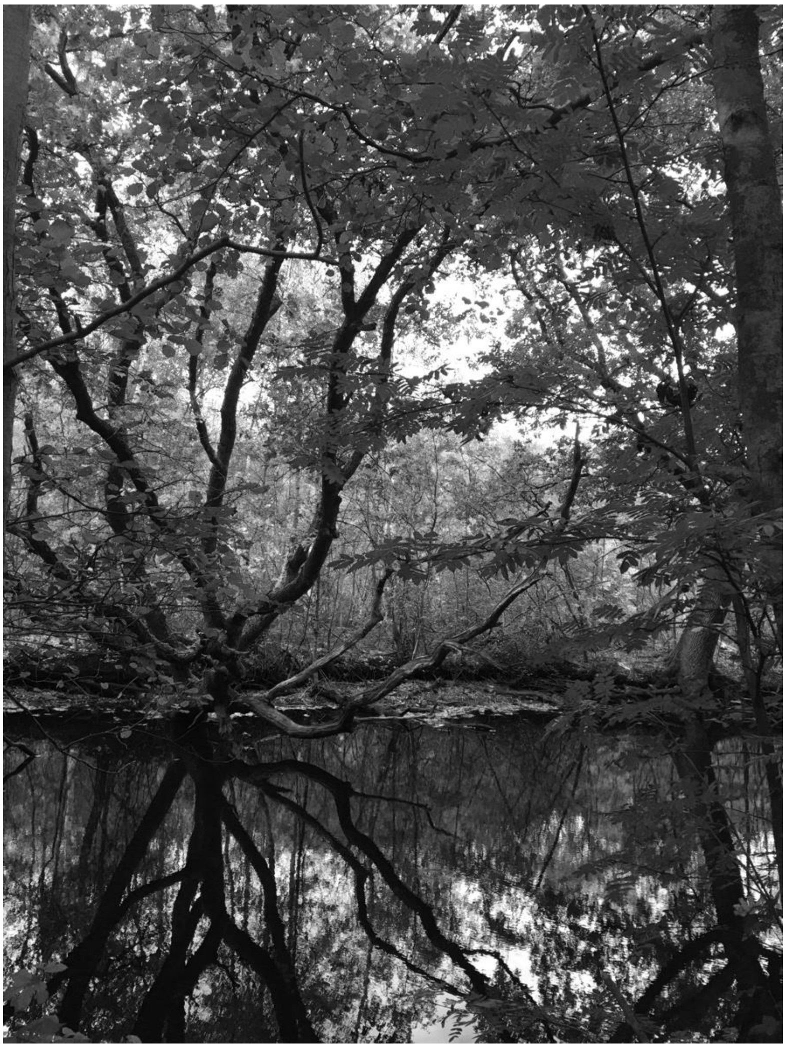
The tree of life and LUCA

Is all life on Earth related? And who is LUCA?



Anja Spang

Department of Marine Microbiology and Biogeochemistry, NIOZ, Royal Netherlands Institute for Sea Research, Den Burg, The Netherlands

What is the tree of life? And who is LUCA?

Storyline

All cellular life on Earth is related by common decent, that is, has evolved from a shared last common ancestor. In this topic framework, you will learn what major types of cellular life can be found on Earth, how all these forms of life are related to each other, how life has evolved over time, who the Last Universal Common Ancestor – LUCA – was, and how eukaryotic cells originated from superficially simpler-appearing prokaryotic cells. I hope you enjoy this journey into the deep origins of cellular life and our microbial ancestors.

The Microbiology and Societal Context

The microbiology: living organisms and cells; overview of the cellular life forms on Earth; the last universal common ancestor of all life; the origin of cellular complexity and eukaryotic life which includes us humans. *Sustainability issues:* future of life; predicting evolution; evolution, diversification, and extinction in the age of climate change.

The Tree of Life (TOL): the Microbiology

1. ***What cellular life is there?*** Cellular life on our planet is extremely diverse and comprised of many different organisms, all of which are related to each other and trace their lineage back to a shared common ancestor (Box 1). Broadly, these organisms can be divided into three major groups referred to as **domains**: the Archaea, the Bacteria and Eukarya (also Eukaryota or **eukaryotes**).

Maybe you have heard about Bacteria and eukaryotes before? But do you know who the Archaea are? For a long time, it was thought that there are complex cells, such as those of eukaryotes (see Section 4), as well as the simpler cells without **nucleus**, the **prokaryotes,** which were synonymous with Bacteria. To a large degree, this was due to the way scientists studied cells: they looked at them in the microscope and saw two broadly distinct types of cells.

However, during the last half century, our tools to study cells have been rapidly improving, which has changed our ability to distinguish between cells that superficially look alike. In particular, Carl Woese pioneered the use of genetic information (e.g the **ribosomal RNA**) to study cellular life and showed that not all the simpler appearing cells are indeed Bacteria. Instead, those cells seemed to be comprised of two distinct types of organism groups: the previously known Bacteria as well as the so-called Archaea (named after the Greek word 'archaía', which means ancient). While the scientific community was at first sceptical to this finding, subsequent research has confirmed that Archaea indeed form a separate domain of life besides the Bacteria and eukaryotes: they have their own unique features, features like those of Bacteria as well as features in common with eukaryotes to the exclusion of Bacteria. How is this possible? We can learn about these fundamental cell types by studying their evolutionary relationships.

2. ***What is the tree of life and how did our perception of it change over time?*** Maybe when hearing the term "tree of life", you remember the movie of the same name by Terrence Malick featuring Brad Pitt as main actor? Or you may think about the tree of life as a symbol in folklore and fiction or within the context of mythology, religion, or philosophy? Indeed, the "tree of life" has been a common and useful metaphor in various cultures long before the Gregorian calendar was introduced.

The tree of life (TOL) concept has also proven to be a powerful framework to depict the relationships of cellular life over time and was introduced into biology from around the time Charles

Darwin published his theory of evolution in "On the Origin of Species" (1859) and suggested that all organisms derive from a common ancestor.

The first TOLs published emphasized the diversity and relatedness of multicellular life and were predominantly based on **morphological** characteristics. However, the discovery of DNA as the molecule of heredity as well as the finding that all cellular life shares a **ribosome** responsible for the synthesis of proteins has provided important data allowing to reconstruct a much more accurate and less biased TOLs. Instead of relying on **morphology**, scientists can now compare ribosomal RNA or genes with heredity information (i.e. genes that are inherited from parent to offspring) across all cellular life.

This allows to reconstruct so-called **phylogenetic trees** whose tips represent extant (currently living) organisms, while the various nodes depict the ancestors. The branches (of differing length) connecting the various nodes and tips indicate the amount of change that separates ancestors from extant organisms. It is precisely such genetic information that allowed Carl Woese to separate the prokaryotes into two distinct groups, the Archaea and Bacteria (see Section 1).

The sequencing revolution, starting around the end of the 20th and the beginning of the 21st century, allowed to decipher the genetic material (DNA) of both eukaryotes, Archaea and Bacteria. While it was originally only possible to sequence **genomes** of cultivated organisms, which represent only a tiny fraction of all organisms, the development of better tools has not only made it feasible but common practice to obtain genomes of all the organisms present in a sample taken from any environment on Earth (e.g. a gram of soil), *without the need to cultivate them*. In this way, researchers were able to discover a large diversity of previously unknown Archaea, Bacteria and unicellular eukaryotes that cannot easily be cultivated in the laboratory or identified by eye (probably far less than 20% of prokaryotes have been cultivated so far).

The reconstruction of a new TOL based on this large amount of sequencing data has led to new fundamental insights into life's biodiversity: First, this has revealed that Archaea and Bacteria are extremely diverse and form many more branches in the tree of life than eukaryotes. Furthermore, in line with other data, it has revealed that eukaryotes are likely derived from a **symbiotic** event between an archaeon and a bacterium (see Section 4), suggesting that an archaeal and a bacterial branch have merged to form a new root at the basis of eukaryotic diversification (Box 1).

While this in a strict sense contradicts the TOL concept suggesting that all life has evolved vertically from a shared common ancestor, the TOL remains a powerful framework to explain the evolutionary history of life. But it is a tree more reminiscent of a network that allows branches to merge or certain characters to move between branches through genetic exchanges, so-called **horizontal gene transfer**. As a result of the merging of branches at the base of eukaryotic origins, some researchers have suggested to refer to Archaea and Bacteria as **primary domains** of life descending from a common ancestor (LUCA, see next Section) as well as a **secondary domain** including all eukaryotes.

3. ***Who was the Last Universal Common Ancestor?*** All known cellular organisms on Earth, that is, all the diverse species of Bacteria and Archaea, as well as eukaryotes derived from the former, ultimately descended from a single ancestor or a population of ancestral cells, the **L**ast **U**niversal **C**ommon **A**ncestor referred to as **LUCA**. In fact, the reconstruction of the **TOL** based on genetic information such as that introduced by Carl Woese (see Section 2) testifies to the existence of LUCA, which corresponds to the deepest node in this tree.

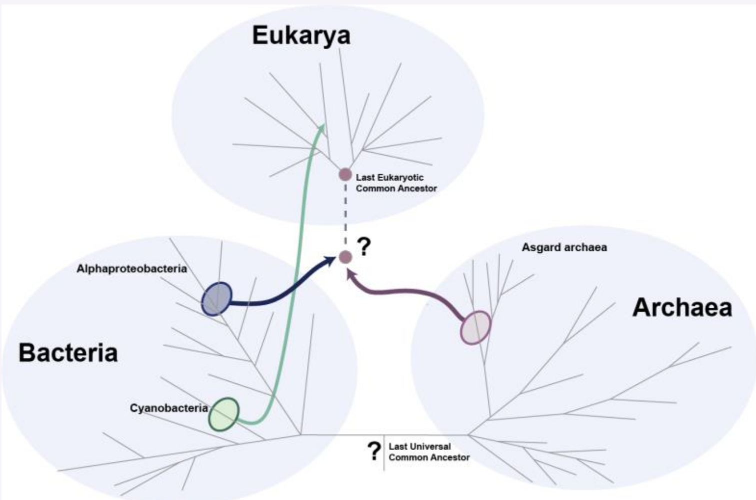
It is important to note that LUCA was not the first cell or cell-like organism on Earth, nor was it the sole cell/cell population around at that time. It was likely surrounded by many other cell-like life forms. However these, like other hominids that lived at the time of early *Homo sapiens*, did not perpetuate and give rise to any evolutionary lineages that currently exist today.

Besides the TOL, other features shared by all cellular life on Earth suggest a common ancestry: for example, all cells - those of Bacteria, Archaea and eukaryotes - share the same genetic code and basic principles of its decoding: **transcription** (**transfer RNA**, **messenger RNA**) and **translation** (e.g. **ribosomes** including **ribosomal RNA**), which mediate the synthesis of proteins from **genes** encoded in the DNA. Furthermore, all organisms are composed of cells surrounded by a **lipid bilayer** membrane providing a compartment for the genetic material and metabolism of the cell. While the lipid moieties of Archaea and Bacteria differ somewhat, it is likely that LUCA too, already had a lipid membrane.

What else do we know about LUCA? Because LUCA lived more than 4 billion years ago and did not leave any known **fossil** or **biomarker** information, we do not have any direct trace of this organism and it is therefore extremely challenging to accurately predict its characteristics. Nevertheless, using novel bioinformatic approaches, scientists can improve estimates on the time when LUCA lived and use information from extant life to reconstruct the nature of the various hypothetical ancestors in the tree of life down to its root, that is: as far back in time as LUCA.

Furthermore, the study of biological records and biochemistry can help to predict environmental conditions at the time of LUCA and provide constraints on how this cellular life form evolved from even simpler precursors. Findings from the application of such approaches suggest that LUCA may have been an **anaerobic**, **thermophilic** or even **hyper-thermophilic** organism optimally thriving at temperatures far above 50°C. It likely also already harboured several central metabolic pathways including enzymes for the synthesis of **ATP** – the energy currency of the cell – and was potentially able to transform inorganic carbon into cellular material – biomass. Finally, it may have lived in an environment that resembles hydrothermal vents and was rich in hydrogen and carbon dioxide. Yet all these predictions remain hypothetical and, while many features and secrets of LUCA may forever remain unknown to science, much can still be learned about our common ancestor using the wealth of recently generated data combined with ever improving methodologies.

Box 1



A depiction of the tree of life.
Archaea and Bacteria form two primary domains of life that evolved from the last universal common ancestor (LUCA) billions of years ago. It is currently thought but not confirmed that LUCA lies between the Archaea and Bacteria. The cells of both archaea and bacteria are unicellular and do not contain compartments. In contrast, eukaryotes, which include both unicellular protists and algae but also multicellular organisms such as plants, fungi, and animals, including us humans, have likely evolved later. In fact, the prevailing theory suggests that eukaryotic cells evolved from a symbiosis between an archaeaon or a sister-group of archaea (Asgard archaea) and a bacterium (Alphaproteobacterium) giving rise to the last eukaryotic common ancestor (LECA). Later, the acquisition of another bacterial symbiont (Cyanobacterium) by a eukaryotic lineage, led to the evolution of photosynthetic eukaryotic cells, including those of algae and plants. Therefore, both Bacteria but also the much lesser studied Archaea are very important for our understanding of the evolution of life on Earth.

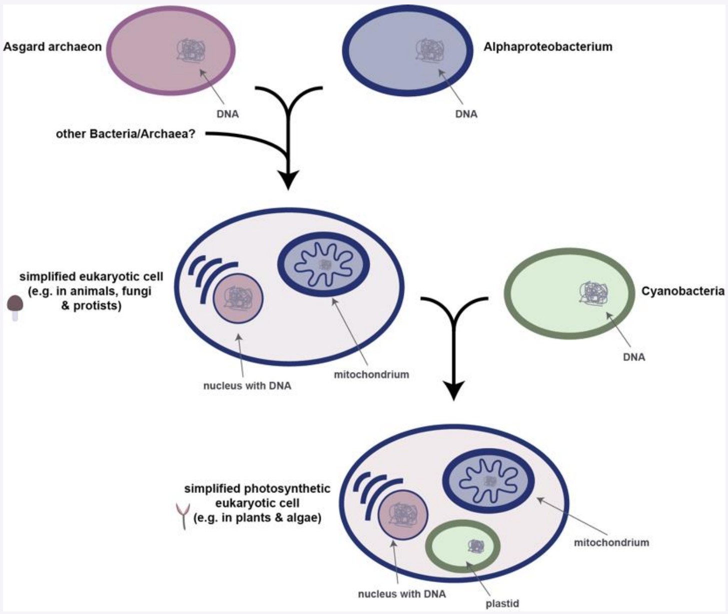
4. ***How did eukaryotes like us evolve?*** As indicated above, eukaryotic cells are more complex than those of Archaea and Bacteria and are characterised by intracellular compartments referred to as **organelles** (**Box 2**). These include a **nucleus** containing the genetic content of the organism, as well as

mitochondria, the ATP-generating powerhouse of eukaryotic cells. Mitochondria are responsible for **aerobic respiration** and the generation of cellular energy, and thereby fuel the **metabolism** of eukaryotes. Some eukaryotes have **mitochondria-related organelles** (MROs), that is, compartments that evolved from mitochondria but no longer respire oxygen. Some of these MROs produce hydrogen instead and allow the corresponding eukaryotes to inhabit anoxic environments, while others have lost most metabolic functions.

Some eukaryotic cells in addition have **plastids**, which are small compartments that enable the harvesting of sunlight for energy acquisition. Besides, these and additional organelles, eukaryotic cells also have an elaborate cytoskeleton – a cellular network of structural proteins that functions as an internal skeleton – and intermembrane vesicle transport systems linking the different compartments with each other.

Do you have any ideas where these organelles may come from? One of the currently most strongly supported hypotheses suggests that at some point in time, probably as far back as 2 billion years ago, there were some archaea (related to the so called **Asgard archaea**) and some bacteria (related to the **Alphaproteobacteria**) that lived closely together. At some point these archaeal and bacterial cells engaged in **symbiotic** interactions that led to the uptake of the bacterium into the archaeal cell (**Box 1 and 2**). The internalised bacterial cell(s) evolved subsequently into mitochondria which metabolise sugars to provide energy for growth. Mitochondria have a small genome that contains genes originally derived from the alphaproteobacterial ancestor, though some to all of these genes have been lost in MROs. Other genes are now located in the nucleus, together with genes derived from the archaeal host and many new genes that evolved later.

Box 2



A schematic illustration of the origin of the eukaryotic cell.
Increasing evidence suggests that eukaryotic cells emerged through a symbiosis between at least two prokaryotic partners, an Asgard archaeon (purple cell with archaeal lipid membrane) and an Alphaproteobacterium (blue cell with a bacterial membrane). Some researchers think that a third partner may have been involved as indicated by an additional arrow, though this remains unresolved.
The Alphaproteobacterium evolved into the mitochondrion (or organelles derived from the mitochondrion) and often still have a small DNA genome derived from the Bacterium. The main genome of eukaryotic cells is located in the nucleus and to a certain degree derives from the archaeal ancestor. The eukaryotic cell membrane is similar to those of Bacteria (indicated in dark blue). Animals, fungi and many unicellular eukaryotes share this same cellular structure. However, plant and algae cells subsequently evolved further when a eukaryotic cell engulfed a photosynthetic cyanobacterium, which gave rise to eukaryotic plastids. These plastids also have small DNA genomes derived from the Cyanobacterial ancestor and allow certain eukaryotes to use sunlight as energy source.

Similarly, some of these eukaryotic cells later on took up other bacteria (so called blue-green algae or **Cyanobacteria**), which evolved into plastids and provide algal and plants cells with the ability to harness energy from sunlight through a process called **photosynthesis**. In fact, some eukaryotes have engulfed eukaryotic algae (i.e. eukaryotic cells with plastids) leading to highly complex internal organelles. Furthermore, some of the unicellular eukaryotic cells evolved into multicellular organisms such as plants and animals including us humans.

Thus, all the visible biosphere today is ultimately derived from archaea and bacteria, microorganisms which still comprise the largest number of cells on Earth and have dominated organismal diversity on Earth since the earliest stages of the evolution of cellular life.

Potential Implications for Decisions

What do we not currently know that is important for future policy/behaviour? Recently, many new groups of Archaea, Bacteria and unicellular Eukaryotes have been discovered in all environments on Earth, including soils, sediments, lakes, the ocean, as well as associated with plants, animals and humans. Yet, we still know extremely little about the functions of all these organisms and how they affect the cycling of nutrients and climate change. A tiny fraction of these organisms could also represent unknown or emerging pathogens, that may cause disease in humans or other eukaryotes. Furthermore, much remains to be learned about major events in cellular evolution and the details of the earliest divergence of Archaea and Bacteria, as well as the origin and subsequent diversification of the eukaryotic cell. Will we one day be able to fully reconstruct the character evolution and timing along the various branches of the TOL using ever improving technologies and analytic tools? And/or will we be able to test some of our hypotheses in the laboratory or computationally and replay some parts of these ancient events to further illuminate our deepest origins? And perhaps predict the future of life? Undoubtedly: as long as humans or similarly intelligent self-conscious life inhabit Earth, we will keep pondering questions about how it all began and will continue into the future.

The evidence base, further reading and teaching aids

Videos:
On the tree of life, eukaryogenesis and Archaea:
https://www.youtube.com/watch?v=hw-ij3822DY
https://www.youtube.com/watch?v=rd37jBXfM4k
On LUCA:
https://www.youtube.com/watch?v=vN0nEHYRD3E

Literature
Review about the tree of life:
Spang, A., Mahendrarajah, T. A., Offre, P. and Stairs, C.W. 2022. Evolving perspective on the origin and diversification of cellular life and the virosphere. Genome Biology and Evolution. Review.

Review about the domains of life:
Williams TA, Foster PG, Cox CJ, Embley TM. An archaeal origin of eukaryotes supports only two primary domains of life. Nature. 2013 Dec 12;504(7479):231-6.

Review on the origin of the eukaryotic cell:
Eme L, Spang A, Lombard J, Stairs CW, Ettema TJG. Archaea and the origin of eukaryotes. Nat Rev Microbiol. 2017 Nov 10;15(12):711-723.

Articles on the last universal common ancestor:
Woese, C. The universal ancestor. Proc. Natl Acad. Sci. USA 95, 6854–6859 (1998).

Weiss MC, Sousa FL, Mrnjavac N, Neukirchen S, Roettger M, Nelson-Sathi S, Martin WF. The physiology and habitat of the last universal common ancestor. Nat Microbiol. 2016 Jul 25;1(9):16116.

Glossary

Aerobic respiration. The oxidation of growth substrates (e.g. sugar, proteins, fats) using oxygen to end products (e.g. CO2 and H2O), via a series of membrane embedded enzymes generating an electron gradient that can be harnessed for the production of ATP (the energy provider of the cell).
Anaerobe. An organism that does not require oxygen for growth. In fact, oxygen is toxic to many (though not all) anaerobes.
Alphaproteobacteria. A group of Bacteria which is thought to have given rise to the mitochondria, the small organelles within eukaryotic cells that are responsible for providing energy for growth.
Asgard archaea. A recently discovered group of Archaea, which is more closely related to eukaryotes than any other so far known archaeal lineage. It is suggested that Asgard archaea once interacted with Alphaproteobacteria and together gave rise to eukaryotes.
ATP. The molecule Adenosine triphosphate is the predominant universal energy carrier in all cellular life on Earth.
Biomarker. Biological molecules that are indicative for certain life forms of life processes. Their presence in geological samples can indicate the existence of a certain form of life or metabolic process at the time of burrial.
Cells. Are the small structural units of organisms. All cellular life is comprised of cells that contains genetic material with the information for growth and replication. See also: <https://www.nature.com/scitable/topicpage/what-is-a-cell-14023083/>
Cyanobacteria. A group of Bacteria whose members are able to obtain energy for their living from sunlight through the process called photosynthesis. Some early eukaryotic cells have taken up cyanobacterial cells, that allowed these eukaryotes to use sun light for growth as well. For examples, algae and plants, which live from the energy of sunlight, have tiny remnants of cyanobacteria within their cells referred to as plastids, which are responsible for photosynthesis.
DNA, Deoxyribonucleic acid and RNA, Ribonucleic acid. DNA is a polymer composed of repeating subunits of the same basic molecule, i.e. so called nucleotide monomers. Each of those nucleotides is itself composed of one of four nitrogen-containing nucleobases, a sugar called deoxyribose, and a phosphate group. In RNA polymers, the sugar is ribose instead of deoxyribose and one nucleobase is different. While DNA occurs as large double-stranded helices that comprise the genomes of cellular organisms, the various RNA polymers of cellular organisms, such as the messenger, transfer and ribosomal RNAs (mRNA, tRNA, rRNA) are generally single-stranded and serve as intermediates between the DNA (genetic code) and proteins.
Fermentation. The production of ATP by substrate-level phosphorylation rather than using respiratory processes.
Fossils. Preserved remains or traces of organisms in the geological record such as bones, shells, exoskeletons, stone imprints etc.
Genes. A DNA sequence that encodes the information of a gene product, i.e. RNA or a protein. Proteins consist of amino acids, carry out metabolic reactions in the cell, and are responsible for the functions characterizing an organism.
Genome. The genetic information of an organisms comprising all of its genes. It can be organised into one contiguous strand (usually one circular chromosome in prokaryotes) or multiple (often linear) chromosomes (e.g. in eukaryotes).

Horizontal gene transfer. The exchange of genes between organisms (living at the same time in the same environment) that are distantly related to each other. Such horizontal gene exchange can blur some of the signal of vertical evolution, i.e. of genes that have evolved through transfer from parent to offspring over the course of evolution. The latter type of genes are ideal markers to reconstruct a tree of life.
Hydrogen, H_2. A gas which represents an important food source for microbial organisms, especially in environments with little or no oxygen such as in sediments.
Hyperthermophile. An organism with an optimal growth temperature around or above 80 °C.
Lipid bilayer. A membrane composed of two distinct layers each of which contains various individual lipid moieties. Such membranes, amongst others, enclose cells in which case they generally also contain molecules other than lipids such as transporters or membrane-bound enzyme complexes.
Metabolism. Life-sustaining chemical reactions within organisms that allow the production of ATP (the energy provider of the cell) and its use to synthesize building blocks of the necessary cellular functions such as growth and replication.
Mitochondria. Membrane bound organelles/compartments inside eukaryotic cells, which are thought to be derived from Alphaproteobacteria. They function as the powerhouse or engine of eukaryotic cells, i.e. they provide eukaryotic cells with the energy needed for growth.
Morphology. Form and structure of life forms/organisms.
Multicellular. Organisms consisting of many cells, all with the same genetic information.
Nucleus. Another membrane-bound organelle of eukaryotic cells, which contains the genetic material in form of a genome. Most eukaryotic cells have one nucleus (but several mitochondria).
Organelles. A specialized subunit or compartment within a cell, which is usually enclosed by a membrane of its own.
Parasite. A symbiont, that has negative effects on its host organism, like for example a virus or bacterium that causes harm for the host organism.
Pathogen. Any parasite that can cause disease.
Phylogenetic tree. A simple branching diagram (or network) similar to a family tree generated on basis of comparisons of morphological characteristics or DNA/protein sequence data from extant organisms, that illustrates the relationships between genes or organisms. The root is the common ancestor, while the tips represent the characters or sequences of extant organisms. Internal nodes reflect intermediates or hypothetical ancestors for a subset of the tips. The branch lengths are relative to the amount of change that characterizes a sequence and is estimated based on a model of evolution specifying the parameters of the evolutionary process.
Plastid. Organelles in eukaryotic algae that enable the harvesting of sunlight for energy conservation and are originally derived from Cyanobacteria.
Primary domain of life. The domains of life that directly (vertically) evolved from the LUCA, i.e. Archea and Bacteria.
Protein. Large molecule that consists of different amino acids arranged in a specific sequence that is dictated by its gene. All the genes of the cell are organized in chromosomes and collectively make up the cell's genetic information or genome. The different proteins of an organism determine its function and biology. For example, there are structural proteins that are used to build the cell, while others are enzymes that carry out cellular reactions that metabolize food sources or replicate the genome etc.
Ribosome. A macromolecular machine present in all cellular life that is responsible for the synthesis of proteins from genes encoded in the genome. It consists both of a large number of individual ribosomal proteins as well as ribosomal RNA components.
Secondary domain of life. A domain of life that did not vertically evolve from LUCA but emerged through the fusion of two distinct representatives of the primary domains of life. Most recent models suggest that eukaryotes have evolved through such a fusion event such that they have been referred to as secondary domain of life.

Symbiont. An organism living in a symbiotic relationship, i.e. in a long-term interaction with another organism of a different kind. When both partners have a positive effect from the symbiotic interaction, the relationship is mutualistic. However, if only one partners profits and causes harm to the other partner, the interaction is parasitic.
Thermophile. An organism with an optimal growth temperature between 50 and 80°C.
Transcription. The process in which the information encoded in a gene, i.e. a fragment of the DNA, is transcribed into messenger RNA.
Translation. The process in which the information encoded in the messenger RNA is translated into a protein.
RNA (tRNA, mRNA and rRNA). see DNA.
Tree of life. A tree that depicts the relationship of all living cellular organisms to each other.
Unicellular. Organisms consisting only of one cell.