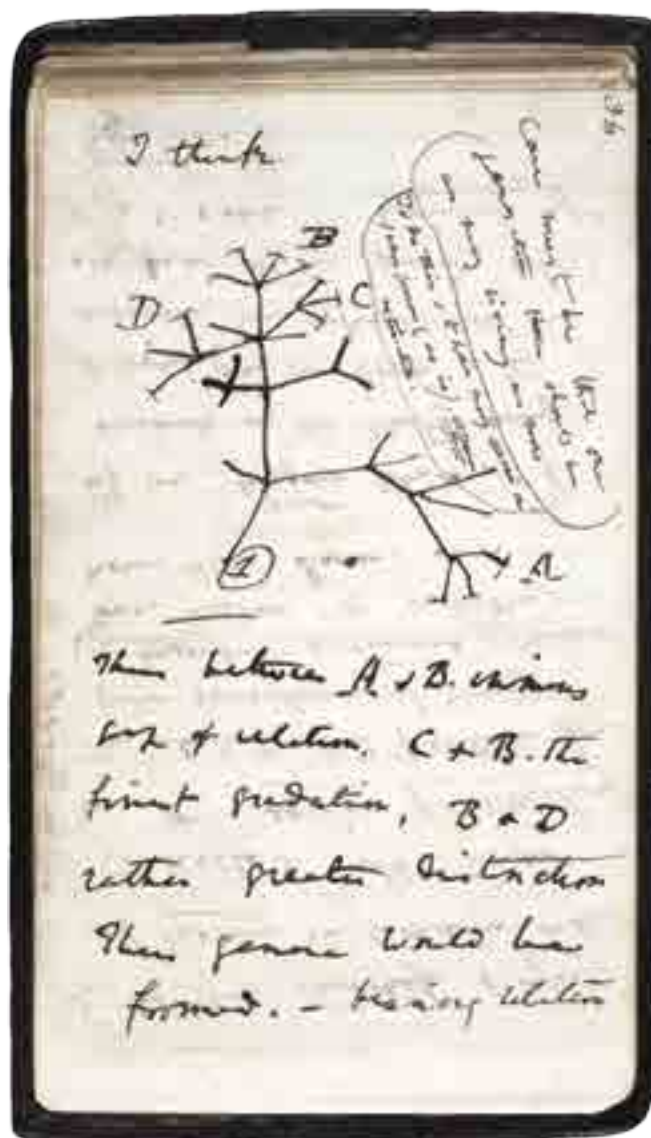# THE TREE OF LIFE:
THE POWERFUL MATHEMATICAL IDEA AT THE HEART OF EVOLUTION

MUKUND THATTAI

The way in which all living things — from microscopic bacteria to human beings to giant sequoia — are related to one another follows a deep and unexpected mathematical pattern, which we now know as the 'tree of life'. This pattern was discovered by naturalists over centuries, but it was Charles Darwin and Alfred Russel Wallace who realised that it held the key to understanding the origin and diversity of life on Earth.

We humans, like many other animals, have an instinct for finding patterns in the world around us. Our survival depends on this instinct — it allows us to separate friend from foe, to track the rhythms of the seasons, to plan ahead based on past experience. Science, or natural philosophy as it was once known, is based on this same instinct — it is an attempt to reduce the overwhelming diversity of the observed universe into reliable patterns, which we explain in terms of laws of nature. Ironically, our hard-wired skills of pattern detection are often at odds with the scientific ideal — we see patterns where there are none; we see structure in randomness. The history of science is the history of how we learned to distinguish true patterns from mere illusions.

Box 1. Observations, patterns, and explanations:

Most successful scientific theories have developed in three stages. The first to come are the **observations:** an enormous list of facts about the world as we find it; or of the world perturbed through experiments. These observations are often confusing and chaotic — it is unclear how they relate to one another; and difficult to predict what the next observation will point to. It is like watching an artist working on a canvas, filling part of the canvas here, another part there — the colours and shapes do not seem to make any sense, and we cannot guess what the subject of the painting is. As enough observations accumulate over the course of time, however, broad **patterns** start to become evident. There is often a singular moment in which the canvas suddenly gels into view, and we realise what we are looking at. In science, this realisation can often be stated in mathematical form, a simple set of rules or equations that summarise the broad structure of the observations. This, then, is the turning point. The final step is to provide an **explanation** of what we are seeing. This is the step where

the viewer grasps the meaning of the painting, the idea the artist is trying to convey. For a scientist, this goes beyond a mere mathematical summary and involves looking for deep causes for the patterns. The greatest triumph of science has been to expose these deep causes as a handful of inviolate laws of nature, and not the capricious actions of some hidden cosmic artist. This is what allows science to be predictive rather than merely descriptive.

A few examples (Table 1) will make this process clearer. In the 1500s, the nobleman Tycho Brahe recorded the motions of planets across the sky with unprecedented accuracy. By using Brahe's data, the young mathematician and astronomer Johannes Kepler discovered his famous elliptical patterns of planetary motion. It took the genius of Isaac Newton to realise the meaning of Kepler's laws in terms of a deeper and more universal theory of mechanics, published in his book *Principia Mathematica* in 1687. Newtonian mechanics heralded the birth of the modern scientific era.

Similarly, in 1789, Antoine Lavoisier identified 33 chemical elements based on his study of chemical reactions, but was unable to find a simple description of their properties. In 1869, Dmitri Mendeleev showed that an element's chemical properties depended not on its atomic weight, but on its atomic number — its numerical position in the periodic table. However, the meaning of this pattern only became clear with the discovery of sub-atomic particles in the 1900s — it was the number of protons and electrons, not neutrons, which determined an atom's chemical properties.

The 1900s saw multiple scientific revolutions that overthrew centuries-old theories. The mathematical patterns in Maxwell's equations, which summarised nearly a century of observations about electricity and magnetism, led Einstein to discover special relativity. It was Rydberg's mathematical pattern describing the hydrogen spectrum that set the stage for a quantum-mechanical explanation of the universe. Relativity and quantum mechanics were combined through the 20th century to form our most precise theory of the nature of matter, known as the Standard Model. This model, too, is based on deep mathematical patterns of nature, known as symmetries.

The key point in all these cases is this — once a summary of data is available in the form of a mathematical pattern, the scientist can step away from observational details and begin searching for a deeper, and often simpler, explanation.

Observation	Mathematical patterns	Explanation
Planetary motion	Kepler's laws, $T^2 = R^3$	Newton's mechanics
Chemical properties of elements	Periodic table	Atomic theory
Electricity and magnetism	Maxwell's equations	Special relativity
Spectrum of hydrogen	Rydberg's law $\frac{1}{\lambda} = R\left(\frac{1}{n_1^2} - \frac{1}{n_2^2}\right)$	Quantum mechanics
Classification of living things	Tree of life	Evolution

Table 1. The development of scientific theories can be broken into three stages: observations, patterns, and explanations.

L'HOMME.
Orang-Outang.
Singe.
QUADRUPEDES.
Ecureuil volant.
Chauvesouris.
Autruche.
OISEAUX.
Oiseaux aquatiques.
Oiseaux amphibies.
Poissons volans.
POISSONS.
Poissons rampans.
Anguilles.
Serpens d'eau.
SERPENS.
Limaces.
Limaçons.
COQUILLAGES.
Vers à tuyau.
Tignes.
INSECTES.
Gallinsectes.
Taenia, ou Solitaire.
Polypes.
Orties de Mer.
Sensitive.
PLANTES.
Lychens.
Moississures.
Champignons, Agarics.
Truffes.
Gommes & Coralloïdes.
Lithophytes.
Amianthe.
Talcs, Gyps, Sélénites.
Ardoises.
PIERRES.
Pierres figurées.
Cristallisations.
SELS.
Vitriols.
METAUX.
DEMI-METAUX.
SOUFRES.
Bitumes.
TERRES.
Terre pure.
EAU.
AIR.
FEU.
Matieres plus subtiles.

The classification of living things

Fig. 1. The Great Chain of Being.
Credits: Charles Bonnet, Wikimedia Commons. URL: https://commons.wikimedia.org/wiki/File:BonnetChain.jpg. License: CC-BY.

There have been many attempts to classify the diversity of life on Earth. One of the earliest, and most common, ways was based on quasi-religious grounds, and known as the Great Chain of Being (refer **Fig. 1**). Most commonly associated with Christian scholarship, similar attempts are seen in the Hindu *puranas*, and in ancient Greek and Egyptian philosophies. All versions of the Great Chain place living things on a strict ladder — at the bottom you have minerals and non-living matter; the next to come are the simplest forms of life, for example, the microbes (a modern addition); then plants; then animals; then, above all of these, human beings; and, sometimes, angels and deities above humans. This is a beguiling pattern, and fits with our natural inclination to place ourselves on top. Unfortunately, it is incorrect. It is not based on rigorous observations, but rather arises from our desire to make the universe fit with our own preconceptions.

A scientific approach to classification, firmly grounded in agnostic observations, is known as taxonomy. We first gather and record a huge volume of detailed information about the properties of every known living thing. We then start with the difficult task of separating these organisms into groups based on their properties. But, when we do this, we immediately encounter a problem: different authorities use different yardsticks; different traits by which to make the groupings. Some choose complexity, some choose size, some choose the mode of living; and some choose habitats. Which one of these is correct would seem to be a matter of faith or opinion, rather than fact; and each choice produces a different taxonomy.

Classification games reached a fever pitch in Europe in the 1600s. It was a time when European power reached across the world. Collections of exotic animals and plants, known as menageries, were brought from the farthest corners of their empires to their imperial capitals for the amusement of the citizenry (refer **Fig. 2**). I like to imagine a great hall in which, neatly stuffed and mounted, specimens of every type of animal is scattered across the floor. Amateur taxonomists wander the hall, moving these specimens here and there, and generally trying to out-do one another in terms of their grouping. How would this play out? One person might arrange everything based on colour, only to be upset the next day when someone else rearranges everything based on size. Everyone is at loggerheads with one another. Then, something remarkable happens. Some of our taxonomists, just for fun, start to use increasingly obscure traits as the basis of classification. Instead of size or colour, these taxonomists look at the number of bones in the ear, the arrangement of holes in the hip, the layering of muscles on the toes. Of course, some of these people still disagree with one another, and leave the hall in frustration. But slowly, a large group builds up in the hall, each person quietly working in their corner while the broader classification remains largely unchanged.

What is going on? How is it that a huge number of people using completely independent yardsticks suddenly start to agree? This is where we first notice a deep mathematical pattern. Let's consider any three animals in the hall, say X, Y, and Z; and two taxonomists A and B each using different traits for classification. We ask A, and find she believes that {X, Y} form one group and {Z} another, which we write as {{X, Y}, Z}. Suppose that B believes {Y, Z} form one group and {X} another, which we write as {{X}, {Y, Z}}. In this case, A and B will never be able to agree. Suppose, instead, that B believes {X}, {Y}, and {Z} represent three distinct

Fig. 2. A European menagerie of exotic animals.

groups, which we write as {{X}, {Y}, {Z}}. No problem, says A: all B has done is further sub-divide her classification. For example, A might be thinking of the animals as insects {X, Y} and birds {Z}, while B might be thinking of beetles {X}, bees {Y}, and birds {Z}. That is, we can write these groups down in a *nested list* {{{X}, {Y}}, {Z}}, and both A and B will be happy. More generally, two classifications A and B are said to be **consistent** if, for every choice of three objects X, Y, and Z, disagreements do not occur. What happened in our game (refer **Fig. 3**) is that thousands of taxonomists found their preferred groupings consistent with one another and, over the course of time, divided the entire hall into a series of nested groups and sub-groups.
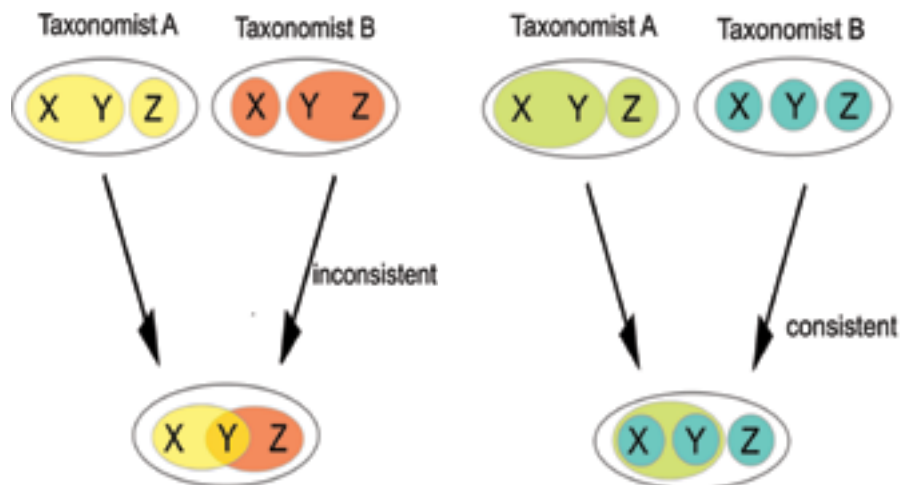

Fig. 3. The classification game.
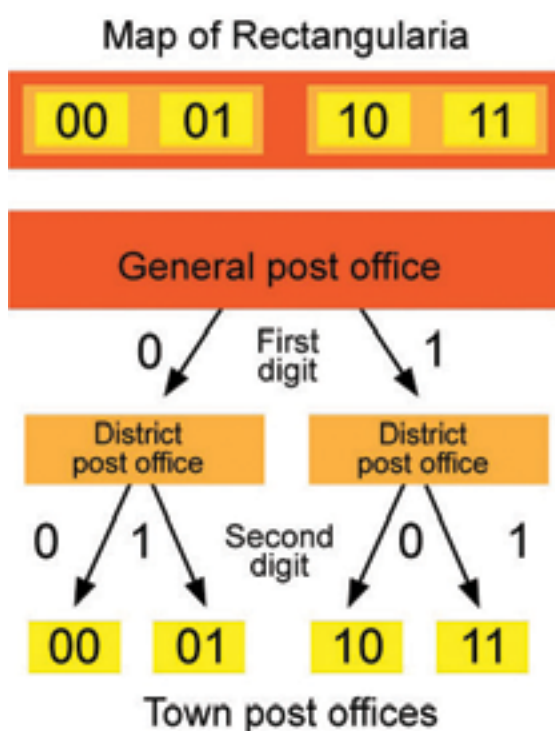
Map of Rectangularia

Fig. 4. The nested postal system of Rectangularia.
Credits: Mukund Thattai. License: CC-BY-NC.

Now comes the final piece of the puzzle. A process very much like our imaginary game played out over the course of the 1700s. When the dust settled, it turned out there was a single **unique** solution to the classification problem. There was one giant set of thousands and thousands of traits (mostly very obscure ones) each consistent with one another; and with traits (size and colour, for example) that were inconsistent, ignored. This was a fact, not a matter of faith or opinion. The ultimate expression of this fact is found in Carl Linneaus's book, Systema Naturae, published in 1735. From that date onwards, the Linnaean system of taxonomy became the single correct and accepted biological classification system. Whenever a new species was discovered, there were typically initial disagreements about where to place it but, eventually, the preponderance of evidence based on multiple traits would reveal its correct placement.

Darwin's and Wallace's insight: nested groups are also trees

By the time Charles Darwin set sail on the HMS Beagle in 1831, the classification of all living beings into nested groups was well established. Many naturalists, including Charles's grandfather Erasmus Darwin, had already realised how surprising this mathematical pattern was. Let's pause to consider this, by trying to classify some other type of objects. Furniture can be classified by size, shape, material, colour, and use; but you will never find any agreement among the different groupings. Words can be classified into nouns, verb, adjectives, and so on, but this system is not nested; it operates at a single level. Sounds can be classified by pitch and volume, and in a more modern sense by spectral components, but this is an additive system, not a nested one.

We can flip the problem around, and ask what kinds of things are usually found in nested groups. Here we have a very familiar example. Countries are broken up into postal codes, allowing mail to be efficiently delivered. Typically the left-most digits of the postal code represent large subdivisions, and the right-most digits represent small sub-divisions (refer **Fig. 4**). If we dig a little deeper into what these digits actually mean, we will find that each one is associated with a real object, namely a post office. Starting from the General Post Office, each digit going from left-to-right is the name of a district level, a town-level, and a street-level post office. Suddenly it becomes clear that a nested list is really a tree in disguise!

Each node of the **tree** is a post office, and arrows show how mail flows from higher to lower levels.

What does this mean for biology? Taxonomists have always grouped all the living things known at a certain moment, implicitly assuming that this set never changed. Imagine a world in which all these species were designed at the moment of creation and remained unchanged. It would then be hugely surprising to find a single unique nested classification system. A world of fantasy animals whose traits were all jumbled together would look more like furniture, and be unclassifiable.

Darwin and Wallace independently suggested that the Linnaean classification system be seen as a tree rather than as nested groups. Darwin's study of finches, and Wallace's identification of biogeographic regions, both emphasized the role of the passage of time. This made a profound difference, because it provided a mechanistic explanation for the tree-like pattern. If we think of the arrows of the tree as representing the flow of time, then the relationships between all living things **today** (the nested list) tells us a great deal about the **past** (the ancestral limbs of the tree)! If animals and plants could change over time, accumulating trait differences from parents to offspring, then the nested grouping of present-day traits is easily explained. Of course, this is only the beginning. We would have to provide some explanation for how these traits changed, how heritable they were, or how certain traits were selected over others in each generation. It is only in the modern biological era that the molecular mechanisms underlying heritability and genetic encoding were elucidated. This collection of ideas together represents the modern synthesis of the theory of evolution by natural selection, and it began by recognising the tree of life.

"I think..."

In Darwin's notebook, a sketch dating from 1837 represents his first depiction of a tree of living things. It is famously captioned *"I think"*. Darwin's 'Origin of Species', published in 1859, contains a single figure, also representing the tree of life (refer **Fig. 5**). So what did Darwin think?

We can never know for sure. But, it is reasonable to guess that Darwin had just realised the connection between nested groups and trees. Both represent equally valid ways of summarising a large volume of observations about the traits of living things. However, the nested groups suggest immutability (unchanging), while the tree suggests **transmutation**, the gradual changing of one species to another. Once this was clear, Darwin immediately realised that not all branches of the ancient tree survive to the present day — there must have been strange forms of life

in the past that have vanished since then, without a trace. Most importantly, Darwin had a further insight. Just like the nodes of the post-office tree represent real physical buildings, the internal nodes of the tree of life represent something real as well. Each of these nodes is an ancestral creature — microbe, plant and animal — that must have lived and died billions of years in the past. In other words, this nested taxonomic classification implies the existence of intermediate forms in the fossil record, buried in layers of rock as a record of the past.

Darwin's and Wallace's theory of evolution has been put to the test, and passed every challenge. The tree of life (and the process of natural selection), which was first developed to explain the diversity of plants and animals, is now known to apply to every form of cellular life on earth — including prokaryotic bacteria, archaea and single-celled microbial eukaryotes. There have

been a few interesting deviations — we know now that cells can exchange DNA; and that two species can hybridize to make a third. But these processes are just embellishments on the central tree. The global set of consistent traits now ranges from classical macroscopic measurements to molecular-level information. Literally each base pair of a cell's genome stands witness to the process of evolution. Beginning from Carl Woese's first attempts at molecular classification in the 1970s, it has now become routine to find the location of an organism in the tree of life by the DNA evidence alone.

Understanding life's history teaches us that each living form is precious and represents a single unbroken path reaching into the past. Evolution is on-going — the processes that generated life's present diversity continue to operate. Evolution can happen in a matter of hours or billions of years; humans continue to evolve, just like all
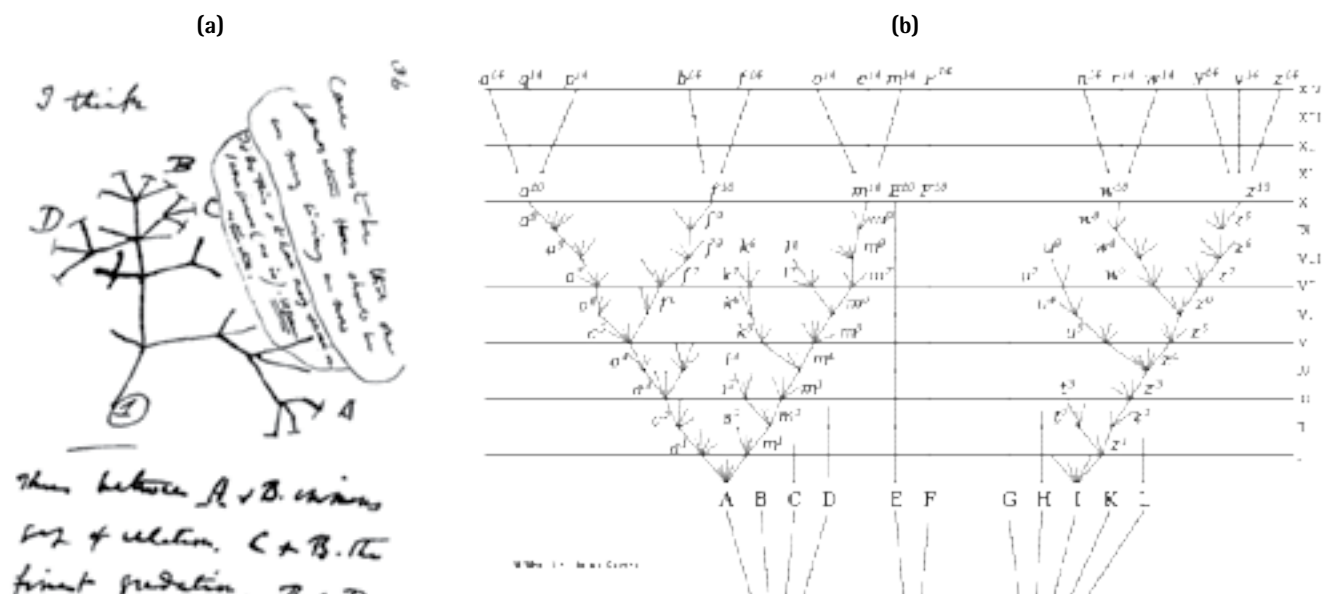

(a)

(b)

Fig. 5. Darwin's Trees.

(a) From his first notebook on Transmutation of Species (1837).

Credits: Trockennasenaffe, Wikimedia Commons. URL: https://commons.wikimedia.org/wiki/File:Darwin_tree. png. License: CC-BY-SA.

(b) In On the Origin of Species (1859).

Credits: Charles Darwin, Wikimedia Commons. URL: https://commons.wikimedia.org/wiki/File:Origin_of_ Species.svg. License: CC-BY-SA.

other organisms. New species continue to arise; but many more are wiped out by human activity in the sixth great extinction of life on Earth. This scale of extinction is unprecedented and irreversible. We have become poor custodians of the tree of life — preserving the diversity of living things is the single greatest challenge of our age.

— ◦ ◇ ◦ —

Mukund Thattai is a researcher at the National Centre for Biological Sciences. Trained in physics, he now studies how cells evolved over billions of years. He is deeply involved in public engagement, working with artists and theatre practitioners to explore the practice of biology and its impact on society. He can be contacted at thattai@ncbs.res.in.