the histones, leaving behind a record of where the polymerase has been. This information may aid in transcribing a gene over and over again once it has become active for the first time, and may also be used to coordinate transcription elongation with RNA processing [1].

Transcription Creates Superhelical Tension

DNA Supercoiling is a conformation that DNA adopts in response to superhelical tension, which can be created by opening or unwinding a double-helical DNA molecule [1].

Effect of RNA Polymerase on DNA Supercoiling is to create superhelical tension as it moves along a stretch of DNA that is anchored at its ends, generating positive superhelical tension in the DNA in front of it and negative helical tension behind it [1].

Consequences of DNA Supercoiling include making the DNA helix more difficult to open, but also facilitating the partial unwrapping of the DNA in nucleosomes in eukaryotes [1].

Removal of Superhelical Tension is achieved by DNA topoisomerase enzymes in eukaryotes, while in bacteria, DNA gyrase uses the energy of ATP hydrolysis to pump supercoils continuously into the DNA, maintaining the DNA under constant tension [1].

Transcription Elongation in Eukaryotes Is Tighly Coupled to RNA Processing

Eukaryotic mRNA Synthesis involves several steps beyond transcription, including covalent modification of the ends of the RNA and removal of intron sequences by RNA splicing [1].

Modification of Eukaryotic mRNAs includes capping on the 5′ end and polyadenylation of the 3′ end, allowing the cell to assess the integrity of the mRNA molecule before exporting it from the nucleus and translating it into protein [1].

RNA Splicing joins together the different portions of a protein-coding sequence, enabling eukaryotes to synthesize several different proteins from the same gene [1].

Coupling of RNA Processing to Transcription Elongation is achieved through the phosphorylation of the RNA polymerase II tail, which allows a new set of proteins to associate with the RNA polymerase and function in transcription elongation and RNA processing [1].

RNA Capping Is the First Modification of Eukaryotic Pre-mRNAs

5′ Capping of Eukaryotic mRNAs involves the addition of a modified guanine nucleotide to the 5′ end of the new RNA molecule, which is performed by three enzymes acting in succession [1].

Mechanism of 5′ Capping includes the removal of a phosphate from the 5′ end of the nascent RNA, the addition of a GMP in a reverse linkage, and the addition of a methyl group to the guanosine [1].

Function of the 5′-Methyl Cap is to signify the 5″ end of eukaryotic mRNAs, helping the cell to distinguish mRNAs from other types of RNA molecules, and to facilitate further processing and export of the mRNA [1].

Binding of the 5′-Methyl Cap to a protein complex called CBC (cap-binding complex) in the nucleus, which helps the mRNA to be further processed and exported [1].

RNA Splicing Removes Intron Sequences from Newly Transcribed Pre-mRNAs

Eukaryotic Gene Structure is characterized by the presence of noncoding intervening sequences (introns) that interrupt the protein-coding sequences (exons) [1].

RNA Splicing is the process by which intron sequences are removed from the newly synthesized RNA, and the exons are joined together through two sequential phosphoryl-transfer reactions known as transesterifications [1].

Mechanism of RNA Splicing involves a complex machinery consisting of five additional RNA molecules and several hundred proteins, which hydrolyzes many ATP molecules per splicing event to ensure accurate and flexible splicing [1].

Advantages of RNA Splicing include facilitating the emergence of new and useful proteins over evolutionary time scales by allowing genetic recombination to combine the exons of different genes, and enabling eukaryotes to increase the coding potential of their genomes by producing a corresponding set of different proteins from the same gene through alternative splicing [1].

Nucleotide Sequences Signal Where Splicing Occurs

Recognition of Splicing Sites requires the splicing machinery to identify three portions of the precursor RNA molecule: the 5′ splice site, the 3′ splice site, and the branch point in the intron sequence [1].
Consensus Sequences for Splicing are short nucleotide sequences that are similar from intron to intron and provide cues for where splicing is to take place, but can accommodate extensive sequence variability [1].
Challenges in Deciphering Genome Sequences arise from the high variability of splicing consensus sequences and the possibility of alternative splicing, making it difficult to predict protein sequences solely from a genome sequence and identify all of the genes in a complete genome sequence [1].

RNA Splicing Is Performed by the Spliceosome

Role of RNA Molecules in RNA Splicing involves specialized RNA molecules, known as snRNAs (small nuclear RNAs), that recognize the nucleotide sequences that specify where splicing is to occur and catalyze the chemistry of splicing [1].
Composition of the Spliceosome includes five snRNAs (U1, U2, U4, U5, and U6) complexed with at least seven protein subunits to form an snRNP (small nuclear ribonucleoprotein), which forms the core of the spliceosome [1].
Mechanism of Spliceosome Assembly involves the recognition of the 5′ splice junction, the branch-point site, and the 3′ splice junction through base-pairing between the snRNAs and the consensus RNA sequences in the pre-mRNA substrate, and the dynamic assembly and rearrangement of the spliceosome components during the splicing reaction [1].
Spliceosome Structure and Function is complex and dynamic, with some scientists believing that the spliceosome is a preexisting, loose assembly of all the components that captures, splices, and releases RNA as a coordinated unit, undergoing extensive rearrangements each time a splice is made [1].

The Spliceosome Uses ATP Hydrolysis to Produce a Complex Series of RNA-RNA Rearrangements

Energy Requirements for RNA Splicing reveals that ATP hydrolysis is not required for the chemistry of RNA splicing itself, but is necessary for the assembly and rearrangements of the spliceosome [1].
Role of ATP Hydrolysis in Spliceosome Assembly involves the use of energy from ATP hydrolysis to break existing RNA-RNA interactions and allow the formation of new ones, enabling the splicing signals on the pre-RNA to be examined by snRNPs several times during the course of splicing [1].

Purpose of Spliceosome Rearrangements is to increase the overall accuracy of splicing by allowing the spliceosomes to check and recheck the splicing signals, and to create the active sites for the two transesterification reactions [1].

Catalytic Sites of the Spliceosome are formed by both protein and RNA molecules, with the RNA molecules catalyzing the actual chemistry of splicing [1].

Disassembly of the Spliceosome requires another series of RNA-RNA rearrangements that require ATP hydrolysis, returning the snRNAs to their original configuration so that they can be used again in a new reaction [1].

Exon Junction Complex (EJC) is a set of proteins that bind to the mRNA near the position formerly occupied by the intron, marking the site of a successful splicing event and influencing the subsequent fate of the mRNA [1].

RNA Can Be Processed by Dicer to Produce siRNAs

Energy Requirements for RNA Splicing reveals that ATP hydrolysis is not required for the chemistry of RNA splicing itself, but is necessary for the assembly and rearrangements of the spliceosome [1].

Role of ATP Hydrolysis in Spliceosome Assembly involves the use of energy from ATP hydrolysis to break existing RNA-RNA interactions and allow the formation of new ones, enabling the splicing signals on the pre-RNA to be examined by snRNPs several times during the course of splicing [1].

Purpose of Spliceosome Rearrangements is to increase the overall accuracy of splicing by allowing the spliceosomes to check and recheck the splicing signals, and to create the active sites for the two transesterification reactions [1].

Catalytic Sites of the Spliceosome are formed by both protein and RNA molecules, with the RNA molecules catalyzing the actual chemistry of splicing [1].

Disassembly of the Spliceosome requires another series of RNA-RNA rearrangements that require ATP hydrolysis, returning the snRNAs to their original configuration so that they can be used again in a new reaction [1].

Exon Junction Complex (EJC) is a set of proteins that bind to the mRNA near the position formerly occupied by the intron, marking the site of a successful splicing event and influencing the subsequent fate of the mRNA [1].

Other Properties of Pre-mRNA and Its Synthesis Help to Explain the Choice of Proper Splice Sites

Challenges in Splice-Site Selection arise from the enormous size variation of intron sequences, which could lead to frequent splicing mistakes if not addressed by fidelity mechanisms [1].

Strategies to Increase Splicing Accuracy include coupling splicing to transcription, which helps the cell keep track of introns and exons, and "exon definition", which involves marking off exon sequences with SR proteins to increase the accuracy of splicing component deposition [1].

Mechanism of Exon Definition involves the assembly of SR proteins on exon sequences, which recruit U1 and U2 snRNAs to mark the downstream and upstream exon boundaries, respectively, and increase the accuracy of splicing.

Role of SR Proteins is to bind preferentially to specific RNA sequences in exons, termed splicing enhancers, and help to discriminate exon sequences from intron sequences [1].

Timing of Splicing Events reveals that the marking of exon and intron boundaries and the assembly of the spliceosome begin on an RNA molecule while it is still being elongated by RNA polymerase, but the actual chemistry of splicing can take place later, allowing for the removal of intron sequences in a non-linear order [1].

Chromatin Structure Affects RNA Splicing

Influence of Chromatin Structure on RNA Splicing reveals that the way a gene is packaged into chromatin can affect how the RNA transcript of that gene is ultimately spliced [1].

Nucleosome Positioning and RNA Splicing suggests that nucleosomes tend to be positioned over exons, acting as "speed bumps" that allow proteins responsible for exon definition to assemble on the RNA as it emerges from the polymerase [1].

Effect of Chromatin Structure on Transcription Rate shows that changes in chromatin structure can alter splicing patterns by affecting the rate at which RNA polymerase moves along DNA [1].

Direct Influence of Chromatin Structure on RNA Splicing involves specific histone modifications attracting components of the spliceosome, which can be transferred to the emerging RNA, affecting the final pattern of splicing [1].

RNA Splicing Shows Remarkable Plasticity

Factors Influencing Splice Site Choice include the strength of the three signals on the RNA (the 5′ and 3′ splice junctions and the branch point), co-transcriptional assembly of the spliceosome, chromatin structure, and exon definition [1].

Flexibility of Splicing allows for the splicing machinery to pick out the best possible pattern of splice junctions, even if the optimal one is damaged by mutation, and can lead to changes in splicing patterns caused by random mutations [1].

Consequences of Mutations Affecting Splicing can be severely detrimental to the organism, and it is estimated that 10% of point mutations that cause inherited human diseases produce aberrant splicing of the gene containing the mutation [1].

Regulation of RNA Splicing allows the cell to easily regulate the pattern of RNA splicing, and alternative splicing can give rise to different proteins from the same gene, enhancing the coding potential of genomes [1].

Regulated RNA Splicing involves the cell regulating the splicing patterns so that different forms of the protein are produced at different times and in different tissues, and is a common strategy to enhance the coding potential of genomes [1].

Spliceosome-Catalyzed RNA Splicing Probably Evolved from Self-splicing Mechanisms

Evolution of the Spliceosome suggests that RNA molecules were used as major catalysts in early cells, and RNA-catalyzed splicing reactions played critical roles in these cells [1].

Evidence for RNA-Catalyzed Splicing includes the existence of self-splicing RNA introns, which can catalyze their own splicing in the absence of proteins or other RNA molecules [1].

Mechanism of Self-Splicing Introns involves the RNA molecule folding into a specific three-dimensional structure that brings the intron/exon junctions together and catalyzes the two transesterification reactions [1].

Evolution of Pre-mRNA Splicing is thought to have evolved from a simpler, ancestral form of RNA self-splicing, with the basic chemistry of some self-splicing reactions being similar to pre-mRNA splicing [1].

RNA-Processing Enzymes Generate the 3′ End of Eukaryotic mRNAs

Processing of the 3′ End of Eukaryotic mRNAs involves the recognition of signals encoded in the genome by RNA-binding proteins and RNA-processing enzymes [1].
Role of CstF and CPSF is to bind to their recognition sequences on the emerging RNA molecule and assemble with additional proteins to create the 3′ end of the mRNA [1].
Mechanism of 3′-End Processing involves the cleavage of the RNA from the polymerase, followed by the addition of a poly-A tail by poly-A polymerase, and the assembly of poly-A-binding proteins onto the tail [1].
Synthesis of the Poly-A Tail is a non-template-dependent process, where poly-A polymerase adds approximately 200 A nucleotides to the 3′ end of the mRNA, using ATP as the nucleotide precursor [1].
Termination of Transcription is thought to be caused by the continued degradation of the newly synthesized RNA by a $5′ \rightarrow 3′$ exonuclease carried along on the polymerase tail, which eventually causes the RNA polymerase to release its grip on the template[1].

Mature Eukaryotic mRNAs Are Selectively Exported from the Nucleus

Orderly RNA Processing in the Nucleus ensures that only mature mRNA is exported to the cytosol, while debris such as excised introns and aberrantly processed pre-mRNAs are retained and degraded in the nucleus by the nuclear exosome [1].
Protein Markers of mRNA Maturation include cap-binding complexes, exon junction complexes, and poly-A-binding proteins. These proteins mark the completion of capping, splicing, and polyadenylation, respectively [1].
Distinguishing Mature mRNAs from RNA debris is crucial, as incomplete processing is indicated by the presence of certain proteins (e.g., snRNPs), which would signify incomplete splicing. Properly processed mRNAs are transported to the cytosol [1].
mRNA Export Process involves binding of specific nuclear transport receptors to mRNA molecules. These receptors escort mRNA through the nuclear pore complex, ensuring that only processed mRNA exits the nucleus [1].
Role of hnRNP Proteins in RNA processing is to help unwind hairpin helices in RNA and to distinguish mature mRNA from RNA processing debris. They guide mRNAs through nuclear pore complexes for export [1].

Noncoding RNAs Are Also Synthesized and Processed in the Nucleus

Abundance of rRNA – Ribosomal RNA (rRNA) makes up 80% of total RNA in dividing cells, while only 3–5% is mRNA [1].
Specialized RNA Polymerases – Eukaryotes use RNA polymerase I to synthesize rRNAs, while RNA polymerase III makes 5S rRNA [1].
rRNA Modifications – rRNA undergoes specific chemical modifications, guided by small nucleolar RNAs (snoRNAs), which help with folding and processing [1].
Multiple rRNA Gene Copies – Human cells have around 200 copies of rRNA genes to meet the high demand for ribosomes [1].

The Nucleolus Is a Ribosome-Producing Factory

Nucleolus Function – The nucleolus processes rRNAs and assembles ribosome subunits, without being membrane-bound [1].

Composition – It consists of rRNA genes, precursor and mature rRNAs, processing enzymes, snoRNPs, and ribosomal proteins, allowing rapid ribosome assembly [1].

Nucleolus Formation – rRNA genes on 10 chromosomes contribute to nucleolus formation during interphase, and it reforms after mitosis [1].

Size and Activity – The nucleolus size reflects ribosome production levels and can occupy up to 25% of the nuclear volume in highly active cells [1].

Other Functions – It also assembles noncoding RNA-protein complexes like U6 snRNP, telomerase, and signal-recognition particles, and processes tRNAs [1].

The Nucleus Contains a Variety of Subnuclear Aggregates

Cajal Bodies and Interchromatin Granule Clusters – These nuclear structures lack membranes and are involved in RNA processing and assembly of macromolecules [1].

Cajal Body Function – Cajal bodies facilitate snRNP and snoRNP maturation, recycling, and resetting after RNA splicing [1].

Interchromatin Granule Clusters – These clusters store mature RNA processing components, such as snRNPs, ready for mRNA production [1].

Dynamic Nature – These structures change during the cell cycle, and their main role is to concentrate components, accelerating assembly processes like U4/U6 snRNP formation [1].

Splicing Factories – RNA splicing occurs co-transcriptionally, with transcription and splicing components creating "factories" for efficient processing near interchromatin granule clusters [1].

Summary - DNA to RNA

Before protein synthesis begins, mRNA must be produced by transcription. Bacteria use a single RNA polymerase, while eukaryotes use three: RNA polymerases I, II, and III. RNA polymerase II synthesizes mRNA with the help of general transcription factors and activators. It also relies on chromatin remodeling complexes to initiate transcription on chromatin templates [1].

During transcription, the nascent RNA undergoes processing: capping, splicing, and polyadenylation, co-ordinated by proteins bound to RNA polymerase II's C-terminal tail. Only properly processed mRNAs are exported to the cytosol for translation into proteins [1].

In eukaryotes, RNA polymerase I synthesizes rRNAs, which are processed and assembled into ribosomal subunits in the nucleolus. Other subnuclear structures, like Cajal bodies, aid in RNA processing, assembly, and recycling. These "factories" concentrate components, making the processes efficient [1].

From RNA to Protein

RNA as Final Product – Some genes produce RNA as their final product, such as the RNAs found in snRNPs and ribosomes [1].

mRNA as Intermediate – Most genes, however, produce mRNA molecules, which serve as intermediaries in protein production [1].

The Coding Problem – In the late 1950s, biologists sought to understand how the linear sequence of nucleotides in RNA is translated into the linear sequence of amino acids in proteins [1].

Cracking the Code – Over time, the genetic code was deciphered step by step, and in 2000, the ribosome's structure, responsible for decoding mRNA, was revealed in atomic detail [1].

An mRNA Sequence Is Decoded in Sets of Three Nucleotides

Translation Process – After transcription and processing, mRNA is used to synthesize proteins, translating nucleotide sequences into protein sequences [1].

Genetic Code – Deciphered in the early 1960s, the genetic code translates triplet codons (64 possible) into 20 amino acids, indicating redundancy where multiple codons can specify the same amino acid [1].

Universal Code – This genetic code is universal across organisms, with minor variations in mitochondrial DNA [1].

Reading Frames – An RNA sequence can be read in three frames, but only one frame encodes the correct protein, determined by a specific signal at the start of translation [1].

tRNA Molecules Match Amino Acids to Codons in mRNA

Adaptor Function – Codons in mRNA do not directly bind to amino acids. Instead, **tRNAs**, small RNA molecules of about 80 nucleotides, serve as adaptors that recognize codons and carry the corresponding amino acid. tRNAs fold into a **cloverleaf** structure, which further folds into an **L-shaped** configuration stabilized by hydrogen bonds [1].

Anticodon and Amino Acid Attachment – The anticodon pairs with the mRNA codon, while the **3′ end** of the tRNA binds the specific amino acid. The genetic code's redundancy means some amino acids have more than one tRNA, and **wobble base-pairing** allows certain tRNAs to pair with multiple codons by tolerating mismatches at the third codon position [1].

tRNAs Are Covalently Modified Before They Exit from the Nucleus

Role of tRNA – Transfer RNAs (tRNAs) are small RNA molecules (about 80 nucleotides) that act as adaptors, recognizing codons in mRNA and matching them with the corresponding amino acids [1].

Structure of tRNA – tRNA molecules fold into a cloverleaf structure, with key regions forming an anticodon that pairs with mRNA codons and a 3′ end where the amino acid is attached [1].

tRNA Variety – Humans possess nearly 500 tRNA genes, featuring 48 different anticodons, reflecting the redundancy of the genetic code, where multiple codons can specify a single amino acid [1].

Wobble Base-Pairing – Some tRNAs can base-pair with more than one codon due to "wobble" base-pairing, allowing for flexibility in the third nucleotide of codons, facilitating the matching of 20 amino acids to 61 codons [1].

Specific Enzymes Couple Each Amino Acid to Its Appropriate tRNA Molecule

Synthetase Function – **Aminoacyl-tRNA synthetases** are enzymes that covalently attach each amino acid to its corresponding tRNA. Most cells have 20 synthetases, one for each amino acid, though some bacteria use fewer, with additional enzymes modifying incorrectly attached amino acids to match their tRNA [1].

Energy and Accuracy – The attachment of amino acids to tRNAs is powered by ATP hydrolysis, creating a high-energy bond between the tRNA and the amino acid, which drives peptide bond formation during

protein synthesis. Both tRNAs and synthetases play crucial roles in the decoding process, ensuring the correct amino acid is added to the growing polypeptide chain [1].

Two Adaptors – The translation of the genetic code involves two sets of adaptors: tRNAs and synthetases. Their combined action ensures that each codon in mRNA is matched with the correct amino acid, maintaining the accuracy of protein synthesis [1].

Editing By tRNA Synthetases Ensures Accuracy

Two-Step Mechanism – Aminoacyl-tRNA synthetases select the correct amino acid through a two-step mechanism, favoring the amino acid with the highest affinity for the active-site pocket while excluding larger amino acids [1].

Second Discrimination Step – After the amino acid is linked to AMP, the synthetase uses an editing pocket to exclude the correct amino acid and remove incorrectly linked ones, achieving an accuracy of one mistake in approximately 40,000 couplings [1].

tRNA Recognition – Synthetases recognize the correct tRNAs through structural and chemical complementarity, often binding directly to the tRNA anticodon, or recognizing features of the acceptor stem [1].

Amino Acids Are Added to the C-Terminal End of a Growing Polypeptide Chain

Peptide Bond Formation – Protein synthesis forms a peptide bond between the growing polypeptide's carboxyl group and an incoming amino acid's amino group [1].

Synthesis Direction – Proteins are synthesized from the N-terminal to the C-terminal end [1].

tRNA Role – The carboxyl end stays activated by covalent attachment to tRNA (peptidyl-tRNA), allowing the rapid formation of a new bond with each amino acid addition [1].

Activation Energy – Each added amino acid carries the energy needed for the next addition, illustrating "head growth" polymerization [1].

The RNA Message Is Decoded in Ribosomes

Overview of Protein Synthesis: Proteins are synthesized based on mRNA information, with ribosomes ensuring accuracy (about 1 mistake every 10,000 amino acids). Eukaryotic cells contain millions of ribosomes composed of over 50 proteins and several rRNAs [1].

Ribosome Assembly: Ribosomal subunits assemble in the nucleolus, where rRNAs associate with proteins. The small subunit matches tRNAs to mRNA codons, while the large subunit catalyzes peptide bond formation [1].

Translation Process: Ribosomal subunits join at the mRNA's 5' end to initiate synthesis, pulling mRNA through three nucleotides at a time. Eukaryotic ribosomes add 2 amino acids per second, while bacterial ribosomes add about 20 [1].

Cycle of Amino Acid Addition: Each amino acid addition involves four steps: 1. **tRNA Binding**: A tRNA binds to the A site. 2. **Peptide Bond Formation**: The polypeptide chain transfers from the P site tRNA to the A site tRNA. 3. **Large Subunit Translocation**: The large subunit shifts, moving tRNAs to the E and P sites. 4. **Small Subunit Translocation**: The small subunit moves the mRNA three nucleotides forward, ejecting the spent tRNA.

This cycle repeats, elongating the polypeptide chain from its amino to carboxyl end [1].

Elongation Factors Drive Translation Forward and Improve Its Accuracy

Role of Elongation Factors – Elongation factors **EF-Tu** and **EF-G** (or **EF1** and **EF2** in eukaryotes) enhance translation efficiency by hydrolyzing **GTP** to **GDP** during each cycle, ensuring that translation proceeds rapidly and accurately [1].

Accuracy of tRNA Binding – **EF-Tu** increases accuracy by binding **GTP** and aminoacyl-tRNAs, allowing the ribosome's **16S rRNA** to assess the correctness of codon–anticodon matches. Correct matches trigger **GTP hydrolysis**, promoting the incorporation of the amino acid into the growing polypeptide chain [1].

Proofreading Mechanisms – After GTP hydrolysis, the ribosome has a second chance to prevent incorrect amino acid addition due to a time delay in tRNA positioning. Incorrectly matched tRNAs dissociate more quickly, enhancing translation accuracy to 99.99% [1]. If a wrong amino acid is incorporated, misreading in the A site can lead to premature termination by release factors, ensuring flawed proteins are degraded [1].

Many Biological Processes Overcome the Inherent Limitations of Complementary Base-Pairing

Induced Fit Mechanism – Many biological processes, such as DNA replication, transcription, and translation, rely on **complementary base-pairing**, but the specificity achieved far exceeds what can be explained by hydrogen bonding alone. One mechanism enhancing accuracy is **induced fit**, where the ribosome or RNA polymerase folds around a correct base-pair match, completing the reaction only when the match is correct [1].

Kinetic Proofreading – Another principle is **kinetic proofreading**, where an irreversible step, like GTP hydrolysis, triggers a time delay during which incorrect codon–anticodon pairs are more likely to dissociate before catalysis. This increases the specificity of base-pairing at the cost of ATP or GTP hydrolysis [1].

Accuracy in Translation Requires an Expenditure of Free Energy

Accuracy vs. Speed – Translation balances accuracy (1 error per 10^4 amino acids) with speed (20 amino acids/second in bacteria) [1].

Energy Demand – Protein synthesis consumes more energy than any other cellular process [1].

Energy Cost per Peptide Bond – Four phosphate bonds are split for each peptide bond: two for charging tRNA, two during synthesis [1].

Proofreading Cost – Extra energy is used for error correction by tRNA synthetases and ribosome proofreading mechanisms [1].

The Ribosome Is a Ribozyme

Ribosome Composition – Ribosomes are two-thirds RNA and one-third protein. Ribosomal RNA (rRNA) is responsible for structure, tRNA positioning, and peptide bond catalysis [1].

rRNA Structure – rRNAs form a compact, 3D core that defines ribosome shape, while ribosomal proteins stabilize the structure and assist in assembly [1].

Catalytic Site – The active site for peptide bond formation is composed of rRNA, not proteins, which use hydrogen bonding to position reactants precisely [1].

Ribozyme Role – Ribosomes are ribozymes, RNA-based catalysts possibly tracing back to early cellular evolution [1].

Nucleotide Sequences in mRNA Signal Where to Start Protein Synthesis

Initiation Importance – Initiation sets the reading frame for translation and determines the protein synthesis rate [1].

AUG Codon – Translation begins at the AUG codon with initiator tRNA carrying methionine in eukaryotes or formylmethionine in bacteria [1].

Initiation in Eukaryotes – The small ribosomal subunit binds to the 5' cap of mRNA and scans for the first AUG codon [1].

Leaky Scanning – Eukaryotic ribosomes sometimes skip the first AUG, allowing production of multiple proteins from a single mRNA [1].

Bacterial Initiation – Bacteria use the Shine-Dalgarno sequence to position the ribosome at the start codon; bacterial mRNAs are often polycistronic [1].

Stop Codons Mark the End of Translation

Stop Codons – Translation ends at stop codons (UAA, UAG, UGA), which do not specify amino acids but signal the ribosome to stop [1].

Release Factors – Proteins called release factors bind to the ribosome at the stop codon, catalyzing the addition of a water molecule, releasing the polypeptide chain [1].

Ribosome Dissociation – After release, the ribosome separates into its large and small subunits, which can then assemble on another mRNA for new protein synthesis [1].

Polypeptide Tunnel – The nascent polypeptide moves through a water-filled tunnel in the ribosome, primarily formed by 23S rRNA, which allows the peptide to slide easily [1].

Protein Folding – As proteins exit the ribosome, they must fold into their functional three-dimensional structures, a process discussed later in the chapter [1].

Proteins Are Made on Polyribosomes

Polyribosomes – Multiple ribosomes translate a single mRNA simultaneously, forming polyribosomes (or polysomes) spaced about 80 nucleotides apart, increasing protein synthesis efficiency [1].

Multiple Initiations – As one ribosome moves along the mRNA, it allows new ribosomes to initiate translation, enabling rapid production of protein molecules [1].

Bacterial Translation – In bacteria, ribosomes can begin translation while the mRNA is still being synthesized, tightly coupling transcription and translation [1].

Eukaryotic Translation – Eukaryotic mRNAs undergo processing, but their 5' and 3' ends interact, allowing ribosomal subunits to quickly reinitiate translation after dissociation [1].

There Are Minor Variations in the Standard Genetic Code

Common Ancestry – The genetic code is largely universal across life forms, supporting the idea of a common ancestry among all organisms [1].

Exceptions – Rare exceptions exist, such as the translation of CUG as serine in Candida albicans and AUA as methionine in mammalian mitochondria, differing from the standard leucine and isoleucine translations, respectively [1].

Translation Recoding – Some cells can alter the meaning of the genetic code at specific sites through translation recoding, utilizing additional nucleotide sequences in mRNA to change codon interpretations [1].

Selenocysteine – A 21st amino acid, selenocysteine, is incorporated into proteins through a specialized tRNA that base-pairs with the UGA stop codon, requiring a specific mRNA sequence for this recoding [1].

Inhibitors of Prokaryotic Protein Synthesis Are Useful as Antibiotics

Source of Antibiotics – Many antibiotics, produced by fungi, inhibit bacterial protein synthesis, reflecting coevolution for environmental niches [1].

Targeting Ribosomes – These drugs exploit structural differences between bacterial and eukaryotic ribosomes, allowing high dosages with low toxicity to humans [1].

Mechanisms of Action – Antibiotics interfere with ribosomal function by blocking specific sites, like the exit channel, or by lodging in ribosomal RNAs [1].

Research Applications – Compounds like chloramphenicol, cycloheximide, and puromycin are widely used in cell biology studies due to their targeted inhibition of protein synthesis [1].

Quality Control Mechanisms Act to Prevent Translation of Damaged mRNAs

mRNA Processing – In eukaryotes, mRNA undergoes transcription and processing in the nucleus before being transported to the cytosol for translation [1].

Surveillance Systems – Mechanisms like nonsense-mediated mRNA decay eliminate defective mRNAs, preventing translation of truncated proteins [1].

EJC Functionality – Exon junction complexes (EJCs) are displaced by the ribosome; if a stop codon is reached too early, the mRNA is rapidly degraded [1].

Evolutionary Significance – This quality control may have facilitated eukaryotic evolution by selecting for full-length proteins from rearranged genes [1].

Impact on Genetic Disorders – Nonsense-mediated decay helps mitigate symptoms of genetic disorders by degrading aberrant mRNAs, preventing toxic protein production [1].

Some Proteins Begin to Fold While Still Being Synthesized

Folding Begins Early – Protein folding can start during synthesis, beginning with the N-terminal end as it emerges from the ribosome [1].

Folding Process – Proteins fold to their lowest free energy conformation, burying hydrophobic residues and forming noncovalent interactions [1].

Molten Globule State – Some proteins form a dynamic "molten globule" structure before undergoing slower side-chain adjustments to reach the correct tertiary structure [1].

Evolutionary Selection – Proteins are evolutionarily optimized not only for their final conformation but also for rapid folding [1].

Completion Timing – Much of a protein's folding may be complete by the time its synthesis finishes [1].

Molecular Chaperones Help Guide the Folding of Most Proteins

Role of Chaperones – Molecular chaperones assist proteins that do not fold correctly during synthesis, preventing misfolding and aggregation [1].

Kinetic Trapping – Proteins can become "kinetically trapped" in incorrect configurations without chaperones, leading to nonfunctional structures [1].

Recognition Mechanism – Chaperones recognize misfolded proteins by their exposed hydrophobic surfaces, typically buried in correctly folded proteins [1].

Prevention of Aggregates – By binding to these hydrophobic surfaces, chaperones prevent harmful aggregates from forming [1].

Refolding Opportunity – Chaperones release proteins in a manner that allows them to attempt proper folding again [1].

Cells Utilize Several Types of Chaperones

Chaperone Families – Hsp60 and hsp70 are key molecular chaperones that help proteins fold correctly, preventing aggregation [1].

Heat-Shock Response – Increased production of chaperones like hsp70 occurs during stress (e.g., heat), boosting protein repair [1].

Hsp70 Function – Acts early during protein synthesis by binding to unfolded proteins, helping them refold with each cycle of ATP hydrolysis [1].

Hsp60 Function – Functions post-synthesis in a barrel-shaped chamber where proteins fold in isolation, protected from aggregation [1].

ATP Hydrolysis – Both hsp60 and hsp70 use ATP to ensure proper folding, recycling proteins through binding and release [1].

Exposed Hydrophobic Regions Provide Critical Signals for Protein Quality Control

Role of Hydrophobic Regions – Exposed hydrophobic regions signal misfolding, indicating the need for quality control mechanisms [1].

Chaperone Involvement – Hsp70 assists during synthesis, and hsp60 helps with post-synthesis folding, recognizing abnormal hydrophobic patches [1].

Protein Repair or Destruction – Chaperones provide misfolded proteins extra folding attempts while preventing aggregation; persistent issues trigger degradation [1].

Proteolytic Pathway – Failed refolding leads to proteolysis via the proteasome, marking and destroying misfolded proteins to prevent damage [1].

The Proteasome Is a Compartmentalized Protease with Sequestered Active Sites

Proteasome Function – The proteasome degrades misfolded or abnormal proteins, competing with chaperones for recognition [1].

Structure – Composed of a central 20S core with protease-active sites inside and 19S caps that regulate entry, the proteasome unfolds and degrades proteins [1].

Targeting for Destruction – Proteins are tagged for destruction by ubiquitin chains, specifically linked at lysine 48, marking them for the proteasome [1].

Unfolding Mechanism – ATP hydrolysis drives the unfolding and threading of target proteins through the 19S cap into the 20S core, ensuring full degradation [1].

Regulation – The proteasome distinguishes between misfolded proteins and normal proteins, destroying the former while sparing those with proper conformations [1]

Many Proteins Are Controlled by Regulated Destruction

Functions of Proteolytic Mechanisms – Proteolytic pathways recognize and eliminate misfolded proteins while also controlling the lifetimes of normal proteins, allowing their concentrations to change with cellular states [1].

Short-Lived Proteins – Some proteins are conditionally short-lived, remaining stable under certain conditions but becoming unstable with changes in the cell's state. For instance, mitotic cyclins are long-lived during most of the cell cycle but are degraded at the end of mitosis [1].

Mechanisms of Regulated Destruction – Regulated destruction can occur through [1]:

- E3 Activation: Ubiquitin ligase activity can be turned on via phosphorylation or allosteric changes, such as in the anaphase-promoting complex (APC) during mitosis.
- Creation of Degradation Signals: Intracellular or environmental signals can create degradation signals through phosphorylation or subunit dissociation.
- N-terminal Cleavage: Cleaving a peptide bond may expose a new N-terminus recognized by specific E3 proteins for degradation

Acetylation and Destruction Signals – Many human proteins are acetylated at their N-terminus, and this modification acts as a signal for degradation, becoming exposed as proteins age or misfold [1].

Summary - RNA to Protein

Translation Process
- Location: Translation occurs in the cytosol on ribosomes, large ribonucleoprotein assemblies. tRNA and Codons: Each amino acid attaches to a tRNA that recognizes specific mRNA codons through base-pair interactions.
- Initiation: Translation begins at the start codon (AUG) with the help of an initiator tRNA, followed by elongation and termination steps.
-

Elongation Steps
- Aminoacyl-tRNA Binding: Sequential binding of aminoacyl-tRNAs to corresponding codons occurs, adding amino acids to the C-terminal end of the polypeptide.
- Sequential Steps: Four steps: binding, peptide bond formation, and two translocation steps.
- Role of Elongation Factors: GTP hydrolysis drives reactions and enhances accuracy during amino acid selection.

Termination of Translation
- The mRNA is read in the 5'-to-3' direction until a stop codon is reached, where a release factor binds to terminate translation and release the polypeptide.

Ribosomal Structure
- Eukaryotic vs. Bacterial Ribosomes: Similarities exist despite differences in rRNA and protein components. rRNA plays a crucial role in ribosome structure and function.

Chaperones and Protein Folding
- Molecular Chaperones: Hsp60 and hsp70 chaperones recognize exposed hydrophobic patches, preventing aggregation and assisting proper folding of polypeptides.

- Quality Control Mechanism: Misfolded proteins are marked with ubiquitin by a ubiquitin ligase, leading to proteolytic degradation by the proteasome. This mechanism also regulates the lifetimes of normally folded proteins based on specific signals

The RNA World and the Origins of Life

Interdependence of Nucleic Acids and Proteins – The complexity of heredity involves DNA, RNA, and proteins, creating a paradox about their origins [1].

RNA World Hypothesis – Proposes that an RNA world existed prior to modern cells, where RNA stored genetic information and catalyzed reactions [1].

Transition to DNA and Proteins – Over time, DNA became the genetic material and proteins the primary catalysts, though RNA still plays crucial roles [1].

Unique Properties of RNA – RNA's ability to serve as both genetic information carrier and ribozyme is central to the RNA world hypothesis, suggesting its importance in early life [1].

Single-Stranded RNA Molecules Can Fold into Highly Elaborate Structures

RNA Folding – RNA molecules fold uniquely through complementary base-pairing and hydrogen bonds, driven by their nucleotide sequences [1].

Conserved Structural Motifs – RNA forms conserved motifs used repeatedly in larger structures, allowing for specialized shapes [1].

Catalytic Properties – RNA molecules, like proteins, can act as catalysts (ribozymes) by positioning metal ions at their active sites, enabling diverse biochemical reactions [1].

RNA World Insights – Lab experiments selecting RNA with specific properties suggest that RNA can catalyze various reactions, though proteins offer more catalytic flexibility with 20 amino acids [1].

RNA Can Both Store Information and Catalyze Chemical Reactions

Self-Replication Ability – RNA can guide the formation of copies of its sequence through complementary base-pairing, similar to DNA replication and transcription [1].

Catalyst Requirement – Efficient RNA synthesis requires catalysts to promote polymerization, as RNA formation without them is slow and error-prone [1].

RNA as Catalysts – RNA molecules likely served as early catalysts for template-dependent RNA synthesis, supporting the RNA world hypothesis [1].

Experimental Advances – While self-replicating RNA systems have not been found in nature, laboratory experiments show progress, suggesting such systems could have been possible [1].

How Did Protein Synthesis Evolve?

Complexity of Protein Synthesis – The intricate interplay of RNA and proteins in modern protein synthesis makes its evolution hard to conceptualize [1].

Primitive Synthesis – Early, noncoded peptide synthesis likely occurred without ribosomes, possibly catalyzed by RNA molecules in the RNA world [1].

Genetic Code Origins – Lab-made ribozymes can perform aminoacylation, hinting that tRNA-like adaptors may have initiated the genetic code [1].

Protein Dominance – Once coded synthesis began, proteins became dominant due to their versatility with 20 subunits versus RNA's 4 [1].

All Present-Day Cells Use DNA as Their Hereditary Material

RNA Preceded DNA – RNA likely stored genetic information in early cells before DNA emerged, supported by the simpler formation of ribose compared to deoxyribose [1].
DNA Stability – DNA, with its deoxyribose backbone, is chemically more stable, allowing for longer chains and better preservation of genetic information [1].
Structural Advantages – The double-helix structure and the use of thymine instead of uracil make DNA more resistant to damage and easier to repair than RNA [1]

Summary - The RNA World and the Origins of Life

RNA as the Initial Catalyst – Early life likely began with RNA molecules capable of self-replication, acting as the first autocatalytic systems [1].
DNA as a Later Addition – With the evolution of protein catalysts and more complex cells, DNA replaced RNA as a more stable genetic storage molecule [1].
Increased Genetic Information – DNA's stability allowed for the storage of larger amounts of genetic information, essential for the complexity of evolving life forms [1].

5.2   2nd Lecture

5.2.1   Article 1 - Classification of Yeasts

Introduction – The History of Yeasts

Definition of Yeasts – Yeasts are unicellular fungi reproducing through budding or fission, significant in food and beverage fermentation and spoilage [3].
Yeast Taxonomy Evolution – Changes in yeast taxonomy, especially in Saccharomyces species, are driven by molecular methods and increasing interest in yeast biotechnology [3].
Role of Saccharomyces in Fermentation – Saccharomyces species play a central role in industrial fermentation, though taxonomical updates may complicate their practical identification [3].
S. cerevisiae Genome Sequencing – The complete sequencing of S. cerevisiae in 1996, with its 16 chromosomes and 13 million base-pairs, marked a significant achievement in yeast research [3].

Classification of Yeasts

Traditional Methods – Yeast classification is traditionally based on morphological, physiological, biochemical, and ecological traits [3].
Modern Techniques – Advances include nuclear DNA homologies, rRNA-DNA homology, protein patterns, and cell wall composition analysis [3].
Taxonomic Groups – Current yeast taxonomy classifies yeasts into two phyla: Ascomycota and Basidiomycota, covering both teleomorphic and anamorphic species [3].
Food-Related Yeasts – Yeasts involved in food production and spoilage are typically ascomycetes or anamorphic basidiomycetes [3].

Procedures for classification and identification of yeasts

Culture characteristics

Cell Morphology – Cell shape (e.g., spheroidal, ovoid) and arrangement (single, pairs, aggregates) are used for yeast identification [3].

Growth in Liquid Media – Yeast growth can form sediments, rings, or pellicles with characteristic smells. Hyphal formation may result in surface growth [3].

Colony Morphology – Colony characteristics on agar include texture, color, surface, elevation, and margin, which aid in species identification [3].

Characteristics of Vegetative Reproduction

Types of Reproduction – Yeasts reproduce vegetatively by budding, fission, or forming arthroconidia [3].

Holoblastic Budding – Budding occurs with the formation of a small evagination, leading to the formation of a new cell. After separation, a chitin scar remains [3].

Enteroblastic Budding – In rare cases, the bud forms through a rupture in the parent cell wall [3].

Budding Patterns – Budding can be monopolar, bipolar, or multilateral, with multilateral being the most common [3].

Fission and Arthroconidia – Fission occurs in genera like *Schizosaccharomyces* and arthroconidia form from the end of hyphae [3].

Formation of Pseudomycelium and True Mycelium

Budding and Chains – In vegetative reproduction, buds may separate or remain attached, forming clusters that lead to pseudomycelium [3].

Pseudomycelium – Characterized by filamentous structures from budding without distinct septa; terminal cells are often smaller with clear constrictions [3].

True Mycelium – Formed by continuous growth with distinct septa; terminal cells are longer than adjacent cells, showing minimal constriction at septa [3].

Distinguishing Forms – Differentiating between pseudohyphae and true hyphae can be challenging and influenced by cultural conditions [3].

Mycelium Formation Demonstration – Mycelium growth can be observed on Yeast Morphology agar under aerobic and anaerobic conditions [3].

In the table 5.5, some yeasts relevant to food and beverage production and spoilage are listed based on their ability to form pseudomycelium and true mycelium [3].

Characteristics of Ascospore Formation

Sexual Reproduction in Yeasts – In yeasts relevant to food and beverage production, sexual reproduction occurs through ascospores; genera forming basidia are not significant in this context [3].

Importance of Sporulation – Sporulation is crucial for yeast taxonomy and genetic resilience; sporogenous yeasts can better withstand environmental stresses and are often implicated in spoilage, e.g., *S. cerevisiae* var. diastaticus [3].

Mechanisms of Ascus Formation – Asci can form through various mechanisms: (1) haploid cell mitosis with a vestigial bud, (2) conjugation of two free cells, or (3) reduction division of a diploid cell [3].

Table 5.5: formation of pseudomycelium and true mycelium for yeasts relevant to the production and spoilage of foods and beverages.

PSEUDOMYCELIUM	TRUE MYCELIUM
Debaryomyces	*Galactomyces (Geotrichum)*
Dekkera (Brettanomyces)	*Schizosaccharomyces*
Hanseniaspora (Kloeckera)	*Trichosporon*
Kazachstania	
Kluyveromyces	
Phaffia	
Saccharomyces	
Saccharomycodes (poorly)	
Torulaspora	
Zygosaccharomyces	
PSEUDO- AND TRUE MYCELIUM	
Candida	
Pichia	
Rhodutorula	
Saccharomycopsis	
Yarrowia	

Variability in Ascospores – Ascospores may vary in shape and surface texture, with numbers per ascus typically ranging from 1 to 8, influenced by the yeast genus [3].

Sporulation Conditions – Optimal sporulation requires specific media and conditions; factors like nutrient availability, oxygen, and temperature affect spore quality and production [3].

Table 5.6 shows a list with some of the typical ascospore forms for the various genera of interest [3].

Table 5.6: A list with the typical ascospore form for the various genera of interest in food and beverage production and spoilage.

Spherical, Ellipsoidal, or Reniform	Hat- or Saturnshaped (Spherical, Ellipsoidal or Reniform may occur)
Debaryomyces	*Dekkera (Brettanomyces)*
Galactomyces (Geotrichum)	*Hanseniaspora (Kloeckera)*
Kazachstania	*Pichia*
Kluyveromyces	*Saccharomycopsis*
Saccharomyces	*Yarrowia*
Saccharomycodes	
Schizosaccharomyces	
Torulaspora	
Zygosaccharomyces	

Biochemical characteristics

Fermentation of Carbohydrates

Fermentative Yeasts – Most yeasts of interest in food and beverages, except *Rhodotorula*, ferment glucose and other sugars [3].

Fermentation Tests – CO_2 production is measured in Durham tubes using 2[w/V]% sugar in nutrient-rich media to evaluate fermentation [3].

Nutritional Medium – A standard medium contains 4.5 g yeast extract and 7.5 g Bacto peptone per 1000 ml, with pH indicators and additional vitamins for specific yeasts [3].

Yeast Inoculum – Defined inoculum from 24-48 hour growth on Yeast Morphology Agar at 25°C ensures reliable fermentation results [3].

Pattern Instability – Fermentation patterns can change due to mutations linked to prolonged exposure to specific carbohydrates [3].

Assimilation of Carbon Compounds

Testing Method – Growth is assessed in liquid basal media, like Difco Yeast Nitrogen Base, where the tested compound is the sole carbon source, requiring standardized inocula free from interfering carbon compounds [3].

Standardized Procedures – Tests typically include up to 30 carbon compounds, with incubation at 25°C and weekly growth evaluations over a month, including a blank control [3].

Auxanographic Tests – For rapid results, auxanographic tests using commercial kits like API 20 Auxanogram can provide useful data in 2-3 days, benefiting food and beverage microbiology despite their clinical origin [3].

Importance in Classification – Assimilation of carbon compounds is a crucial factor in distinguishing yeast species for classification purposes [3].

Assimilation of Nitrogen Compounds

Testing Method – Nitrogen assimilation is assessed similarly to carbon compounds in a nitrogen-free liquid medium, such as Difco Carbon Base, requiring careful inoculum preparation [3].

Key Nitrogen Sources – Useful sole nitrogen sources include nitrate, nitrite, ethylamine hydrochloride, cadaverine dihydrochloride, L-lysine, creatinine, and creatine, with potassium nitrate being the most significant for genus-level discrimination [3].

Stability and Limitations – Nitrate utilization appears stable at the species level, with genera such as *Saccharomyces*, *Kluyveromyces*, *Kazachstania*, and *Debaryomyces* being nitrate-negative, while *Candida* and *Pichia* contain both nitrate-positive and nitrate-negative species [3].

Other Characteristics

Vitamin Requirements – Growth in vitamin-free media is rare among yeasts and can be a key identification criterion [3].

Growth Temperature – Growth at 37°C helps differentiate ale from lager yeasts, as lager yeasts do not grow at this temperature [3].

Growth Conditions – Identification also considers growth on 50% [w/w] glucose yeast extract agar or in 10% [w/w] NaCl with 5% [w/w] glucose in Yeast Nitrogen Base [3].

Resistance and Pigments – Actidione (cycloheximide) resistance is crucial for detecting wild yeasts, along with the formation of pigments and esters [3].

Coenzyme Q System – Coenzyme Q, varying in isoprene units, is noted in some species, with ascomycetous yeasts showing 6 to 10 units and basidiomycetous yeasts showing 8 to 10 [3].

Yeast strain characterization by technological properties

Industrial Yeasts – *Saccharomyces cerevisiae* (ale) and *Saccharomyces pastorianus* (lager) are characterized by technological traits like cell multiplication, flocculation, and attenuation rather than taxonomy [3].

Characterization System – Brewing yeasts are identified by a six-digit code predicting fermentation performance. Similar systems apply to wine yeasts [3].

Debaryomyces hansenii – NaCl tolerance, pH regulation, and proteolytic activity are key in meat and cheese fermentation [3].

Yeast Strain Characterization by Molecular Methods

Molecular Methods – Various techniques like species-specific primers, RAPD, RFLP, PFGE, and sequence-based methods are used to identify and differentiate yeast strains. Sequencing of the D1/D2 region of the 26S rRNA gene is effective for species identification [3].

PCR Techniques – Species-specific primers help identify individual species but can produce uncertain results with many species. RAPD amplifies random DNA segments and is used for differentiating brewing yeasts, but reproducibility between labs can be problematic [3].

RFLP and PFGE – RFLP uses restriction enzymes to analyze DNA fragments, applicable for differentiating lager and ale yeasts. PFGE (karyotyping) separates large DNA fragments and is useful for wine yeast identification, though labor-intensive [3].

Rep-PCR and DGGE – Rep-PCR groups isolates based on repetitive DNA elements, helpful for partial characterization, e.g., in cocoa fermentation. DGGE separates DNA fragments for population dynamics studies, with sensitivity as low as 10^2 [cfu/ml], useful in wine and coffee fermentation [3]

Characteristics of food borne yeast genera

This section includes approximately 10 pages of tables (Tables VI to XXIII in the Reading material (RM)) detailing the taxonomic characteristics of food-borne yeasts relevant to food and beverage production and spoilage, identifying them at the genus level. For species-level identification, refer to the previous criteria and The Yeasts, a Taxonomic Study, 5th edition (Kurtzman, Fell, and Boekhout, 2011) [3].

5.3   3rd Lecture

Introduction

Bioinformatics plays a key role in classification and identification, historically through the 16S rRNA gene, now shifting towards whole genome sequences [2].

16S rRNA gene analysis remains a gold standard, but the use of full genomic sequences is growing [2].

The **phenotype** is still essential for classification. New species must show at least two phenotypic differences from known species, though predicting phenotypes from gene sequences is still limited [2].

Identification links bacterial strains to species, though traditional methods are slow and sometimes inaccurate [2].

16S rRNA and other genes, such as *rpoB*, are used to identify difficult isolates with close genetic relationships [2].

5.3.1 Classification of Prokaryotes

Classification Based on 16S rRNA Gene Sequence Comparison

16S rRNA sequence comparison is fundamental for classifying bacteria at species, genus, and higher taxonomic levels [2].

The **16S rRNA gene** is used because it forms the backbone of ribosomes, making it universal across organisms [2].

Gene sequence similarity is less than 97% between species and less than 95% between genera [2].

Other methods, such as DNA-DNA hybridization, are used when species have closer than 97% similarity [2].

Members of **Archaea** require different primer sets due to their distinct 16S rRNA genes [2].

PCR amplification with conserved primers can be applied to all bacteria, confirmed by agarose gel electrophoresis [2].

Traditional **Sanger sequencing** is reliable, though newer methods like PacBio and Nanopore are emerging [2].

The **16S rRNA gene sequence** is used for database comparison, alignment, and phylogenetic analysis [2].

DNA-DNA hybridization is used to classify bacterial strains at the species level by measuring the reassociation of denatured DNA from two strains [2].

Hybridization above 70% indicates the strains belong to the same species; below 70% suggests they are different [2].

Whole genome sequence comparison has replaced traditional DDN, measured by ANI (average nucleotide identity) [2].

An **ANI of 95%** correlates with the conventional 70% DDH cutoff, though results between 93-96% require cautious interpretation [2].

Tools like **GGDC** and **JSpecies** calculate DDN or ANI from genome comparisons [2].

GC content offers taxonomic value, with species within the same genus differing by no more than 5% [2].

Classification Based on Multilocus Sequence Analysis (MLSA)

MLSA complements *16S rRNA* analysis by using concatenated DNA sequences of conserved genes [2].

Traditionally, up to 10 genes are used, but whole genome comparisons now allow analysis of all conserved genes or proteins [2].

For genomic **MLSA**, a minimum of 31 genes or proteins is recommended [2].

MLSA is applied at species and higher taxonomic levels [2].

5.3.2 Classification of the Taxonomic Hierarchy

Classification of Species

Taxonomic Unit – Species is the main taxonomic unit, comprising prokaryotes with 16S rRNA gene sequence similarities above 97% [2].

Confirming Diversity – DNA-DNA hybridization confirms diversity when 16S rRNA similarities exceed 97%, requiring over 70% reassociation [2].

Genomic Comparisons – Whole genomic sequences are replacing experimental DNA-DNA hybridization for grouping species [2].

Gene Sequencing – Housekeeping genes (*recN*, *rpoA*, *thdF*) can define species, with *recN* alone being sufficient [2].

Classification of Genera

Less Strict Rules – Classification at the genus level has more lenient criteria, typically requiring less than 95% 16S rRNA gene sequence similarity [2].

Metagenomic Analysis – A 96% cutoff is used in metagenomic studies based on partial 16S rRNA gene sequences [2].

Protein Comparison – Average nucleotide identity (ANI) can also be applied to protein comparisons, with conserved protein percentages (POCP) verifying genus classification [2].

Protein Similarity – Different genera should share less than 50% of their proteins to maintain distinct classification [2].

Classification of Families, Orders, Classes, and Phyla

Lenient Rules – The classification of higher hierarchical units like families, orders, classes, and phyla is less strict, often formed to accommodate new genera, which may create new orders [2].

16S rRNA Identity – Analysis indicates a minimum of 87.5% 16S rRNA gene sequence identity between type strains of different families [2].

Monophyletic Groups – While no strict sequence limits exist, families, orders, classes, and phyla should be monophyletic [2].

Whole Genomic Comparisons – Efforts are underway to define these units based on whole genomic comparisons, emphasizing conserved genes [2].

Nomenclature Rules – Families, orders, and classes are governed by the International Code of Nomenclature of Prokaryotes, while phyla are not, and new names must include the type genus [2].

An example of the taxonomic hierarchy for prokaryotes is shown in table 5.7 for the bacteria *Escherichia coli* [2].

Table 5.7: Prokaryotic taxonomic categories have validated names written in cursive (ICNP). New classes follow the genus (ICNP rule 8)

Prokaryotes: Bacteria (renum nov.) and Archaea (regnum nov.)		
'Phylum'	[Proteobacteria]	
Class (klasse)	[Gammaproteobacteria]	
Order (orden)	[*Enterobacteriales*]	
Family (familie)	[*Enterobacteriaceae*]	Prokaryotic Code
Genus (slægt)	[*Escherichia*]	
Species (art)	[*Escherichia coli*]	
Subspecies (underart)		

5.3.3   Rules for the Naming of a New Prokaryote

Nomenclature Authority – Prokaryotes are named according to the International Code of Nomenclature of Prokaryotes; names published in IJSEM are valid, while others may be validated through acceptance on the VL [2].

Synonyms – Homotypic synonyms share the same type strain (e.g., Klebsiella mobilis and Enterobacter aerogenes), while heterotypic synonyms have different type strains, with the first published name typically having priority [2].

Naming Format – Prokaryotic names are italicized and consist of a genus and specific epithet, following rules established for bacteria and referenced in the Approved Lists and LPSN [2].

Bacterial Species Names Are Linked to the Type Strain

Type Strain Definition – The type strain is a well-characterized strain representing a species and must be deposited in at least two public culture collections for accessibility [2].
Name Bearer Role – The type strain acts as the "name bearer" of the species, maintaining its designation even if transferred to a new genus, such as the case of Streptococcus faecalis to Enterococcus faecalis [2].
Importance in Taxonomy – Although the type strain may lose virulence over time, it remains essential for classification and identification purposes [2].
Examples – The subchapter includes examples of "An Old Bacterial Name that Never Changed" and "A Taxon with Many Reclassifications and Changes in Genus Name". These examples can be found in the RM [2].

5.3.4   The Benefits of Sequence-Based Identification

DNA sequences enable global comparison. They serve as electronic public information, often replacing live strain analysis [2].
Enhances species-level identification. DNA sequence-based methods support molecular epidemiology and population genetics [2].
Annotations predict phenotypes. Predictive capabilities of DNA annotations are crucial for understanding phenotypic properties [2].

16S rRNA Sequence-Based Identification

EzBioCloud for precise identification. Comparison in EzBioCloud enables identification against validated bacterial type strains, often outperforming GenBank BLAST searches due to strain status uncertainty [2].
Phylogenetic analysis enhances identification. Phylogenetic analysis may be necessary for distinguishing closely related species [2].
MALDI-TOF MS as an alternative method. This method offers fast, cost-effective bacterial identification, although it has limitations due to database size and species resolution [2].
16S rRNA identification is a gold standard. It serves as a benchmark for developing the MALDI-TOF database [2].

16S rRNA-Based Identification Without Culture

Prokaryotes characterized without culture. In the late 1980s, rRNA was extracted, reverse transcribed to DNA, cloned, and sequenced [2].
PCR amplifies 16S rRNA genes. PCR was used to amplify 16S rRNA genes from environmental DNA, followed by cloning and sequencing [2].
High-throughput sequencing advances metagenomics. High-throughput sequencing of 16S rRNA target regions V1–V3 enables deep sequencing of environmental samples [2].
Short-read length limits accuracy. Current short-read lengths in 16S rRNA-targeted metagenomics restrict species-level identification accuracy [2].

5.3.5 Activity: Identifying Bacterial Isolates

PCR amplification for difficult isolates. When phenotypic identification fails, amplify the 16S rRNA gene and sequence the DNA [2].

Challenges with traditional BLAST. Traditional nucleotide BLAST in databases like GenBank is complicated by unreliable species names associated with sequences [2].

Type strain sequences for accuracy. The ideal reference is the type strain's sequence, but locating it in BLAST output can be challenging [2].

EzBioCloud for streamlined identification. A server by Professor Chun searches 16S rRNA sequences of type strains only [2].

Expected result from the activity. Download sequence KX858032 from NCBI and use EzBioCloud to identify it; the expected top hit shows 96.1% similarity, indicating a potential new species [2].

5.3.6 Take-Home Messages

This part is copy paste from the RM to ensure that the key points are not missed [2].

- Bioinformatics is important to classification and identification since all prokaryotes have been classified based on the phylogeny of the 16S rRNA gene sequence and this character also to a great extent has been the gold standard for identification.
- The classification of prokaryotes at the species level based on DNA-DNA hybridization can be estimated in silico by comparison of whole genomics sequences.
- Whole genomic sequences are increasingly used also for the classification of genera, families, orders, classes, and phyla by phylogenetic comparison of common core genes or predicted proteins.
- The *rpoB* gene sequence is frequently used for identification since it allows the separation of some species that cannot be identified by 16S rRNA gene sequence comparison.
- Identification of prokaryotes that cannot be cultured is possible by 16S rRNA gene amplicon sequencing which is further described in 7 Chap. 8 (in the respective RM).

Chapter 6
Exam

6.1 Regarding the Curriculum

The questions asked will cover the course curriculum in general. Be aware that we specifically expect you to be able to describe the principles behind molecular methodologies for identification (species) and typing (subspecies/strain) of both pro- and eukaryotes.

6.2 Guideline for exam questions

You need to prepare for all of the 13 exam questions before the exam. You are allowed to use notes when presenting the exam question and to bring printed versions of your lab group slides. It is up to you to organize the oral presentation (no power point or shared screen are allowed). We recommend that you prepare keywords, and practice what you want to say based on the keywords. Emphasize the points you find most relevant and would like to discuss. Please notice, that it is not allowed to read aloud from a written text.

The grade will primarily reflect the your ability to analyze and discuss the exam question and other topics related to the curriculum.

6.3 Exam Questions

6.3.1 AAB

Question

Acetic acid bacteria (AAB) are important in several food and biotechnological processes. Define AAB and name the most important AAB relevant to food. Describe at least two food processes where AAB play an essential role and explain their functions in these products/processes.

Answer

AAB

AAB belong to the family *Acetobacteraceae* and are primarily Gram-negative, with a few Gram-variable exceptions. They are obligate aerobes that require oxygen for growth and are capable of oxidizing ethanol into

acetic acid, a key trait in food and biotechnological processes.

Characteristics

- Gram-negative, ellipsoid to rod-shaped.
- Catalase positive, oxidase negative.
- Mesophilic, with optimal growth between 25-30°C, though some species can tolerate higher temperatures.
- pH optimum is between 5-6.5, though some AAB can grow at pH as low as 3.
- Oxidize sugars and alcohols to organic acids, primarily ethanol to acetic acid (*Asaia* spp. being the exception, as it does not oxidize ethanol).

Notable genera

Acetobacter (found in sugar-rich environments) and *Gluconobacter* (found in alcohol-rich environments).

Classification

AAB are classified as follows:
- Phylum: *Proteobacteria*
- Class: *Alpha Proteobacteria*
- Family: *Acetobacteraceae*
- Genera: *Acetobacter*

Phenotypical characteristics differentiations

Here are some common characteristics that can be used to differentiate AAB from each other:

Table 6.1: Differential characteristics of AAB

Characteristics	
Flagellation	Oxidation of ethanol to acetic acid
Oxidation of acetic acid to CO_2 and H_2O	Growth on 0.35% acetic acid-containing medium
Growth on methanol	Growth on D-mannitol
Growth in the presence of 30% D-glucose	Production of cellulose
Production of levan-like mucous substance from sucrose	Fixation of molecular nitrogen
Ketogenesis (dihydroxyacetone) from glycerol	

Acid production from:
D-mannitol
Glycerol
Raffinose
Cellular fatty acid type
Ubiquinone type
DNA base composition [mol% G+C]

Identification

Identification of AAB can be done using phenotypic tests, which are tedious and difficult to perform. Therefore many prefer to use molecular biology-based tests instead. PCR-RFLP-analysis, Rep-PCR, 16S rRNA gene

sequencing, MALDI-TOF, MLSA, and full genome sequencing are some of the methods used for identification.

Food processes

6.3.2 AAB

Question

Name at least one suitable substrate for isolation of AAB and explain the selective and/or indicative principle(s) behind the substrate. Name at least three phenotypic features of AAB that can be used to differentiate AAB from other microorganisms (LAB, yeasts) occasionally growing on AAB-selective substrates. Give an overview of methods that can be used to identify the different isolates to species level.

6.3.3 Yeast

Question

Give 2-3 examples on foods and/or beverages that are fermented by yeasts. Explain whether single or multiple species participate in the fermentation and whether the yeast(s) are added as starter culture(s) or occur spontaneously. Describe the functions of the yeasts in the products. Explain how yeast can be identified to species level and give examples of typing techniques for strain characterization.

6.3.4 Yeast

Question

Describe the micro- and macromorphological characteristics of the species Saccharomyces cerevisiae and explain how this species differs genetically from Saccharomyces pastorianus. Give examples on food and/or beverages that favor the growth of Saccharomyces spp. and mention potential yeast interactions with other types of microorganisms.

6.3.5 Bacillus

Question

Identify the most commonly occurring species of aerobic, endospore-forming microorganisms found in alkaline fermented foods. Mention at least 2-3 examples of desired technological properties of Bacillus spp. and how these technological properties influence the quality of the final product. Discuss shortly the implications of spontaneous as compared to controlled fermentations for the quality and safety of the final product.

6.3.6 LAB

Question

The former Lactobacillus genus has been split into 25 new genera. Among these new genera are *Companilactobacillus*, *Lactiplantibacillus*, *Lactobacillus*, *Lacticaseibacillus*, *Levilactobacillus*, *Lentilactobacillus*, and *Limosilactobacillus*. From these genera, mention two species of relevance for fermentation of food or beverages, one of them being a specialist and the other a generalist; describe their habitat, typical phenotypes, and their

contribution to the fermentation. Describe how you would identify the species by use of phenotypic and molecular methodologies.

6.3.7 LAB

Question

Explain the strain definition for bacteria. Describe the methods you will use for strain typing and the principles behind these methods (select 2-3 methods). Additionally, explain how these methods can be used to determine whether different strains of *Lacticaseibacillus casei* used as ripening cultures during cheese production are able to grow in the cheese.

6.3.8 LAB

Question

Give an overview of the antimicrobial compounds produced by LAB. How do you detect the inhibitory compounds and how do you investigate which types of inhibitory compounds are produced? How will you determine that an inhibitory compound is a bacteriocin? Classify bacteriocins and give examples of their inhibition spectrum and mode of action.

6.3.9 Moulds

Question

Mention 2-3 mould genera and methodologies for their identification to species level. Give examples (2-3) of how moulds are used in fermentation of foods and beverages. Discuss the positive and negative characteristics of moulds in food production.

6.3.10 Microbial Interactions/Quorum Sensing

Question

Describe the principles behind quorum sensing (QS) and give examples of microbial properties (traits) that are under QS control. Explain how QS differs from other types of microbial interactions. Mention 2-3 types of microbial interactions which are not under QS control and explain their mode of action.

6.3.11 Probiotics

Question

Give a definition of probiotics and name 2-3 species where strains are commonly applied as probiotics. Give examples on properties that a potential probiotic candidate can fulfil. Line out diverse ways of investigating these probiotic properties. Probiotics might have beneficial effects against diarrhea. What are the proposed mechanisms behind this effect?

6.3.12   Plant-based Fermented Foods/Health

Question

Describe how microorganisms, through fermentation of plant-based foods, can benefit nutritional quality and human health. Include at least 2-3 different effects and discuss their underlying mechanisms.

6.3.13   Sequencing Methods

Question

Define culture-independent and culture-dependent techniques for microbial characterization of fermented food and describe advantages and disadvantages of the two approaches. Give 1-2 examples on how you can apply Nanopore amplicon sequencing for species identification. Discuss the advantages and disadvantages of amplicon sequencing compared to whole genome sequencing.

Bibliography

[1] B. Alberts. *Molecular Biology of the Cell*. W.W. Norton, 2017. ISBN: 9781317563747. URL: https://books.google.dk/books?id=2xIwDwAAQBAJ.

[2] Henrik Christensen. *Introduction to Bioinformatics in Microbiology*. eng. Second edition. Learning Materials in Biosciences Series. Cham, Switzerland: Springer, 2023. ISBN: 3-031-45293-3.

[3] Mogens Jakobsen and Lene Jespersen. "Classification of Yeasts". In: *Food Microbiology* 1 (2000). Revised 2012, pp. 1–28.

[4] Jyoti P. Tamang, Koichi Watanabe, and Wilhelm H. Holzapfel. "Review: Diversity of Microorganisms in Global Fermented Foods and Beverages". eng. In: *Frontiers in microbiology* 7 (2016), pp. 377–377. ISSN: 1664-302X.

[5] Benjamin E. Wolfe and Rachel J. Dutton. "Fermented Foods as Experimentally Tractable Microbial Ecosystems". In: *Cell* 161.1 (2015), pp. 49–55. ISSN: 0092-8674. DOI: https://doi.org/10.1016/j.cell.2015.02.034. URL: https://www.sciencedirect.com/science/article/pii/S0092867415002007.

Chapter A
Appendix