The natural universe

Part I – The deep past: From the Big Bang to Homo sapiens

John O'Neall

Table of Contents

1. Introduction: Why?

1.1. A bit of autobiography

It was some time around 2 or 3 in the morning (or, rather, the night) and I was sleeping soundly, like most 12-or-so-year-old boys do at that hour, when I was awakened by a gentle nudge and a voice speaking softly, “Come on, get up.” It was difficult, but I managed to get my eyes open and saw it was Daddy – my sister and I continued calling him by that name even when he was 102 years old. “It’s almost time for the eclipse,” he insisted.

Oh right, there was to be an eclipse of the moon and he had got himself up at a horrible hour and then waked me, as promised, so we could watch it together. In our PJs – It was warm in Florida at the time – we two, accompanied by our pooch Suzy, walked out onto the front lawn. The moon shone beautifully, but soon a slightly rounded shadow came over it from the left, spread slowly across and then covered it completely. What a thrill that was to see! So it was true, what I had been reading in the *Golden Book of Astronomy*. Proven by my first experiment – well, sort of.

And it was all science. Daddy was a lapsed (very lapsed) Quaker, so my sister and I were fairly protected from Mother’s non-Calvinist version of Presbyterianism. For us, the eclipse was a completely natural thing.

Another time, Daddy took me fossil hunting in the mostly dried-up bed of a stream in southern Ohio, near where the family farm had been before Grandpa and Grandma moved to Florida. We found lots of trilobites, some ammonites and some other strange-looking creatures resembling branching tubes – probably some kind of prehistoric sponge. Decades later, we went back there with my wife, Siv, and found more fossils, though not so many.

I was (and still am) a reader and that overlapped with science. As a kid, I regularly read three magazines – “Popular Science”, “Popular Mechanics” and “High Fidelity” – all purchased from the local drugstore for less than the current price of a postage stamp. The last, I perused to read about musical works and identify those I might like to purchase on records (LPs, in those days). Anything which sounded as if it would feature lots of brass and drums would be high on my list of candidates. And I learned a lot about music by playing the clarinet and, later, the alto saxophone, in the junior-high and high-school bands.

Toward the end of my high-school years, I was subjected to a battery of aptitude tests which purported to determine that although I loved science, I loved music more and therefore should study music. True to adolescent behavior, I chose to study physics. I have read since that young minds work that way because the decision-making frontal cortex is still far from fully formed (a subject we will probably get around to later on). After getting a doctorate in particle physics and then working five years in two different labs, I switched from physics to computing. But I have continued to be fascinated by science and music ever since. Now, in retirement, I have time to read, to refresh my memory and to catch up some on the huge quantity of discoveries that have been made since I dropped physics – discoveries not just in physics, but also, notably, in paleontology and biology. In fact, I learned almost no biology until I started reading about it at age 70.[1] And the things I have learned about are frankly mind-blowing.

I realized this the day I first read about the electron transport chain (or ETC), or oxidative phosphorylation. That's a mouthful, but the basic principals are simple as 2+2=4.

If you want to know already, the ETC is a series of complex reactions which take place in the inner membrane of organelles (thingies inside cells) called mitochondria. Among other things, what these reactions do is pump positively-charged protons across the membrane to set up an electric potential – a battery! And when the battery is sufficiently charged, it discharges to power a kind of mill, rather like the pepper mills in our spice racks, set in the wall of the mitochondrion. Unlike a mill which is driven by water to power a device for grinding up grain, this mill is driven by the electric potential to power a machine which puts things together – phosphate and something called ADP (the "D" is for "di", meaning two, in this case, phosphates) – into a similar thing with three phosphates and called, logically enough, ATP (where the "T" is for "tri", of course). And ATP is a miniature batter in itself, which can be shuttled around and used to power other processes, such as contraction of muscles. Simple, *n'est-ce pas*? It’s just physics. (Well, not just...)

1 I’m getting to know some more music too, especially chamber music and jazz.

The discovery – at least, I saw it that way personally – was for me a moment of illumination and surprise and even jubilation. It was so simple, so practical.

Since then, I have come across other such "eureka" moments in science reading:

- the gorgeous, lacy filament structure of the distribution of galaxies on an unimaginably large scale;
- the regulation of protein expression by transcription factors – in plain English, how a cell's environment tells it which DNA recipes to use to make proteins, so a heart does not manufacture liver proteins, for instance;
- the fact that what we see out there – cats and beautiful landscapes and cabbages – are not what is really there, but our brain constructs that impression. (Should that then be called an illusion? Some scientists think so.);
- the mechanism of plate tectonics and the movement of the plates around the Earth (but not other planets) since their formation, including the formation at least once of one single giant supercontinent, Pangea;
- the way stars are formed and maintain themselves in equilibrium (balance) between outward pressure from the "burning" interior and the inward gravitational pull of the huge amount of matter composing the star;
- the sodium-potassium pump, another thingy in the membrane of cells, which pumps sodium ions out of and potassium ions into the cell – but in differing numbers, so as to set up another electric potential across the cell wall, like a battery, and how this can lead to a burst of electricity (an "action potential") which then travels down a neuron and initiates the release of chemical substances which tell another neuron to do the same thing, and so on in series until a circuit sends a message to, say, a muscle or maybe commits something to memory. (By the way, what that may mean is also fascinating);
- the ingenious way retinal cells massage light signals to limit traffic over the optic nerves;
- and lots more.

I also am impressed by the way scientific research is carried out and shared. In spite of some personal or national differences, science comes as close as there is to a societal enterprise. What with peer-reviewing and repeating of key experiments, we can be pretty damn sure about what we have learned. Which is not saying we know everything. But even if another theory one day supplants the current one in some field, it will have to start by explaining the same things the current one does, so we really will not have lost anything – except some ignorance.

1.2. Why this document?

My love of science, by which I now mean the findings of science[2] concerning … well, most everything, began early and was nurtured not only by my parents but by my friends and teachers. I feel very privileged not only by that, but about the situation and place in which I was born, where I was able to grow up having the time, interest, encouragement and resources to live through what was – as I now remember – a delightful and satisfying childhood. What surprises me though is that lots of people now have similar privileges, but – for some reason, education or parental encouragement or something else I can only refer to as *Zeitgeist*, mentality of the age – never learn much about the natural world. Not only that, but many don't seem to care. Yet, as has often been pointed out, current findings in science can help us to make a paradise or a hell of whatever may be left of the Earth after what we are doing to it.

Many scientific endeavors are being prepared or carried out which may affect our future in quite a radical way – by public and private enterprise and by the governments of the world. Think of the use of pesticides, nuclear power, weapons of global mass destruction, environmental tampering by destruction of forests and farmland, over-use of antibiotics, nanotechnology, genetic engineering and still more. There should be informed discussion of these subjects, with all of us participating. The risks we run are exacerbated by our political structure: Our political representatives may not be able to make informed decisions because of their ignorance, and we may not pick informed representatives because of our ignorance, or indifference or

2 I am talking about the findings of science, whereas the basic importance of science depends on its methods. It's not just one more postmodernist story.

vested interests. Added to that, it is not obvious that the powers that be even want "us" to participate in any discussion.[3] The way the governments of Europe ignored the referendum-expressed voice of the people concerning the Lisbon Treaty around 2005 provides clear evidence of their mistrust of us.

Carrying out informed, intelligent decisions on the uses of the findings of science and technology depends on many factors, the two most important being, in my opinion, educational and political. Alas, I have little influence on the latter, which seems to have fallen mainly into the hands of an extremely restricted group of people more interested in short-term gain and economic power than in conserving something of the Earth for their own children and grandchildren, not to mention the preterite[4], those who are not members of the elite group. But what about the former? Since making meaningful decisions as to how to use the results of scientific inquiry requires being aware of those results, I would like to make whatever contribution possible to spreading knowledge of those facts. And that is why I intend to write this document.

All that is the *raison d'être* of this document-- to get ready to assess these subjects for ourselves. In order to get through it all, we will firmly and resolutely stick to the point of view that our best — and most certainly only — means of learning about the universe is through science, by which we mean a specific method developed to help us apprehend "reality" without kidding ourselves. That method is based on incessant questioning and constant testing.

For millennia, men have invented mythologies in order to furnish superstitious explanations of things. One far-out one begins with the Hindu god Vishnu asleep on his coiled-up snake Ananta on the calm cosmic waters.

Figure 1.1: Vishnu napping on the serpent Ananta. Note the lotus stem and Brahma.[5]

Suddenly, a lotus plant sprouted from his navel. When the lotus flower opened, the god Brahma was revealed sitting on it. Brahma created male and female by splitting off parts of himself. In this way, he went on to create all living things, restoring himself after each split. When Brahma went to sleep at night, everything was destroyed, but Brahma recreated it anew the next morning. Since a day of Brahma's time is billions of years of human time, no one notices. But even Brahma lived out his life and then the god Shiva Nataraja, the King of the Dance, performed his cosmic dance and everything was completely destroyed. Until Vishnu yawned and stretched and another lotus started to grow out of his navel...

Isn't that great? A lot more interesting than the Judeo-Christian creation myth. Unfortunately, it isn't true either.

1.3. The program

So what does science have to say about the world we live in and about us ourselves?

Many popular books talk about the history of science and scientists, even though their titles may suggest differently. As interesting — even fascinating — as such topics are, it seems to me it ought to be possible to grasp the basic scientific world-view without plowing through all those false starts and errors, tests and

3 Am I a bit paranoid here?

4 The non-elect in religions touting predestination. I picked up the term from Thomas Pynchon's incredible novel *Gravity's Rainbow*.

5 Carving on a rock along the Tungabhadra River in Hampi, Karnataka, India. Photo by author

experiments, names and dates. Why not simply present current scientific notions without going into the gritty details of why and how we have come to accept them as valid? A simple description. Afterwards, there are lots of books for those who want to know more. (Some are listed in the bibliography.)

That is what I propose to do: Take a look at what we are, where we are and how it works, and all that from the point of view of scientific knowledge. I will try to make it simple and will not go into the math, although one or two formulas may pop up. You will certainly recognize one of them – you might even have it on a T-shirt.

In order to understand the what and the where, we start with the wherefrom and work our way up to the "Pale Blue Dot"[6] we call Earth. We will go through ideas from the following subjects:

- ***cosmology*** and ***astronomy*** – the formation of matter; the birth, life and death of stars, and the subsequent formation of other stars and of solar systems and planets;
- ***geology*** – the evolution of the Earth itself, its surface and atmosphere, plate tectonics, what is below the surface and beyond the atmosphere, climate and the evolution of fauna and flora;
- ***biology*** and ***molecular biology***, ***paleontology*** – the study of life and its development, to learn about the evolution of living organisms, including man;
- ***physiology*** – the study of cells, tissues, organs and organisms;
- ***neuroscience*** – the nervous system and the body's control center, the brain.

Our path will go by way of certain landmarks, theories without which we would never find our way to the understanding which we have today: Quantum Mechanics, Special and General Relativity, Plate Tectonics and Evolution through natural selection.[7] As a (very) former physicist and *informaticien*, I am naturally interested in energy and communication and these will be two guiding threads along our route.

So there it is. We want to explore what we currently know about "it all", starting some 13.7 billion (10^9) years ago. There will be few, if any, explanations of how we came to "know" it, or detailed analysis of the gaps in our knowledge, although some of these will be mentioned in order to keep things "fair and balanced". Names of scientists will be avoided. We will just look at the currently prevalent world-view of scientific thought as I understand it As such, it is certain to be somewhat out-of-date and subject to change – and maybe soon.

The following table may be useful, providing a sort of schematic outline to subjects.

Time scale	Space scale	Subject
13.7 Gya to future	The Universe	From the Big Bang to galaxies and beyond
4.7 Gya to today	The solar system and Earth	The origin and history of life
Today	Earth	Us, Homo sapiens

The first two parts are also the initial subject of the relatively recent discipline of "Big History".[8] I think I came up with my ideas before I read about that, but we shall see that the brain often plays tricks with our memories.

Now a word of warning. As already pointed out, it is true that scientists do not understand everything. Probably every branch of science contains unsolved problems, many of which lead to heated discussions among the advocates of different hypotheses. Here are just a few outstanding questions.

- Physics: What are dark matter and dark energy? How do we mate quantum mechanics, the science of the very small, with General Relativity, the science of gravity and the very large? Is information conserved or lost in black holes?
- Geology: What is the driving force of plate tectonics? Is it mantle plumes?

6 A photograph of Earth taken by Voyager I from a record distance of about 6 billion km inspired Carl Sagan to use this term. See https://www.youtube.com/watch?v=p86BPM1GV8M.

7 Please appreciate the capital letters!

8 Christian, David. *Origin story.*

- Biology: How did life begin? And where? On the sea bottom, in volcanoes or … out in space?
- Neurobiology: What (where) is consciousness? When are we conscious of an impending act? Do we have what is commonly (and somewhat vaguely) called "free will", or are all our acts determined by previous causes?[9]

That does not mean scientists do not know anything with certainty. Scientists know a helluva lot and generally reveal the limits of their knowledge in terms of the standard deviation, a quantitative measure of the uncertainty in measured results. The lack of answers to the above questions should not lead you to doubt scientists most of the time. For instance, in spite of the lack of a theory of the simutaneously very, very small (QM) and the very, very massive (GR), the so-called Core Theory of physics explains to a great degree of precision everything that matters to us in our everyday lives. After all, you will never see a quark (nor will anyone else) and you most definitely do not want to spend a weekend visiting a black hole.

It hardly needs mentioning that all errors or oversights are the complete responsibility of the author – me.

Now we do need to adopt some terminology. We will be speaking often of events which took place millions or billions of years ago, but I don't want to have to type that out every time. So the following notation will be handy and fairly obvious:

- My = million years
- Gy = billion years (G for giga, meaning 10^9, or 1,000,000,000)
- Mya = million years ago
- Gya = billion (G for giga[10]) years ago.

Giga (10^9) comes from the metric system, which is used by all scientists.[11] So by us, too.

That's the dry part. What's better is that many of the subjects we will look at are brilliant – unexpected and fascinating and mind-blowing! So let's get started. It all began 13.7 billion years ago. Oops, I mean 13.7 Gya... At this stage, it's all about energy…

9 Actually, there is a pretty strong consensus on this subject.

10 In case you weren't paying attention two lines ago – or have already forgot.

11 I did once meet a British astronomer who poo-pooed the centimeter-gram-second system of units and insisted we should use the rod-stone-fortnight system. In fact, that might be better for astronomical scales.

2. Science marches in – basic theories

Four theories of science are essential to our understanding of what our "environment" is and how it and we got that way. These are theories in the scientific sense, bodies of accepted knowledge – not at all the same thing as hypotheses. A theory like plate tectonics or General Relativity is a very Big Deal and not "just a theory." All four of our theories have been tested by innumerable experiments which have found them to be true to nature. This does not mean that they will never be improved upon, but whatever other theory or extension does so must also explain why they are true in current applications.

We generally consider physics to be the most "basic" of sciences, because it is self-contained and explained by no other science. It in turn explains chemistry – and therefore geology – and parts of biology. Thermodynamics, quantum mechanics and relativity lie within the domain of physics. They are explained here, because they are necessary to an understanding of the first subject, cosmology. Evolution through natural selection is also explained here. Plate tectonics only directly concerns geology and so fits well into that chapter. Chemistry is a fairly vast subject, more than a theory, and will be discussed wherever appropriate, with physics as well as with biology.

2.1. Thermodynamics

Thermodynamics is the science of energy in its various forms and transformations, so it is one of our most important and interesting topics.

2.1.1. The importance of energy

Energy is defined classically as the capacity for doing work. (When we speak of "classical" physics, we mean that known before quantum mechanics or relativity.) More generally, energy is the capacity to do things or make things happen. Among forms of "work" for which energy is required are lifting a heavy object, moving electrons through a wire (electric current), setting up electrochemical circuits in the neurons of the brain (thought) or effecting contraction of muscle fibers (to lift that heavy object).

It used to be that physicists talked about forces, but those days are mostly gone. Aside from physics 101 classes, you will rarely see a physicist write a formula for a force. Instead, they use energy, which is the source of forces. They still talk about the four basic forces – strong, weak, electromagnetic and gravitational – but they do calculations using the method of Lagrange or Hamilton, the equations of which are expressed in terms of energies rather than forces. In fact, a force is the rate of change of a kind of energy called potential energy. So energy really is the fundamental subject of modern physics. Our currently ultimate theory is quantum field theory, according to which everything is composed of energy fields.

2.1.2. The laws of thermodynamics

Thermodynamics is stated in the form of three laws.

1. The total energy of the universe does not change; it is always conserved.[12]
2. "The second law of thermodynamics says that energy of all kinds in our material world disperses or spreads out if it is not hindered from doing so. Entropy is the quantitative measure of that kind of spontaneous process: how much energy has flowed from being localized to becoming more widely spread out (at a specific temperature)."[13]
3. The entropy of a system at a temperature of absolute zero[14] is zero.

Law number 1 is the justly famous law of conservation of energy. Law number 2 explains why physical processes don't play out backwards in time. It is often expressed as: In a physical process, the *entropy* of the universe always increases or stays the same. Explanation coming. Number 3 says that at absolute zero,

12 There is a problem with this concept in General Relativity. In curved space, we don't really know how to compare energy at different points in spacetime. For the moment, just forget I said that.

13 Lambert, Frank, Entropy is simple – if we avoid the briar patches: http://franklambert.net/entropysimple.com/ or https://www.esalq.usp.br/lepse/imgs/conteudo_thumb/Entropy-Is-Simple---If-We-Avoid-The-Briar-Patches.pdf.

14 Absolute zero is 0 Kelvin, or -273.15° Celsius.

a system is in its state of minimum energy.

Over time, thermodynamics has evolved, adapting itself to new discoveries. First elaborated to deal with steam engines and boring cannons[15], it has evolved to describe such phenomena as electromagnetism and even to apply the entropy concept to information, although its relevance to this last subject is contested.

Actually, after the discovery of these three laws, someone noticed that there should be another, yet more fundamental law, making the laws of thermodynamics a trinity of four. The new one is called the zero-th law.

0. If A is a system in thermal equilibrium with system B, and B is in thermal equilibrium with system C, then systems A and C are also in thermal equilibrium with each other.

This provides a way to measure the thermal state of a system. If system B is a thermometer, this says that if A and C are each in thermal equilibrium with it, they both have the same temperature.

2.1.3. Entropy and the second law

The second law is inseparable from the concept of entropy and that is where complications set in. ***Entropy*** is not really defined in words, but is a mathematical construct. You cannot hold a grain of entropy in your hand or feel it, like something warm or cold. There is no such a thing as a simple entropy meter. Its change commonly is calculated as the quotient of two other quantities (energy dispersed / temperature) or as the logarithm of one more not-really-obvious quantity. The connection between the two formulas is not immediately apparent and neither is intuitive. This has led to the use of analogies to explain a fundamentally mathematical concept, a procedure prone to error, doing which often leads to mistaken ideas and disagreement on the utility of such ideas.[16] The same problem of expressing in words a mathematical quantity will come back to haunt us when we consider the tenets of quantum mechanics. But don't be discouraged. Some of the analogies are not entirely bad.

There are several ways of looking at entropy and each way has its proponents and opponents.

- Entropy as a measure of disorder.
- Entropy as a measure of the statistical probability of a state.
- Entropy as a measure of dispersal of energy.
- Entropy as a measure of thermal energy which is unavailable to do work. This is really a restatement of energy dispersal.

The erstwhile common attempt to explain entropy was as a kind of "disorder", claiming that nature seeks disorder, and therefore increased entropy. Entropy thus serves to predict which direction a process will take. An egg breaks to make a disordered mess, the mess never reforms an ordered egg. But disorder is a word used in everyday life with a meaning which does not correspond well to that used in science and so its use is often ambiguous or vague. What about an explosion of hydrogen and oxygen in the presence of a spark? We will see that such an event is a reason why the disorder analogy is strongly contested by some scientists.

From a statistical viewpoint, entropy is understood as follows.

> *Entropy is a measure of the probability that something composed occurs in a certain state, i. e., with a certain configuration or order of its components.*

Physical, chemical and biological sciences explain that everything is composed of other smaller things, except for the most elementary particles, quarks and electrons[17]. To understand the word "composed", one can imagine an assembly of molecules or grains of something. If we interchange two identical grains, we have a different system but one which we cannot distinguish from the first. These indistinguishable states are referred to as micro-states. One macro-state (like a bowl of sugar) is composed of many different, indistinguishable micro-states, because we do not know which individual grain is which – nor do we care. The number of interchangeable micro-states can be quite large. The probability of occurrence of a macro-

15 The *double entendre* is intended.

16 Haglund, Jeppson and Strômdahl. Different senses of entropy – implications for education. https://www.mdpi.com/1099-4300/12/3/490

17 Buddhists would agree on this (First Seal).

state depends on its number of indistinguishable micro-states. Its entropy is defined so as to take this multiplicity into account. I can't resist writing the formula for this because it is so simple.

$$S = k \log W, \qquad (2.1)$$

where S is the entropy, k is the so-called Boltzmann constant, and W is the number of possible micro-systems which correspond to the same macro-system.[18,19] In other words, the greater the number of possible micro-systems, the greater its probability and its entropy. Greater entropy corresponds to a higher probability. It was this idea of the greater number of micro-states which originally suggested the notion of increasing disorder, since more component particles can be exchanged without our noticing, and this was construed as disorder. Note that this equation is born from statistical mechanics and so represents a probability. Also, it says nothing about the rate at which such a change of state takes place.

But there is a third way of looking at entropy, contained in the way we state the second law: Energy spreads out if it can. This statement is perfectly intuitive. Energy dispersed becomes unavailable to do useful work for a process. It is a question of the quality of the energy. An example of such dispersed energy is that which is lost to friction as heat, which keeps an object from sliding infinitely far on a plane surface. From this point of view, it is not the arrangement of the matter left behind which counts, but the energy dispersed as it pushes the matter into that arrangement.

Back to entropy. The change of state depends on the existence of another state of the system with greater probability due to a greater possible number of micro-states. Then equation (2.1) defines the entropy as a quantitative measure of this probability, which can be calculated (laboriously) by the methods of statistical mechanics. The statistical law is true whatever our interpretation of the entropy. Lambert states this clearly:

> *Thermodynamic entropy change consists of two factors. Entropy change is enabled in chemistry by the motional energy of molecules (or from bond energy change in a chemical reaction) but thermodynamic entropy is only actualized if the process itself (expansion, heating, mixing, reaction) makes available a larger number of microstates, a maximal Boltzmann probability at the specific temperature.*[20]

There you have it. Thermodynamic change depends in a probabilistic way on a greater number of micro-states being available to disperse into and this increase in probability is equivalent to an increase in entropy as defined by equation (2.1).

For the mathematically oriented, the equation of thermodynamics for a reversible change, ignoring chemical energies, is[21]

$$\Delta U = T\Delta S - p\Delta V, \qquad (2.2)$$

where U = total energy, T = temperature, S = entropy (Why S?), p = pressure and V = volume. This can be interpreted as energy change ΔU is equal to work done by useful energy, $-p\Delta V$, plus energy lost, $T\Delta S$. So the entropy is both a measure of the probability of the final state by (2.1) and a parameter of the energy lost by (2.2). Entropy is thus the link between probability and energy dispersal.

The thermodynamics studied in undergraduate classes is usually ***equilibrium thermodynamics***. This requires that any changes which take place do so very slowly, so that they remain almost in equilibrium. We pass from one equilibrium state to another so (infinitesimally) close to it that the intervening disequilibrium can be ignored. Obviously, this does not apply when, for instance, you allow molecules of oxygen O_2 and hydrogen H_2 to mix together in the presence of a spark. The resulting explosion is far from equilibrium and so the resultant gas of H_2O molecules, of which there are only 2/3 the original number of molecules, possesses a smaller number of micro-states and therefore seemingly lower probability and entropy. We must take into account other factors – in particular, the large amount of binding energy of the H_2 and O_2 molecules, which is liberated as heat during the explosion. The lost energy is why many physicists and chemists prefer to speak of entropy as a measure of energy dispersal.[22]

18 W stands for the German word *Wahrscheinlichkeit*, probability. The value of k is 1.38 $\times 10^{-23}$ m^2 kg s^{-2} K^{-1}.

19 If you have forgotten, the logarithm to base 10 of a number is the power of 10 which corresponds to that number. E.g., log(100)=2, because $10^2 = 100$.

20 Frank Lambert, Entropy is simple. http://www.esalq.usp.br/lepse/imgs/conteudo_thumb/Entropy-Is-Simple---If-We-Avoid-The-Briar-Patches.pdf

21 This is hyper-simplified and sweeps all sorts of things under the rug.

22 Lambert, op. cit.

How about perfume spreading spontaneously through a room, or cream through a cup of coffee? There is no loss of energy, so how does the second law apply? There are answers on two levels. Qualitatively, rapidly moving molecules will quite naturally spread to fill the space allowed. On a deeper level, one must consider the motion of those rapidly-moving molecules. It may be due to translation, rotation, vibration about a fixed point or stretching, to name a few. Each of these types of movement has discrete allowed energy levels. The more such energy levels available, the more quantum states possible and so the higher the probability. Quantum mechanics tells us that expanding the space available to the molecules provides a denser packing of energy levels and so more of them. In the simplest approximation, in an infinite square potential well of width a, a particle of mass m will have energy eigenvalues[23]

$$E_n = \frac{n^2\pi^2\hbar^2}{2ma^2},$$

so the energy-level spacing is inversely proportional to the square of the space available. The more space is available, the greater number of energy levels will be available, the more energy can be dispersed, the greater the number of micro-states, the higher the probability and entropy.

The second law also applies to living creatures, such as us. Our bodies' cells, tissues and organs are highly ordered biological objects. The process of digestion (cellular respiration) liberates the binding energy of the food we ingest. For carbohydrates, for instance,

$$C_6H_{12}O_6 + 6O_2 \rightarrow 6CO_2 + 6H_2O + \text{energy}. \qquad (2.3)$$

The liberated binding energy of the carb (such as glucose) is stored in the molecule ATP, which then carries it to cells which use it to maintain the metabolism of the body. But with age, things don’t work so well any more. This is because such problems as worn-out joints, weakened muscles or hardening arteries function less efficiently, losing energy which is then dispersed from us. Old entropy's got us![24] (All this will be covered in later chapters.)

But the notion of entropy is not all negative. We just saw that chemical bonds generally contain a lot of energy. This means that they are difficult to break. The initial energy necessary to break the bonds and start the reaction is called, quite logically, the ***activation energy***. It is this non-zero activation energy which keeps our bodies from deteriorating almost instantly. This is why, on a molecular-biological level, enzymes (organic catalysts) are necessary to start the reactions of our metabolism. And enzymes are almost all proteins, which means that they are produced by ribosomes in our cells starting from recipes in our DNA and using proteins in the cytoplasm. But that story will come in due time.

The universe itself is not exempt from the Second Law. Galaxies, stars, our solar system and our dear, mistreated Earth[25] are all ordered at the expense of nuclear energy from the stars, including our Sun. One day, energy dispersal will prevail there too. Dispersal of the energy which now holds things together will mean further expansion of all parts of the universe. Even if the human species, should they still exist in some form, could find refuge on another planet before our sun explodes in some five billion years, the entire universe will ultimately become dispersed, cold, dark and diffuse.[26] *Carpe Diem*!

It appears that there have been at least 30 different versions of the second law over the last 130 years.[27] Here is ours:

- The second law of thermodynamics says that energy of all kinds in our material world spontaneously disperses or spreads out if it is not hindered from doing so.
- Entropy change measures the dispersal of energy: how much energy is spread out in a particular process, or how widely spread out it becomes (at a specific temperature).

23 Griffiths and Schroeder, Introduction to quantum mechanics, 33.
24 To be sung to the tune of “Ol’ rockin' chair’s got me...”.
25 The pale blue dot from space and so well described by Carl Sagan, www.youtube.com/watch?v=rBTVgTrTo8k.
26 This scenario is not quite certain, so there are other possibilities, but this is the most serious one, as well as the gloomiest.
27 Lambert, op. cit.

2.2. Quantum mechanics

Quantum mechanics is the theory of what happens at very small dimensions, on the order of 10^{-30} meters[28] or less! It is the theory which explains atoms and elementary particles.

According to quantum mechanics, what is "out there" is a vast amount of space. This space is filled with particles so small that the distance between them is huge compared to their own sizes. Not only that, but they don't behave like tiny pebbles of matter, but like waves, or something else which acts sometimes like waves and sometimes like particles. The modern interpretation of this is in terms of ***fields***, things which have a value (and very often a direction or other properties) at every point in space. For instance, every point in space has a specific temperature, so temperature is a field. It has only one value at each point and is called a ***scalar*** field. A magnetic field has not only a magnitude but also a direction at each point and is called a ***vector*** field.

"Every particle and every wave in the Universe is simply an excitation of a quantum field that is defined over all space and time."[29] Admittedly, this is not easy to understand.

Nobody can actually measure simultaneously where a particle is and how fast it is moving (or how much energy it possesses and when). This effect is referred to as ***indeterminacy***, or the ***Uncertainty Principle***, one of the more uncomfortable and, simultaneously, fruitful results of the theory.

As a result of this indeterminacy, energy need not be conserved for very short periods of time, regardless of thermodynamics, giving rise to all sorts of unexpected phenomena, such as radiation from black holes. But that is another subject.

QM is explained by a mathematical formalism based on an equation, generally referred to as the ***Schrödinger equation***, although it exists in several forms (differential, matrix, bra-ket, tensor). Solutions to this equation are called ***wave functions***. A wave function serves to predict the ***probability*** that the system under study be in a given state. It gives only a probability for the state. (In fact, the probability is not the wave function itself, but its complex square.) This knowledge only of probabilities really irks some people and nobody really understands what it means (*dixit* Richard Feynman, one of the greatest of quantum theorists). But the mathematics works. Remember, we are talking about very small scales. The probabilities that a particle be located at a point span an extremely small range, completely undetectable on the scale of things like tables or pizzas or cats.

Here is the equation, but you are not expected to understand it – or memorize it.

$$i\hbar\frac{\partial}{\partial t}\Psi(x,t) = \Big(\frac{-\hbar^2}{2m} + V(x,t)\Big)\Psi(x,t)$$

Figure 2.1: Time-dependent non-relativistic Schrödinger equation

According to QM, some parameters of a system, such as energy or wavelength, can only take on certain values; any values in between are not allowed. Such allowed values are called ***eigenvalues***. The eigenvalues are separated by minimal "distances" called ***quanta*** (quantum, in the singular) and the system is said to be quantized. We will see a good example of them when we look at atomic structure.

An important result of QM is that certain particles known as ***fermions*** are constrained so that two of them can never occupy exactly the same QM state. This phenomenon, called the ***Exclusion Principle***, is at the root of solid-state physics and therefore of the existence of transistors and all the technologies dependent thereupon – portable computers, mobile telephones, space exploration and the Internet, just as to mention a few examples. So QM has indeed revolutionized modern life, for the better and for the worse.

The Exclusion Principle is also responsible for the fact that electrons in a collapsing super-dense star cannot all be in the same state, so there is a pressure effectively keeping them from being compressed any further. We will read more about that in the cosmology chapter.

In fact, and most important, without the exclusion principle, fermions would all settle down into the same state, the one of lowest energy, and everything would be the same. Matter as we know it would not exist and

28 You should know engineering notation, where 10^{-30} (ten to the minus 30th power) means 0. followed by 29 zeroes, followed by 1: 0.000000000000000000000000000001. Similarly, 10^{17} = 100,000,000,000,000,000.

29 Blundell and Lancaster, 1.

so neither would we.

2.3. Relativity

Just as QM is the theory of the very small, Relativity is the theory of very large. – and very heavy

2.3.1. Special relativity

In fact, there are two theories of relativity. The simpler, first one, called Special Relativity (***SR***), is about space-time, relative movement and the speed of light, i.e., everything. It says:

1. The equations of physics are the same (***invariant***, in math-physics speak) for all observers who are moving with uniform steady motion relative to each other. Such observers are said to be in ***inertial systems***. They are trundling along under their own inertia, without any external forces pushing or accelerating them.
2. The speed of light in a vacuum is a constant, the same for all inertial observers.

The first statement, the principle of relativity, means that if someone goes by you in a high-speed train at a constant high speed on a straight track, then either you or she can consider herself to be stationary and the other moving. If you have traveled by train, you may have wondered sometimes whether your train was advancing or the one next to you was moving backwards. Relativity!

The second statement means that if both you and your friend in the train measure the speed of a light beam, you both will find the same answer, 299,792,458 meters/sec. This is not at all intuitive.

From these two statements, lots of things follow, including.

- the equivalence of mass and energy, expressed by everybody's favorite (and perhaps only) equation, $E = mc^2$;
- the fact that only bodies without mass can travel at the speed of light;
- curious distortions like the faster someone is moving relative to us,
    - the heavier she gets,
    - the shorter she gets in the direction of motion and
    - the slower her clock runs, including her body clock, the heart.

This last point is the origin of the famous twin "paradox". If your twin takes off in a space ship and travels fairly fast compared to the speed of light, then when she gets back to Earth, she will be younger than you, the twin she left behind. Strange, but this has been tested (with particles) and found to be true. The reason you cannot take the inversely “relative” view and claim that she should see you as the younger is that she has accelerated and decelerated in her flight and so has not remained inertial (without acceleration).

One of the strangest and most important results of SR is the banishment of the notion of simultaneity. One can no longer speak of the order of events in time, as this order depends on the observer, who sees effects on spatial and temporal “lengths” (displacements), the last two points in the preceding list. One can no longer define simultaneous events!

I can't help but describe this through a thought experiment, because it's so much fun. Say I have a garage which is a former barn, with doors at both ends. It's just long enough for my VW Beatle. Suppose you have a stretch limo.[30] If you drive your limo sufficiently fast, then it will look to me to be shorter, even short enough to fit in the barn. So I will see things in the following order:

- The front of the limo enters the barn.
- The back of the limo enters the barn.
- The front leaves the barn, and then the back.

But from your point of view, it's not the limo which gets shorter, but the barn, so there's no way you can fit the

30 This experiment is more often imagined with a pole vaulter carrying his pole parallel to the ground. The limo version comes from Leonard Susskind.

limo there at one time. You will see events in the following order:

- The front of the limo enters the barn.
- The front of the limo leaves the barn.
- The back of the limo enters the barn.

The second and third events have changed order! So saying which comes first depends on the observer.

All this comes out of the mathematics for describing relative motion, which is in turn due to the fact that space and time are not independent, but form a four-dimensional “thing” called ***space-time***. What we see as space and time are aspects of space-time and can appear differently for different observers. And they can get mixed up together if you are traveling at high speeds.

2.3.2. General relativity

The other Relativity theory is called General Relativity (***GR***). Whereas Special Relativity is the theory of space-time and light, General Relativity is the theory of gravity. GR says that space-time is curved and that it is more curved where gravitation is stronger. In fact, gravity *is* the curvature of space-time, not a force. Particles not under the influence of any of the three other forces simply follow their noses, advancing in a direction which looks to them like straight ahead. Or, you can think of a plane surface with a depression in it. Put a ball on it and the ball will roll into the depression. Try to visualize that in 4 dimensions (Good luck!) and you've got GR. Uh, sort-of...

There are also lots of things which follow from GR. One of the more interesting is that gravity can change the direction of light, since light travels through the space deformed by gravity. This and many other predictions have been observed to occur just as GR predicts.

Curiously enough, clocks run slower in a stronger gravitational field. A clock runs faster on top of a high building or in a satellite than it does on the surface of the Earth. For a satellite, this is opposite to the effect of lower speeds of fast-moving clocks in SR The two corrections are in opposite directions but do not cancel each other completely, so both must be taken into account in order for GPS systems to function correctly.

More recently, it has been discovered that not only is space-time curved, but space is expanding – faster and faster. This is because a gravitational field exerts not just a force, but a pressure. This pressure is considered to come from a term in the equation of GR called the ***Cosmological Constant***. Unlike the force of gravity, which is always attractive, the pressure can be negative, in which case it is not attractive, but repulsive. It is the outward force of the negative pressure which drives the expansion of space. For the last 7 Gy[31], the expansion of space has been accelerating. More about that in the cosmology chapter. Just one word: Space is expanding – empty space, not things, such as galaxies, bananas or you.

One more formula, just to appreciate, Einstein’s field equation for the curvature of space.

$$G_{\mu\nu} + \Lambda g_{\mu\nu} = \kappa T_{\mu\nu}.$$

The right-hand side contains all of matter and energy. The first term on the left is the curvature of space. And the second left-hand term contains Λ, the cosmological constant, which causes the Universe to expand. Simple, isn’t it? (It’s not.)

2.3.3. QM and GR – an unhappy marriage

As far as they go, SR and GR are as true to nature as can be detected within their separate domains, and so constitute bodies of knowledge – theories.

The problem is that the equations of QM do not work on the huge scales of GR, nor do those of GR on the infinitesimal scales of QM. This is certainly one of the biggest problems in contemporary physics, the resolution of which is the grail now pursued by theorists.

In summary:

> We live in a world where stuff sometimes behaves like a particle and sometimes like a wave and

31 Gy means giga-years, or 10^9 years or 10 billion years.

where only probabilities can be calculated, although the range of probabilities is undetectable on our scale of things. All this takes place in a curved four-dimensional space-time which is expanding at an accelerating rate! And this space is not an empty vacuum, it is something.

Got that?

2.4. The standard model of elementary particles

This is as good a place as any to mention the Standard Model. It is not a theory like QM or GR, but it uses QM to build a “model”, i. e., a proposed explanation, of the constitution and properties of the elementary particles of which everything else is constituted. Nevertheless, it is more than that as it explains the world around us quite well.

2.4.1. The particle zoo

At this point, you might think that the elementary particles are the proton, the neutron and the electron, but that is both incomplete and too simple. Incomplete, because there are lots of other particles and too simple because – among other reasons – neutrons and protons (as well as mesons) are not elementary, they are composed of quarks.

Warning: This presentation necessarily presents lots of new vocabulary. In addition, it sounds a bit like numerology. But it works – with one notable exception, the force of gravity. However, gravity is thought to be negligible at atomic scales (except when discussing black holes, which we are not).

Basically, there are two type of particles, fermions and bosons, based on their spin. ***Spin*** is analogous to a particle's angular momentum due to its spinning on an axis, but is an inherent property of a particle, not due to its orientation in ordinary space. If you don’t understand that, it doesn’t matter.

Bosons have spin which takes on only integral values – for instance, 0 or 1. They are virtual particles which carry the forces between other particles.

Fermions all have half-integral values of spin – ½, 1½ and so on. They are the basic components of matter, the blocks from which matter is built, and are constrained by the Exclusion Principle. There are two types of fermions – ***quarks*** and ***leptons***. They all are shown in Figure 2.2, the particle zoo, which is composed of six quarks (shown in purple), six leptons (green) and four bosons (orange). (The Higgs is special and will be discussed later – maybe.)

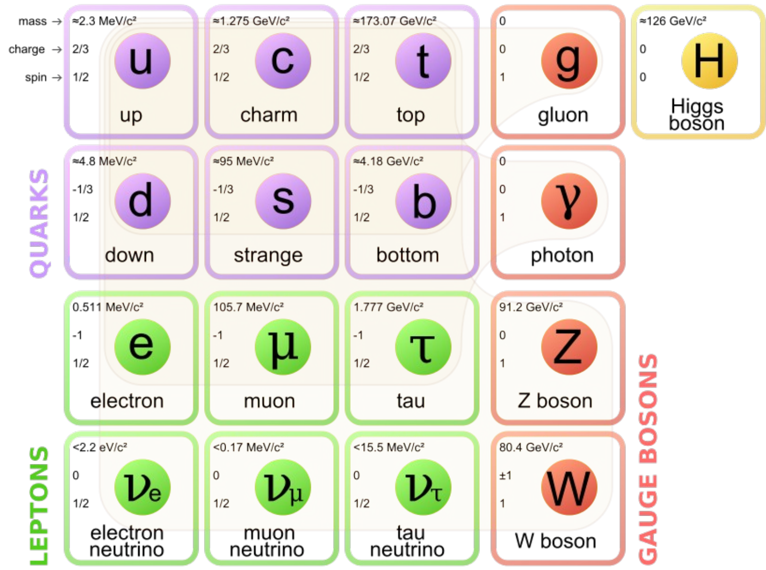


Figure 2.2: Elements of the standard model[32]

Here's the part which smacks of numerology. No marks will be taken off if you now skip to the list of forces below. The quarks are said to differ by their ***flavor*** – up, down, charm, strange, top or bottom – which obviously is just an arbitrary term and has nothing to do with taste. They are arranged in three columns called ***generations***. Quarks have charges +2/3 or -1/3; leptons, 0 or 1. Each of the six flavors of quark exists in three versions indicated (by analogy) by the colors red, green and blue[33], for a total of 18. When forming matter particles, the quarks must group together in such a way that the result is "colorless". So they must occur in the combination $R+G+B$ for ***baryons*** like the neutron or proton, and $R+\bar{R}$, where the bar above the character indicates an antiparticle, for ***bosons***. Particles composed of quarks are called ***hadrons***.

Note that the mass of, for instance, a proton (about 938 MeV/c^2) is far greater than the sum of the masses of its constituent quarks (2x2.3 MeV/c^2 + 4.8 MeV/c^2).[34] The excess, almost 99% of the proton's mass, is potential energy of the strong force which binds the quarks together into the proton. Splitting the proton into quarks would require furnishing this amount of energy, which explains why physicists need such powerful particle accelerators.

Matter is made up of atoms with nuclei containing protons and neutrons (also called ***nucleons***), with electrons forming a negatively-charged cloud around the nucleus. Such long-lived particles are made up of quarks from the first column, called the first generation. A proton is composed of two up quarks and a down quark. The former have a charge of +2/3, the latter of -1/3, so the total is +1. A neutron is composed of an up and two downs, for a total charge of +2/3-1/3-1/3 = 0. And so on. (This is the numerology part.)

As stated, bosons are virtual, force-carrying particles. Modern physics recognizes four forces in nature.

1. Gravity is theorized to be conveyed by a yet-to-be observed particle called the ***graviton*** (not shown because never observed and because gravity is not part of the Standard Model). It is a weak force (the weakest) but works across enormous, interstellar distances. We shall see that it is responsible for the formation of stars, galaxies and planets – no less.

2. The electromagnetic force between charges or magnets is conveyed by the ***photon***, which is the particle of light and has no mass. Like gravity, it has infinite range, but is stronger. Since its sources can be either positive or negative charges, the two cancel each other out over large distances, making the overall force weaker. The quantum theory of the electromagnetic force is called ***quantum electrodynamics***, or ***QED***.

3. The strong force is the strongest, but is very short-ranged. It holds quarks together in the nucleus in

32 Wikipedia Creative Commons, https://en.wikipedia.org/?title=Standard_Model#/media/File:Standard_Model_of_Elementary_Particles.svg.

33 Not the same as the colors on the chart.

34 One MeV is one million electron volts, an electron volt is the energy added to an electron accelerated by a one-volt electric potential (equivalent to a force). It is commonly used in accelerator physics.

spite of the repulsive electric forces between their charges. It is conveyed by the appropriately-named ***gluon***. The quantum theory of strong interactions is called ***quantum chromodynamics***, or ***QCD***.

4. The weak force is weaker than the strong or EM forces, but is still stronger than gravity. It also is a very short-range force, responsible for decays of various radioactive particles. Such decay is responsible for the existence of the Periodic Table and so for the elements of which our Earth and we are made. The weak force is conveyed by the W and Z bosons.

So there two infinite-range forces, gravity and EM, and two short-range ones, the strong and weak. By order of strength, from strongest to weakest, they are strong, EM, weak and gravity.

EM should be as far-reaching as gravity, but in fact, as you get farther away from, say, a positive charge, the more it is shielded by negative charges, so that the net charge seen is about zero. Gravity only has an attractive "charge", so its effects extend over a very large range, over galaxies and more. And, as we shall soon see, that is a Good Thing!

2.4.2. The Core Theory

It is important to understand that quantum theory alone does not explain the world. It is a framework for expressing theories about the world. It can, for instance, be used to explain atoms and the periodic table, but only by adding information, e.g., that a hydrogen atom is composed of a proton with one electron moving about near it, both under the sway of the electromagnetic force.

Remember what was said about fields in section 2.2, that modern physics considers everything to be made up of fields. This is the idea behind ***quantum field theory*** (***QFT***). Fields are where the buck stops, or like the bottom turtle, they are not made up of anything else. The particles which we see as the constituents of all the stuff around us are vibrations in quantum fields, fermion fields for matter, boson fields for forces. Although the separation of GR and QM keeps us from understanding completely the Big Bang or black holes, most of the time who cares? The stuff around us in our everyday world is just ordinary matter and not composed of black holes, so we can effectively describe the world we live in with the standard model, based on QFT and the four forces, plus GR. This is what some physicists call the ***Core theory***.

Core Theory = QFT + GR

And It works, providing the physical laws underlying chemistry, biology, astrophysics, engineering and much of cosmology.

2.5. Evolution – the modern synthesis

Evolution is most simply defined as descent with modification. Biologists' understanding of evolution since the 1940s is called the modern synthesis[35], the synthesis being between observed biological evolution through natural selection and the science of genetics.[36]

2.5.1. Population genetics

The modern synthesis employs mathematical (statistical) methods to study what is called a ***population***, the collection of those members of a particular species living in a specified area. ***Population genetics*** considers the collective ***gene pool*** of such a population, i.e., the sum of all ***alleles*** – versions of genes[37] – within the population. A given gene may have more than two alleles, such as those for human blood antigens (A, B and O). The link with what we actually see comes about when these alleles are expressed as ***phenotypes***, i.e., observable traits in the organisms of the species. The underlying genetic composition which determines the phenotype is the ***genotype***. Modern evolutionary scientists define evolution as the change in the distribution (frequency) of the alleles in the gene pool of a population across generations in

35 Or neo-Darwinian synthesis.

36 Or between morphology and genetics.

37 More specifically, gene variants at the same positions on homologous – corresponding – chromosomes. All this gene business will be explained in a later chapter.

time.[38] A species is for them, as it was for Darwin, "...an aggregate of populations that are connected by the migration of individuals from one population to another."[39]

The modern synthesis merges genetics and morphology and provides mathematical techniques for a quantitative study of evolutionary change. It recognizes four mechanisms of evolutionary change.[40]

1. ***Natural selection*** operates when random changes in genes, such as those occurring during meiosis (explained later), improve the ability of the individuals possessing those genes to survive long enough to reproduce successfully and pass the mutated genes on to their offspring. One then says that differential reproduction has taken place. The superior survival rate of these individuals because of this ***adaptive*** trait will assure that, over time, their relative numbers will increase. In terms of population genetics, the frequency of their alleles will increase in the gene pool.

2. ***Mutation*** is a random change in DNA, the genetic material (much more about that later on). Mutations are small and slow. They may disappear quietly without leading to evolutionary change, or lead to the death of the organism (and the allele). Or they may accumulate very slowly. Mutation is the ultimate generator of variation in alleles and so makes possible the other three mechanisms of evolutionary change.

3. ***Genetic drift*** occurs when a chance variation (due to mutation) reproduces itself and then gradually becomes important in the gene pool. Drift dominates natural selection in very small populations. According to statistics, the range of values of a trait will be well represented by the normal distribution (the famous "bell curve") in a large population of samples. Since a large population furnishes numerous possible mates, any variation probably will be quickly diluted. However, in a small sample, the measured "average" may vary significantly. Therefore, genetic drift takes place more often in smaller populations. Genetic drift occurs most frequently when a sub-group of a population becomes physically separated from the main group, say, due to emigration to an island or environmental isolation by the forming of a new body of water or a mountain ridge – or even just due to large distances between extremities of a vast landmass. The small group may then change sufficiently to form a different species. Such ***speciation*** (formation of a new species) due to an external barrier to gene flow is called ***allopatric***[41] (Greek for "another country"). Since it is due to chance (statistical variation), genetic drift need not be adaptive. Natural selection, on the contrary, is adaptive and so can not bring about evolution of a lasting trait which is bad for the organism.

4. ***Gene flow*** (or ***migration***) is similar to genetic drift, but involves the immigration of individual organisms or alleles into one population from another. Pollen blown by wind may be an example of gene migration. The result is that genes flow from one population to another.

In all these cases, change is a random, chance occurrence, whereas natural selection is a law which operates only through such changes. In spite of the terms often used to express evolutionary change ("Flowers have colors in order to attract bees."), it is not teleological (goal-oriented). One should say, "Colored flowers attract bees and so survive better."

It is important to understand that natural selection works on phenotypes, not genotypes (or genomes). An allele is selected via the phenotype because of its difference with another allele. A gene is but the recipe for a protein. If the protein leads to a fitter form of a phenotype, it will be selected. A long chain of events takes place between the confection of that protein and a phenotype, and this chain will involve proteins and enzymes produced by other genes. The path is neither straight nor short. The result may be more than just an anatomical or neural feature, it may be a form of behavior. The dams of beavers and the nests of bees are the results of behavior which are due to particular genotypes. They are extended phenotypes, in that they are not physically part of the animal concerned. In an even larger sense, a beaver lake exists because of a dam built by beavers in accordance with their genetically inherited behavior and it provides the environment for the beaver's specific way of life, which enables its survival. It is therefore an example of an ***extended phenotype***.[42] If a global catastrophe eliminates the lake, it will eliminate the beavers too.

38 Reznick, 96

39 *Ibid.*

40 Sometimes, they may mention non-random mating as a fifth mechanism.

41 Also referred to as geographic or vicariant.

42 Dawkins (2004), 193-6.

In addition, sex is a source of gene mixing, as will be discussed in a later chapter. But the oddest source of change is shuffling of the genetic material when antibodies are created in our bodies. More on that later, too.

Like quantum mechanics, evolution injects the theme of randomness into our understanding of the workings of the universe, but on a completely different scale. The comparison stops there.

2.5.2. Species

The notion of species has been much discussed in the history of biology before arriving at the current definition of a ***species*** as "groups of interbreeding natural populations, which are reproductively isolated from other such groups".[43] This means that members of two different groups do not mate to produce viable offspring for one of two reasons: Either (1) they are geographically isolated by necessity or by choice; or (2) their genetic differences are such that they cannot produce viable offspring, or "fertile hybrids"[44] Either they are unable to interbreed or their offspring are less fit.

However, not much is known about the sexual reproduction of, e.g., bacteria or fossils. So in many cases, biologists must fall back on observation of physical or genetic similarities for distinguishing species. It also happens that observed differences do not amount to differences in species. The subject is complicated.

Speciation is the formation of new species. It may occur that a new biological niche becomes available (or an old one vacant), as can happen on a newly colonized island. Under these conditions, one species may diversify rapidly into a number of similar species, a process called an ***adaptive radiation***.

In the case of genetic drift, members of the changed smaller group may rejoin the larger one (thus becoming ***sympatric***), bringing their variations with them. The change may take place over a time scale less than the difference of the dates of adjacent rock strata, in which case there would be no intermediate fossils visible between the two states, hiding any continuity in the evolutionary process. The appearance of evolutionary stasis punctuated by discrete changes has been called ***punctuated equilibrium***, derisively abbreviated as "punk-eek".[45]

Evolution can give rise to two kinds of similar structures. Characters[46] of different species which are inherited from a common ancestor are called ***homologous***. An example is the presence of the four limbs of tetrapods. Similar characters which are not inherited from a common ancestor are ***analogous***, like the wings of bats and birds, and are the result of ***convergent evolution***, the evolution of similar adaptations in species which are at most distantly related.

2.6. Paleontology – fossils and classification

The goal of evolutionary classification is to demonstrate and explain relations among species. The method uses hierarchical grouping of ***species*** into larger groups called ***genera*** (singular, ***genus***), genera into ***families*** into ***orders*** into ***classes*** into ***phyla*** into ***kingdoms***.[47] This amounts to a road map or genealogy of evolutionary relations. Since two species may be related by being descendants of a long-extinct ancestor, information about that extinct ancestor is necessary. This is obtained primarily through the study of fossils.

These classifications will be discussed in more detail in following chapters.

2.6.1. Fossils and fossilization

Fossils are central to our understanding of past species, but they come with problems.

How fossils are formed

If a dead organism does not decay and is not destroyed by predators, it may be covered, all or partially, by

43 Dobzhanksy, cited by Reznik (2010), 144.
44 Coyne and Orr, cited by Reznidk (2010), 144.
45 Or, more disparagingly, "evolution by jerks". Some biologists have a strange sense of humor.
46 The word "character" refers to a heritable trait and may be morphological, genetic or behavioral. I would have said characteristics, but nobody asked me.
47 There are many mnemonics for remembering these, of which I prefer: "King Phillip Came Over For Good Spaghetti".

sediments. The sediment continues accumulating and, under pressure, eventually may form rock. Shells may dissolve and leave a hole in their own form in the sediment – a mold. Dissolved minerals such as silica may move through the porous rock, fill in the hole and harden, taking on the shape of the organism. Or the dissolved minerals may fill in the pores of the organism itself, leaving a detailed record. The result is a fossil, entombed within the sedimentary rock.

Fossils may also be found where organisms are frozen (as in Siberia) or desiccated (as in deserts). They may be preserved in amber from pine resin or in tar if they fall into a petroleum swamp. They may leave only impressions, tracks or even footprints which are then preserved, for instance, by falling ash. So a fossil can be an impression or a trace or even preserved remains of an organism.

From the number of species alive today and the "turnover" seen in fossils, an estimation of the number of species which have existed during Earth's history gives numbers in the billions. But the fossil record only contains some hundreds of thousands, so the fraction of the number of past species in the fossil record is much, much less than 1%.[48] Where have all the fossils gone?

What can fossilize

Some fossils have indeed gone, but many others were never formed. Organisms consist of parts of differing hardness. Teeth are the hardest and decay slowly enough that they may be fossilized easily. Bones decay faster than teeth, but still slowly enough to be fossilized. However, soft body parts usually decay before adequate sedimentation can take place. Only if sedimentation occurs rapidly can soft parts be preserved, as in the case of the mudslide which preserved the fossils of the Burgess Shale fossils or those at Chengjiang, China.

Many fossils are unable to form because of their environment. The ideal habitat for fossilization is a shallow basin which is slowly subsiding, providing good conditions for marine animals and for an appropriate sedimentation rate. This is no help for terrestrial animals.

Even if fossils are formed, they may be destroyed by subsequent movements of the earth, during which heat and pressure may deform, break, crush or burn them up. No fossils are found in igneous rocks and almost none in metamorphic rocks.

Fossils are often hard to find. Most of the earth has not yet been searched for fossils. They often are found when their enclosing sediments erode or are exposed by geological activity, such as at the East African Rift. Such conditions do not take place everywhere.

For all these reasons and more, the fossil record is a very incomplete history of life on Earth. So it is surprising that we have managed to learn so much from it.

2.6.2. Classification – taxonomy and cladistics

According to the theory of evolution, all living organisms of one species have evolved from organisms of a different species, and those from still different species, all the way back to the original Ur-form of life, the first cell or amino acid or whatever it was.[49]

The branch of biology called ***taxonomy*** is the science of defining groups of organisms based on shared characteristics in order to show up their shared evolutionary history. The result may be displayed as a family tree or ***phylogeny***. Biologists use ***cladistics***, or ***phylogenetic systematics***, to diagram the evolutionary steps between species. An example of such a diagram is in Figure 2.3. A clear distinction is made between primitive characters and advanced or derived characters. Only groups defined by common derived characters are considered valid ***monophyletic*** groups or ***clades***, which are groups descended from a common ancestor and which include all descendants of the common ancestor. Monkeys and apes without humans do not constitute a monophyletic group, nor do reptiles without birds. It is important to understand that a phylogeny is a tree, not a ladder, and implies nothing about whether an organism is "advanced" or "primitive".

48 Reznick (2010).
49 More on this in the geology chapter.

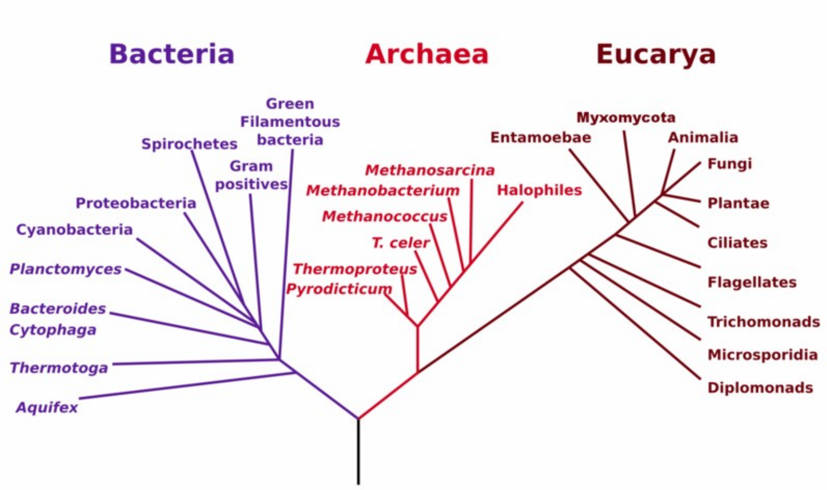


Figure 2.3: A phylogenetic tree of life, from Wikipedia[50]

Although these phylogenies are based on both morphological and molecular data, new data often lead to slightly different relations, so the tree changes somewhat as time goes by. This is due mainly to the incomplete record presented by the fossils.

Such trees can also be derived from the comparison of genomes. This is such a huge task that one most often studies the evolution of a single gene. But the choice of gene results in significantly differing trees, so that compilations also have been made of differing trees.

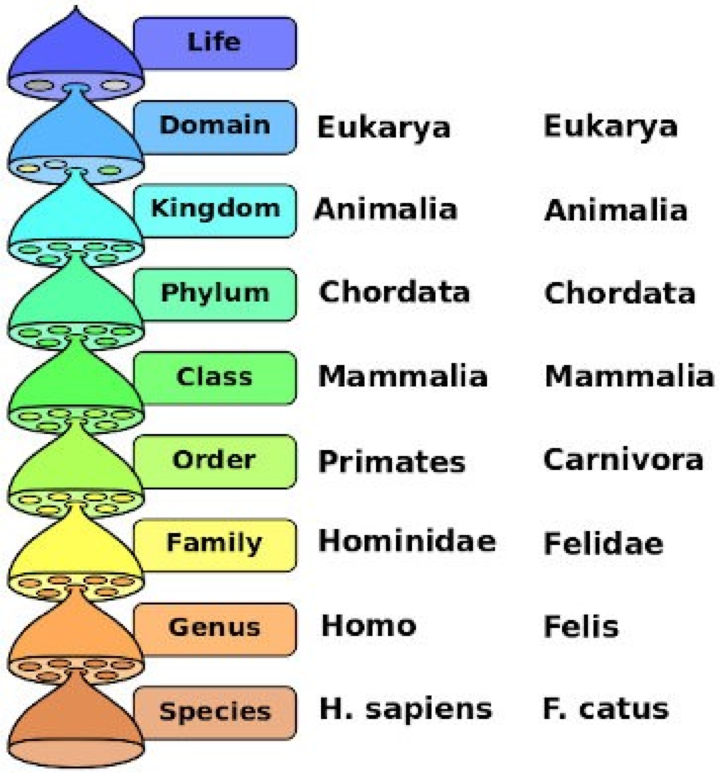


Figure 2.4: Classification of modern humans and house cats, after Wikipedia[51]

50 Phylogenetic tree, https://en.wikipedia.org/wiki/Phylogenetic_tree

51 File:Biological classification L Pengo vflip.svg, https://commons.wikimedia.org/wiki/File:Biological_classification_L_Pengo_vflip.svg.

3. What atomic physics and chemistry tell us

I am, reluctantly, a self-confessed carbon chauvinist. Carbon is abundant in the Cosmos. It makes marvelously complex molecules, good for life. I am also a water chauvinist. Water makes an ideal solvent system for organic chemistry to work in and stays liquid over a wide range of temperatures. But sometimes I wonder. Could my fondness for materials have something to do with the fact that I am made chiefly of them?

– Carl Sagan, *Cosmos*

The early stages of the universe and the lives of stars are the matter of physics and astronomy and their offspring, astrophysics and cosmology. By the time the first living things showed up on Earth, processes were occurring which require our knowing about the phenomena described by the science of chemistry. QM is the basis of atomic physics and that is the basis of chemistry, so we are ready for it.

Even to begin a comprehensive survey of chemistry is well beyond the scope of this study. We will illustrate its usefulness and some of its fruits by considering two subjects of great importance not only to Carl Sagan but to all of us – carbon and water.

In order to do that, it is necessary to know about about atoms, electrons, chemical bonding and energy states.

3.1. Atomic energy levels and chemical bonding

Atomic structure is the basis of chemistry. It is explained by Quantum Mechanics and Electromagnetics. In QM, the properties of a system, that is, a given object or set of objects, such as an atom, are given by the solution to the Schrödinger equation for the system, which depends on EM, as electrons are attracted to protons in the nucleus by EM forces. For atoms, there are a set of solutions, corresponding to different energy states of the atoms. What follows may smack of numerology.

Consider the hydrogen atom, composed of one negatively-charged electron in orbit around a nucleus containing one positively-charged proton. (This is an experimental result.) Look out, the orbit is not a well-defined path around the nucleus like those animations you see in TV ads, but rather a cloud of probability which indicates the likelihood that the electron will be found at any point in the cloud. This is due to the probabilistic character of QM and the Uncertainty Principle. The different solutions to the Schrödinger equation express the possible energy values of the atom. Each one is specified by a set of integer numbers called ***quantum numbers***. In the case of the hydrogen atom, they are the following:

1. The ***principal quantum number***, designated by the symbol n, takes on integer values from 1 on up, but in practice only to 7. It indicates the ***shell***, or level of the cloud, in which the electron is found. The values 1-7 are often indicated by the letters K, L, M...Q. (No, I don't know why.)
2. The ***orbital quantum number***, l, indicates a level within the shell which is called the ***subshell***, It can take on values from 0 up to n-1. The values 0-3 are often referred to as s, p, d and f[52].
3. The ***orbital magnetic quantum number***, m, refers to the magnetic orientation of the electron. It can range from -l up through +l.
4. The electron ***spin***, m_s, can take on only two values, ½ or -½.

So the only allowed values for the quantum numbers are

n = 1, 2, 3, …

l = 0, 1...n-1 (for a given value of n)

m = -l, -l+1,...+l (for a given value of l)

52 The notations s, p, d and f come from spectroscopy and are abbreviated forms of sharp, principal, diffuse and fundamental.

because those are the ones for which the Schrödinger equation has solutions.

The QM exclusion principle forbids two electrons to occupy the same state. So each set of values (n, l, m, m_s) can correspond to only one electron. The result is illustrated in the Table 1.

The fact that the quantum numbers do not vary continuously from, say, 0 to 0.001 and then 0.002 and on, but but jump from one integer value to another means that the energy of the electron in the electric field of the nucleus also takes on non-continuous values. These are called ***quantum states*** and are a feature (or if you prefer, a peculiarity) of QM.

The chemical properties of an atom depend only on the number of electrons. Unless the atom is ionized or chemically combined with another, the number of electrons is equal to the number of protons, which is called the ***atomic number.*** All atoms except hydrogen have nuclei which also contain neutrons. As everyone knows, the number of neutrons can vary and atoms of an element with different numbers of neutrons are called ***isotopes***. As we shall see, these are extraordinarily useful due to the fact that the properties of the element can vary from isotope to isotope.

n (shell)	l (subshell)	m (orbital)	Max no. electrons
1	0	0	2
2	0 1	0 -1, 0, +1	2 6
3	0 1 2	0 -1, 0, _1 -2, -1, 0, 1, 2	2 6 10
4	0 1 2 3	0 -1, 0, +1 -2, -1, 0, 1, 2 -3, -2, -1, 0, 1, 2, +3	2 6 10 14

Table 1: Atomic quantum numbers

The table summarizes the allowed values of quantum numbers for the first four shells. In specifying which subshells are occupied by the electrons in an atom, one often uses the format

$nl^{\#}$

where l is specified as s, p, d or f and # is the number of electrons in the subshell. In its minimum energy state, called the ground state, the carbon atom (atomic number = 12, nucleus contains 6 protons and 6 neutrons) has the following electron configuration:

^{12}C: $1s^2 2s^2 2p^2$

which indicates the maximum number of two electrons in shell 1, again in subshell s of shell 2 and the remaining two in subshell p of shell 2. Similarly, oxygen (atomic number = 16, 8 each of protons and neutrons) is

^{16}O: $1s^2 2s^2 2p^4$

the meaning of which should now be clear.

What is interesting is that, for energetic reasons, each atom would like to have its outside subshell filled (or empty). If a few electrons are missing, it wants more; if most are missing, it might be willing to give up the rest in order to have an empty outside shell, referred to as the ***valence shell***. (The number of electrons in this outer subshell is called the ***valence.***[53]) For instance, hydrogen

^{1}H: $1s^1$

wants two electrons or none in its 1s shell, so it could give up its electron or gain one. What happens is, two H atoms share their electrons to make a molecule of H_2, so each has two electrons half the time. Better than

53 Officially, the maximum number of univalent atoms (originally hydrogen or chlorine atoms) that may combine with an atom of the element under consideration, or with a fragment, or for which an atom of this element can be substituted.

nothing.[54]

Since oxygen already has shell 2 half-filled, it would usually prefer to gain electrons to fill it. And carbon… but carbon is special and will be considered in a moment.

Look at sodium (Na, atomic number 11) and chlorine (Cl, atomic number 17):

Na: $1s^2 2s^2 2p^6 3s^1$

Cl: $1s^2 2s^2 2p^6 3s^2 3p^5$

Sodium could happily give up that 3s electron and chlorine could use it to fill up its 3p valence shell. And this is what happens in table salt, NaCl. If you put salt in water, it separates (for reasons which will be discussed shortly) into charged ions, Na^+ and Cl^-, because chlorine is greedy and keeps that negative 3s electron it took away from sodium. This attraction for electrons is called ***electronegativity***. This is very important in biochemical reactions in cells, as we shall see.

In brief, it turns out that elements with ***two, ten or eighteen electrons*** are particularly stable.

Chemistry is the study of chemical systems (atoms, molecules) and chemical bonding between such objects. In the case of NaCl, the sodium and chlorine have opposite electrical charge and the attractive electric force is what holds the molecule together. This is called ***ionic*** binding. Sometimes, when atoms cannot decide which has more right to an electron, the electron is shared between them, making both atoms relatively happy. Binding based on shared electrons is called ***covalent*** binding; it is a sort of consensus situation, if we may go on with the anthropomorphisms.

Elements with the same number of electrons in their outer shells have similar chemical properties. So they are arranged in columns in that wonderful physical/chemical tool, the periodic table of the elements (Figure 3.1).

Columns in the table are called ***groups***; rows, ***periods***.

Group→ ↓Period	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	1 H																	2 He
2	3 Li	4 Be											5 B	6 C	7 N	8 O	9 F	10 Ne
3	11 Na	12 Mg											13 Al	14 Si	15 P	16 S	17 Cl	18 Ar
4	19 K	20 Ca	21 Sc	22 Ti	23 V	24 Cr	25 Mn	26 Fe	27 Co	28 Ni	29 Cu	30 Zn	31 Ga	32 Ge	33 As	34 Se	35 Br	36 Kr
5	37 Rb	38 Sr	39 Y	40 Zr	41 Nb	42 Mo	43 Tc	44 Ru	45 Rh	46 Pd	47 Ag	48 Cd	49 In	50 Sn	51 Sb	52 Te	53 I	54 Xe
6	55 Cs	56 Ba	*	72 Hf	73 Ta	74 W	75 Re	76 Os	77 Ir	78 Pt	79 Au	80 Hg	81 Tl	82 Pb	83 Bi	84 Po	85 At	86 Rn
7	87 Fr	88 Ra	**	104 Rf	105 Db	106 Sg	107 Bh	108 Hs	109 Mt	110 Ds	111 Rg	112 Cn	113 Uut	114 Fl	115 Uup	116 Lv	117 Uus	118 Uuo
		*	57 La	58 Ce	59 Pr	60 Nd	61 Pm	62 Sm	63 Eu	64 Gd	65 Tb	66 Dy	67 Ho	68 Er	69 Tm	70 Yb	71 Lu	
		**	89 Ac	90 Th	91 Pa	92 U	93 Np	94 Pu	95 Am	96 Cm	97 Bk	98 Cf	99 Es	100 Fm	101 Md	102 No	103 Lr	

Figure 3.1: The periodic table of the elements, from Wikimedia Commons[55]

It is easy to see that each element in the first column is like hydrogen in having one electron in its valence

54 To continue the anthropomorphisms, this is a kind of solidarity in which humans are often lacking.

55 https://commons.wikimedia.org/wiki/File:14LaAc_periodic_table_IIb.jpg

shell.

H: $1s^1$

Li: $1s^2 2s^1$

Na: $1s^2 2s^2 2p^6 3s^1$

K: $1s^2 2s^2 2p^6 3s^2 3p^6 4s^1$

… and so on.

The extra elements in the middle are rule-breakers. Instead of filling one subshell before moving on to the next, they start one, add a small number (often only one) of electrons to the next, then go back to finish filling the next-to-last.

The subshell configurations we have been giving are for the lowest energy state of the atom, called the ***ground state***, in which subshells are filled from the "bottom" up (with some exceptions, as just mentioned). But if that hydrogen electron is struck by a photon, enough energy may be transferred from the photon to the 1s electron to push it into a higher-energy subshell. The atom is then said to be in an ***excited*** state. The electron may then re-descend spontaneously to the lower subshell, emitting a photon of energy equivalent to the difference in energy levels of the subshells. In QM, photons behave like waves whose energy is a function of their frequency, so the frequency – equivalently, the color – of the light emitted is characteristic of the difference in energy of the two subshells. Any atom's subshells will therefore correspond to a given set of photon frequencies emitted and these are seen as colors, although not all these colors will be visible to a human eye. The set of frequencies constitute the ***spectrum*** of the atom and may be used to analyze the identity of a light source. In this way, we can identify the chemical components of light-emitting objects like distant stars.

There are two other types of bonding. We will consider hydrogen bonds very shortly in the discussion of water. The fourth form is due to the shifting electron density distribution around an atom. At times, this may form a temporary dipole even in a neutral atom. This may in turn induce a dipole in nearby atom in such a way that the two dipoles attract each other very weakly. This is ***London***, or ***van der Waal's, bonding***.

3.2. Carbon

Now we are ready to understand how it is that carbon is such a versatile element. It is at the basis of all organic chemistry and, in particular, biochemistry. The functioning of all living things depends on water and on the versatility of the carbon atom.

In living organisms, the four most abundant elements by number of atoms are hydrogen, oxygen, nitrogen and carbon. Together, they comprise over 99% of the mass of most cells.

We saw that the carbon atom's electron-shell configuration was

^{12}C: $1s^2 2s^2 2p^2$

so it has four electrons in its valence shell (n=2). That enables it to share its four electrons with four others from other atoms. The bonds tend to be equally spaced around the carbon atom in the form of a tetrahedron, like those little creamer packets you get in cheap restaurants. For instance, a carbon atom can bond with four hydrogens, sharing each of its four valence electrons with one hydrogen, so each hydrogen has two and the carbon has eight and everybody is happy. This is called methane and looks like this.

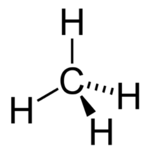


Figure 3.2: Methane molecule, CH_4, from Wikimedia Commons[56]

You should see one of the lower-right-hand hydrogens as pointing up out of the page; the other, down into it.

56 Methane molecule, CH_4 by Patricia.fidi via Wikimedia Commons, https://commons.wikimedia.org/wiki/File:Methane-2D-stereo.svg.

The angles between any two adjacent connecting lines (which of course are only imagined by us) are about 109.5°. Carbon's versatility in binding is illustrated by the examples in this diagram.

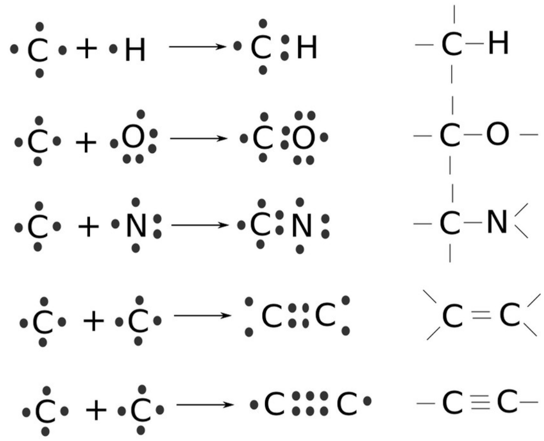


Figure 3.3: Versatility of carbon bonding, after Lehninger.

The dots represent valence electrons and the right-hand column is another way of looking at the product in terms of bonds rather than electrons. Each line between atoms is a shared pair of electrons. Note the double and triple inter-carbon bonds in the last two examples. This large number of ways of bonding is the key to carbon's versatility. In fact, compared to the huge number of such molecules possible, only a relatively small number of the same biomolecules occur in living organisms. This is the first example we see of nature using the same set of techniques or tools all over the biosphere.

Single bonds between carbons also exist, of course, and have the particular advantage that the carbons and whatever is bonded to them can rotate around the axis linking the two carbons. This is more important than one might think. It turns out that some proteins function differently in their left-handed and right-handed versions. Since rotation can change the shape of the molecule, this enables biomolecules with hundreds of atoms to take on specific shapes with definite mechanical or fluid properties. We will see some of this in the biochemistry chapter.

3.3. Water

Water is tremendously important to us if only because around 70% of the surface of the globe is covered with it. Each of us is 55-75% water (by weight) and life on Earth almost certainly arose in water. Two properties of water are of fundamental importance for biochemistry and, therefore, for life:

- the attractive force between water molecules and
- the tendency of water to ionize slightly.

3.3.1. Polarization and hydrogen bonds

As we saw, the electron configuration of oxygen's eight electrons is:

^{16}O: $1s^2 2s^2 2p^4$

So it needs two more electrons in order to fill its valence shell. As everybody knows, it bonds with two atoms of hydrogen to make H_2O. Each hydrogen atom shares its electron with the oxygen, making two covalent bonds. The oxygen atom now has the desired eight electrons in its valence shell. The resulting arrangement is tetrahedronal (triangular faces).

Oxygen is more electronegative[57] than hydrogen, meaning it has a stronger attraction for electrons, so the electrons spend more time in the vicinity of the oxygen, making that end of the molecule slightly more negative. The molecule is said to be ***polarized***. This is illustrated in Figure 3.4, where the small Greek letter δ designates a small electric charge.

57 Electronegativity depends on the number of electrons and on the distance of the valence electrons from the nucleus

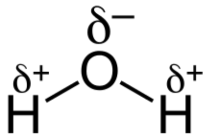


Figure 3.4: Polarization of water molecule, from Wikimedia Commons[58]

Since one end of the molecule is more negative than the other, the negative end of one molecule is electrostatically attracted to the positive end of another and this forms a weak bond called a ***hydrogen bond***. In this image, one sees the proposed tetrahedral form of the molecule as well as the hydrogen bonds between molecules.

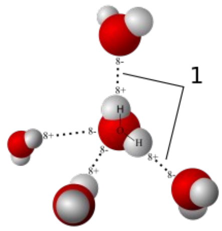


Figure 3.5: Model of hydrogen bonds (dashed lines) between five water molecules[59]

Hydrogen bonds are strongest when the hydrogen atom and the two atoms sharing its molecules are in a straight line, i.e., when the angle between the OH lines, for instance, is 180°.[60] The straighter the line (the greater the angle), the stronger the bond. The tetrahedron of bonds between water molecules, as shown in Figure 3.5 maximizes the angles and so represents the strongest bonding, This directionality is responsible for the geometric structure of hydrogen-bonded molecules into crystals.

Hydrogen bonds do not only occur in water. They also form between an electronegative atom and a hydrogen atom covalently bonded to another electronegative atom, be it of the same or a different molecule.

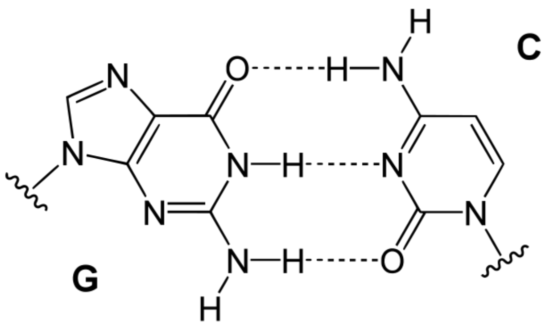


Figure 3.6: Hydrogen bonding (dashed lines) between guanine and cytosine, two of the four types of base pairs in DNA, from Wikimedia Commons[61]

Hydrogen bonds are much weaker than covalent bonds, typically on the order of a twentieth of their strength. But when there are many of them, their combined strength can be great indeed. A striking example is DNA, in which the opposing strands are held together by hydrogen bonds between the bases, as in Figure 3.6. But more on that later.

If the molecules are rushing about (as in water), they are relatively independent and the substance is a liquid. Hydrogen bonds are constantly formed and broken, forming so-called "flickering clusters". Heat them some more and they separate entirely and the water becomes a gas – water vapor, or steam. The hydrogen bonds between water molecules hold them together pretty well, though, and this accounts for the rather high boiling temperature of water. Chill them down to a temperature where they do not move much any more and

58 Polarization of water molecule, by Jü via Wikimedia Commons, https://commons.wikimedia.org/wiki/File:H2O_Polarization_V.1.svg.

59 Image by Qwerter via Wikimedia Commons, https://commons.wikimedia.org/wiki/File:3D_model_hydrogen_bonds_in_water.svg.

60 Lehninger, 46.

61 Image by Yikrazuul via Wikimedia Commons, https://commons.wikimedia.org/wiki/File:3D_model_hydrogen_bonds_in_water.svg.

the hydrogen bonds assemble the molecules into a solid lattice or crystal – ice.

3.3.2. Ionization, plus hydrophobic and hydrophilic molecules

Because of its polarization, water can pull apart loosely linked ionic molecules, such as table salt, NaCl, where the positive sodium Na^+ is attracted by the negative end of the water molecule and the negative chlorine Cl^- by the positive end. This is what makes water a good solvent. One can see the advantage of this from another angle. Remember entropy? Nature wants higher entropy, meaning more disorder. But NaCl forms a highly ordered crystal structure. When the molecules are pulled apart in water, a more disordered state is achieved and entropy increases.

On the other hand, non-polar molecules are not soluble. They are called ***hydrophobic***, because they "fear" water. NaCl "likes" it and so is called ***hydrophilic***. This has some amazing and important consequences.

The behavior of solvents in aqueous solutions is a very important subject in biochemistry -- and a fairly vast one. Let us look at one interesting and essential type of compound: **Ampiphatic** compounds have some regions that are polar or charged, therefore hydrophilic, and others that are not polar or charged and so are hydrophobic. In the figure below, we consider molecules illustrated as having a green hydrophilic head and long, yellow hydrophobic tails. When they are dissolved in water, the hydrophobic parts flee the water and tend to group together (like people grouped together facing outwards in the midst of a pack of threatening wolves), leaving the hydrophilic parts on the outside turned towards the water. The result is a spherical blob called a ***micelle***, as shown in part 2 of Figure 3.7.

One can understand this from thermodynamics, too. The water molecules which isolate hydrophobic compounds are highly ordered around them. Hiding the hydrophobic parts of an ampiphatic compound on the interior of the micelle reduces the amount of ordering of water molecules and therefore represents a state of higher entropy.[62]

There is another possibility. Think of the micelle opened up, like an orange, and flattened out and another one placed against it, so that the hydrophobic ends are against each other and isolated from the water by the hydrophilic ends on the outside, as in part 1 of Figure 3.7.

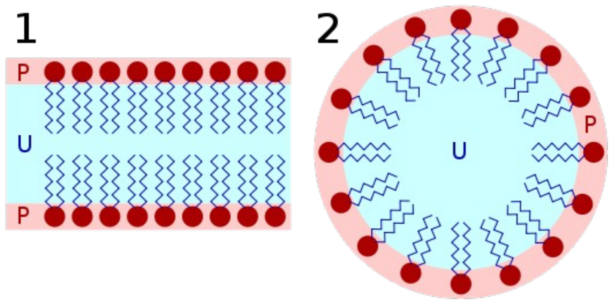


Figure 3.7: Lipid bilayer and micelle, from Wikimedia Commons[63]

This ampiphatic substance could be a lipid (organic fat), in which case this is a ***lipid bilayer***, which is what forms ***cell membranes***. So we are ready to start looking at ***cells*** in the physiology chapter. And all that is due to electrostatics, QM and thermodynamics -- it's all simple physics.

3.3.3. Diffusion and osmosis

Collective or intensive properties like pressure or boiling and melting points, which are independent of the amount of a substance, are called ***colligative properties***. The concentration of a solute is such a property. A solution wants to have the same concentration everywhere, as this represents the state of highest entropy (randomness). So in case of non-equilibrium, a solute will migrate from any region of higher concentration to regions of lower concentration -- just like heat energy flows from hotter to cooler, and for similar reasons. When both regions are mixed and at the same concentration, the result is less ordered and so of higher

62 Lehninger, 49.

63 Image by Stephen Gilbert via Wikipedia Commons https://commons.wikimedia.org/wiki/File:Lipid_bilayer_and_micelle.svg.

entropy. This net movement of molecules from a region of higher concentration to one of lower concentration is called ***diffusion***.

Another very important colligative property is osmotic pressure. This is a bit trickier to understand. Normally, one expects a solute to diffuse from a region of higher concentration to one of lower concentration in order to bring about equal concentrations of the solute. But if two such regions are separated by a membrane which the solute cannot cross but the solvent, such as water, can, then the opposite happens. Some of the solvent flows from the region of lower concentration, i.e., where there is less solute, to the region of higher concentration, which has the effect of diluting the latter and lowering its concentration. At the same time, the solute concentration on the other (source) side goes up. This process of movement of a solvent across a semi-permeable membrane in order to equalize concentrations of a solute on either side is called ***osmosis***. The movement is associated with a pressure driving the water across the membrane and this is called ***osmotic pressure***.

So, in diffusion, the solute migrates; in osmosis, the solute cannot cross the membrane, so the solvent migrates.

Important: The concentration of solute depends not on its mass but only on the number of atoms or ions.

If the membrane is a cell membrane, then water flows into or out of the cell, depending on the solute concentration inside and outside. Cells usually have a higher solute concentration of biomolecules inside than out, which drives water into the cells. If unchecked, the inflow of water could cause the cell to expand until it exploded, but nature has come up with mechanisms to prevent this catastrophe, including reinforcement of cell walls and pumps to remove water from the cell.

In plants, osmotic pressure stiffens cells which have reinforced cell walls, giving the plant rigidity to support it standing up. The opposite thing happens when a salad leaf wilts.

3.3.4. Buffering – acids and bases

Water is naturally somewhat ionized.

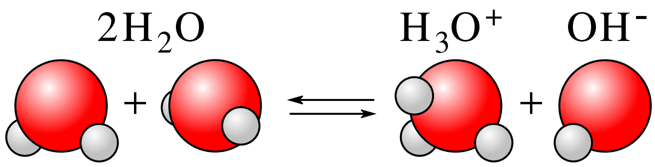


Figure 3.8: Auto-ionization of water, from Wikimedia Common[64]

There are no free protons in water (even though we often will write them as such), but rather ***hydronium*** ions, H_3O^+, with an extra proton. Acids and bases are defined such that

- an ***acid*** is a proton donor (furnishes H^+) and
- a ***base*** as a proton acceptor (consumes H^+).

Water is to some weak extent both: H_3O^+ is a donor and OH^- is an acceptor. In equilibrium at 25°C, water is constantly ionizing and rejoining, so the reaction

$H_2O \leftrightarrow H^+ + OH^-$

is characterized by a fixed equilibrium constant

$$K_{eq} = \frac{[H^+][OH^-]}{[H_2O]} = \frac{[H^+][OH^-]}{55.5}.$$

The ***ion product*** of water

$$K_W = [H^+][OH^-] = 1.0x10^{-14}$$

is thus constant also. Exactly equal concentrations of both ions then means that each has a concentration of $1x10^{-7}$.

64 Image by Cdang via Wikimedia Commons, https://commons.wikimedia.org/wiki/File:Autoprotolyse_eau.svg.

The degree of acidity is frequently indicated by the ***pH*** value, where

$$pH = \log(1/[H^+]) = -\log([H^+])$$

where $[H^+]$ is the concentration of H^+ in moles per liter.[65] As we have just seen, water at 25°C has a pH of 7; pH < 7 means more H^+ and therefore more ***acidic***; pH > 7 means basic, or ***alkaline***.[66] As for all chemical transformations, there is an equilibrium point for the above reactions. This is also true for any other weak acid dissolved in water, i.e., in aqueous solution.

Consider acetic acid,

$$CH_3COOH \leftrightarrow CH_3OOO- + H^+$$

which occurs in an equilibrium state of acetic acid itself (an acid, therefore a proton donor) and CH_3OOO^- (a base, or proton acceptor). These two substances constitute a ***conjugate acid-base pair***. When this weak acid is dissolved in water, two equilibria must be established at the same time, for water and for acetic acid, here represented simply as HAc.

$$H_2O <-> H^+ + OH^-$$

$$HAc <-> H^+ + Ac^-$$

Now if we add a small quantity of a basic substance, say NaOH, to this solution, the base will decompose into Na^+ and OH^-. The latter, being a proton acceptor or base, will therefore increase the pH of the solution. But the concentration of HAc will adjust in turn so as to decrease the pH and the alkalinity. The resulting overall increase in alkalinity will be, in the best of cases, less than expected just considering the addition of a small quantity of strong base.

Seen from the point of view of radicals, the OH^- from the strong base will combine with some of the protons from the water and acetic acid. But then the acetic acid will be out of equilibrium, so it will produce more free protons to re-establish its equilibrium, thereby attenuating the effects of the added NaOH. A similar but opposite mechanism acts to maintain pH if a small quantity of strong acid is added.

This ability to reduce induced acidity is called ***buffering***. A ***buffer*** is an aqueous system which resists changes in acidity from a small amount of added base or acid. It is composed of a weak acid and its conjugate base. It is important as the mechanism by which living beings adjust the acidity of cells. If body acidity is not within rather strict limits, enzymes will not function and so neither will we. The body uses a buffer system based on the conjugate pair carbonic acid and bicarbonate:

$$H_2CO_3 <-> H^+ + HCO_3^-$$

If blood acidity starts to become too high, bicarbonate leaps in and absorbs protons. If it becomes too low, carbonic acid supplies them.[67] This is one of many regulative mechanisms the body has for maintaining the proper equilibrium of certain solutions and processes necessary in order to stay alive. We will see more.

3.3.5. The global water cycle

Let us briefly leave the microscopic considerations of chemistry and look at water on the scale of the Earth. Water circulates through the ground, streams, oceans and lakes and the atmosphere in what is called the water cycle.

65 A mole is the mass of a substance containing the same number of fundamental units (atoms, molecules, etc.) as there are atoms in 12.000 g of ^{12}C. This number is 6.023×10^{23}, which is called ***Avogadro's number***, designated by N_A.

66 The opposite of acidic is alkaline. Why?

67 It's really a tad more complex because of another equilibrium: $H_2C0_3 \leftrightarrow CO_2 + H_2O$. See Lehninger, 63.

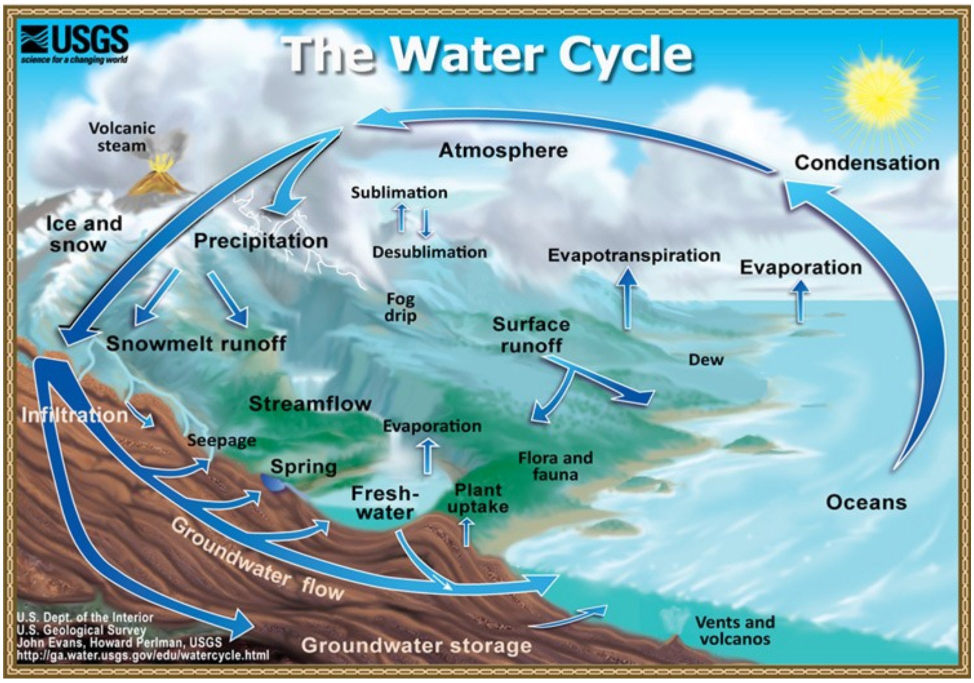


Figure 3.9: The water cycle[68]

This is just one of a number of transforming processes which assure the distribution of an essential component of life on Earth. The diagram is pretty much self-explanatory.

That's it for the introductory material. Now let's look at the history of it all. That starts in the past. Way back in the past, about 13.7 billion years ago (Gya).

68 Image from USGS, http://water.usgs.gov/edu/watercycle.html.

4. What cosmology and astronomy tell us

Cosmology is the study of the origin and evolution of the cosmos, where by cosmos we mean no less than everything. This is a subject which has always fascinated mankind and has given rise to many fanciful and often farcical stories. It is one of the principle reasons for the invention of religion.[69] We looked at one such story in the introduction, the cyclic creation, maintenance and destruction of the cosmos by Brahma, Vishnu and Shiva. Now let's get serious – or, rather, natural.

4.1. *What we know about the Big Bang*

We know from astronomical observation that the universe – space – is expanding. The stars, galaxies and other distant celestial bodies are moving away from each other and so from us (and we, from them). Observation of type 1a supernovae (explained later) has shown that for the last 7 Gy the expansion has been accelerating slowly.

If we do a backwards extrapolation of the measured expansion, we find that about 14 Gya all spatial objects were in the same place. This figure has recently been refined to 13.8 Gya.[70] The entire visible universe once occupied a very small region, smaller than the size of an atom.

It is important to understand that we can see only part of the cosmos, the visible universe, what we call “our Universe”. It is limited in size because of the time light takes to get from a distant point to us and because of the age of the universe. Light travels with a finite speed, so the more distant an object or event, the more time light takes to reach us and the farther back in the past the event we are seeing. The most distant objects we can see are therefore the oldest, those which emitted their light at the time of the Big Bang, about 14 Gya. Because space has been expanding in the meantime, such objects are now about 46 Giga light-years away, which is thus the current radius of the visible universe.[71] It is getting bigger with time.

4.2. *Current hypothesis – inflation*

Astronomers, cosmologists and physicists – the folks who worry about this sort of thing -- generally accept the Inflationary Big Bang as the explanation of the origin of our universe, simply because it is capable of accounting for several facts left unexplained by the original, non-inflationary Big Bang hypothesis.[72] Or so they say...

The inflationary Big Bang framework says that the universe has evolved in a two-step process:

- a brief and extremely rapid expansion, called ***inflation***, followed by
- subsequent, slower expansion powered by the negative pressure of gravity (the cosmological constant).

In a nutshell, our universe started out smaller than the current size of an atom. It was roughly uniform but with tiny quantum-mechanical fluctuations (often referred to as “jitters”). Almost instantly, on a human time scale, it inflated (expanded) enormously to about the size of a grapefruit.[73]. The rapid expansion isolated the quantum jitters before they could resolve. When inflation stopped, a mainly uniform space had been filled with matter formed from the spread-out quantum fluctuations. Since that time, it has expanded well beyond what we can see of it.

It must be stressed that when we speak of the ***expansion of space***, we mean the “fabric” of space itself. The expansion is a property of space only: As space expands, the push to expand remains the same at each (new) point and does not dilute away. Only space expands and only on very large scales. On smaller

69 The other principal reason is the control it can give over people.

70 According to the WMAP project, the age of the universe is 13.77 ± 0.059 billion years. http://map.gsfc.nasa.gov/universe/uni_age.html. The ESA Planck probe finds 13.799 ± 0,021 billion years. We will settle for 13.8 Gya.

71 Carroll (2010), 387n38.

72 Some scientists contest inflation, even hotly. See, e.g., Hossenfelder, “Is the inflationary universe a scientific theory? Not anymore.”. http://backreaction.blogspot.fr/2017/10/is-inflationary-universe-scientific.html

73 Tegmark (2014), 117.

scales, the other physical forces, such as gravity, keep particles and planets together. Although the space around our galaxy, the Milky Way, is expanding, the galaxy is not. Nor are the stars, nor the planets, nor you nor I.[74]

One of the strong points of the inflationary Big Bang is that it explains how the expansion started. It was due to the properties of the inflaton field. Read on.

The origin of inflation is thought to be due to a field called the ***inflaton field*** (or ***false vacuum***). This field was originally in a metastable state of high energy in which it exerted a huge outwardly-directed force which caused space to expand enormously, in fact exponentially, doubling its size every 10^{-37} seconds.

Such an increase in size rapidly becomes huge. At that scale, doubling its size every 10^{-37} seconds, one hundred doublings would have taken about 10^{-35} seconds, but would have resulted in a total increase by a factor of $2^{100} = 1.3x10^{30}$.[75]

Here's more detail. The inflaton field was in a state of unstable equilibrium. Anyone who has ever been delicately balanced, say on a taut rope, so that if she leans one way or the other she starts to fall, understands what it means to be in unstable equilibrium. It's why the police have suspected drunks try to walk a straight line. At about 10^{-35} seconds of age[76], the field "lost its balance" and started to "fall" (metaphorically) off its high value. After about 10^{-32} seconds, space had "fallen" as far as it could in the inflaton field, effectively hitting "bottom". More precisely, it reached its point of minimum energy in the inflaton field. The inflation stopped but the universe went on expanding, but more slowly, under the outwardly-directed pressure of gravity (the cosmological constant). The energy released by the "falling" inflaton field provided the energy and matter that constitute the universe today. As each bit of matter exerted a gravitational attraction on every other bit, the rate of expansion slowed down. When the universe attained the age of about 7 billion years, the expansion rate began accelerating. More on that later.

At its pre-inflation size, the universe was so small that it must be explained using quantum mechanics, according to which states of energy normally forbidden by the Law of Conservation of Energy can exist for very short periods of time, as explained in section 2.2. So there were a great number of relatively very tiny places where the energy was higher or lower than the average. The rapid inflationary expansion caught these energy fluctuations unaware, so to speak, and scattered them out over macroscopic scales before they could merge and even out the energy. They gave rise to particles and antiparticles. This tiny non-uniformity has been observed in the cosmic microwave background radiation (explained shortly).

The inflationary Big Bang solves the so-called ***horizon problem***: How the universe managed to reach large-scale uniformity of temperature without there having been enough time for light to traverse it. Before the inflaton field started "falling", the universe was so tiny that light could travel from any part to any other, so it managed to reach a state of internal equilibrium in which it was all at practically the same temperature everywhere. As the universe inflated, this region of almost constant temperature did too, so the universe we see today is approximately homogeneous, at least on a large scale, a fact which is confirmed by satellite observation.[77] Objections to the need for inflation in order to explain this include "What's to explain? It was uniform to start with." to the reminder that understanding the first moments of the Universe really requires a theory of quantum gravity and we don't have that.[78]

Inflation also resolves the so-called ***flatness problem***, since it was so fast that the resulting universe looks flat. (Think of the surface of a balloon which is rapidly inflated to a very large size.[79]) In cosmology, flatness is very important, as it means the expansion will continue; otherwise, the universe would begin to collapse back together, terminating in a "Big Crunch". (More on flatness later, too.) Objectors claim that this is an esthetic problem and therefore not to be taken seriously.

74 If you are expanding, you cannot blame it on universal inflation.

75 Greene (2004), 308, Guth (1997), 173. S Max Tegmark says the doubling may have occurred every 10^{-38} seconds.

76 Greene (2004), 285.

77 NASA WMAP project, http://map.gsfc.nasa.gov/universe/bb_cosmo_fluct.html. These results have been refined by the more recent Planck probe.

78 Hossenfelder, Sabine. |Is the inflationary universe a scientific theory? Not anymore." http://backreaction.blogspot.com/2017/10/is-inflationary-universe-scientific.html

79 Better yet, glue some 2-centime pieces to the balloon, before you inflate it more. You will see how the coins move apart without getting any larger

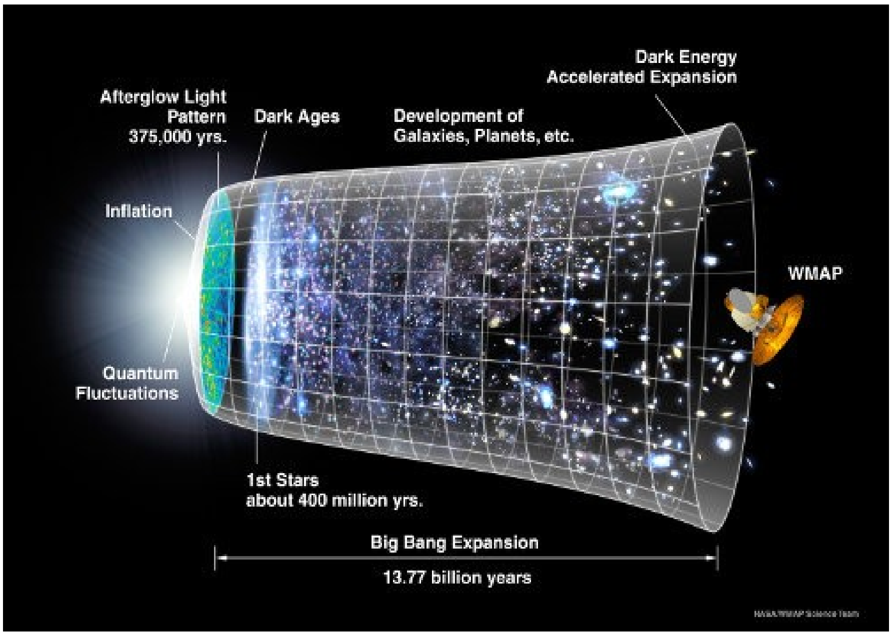


Figure 4.1: Time line of the inflationary Big Bang[80]

There are several reasons to think inflationary exponential expansion took place. It explains why expansion took place (the inflaton field), as well as why space has a uniform temperature everywhere on the average. And it shows how perfectly "normal" quantum-mechanical phenomena (briefly-existing spots of non-conserved energy) gave rise to packets of energy and eventually particles. But all this comes at the cost of yet another hypothetical field.

This much is agreed on by almost all versions of the inflationary Big Bang theory, but there are several versions of it – far too many, according to critics.

4.3. Infinite expansion and the inflationary multiverse

No one knows for sure whether our universe is infinite or not. Period.

There are many versions of inflation. In some models, expansion goes on forever. In small, local regions, the inflaton field may decay non-uniformly into “bubbles” (like boiling water). Each bubble may fall off its metastable energy peak to give up energy to form … a universe! Each one looks like ours to the extent that it has been born out of an inflationary burst, but the physics may vary from one to another. (This process has been compared to cell division[81], but does this suggest that the inflationary multiverse should be regarded as an evolutionarily evolving population of universes? I think not.)

Such universes within the inflationary multiverse have been called “***pocket***” or ***“bubble” universes***. Each one is infinite. Although the pocket may seem finite from the outside, time is transformed into space inside due to the effects of relativity, making the interior infinite!

Since inflation continues forever in this scenario, the Big Bang at the origin of our universe was really the moment when inflation in our part of the multiverse stopped.

The multiverse hypothesis is a suggestion on the extreme edge of cosmology[82], but it is taken seriously by many cosmologists.

4.4. L'après Big Bang – nucleosynthesis and background radiation

Once the inflaton field's energy was spent, inflationary expansion ended. The universe went on growing under the pressure of the cosmological constant, but the gravitational attraction of all that mass limited the

80 WMAP project, http://map.gsfc.nasa.gov/media/060915/.

81 Guth (1997), 251.

82 For more multiverse hypotheses, see Hossenfelder,”The multiverse hypothesis: Are there other universes besides our own?” https://backreaction.blogspot.com/2019/06/the-multiverse-hypothesis-are-there.html

expansion. As the universe expanded, it cooled.[83] It was then that the first step in the formation of chemical elements began.

Big-Bang nucleosynthesis, the formation of the light chemical elements, hydrogen, helium and lithium, took place in the period from about 1 sec (or less) to 3-4 minutes after the Big Bang. Here's how.

At .1 sec ABB, it was too hot for atomic nuclei to form. The universe was filled with rapidly moving protons and neutrons, electrons and positrons, neutrinos and antineutrinos and lots of radiation (photons). Protons and neutrons constantly changed into one another.[84]

Remember the four basic forces of physics? At such high temperatures, their impact was negligible. But as temperatures fell and particles moved more slowly, the strong force was able to exert its influence – by forming nuclei (but not yet atoms).

At around one second of age[85], the universe's temperature was "cool" enough that the frenetic activity of all those particles calmed some. Protons stopped changing into neutrons but since neutrons are slightly heavier than protons, they still decayed into protons and electrons – and still do. Neutrons which had not decayed combined with protons to make an isotope of hydrogen called deuterium (^{2}H, one proton + one neutron). Electrons and positrons annihilated to form huge numbers of photons, visible today as the ***Cosmic Microwave Background radiation***. By 30 seconds of age, about half the electrons and positrons had mutually annihilated.

At about 3 minutes of age, the universe became an element-producing nuclear furnace. Deuterium nuclei combined to form helium (^{4}He). In this way, the lighter elements, including small amounts of lithium (^{6}Li), were formed. This was the era of ***Big-Bang nucleosynthesis***, the first of three steps in the formation of the elements of the universe. At the end of this period, there were about seven protons for every neutron.[86] Free neutrons tend to combine with protons to make ^{4}He, so the Universe was composed of 75% protons and 25% ^{4}He by mass and still is today, give or take a small amount of trace elements.[87].

4.5. The Cosmic Microwave Background Radiation (CMB)

As the universe continued expanding and cooling, more heat was given off in the form of radiation, i.e., photons. The photons had such high energy that they continually knocked electrons loose from any protons they might bond with, keeping all the atoms ionized. So the matter was in the form of a plasma, a "soup" of rapidly-moving free protons, neutrons and electrons. Just as the photons knocked electrons loose from atoms, so the electrons scattered the photons, which were thus like light in a cloud or fog. In a word, the universe was opaque.

After about 380,000 years, the temperature of the universe descended to around 3000 Kelvin and the electromagnetic force was able to bind electrons with nuclei to form atoms, a process cosmologists call ***recombination***. The binding of electrons in atoms was equivalent to a ***phase change*** (like, say, water freezing to ice) and is considered as such by physicists. Photons now were no longer deflected by electrons and the universe became transparent, although the number of photons was so low that it was still dark as there were not yet any stars to make more light. The photon radiation was emitted in all directions and is still "visible" today.[88] This ***Cosmic Microwave Background Radiation (CMB)*** is light from 13.3 Gya.

When we look up into the sky, we are looking into the past, because of the time light from distant celestial objects takes to reach us. But because the universe was opaque until 380 Ky ABB, we can see no farther back than that. The most distant light we can detect is composed of the remnants of radiation emitted then, now reduced by the expansion of space to microwave frequencies, the CMB.

The great uniformity in the distribution of this radiation lends support to the ***cosmological principle***, the

83 Think of how a bicycle tire gets hot as you pump air into it; this is just the opposite phenomenon.

84 The ratio of protons to neutrons was 1.61, due to the equilibrium rate of the reactions.

85 Guth (1997), 94.

86 This number can be calculated (result, about 7.3) from statistical mechanics (Boltzmann distribution), the neutron decay rate and the thermal (temperature) history of the early Universe. See Liddle, 94-95.

87 Take 14 protons and 2 neutrons, make as much He as possible, and you get 12 protons (total atomic mass ~12) and one He (atomic mass ~4), which gives the 75:25 ratio.

88 Since it is in the microwave frequency, we cannot really see it.

idea that the universe is uniform or homogeneous on very large scales, *i.e.*, on the order of 10^6 light-years.

About 200 My after the Big Bang, something happened to dispel the darkness.

4.6. The life of stars

Before about 200 My ABB[89], only simple atoms existed, mostly single protons (hydrogen) and a small number of proton-neutron pairs (helium).They formed clouds of particles which are referred to by the Latin word, ***nebulae***. The coolest and densest of these clouds of gas and molecules, called ***giant molecular clouds***, are propitious to star formation and are where stars are forming today. The composition of nebulae has changed since the first stars formed. We will see how.

The life cycle of stars follows a series of steps.

4.6.1. Formation of protostars

After about 200 My, the universe had cooled enough that particles in molecular clouds moved more slowly and could no longer resist their mutual gravitational force. Dust particles stuck together and formed tiny clumps which in turn collided and adhered together to form larger ones. In this way, larger clumps of gaseous matter were built up. As the density of such a clump increases, it becomes warmer. If its mass exceeds the so-called ***Jeans mass***[90], the pressure due to the kinetic energy of the particles in the clump can no longer resist the pull of its own gravity and it begins to compact. The matter composing it becomes so dense that any generated radiation (in the form of photons) Is scattered within the clump and cannot escape, so the clump becomes opaque. It is now a ***black body***, absorbing all the radiation incident upon it. The trapped photons cause the clump to heat even faster. The temperature increases and the clump starts to emit light. The clump is now a ***protostar***.

The opposing forces of pressure and gravity have opposite thermodynamic effects. Gravity pulls matter into a more organized state, so one of lower entropy. However, the radiation from the outward pressure disperses more than enough energy into space to guarantee a global increase in entropy.

4.6.2. Main-sequence stars – hydrostatic equilibrium and the HR diagram

The evolution of stars differs for different mass stars. We will simplify by considering only general behavior of several broad groups.

A protostar may or may not go on to become a star. If its mass is too low (<8% of the mass of our Sun, or 0.08 $M_{\odot}$), its temperature never becomes high enough for fusion to begin and the protostar becomes what astronomers call a ***brown dwarf***, supported against collapse by ***degeneracy pressure*** due to the Exclusion principle.

Let's consider larger protostars, what we might call "typical" stars, in the range of about 1-8 $M_{\odot}$, so our Sun is included. When the temperature rises to 10^7 Kelvin, there begins to take place a series of nuclear fusion reactions referred to as the ***proton-proton chain***[91], which produce helium from protons. First, two protons fuse together to form deuterium (^{2}H, with one proton and one neutron in its nucleus), then ^{3}He (two protons and one neutron) and finally ^{4}He (two protons and two neutrons). When fusion begins, the protostar has become a ***star***. This first occurred only when the universe was some 200 million or so years old.

The fusion process is often referred to as "burning" of hydrogen, but it has nothing to do with ordinary terrestrial burning, or oxidation. It is really four protons forming two protons and two neutrons held together to form a helium nucleus. Since the He nucleus weighs slightly less than the initial four protons and neutrons, the extra mass is converted into a great deal of energy[92] due to the large size of the speed of light, c, in the famous equation from SR,

$$E = mc^2.$$

89 After Big Bang.

90 The Jeans mass depends on temperature and density, but is typically many thousands of solar masses.

91 De Palma, https://www.e-education.psu.edu/astro801/content/l5_p4.html.

92 If only we could harness this fusion energy here on Earth.

Some of this energy is emitted in the form of light radiated by the the stars, so from the moment fusion began, the universe was not only transparent, there was light (i.e., photons) in it.

Astronomers visualize the properties and lives of stars by a two-dimensional plot called the ***Hertzsprung-Russell diagram*** (hereafter abbreviated ***HR***). The HR is somewhat analogous to the Periodic Table of the elements in that it arranges stars by their properties in such a way as to show similarities among them. It is a scatter plot (points in a 2-dimensional space) of luminosity (or brightness, on the Y-axis) versus temperature (or spectral class, on the X-axis). Stars close to one another on the diagram are found to share many properties. As stars evolve, they may move from one point to another on the diagram, unlike the periodic table, where elements maintain their place. About 90% of stars fall on the diagonal – called the ***main sequence*** – between hotter, more luminous stars in the upper-left-hand corner and cooler, dimmer ones in the lower-right-hand corner. Our Sun lies near the center of the main sequence – for the moment (where the "moment" will last about another 5 Gy). A star's position in the main sequence is mainly a function of its mass and does not change until it leaves the main sequence. The current luminosity of the Sun is taken as the standard with a value of 1.

[?? CHECK OUT SCHUTZ, 10.7 AND ON ...]

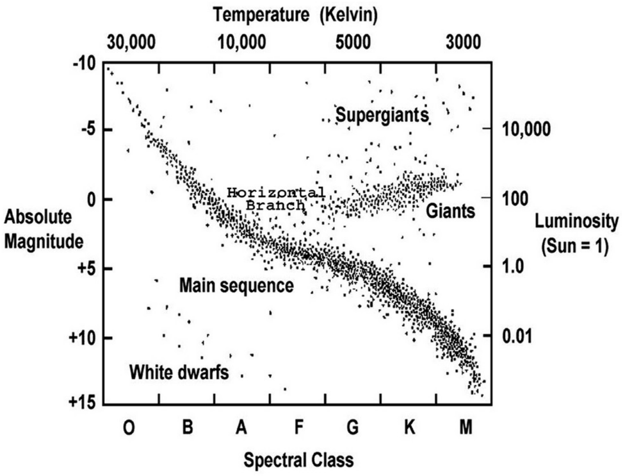


Figure 4.2: HR diagram, from NASA[93]

Main-sequence stars (***dwarfs***) are stars "burning" H to He, thereby maintaining themselves in ***hydrostatic equilibrium*** between the outwardly-directed radiation and thermal pressure due to the fusion in the core and the inwardly-directed force of gravity due to the star's mass. They all burn at a core temperature about that of the Sun, 10 million K, and all share about the same ratio of mass to radius, so heavier stars are bigger stars. Because such a star is a bound system, it is subject to the virial theorem. (See annex A for details.) A shrinking star increases the magnitude of its gravitational energy (reduces it, since it is negative). The virial theorem requires the star to convert half of this energy to kinetic energy, so the star heats up, causing it to expand some. The other half is radiated as star light. The star thus auto-corrects its rate of fusion as if it had an internal thermostat. An increase in burning increases the temperature and kinetic energy, which pushes the volume outwards, then the expansion decreases pressure and burning; and vice versa. As long as its nuclear fuel holds out, a star is obliged by the virial theorem to maintain its size and temperature and so to remain at its position on the main sequence.

Larger stars burn faster and so have briefer lives than smaller stars. But in all stars, the hydrogen fuel in the core is eventually diminished. The pressure then is reduced to where it can no longer counteract the force of gravity and the core begins to contract.

4.6.3. Red giant phase

When core fusion in a main-sequence star slows because of diminished fuel[94], the core contracts slowly, due to degeneracy or gas pressure, and so heats up. This additional kinetic energy now comes not from fusion but from conversion of gravitational energy. Since there is still some hydrogen outside the core, the heat thus

93 INASA/CXC/SAO, http://chandra.harvard.edu/edu/formal/stellar_ev/story/index3.html.

94 When He reaches about 10% of the core mass by weight. King, 54.

generated causes hydrogen in a thin layer on the outside of the core to begin to fuse into helium in ***hydrogen shell fusion***, as opposed to the hydrogen core fusion which has taken place until now. The rate of fusion in this stage is much higher than in the main sequence, so the stage does not last nearly as long, somewhat less than 10% of the star's main-sequence lifetime. Observation of the outer envelope, which is all we can see, will first show a slight cooling with only a small increase in brightness. But the outer layers of the star, being farther from the center, are less constrained by gravity than the core, so the increased pressure from the shell fusion will bring about an increase in size of the envelope of a hundredfold or more at approximately constant temperature – the short horizontal path in Figure 4.3. The star will come to be orange to red in color. It will now be a ***red giant***. When the Sun reaches this stage, it will probably engulf the inner four planets – among them, the Earth. During this phase, fusion occurs only in the shell, so the core is far from equilibrium and continues to contract. For stars less than about 3 $M_{\odot}$, the core becomes a degenerate gas. (More details in Annex A.) In spite of being much bigger, the red giant has less mass than its parent star because of the energy lost through expansion and radiation.

Of the stars we see in the night sky, many are red giants, a notable example being Betelgeuse in Orion.

All stars pass through a red-giant phase. What happens afterward depends on the mass of the star.

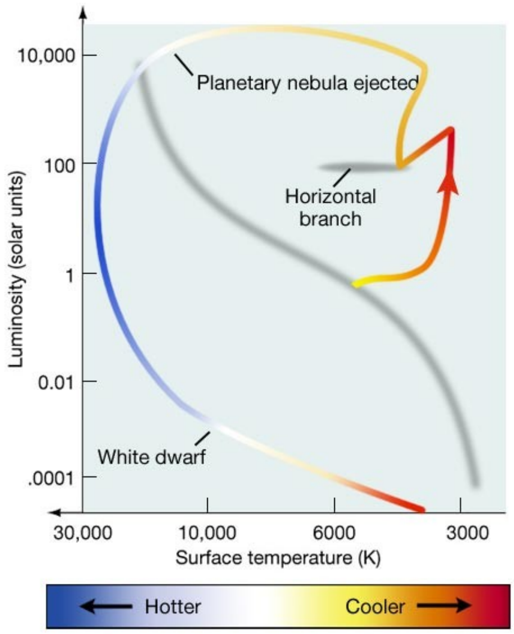


Figure 4.3: Complete evolutionary track of a Sun-like star[95]

4.6.4. Triple-alpha phase, supergiants, white dwarfs

In red giants from stars less than about 8 $M_{\odot}$, the core still is not in equilibrium and continues to contract until it hits heats up to 100 million K, where the pressure is sufficient to maintain it. As a result, the helium from the preceding hydrogen core fusion begins to fuse into heavier elements, ^{8}Be, ^{12}C and ^{16}O. Since ^{12}C formation requires three alpha particles (^{4}He nuclei), the process is called the ***triple-alpha process*** and the star is in the ***core helium fusion*** phase. The star now has two layers, the inner one fusing to heavier elements than the outer one. In lighter stars, less than 2 $M_{\odot}$, with cores which are already degenerate, the onset of He fusion is so rapid it causes a flash, the **helium flash**.[96] The star is now on the so-called ***horizontal branch*** of the HR diagram.

For the Sun, this stage will last less than about 1 billion years, during which it will "burn" ^{4}He into ^{12}C and ^{16}O. A star of about its size will live for a total of around 10 Gy[97]. Since the Sun has already lived for almost 5 Gy

95 Image from Pennsylvania State University Astro801 project, Creative Commons, https://www.e-education.psu.edu/astro801/content/l6_p2.html.

96 Atronomy 801,

97 Most authors would say 10 billion years, but in order to be consistent with later chapters, we will say 10 Gy.

(as we shall see in the section 5.2.5, it has about 5 Gy left. We have fewer...

The core helium is used up faster than the hydrogen was. When this occurs, the core contracts again. The increased temperature causes a thin shell of remaining helium just inside the hydrogen-fusion shell to go on fusing. The core now has three layers: an inner layer of carbon and oxygen, a middle layer of helium fusion and an outer shell of hydrogen fusion. The star expands again, even more than before, and becomes a ***supergiant***.

Figure 4.4: The Cat's Eye Nebula, NGC 6543, a planetary nebula formed by several successive pulses.[98]

For a star not much larger than the Sun, the mass of its shell is now so spread out that the core is incapable of retaining it against the outward force of radiation. The loss of mass (by a process like solar "wind") and the reduced supply of nuclear fuel will lower the core's pressure to where gravity shrinks it to approximately the size of the Earth. Its outer layers will be cast outwards into an expanding shell of gas and for a while it will remain a hot core – a gas composed of atomic nuclei and free electrons – surrounded by a ***planetary nebula*** (cloud) of ionized gas. (The name is a historic misnomer, as a planetary nebula has nothing whatsoever to do with planets.) UV radiation from the cooling core ionizes the gas of the nebula, lighting it up (Figure 4.4). The star now consosts of two separated remnants, the core and the envelope. All that will be left of the core will be a relatively cool and almost invisible ***white dwarf***, composed of a ***degenerate electron gas*** held up by the QM Exclusion Principle. Eventually, the white dwarf will cool into a hunk of cold matter – a less than glamorous end to quite a stellar career. This may however take longer than the current age of the Universe. This will be the fate of any white dwarf less than about 1.4 times the mass of the Sun, a mass known as the ***Chandrasekhar limit***. This corresponds to an initial mass of about 7 solar masses.

4.6.5. Supernovae, neutron stars and black holes

For stars whose mass is initially greater than about 8-10 $M_\odot$, the sequence goes further. In later stages, ^{12}C and ^{16}O may fuse to ^{20}Ne, ^{23}Na, ^{24}Mg and ^{28}Si. As such massive stars pass from step to step of the fusion process, burning successively carbon and oxygen, then other elements, each new step takes place in the core's center, pushing the previous step's reactants outwards. So the star comes to have a number of layers, like an onion, with different fusion reactions taking place in different layers. Each successive step takes less time, the final fusion of Si into Fe lasting only about four days. (This is the ***second step of nucleosynthesis***.) Once the Si is converted to Fe, the core is inert so the pressure drops and nothing can keep the star from collapsing. The star now has left only a few seconds of life.

The outer layers fall towards the center. The core's density and temperature shoot up. Protons and electrons combine to form neutrons and neutrinos, the neutrinos inheriting most of the enormous quantity of gravitational energy released by the contracting core. The infalling shell meets the energetic outgoing neutrinos. Although only a small fraction of neutrinos interact with the shell, their enormous energy is sufficient to literally blow the shell away, causing the whole structure to explode and become a ***Type II supernova*** (***SNII***), or ***core-collapse supernova***. Such an enormous amount of energy is released that the light is as bright as that of several billion Suns. For some days or weeks, it may be the brightest object in

98 ESA/Herschel/PACS/MESS Key Programme Supernova Remnant Team; NASA, ESA and Allison Loll/Jeff Hester (Arizona State University). https://www.jpl.nasa.gov/spaceimages/images/largesize/PIA17563_hires.jpg

the sky, depending on how far away it is. Light from the supernova gets brighter over a period of about 20 days and then decreases slowly over a year or so.

Figure 4.5: The Crab Nebula, remains of a supernova explosion, seen by Chinese astronomers in 1054. From NASA[99].

Meanwhile, back in the core, since neutrons are not mutually repulsed by electromagnetic forces, they are pulled together until they form a smooth "pudding" of the density of a nucleus (!) and a size of several tens of kilometers. They are held apart by neutron degeneracy pressure, due to the Exclusion principle, and the hard core of the strong-force potential, which becomes repulsive at very close distances. For larger neutron stars, above about 0.7 $M_{\odot}$, the strong force is more important. The result is now a ***neutron star***.

Uncertainty means some neutrons will have enough energy to escape the gravitational pull unless there is a certain minimum mass. From this fact, the calculated minimum mass is around 0.1 $M_{\odot}$. But of course at any mass below the Chandrasekhar mass (about 1.4 $M_{\odot}$), the star forms a white dwarf, so that is the effective minimum mass of a neutron star.

Recent studies have discovered that neutron stars have an internal structure in at least three layers. From outside to inside:

- A gaseous "atmosphere" several centimeters thick, composed of hydrogen and helium.
- A crust, about a kilometer thick, of nuclei of heavy elements, especially iron, arranged in a crystal structure.
- Perhaps one or two more layers whose constitution is not agreed on.

Deeper and deeper in the inmost layer(s), the pressure goes up and the nuclei are composed more and more of neutrons. At some point, there are no more nuclei, but a plasma of neutrons and electrons. One hypothesis is that, still farther in, the neutrons decompose into quarks and gluons. If the quarks recombine in pairs to form bosons, to which the Exclusion Principle does not apply, the density could increase farther to where the pairs form a superfluid of zero viscosity.

GR considerations lead to a maximum possible mass of 2-3 $M_{\odot}$, above which gravity is so strong that the star collapses to form a ***black hole***. Fortunately, the maximum neutron-star mass is greater than the Chandrasekhar mass or no stars producing heavier elements would form – and we need them.

Before the supernova stage, some of the neutrons created by the combination of electrons and protons react with the Fe to create heavier, radioactive elements which subsequently are expelled into the supernova. These radioactive elements then decay with emission of gamma rays. It has been proposed that such gamma rays from a huge supernova may have been the cause of the Ordovician-Silurian extinction around 447-443 Mya (see the geology chapter, section 5.6.2). Be that as it may, many of these atoms combine by simple chemical reactions and form a new, but richer, dust cloud in space, richer because it contains all those heavier elements produced by the first-generation stars. These elements include C and O which are

99 NASA, https://www.nasa.gov/sites/default/files/images/148387main_image_feature_567_ys_full.jpg

necessary for life as we know it. This cloud may be the source for new stars. This is the ***third and final step in nucleosynthesis***.

The few best measured neutron stars are of about 1.4 $M_\odot$ and this may be true of all of them.

So the core of a massive star remains as a neutron star. If it had been a smallish star like the sun, it would have formed a white dwarf; if very massive, a black hole.

4.6.6. One more time...

A minimal outline of stellar evolution would run like this.

1. Main-sequence, hydrostatic equilibrium, gravity vs core hydrogen fusion.
2. Red giant, hydrogen shell fusion only.
3. Core He fusion, triple-alpha process, two fusion layers.
4. Supergiant, three layers – C and O core, He fusion, H fusion.
5. Then there's a choice, by mass:
   A) Sun-size stars: planetary nebula + white dwarf.
   B) Larger stars > 8 $M_\odot$, more phases and layers, up to Fe core. Either
      a) core-collapse supernova (SNII) + degenerate neutron star, or
      b) above 2 $M_\odot$, black hole.

Stars can end their lives in one of three forms:

- white dwarfs,
- neutron stars, or
- black holes.

Within the matter ejected into space as nebula, the whole process can start over again with agglomeration of dust to form protostars, but with two important differences.

- The nebula dust now contains heavier elements like carbon and oxygen.
- Exploding stars, supernovae, send out shock waves which may provoke the formation of new protostars.

To resume, nucleosynthesis takes place in three steps:

1. In Big-Bang nucleosynthesis, the lightest atoms, H, He and Li are formed.
2. In stellar fusion, heavier elements are forged, through C, O and up to and including Fe.
3. In supernovae, still heavier elements are formed, including radioactive ones.

4.7. Some more exotic stellar beings

Several special cases deserve mention.

4.7.1. Variable stars – Cepheids

The main-sequence expansion to a red giant does not necessarily go smoothly. Some stars will be in unstable equilibrium, like a wobbly taught-wire walker. Pressure may cause them to overshoot and become too big, then gravity will take over and they may become too small. And so on, back and forth. Such stars are called ***Cepheid variables*** (or RR Lyrai variables). The periodically varying temperature changes the ionization of He in the star and this in turn affects its transparency (or opaqueness), thus changing the amount of light it can omit – its luminosity. Cepheids are important to astronomers because there exists a strong, direct relation between the absolute luminosity and the period of pulsation. Comparison of absolute to observed luminosity enables calculation of the distance to the stars. It also leads to a determination of the age of the Universe of about 13.6 Gy, in agreement with the CMB results.[100] Cepheids occupy a specific region of the HR diagram.

4.7.2. Type I supernovae and the expansion of space

Many – maybe most – stars in the universe are binary, meaning that two relatively nearby stars orbit around the center of gravity of the pair. If one of them is a white dwarf below the Chandrasekhar limit, it may pick up matter radiated by its twin. If enough such matter is accreted to its surface, it may become sufficiently massive, compared to the Chandrasekhar limit, that the electron degeneracy[101] can no longer support it and it collapses and explodes into a ***Type Ia supernova*** (***SNIa***). Such objects are extraordinarily bright[102]. They are important to cosmologists because their common origin suggests that they all burn with the same luminosity, making them standard “candles” which can be used to estimate their distance from the Earth (by their apparent brightness) and their speed (from the red shift of their spectra). In this way, astronomers have measured the speed and distance of distant stars and galaxies and have made an extraordinary discovery: The expansion of space has been accelerating for the last 7 billion years.

Remember: SNII are nuclear core-collapse supernovae which leave behind neutron stars or black holes. SNIa supernovae are due to gravitational collapses of old, former white dwarfs and leave nothing behind.

4.7.3. Pulsars

Neutron stars often rotate rapidly, with periods of rotation down to milliseconds, and have strong magnetic fields. Radio-frequency radiation is expelled along the poles and appears to distant astronomers as if the star flashes as it spins, rather like a lighthouse. Such stars are called ***pulsars***. The axis of rotation is not in general orthogonal to the line of sight of the astronomers, so many, if not most, pulsars go unobserved. All the energy of the pulsation comes from the pulsars’ rotation, which thus must and does slow down. This has been well verified, as it is easy to ascertain by measuring periods of many pulsations and calculating the average. Measuring the change in pulsation rate enables calculation of the pulsar’s age. One such pulsar is in the Crab nebula (Figure 4.5), whose birth was observed by Chinese astronomers in 1054 BCE.[103] The calculated age of the pulsar – 1000 years – agrees with this date. The result also confirms that pulsars are rotating neutron stars produced in supernovae.

4.7.4. Black holes

Black holes were originally predicted as solutions of the equation of GR assuming a single, strong source of gravitation in a universe which is homogeneous and isotropic (the ***Cosmological Principle***). They have since been found. Their gravity is so strong that anything, even light, passing a certain radius from the hole can never again escape the inward pull of gravity. The virtual sphere of this radius around the center is called the ***event horizon***. Of course one can not “see” a black hole because they are black against a black background. But, very recently, an international team has “photographed” a black hole as a black spot set off against the radiation around it. (Figure 4.6)

Black holes are one of those subjects where GR confronts QM – a current dilemma in physics. Some understanding is based on applying ideas like the Uncertainty Principle to black holes. This method shows, for instance, that a black hole radiates and so loses energy. Application of thermodynamic principles has led to a formula for the entropy of a black hole, which is proportional to the surface area of its event horizon. It is now thought that there is an enormous black hole at the center of every galaxy.

100 King, 103.
101 Another name for the Exclusion Principal, in this case applied to electrons.
102 The most recently viewed one in our galaxy was observed in 1572 by the young Danish astronomer Tycho Brahe.
103 King, 75.

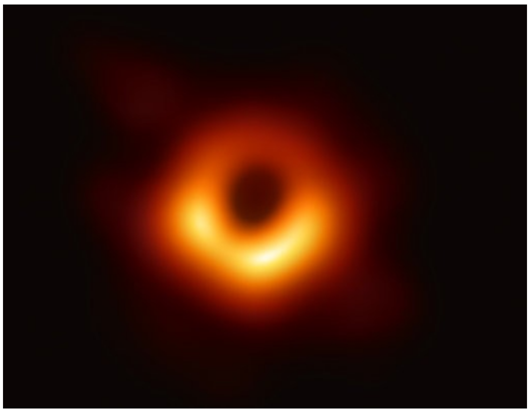


Figure 4.6: Photo of a black hole from NASA. Credits: Event Horizon Telescope collaboration et al.[104]

4.7.5. Quasars

Quasar stands for ***quasi-stellar object***, also referred to as a ***QSO***. Quasars are enormously bright, perhaps brighter than an entire galaxy but with a size less than that of our Solar System. A quasar is an extreme example of an ***active galactic nucleus*** (***AGN***), a compact region at the center of a galaxy, called an active galaxy, with an unusually high luminosity on some region of the EM spectrum (X-rays, UV, microwave, etc.). This radiation is not produced by a star but by energy emitted as matter falls into an accretion disc around a supermassive black hole at the galaxy's center.

4.8. Galaxies, clusters, super-galaxies

Today, there are hundreds of billions of galaxies, each containing hundreds of billions of stars. And that is just for the observable universe.

Space is really, really big. Not only are stars grouped into galaxies, galaxies are grouped into clusters and clusters into ***superclusters***. Stars in a cluster generally have about the same age and were probably formed from the same nebula. Our galaxy, the Milky Way, is a member of a group — too small to be called a cluster — called the ***Local Group***.

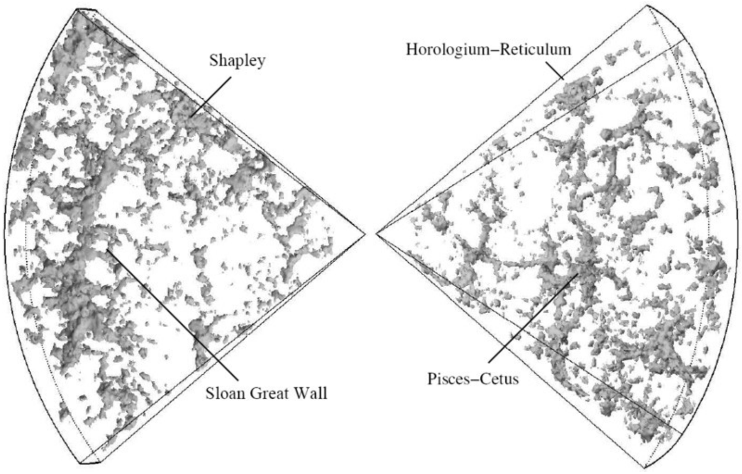


Figure 4.7: Cosmic structures from reconstruction of 2dF galaxy redshift survey, via Wikimedia Commons.[105]

Studies of the large-scale structure of the universe — on scales where galaxy clusters look tiny — indicate that galaxies are spread out in a filament-like structure. It is sometimes compared to soap bubbles, where the galaxies are distributed mainly on the surfaces of the bubbles.

Although the interior of the "bubbles" is less bright, so that they are also called voids, there are galaxies

104 NASA, https://www.nasa.gov/mission_pages/chandra/news/black-hole-image-makes-history.
105 Wikimedia Commons, https://commons.wikimedia.org/wiki/File:2dfdtfe.gif.

inside them, including our Local Group.

4.9. Large-scale structure and geometry of the universe

Assuming the validity of the ***Cosmological Principle***, that the distribution of objects in the universe is homogeneous and isotropic (the same everywhere and in every direction), the equations of GR have different solutions for the geometry of the universe. These solutions correspond to three possible sorts of curvature of the universe (in the GR sense).

The geometry depends on a parameter called the critical density of the universe, ρ_{crit}. One often refers to Ω_0, the ratio of the observed density to the critical density,.

$\Omega_0 = \rho_{obs}/\rho_{crit}$.

- $\Omega_0 = 1$, means the universe is flat and the sum of the angles of a triangle is 180°.
- $\Omega_0 < 1$ means the universe has negative curvature and the sum of the angles of a triangle is less than 180°.
- $\Omega_0 > 1$ means the universe has positive curvature and the sum of the angles of a triangle is greater than 180°.

For Ω_0 >= (greater than or equal to) 1, space is infinite. If space is infinite now, it always has been, including at the Big Bang. That was indeed the universe in a grain of sand!

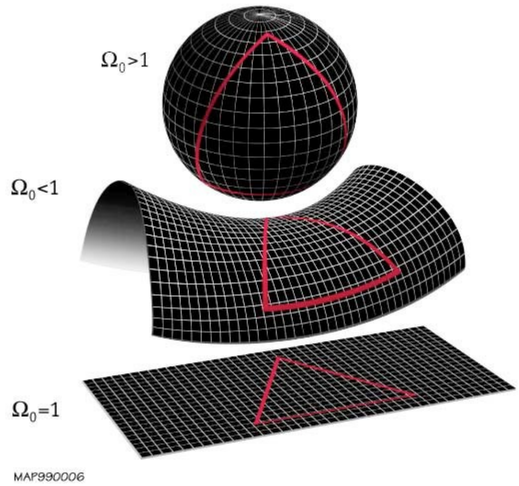


Figure 4.8: Geometry of the universe[106], from WMAP via NASA

4.10. Different forms of matter and energy

It is now thought that the universe is composed of three sorts of matter and energy.

Ordinary matter is the classic one and once was thought to be the only kind. This is the matter we are familiar with and which is composed mainly of protons, neutrons and electrons. Contrary to previous ideas, we now know that it only makes up 4.6% of the energy density of the universe[107]. The rest is constituted of dark matter and dark energy.

It has been observed that the outer parts of galaxies rotate so fast that the amount of matter in the galaxy is insufficient to generate a gravitational force capable of keeping them from spinning off into space. But ***dark matter*** could counteract that centrifugal force. The mass of dark matter provides the additional gravitational pull needed to hold the galaxy together. Similar phenomena, seen for galaxies in clusters and for the

106 From WMAP via NASA, http://map.gsfc.nasa.gov/universe/bb_concepts.html.

107 All percentages in this section come from WMAP, http://map.gsfc.nasa.gov/universe/uni_matter.html.

strength of gravitational lensing, support the existence of dark matter.[108]

Dark matter must be composed of something which interacts only weakly with ordinary matter and is not visible to us. Other than that, its composition is still unknown. It may be non-baryonic matter left over from the Big Bang. Hypothetical dark-matter particles are referred to as ***WIMPs***, Weakly Interacting Massive Particles. Recent studies suggest several types of such particles. Dark matter constitutes 24% of the energy density of the universe[109].

Dark energy is stranger yet. It has been observed from SNIa red shifts that the expansion rate of the universe has been increasing over about the last 7 billion years. The equations of GR contain what at first seemed an arbitrary constant[110], which has the interesting – in this case, very handy – property that it causes space to be suffused with a dark energy of negative pressure which exerts a push, not a pull, on matter. After inflation ended, ordinary gravity attracted matter together and slowed down the expansion. But expansion continued nonetheless. As matter became more diffuse in space, gravity was no longer a match for the dark-energy pressure, which is thought to be a property of space and so is constant and not diluted as space expands. Since about the time the universe celebrated its 7 billionth birthday, the dark-energy push has been stronger than the attraction of gravity and the expansion of the universe has accelerated. Dark energy has been calculated to furnish 71.4% of the energy density of the universe. Almost amazingly, this is just the total energy density required in order for Ω_0 to be approximately equal to 1,[111] in which case the universe is “flat”.

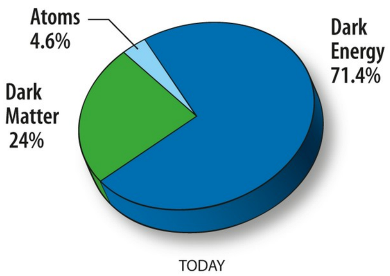


Figure 4.9: Constitution of the universe by types of matter and particles[112]

The bad news is that theoretical calculations of the cosmological constant come up with a value far, far too large to be true. Work continues.

In summary, dark matters attracts and can be diluted by the expansion of space. Dark energy repulses and never dilutes.

It is quite astonishing that only of 4.6% of the universe is constituted of the stuff of which we are made and which we can actually examine.

4.11. Formation of our solar system

About 10 Gya, a new protostar formed in the outer fringe of the the galaxy we call the Milky Way. The agglomeration of dust was probably initiated by shock waves from a nearby exploding supernova. The protostar was turning, and since the centrifugal force was greatest in the plane perpendicular to the axis of rotation, matter surrounding the forming Sun was compressed into a flat disk of dust formed around the big lump in the middle.[113] Increased density meant increased pressure and, therefore, temperature. The lump condensed further and became the proto-Sun. The matter in the disk also clumped together into rotating chunks. Some of these had diameters of 1 km or so and are called ***planetesimals***. The passage from a cloud of dust and gas to a star surrounded by planetesimals probably only took about ten million years.[114] The largest of these chunks formed the planets; the other, smaller pieces became asteroids and comets.

108 Astronomers Measure the Density of Dark Matter in Galaxy Clusters. SciTechDaily. http://scitechdaily.com/astronomers-measure-the-density-of-dark-matter-in-galaxy-clusters/

109 Ibid.

110 Deemed the Cosmological Constant.

111 WMAP finds a value of 1.02+-0.02

112 NASA Universe 101, http://map.gsfc.nasa.gov/universe/uni_matter.html.

113 In a plane perpendicular to its total angular momentum vector.

114 MacDougall 2011, 29.

Heavier elements like iron and silica were attracted by the Sun's gravity and so formed the closer planetesimals which would become the four rocky planets – Mercury, Venus, Earth and Mars. Lighter gases like hydrogen and helium were volatilized by the heat and blown farther out by the solar wind and came to form the giant gas planets, Jupiter, Saturn, Uranus and Neptune. When the centrifugal force of their motion around the Sun balanced out gravity, they were maintained in elliptical orbits at nearly constant radial distances[115] from the Sun. Slowly, the protoplanets swept their orbits clean, merging the encountered rocks with their surfaces. The subsequent history of the Earth is the subject of geology and so of the next chapter.

4.12. Future of the solar system and the universe

According to currently prevailing cosmological theory[116], the ultimate state of the universe is quite literally not bright.

As we have seen, in 5 billion or so years, the Sun will flare up into a red giant and engulf the Earth, including us, unless we first can escape to another solar system, a voyage of at least several generations. But even that would be a temporary respite. Meanwhile, our solar system finally either will be absorbed by the Sun or burnt to ashes before settling into a permanent freeze.

On a broader scale, galaxies will disperse as stars explode into dust floating throughout expanding space-time. As the Universe goes on expanding under the impulsion of Dark Energy, the dust eventually will be too diffuse for another star to form from it under the pull of gravity. The universe will then become a thin "soup" of particles diffused randomly throughout an enormous and ever-expanding void[117]. The expansion of space will prevent any communication between celestial objects because of the speed limit imposed by SR – the speed of light.[118] There will be only cold and darkness, with no starlight to warm and delight us and no "us" there to be warmed or delighted. It will be totally silent, not a bird will sing, nor a cat meow. "The Big Freeze".

This will be the fate of the universe if it is indeed flat, as is now thought. If the energy density is greater than it is thought to be, then the universe may eventually collapse back onto itself in a "Big Crunch" – before maybe starting over. Perhaps that can make us feel better...

Nevertheless, the understanding of the universe around us underscores and nourishes our astonished appreciation of its beauty. The knowledge that time for life and love and beauty, although vast, is limited should remind us to appreciate that we are here now – once – so it is our only chance to make the best possible use of our limited lifetimes – individual or collective. Perhaps we even will manage to develop ways of living which will leave the billions of years of life still left on Earth a place pleasing to our descendants – whatever they may be.

115 Elliptical radii, major and minor, to be sure.

116 There are other, less gloomy, prognostics.

117 It will also possess much greater entropy than it does today.

118 This does not violate Special Relativity. It is not a signal moving faster than light, it is space expanding. Besides, how does one define velocity in expanding space...?

5. What geology tells us

The discoveries made by geologists and paleontologists about Earth's history make one of the most fascinating and absorbing stories ever. That is why so many folk traditions have imagined their own versions of the story in the days before the empirical sciences. Even though the details of some of the mechanisms behind the dynamics of Earth science are not yet completely understood, the knowledge gained explains very satisfyingly – and in great detail – how and why the Earth and its inhabitants, including humans, have come to be what they are today. It also gives hints as to what may happen tomorrow. So read on.

Eon	Era	Period		Epoch	Mya
Phanero-zoic	Cenozoic (mammals)	Quaternary		Holocene	0.01
				Pleistocene	1.8
		Tertiary	Neogene	Pliocene	5.3
				Miocene	23
			Paleogene	Oligocene	35
				Eocene	55
				Paleocene	65.5
	Mesozoic (dinosaurs)	Cretaceous (S. America, Africa)			146
		Jurassic (Alps/Himalaya)			200
		Triassic (Pangea)			251
	Paleozoic (inverte-brates)	Permian			299
		Carboniferous	Pennsylvanian		320
			Mississippian		359
		Devonian (fish)			416
		Silurian			444
		Ordovician			488
		Cambrian (Pannotia, Rodinia)			542
Pre-cambrian	Proterozoic (O2-rich atmosphere)	Ediacaran			635
					2500
	Archean (appearance of life)				3800
	Hadean				4700

Figure 5.1: Geological time scale. Red lines represent mass extinctions.

There is still discussion as to the standard to be used for “million years ago”: “Mya” or “MA”. I will use the more intuitive Mya = million years ago, Gya = billion (thousand million, or giga) years ago. “My” then defines a duration of a million years.

5.1. Some geophysics (Earth physics)

First, let us look at the composition of the Earth.

5.1.1. Minerals and rocks

Almost all rocks are made up of minerals, which are naturally occurring, inorganic, crystalline solids. Geologists distinguish them by their color, luster (how they reflect light), transparency (compare diamond and coal), streak (powdered form), hardness, tenacity, and cleavage and fracture (how they chip or break).

The mass of the Earth is constituted of the following proportions of elements by mass[119]:

- 32.1% iron
- 30.1% oxygen
- 15.1% silicon
- 13.9% magnesium
- 8.8% other elements, especially aluminum and calcium.

The most common elements in rocks at the Earth's surface are silicon and oxygen. These form ***silicate***, a mineral having a tetrahedral structure rather like that of methane, as we saw in the section 3.2. This image shows two of them bound together by sharing an oxygen between two tetrahedra.

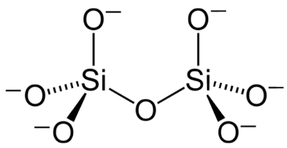


Figure 5.2: Silicate double tetrahedra,by Ben Mills via Wikimedia Commons[120]

This silicate binding can be extended to give the many different crystalline structures, such as rings and chains, which make up the rocks around us. For instance, quartz crystals have the following structure.

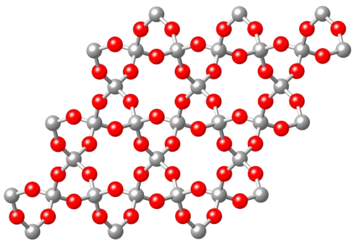


Figure 5.3: Ball-and-stick model of part of the crystal structure β-quartz, a form of silicon dioxide, SiO_2, by Ben Mills via Wikimedia Commons[121]

Beryl has a ring structure, shown here both as atoms and in a tetrahedron representation.

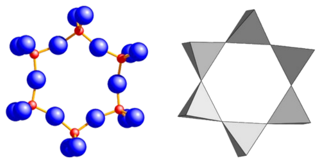


Figure 5.4: Unbranched 6er single ring of beryl from Brown and Mills via Wikimedia Commons[122].

Non-silicate minerals (e.g., carbonates, sulfates, sulfides, oxides) make up only 5-8% of the Earth's crust.

119 Marshak, 142.

120 https://commons.wikimedia.org/wiki/File:Silicate-double-tetrahedra-2D.png.

121 https://commons.wikimedia.org/wiki/File:Beta-quartz-CM-2D-balls.png#/media/File:Beta-quartz-CM-2D-balls.png.

122 https://commons.wikimedia.org/wiki/File:Beryll.ring.combined.png.

5.1.2. Rock cycle and types

As is well known, there are three types of rocks:

- ***igneous*** — formed when a molten rock precursor (magma or lava) cools;
- ***sedimentary*** — formed from mineral or organic particles, such as weathered or eroded rocks or shells, which settle on the Earth's surface or at the bottom of a body of water;
- ***metamorphic*** — formed by the transformation of preexisting rocks by intense heat or pressure,

Rocks may be ***intrusive*** (formed beneath the surface of the Earth) or ***extrusive*** (formed above the surface). Among *igneous* rocks, there are several types based on the amount of silica in them:

- ***felsic*** rocks contain > 65% silica and are generally light-colored;
- ***mafic*** rocks contain 45-55 % silica and are generally dark-colored;
- ***ultramafic*** rocks contain <45% silica.

Some common examples:

- granite — intrusive, igneous and felsic
- rhyolite — extrusive, igneous and and felsic
- basalt — extrusive, igneous and mafic
- sandstone – extrusive and sedimentary
- quartzite – intrusive and metamorphic

Rocks and minerals are formed in a series of cycles. Volcanic activity causes hot magma to pierce the Earth's surface, where it cools and hardens into various sorts of igneous rocks. Other material may be coughed up from below ground at the same time as igneous or metamorphic rock. As these rocks are weathered by ice, water and wind, they are broken down into smaller grains which are then carried to the sea or lake floors, where they accumulate, eventually to form sedimentary rocks. These rocks may then be pushed by subduction down to a hotter, denser level where they are converted into igneous or metamorphic rocks and the cycle continues[123]. Such cycles have been active for 4 billion years.

5.1.3. Interior structure of the Earth

The Earth consists of several layers (Figure 5.5). From the center:

1. The core, around 2500 km in radius, is composed of heavy metals such as iron and nickel and has two layers:
   - the inner core, believed to be solid;
   - the outer core, shown by seismic waves to be a very viscous liquid;
2. The layer outside the core is called the ***mantle*** and constitutes about 2/3 of the Earth's mass. It is around 2800 km in radius and is composed mostly of light elements. Seismic studies show the boundary between core and mantle to be anything but smooth. The mantle has three layers:
   - The ***mesosphere***: Although temperatures here are high enough to melt the rock, the intense pressure keeps it solid.
   - The ***asthenosphere***: This layer is also solid, but can move in a very slow plastic flow.
   - The ***lithosphere***: This is the outermost layer of the mantle and comprises both the outermost mantle layer and the crust. The lithosphere is rigid and relatively brittle. The lithosphere differs from the crust by its mineral composition.
3. Outside the so-called Moho discontinuity in the lithosphere lies the Earth's crust, attached to

123 This is only one of the cycles of nature. It may be compared to, e.g., the water cycle (clouds – rain - bodies of water – evaporation – clouds) or the nitrogen cycle (fixation – assimilation – nitrification - denitrification).

the outer mantle. Above the Moho, the rocks are lighter and felsic, constituted primarily of silica, the principal constituent of surface rocks. There two kinds of crust:

- The thicker continental crust is about 20 km in depth (higher for mountains) and is composed mainly of granites, relatively light felsic rocks. The existence of dry land today is due to the existence of this relatively light rock.
- The thinner oceanic crust is typically less than 10 km thick, and is composed mainly of basalt, which is denser than continental crust.

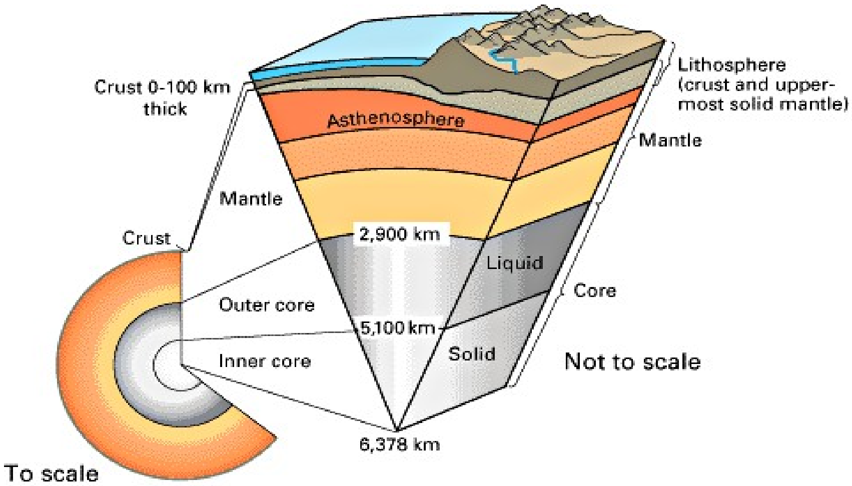


Figure 5.5: The Earth's core, from USGS[124]. Distances are depths, measured from the surface.

The relative densities of the two crust types are important in plate tectonics.

5.1.4. The Earth's magnetic field

Because of its iron core, the Earth is like a bar magnet and has a ***dipole*** magnetic field with north and south poles. This is thought to be the consequence of a dynamo effect, wherein charged particles in the hot core move about because of convection due to the Earth's rotation. Moving charges generate a magnetic field.

The polarity changes randomly in geologic time; the poles actually reverse. If this were to happen today, magnets would indicate Antarctica to be in the north! Fortunately for human navigation, this does not happen frequently, the last time having been about 780 thousand years ago. Geologists find a record of this reversing polarity in the phenomenon known as magnetic striping – variations in direction of the magnetic field of the rock on the ocean floor (Figure 5.6).

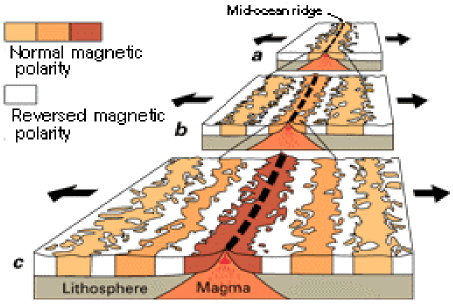
As it “burns”, the sun continually spews out charged particles, mostly electrons and protons, creating a ***solar wind***. Fortunately for us, the magnetic field due to the Earth's molten metal core deflects these particles so that most of them pass around the Earth without doing any harm. Otherwise, they could actually blow away the Earth's atmosphere. Mars is much smaller than Earth and has lost its original molten core and magnetic field, though some residual magnetism has been found trapped in rocks.[125] But the solar wind has stripped away the atmosphere of Mars, leaving the red desert that space probes are exploring today.

As the solar wind passes over the Earth’s poles, where the magnetic field is most intense, they sometimes ionize the air and produce the beautiful wavy colors of the ***aurora***.

The Earth's magnetic field can deflect only a limited amount of the solar wind. In 1859, an enormous solar storm called the Carrington storm released so much electromagnetic energy and solar matter that aurora were seen as close to the Equator as the Caribbean and Santiago, Chile. Telegraph communication was often interrupted and operators reported sparks jumping from their equipment. If that happened again now, things like GPS satellites and telecommunications equipment could suffer.

124 U. S. Geological Survey, “This dynamic Earth”. http://pubs.usgs.gov/gip/dynamic/inside.html

125 “NASA’s Mars lander discovers quakes and a surprising magnetic field”, https://www.washingtonpost.com/science/2020/02/24/nasas-mars-lander-discovers-quakes-surprising-magnetic-field/



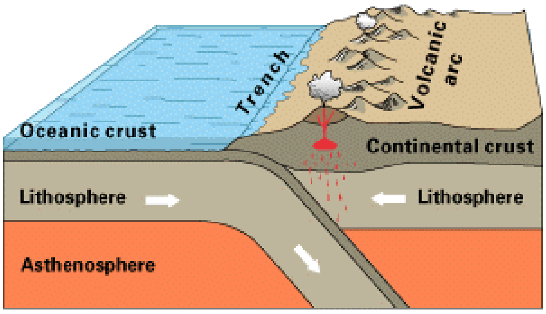


Figure 5.6: Ocean-floor magnetic striping, from USGS[126]

Figure 5.7: Subduction from USGS[127]

5.1.5. Plate tectonics

Plate tectonics (from the Greek *tecton*, builder or carpenter) is the theory which explains how plates of continental crust move about on the surface of the Earth. It answers the question of how and why they move. It explains the existence of volcanoes and mountains and furnishes the background for all Earth science. It has been said to be the theory of “how the Earth works”. It has transformed and integrated modern geology just as quantum mechanics and relativity have modern physics or evolution, biology.

The idea of plates, continent-sized hunks of lithosphere moving on the surface of the Earth, was first suggested by the comparison of the eastern border of South America with the western border of Africa. Since then, understanding of the Earth's interior has explained how this works, although some details of the driving mechanism (especially, mantle plumes and hot spots) are still subjects of discussion.

The Earth's lithosphere is composed of plates, seven main plates and numerous smaller ones (Figure 5.8).

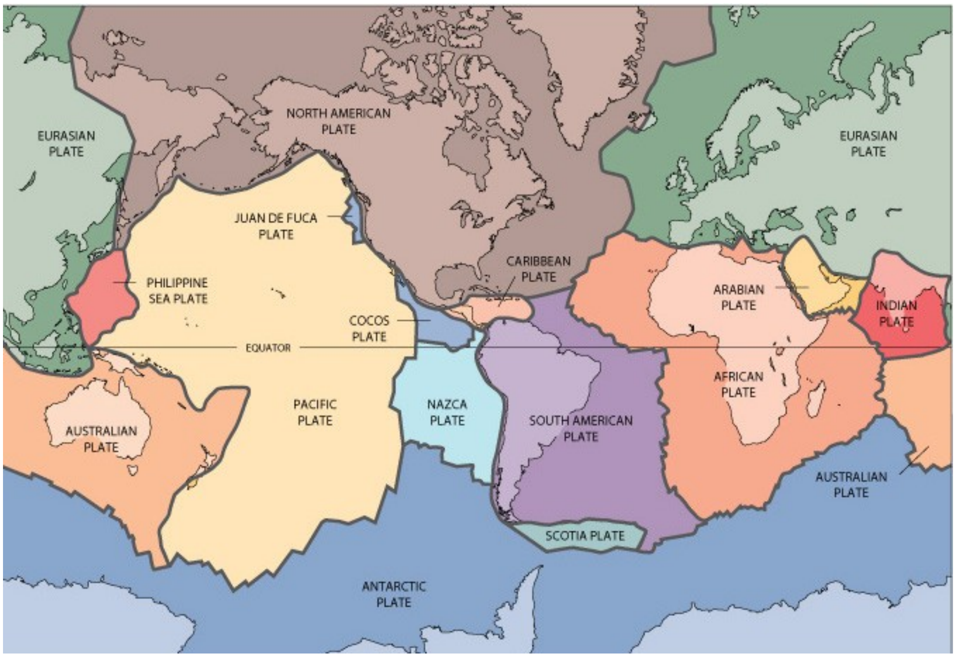


Figure 5.8: The Earth's tectonic plates from USGS[128]

The plates are essentially huge hunks of rock centered around very ancient cores called ***cratons***, most of which were formed during the Precambrian eons, perhaps as early as 4.3 Gya. They were certainly in existence by around 3.8 Gya, according to the age of sediments in western Greenland. Recent studies

126 U. S. Geological Survey, “This dynamic Earth”. http://pubs.usgs.gov/gip/dynamic/developing.html
127 U. S. Geological Survey, “This dynamic Earth”. http://pubs.usgs.gov/gip/dynamic/understanding.html
128 U. S. Geological Survey, “This dynamic Earth”. http://pubs.usgs.gov/gip/dynamic/slabs.html

suggest that plate tectonics began about 3 Gya.[129] The oceanic plates (or parts of plates) are thinner (less than 15km), but are composed mostly of very heavy basalt. The continental plates are thicker (up to 200km) but are composed mainly of granitic rocks, which are lighter. The plates essentially float on the asthenosphere, the next layer down, which is not a liquid, but can flow very slowly due to its high temperature.

It is currently thought that thermal convection due to the Earth's hot core causes the plates to move very slowly about. This movement has caused the configuration of land – the continents – to change over geological history.

When a plate of dense, oceanic rock collides edge-on with a plate of less dense continental rock, it is pushed down under the edge of the continental plate, a process called ***subduction*** (Figure 5.7). The descending ocean plate may create a deep trench at the boundary. Because of the heat from the rocks' rubbing together and because water vapor released by the oceanic crust lowers the melting point of the rocks in contact with it, much of the rock melts, creating ***magma***. Some of this rises through faults in the crust, creating a line of volcanoes along the edge of the continental plate above the subduction zone. The western coast of South America and much of that of North America are regions where the Pacific plate is being subducted under the American plates, forming the Andes and other mountains.

The Indian plate, which was once a separate continent south of Asia, moved northward and swung around to plow slowly into the southern edge of the Eurasian plate around 55 Mya. The impact between the two equally dense continental plates brought about no subduction, but instead pushed up the string of mountains known as the Himalayas. The process has not ended and the Himalayas are still growing.

Volcanic processes deep in the Earth below the ocean beds can break through the thinner oceanic crust, letting magma escape. This magma usually forms basalt on cooling, so the Earth's ocean floors are paved with dense basalt.

Right down the middle of the Atlantic Ocean, there is a ridge formed by magma pushed up by volcanic processes. As the magma flows out in east and west directions, perpendicularly to the ridge, the plates on either side grow at a rate of about 2.5 cm per year, so that their outer parts are pushed further east and west. In this way, the plates on either side of the Atlantic grow outwards from the mid-ocean ridge. In plate tectonics, this movement of two plates away from each other is called a ***divergent boundary***, or ***rift***.[130]

Running along the middle of the sub-oceanic chain, there is a trench where the magma comes out. Not all magma moves to the side immediately. It tends to pile up, so that the central line of activity is higher than the surrounding area, forming an underwater mountain chain, such as the one in the center of the Atlantic.

All the world's oceans have come about in such a rifting process. In the case of the Atlantic Ocean, the oceanic and continental crusts on each side form single lithographic plates, so the above-water parts of the continental plates of the Americas and of Africa and Eurasia seem to move apart[131]. Since the surface area of the Earth remains the same, plates which are added to in one place must be destroyed elsewhere. As the Atlantic gets wider, so must the Pacific get smaller. This illustrates the widening or disappearance of oceans.

Magma also may break through the ocean floor at a single spot, know as a ***hot spot***, bringing about the formation of a volcanic island above it. As the oceanic plate slides across the hot spot, a series of volcanoes may be formed, as is the case for the Hawaiian Islands or Icelandic volcanoes.

When molten magnetic rock solidifies, magnetic dipoles in it conserve the orientation of the Earth's magnetic field. When the field reverses direction, so does the magnetization recorded in the rocks. The symmetric distribution about the trench of the magnetized rocks points clearly to spreading of the sea floor in a context of reversals of magnetic field polarity (Figure 5.6).

Plates may converge on each other, as in subduction along the Pacific shore of South America, or diverge from each other, as is probably happening in the Great African Rift, or slide alongside each other, as in the San Andreas Fault in California. Sometimes, two zones of continental crust may converge, piling up mighty mountain chains, as has been the case with the Pyrenees, the Alps or the Appalachians and is still going on in the Himalayas.

129 “New study zeros in on plate tectonics' start date', http://cmns.umd.edu/news-events/features/3404.

130 There is still some disagreement as to whether the rifting is pushing the plates apart or whether the moving plates are pulling away at the rift.

131 The mechanism explaining this phenomenon is still a subject of discussion.

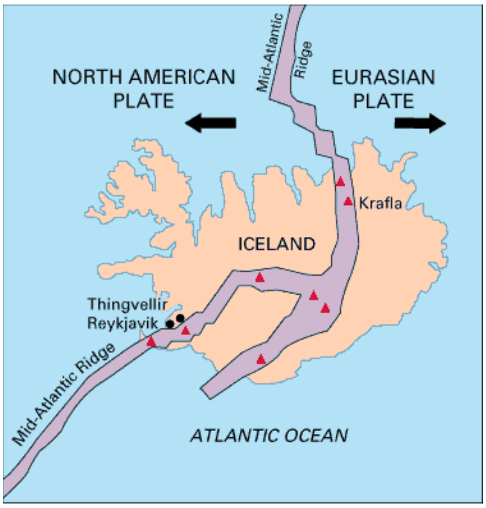


Figure 5.9: Iceland sits astride the joint of two plates, from USGS[132]

Figure 5.10: Looking out over the mid-Atlantic Ridge, the the rift between the American and European plates, at Þingvellir, Iceland. Photo by author.

Iceland is a fascinating geological showplace where the mid-ocean mountain range has been pushed up above the water level by a local hot spot. Over time, the country has moved relative to the hot spot, leaving volcanic craters along the path. The whole island country sits astride the rift between the North American and European plates.

Plate tectonics tells us that the continents we know have been in different configurations over the eons, sometimes separate, sometimes joined together into different large continents. At least once, they were joined together into a single supercontinent called ***Pangea***, about which more later. Such changes of tectonic configuration have an important influence on climate and the evolution of forms of life.

In our solar system, plate tectonics is unique to Earth; no continental plates are known on other planets. The Earth is lucky to have just the right mix of a solid surface and a flowing interior.

5.1.6. The Earth's atmosphere

Life on Earth depends on its atmosphere, a thin layer of gases, mostly nitrogen and oxygen, as well as some hydrogen and carbon and minute quantities of other elements. The atmosphere protects us and provides an environment suitable for life. It in turn depends on several other factors: the Earth's temperature, its size (and, therefore, gravity) and its distance from the sun. In fact, in the beginning of Earth's existence, the sun was not so bright, so temperatures were lower than one might expect.

What the atmosphere brings to the Earth

The atmosphere is important not only for its breathable gases but also for the protection it affords. It is essential in acting as a greenhouse gas to conserve the warmth from the sun. This effect can be overdone, as has been the case on Venus, which is much too hot to sustain life. The atmosphere also protects us from dangerous ultraviolet rays. And it shields us from countless meteorites, which are burned up from friction in the atmosphere before they can strike the ground and make huge craters, as has occurred on the atmosphere-free planet Mercury or on the Moon.

What the Earth brings to the atmosphere

The atmosphere is held onto the Earth by the gravitational force of the Earth itself. Earth's mass and

132 “This dynamic Earth”, USGS at http://pubs.usgs.gov/gip/dynamic/understanding.html

therefore its gravity is just right to hold onto the heavier gases nitrogen and oxygen that we breathe. In contrast, Mercury's gravity is too weak to hold on to much of anything, so it has lost all its atmosphere. Jupiter, which is much larger than Earth, retains light gases as well, but packs them all down into concentrations such that its atmosphere is enormously dense. Tremendous storms rage across its surface, the most obvious one being the famous Red Spot.

The Earth's size and its distance from the sun are delicately balanced. If the Earth were closer to the sun and therefore hotter, or if it were smaller and had less gravity, the movement of the atmosphere's molecules would be great enough to allow them to escape the Earth's gravitational attraction. Similarly, the relation between the production of heat by Earth's core and the size of the Earth is crucial to maintenance of a temperature conducive to life. The temperature of the Earth's surface also depends on its distance from the sun (closer would mean hotter) and its rotation (toasting each side evenly). The planet Venus is much closer to the Sun and its surface is far too hot for any sort of life to exist.

The Earth's temperature and gravity guarantee conditions necessary to the existence of water in three phases – gas, liquid and solid – which are the source of all Earth's weather and the rain cycles which irrigate the land. On Titan, one of Jupiter's moons, analogous conditions hold for different reasons, namely that its atmospheric molecules are much cooler than Earth's, so its lesser gravity is adequate to retain them. But the greatly reduced temperature relative to Earth's means that the phases there are of methane, not oxygen. All the water on Titan is frozen hard as steel.

All these things – the hot magnetic core, the atmosphere, the distance from the sun, the size and the orientation of the Earth – work together to make life possible on our planet. Without any one of them, things would be a lot different here. The Earth – and we, ourselves – are privileged to be in a set of conditions propitious for life.

5.1.7. Milankovitch cycles and climate

Two major influences on the evolution not only of life on Earth but of its geology and climate are sunshine and carbon.

The effect of sunshine on the Earth depends on several parameters.

The ***eccentricity*** of the Earth's elliptic orbit is a measure of how much the orbit deviates from a circle. The gravitational attraction of the largest planet, Jupiter, causes the shape of the Earth's orbit to vary so that at some times the annual maximum distance of the Earth from the Sun is greater than at others. This variation affects the magnitude of seasonal variations and influences the rise and fall of glaciation. There are several components with different periods which combine into an overall period on the order of 100,000 years, which matches the approximately 100,000-year cycles of glaciation observed since the beginning of the Quaternary (Figure.6.4). The eccentricity can vary from 0.005 (nearly circular) to 0.06. Currently, it is at a modest 0.0167, which corresponds to only a 3% variation in the distance from the Earth to the Sun and a fairly limited seasonal climatic variation from this effect.

The angle, called the ***obliquity*** or tilt, of the Earth's rotational axis relative to the plane of its orbit, is the source of the seasons. It varies by about 2.5 degrees over a period of about 41,000 years and this effect, like the varying eccentricity, modifies the intensity of seasonal variation. Increased tilt leads to more ***insolation*** (rate of incident sunlight energy per unit area) in summer and less in winter, but in a non-uniform way. The currently decreasing tilt should bring about milder seasonal variation, but could perhaps lead to another ice age because of less overall insolation in summer and at higher latitudes where glaciers form.

The rotational orbit of the earth changes slowly with respect to the "fixed" stars, wobbling like a top which is slowing down. A complete cycle has a period of about 26,000 years. This variation is called ***axial precession*** (or just precession, or wobble). It is due mainly to the gravitational pull (tidal forces) of the Moon and the Sun on the Earth's equatorial bulges. One result is the change in which star is the true North Star, which is only temporarily Polaris.

Taking all these factors and others into account leads to the calculation of so-called ***Milankovitch climate cycles***, which agree with some temperature-variation results from geology, notably the 100,000-year cycle, but leave other observations unexplained. In the 1920s, Milankovitch did his computing by hand. Today, computer models incorporating these cycles, continental configurations, atmospheric composition (CO_2) and many other factors are used to predict the future of global climate. Partly because they employ such long periods of time, but also because they are clearly not the only factors influencing climate on the Earth, Milankovitch cycles are incapable of explaining the current period of rapid warming since the mid-19th century. This is one reason to suspect human beings to be responsible for it.[133]

5.1.8. The carbon cycle

Carbon is not only the stuff of life but a major regulator of temperature, and so, climate. Its change of form and presence make up the ***carbon cycle***.[134] It is a good deal more complex than the water cycle already presented. In fact, there are several cycles and they interact one with another.

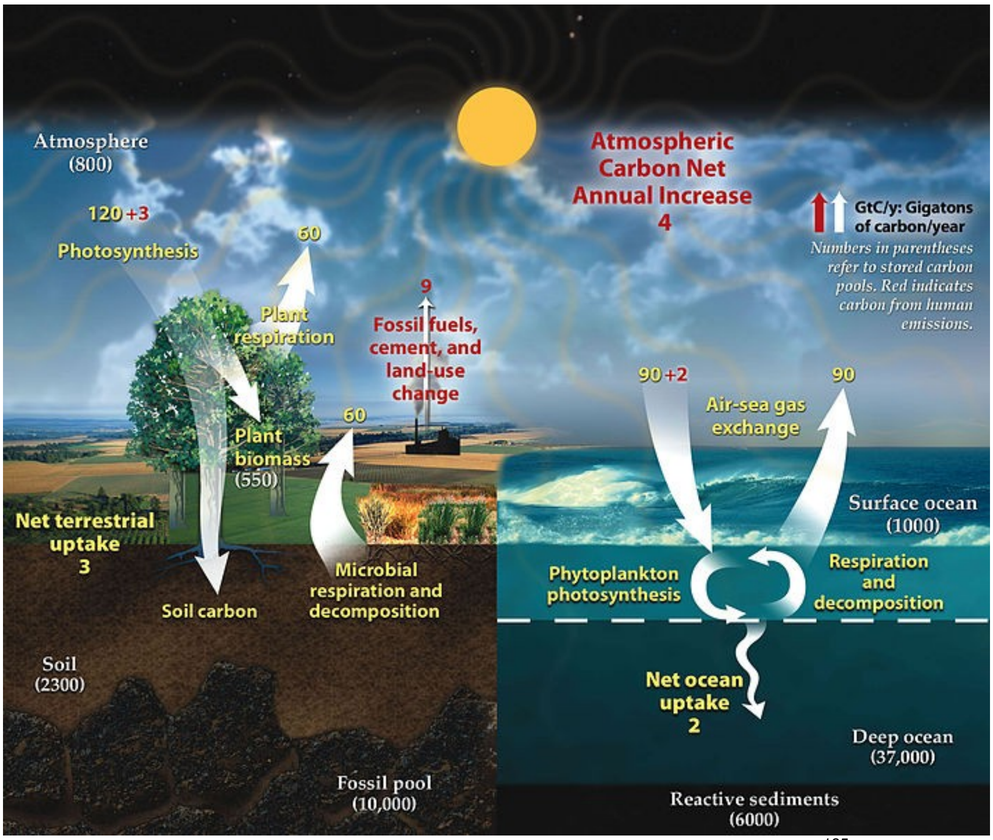


Figure 5.11. The carbon cycle, from Wikimedia Commons[135]

Two major atmospheric components contain carbon: carbon dioxide (CO_2) and methane (CH_4). Both are greenhouse gases, methane being the more powerful by far. Carbon and global temperatures have important influence one on the other.

Geologists consider that carbon is exchanged among several types of carbon reservoirs:

- the atmosphere;
- the terrestrial biosphere;
- oceans;
- soil and sediments;
- Earth's interior (mantle and crust).

The carbon cycle consists of a number of processes occurring in these reservoirs and exchanging carbon among them. Several such processes take place on land:

133 “Why Milankovitch cycles can’t explain Earth’s current warming”. https://climate.nasa.gov/blog/2949/why-milankovitch-cycles-cant-explain-earths-current-warming/

134 Principal source: MacDougall 2011, 170-173.

135 Carbon cycle, https://commons.wikimedia.org/wiki/File:Carbon_cycle.jpg

- Photosynthesis absorbs atmospheric CO_2 and converts it into a reduced form as organic carbon in the biosphere (ejecting oxygen into atmosphere).[136]
- Plant death releases carbon either to the atmosphere as CO_2 or to the soil or sediment.
- Carbon in oxygen-poor soil may be transformed by bacteria and released to the atmosphere as methane, or it may be stored as sediments, peat or coal.

Processes of the carbon cycle in the sea include the following:

- Photosynthesizing plankton near the surface store carbon from the atmosphere as organic carbon.
- Plankton also use dissolved carbon to make shells ($CaCO_3$), thereby storing carbon.
- When the plankton die, shells and organic carbon sink, removing carbon from the upper sea. Water in the surface layer is then relatively depleted in CO_2 and so dissolves more from the air. Some deep-sea carbon dissolves, while some sinks as sediment. This step constitutes a "biological pump", pumping carbon from the atmosphere, through the upper sea to the lower.

The sequence

$$CO_2 \rightarrow \text{photosynthesis} \rightarrow O_2 \rightarrow \text{respiration} \rightarrow CO_2$$

constitutes the ***short-term carbon cycle***. Plants may “store” carbon for years or decades, whereas oceans store it for much longer periods, perhaps thousands of years.

In the atmosphere, CO_2 content is in a delicate balance which can be influenced by external phenomena. For instance, an increase in seawater temperature causes more CO_2 to be released into the atmosphere; decreased temperature enhances absorption.

Over the ***long term*** (measured in millions of years), volcanic activity adds CO_2 to the atmosphere and the resulting greenhouse effect increases atmospheric temperature. CO_2 dissolves in rainwater to make carbonic acid (H_2CO_3). Both the acid and the increase in temperature cause increased ***chemical weathering*** of rocks and this removes CO_2 from the atmosphere.[137] The rock particles are washed down to the sea and laid down as sediment which may later be subducted and eventually ejected by volcanoes. The cycle then repeats. Chemical weathering of rocks following the formation of the Himalayas is thought to have contributed to global cooling in the past.

So a long-term equilibrium between volcanic activity and chemical weathering (and other factors) maintains a balance of atmospheric CO_2 over long periods of time.

Human-caused emissions (fossil-fuel burning, livestock, etc) release CO_2 and other greenhouse gases into the atmosphere, leading to rises in global temperatures.

5.1.9. How do we know all this?

Geological clues come from different sources.

- Volcanic and other magma activity, along with measurement of seismic waves (which may be artificially induced) can show the configuration of the Earth's mantle.
- Various rock formations, folds and magnetic properties give evidence of the action of plate tectonics. Evidence for plate tectonics also comes from fossils in the plates, especially along edges of plates which were formerly contiguous. The mid-oceanic trenches have been measured by sonar-equipped ships and by satellite.
- Quantities as diverse as the distribution of elements (e.g., iridium) and isotopes (e.g, oxygen) can give evidence about oxygen concentration or water temperatures.
- Measurement of remaining quantities of radioactive isotopes of certain elements with known decay rates can be used to determine the time which has passed since the element was incorporated into an object, rock or fossil. This method has been used in combination with fossil dating to determine the sequence of geological events.

136 See Part II, Chapter 7, Oxidation-reduction and electron carriers.
137 For more details, see Ward and Kirschvink, 22-23.

- Fossils show the evolution of living beings and can be used to establish relative dates of the rocks in which they are found.

Now on to the first eon.

5.2. The Hadean Eon – early Earth

The Earth's first 700 million years are referred to as the ***Hadean Eon***. It was during this eon that the Earth's constituent rocks formed and differentiated and separated into layers, preparing the Earth for the onset of plate tectonics around or just after the end of the eon. A giant impact with a small planet led to the formation of the Moon. And water appeared on the Earth.

5.2.1. Differentiation of the Earth's minerals

Before its orbital path was swept clean, the Earth was bombarded by rocks circulating through space. These impacts transferred a huge amount of kinetic energy to the young planet and this was transformed into thermal energy, to which was added more heat generated by radioactive decay of short-lifetime elements. The surface of the young Earth was so hot that rocks melted to magma. It was indeed like Hell, as the name Hadean suggests.

The elements forming the Earth had been produced in exploding stars. At first, these different sorts of matter were more or less homogeneously mixed, with one part of the Earth similar to another. Then, because elements like iron and nickel are denser than the others, they were attracted to each other by gravity, which eventually pulled them down through the Earth's other matter, falling like sand through water and coalescing in the center of the new planet. Lighter material such as silicon remained closer to the surface. This process of ***differentiation*** occurred during several tens of millions of years. It left Earth with a hot iron-rich core, whose existence would be essential to its subsequent history. The temperature of the core is due both to the tremendous pressure of the outer minerals and also to heat generated by radioactive decay of core isotopes.

Another result of this differentiation was the formation of up to three hundred different types of minerals.[138] The principal elements in the mantle are as follows:

Element	Approx. percentage by weight[139]
Oxygen (O)	46.6
Silicon (Si)	27.7
Aluminum (Al)	8.1
Iron (Fe)	5.0
Calcium (Ca)	3.6
Sodium (Na)	2.8
Potassium (K)	2.6
Magnesium (Mg)	2.1
Others	1.5

Table 2: Principle elements of Earth's mantle

Volcanoes spewed out lava, rocks and gases, thus establishing the early Earth's atmosphere, with some gas perhaps arriving on rocks from space. Its composition is not known precisely, but it probably contained water, methane, ammonia, carbon dioxide, sulfur dioxide, hydrogen and nitrogen. Carbon dioxide was present in relatively large amounts, as on Mars and Venus. When it later dissolved in the oceans, it would combine with calcium to make calcium carbonate which precipitated out to form limestone, thereby reducing the quantity of CO_2 in the atmosphere.[140] But it seems that there was almost no separate oxygen at all for

138 Hazen, 19.

139 Source: http://www.indiana.edu/~geol105/1425chap5.htm.

140 MacDougal, 1998 27.

about two billion years.

5.2.2. Formation of the Moon – the Giant-impact Hypothesis[141]

One collision was especially violent. Most scientists now accept the ***Giant-Impact Hypothesis***, according to which a proto-planet about the size of Mars, now referred to as ***Theia***[142], crashed into the Earth – whether obliquely or directly is not certain – sometime between 30 and 100 My after Earth's formation. The collision was so violent that the two bodies merged together, sharing their matter. The greater gravitational field of the Earth captured most of the iron and added it to its core. The high temperatures generated by the collision caused a significant portion of the total mass to be vaporized from the combined mantle and thrown off into a disk of matter orbiting the Earth. The ejected particles collided and accreted and became, probably, two orbiting moonlets, before merging about 5 My later to form the Moon. Such a scenario would explain the differing compositions of the near and far sides of the Moon.[143]

The hypothesis explains several observations. Aside from Earth's iron and volatiles, it and the Moon have similar composition because Earth and Theia did, indicating that they were formed close to each other. Light, volatile elements like hydrogen nitrogen, carbon and sulfur, were dissipated by heat from the fragments which formed the Moon. The giant-impact hypothesis explains the similar ratios of oxygen isotopes on Earth and Moon. It also explains how the Moon's rotation in its orbit keeps the same side toward the Earth. The impact has also been evoked to explain the inclination of the Earth's rotational axis relative to its orbital plane, caused when the impact tipped over the Earth's axis.

Originally, the Moon was much closer to the Earth than now. The Earth's orbital rotation was much faster, so a day was only five hours long, but the Moon took 84 hours to go around the Earth. Gravitational tides resulting from the pull of the two bodies on each other deformed their surfaces, stirring them into oceans of magma. The process also brought about a slowing of the rotation of the Earth and an increase in distance between the two bodies. In this way, the Earth's rotational angular momentum decreased as the Moon's orbital angular momentum increased, thus conserving total angular momentum. Some of Earth's angular momentum was effectively transferred to the Moon. Both ancient tidal sediments and coral layers indicate that an Earth day was significantly shorter hundreds of millions of years ago. Mirrors left on the Moon by astronauts have enabled the observation that the Moon still is slowly receding. The giant-impact hypothesis seems to offer the best explanation for the Moon's formation, but study continues.

As the Earth-Moon distance increased, their surfaces cooled and eventually hardened. Later, around 4 Gya, in a sequence of events known as the ***Late Heavy Bombardment***, the Moon was struck by a huge number of asteroids, leaving it pock-marked with the craters we see today. Such asteroids probably also struck the Earth, destroying its mantle and oceans. This would explain why the oldest rocks found on Earth date only to 3.85 Gya.[144]

5.2.3. Continued differentiation and formation of minerals

After the Earth-Theia impact, the surface of the Earth again was molten. As it slowly cooled, crystals of ***olivine*** (magnesium silicate) formed. Being heavier than the surrounding magma, they sank into it and continued growing into larger crystals, eventually forming the green rock ***dunite***. Then pyroxene, a chain silicate[145], formed and mixed with the olivine near the Earth's surface to constitute greenish-black ***peridotite***, starting about 4.5 Gya and continuing for hundreds of millions of years. As the peridotite hardened, it formed the Earth's first, temporary crust. But as it became denser, it cracked and broke and sank, pushing more magma to the surface to cool and continue the process as the mantle too slowly solidified.[146] Magnesium-silicate peridotite came to constitute much of the Earth's mantle.

As peridotite was subjected to heat and pressure in the mantle, it began to melt. Not all of it melted at once,

141 This section is based principally on chapter 2 of Hazen, *The story of Earth*. He refers to the event as the "Big thwack".

142 In Greek mythology, Theia was a Titaness and was mother of Helios (the Sun), Selene (the Moon) and Eos (the Dawn).

143 Levett, Richard. "Early Earth may have had two moons." Nature. http://www.nature.com/news/2011/110803/full/news.2011.456.html#B1

144 Marshak, 435.

145 We will see what that means in the next paragraphs.

146 Details from Hazen, *op. cit.*, chapter 3.

so the first fraction to melt had a different constitution and contained much less magnesium and more of other elements. The result was a lighter magma which in turn rose to the surface and accumulated in cracks and pockets and finally exited through volcanoes or rifts as ***basalt***. Since basalt is lighter than the rest of the mantle, it piled up into volcanic mountains. Basalt exists in many different forms, but two minerals are essential: plagioclase feldspar and the same pyroxene that exists in peridotite.

Although peridotite is still the principal component of the Earth's upper mantle, it is rare on the surface of the Earth. Basalt has remained and continues to be produced, as in Fig. 5.12.

Figure 5.12: Black basalt field at the Krafla caldera in Iceland, photo by author

Figure 5.12, with its black basalt "sea" and volcanic cone of lighter rock (rhyolite), gives an idea of how Earth may once have looked. But this Icelandic landscape is the result of volcanic activity known as the "Krafla Fires" which occurred from 1975 to 1984. The Earth still has some seismic tricks in its mantle.

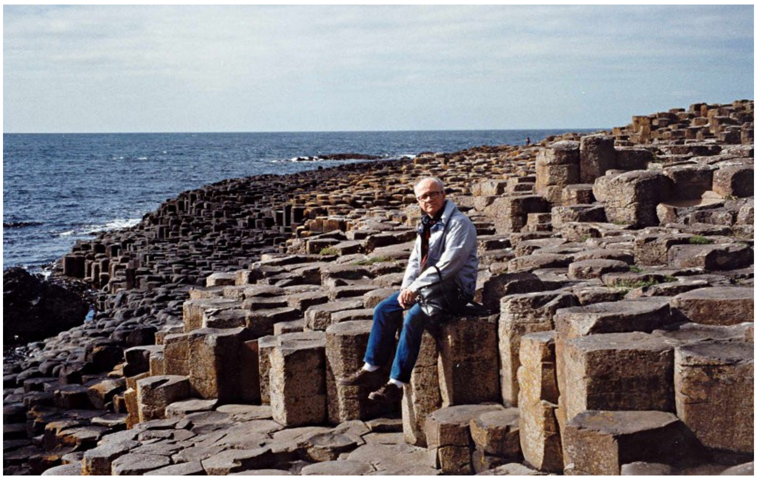


Figure 5.13: Basalt columns at Devil's Causeway, North Ireland. Photo of author by Siv O'Neall.

On cooling, thick basaltic lava flows may crack so as to form vertical columns, often hexagonal in cross section, as in Figure 5.13.

5.2.4. Formation of oceans and first continents

Volcanoes had been spewing forth water vapor from the Earth's interior for a long time before the Earth-Thea collision blew it away. But afterwards, the process started over and soon there was liquid water on Earth again.

Zircon, or zirconium silicate, is an extremely hard mineral, resistant to alteration, and often is used as a gemstone. It can last for a very long time, it may contain uranium – which means it can be dated – and it contains oxygen. High proportions of the heavy-oxygen isotope, ^{18}O, indicate that crystals were probably formed in water. Such is the case for the world's oldest known zircon, dated to 4.4 Gya, which leads many scientists to consider that there was already water on the Earth at that time. Continents had not yet formed, but convection in the mantle gave rise to volcanic heat rising in certain places, hot spots, which ejected heat and matter through volcanoes at the surface. Those convection currents were not circulating air, but circulating rock, and on a huge scale. The Earth of 4.4 Gya would have been covered with an ocean studded with such volcanic islands. Some scientists disagree with this scenario.

Evidence of sediments from at least 4 Gya indicates clearly that large bodies of water existed by then. It was water which brought about the next major step in differentiation of the Earth's minerals. As contradictory as that may seem, basalt heated by underlying magma melts at a lower temperature when it is cooled by water.[147] Just as the sinking of peridotite isolated basalt at the surface, so did melting basalt isolate silicon in turn. Minerals with high silicon content were even lighter than basalt and so floated to the surface to form the first granite. Granite is composed of four constituents: quartz, two different feldspars and either pyroxene or mica.

As more basalt sank into the mantle and more, lighter granite rose to the surface, the increasing amounts of granite coalesced into large islands. The islands were perhaps sometimes fused together by heat from asteroid collisions and grew into floating plates. Any remaining basalt was lower than the granite plates and so found itself on the floor of the ocean where it is predominantly found still today. By 3 Gya, plate tectonics had started.

So we have seen the following steps in differentiation of the Earth's minerals:

- Settling of heavy iron to the core leaves lighter elements behind in the mantle.
- Melting magma forms olivine which forms dunite and sinks into the mantle.
- Remaining olivine combines with pyroxene to form peridotite, which forms the Earth's first, temporary crust before it too cracks and sinks into the mantle.
- Melting peridotite combines with feldspar to produce basalt, which floats to the surface and forms the next layer of crust. Basalt is still being formed along, for instance, the mid-Atlantic ridge.
- Melting basalt sinks beneath silicon-rich granite, which floats on the basalt mantle.

The crust is still formed of basalt at its lower levels, beneath the oceans, and granite higher up on the land.

5.2.5. Age of the Earth

Because of the Earth's continued volcanic activity, no rocks from the Earth's beginnings exist. Radioactive dating of the oldest rocks gives an age greater than 3.8 Gy.[148] Zircon crystals from western Australia have been found with ages up to 4.4 Gy.

The fact that all the planets and their moons not only follow their orbits but rotate in the same direction strongly suggests their all having the same origins. Meteorites which would thus have been formed at the same time have been dated and all were formed around 4.5 Gya. Meteorites called ***chondrites*** have the same chemical composition as the Sun and are formed of collections of grains which indicate they have never been melted. So they are considered to have the same composition as the Earth before differentiation. They should also be only slightly older than the Earth. Different methods of radioactive dating indicate that these meteorites are at least 4.6 Gy of age.

147 This is like salt melting ice on your driveway in the winter.

148 The principal source for all these figures is an excellent review paper, "The scientific age of the Earth", on-line at http://www.talkorigins.org/faqs/dalrymple/scientific_age_earth.html. Another excellent source is the US Geological Survey at

Rocks brought back from NASA's Apollo mission to the Moon have been dated to between 4.5 and 4.6 Gya.

The best measurements of the Earth's age come from assuming that isotopes of lead (Pb) were in equal ratios in the solar nebula and therefore in planetesimals and meteorites at the time of the planets' formation. From this, a calculation based on various rocks and meteorites results in

Age of Earth: 4.54 Gy ± 1%.[149]

5.3. The Archean Eon – the appearance of life forms

The period from 3.8 to 2.5 Gya is referred to as the Archean Eon.[150] During the Archean, plate tectonics got under way. Life began. Stress that: Life began. The famous ozone layer formed, protecting the planet from UV radiation, and bacterial photosynthesis started adding oxygen to the atmosphere, fundamentally changing the environment of the Earth's surface.

5.3.1. Geology and atmosphere

Over the period of about 3.2-2.7 Gya, rocks on the surface came together to form the first ***cratons***, which would become the central cores of continental plates. Sediments from the eon, which indicate that the rock cycle (volcanism-sedimentation-metamorphism) was in action, provide evidence for the existence of continents and oceans. The oldest existing continental rocks date from the Archean at about 4 Gya.[151] By the end of the Archean, plate tectonics was under way.

Solar energy received at the surface of the Earth was about 20 to 25 % lower than at present, which could have made the planet too cold for life to be established[152], but the CO_2 retained heat beneath the atmospheric layer, causing a greenhouse effect which slowly raised atmospheric temperatures. Sunlight striking the water vapor caused photochemical dissociation, the breaking up of the water molecules and the bonding together of the resulting oxygen atoms to create ozone, or O_3. In time, the ozone came to protect the surface of the Earth from ultraviolet radiation from the sun. At the same time, it prevented further chemical dissociation, which therefore has not played an important role in the oxygenation of the atmosphere. Further increase of atmospheric oxygen had to wait for photosynthesis, as described below.

5.3.2. Life and atmosphere

Arguments over what was the earliest form of life (and who discovered it) probably are not over yet. Currently, the oldest fossils would be of bacteria from Australia, dating from 3.4 Gya.[153] What makes them interesting is the fact that their metabolism was based on sulfur rather than oxygen, which was not yet common in the atmosphere.

For comparison, the oldest fossil evidence for cyanobacteria dates from 2.22 Gya[154]; for eukaryotes, 1.78-1.68 Gya.[155]

Life may be defined as "… a self-sustaining chemical system capable of incorporating novelty and undergoing Darwinian evolution."[156] There are three main hypotheses to explain its origin on Earth.

The "primordial soup" hypothesis considers life to have been brought about using energy from electricity (lightning) in a mixture of gases including water and methane. Such production of organic molecules has

149 U.S. Geological Survey, "Age of the Earth", http://pubs.usgs.gov/gip/geotime/age.html.

150 The International Commission on Stratigraphy, apparently the expert in these matters, places the beginning of the Achean at 4..0 Gya, but I have no book which says other than 3.8.

151 Spooner, 255.

152 "Climate puzzle over origins of life on Earth", http://www.manchester.ac.uk/discover/news/article/?id=10798.Most of these dates come from Benton.

153 Microfossils of sulphur-metabolizing cells in 3.4-billion-year-old rocks of Western Australia, Nature Geoscience https://www.nature.com/articles/ngeo1238.

154 Ward and Kirschvink, 81.

155 j. Brocks, cited by Ward and Kirschvink, 75.

156 Gerald Joyce, NASA, quoted by Hazen, 130.

been demonstrated in the laboratory, but fails to convince many scientists because it depends greatly on the composition of the atmosphere at the time. In particular, it is now thought that CO_2 was far more prevalent than methane. More recent experiments with different mixtures have produced similar organic compounds. The discovery of amino acids on meteorites adds weight to the hypothesis that varying atmospheric conditions could lead to production of organic molecules.[157]

The hydrothermal- vent hypothesis exists in two varieties. The first supposes that life came into being in “black smokers”, hydrothermal vents formed along undersea ridges such as the Mid-Atlantic Ridge. Sea water leaks down through fissures in the rock and is super-heated by magma. The super-heated water may attain a temperature of 400°C, but the immense pressure keeps it from boiling. When the mineral-laden water rises and hits the relatively colder sea water, dissolved minerals are liberated, emitting sulfur-bearing black molecules which look like, but are not, smoke and which pile up to produce “chimneys”. White smokers, which carry barium, calcium and silicon, also exist. It was thought that the reaction of hydrogen sulfide from the vent with water would provide the energy necessary for the formation of life. Although life does abound in these vents, it is not at all like ours. One finds, for instance, extremophile organisms which live in darkness and obtain their nourishment from hydrogen sulfide.

The second hydrothermal-vent model proposes that life originated in ***alkaline hydrothermal vents***. In places on the ocean floor, peridotite rock, which is normally found deep in the Earth's mantle, has been pushed up to within several kilometers from the surface by faulting. The rock contains olivine, a Mg-Fe silicate, which reacts with sea water which has filtered down through the rock and forms the minerals serpentine and magnetite. In this process, called ***serpentinization***, the ferrous iron (Fe^{2+}) in olivine is oxidized to ferric iron (Fe^{3+}), releasing hydrogen gas and heat, as well as reducing gases methane and hydrogen sulfide.[158] The heat energy is generated by chemistry and does not come from hot magma, as is the case with “black smokers”.

Figure 5.14: “Nature Tower”, an alkaline “chimney” in the Lost City group.[159]

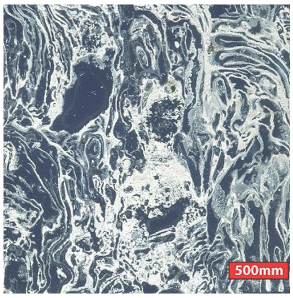


Figure 5.15: Whorls and pores in a thin section of a Lost City chimney.[160]

The result is an alkaline solution (pH = 9-11) whose maximum temperature is 100°C, rich in calcium and H_2. On contact with colder sea water, the calcium precipitates out, forming white structures like chimneys.[161] Rising fluids are very alkaline (basic) and thereby precipitate out more calcium carbonate and other alkaline substances when they hit the cold sea water and these build up on the pile of rock already started. Soon “reverse stalactites” are produced by the carbonate left behind by the thermally rising water. Eventually, small cracks and “cells” form within the rock.

So the rising “chimneys”, which may reach many meters in height, are associated with a source of energy, gases like those in the “primordial soup”, and small cell-sized alveoli or compartments. Such an environment

157 Prothero (2007), 147-52.

158 “The Lost City 2005 expedition”, NOAA, http://oceanexplorer.noaa.gov/explorations/05lostcity/background/serp/serpentinization.html.

159 From National Oceanic and Atmospheric Administration (NOAA), http://oceanexplorer.noaa.gov/explorations/05lostcity/logs/july31/july31.html.

160 Image from NOAA, http://oceanexplorer.noaa.gov/explorations/05lostcity/background/chimney/media/thinsection1.html.

161 These results are for the Lost City Hydrothermal vents. Results differ some for other vents.

may well be suited to abiotic hydrocarbon production.[162] The compartments contain and protect their contents as well as ensuring their concentration, making excellent conditions for the production of inorganic precursors to organic life. From these, prokaryotes and archea could have evolved independently around 3.8 Gya and eukaryotes later, about 1.8 Gya.

A third hypothesis for the origin of life, the ***volcanic-pool hypothesis***, is more recent.[163] It posits the combination of simple molecular building blocks, perhaps from space, using thermal energy from volcanic pools, like those at Yellowstone or in Iceland. As external conditions change, they could evolve in a Darwinian manner as they survive through wet, dry and moist cycles in land-based hot springs. Such organisms have been called ***progenotes***.[164]

In any case, the appearance of cell membranes meant that different environments and molecules could be separated from each other, a kind of biological differentiation. This led in turn to the the formation of simple cells, called ***prokaryotic*** cells. As we have already seen, the earliest clear occurrence of life is in the form of microscopic cells in Archean sediments in Australia, dating from about 3.4 Gya. ***Cyanobacteria*** existed by 2.22 Gya and are still alive in many places on Earth today. The "cyan" in their name refers to their blue-green color. They are the oldest currently-living beings. But they are extremely important for another reason.

Some of these bacteria mixed with sand to make microbial mats. As the sandy mixture became muddy, the cyanobacteria migrated upwards and the process repeated, resulting in lumpy layers of colonies called ***stromatolites***. Stromatolites thrived over the period from about 3.5 Gya to 0.5 Gya, but are still found in a few places such as Shark Bay, Australia, or the Pacific Coast of Baja California. They survive only in especially salty water (twice the sea's normal saltiness) or in places with especially strong currents, as both conditions limit predators such as snails which otherwise would devour them.

Figure 5.16: Stromatolites in limestone near Saratoga Springs, NY, by M. C. Ryget via Wikimedia Commons[165]

Figure 5.17: Living stromatolites in Shark Bay, Australia, by Paul Harrison via Wikimedia Commons[166]

Cyanobacteria have been called the "working-class heroes of the Precambrian Earth"[167] and were fundamental to the development of life. The importance of these organisms cannot be stressed too much, as they were the first organisms to carry out ***photosynthesis***, the use of energy from the sun to convert carbon

162 https://www.researchgate.net/profile/Marvin_Lilley/publication/5613067_Abiogenic_hydrocarbon_production_at_lost_city_hydrothermal_field/links/0c960520e90a17f539000000.pdf

163 Van Kranendonk, Martin J., Deamer, David, and Djokic, Tara, "Life springs", Scientific American, August 2017, 22.

164 There is some disagreement as to whether progenote simply means LUCA (last universal common ancestor). Others, more specifically, call it '"a theoretical construct, an entity that, by definition, has a rudimentary, imprecise linkage between its genotype and phenotype (Woese, 1987)"—a creature still experiencing progressive Darwinian evolution, in other words.' From https://www.ncbi.nlm.nih.gov/books/NBK232215/.

165 https://commons.wikimedia.org/wiki/File:Stromatolites_hoyt_mcr1.JPG

166 https://commons.wikimedia.org/wiki/File:Stromatolites_in_Sharkbay.jpg

167 Knoll (2003), 42.

dioxide into nutrients and free oxygen, which is returned to the atmosphere.[168] Over hundreds of millions of years during the Archean and Proterozoic Eons, as cyanobacteria used photosynthesis to recover the energy necessary for their own metabolism, they brought about the gradual transformation of atmospheric CO_2 into the oxygen necessary for other forms of life.[169] At the same time, the greenhouse effect was reduced and this caused global cooling. Much CO_2 was also dissolved in the seas, where it combined with calcium to form calcium carbonate, which in turn solidified to form limestone. Limestone, ocean water and corals are huge stores of carbon dioxide (***carbon sequestration***).

Interestingly, thousands of the minerals found on Earth today are due to oxidation by oxygen dissolved in water. So oxygen has not only allowed life to begin, but has also thoroughly changed our mineral environment. The biosphere and the geosphere have evolved together.

Photosynthesis took place in the top layer of stromatolites and each layer lived off the layer above. As such, they represented an early symbiosis or way of living together – an example of what we now call ecology. Notice that ecology (water, atmosphere) led to biology (stromatolites), which in turn influenced ecology (atmospheric oxygen).

5.4. The Proterozoic Eon – the dance of the continents

The Proterozoic Eon ran from 2.5 Gy to 542 Mya. It has been so named because of the appearance of more complex organisms during this time. Continents formed and unformed and formed again or differently. Increased atmospheric oxygen modified the mineral composition of the Earth's crust and, especially, allowed oxygen-based life to flourish. Eukaryotes came into being and, at the end of the eon, the first multi-celled organisms were born.

5.4.1. Geology and atmosphere

During this eon, plate tectonics came into its own, with cratons moving about on the surface of the Earth in what has been called a "stately dance", i.e., a slow one. It was like some kind of round, with one continent dancing for a while with another, then separately, then with a third. At least five times, they all came together to form a single supercontinent. As they slowly smashed into each other, they pushed up mountains, a process geologists call ***orogeny***. As they rifted and came apart, seas formed between them.

There is rather weak evidence for a perhaps small-continent-sized landmass dubbed ***Vaalbara*** 3.3 Gya. It is much more certain that there existed a continent-sized landmass called ***Ur*** about 3.1 Gya, made up of cratons from what now are South Africa, Australia, India and Madagascar. Ur was not a supercontinent, but initially it's about all there was. Ur lasted for about 300 Gy, undergoing various combinations with other continents, until the breakup of Pangea.

It should be remembered that these reconstructions of ancient cratons or continents from geological and other data are to varying extents uncertain as to the details.

About 2.7 Gya, the first supercontinent came into being – ***Kenorland*** (also called Superia). Since the atmosphere was essentially devoid of oxygen at the time, only acid rain fell, and this eroded and dissolved the land, leading to the deposit of sediments along the continent's coasts. About 2.4 Gya, just as oxygen started accumulating in the atmosphere, Ur broke away and Kenorland began its fragmentation.

By 2 Gya, there were at least five separate cratons:

- the Laurentian supercraton, the geological core of North America;
- Ur, composing current India, western Australia and South Africa;
- Baltica and Ukrainian cratons, making up eastern Europe;
- cratons comprising most of what are now South America, China and Africa.

By about 1.8 Gya, all these cratons had collided and coalesced to form the supercontinent ***Columbia***[170]. Columbia was situated on the equator, so its interior was hot and dry. There were therefore few or no ice

168 Photosynthesis will be described in chapter 8.
169 The capability of stromatolites to accomplish this task alone has been questioned and other mechanisms suggested.
170 Also called Nena, Nuna or Hudsonland.

caps and ocean levels were relatively high.

Around 1.6 Gya, Ur split off from Columbia and a new sea formed between them. They were still situated at the equator, so ice remained low and ocean levels high.

About 1.2 Gya, a new supercontinent now called ***Rodinia*** was forming, again near the equator, so its interior was again hot, dry and lifeless and no sediments were formed. Evidence for Rodinia comes from the so-called ***Grenville orogeny***, rocks of which are found in the cratons of all current continents. Also, the absence of sedimentary rocks from this period suggests an absence of shallow seas, which would have been the case if there was only one supercontinent. Rodinia now was surrounded by a single superocean called ***Mirovia***. Around 850 to 800 Mya, Rodina broke apart and then, somewhere around 700 Mya, it may have reformed with the pieces in a different order to form another supercontinent, ***Pannotia***, which "only" lasted about 60 million years before It broke up in turn.

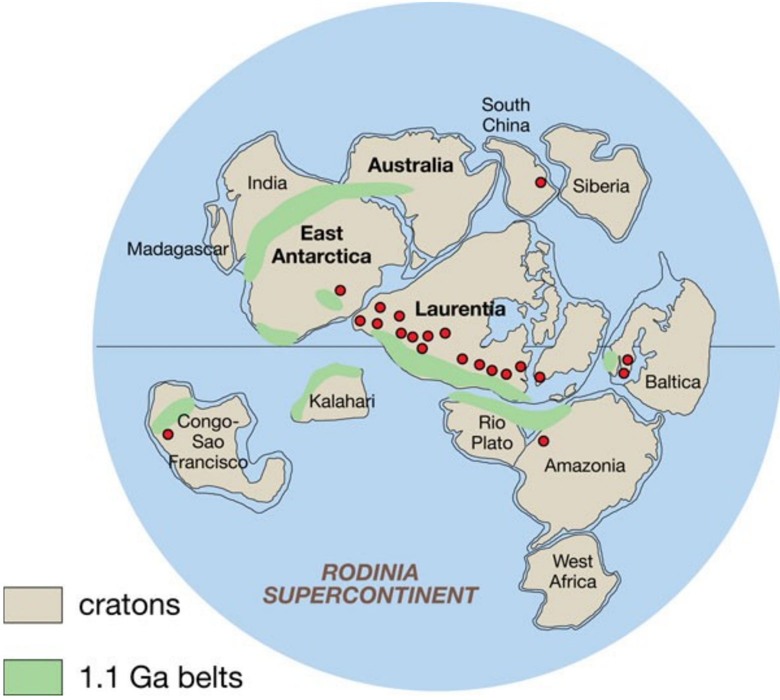


Figure 5.18: Reconstruction of the supercontinent Rodinia, by John Goodge [Public domain], via Wikimedia Commons[171]

The next supercontinent, ***Pangea***, formed only later, in the Phanerozoic Eon.

It was during the Proterozoic and the beginning of the Phanerozoic Eons that the oxygen content of the Earth's atmosphere began to increase significantly. Alternating layers of red, iron-containing minerals and silica minerals called ***banded iron formations*** (or ***BIF***s) indicate fluctuations in the oxygen levels of oceans about 2.5-1.8 Gya. Iron(II), or Fe^{2+}, is soluble in water, but is oxidized by atmospheric oxygen to iron(III), or Fe^{3+}, which precipitates. Since BIFs exist in sedimentary rocks, it is thought that fluctuating levels of oxygen in the sea water led to the bands of minerals of alternating colors.

Later formations called ***red beds***, which are sedimentary sandstone or shale, exist from 1,8 Gya. Their red color is due to the mineral hematite, Fe_2O_3, formed by the oxidation of iron, but this time on land. So by this time, the air must have contained enough oxygen to oxidize iron. Red beds are also common in rocks from the Phanerozoic Eon.

171 https://commons.wikimedia.org/wiki/File%3ARodinia_reconstruction.jpg

Figure 5.19: The Lal Qila, or Red Fort, in Delhi is built of red-bed sandstone. Photo by Siv O'Neall.

In summary:

- From 2.5-1.8 Gya, fluctuating oxygen content in seawater formed BIFs.
- Since 1.8 Gya, increasing atmospheric oxygen has oxidized Fe to hematite.

Atmospheric oxygen also allowed the formation of new types of minerals, so once more life influenced geology. So life has affected geology, as well as the opposite.

5.4.2. Life

Once the great oxidation event had taken place, life went through a long, slow period referred to as the ***boring billion***. This period nevertheless included the oxidation of the atmosphere and the evolution of eukaryotes – not exactly boring to us.

Evolution and the atmosphere

As shown by fossil evidence, stromatolites thrived in the Proterozoic and continued their conversion of atmospheric CO_2 into O_2. Several types of indirect evidence, based on the presence of certain molecules in rocks, indicate that around 2 Gya, the content of free oxygen in the atmosphere increased significantly. More recent data indicate that this increase began about 2.4 Gya in what is called the ***Great Oxidation Event*** (***GOE***).[172]

Some time after that, aerobic organisms evolved which removed oxygen from the air, eventually leading to the current equilibrium. In spite of evidence for important fluctuations in oxygen levels over the millenia since the GOE, the average oxygen content of the atmosphere has been increasing for the last two billion years. It is now at about 21%, compared to less than 1% at the beginning of the Proterozoic. (See Figure 5.20.)

Scientists wonder why evolution was so slow during the Boring Billion. One hypothesis holds that there was competition during this time between bacteria which did ***oxygenic photosynthesis*** and others which were based on sulfur and used H_2S rather than H_2O as a source of electrons. This could have been due to conditions such as ocean stratification, where there was no oxygen available in the deep lower layer.

Another model which has been proposed concerns sudden "bursts" of life – meaning in fact bursts of fossils which have been found. The idea is that a great increase in oxygen reduced greenhouse gases in the atmosphere. Also the sun was relatively weak relative to now. This would have made for much lower temperatures which, along with other effects, brought about one of those ***global glaciations*** referred to as a "***Snowball Earth***"[173]. There is evidence for such a glaciation, called either the ***Huronian*** or the **Makganyene**, just after the GOE about 2.22 Gya. Firmer evidence exists for two others, one around 710 Mya, the ***Sturtian***, the second around 640Mya, the ***Marinoan***. There were probably more, smaller ones.

Life – including the cyanobacteria producing all that oxygen – was greatly reduced by the glaciations, but it

172 Not a good term, in my opinion. How can you talk about an event which lasts millions of years?

173 Personally, I prefer "global glaciation" to Snowball Earth. The latter may be catchier, but try to imagine a snowball whose interior is hotter than 5000 K.

did survive, perhaps in pockets of geothermal heat. Such pockets would have been ideal locations for genetic drift and the production of new species.

So the basic idea is[174]:

increased oxygen →

lower temperatures →

glaciation →

rapid evolutionary adaptation →

new species

The GOE did not bring about evolution only of organisms. It also contributed to the formation of minerals, perhaps 2500 of the 4500 or so minerals still around today.[175]

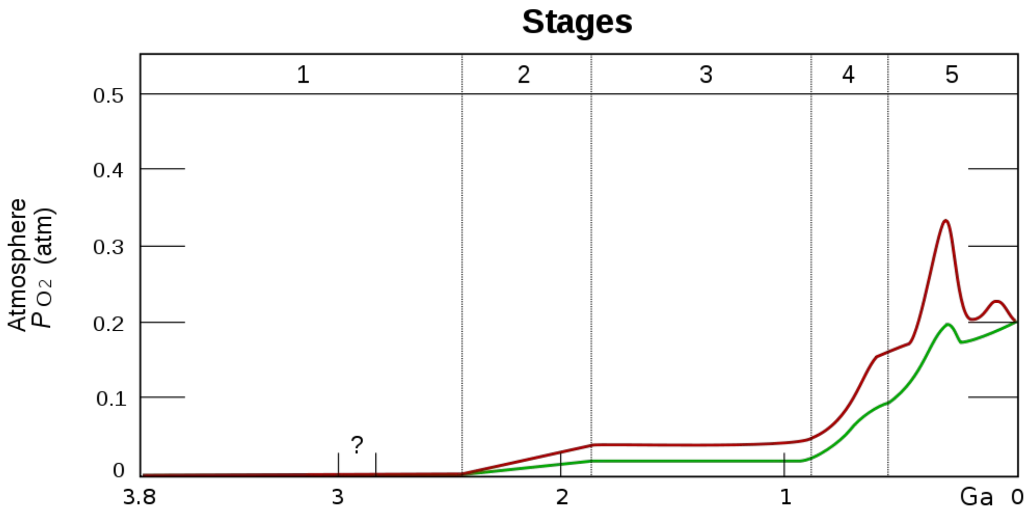


Figure 5.20: Estimated evolution of atmospheric O_2 percentage, by Heinrich D. Holland via Wikimedia Commons.[176] The red and green lines are ranges of estimates.

With the atmosphere richer in oxygen, other forms of life evolved. More complex cells called ***eukaryotes*** appeared about 1.8 Gya. Such cells incorporate smaller components called ***organelles***. Examples are the cell nucleus and the ***mitochondria***[177] essential to the generation of energy for the cell. It is now widely accepted that organelles within eukaryotes are bacteria which entered the original cell, be it prokaryote or some sort of proto-eukaryote, and stayed – a process referred to as ***endosymbiosis***.

Prokaryotes reproduce by a process of ***mitosis***, duplication and division, after which each "child" organism is essentially a clone of the "parent". ***Eukaryotes*** also duplicate themselves by mitosis, but they reproduce by ***meiosis***, a process in which selections of genes from each parent are combined.[178] This method of reproduction leads more rapidly to greater diversity of genes and, so, to the formation of new species. Only eukaryotes form multicellular organisms, a necessity for more advanced forms of life.

Taking into account biochemistry and evolutionary history, biologists now usually divide life into three domains: ***bacteria*** and ***archaea***, (both prokaryotes), and ***eukarya***, the last two being descended from the first in a yet-to-be-agreed-on order. Current eukarya include plants and animals – such as us. One proposed Tree of Life is shown in Figure 5.21.

174 Ward and Kirschvink.

175 Great oxidation event, Wikipedia, https://en.wikipedia.org/wiki/Great_Oxygenation_Event.

176 https://commons.wikimedia.org/wiki/File:Oxygenation-atm-2.svg

177 Singular, mitochondrion.

178 The subject of reproduction through mitosis and meiosis will be discussed in more detail in the chapter on biochemistry and cellular biology.

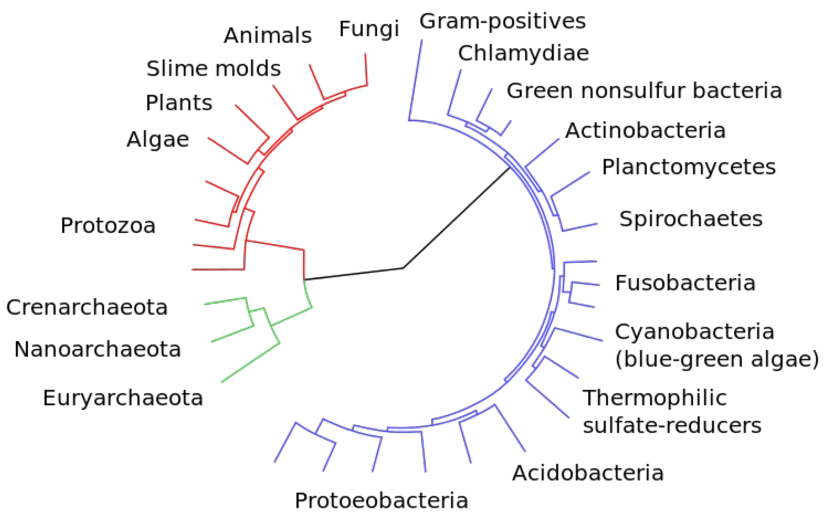


Figure 5.21: A tree of Life. Eukaryotes are colored red, archaea green and bacteria blue. From Wikimedia Commons[179]

Such trees of life depend on comparisons of certain genes across species and the choice of genes has an influence on the resulting tree. So this is one among many. It is also argued that life forms not a tree, but a network, or mesh.

Phylogenetic studies show that eukaryotes form, usually, five ***supergroups***, all of which evolved from a last eukaryotic common ancestor (LECA).[180] The members of each group have then evolved independently of the other groups. Only eukaryotes have evolved to form complex life and they all conserve properties of the common ancestor. Eukaryote cells are all very similar, all of them having, for example, common properties of cellular respiration, sexual reproduction and DNA contained in cell nuclei. The LECA must possess all the traits of the supergroups, but little else is known about its evolution.

Climate instability and glaciations

The Earth's climate now entered a period of great instability. The initial cause may have been imbalances in the geosphere and biosphere. The period of existence of a single continent, Rodinia (or Pannotia), surrounded by a single ocean under an atmosphere still low in oxygen was coming to an end around 750 Mya. New coasts brought more shallow coastal seas and bays, which in turn allowed more algal blooms. These may have gobbled up CO_2 as did chemical weathering of rocks, leading to a global cooling.

Whatever may have been the cause, Earth now embarked on an unstable period of glaciations, a second period of "Snowball Earths". Evidence for glaciers between 740 and 580 Mya ago comes from all around the Earth. Life forms were severely diminished but some survived, perhaps in warm, underwater hydrothermal vents. Eventually, CO_2 pumped into the atmosphere from volcanoes and methane, CH_4, perhaps manufactured by methanogen bacteria, conspired with other feedback effects to bring about rising temperatures and end each glaciation. Then it started all over again. Over 150 million years, at least three cycles of ice age followed by global warming occurred.

- The Sturtian glaciation peaked about 720 Mya;
- the Marinoan glaciation, about 650 Mya; and
- the Gaskiers glaciation, about 580 Mya.

Characteristic rocks left behind by retreating glaciers attest to these cycles of cold and hot. Between the second and third cycles, oxygen levels reached levels near those of today and animal life took off.

179 Wikimedia Commons, https://commons.wikimedia.org/wiki/File:CollapsedtreeLabels-simplified.svg.
180 Lane (2016), 40.

Ediacaran fossils

Fossils usually only show the harder body parts of the fossilized organisms. But from the end of the Proterozoic, around 635-542 Mya, fossils were discovered which included softer body parts of strange and complex organisms. Although the first was discovered in Charnwood Forest, England, they are named after the ***Ediacaran*** Valley in Australia where they abound. They have since been found all around the world, for instance in the interestingly named Mistaken Point, Newfoundland.

Figure 5.22: Charnia, from Charnwood Forest, by Verisimilus via Wikimedia Commons[181]

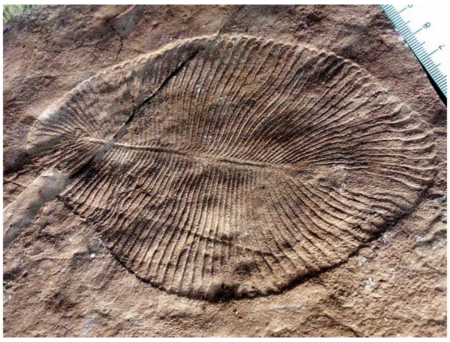


Figure 5.23: Dickinsonia costata, by Verisimilus via Wikimedia Commons[182]

The Ediacaran fossils are difficult to interpret. They seem to be generally flat, multi-sectioned organisms, often described as "quilted", without any internal structure. Charnia, for instance, seems to be a flat, fractal construction without any central stalk. Ediacaran fossils do not resemble modern organisms and are generally considered to represent an evolutionary dead end in spite of their being complex, multi-celled organisms. In any case, since they date from as much as 575 Mya, they do show that multi-cellular life existed before the Cambrian. After the Ediacarans had lived alone for up to 90 million years, they disappeared forever as small shelled organisms and trilobites took over.

5.5. The Phanerozoic Eon – rise of complex organisms

The ***Phanerozoic Eon*** is divided into three eras:

- the Paleozoic (542-251 Mya),
- the Mesozoic (251-65.5 Mya) and
- the Cenozoic (65.5 Mya to today, unless you count a 4th, the Anthropocene).

5.6. The Paleozoic Era

During the Paleozoic, the buildup of cratons and mountains continued; glaciers and shallow seas were formed. Life spread from the sea to occupy the land; and fishes, reptiles and primitive mammals evolved. Two important advances were vascular plants and the amniotic egg. A number of important extinction events forced reboots of evolution, giving rise to new organisms.

Geologists have found a huge increase in the number, variety and, especially, the complexity of fossils dating from around 542 Mya in western Canada and in China. This date has therefore been adopted as the beginning of the Paleozoic Era, which is considered to run from 542 to 251.902 Mya.[183] It is itself broken down into six subdivisions called ***periods***, named as follows:

- Cambrian (542-500 Mya),

181 https://commons.wikimedia.org/wiki/File:Charnia_Spun.jpg
182 https://commons.wikimedia.org/wiki/File:DickinsoniaCostata.jpg
183 Wikipedia, "Paleozoic". https://en.wikipedia.org/wiki/Paleozoic.

- Ordovician (500-440 Mya),
- Silurian (440-410 Mya),
- Devonian (410-360 Mya),
- Carboniferous (360-290 Mya) and
- Permian (290-251 Mya). (See Figure 5.1.)

5.6.1. Geology

Around the beginning of the Paleozoic, as tectonic plates continued moving, Rodinia broke up into ***Gondwana*** and ***Laurentia***. About 300 Mya, the sea between them shrank and they collided to form the supercontinent, ***Pangea***[184] (Figure 5.30). The superocean which surrounded it is called ***Panthalassa***. Finally, a supercontinent was not located right at the equator, as about ¾ of Pangea was in the southern hemisphere. So life forms could inhabit the continental interior.

5.6.2. Life in the sea

The extraordinary increase in the number of multi-cellular animal phyla which took place at the beginning of the Paleozoic Era (Cambrian Period) has been referred to as the ***Cambrian Explosion***. It is seen today especially as an explosion of fossils. In fact, the word "explosion" is an exaggeration which has led at least one scientist to react and call it the Cambrian "slow fuse".

For 2 billion years after the appearance of life on Earth before or around 3.5 Gya, only single-celled prokaryotes existed, cyanobacteria diligently working to survive, at the same time increasing the oxygen content of the atmosphere. Then the enigmatic fossils of the Ediacaran fauna show that multi-celled, invertebrate organisms came and, it seems, went between about 600 and shortly after 545 Mya.

The next logical step, the development of some sort of skeleton or carapace, came about in the early Cambrian, about 545-520 Mya, in the form of "***small shelly fossils***" (***SSF***s), or just "little shellies". These tiny creatures had shells of calcium phosphate, presumably because atmospheric conditions did not yet favor the calcium carbonate shells of today. So for about 25 My, the so-called Cambrian Explosion was represented simply by small shelled creatures – not much of an explosion!

Sponges, considered to be the most primitive animals alive today, had appeared in the late Ediacaran (end of the Protozoic). Radially symmetric ***echinoderms*** of the early Cambrian were the ancestors of today's starfish and sea urchins. From about 530 Mya, other invertebrates like brachiopods and worms started to leave fossil traces. ***Brachiopods***, which were shellfish with hard upper and lower valves (as opposed to the left and right valves of modern oysters and scallops, to mention the most edible of them), grew wild on the sea floors.

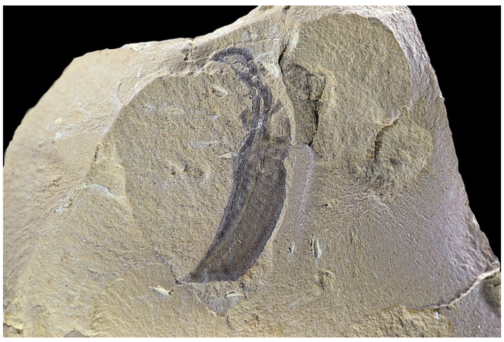


Figure 5.24: Haikouella lanceolata, from the Chengjian fossils, by Didier Descouens via Wikimedia Commons[185]

The phyla whose discovery gave rise to the term "explosion" showed up somewhat later. Extraordinary fossils including soft parts of the animals were deposited in two remarkable sites. The chronologically

184 Also spelled Pangaea.

185 Haikouella lanceolata, Chengjiang County, Yunnan Province, China. https://en.wikipedia.org/wiki/File:Haikouella_lanceolata_China.jpg

earliest was Chengjiang, China, with fossils dating around 515 Mya. Among the Chengjiang finds is the oldest fish, which is also the oldest vertebrate, dated to about 500 Mya.[186]

Probably the most famous of the Cambrian fossils are those of the Burgess Shale field of about 505 Mya (Middle Cambrian), now in Canada.[187] Some of them were quite strange and are still the subject of study and hypotheses.

Figure 5.25: Reconstruction of *anomolocaris canadensis*, one of the more bizarre Burgess Shale fossils, by PaleoEquil, via Wikimedia Commons[188]

Figure 5.26: *Hallucogenia*, another Burgess Shale fossil, after PaleoEquii,via Wikimedia Commons[189]

After the strange Ediacaran fossils, there exist fossil tracks of mostly worm-like creatures from 555 Mya. But the organisms represented by the ***Cambrian***-period fossils were of a new kind. During the early Paleozoic, continents were under shallow seas for periods of several million years at a time, so life was dominated by creatures of the seas, including reef builders. These organisms had no internal skeletons, meaning they were ***invertebrates***, but they did have a hard ***exoskeleton*** or carapace. The support this gave was advantageous in several ways: It shielded them from the sun, allowed them to retain moisture, gave support for a muscle system and protected them to some extent from predators. Many types of these creatures existed in the Paleozoic seas. Possession of a support structure enabled Cambrian organisms to grow to be larger and more complex than their predecessors. From tiny creatures, larger ones evolved.

Figure 5.27: Small trilobite, 5 cm (Ohio)

Figure 5.28: Larger trilobite, ~40 cm (Lourinha, Portugal)

An ***arthropod*** is "...an invertebrate animal having an exoskeleton, a segmented body, and paired jointed appendages." [190] One group of arthropods, ***trilobites***, became a dominant form of marine life from about 520 Mya.They existed in thousands of different species on every continent for some 270 million years, so long that they have been referred to as the "mascots" of the Paleozoic. They ranged in size from several

186 The last few paragraphs are based on Prothero (2007), 161-70.

187 Burgess shale fossils and their importance. http://www.burgess-shale.bc.ca/discover-burgess-shale/burgess-shale-fossils-and-their-importance

188 Wikimedia Commons, https://commons.wikimedia.org/wiki/File:Anomalocaris2019.jpg.

189 Wikimedia Commons, https://commons.wikimedia.org/wiki/File:Hallucigenia_reconstructions.jpg.

190 Wikipedia, https://en.wikipedia.org/wiki/Arthropod.

millimeters to over 50 centimeters. Some had eyes with many crystalline lenses, rather like fly eyes. Over time thousands of species of trilobites existed in shallow seas on every continent. Near the end of the Cambrian, there were three trilobite mass extinctions due to climate change and other factors (continental movements, evolution of predators). But trilobites survived – until the end-Permian extinction.

At the end of the ***Ordovician*** and the beginning of the ***Silurian***, two mass extinctions took place, separated by around 4 million years. They are referred to as the ***Ordovician-Silurian extinction events***. Since most life was in the sea, it was this sea life which suffered, It is estimated that 60% of marine invertebrates were destroyed. The extinctions were probably largely caused by climate change due to movement of the continents. Gamma rays from a supernova have also been proposed as a possible cause. (Section 4.6.5.)

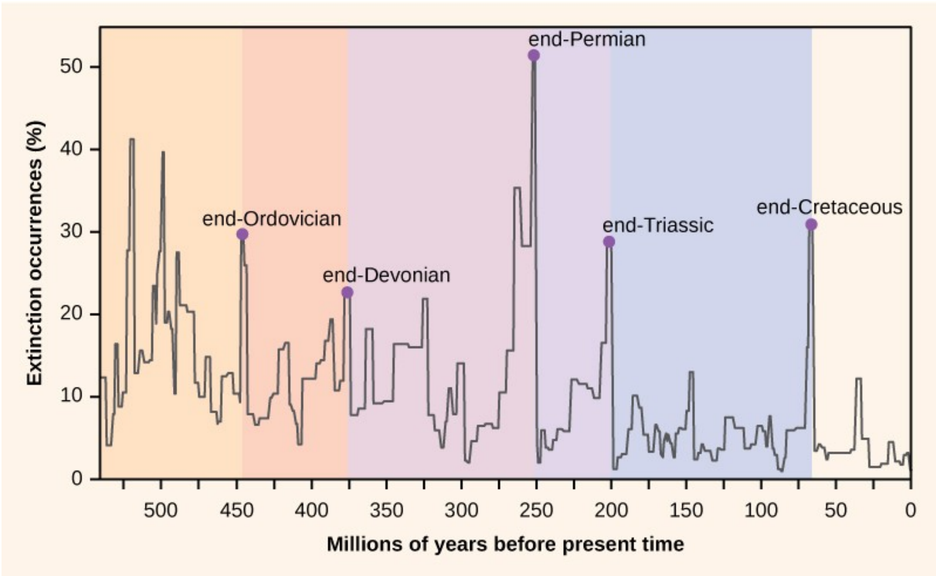


Figure 5.29: Mass extinctions, from Openstax College[191]

In the ***Silurian*** period, ***eurypterids*** (looking like scorpions or crayfish) developed which were capable of living in salt or fresh water, an important step in animal evolution. The ***ammonoids*** and ***nautiloids*** whose fossils we find so beautiful appeared toward the end of the Paleozoic,

The first fossil evidence of fishes show species which had spinal cords (making them ***chordates***) but no internal skeletons or jaws. The latter evolved from gills only later. Fish became numerous in the **Devonian** Period, which is often referred to as the “Age of fishes”. Although many types later became extinct, some of their ancestors survive even today: cartilaginous fish, like sharks or rays; fish with bones, like today's trout or bass; and lobe-finned fish, like today's lungfish.

At the end of the Devonian, another series of extinctions referred to collectively as the ***Late Devonian mass extinction*** took place. Individual events may have been separated by over millions of years. Mostly marine life was affected and trilobites were almost finished off.

5.6.3. Life on land

Plants developed first in water. Their migration onto land seems to have taken place by around 480 Mya (spores) or 425 Mya (actual plant fossils), in the Silurian.[192] The migration of plants to land was facilitated by the development of a cellulose-based support structure which afforded them the ability to transport water and nutrients from the soil to replace moisture eliminated from the plant's upper parts. Such ***vascular plants*** were flourishing by 375 Mya, and represented an important advance in adaptation to life on land.

With the advent of woody stems, plants developed to the point where the ***Carboniferous*** Period was one of dense areas of vegetation, tree-like plants and swamps. Carboniferous plants were all seedless and so had no flowers. This plant material decayed and was eventually transformed by heat and pressure into the fossil

191 http://cnx.org/contents/GFy_h8cu@10.12:lvk44_Wx@5/The-Biodiversity-Crisis
192 McDougall (2011), 237.

fuels we are busily burning up in a tiny fraction of the time it took to make them.

Such carbon sequestration led to higher oxygen levels in the atmosphere. The oxygen content of the Carboniferous atmosphere was 50-100% greater than now (Figure 5.20) and this had an effect on evolution. By or around 400 Mya, the first insect had appeared. During the Carboniferous, giant insects evolved, including a dragonfly with a 65 cm wingspan. Later, when oxygen levels came back down, the giant insects disappeared.[193]

During the ***Silurian***, tiny arthropods appeared on land. They did not have a digestive system capable of making them herbivores, but lived off decayed matter. During the ***Devonian***, skeletal changes which permitted animals to support themselves out of water facilitated the transition from fishes to ***tetrapods*** (four-limbed animals, including birds). The first land-based tetrapods were still aquatic or amphibious animals and probably lived mainly in ponds. But they were capable of breathing air, so they could move to another pond in times of drought. They also laid their eggs in water, which furnished nutrients for the young, which were essentially fish (like tadpoles). The oldest such fossils date from about 375 Mya.[194]

So first, plants moved onto the land. They were followed quickly by small arthropods, which ate decayed matter from the plants. And then tetrapods followed and ate plants and arthropods. It is all about getting enough to eat.

A very important evolutionary step was the development of the ***amniotic*** egg. This protected the young of land-dwelling animals inside a protective cover and provided the nutrients that young amphibians could only get in water. This development contributed greatly to the evolution of ***amniotes*** (the first of which resembled small lizards), which now could leave the water completely. These animals were the predecessors of two groups, ***synapsids*** (which would evolve into mammals) and ***sauropsids*** (early reptiles). The ancestors of mammals and reptiles both diverged from a common ancestor during the late ***Carboniferous***, around 315 Mya, when life on land and sea reached a new peak of development and diversity.

There is disagreement as to whether synapsids at this stage should be considered reptiles or mammals, although there are anatomical features which differentiate them. When considering the evolution of mammals, one usually speaks of a group called "***mammal-like reptiles***" (or sometimes “***stem mammals***”).

The earliest-known synapsids are the ***pelycosaurs***, the best known of which is the sail-backed dimetrodon, which flourished around 295-272 Mya and could grew up to almost 5 meters in length. Their skull has only one dermal opening permitting the attachment of muscles to the jaw, a characteristic of mammals but not of dinosaurs, which have two.

The next step along the way from reptiles to mammals came about 275 Mya, when pelycosaurs were replaced by ***therapsids***, a sub-order of synapsids. These animals were more like modern mammals, for instance, having their limbs positioned more beneath the body. They were dominant on land before the arrival of the dinosaurs.

Next, arriving around 260 Mya, came the ***cynodonts***. Officially members of the therapsid group, they were anatomically even more mammal-like than earlier therapsids, especially the skulls, and they may have had skin and been warm-blooded. Some of them survived the end-Permian mass extinction and were our ancestors.

From one mammal-like reptile to another, they became more and more mammalian. For instance, their limbs moved more and more under their bodies so their bellies were higher off the ground. Also, their double, reptilian jaw bones morphed into one, mammalian bone with the other one moving into the three ear bones.

But we have erred now from the current period, the Paleozoic, into the Mesozoic, so let's get back to where we were. We’ll look more at the Mesozoic in section 5.7.3.

Tectonically, the plates which today would be Europe and North America were then situated in tropical climates near the equator. Because no land mass was over either pole, polar ice caps were limited and the Earth's temperature gradient was less pronounced. On land, huge tree-like plants grew in swamps and life reached all the continents. Insects and tetrapods swarmed through the undergrowth. But no bird sang and no flowers lent color to the scene.

193 Fortunately, in my opinion.
194 McDougall (2011), 236.

Near the end of the ***Carboniferous***, as Gondwana (the southern continent comprising today's South Africa, South America, Antarctica, Australia and India) approached the poles (Figure 5.30), there was a period of glaciation which lasted into the Permian. Remaining glacial features on these continents provide evidence for plate tectonics, as some of these continents now occupy much warmer latitudes[195].

The ***Permian*** Period was dominated by the existence of the supercontinent Pangea. Around the equator, the Carboniferous swamps had given way to deserts and these arid conditions were well suited to the development of reptiles.

To the east, projecting into the continental land mass, was the Tethys Sea, which was swarming with life. This was also true of the Zechstein Sea in the north, the area of current northern Europe. Parts of the Zechstein evaporated, leaving behind minerals (*evaporites*) which helped furnish raw materials for the Industrial Revolution – plaster of Paris, gypsum and substances used for the production of acids and ammonia.

The distribution and variety of organisms today is a result of the existence and subsequent breakup of Pangea. During its existence, no waterways blocked migration routes, so animals, at least those who could support the aridity of the interior, were free to move about to new habitats. The later breakup of Pangea was an equal boon to evolution as organisms isolated from one another tend to evolve in different ways from similar beginnings. Simply put, “isolation begets diversity.”

The time of Pangea was one of much development in the forms of life. By its end, dinosaurs and early mammals had developed. Many insects existed, including cockroaches, which are still with us, alas.

5.6.4. The end-Permian extinction

The Paleozoic Era ended with the greatest of all the mass extinctions, the ***end-Permian extinction*** (or Permian-Triassic extinction), sometimes referred to as the ***Great Dying***. It is estimated that 96% of sea and 70% of land species disappeared[196]. The date of the extinction marks the end of the Paleozoic and the beginning of the Mesozoic Era, officially 251.902 Mya.

Large igneous provinces (***LIPs***), or ***flood basalts***, may cover up to millions of square kilometers and be several km thick. LIPs are formed by relatively continuous, non-explosive volcanic activity, but nevertheless form quickly on the geological timescale, in less than a million years. They are thought to have formed over deep plumes of magma and so can form on land or under seas, independently of plate boundaries

A strong case has been made for the model in which LIPs bring about major climate change due to temperature increases caused by emitted CO_2 and methane, which then lead to mass extinctions.[197] At least three major LIPs occurred at the same time as three major mass extinctions:[198]

- The Siberian flood basalts, or Siberian traps[199], remain from a major LIP which occurred around 252 Mya – at the time of the end-Permian extinction.
- The Central Atlantic Magnetic Province (CAMP) occurred about 200 Mya – at the time of the Triassic-Jurassic mass extinction.
- The Deccan traps in India were formed about 65 Mya at the time of the K-T extinction.

So the end-Permian extinction was probably initiated by volcanic eruptions in Siberia which increased the amount of CO_2 and methane in the atmosphere, disrupting the carbon cycle and bringing about a “runaway greenhouse phenomenon”.[200] This in turn would have caused oceans to release dissolved oxygen. It could have caused acid rain which killed land plants vital to survival of animals. The CO_2 would have been absorbed by the oceans and led to their acidification, wiping out many marine organisms.

It took around 20 million years for life to recover, far longer than after the other known mass extinctions.

195 Benton, p. 90.
196 MacDougall 1998, p. 321.
197 See Annex B for more detail on LIPs and OAEs (oceanic anoxic events).
198 Eight others have been asserted. “Flood basalt volcanism during the past 250 million years”, https://science.sciencemag.org/content/241/4866/663.
199 “Traps” from the swedish word for staircase, “trappa”.
200 Benton, p. 118.

When life did attain its previous diversity, its forms had changed. The few remaining trilobites had been completely eliminated.

5.7. The Mesozoic Era – age of reptiles

The Mesozoic began after the End-Permian mass extinction, 251 Mya, and ended in the less catastrophic but better-known Cretaceous-Tertiary ("K-T") mass extinction, 65 Mya. Like the Paleozoic, it is subdivided into periods:

- the Triassic (251-200 Mya),
- the Jurassic (200-145 Mya) and
- the Cretaceous (145-65 Mya).

Life in the Mesozoic was dominated by reptiles, including dinosaurs, so it is often called the "Age of Reptiles".

5.7.1. Geology

As the Mesozoic opened, there was only one continent, Pangea. Pangea did not just form and then sit stationary for millions of years; tectonic activity continued throughout the era. Evidence from fossils (types of organisms) and sedimentary rocks (formation from sand dunes) indicate that the interior of Pangea was quite arid.

As plates shifted, rifts developed and sea water periodically spilled into the rifts. In between periods of flooding, evaporating water left behind salts (*evaporites*). Measurement of the ages of these salts gives geologists a calendar of the opening of the rifts and therefore the development of the inter-continental oceans.

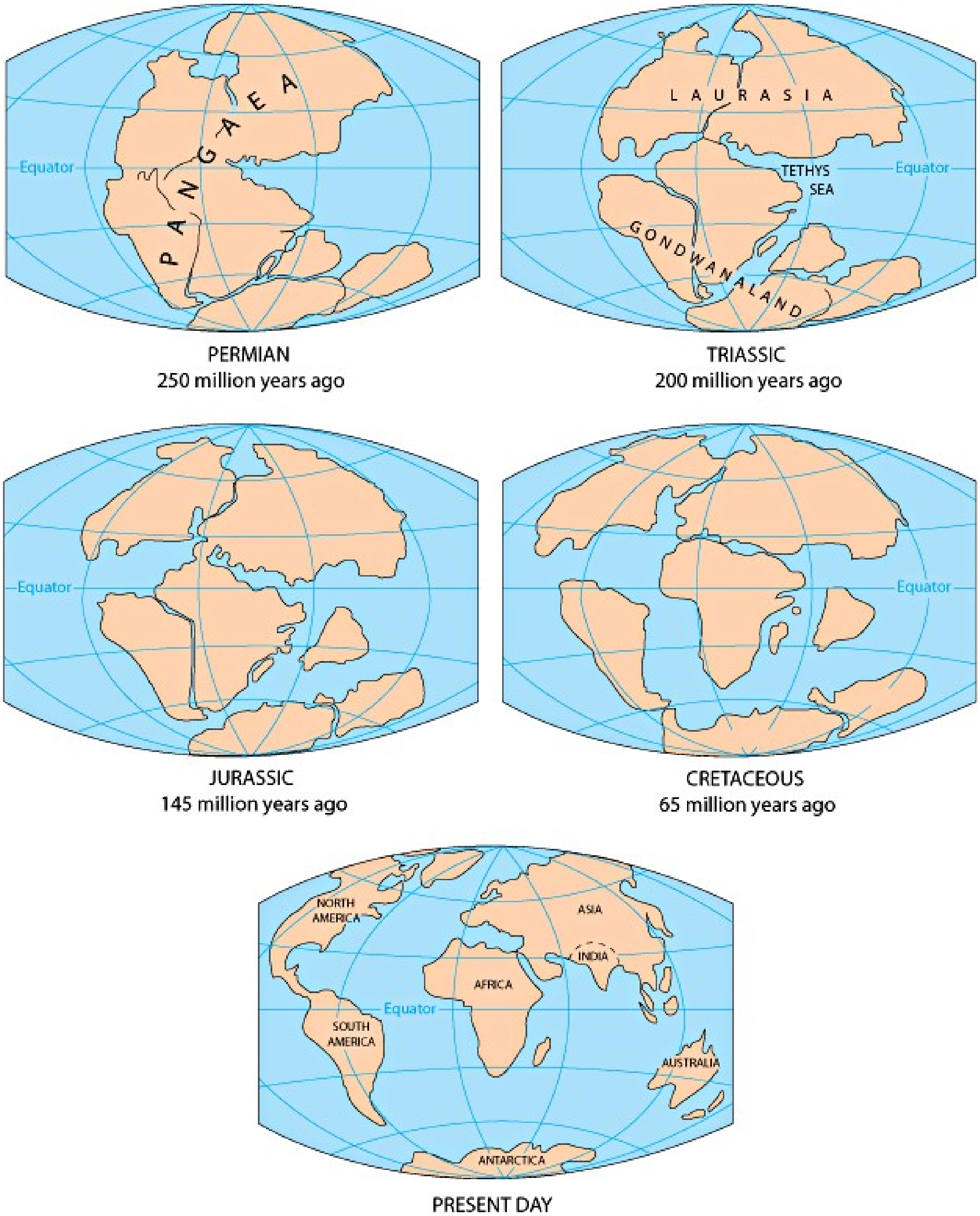


Figure 5.30: Movement of the continents[201]

The main geological event of the Mesozoic was the breakup of Pangea into seven major continents. It began during the Jurassic, around 175 Mya, when what would become the Atlantic Ocean penetrated into Pangaea from the west and the Tethys Sea, from the east, separating the supercontinents ***Laurasia*** in the north and ***Gondwana*** (or Gondwanaland) in the south. During the Cretaceous, around 65 Mya, Laurentia rifted from Laurasia and the newly born North Atlantic opened up. A huge mountain chain which had stretched diagonally across the great continent left its remains in the Appalachians of North America and the mountains of Ireland, northern Scotland and western Scandinavia.

201 U. S. Geological Survey, “This dynamic earth”. http://pubs.usgs.gov/gip/dynamic/historical.html

Only somewhat later in the Jurassic, though the exact timing is still under study, Gondwana rifted and South America and Africa began to separate, the South Atlantic Ocean opening between them. Antarctica and Australia broke loose from Gondwana and India, from Africa. Only in the last 20 My or so have North and South America joined to form the minor supercontinent America.

Note that Laurentia and Gondwana were originally formed at the breakup of Rodinia, then fused to form Pangea, before again breaking apart at the end of Pangea and fragmenting into seven continents.

The resulting changes in continental and oceanic configurations brought about changes in the climate. It is thought that the climate during this era was generally warm, with a reduced temperature gradient between the equator and the poles. This spurred evolution, for instance, through the creation of new econiches.

5.7.2. Life in the sea

The ancient “mascot” of the Paleozoic, the trilobite, did not survive the end-Permian extinction. This diverse group of arthropods had proliferated in the seas for over 250 million years. In terms of longevity, they beat out the soon-to-become-dominant dinosaurs. Nor were there any reefs during the early or middle Triassic. It is amazing anything did survive. The extinction was like pushing the restart button for life on Earth. But survive, it did.

An expansion in the numbers of plankton brought about abundance at the bottom of the food chain. Skeletons of these tiny beings, which precipitate to the sea bottom after death, contain calcium carbonate and silica. They pile up on the sea bed to form chalk, hence the name, cretaceous. The cliffs of Dover have their origin in these minuscule creatures. Elsewhere, along the edges of the Tethys Sea, organic material decayed and metamorphosed to produce the oil found today in places like Russia, the middle East or along the Gulf of Mexico – the source of so much pollution and conflict.

Marine reptiles of new sorts appeared. The ***ichthyosaur***, a giant fish-like marine animal was a land reptile which had returned to the sea but continued bearing its young alive. However, it was not the ancestor of current whales or dolphins.

5.7.3. Life on land

Some land plants survived the extinction. Forests of conifer and ginkgo trees, among others, grew up. Flowering plants (***angiosperms***) showed up only late in the Mesozoic, around 100 Mya, but quickly spread to all environments. Such plants produce seeds which, like the amniotic egg for animals, are furnished with a protective skin and and plentiful nourishment. Through nourishment and pollination, plants and insects have co-evolved ever since the Cretaceous. Plants developed colors and structures that protected them from all but specific insects. Insects simultaneously developed so as to live off certain flowers. Flowering plants and insects form an inseparable symbiosis. Think ecology.

The acknowledged stars of the Mesozoic were the reptiles, who came to dominate the sea, the land and the air for almost 200 million years. Dinosaurs diversified by occupying econiches left vacant by the animals killed off in the end-Permian extinction. They evolved to eat vegetation, meat, fish and even insects. They lasted so long and went so far that, had they had historians or philosophers, these might well have thought that they were the end product of evolution and, so, eternal. How about that?

So-called stem reptiles[202] already had been around at the end of the Permian, having evolved from amphibians. Although their numbers were greatly diminished by the end-Permian extinction, they made a comeback in the early Mesozoic.

At that time, there were principally two types of reptiles which had survived the extinction. The ***cynodont*** (of the ***therapsid*** group) was one of the lucky survivors from the late Permian. We mentioned their descent from other therapsids in section 5.6.3. It was a mammal-like reptile and had skin instead of scales. It probably had hair and whiskers and may have been endothermic (warm-blooded). In spite of being a reptile, details of its skull and jaw were mammal-like. It remained small and furtive throughout the Mesozoic, living in burrows by day and sneaking out at night to scavenge or hunt for insects. Among its descendants is the species *Homo*.

202 Or *captorhinids*; ancestors of archosaurs.

The other surviving reptiles were the ***archosaurs***[203], whose current descendants include turtles, snakes, lizards, crocodilians and birds. They also evolved into dinosaurs.

The oldest dinosaur fossils date from around the early Triassic, 240 Mya. The fossil record of dinosaurs is extremely incomplete, but it is known that they spanned the globe, largely due to their amniotic egg and scaly skin, which enabled them to live far from the sea. As Pangea started breaking up in the Jurassic, the climate grew moister, making a much greater area habitable by animals.

The boundary between these two periods was marked by the ***Triassic-Jurassic mass extinction***. This extinction occurred in two or three steps across some 18 million years. Although various causes are ascribed (see section 5.6.4) , it is known that at least half the species on Earth disappeared, allowing dinosaurs to take over as the dominant species. So the Jurassic was to be the age when dinosaurs ruled.

Figure 5.31: Dinosaur eggs from Museu da Lourinhã, Portugal. Photo by author.

It is not certain whether dinosaurs were warm or cold-blooded. The biggest ones would have needed high blood pressure in order to pump blood from the heart to the brain, many meters away, maybe even in an upward directions; this would have implied a mammal-like circulatory system, in which case they would have been warm-blooded.

The first dinosaurs were quite small, about 1 meter long. Many of them were bipedal. In the course of their long reign of almost 200 million years, they diversified and evolved. Many grew larger or developed defensive armor, like the ***stegosaurus*** or the ***triceratops***. Dinosaurs did evolve quite a lot, and not all species of them lived simultaneously. The fight between Tyrannosaurus rex and Stegosaurus to the music of Stravinsky's *Le sacre du printemps* in the original movie “Fantasia” could not have taken place, as these animals lived at different times.

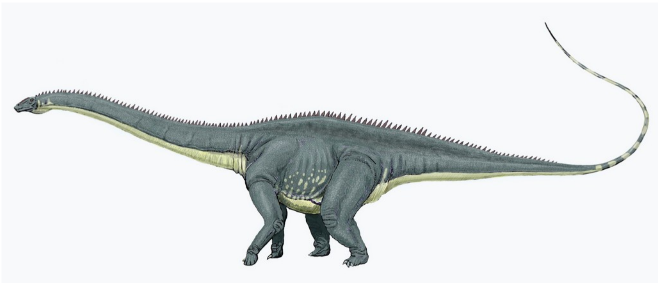
Paleontologists identify two types of dinosaurs, based on the bones of their hip and pelvis: ***ornithischian*** (“bird-hipped”) and ***saurischian*** (“lizard-hipped”). The bird-hipped dinosaurs could be either bipedal or quadrupedal and were all herbivores. The *stegosaurus* was one of them. The lizard-hipped dinosaurs are in turn divided into two groups: ***sauropods*** and ***theropods***. Sauropods were the commonest and were huge, including some of the largest land animals ever, ***diplodocus***. Theropods were all bipedal and carnivorous and included the infamous ***tyrannosaurus***.

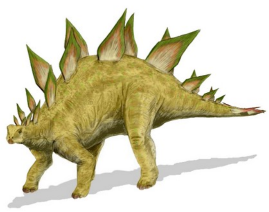
A very limited number of examples of dinosaurs is indicated in the following table.

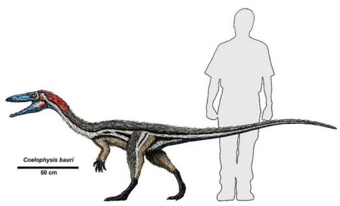
203 Again, there is disagreement as to whether the archosaurs are the animals discussed here or others which appeared only in the early Triassic.

Name	Length/height	Weight (kg)	Diet	Pedalism	Time
coleophysis	3m/2m	27	carn	2	late Triassic
megalosaurus	9m/-	n/a	carn	2	mid Jurassic
stegosaurus	9m/-	n/a	herb	4	late Jurassic
archaeopteryx	0.5m/0.2m	0.4-0.5	carn	2, flying	late Jurassic
diplodocus	26m/8m	20000-25000	herb	4	late Jurassic
triceratops	9m/3m	5500	herb	4	late Cretaceous
tyrannosaurus	12m/5.6m	7000	carn	2	late Cretaceous

Dinosaurs (not at all to scale; that would have been impossible) in the following table are, clockwise from the top[204]: Diplodocus, Triceratops, Tyrannosaurus Rex, Stegosaurus and Coelophysis







Birds only evolved in the late Jurassic, 140-150 Mya. One of their reptilian fore-fingers grew long enough to support a web of skin which helped the animal to plane in the air – the first wing. The early ***pterosaur*** ("winged lizard") was a flying reptile, not yet a bird. Birds and pterosaurs evolved from reptiles along separate lines. The transition between reptiles and birds is thought to be ***archaeopteryx***, which had a reptilian skeleton but possessed feathers formed for flight. This made it a better flyer than the pterosaur. During the Cretaceous, these animals developed hollow bones like modern birds. It is now known that a number of dinosaurs had feathers, which were probably for insulation as they did not have the asymmetric

204 All images from Wikimedia Commons: Dipodocus (https://commons.wikimedia.org/wiki/File:SeismosaurusDB.jpg#/media/File:SeismosaurusDB.jpg), Triceratops (https://commons.wikimedia.org/wiki/File:Triceratops_BW.jpg), Tyrannosaurus Rex (https://commons.wikimedia.org/wiki/File:Tyrannosaurus_rex_mmartyniuk.png), Stegosaurus (https://commons.wikimedia.org/wiki/File:Stegosaurus_BW.jpg) and Coelophysis (https://commons.wikimedia.org/wiki/File:Coelophysis_size.jpg).

form needed to act as an airfoil.

The world was now enormously different from the Paleozoic, enhanced (from our point of view) by flowers and bird song.

5.7.4. K-T extinction

At the end of the Cretaceous and just before the Tertiary, there was another great extinction. Though much less serious than the end-Permian extinction, this one has captured imaginations more. This is partly due to the fact that it brought about the end of the dinosaurs, except for birds, thereby permitting the rise of mammals and the eventual arrival of humans. It is also due to the explanations offered for the extinction.

One explanation depends on volcanic explosions in the Deccan Traps, in what is now central India. (see section 5.6.4) The other explanation cites the crash of a mammoth asteroid into the Earth on the coast of what is now Yucatan, forming the Chicxulub Crater. Both events occurred very near the K-T border. The asteroid would have caused a massive tsunami. Both events would have flung debris and gases (sulfur dioxide and carbon dioxide) into the air, darkening the planet and reducing temperatures.[205] Cooling surface waters would have brought up nutrients from below and favored plankton blooms, affecting plant and animal life.

The result was the disappearance of some three quarters of plant and animal species on Earth, including ammonites and dinosaurs. Fossil evidence shows that the extinction was world-wide. But some members of each group of organisms managed to survive into the Cenozoic Era.

Time	Approx. date (Mya)	Probable cause(s)[206]	Principal victims[207]
Ordovician-Silurian	450-440	Climate change (ice age) due to continental movement	60-70% of all species (2nd worst), mainly in sea
Late Devonian	375-360	Asteroid impacts,changes in sea level and chemistry	At least 70% of all species, worst in shallow seas
Permian-Triassic (End-Permian, “The Great Dying”)	252	Atmospheric change due to volcanic eruptions – and much more	90-96% of all species (worst), including trilobites and other marine creatures, and insects
Triassic-Jurassic	200	Massive lava floods	70-75% of all species
Cretacious-Tertiary (“K-T”)	65	Volcanic explosions and basalt floods, asteroid impact	70-75% of all species, including all dinosaurs except birds

Table 3: Five mass extinctions

Scientists usually count five major mass extinctions, which are summarized in Figure 5.29 and in the preceding table.

5.8. The Cenozoic Era – mountains and mammals

5.8.1. Geology and atmosphere

The Cenozoic Era is considered to be divided into several periods:

- the “Tertiary” (65-1.8 Mya), commonly broken down into

205 According to one study, temperatures would have been reduced by at least 26°C, which would have brought about temperatures around -15C. http://onlinelibrary.wiley.com/doi/10.1002/2016GL072241/abstract or http://passeurdesciences.blog.lemonde.fr/2017/01/23/les-dinosaures-sont-morts-de-froid/.

206 Causes from Wikipedia, Ibid., and BBC Nature. “Big five mass extinctions”, http://www.bbc.co.uk/nature/extinction_events.

207 Figures from Wikipedia, “Extinction event”, https://en.wikipedia.org/wiki/Extinction_event

    - the Paleogene (65-23 Mya) and
    - the Neogene (23-1.8 Mya);
- the Quaternary (from 1.8 Mya).

Each period is broken down into two or three epochs. (See Figure 1.) Some scientists consider that we now are living in a new epoch, the ***Anthropocene***, dating vaguely from the beginning of the Industrial Revolution (or perhaps the beginning of agriculture), but the term has not yet been accepted by any official geo-chronological body.

Mountains

The Cenozoic Era has been called the age of mammals, but it could equally well be called the age of *orogeny*, or mountain-building. The band of mountains running from North Africa to Indonesia, including the Atlas, Pyrenees, Alps, Taurus and Himalayas, was formed during the Cenozoic.

During the Cretaceous, Eurasia and Africa were separated by the Tethys Ocean (Figure 5.30). When the Atlantic started to spread out and Africa broke loose from Antarctica, it floated northwards towards Eurasia. Initially, small bits of continents within the Tethys (now Spain and Italy) were pushed up and joined to southern Europe, forming the first Alps and Pyrenees. As Africa continued its journey north, the Tethys was closed and the Mediterranean formed. Subduction due to Africa's continued northward push is responsible for the volcanoes around southern Italy, Sicily and Greek islands like Thira (Santorini).

In a similar way, the Arabian plate moved toward Turkey and pushed up the Taurus Mountains. This plate too is still moving and is responsible for earthquakes in the Middle East.

With the breakup of Gondwanaland, what is now India started moving north. Around 55 Mya, it entered into collision with Asia, administering the *coup de grace* to the Tethys. The crust of India was too light to be subducted, but some of it slid in under Asia, making the continental crust there almost twice its maximum elsewhere. The result was the chain of the mighty Himalayas and the high Tibetan plateau. In both places, at thousands of meters above sea level, there are sedimentary rocks from the floor of the Tethys, with accompanying fossils, such as the ammonites called ***shaligrams***, which are worshiped by Hindus as an incarnation of the god Vishnu (him again!). India's encroachment on Asia is still going on at the rate of around 2 cms per year. This process leads to buildups of tension in the rock which, when released, cause enormous earthquakes in China.

Figure 5.32: A shaligram (ammonite) from the Himalayas

As summer heat warms up the high Tibetan plateau, moist tropical air is drawn in. When this precipitates, it brings about the monsoon. This weather phenomenon is both good (irrigation) and bad (flooding and mudslides) for India and Bangladesh.

Since the Earth's surface maintains a fixed area, the expansion of the Atlantic Ocean causes the squeezing up of the Pacific. As the Pacific plate has subducted almost all the way around its perimeter, volcanic activity has arisen there – the so-called Pacific Ring of Fire.

Climate and the PETM

Geologists originally distinguished the boundaries between different geological ages by relatively sudden observed changes, such as significant modification of the fossil record. Only later did they learn why such changes had occurred. The Paleocene-Eocene boundary is at the "moment" of the ***Paleocene-Eocene Thermal Maximum***, or ***PETM***, also referred to as the ***Great Warming***, which can be seen clearly as a narrow peak on the graph of temperature change in Figure 6.3.

Data from sediment cores of about 55 Mya show a strong "isotope excursion" over a period of around 110,000 years and a lack of carbonate deposits (from dead organisms) over about 60,000 years.[208] These data indicate that a sudden, huge amount of carbon was added to the atmosphere – as CO_2 and methane. Ocean currents were also disrupted. These changes occurred at the same P-E boundary where it was already known that many foraminifera species (of ocean plankton) became extinct. It has since been found that the isotope excursion was world-wide, on land, in oceans and in the atmosphere. It also shows that much of the carbon was ejected in the form of methane, but there is still disagreement as to its source. Nevertheless, the picture which has emerged is that the normal equilibrium of the carbon cycle maintained by volcanic eruption and chemical weathering (see section 5.1.8) was overpowered by short-term processes. The result of all this greenhouse gas was an increase in atmospheric and oceanic temperatures between 5°C and 10°C.[209]

At the beginning of the Oligocene Epoch, about 34 Mya, temperatures started falling steeply, leading to the ice ages of the Quaternary. Temperatures can be deduced from the shape of fossil leaves and from the isotopes of oxygen in limestone formed in the sea. The two methods agree well on the high temperatures of the Eocene epoch and the sharp drop at the beginning of the Oligocene. In the next chapter, we will see the important influence of these climate variations on primate development.

The origins of these temperature shifts are found in plate tectonics. The existence of a large antarctic continent meant that sea waters had to flow around the whole continent. But when Australia and South America separated from Antarctica in the Cenozoic, it became isolated in a circuit of cold waters and a permanent ice cap gradually formed, around 10 Mya. In a similar way, the closing of the Atlantic-Pacific connection by the formation of the Isthmus of Panama 3-4 Mya brought about reinforcement of the warming Gulf Stream, good news for Europe. But this led moisture up to the north, where it precipitated and soon formed a polar ice cap there. Ice caps bring about global cooling. Plate tectonics influences climate which, in turn, influences evolution.

The still-controversial ***uptake hypothesis*** suggests that the chemical weathering of recently-exposed rock removes CO_2 from the atmosphere. If so, then formation of the Himalayas might in this way have contributed to global cooling.

Since the beginning of the Quaternary, there have been a series of glaciations, or ice ages, with periods of roughly 100,000 years, explained as Milankovitch cycles. This repeated glaciation has shaped the Earth through erosion and displacement of rocks. The last great ice age started about 130 Kya and left northern Europe and America only around 12-15,000 years ago. We are now living near the peak of a relatively warm interglacial period, but there is no reason to expect that to last indefinitely.

5.8.2. Life

As the Mesozoic was the age of reptiles, so was the Cenozoic the age of mammals.

During the Tertiary, mammals took over the econiches left empty by the then-extinct dinosaurs, just as these had occupied econiches left vacant after the end-Permian extinction. Like the dinosaurs, mammals rapidly grew in numbers and varieties. They currently range in size from mice to elephants, and of course whales. During the ice ages of the early Quaternary, some grew to be enormously big: a huge rhinoceros-like animal; elephants and mammoths; whales, which include the largest animal ever, bigger even than the largest dinosaurs; various large animals like camels, elks, sloths, bears and so forth. Although their size made them well adapted to cold climates, most of these megafauna died out about 13 Kya, for reasons which are still not understood.

The Earth has been around for 4700 million years and mammals since something like 250 million years but only in great numbers for 65 million years. Man has only been around for at most 4 million years – and even then he did not bear much resemblance to us.

But that is for the next section (What paleontology tells us).

208 MacDougall 2011, 174.
209 MacDougall 2011, 178.

5.9. And now...

Plate movements are not finished. At this moment, Africa is on the move. The entire continent is moving slowly northward, causing volcanic activity on mainland Italy (Vesuvius) and Sicily (Etna). One day, Africa will smash (slowly) into southern Europe, raising new mountain ranges and obliterating the Mediterranean Sea. Meanwhile, within Africa, the Great Rift Valley is threatening to break off all of East Africa into a new (small) continent, with a new ocean in between it and the rest of Africa. The dance continues.

Temperature and sea-level variations will certainly continue. In the long run, whatever that may turn out to be, there is no reason to expect the climate to remain friendly to the human species. Bacteria, not eukaryotes, are the basis and the necessary ingredient for life on Earth. We depend on them, not the other way around. One day we will go the way of the trilobites and the dinosaurs – and probably much sooner than they.

6. What paleontology and evolution tell us

“In short, paleontology is the study of what fossils tell us about the ecologies of the past, about evolution, and about our place, as humans, in the world.”[210] Paleontology, the study of the evolution of ancient life, draws information not only from the discovery and study of old bones, but also from archeology, genetics, linguistics, climatology and other fields. Interpretations of existing data differ and can change with each new discover of fossils. Since new bones are discovered quite often, paleontology is constantly a Work in Progress.

That explains why this article may well be the one with the most occurrences of words like “maybe”, “perhaps” or “thought” (as in “thought to be...”), indicating uncertainty in the understanding of some findings. This situation casts no doubt on the overall results showing the evolution of our species, Homo.

6.1. The evolution of man and his family bush

It would be nice to be able to draw a family tree for mankind. There exists much evidence for numerous intermediate species between man and his ***last common ancestor*** (***LCA***) with chimpanzees. But it is currently impossible to distinguish a linear sequence of species on such a tree. To continue the metaphor, the tree really looks more like a bush, with twigs sticking out in all directions, masking the underlying branches. Nevertheless, it is convenient to group together some twigs whose similar characteristics indicate they may sprout from a common branch.

The problem is complicated by the fact that different genes may have reached a given "destination" via different pathways. Perhaps a tree for a given species is best represented by a "democratic majority of the genome".[211]

Before going further, some vocabulary is needed. A ***primate*** is a mammal of the order Primates (logically enough), mostly arboreal, ranging in size from lemurs to gorillas, and including, among others, monkeys, chimpanzees, gibbons and man. ***Hominins*** are species on the main human twig of the bush of evolution members of the family ***Hominidae***.[212] Members of the chimpanzee twig are called ***panins***.

There are some points of which we are quite certain:

1. Man has evolved from some creature which was the common ancestor of both man and the chimpanzee, which genetic analysis shows to be the current species closest to us. We are not descended from monkeys; we and monkeys have a common ancestor from which both are descended.
2. Among all the forms of primates which have preceded modern man, it is difficult to distinguish a unique, linear sequence of forms, each one evolved from the one before. Nevertheless, overall changes show clearly that evolution has taken place.
3. The genetic notion of “molecular clocks” indicates that hominin evolution has taken place for up to 7 million years, often during periods of extreme climate change (Figure 6.3) which some species survived better than others.[213]
4. Astounding as it may appear to us now, at most times in our evolutionary history, different forms of man existed at the same time. The best known example is that of Neanderthals and Cro-Magnons. They lived near each other in western Europe and even shared some genes, so it is clear that “social” interaction took place between the species. Imagine living near a group of animals of another species, another kind of animal, a sort of ape with which you could communicate (and even copulate). Would we try to enslave or annihilate them (or use them for experiments)?

210 University of California Museum of Paleontology, http://www.ucmp.berkeley.edu/paleo/paleowhat.html.

211 Dawkins (2004), 142.

212 Wikipedia lists six classifications for humans beneath the family Hominidae: subfamily Homininae, tribe Homini, subtribe Hominina, genus Homo, species, H. Sapiens, subspecies H. s. sapiens. Who can possibly remember and distinguish those three different endings for homini – ai,i and a?

213 The so-called genetic clock calculates duration based on the number of genetic changes taken place multiplied by an approximate time per change.

5. “It” (the evolution of primates from earlier forms into man) all started in Africa.

6.2. Characteristics of hominins

The following criteria are generally taken to show that a given fossil is more like a hominin than a panin.

- More perfected ***bipedalism***, a greater ability to walk upright on the two hind legs. A number of factors are associated with this ability:
    - a more vertical trunk, wider hips, straighter and lockable knees, lower limbs longer than upper, feet suited for walking rather than for climbing in trees;
    - relatively forward placement of the foramen magnum, the hole where the spinal column enters the skull, due to the erect posture;
    - greater height;
- greater cranial volume and brain size (larger for hominins), which is correlated with increased pelvic size necessary for such large-brained babies to be born;
- less prognacious (less flat face;
- skull, jaw and dental structure (related to diet):
    - teeth in a parabolic row;
    - smaller and less protruding canines, relatively larger incisors and larger chewing teeth (molars);
    - more robust mandibles (lower jaws).

Two somewhat linked developments are bipedalism and increased brain size. Bipedalism seems to have prepared the way for bigger brains (as we shall see shortly).

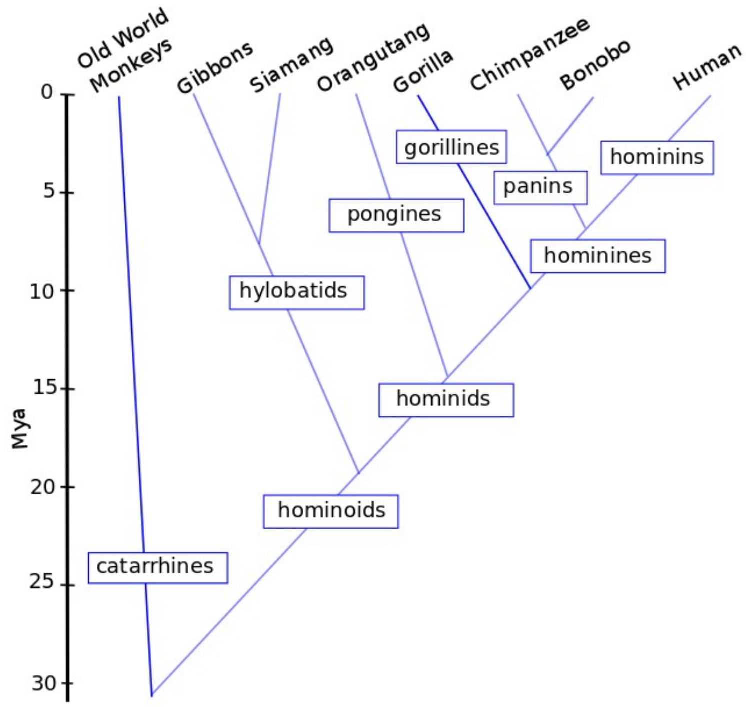


Figure 6.1. Hominoid families with dates, diagram by author

Either such taxonomic[214] characteristics or genetic analysis may be used to classify different families and species of primates as shown in Figure 6.1. Results from the two methods are not necessarily the same.

Another way to see this is in Table 4 The difference between the table and the diagram in the placement of

214 Taxonomy is the practice and science of classification.

the Gorillini may indicate the method of analysis used (taxonomic or genetic)[215].

Super-family	Hominoidea (hominoids)					
Family	*Hominidae* (hominids)					*Hylobatidae* (hylobatids)
Sub-family	*Homininae* (hominines)			*Ponginae* (pongines)	*Gorillinae* (gorillines)	
Tribe	*Hominini* (hominins)		*Panini* (panins)			
Sub-tribe	*Hominia* (hominans)	*Australopithecina* (australopiths)				
Genus	*Homo* (*heidelbergensis,* floresiensis, erectus, ergaster, neanderthalensis, *sapiens*)	Ardepithecus Australopithecus Kenyanthropus Orrorin Paranthropus Sahelanthropus	*Pan* (chimps, bonobos)	*Pongo* (orangutan)	Gorilla	*Hylobates* (gibbon, siamang

Table 4: Hominoidea super-family[216]

6.3. Groups of hominins

Species may be grouped together according to some common characteristics. Figure 6.2 indicates the time period of most currently known fossil hominins. Different colors indicate different groups.

In Figure 6.2, a significant number of hominin species are grouped in two different ways. One grouping[217] is indicated by the background colors:

- beige – possible and probable hominins
- blue – archaic and transitional Homo
- green – pre-modern Homo
- pink – Homo group

The color of the vertical bars representing the time when the species lived represents the grouping of the Smithsonian Museum of Natural History[218]:

- brown – Ardipithecus group
- green – Australopithecus group
- magenta – Paranthropus group
- blue – Homo grouping
- pale green – not grouped by Smithsonian (considered controversial)

215 There is disagreement about placing gorillas under hominoids or hominids. See www.hominides.com/html/dossiers/hominoide.php.

216 Based on Wood, 2005.

217 Based on Wood, 2005,

218 http://humanorigins.si.edu/evidence/human-family-tree

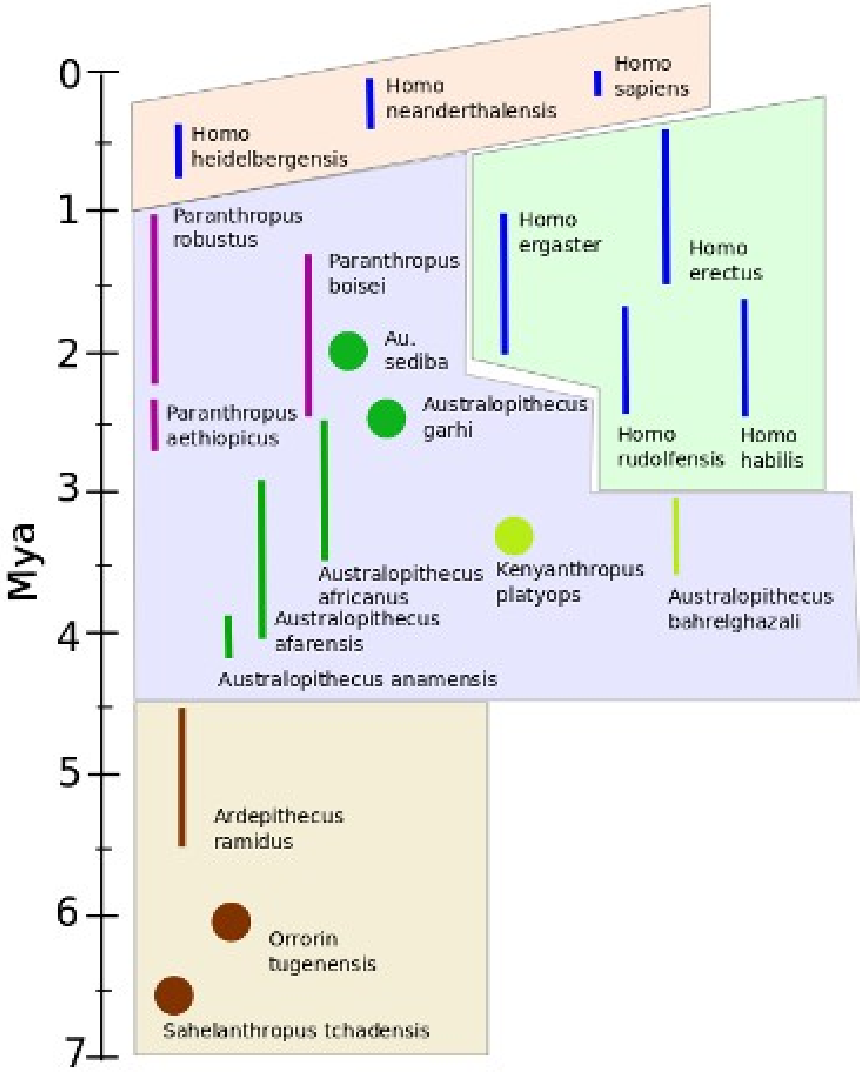


Figure 6.2. Timeline and grouping of principal fossil hominid species

While general characteristics of different species among the Australopiths and others evolve across the ages, different parameters do not always evolve together. For instance, *Au. anamensis* has chimp-like canines but fairly evolved bipedalism, whereas *Pa. aethiopicus* has smaller canines but its foramen magnum is near the back. Nevertheless, from the bottom of the bush to the top, overall evolution does occur and near the top we find our modern species panins and hominins – in particular, us.

Many paleontologists think that *H. erectus* is a later and Asian version of *H. ergaster;* others think they are different. In either case, there were at least four species of hominins living around 2 Mya,

Before considering these groups in detail, it is necessary to consider the preceding rise of mammals and the role of climate in evolution.

6.3.1. Geology, climate and evolution

Global temperature is a function of many variables, but there are two main ones:

1. how much energy is received from the sun and
2. how much of it is trapped by the oceans and the atmosphere, rather than being reflected back out into space.

Considerations of energy received must take into account solar activity.

As already discussed in Section 5.1.7, the energy falling onto the Earth's surface depends on its orbit – its axial tilt (the angle of its rotational axis relative to its orbital axis), the precession of the orbit[219] and its eccentricity (the changing shape of the orbit), which modifies the distance of the Earth from the Sun. All these factors contribute to the climatic variations called ***Milankovitch cycles*** (section 5.1.7).

How much energy is retained by the Earth depends on the distribution of land and sea, the properties of the land's surface (reflective or absorbing) and the composition of the atmosphere (the much-discussed greenhouse effect and the ozone barrier). ***Albedo*** is a measure of the reflection, from 0 (none) to 1 (all).

The period when primates developed, the beginning of the Eocene epoch (55 Mya), was the warmest moment in the Tertiary and the warmth ostensibly spurred growth and evolution. Since the Eocene peak, global temperatures have been gradually decreasing, with short-term fluctuations superimposed on the general background. Figure 6.3 shows the general behavior that has been observed. Note that what is shown on the vertical axis is not the temperature, but a “proxy”, something which varies with the temperature.

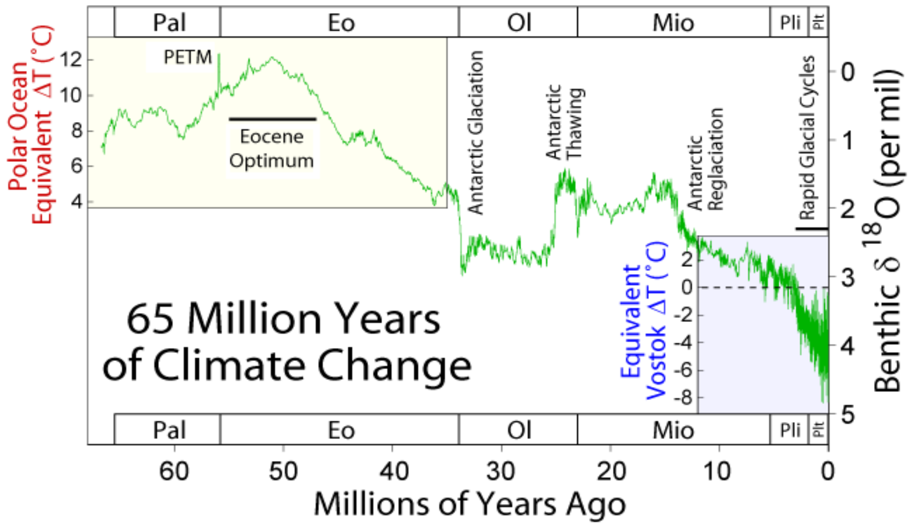


Figure 6.3. 65 million years of climate change, from Wikimedia Commons[220]

At the beginning of the Oligocene (33.9 Mya), a period of rapid cooling brought to an end the warmth of northern forests, with disastrous effects which almost wiped out our ancestral line.

As we have seen, these changes in temperature are to a great extent due to geology – the movement of tectonic plates. As plates have moved, oceans have opened (such as the separation between Antarctica and Australia or South America) or closed (Tethys Sea, Isthmus of Panama). This opening and closing of channels changed sea currents (e.g., the Gulf Stream) and led to formation of the Antarctic and Arctic ice caps[221], which in turn brought about lowering of global sea levels. The ice caps themselves reflect solar energy back into space, causing further cooling. Coming together of continents has created mountain chains (Africa pushed up the Alps; India, the Himalayas) which have altered meteorological conditions, especially rain patterns (such as the Asian monsoon). During the latter part of the ice age, melting continental ice sheets have caused sea levels to rise. Geology and climate and, hence, evolution all evolve together.

Figure 6.4 shows the general lowering of temperatures over the last 5 My, as well as the cyclic character of temperatures. The relative increase over the last 10,000 years began at the end of the last great Ice Age, which started some 130 Kya and only ended about 10 Kya. We are currently in a warm, interglacial period. There is no reason to expect this warmth to continue very long (on a geological time scale).

219 The elliptic orbit depends on two foci, one of which is at the sun. The other rotates slowly around the sun.

220 Diagram from Wikipedia, commons.wikimedia.org/wiki/File:65_Myr_Climate_Change.png, under GNU Free Documentation License. This is not temperature, but is indicative of it. See the site for explanation of the data and their interpretation.

221 This paragraph is only a summary, ignoring chronology. Formation of the antarctic ice cap coincided with the drop in temperatures at the beginning of the Oligocene, c. 35 Mya, whereas the Isthmus of Panama was closed c. 4-3 Mya and the arctic ice cap formed around 2.5 Mya.

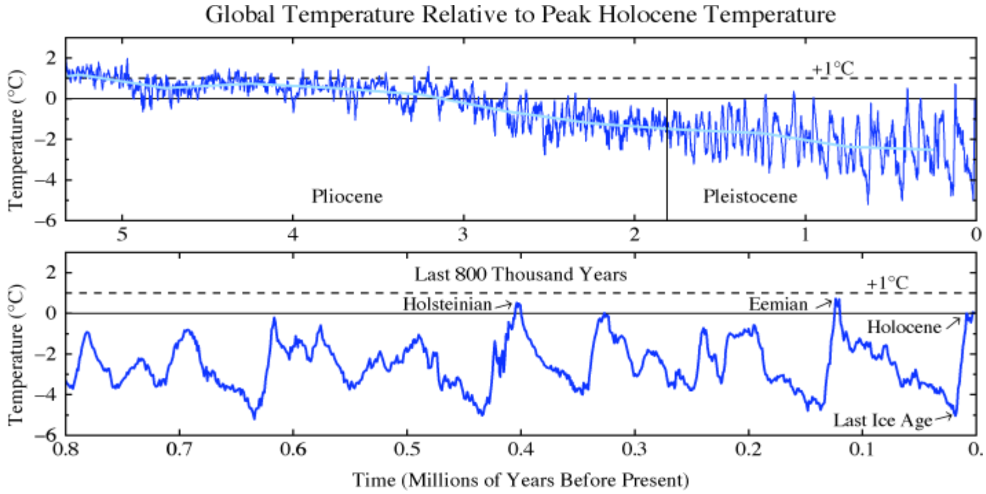


Figure 6.4. Global temperature over 6 My[222]

6.3.2. Rise of mammals and early hominins

Life appeared on Earth about 3.5 Gya. At first, bacteria dominated. In numbers they still do, but they are not alone. We will now take up where we left off after the K-T extinction, at the beginning of the first epoch of the Cenozoic Era.

Paleocene Epoch, 65 Mya

Although the earliest true mammals evolved during the late Triassic, they remained small and relatively inconspicuous until around 65 Mya, the time of the K-T extinction and the disappearance of the dinosaurs. It Is not clear how mammals managed to survive. One theory, based on the hypothesis that dinosaurs were killed by extreme heat after a huge asteroid struck the Earth, would have it that mammals stayed safe either in their burrows or under the sea until the worst heat had passed. In this case, the heat would have had to pass quickly. Perhaps a more likely suggestion is that particles in the air, whether from asteroid impact or volcanic activity, reduced incident sunlight so that photosynthesis was diminished. Animals which depended on plants for food (or on animals dependent on plants) died out, whereas those which ate organisms like insects or worms, which in turn fed on detritus, survived. Possibly both processes – and others – contributed. Be that as it may, after the dinosaurs died out, surviving mammals could creep out of their holes to occupy the old econiches as well as new ones. The number of mammals and mammal species underwent an extraordinary increase.

Interestingly, genetic and fossil evidence indicate that mammal species had already increased greatly before the K-T extinction, which allowed them to flourish afterwards.[223]

Mammals are characterized by:

- having differentiated teeth, i.e., teeth differently shaped to fulfill different functions in different parts of the mouth (e.g., incisors, canines and molars);
- being endothermic, or warm-blooded, which allows them to adjust their body temperatures according to external conditions, an advantage for adaptation to different climates;
- bearing their young alive (in most cases);
- producing milk to feed their young (in mammary glands);
- having fur or hair on their bodies.

Three orders of mammals which survived are still around today. ***Monotremes*** are rather rare egg-laying mammals, such as the platypus. Remote ancestors of mammals all laid eggs and these still do. ***Marsupials***, such as kangaroos and koalas, do not lay eggs, but their young are born underdeveloped and must be protected in the mother's pouch as they grow and develop. ***Placental mammals***, such as humans, protect

222 From NASA Goddard Institute for Space Studies, http://www.giss.nasa.gov/research/briefs/hansen_15.
223 Dawkins (2004), 180.

their young within the mother's body. They are the most divers and wide-ranging of contemporary mammals.

At the beginning of the Tertiary Period, climates were warm and forests spread across all the continents. ***Angiosperms***, plants fertilized by insects, grew everywhere. Plants, insects and animals evolved together. Among animals, one group, the ***archonta***, were the ancestors of today's bats, flying squirrels, tree shrews and primates.

Eocene Epoch, 55 Mya

During the warmth of the early Eocene Epoch, plant life abounded. Placental mammals and the first primates appeared. As forests developed, mammals evolved to inhabit them, ranging over what are now Europe, North America, Asia and Africa. As greenery spread, photosynthesis increased and therefore so did the amount of oxygen in the atmosphere.

There are numerous candidate fossils for the first primates. They include ***Purgatorius*** (Montana, North America, c. 65 Mya, at the end of the Paleocene), ***Altiatlasius*** (Morocco, c. 55-56 Mya), ***Teilhardina*** (North America and Asia, c. 56 Mya), ***Notharctus*** (North America, c. 50 Mya) and ***Eosimias*** (China, c. 45 Mya). ***Altiatlasius*** is generally considered to be the oldest known primate, even though only a few molars and a piece of jaw have been found. Early primates were small, squirrel-like creatures, but their paws could grasp and their eyes looked forwards, thereby giving them improved stereoscopic vision, so they were well able to walk or run or climb in trees.

Oligocene Epoch, 33.9 Mya

Eocene warmth gave way to cooling and subsequent formation of the Antarctic ice cap. Glaciations accelerated reduction of temperatures and absorption of water, bringing about increased aridity. Forests regressed, leaving grasslands behind. The necessity of adaptation to new conditions brought about the disappearance of fauna in higher latitudes and the appearance of new groups. Mammals became larger, sometimes huge. It was at this time that the first modern monkeys evolved in Africa or in Asia, exactly where being a matter of some controversy. One such creature, the ***Aegyptopithecus***, or Egyptian Monkey, lived in the forests of what is now Egypt around 35-33 Mya and possessed the gross anatomical traits of later monkeys. He weighed around 6 kg and lived mostly in trees, with opposable thumbs on all four feet, ideal for holding onto tree limbs.

In Western Europe, a sudden change in fauna known as the ***Grande Coupure*** involved the extinction of many species. The fossil trace of hominoids is then lost until the Miocene.

Miocene Epoch, 23 Mya

Around 23 Mya, an increase in temperatures was followed by a subsequent division of monkeys into two lines: One, the ***Cercopithecoidea***, now taxonomically considered a super-family, gave rise to the Old World monkeys (***catarrhines***) remaining in Africa and Asia today. The other was ***Hominoidea***, which gave rise to tail-less gibbons, orangutans, gorillas, chimpanzees and humans. (See Figure 6.1) Hominoids flourished, beginning what has been called the "Golden age of hominoids". One early hominoid, ***Proconsul***, 25-23 Mya, rather resembled a monkey but was definitely not one, being tailless and having a larger brain and ape-like teeth. Later, 17 Mya, ***Morotopithecus*** could move in a vertical stance, suspended from tree branches (***brachiation***). This technique may later have been transferred to the ground to give upright walking.

Several hominoid fossils have been found dating from around 14 Mya. One, ***Kenyapithecus*** sometimes called ***Afropithecus***), had thickly enameled teeth and may have been an ancestor of modern hominoids.

Subsequent, faster glacial fluctuations reduced the size of northern forests and brought about a rapid diminution of the relative numbers of tailless hominoids compared to cercopithecoids. In fact, hominoids barely survived; 8 Mya, only a few species remained. In Asia, only the ancestors of todays orangutans and gibbons made it. The only other hominoid survivors were in Africa and are dealt with in the next section. Today there remain only five hominoid species for 80 species of Old World Monkeys. Even in Africa, they were greatly outnumbered by cercopithecoids.

After *Kenyapithecus*, few fossil hominoids occur before the arrival of the australopiths, leaving a 7-My fossil-less period known as the “African ape gap”. Nevertheless, more fossils are turning up all the time. *Samburapithecus*, from 9.5 Mya, probably lived before the LCA[224], which the molecular-clock method dates to between 6 an 8 Mya.

6.3.3. Possible and probable hominins

Detailed descriptions of the various hominin species and their characteristics do not make for an exciting literary experience for many people, but this is unavoidable. There is much lack of certainty in the details of this domain, but a general evolutionary trend is perfectly clear.

As Figure 6.3 shows, the climate since c. 10 Mya has been one of cooling and therefore drying. Some forests have given way to grasslands, like present-day African savannas. At the time when today's hominin fossils were alive and thriving, the area was what paleontologists call a “mosaic” – a mixed regions of woods, savannas and riverside forests. So creatures living there would have benefited from living both in the trees and on the ground. So that is what they did.

If you prefer to ignore the details, skip from here to paragraph 6.4, “Homo sapiens”.

To date, four species have been discovered of which one may be the first hominin. All are somewhat controversial.

The oldest, ***Sahelanthropus tchadensis*** (also referred to as “Toumai”), dated about 7-6 Mya[225], lived in what is now Chad, which was then a mosaic environment and not yet a desert.[226] Only a cranium and several mandibles have been found. The cranium is no bigger than a present-day chimp's, with small brain and large brow ridges, but its sloping face, its small canines and enlarged cheek teeth arranged in the shape of a C, and the position of the foramen magnum (the relative importance of which is disputed in this case) suggest that it was a hominin, not an ape, and capable of walking upright or living in trees. If this is indeed true, it came (probably just) after the LCA, so the fact of its having lived west of the Great African Rift would tend to negate the thesis (“East Side Story”) that hominins evolved in eastern savannas where their upright posture, enabling them to see over the grass, could have been an adaptive trait. In that case, it would have come about as the result of tectonic processes whereby the rift led to the formation of the East African savannas. All hominoids, chimps and all, are capable of walking upright, but can not do so as fast or as far as humans.

Of ***Orrorin tugenensis***, who lived in Kenya about 6.2-5.8 Mya, only some teeth, a femur and some phalanges have been found. These fragments indicate that he could live well in trees but was capable of frequently walking upright. His teeth were thickly enameled like those of later australopiths, but smaller. The thick enamel may not be unique to hominin-like creatures and some paleontologists question even whether *Orrorin* was a hominin.

Two forms of ***Ardepithecus*** have been found. One, ***Ardepithecus kaddaba*** dates from 5.8-5.2 Mya and rather resembles a chimpanzee, but was capable of walking upright.

The more recent one, ***Ardepithecus ramidus,*** (“Ardi”) dates from around 4.4 Mya in what is now Ethiopia. Although his foramen magnum was somewhat farther forward than that of chimps, he still was more like a chimpanzee than a hominin, although he had a U-shaped jaw, unlike apes. A reconstruction of a crushed pelvis indicates Ardi was at home in trees or on the ground, probably more the former, as his hands were those of a tree dweller. Nor could he run long distances upright. He may represent an additional line of hominoids to pangins and hominins, or he might lie along the line to Australopithecines. Whether or not he was a direct ancestor of man is debated. We will shortly have more to say more about Ardi.

6.3.4. Archaic and transitional Homo

This rather large group includes the sub-tribe ***Australopithecina*** and its two genera, the earlier ***Australopithecus*** and the later ***Paranthropecus***. Although the skulls and overall size of australopiths were

224 The last common ancestor of chimps and humans, if you have forgotten.

225 Some authors cite slightly different dates. In general, we use those of the Smithsonian Human Origins Program, http://humanorigins.si.edu/research.

226 It goes without saying that the statement that a species “lived” in a place at some time means that its fossils have been found at that place and dated to that time.

chimp-like, their front teeth were relatively smaller and their dentitions starting to approach the parabolic shape of those of humans. Also, australopiths could walk upright. They therefore show both ape-like and hominin characteristics, putting them on the road from the former to the latter.

Paranthrops had huge teeth and massive jaws.

Australopithecus hominins

Australopithecus lived during a warm climatic period in a mosaic environment of forest and savanna in east and south Africa between about 4.2 and 2.5 Mya, a period of well over a million years. Members of this genus generally had well developed mandibles and teeth for eating tough roots and plants and they were used to walking upright as well as moving about in trees. Their brains were not significantly bigger than chimps' brains, although they may have been organized differently. There are at least five species considered to be members of the genus *Australopithecus;* they are a rather heterogeneous lot.

Australopithecus anamensis lived in current-day Kenya about 3.85-2.95 Mya. He could walk upright or move about in trees. His teeth resembled those of humans more than those of chimps. Opinions about him differ:

- Some paleontologists think *Au. anamensis* may be the ancestor of *Au. garhi* who in turn would be ancestor of *Homo*.
- Fossils bones of *Au. anamensis* also have been found in Ethiopia, where he lived at the same time as *Ardipithecus ramidus*. This suggests that he may have been a descendant of Ardi and ancestor of *Au. afarensis*, as is also suggested by his parabolic jaw structure.
- Others wonder if he is not simply the same species as *Au. Afarensis*.

The diverging opinions about *Au. Anamensis* are good example of the current state of uncertainty concerning hominid fossils.

Australopithecus afarensis lived in Ethiopia and Kenya roughly 3.7-3.0 Mya and so lived over a period of about a million years. Many bones of this species have been found, including the so-called “First Family”. With an ape-like face and cranium and long arms for climbing in trees, but small canines and definite bipedal capabilities in the knee and hip bones, she possessed a mixture of ape-like and hominin characteristics which enabled her to survive important environmental changes. Fossil remains of over 300 individuals have been found. Not only was she one of the longest-lived hominids, she is also the best known, because of Lucy. If there is a common “spokesperson” for australopiths in general and *afarensis* in particular, it is certainly Lucy, the 40% complete skeleton of a young *afarensis* woman found in Ethiopia. Her fame as a hominin fossil has made the tour of the world.

Figure 6.5. Lucy's skeleton[227]

Figure 6.6: Partial copy of Laetoli footprints[228]

Although Lucy could and certainly did walk upright, she certainly had an awkward, swaying gait. Her brain was somewhat larger than a chimp's and she probably used natural objects which she found, like sticks and stones, for tools.

Footprints of two hominins from the same period, about 3.6 Mya, have been found at Laetoli, in Kenya. The big toe of the walkers is in line with the foot and the walk is a heel-first-toes-last walk which suggests either modern humans or Australopithecus to different studies. Since the observed short stride corresponds to the short legs of *Au. afarensis*, some of whose bones have been found nearby in the same sediment layer, it is generally accepted that the prints are of that species. Even so, some scientists insist the footprints are those of Au. anamensis.

On the basis of these footprints, it has been suggested that bipedalism did not evolve from quadripedalism, but had always been possible and was instead lost by monkeys and others who became uniquely quadripedal[229]. The least one can say is that this idea does not seem to have caught on very much.

Another skull from about 3.5 Mya, which has been named ***Kenyanthropus platyops***, is controversial. One paleontologist thinks it is just a deformed skull of *Au. afarensis*; another, an ancestor of *Homo rudolfensis*. However, a recent discovery[230] of what would be the oldest known stone tool dates back to 3.3 Mya. It was found near the site of *Kenyanthropus*, so it may have been used by him, whatever he was.

Australopithecus africanus is similar to *Au. afarensis* in possessing both ape-like and human-like features, so much so that some paleontologists consider them to be the same species. It lived in east and south Africa 3.3-2.1 Mya. It had teeth more like those of humans than of australopiths, with smaller canines arranged in a semicircle. It had a flat face and the foramen magnum placed as for an upright posture. Its brain was bigger than that of Au. afarensis and it was both bipedal and arboreal. It is considered by some to be an ancestor of Homo.

Australopithecus bahrelghazali (called "**Abel**" by his discoverers) lived about 3.5-3 Mya in Chad, west of the Rift. Many paleontologists think he is just a geographical variant of *Au. afarensis*. His discoverer, of course, does not agree.

Of ***Australopithecus garhi***, who lived about 2.5 Mya in Ethiopia, only pieces of a skull have been found. Although a nearby partial skeleton may go with the skull, this is not yet proven. Although it might just be an

227 Cast from Museum national d'histoire naturel, Paris. Photo from Wikipdedia Commons, commons.wikimedia.org/wiki/File:Lucy_blackbg.jpg.

228 Replica in National Museum of Nature and Science, Tokyo. Photo from Wikipedia Commons, http://commons.wikimedia.org/wiki/File:Laetoli_footprints_replica.jpg.

229 Yvette Levoisin, "L'homme de descend pas d'un primate arboricole!". http://www.hominides.com/html/references/bipede-homme-primate-deloison.php

230 "Wrong Turn Leads to Discovery of Oldest Stone Tools", http://news.nationalgeographic.com/2015/05/150520-oldest-stone-tools-discovery-harmand-archaeology/

Au. afarensis or a female *Paranthropus aethiopicus,* it was considered to represent a separate species because of the previously unknown combination of a small brain and large molars. Bones found nearby indicate that a sharp-edged tool had been used to remove meat.

The most recent australopith, both for its life period and its discovery, is ***Australopithecus sediba***, who lived in south Africa between 1.977 and 1.98 Mya.[231] *Au. sediba* has certain details of its teeth, arm and leg length and upper chest like earlier australopiths, but other tooth traits and lower chest like humans. Its discoverers think it is descended from *Au. africanus*, and that it shows features more like Homo than like any other australopith. For them, it could help understand the transition from late australopiths to direct ancestors of humans, in which case it could link the origin of humans to South rather than East Africa. As usual, not everyone agrees. A study of its teeth finds it to be distinct from east African australopiths but close to south African *Au. africanus*. On the other hand, a study of its jaw finds it to be distinct from *Au. africanus*. Although it was bipedal, it had what is described as a hyper-pronating gait, meaning that its feet rolled inward at the end of each step.

So there are two lines suggested – and debated:

- one in east Africa (*Ardipithecus ramidus → Au. anamensis → Au. afarensis → Homo*);
- one in south Africa (*Au. africanus → Au. sediba → Homo*).

Paranthropus hominids

Though this group was originally thought to be australopiths, they are now considered to be a separate species because of their more robust frame, in particular, their chewing apparatus – jaw bone and muscle, and teeth. They are a more homogeneous group than australopiths – so far.

Paranthropus aethiopicus lived 2.7-2.3 Mya in East Africa. A jawbone and a skull have been found, the latter with a protruding face, strong jaw and well developed sagittal crest. But its most striking feature is a set of huge, thickly-enameled megadont teeth in a powerful jaw attached via large zygomatic arches to a sagittal crest in order to permit the chewing of tough, fibrous foods. This guy could eat really tough things like roots. Some paleontologists think he was a robust form of Australopithecus, maybe intermediate between *Au. afaransis* and *P. robustus*.

Paranthropus boisei (originally called ***Zinjanthropus boisei,*** "Zinj" for short) lived 2.3-1.2 Mya in East Africa. His skull has a massive jaw, megadont teeth – even bigger than those of *P. robustus* – and flaring cheekbones to hold his strong chewing muscles. He is often referred to as "Nutcracker man" and one study indeed finds that he ate nuts. His brain was bigger than that of his predecessors and increased gradually in size over time, as he flourished for about 1 million years. One hypothesis is that he evolved from *P. aethiopicus*. He is generally considered to be a side branch of our evolutionary tree because he lived in east Africa at the same time as the first Homo species.

231 "Australopithecus sediba – new analyses and surprise", Smithsonian Human Origins Program,

Figure 6.7. Skulls of Au. Boisei and of Homo habilis. Both skulls photographed by author at Olduvai Gorge Museum, Oct 2012.

Paranthropus robustus, lived 1.8-1.5 Mya in South Africa. He had an imposing, wide face, with large zygomatic arches, a sagittal crest and robust, almost megadont jaws for chewing tough fibrous foods. He may have been the user of bone tools found nearby.

6.3.5. General appearance of Australopithecus and Paranthropecus

If you met an Australopith in the street, you would wonder why he was loose. Even if he were wearing a suit, you would probably call the nearest zoo or circus to inform them that one of their stars had escaped. They rarely reached 1.4m in height. Their brains were small and they were certainly covered in fur and had chimp-like faces with protruding muzzles. Even if the one you saw walked on two legs, you would have thought he was walking on his hind legs, because his “arms” more resembled front legs, with his hands pretty much like his feet. And his way of walking would probably make you wonder how long he had been down out of his tree – or whether you were out of yours.

Paranthropus specimens were slightly taller than early Australopiths. Still, if you met one in the street, you would take one look at his massive jaw and cheek bones and choose a different street.

6.3.6. Pre-modern Homo

A recently found jawbone dating from 2.75-2.8 Mya, initially referred to as the ***Ledi jaw***, after the Ledi-Geraru site in the Afar region of Ethiopia where it was found, pushes back the date of the oldest fossil of genus *Homo.*[232]

Homo habilis fossils have been found in East and South Africa and dated to 2.4-1.4 Mya. His brain was slightly larger (550-680 cm^3) than those of his predecessors. It has been claimed that the cranium shows evidence for a developed Broca's area[233], suggesting that he may have used spoken language to communicate, although Broca's area is hardly a sufficient condition for speech. His face and teeth were smaller than those of australopiths, so maybe he was more carnivorous, but his body was more ape-like. He may have used stone tools, but it is difficult to know if these were made by him or by other species in the same region. Whatever the case, the supposition that he had done so resulted in his being considered the first homo. Nevertheless, some researchers find little to distinguish between *H. habilis* and australopithecus. It also is not clear that *H. habilis* was an ancestor of *H. erectu*s, as the time periods of the two overlapped by a half million years.

Homo rudolfensis lived in East Africa around 1.9-1.8 Mya. His brain was larger and his face longer than those of *H. habilis*, but his chewing teeth were larger, more like those of Paranthropus. Some paleontologists think he was an *Australopithecus*; some others, the same species as *H. habilis*. No good example of his skeletal structure has been discovered. A recent find of a cranium and jawbones[234] supports the idea of *H. rudolfensis* as a *Homo*.

Homo ergaster (considered by many paleontologists to be early African *H. erectus*) lived 1.9-1.5 Mya in East Africa. He was the first hominin to have a body silhouette similar to that of modern humans and his proportions indicate that he lived on the ground. He was 1.7 m tall and capable of running and of walking long distances. His brain, at 850 cm^3, was bigger than that of *Paranthropus*, but not yet that of a modern human, one more piece of evidence that larger brain size followed bipedalism on the evolutionary time scale, rather than preceding it. It is easy to imagine *H. ergaster* as the first “naked ape”. H. ergaster used primitive tools of the “Olduwan” culture.[235]

232 “Oldest human fossil found, redrawing family tree:, http://news.nationalgeographic.com/news/2015/03/150304-homo-habilis-evolution-fossil-jaw-ethiopia-olduvai-gorge/.

233 See the neuroscience chapter.

234 “New fossils confirm diverse species at the root of our lineage”. Smithsonian Human Origins Program,http://humanorigins.si.edu/research/whats-hot/australopithecus-sediba-%E2%80%93-new-analyses-and-surprises. http://humanorigins.si.edu/research/whats-hot/new-fossils-confirm-diverse-species-root-our-lineage.

235 Tools will be discussed in more detail in section 6.5.

Homo erectus lived from 1.9 to 0.14 Mya (or even later), making him the longest lived human species to date. Modern humans are far from having achieved such a long species lifetime. Like *H. ergaster*, *H. erectus* had a habitually upright posture and his or, in this case, her wide pelvis would have permitted birth of larger-brained babies.

H. ergaster/erectus (hereafter referred to simply as *H. erectus*) needed good nutrition in order to provide energy to that enlarging brain, so he may have mastered the use of fire and cooked his food, thus making food digestible even with his relatively smaller teeth. He was generally carnivorous, and the richer nutritive value of meat meant he could have a shorter digestive tract, which in turn made energy available faster. Hunting and butchering were enhanced by the use of thinner, double-sided stone cutting tools, which the archaeological record shows started around 1.76 Mya – the ***Acheulean*** technology. Note that 1.9 Mya, *H. erectu*s coexisted with *H. rudolfensis*, *H. habilis* and *P. boisei*, and by .143 Mya, with *H. sapiens*.

H. erectus was the first species to expand beyond Africa, starting as early as 1.8 Mya. This was the first of many surges of expansion of hominins from Africa. Remains have been found in Asia ("Java man" in Indonesia, "Peking man" in China, Georgia), Africa and maybe Europe, although it is not impossible that some of these were different species. Migration will be discussed more later.

Several recent discoveries witness the work-in-progress character of paleoanthropology.

A skull discovered in 2001 from Dmanisi, Georgia, dated at 1.8-1.7 Mya possesses certain features of *Homos habilis*, *erectus* and *rudolfensis*, suggesting that they may not be distinct species[236]. This also is the oldest good fossil evidence for hominins outside Africa. Oldowan tools were found on the site.

Two more recent discoveries either clear up or confuse questions concerning *H. erectus*, depending on one's point of view. One is the discovery in Buia (Eritrea) in 1998 of a nearly complete cranium, some pelvic bones and two incisors dated to 1.4-0.6 Mya. The long, ovoid brain case, wide cheekbones, massive brow ridges and medium-sized brain (~750-800 cc, a preliminary result) make him look like *H. erectus*. However, the parietal bones of the cranium are claimed to represent a more modern trait. The discoverers see him as a link between *H. erectus* and *H. sapiens* and claim that the date of first *H. sapiens* morphology has been pushed back to around 1 Mya.

The second discovery comes from the Afar region of Ethiopia, where in 1997 a crushed skull was found in the Dakanihylo sedimentary layer, also dated to around 1 Mya. The reconstructed skull of this ***Daka man*** has a long, sloping forehead, massive brow ridges and brain case shaped rather like that of Buia man, giving him a resemblance to *H. erectus* specimens found at that time in far-away Asia. Since *H. erectus* first appeared around 1.8 Mya in Africa, the discoverers of Daka man claim this to be evidence that by 1Mya, he had become one single world-wide species.

From 1.8 to 0.7 Mya, during the Pleistocene, glacial cycles increased in intensity (Fig 6.4). It was during this period that *H. habilis* and *H. rudolfensis* died out. The paranthropi soon followed suit, leaving only the genus *Homo* to spread to three continents.

6.3.7. Modern Homo

Homo heidelbergensis, who lived between about 0.6 and 0.1 Mya, is considered by many paleontologists to be the first modern human. Fossils attributed to this species have been found in Germany, Greece, Ethiopia, Gambia the U.K. and Spain, although he certainly originated in Africa. Some paleontologists think some specimens represent different species, such as *H. antecessor* (from Spain), *H. cerpanensi* (Italy) or *H. rhodesiensis* (Zambia). His large brain capacity of 1000-1300 cm^3 confirms him as a Homo, although he had very large brow ridges and a flat face. He used fire and wooden spears, hunted large animals and built shelters of wood or rock. Some paleontologists think he was our ancestor, but wonder who was his. Many consider him to be a transition species between *H. erectus* and *H. neanderthalensis*.

Neanderthals

Homo neanderthalensis lived about 400-40 Mya over a large region extending from Western Europe into

236 "Complete skull from Dmanisi". Smithsonian Human Origins Program, http://humanorigins.si.edu/research/whats-hot/complete-skull-dmanisi.

Asia, but concentrated mostly in Europe and the Near East. His body was shorter and more robust than that of modern humans, well adapted for cold, mountainous environments. His brain was even larger than ours, 1500-1750 cm^3, but he weighed more, so his encephalization (ratio of brain mass to body mass) was similar to ours. Neanderthals hunted large animals but also ate plants. They controlled fire, used sophisticated tools, lived in shelters and wore clothes they made themselves. Indeed, it is now thought that Neanderthals participated in about the same activities as *H. sapiens*. Discovery first of a Neanderthal hyoid bone (a throat bone necessary for enunciation) and then of his possessing the FOXP2 gene required for speech and language indicate that he may well have used language. Rock paintings found in Spain and dating from 43.5-52.5 Kya show that Neanderthals were artistic.

Much has been written about *Homo neanderthalensis* – entire books, including at least one novel sympathetic to him (William Golding, *The Inheritors*). Neanderthals are usually portrayed as ugly, ignorant and primitive and lacking in culture. They coexisted with *H. sapiens* for over 100,000 years and, in comparison with them, they were primitive and ignorant. But "ugly" is is in the eye of the beholder. And culture they had, though it is debated as to how much. There is evidence that they buried their dead and scratched designs into shells, an early example of art. It has recently been discovered that intercourse did take place between Neanderthals and the ancestors of modern humans, as modern Europeans and their descendants have about 1-4% of their genes from Neanderthals. They therefore might be considered the same species as *H. sapiens* in spite of morphological differences. That would make them *Home sapiens neanderthallens*is, as compared to our designation as *Homo sapiens sapiens*. And this raises the question of what is a species.

The Earth was now subject to Milankovitch-cycle glacial periods which came all the way down into northern continental Europe (Figure 6.4). During interglacial periods, such as those 500 and 400 Kya, Neanderthals penetrated northwards. During those of 320 and 220 Kya, they made it as far north as England and Wales. Between 120 and 70 Kya, they advanced as far as Siberia. But as the Earth became cooler after 100 Kya, they were forced to move south again. Their last holdouts were in Croatia, Russia and Gibraltar, not later than 28 Kya.

Neanderthals evolved during their long stay on Earth, as did their contemporary African Homos. Morphological traits distinguishing Neanderthals from modern man are more and more accented, the farther west they are found. Genetic studies of Neanderthals from across Eurasia suggest distinguishing three different groups according to where they lived – western Europe, the Mediterranean and the East. So as modern man moved west from the Middle East, he met populations showing more distinct, because more developed, Neanderthal morphology. Neanderthal extinction followed the same East-West gradient, those in the east disappearing before those farther west.

Evolution and variation within a species complicates distinguishing the species from others. So stating that such a fossil is such a species is rather like taking a snapshot of a moving object.

Theories abound as to what brought about the demise of the *H. neanderthalensis* species. The principle accused are the rapidly fluctuating climate, competition for resources and physiological differences from their fellow man, *H. sapiens*. Opinions on this question are often influenced by species identification or political correctness. The jury is still out.

Denisovans

Less well known is the recently found *Denisovan* species, discovered only in 2008. First, a young girl's finger bone was found in the Denisova cave in Southern Siberia. Since then, a toe and some teeth have also been found. While this may not sound like much, the important thing is that scientists have been able to make analyses of both nuclear and mitochondrial DNA of the species. Denisovans were found to be genetically closer to Neanderthals than to Homo, but distinct enough to deserve being considered a separate species. More recent genetic studies suggest that Denisovans and Neanderthals had a common origin about 1 Mya which may have been *H. heidelbergensis*. Denisovans and Neanderthals then diverged around 640 Kya, after leaving Africa.

In 2018, the right half of a human mandible which had been lying around in a Chinese museum was found by genetic analysis to be Denisovan also. This jawbone, the largest Denisovan fossil yet found, was discovered in 1980 by a meditating Buddhist monk in a high-altitude cave. Uranium dating attributes an age of about 160 Ky to it. The cave where it was found is situated near the city of Xiahe on the Tibetan plateau at an altitude of 3300 m, the highest ever for a fossil hominin. It is thought that Denisovans may have given present-day Tibetans a gene which facilitates living at high altitudes.

Denisovan genes have been found in the Melanesian people of Papua, New Guinea. Apparently, the people who migrated there first shared genes with the Denisovans before moving on to New Guinea 45 Kya. And Denisovan genes make up around 1% of the genes of modern Europeans.

6.4. *Homo sapiens*

6.4.1. Bipedalism, temperature and a larger brain

To simplify only a little, all the characteristics of modern humans are due to two main traits: bipedalism and bigger brains. That is the correct chronological order, as small-brained Australopiths were already bipedal to some extent and even completely bipedal *H. ergaster* had a brain of only about 850 cm^3. In some ways, bipedalism provided conditions necessary for an enlarged brain.

Bipedalism led to a non-grasping foot, simplified ankle and knee joints, a narrow, vertical and bowl-shaped pelvis (to support innards), related modifications to the hip joint and femur (the thing old folks break so easily), and a vertical, S-shaped spine (which pains many of us). It also freed up the hand, which could then develop other skills, such as making and using tools. Upright walking is thought to have brought modifications in the spatial relations of throat components (pharynx, larynx) necessary for speech, but keeping us from breathing and swallowing simultaneously.

Bipedalism and temperature regulation in the body are so associated that they probably developed together. Under the hot African sun, being upright meant that less light hit the body during the hottest time of day. So fur could disappear (except on top of the head), probably starting with H. ergaster.[237] Moreover, the absence of fur allowed improved temperature control by sweat glands in the skin coupled with blood circulation, together constituting a natural heat pump capable of cooling or warming the body depending on external conditions. For instance, vasodilation increases blood flow to carry off heat, as does sweating through condensation. Shivering, on the other hand, is rapid contraction and decontraction of muscles and this generates heat (and entropy). These processes allow maintenance of the strict temperature range needed by enzymes responsible for metabolic processes. Since a significant part of this is controlled by the brain, bipedalism indirectly profited from a bigger brain. Also, upright walking requires less energy than moving about on all fours, leaving more energy for the brain. And since the upright posture allowed faster movement, men became better hunters and were able to obtain more meat protein, which provided more energy, from which – again – the brain gained. Many of these features reinforced each other.

In addition to an improved nervous system and cognitive ability, a larger brain contributed to changes in the face and skull structure. Food not only nourished the larger brain, it also played a roll in the evolution of the jaw and teeth needed for mastication. The chin, unique to *H. sapiens*, attaches muscles used for fine lip movements necessary for speech. The brain developed a basic language function, constituting a default form of grammar found the world over.

Bipedalism also is important because before man could move "out of Africa", he had first to move out of the trees! Otherwise, he could never have crossed the very different landscapes which he encountered – forests, grasslands and sometimes snow or even seas.

So changes in posture, internal organs, brain size and interaction with the environment followed one upon another in a continual evolution towards our present (temporary) state.

6.4.2. Where, when and how?

To answer the first two questions, where and when, the oldest fossil evidence of *H. sapiens* is from Morocco and dates from around 300 Kya.[238] Together with the next-oldest known, from Ethiopia about 200 Kya, they indicate that "… the evolutionary processes behind the emergence of H. sapiens involved the whole African continent…" and not only East Africa, as once supposed.

237 Picq (2013), 107.

238 Hublin et al. New fossils from Jebel Irhoud, Morocco and the pan-African origin of Homo sapiens. https://www.nature.com/articles/nature22336

Fossils, archaeology (tools and art), genetics and language studies confirm (with some reservations) the "Out of Africa" model, which holds that hominoids all developed in Africa and then expanded to the rest of the world. The alternative model, the "multiregional" hypothesis, posits that hominoids migrated long before and then developed into local variants on site. In fact, both happened, with speciation taking place both before and after migration from Africa into Eurasia. A number of migrations – perhaps many -- took place, such as the one by *H. erectus* around 2 Mya and the more recent ones around 60-45 Kya by modern humans. There is "very little doubt" that we all are descended from this later migration.[239] It seems that when modern humans, or archaic *H. sapiens*, wandered out to the rest of the world, they met and to some extent mixed genes with locally variant species evolved from the earlier migrants. This interbreeding explains the small percentages of Neanderthal and Denisovan genes in those of modern humans.

The process is illustrated in Figure 6.8.

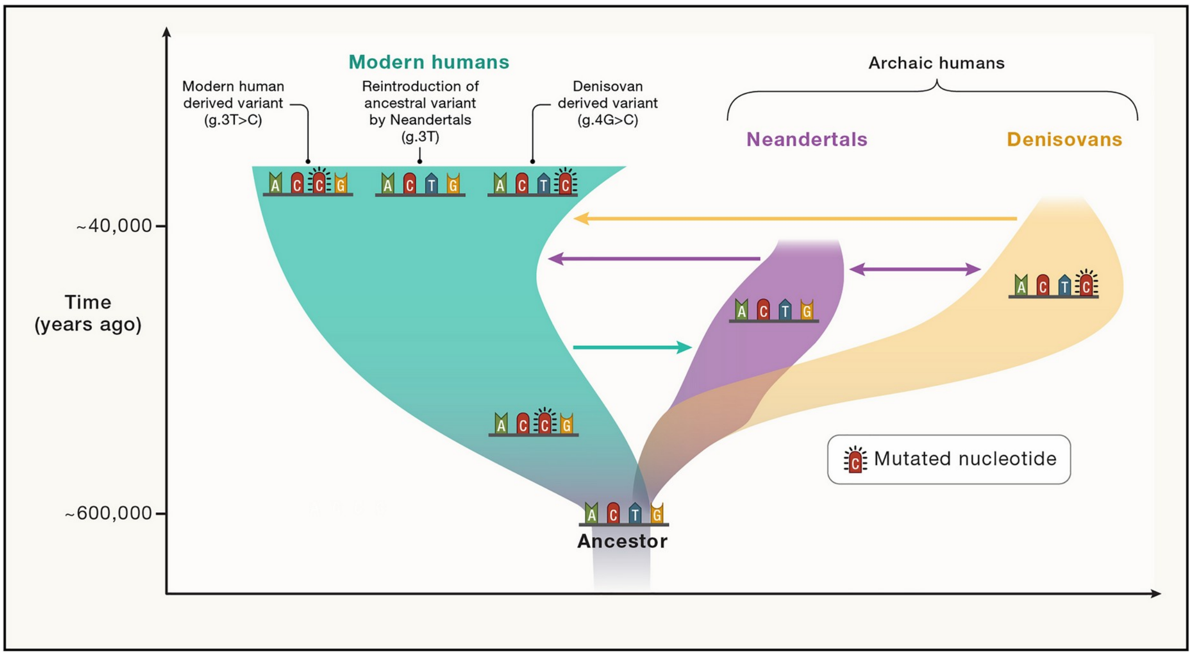


Figure 6.8. Schematic illustration of the history of archaic and modern humans and DNA sequence evolution. From Zeberg et al.[240]

As usual in paleontology, there is dispute about which was the first modern human fossil to be found. Modern Homo bones known as the "Red Lady", because they were smeared with ochre, were discovered in Wales in 1822-23. But, deserving or not, ***Cro-magnon*** man generally wins the prize, regardless of whether he actually was found in a *cros* (occitan for *creux*, shelter) on the farm of Mr Magnon[241] in south-central France in 1868. These early *H. sapiens* were more robust than modern humans, but otherwise resembled them closely. Actually, some of them had bigger brains than we do.

239 J. Coyne. Evidence that modern humans left Africa much earlier than we thought. https://whyevolutionistrue.com/2024/07/12/evidence-that-modern-humans-left-africa-much-earlier-than-we-thought/

240 Zeberg, Jakobsson and Pääbo. The genetic changers that shaped Neanderthals, Denisovans, and modern humans. ScienceDirect. https://www.sciencedirect.com/science/article/pii/S0092867423014034#:~:text=When%20modern%20humans%20started%20spreading,with%20Neandertals14%20and%20Denisovans.&text=As%20a%20result%2C%20all%20people,variants%20that%20come%20from%20Neandertals.

241 Maybe. Could also come from Cròs-Manhon, where Manhon is either occitan for "big" or a proper name.

Figure 6.9: Painting in the Grotte de Lascaux, by Prof saxx via Wikimedia Commons[242]

Fossil remains of H. sapiens have been found in Romania from 35 Kya; southeast Asia, maybe 40 Kya; and in the New World in Alaska, c. 12 Kya, and the Clovis Culture in North America, 11 Kya. Evidence for the existence of pre-Clovis cultures in Pennsylvania dating from 14 and possibly up to 20 Kya is controversial. (See section 6.6.4)

By 40 Kya, Homo's skill at tool-making had increased to the point where he began to make works of art – cave paintings and engravings, and carved bones.

6.5. More about tools – the Paleolithic

Establishing a chronology for tool fabrication and use is difficult for at least two reasons. First, the only tools which have remained over time are the hard ones made of stone or fossilized bone. Tools made of softer materials such as wood or bone have not survived. Second, finding tools near a fossilized hominin remain does not necessarily prove that the tools were made or used by that particular hominin. Still, we must do our best with the facts available.

The *Paleolithic*, or *Old Stone Age*[243], is taken as extending from the first known appearance of stone tools about 3.3 Mya in Kenya[244] and extending to the end of the last Ice Age about 10 Kya. It is divided into three periods.

- The first is the Early Paleolithic (also called the Lower Paleolithic, because the corresponding geological layer is located below the others), which is divided into two overlapping periods according to prevalent technologies.
    - The first period, the ***Oldowan*** (after Olduvai Gorge), extended from at least 2.6 Mya to about 1 Mya. During this time, hammerstones were used to knock sharp-edged chips off core rocks to make choppers. They were made and used by late australopithecines and maybe by paranthropus and *H. ergaster/habilis*. It was also during this time that the first expansions of Homo from Africa took place.
    - The second period, from 1.7 Mya to about 250 Kya, was that of the ***Acheulean*** technology, which spread from Africa into the Middle East and on to India, south of the Movius Line (explained below). This technique made thinner, double-sided chips to use for such items as hand axes. Such technology required planning by the toolmaker and therefore augmented brain

242 Wikimedia Commons. https://commons.wikimedia.org/wiki/File:Lascaux_painting.jpg

243 In this document, we will generally ignore geographically-based chronological differences between, e.g., Lower Paleolithic in Europe and Old Stone Age in Africa.

244 Before 2018, the oldest was 2.6 Mya in Ethiopia.

power. Hominins of the period were *H. erectus* and, later, *H. heidelbergensis*.

Oldowan chopper from Ethiopia, by Didier Descouens[245] via Wikimedia Commons[246]

Biface from Saint Acheul, France, from Wikimedia Commons[247]

- During the Middle Paleolithic, about 250-30 Kya, the ***Mousterian*** technology introduced the technique of making fine flakes of stone which could be attached to sticks to make spears. Fire came into general use. Principal hominins were Neanderthals and earliest modern humans.
- The Late Paleolithic, about 40-10 Kya, saw the use of bone, antlers and ivory to make still finer tools such as needles or harpoons. Hunting and fishing thus improved. Symbolic art, musical instruments and throwing devices dating from this period were made and used by anatomically modern humans. The oldest known musical instruments are bone flutes from 35 Mya, but most likely there had been previous instruments of less survivable material.

Figure 6.10: Aurignacian bone flute, by José-Manuel Benito Álvarez via WIkimedia Commons[248]

As mentioned, examples of *H. erectus* have been dated in Asia up to around 2 Mya. The tools found with them were those of the Oldowan technique. Acheulean tools date back to to 1.76 Mya and have been found only south of a line running from present-day Denmark to the Gulf of Bengal, the so-called ***Movius Line***, named after the paleontologist who first noted this distribution. The explanation for this geographical anomaly is generally agreed to be that *H. erectus* first took his Oldowan tools with him when he migrated east before around 2 Mya. Later, the Acheulean developed in East Africa (even though it is named after a site in France where it was first discovered) and subsequent migrations carried it to most of the rest of Africa and the Near East (Georgia), from where it moved east into India and northwest into Europe around

245 https://commons.wikimedia.org/wiki/User:Archaeodontosaurus
246 https://commons.wikimedia.org/wiki/File:Pierre_taill%C3%A9e_Melka_Kunture_%C3%89thiopie_fond.jpg
247 https://commons.wikimedia.org/wiki/File:Biface_de_St_Acheul_MHNT.jpg
248 https://commons.wikimedia.org/wiki/File:Flauta_paleol%C3%ADtica.jpg

600 Kya.

Similar considerations hold for the arrival of modern man in Europe. *H. erectus* seems to have arrived in southern Europe over 1 Mya; chipped stone tools have been found from 1.2 Mya. Although the Acheulean technology originated in Africa around 1.4 Mya, there are no examples of it in Europe from this time, This fact is explained if we accept that early hominins in Europe descended from those in Georgia 1.7 Mya, as this was a logical station between eastern Africa and Europe.[249] Because of glacial periods which affected Europe over this period, there were most likely numerous attempts at settling during warm periods, many of which failed due to climate extremes. The settlement of Europe was not a one-time event. Only for the last 600 Kys has northern Europe (north of the Alps) been permanently settled and these settlers brought the Acheulean technology with them.

In the cases of both Asia and Europe, absence of Acheulean tools for the earlier peoples leads to the recognition of different waves of migration across hundreds of thousands of years.

6.6. Ancient-population genetics, language, culture and migrations[250]

Since the initial sequencing of the human genome in 2001, rapid advances in gene-sequencing technology and breakthroughs in isolation, filtering and analysis of DNA from ancient bones have led to the sequencing of many truly ancient genomes. Previously-held theories about human origins and migration have had to be modified or abandoned and new ones have been suggested. Not only has DNA from previously known human forms such as Neanderthals been sequenced, at least one new form has been discovered, the Denisovans.

The first really big surprise was the discovery that the DNA of modern humans contains small amounts of DNA from Neanderthals and Denisovans. We are related, which is why it is not clear whether the word "species" should be employed here.

6.6.1. Out of Africa – several times

Evidence from paleontology, archaeology and genetics all concur that modern man, Homo sapiens, originated in Africa around 150-200 Kya from a line that split from that of chimpanzees and bonobos about 5-6 Mya. During that period, pre-modern Homo began to push out from Africa to the rest of Eurasia.

As was already determined by archaeological studies, *H. erectus* was the first humanoid species to expand out of Africa, before or about 1.8 Mya. But the mingling of modern humans with two now-extinct species modifies somewhat the out-of-Africa hypothesis, which says that all modern humans have evolved from a migration out of Africa around 60-50 Kya. The mingling with far older Neanderthals and Denisovans contradicts that notion, in the detail if not in the basic idea. It also shows that some of the branches of the tree model have mingled together briefly.

Analysis of ancient DNA has suggested the following sequence of events, correlating skeletal and genetic evidence[251]:

1) At least 1.8 Mya, a first expansion of humans to Eurasia. *H. erectu*s. (Skeletal data, Dmanisi, Georgia.)
2) Around 1.4 to 0.9 Mya, a split of the lineage of modern humans, Neanderthals and Denisovans from a "super-archaic" lineage whose presence is proposed in order to explain an observed characteristic of Denisovan DNA.[252]
3) At 770 to 550 Kya, split of ancestors of modern humans from those of Neanderthals and Denisovan.
4) At 470 to 380 Kya, split of ancestors of Neanderthals from those of Denisovans.

Although these dates may change or be clarified by further refinements in methods, their sequence is

249 Hublin, 95.

250 This history of human migration is severely condensed. It's predominant source is David Reich's fascinating book, *Who we are and how we got here,* chapters 3-6. Interested readers are encouraged to consult Reich's book.

251 This whole section based on Reich, especially his chapter 3.

252 Reich, 67.

expected to be correct.

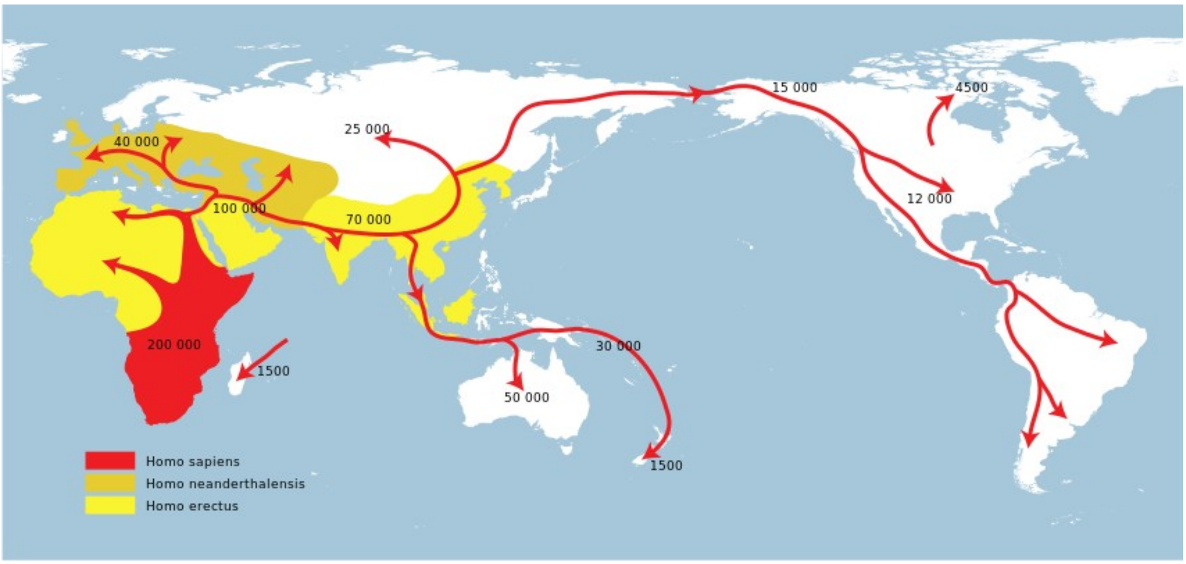


Figure 6.11. Spreading H. sapiens, from NordNordWest via Wikipedia.[253]

All these events were discovered from genomes extracted from evidence found in Eurasia. Whether these events took place in Africa, followed each time by a migration into Eurasia, or in Eurasia, followed by one unique movement back into Africa, sometime around 300 Kya, is not known, although the former hypothesis is favored by the majority of scientists. The latter hypothesis would agree with the suggestion that the so-called *H. antecessor*, from about 1 Mya and discovered in Atapuerca, Spain, could be an ancestor of both Neanderthals and modern humans, as he exhibits traits of both. In either case, archaeological evidence shows that Neanderthals were evolving in Europe (Spain) around 430 Kya.

It has long been considered that Neanderthals evolved outside of Africa, as no evidence for them has been found in Africa. Nevertheless, the Neanderthal Y-chromosome has a greater resemblance to that of living humans than to the rest of the Neanderthal genome. Recent research has found evidence interpreted as indicating the presence of "...tiny pieces of Neanderthal-like DNA scattered across all 12 of the populations [across Africa] they studied."[254] Their interpretation of their data is that c. 250 Kya, Africans with this Neanderthal-like DNA migrated into Eurasia where they interbred with the Neanderthals already there. According to this scenario, it was the Neanderthals who inherited the DNA from the newly-arrived Africans. The surprise is the early date for a migration of modern humans, posited in order to account for the Neanderthal Y-chromosome.

A mathematical analysis, in 2002, of 13 haplotypes across large genetic studies had already concluded that there were three major migrations out of Africa -- and one back in:[255]

- The H. erectus migration of around 1.7 Mya, supported by fossil evidence as well as archaeological remains from the Acheulean technology.
- The most recent migration of 150-80 Kya (according to this study) attested by chromosomal (mitochondrial and Y) evidence.
- An intermediate migration around 420-840 Kya.
- The study also finds there was a major back-migration from Asia into Africa around 50 Kya.

There are still questions concerning these models of human evolution, as well as other models. Study continues. And news insights are expected.

Humans may have reached Australia 60 Kya. But the migration which would have the greatest effect, of *H.*

253 Wikimedia Commons, commons.wikimedia.org/wiki/File:Spreading_homo_sapiens_la.svg.

254 J. Coyne. Evidence that modern humans left Africa much earlier than we thought. https://whyevolutionistrue.com/2024/07/12/evidence-that-modern-humans-left-africa-much-earlier-than-we-thought

255 Templeton, "Out of Africa again and again." www.nature.com/articles/416045a. Also diagram at cogweb.ucla.edu/ep/Templeton_02.html. Cited by Dawkins (2004), 60.

*sapien*s from Africa, took place around 70-50 Kya. Modern humans mixed genes with Neanderthals around 54-49 Kya[256], and with Denisovans around 49-44 Kya. Around this time, 45-40 Kya, modern humans expanded into western Europe from the Near East, which a glance at a map shows to be an obligatory transit point en route by land from Africa to Eurasia.

Eruption of a supervolcano, Campi Flegrei, near Naples, about 39 Kya spread ash over a vast area of perhaps 3.7 million square kilometers. Most modern human cultures existing before the eruption were gone afterwards. This includes the Neanderthals, but the role of the volcano in their destruction is debated.[257] So the first migration of modern humans into western Eurasia came to naught.

6.6.2. Later movements in Eurasia

The next migrants to spread to Europe, probably from the Near East, were more successful. They remained there as hunter-gatherers for 20 thousand years, over the period 37-14 Kya. Archaeologists identify them with the ***Aurignacian culture***.

Around 33-22 Kya, an eastern European group moved back west, bringing with them the ***Gravettian culture***, including statuettes of voluptuous women referred to today as "Venus", rudimentary flutes and amazingly beautiful cave art.

These *Homo sapiens* replaced the Neanderthals and *H. erectus* and, about 16-13 Kya, they made it to the Americas via Siberia. Man became the dominant species on Earth, for better and for worse.

Evidence indicating genetic affinity between some Europeans, such as the French, and Native Americans, has led geneticists to propose the existence of a culture referred to as the Ancient North Eurasians, since overlaid genetically by other mixtures. Discovery of the fossil of Mal'ta boy, from about 24 Kya near Lake Baikal in Russia, confirmed that hypothesis. Archaeology corroborated a prediction of genetics.

Expanding glaciation pushed humans into southern Europe, and as it receded, people flowed northward, those from Spain, around 19-14 Kya, bringing with them the ***Magdalenian culture***. In Goyet Cave in Belgium, archaeologists have discovered remains from all three cultures – the Aurignacian, Gravettian and Magdalenian.

Meanwhile, back in the Near East, farming began in Anatolia (southeastern Turkey) and northern Syria about 12-11 Kya. Farmers soon expanded into Europe (from Turkey), East Africa (from Israel and Jordan) and into the steppe north of the Black and Caspian Seas (from Iran). From 6-4.5 Kya, they spread across Europe, leading to genetic mixing between the new arrivals and previously established peoples, although both cultures went on coexisting. The mixture, in which farmers predominated, led to the ***Funnel Beaker culture***.

At about the same time, farmers expanded from Iran into north and south India, where they mixed with indigenous populations ~9-4 Kya.

In the steppes, from 7-5 Kya, the arrival of peoples from Armenia and Iran merged to form the ***Yamnaya***. The Yamnaya are characterized by the use of the wheel and the horse as well as by their form of burial in kurgans, large mounds of earth. The Russian word for kurgan is Ямная, yamnaya, and so the people are called today. The horse and wheel enabled their herding, nomadic culture to spread across the Pontic steppe, between and north of the Black and Caspian Seas.

The situation 5 Kya in the steppe was one of a Yamnayan genetic structure, a mixture composed in about equal proportions of an Iranian-related population and an Eastern European hunter-gatherer population. In Europe, the genetic makeup was largely derived from farmers from Anatolia along with a minority component of indigenous European hunter-gatherers. This soon changed when the Yamnaya migrated en masse both westward and eastward.

The language of the Yamnaya originated perhaps from the Caucausus, between the Black and Caspian Seas, or south of this in Iran (Iranian being today an Indo-European language). It has been

256 Reich, 39. Where the mix took place is not known to any degree of certainty. Riech thinks it may have been in the Near East. Reich, 41-2.

257 "Supervolcano cleared in Neanderthals' demise", LiveScience, https://www.livescience.com/48432-supervolcano-cleared-in-neanderthals-extinction.html.

identified as ***Proto-Indo-European*** (***PIE***) and the Yamnaya region of the steppe is generally considered the *Urheimat*, the original homeland, of the PIE. Most Indo-European languages today possess a large vocabulary for wagons – axles, wheels, harness poles and the like. So the PIE cannot be much older than wagons, i.e., about 6 Ky. It was the Yamnaya who carried it into Europe and India. Curiously, the geographic components of its name designate the destinations of this language family, not its origins. Indo-European languages are the majority languages in Europe. Indeed, Europeans are mostly Yamnaya: "...the single most important source of ancestry across northern Europe today is the Yamnaya or groups closely related to them."[258] The period beginning about 4.9 Kya also saw the beginnings in Europe of the ***Corded Ware culture***, so named because cords were used to impress patterns into the soft clay of pots. In Germany, people of the Corded Ware culture have about ¾ of their ancestry from the Yamnaya, the rest from the descendants of the first farmers in the region. So linguistic, genetic and archaeological evidence concur in attributing the genetic linguistic and cultural composition of modern Europe to a mass migration from the steppe.

6.6.3. Peopling of the Indian subcontinent

Around 9 Kya, Iranian farmers migrated into India and mixed with South Asian hunter-gatherers to form the population of the Ancestral South Indians (ASI), of which the arriving Iranian farmers made up about 25%. As the Yamnaya expanded into India around 5 Kya, they mixed along the way with descendants of the Iranian farmers to form the Ancestral North Indians (ANI), composed of about half steppe people and half farmers. Indians today are all mixtures of these groups, dating to ~4-3 Kya. Indo-European speaking groups are more recent mixtures on average than Dravidian-speaking groups, who live more in the south, so later arrivals from the steppe apparently did not penetrate as far south as earlier ones. The ANI ancestry has been found to descend more through males, so the distribution of the two groups is compatible with the arrival of an Indo-European speaking people taking power ~4 Kya, about the time of the collapse of the Indus Valley civilization and the composition of the Rig Veda. This power has since been reinforced by the caste-related endogamy inspired by the Veda and later Hindu scriptures.

6.6.4. Migration into the Americas

Genetic separation of the ancestors of Native Americans from Siberians occurred about 23-20 Kya, but migration into the Americas had to wait. It was long thought that movement from Siberia into the Americas could only take place after about 13 Kya along an "ice-free corridor", attested by plant and animal remains. The need to explain findings of human remains at Paisley Caves, Oregon, from about 14 Kya, and from Monte Verde, Chile, older yet, led geologists and archaeologists more recently to realize that at least parts of a coastal route were ice-free after 16 Kya. These routes are compatible with a recent explanation of migration of Native Americans from Siberia in several waves, three or four according to some researchers. It remains true even now, though, that the number of waves of migration is still a hotly debated subject among archaeologists and geneticists. In any case, they all came from the same, common ancestral population in eastern Eurasia, hence their physical and genetic resemblance. This ancestral population has been called the First Americans. It is members of this group who "... have made a dominant demographic contribution to all present-day indigenous peoples in the Americas."[259]

The first wave came between 16 and 14 Kya and moved south along the west coast of both American continents, producing the populations whose remains have been found in Oregon and southern Chile. Presumably, more came after 13 Kya.

Native American languages are often divided into three groups, two groups being the Eskimo-Aleut languages spoken in the north of Asia and North America, and the Na Dene spoken in inland Canada and the southwest United States. The third group, "all the rest", is called Amerind. Genetic and linguistic data support the notion of three migrations, corresponding to the three language groups. But Amerind remains the basis of the languages of all three groups, composing more than half of each of the other groups – more proof of a common ancestral population.

A fourth wave composed of a "ghost" population named "Population Y" has been proposed, date of arrival

258 Reich, 119.
259 Reich, 176.

unknown, in order to explain apparent genetic links to Australians, New Guineans and Andaman Islanders, but this population has been put into question recently by the very researchers who first proposed it.[260] In any case, these genes, if they are real, are found only in Amazonia and no longer exist in unmixed form, suggesting that elsewhere they succumbed to genetic mixing with later arrivals.

6.7. Overall summary

After dinosaurs disappeared from the surface of the Earth about 65 Mya, the number and size of mammal species took off, eventually ranging in size from tiny mouse-like creatures up to elephants and whales. During a particularly warm period of the mid-Eocene, primates came into existence, at first small squirrel-like creatures. About 23 Mya, the primate line split into Old-World Monkeys (catarrhines) and hominoids and the latter split into hylobatids (gibbons and the like) and hominids. From hominids sprang pongines (orangutans) and hominines and the latter begat panins (chimps and bonobos) and hominins. The first hominins were the precursors of man, but not all of them were his ancestors. No direct line from the LCA[261] of chimps and hominins can be distinguished; the tree of life is rather a bush, with the branches hidden and many twigs becoming dead ends or even merging together again. However, there is clear evidence for over twenty species intermediate between the time of the LCA and modern Homo.

First, there were some fairly difficult-to-classify species found in East and Central Africa, dating from 7 to 4.5 Mya. Though the consensus seems to be that these represented steps in a generally man-like direction, they are all subject to controversy as to whether they are hominins or on another line. Some may have lived before the LCA.

Then, from 4.5 to 2.5 Mya, there evolved a fairly heterogeneous group called australopithecines, one genus of which was australopithecus. Although they possessed varying degrees of bipedalism and stronger chewing teeth, their brains were still about the size of a chimp's. Up to at least four of these species lived at the same time. A second group, which followed up to about 1 Mya was paranthropus, sometimes classed as a genus of australopithecines. These were more robust versions of australopithecus, with a chewing apparatus capable of masticating tough roots and nuts.

From around 2.5 Mya, true Homo, the same genus as modern humans, appeared on the scene. Members of these species tamed fire and invented cooking, made and used stone tools and hunted large animals. The first one to really look more or less like a modern human, *Homo ergaster*, had long legs, an upright body and could walk long distances and even run. Another (or maybe they were the same), *Homo erectus*, migrated out of East Africa and as far as Asia around 2 Mya, taking along Oldowan tool technology. More migratory surges took place until finally, around 60 Kya years ago, another, more modern species made their way into the Middle East and from there to Europe, Asia and, eventually, the Americas, this time with the Acheulean technology. Along the way, they mixed their genes with those of the local populations evolved from earlier H. erectus – Neanderthals and Denisovans and perhaps others. Nevertheless, most of our genes originated in Africa over ~60 Kya..

H. erectus inhabited the planet for about 2 million years. Will we?

260 “Our failure to find significant evidence of Australasian or Paleolithic East Asian affinities in any of the ancient Central and South American individuals raises the question of what ancient populations could have contributed the Population Y signal in Surui and other Amazonian groups and increases the previously small chance that this signal—despitethe strong statistical evidence for it—was a false-positive.” Posth et al., “Reconstructing the deep population history of central and south America”. https://reich.hms.harvard.edu/sites/reich.hms.harvard.edu/files/inline-files/PIIS0092867418313801_1.pdf

261 Last Common Ancestor

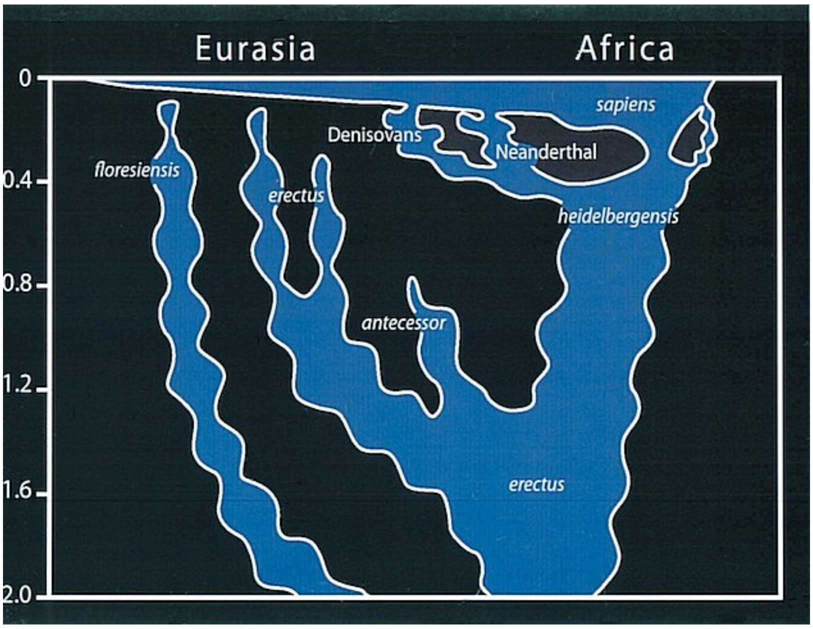


Figure 6.12. A tree diagram of human evolution based on both fossil evidence and DNA research, according to Chris Stringer, via Wikipedia.[262]

The spread of H. erectus as determined from both fossil and DNA evidence is diagrammed in Figure 6.12

The best-known of the Homo which became extinct is *Homo neanderthalensi*s. Neanderthals have been the victims of much bad press. Compared to the *H. sapiens*, with whom they shared the environment during their last 100,000 years or more, they were primitive. But they made and used tools and their own clothing, they probably buried their dead and they made decorations which may be considered an early form of art. And they left some of their genes in us.

The demise of the Neanderthals left, for the first time, only one species of hominin on Earth – *H. sapiens sapiens*.

6.8. Table of hominins

In the following table, (+) indicates more human-like; (-), less.

262 Homo-Stammbaum, version Stringer, https://commons.wikimedia.org/wiki/File:Homo-Stammbaum,_Version_Stringer.jpg.

Table 5: Principle hominin attributes

Name	Sa. tchadensis	Or. tugenensis	Ar. kadabba	Ar. ramidus
Date (Mya)	6-7	6	5.7-5.2	4.5-5.5
Place	C. Africa (Chad)	E. Africa (Kenya)	E. Africa (Ethiopia)	E. Africa (Ethiopia)
Found	cranium, jawbones	teeth, femur, some phalanges	jaw + hand, feet, toe, arm bones	skull, teeth, hands, feet, pelvis
Height (m)	1.15-1.25 (-)	1.15-1.25	uncertain, ~chimps	1.1-1.2
Weight (kg)	25-50	30-45	uncertain, ~chimps	50
Brain (cm^3)	310-370 (small)			
Canines	small (+)		small (+), but projecting (-)	small (+)
Molars	med. enamel	small, thick enamel (+)		med. enamel
Jaw		–		
Face	sloping (-), short (+)	–		less projecting (+)
Bipedalism	FM forward (+), prob. both	femur ==> both (+)	toe ==> both (+)	both (+)
Pelvis	–	–		crushed, hypothesis is bipedal
Hands	–	–		arboreal, but flexible wrists (+)
Feet	–	–		==> bipedal + arboreal
Environment	diverse forest + savanna			wooded
Tools				
Remarks	human and ape-like features			human and ape-like features

Name	Au. anamensis	Au. afarensis	Ke. platyops	Au. africanus
Date (Mya)	4.2-3.9	4.1-2.9	3.3-3.5	3.5-2.5
Place	E. Africa (Kenya, Tanzania)	E. Africa (Ethiopia, Kenya)	E. Africa	S. Africa
Found		>300 individuals, one >40% complete skeleton (Lucy)	only deformed skull	
Height (m)	1.4	1.05-1.35	1.15-1.40	
Weight (kg)	50	30-45	50?	
Brain (cm^3)		~450		450-530; developed parietal lobes
Canines	robust (-)	small (+)	small (+)	
Molars	large			
Jaw	robust			
Face	prognacious (-)	flat face (+), prognacious (-)	flat (+)	sloping, prognacious (-)
Bipedalism	more than later Au.	both (+)		both (+)
Pelvis		==> bipedal, swaying gait		==> bipedal
Hands		fingers ==> arboreal		==> arboreal
Feet		supposed made Laetoli footprints		==> bipedal
Environment	wooded	moist wooded	wooded savanna	wooded savanna
Tools		probably used found objects		
Remarks	ancestor of Au. garhi?	survived 10^6 years, hyoid bone found	controversial	

Name	Au. bahrelghazali ("Abel")	Au. garhi	Pa. aethiopicus	Au. sediba
Date (Mya)	3.5-3	3-2.5	2.7-2.3	1.98
Place	C. Africa (Chad)	E. Africa (Ethiopia)	E. Africa	S. Africa
Found	only lower jaw	cranium + skull fragments	jawbone, "black" skull	2 skeletons + much more[263]
Height (m)	1.0-1.2?	1.2	uncertain, 1.5?	
Weight (kg)	30-40?	?	uncertain, 50?	
Brain (cm^3)	–	small	~420	
Canines			small	
Molars		large	large	
Jaw				
Face			protruding, sagittal crest	
Bipedalism	–	larger femur ==> longer stride; bipedal (+)		previously unknown form of bipedalism
Pelvis	–			
Hands	–			
Feet	–			bipedal, hyper-pronating
Environment	wooded savanna	savanna		
Tools		may have used nearby tools		
Remarks	1st Au. west of RIft		Prob. intermediate A. afarensis -P. robustus/boisei	transitional Au. to H.

263 Some of the most complete sets of hominoid remains yet discovered.

Name	Pa. boisei (“Zinj”)	Pa. robustus	H. habilis	H. rudolfensis
Date (Mya)	2.4-1.2	1.8-1.2	2.4-1.6	1.9-1.8
Place	E. Africa (Ethiopia, Kenya, Tanzania, Malawi)	S. Africa	E. and S. Africa (Kenya, Tanzania, Sterkfontein)	E. Africa (Kenya, Malawi)
Found	skulls	jaw, teeth, other bones	skull, femur	
Height (m)	1.5-1.2	1.12-1.45	1.15-1.30	
Weight (kg)	30-55	25-45	30-40	
Brain (cm^3)	500-600		550-680	650-750
Canines				
Molars	megadont	robust		robust
Jaw	robust	robust		robust
Face		wide		
Bipedalism				
Pelvis	lower members ==> bipedal			
Hands			lower members robust ==> bipedal and arboreal	
Feet				
Environment			wooded savanna	
Tools			primitive stone tools, cirular huts	
Remarks	“Nutcracker man”		lived period increased climatic fluctuations	= H. habilis?

Name	H. ergaster	H. erectus	H. heidelbergensis	H. neanderthalensis
Date (Mya)	2-1	1.59-0.143	0.8-0.3	.35-.035
Place	Africa, S. Europe, Central Asia	E. and C. Asia, into Africa and Europe	Africa, Europe, W. Asia	Africa, Europe and W., C. and S. Asia
Found	many, almost complete skeleton	many fossils		
Height (m)	1.55-1.75	1.50-1.65		
Weight (kg)	50-65	45-55		
Brain (cm^3)	850	~800		
Canines				
Molars				
Jaw				
Face				
Bipedalism	long lower members ==> excellent walker and runner			
Pelvis				
Hands				
Feet				
Environment	savanna	savanna and forests		
Tools	used fire, made Acheulean biface tools	used fire, made Acheulean biface tools		
Remarks				

7. Annex A: Stellar evolution – the virial theorem

7.1. The virial theorem and equilibrium states in stars

For a bound system, one can define something called a ***virial*** as

$$G = \sum_i p_i \cdot r_i.$$

Then by taking the time derivative of this quantity, calculating averages and assuming suitable boundary conditions on G, one can prove the ***virial theorem***, which states that for a bound state of particles with mean total kinetic energy T and mean potential energy V,[264]

$$2T + V = 0. \quad (7.1)$$

The total energy of the system is then

$$E = T + V = -T. \quad (7.2)$$

That the energy be negative is logical for a bound system. This is easily verified for the case of a circular orbit. Equating the gravitational and centrifugal forces, $mv^2/r = GmM/r^2$, leads to

$$T = \frac{1}{2}mv^2 = \frac{GmM}{2r} = -\frac{1}{2}V \quad (7.3)$$

as in (7.1). Note that if the orbit shrinks a bit, so its radius decreases, the increased gravitational force means the orbiting body must speed up in order to resist the pull of the gravitational source. Shrinking the orbit increases the kinetic energy of the orbiting body, which is equivalent to increased temperature.

A star can be represented as a bound state of particles. As the star's fuel of hydrogen decreases, core pressure is reduced and the star shrinks some. Since this puts its particles closer together, its gravitational potential energy, already negative, increases in magnitude, becoming more negative. By the virial theorem, the gravitational energy lost causes T to increase by half that amount. But that leaves ½ΔV left over, which must be radiated by the star. The important point is the increase in internal kinetic energy, which means that when a star shrinks and loses energy by radiation, it actually gets hotter.[265] This is analogous to the increase in kinetic energy of the orbiting body as the orbit size decreases.

To remember: Because of the virial theorem, a shrinking star loses gravitational energy, half of which is radiated (star light) and half converted to kinetic energy, so the star heats up.

From a more physical point of view, that T be less than the magnitude of V says that the internal kinetic energy of the star is not enough to escape the gravitational forces holding it together. As long as the virial theorem holds, the star is bound by gravity. If the star shrinks a little, its temperature goes up and so does the pressure, which will cause it to expand some. Conversely, if the star expands a bit, its central temperature goes down and nuclear burning slows and reduces pressure.[266] In either case, the star automatically corrects the deviation from the ideal size and temperature. The star "swaps" energy back and forth between heat and gravity and remains bound. In this way, it auto-regulates its core temperature (or kinetic energy).

Mathematically, this inverse relation of temperature and size reflects the statement that the kinetic energy is proportional to the ratio of the mass and the size:

$$T \propto GM/R, \quad (7.4)$$

as in the orbital example (7.3). A more exact calculation shows that the temperature of the Sun's core is at about 10 million K.

The energy lost to radiation must be and is replaced by energy from nuclear fusion. Consideration of the rate

264 This statement lacks rigor. For a derivation and discussion, see, for instance, http://www.math.ucr.edu/home/baez/virial.html.

265 King, 23-25.

266 Yes, we are employing the terms temperature and kinetic energy interchangeably, also gravity and potential energy. This is imprecise, but reasonable.

of fusion via the proton-proton chain shows that the temperature range at which the reactions are able to take place is quite limited.[267] So main-sequence stars (***dwarfs***) all burn at a core temperature about that of the Sun – 10 million K – and therefore all share about the same ratio of mass to radius. Heavier stars are bigger stars. Given that a star's luminosity is a function of its temperature, this then fixes a relation between the star's luminosity and its mass, which turns out to be $L \propto M^3$ for a Sun-sized star.[268] This enables astrophysicists to calculate mass from observations of luminosity.

Note that supporting the star against collapse requires a net force of pressure away from the center, which means the pressure is greater closer to the center. Only in the core of the star is the pressure high enough for nuclear "burning" to occur. The "weight" of the star above a given radius increases as one approaches the center, just like water pressure in oceans increases with depth.

7.2. Stellar evolution

So main-sequence stars are bound by gravitation and use energy from hydrogen fusion to replace the energy radiated in order to follow the dictates of the virial theorem. This allows – rather, requires – them to maintain their size (radius) and therefore their temperature. But eventually, the hydrogen fuel becomes insufficient to maintain the needed pressure. In the center of the core is an increasing mass of ^4He, the product of hydrogen fusion, which does not yet undergo fusion. Because of the relation (7.4), at constant temperature the mass and radius are proportional and so density increases like M^{-2}, i.e., the density of the core decreases as the mass increases. Also, ^4He is heavier than H and it is the number of atoms used to calculate the pressure, not their mass.[269] As the density decreases, so does the pressure. When the star's fuel of hydrogen decreases to where the He in the core reaches about 10% of its mass, the pressure is too low to support the star's weight. Gravity starts to effectively cause the star to shrink, but the shrinking is slowed, either by exclusion-principle degeneracy pressure, for stars less than about twice the mass of the Sun, or by gas pressure, for larger stars.

The shrinking is slowed enough that the star remains close to hydrostatic equilibrium. Under these conditions, the twin requirements of the virial theorem and conservation of the star's total energy (T+V) necessitate the conservation of T and V separately. As the star contracts, its gravitational energy V increases, so it must lose some of that by expanding (the whole star). As the core contracts it becomes hotter, so conservation of T requires that the envelope around it become cooler. The star is now in an extreme imbalance of thermal and gravitational energy. Since gravity is stronger in the core than in the envelope (tidal forces), the envelope expands. The star expands enormously and the envelope cools. A shell of hydrogen around the core heats up and starts burning, increasing the star's luminosity. Since luminosity is given by

$$L = 4\pi R^2 \sigma T^4,$$

this also requires the star to become much larger. The star becomes cooler, larger and redder – a ***red giant***.

For more massive stars, the temperature and pressure within the core may continue expanding until ^4He starts to fuse into ^{12}C and ^{16}O, ***core helium burning***, which makes the He core heat up and expand. Now conservation of gravitational and thermal energy requires the core to shrink. The core goes through successive phases of fusion of ^{12}C and ^{16}O to ^{20}Ne, ^{23}Na, ^{24}Mg and ^{28}Si. Forming layer after layer of elements like an onion, with lighter elements outside (on top) and heavier ones within, the star warms and expands, cools and shrinks, dancing back and forth on the HR diagram, until the core contains only Fe and fusion comes to an end. By this time, the star will have expanded to become a ***supergiant***.

By the time it reaches supergiant status, much of the star's mass is in a huge envelope around the core. Since it is at a great distance from the center, it is not strongly bound by gravitation, so it is relatively easy for the star to lose mass by ejecting the envelope into the space around the cooling core. The result is a ***planetary nebula***, a roughly spherical cloud of dust and matter surrounding a cold core called a ***white dwarf***, held up only by electron degeneracy, i.e., the exclusion principle.

267 King, 41.
268 King, 43-44.
269 By the ideal gas law, PV = nRT, where n/V is the number density.

Once the star is bound by gravity, the virial theorem comes into play. It "...connects gravity with thermodynamics..."[270] and forces stars to grow through a series of evolving states to become red giants and then supergiants.

8. Annex B: LIPS, OAEs and mass extinctions

An understanding of how geology and biology together can influence life on the Earth can be gained by studying the period of the Cennomanian-Turonian (C-T) boundary in the mid-Cretaceous, 94-93 Mya, a time of significant extinctions of marine organisms.. A number of factors were at work.[271]

- Intense volcanic activity at various times and places in the history of the Earth has produced huge quantities of flowing lava called ***flood basalts***, for which geologists have coined the acronym ***LIPs*** (***large igneous provinces***). They may cover up to millions of square kilometers and be several km thick. LIPs are formed by relatively continuous, non-explosive volcanic activity, but nevertheless form quickly on a geological timescale, in less than a million years. They are thought to have formed over deep plumes of magma and so can form on land or under seas, independently of plate boundaries. One such LIP, far from the largest, is the Caribbean LIP of 94-93 Mya. It formed between the North American and South American plates, which of course had quite a different configuration than they do today. So though the hot spot which formed them is probably now beneath the Galapagos Islands, the LIP, mostly still underwater, lies mainly beneath the Caribbean Sea.
- During the Cretaceous, the amount of CO_2 in the atmosphere was abnormally high, bringing higher temperatures to the whole planet. Increased volcanic activity pushing up magma as the seafloor opened to form the Atlantic Ocean brought much CO_2 with it, overwhelming its rate of removal by rock weathering.
- ***Black shales*** are dark-colored Cretaceous rocks rich in carbon which form when large amounts of plant and animal life near the surface of the oceans die and descend to the sea floor. The shales only can form when the deep water contains no or almost no oxygen, at variance with the almost continuous oxygenation of seas since the Proterozoic as shown by core samples. The periods when such low-oxygen conditions hold are referred to as ***oceanic anoxic events***, or ***OAEs***. Evidence for the existence of OAEs has been discovered worldwide, in all oceans and also on land. They arise and disappear abruptly and are known to be associated with times of global environmental change. There were three OAEs during the Cretaceous and one of them was in the period 94-93 Mya, circumstantial evidence associating it with the Caribbean LIP.
- Isotope ratios of osmium (Os) in sedimentary rocks depend on whether the sediment comes from continental rocks or sea-floor volcanic activity. The data show that during the C-T OAE, osmium in sediments abruptly comes almost all from underwater volcanoes and not from continental weathering. The only candidate for the undersea volcanoes was the Caribbean LIP.

So the Caribbean LIP has been found to be strongly correlated in time with the strongest of the Cretaceous ocean anoxic events, with an onset of seafloor volcanic activity and with a time of significant extinctions of marine organisms. Evidence also shows that global temperatures were significantly higher at this time, by 3-8°C. Carbon-isotope studies in sedimentary rocks show substantial increases in atmospheric CO_2 at the time of each of the main LIPs mentioned below, so the main influence of the LIP was increased CO_2 from the volcanic activity and that brought about global temperature increases. It all seems to hang together. Extrapolating from the Caribbean case, it seems clear that LIPs can bring about major climate change leading to mass extinctions.

Other related factors are the following.

- So-called ***green sulfur bacteria*** live in the ocean by anaerobic photosynthesis, using sulfur dioxide (SO_2) as an electron source. Biomarkers for these bacteria therefore indicate the presence of sulfur dioxide. Their presence in black shales at the same time as the C-T OAE suggests a significant presence of CO_2 in the oceans, so some of the extinctions may have been due to this toxic gas.

270 King, 59,
271 Most of the following discussion is based on MacDougall 2011, 188-202.

- A recent study[272] indicates that the eruptions in the Siberian Traps increased the amount of nickel in the Earth's crust and this was a nutrient for a microbe, a methane-producing archaea called Methanosarcina, which had undergone a genetic change at about that time. It is suggested (claimed, even) that the microbe emitted vast amounts of methane into the atmosphere and so changed the climate.

272 “Ancient whodunit may be solved: The microbes did it!|” March 2014: MIT News, newsoffice.mit.edu/2014/ancient-whodunit-may-be-solved-microbes-did-it.

Figure Index