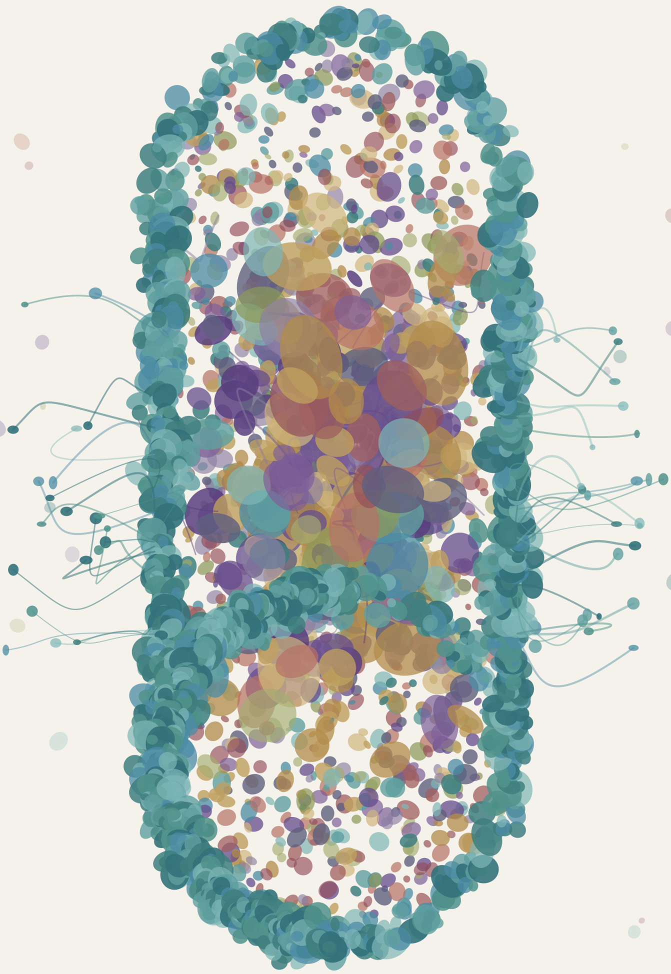
THE

INHABITED BODY

Primer

A Beginner's Guide to the Biology You Need



Dr Horst Herb

BACTERIA · ARCHAEA · EUKARYA

The Inhabited Body — Primer

A Beginner's Guide to the Biology You Need

Horst Herb

2026-04-14

Contents

The Inhabited Body — Primer

A Beginner's Guide to the Biology You Need

Horst Herb

Introduction: Why This Primer Exists

You are about to read a book about the trillions of microorganisms that live inside and upon your body — how they got there, what they do, and why they matter for your health, your mood, and perhaps even who you are. That book, *The Inhabited Body*, assumes no specialist knowledge. Every technical term is explained on first use. Every complex idea is introduced with an analogy before the details arrive.

But here is the reality: some ideas in biology need more than a sentence of explanation. They need a running start.

If you already know what DNA is, how a virus differs from a bacterium, and what it means when scientists talk about "sequencing a genome," you may not need this primer at all. Feel free to jump straight to Chapter 1 of the main book. You can always come back here if you hit a term or concept that catches you off guard.

If, on the other hand, the last time you thought about biology was in secondary school — or if terms like *mRNA*, *archaea*, and *metagenome* feel unfamiliar — this guide was written for you.

What This Primer Is

This is a short, self-contained introduction to the biology you need in order to follow the arguments in *The Inhabited Body*. It covers six topics:

1. **The living world** — the major categories of life on Earth, from bacteria to fungi to the strange entities (viruses, prions) that blur the boundaries of "alive."
2. **The cell** — what cells are, how they work, and why the cells of bacteria, archaea, and humans are built on fundamentally different plans.
3. **The code of life** — DNA, RNA, proteins, and the molecular machinery that turns genetic instructions into functioning organisms.
4. **Fungi** — a kingdom of life that is often overlooked but increasingly important to the microbiome story.
5. **Viruses and their relatives** — including the bacteriophages that shape microbial communities in ways we are only beginning to understand.
6. **How we study what we can't see** — the tools and techniques that underpin every claim in the main book, from microscopy to metagenomics.

Each chapter is written to stand alone. You can read them in order — they are arranged in a "zoom in" sequence, from the broadest categories of life down to the molecular details — or you can dip into whichever chapter addresses a gap you have noticed. Cross-references to the main book are provided at the end of each chapter, so you can see exactly where each concept becomes relevant.

What This Primer Is Not

This is not a textbook. It does not attempt to be comprehensive. There are entire university courses on each of the topics covered here, and this primer compresses them ruthlessly. The goal is not mastery but orientation — enough understanding that, when the main book describes a bacterium exchanging genes with its neighbour, or a phage altering the immune response, or a metagenomic study revealing an unexpected community of organisms, you know what is being talked about and why it matters.

Where simplification risks distorting the science, we say so. Where a topic is more complicated than the explanation suggests, we flag it. Science is full of caveats, exceptions, and ongoing debates, and pretending otherwise would be doing you a disservice.

A Note on Language

Biology has a terminology problem. Many of its key words — *gene*, *species*, *alive* — mean slightly different things depending on context, and even experts disagree about precise definitions. In this primer, we use each term in its most common sense and explain it in plain language. Technical terms are **bolded** on first use and defined in the glossary at the back.

Where a word has a Greek or Latin root that helps make sense of it, we mention it — not because etymology is essential, but because knowing that *prokaryote* means "before the kernel" and *eukaryote* means "true kernel" makes the distinction between these two types of cell easier to remember. You will find, over time, that the jargon stops being jargon and starts being a useful shorthand.

How to Use This Guide

If you are reading *The Inhabited Body* from the beginning, you may want to read this primer first, or at least skim the chapter summaries in the table of contents to see which topics you are already comfortable with.

If you are already reading the main book and have hit a wall, the cross-references at the end of each primer chapter will tell you which primer chapter covers the concept you are struggling with. The glossary can also serve as a quick-reference tool.

If you are simply curious about microbiology, this primer works perfectly well on its own. It covers the fundamentals of how life is organised, how cells work, and how scientists study the invisible world — topics that are interesting in their own right, whether or not you go on to read about the human microbiome.

However you use it, we hope it makes the science in *The Inhabited Body* not just accessible but genuinely enjoyable. The microbiome is one of the most exciting frontiers in modern biology. Understanding it starts here.

Let us begin with the broadest question of all: what kinds of life exist on Earth?

Primer Chapter 1: The Living World — Life's Major Domains

This chapter is part of the companion primer to The Inhabited Body. *It introduces the broadest categories of life on Earth, explains where viruses and other non-cellular entities fit, and provides the conceptual map that later chapters — both in this primer and in the main book — will build upon.*

What Does It Mean to Be Alive?

Before we can talk about the microbes that share your body, we need to agree on what "alive" means. This turns out to be surprisingly difficult. Biologists have argued about it for centuries, and the argument is not settled.

Most working definitions of life include a short list of properties. A living thing takes in energy from its environment and uses it to maintain itself — a process called **metabolism**. It grows. It reproduces, making copies of itself that are similar but not always identical. It responds to its surroundings. And — critically — it carries **hereditary information**, a set of instructions that can be passed from one generation to the next and that can change over time, allowing populations to evolve.

These criteria work well for most of the organisms you are likely to think of. A dog is alive. A fern is alive. The bacterium swimming through your gut right now is alive. Each of these takes in energy, grows, reproduces, and carries a genome — a complete set of genetic instructions encoded in DNA.

But what about a virus? A virus carries genetic information. It evolves, sometimes with breathtaking speed. Yet it cannot reproduce on its own. It has no metabolism. Outside of a host cell, a virus is essentially an inert particle — a set of instructions wrapped in a protein coat, waiting. It does nothing until it encounters a cell it can hijack. Is it alive?

The honest answer is that there is no consensus. Some biologists treat viruses as living. Others treat them as biological entities that are not, strictly speaking, alive but are undeniably part of the living world — much as a computer virus is not a living thing but is certainly a participant in the digital ecosystem. For the purposes of this book, we will take the pragmatic view: viruses matter enormously to human biology and to the microbiome, regardless of whether we grant them the label "alive." We will meet them properly in Primer Chapter 5.

For now, let us focus on the organisms that everyone agrees are living — the cellular life forms — and ask a different question: how do we organise them?

Two Kingdoms, Five Kingdoms, Three Domains

For most of recorded history, people divided living things into two groups: animals and plants. It was an intuitive split. Animals moved; plants did not. Animals ate other organisms; plants drew sustenance from soil and sunlight. Aristotle used this framework. Linnaeus, the father of modern taxonomy, formalised it in the eighteenth century. For everyday purposes, it worked.

It did not work for microbes.

When Antonie van Leeuwenhoek peered through his handmade microscopes in the 1670s and saw what he called *animalcules* — tiny organisms teeming in pond water, in dental scrapings, in rainwater — the two-kingdom system immediately began to creak. Were these creatures animals? They moved, some of them, but they did not eat in any recognisable way. Were they plants? They had no leaves, no roots, no flowers. They were something else, and for the next three centuries, biologists struggled to fit them into a classification scheme that had not been designed for them.

By the mid-twentieth century, the prevailing view was that life could be divided into five kingdoms: Animals, Plants, Fungi, Protists (a grab-bag of single-celled organisms with complex cells), and Monera (bacteria and their relatives — single-celled organisms with simpler cells). This was better, but it still had problems. The kingdom Monera, in particular, lumped together organisms that looked similar under a microscope but turned out, at the molecular level, to be as different from each other as either was from you.

The revolution came from an unlikely source: a physicist-turned-microbiologist named Carl Woese, working at the University of Illinois. In the late 1970s, Woese had the idea of comparing organisms not by their physical appearance — which can be misleading, especially among microbes — but by the sequence of a particular molecule found in every living cell: **ribosomal RNA**, or rRNA. This molecule is part of the cellular machinery that translates genetic instructions into proteins. Because it performs the same essential job in every organism on Earth, it changes only slowly over evolutionary time. By comparing the sequence of rRNA between different species, Woese reasoned, you could build a family tree of all life — one based not on what organisms *looked* like, but on how they were actually related [REF:woese1977].

What he found upended biology.

The organisms that had been lumped together as "bacteria" actually fell into two groups that were profoundly different from each other — as different, at the molecular level, as either was from animals or plants. One group was the true bacteria, which Woese called the **Bacteria** (or Eubacteria). The other was a previously unrecognised lineage that he called the **Archaea** — from the Greek for "ancient things" — because some of the first species discovered lived in extreme environments reminiscent of early Earth: boiling hot springs, highly acidic pools, oxygen-free muds.

In 1990, Woese, Otto Kandler, and Mark Wheelis formally proposed reorganising all of life into three great **domains**: Bacteria, Archaea, and Eukarya [REF:woese1990]. The domains sat above kingdoms in the hierarchy of classification — a higher, more fundamental level of organisation. Animals, plants, fungi, and protists were all Eukarya. Every bacterium you had ever heard of belonged to Bacteria. And the Archaea were something else entirely — a third form of cellular life, hiding in plain sight.

The proposal was not warmly received. Woese was called a crank. One prominent biologist publicly dismissed the three-domain system as unnecessary. But the data accumulated, and by the mid-1990s, the three-domain framework had become the standard way that biologists understood the deepest divisions of life. Today, it remains the most widely used classification at this level, though as we shall see, even it has been challenged by newer discoveries.

Think of the three domains as three great continents on the map of life. Everything alive on Earth —

every organism that metabolises, reproduces, and carries genetic information in the form of DNA — belongs to one of these three groups. The differences between them are not superficial. They go down to the very chemistry of their cells.

Domain 1: Bacteria — The Familiar Strangers

When most people hear the word "microbe," they think of bacteria. This is understandable. Bacteria are the microorganisms we encounter most often in daily life and in the news — from the *Escherichia coli* in our gut to the *Streptococcus* behind a sore throat to the *Lactobacillus* in a pot of yoghurt.

Bacteria are single-celled organisms. Each cell is small — typically between 0.5 and 5 micrometres in length, which means you could line up roughly a thousand of them across the head of a pin. They have no nucleus — no membrane-bound compartment to house their DNA. Instead, their genetic material floats freely within the cell in a concentrated region called the **nucleoid**. This is one of the defining features of what biologists call a **prokaryotic** cell (from the Greek *pro*, "before," and *karyon*, "kernel" or "nut" — a cell without a true kernel).

Despite their small size and apparent simplicity, bacteria are extraordinarily diverse. They come in several characteristic shapes — spheres (*cocci*), rods (*bacilli*), spirals (*spirilla*), and comma-shaped curves (*vibrios*) — but the real diversity lies in their metabolism. Some bacteria eat sugar. Some eat iron. Some eat rock. Some harvest energy from sunlight. Some thrive in boiling water; others flourish in Antarctic ice. Some require oxygen; others are killed by it. The metabolic versatility of bacteria is, without exaggeration, unmatched by any other domain of life.

How many bacterial species exist on Earth? This is a question with no settled answer. Estimates vary wildly depending on how you define a "species" (a surprisingly contentious issue among microbiologists) and how you count. A major census published in 2019 by Louca and colleagues analysed over 1.7 billion genetic sequences and estimated that there are between 0.8 and 1.6 million bacterial and archaeal species globally — of which roughly 690,000 distinct types were detected [REF:louca2019]. Other analyses, using different mathematical models, have proposed numbers as high as one trillion species [REF:locey2016]. The truth is that we do not know. What we do know is that only a tiny fraction of microbial species — perhaps 15,000 bacteria and a few hundred archaea — have been formally named and described in laboratory cultures.

For the purposes of this book, the important point is this: when we talk about the human microbiome, most of the organisms we are discussing are bacteria. They dominate the gut, the skin, the mouth, and most other body sites by both numbers and diversity. Understanding what bacteria are, how they live, and how they differ from other organisms is essential groundwork for everything that follows.

Domain 2: Archaea — The Hidden Third

If bacteria are the familiar strangers, archaea are the strangers you have never met — even though they have been living inside you all along.

Archaea look like bacteria. Under a microscope, you often cannot tell them apart. They are single-celled, prokaryotic (no nucleus), roughly the same size, and come in similar shapes. For decades, they *were* classified as bacteria. It took Woese's molecular revolution to reveal that beneath the surface, archaea are profoundly different.

The differences are chemical. Bacterial cell membranes are made of fatty acids linked to a glycerol backbone by a type of chemical bond called an **ester linkage** — the same basic membrane chemistry that your own cells use. Archaeal membranes are different: they use branched hydrocarbon chains joined to glycerol by **ether linkages**, a more stable arrangement that may help explain why many archaea can tolerate extreme conditions. Bacterial cell walls typically contain a molecule called **peptidoglycan** — a mesh-like polymer that gives the cell structural rigidity. Archaea lack peptidoglycan entirely. Their walls are built from different materials, including a variant sometimes called pseudopeptidoglycan, but the chemistry is distinct.

These are not trivial differences. They are roughly equivalent to discovering that two buildings that look identical from the outside are constructed from entirely different materials — one from brick and mortar, the other from interlocking carbon fibre. The external appearance is similar, but the engineering is fundamentally different.

The early archaea discovered by researchers tended to live in extreme environments — volcanic hot springs, ultra-salty lakes, the oxygen-free depths of swamp mud — which gave rise to the popular image of archaea as "extremophiles," organisms that thrive where nothing else can. This reputation, while not entirely wrong, is misleading. As molecular detection tools improved, researchers found archaea almost everywhere: in soil, in ocean water, in the guts of cattle, and — of direct relevance to this book — in the human body.

In the human microbiome, archaea are a minority but a consistent presence. The most commonly detected human-associated archaeon is *Methanobrevibacter smithii*, a methane-producing organism (a **methanogen**) found in the colon of most adults. *M. smithii* does not eat food directly. Instead, it consumes the waste products of bacterial fermentation — particularly hydrogen and carbon dioxide — and converts them to methane. In doing so, it removes hydrogen from the environment, which actually helps the bacteria around it ferment more efficiently. It is a partnership: the bacteria feed the archaea, and the archaea, by removing a waste product, help the bacteria work better. We will encounter this kind of metabolic cooperation repeatedly throughout the main book.

Perhaps the most profound insight from Woese's work — one that is still being refined today — is that archaea appear to be more closely related to us than to bacteria. The molecular machinery that archaea use to read their DNA and build proteins shares key features with the equivalent machinery in human cells. Some researchers now argue that eukaryotes — your domain, the domain of all complex life — actually evolved from within the archaea, making the archaea not merely our distant cousins but, in a sense, our ancestors [REF:spang2015]. If this view is correct, the three-domain tree of life is really a two-domain tree, with Eukarya as a branch of Archaea. The debate is ongoing, but the direction of the evidence is clear: archaea are far more important to the story of life — including human life — than their obscurity in popular science would suggest.

Domain 3: Eukarya — The Complex Ones

The third domain is the one you belong to. **Eukarya** — from the Greek *eu*, "true," and *karyon*, "kernel" — encompasses every organism whose cells contain a true nucleus: a membrane-bound compartment that houses the DNA. This includes all animals, all plants, all fungi, and a vast assortment of single-celled organisms collectively (and somewhat unsatisfactorily) known as **protists**.

Eukaryotic cells are, as a rule, much larger and more complex than prokaryotic cells. A typical human cell is roughly 10 to 30 micrometres across — ten to sixty times the diameter of a typical bacterium, and perhaps a thousand times its volume. Inside, eukaryotic cells are divided into specialised compartments called **organelles**, each surrounded by its own membrane. The nucleus holds the DNA. The **mitochondria** — often called the "powerhouses of the cell" — generate most of the cell's energy. Plant cells have **chloroplasts**, which capture sunlight. There are internal transport systems, waste-processing centres, protein-packaging facilities. A eukaryotic cell is, by comparison to a bacterium, a small city.

How did this complexity arise? The leading explanation — now supported by overwhelming evidence — is **endosymbiosis**: the idea that certain key organelles in eukaryotic cells were once free-living prokaryotes that were engulfed by a larger cell and, over billions of years, became permanently integrated. Mitochondria are descended from ancient bacteria — almost certainly from a group related to modern **Alphaproteobacteria**. Chloroplasts are descended from ancient **cyanobacteria**, the photosynthetic bacteria that first oxygenated Earth's atmosphere. The host cell that engulfed them was, based on current evidence, most likely an archaeon or something very close to one.

This means that every cell in your body is, in a sense, a partnership. Your mitochondria still carry their own small genome — a remnant of their bacterial past. They still divide independently within the cell, on their own schedule. You are, at the cellular level, a collaboration between domains of life that diverged billions of years ago.

For the microbiome story, the relevant eukaryotes are primarily the fungi and the protists. Fungi — including yeasts, moulds, and mushrooms — are a kingdom of their own within Eukarya, and they are significant members of the human microbiome, particularly on the skin and in the gut. We will give them their own primer chapter (Primer Chapter 4). Protists — organisms like the *Blastocystis* species found in the gut of many healthy people — are rarer in the human microbiome but not absent, and their role is only beginning to be explored.

The Outliers: Viruses, Prions, and Other Entities

The three-domain system organises all *cellular* life. But the biological world contains entities that do not fit neatly into any domain because they are not cells at all.

Viruses

Viruses are the most numerous biological entities on Earth. Their estimated global count — roughly 10^{31} individual viral particles, or **virions** — exceeds the number of all cellular organisms combined. Yet they are not included in the three-domain tree because they lack the defining features of cellular life: they have no metabolism, no ribosomes, no capacity to grow or reproduce independently.

A virus is, at its simplest, a piece of genetic material — either DNA or RNA, but typically not both — wrapped in a protein shell called a **capsid**. Some viruses have an additional outer layer, an **envelope**, made of lipids stolen from the host cell. That is it. No cytoplasm, no membranes of their own, no energy-generating machinery. A virus reproduces by injecting its genetic material into a host cell and commandeering that cell's machinery to make copies of itself.

Viruses infect every domain of life. Some infect animals (influenza, SARS-CoV-2). Some infect plants (tobacco mosaic virus). Some infect bacteria — and these bacteria-infecting viruses, called **bacteriophages** (or simply **phages**), are of enormous importance to the microbiome. Phages are, by some estimates, the most abundant biological entities in the human gut. They shape bacterial communities by killing some species and sparing others, and they shuttle genes between bacteria in ways that can change what those bacteria do. We will devote considerable attention to phages in the main book (Chapters 16–18).

Proviruses and Prophages

Not all viruses immediately destroy their host. Some integrate their genetic material into the host's genome and stay there — quietly, sometimes for generations. When a virus does this to a bacterium, the integrated viral DNA is called a **prophage**, and the virus is said to be in its **lysogenic** cycle. The bacterium goes about its business, replicating the prophage DNA along with its own every time it divides. Under certain conditions — stress, DNA damage, changes in the environment — the prophage can reactivate, excise itself from the bacterial genome, and resume making new viral particles, typically destroying the host cell in the process.

When a virus integrates into the genome of a more complex organism — including a human — it can become an **endogenous virus**. The human genome contains thousands of fragments of ancient retroviruses, called **human endogenous retroviruses** (HERVs), that inserted themselves into our ancestors' DNA millions of years ago. Most are now broken and inert, but a few have been co-opted for essential functions. The protein **syncytin**, for instance — critical for the formation of the human placenta — is derived from an ancient retroviral envelope gene. We are, quite literally, built with viral spare parts. This story is told in detail in Chapter 2 of the main book.

Prions

At the far edge of biology sit **prions** — infectious agents that contain no genetic material at all. A prion is a misfolded protein: a normal cellular protein (called PrP) that has adopted an abnormal three-dimensional shape. The misfolded version can induce normally folded copies of the same protein to refold into the abnormal shape, creating a chain reaction that spreads through brain tissue. Prion diseases — including Creutzfeldt-Jakob disease in humans, BSE ("mad cow disease") in cattle, and scrapie in sheep — are invariably fatal and currently untreatable.

Prions are not part of the microbiome story, but they are worth mentioning here because they illustrate how broad the category of "biological entity" truly is. Life — or at least biological agency — does not require a cell, does not require a genome, and does not require metabolism. It requires only the ability to propagate information. A prion does exactly that, using protein shape rather than DNA as its information carrier.

Viroids and Obelisks

Finally, there are entities even simpler than viruses. **Viroids** are small, circular RNA molecules — no protein coat, no capsid, just naked RNA — that can infect plant cells and cause disease. They were discovered in 1971 and remain the smallest known infectious agents.

In 2024, a team led by Ivan Zheludev reported the discovery of a new class of viroid-like RNA elements in the human gut, which they named **obelisks** [REF:zheludev2024]. These are tiny circular RNA molecules, found within bacteria in the gut microbiome, that appear to encode a single protein. They are not viruses, not viroids in the classical sense, and not part of any known category of biological entity. They are something new — a reminder that the catalogue of life's forms is still far from complete.

The Map of Life: Putting It All Together

Let us step back and survey the landscape. All cellular life on Earth belongs to one of three domains:

- **Bacteria**: single-celled, no nucleus, peptidoglycan cell walls, ester-linked membrane lipids. The dominant organisms in most human microbiome sites.
- **Archaea**: single-celled, no nucleus, no peptidoglycan, ether-linked membrane lipids. A minority presence in the human microbiome, but metabolically important — particularly the methanogens.
- **Eukarya**: cells with a true nucleus and internal organelles. Includes all animals, plants, fungi, and protists. The fungi are significant members of the human microbiome.

Beyond the cellular domains, the biological world includes:

- **Viruses**: non-cellular, dependent on host cells for reproduction. Include bacteriophages, which profoundly influence the microbiome.
- **Proviruses/prophages**: viral DNA integrated into host genomes.
- **Endogenous retroviruses**: ancient viral sequences embedded in the genomes of complex organisms, including humans.
- **Prions**: misfolded proteins that can propagate their shape — infectious agents with no genome.
- **Viroids and obelisks**: minimal RNA elements, some of which inhabit the human gut microbiome.

This map is not complete — no map of life ever is — but it gives us the coordinates we need. In the chapters that follow, we will zoom in on each of these groups, examine their biology in more detail, and begin to understand why they matter to the story of human health.

Where This Matters in *The Inhabited Body*

- **Chapter 1** introduces the microbiome as a community of bacteria, archaea, fungi, and viruses — the cast of characters described here.
- **Chapter 2** tells the story of endogenous retroviruses and how viral DNA became part of the human genome.
- **Chapters 16–18** explore bacteriophages, their role in the microbiome, and the revival of phage therapy.
- **Chapter 4** describes the molecular tools (including 16S rRNA sequencing, the technique Woese pioneered) used to study microbial diversity.

Chapter References

- [REF:woese1977] Woese, C.R. & Fox, G.E. (1977). Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proceedings of the National Academy of Sciences*, 74(11), 5088–5090. DOI: 10.1073/pnas.74.11.5088

- [REF:woese1990] 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*, 87(12), 4576–4579. DOI: 10.1073/pnas.87.12.4576
- [REF:louca2019] Louca, S., Mazel, F., Doebeli, M. & Parfrey, L.W. (2019). A census-based estimate of Earth's bacterial and archaeal diversity. *PLOS Biology*, 17(2), e3000106. DOI: 10.1371/journal.pbio.3000106
- [REF:locey2016] Locey, K.J. & Lennon, J.T. (2016). Scaling laws predict global microbial diversity. *Proceedings of the National Academy of Sciences*, 113(21), 5970–5975. DOI: 10.1073/pnas.1521291113
- [REF:spang2015] Spang, A. et al. (2015). Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature*, 521(7551), 173–179. DOI: 10.1038/nature14447
- [REF:zheludev2024] Zheludev, I.N. et al. (2024). Viroid-like colonists of human microbiomes. *Cell*, 187(23), 6521–6536. DOI: 10.1016/j.cell.2024.09.033
- [REF:lederberg2001] Lederberg, J. & McCray, A.T. (2001). 'Ome sweet 'omics — a genealogical treasury of words. *The Scientist*, 15(7), 8.
- [REF:sender2016] Sender, R., Fuchs, S. & Milo, R. (2016). Revised estimates for the number of human and bacteria cells in the body. *Cell*, 164(3), 337–340. DOI: 10.1016/j.cell.2016.01.013

Primer Chapter 2: The Cell — Life's Basic Unit

This chapter is part of the companion primer to The Inhabited Body. *It explains what cells are, how they work, and why the cells of bacteria, archaea, and humans are built on fundamentally different plans — differences that matter enormously for medicine and for understanding the microbiome.*

The Idea That Changed Everything

In 1665, an English polymath named Robert Hooke pointed a crude microscope at a thin slice of cork and saw something no one had seen before: a lattice of tiny, empty compartments. They reminded him of the small rooms — *cellae* — in which monks lived, and so he called them **cells** [REF:hooke1665].

Hooke did not know what he was looking at. The compartments in cork are the remains of dead plant tissue, and his microscope was far too weak to reveal anything of their internal machinery. But the name stuck, and over the next two centuries a series of increasingly powerful observations converged on a single, transformative insight.

In the late 1830s, the botanist Matthias Schleiden and the physiologist Theodor Schwann independently proposed that all living things are composed of cells — that the cell is the fundamental unit of life [REF:schwann1839]. In 1855, the physician Rudolf Virchow added a crucial corollary: every cell arises from a pre-existing cell. *Omnis cellula e cellula* — "all cells from cells" [REF:virchow1855].

This is **cell theory**, and it remains one of the foundational ideas in biology. It tells us that whether you are looking at a human being composed of roughly 37 trillion cells, or a bacterium that *is* a single cell, the same basic logic applies. Life is cellular. To understand living things — including the trillions of microbes that share your body — you need to understand cells.

What Every Cell Has in Common

Cells come in a staggering variety of sizes, shapes, and specialisations. A human nerve cell can be a metre long. A red blood cell is a flattened disc barely 7 micrometres across. A bacterium might be a hundred times smaller still. Yet despite this diversity, every cell on Earth shares a handful of features. Think of these as the non-negotiable requirements for membership in the club of life.

A boundary. Every cell is surrounded by a **cell membrane** (also called the **plasma membrane**) — a thin, flexible barrier that separates the inside of the cell from the outside world. This is not a passive wall. The membrane is studded with proteins that act as gates, sensors, and pumps, controlling what enters and what leaves. Without a membrane, a cell would simply dissolve into its surroundings, like a drop of ink in water.

The membrane is made of **phospholipids** — molecules with a water-loving (hydrophilic) head and two water-fearing (hydrophobic) tails. In water, these molecules spontaneously arrange themselves into a double layer — a **lipid bilayer** — with the tails pointing inward, away from the water on both sides. This arrangement is so energetically favourable that it forms automatically. You do not need a factory to build a cell membrane. You just need the right molecules and some water.

An interior. Inside the membrane is the **cytoplasm** — a gel-like substance made mostly of water, along with salts, organic molecules, and the molecular machinery the cell needs to function. Everything the cell does — breaking down food, building new components, responding to signals — happens in this crowded interior environment.

A set of instructions. Every cell carries its genetic information in the form of **DNA** — the long, double-stranded molecule we will explore in detail in Primer Chapter 3. DNA contains the instructions for building every protein the cell needs. It is the cell's blueprint, its recipe book, and its operating manual, all in one molecule.

Protein-building machinery. Instructions are useless without the equipment to follow them. Every cell contains **ribosomes** — tiny molecular machines that read the instructions encoded in DNA (via an intermediate molecule called messenger RNA) and assemble the corresponding proteins, one amino acid at a time. Ribosomes are so fundamental to life that their structure is nearly identical across all three domains. As we saw in Primer Chapter 1, it was precisely this deep conservation that allowed Carl Woese to build a universal family tree of life by comparing ribosomal RNA sequences.

These four features — a membrane, a cytoplasm, DNA, and ribosomes — are universal. Every cell that has ever lived on Earth possesses them. Beyond this shared foundation, however, cells diverge dramatically. The most important divergence, for the purposes of this book, is the one between **prokaryotic** cells and **eukaryotic** cells.

The Prokaryotic Cell: Streamlined and Successful

The word **prokaryote** comes from the Greek *pro* ("before") and *karyon* ("kernel" or "nut"). A prokaryotic cell is one that lacks a **nucleus** — a membrane-bound compartment to house its DNA. Instead, the DNA sits in an open region of the cytoplasm called the **nucleoid**, with no membrane separating it from the rest of the cell's contents.

All bacteria and all archaea are prokaryotes. They are, by a wide margin, the most numerous cells on Earth and the dominant organisms in the human microbiome. Understanding their architecture is essential.

[Figure P2.1: The prokaryotic cell — see diagram above]

Size

Prokaryotic cells are small. A typical bacterium is between 0.5 and 5 micrometres in length — roughly a thousandth the width of a pinhead. To put this in perspective: if a human cell were the size of a football stadium, a bacterium would be about the size of a car parked in the middle of the pitch. This is not just an interesting fact. Size constrains biology. A small cell has a large surface area relative to its volume, which means nutrients can diffuse in and waste products can diffuse out quickly, without the need for elaborate internal transport systems. This is one reason why prokaryotes can get away with a simpler internal organisation.

The Cell Envelope

Most prokaryotic cells are surrounded not only by a cell membrane but also by one or more additional layers collectively called the **cell envelope**. The most important of these is the **cell wall** — a rigid or semi-rigid structure that sits outside the membrane and gives the cell its shape. Without a cell wall, a bacterium would swell and burst under osmotic pressure, much as a balloon pops when you overfill it.

In bacteria, the cell wall is made of **peptidoglycan** — a mesh-like polymer of sugars and short amino acid chains that wraps around the cell like a chain-link fence wrapped around a water balloon. Peptidoglycan is unique to bacteria. It is not found in archaea, in eukaryotes, or in any other form of life. This uniqueness makes it an excellent target for antibiotics. Penicillin, the first antibiotic ever discovered, works by disrupting peptidoglycan synthesis — it weakens the cell wall so that the bacterium swells and bursts. Because human cells have no peptidoglycan, penicillin harms bacteria without harming us. This is the principle of **selective toxicity**, and it underpins much of antimicrobial medicine.

A Clinical Landmark: The Gram Stain

In 1884, the Danish bacteriologist Hans Christian Gram developed a staining technique that would become one of the most widely used tools in clinical microbiology [REF:gram1884]. The procedure is simple: bacteria are stained with a violet dye, treated with iodine to fix the dye, washed with alcohol, and then counterstained with a pink dye. Bacteria that retain the violet dye are called **Gram-positive**; those that lose it and take up the pink counterstain are called **Gram-negative**.

The difference comes down to cell wall architecture. Gram-positive bacteria have a thick layer of peptidoglycan — up to 40 layers deep — that traps the violet dye and resists the alcohol wash. Gram-negative bacteria have only a thin peptidoglycan layer, but they compensate with an additional **outer membrane** — a second lipid bilayer sitting outside the cell wall. The alcohol wash penetrates the thin peptidoglycan of Gram-negatives, washes out the violet dye, and the cells pick up the pink counterstain instead.

This distinction is not merely academic. The outer membrane of Gram-negative bacteria contains molecules called **lipopolysaccharides** (LPS), which are potent triggers of the human immune system. When large quantities of LPS enter the bloodstream — as happens in severe Gram-negative infections — the immune response can spiral out of control, leading to sepsis. The outer membrane also acts as a barrier to many antibiotics, which is one reason why Gram-negative infections are generally harder to treat than Gram-positive ones.

For a clinician, knowing whether an infection is Gram-positive or Gram-negative is often the first step in choosing the right antibiotic. For the microbiome researcher, the Gram stain provides a quick shorthand for grouping bacteria and predicting how they interact with the immune system.

Inside the Prokaryotic Cell

The interior of a prokaryotic cell is, compared to a eukaryotic cell, relatively unstructured. There are no membrane-bound compartments, no elaborate internal transport systems. But “unstructured” does not mean “simple.” The cytoplasm of a bacterium is densely packed with molecular machinery.

The **nucleoid** contains the cell’s main chromosome — typically a single, circular molecule of DNA. Unlike eukaryotic chromosomes, bacterial chromosomes are not wrapped around histone proteins (with a few exceptions). The DNA is, however, tightly coiled and organised by specialised proteins that ensure it fits within the tiny cell.

Scattered throughout the cytoplasm are the cell's **ribosomes** — typically thousands of them in a rapidly growing cell. Prokaryotic ribosomes are slightly smaller than their eukaryotic counterparts (designated **70S** versus **80S**, where "S" stands for the Svedberg unit, a measure of sedimentation rate during centrifugation). This size difference is medically important: several classes of antibiotics — including tetracyclines, aminoglycosides, and macrolides — specifically target the 70S ribosome, blocking protein synthesis in bacteria without affecting the 80S ribosomes in human cells.

Many bacteria also carry **plasmids** — small, circular DNA molecules that are separate from the main chromosome. Plasmids replicate independently and often carry genes that provide advantages in particular environments: antibiotic resistance, the ability to break down unusual food sources, or toxin production. Critically, plasmids can be transferred between bacteria — even between different species — through a process called **conjugation**. This is one of the primary mechanisms of **horizontal gene transfer**, a phenomenon we explore in detail in Chapter 2 of the main book. It is also one of the main ways antibiotic resistance spreads through bacterial populations.

On the outside, many prokaryotes sport additional structures. **Flagella** (singular: flagellum) are long, whip-like appendages that spin like a propeller to drive the cell forward. **Pili** (singular: pilus) are shorter, hair-like projections used for attachment to surfaces or for transferring DNA during conjugation. Some bacteria also produce a **capsule** — a slimy outer layer of sugars that helps them evade the immune system and stick to surfaces, including human tissues.

The Eukaryotic Cell: Compartmentalised Complexity

The word **eukaryote** comes from the Greek *eu* ("true") and *karyon* ("kernel"). A eukaryotic cell is one that has a true nucleus — a membrane-enclosed compartment that houses its DNA, separating the genetic material from the rest of the cytoplasm.

All animals, plants, fungi, and protists are eukaryotes. Your body is made of eukaryotic cells. So are the fungi in your microbiome — *Candida*, *Malassezia*, *Saccharomyces*, and their relatives.

Eukaryotic cells are, on average, ten to a hundred times larger than prokaryotic cells. A typical animal cell is 10 to 30 micrometres in diameter — big enough that you could fit a dozen bacteria inside with room to spare. But the really significant difference is not size. It is internal organisation.

[Figure P2.2: The eukaryotic cell — see diagram above]

A Cell of Many Rooms

If a prokaryotic cell is like a studio apartment — one open room where everything happens — then a eukaryotic cell is like a house with many rooms, each dedicated to a particular function. These rooms are the cell's **organelles** — membrane-bound compartments that create specialised chemical environments within the cell.

The nucleus is the largest and most prominent organelle. It contains the cell's DNA, organised into structures called **chromosomes** — long DNA molecules wrapped around spool-like proteins called **histones**. The nucleus is enclosed by a double membrane perforated with **nuclear pores**, which act as gatekeepers controlling the flow of molecules between the nucleus and the cytoplasm. Inside the nucleus, a dense region called the **nucleolus** manufactures the ribosomal RNA that will be assembled into ribosomes.

Mitochondria are the cell's power stations. These oval-shaped organelles take in oxygen and nutrients and convert them into **ATP** (adenosine triphosphate) — the molecule that serves as the universal energy currency of the cell. Almost every energy-requiring process in your body — from contracting a muscle to firing a nerve impulse to copying DNA — is powered by ATP generated in your mitochondria.

As we discussed in Primer Chapter 1, mitochondria have a remarkable origin. They are descended from free-living bacteria — specifically, ancient **alphaproteobacteria** — that were engulfed by an ancestral cell roughly two billion years ago [REF:margulis1967]. Over vast stretches of evolutionary time, the engulfed bacterium lost most of its genes to the host cell's nucleus, but it retained its own small circular chromosome and its own ribosomes (which are bacterial-type 70S ribosomes, not eukaryotic 80S). This is the theory of **endosymbiosis**, first championed by Lynn Margulis in 1967, and it is now one of the best-supported ideas in biology [REF:sagan1967].

The endosymbiotic origin of mitochondria has a practical consequence that matters for the microbiome story. Because mitochondria retain bacterial-type ribosomes, antibiotics that target bacterial ribosomes — such as aminoglycosides and chloramphenicol — can also damage mitochondria at high doses. This is one reason why certain antibiotics carry a risk of side effects involving tissues with high energy demands, such as the inner ear (hearing loss) and the kidneys.

The endoplasmic reticulum (ER) is a vast network of folded membranes that extends from the nucleus throughout the cytoplasm. The "rough" ER is studded with ribosomes and specialises in synthesising proteins destined for export from the cell or for insertion into membranes. The "smooth" ER lacks ribosomes and is involved in lipid synthesis and detoxification.

The Golgi apparatus (named after the Italian physician Camillo Golgi, who first described it in 1898) acts as the cell's post office. Proteins arriving from the ER are sorted, modified, packaged into small membrane-bound sacs called **vesicles**, and dispatched to their final destinations — other organelles, the cell membrane, or the world outside the cell.

Lysosomes are the cell's recycling centres. These small, acidic vesicles contain powerful digestive enzymes that break down worn-out organelles, invading bacteria, and cellular debris. When a white blood cell engulfs a bacterium, it is lysosomes that digest it. Defects in lysosomal enzymes cause a group of inherited diseases known as lysosomal storage disorders, in which undigested material accumulates within cells.

The Cytoskeleton

Eukaryotic cells have an internal scaffolding called the **cytoskeleton** — a network of protein filaments that gives the cell its shape, enables it to move, and provides tracks along which organelles and vesicles are transported. The cytoskeleton is a dynamic structure, constantly being assembled and disassembled as the cell's needs change. It has no equivalent in most prokaryotes (although some bacteria have simpler cytoskeletal proteins that were discovered more recently [REF:shih2006]).

A Word About Plant Cells

The eukaryotic cell described above is an animal cell — the kind that makes up your body. Plant cells share all of these organelles but add two features of their own: a rigid **cell wall** made of cellulose (not to be confused with the peptidoglycan wall of bacteria), and **chloroplasts**, the organelles responsible for photosynthesis. Like mitochondria, chloroplasts are descended from ancient bacteria — in this case, photosynthetic cyanobacteria — acquired through endosymbiosis. Plant cells are not part of the human microbiome, so we will not dwell on them here, but it is worth knowing that the endosymbiotic trick of swallowing a bacterium and keeping it as a permanent internal partner happened at least twice in the history of life.

Archaea: Prokaryotic Architecture, Unique Chemistry

Archaea, as we noted in Primer Chapter 1, are prokaryotes — they lack a nucleus and membrane-bound organelles, and in size and general appearance they resemble bacteria. Under an ordinary microscope, you might not be able to tell them apart. But at the molecular level, archaea are profoundly different, particularly in their membranes.

Bacterial and eukaryotic cell membranes are built from phospholipids with **ester-linked** fatty acid tails — ordinary straight-chain fatty acids attached to a glycerol backbone by ester bonds. Archaeal membranes use **ether-linked** isoprenoid chains instead — branched hydrocarbon tails attached by ether bonds, which are more chemically stable. Some archaeal membranes go further: instead of a bilayer (two separate layers of lipids), they form a **monolayer** — a single sheet of lipids that spans the entire membrane. This monolayer is extraordinarily tough and is found in archaea that live in extreme environments, such as the boiling acidic springs of Yellowstone National Park.

For the human microbiome, the most important archaeal group is the **methanogens** — organisms that produce methane as a metabolic by-product. The dominant human-associated methanogen, *Methanobrevibacter smithii*, lives in the gut and plays a supporting role in the microbial ecosystem. It consumes hydrogen gas produced by bacterial fermentation, which prevents hydrogen from building up and inhibiting the very fermentation reactions that produce it. Think of *M. smithii* as a waste-removal service that keeps the production line running.

Archaea have no peptidoglycan — their cell walls, where present, are made of different materials, most commonly a protein lattice called an **S-layer** (surface layer). This means that antibiotics targeting peptidoglycan, such as penicillin, have no effect on archaea. It also means that the Gram stain — designed to detect differences in peptidoglycan thickness — does not apply to archaea in any meaningful way.

Why the Difference Matters: A Medical Perspective

The structural differences between prokaryotic and eukaryotic cells are not merely academic. They are the foundation of antimicrobial medicine. Almost every antibiotic in clinical use exploits a feature that bacterial cells have and human cells lack:

Peptidoglycan synthesis — targeted by penicillins, cephalosporins, carbapenems, and vancomycin. Human cells have no peptidoglycan, so these drugs are selectively toxic to bacteria.

70S ribosomes — targeted by tetracyclines, aminoglycosides, macrolides (such as erythromycin and azithromycin), and chloramphenicol. Human cytoplasmic ribosomes are 80S and are not affected. (Mitochondrial ribosomes, being bacterial in origin, are 70S — which is why some ribosome-targeting antibiotics carry mitochondrial toxicity at high doses.)

Unique metabolic pathways — for example, the folic acid synthesis pathway, targeted by trimethoprim and sulfonamides. Humans obtain folic acid from their diet and lack this pathway entirely.

This principle — find a target that the pathogen has and the patient does not — is the reason we can treat bacterial infections without poisoning the patient. It is also the reason viral infections are so much harder to treat: viruses hijack the host cell's own machinery, which means there are far fewer unique targets to aim at.

For the microbiome, the implication is equally important but often overlooked. Antibiotics do not distinguish between pathogenic bacteria and beneficial bacteria. A course of amoxicillin prescribed for a throat infection will also kill susceptible bacteria in the gut, the skin, and every other microbiome site. The structural similarity of all bacteria — the shared peptidoglycan, the shared ribosomes — means that our most powerful medical tools are also blunt instruments when it comes to the microbiome. We will return to this problem in Chapter 19 of the main book.

The Factory Analogy

It can help to pull these ideas together with a single analogy. Imagine two kinds of factory.

The prokaryotic factory is a single open warehouse. There is one large room. The blueprint archive (the DNA) sits on a table in the middle of the floor — anyone can walk up and read it. The assembly stations (ribosomes) are scattered around the room. Raw materials come in through the loading dock (the membrane), and finished products leave the same way. The whole operation is lean, fast, and efficient. It does not need much space. It can be duplicated quickly — some bacteria can divide every twenty minutes, effectively copying the entire factory in less time than it takes to eat lunch.

The eukaryotic factory is a large, multi-storey building. The blueprints are locked in a secure archive room (the nucleus), and only authorised copies (messenger RNA) are allowed to leave. Assembly stations sit on a dedicated production floor (the rough ER). Finished products are sent to a quality-control and packaging department (the Golgi apparatus) before being shipped out. The building has its own power plant (mitochondria), a recycling department (lysosomes), and a structural framework (the cytoskeleton) that keeps everything in place and moves materials between floors. It is slower, larger, and more expensive to run — but it can manufacture products of extraordinary complexity.

Neither design is "better" in an absolute sense. The prokaryotic model has dominated life on Earth for over three billion years. The eukaryotic model enabled the evolution of multicellular organisms — including you. Both are present in your microbiome, and understanding their architecture is the first step toward understanding how they interact with your body and with each other.

A Note on Scale

Numbers in biology can become abstract very quickly, so it is worth pausing to anchor a few of them.

A single bacterium is roughly 1 to 2 micrometres long. A micrometre is one millionth of a metre, or one thousandth of a millimetre. The full stop at the end of this sentence is about 500 micrometres across — large enough to fit several hundred bacteria side by side.

A typical human cell is about 10 to 30 micrometres in diameter. Your body contains roughly 37 trillion of them [REF:bianconi2013].

The bacterium *Escherichia coli*, which lives in your gut and is probably the most studied organism on Earth, is about 2 micrometres long and 0.8 micrometres wide. Its entire genome — the complete set of genetic instructions for building and running the cell — is about 4.6 million base pairs of DNA, encoding roughly 4,300 genes [REF:blattner1997].

Your genome, by comparison, is about 3.2 billion base pairs long and contains approximately 20,000 protein-coding genes [REF:international2004]. Your genome is 700 times larger than *E. coli*'s — but you have only about five times as many protein-coding genes. Much of the difference is accounted for by non-coding DNA, regulatory sequences, and the remnants of ancient transposable elements and endogenous retroviruses (as we discussed in Chapter 2 of the main book).

Where This Matters in *The Inhabited Body*

- **Chapter 1** discusses the revised estimates of cell numbers in the human body — roughly 37 trillion human cells, a similar number of microbial cells.
- **Chapter 2** explores horizontal gene transfer, plasmids, and how genetic material moves between bacterial cells — and even between bacteria and the human genome.
- **Chapter 4** introduces the tools (microscopy, culturing, sequencing) used to study cells too small to see with the naked eye.
- **Chapter 12** explains how the immune system distinguishes between the eukaryotic cells of the body and the prokaryotic cells of the microbiome — a distinction that depends on the structural differences described in this chapter.
- **Chapter 19** examines how antibiotics exploit the structural differences between prokaryotic and eukaryotic cells — and the collateral damage they inflict on the microbiome.

Chapter References

- [REF:hooke1665] Hooke, R. (1665). *Micrographia: or Some Physiological Descriptions of Minute Bodies Made by Magnifying Glasses*. London: Royal Society.

- [REF:schwann1839] Schwann, T. (1839). *Mikroskopische Untersuchungen über die Uebereinstimmung in der Struktur und dem Wachsthum der Thiere und Pflanzen*. Berlin: Verlag der Sander'schen Buchhandlung. [English translation: *Microscopical Researches into the Accordance in the Structure and Growth of Animals and Plants*, 1847.]

- [REF:virchow1855] Virchow, R. (1855). Cellular-Pathologie. *Archiv für pathologische Anatomie und Physiologie und für klinische Medicin*, 8(1), 3–39. DOI: 10.1007/BF01935312

- [REF:gram1884] Gram, H.C. (1884). Ueber die isolirte Färbung der Schizomyceten in Schnitt- und Trockenpräparaten. *Fortschritte der Medicin*, 2, 185–189.
- [REF:sagan1967] Sagan, L. [Margulis] (1967). On the origin of mitosing cells. *Journal of Theoretical Biology*, 14(3), 225–274. DOI: 10.1016/0022-5193(67)90079-3
- [REF:margulis1970] Margulis, L. (1970). *Origin of Eukaryotic Cells*. New Haven: Yale University Press.
- [REF:shih2006] Shih, Y.-L. & Rothfield, L. (2006). The bacterial cytoskeleton. *Microbiology and Molecular Biology Reviews*, 70(3), 729–754. DOI: 10.1128/MMBR.00017-06
- [REF:bianconi2013] Bianconi, E. et al. (2013). An estimation of the number of cells in the human body. *Annals of Human Biology*, 40(6), 463–471. DOI: 10.3109/03014460.2013.807878
- [REF:blattner1997] Blattner, F.R. et al. (1997). The complete genome sequence of *Escherichia coli* K-12. *Science*, 277(5331), 1453–1462. DOI: 10.1126/science.277.5331.1453
- [REF:international2004] International Human Genome Sequencing Consortium (2004). Finishing the euchromatic sequence of the human genome. *Nature*, 431(7011), 931–945. DOI: 10.1038/nature03001
- [REF:alberts2022] Alberts, B. et al. (2022). *Molecular Biology of the Cell*, 7th edition. New York: W.W. Norton.

Primer Chapter 3: The Code of Life — DNA, RNA, and Proteins

This chapter is part of the companion primer to The Inhabited Body. *It explains what DNA is, how cells read and use genetic information, and why understanding these processes matters for the microbiome. If you have ever wondered what mRNA actually does — or why it featured so prominently during the COVID-19 pandemic — this chapter is for you.*

A Book Inside Every Cell

Imagine that every building in a city — every house, every office tower, every garden shed — contained a complete copy of the same enormous instruction manual. This manual tells the city how to build every structure, run every service, and respond to every emergency. No building uses every page. A fire station reads the chapters on firefighting; a bakery reads the chapters on baking. But every building holds the entire book, just in case.

This is roughly how your body works. Almost every one of your 37 trillion cells carries a complete copy of your **genome** — the full set of genetic instructions for building and maintaining you. The genome is written in a molecule called **DNA**, and it is astonishingly long. If you could unwind the DNA from a single human cell and stretch it into a line, it would extend about two metres. If you lined up the DNA from all the cells in your body, it would cover the distance from the Earth to the Sun more than 40 times over — roughly 6 billion kilometres of molecular thread [REF:piovesan2019].

But length is not the same as complexity. The genome's real power lies not in how much DNA there is, but in how the information within it is organised, read, and acted upon. Understanding this process — how a chemical molecule becomes a living, functioning organism — is the purpose of this chapter.

What DNA Actually Is

In April 1953, James Watson and Francis Crick published a brief paper in *Nature* that would transform biology. In barely more than a page, they described the structure of **deoxyribonucleic acid** — DNA — as a **double helix**: two strands wound around each other like a twisted ladder [REF:watson1953].

The discovery did not come from nowhere. It built on decades of work by many scientists, including the X-ray crystallography of Rosalind Franklin and Maurice Wilkins, the biochemistry of Erwin Chargaff, and the earlier demonstration by Oswald Avery and colleagues that DNA, not protein, was the molecule of inheritance. But Watson and Crick's model brought everything together in a single, elegant structure — one whose shape immediately suggested how it might work.

The Twisted Ladder

Think of DNA as a ladder that has been gently twisted along its length. The **rungs** of the ladder are pairs of small molecules called **bases** (or nucleotide bases). There are only four of them:

- **Adenine** (A)
- **Thymine** (T)
- **Guanine** (G)
- **Cytosine** (C)

These four letters are the entire alphabet of the genetic code. Every instruction for building and running a living organism — from a bacterium to a blue whale — is written using just A, T, G, and C.

The bases pair in a strict, predictable way: A always pairs with T, and G always pairs with C. These are called **base pairs**. This pairing rule — known as **complementary base pairing** — is one of the most important principles in molecular biology. It means that if you know the sequence of one strand of the DNA ladder, you can immediately deduce the other. The two strands are mirror images, like a photograph and its negative.

The **sides** of the ladder (the vertical rails to which the bases are attached) are made of alternating sugar and phosphate molecules. The sugar in DNA is **deoxyribose** — hence the "deoxyribo" in deoxyribonucleic acid. Each unit of one base plus one sugar plus one phosphate group is called a **nucleotide**. DNA is simply a chain of nucleotides — a polymer — and the human genome is about 3.2 billion nucleotides long [REF:international2004].

Why the Double Helix Matters

The double-helix structure is not just aesthetically pleasing. It solves two fundamental problems of biology.

First, it explains **how genetic information is stored**. The information in DNA is encoded in the sequence of bases along the strand — just as the information in a book is encoded in the sequence of letters on the page. The sequence ATGCCTAA means something different from AATCCGTA, just as "cat" means something different from "act." Change the sequence and you change the instructions.

Second, it explains **how genetic information is copied**. When a cell divides, it must duplicate its DNA so that each daughter cell receives a complete copy. The base-pairing rules make this straightforward: the two strands separate (like unzipping a zip), and each strand serves as a template for building a new partner. Wherever there is an A on the old strand, a T is placed on the new one; wherever there is a G, a C is placed. The result is two identical double helices where there was one before.

This process is called **DNA replication**, and it occurs every time a cell divides. In a rapidly growing bacterium like *Escherichia coli*, it happens once every 20 minutes or so. In your body, it happens millions of times a day. The accuracy is remarkable — the cellular machinery makes only about one error per billion nucleotides copied — but it is not perfect. The occasional mistake that slips through is a **mutation**, and mutations are the raw material of evolution.

From DNA to Protein: The Central Dogma

Knowing the structure of DNA was a monumental achievement, but it immediately raised a deeper question: how does a cell actually *use* the information in its DNA? The answer, worked out over the decade following Watson and Crick's discovery, was formalised by Crick in 1970 as the **central dogma of molecular biology** [REF:crick1970]:

DNA ☐ RNA ☐ Protein

This is not a dogma in the religious sense (Crick later admitted the name was poorly chosen). It is a statement about the flow of information in living systems. Genetic information flows from DNA to **RNA** to **protein**. Let us unpack what this means, step by step.

The Recipe Book Analogy

Imagine your genome as a vast recipe book locked in a vault (the cell's **nucleus** — recall from Primer Chapter 2 that eukaryotic cells keep their DNA inside a membrane-bound nucleus). The book is far too valuable to take into the kitchen. So when the cell needs to make a particular dish — say, a digestive enzyme — it does not haul out the entire book. Instead, it makes a **temporary copy** of just the relevant recipe. That copy is carried out of the vault into the kitchen, where it is read by the cooks and used to assemble the dish. When the dish is done, the copy is discarded.

In this analogy:

- The **recipe book** is the DNA (the genome).
- The **temporary copy** is **messenger RNA (mRNA)**.
- The **kitchen** is the **ribosome** (the protein-building machine introduced in Primer Chapter 2).
- The **dish** is a **protein**.

This analogy is imperfect — biology always is — but it captures the essential logic. DNA stores the information. RNA carries a working copy. Ribosomes read the copy and build the protein.

Step One: Transcription — Copying the Recipe

The process of making an RNA copy of a gene is called **transcription** (literally, "writing across"). It works like this:

1. An enzyme called **RNA polymerase** binds to the DNA at the start of a gene.
2. It unwinds a short stretch of the double helix, exposing the bases on one strand.
3. It reads along the strand, building a complementary **RNA** molecule one base at a time — much like DNA replication, except that RNA uses the sugar **ribose** instead of deoxyribose and substitutes the base **uracil** (U) for thymine (T).
4. When it reaches the end of the gene, it releases the newly made RNA molecule and the DNA snaps back together.

The result is a single-stranded molecule of **messenger RNA (mRNA)** — a disposable working copy of one gene. In eukaryotic cells, the mRNA is processed (trimmed, modified, and polished) before being exported from the nucleus into the cytoplasm, where the ribosomes await.

In bacterial cells — which, as we saw in Primer Chapter 2, have no nucleus — transcription and the next step (translation) happen almost simultaneously. An mRNA molecule can begin being read by ribosomes even before it is fully transcribed. This difference in timing between bacteria and human cells has important consequences for antibiotic design, as we will see in later chapters.

A Note on RNA

RNA is chemically very similar to DNA, but there are three key differences:

- RNA uses the sugar **ribose** (DNA uses deoxyribose — which is ribose with one fewer oxygen atom, hence "deoxy").
- RNA uses the base **uracil (U)** where DNA uses **thymine (T)**. Both pair with adenine.
- RNA is usually **single-stranded**, whereas DNA is double-stranded.

These differences make RNA less chemically stable than DNA — which is exactly the point. RNA is meant to be temporary. It is the Post-it note of molecular biology: carry the message, deliver it, then get recycled. DNA, by contrast, is the archival copy — built for permanence.

Step Two: Translation — Building the Dish

Once the mRNA reaches a ribosome, the second stage begins: **translation** (so called because the information is "translated" from the language of nucleic acids into the language of proteins). This is where the genetic code — the set of rules that links base sequences to amino acids — comes into play.

The Genetic Code: A Three-Letter Language

In the early 1960s, a series of brilliant experiments — most famously by Marshall Nirenberg and Heinrich Matthaei at the US National Institutes of Health — cracked the **genetic code** [REF:nirenberg1961]. Here is how it works.

The mRNA is read in groups of three bases at a time. Each group of three — called a **codon** — specifies one **amino acid**. Amino acids are the building blocks of proteins, rather like beads on a necklace. There are 20 standard amino acids used by living cells, and the sequence of codons in an mRNA determines the sequence of amino acids in the resulting protein.

With four bases and groups of three, there are $4 \times 4 \times 4 = 64$ possible codons. But there are only 20 amino acids, so most amino acids are specified by more than one codon. The codon GCU, for example, specifies the amino acid alanine — but so do GCC, GCA, and GCG. This redundancy is called **degeneracy** of the code, and it provides a buffer against mutation: a change in the third position of a codon often has no effect on the resulting protein.

Three of the 64 codons do not specify any amino acid. Instead, they act as **stop signals** — punctuation marks that tell the ribosome "the protein ends here." One particular codon, AUG, serves double duty: it specifies the amino acid methionine *and* acts as the **start signal** for translation. Almost every protein begins with methionine.

The genetic code is very nearly universal. With minor exceptions, every organism on Earth — from the bacteria in your gut to the cells of your brain — uses the same code. A codon that specifies alanine in a human cell specifies alanine in a bacterium. This universality is powerful evidence that all life on Earth shares a common ancestor.

How Translation Works: A Factory Floor

Picture a ribosome as a small but precise factory machine, sitting on the mRNA like a train on a track. As the ribosome moves along the mRNA, it reads each codon and recruits the matching amino acid.

The matchmaking is done by a third type of RNA molecule: **transfer RNA (tRNA)**. Each tRNA carries a specific amino acid on one end and has a three-base **anticodon** on the other — a sequence complementary to the mRNA codon. When the anticodon matches the codon currently sitting in the ribosome, the tRNA delivers its amino acid, which is linked to the growing protein chain. The empty tRNA is then ejected, the ribosome moves forward by three bases, and the process repeats.

A typical protein is between 100 and 1,000 amino acids long. A bacterial ribosome can assemble about 20 amino acids per second — meaning a medium-sized protein takes about 15 seconds to build. Multiple ribosomes can read the same mRNA simultaneously, strung along it like beads on a string, each producing its own copy of the protein. This cluster of ribosomes on an mRNA is called a **polysome**, and it allows cells to produce many copies of a protein quickly.

When the ribosome encounters a stop codon, translation halts. The newly made protein is released and begins to fold into its functional three-dimensional shape — a process we will return to shortly.

Proteins: The Workers of the Cell

If DNA is the blueprint and RNA is the messenger, then **proteins** are the workers — the molecules that actually *do* things. The importance of proteins to life cannot be overstated. Consider just a few of their roles:

Enzymes speed up chemical reactions. Every metabolic process in your body — digesting food, synthesising neurotransmitters, detoxifying drugs — depends on enzymes. Without them, the chemistry of life would be too slow to sustain a cell.

Structural proteins provide physical support. Collagen gives your skin its strength. Keratin makes up your hair and nails. The bacterial cell wall (discussed in Primer Chapter 2) is reinforced by enzymes that build and maintain it.

Transport proteins move molecules around. Haemoglobin carries oxygen in your blood. Membrane channels shuttle ions and nutrients in and out of cells.

Signalling proteins carry messages. Hormones like insulin, cytokines that coordinate the immune response, and neurotransmitters that relay nerve impulses are all proteins (or are made by proteins).

Defence proteins protect the organism. Antibodies, which recognise and neutralise invaders, are proteins. So are many of the antimicrobial peptides that your body deploys against pathogenic bacteria.

Protein Folding: Shape Is Everything

A newly made protein is just a long chain of amino acids — a string of beads. To function, it must fold into a specific three-dimensional shape, dictated by the sequence of its amino acids. Hydrophobic amino acids tend to cluster together in the interior (away from the surrounding water), while hydrophilic ones face outward. The chain twists into local structures — **alpha helices** (like a corkscrew) and **beta sheets** (like a pleated fan) — which then pack together into a compact, functional form.

This folding happens in milliseconds, and when it goes right, it produces a molecular machine of astonishing precision. But when folding goes wrong, the consequences can be severe. Misfolded proteins are the basis of diseases like Alzheimer's, Parkinson's, and the prion diseases discussed in Primer Chapter 1. Your cells have elaborate quality-control systems — **chaperone** proteins that help other proteins fold correctly, and **proteasomes** that shred and recycle those that fail.

Gene Regulation: Not Every Recipe Is Cooked

Here is a puzzle. Almost every nucleated cell in your body carries the same genome — the same 20,000-odd protein-coding genes. (There are exceptions: mature red blood cells have shed their nucleus entirely, immune cells deliberately rearrange certain genes to produce diverse antibodies, and researchers have documented small-scale somatic mutations that accumulate differently across tissues over a lifetime. But the vast majority of your cells share an essentially identical copy of the master plan.) Yet a liver cell looks and behaves utterly differently from a neuron, which looks and behaves utterly differently from a white blood cell. If they all have the same instruction manual, how can they be so different?

The answer is **gene regulation** — the system that controls which genes are switched on (expressed) and which are switched off (silenced) in any given cell at any given time. A liver cell and a neuron contain the same recipes, but they are reading different pages.

This insight traces back to the pioneering work of François Jacob and Jacques Monod, who in 1961 described how bacteria regulate gene expression in response to their environment [REF:jacob1961]. They showed that *E. coli* only produces the enzymes for digesting the sugar lactose when lactose is actually present — a feat of molecular economy that earned them the Nobel Prize.

Gene regulation is extraordinarily sophisticated. In your cells, it involves a cast of hundreds of specialised proteins called **transcription factors** that bind to specific DNA sequences near genes and either promote or block transcription. Some transcription factors activate genes; others repress them. The combination of active transcription factors in a cell — determined by signals from the environment, from neighbouring cells, and from the cell's own developmental history — defines the cell's identity.

This is why a liver cell stays a liver cell and a neuron stays a neuron, even though both carry the same genome. They are reading different chapters of the same book.

A Dimmer Switch, Not an On/Off Toggle

It is tempting to think of genes as being simply "on" or "off," like a light switch. In reality, gene expression is more like a dimmer: genes can be expressed at high levels, low levels, or anything in between. The rate of transcription can be adjusted up or down, the stability of the mRNA can be altered, and the efficiency of translation can be tuned. Even after a protein is made, it can be chemically modified, redirected, or tagged for destruction.

This fine-tuning is essential. Consider insulin: your body does not simply make insulin or not. It adjusts insulin production moment by moment in response to blood sugar levels. Too much insulin is as dangerous as too little. The same is true for almost every protein in your body — and, as we will see in later chapters, for many of the proteins produced by your microbial residents as well.

Beyond the Gene: The "Junk DNA" Revolution

When the Human Genome Project published its first results, one of the most surprising findings was how little of the genome actually codes for proteins. Of the roughly 3.2 billion base pairs in the human genome, only about 1.5 per cent consists of **exons** — the segments of genes that are translated into protein [REF:venter2001]. The rest was initially dismissed as "junk DNA" — evolutionary debris with no function.

That view has been substantially revised. The ENCODE project, a massive international effort to catalogue the functional elements of the human genome, reported in 2012 that roughly 80 per cent of the genome shows some form of biochemical activity — including regulation, transcription of non-coding RNAs, and other functions [REF:encode2012]. (The precise interpretation of this figure remains debated among scientists, but the days of "junk DNA" as a default label are over.)

Among the non-coding regions that have turned out to be important:

Promoters and enhancers are DNA sequences that control when, where, and how strongly a gene is transcribed. They do not code for proteins themselves, but without them, the protein-coding genes would be silent.

Introns are sequences *within* genes that are transcribed into RNA but then cut out before the mRNA is translated. They were once thought to be useless, but many introns contain regulatory elements, and the process of removing them — called **splicing** — allows a single gene to produce multiple different proteins by including or excluding different combinations of exons. This is called **alternative splicing**, and it vastly increases the protein diversity that a genome can generate. The human genome has about 20,000 protein-coding genes, but through alternative splicing it can generate a far larger number of distinct protein variants — though exactly how many of these splice variants produce stable, functional proteins remains an active area of research.

Non-coding RNAs are RNA molecules that are transcribed from DNA but never translated into protein. Far from being waste products, many of them play critical roles. **MicroRNAs** (miRNAs), for example, are tiny RNA molecules — only about 22 bases long — that regulate gene expression by binding to messenger RNAs and preventing their translation or promoting their destruction [REF:schmiedel2015]. **Long non-coding RNAs** (lncRNAs) participate in a bewildering variety of regulatory functions, many of which are still being discovered.

Epigenetics: Writing in the Margins

There is one more layer of gene regulation that deserves special attention, because it has profound implications for the microbiome story.

Epigenetics (literally, "above genetics") refers to changes in gene expression that do not involve altering the DNA sequence itself. Think of it this way: if the genome is a book, epigenetics is the system of bookmarks, highlights, and sticky notes that tell the reader which pages to read, which to skip, and which to read extra carefully. The underlying text does not change, but its interpretation does.

The two best-understood epigenetic mechanisms are:

DNA methylation. Small chemical groups called **methyl groups** can be attached to cytosine bases in the DNA. When a gene's promoter region is heavily methylated, the gene is typically silenced — like placing a "do not read" sticker over a recipe. Methylation patterns are established during development and can be maintained through cell division, ensuring that a liver cell's daughter cells are also liver cells.

Histone modification. In eukaryotic cells, DNA is not floating free. It is wound around spool-like proteins called **histones**, forming a structure called **chromatin**. The tightness of this winding determines whether a gene is accessible or hidden. Chemical modifications to the histone tails — adding or removing acetyl groups, methyl groups, phosphate groups, and others — can loosen or tighten the chromatin, making genes more or less available for transcription. Think of this as the difference between a tightly shelved book (hard to read) and one lying open on the desk (easy to read).

Why Epigenetics Matters for the Microbiome

Here is where the story connects to *The Inhabited Body*. Epigenetic changes are not fixed at birth. They can be influenced by environmental factors throughout life — including diet, stress, toxin exposure, and, crucially, the metabolites produced by your gut microbiome.

Short-chain fatty acids such as **butyrate**, produced by bacterial fermentation of dietary fibre in the colon, are potent inhibitors of histone deacetylases — enzymes that tighten chromatin and silence genes. By inhibiting these enzymes, butyrate keeps chromatin in a more open, transcriptionally active state, influencing the expression of genes involved in immune regulation, gut barrier integrity, and even cancer suppression. This means that the bacteria in your gut are not merely passive passengers. Through the metabolites they produce, they are actively writing in the margins of your genetic instruction manual — influencing which genes your own cells read, and how loudly.

We will explore this extraordinary dialogue in detail in Chapters 7 and 11 of the main book. For now, the key point is that understanding DNA, RNA, and protein is not just about understanding your own cells. It is about understanding the molecular language through which your microbial residents communicate with you.

---

mRNA: From Obscurity to Headlines

For decades, mRNA was the quiet middle child of molecular biology — overshadowed by DNA (which stored the information) and protein (which did the work). Then came the COVID-19 pandemic, and mRNA became a household word.

The **mRNA vaccines** developed by Pfizer-BioNTech and Moderna work by delivering synthetic mRNA into your cells. This mRNA encodes the spike protein of the SARS-CoV-2 virus — the protein the virus uses to latch onto and enter your cells. Your ribosomes read the synthetic mRNA and produce the spike protein, just as they would any other protein. Your immune system then recognises the spike protein as foreign and mounts a defence. When the real virus arrives, your immune system is already primed and ready.

The synthetic mRNA does not enter the nucleus and cannot alter your DNA. It is read, used, and degraded within hours to days — just like any other mRNA. It is the epitome of the disposable messenger.

The key breakthrough that made mRNA vaccines possible came from the work of Katalin Karikó and Drew Weissman, who discovered in 2005 that chemically modifying certain nucleosides in synthetic RNA prevented it from triggering a dangerous inflammatory response [REF:kariko2005]. This finding — which earned them the Nobel Prize in Physiology or Medicine in 2023 — solved a problem that had stymied mRNA therapeutics for years: the immune system's tendency to attack foreign RNA as if it were a sign of viral infection.

This detail circles back to the microbiome in an interesting way. The innate immune system distinguishes self from non-self RNA in part by detecting the presence or absence of nucleoside modifications. Mammalian RNA is heavily modified; bacterial and viral RNA typically is not. This means that when bacterial RNA leaks from damaged microbes in the gut, it can trigger immune responses through **Toll-like receptors** (TLR3, TLR7, and TLR8) — the same receptors that Karikó and Weissman's work was designed to evade. The conversation between microbial RNA and the immune system is an active area of research that we revisit in Chapters 12 and 13 of the main book.

How Bacteria Do It Differently

Everything described so far — the nucleus, the introns, the histones, the elaborate processing of mRNA — applies to eukaryotic cells. Bacteria and archaea, as we saw in Primer Chapter 2, are prokaryotes, and they handle their genetic information differently. These differences are not merely academic curiosities. They are the reason many antibiotics work, and they are essential for understanding how the microbiome functions.

No Nucleus, No Waiting

Bacterial DNA sits in the cytoplasm, not locked away in a nucleus. This means there is no barrier between where mRNA is made (transcription) and where it is read (translation). In bacteria, ribosomes begin translating an mRNA while it is still being transcribed — the ribosome latches on to the leading end of the mRNA even as RNA polymerase is still extending the trailing end. This coupling of transcription and translation gives bacteria extraordinary speed. A bacterium can go from receiving an environmental signal to producing a new protein in under a minute.

Operons: The Bacterial Efficiency Trick

Bacteria organise many of their genes into **operons** — clusters of genes that are transcribed together as a single mRNA. The genes in an operon typically encode proteins that work together in the same metabolic pathway. For example, the *lac* operon in *E. coli* contains three genes needed for digesting lactose, all under the control of a single promoter. When lactose is absent, a repressor protein blocks transcription of the entire operon. When lactose appears, the repressor releases, and all three genes are transcribed together. It is an elegantly efficient system — like printing all the pages of a multi-step recipe on a single sheet of paper.

Eukaryotic cells, by contrast, almost never organise their genes into operons. Each gene typically has its own promoter and is transcribed independently. This gives eukaryotes more precise control over individual genes, but at the cost of the streamlined efficiency that operons provide.

Fewer Introns, Faster Processing

Bacterial genes almost never contain introns. The mRNA produced by transcription is the mRNA that gets translated — no splicing required. This is another reason bacterial gene expression is so fast. It also means that bacterial genomes are far more compact than eukaryotic ones. The *E. coli* genome has about 4,300 genes packed into 4.6 million base pairs. The human genome has about 20,000 protein-coding genes spread across 3.2 billion base pairs. Gene for gene, bacteria use their DNA about 70 times more efficiently.

Different Ribosomes, Different Targets

As mentioned in Primer Chapter 2, bacterial ribosomes (70S) are structurally different from eukaryotic ribosomes (80S). This difference is the basis for several classes of antibiotics. Drugs like **erythromycin**, **tetracycline**, and **chloramphenicol** bind specifically to the bacterial 70S ribosome and block translation, killing the bacterium or halting its growth. Because these drugs do not bind to the human 80S ribosome, they do not directly harm human cells — a crucial application of the principle of **selective toxicity** introduced in Primer Chapter 2.

However — and this is a point we will return to repeatedly in *The Inhabited Body* — these antibiotics cannot distinguish between pathogenic bacteria and beneficial ones. When you take an antibiotic for a throat infection, the drug also disrupts protein synthesis in the trillions of helpful bacteria in your gut, on your skin, and throughout your body. The collateral damage can be profound.

Horizontal Gene Transfer: Sharing the Recipe Book

In animals and plants, genes are passed **vertically** — from parent to offspring. Bacteria, however, have a second trick: **horizontal gene transfer** (HGT), in which genes are passed between unrelated organisms, even across species boundaries. If vertical gene transfer is like inheriting your grandmother's recipe book, horizontal gene transfer is like a stranger at a bus stop handing you a recipe they tore from a magazine.

There are three main mechanisms, all introduced briefly in Primer Chapter 2 and explored in depth in Chapter 2 of the main book:

Transformation — a bacterium picks up free-floating DNA from its environment (often released by dead cells nearby).

Transduction — a bacteriophage (a virus that infects bacteria) accidentally packages a fragment of one bacterium's DNA and injects it into another.

Conjugation — two bacteria form a physical bridge (a **pilus**), and one transfers a plasmid or other DNA directly to the other. This is sometimes called "bacterial sex," although it involves gene transfer rather than reproduction.

Horizontal gene transfer is how antibiotic resistance spreads so rapidly through bacterial populations. A single bacterium that acquires a resistance gene — whether by mutation or by receiving a plasmid — can share that gene with its neighbours in hours. This is why antibiotic resistance is one of the most urgent public health challenges of our time, as we will discuss at length in Chapter 19 of the main book.

Tying It All Together

The flow of genetic information — from DNA to RNA to protein — is the central process of all cellular life. But understanding the code of life is not just about memorising the steps. It is about appreciating the system as a whole: a dynamic, regulated, responsive network that allows cells to adapt to their environment, communicate with their neighbours, and maintain their identity.

For the microbiome, these molecular details are not abstractions. They are the mechanisms through which:

- Bacteria produce the enzymes that ferment dietary fibre into short-chain fatty acids.
- Gut microbes manufacture vitamins, neurotransmitters, and immune-modulating signals.
- Pathogenic bacteria deploy toxins, adhesins, and drug-resistance proteins.
- Your own immune cells recognise microbial molecules (including RNA and proteins) and calibrate their response.
- The metabolic products of microbial gene expression influence the epigenetic regulation of your own genes.

Every one of these processes depends on the molecular machinery described in this chapter: DNA storing the instructions, RNA carrying the message, ribosomes building the protein, and regulatory systems deciding which instructions to follow and when.

Where This Matters in *The Inhabited Body*

- **Chapter 2** explores horizontal gene transfer and how genes move between microbes — and between microbes and the human genome.
- **Chapter 4** introduces metagenomics and other tools for reading microbial DNA directly from environmental samples.
- **Chapter 7** examines how microbial metabolites — the products of microbial gene expression — influence human metabolism and health.
- **Chapter 11** discusses the gut-brain axis, where microbial-derived neurotransmitters (proteins and small molecules built from genetically encoded pathways) affect mood and cognition.
- **Chapter 12** explains how the immune system distinguishes self from non-self — a distinction that depends on recognising microbial DNA, RNA, and protein.
- **Chapter 19** examines antibiotic resistance: the spread of resistance genes through horizontal gene transfer, and the impact of antibiotics on beneficial microbes.

Chapter References

- [REF:piovesan2019] Piovesan, A. et al. (2019). On the length, weight and GC content of the human genome. *BMC Research Notes*, 12, 106. DOI: 10.1186/s13104-019-4137-z
- [REF:watson1953] Watson, J.D. & Crick, F.H. (1953). Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. *Nature*, 171(4356), 737–738. DOI: 10.1038/171737a0
- [REF:crick1970] Crick, F. (1970). Central dogma of molecular biology. *Nature*, 227(5258), 561–563. DOI: 10.1038/227561a0
- [REF:nirenberg1961] Nirenberg, M.W. & Matthaei, J.H. (1961). The dependence of cell-free protein synthesis in *E. coli* upon naturally occurring or synthetic polyribonucleotides. *Proceedings of the National Academy of Sciences*, 47(10), 1588–1602. DOI: 10.1073/pnas.47.10.1588
- [REF:crick1961] Crick, F.H., Barnett, L., Brenner, S. & Watts-Tobin, R.J. (1961). General nature of the genetic code for proteins. *Nature*, 192, 1227–1232. DOI: 10.1038/1921227a0
- [REF:jacob1961] Jacob, F. & Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. *Journal of Molecular Biology*, 3, 318–356. DOI: 10.1016/s0022-2836(61)80072-7
- [REF:kariko2005] Karikó, K., Buckstein, M., Ni, H. & Weissman, D. (2005). Suppression of RNA recognition by Toll-like receptors: the impact of nucleoside modification and the evolutionary origin of RNA. *Immunity*, 23(2), 165–175. DOI: 10.1016/j.immuni.2005.06.008
- [REF:encode2012] ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature*, 489(7414), 57–74. DOI: 10.1038/nature11247
- [REF:venter2001] Venter, J.C. et al. (2001). The sequence of the human genome. *Science*, 291(5507), 1304–1351. DOI: 10.1126/science.1058040
- [REF:international2004] International Human Genome Sequencing Consortium (2004). Finishing the euchromatic sequence of the human genome. *Nature*, 431(7011), 931–945. DOI: 10.1038/nature03001
- [REF:schmiedel2015] Schmiedel, J.M. et al. (2015). MicroRNA control of protein expression noise. *Science*, 348(6230), 128–132. DOI: 10.1126/science.aaa1738
- [REF:alberts2022] Alberts, B. et al. (2022). *Molecular Biology of the Cell*, 7th edition. New York: W.W. Norton.

Primer Chapter 4: The Fungal Kingdom — Your Closest Strangers

This chapter is part of the companion primer to The Inhabited Body. *It introduces the kingdom Fungi — what they are, how they live, how they differ from bacteria, plants, and animals, and why their biology matters for understanding the human mycobiome. Later chapters in the main book will explore their specific roles at different body sites.*

Not a Plant, Not an Animal

For most of recorded history, fungi were classified as plants. It is easy to see why. Mushrooms sprout from the ground. Moulds grow on bread. Neither runs away when you approach. Carl Linnaeus, who gave us modern taxonomy, placed fungi firmly in the plant kingdom in the eighteenth century, and there they remained for the better part of two hundred years.

This was a mistake — and not a small one.

Fungi do not photosynthesise. They have no chloroplasts, no chlorophyll, no capacity whatsoever to capture energy from sunlight. Every fungus that has ever lived has obtained its energy by breaking down organic matter produced by other organisms — or, in some cases, by directly parasitising a living host. In this sense, fungi eat more like animals than they grow like plants.

But the most surprising correction came from molecular biology. When researchers in the 1990s began comparing DNA sequences across the major kingdoms of life, they discovered something that upended centuries of intuition: fungi are more closely related to animals than they are to plants [REF:baldauf1993]. You share a more recent common ancestor with the mould on your shower curtain than that mould shares with the grass in your garden.

The evidence for this relationship is now overwhelming. Fungi and animals both belong to a supergroup of eukaryotic life called the **Opisthokonta** — named for the single, rear-mounted flagellum (from the Greek *opisthen*, “behind,” and *kontos*, “pole”) found in the swimming cells of ancestral members of the group [REF:cavalier-smith1987]. Chytrid fungi still produce flagellated spores that swim with this characteristic posterior tail. Animal sperm cells do the same. Plants, by contrast, belong to an entirely different branch of the eukaryotic tree.

The last common ancestor of animals and fungi is estimated to have lived roughly one billion years ago, most likely as a single-celled, flagellated organism in an aquatic environment [REF:parfrey2011]. From that shared starting point, the two lineages diverged spectacularly — one branch developing nervous systems, muscles, and internal skeletons; the other evolving into networks of filaments that digest the world from the outside in. But at the molecular level, the kinship is still visible. Fungi and animals share key metabolic enzymes, signalling pathways, and gene families that are absent in plants. Both store surplus energy as **glycogen** rather than starch. And both build important structural molecules from **chitin** — in fungi, it forms the cell wall; in animals, it forms the exoskeletons of insects and crustaceans.

This evolutionary closeness has a practical consequence that matters enormously for medicine: because fungal cells are biochemically similar to animal cells, it is much harder to develop drugs that kill fungi without also harming the patient. Antibacterial drugs exploit the many differences between bacterial and human cells — different ribosomes, different cell wall chemistry, different membrane structures. Antifungal drugs have far fewer differences to exploit. This is one of the reasons that fungal infections remain disproportionately difficult to treat, particularly in immunocompromised patients, and why the repertoire of clinically available antifungal drugs is so much smaller than the antibiotic arsenal.

What Is a Fungus?

Defining "fungus" with a crisp one-sentence definition is harder than it sounds, because the kingdom encompasses an extraordinary range of forms. But most true fungi share a recognisable set of features.

They are eukaryotes. Like animal and plant cells, fungal cells have a nucleus enclosed by a membrane, along with other membrane-bound organelles including mitochondria and an endoplasmic reticulum. This places them squarely in the domain Eukarya, alongside all the other complex-celled organisms, and distinguishes them from the bacteria and archaea that dominate the rest of the microbiome.

They have cell walls made of chitin and glucan. This is one of the defining chemical signatures of the kingdom. Plant cell walls are built from **cellulose**; bacterial cell walls are built from **peptidoglycan**. Fungal cell walls are built from a scaffold of **chitin** (a tough polymer of *N*-acetylglucosamine — the same molecule found in crab shells) reinforced with **β-glucans** (polymers of glucose with characteristic β-1,3 linkages) and coated with a layer of **mannoproteins** [REF:gow2017]. The cell wall typically accounts for about 40 per cent of the total cell volume — it is not a thin skin but a massive, load-bearing structure that protects the cell from osmotic stress and environmental insult.

Think of it as a building analogy: the chitin fibres are the steel reinforcing bars; the β-glucan matrix is the concrete that surrounds them; and the mannoprotein coat is the painted plaster on the outside. Together, they create a structure that is rigid enough to maintain shape, yet flexible enough to allow growth.

They are heterotrophs. Fungi cannot make their own food. They obtain carbon and energy by secreting digestive enzymes into their environment and then absorbing the resulting small molecules across their cell membranes — a strategy called **absorptive nutrition** (sometimes called **osmotrophy**). In effect, fungi digest their food externally before absorbing it, which is the reverse of what animals do: we swallow first and digest later. A useful analogy is to imagine dissolving your meal in acid outside your body and then soaking up the nutrients through your skin.

Their membranes contain ergosterol. Animal cell membranes use **cholesterol** to regulate membrane fluidity and stability. Fungal membranes use a related but structurally distinct molecule called **ergosterol** [REF:naranjo-ortiz2019]. This difference is one of the few biochemical handles that antifungal drugs can grip: the azole class of antifungals (fluconazole, voriconazole, itraconazole) works by blocking the enzyme that synthesises ergosterol, destabilising the fungal membrane while leaving human cholesterol production largely intact. The polyene antifungal amphotericin B goes further — it binds directly to ergosterol molecules in the membrane, punching holes that cause the cell to leak and die. The specificity is imperfect, however, which is why amphotericin B is notorious among clinicians for its toxicity: at high doses, it can also interact with human cholesterol, damaging kidney and liver cells. The old clinical nickname for the drug — “amphoterrible” — reflects this narrow therapeutic window.

Shape-Shifters: Yeasts, Moulds, and Mushrooms

Fungi come in three basic body plans, but many species can switch between them depending on environmental conditions — a flexibility that is central to their success, both as ecological recyclers and as human pathogens.

Yeasts

Yeasts are single-celled fungi. Each cell is a small, rounded or oval unit, typically 3 to 10 micrometres in diameter — larger than most bacteria, but still microscopic. Yeasts reproduce primarily by **budding**: a small daughter cell grows as an outgrowth from the mother cell, gradually enlarging until it pinches off and becomes independent. Some yeasts can also reproduce sexually, forming spores under specific environmental conditions.

The most familiar yeast is *Saccharomyces cerevisiae* — baker’s yeast, brewer’s yeast — the organism that gives bread its rise and beer its alcohol. It was also the first eukaryotic organism to have its entire genome sequenced, in 1996, making it one of the most thoroughly studied organisms in biology [REF:goffeau1996]. But in the context of the human microbiome, the yeasts that matter most are different species: *Candida albicans*, a commensal resident of the gut and mucous membranes that can become an aggressive pathogen when the immune system falters; *Malassezia*, the dominant fungal genus on human skin; and *Saccharomyces boulardii*, a yeast that has been studied as a probiotic — though as we will discuss in the main book, the evidence base for fungal probiotics is considerably thinner than the marketing suggests.

Moulds

A mould is a fungus that grows as a network of branching, thread-like filaments called **hyphae** (singular: **hypha**). Each hypha is a cylindrical tube, typically only a few micrometres wide, enclosed by a rigid cell wall. As the hypha grows — always at its tip — it extends into new territory, secreting enzymes ahead of itself to digest whatever substrate it is growing on. When many hyphae branch and interweave, they form a tangled mat called a **mycelium** (plural: **mycelia**), which is the main body of the organism.

A mycelium can be enormous. The largest known organism on Earth is a single individual of *Armillaria solidipes* (the honey fungus) in Oregon’s Blue Mountains, whose mycelium extends over roughly 9 square kilometres and is estimated to be several thousand years old. What you see when you spot a mushroom in the forest is not the organism itself — it is a temporary reproductive structure, like a flower on a plant. The real organism is the vast, hidden mycelium beneath the soil.

In the human body, the moulds of greatest medical importance include *Aspergillus* species (particularly *Aspergillus fumigatus*, a major cause of invasive pulmonary aspergillosis in immunocompromised patients), *Mucor* and *Rhizopus* (agents of the devastating infection mucormycosis, which gained worldwide attention during the COVID-19 pandemic), and various *Fusarium* species.

Mushrooms

Mushrooms are the large, visible fruiting bodies produced by certain fungi — specifically, by members of the phyla Basidiomycota and some Ascomycota — for the purpose of sexual reproduction and spore dispersal. The cap-and-stem structure that most people picture when they think of a mushroom is essentially a sophisticated spore-launching platform. Some mushrooms release billions of spores per day, carried by air currents to colonise new habitats.

Mushrooms are not significant members of the human microbiome. We mention them here only to complete the picture of fungal diversity and to make a point: the kingdom Fungi is much more than the moulds and yeasts that colonise the human body. It includes organisms that range from single-celled yeasts a few micrometres across to mycelial networks that span kilometres — a range of scale that rivals anything in the animal or plant kingdoms.

Dimorphism: The Switch

Some of the most clinically important fungi are **dimorphic** — they can switch between a yeast form and a mould form depending on their environment. *Candida albicans*, the most common fungal pathogen of humans, is a textbook example. In its yeast form, *C. albicans* exists as harmless, rounded, budding cells on the mucosal surfaces of the gut, mouth, and vagina. But when conditions change — a shift in pH, a weakened immune response, a disruption of the competing bacterial community — it can switch to a hyphal form, extending invasive filaments that penetrate tissue, evade immune cells, and cause disease [REF:sudbery2004].

This morphological switching is not merely a change of shape. It is accompanied by changes in gene expression, surface proteins, and virulence factors. The hyphal form expresses adhesins that help it stick to human cells, secretes enzymes that degrade tissue, and forms **biofilms** — structured microbial communities encased in a self-produced matrix — on medical devices such as catheters and prosthetic heart valves. Understanding what triggers the yeast-to-hypha transition, and how to prevent it, remains one of the central questions in medical mycology.

How Fungi Feed: Decomposers, Mutualists, and Parasites

The way an organism obtains its food determines its ecological role. Fungi fall into three broad nutritional strategies, and understanding these helps explain why they are found in almost every habitat on Earth — including the human body.

Saprotrophs: The Recyclers

Most fungi are **saprotrophs** (from the Greek *sapros*, “rotten”) — they decompose dead organic matter. Fallen leaves, dead wood, animal carcasses, food scraps: saprotrophic fungi secrete cocktails of powerful enzymes — cellulases, ligninases, proteases, lipases — that break down complex organic molecules into simpler compounds that the fungus can absorb. Without saprotrophic fungi, dead plant material would accumulate indefinitely and the global carbon cycle would grind to a halt. Fungi are the only organisms on Earth that can efficiently decompose **lignin**, the tough structural polymer that gives wood its rigidity. Before fungi evolved this ability — roughly 300 million years ago — dead trees simply piled up rather than rotting. This is the period that gave us the vast coal deposits of the Carboniferous era.

Mutualists: The Partners

Some fungi form intimate partnerships with other organisms from which both partners benefit. The most ecologically important of these are the **mycorrhizae** (from the Greek *myco*, "fungus," and *rhiza*, "root") — associations between fungi and plant roots that have existed for at least 450 million years, since the earliest land plants colonised the terrestrial environment. In a mycorrhizal partnership, the fungal mycelium extends into the soil far beyond the reach of the plant's own roots, absorbing water and mineral nutrients (particularly phosphorus) and delivering them to the plant. In return, the plant provides the fungus with sugars produced by photosynthesis. The vast majority of land plants — roughly 90 per cent of species — form mycorrhizal associations, and many cannot survive without them.

Lichens are another form of mutualism: composite organisms formed by a fungus (usually an ascomycete) living in intimate association with a photosynthetic partner — either a green alga or a cyanobacterium. The fungus provides structure and protection; the photosynthetic partner provides food. Lichens can colonise bare rock, arctic tundra, and desert surfaces where neither partner could survive alone.

In the human body, the relationships between fungi and their host are more complex and ambiguous — often described as **commensalism** (the fungus benefits; the host is apparently unaffected) rather than true mutualism. Whether gut fungi actively benefit their human hosts — for example, by training the immune system or competing with pathogens — is an active area of research that we will examine in Chapter 10 of the main book.

Parasites and Pathogens

Some fungi obtain their nutrition from living hosts, causing disease in the process. Fungal pathogens infect plants, insects, amphibians, and mammals. In humans, fungal infections range from superficial and merely annoying (athlete's foot, ringworm, dandruff) to invasive and life-threatening (invasive aspergillosis, disseminated candidiasis, cryptococcal meningitis).

Globally, fungal diseases are estimated to kill more than 1.5 million people each year — a toll that rivals tuberculosis and exceeds malaria in some estimates — yet they receive a fraction of the research funding and public attention [REF:bongomin2017]. This neglect is slowly changing, in part because the COVID-19 pandemic brought fungal superinfections — particularly mucormycosis in India and invasive aspergillosis in intensive care patients — to sudden and devastating public visibility.

Fungal Reproduction: Spores, Sex, and Survival

Fungi reproduce by both asexual and sexual means, and many species alternate between the two depending on environmental conditions. Understanding the basics of fungal reproduction helps explain why fungi are so successful at colonising new environments, including the human body.

Asexual Reproduction

Most fungi can reproduce asexually — that is, without the genetic mixing that comes from sex. The simplest method is **budding**, as seen in yeasts: a new cell simply grows from the surface of an existing one. In filamentous fungi, asexual reproduction more commonly involves the production of **spores** — tiny, lightweight cells that are released into the environment and can survive hostile conditions until they find a suitable place to germinate and grow.

Asexual spores come in various forms. *Aspergillus* produces chains of spores called **conidia** (singular: **conidium**) at the tips of specialised structures, releasing thousands of them into the air. A single *Aspergillus fumigatus* colony can produce billions of conidia, and every human being inhales several hundred of them each day. In healthy individuals with intact immune systems, these spores are cleared by alveolar macrophages in the lungs without incident. In patients whose immune systems are compromised — by chemotherapy, organ transplantation, high-dose corticosteroids, or advanced HIV — the spores can germinate in lung tissue, producing invasive hyphae that destroy the surrounding parenchyma.

Sexual Reproduction

Sexual reproduction in fungi involves the fusion of genetic material from two compatible individuals, producing offspring with new genetic combinations. The details vary enormously across the kingdom, but the general pattern involves three steps: first, the fusion of two cells (**plasmogamy**); then, the fusion of their nuclei (**karyogamy**); and finally, **meiosis** — the reductive cell division that shuffles the genetic deck and produces haploid spores.

For the non-specialist reader, the key takeaway is this: sexual reproduction generates genetic diversity, which helps fungal populations adapt to new challenges — including, in the clinical setting, the selective pressure of antifungal drugs. Drug-resistant strains of *Candida auris*, which has emerged as a globally significant pathogen in the last decade, owe some of their adaptability to mechanisms of genetic exchange and recombination [REF:lockhart2017].

The Major Phyla: A Brief Tour

The kingdom Fungi is currently divided into approximately 19 phyla, a number that continues to shift as molecular phylogenetics reveals new relationships [REF:tedersoo2018]. For the purposes of this primer, we will focus on the groups most relevant to human biology.

Ascomycota — The Sac Fungi

The Ascomycota is the largest phylum of fungi, containing roughly 64,000 described species — more than half of all named fungi. The name comes from the **ascus** (plural: **asci**), a sac-like structure in which sexual spores are produced. This is an enormously diverse group. It includes yeasts (*Saccharomyces*, *Candida*, *Pichia*), moulds (*Aspergillus*, *Penicillium*, *Fusarium*), the organisms that produce truffles and morels, and most lichen-forming fungi.

For the human microbiome, the Ascomycota is the most important phylum. *Candida albicans* and its relatives, *Aspergillus fumigatus*, and *Pneumocystis jirovecii* (the cause of a devastating pneumonia in AIDS patients — itself an organism that was misclassified as a protist until molecular studies placed it firmly among the ascomycetes) are all members of this group.

Basidiomycota — The Club Fungi

The Basidiomycota contains roughly 32,000 described species and includes most of the organisms that produce conspicuous mushrooms, bracket fungi, and puffballs. The name comes from the

basidium (plural: **basidia**), a club-shaped structure on which sexual spores are produced.

In the human microbiome, the most significant basidiomycete is *Malassezia* — a genus of lipid-dependent yeasts that dominate the fungal community on human skin. *Malassezia* is associated with seborrhoeic dermatitis, dandruff, pityriasis versicolor, and may play a role in atopic eczema. Despite being one of the most abundant organisms on the human body surface, *Malassezia* was historically difficult to study because its strict requirement for exogenous lipids made it hard to grow in standard laboratory culture media — another example of how technical limitations have shaped what we know about the microbiome. The genus *Cryptococcus*, which causes cryptococcal meningitis (a leading killer of people with advanced HIV infection), is also a basidiomycete.

The "Basal" Lineages

Below the Ascomycota and Basidiomycota (which together form a clade called the **Dikarya**, meaning "two nuclei," because their cells often contain two genetically distinct nuclei for a period before sexual fusion), there are a number of earlier-diverging fungal lineages. These include the **Mucoromycota** (which contains the agents of mucormycosis — *Mucor*, *Rhizopus*, *Rhizomucor*) and the **Chytridiomycota** (aquatic fungi that retain the ancestral posterior flagellum). These groups are less frequently encountered in the human microbiome but are not absent from it, and the mucormycosis agents can cause catastrophically aggressive infections in diabetic and immunocompromised patients.

The Cell Wall as a Battleground

We return to the fungal cell wall — not because it is architecturally interesting (though it is) but because it is the primary interface between the fungus and the human immune system, and the principal target of most antifungal drugs. Understanding its structure helps explain both how our bodies detect fungal invaders and why treating fungal infections is so difficult.

The immune system detects fungi primarily through **pattern recognition receptors** on the surface of innate immune cells — macrophages, dendritic cells, and neutrophils. The most important of these, for fungi, is a receptor called **Dectin-1**, which recognises the β-1,3-glucan that forms the structural core of the fungal cell wall [REF:brown2006]. When Dectin-1 binds β-glucan, it triggers a cascade of inflammatory signals that recruit more immune cells and activate antifungal killing mechanisms. A related receptor, **Dectin-2**, recognises mannan structures on the outer surface of the wall.

But here is the catch: in many pathogenic fungi, the β-glucan layer is hidden beneath the outer coat of mannoproteins. This **masking** means that the immune system's primary sensor cannot see the target. *Candida albicans*, for instance, actively remodels its cell wall surface to conceal its β-glucan from Dectin-1, and only certain conditions — damage to the wall, exposure to antifungal drugs, or the probing of neutrophil enzymes — strip away the masking layer and expose the glucan underneath. This is a form of immune evasion, and it explains, in part, why some fungal infections are so difficult for the body to control.

The clinical classes of antifungal drugs map neatly onto the cell wall and membrane:

- **Azoles** inhibit lanosterol 14α-demethylase, a key enzyme in the ergosterol biosynthesis pathway, starving the membrane of its essential sterol. This is a large class: it includes over-the-counter topical agents that many readers will have used — clotrimazole and miconazole for thrush or athlete's foot — as well as the systemic agents fluconazole, voriconazole, itraconazole, and posaconazole that are used for invasive infections.

- **Allylamines** (most notably **terbinafine**, widely sold as Lamisil) block a different enzyme in the same ergosterol pathway — **squalene epoxidase** — which catalyses an earlier step in sterol synthesis. The result is twofold: ergosterol levels fall, and the precursor squalene accumulates to toxic concentrations within the cell. Terbinafine is probably the antifungal most familiar to the general public, used extensively for fungal nail infections and ringworm.
- **Polyenes** (amphotericin B, nystatin) bind directly to ergosterol molecules already in the membrane, forming pores that cause the cell to leak and die.
- **Echinocandins** (caspofungin, micafungin, anidulafungin) inhibit the enzyme β-1,3-glucan synthase, dismantling the structural scaffold of the cell wall.

Notice that three of these four classes — azoles, allylamines, and polyenes — all converge on the same molecular target: ergosterol. They simply attack it at different points (synthesis versus the finished molecule in the membrane). Only the echinocandins target something entirely different (the cell wall glucan). That is essentially the entire clinical toolkit for systemic and serious fungal infections — four classes of drugs, but really only two molecular targets (ergosterol and β-glucan). Compare this to the dozens of antibiotic classes available for bacterial infections, and the vulnerability of the antifungal arsenal becomes starkly apparent. The recent emergence of *Candida auris* — a multidrug-resistant yeast first identified in 2009 that can resist all three classes simultaneously — has been described by public health agencies as a serious and urgent threat [REF:lockhart2017].

How Many Fungi? And How Little We Know

How many fungal species exist on Earth? As with bacteria, the honest answer is that we do not know. Current estimates range from 2.2 to 3.8 million species, with some analyses using environmental DNA data suggesting the number could be as high as 12 million [REF:hawksworth2017]. As of 2024, roughly 155,000 species have been formally described — meaning that somewhere between 92 and 97 per cent of all fungal species have never been given a name [REF:hawksworth2017].

This matters for the human microbiome because when researchers sequence fungal DNA from human body sites using ITS amplicon sequencing (the fungal equivalent of the bacterial 16S approach, as discussed in Chapter 4 of the main book), a substantial proportion of the sequences they recover cannot be confidently matched to any known species. The fungal reference databases — primarily **UNITE** (a curated database of ITS sequences) — are less complete than their bacterial counterparts, and the gap is closing slowly. The human mycobiome is still, in many respects, a partially mapped territory.

Why This Matters for the Microbiome Story

The key messages of this primer chapter — the ones to carry forward into the main book — are these:

Fungi are eukaryotes, not bacteria. They are vastly more complex at the cellular level than the bacteria that dominate most microbiome discussions. They have nuclei, organelles, and genomes that are orders of magnitude larger. They are, in evolutionary terms, our distant relatives.

Fungi are biochemically similar to human cells. This makes them hard to kill selectively — a fact that shapes every aspect of antifungal medicine.

The fungal cell wall is both a shield and a signal. Its β-glucan and chitin content defines the kingdom, provides the target for the major antifungal drugs, and serves as the primary molecular pattern by which the immune system detects fungal presence.

Most fungal diversity is undescribed. Our ability to identify and understand the fungi in the human body is limited by reference databases that are still under construction.

Fungi are a minority by numbers but not by influence. In most human body sites, fungi represent less than 0.1 per cent of the total microbial community by cell count. But their cells are much larger than bacterial cells, they interact with bacteria through cross-kingdom signalling, and they expand rapidly into niches vacated by bacteria during antibiotic therapy. Their influence on human health is disproportionate to their census numbers — a theme we will return to, at length, in Chapter 10.

In the next primer chapter, we will turn to the viruses — entities that are not cells at all, yet profoundly shape every microbial community in and on the human body.

References Cited in This Chapter

- [REF:baldauf1993] Baldauf, S.L. & Palmer, J.D. (1993). Animals and fungi are each other's closest relatives: congruent evidence from multiple proteins. *Proceedings of the National Academy of Sciences*, 90(24), 11558–11562. DOI: 10.1073/pnas.90.24.11558
- [REF:cavalier-smith1987] Cavalier-Smith, T. (1987). The origin of Fungi and pseudofungi. In: Rayner, A.D.M., Brasier, C.M. & Moore, D. (eds) *Evolutionary Biology of the Fungi*. Cambridge University Press, pp. 339–353.
- [REF:parfrey2011] Parfrey, L.W., Lahr, D.J.G., Knoll, A.H. & Katz, L.A. (2011). Estimating the timing of early eukaryotic diversification with multigene molecular clocks. *Proceedings of the National Academy of Sciences*, 108(33), 13624–13629. DOI: 10.1073/pnas.1110633108
- [REF:gow2017] Gow, N.A.R., Latge, J.-P. & Munro, C.A. (2017). The fungal cell wall: structure, biosynthesis, and function. *Microbiology Spectrum*, 5(3), FUNK-0035-2016. DOI: 10.1128/microbiolspec.FUNK-0035-2016
- [REF:naranjo-ortiz2019] Naranjo-Ortiz, M.A. & Gabaldón, T. (2019). Fungal evolution: diversity, taxonomy and phylogeny of the Fungi. *Biological Reviews*, 94(6), 2101–2137. DOI: 10.1111/brv.12550
- [REF:goffeau1996] Goffeau, A. et al. (1996). Life with 6000 genes. *Science*, 274(5287), 546–567. DOI: 10.1126/science.274.5287.546
- [REF:sudbery2004] Sudbery, P., Gow, N. & Berman, J. (2004). The distinct morphogenic states of *Candida albicans*. *Trends in Microbiology*, 12(7), 317–324. DOI: 10.1016/j.tim.2004.05.008

- [REF:bongomin2017] Bongomin, F., Gago, S., Oladele, R.O. & Denning, D.W. (2017). Global and multi-national prevalence of fungal diseases — estimate precision. *Journal of Fungi*, 3(4), 57. DOI: 10.3390/jof3040057
- [REF:lockhart2017] Lockhart, S.R. et al. (2017). Simultaneous emergence of multidrug-resistant *Candida auris* on 3 continents confirmed by whole-genome sequencing and epidemiological analyses. *Clinical Infectious Diseases*, 64(2), 134–140. DOI: 10.1093/cid/ciw691
- [REF:tedersoo2018] Tedersoo, L. et al. (2018). High-level classification of the Fungi and a tool for evolutionary ecological analyses. *Fungal Diversity*, 90, 135–159. DOI: 10.1007/s13225-018-0401-0
- [REF:hawksworth2017] Hawksworth, D.L. & Lücking, R. (2017). Fungal diversity revisited: 2.2 to 3.8 million species. *Microbiology Spectrum*, 5(4), FUNK-0052-2016. DOI: 10.1128/microbiolspec.FUNK-0052-2016
- [REF:brown2006] Brown, G.D. (2006). Dectin-1: a signalling non-TLR pattern-recognition receptor. *Nature Reviews Immunology*, 6(1), 33–43. DOI: 10.1038/nri1745
- [REF:li2021] Li, Y. et al. (2021). A genome-scale phylogeny of the kingdom Fungi. *Current Biology*, 31(8), 1653–1665.e5. DOI: 10.1016/j.cub.2021.01.074

Primer Chapter 5: Viruses and Their Strange Relatives

This chapter is part of the companion primer to The Inhabited Body. *It explores the biology of viruses in depth — with particular attention to bacteriophages, the viruses that infect bacteria. Phages are among the most important and least appreciated players in the human microbiome. This chapter also covers retroviruses, the concept of the virome, and the recently discovered obelisks.*

The Most Abundant Entities on Earth

If biology were a popularity contest, viruses would win by a margin so vast it defies comprehension. The estimated number of viral particles on Earth — roughly 10^{31}, or ten million trillion trillion — exceeds the combined total of every other biological entity [REF:suttle2007]. There are more viruses in a litre of seawater than there are people on the planet. There are more viruses in the human gut alone than there are stars in the Milky Way.

And yet, as we noted in Primer Chapter 1, viruses are not "alive" by most standard definitions. They have no metabolism. They cannot grow. They cannot reproduce on their own. A virus is a set of instructions — a genome made of DNA or RNA — wrapped in a protein shell. It is a biological message in a bottle, inert and purposeless until it reaches the right cell. Then it becomes something else entirely: a hijacker of extraordinary efficiency, commandeering the cell's own machinery to produce copies of itself.

In Primer Chapter 1, we introduced viruses briefly and placed them on the map of life alongside bacteria, archaea, and eukaryotes. Now it is time to look inside the bottle.

Anatomy of a Virus

Despite their dizzying diversity — there are viruses that infect every domain of cellular life, and probably every species — all viruses share a few basic structural features.

The Genome

Every virus has a genome: a set of genetic instructions encoded in nucleic acid. But unlike cells, which invariably use double-stranded DNA, viruses have explored almost every possible nucleic acid configuration. Some carry double-stranded DNA (like the herpes simplex virus). Some carry single-stranded DNA (like the parvoviruses). Some carry double-stranded RNA (like the rotaviruses). And some carry single-stranded RNA, which can be "positive-sense" (ready to be read directly by the host's ribosomes, like SARS-CoV-2) or "negative-sense" (needing to be copied into a complementary strand first, like the influenza virus).

This variety of genome types is one reason viruses are so diverse and so difficult to classify. In 1971, the virologist David Baltimore proposed a classification system based not on what viruses look like but on how they handle their genetic information — specifically, how they produce messenger RNA. The **Baltimore classification** groups all viruses into seven classes based on their genome type and replication strategy. It remains one of the most widely used frameworks in virology.

The Capsid

Surrounding the genome is the **capsid** — a shell made entirely of protein. Capsid proteins are typically encoded by the virus's own genome and self-assemble into one of a few highly symmetrical shapes.

Think of virus shapes like packaging options. Some viruses are **icosahedral** — roughly spherical, with twenty flat triangular faces, like a microscopic football. This is an extraordinarily efficient shape: it encloses the maximum volume for the minimum number of protein subunits. Many common viruses, from those that cause the common cold (rhinoviruses) to those that cause polio, use this design.

Other viruses are **helical** — their capsid proteins wrap around the genome in a spiral, producing a tube-like or rod-like structure. The tobacco mosaic virus, the first virus ever identified (in 1892 by Dmitri Ivanovsky), is a classic example.

Then there are the **complex** viruses, which combine elements of both designs or adopt entirely unique architectures. The most striking examples are the bacteriophages — the viruses that infect bacteria — many of which look like nothing else in biology: an icosahedral head sitting on top of a cylindrical tail, with spidery tail fibres splayed out at the base, like a lunar lander built from proteins. This structure is not decorative. It is a precision injection device, engineered by billions of years of evolution to dock onto a bacterial surface and deliver a viral genome with mechanical efficiency.

The Envelope

Some viruses — but not all — surround their capsid with an additional outer layer called the **envelope**. The envelope is a lipid bilayer, similar to a cell membrane, that the virus acquires by budding through the membrane of its host cell on the way out. Studded in this stolen membrane are viral proteins — often glycoproteins — that the virus uses to recognise and bind to new host cells.

The **spike protein** of SARS-CoV-2, which became famous during the COVID-19 pandemic, is one such envelope protein. It is the molecular key that the virus uses to unlock the ACE2 receptor on human cells.

Enveloped viruses tend to be more fragile outside the body than non-enveloped ones, because their lipid membrane is easily disrupted by soap, alcohol, and drying. This is why washing your hands with soap is effective against influenza and coronaviruses (both enveloped) but less so against norovirus (non-enveloped).

The Viral Life Cycle: Two Strategies

Once a virus has found its target cell, it faces a fundamental choice — one that has profound consequences for the microbiome.

The Lytic Cycle: Smash and Grab

In the **lytic cycle**, the virus is an unambiguous predator. It attaches to the cell surface, injects (or otherwise delivers) its genome, and immediately takes over the cell's machinery. The host cell becomes a virus factory, churning out copies of the viral genome and viral proteins. These components assemble into new virions inside the cell. When enough new virions have accumulated, the cell is lysed — broken open — releasing a burst of progeny viruses into the environment. The host cell is destroyed.

The analogy here is a factory raid. The virus breaks in, fires the management, reprograms the production lines to manufacture copies of itself, and then demolishes the building on the way out.

The lytic cycle is fast and violent. A single bacteriophage infecting an *E. coli* cell can produce 100 to 200 new phage particles in as little as 20 to 30 minutes. Multiply this across the trillions of phage infections happening every day in your gut, and you begin to appreciate the scale of viral predation within the microbiome.

The Lysogenic Cycle: The Sleeper Agent

The second strategy is subtler and, in many ways, more interesting. In the **lysogenic cycle**, the virus does not immediately destroy its host. Instead, it integrates its genome into the host cell's DNA and goes quiet.

The integrated viral genome — called a **prophage** in bacteria or a **provirus** in other organisms — is replicated along with the host's own DNA every time the cell divides. The host cell is unharmed. It may not even "know" it is infected. The viral genome rides along as a silent passenger, copied faithfully from one generation to the next, potentially for thousands of bacterial generations.

Think of this as a sleeper agent embedded in a foreign government. The agent lives a normal life, does normal work, raises no suspicion. But the instructions for a mission are still encoded, waiting. Under certain conditions — ultraviolet light, DNA damage, starvation, exposure to certain chemicals — the prophage can reactivate, excise itself from the bacterial chromosome, and switch to the lytic cycle, producing new virions and killing the host cell.

This switch from lysogeny to lysis is called **induction**, and it is not random. Many of the triggers that induce prophages are signals of stress — conditions in which the host bacterium is likely to die anyway. From the virus's perspective, it makes sense to abandon a sinking ship and seek a new host.

Why Lysogeny Matters for the Microbiome

Lysogeny is not merely a curiosity of virology. It is central to the microbiome story, for three reasons.

First, **prophages are everywhere.** Surveys of bacterial genomes have revealed that the majority of sequenced bacteria carry at least one prophage, and many carry several. Your gut bacteria are riddled with dormant viral genomes. Some of these prophages have been silent for so long that they have decayed — accumulated mutations that prevent them from ever reactivating — and are now permanent fixtures of their host's genome, indistinguishable from "normal" bacterial DNA except by careful sequence analysis.

Second, **prophages can change what their host does.** When a virus integrates into a bacterium's chromosome, it sometimes brings genes that have nothing to do with viral replication — genes that can give the host new capabilities. This is called **lysogenic conversion**, and its consequences can be dramatic. The cholera toxin, which causes the devastating diarrhoea of cholera, is encoded not by *Vibrio cholerae* itself but by a prophage (CTXφ) integrated into its genome. Without the phage, the bacterium is harmless. Similarly, the Shiga toxin produced by pathogenic strains of *E. coli* is carried by a prophage. In a very real sense, some of the most dangerous bacterial toxins are viral inventions.

Third, **prophage induction reshapes microbial communities.** When environmental conditions change — a shift in diet, a course of antibiotics, an inflammatory flare — dormant prophages across the gut can be induced simultaneously, triggering a wave of bacterial lysis. This mass killing event releases a flood of bacterial DNA, cell contents, and new phage particles into the gut environment, altering the composition of the microbial community. Researchers are beginning to suspect that prophage induction may be one of the mechanisms through which antibiotics cause lasting disruption to the microbiome — not just by directly killing bacteria, but by activating the viral time bombs already embedded in their genomes.

Bacteriophages: The Invisible Regulators

Bacteriophages — viruses that infect bacteria — are the most abundant biological entities in the human body. They outnumber the bacteria they infect by an estimated ratio of roughly ten to one. Yet until recently, they were almost entirely overlooked in microbiome research, which tended to focus exclusively on bacteria [REF:shkoporov2019].

The reason for this neglect is partly technical. Standard metagenomics pipelines (introduced in Primer Chapter 6 and discussed in detail in Chapter 4 of the main book) are designed to capture bacterial DNA. Phages, with their small genomes and high mutation rates, often slip through the cracks. Many phage sequences in metagenomic datasets are classified as "unknown" or "uncharacterised" — dark matter in the microbial universe.

But the picture is changing rapidly. Advances in viral metagenomics have revealed an astonishing diversity of phages in the human gut — a community now referred to as the **phageome** (or, more broadly, the gut **virome**, which includes all viruses, not just phages). Some findings:

The human gut virome is **highly individual**. No two people share the same phage community, even identical twins. Your phageome is as unique as your fingerprint — arguably more so.

The phageome is **relatively stable over time** within an individual, despite the rapid turnover of individual phage particles. This suggests that phages and their bacterial hosts co-exist in a dynamic equilibrium — a perpetual arms race in which neither side wins outright.

Phages shape bacterial community structure through what ecologists call **"kill the winner" dynamics**. When a particular bacterial species becomes very abundant (the "winner"), it becomes a more visible target for phages that specialise in infecting that species. The phages multiply, drive down the winner's numbers, and create space for other bacterial species to expand. This prevents any single species from dominating and helps maintain the diversity of the microbiome. It is a biological thermostat, keeping the ecosystem in balance.

Retroviruses: Information Flowing Backwards

Most viruses follow the standard flow of genetic information described in Primer Chapter 3: DNA ☐ RNA ☐ protein. But one group of viruses breaks this rule in a spectacular way.

Retroviruses carry their genome as single-stranded RNA. When a retrovirus infects a cell, it uses a special enzyme called **reverse transcriptase** to convert its RNA genome into DNA — the reverse of the normal flow. This viral DNA is then integrated into the host cell's genome by another viral enzyme called **integrase**. Once integrated, the viral DNA (now called a **provirus**) is transcribed and translated using the host cell's own machinery, producing new viral RNA genomes and viral proteins that assemble into new virions.

The discovery of reverse transcriptase in 1970, independently by David Baltimore and by Howard Temin and Satoshi Mizutani, was a landmark in molecular biology [REF:baltimore1970] [REF:temin1970]. It overturned the assumption that genetic information could only flow from DNA to RNA, never the reverse. Both groups shared the Nobel Prize in 1975.

The most infamous retrovirus is **HIV** (human immunodeficiency virus), the cause of AIDS. HIV specifically infects a type of immune cell called the **CD4+ T cell** — a critical coordinator of the immune response. By destroying these cells over years, HIV progressively dismantles the immune system, leaving the body vulnerable to infections that a healthy immune system would normally control [REF:barresinoussi1983].

HIV illustrates the insidiousness of the retroviral strategy. Because the provirus is integrated into the host cell's DNA, it is copied every time the cell divides. It becomes a permanent part of that cell's genome. This is why HIV cannot be cured by conventional antiviral drugs: even if every active virus particle in the body is suppressed, the proviral DNA persists in long-lived "reservoir" cells, ready to reactivate if treatment is stopped.

Endogenous Retroviruses: Your Viral Ancestry

As we mentioned briefly in Primer Chapter 1, retroviral integration is not always a dead end. If a retrovirus happens to infect a germ cell — a sperm or egg — its proviral DNA can be passed to the next generation. And the next. And the next. Over millions of years, the human genome has accumulated thousands of such insertions. These ancient sequences are called **human endogenous retroviruses** (HERVs), and they make up roughly 8 per cent of the human genome — far more than the 1.5 per cent that codes for proteins.

Most HERVs are fossils: mutated beyond recognition, incapable of producing functional viruses. But some fragments have been domesticated — co-opted by the human genome for its own purposes. The protein **syncytin**, derived from an ancient retroviral envelope gene, is essential for the formation of the human placenta. Without this repurposed viral protein, human pregnancy as we know it would not be possible.

This is a profound twist in the story of host and parasite. Over evolutionary time, the boundary between "our" DNA and "their" DNA has become thoroughly blurred. We carry viral genomes within us not as parasites but as integral components of our own biology.

The Virome: Thinking Beyond Bacteria

For most of the history of microbiome research, "microbiome" effectively meant "bacteriome" — the community of bacteria living in and on us. Viruses, fungi, archaea, and protists were afterthoughts. But as sequencing technologies have improved, it has become clear that the complete picture must include the **virome** — the total community of viruses associated with the human body.

The human virome includes:

Bacteriophages — by far the largest component, numbering in the trillions and targeting the bacteria of the microbiome.

Eukaryotic viruses — viruses that infect human cells. Many of these are present in healthy people without causing symptoms. Torque teno virus (TTV), for example, is found in more than 90 per cent of the human population and appears to cause no disease. Some researchers speculate that these persistent, seemingly benign viral passengers may actually modulate the immune system in ways we do not yet understand.

Endogenous viruses — the ancient retroviral sequences embedded in the human genome, discussed above.

Plant-derived and dietary viruses — sequences from viruses that infect the foods we eat, which pass through the gut without infecting human cells but can still interact with the immune system.

The virome is not static. It changes with diet, age, geography, and health status. Alterations in the virome have been associated with inflammatory bowel disease, type 1 diabetes, and malnutrition, though in most cases it remains unclear whether the viral changes are a cause or a consequence of disease. Untangling cause from correlation is one of the major challenges facing virome research — and microbiome science more broadly.

Phage Therapy: An Old Idea Whose Time May Have Come

The idea of using phages to treat bacterial infections is almost as old as the discovery of phages themselves. Félix d'Hérelle, who coined the term "bacteriophage" in 1917, began experimenting with phage therapy almost immediately. In the decades before antibiotics became widely available, phage therapy was practised in parts of Europe and the Soviet Union, particularly at the Eliava Institute in Tbilisi, Georgia, where it continues to this day.

The rise of antibiotics in the 1940s pushed phage therapy to the margins of Western medicine. Antibiotics were easier to manufacture, store, and administer. They worked against a broad range of bacteria, while phages were highly specific — a single phage might infect only one strain of one species. This specificity, which seemed like a disadvantage in the age of antibiotics, is now being reconsidered as a strength.

The logic is straightforward. Broad-spectrum antibiotics are blunt instruments. They kill pathogenic bacteria, but they also devastate the beneficial microbiome, causing side effects ranging from diarrhoea to life-threatening *Clostridioides difficile* infection. Phages, by contrast, are precision weapons. A phage that targets pathogenic *E. coli* will leave *Bacteroides*, *Lactobacillus*, and the rest of the gut community untouched.

As antibiotic resistance continues to erode the effectiveness of our existing drugs, phage therapy is experiencing a renaissance. Clinical trials are underway for phage treatments of wound infections, urinary tract infections, and lung infections in cystic fibrosis patients. Regulatory frameworks are being developed. And new techniques for engineering phages — modifying their host range or enhancing their killing ability — are expanding the therapeutic toolkit.

We will discuss phage therapy in depth in Chapters 16–18 of the main book. For now, the key point is that the very same specificity that makes phages so important for regulating the natural microbiome also makes them promising tools for medicine — if we can learn to deploy them wisely.

Obelisks, Viroids, and the Unknown Unknowns

At the margins of virology — indeed, at the margins of biology itself — lie entities that are simpler than viruses, stranger than anything in a textbook, and, in some cases, discovered so recently that we have barely begun to understand them.

Viroids, introduced briefly in Primer Chapter 1, are tiny circular RNA molecules — typically just 250 to 400 nucleotides long — that infect plant cells and cause disease. They have no capsid, no envelope, no proteins of any kind. They are naked RNA: the smallest known infectious agents. Despite their simplicity, viroids can replicate inside host cells by hijacking the cell's own RNA polymerase. They were discovered in 1971 by Theodor Diener, and for decades they were thought to be exclusively a problem for plants — causing diseases in potatoes, coconut palms, avocados, and other crops.

Then came the **obelisks**. In 2024, Ivan Zheludev and colleagues reported the discovery of a new class of viroid-like RNA elements in the human gut microbiome [REF:zheludev2024]. These agents — which the researchers named obelisks for their predicted rod-shaped RNA secondary structure — are small circular RNA molecules of roughly 1,000 nucleotides. Like viroids, they have no protein coat. Unlike classical viroids, they encode a protein — a single, novel protein the researchers called "Oblin," which belongs to no known protein family.

Obelisks are not rare. They were detected in roughly 7 per cent of gut metatranscriptomes and in about 50 per cent of oral metatranscriptomes analysed by the team. Large-scale searches identified nearly 30,000 distinct obelisk sequences from samples spanning all seven continents. In at least one case, obelisks were shown to persist in an individual for more than 300 days, and one obelisk was traced to a specific bacterial host species, *Streptococcus sanguinis*, a common member of the oral microbiome.

What do obelisks *do*? We do not know. Whether they are parasites, commensals, or something else entirely remains an open question. They do not fit neatly into any existing category: they are not viruses (no capsid), not classical viroids (they encode a protein), and not plasmids (they are RNA, not DNA). They are, for now, a category unto themselves — a reminder that even in the most intensively studied environment on Earth (the human body), there remain biological entities that we have only just begun to notice.

Why This All Matters for the Microbiome

At first glance, a chapter on viruses might seem tangential to a book about the human microbiome. It is not. Viruses — particularly bacteriophages — are inseparable from the story of microbial communities.

Phages regulate bacterial populations through predation, keeping any single species from monopolising resources. They drive bacterial evolution by transferring genes between hosts — including genes for antibiotic resistance, toxin production, and metabolic capabilities. Prophages embedded in bacterial genomes can alter what bacteria do, turning harmless commensals into dangerous pathogens — or, occasionally, conferring advantages that help their hosts thrive.

The virome interacts with the immune system in ways we are only beginning to understand. Phage particles can cross the gut epithelium and enter the bloodstream, where they are recognised by immune cells. Phage DNA, once inside immune cells, can trigger innate immune responses through pattern recognition receptors — the same receptors discussed in the context of mRNA vaccines in Primer Chapter 3.

And at the very edge of our knowledge, entities like obelisks hint at an entire stratum of biological complexity that we have not yet mapped. The microbiome is not just bacteria. It is an ecosystem — and like all ecosystems, it includes predators, parasites, symbionts, and entities that defy easy classification.

Where This Matters in *The Inhabited Body*

- **Chapter 2** explores horizontal gene transfer, including the role of phage-mediated transduction in shuttling genes between bacterial species.
- **Chapter 4** discusses metagenomic and metatranscriptomic methods for studying the virome — and the technical challenges that have historically caused viral sequences to be overlooked.
- **Chapters 16–18** are devoted to bacteriophages and the microbiome: their ecological role, their impact on bacterial evolution, and the revival of phage therapy as a clinical strategy.
- **Chapter 12** examines how the immune system detects and responds to viral components, including phage DNA that crosses the gut epithelium.
- **Chapter 19** discusses how antibiotics disrupt the microbiome — a process in which prophage induction may play an underappreciated role.

Chapter References

- [REF:suttle2007] Suttle, C.A. (2007). Marine viruses — major players in the global ecosystem. *Nature Reviews Microbiology*, 5(10), 801–812. DOI: 10.1038/nrmicro1750

- [REF:shkoporov2019] Shkoporov, A.N. & Hill, C. (2019). Bacteriophages of the human gut: the “known unknown” of the microbiome. *Cell Host & Microbe*, 25(2), 195–209. DOI: 10.1016/j.chom.2019.01.017

- [REF:baltimore1970] Baltimore, D. (1970). RNA-dependent DNA polymerase in virions of RNA tumour viruses. *Nature*, 226(5252), 1209–1211. DOI: 10.1038/2261209a0

- [REF:temin1970] Temin, H.M. & Mizutani, S. (1970). RNA-dependent DNA polymerase in virions of Rous sarcoma virus. *Nature*, 226(5252), 1211–1213. DOI: 10.1038/2261211a0

- [REF:barresinoussi1983] Barré-Sinoussi, F. et al. (1983). Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science*, 220(4599), 868–871. DOI: 10.1126/science.6189183

- [REF:zheludev2024] Zheludev, I.N. et al. (2024). Viroid-like colonists of human microbiomes. *Cell*, 187(23), 6521–6536. DOI: 10.1016/j.cell.2024.09.033
- [REF:alberts2022] Alberts, B. et al. (2022). *Molecular Biology of the Cell*, 7th edition. New York: W.W. Norton.

Primer Chapter 6: How We Study What We Can't See

This chapter is part of the companion primer to The Inhabited Body. *It introduces the major techniques scientists use to study microbial communities — from the earliest microscopes to modern DNA sequencing and beyond. The goal is not to make you a microbiologist, but to give you the vocabulary and conceptual framework you need to follow the evidence presented in the main book — and to understand its limitations.*

The Invisible Majority

Everything we have covered in this primer so far — the three domains of life, the architecture of cells, the code written in DNA, the strange biology of fungi and viruses — has been building toward a practical question: how do scientists actually *study* these things?

For most of human history, the answer was: they could not. Bacteria are typically between 0.5 and 5 micrometres long. A micrometre is one-thousandth of a millimetre. Line up a thousand bacteria end to end, and they might stretch across the head of a pin. Viruses are smaller still — most are between 20 and 300 nanometres, which is to say, between twenty and three hundred *billionths* of a metre. You cannot see a bacterium with the naked eye, and you cannot see a virus even with a standard laboratory microscope. For the overwhelming majority of our species' existence, the microbial world was not merely unknown — it was unknowable.

The history of microbiome science is, fundamentally, a history of tools. Every major advance in our understanding of the microbial world has followed an advance in our ability to observe it. This chapter walks through those tools — not in exhaustive technical detail, but in enough depth that you understand how the claims in the main book are actually evidenced, and where those methods have blind spots.

Seeing the Unseen: Microscopy

The story begins with glass.

In the 1670s, a Dutch cloth merchant named Antonie van Leeuwenhoek began grinding tiny glass lenses — smaller than a raindrop, but polished to extraordinary precision — and mounting them in handheld brass frames. Using these simple, single-lens instruments, he achieved magnifications of up to 270 times and became the first human being to see bacteria. He called them *animalcules* — "little animals" — and reported their existence, with meticulous drawings, in a series of letters to the Royal Society of London [REF:lane2015].

Van Leeuwenhoek's microscopes were **light microscopes** (also called **optical microscopes**): they used visible light, focused through glass lenses, to magnify small objects. The principle remains the foundation of microscopy today, though modern light microscopes are vastly more sophisticated. A good research-grade light microscope can magnify up to about 1,000 times and resolve structures down to roughly 0.2 micrometres — enough to see individual bacteria and to distinguish their basic shapes (rods, spheres, spirals), but not enough to see their internal structures in any detail.

Think of it like binoculars. Binoculars let you see a bird on a distant branch — its shape, its colour, perhaps whether it has a crest or a long tail. But they will not show you the structure of its feathers, the pattern of blood vessels in its wing, or the food in its stomach. Light microscopy gives you the equivalent of a bird's silhouette. You can tell that something is *there*, and you can make out its rough shape, but the fine details remain hidden.

To see finer structures — the internal architecture of cells, the components of viruses — scientists use **electron microscopy**, which replaces light with beams of electrons. Because electrons have much shorter wavelengths than visible light, electron microscopes can achieve resolutions thousands of times higher. **Transmission electron microscopy** (TEM) fires electrons through an ultra-thin slice of a specimen and can resolve structures down to a few nanometres — enough to see individual proteins, the layers of a cell wall, or the intricate injection apparatus of a bacteriophage. **Scanning electron microscopy** (SEM) bounces electrons off the surface of a specimen, producing dramatic three-dimensional images — the iconic photographs of bacteria clinging to intestinal villi, or phages perched on the surface of a bacterial cell like lunar landers, are almost always SEM images.

The trade-off is that electron microscopy generally requires specimens to be dead, dehydrated, and coated in metal. You are looking at a snapshot, not a living process. And while microscopy can show you *that* microbes are present, it usually cannot tell you *which* microbes they are. A rod-shaped bacterium viewed under a microscope could be any one of thousands of species. To put names to faces, scientists needed different tools.

> **Where This Matters:** Chapter 4 of *The Inhabited Body* discusses how modern imaging techniques, including fluorescence microscopy with species-specific probes (FISH), are now being used to visualise the spatial organisation of microbial communities — revealing, for example, that the gut microbiome is not a random soup but a structured landscape where different species occupy distinct niches.

Growing the Invisible: Culture

If microscopy was the first breakthrough, the second was learning to *grow* microbes outside the body.

In the late nineteenth century, the German physician Robert Koch and his colleagues developed a set of techniques that would define microbiology for the next hundred years. They invented solid nutrient media — the familiar **agar plate**, a gel-filled dish on which bacteria can grow as visible colonies — and a rigorous experimental framework, now known as **Koch's postulates**, for proving that a specific microbe causes a specific disease [REF:blevins2010].

The logic of **culturing** is straightforward. You take a sample — a swab from a patient's throat, a drop of blood, a smear of soil — and you spread it on a nutrient medium. Each individual bacterium that can grow on that medium divides, again and again, until it forms a visible clump: a **colony**. Each colony is a clone — millions of genetically identical cells descended from a single ancestor. You can then pick that colony, transfer it to a fresh plate, and study it in isolation: stain it, look at it under a microscope, test which antibiotics kill it, analyse its metabolic capabilities.

This approach was spectacularly successful for identifying pathogens. The bacteria that cause tuberculosis, cholera, plague, diphtheria, and dozens of other infectious diseases were all identified using culture-based methods. Clinical microbiology laboratories in hospitals around the world — including, quite likely, the one that processes your doctor's samples — still rely on culture as a cornerstone of diagnosis.

But for studying *communities* of microbes — the diverse ecosystems that make up the microbiome — culture had a devastating blind spot.

The problem is that most microbes cannot be grown using standard laboratory techniques. A bacterium will only form a colony if the medium, the temperature, the atmosphere, and the incubation time are all within its tolerance range. Many gut bacteria are **obligate anaerobes** — they are killed by oxygen. Others require nutrients that standard media do not provide, or depend on chemical signals from neighbouring species, or grow so slowly that they are overgrown by faster-dividing competitors before they become visible.

In 1985, the microbiologists James Staley and Allan Konopka gave this problem a name: the **great plate count anomaly** [REF:staley1985]. When you examined an environmental sample under a microscope, you could count the cells directly. When you tried to grow them, only a tiny fraction — typically less than one per cent — would form colonies. The rest were invisible to culture.

The great plate count anomaly meant that, for most of the twentieth century, microbiologists could only study the minority of microbial species that happened to grow well in their laboratories. The majority of microbial diversity — including the majority of the species living inside the human body — was hidden. Not merely uncharacterised. *Unknown*.

The solution to this problem would come not from better culture media, but from a completely different approach: reading DNA.

> **Where This Matters:** Chapters 4 and 19 of *The Inhabited Body* explore the culture problem in detail — including the recent revival of culture through a technique called *culturomics*, which we introduce at the end of this chapter.

Molecular Barcodes: 16S rRNA Sequencing

In the late 1970s, Carl Woese and George Fox made a discovery that would reshape both our understanding of life and our ability to study it. By comparing the sequences of a gene called **16S ribosomal RNA** (16S rRNA) across a wide range of organisms, they discovered the three-domain tree of life we introduced in Primer Chapter 1 [REF:woese1977].

But the 16S rRNA gene turned out to have a second, even more practical significance. Because every bacterium and archaeon possesses this gene, and because it changes slowly enough to retain similarities between related species yet fast enough to distinguish between unrelated ones, the 16S rRNA gene functions as a **molecular barcode** — a universal identifier for prokaryotic life.

Think of it like an ISBN on a book. Every published book has an ISBN — a standardised code that identifies it uniquely. If someone handed you a torn page from an unknown book, you could not identify it. But if that page happened to contain the ISBN, you could look it up immediately. The 16S rRNA gene is the microbial world's ISBN. If you can read that gene from an organism — even from an organism you have never seen, never grown, and know nothing else about — you can identify it.

The technique works as follows. You take a sample — a gram of stool, a swab from the skin — and extract all the DNA it contains. You then use a laboratory method called the **polymerase chain reaction** (PCR) to make millions of copies of just the 16S rRNA genes in the sample, using short synthetic DNA sequences called **primers** that bind to the conserved (unchanging) regions of the gene. Finally, you **sequence** those copied genes — read out their nucleotide letters — and compare each sequence against a database of known 16S sequences to identify which organisms were present.

This is **16S rRNA gene sequencing**, also known as **amplicon sequencing**, and it revolutionised microbiology in the 2000s. For the first time, scientists could survey entire microbial communities without growing a single organism. The Human Microbiome Project — the landmark study we will encounter in Chapter 1 of the main book — used this technique to map the microbial communities at 18 body sites in 242 healthy adults, producing the first comprehensive atlas of the healthy human microbiome [REF:hmp2012].

But 16S sequencing has important limitations. It identifies organisms primarily at the genus level — it can tell you that a sample contains *Bacteroides*, but often cannot distinguish between closely related species. It only detects bacteria and archaea — not viruses (which have no ribosomes) or most fungi (whose ribosomal genes require different primers). And because it relies on PCR amplification, it can introduce biases: species whose 16S genes happen to amplify efficiently with the chosen primers will be overrepresented, while those with mismatches in the primer-binding sites may be underrepresented or missed entirely.

Most importantly, 16S sequencing tells you *who* is present but not *what they are doing*. Knowing that a sample contains a particular species of *Faecalibacterium* tells you nothing about which genes that species is expressing, which metabolites it is producing, or how it is interacting with its neighbours. For that, scientists needed to go beyond the barcode.

> **Where This Matters:** Chapter 4 of *The Inhabited Body* covers 16S sequencing in technical detail, including the shift from operational taxonomic units (OTUs) to amplicon sequence variants (ASVs), and the critical issue of how DNA extraction methods can bias community profiles.

Reading All the Genes: Metagenomics

If 16S sequencing reads a single gene from each organism in a sample, **metagenomics** reads *everything*.

Instead of targeting one specific gene, metagenomic sequencing takes all the DNA in a sample — from every bacterium, archaeon, fungus, virus, and human cell present — and chops it into millions of short fragments. These fragments are then sequenced and computationally reassembled, or at least matched against databases of known organisms and genes.

The analogy commonly used is a paper shredder. Imagine taking a library of a thousand different books — some in English, some in Mandarin, some in Arabic, some in languages you have never seen — shredding them all together, and then trying to figure out which books were in the pile and what they were about, using only the resulting strips of text. Metagenomic analysis does something similar, except the "books" are genomes and the "strips" are sequence reads of 150 to 300 DNA letters each.

The power of metagenomics is that it provides not just taxonomic identification but **functional information**. You can identify which metabolic genes are present in a community, which antibiotic resistance genes are circulating (the **resistome**), and which virulence factors are encoded. You can detect organisms at the species or even strain level. And you can find things that 16S sequencing misses entirely — viruses, novel organisms with no close relatives in databases, and genes of unknown function that may represent entirely new biology.

The landmark MetaHIT study, published in 2010, demonstrated this power dramatically. By sequencing all the DNA in stool samples from 124 European adults, Junjie Qin and colleagues assembled a catalogue of 3.3 million microbial genes — roughly 150 times the number of genes in the entire human genome [REF:qin2010].

The trade-offs are cost, complexity, and noise. Metagenomic sequencing generates enormous datasets that require significant computational infrastructure to analyse. Samples from the human body inevitably contain human DNA alongside microbial DNA — in a nasal swab, over 90 per cent of the DNA may be from the host, meaning you must sequence vastly more to capture the microbial fraction. And, like 16S sequencing, standard metagenomics reveals which genes are *present* but not necessarily which ones are *active*.

Beyond DNA: Metabolomics and Multi-Omics

DNA tells you what an organism *could* do. To understand what it *is* doing, you need to look at its outputs.

This insight has driven the development of several complementary approaches, collectively known as **multi-omics**.

Metatranscriptomics extracts and sequences the RNA — specifically, messenger RNA (mRNA) — from a microbial community. As we discussed in Primer Chapter 3, mRNA is produced only when a gene is being actively used. The metatranscriptome is therefore a snapshot of gene expression: not the parts list, but the set of instructions currently being executed. It has revealed, for example, that gene expression in the gut microbiome shifts dramatically in response to meals, medications, and even the time of day — even when the underlying DNA composition of the community barely changes.

Metabolomics takes an entirely different approach. Rather than reading nucleic acids, it uses analytical chemistry — typically **mass spectrometry** or **nuclear magnetic resonance (NMR) spectroscopy** — to identify and quantify the small molecules (**metabolites**) present in a sample. These metabolites are, in a sense, the final output of all the genomic activity: the short-chain fatty acids, bile acid derivatives, vitamins, and signalling molecules that the microbiome actually produces and that interact with the host body.

When you read in the main book that the gut microbiome "communicates with the brain" or "influences cardiovascular risk," it is largely through metabolites that this communication occurs. Butyrate, a short-chain fatty acid produced by bacterial fermentation of dietary fibre. Trimethylamine, produced from dietary choline and linked to heart disease. Indole derivatives, which modulate intestinal barrier function. Metabolomics is the tool that identifies and measures these molecules.

The most powerful modern studies combine multiple layers — metagenomics *and* metabolomics, for instance, or metatranscriptomics *and* proteomics — to build integrated pictures of what a microbial community is and what it is doing. This **multi-omics** approach is computationally demanding but is beginning to reveal how microbial communities function as coordinated systems rather than as collections of unrelated individual species [REF:knight2018].

> **Where This Matters:** Multi-omics approaches are central to many findings discussed throughout *The Inhabited Body*, particularly in Chapters 7 (diet and the microbiome), 10 (the gut-brain axis), and 13 (microbiome and metabolism).

The Return of Culture: Culturomics

Given everything we have said about the limitations of culture, it might seem surprising that one of the most exciting recent developments in microbiome science is… better culturing.

In 2012, a team led by the French microbiologist Didier Raoult introduced an approach they called **culturomics** [REF:lagier2018]. The strategy is conceptually simple but operationally ambitious: instead of growing a sample on one or two standard media under a single set of conditions, you use dozens or even hundreds of different culture conditions — different nutrient media, different atmospheric compositions (aerobic, anaerobic, microaerobic), different temperatures, different incubation times (days, weeks, sometimes months) — to coax into growth the many species that standard methods miss.

Every colony that appears is then identified, not by traditional biochemical tests, but by **MALDI-TOF mass spectrometry** — a rapid technique that identifies organisms by the unique pattern of proteins they contain, producing a result in minutes rather than the days required by classical identification. Colonies that cannot be matched to any known species are flagged as potentially novel and subjected to genome sequencing.

Think of it as fishing with every kind of net, bait, and lure you can find, in every part of the lake, at every time of day — instead of casting the same hook from the same spot and concluding that the lake only contains one kind of fish.

The results have been remarkable. Culturomics studies have isolated hundreds of bacterial species from the human gut that had never previously been grown in culture — including some that were known only from their DNA sequences in metagenomic databases, and others that were entirely new to science [REF:lagier2018]. By 2018, the approach had added more than 200 new species to the catalogue of human-associated bacteria.

Culturomics and metagenomics are complementary. Metagenomics tells you who is there; culturomics lets you get a living specimen. A living culture can be characterised in far more detail than a sequence: you can test its antibiotic susceptibility, study its metabolism, observe its behaviour under controlled conditions, and — critically — use it in animal experiments to test its effects on the host. You cannot do any of these things with a sequence alone.

> **Where This Matters:** Chapter 4 of *The Inhabited Body* covers culturomics in detail. Later chapters, particularly those on probiotics (Chapter 20) and faecal microbiota transplantation (Chapter 21), depend heavily on having living cultures of gut bacteria — something that culturomics has made far more feasible.

Germ-Free Animals: Testing Cause and Effect

The tools described above — sequencing, metabolomics, culturomics — can reveal correlations: that a particular species is more abundant in people with a certain disease, for instance, or that a particular metabolite is elevated after a dietary change. But correlation is not causation. To test whether a microbe actually *causes* an effect, you need an experiment. And the most powerful experiment in microbiome science involves an animal with no microbes at all.

Gnotobiotic animals (from the Greek *gnotos*, "known," and *bios*, "life") are raised from birth in completely sterile environments — sealed isolators with filtered air and sterilised food and water. The most commonly used are **germ-free mice**: animals that harbour no bacteria, no viruses, no fungi — nothing. Their microbiome is a blank slate.

This blank slate is extraordinarily useful. If you want to know whether a specific bacterium affects body weight, you can colonise a group of germ-free mice with that species — and only that species — and compare their weight gain to that of mice that remain germ-free. If you want to know whether the gut microbiome from an obese human can transfer obesity to a mouse, you can transplant the entire faecal community from an obese donor into germ-free mice and watch what happens. (It can, and it does — a landmark experiment we discuss in Chapter 8.)

Germ-free animals have revealed that the microbiome influences immune system development, brain chemistry and behaviour, bone density, fat storage, and dozens of other physiological processes. Without this experimental model, we would know that the microbiome *correlates* with many aspects of health. With it, we can begin to demonstrate *causation*.

> **Where This Matters:** Gnotobiotic animal studies are referenced throughout *The Inhabited Body*, especially in Chapters 8 (obesity), 10 (gut-brain axis), 11 (immune development), and 14 (autoimmune disease). Chapter 4 discusses the model's limitations and ethical considerations.

How to Read Microbiome Science: A Sceptic's Checklist

Understanding the tools is important not just for appreciating what scientists have discovered, but for evaluating the claims you encounter — in this book, in the news, and in the marketing materials of probiotic companies.

Here is a brief checklist of questions worth asking when you encounter a microbiome study:

What method was used? A 16S study can tell you who is present but cannot identify viruses or provide functional information. A metagenomic study is more comprehensive but more expensive and computationally complex. A study based only on culture may have missed the majority of species. The method shapes what can — and cannot — be concluded.

How big was the study? Early microbiome studies often involved fewer than twenty participants. Modern cohort studies can include thousands. Small studies may detect real patterns, but they are also more susceptible to false positives — apparent associations that do not replicate in larger samples.

Was it correlational or causal? A study showing that people with disease X have more of bacterium Y does not prove that bacterium Y causes disease X. It might be a consequence of the disease, a side effect of its treatment, or a coincidence. Causal claims require interventional experiments — ideally in gnotobiotic animals or in controlled human trials.

Were confounders controlled? Diet, medication, age, geography, and dozens of other factors influence the microbiome. A study comparing the microbiomes of healthy people and people with a disease must account for these variables, or any observed differences might reflect the confounders rather than the disease itself.

Can it be replicated? A finding reported in one study, from one laboratory, using one set of methods, is preliminary. It becomes robust only when independent teams, using different methods and different populations, find the same result.

This is not a counsel of despair. The microbiome field has produced a remarkable body of rigorous, replicated findings that have genuinely advanced our understanding of human biology. But it has also produced a considerable volume of preliminary, overhyped, or poorly designed research — and an even larger volume of commercial claims that far outstrip the evidence. A little healthy scepticism serves the reader well.

> **Where This Matters:** Chapter 4, Section 4.11 of *The Inhabited Body* provides a more detailed guide to reading microbiome science critically, including common statistical pitfalls and the problem of "p-hacking" in large omics datasets.

The Tools That Shaped the Story

The technologies we have surveyed in this chapter — microscopy, culture, 16S sequencing, shotgun metagenomics, metabolomics, culturomics, germ-free animal models — are not merely background information. They are the reason the story in this book can be told at all.

Every claim about the microbiome's role in health and disease rests on one or more of these tools. When you read that certain gut bacteria produce molecules that influence mood, that claim was established through metabolomics. When you read that the infant gut is colonised in a specific sequence, that finding came from 16S or metagenomic sequencing of serial stool samples. When you read that a faecal transplant can cure a recurrent infection, that was demonstrated in clinical trials using culture and molecular diagnostics to track outcomes.

Understanding the tools also helps you understand the gaps. The virome is less well characterised than the bacteriome because viruses lack the universal barcoding gene that 16S provides. The mycobiome has historically been overlooked because standard sequencing protocols targeted bacteria. The functional activity of the microbiome is harder to study than its composition because metabolomics and metatranscriptomics are technically demanding and expensive.

If Primer Chapters 1 through 5 gave you the biological vocabulary you need to understand the microbiome, this chapter gives you the methodological vocabulary — the ability to understand not just *what* scientists have found, but *how* they found it, and how confident you should be in the findings.

You are now ready for the main book.

Chapter References

- [REF:lane2015] Lane, N. (2015). *The Unseen World: Reflections on Leeuwenhoek*. Philosophical Transactions of the Royal Society B, 370(1666), 20140344. DOI: 10.1098/rstb.2014.0344
- [REF:blevins2010] Blevins, S.M. & Bronze, M.S. (2010). Robert Koch and the 'golden age' of bacteriology. *International Journal of Infectious Diseases*, 14(9), e744–e751. DOI: 10.1016/j.ijid.2009.12.003
- [REF:staley1985] Staley, J.T. & Konopka, A. (1985). Measurement of in situ activities of non-photosynthetic microorganisms in aquatic and terrestrial habitats. *Annual Review of Microbiology*, 39, 321–346. DOI: 10.1146/annurev.mi.39.100185.001541
- [REF:woese1977] Woese, C.R. & Fox, G.E. (1977). Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proceedings of the National Academy of Sciences*, 74(11), 5088–5090. DOI: 10.1073/pnas.74.11.5088
- [REF:hmp2012] Human Microbiome Project Consortium (2012). Structure, function and diversity of the healthy human microbiome. *Nature*, 486(7402), 207–214. DOI: 10.1038/nature11234
- [REF:qin2010] Qin, J. et al. (2010). A human gut microbial gene catalogue established by metagenomic sequencing. *Nature*, 464(7285), 59–65. DOI: 10.1038/nature08821
- [REF:knight2018] Knight, R. et al. (2018). Best practices for analysing microbiomes. *Nature Reviews Microbiology*, 16(7), 410–422. DOI: 10.1038/s41579-018-0029-9
- [REF:lagier2018] Lagier, J.-C. et al. (2018). Culturing the human microbiota and culturomics. *Nature Reviews Microbiology*, 16, 540–550. DOI: 10.1038/s41579-018-0041-0

The Inhabited Body — Glossary

A living document. Terms are added as chapters are drafted. Terms marked with [P1], [P2], etc. indicate the primer chapter where they are first introduced. Terms marked with [Ch1], [Ch2], etc. indicate the main book chapter where they are first introduced.

A

Alphaproteobacteria [P1] — A class of bacteria believed to be the evolutionary ancestors of mitochondria, the energy-producing organelles in eukaryotic cells.

Archaea [P1] — One of the three domains of cellular life, alongside Bacteria and Eukarya. Archaea are single-celled prokaryotic organisms that resemble bacteria in size and shape but differ fundamentally in their cell membrane chemistry, cell wall composition, and molecular biology. In the human microbiome, the most commonly detected archaeon is the methanogen *Methanobrevibacter smithii*.

Bacteroidota [Ch5] — A major phylum of bacteria (formerly called Bacteroidetes), dominant in the human gut. Key genera include *Bacteroides* (versatile carbohydrate degraders associated with Western diets) and *Prevotella* (associated with high-fibre, plant-rich diets).

Bacillota [Ch5] — A major phylum of bacteria (formerly called Firmicutes), typically the most abundant phylum in the adult human gut. Includes important genera such as *Faecalibacterium*, *Roseburia*, *Ruminococcus*, *Clostridium*, and *Lactobacillus*. Many Bacillota are specialist fibre degraders and major producers of short-chain fatty acids.

B

Bacillus (pl. bacilli) [P1] — A rod-shaped bacterium. Not to be confused with the genus *Bacillus*, which is a specific group of rod-shaped, spore-forming bacteria.

Bacteria [P1] — One of the three domains of cellular life. Bacteria are single-celled prokaryotic organisms characterised by ester-linked membrane lipids and, in most species, cell walls containing peptidoglycan. They are the dominant organisms at most human microbiome sites.

Bacteriocin [Ch5] — A small antimicrobial peptide produced by bacteria that kills or inhibits closely related bacterial species. Bacteriocins are one mechanism by which established gut communities resist colonisation by newcomers.

Bacteriophage (phage) [P1] — A virus that infects bacteria. Bacteriophages are the most abundant biological entities on Earth and play a major role in shaping microbial communities, including the human microbiome. See also: *Prophage*, *Lysogeny*.

Butyrate [Ch5] — A four-carbon short-chain fatty acid produced by bacterial fermentation of dietary fibre in the colon. Butyrate is the primary energy source for colonocytes (the cells lining the colon), promotes anti-inflammatory regulatory T cell differentiation, and helps maintain the low-oxygen environment that anaerobic gut bacteria require. See also: *Short-chain fatty acids*.

C

Capsid [P1] — The protein shell that surrounds and protects the genetic material of a virus.

Chloroplast [P1] — An organelle found in plant and algal cells, responsible for photosynthesis. Chloroplasts are descended from ancient cyanobacteria that were engulfed by an early eukaryotic cell (see *Endosymbiosis*).

Coccus (pl. cocci) [P1] — A spherical bacterium. Examples include *Staphylococcus* and *Streptococcus*.

Cyanobacteria [P1] — Photosynthetic bacteria responsible for producing much of the oxygen in Earth's atmosphere. Ancient cyanobacteria gave rise to chloroplasts through endosymbiosis.

Colonisation resistance [Ch5] — The ability of an established microbial community to prevent the colonisation of new organisms, including pathogens. Mechanisms include nutrient competition, production of antimicrobial compounds (bacteriocins), bile acid modification, and immune priming. First described by van der Waaij et al. (1971).

Colonocyte [Ch5] — An epithelial cell lining the colon. Colonocytes derive 60–70% of their energy from butyrate, a short-chain fatty acid produced by gut bacteria.

Competitive exclusion [Ch5] — An ecological principle stating that two species competing for exactly the same resource in a stable environment cannot coexist indefinitely — one will outcompete the other. In the gut, competitive exclusion is a key mechanism of colonisation resistance.

D

DNA (deoxyribonucleic acid) [P1] — The molecule that carries genetic information in all cellular life and many viruses. DNA consists of two complementary strands wound into a double helix. See Primer Chapter 3 for a detailed explanation.

Domain [P1] — The highest level of biological classification. Life is divided into three domains: Bacteria, Archaea, and Eukarya.

E

Endogenous retrovirus (ERV) [P1] — A retroviral sequence that has been integrated into the germline genome of a host organism and is inherited by subsequent generations. The human genome contains thousands of human endogenous retroviruses (HERVs), some of which have been co-opted for essential functions. See also: *Syncytin*.

Endosymbiosis [P1] — The evolutionary process by which one organism lives inside another, with both partners eventually becoming mutually dependent. The origin of mitochondria and chloroplasts in eukaryotic cells is explained by endosymbiosis: free-living bacteria were engulfed by an ancestral cell and, over billions of years, became permanent organelles.

Envelope (viral) [P1] — An outer lipid layer surrounding the capsid of some viruses, derived from the membrane of the host cell.

Ester linkage [P1] — A type of chemical bond found in the cell membranes of bacteria and eukaryotes, linking fatty acids to glycerol. Contrast with the *ether linkages* found in archaeal membranes.

Ether linkage [P1] — A type of chemical bond found in the cell membranes of archaea, linking branched hydrocarbon chains to glycerol. Ether linkages are more chemically stable than ester linkages, which may contribute to the ability of some archaea to survive extreme conditions.

Dysbiosis [Ch5] — A disruption in the composition or function of a microbial community, often associated with disease. Markers of gut dysbiosis include reduced diversity, loss of beneficial species, and expansion of Pseudomonadota (Proteobacteria). The term is widely used but somewhat imprecise, as no single "healthy" community composition has been definitively established.

Enterochromaffin cell [Ch5] — A specialised epithelial cell in the gut lining that produces serotonin. Approximately 90% of the body's total serotonin is produced by enterochromaffin cells, and their serotonin synthesis is regulated in part by gut bacteria.

Enteric nervous system (ENS) [Ch5] — The network of approximately 500 million neurons embedded in the walls of the gastrointestinal tract. Sometimes called the "second brain," the ENS can coordinate digestion, motility, and local immune responses independently of the brain. Organised into two main plexuses: the myenteric plexus (controls peristalsis) and the submucosal plexus (regulates secretion and blood flow).

Enterotype [Ch5] — A proposed classification of the human gut microbiome into distinct community types characterised by the dominance of particular genera (e.g., *Bacteroides* or *Prevotella*). First described by Arumugam et al. (2011). The concept remains debated: current evidence suggests continuous gradients rather than strictly discrete types.

Eukarya (Eukaryota) [P1] — The domain of life encompassing all organisms whose cells contain a true nucleus and membrane-bound organelles. Includes animals, plants, fungi, and protists.

Eukaryotic cell [P1] — A cell that possesses a membrane-bound nucleus and internal organelles (e.g., mitochondria, endoplasmic reticulum). All animals, plants, fungi, and protists are composed of eukaryotic cells. Contrast with *Prokaryotic cell*.

Extremophile [P1] — An organism that thrives in extreme environmental conditions, such as very high or very low temperatures, high salinity, high acidity, or high pressure. Many extremophiles are archaea, though extremophilic bacteria and eukaryotes also exist.

Free fatty acid receptor (FFAR) [Ch5] — A family of cell-surface receptors (including FFAR2/GPR43 and FFAR3/GPR41) that are activated by short-chain fatty acids. Expressed on immune cells, adipocytes, enteroendocrine cells, and neurons, these receptors provide a direct molecular link between microbial metabolism and host physiology.

G

Goblet cell [Ch5] — A specialised epithelial cell that secretes mucus. In the colon, goblet cells produce the MUC2 mucin that forms the two-layer mucus barrier separating the microbial community from the underlying tissue. "Sentinel" goblet cells at crypt entrances detect bacterial contact and trigger protective mucus secretion.

H

HERV (human endogenous retrovirus) [P1] — See *Endogenous retrovirus*.

L

LUCA (Last Universal Common Ancestor) [P1] — The hypothetical single-celled organism from which all current cellular life on Earth is descended. LUCA is not a specific species but a theoretical ancestor of all three domains.

Lysogeny (lysogenic cycle) [P1] — A mode of viral reproduction in which the virus integrates its genetic material into the host cell's genome (becoming a *prophage* in bacteria) and is replicated passively as the host divides. Under stress, the virus may exit lysogeny and enter the *lytic cycle*, producing new viral particles and destroying the host cell.

Ileocaecal valve [Ch5] — The one-way valve separating the ileum (terminal small intestine) from the caecum (beginning of the large intestine). Passage through this valve marks the transition from the relatively sparse microbial environment of the small intestine to the densely colonised colon.

M

Microbiota-accessible carbohydrates (MACs) [Ch5] — Complex carbohydrates — including cellulose, hemicellulose, resistant starch, inulin, and pectins — that human enzymes cannot digest but which are fermented by gut bacteria, primarily in the colon. The term, introduced by the Sonnenburgs, is increasingly preferred over "dietary fibre" in microbiome research because it specifies that the relevant property of these molecules is their accessibility to microbial metabolism.

MUC2 [Ch5] — The principal mucin glycoprotein of the colonic mucus layer. One of the largest molecules produced by the human body (molecular weight >5 million daltons), MUC2 is secreted by goblet cells and forms the structural basis of both the inner (bacteria-free) and outer (bacteria-colonised) mucus layers of the colon.

Metagenome [Ch1] — The collective genetic material of all microorganisms in a particular environment. The human gut metagenome, for example, contains between two and twenty million genes — far more than the approximately 20,000 protein-coding genes in the human genome.

Methanogen [P1] — An organism that produces methane as a metabolic by-product. In the human gut, the most common methanogen is the archaeon *Methanobrevibacter smithii*, which consumes hydrogen and carbon dioxide produced by bacterial fermentation.

Microbiome [Ch1] — The entire habitat of microorganisms associated with a particular environment, including the organisms themselves, their collective genetic material, their metabolic products, and the environmental conditions they create and respond to. Often used interchangeably with *microbiota* in popular writing.

Microbiota [Ch1] — The community of living microorganisms (bacteria, archaea, fungi, protists, and viruses) inhabiting a particular environment, such as the human gut.

Mitochondrion (pl. mitochondria) [P1] — A membrane-bound organelle found in most eukaryotic cells, responsible for generating the majority of the cell's energy through aerobic respiration. Mitochondria are descended from ancient alphaproteobacteria that were engulfed by an ancestral cell (see *Endosymbiosis*) and retain their own small genome.

N

Nucleoid [P1] — The region within a prokaryotic cell where the DNA is concentrated. Unlike a eukaryotic nucleus, the nucleoid is not surrounded by a membrane.

Nucleus [P1] — A membrane-bound organelle in eukaryotic cells that contains the cell's DNA. The presence of a nucleus is the defining feature of eukaryotic cells.

O

Obelisk [P1] — A recently discovered class of viroid-like circular RNA elements found in the human gut microbiome. Obelisks encode a single protein and do not fit into any previously known category of biological entity. First described by Zheludev et al. in 2024.

Organelle [P1] — A specialised, membrane-bound structure within a eukaryotic cell that performs a specific function. Examples include the nucleus, mitochondria, and chloroplasts.

P

Obligate anaerobe [Ch5] — An organism that cannot survive in the presence of oxygen. Most of the dominant bacteria in the human colon are obligate anaerobes, which is why the low-oxygen environment of the colonic lumen is essential to maintaining normal gut community composition.

Peyer's patches [Ch5] — Organised clusters of immune tissue (lymphoid follicles) embedded in the wall of the ileum. Peyer's patches sample the microbial contents of the small intestine and are major sites of immune surveillance and education.

Peptidoglycan [P1] — A polymer of sugars and amino acids that forms a mesh-like layer in bacterial cell walls, giving the cell structural rigidity. Absent from archaeal cell walls. The target of several classes of antibiotics, including penicillins and cephalosporins.

Prion [P1] — An infectious agent consisting of a misfolded form of a normal cellular protein (PrP). Prions can induce correctly folded copies of the same protein to adopt the abnormal shape, causing a chain reaction that leads to progressive, fatal brain diseases such as Creutzfeldt-Jakob disease (CJD) in humans.

Prokaryotic cell [P1] — A cell that lacks a membrane-bound nucleus and other internal membrane-bound organelles. Bacteria and archaea are prokaryotes. Contrast with *Eukaryotic cell*.

Prophage [P1] — The DNA of a bacteriophage that has been integrated into a bacterial host's genome during the *lysogenic cycle*. The prophage is replicated along with the host's DNA during cell division and can reactivate under certain conditions.

Protist [P1] — An informal term for any eukaryotic organism that is not an animal, plant, or fungus. Protists are a diverse and largely unrelated collection of organisms, mostly single-celled, that include amoebae, algae, and parasites such as *Plasmodium* (malaria) and *Blastocystis*.

R

Ribosomal RNA (rRNA) [P1] — A structural component of ribosomes, the cellular machinery that translates genetic instructions into proteins. The gene encoding 16S rRNA (in prokaryotes) or 18S rRNA (in eukaryotes) is highly conserved across all life and is widely used for identifying and classifying microorganisms. Carl Woese's comparison of 16S rRNA sequences led to the discovery of the three domains of life.

S

Short-chain fatty acids (SCFAs) [Ch5] — Small organic acids — principally acetate (2 carbons), propionate (3 carbons), and butyrate (4 carbons) — produced by bacterial fermentation of dietary fibre in the colon. SCFAs are produced in quantities of ~500–600 mmol/day and serve as the primary energy source for colonocytes, regulate immune function, influence distant organs via the bloodstream, and help maintain the anaerobic environment of the gut. They are arguably the most important class of molecules produced by the gut microbiome.

Spirillum (pl. spirilla) [P1] — A spiral-shaped bacterium.

Syncytin [P1] — A protein essential for the formation of the human placenta, derived from the envelope gene of an ancient endogenous retrovirus. Syncytin mediates the fusion of cells in the outer layer of the placenta, enabling nutrient exchange between mother and foetus.

V

Vibrio [P1] — A comma-shaped or curved-rod bacterium. The genus *Vibrio* includes the species *Vibrio cholerae*, which causes cholera.

Viroid [P1] — A small, circular RNA molecule that can infect plant cells and cause disease. Viroids are the smallest known infectious agents and consist of naked RNA with no protein coat.

Virion [P1] — A complete, mature viral particle outside of a host cell. A virion consists of the viral genome (DNA or RNA) enclosed in a protein capsid, and in some cases an outer lipid envelope.

Virus [P1] — A non-cellular biological entity consisting of genetic material (DNA or RNA) enclosed in a protein coat (capsid). Viruses cannot reproduce independently and must hijack the machinery of a host cell. They infect organisms across all three domains of life.

This glossary will be expanded as subsequent primer and main book chapters are drafted.

The Inhabited Body — Primer Bibliography

References for companion primer chapters, sorted by chapter. Reference identifiers use the format REF:shortname for in-text citation. All DOIs verified via PubMed and/or web search, April 2026.

Primer Ch. 1 — The Living World: Life's Major Domains

- [REF:woese1977] Woese, C.R. & Fox, G.E. (1977). Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proceedings of the National Academy of Sciences*, 74(11), 5088–5090. DOI: 10.1073/pnas.74.11.5088
- [REF:woese1990] 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*, 87(12), 4576–4579. DOI: 10.1073/pnas.87.12.4576
- [REF:louca2019] Louca, S., Mazel, F., Doebeli, M. & Parfrey, L.W. (2019). A census-based estimate of Earth's bacterial and archaeal diversity. *PLOS Biology*, 17(2), e3000106. DOI: 10.1371/journal.pbio.3000106
- [REF:locey2016] Locey, K.J. & Lennon, J.T. (2016). Scaling laws predict global microbial diversity. *Proceedings of the National Academy of Sciences*, 113(21), 5970–5975. DOI: 10.1073/pnas.1521291113
- [REF:spang2015] Spang, A. et al. (2015). Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature*, 521(7551), 173–179. DOI: 10.1038/nature14447
- [REF:zheludev2024] Zheludev, I.N. et al. (2024). Viroid-like colonists of human microbiomes. *Cell*, 187(23), 6521–6536.e18. DOI: 10.1016/j.cell.2024.09.033
- [REF:lederberg2001] Lederberg, J. & McCray, A.T. (2001). 'Ome sweet 'omics — a genealogical treasury of words. *The Scientist*, 15(7), 8. URL: https://www.the-scientist.com/ome-sweet-omics-a-genealogical-treasury-of-words-54889
- [REF:sender2016] Sender, R., Fuchs, S. & Milo, R. (2016). Are we really vastly outnumbered? Revisiting the ratio of bacterial to host cells in humans. *Cell*, 164(3), 337–340. DOI: 10.1016/j.cell.2016.01.013

Primer Ch. 2 — The Cell: Life's Basic Unit

- [REF:hooke1665] Hooke, R. (1665). *Micrographia: or Some Physiological Descriptions of Minute Bodies Made by Magnifying Glasses*. London: Royal Society. URL: https://www.gutenberg.org/ebooks/15491
- [REF:schwann1839] Schwann, T. (1839). *Mikroskopische Untersuchungen über die Uebereinstimmung in der Struktur und dem Wachsthum der Thiere und Pflanzen*. Berlin: Verlag der Sander'schen Buchhandlung. [English translation: *Microscopical Researches into*

the Accordance in the Structure and Growth of Animals and Plants, 1847.] URL: https://archive.org/details/b29287455

- [REF:virchow1855] Virchow, R. (1855). Cellular-Pathologie. *Archiv für pathologische Anatomie und Physiologie und für klinische Medicin*, 8(1), 3–39. DOI: 10.1007/BF01935312
- [REF:gram1884] Gram, H.C. (1884). Ueber die isolirte Färbung der Schizomyceten in Schnitt- und Trockenpräparaten. *Fortschritte der Medicin*, 2, 185–189. DOI: 10.1055/s-0029-1209285
- [REF:sagan1967] Sagan, L. [later Margulis] (1967). On the origin of mitosing cells. *Journal of Theoretical Biology*, 14(3), 225–274. DOI: 10.1016/0022-5193(67)90079-3
- [REF:margulis1970] Margulis, L. (1970). *Origin of Eukaryotic Cells*. New Haven: Yale University Press. ISBN: 978-0-300-01353-5.
- [REF:shih2006] Shih, Y.-L. & Rothfield, L. (2006). The bacterial cytoskeleton. *Microbiology and Molecular Biology Reviews*, 70(3), 729–754. DOI: 10.1128/MMBR.00017-06
- [REF:bianconi2013] Bianconi, E. et al. (2013). An estimation of the number of cells in the human body. *Annals of Human Biology*, 40(6), 463–471. DOI: 10.3109/03014460.2013.807878
- [REF:blattner1997] Blattner, F.R. et al. (1997). The complete genome sequence of *Escherichia coli* K-12. *Science*, 277(5331), 1453–1462. DOI: 10.1126/science.277.5331.1453
- [REF:international2004] International Human Genome Sequencing Consortium (2004). Finishing the euchromatic sequence of the human genome. *Nature*, 431(7011), 931–945. DOI: 10.1038/nature03001
- [REF:alberts2022] Alberts, B. et al. (2022). *Molecular Biology of the Cell*, 7th edition. New York: W.W. Norton. ISBN: 978-0-393-88482-1.

Primer Ch. 3 — The Code of Life: DNA, RNA, and Proteins

- [REF:piovesan2019] Piovesan, A. et al. (2019). On the length, weight and GC content of the human genome. *BMC Research Notes*, 12, 106. DOI: 10.1186/s13104-019-4137-z
- [REF:watson1953] Watson, J.D. & Crick, F.H. (1953). Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. *Nature*, 171(4356), 737–738. DOI: 10.1038/171737a0
- [REF:crick1970] Crick, F. (1970). Central dogma of molecular biology. *Nature*, 227(5258), 561–563. DOI: 10.1038/227561a0
- [REF:nirenberg1961] Nirenberg, M.W. & Matthaei, J.H. (1961). The dependence of cell-free protein synthesis in *E. coli* upon naturally occurring or synthetic polyribonucleotides. *Proceedings of the National Academy of Sciences*, 47(10), 1588–1602. DOI: 10.1073/pnas.47.10.1588
- [REF:crick1961] Crick, F.H., Barnett, L., Brenner, S. & Watts-Tobin, R.J. (1961). General nature of the genetic code for proteins. *Nature*, 192, 1227–1232. DOI: 10.1038/1921227a0

- [REF:jacob1961] Jacob, F. & Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. *Journal of Molecular Biology*, 3, 318–356. DOI: 10.1016/s0022-2836(61)80072-7
- [REF:kariko2005] Karikó, K., Buckstein, M., Ni, H. & Weissman, D. (2005). Suppression of RNA recognition by Toll-like receptors: the impact of nucleoside modification and the evolutionary origin of RNA. *Immunity*, 23(2), 165–175. DOI: 10.1016/j.immuni.2005.06.008
- [REF:encode2012] ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature*, 489(7414), 57–74. DOI: 10.1038/nature11247
- [REF:venter2001] Venter, J.C. et al. (2001). The sequence of the human genome. *Science*, 291(5507), 1304–1351. DOI: 10.1126/science.1058040
- [REF:international2004] International Human Genome Sequencing Consortium (2004). Finishing the euchromatic sequence of the human genome. *Nature*, 431(7011), 931–945. DOI: 10.1038/nature03001
- [REF:schmiedel2015] Schmiedel, J.M. et al. (2015). MicroRNA control of protein expression noise. *Science*, 348(6230), 128–132. DOI: 10.1126/science.aaa1738
- [REF:alberts2022] Alberts, B. et al. (2022). *Molecular Biology of the Cell*, 7th edition. New York: W.W. Norton. ISBN: 978-0-393-88482-1.

Primer Ch. 4 — The Fungal Kingdom

- [REF:baldauf1993] Baldauf, S.L. & Palmer, J.D. (1993). Animals and fungi are each other's closest relatives: congruent evidence from multiple proteins. *Proceedings of the National Academy of Sciences*, 90(24), 11558–11562. DOI: 10.1073/pnas.90.24.11558
- [REF:cavalier-smith1987] Cavalier-Smith, T. (1987). The origin of Fungi and pseudofungi. In: Rayner, A.D.M., Brasier, C.M. & Moore, D. (eds) *Evolutionary Biology of the Fungi*. Cambridge University Press, pp. 339–353. [No DOI available for this book chapter.]
- [REF:parfrey2011] Parfrey, L.W., Lahr, D.J.G., Knoll, A.H. & Katz, L.A. (2011). Estimating the timing of early eukaryotic diversification with multigene molecular clocks. *Proceedings of the National Academy of Sciences*, 108(33), 13624–13629. DOI: 10.1073/pnas.1110633108
- [REF:gow2017] Gow, N.A.R., Latge, J.-P. & Munro, C.A. (2017). The fungal cell wall: structure, biosynthesis, and function. *Microbiology Spectrum*, 5(3), FUNK-0035-2016. DOI: 10.1128/microbiolspec.FUNK-0035-2016
- [REF:naranjo-ortiz2019] Naranjo-Ortiz, M.A. & Gabaldón, T. (2019). Fungal evolution: diversity, taxonomy and phylogeny of the Fungi. *Biological Reviews*, 94(6), 2101–2137. DOI: 10.1111/brv.12550
- [REF:goffeau1996] Goffeau, A. et al. (1996). Life with 6000 genes. *Science*, 274(5287), 546–567. DOI: 10.1126/science.274.5287.546
- [REF:sudbery2004] Sudbery, P., Gow, N. & Berman, J. (2004). The distinct morphogenic states of *Candida albicans*. *Trends in Microbiology*, 12(7), 317–324. DOI: 10.1016/j.tim.2004.05.008

- [REF:bongomin2017] Bongomin, F., Gago, S., Oladele, R.O. & Denning, D.W. (2017). Global and multi-national prevalence of fungal diseases — estimate precision. *Journal of Fungi*, 3(4), 57. DOI: 10.3390/jof3040057
- [REF:lockhart2017] Lockhart, S.R. et al. (2017). Simultaneous emergence of multidrug-resistant *Candida auris* on 3 continents confirmed by whole-genome sequencing and epidemiological analyses. *Clinical Infectious Diseases*, 64(2), 134–140. DOI: 10.1093/cid/ciw691
- [REF:tedersoo2018] Tedersoo, L. et al. (2018). High-level classification of the Fungi and a tool for evolutionary ecological analyses. *Fungal Diversity*, 90, 135–159. DOI: 10.1007/s13225-018-0401-0
- [REF:hawksworth2017] Hawksworth, D.L. & Lücking, R. (2017). Fungal diversity revisited: 2.2 to 3.8 million species. *Microbiology Spectrum*, 5(4), FUNK-0052-2016. DOI: 10.1128/microbiolspec.FUNK-0052-2016
- [REF:brown2006] Brown, G.D. (2006). Dectin-1: a signalling non-TLR pattern-recognition receptor. *Nature Reviews Immunology*, 6(1), 33–43. DOI: 10.1038/nri1745
- [REF:li2021] Li, Y. et al. (2021). A genome-scale phylogeny of the kingdom Fungi. *Current Biology*, 31(8), 1653–1665.e5. DOI: 10.1016/j.cub.2021.01.074

Primer Ch. 5 — Viruses and Their Strange Relatives

- [REF:suttle2007] Suttle, C.A. (2007). Marine viruses — major players in the global ecosystem. *Nature Reviews Microbiology*, 5(10), 801–812. DOI: 10.1038/nrmicro1750
- [REF:shkoporov2019] Shkoporov, A.N. & Hill, C. (2019). Bacteriophages of the human gut: the "known unknown" of the microbiome. *Cell Host & Microbe*, 25(2), 195–209. DOI: 10.1016/j.chom.2019.01.017
- [REF:baltimore1970] Baltimore, D. (1970). RNA-dependent DNA polymerase in virions of RNA tumour viruses. *Nature*, 226(5252), 1209–1211. DOI: 10.1038/2261209a0
- [REF:temin1970] Temin, H.M. & Mizutani, S. (1970). RNA-dependent DNA polymerase in virions of Rous sarcoma virus. *Nature*, 226(5252), 1211–1213. DOI: 10.1038/2261211a0
- [REF:barresinoussi1983] Barré-Sinoussi, F. et al. (1983). Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science*, 220(4599), 868–871. DOI: 10.1126/science.6189183
- [REF:zheludev2024] Zheludev, I.N. et al. (2024). Viroid-like colonists of human microbiomes. *Cell*, 187(23), 6521–6536.e18. DOI: 10.1016/j.cell.2024.09.033
- [REF:alberts2022] Alberts, B. et al. (2022). *Molecular Biology of the Cell*, 7th edition. New York: W.W. Norton. ISBN: 978-0-393-88482-1.

Primer Ch. 6 — How We Study What We Can't See

- [REF:lane2015] Lane, N. (2015). The unseen world: reflections on Leeuwenhoek (1677) 'Concerning little animals'. *Philosophical Transactions of the Royal Society B*, 370(1666), 20140344. DOI: 10.1098/rstb.2014.0344
- [REF:blevins2010] Blevins, S.M. & Bronze, M.S. (2010). Robert Koch and the 'golden age' of bacteriology. *International Journal of Infectious Diseases*, 14(9), e744–e751. DOI: 10.1016/j.ijid.2009.12.003
- [REF:staley1985] Staley, J.T. & Konopka, A. (1985). Measurement of in situ activities of non-photosynthetic microorganisms in aquatic and terrestrial habitats. *Annual Review of Microbiology*, 39, 321–346. DOI: 10.1146/annurev.mi.39.100185.001541
- [REF:woese1977] Woese, C.R. & Fox, G.E. (1977). Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proceedings of the National Academy of Sciences*, 74(11), 5088–5090. DOI: 10.1073/pnas.74.11.5088
- [REF:hmp2012] Human Microbiome Project Consortium (2012). Structure, function and diversity of the healthy human microbiome. *Nature*, 486(7402), 207–214. DOI: 10.1038/nature11234
- [REF:qin2010] Qin, J. et al. (2010). A human gut microbial gene catalogue established by metagenomic sequencing. *Nature*, 464(7285), 59–65. DOI: 10.1038/nature08821
- [REF:knight2018] Knight, R. et al. (2018). Best practices for analysing microbiomes. *Nature Reviews Microbiology*, 16(7), 410–422. DOI: 10.1038/s41579-018-0029-9
- [REF:lagier2018] Lagier, J.-C. et al. (2018). Culturing the human microbiota and culturomics. *Nature Reviews Microbiology*, 16, 540–550. DOI: 10.1038/s41579-018-0041-0

Verification Notes

References verified against PubMed metadata (April 2026). All DOIs confirmed to resolve to the cited titles with matching authors, journals, volumes, and page numbers.