

# Decoding Indoor Fungal Microbiomes with DNA based Technology and Bioinformatics

---

Lesson plan created by Carlos M. Rodríguez Minguela

## Title

Decoding Indoor Fungal Microbiomes with DNA-based Technology and Bioinformatics

## Overview

This four-lesson module presents indoor fungal microbiome analysis as a connected storyline, moving from why indoor microbial communities matter to how scientists identify and interpret them using DNA-based tools. Students first explore the relevance of built environments to air quality and health, then examine DNA barcoding, sequencing, database comparison, and bioinformatics as practical methods for studying fungi that are often invisible or difficult to identify by appearance alone. The module supports both conceptual understanding and hands-on data interpretation. Using curated sequence datasets, DNA sequence alignments, and phylogenetic analyses, students identify unknown organisms, recognize genetic relationships, and connect sequence matches to biological meaning. These lessons also place indoor fungal identification within broader questions about microbial interactions, biological control, and emerging tools for understanding healthier built environments. Beyond the classroom, the module connects microbiome science to real workplace needs in biopharma/medical device manufacturing, quality control, and environmental health, where environmental microbial communities must be monitored and understood. Through DNA-based identification and data interpretation, students build awareness of skills that support careers focused on safer products, cleaner facilities, and healthier indoor environments.

## Key Search Words

Indoor fungal microbiome, fungal ITS barcoding, multiple sequence alignment, indoor fungal biocontrols, genetic classification of microbes, phylogenetic trees

## Learning Objectives

- Explain why the indoor microbiome matters by connecting indoor microbial communities to air quality, moisture, mold growth, allergens, and potential health concerns.
- Describe the major steps of DNA barcoding by following the workflow from sampling and DNA extraction to PCR, sequencing, database comparison, and organism identification.
- Explain why the fungal ITS region is useful for identification by describing how conserved regions allow PCR amplification and variable regions help distinguish fungal species.
- Use multiple sequence alignment and phylogenetic trees to identify unknown organisms by comparing unknown barcode sequences with closely related reference sequences.
- Connect sequence-based identification to biological meaning by researching the habitat, function, health relevance, or ecological role of the identified organism.

## Curriculum Alignment

This lesson supports several core learning expectations identified in the American Society for microbiology (ASM) Curriculum Guidelines for Undergraduate Microbiology

<https://asm.org/asm/media/education/asm-curriculum-guidelines.pdf>:

1. **“The evolution of microbes is impacted by their interactions with the environment and a variety of ecological forces, including other microbes, humans, and habitats.”** Students examine how **indoor environments and interactions among microorganisms** influence fungal communities, including relationships involving fungi, other microbes, and mycoviruses.

2. **“Phylogenetic trees best reflect the evolutionary relatedness of all organisms although microbial lineages may be difficult to define due to horizontal gene transfer or lack of conserved genes.”**  
Students use **fungal ITS sequence alignments and phylogenetic trees** to visualize genetic relationships and identify unknown fungal sequences.
3. **“Microbes have unique genomes, structures, and/or biochemical characteristics that distinguish them from each other.”**  
Students compare **fungal ITS barcode sequences** and use sequence variation to distinguish fungal species and investigate their characteristics.
4. **“Most microbial life is currently unculturable and therefore both cultivation-dependent and cultivation-independent techniques are used to identify microbial populations and their potential metabolic pathways.”**  
Students compare **culture-dependent and culture-independent approaches** and learn how DNA-based methods can detect fungi without requiring cultivation.
5. **“Non-cellular infectious agents, such as viruses, prions, viroids, and satellites, are dependent on host cell processes in order to replicate.”**  
Students explore **mycoviruses as intracellular viral agents that depend on fungal host cells** and investigate their potential effects on fungal growth and virulence.
6. **“Microbes are ubiquitous, found in diverse and dynamic ecosystems, where they use available resources and often form complex communities.”**  
Students investigate the **built environment as a microbial ecosystem** and examine the diversity and ecological roles of fungi within indoor microbiomes.
7. **“Humans leverage microbes and their products to address problems and improve quality of life.”**  
Students examine the potential use of **beneficial fungi and mycoviruses as biological-control tools** for managing problematic fungi in built environments.
8. **“The extent of microbial diversity is largely unknown, and exploration of this diversity is critical to understanding microbes and their role in the biosphere.”**  
Students use **DNA barcoding, bioinformatics, and scientific literature** to explore fungal diversity and investigate the ecological roles of fungi associated with indoor environments.
9. **“Apply scientific methods: Investigate microbial systems.”**  
Students investigate the **fungal microbiome of built environments** and examine how fungal species can be identified and compared using molecular and bioinformatics approaches.
10. **“Apply scientific methods: Document and communicate the methods, results, and conclusions.”**  
Students document their analysis workflow and communicate conclusions derived from **DNA barcoding, sequence alignment, and phylogenetic-tree interpretation**.
11. **“Apply scientific methods: Collaborate, give and receive feedback, update method and reassess conclusions.”**  
Students work in **pairs or small groups** to compare results, discuss interpretations, provide feedback, and refine conclusions based on evidence.
12. **“Apply scientific methods: Use quantitative reasoning and computational skills, such as mathematical reasoning, graphing, and statistics to evaluate and interpret data in microbiology.”**  
Students use **computational bioinformatics tools, including MAFFT**, to align fungal ITS sequences and interpret patterns of sequence similarity, variation, and phylogenetic relationships.
13. **“Identify, interpret, and evaluate credible sources of information and cite them appropriately.”**  
Students use **GenBank, PubMed, Google Scholar, and peer-reviewed scientific literature** to research identified fungi, evaluate reliable scientific information, and support their conclusions with appropriate citations.

## Classroom time required

Classroom time required is the amount of time a teacher will need to schedule for implementation of this lesson plan.

### Lecture Content: Lessons 1 and 2

Approximately one hour 15-minute lecture period

#### Main topics:

- Microbiomes and the built environment: 10 minutes
- DNA barcoding and the fungal ITS region: 20 minutes
- Culture-based and culture-independent approaches: 20 minutes
- Overview of sequence alignment and phylogenetic analysis: 25 minutes

### Lab Content: Lessons 3 and 4

Approximately one 2-hour-and-40-minute laboratory period

#### Main activities:

- Review and activity setup: 10 minutes
- Fungal ITS sequence alignment using MAFFT: 10 minutes
- Phylogenetic-tree construction and organization: 15 minutes
- Identification of unknown fungal sequences: 10 minutes
- Literature research and interpretation: 60 minutes
- Worksheet completion and closing discussion: 40 minutes
- Transitions, technical assistance, and troubleshooting: 15 minutes
- 

Alternatively, the individual lessons, activities, or slides can be partially incorporated into introductory biology courses to reinforce foundational concepts, including DNA structure and replication, genetic variation, PCR, evolution, microbial diversity, ecological interactions, and detection/classification/identification of microbes. Instructors may select the components most relevant to their course and adjust the depth, instructional time, and level of complexity based on student preparation and learning objectives.

## Materials & Technology

The instructional materials for this unit consist of PowerPoint presentations and accompanying student worksheets for each of the four lessons (described below). The PowerPoint presentations include instructor notes on every slide to support lesson preparation, facilitate classroom instruction, and provide smooth transitions between topics. The worksheets reinforce the lesson content through guided practice, bioinformatics activities, and critical-thinking exercises. Editable lesson files and PDF versions will be available to both instructors and students.

The activities are designed for students to work collaboratively in pairs or small groups, promoting discussion, problem-solving, and interpretation of bioinformatics data. Each student should have access to a computer with a reliable internet connection to complete the bioinformatics activities. Internet access is required to use the MAFFT multiple sequence alignment platform and to access peer-reviewed scientific literature that supports sequence analysis, phylogenetic interpretation, and organism identification.

### Lesson 1: Building the Foundation for Indoor Fungal Microbiome Analysis

Lesson 1 introduces students to the microbiome as a microbial ecosystem, then narrows the focus to the built environment microbiome, the microbial communities found in homes, classrooms, offices, and other indoor spaces. It begins with distinction between microbiota and microbiome: microbiota refers to the organisms present, while microbiome includes the organisms plus their genes, molecules, interactions, viruses, and surrounding environmental conditions. This foundation helps students understand that microbiomes are not just lists of microbes, but dynamic communities that interact with each other and with their environment.

The section then transitions into the importance of studying the indoor microbiome. Students are encouraged to think about indoor spaces as living microbial environments that can affect the air we breathe and the surfaces we touch. The slides highlight that moisture, poor ventilation, mold growth, bacteria, allergens, and microbial particles can influence indoor air quality and may contribute to irritation, allergies, asthma triggers, odors, or infection risks. This makes the topic immediately relevant to students because it connects microbiology to everyday places where people spend much of their time.

After establishing why indoor microbiomes matter, the material introduces DNA barcoding as a practical tool for identifying microorganisms. The slide explains that a short, standardized DNA region can function like a biological barcode, allowing scientists to distinguish one microbial species from another. This is an effective teaching bridge because it connects a familiar idea, a barcode used to identify a product to a molecular tool used to identify microbes.

The section also introduces two complementary ways to study microbiomes using DNA barcoding. In the culture-based approach, microorganisms are first grown as isolated colonies in the laboratory and then individually barcoded. This approach is useful because it provides living isolates that can be studied further. In the culture-independent approach, DNA is extracted directly from the total sample, allowing scientists to detect both culturable and non-culturable organisms and obtain a broader view of the microbial community.

Together, these slides provide a launch point for a lecture or activity because they move from concept to relevance to method. Students first learn what a microbiome is, then why indoor microbiomes deserve attention, and finally how DNA-based technologies help scientists identify and compare microbes in complex environmental samples. The workflow is developed step by step, beginning with the idea that a DNA barcode works like a biological identification tag: a short, standardized DNA sequence can be compared with reference databases to connect an unknown fungus with known species information.

## **Lesson 2: An overview of the DNA barcoding workflow**

This section guides students through the core logic of DNA barcoding as a practical method for identifying fungi from indoor environments. The slides then expand this idea by explaining how barcoding fits into a broader workflow that includes sampling, PCR amplification, sequencing, database comparison, and interpretation of results.

The material explains why specific genetic regions are used, especially the fungal ITS region, and how conserved and variable regions work together. Conserved regions allow primers to bind and amplify DNA across many fungi, while variable regions provide the sequence differences needed to distinguish one fungal species from another. This gives students a clear molecular reason for why DNA barcoding can be used for fungal identification.

The section also prepares students to understand why bioinformatics is necessary. Once multiple unknown isolates are sequenced and compared to databases, the result is no longer a single identification problem; it becomes a larger and more complex sequence dataset. The slides introduce multiple sequence alignment to organize those sequences so similarities and differences can be compared systematically. This sets up the next concept: aligned sequences can be used to construct phylogenetic trees that summarize genetic relationships among many isolates at once.

For instructors, this sequence of slides provides a smooth bridge from wet-lab concepts to computational biology. Students first learn how a barcode sequence is generated and matched to a database, then they see why alignment and tree-building are useful when many sequences must be interpreted together. The material is especially useful for microbiology, biotechnology, environmental health, and general biology courses because it connects indoor microbiome context with skills students can apply in a bioinformatics activity. Students also follow the progression from fungal sample → DNA barcode → database match → complex dataset → multiple sequence alignment → phylogenetic tree. This makes the later MAFFT activity more meaningful because students understand not only what they are doing, but why alignment and phylogenetic analysis are needed to interpret microbial diversity.

## **Lesson 3: Unknown Microbes, and the Tree of Life: Revealing Genetic Relationships with MAFFT**

This section turns the lesson from concept-building into an active bioinformatics experience. Students move beyond learning what DNA barcoding is and begin using sequence data the way microbiologists do: by aligning sequences, comparing genetic similarities, and building phylogenetic trees to interpret relationships among microorganisms.

The section begins with Bioinformatics Activity 1. In this activity, students will use small subunit RNA barcode gene sequences to explore genetic relationships among representatives of the three domains of life: Bacteria, Archaea, and Eukarya, represented here by fungi.

Students will work with a sequence dataset collected from GenBank and checked using the BLAST search tool. The dataset includes known reference sequences from bacteria, archaea, and fungi, along with an assumed unknown bacterial sequence obtained from a clinical sample.

The sequence dataset will be uploaded into an online bioinformatics pipeline, MAFFT. The pipeline will first build a multiple sequence alignment, which lines up the DNA sequences so students can compare conserved and variable

positions. Then, the aligned sequences will be used to construct a phylogenetic tree, which helps visualize genetic relationships among the organisms.

Students will use the phylogenetic tree to determine which known bacterial sequence is most closely related to the unknown clinical isolate. Based on sequence similarity, they will make a likely identification of the unknown bacterium. After identifying the unknown isolate, students will conduct internet searches of peer-reviewed scientific literature to learn more about the organism, including its biological role, clinical or environmental significance, and why its identification may matter in a real-world context.

The slides also provide a practical walkthrough of the MAFFT workflow: loading sequences, accessing the alignment, viewing the phylogenetic tree, adjusting the appearance of the tree, rerooting it, rotating branches, and organizing clusters for clearer interpretation. This makes the activity approachable for students who may be new to bioinformatics because the process is shown step by step rather than assumed.

A major strength of this section is that it teaches students how to read a phylogenetic tree, not to simply generate one. Students learn that branches represent genetic relationships, branch lengths reflect genetic differences, nodes represent shared ancestry, bootstrap values indicate confidence levels on the observed clustering patterns and branch rotation changes the appearance of the tree without changing the underlying relationships. This helps prevent common student misconceptions about phylogenetic trees.

Completion of this activity will allow students to recreate, in a simplified way, the type of sequence-based approach used by Carl Woese et al. 1990, whose comparisons of small subunit ribosomal RNA sequences transformed Biology by showing that living organisms are organized into three major domains: *Bacteria*, *Archaea*, and *Eukarya*. This discovery was important because it changed how scientists understand the diversity of life and showed that genetic information can reveal deep relationships that are not obvious from appearance alone.

#### **Lesson 4: Indoor Fungi Through the Lens of DNA Barcodes**

This section extends the previous bioinformatics activity into a more focused investigation of the indoor fungal microbiome. Students apply the same general logic introduced earlier: barcode sequences are aligned, compared, and organized into phylogenetic trees. Here, however, the dataset is built around fungal ITS barcode sequences, the standard DNA barcode region commonly used to identify fungi.

The activity uses a curated set of fungal ITS sequences, including several unknown sequences. Students use MAFFT to construct a multiple sequence alignment, which organizes the DNA sequences so that similarities and differences can be compared across many fungi at once. The aligned sequences are then used to generate a phylogenetic tree, allowing students to see which unknown isolates cluster with known fungal species.

A central learning goal is interpretation. Students do not simply produce a tree; they use the tree to infer the likely identity of up to five unknown fungal sequences based on their closest genetic matches. Once the unknowns are grouped with their most related reference sequences, students can investigate what those fungi are known for: whether they are commonly found indoors, associated with environmental health concerns, relevant to medicine or biotechnology, resistant to extreme conditions, or involved in interactions with other fungi.

This section also shows why multiple sequence alignment and phylogenetic trees are useful for complex datasets. A database search may provide individual matches, but when many fungal isolates are analyzed together, the results can become difficult to interpret one sequence at a time. Alignment and tree-building help summarize the dataset visually, grouping related organisms and making broader patterns easier to recognize.

The material then uses *Trichoderma harzianum* as a biological example to connect sequence identification with ecological function. Students see that identifying an unknown sequence as closely related to *Trichoderma* is not just a naming exercise; it suggests a possible role. *Trichoderma* species can attack or suppress other fungi, making them important examples of natural biological control.

This example becomes a stepping stone into the next topic: mycoviruses, viruses that infect fungi. The slides transition from fungi that attack other fungi externally to viral parasites that live inside fungal cells. Students are introduced to the idea that some mycoviruses can reduce fungal growth or virulence without necessarily killing the host fungus. This leads into a broader discussion of whether naturally occurring mycoviruses could eventually be studied as potential biological control tools in built environments.

## Safety

There are no safety concerns associated with this activity. All exercises are computer-based and use online bioinformatics tools and scientific literature; no laboratory work or hazardous materials are involved.

## Teacher Preparation for Activity

No laboratory materials, equipment, or other physical preparation are required, as all activities are computer-based. Preparation for this lesson is provided through the detailed instructor notes included with every slide in each of the four PowerPoint lessons, as well as the accompanying study guides and student worksheets, which are designed to support lesson delivery and student learning.

## Student Preparation for Activity

This lesson plan is intended to supplement and reinforce concepts covered in introductory or general biology courses at the community college learning level rather than serve as a stand-alone introduction to molecular biology or evolution. Before implementing the activity, instructors should ensure that students have prior exposure to the Tree of Life, the three domains of life (Bacteria, Archaea, and Eukarya), the basic chemistry and structure of nucleic acids, and the process of DNA replication. Although designed for introductory biology, this lesson can also be readily adapted for more advanced courses, such as Microbiology, where it can reinforce topics in microbial diversity, molecular identification, phylogenetics, and bioinformatics.

## Procedure

Practical activities are designed to be completed in pairs or small groups, using the corresponding worksheet and study guide provided for each lesson. The explanations that teachers provide during class are fully detailed in the instructor notes included with every slide of the four PowerPoint lessons. Likewise, all bioinformatics procedures are explained step-by-step in both the presentations and the accompanying instructor notes.

## Differentiation

The instructor notes provided for each slide may also be used to improve accessibility for students and audience members with special needs.

## Assessment/Check for Understanding

Each lesson is accompanied by a study guide and worksheet containing practice questions in a variety of formats to reinforce key concepts, assess student understanding, and evaluate students' ability to interpret multiple sequence alignments and phylogenetic analyses.

## Required resources

| Resource Title                           | URL                                                                             | How it is Used in this Lesson Plan                                                                                                                                                                                                                                                                   |
|------------------------------------------|---------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| MAFFT Multiple Sequence Alignment Server | <a href="https://mafft.cbrc.jp/">https://mafft.cbrc.jp/</a>                     | Students use MAFFT to generate multiple sequence alignments and construct phylogenetic trees from bacterial and fungal DNA barcode sequences. The alignments and trees are used to compare conserved and variable regions, infer evolutionary relationships, and identify unknown microorganisms.    |
| Google Scholar                           | <a href="https://scholar.google.com/">https://scholar.google.com/</a>           | Students use Google Scholar to locate peer-reviewed scientific literature related to the microorganisms identified during the bioinformatics activities. These articles help students investigate the biology, ecology, health significance, and potential applications of the identified organisms. |
| PubMed                                   | <a href="https://pubmed.ncbi.nlm.nih.gov/">https://pubmed.ncbi.nlm.nih.gov/</a> | Students use PubMed to search peer-reviewed biomedical literature describing the identified microorganisms. The database is used to investigate topics such as pathogenicity, environmental occurrence, antimicrobial resistance, biological characteristics, and public health relevance.           |

## Author comments

The presentations primarily consist of infographic-style slides that consolidate related content into a single, visually organized format. Guided by the corresponding instructor notes, these slides are designed to be discussed interactively using the magnification tool available in PowerPoint Presenter View, allowing the instructor to focus on specific sections of each infographic as individual concepts are explained. This approach reduces the overall number of slides by minimizing the need for multiple slides that contain only small amounts of information. At the end of the discussion, the key concepts remain consolidated in a comprehensive visual summary, helping students recognize relationships among topics, reinforce what they have learned, and use the infographic to review the material covered in class. Individual panels within the infographics can also be captured using the Windows Snipping Tool and pasted into separate slides, where they can be adapted or expanded into standalone slides. The use of this option is left to the instructor's discretion, depending on instructional goals, available time, and student needs. Some slides also contain active hyperlinks that provide direct access to the online resources, databases, tools, and reference materials discussed during the lessons. These links allow instructors and students to quickly navigate from the presentation to relevant resources for demonstrations, classroom activities, and further exploration.

## Sources

### Online Resources

| Website / Resource                                   | Website Link                                                                                                                                                                                                                    | Description and Relevance to the Lesson Plan                                                                                                                                                                                                           |
|------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| EPA – The Indoor Microbiome                          | <a href="https://www.epa.gov/indoor-air-quality-iaq/indoor-microbiome">https://www.epa.gov/indoor-air-quality-iaq/indoor-microbiome</a>                                                                                         | Introduces the indoor microbiome and explains how microorganisms such as bacteria, fungi, and viruses are associated with built environments. It supports lesson content on indoor microbial communities and factors that influence them.              |
| EPA – Biological Contaminants and Indoor Air Quality | <a href="https://www.epa.gov/indoor-air-quality-iaq/biological-contaminants-and-indoor-air-quality">https://www.epa.gov/indoor-air-quality-iaq/biological-contaminants-and-indoor-air-quality</a>                               | Provides an overview of biological contaminants found indoors, including fungi, bacteria, viruses, allergens, and pests. It connects indoor microorganisms with building conditions, moisture, indoor air quality, and potential effects on occupants. |
| NCBI Nucleotide BLAST (BLASTn)                       | <a href="https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&amp;PAGE_TYPE=BlastSearch&amp;LINK_LOC=blasthome">https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&amp;PAGE_TYPE=BlastSearch&amp;LINK_LOC=blasthome</a> | A bioinformatics tool for comparing a DNA sequence with nucleotide sequences in public databases. It can be used to demonstrate how an unknown fungal ITS barcode is compared with reference sequences to help identify the organism.                  |
| NCBI GenBank                                         | <a href="https://www.ncbi.nlm.nih.gov/genbank/">https://www.ncbi.nlm.nih.gov/genbank/</a>                                                                                                                                       | A public database containing annotated DNA sequences submitted by researchers worldwide. Students can use GenBank to locate reference sequences, examine accession records, and obtain sequences for alignment and phylogenetic analysis.              |
| NC BioNetwork – BioWork                              | <a href="https://www.ncbionetwork.org/biowork">https://www.ncbionetwork.org/biowork</a>                                                                                                                                         | Describes a North Carolina workforce-training program for biotechnology, pharmaceutical, and chemical manufacturing. It helps connect lesson skills to local biotechnology and biopharmaceutical career pathways and training opportunities.           |
| LabXchange                                           | <a href="https://www.labxchange.org/">https://www.labxchange.org/</a>                                                                                                                                                           | A free science-learning platform with virtual laboratory simulations, videos, interactives, and instructional resources                                                                                                                                |

## Scientific Literature

Awad-Allah, E. F. A., Shams, A. H. M., Helaly, A. A., & Ragheb, E. I. M. (2022). Effective Applications of *Trichoderma* spp. as Biofertilizers and Biocontrol Agents Mitigate Tomato Fusarium Wilt Disease. *Agriculture*, 12(11), 1950.

Berg, G., Rybakova, D., Fischer, D., Cernava, T., Champomier Vergès, M.-C., Charles, T., Chen, X., Cocolin, L., Eversole, K., Herrero Corral, G., Kazou, M., Kinkel, L., Lange, L., Lima, N., Loy, A., Macklin, J. A., Maguin, E., Mauchline, T., McClure, R., . . . Schloter, M. (2020). Microbiome definition re-visited: Old concepts and new challenges. *Microbiome*, 8, Article 103.

Druzhinina, I. S., Chenthamara, K., Zhang, J., Atanasova, L., Yang, D., Miao, Y., Rahimi, M. J., Grujic, M., Cai, F., Pourmehdi, S., Abu Salim, K., Pretzer, C., Kopchinskiy, A. G., Henrissat, B., Kuo, A., Hundley, H., Wang, M., Aerts, A., Salamov, A., Lipzen, A., LaButti, K., Barry, K., Grigoriev, I. V., Shen, Q., & Kubicek, C. P. (2018). Massive lateral transfer of genes encoding plant cell wall-degrading enzymes to the mycoparasitic fungus *Trichoderma* from its plant-associated hosts. *PLOS Genetics*, 14(4), e1007322.

Fermin, G., Mazumdar-Leighton, S., & Tennant, P. (2018). Viruses of prokaryotes, protozoa, fungi, and Chromista. In P. Tennant, G. Fermin, & J. E. Foster (Eds.), *Viruses: Molecular biology, host interactions, and applications to biotechnology* (pp. 217–244). Academic Press.

García-Pedrajas, M. D., Cañizares, M. C., Sarmiento-Villamil, J. L., Jacquat, A. G., & Dambolena, J. S. (2019). Mycoviruses in biological control: From basic research to field implementation. *Phytopathology*, 109(11), 1828–1839.

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, 30(4), 772–780.

Lu, X., Dai, Z., Xue, J., Li, W., Ni, P., Xu, J., Zhou, C., & Zhang, W. (2024). Discovery of novel RNA viruses through analysis of fungi-associated next-generation sequencing data. *BMC Genomics*, 25, 517.

Myers, J. M., & James, T. Y. (2022). Mycoviruses. *Current Biology*, 32(4), R150–R155.

Nerva, L., & Chitarra, W. (2021). Mycoviruses: A hidden world within fungi. In Ó. Zaragoza & A. Casadevall (Eds.), *Encyclopedia of mycology* (Vol. 1, pp. 134–141). Elsevier.

Woese, C. R., Kandler, O., & Wheelis, M. L. (1990). Towards a natural system of organisms: Proposal for the domains Archaea, Bacteria, and Eucarya. *Proceedings of the National Academy of Sciences of the United States of America*, 87(12), 4576–4579.

Woolford, J. L., Jr., & Baserga, S. J. (2013). Ribosome biogenesis in the yeast *Saccharomyces cerevisiae*. *Genetics*, 195(3), 643–681.

Xu, J. (2016). Fungal DNA barcoding. *Genome*, 59(11), 913–932.

Yao, X., Guo, H., Zhang, K., Zhao, M., Ruan, J., & Chen, J. (2023). *Trichoderma* and its role in biological control of plant fungal and nematode disease. *Frontiers in Microbiology*, 14, 1160551.

## Study Guide & Worksheet

### Lesson 1: Building the Foundation for Indoor Fungal Microbiome Analysis

#### Glossary

Microbiota: The living microorganisms present in a specific environment.

Microbiome: The microorganisms in an environment plus their genes, molecules, interactions, and surrounding environmental conditions.

Microbial community: A group of microorganisms living together in the same environment.

Built environment: Human-made indoor environments such as homes, schools, and offices.

Indoor microbiome: The community of microorganisms found inside buildings.

DNA barcoding: Identification of organisms using a short standardized DNA sequence.

DNA barcode: The DNA region used for species identification.

ITS sequence: Internal Transcribed Spacer; the standard fungal DNA barcode.

Culture-dependent approach: Identifies microorganisms after growing them in culture.

Culture-independent approach: Identifies microorganisms directly from DNA extracted from a sample without culturing.

Pure culture: A culture containing only one microorganism.

Microbial isolate: A single microorganism separated from a mixed sample.

Cell lysis: Breaking open cells to release DNA.

Total DNA: DNA recovered from all organisms in a sample.

PCR: Polymerase Chain Reaction; amplifies a target DNA region.

PCR thermal cycler: Instrument that performs PCR.

DNA sequencing: Determining the order of nucleotides in DNA.

DNA sequencer: Instrument that reads DNA sequences.

BLAST: Bioinformatics program that compares DNA sequences with database sequences.

GenBank: Public database of DNA sequences maintained by NCBI.

Bioinformatics: Use of computers to analyze biological data.

Environmental sample: Sample collected from soil, water, air, or built environments.

Species identification: Determining the identity of an organism.

#### Key Concepts

Microbiota describes who is present; microbiome includes organisms, genes, activities, and environment.

Indoor microbiomes influence indoor air quality and human health.

Moisture and poor ventilation favor microbial growth.

DNA barcoding identifies microorganisms using standardized DNA regions.

Culture-dependent methods study organisms that grow in culture.

Culture-independent methods detect both culturable and non-culturable organisms.

Typical workflow: sample → DNA extraction → PCR → sequencing → BLAST → GenBank identification.

#### Worksheet

##### A. Multiple Choice

1. A microbiome includes:

- A. Only bacteria
- B. Organisms only
- C. Organisms, genes, interactions, and environment
- D. Only fungi

2. DNA barcoding identifies organisms using:

- A. Proteins
- B. Lipids
- C. A short DNA sequence
- D. Cell shape

3. Which approach detects non-culturable microbes?

- A. Culture-dependent
- B. Culture-independent
- C. Microscopy
- D. Staining

4. BLAST is used to:

- A. Amplify DNA
- B. Compare DNA sequences
- C. Culture fungi
- D. Extract DNA

5. GenBank is a:

- A. PCR machine
- B. DNA database
- C. Microscope
- D. Culture medium

#### **B. Matching**

Match: BLAST, GenBank, PCR, Cell lysis, Pure culture

Definitions: DNA amplification; Public DNA database; Break open cells; DNA comparison program; Single-species culture.

#### **C. Fill in the Blank**

The fungal DNA barcode is the \_\_\_\_\_ region.

PCR stands for \_\_\_\_\_.

Cell \_\_\_\_\_ releases DNA from cells.

GenBank stores DNA \_\_\_\_\_.

Microbiota refers to the living \_\_\_\_\_ in an environment.

#### **D. Sequence**

Order the culture-independent workflow:

PCR amplification; DNA sequencing; Cell lysis; DNA extraction; BLAST/GenBank comparison; Species identification.

**Instructor Key**

Multiple Choice : 1-C, 2-C, 3-B, 4-B, 5-B

Matching: BLAST–DNA comparison; GenBank–public DNA database; PCR–DNA amplification; Cell lysis–break open cells; Pure culture–single-species culture.

Fill: ITS; Polymerase Chain Reaction; lysis; sequences; microorganisms.

Sequence: Cell lysis → DNA extraction → PCR amplification → DNA sequencing → BLAST/GenBank comparison → Species identification.

## Lesson 2 Study Guide & Worksheet

### DNA Barcoding Workflow

Based on the uploaded lesson presentation.

#### Glossary

1. **Deoxyribonucleic acid (DNA):** The hereditary molecule that stores genetic information.
2. **Deoxyribonucleic acid (DNA) barcoding:** A method that identifies organisms using a short standardized region of deoxyribonucleic acid.
3. **Barcode region:** The specific deoxyribonucleic acid region used for identification.
4. **Internal Transcribed Spacer:** The most widely used fungal barcode region. The Internal Transcribed Spacer region helps make ribosomes, the cell's protein-making machines.
5. **Primer:** A short deoxyribonucleic acid sequence that binds to a conserved region and starts amplification.
6. **Polymerase Chain Reaction:** A laboratory method that makes millions of copies of a target deoxyribonucleic acid region.
7. **Denaturation:** Step in which the two deoxyribonucleic acid strands separate.
8. **Annealing:** Step in which primers bind to the target sequence.
9. **Extension:** Step in which new deoxyribonucleic acid strands are synthesized.
10. **Conserved region:** A region that is very similar among species.
11. **Variable region:** A region that differs among species and helps distinguish them.
12. **Cell lysis:** Breaking cells open to release deoxyribonucleic acid.
13. **Deoxyribonucleic acid extraction:** Isolation of deoxyribonucleic acid from cells.
14. **Sequencing:** Determining the nucleotide order of deoxyribonucleic acid.
15. **Chromatogram:** Graphical output generated during sequencing.
16. **Reference database:** Collection of known sequences used for comparison.
17. **GenBank:** Public nucleotide sequence database.
18. **Basic Local Alignment Search Tool:** Program that compares a sequence against reference databases.
19. **Multiple sequence alignment:** Arrangement of sequences so similar positions line up.
20. **Phylogenetic tree:** Diagram showing inferred evolutionary relationships.

#### Key Concepts

Extract deoxyribonucleic acid from the sample.

Amplify the barcode region using Polymerase Chain Reaction.

Sequence the amplified barcode region.

Compare the sequence with reference databases to identify the organism.

Internal Transcribed Spacer is the preferred fungal barcode because conserved regions allow primer binding while variable regions distinguish species.

Multiple sequence alignments reveal conserved and variable positions and support phylogenetic tree construction.

#### Worksheet

##### Multiple Choice

1. Which fungal barcode region is most widely used for species identification?
  - A. 18S ribosomal ribonucleic acid
  - B. Internal Transcribed Spacer
  - C. 28S ribosomal ribonucleic acid
  - D. Transfer ribonucleic acid
2. What is the purpose of Polymerase Chain Reaction?
  - A. Sequence deoxyribonucleic acid
  - B. Amplify the barcode region
  - C. Compare databases

D. Culture fungi

3. Why are conserved regions important?

A. They vary greatly

B. They allow primer binding

C. They identify mutations only

D. They replace sequencing

4. Which tool compares an unknown sequence to reference sequences?

A. GenBank

B. Basic Local Alignment Search Tool

C. Microscope

D. Incubator

5. What can be built after multiple sequence alignment?

A. Culture plate

B. Phylogenetic tree

C. Primer

D. Chromosome

### Matching

**Match each term with its definition.**

---

1. Internal Transcribed Spacer

2. Primer

3. Polymerase Chain Reaction

4. Conserved region

5. Variable region

6. GenBank

Definitions:

A. Short sequence that initiates amplification

B. Public sequence database

C. Most common fungal barcode

D. Similar among species

E. Different among species

F. Amplifies target deoxyribonucleic acid

Fill in the Blank

1. The \_\_\_\_\_ is the standard fungal barcode.

2. Polymerase Chain Reaction produces millions of copies of the barcode \_\_\_\_\_.

3. Primers bind to \_\_\_\_\_ regions.

4. Variable regions help distinguish different \_\_\_\_\_.

5. The sequencing instrument produces a \_\_\_\_\_.

### Sequence of Events

Place the DNA barcoding workflow in the correct order:

1. Compare with reference database

2. Extract deoxyribonucleic acid

3. Sequence the barcode

4. Amplify by Polymerase Chain Reaction

5. Identify the organism

## Instructor Key

Multiple Choice: 1-B, 2-B, 3-B, 4-B, 5-B.

Matching: Internal Transcribed Spacer—Most common fungal barcode; Primer—Short sequence that initiates amplification; Polymerase Chain Reaction—Amplifies target deoxyribonucleic acid; Conserved region—Similar among species; Variable region—Different among species; GenBank—Public sequence database.

Fill in the blanks: Internal Transcribed Spacer; region; conserved; species; chromatogram.

Correct workflow: Extract deoxyribonucleic acid → Amplify by Polymerase Chain Reaction → Sequence the barcode → Compare with reference database → Identify the organism.

# Lesson 3 Study Guide and Worksheet

## Unknown Microbes and the Tree of Life: Small-Subunit Ribosomal Ribonucleic Acid Alignment and Phylogeny

Name: \_\_\_\_\_ Date: \_\_\_\_\_ Section: \_\_\_\_\_

### Part I. Glossary

Students should be able to define and apply the following terms after completing the lesson.

- 1. Small-subunit ribosomal ribonucleic acid:** A ribosomal ribonucleic acid molecule that forms part of the small ribosomal subunit and helps the ribosome read messenger ribonucleic acid during protein production.
- 2. Small-subunit ribosomal ribonucleic acid gene:** The deoxyribonucleic acid sequence that encodes small-subunit ribosomal ribonucleic acid; it is widely used to compare organisms because it contains both conserved and variable regions.
- 3. Deoxyribonucleic acid barcode:** A selected deoxyribonucleic acid sequence used to identify an organism or compare it with other organisms.
- 4. Sequence:** The order of nucleotide bases in deoxyribonucleic acid or ribonucleic acid.
- 5. FASTA format:** A plain-text sequence format in which each sequence begins with a description line starting with a greater-than symbol, followed by the nucleotide sequence.
- 6. Multiple sequence alignment:** An arrangement of three or more biological sequences so that corresponding nucleotide positions are placed in the same columns for comparison.
- 7. MAFFT:** A computer program used to create multiple sequence alignments of nucleotide or amino acid sequences.
- 8. Conserved position:** An aligned position containing the same or a very similar nucleotide in most sequences, often indicating an important or slowly changing region.
- 9. Variable position:** An aligned position where nucleotides differ among sequences, providing information that can help distinguish organisms.
- 10. Gap:** A dash inserted during alignment to represent a possible insertion or deletion and to improve positional comparison among sequences.
- 11. Sequence similarity:** The degree to which two biological sequences contain the same nucleotides at corresponding positions.
- 12. Phylogeny:** The inferred evolutionary history and relationships among organisms or sequences.
- 13. Phylogenetic tree:** A branching diagram that represents inferred genetic or evolutionary relationships among organisms or sequences.
- 14. Taxon:** An organism, species, strain, or other biological group represented in a phylogenetic analysis.
- 15. Branch:** A line in a phylogenetic tree representing an evolutionary path.
- 16. Branch length:** A measure of the amount of genetic change represented by a branch; longer branches generally indicate greater sequence difference when the tree is scaled.
- 17. Node:** A branching point in a phylogenetic tree that represents a shared common ancestor.
- 18. Cluster:** A group of sequences positioned near one another in a phylogenetic tree because they are genetically similar.
- 19. Tree topology:** The branching pattern of a phylogenetic tree, independent of how the branches are visually rotated or arranged.
- 20. Root:** The point that represents the ancestral direction of a phylogenetic tree.
- 21. Rerooting:** Changing the displayed root position of a tree using prior biological knowledge or an appropriate comparison group.
- 22. Bootstrap value:** A numerical estimate of how consistently a particular grouping appears when the dataset is repeatedly resampled; values close to one hundred indicate strong support.
- 23. Three domains of life:** The broad evolutionary groups Bacteria, Archaea, and Eukarya.
- 24. Unknown bacterial isolate:** A bacterial sample whose identity has not yet been established and must be inferred from sequence and contextual evidence.
- 25. Peer-reviewed article:** A scientific paper evaluated by qualified experts before publication.
- 26. Primary research article:** A peer-reviewed article that reports original methods, results, and conclusions from a scientific study.

### Part II. Study Guide

- 1. Why this gene is useful:** Small-subunit ribosomal ribonucleic acid genes occur across cellular life. Conserved regions support broad comparisons among organisms, while variable regions help distinguish different groups and closely related organisms.
- 2. Main analysis workflow:** Prepare or obtain sequences in FASTA format, upload them to MAFFT, generate the multiple sequence alignment, inspect conserved and variable positions, construct a phylogenetic tree, improve the tree display, and interpret the clustering pattern.
- 3. Reading an alignment:** Nucleotides placed in the same column are treated as corresponding positions. Conserved columns show similarity, variable columns show differences, and gaps indicate possible insertions or deletions introduced to improve alignment.
- 4. Reading a phylogenetic tree:** The branching pattern shows which sequences are grouped most closely. Branch lengths indicate genetic difference when the tree is scaled. Nodes represent shared ancestors, and bootstrap values estimate support for specific groupings.

**5. Tree appearance versus relationships:** Rotating branches at a node changes only the visual arrangement. It does not change the topology or the inferred relationships. Rerooting changes the displayed ancestral direction and should be guided by accepted biological knowledge.

**6. Identifying the unknown isolate:** The unknown sequence should be evaluated using several lines of evidence: its nearest neighbors in the tree, branch length, bootstrap support, direct sequence-search results, and evidence from peer-reviewed literature.

**7. Limits of interpretation:** A close match for one barcode region supports identification but may not always prove species identity. Sequence quality, database accuracy, gene resolution, contamination, and incomplete reference sampling can affect conclusions.

### Part III. Practice Questions

#### A. Multiple Choice

- Why are small-subunit ribosomal ribonucleic acid genes useful for broad comparisons among organisms?
  - They are found only in bacteria.
  - They contain both conserved and variable regions.
  - They change completely in every generation.
  - They directly encode all cellular proteins.
- What is the main purpose of a multiple sequence alignment?
  - To grow microorganisms in pure culture.
  - To calculate the mass of a ribosome.
  - To place corresponding sequence positions in the same columns.
  - To remove all variable nucleotides.
- What does a variable position in an alignment provide?
  - Evidence that all sequences are identical.
  - Information that can help distinguish organisms.
  - Proof that the sequence is contaminated.
  - A measurement of cell size.
- What does a node represent in a phylogenetic tree?
  - A shared common ancestor.
  - A laboratory instrument.
  - A nucleotide deletion only.
  - A database record number.
- What does a bootstrap value close to one hundred generally indicate?
  - Weak support for a grouping.
  - Strong support for a grouping.
  - A sequence length of one hundred nucleotides.
  - One hundred species were sampled.
- What happens when branches are rotated around an internal node?
  - The underlying relationships change.
  - The sequences are edited.
  - Only the visual arrangement changes.
  - The root is automatically proven.
- Which sequence would usually be considered most closely related to an unknown sequence?
  - The sequence that shares the nearest well-supported node and shortest genetic distance.
  - The sequence with the longest name.
  - The sequence located highest on the page.
  - The sequence from the largest organism.
- Why should identification of an unknown isolate include peer-reviewed literature?
  - To replace sequence analysis completely.
  - To understand the organism's biology, health relevance, and limitations of the identification.
  - To avoid reporting evidence.
  - To ensure every source gives the same conclusion.
- Which statement best describes rerooting?
  - It changes the nucleotide sequence.
  - It changes the displayed ancestral direction of the tree.
  - It removes all unsupported branches.
  - It guarantees species identification.
- What is the most appropriate conclusion from a very close barcode match?
  - The organism is certainly identical in every biological trait.
  - The sequence strongly supports a close relationship, but additional evidence may be needed for definitive identification.
  - The database sequence must be incorrect.
  - The organism belongs to a different domain of life.

#### B. Matching

Match each numbered term with the best lettered description. Each letter is used once.

- Multiple sequence alignment

2. Conserved position
3. Variable position
4. Node
5. Branch length
6. Bootstrap value
7. Rerooting
8. Tree topology
9. MAFFT
10. Peer-reviewed article

- A. A measure of genetic difference represented along a branch.
- B. A scientific publication evaluated by experts before publication.
- C. A program used to align biological sequences.
- D. The branching pattern that describes relationships in a tree.
- E. A position that differs among aligned sequences.
- F. A branching point representing a shared common ancestor.
- G. A numerical estimate of support for a grouping.
- H. The process of placing corresponding sequence positions into shared columns.
- I. A position that is the same or very similar among sequences.
- J. Changing the displayed ancestral direction of a tree.

### C. Fill in the Blank

1. A sequence file in which each entry begins with a greater-than symbol is written in \_\_\_\_\_ format.
2. MAFFT is used to generate a \_\_\_\_\_ sequence alignment.
3. Positions that are similar across many organisms are called \_\_\_\_\_ positions.
4. Positions that differ among organisms are called \_\_\_\_\_ positions.
5. A branching diagram that represents genetic relationships is a \_\_\_\_\_ tree.
6. A branching point in a tree is called a \_\_\_\_\_.
7. Numbers that estimate support for groupings are called \_\_\_\_\_ values.
8. Rotating branches changes the appearance of a tree but not its \_\_\_\_\_.
9. Bacteria, Archaea, and Eukarya are the three \_\_\_\_\_ of life.
10. Scientific papers evaluated by experts are described as \_\_\_\_\_ reviewed.

### D. Sequence of Events

Number the following steps from 1 through 8 to show the correct workflow for the activity.

- \_\_\_\_\_ Inspect the alignment for conserved positions, variable positions, and gaps.
- \_\_\_\_\_ Interpret the unknown isolate using its nearest neighbors and branch lengths.
- \_\_\_\_\_ Upload the FASTA sequence file to MAFFT.
- \_\_\_\_\_ Review peer-reviewed literature about the most likely organism.
- \_\_\_\_\_ Generate and open the phylogenetic tree.
- \_\_\_\_\_ Submit the sequences for alignment.
- \_\_\_\_\_ Reroot and rotate branches as needed to improve biological interpretation and visualization.
- \_\_\_\_\_ Record a final identification statement with supporting evidence and limitations.

## Part IV. Systematic Identification of the Unknown Bacterial Sequence

Complete each step in order. Use the sequence alignment, phylogenetic tree, a reputable nucleotide sequence database, and peer-reviewed scientific literature. Record enough detail that another student could understand and evaluate your conclusion.

### Step 1. Initial Tree-Based Observation

1. What label is used for the unknown sequence in the dataset?  
\_\_\_\_\_
2. Which known sequence or sequences are positioned closest to the unknown isolate?  
\_\_\_\_\_
3. What is the bootstrap value supporting the nearest grouping?  
\_\_\_\_\_
4. Are the connecting branches short, long, or not visibly different? What does this suggest?  
\_\_\_\_\_
5. Based only on the tree, what genus is the unknown isolate most likely to belong to?  
\_\_\_\_\_

## Step 2. Search Peer-Reviewed Scientific Literature

Find at least two peer-reviewed articles. At least one should address health or clinical relevance, and at least one should address biology, ecology, transmission, resistance, or another important characteristic. Prefer primary research articles when possible.

| Article      | Complete citation | Research question or purpose | Key finding relevant to the unknown | Why this source is credible |
|--------------|-------------------|------------------------------|-------------------------------------|-----------------------------|
| 1            |                   |                              |                                     |                             |
| 2            |                   |                              |                                     |                             |
| 3 (optional) |                   |                              |                                     |                             |

## Step 4. Biological and Health Significance

9. Where is this bacterium commonly found?

---

10. Is it normally harmless, beneficial, opportunistic, or pathogenic? Explain using evidence.

---

11. What infections or health effects, if any, are associated with it?

---

12. Which populations or conditions increase the risk of disease?

---

13. Does the literature describe antibiotic resistance, biofilm formation, toxin production, transmission, or persistence? Summarize the most relevant feature.

---

14. Why might identifying this organism matter in a patient, environmental, or public-health setting?

---

## Step 5. Final Evidence-Based Conclusion

Write one well-supported paragraph that includes: the most likely identity of the unknown isolate; the tree and sequence-search evidence; an important biological or health implication; at least one limitation; and one appropriate next step for confirmation or further study.

---

---

---

---

---

---

## Part V. Instructor Answer Key and Guidance

### A. Multiple Choice Key

1. B 2. C 3. B 4. A 5. B 6. C 7. A 8. B 9. B 10. B

### B. Matching Key

1. H  
2. I  
3. E  
4. F  
5. A  
6. G  
7. J  
8. D  
9. C  
10. B

### C. Fill-in-the-Blank Key

1. FASTA  
2. multiple  
3. conserved  
4. variable  
5. phylogenetic  
6. node  
7. bootstrap  
8. topology  
9. domains  
10. peer

### D. Sequence-of-Events Key

1. Upload the FASTA sequence file to MAFFT.
2. Submit the sequences for alignment.
3. Inspect the alignment for conserved positions, variable positions, and gaps.
4. Generate and open the phylogenetic tree.
5. Reroot and rotate branches as needed to improve biological interpretation and visualization.
6. Interpret the unknown isolate using its nearest neighbors and branch lengths.
7. Review peer-reviewed literature about the most likely organism.
8. Record a final identification statement with supporting evidence and limitations.

### E. Expected Interpretation of the Unknown Isolate

Tree-based interpretation: In the lesson tree, the unknown patient isolate clusters within the *Staphylococcus* group and is positioned most closely with the sequence labeled *Staphylococcus aureus*, with strong bootstrap support. The appropriate preliminary conclusion is that the isolate is most consistent with *Staphylococcus aureus* or a very closely related *Staphylococcus* sequence.

Important caution: Students should not claim that a single small-subunit ribosomal ribonucleic acid sequence proves strain identity, antibiotic resistance, virulence, or clinical outcome. These characteristics require additional evidence.

Acceptable confirmation evidence: A strong response records the top database matches, percent identity, query coverage, accession numbers, and whether several closely related species have nearly identical results.

Health-relevance research: Students may find that *Staphylococcus aureus* can colonize human skin and nasal passages and can also cause disease, especially when barriers are disrupted or risk factors are present. Claims about methicillin resistance must be supported by resistance testing or resistance-gene evidence and should not be inferred from the ribosomal sequence alone.

Suggested additional tests: species-specific molecular testing, whole-genome sequencing, biochemical identification, matrix-assisted laser desorption ionization time-of-flight mass spectrometry, antimicrobial susceptibility testing, or detection of relevant resistance and virulence genes.

### F. Research Section Scoring Guide

| Criterion         | Strong                                                                                                                    | Developing                                               | Needs revision                          |
|-------------------|---------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------|-----------------------------------------|
| Sequence evidence | Uses tree placement, bootstrap support, sequence-search results, accession numbers, percent identity, and query coverage. | Uses some sequence evidence but omits important details. | Provides an unsupported identification. |

| Criterion                       | Strong                                                                                          | Developing                                                     | Needs revision                                               |
|---------------------------------|-------------------------------------------------------------------------------------------------|----------------------------------------------------------------|--------------------------------------------------------------|
| Scientific literature           | Uses at least two relevant peer-reviewed sources and accurately summarizes findings.            | Uses limited or partly relevant sources.                       | Uses non-scientific sources or misrepresents evidence.       |
| Health and biological relevance | Explains habitat, health significance, risk factors, and another important trait with evidence. | Provides a general explanation with limited evidence.          | Makes unsupported or inaccurate claims.                      |
| Limitations and next steps      | Clearly identifies limitations and proposes an appropriate confirmation step.                   | Mentions a limitation or next step but not both.               | Treats the identification as certain and gives no next step. |
| Final conclusion                | Integrates evidence, significance, uncertainty, and next step in a clear paragraph.             | Conclusion is partly complete or weakly connected to evidence. | Conclusion is unclear or unsupported.                        |

### G. Instructor Notes for Evaluating Sources

1. Accept sources from established scientific journals, university library databases, PubMed, or Google Scholar when the article itself is peer reviewed.
2. Require students to distinguish primary research articles from review articles and general websites.
3. Encourage students to verify that the organism discussed in the article matches the identification level supported by the sequence evidence.
4. Do not award full credit for claims about antibiotic resistance, virulence, or patient outcome unless the cited article and the student's evidence directly support the claim.
5. Accept different well-reasoned conclusions when students clearly explain database ambiguity or limited species-level resolution.

# Lesson 4 Study Guide and Worksheet

## Indoor Fungi Through the Lens of DNA Barcodes

---

### Part I. Glossary

Students should be able to define and apply the following terms after completing the lesson.

- 1. Internal Transcribed Spacer region:** A non-coding region within the fungal ribosomal ribonucleic acid gene cluster that is widely used as a fungal deoxyribonucleic acid barcode because it contains conserved and variable parts.
- 2. Fungal deoxyribonucleic acid barcode:** A selected fungal deoxyribonucleic acid sequence used to identify or compare fungal organisms.
- 3. FASTA format:** A plain-text sequence format in which each sequence begins with an identifier line starting with a greater-than symbol, followed by the nucleotide sequence.
- 4. Multiple sequence alignment:** An arrangement of three or more sequences so corresponding nucleotide positions appear in the same columns for comparison.
- 5. MAFFT:** A computer program used to align nucleotide or amino acid sequences and generate a multiple sequence alignment.
- 6. Conserved position:** An aligned position containing the same or a very similar nucleotide in most sequences.
- 7. Variable position:** An aligned position where nucleotides differ among sequences and may help distinguish fungal groups or species.
- 8. Gap:** A dash inserted during alignment to represent a possible insertion or deletion and improve positional comparison.
- 9. Phylogenetic tree:** A branching diagram that represents inferred genetic relationships among organisms or sequences.
- 10. Branch:** A line in a phylogenetic tree representing an evolutionary path.
- 11. Branch length:** A measure of genetic difference along a branch when the tree is scaled; longer branches generally indicate greater sequence difference.
- 12. Node:** A branching point representing a shared common ancestor.
- 13. Cluster:** A group of sequences placed near one another because they are genetically similar.
- 14. Tree topology:** The branching pattern of a phylogenetic tree, independent of how branches are visually rotated.
- 15. Root:** The point that represents the ancestral direction of a phylogenetic tree.
- 16. Rerooting:** Changing the displayed root position using biological knowledge or an appropriate outgroup.
- 17. Bootstrap value:** A numerical estimate of how consistently a grouping appears when the dataset is repeatedly resampled; values near one hundred indicate strong support.
- 18. Outgroup:** A sequence known to be outside the main group of interest and used to help root and interpret a phylogenetic tree.
- 19. Unknown fungal isolate:** A fungal sample whose identity has not yet been established and must be inferred from sequence, tree, database, and literature evidence.
- 20. Mycoparasitism:** A biological interaction in which one fungus attacks or obtains resources from another fungus.
- 21. Biological control:** Use of a living organism or biological agent to reduce the growth or damage caused by another organism.
- 22. Mycovirus:** A virus that infects fungi and usually remains inside fungal cells.
- 23. Horizontal transmission:** Transfer of a mycovirus between fungal individuals, commonly through fusion of compatible hyphae.
- 24. Vertical transmission:** Transfer of a mycovirus from a fungus into its spores and then into a new colony.
- 25. Hyphal fusion:** Joining of fungal hyphae that allows cytoplasmic material, including some mycoviruses, to pass between cells.
- 26. Vegetative incompatibility:** A fungal recognition system that can prevent stable fusion between genetically incompatible hyphae and limit mycovirus transmission.
- 27. Hypovirulence:** Reduced ability of a pathogen to grow aggressively or cause disease.
- 28. Peer-reviewed article:** A scientific paper evaluated by qualified experts before publication.
- 29. Primary research article:** A peer-reviewed article reporting original methods, results, and conclusions from a scientific study.
- 30. Environmental monitoring:** Systematic sampling and testing of air, surfaces, water, and other conditions to detect contamination and protect product quality or health.

### Part II. Study Guide

- 1. Purpose of the activity:** Use fungal Internal Transcribed Spacer barcode sequences to compare indoor fungi, identify unknown isolates, and interpret genetic relationships.
- 2. Main workflow:** Upload the provided FASTA file to MAFFT, create the alignment, inspect conserved and variable positions, construct a tree, use the yeast sequences as an outgroup, improve the tree layout, and interpret the unknowns.
- 3. Reading the alignment:** Matching nucleotides in the same column indicate similarity; differences indicate variation; gaps represent possible insertions or deletions introduced to improve alignment.
- 4. Reading the tree:** Nearest neighbors, short branch lengths, and strong bootstrap support provide evidence of close genetic relationships. Tree position alone does not prove every biological trait.

- 5. Unknown sequence interpretation:** Each unknown should be evaluated using tree placement, branch length, bootstrap support, database searches, and peer-reviewed literature.
- 6. Fungal interactions:** *Trichoderma harzianum* can suppress other fungi by competition, antifungal enzymes, and direct attack on fungal hyphae.
- 7. Mycovirus transmission:** Mycoviruses can spread through connected hyphae within a colony, horizontally through hyphal fusion, and vertically through spores. Vegetative incompatibility can limit spread.
- 8. Biocontrol example:** *Cryphonectria hypovirus 1* reduces the virulence of *Cryphonectria parasitica* and has been used to help control chestnut blight in Europe.
- 9. Sequencing and discovery:** Deoxyribonucleic acid barcoding identifies the fungal host, while ribonucleic acid sequencing can reveal mycoviruses living inside the fungus.
- 10. Career connection:** Environmental monitoring and microbial identification are important in cleanrooms, laboratories, pharmaceutical manufacturing, biotechnology, and other regulated facilities.

### Part III. Practice Questions

#### A. Multiple Choice

- Why is the Internal Transcribed Spacer region useful for identifying fungi?
  - It is found only in viruses.
  - It contains conserved and variable regions.
  - It directly produces fungal toxins.
  - It is identical in every fungus.
- What is the purpose of a multiple sequence alignment?
  - To grow fungi on culture media.
  - To place corresponding nucleotide positions in the same columns.
  - To remove all sequence variation.
  - To measure colony diameter.
- What does an outgroup help researchers do?
  - Identify fungal pigments.
  - Root and interpret the tree.
  - Eliminate gaps from the alignment.
  - Increase sequence length.
- Which evidence best supports a close relationship between an unknown and a known sequence?
  - They appear near each other on the page.
  - They share a near node, short branches, and strong bootstrap support.
  - Their names have the same number of letters.
  - They were collected on the same day.
- What is mycoparasitism?
  - A fungus attacking another fungus.
  - A virus infecting a bacterium.
  - A bacterium producing antibiotics.
  - A plant making fungal spores.
- How do many mycoviruses spread horizontally?
  - Through airborne virus particles released after cell lysis.
  - Through fusion of compatible fungal hyphae.
  - Through bacterial conjugation.
  - Through sunlight exposure.
- What can vegetative incompatibility do?
  - Increase fungal barcode length.
  - Limit mycovirus movement between fungi.
  - Convert ribonucleic acid into protein.
  - Guarantee biological control.
- What does hypovirulence mean?
  - Greater ability to cause disease.
  - Reduced growth or disease-causing ability.
  - Complete death of the host fungus.
  - Loss of all fungal chromosomes.
- Why should students examine peer-reviewed articles after identifying an unknown?
  - To understand biological importance, health relevance, and limitations.
  - To replace sequence analysis.

- C. To prove all database records are correct.
  - D. To avoid comparing multiple sources.
10. Which statement best describes the role of environmental monitoring in biomanufacturing?
- A. It helps detect contamination and protect product quality.
  - B. It replaces all cleaning procedures.
  - C. It is used only outdoors.
  - D. It identifies products by color.
11. Which type of sequencing is most useful for detecting hidden ribonucleic acid mycoviruses?
- A. Ribonucleic acid sequencing.
  - B. Protein sequencing only.
  - C. Microscopy alone.
  - D. Colony counting only.
12. Which conclusion is most scientifically appropriate from a close Internal Transcribed Spacer match?
- A. The isolate has every trait of the reference strain.
  - B. The match supports identification, but additional evidence may be needed.
  - C. The unknown must be a virus.
  - D. No literature search is needed.

### B. Matching

Match each numbered term with the best lettered description. Each letter is used once.

- 1. Internal Transcribed Spacer region
  - 2. MAFFT
  - 3. Outgroup
  - 4. Bootstrap value
  - 5. Mycoparasitism
  - 6. Mycovirus
  - 7. Horizontal transmission
  - 8. Vertical transmission
  - 9. Vegetative incompatibility
  - 10. Hypovirulence
  - 11. Biological control
  - 12. Environmental monitoring
- A. Reduced ability of a pathogen to grow aggressively or cause disease.
  - B. A program used to align biological sequences.
  - C. Use of a living organism or biological agent to reduce another organism.
  - D. A virus that infects fungi.
  - E. A fungal barcode region with conserved and variable parts.
  - F. Transfer of a mycovirus through spores to a new colony.
  - G. A comparison sequence used to help root a tree.
  - H. A fungal system that can block fusion and restrict virus spread.
  - I. A fungus attacking another fungus.
  - J. A numerical estimate of support for a tree grouping.
  - K. Transfer of a mycovirus between fungal individuals through hyphal fusion.
  - L. Testing air, surfaces, water, or other conditions to detect contamination.

### C. Fill in the Blank

- 1. A FASTA entry begins with an identifier line starting with a \_\_\_\_\_ symbol.
- 2. The most widely used fungal barcode is the \_\_\_\_\_ Transcribed Spacer region.
- 3. MAFFT produces a multiple sequence \_\_\_\_\_.
- 4. A yeast sequence used outside the main group can serve as an \_\_\_\_\_.
- 5. Numbers near tree nodes that estimate support are called \_\_\_\_\_ values.
- 6. A virus that infects fungi is called a \_\_\_\_\_.
- 7. Transfer of a mycovirus through spores is called \_\_\_\_\_ transmission.
- 8. Fusion between fungal hyphae is called hyphal \_\_\_\_\_.
- 9. Reduced disease-causing ability is called \_\_\_\_\_.
- 10. Scientific articles evaluated by experts are described as \_\_\_\_\_ reviewed.
- 11. Deoxyribonucleic acid barcoding identifies the fungal \_\_\_\_\_, while ribonucleic acid sequencing can reveal viruses inside it.
- 12. Sampling air and surfaces in a cleanroom is an example of environmental \_\_\_\_\_.

D. Sequence of Events

Number each set of steps in the correct order.

1. Fungal barcode analysis workflow

- \_\_\_\_\_ Interpret the unknown using tree, database, and literature evidence.
- \_\_\_\_\_ Upload the FASTA file to MAFFT.
- \_\_\_\_\_ Inspect conserved positions, variable positions, and gaps.
- \_\_\_\_\_ Generate and open the phylogenetic tree.
- \_\_\_\_\_ Submit the sequences for alignment.
- \_\_\_\_\_ Reroot and rotate branches to improve interpretation and visualization.

2. Mycovirus-based biological control workflow

- \_\_\_\_\_ Test whether the infected fungus shows reduced growth or virulence.
- \_\_\_\_\_ Identify a mycovirus in a fungal host.
- \_\_\_\_\_ Introduce or transfer the mycovirus into a target fungus.
- \_\_\_\_\_ Confirm that the mycovirus becomes established.
- \_\_\_\_\_ Compare the infected fungus with an uninfected control.

Part IV. Systematic Identification and Research of the Unknown Fungal Sequences

Complete the following steps for Unknown Isolates 1 through 5. Use the alignment, phylogenetic tree, a reputable nucleotide sequence database, and peer-reviewed scientific literature. The tree suggests the following preliminary nearest matches: Unknown 1 near *Stachybotrys chartarum*; Unknown 2 near *Trichoderma harzianum*; Unknown 3 near *Aspergillus flavus*; Unknown 4 near *Penicillium chrysogenum*; Unknown 5 near *Cladosporium sphaerospermum*. Treat these as hypotheses to evaluate, not final answers.

Step 1. Tree-Based Evidence

| Unknown | Nearest known sequence | Bootstrap support | Branch length | Likely genus | Preliminary species | Confidence and reason |
|---------|------------------------|-------------------|---------------|--------------|---------------------|-----------------------|
| 1       |                        |                   |               |              |                     |                       |
| 2       |                        |                   |               |              |                     |                       |
| 3       |                        |                   |               |              |                     |                       |
| 4       |                        |                   |               |              |                     |                       |
| 5       |                        |                   |               |              |                     |                       |

1. Which unknown has the strongest tree-based support for a species-level identification? Explain.

Step 3. Peer-Reviewed Literature Investigation

For each unknown, find two peer-reviewed articles. One article must address the assigned context, and the second should address biology, ecology, health, biotechnology, resistance, environmental behavior, or another important characteristic. Prefer primary research articles when possible.

| Unknown and context                             | Article citation | Research purpose | Key finding | Why it matters |
|-------------------------------------------------|------------------|------------------|-------------|----------------|
| 1: Environmental health and indoor microbiology |                  |                  |             |                |
| Second article                                  |                  |                  |             |                |
| 2: Fungal biological control                    |                  |                  |             |                |
| Second article                                  |                  |                  |             |                |
| 3: Environmental health and indoor microbiology |                  |                  |             |                |
| Second article                                  |                  |                  |             |                |
| 4: Medicine and biotechnology                   |                  |                  |             |                |
| Second article                                  |                  |                  |             |                |
| 5: Space and radiation biology                  |                  |                  |             |                |
| Second article                                  |                  |                  |             |                |
| Yeast outgroup: Industrial applications         |                  |                  |             |                |
| Second article                                  |                  |                  |             |                |

Step 4. Biological Importance, Health Relevance, and Further Exploration

1. Unknown 1: Where is the likely fungus commonly found indoors? What moisture or building conditions favor it? What health concerns are supported by peer-reviewed evidence?

2. Unknown 2: How does the likely fungus attack or suppress other fungi? What makes it useful for biological control, and what limitations or risks should be considered?

---

3. Unknown 3: What toxins, allergens, infections, or food and indoor-environment concerns are associated with the likely fungus? Which claims require strain-level confirmation?

---

4. Unknown 4: How has the likely fungus contributed to medicine or biotechnology? What products or enzymes are associated with it, and what indoor relevance does it have?

---

5. Unknown 5: What evidence connects the likely fungus with radiation tolerance, space-associated environments, or extreme conditions? What remains uncertain?

---

6. Yeast outgroup: Why is *Saccharomyces* useful as an outgroup in this dataset? What major industrial applications make this yeast important?

---

7. Choose one identified fungus and propose a new research question that could be investigated using deoxyribonucleic acid barcoding, ribonucleic acid sequencing, culture experiments, or environmental sampling.

---

8. Explain how identifying fungi in cleanrooms, laboratories, or production areas could support contamination control and product quality.

---

## Part V. Instructor Answer Key and Guidance

### A. Multiple Choice Key

1. B 2. B 3. B 4. B 5. A 6. B 7. B 8. B 9. A 10. A 11. A 12. B

### B. Matching Key

1. E 2. B 3. G 4. J 5. I 6. D 7. K 8. F 9. H 10. A 11. C 12. L

### C. Fill-in-the-Blank Key

- greater-than
- Internal
- alignment
- outgroup
- bootstrap
- mycovirus
- vertical
- fusion
- hypovirulence
- peer
- host
- monitoring

### D. Sequence-of-Events Key

- Fungal barcode analysis workflow
  - Upload the FASTA file to MAFFT.
  - Submit the sequences for alignment.
  - Inspect conserved positions, variable positions, and gaps.
  - Generate and open the phylogenetic tree.
  - Reroot and rotate branches to improve interpretation and visualization.
  - Interpret the unknown using tree, database, and literature evidence.
- Mycovirus-based biological control workflow
  - Identify a mycovirus in a fungal host.
  - Introduce or transfer the mycovirus into a target fungus.
  - Confirm that the mycovirus becomes established.
  - Compare the infected fungus with an uninfected control.
  - Test whether the infected fungus shows reduced growth or virulence.

### E. Expected Preliminary Interpretation of the Unknowns

- Unknown 1:** Most consistent with *Stachybotrys chartarum* or a very closely related *Stachybotrys* sequence. Students should discuss damp-building ecology and avoid unsupported claims that all exposure produces disease.
- Unknown 2:** Most consistent with *Trichoderma harzianum* or a closely related *Trichoderma* sequence. Students should describe mycoparasitism and biological control through competition, enzymes, and hyphal attack.
- Unknown 3:** Most consistent with *Aspergillus flavus* or a closely related *Aspergillus* sequence. Students may discuss aflatoxin and health relevance but must not infer toxin production from the barcode alone.
- Unknown 4:** Most consistent with *Penicillium chrysogenum* or a closely related *Penicillium* sequence. Students may discuss penicillin production and biotechnology while recognizing current taxonomic complexities and barcode limitations.
- Unknown 5:** Most consistent with *Cladosporium sphaerospermum* or a closely related *Cladosporium* sequence. Students may examine melanization, radiation tolerance, and space-related studies but should distinguish demonstrated findings from speculation.
- Yeast outgroup:** *Saccharomyces* species are outside the main filamentous-fungus group and help root the tree. *Saccharomyces cerevisiae* is important in baking, brewing, fermentation, and biotechnology.  
Important caution: A close Internal Transcribed Spacer sequence match supports identification but does not prove strain identity, toxin production, virulence, resistance, biocontrol performance, or industrial productivity. These traits require additional testing.

### F. Research Section Scoring Guide

| Criterion             | Strong                                                                                                                            | Developing                                      | Needs Revision                                         |
|-----------------------|-----------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------|--------------------------------------------------------|
| Sequence evidence     | Uses tree placement, bootstrap support, branch length, database matches, accession numbers, percent identity, and query coverage. | Uses some evidence but omits important details. | Provides an unsupported identification.                |
| Scientific literature | Uses two relevant peer-reviewed sources per assigned organism or context and summarizes them accurately.                          | Uses limited or partly relevant sources.        | Uses non-scientific sources or misrepresents evidence. |

|                                 |                                                                                                                     |                                                                |                                         |
|---------------------------------|---------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------|-----------------------------------------|
| Biological and health relevance | Explains habitat, significance, health or biotechnology relevance, and an additional important trait with evidence. | Provides a general explanation with limited evidence.          | Makes unsupported or inaccurate claims. |
| Final conclusion                | Integrates identity, evidence, significance, limitation, and next step clearly.                                     | Conclusion is partly complete or weakly connected to evidence. | Conclusion is unclear or unsupported.   |

#### **G. Instructor Notes for Evaluating Sources**

1. Accept sources from established scientific journals, university library databases, PubMed, or Google Scholar when the article itself is peer reviewed.
2. Require students to distinguish primary research articles from review articles and general websites.
3. Confirm that the organism discussed in each article matches the identification level supported by the sequence evidence.
4. Do not award full credit for claims about toxins, pathogenicity, resistance, radiation protection, or biocontrol effectiveness unless the cited evidence directly supports the claim.
5. Accept different well-reasoned conclusions when students explain database ambiguity or limited species-level resolution.
6. Encourage students to identify what evidence would be needed before applying a mycovirus or fungal biological-control organism in an indoor environment.