

---

## 2.1 Microbes Were First: Bacteria Have Existed from Very Early in the History of Life on Earth

With billions of years of evolution before the appearance of animals, prokaryotes shaped and continue to shape both the Earth's biogeochemical landscape and the setting for animal existence (Fig. 2.1) (Knoll 2003).

Bacteria inhabit every environment on the planet. Bacteria fossils discovered in rocks date from at least the Devonian Period, and there are convincing arguments that bacteria have been present since early Precambrian time, about 3.5 billion years ago (Fig. 2.2).

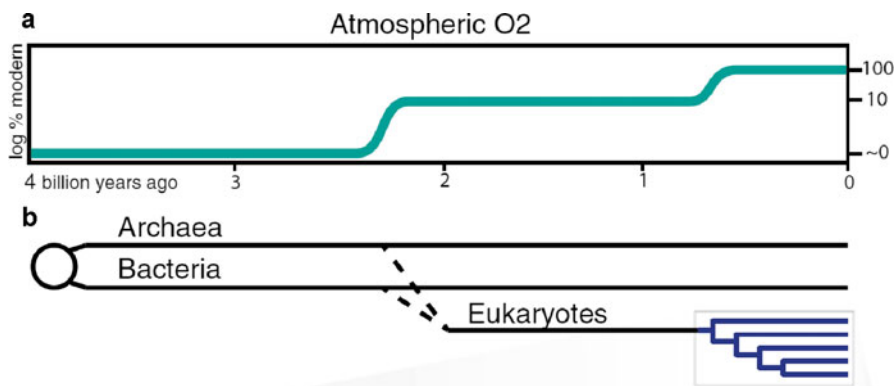
This long history has resulted in ecological interactions among microbes that are broadly diverse and flexible, features enabled by their rapid generation times and large population sizes, in addition to their proclivity for horizontal gene transfer (HGT).

Since animals diverged from their protistan ancestors some 3 billion years after bacterial life originated and as much as 1 billion years after the first appearance of eukaryotic cells, relationships of animals with bacteria were likely already operating when animals first appeared near the end of the Proterozoic Eon. Animal evolution, therefore, is intimately linked to the presence of microbes.

---

## 2.2 Life Did Not Take Over the Globe by Combat, But by Networking

Microbes can be critical determinants of animal population and community structures. Larvae of different marine invertebrate species, for example, are known to respond to either dissolved or surface-bound cues to settlement and metamorphosis. Researchers in the Hadfield laboratory in the Kewalo Marine Lab, a unit of the University of Hawaii, have shown that many of these settlement cues are produced by bacteria (Hadfield 2011). And on the other hand, although most of the microbial



**Fig. 2.1** The evolution of plants and animals in Earth's history has occurred as a patina upon a microbe-dominated landscape. (a) Upper atmospheric oxygen concentration, as a percent of current levels, plotted against geological time. (b) Phylogenetic history of life on Earth, scaled to match the oxygen timeline. Note that the origin of the eukaryotes and the subsequent diversification of animals both correspond to periods of increasing atmospheric oxygen (Source: McFall-Ngai et al. 2013)

**Fig. 2.2** Microscopic view of cyanobacteria, a life-form that first evolved on our planet 3.5 billion years ago became widespread and quite abundant during the Proterozoic (Image Source: Bruce Foulke, U.S. Office of Naval Research)



world operates independently of animals, animals can profoundly impact co-occurring microbes in communities and ecosystems. Examples include the introduction of European earthworms to a North American forest which led to massive declines in soil-microbe biomass and respiration. And rats introduced onto small Pacific islands decimated seabird populations, resulting in decreased guano input, which altered decomposition and nutrient cycling by soil microbes and created a dramatic reduction in ecosystem productivity (Fukami et al. 2006).

Historically, bacteria are seen as pathogens. Starting with Louis Pasteur's discovery of the link between germs and disease, microorganisms were of interest mainly because they were considered to be causes of infectious diseases. In the last 100 years, numerous microorganisms were identified as the causative agents of important human diseases, and bacteriologists, microbiologists, and immunologists are continuing to focus on bacteria as pathogens. This approach has led to enormous

insights in the battle between the invading harmful microbes and the host, as well as enabled the development of efficient strategies to fight infections. However, most bacteria are not harmful to plants or animals, and many of them are beneficial, playing a key ecological role. Louise Taylor and colleagues from the Centre for Tropical Veterinary Medicine at the University of Edinburgh have used the published literature to compile a list of organisms known to be pathogenic to humans, together with the available information on whether they are zoonotic, whether they are regarded as emerging, and on their transmission routes and epidemiologies. Their estimate indicates that only 200 of the millions of bacteria that interact with humans are pathogenic (Taylor et al. 2001).

In fact, eukaryotic cells per se are the descendents of separate prokaryotic cells that joined together in an endosymbiotic event with mitochondria being the direct descendents of a free-living bacterium that was engulfed by another cell (for recent review see Keeling et al. 2015). This became evident in the 1970s when scientists developed new tools and methods for comparing genes from different species. Two teams of microbiologists—one headed by Carl Woese (Woese and Fox 1977) and the other by W. Ford Doolittle at Dalhousie University in Nova Scotia—studied the genes inside chloroplasts of some species of algae. They found that the chloroplast genes bore little resemblance to the genes in the algae’s nuclei. Chloroplast DNA, it turns out, was cyanobacterial DNA. The DNA in mitochondria, meanwhile, resembles that within a group of bacteria that includes the type of bacteria that causes typhus.

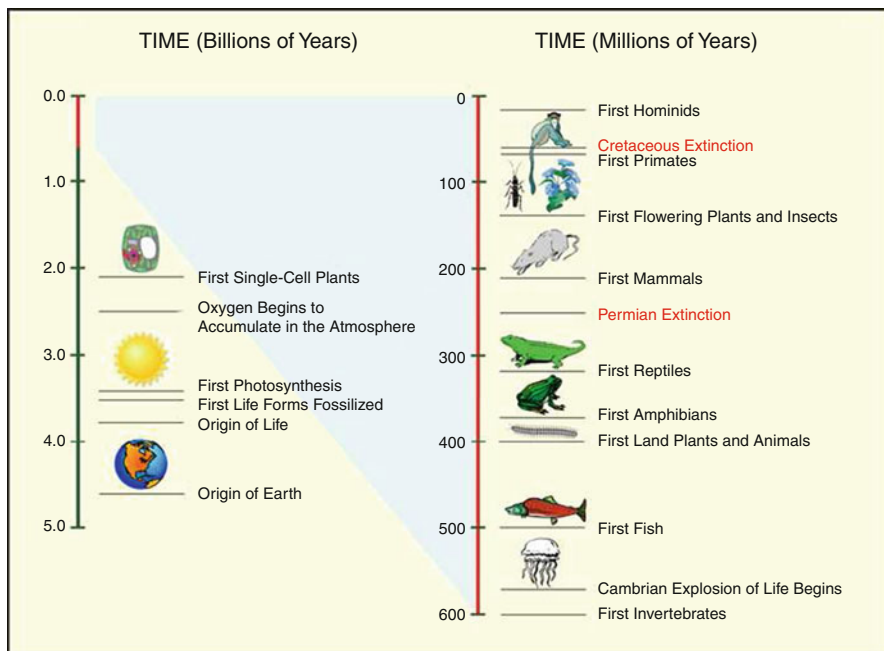
The endosymbiotic theory was advanced and substantiated with microbiological evidence by Lynn Margulis in a 1967 paper. According to Margulis and Dorion (2001), “Life did not take over the globe by combat, but by networking” (i.e., by cooperation). It has become clear that symbiotic events have had a profound impact on the organization and complexity of many forms of life. Algae have swallowed up bacterial partners and have themselves been included within other single cells. Nucleated cells are more like tightly knit communities than single individuals (Doolittle et al. 2013; Doolittle and Zhaxybayeva 2009). Evolution is more flexible than was once believed.

---

## 2.3 The Transformation of the Biosphere at the Ediacaran–Cambrian Boundary

The early Earth was a hostile place, but for most of its 4.6 billion years of its history, life has been present (Fig. 2.3). Prokaryotic life dates back around 4 billion years, its emergence following closely after the end of the period of heaviest bombardments. While the earliest life was unicellular, experiments in multicellularity occurred very early in the history of life on Earth. Some of early Archaean (3.5 BYA) fossils provide clear evidence for early experiments in multicellularity, and by the late Proterozoic (850 MYA), a diverse range of multicellular eukaryotes was present.

Animal life came much later; in fact, fossils from the mid-late Ediacaran (around 580 MYA) represent the earliest evidence for macroscopic multicellular life of any kind and the spectacular radiation of extant animal phyla, known as the Cambrian explosion, at 543 MYA.



**Fig. 2.3** Important events in the evolution of life. Dates for many of the events shown are based on fossil evidence (Image Copyright: Michael Pidwirny) (Taken from The Encyclopedia of Earth, <http://www.eoearth.org/view/article/155427/>)

Why did it take so long for animals to arrive? The answer that this question usually invokes is that levels of oxygen in solution in ancient oceans, which had been increasing gradually, suddenly breached a threshold that could sustain the greater demands (motility, etc.) of animal life. Since the great oxygenation event, the activities of cyanobacteria had brought about steady increases in atmospheric  $O_2$ , so that by the late Vendian and early Cambrian, atmospheric  $O_2$  was probably near present-day levels. The Ediacaran animals were restricted to those patches of shallow water that were relatively well oxygenated; the deeper oceans were in transition from being predominantly anoxic and sulfide-rich to being anoxic but iron-rich and possibly also rich in dissolved organic carbon, but still hostile to all but microbial life.

Reaching a critical atmospheric oxygen level is often nominated as the likely trigger for the emergence of complex multicellular animals, because metazoans have a relatively high demand for oxygen; higher  $O_2$  levels may also have liberated the ancient ecosystem from nitrogen starvation imposed by the sulfide-rich and anoxic state of the oceans.

The idea of a critical level of oxygen triggering animal evolution is, however, not unanimously subscribed to. An attractive alternative idea is that early animals themselves were the agents of change—that by greatly increasing ventilation of the oceans, early animals actually caused most of the observed geochemical perturbations observed during the Neoproterozoic–Paleozoic transition, rather than the

geochemical events enabling/“causing” animal evolution. Zooplankton grazing may have selected for large phytoplankton that were export prone (i.e., upon demise, prone to sinking to the benthos and thus exporting production from the photic zone), effectively transforming what was a cyanobacteria-dominated and anoxic water column that was probably also turbid and stratified into a mixed and aerated system dominated by eukaryotic algae. By causing mixing of the oceans, animals started to transform the environment and thus not only to alter the distribution of microbes but also to open new niches for microbial colonization. Animals, therefore, have thus been major agents of change since their early origins.

---

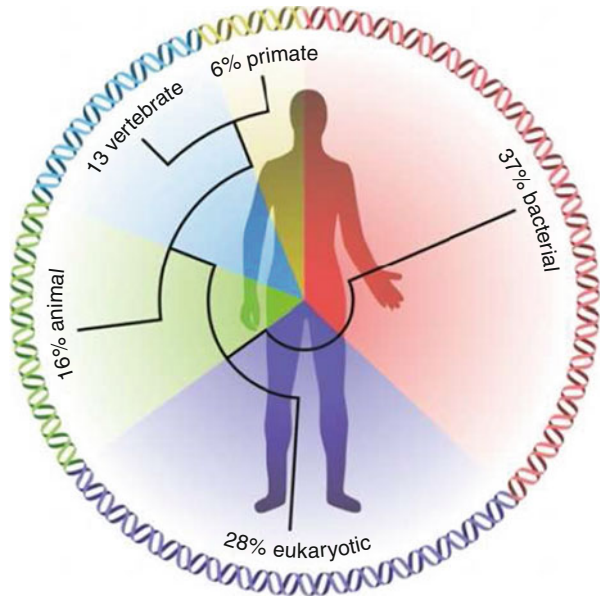
## 2.4 Our Bacterial Ancestry Is Reflected in Our Genomic Signature

Genome sequences provide a unique window into a long history of evolution and dependence on a given habitat. Until quite recently, the consensus view was that animals possessed many genes and pathways not present in other organisms and that an overall increase in genetic complexity (i.e., kinds and numbers of genes) was likely to correlate with increasing morphological sophistication. This view of the ascent of man is almost Copernican in its arrogance, as become clear in the early 2000s with the availability of large molecular datasets for morphologically simple animals.

One of the most intriguing findings to emerge from the sequencing projects in many different animal taxa is that the gene sets of the common metazoan ancestor are surprisingly rich and complex. This flatly contradicts traditional expectations—modern cnidarians are “primitive” animals, with little obvious morphological differentiation, and this is often also assumed to have been the case with Urbilateria. The assumption has been that simple morphology equates to a simple gene set; fewer genes should be required to build a sea anemone than a fly, but this seems not to be true. Sequencing data from cnidarians in particular, but also from sponges, revealed the presence of many genes previously only known from vertebrates and assumed to have evolved in the chordate lineage on the basis of absence from the genomes of the only other metazoans for which data were available at that time, the fruit fly *Drosophila* and the worm *Caenorhabditis*. The cnidarian data in particular have led to the broad (but not universal) acceptance of the idea that genetic complexity came very early in animal evolution, but that widespread gene losses and sequence divergence have occurred and affected some lineages more than others.

While this discovery rapidly remodeled how biologists view the diversity of life, it is only in the last decade that advances in molecular technology have created a seismic shift in biological worldview. Enabled by new genomic approaches, we can now specifically identify microbes and define their activities within communities. The lesson from the comparative genomics of higher animals is that their genomic “dictionaries” share common and deep evolutionary ancestry. This major discovery became possible when a minimum number of genomes had been fully sequenced, the analyses of which are totally reshaping the way biologists view relationships in

**Fig. 2.4** Our bacterial ancestry is reflected in our genomic signature. Image illustrates the relative percentage of human genes that arose at a series of stages in biological evolution. By intertwining of animal and bacterial genomes, a host can expand its metabolic potential, extending both its ecological versatility and responsiveness to environmental change (Source: McFall-Ngai et al. 2013)



the biosphere. Examination of these genomic data reveals that essential all life forms share  $\sim 1/3$  of their genes, including those encoding central metabolic pathways. Many animal genes are either direct descendants of microbial genes or the result of horizontal gene transfer (HGT). For example, 60 % of the  $\sim 23,000$  human genes trace back to at least the origin eukaryotes (Fig. 2.4). The products encoded by these shared genes may provide the foundation for much of the signaling and communication among extant taxa of these two divergent groups.

## 2.5 Genomes of Early Emerging Metazoans, Similar to Humans, Contain a Considerable Fraction of Genes Encoding Proteins of Bacterial Origin

Marine sponges often contain dense and diverse microbial communities, which can constitute up to 35 % of the sponge biomass. Interestingly, metagenomic analysis of sequence reads of the genome of the sponge *Amphimedon queenslandica* (Fig. 2.5) is consistent with existence of a dominant proteobacterial symbiont. From a randomly chosen representative subset of unassembled, filtered reads (120,000 out of 217,873 total), 7720 (6.4 %) were putatively assigned to the bacterial domain of life, and a very small number were assigned to archaea (161, 0.1 %) (Srivastava et al. 2010). The majority of reads map to alpha and gammaproteobacteria. The fraction of reads assigned to alphaproteobacteria exceeds the fraction of all sequenced genomes that belong to alphaproteobacteria. Likewise, we find an excess of reads assigned to Planctomycetes relative to the number of genomes sequenced; however, the total number of putatively bacterial reads assigned to Planctomycetes





**Fig. 2.5** The sea sponge *Amphimedon queenslandica* is the first sponge to have its genome sequenced. This demosponge is found in shallow waters of the Great Barrier Reef and is common around Heron Island (Source: degnanlabs.info)

is small (2.8 %). Thus, metagenomic analysis of sequence reads is consistent with existence of a dominant proteobacterial symbiont.

Evidence for host–microbe interactions has also been published in the course of the draft assembly of the dinoflagellate *Symbiodinium minutum* genome (Shoguchi et al. 2013). Photosynthetic symbionts such as *Symbiodinium* are essential to reef building. The estimated nuclear genome of this species is 1500 Mbp and contains 42,000 protein-coding genes. Similar to observations in *Hydra* and *Nematostella*, the *Symbiodinium* genome provided evidence for the close association with alphaproteobacterium. The largest scaffold in the assembly appears to form a circular DNA of 3.8 Mbp. The GC content of the scaffold was 53 %, compared to 44 % in *S. minutum*. In addition, no expressed sequence tag (EST) data were mapped onto the putative bacterial genome fragment, indicating that the source of this DNA was quite different from *Symbiodinium*. Phylogenetic analysis with the contaminant 16S ribosomal RNA sequence indicated that this organism shows the closest match to *Parvibaculum lavamentivorans* DS-1. *P. lavamentivorans* DS-1 T is the type of species of the novel genus *Parvibaculum* in the novel family Rhodobiaceae (formerly Phyllobacteriaceae) of the order Rhizobiales of alphaproteobacteria. Scanning-electron microscopy shows the presence of bacteria on the surfaces of *Symbiodinium* cells and therefore supports the view that even a protest in fact may be closely associated with bacteria throughout most or all of the host life cycle and thereby should also be considered a holobiont.

The sequence of the *Hydra magnipapillata* genome (Chapman et al. 2010) opened a first window into the *Hydra* holobiont since the genome assembly yielded eight large putative bacterial scaffolds as evidenced by high G+C content, no high-copy repeat sequences typical of *Hydra* scaffolds, and closely spaced single-exon open reading frames with best hits to bacterial sequences. These scaffolds span a total of 4 Mb encoding 3782 single-exon genes and represent an estimated 98 % of

the bacterial chromosome. Phylogenetic analysis of 16S rRNA and conserved clusters of orthologous groups of proteins indicate that this bacterium is a novel *Curvibacter* species belonging to the family Comamonadaceae (order Burkholderiales). About 60 % of annotated *Curvibacter sp.* genes have an ortholog in another species of Comamonadaceae. Notably, the *Curvibacter sp.* genome encodes nine different ABC sugar transporters, compared to only one or two in other species of Comamonadaceae, possibly reflecting an adaptation to life in association with *Hydra*.

Intriguingly, similar findings were reported after looking more closely into the *Nematostella* genome. In 2007, a shotgun genome sequence of *Nematostella vectensis* was obtained, assembled into long scaffolds, and annotated, and a remarkable resemblance of the gene repertoire in the species to the gene sets in more complex metazoa was emphasized, including a higher sequence similarity, a larger content of shared genes, and a higher degree of synteny between *N. vectensis* and vertebrates than between vertebrates and familiar invertebrate model organisms, such as the fruit fly and a soil nematode. Detailed, case-by-case analysis of regulatory and signaling pathways shared by *N. vectensis* and higher animals has confirmed a Premetazoan origin for many of these pathways, either in substantially complete form or in simpler versions that have been elaborated later in metazoan evolution by molecular “tinkering.”

Irena I. Artamonova from the Vavilov Institute of General Genetics in Moscow, Russia, and Arcady R. Mushegian from the National Science Foundation, Arlington, Virginia, USA, have discovered that the *N. vectensis* genome sequence submitted to the sequence databases is very likely to contain many genes of bacterial and bacteriophage origin, derived from prokaryotes closely associated with the sea anemone. The bacterial and phage genes are located in distinct scaffolds, which can be separated from those corresponding to the DNA of the target organism, *N. vectensis*, by several criteria, including the nucleotide composition and the provenance of neighboring genes; the spurious character of introns annotated in these genes; and a rich and apparently randomly drawn repertoire of predicted protein functions, which include several molecular roles considered to be prokaryote specific. The majority of bacterium-like genes from the *N. vectensis* genome project are phylogenetically closer to Bacteroidetes or Proteobacteria homologs than to any eukaryotic homolog. Two distinct bacterial genera are clear leaders in high-sequence similarity of these bacterium-like proteins. They are *Pseudomonas* and *Flavobacterium* (80–85 % and 85–95 % median identity of amino acids to their nearest database homologs, respectively), separated by many hundreds of millions of years of bacterial evolution. Thus, a considerable fraction of *N. vectensis* genes in fact are bacterium-like genes annotated as belonging to *N. vectensis* but encode proteins of bacterial origin. The phylogenetically closest database matches of these bacterium-like sequences originated from several clades of bacteria, but nearly two-third of them belonged to one of two clades: Proteobacteria and Bacteroidetes.

Following the same analytical sequencing approach indicated that the initial set of automatically assigned putative virus proteins consisted of seven bacteriophage proteins and six proteins from eukaryotic viruses.



## 2.6 The CRISPR/CAS System as Window into Ancient Holobionts

The CRISPR/CAS system is a complex resistance mechanism described in bacteria and is the only adaptive and inheritable prokaryotic immune system. CRISPR stands for “clustered, regularly interspaced, short palindromic repeats” and CAS for “CRISPR-associated genes.” In 1987, Ishino and his colleagues discovered an unusual structure of repetitive DNA downstream of the *E. coli inhibitor of apoptosis (iap)* gene, consisting of invariant direct repeats and variable spacing elements. In 2002, Jansen and colleagues coined the term CRISPR and reported that CRISPRs colocalized with specific *cas* genes. CRISPR/CAS systems are exclusively found in prokaryotes and are present in approximately half of all bacteria and almost all archaea (Grissa et al. 2007). Analysis of bacterial and archaeal genome sequences and the associated viral and plasmid sequences led to the realization that CRISPR spacers, which resemble fragments of foreign genetic elements, were derived from invading genomes (Barrangou et al. 2007). This breakthrough, together with the detection of CRISPR locus transcripts with defined length of one or more spacer repeat units, and the predicted nucleic acid-related activities for many of the *cas* genes, led the team of computational biologists around Eugene Koonin at the National Center for Biotechnology Information (NCBI) in Bethesda, to propose that CRISPR/CAS neutralizes invaders via a mechanism reminiscent of RNA interference (Makarova et al. 2011). Soon after, Rodolphe Barrangou, Philippe Horvath, and colleagues provided the first experimental evidence that the CRISPR/CAS system of the lactic acid bacterium *Streptococcus thermophilus* functions as an inheritable, adaptive prokaryotic immune system conferring phage resistance. The bacterial genome integrates a sequence of the viral genome, called a spacer, upon infection; that sequence later serves as a guide for destroying any matching DNA, so that subsequent viral infections are fended off (Horvath and Barrangou 2010).

As pointed out by Artamonova and Mushegian (2013), many of the bacterium-like proteins annotated by the *Nematostella vectensis* genome project have functions that are relevant to the biology of bacteria but have never been reported in eukaryotes. Interestingly, among the examples are proteins with high similarity to Cas1 (one of the CRISPR-associated proteins participating in a prokaryote-specific immunity system), bacterial transcription regulators and bacterial-type protein kinases, the pilin glycosylation protein PglD, OmpA/MotB domain proteins that operate in the outer membrane of gram-negative bacteria, and the aforementioned cyclopeptide biosynthesis enzymes. Even if the genes encoding these proteins were once horizontally transferred to the *N. vectensis* genome, it would be highly unusual for a eukaryote to maintain all of them, in an essentially unmodified form, to perform a set of functions that are without precedent in metazoa.

Artamonova and Mushegian take the finding of CRISPR-associated proteins as computational evidence that argues strongly for the existence of bacteria and viruses closely associated with *N. vectensis*. A substantial, perhaps nearly randomly sampled portion of the *Nematostella* metagenome is already deposited in the databases, apparently having been misannotated as *N. vectensis* genes when in fact this gene

complement should be studied further as evidence of the holobiont organization of the sea anemone. Interestingly, the bacterium-like genes revealed by the *N. vectensis* genome project partition from the host genes similarly to the genes of the known bacterial endosymbiont of *Hydra magnipapillata*. Similar to what has been observed previously with *Hydra magnipapillata* bacterium-like genes (Chapman et al. 2010), the genes of the *N. vectensis* bacteria tend to be intronless open reading frames, which are located in separate genomic scaffolds characterized by distinct GC contents.

All these lines of evidence suggest that of the two possible explanations for the presence of these genes in the data banks, i.e., domestication by an invertebrate of horizontally transferred bacterial genes in the evolutionary past and the occurrence of bacteria in the *N. vectensis* planula used for DNA isolation, the latter seems to be more plausible. Far from these genes being biologically irrelevant sample contamination, their presence in the genome database indicates that, similar to the cases of other marine Anthozoa (Williams et al. 2007) and the completely sequenced freshwater hydrozoan *H. magnipapillata*, the body of the starlet sea anemone is in fact a holobiont, i.e., a consortium of a metazoan animal and bacteria closely associated with it throughout most or all of the host life cycle (see also Sect. 6.4).

---

## 2.7 Origins of Complexity: What Makes an Animal?

Multicellularity has evolved independently many times and appears to be a trivial undertaking in terms of biology. Under artificial selection in the laboratory, normally unicellular bacteria can achieve a kind of multicellularity by “learning” to adhere to each other. And some bacterial lineages even have achieved a degree of cellular differentiation; for example, when *Nostoc*, a cyanobacterium found in a variety of environmental niches (Fig. 2.6), is grown under N-limitation, nitrogen-fixing heterocysts develop at regular intervals along a chain of “normal” vegetative cells, allowing the oxygen-sensitive nitrogen fixing and oxygenic photosynthesis reactions to be partitioned between functionally differentiated cells. Despite billions of years of opportunity, this, however, seems to be as far as bacteria have explored the opportunities provided by multicellularity.

Among eukaryotes, a similar level of complexity has independently been achieved many times. For example, volvocine algae have apparently evolved their version of multicellularity (involving the differentiation of reproductive and ciliated cells) several times independently. Note that the mechanism involved intrinsically limits the volvocine algae this very basic level of multicellularity. As the nearest relatives of metazoans, choanoflagellates are of particular interest; most choanoflagellates are unicellular, but some are not. In the case of *Salpingoeca*, while the mechanism of multicellularity is an elegant reflection of selective forces, the molecules and processes involved are irrelevant in terms of understanding animal origins. In response to the abundance of prey bacteria, the unicellular form of *Salpingoeca* forms aggregates that can better exploit the food source; the

**Fig. 2.6** The nitrogen-fixing bacterium *Nostoc* is a commonly occurring cyanobacterium often found in symbiotic associations (Taken from [profs.com](http://profs.com))



mechanism involves recognition by the choanoflagellate of a specific sulfolipid in the cell wall of a restricted group of bacteria.

This kind of multicellularity is obviously very different both qualitatively and quantitatively from that of animals, but it is important to recognize that intermediate levels of complexity exist. Green plants have achieved a comparable level of complexity to that of animals, with fungi and the red and brown algal lineages at a somewhat less advanced level. That despite multicellularity per se being a relatively simple trick to perform, there have been just five successful independent ventures into “advanced” multicellularity in over 3.5 billion years of opportunity that indicates that the events or innovations involved in founding these lineages were of a fundamentally different nature, that leading to the Metazoa being perhaps the most significant in terms of the evolution of the biosphere.

---

## 2.8 Multicellularity Requires Cooperation of Cells

Below, we will consider possible genetic drivers of the Cambrian and earlier animal diversifications, but first, with so much of a head start, why have prokaryotes failed to evolve beyond a very basic version of multicellularity? One possibility is that the nature of bacterial evolution itself may serve to effectively limit them to this basic level of multicellularity. Cellular specialization demands integration of cells, with greater integration being required for higher levels of multicellularity. However, integration requires cooperation of cells, which means that each cell must be able to “trust” its neighbors—they must operate in fundamentally similar ways, otherwise, for example, effective communication between cells is not possible, and neighbors could “cheat” rather than cooperate in the multicellular endeavor. Multicellularity therefore is contingent on genetic homogeneity—cells can only effectively communicate with (“trust”) its neighbors which have essentially the same genotype. Genetic homogeneity is a hallmark of multicellularity, but is probably quite rare in prokaryotes.

On the contrary, bacteria are genetically promiscuous, exchanging genes with remarkable fluidity, while this facilitates rapid evolution to changing environmental conditions, it limits the ability of bacterial cells to cooperate and integrate and hence achieve complex multicellularity. One corollary of this line of thinking is that the evolution of more complex forms multicellularity should require the evolution of increasingly sophisticated ways of protecting the genome from invading foreign DNA. Both animals and plants do have formidable molecular arsenals that can be deployed against invading foreign DNA, which accounts for the rarity of lateral gene transfers (LGTs) into their genomes. In fact, many putative LGTs into animal genomes are more likely to reflect random but widespread gene loss from a genetically complex ancestor.

The case for very early origins for extant animals is based on molecular phylogenetic arguments and comparing fossil data with extant animals. Molecular phylogenetic approaches assume that it is possible to extrapolate recent rates of sequence evolution back to deep time, but such extrapolations often come with enormous standard deviations, and the nature of the analyses could violate the basic assumptions of molecular phylogenetic approaches. The interpretation of Precambrian fossils is at least equally contentious, with alarming frequency, new discovered early Precambrian fossils or new interpretations of known materials are claimed to either push back the origins of the Metazoa or confirm deep Precambrian origins, the papers (often in high-impact journals) only to be refuted shortly thereafter. Here, we take the conservative view that the Metazoa arose at or close to the Ediacaran/Cambrian boundary and that all or most of the Precambrian fossils represent something else entirely. Before the late Ediacaran, animal life was scarce, simple, and restricted in distribution, but then suddenly the entire game changed—animals were abundant, complex, and more widely distributed and began to have a much greater impact on the biosphere.

---

## 2.9 Genomes of Early Emerging Metazoans Reveal the Origin of Animal-Specific Genes

During animal development, formation of morphologically complex structures from the fertilized egg requires precise cell-to-cell communication. Surprisingly, only a handful of intercellular signaling pathways are used throughout the animal kingdom: wnt,  $\text{tgf-}\beta$ , hedgehog, receptor tyrosine kinase, and notch. The genome of a marine sponge, *Amphimedon queenslandica* (Fig. 2.5), has provided fascinating insights into the origins of animal evolution (Srivastava et al. 2010). The *A. queenslandica* genome harbors an extensive repertoire of developmental signaling and transcription factor genes, indicating that the metazoan ancestor had a developmental “toolkit” similar to that in modern complex bilaterians. Analysis of the *Amphimedon* genome and expression studies done by Maja Adamska’s team at the SARS Center in Bergen (Adamska et al. 2010; Adamska et al. 2011; Fortunato et al. 2014) demonstrate that all major metazoan signaling pathways are used during sponge embryogenesis. Wnt,  $\text{tgf-}\beta$ , and hedgling—related to hedgehog transmembrane protein—genes are expressed in dynamic, partially overlapping patterns in *Amphimedon* embryos.

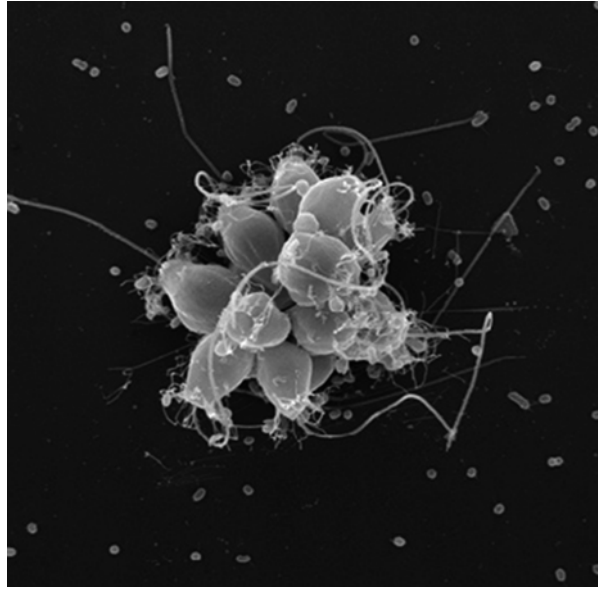
Similarly to higher animals, the wnt pathway appears to be used in establishment of the anterior–posterior axis of the sponge embryo. The origins of many of these and other genes specific to animal processes such as cell adhesion, and social control of cell proliferation, death, and differentiation can be traced to genomic events (gene birth, subfamily expansions, intron gain/loss, and so on) that occurred in the lineage that led to the metazoan ancestor, after animals diverged from their unicellular “cousins.” In addition to possessing a wide range of metazoan-specific genes, the *Amphimedon* draft genome is missing some genes that are conserved in other animals, indicative of gene origin and expansion in eumetazoans after their divergence from the demosponge lineage and/or gene loss in *Amphimedon*.

While the sponge and cnidarian data pushed back the evolutionary origins of developmental signaling pathways and many transcription factors to common metazoan (or eumetazoan) ancestors (Shinzato et al. 2011), it is now clear that many “animal-specific” genes have older origins yet and are not restricted to animals. Clear evidence for more widespread occurrence of “animal-specific” genes came from sequencing of transcripts and then the genome of the choanoflagellate, *Monosiga brevicollis*. Choanoflagellates promise to provide insights into the origins of multicellularity because they sometimes form colonies in which cells maintain their individuality but can also share small molecules. Nicole King’s lab at UC Berkeley has shown that choanoflagellate genomes do encode a number of “animal-specific” genes, including receptor tyrosine kinases, cadherins, and immunoglobulins (Alegado and King 2014; Alegado et al. 2012). Disappointingly, sequencing the genome of the “multicellular” choanoflagellate, *Salpingoeca rosetta* (Fig. 2.7), added very little to understanding metazoan origins; it seems that choanoflagellates have independently invented a mechanism (or mechanisms) for “multicellularity.” Thus, although they tell us more about how multicellularity can be achieved, choanoflagellates cannot inform us about how the metazoan ancestor made this transition.

The recent focus of genome sequencing efforts on other unicellular holozoans highlights just how old some “animal specific” genes actually are. Surprisingly, the genome of the filasterean holozoan *Capsaspora*, which may have been secondarily reduced given its endosymbiotic (or parasitic) lifestyle, encodes a number of “metazoan” transcription factors that have been lost from choanoflagellates, including the RUNX, NFκ, and the T-box protein Brachyury. As efforts to sequence the genomes of a truly representative range of close outgroups and early diverging ingroups of the Metazoa continue, it is likely that the list of animal-specific genes will continue to shrink. In reality, only a small number of gene families or protein domains are likely to be restricted to the animal kingdom.

Although stochastic and widespread, gene loss accounts for many putative LGTs into animal genomes, a few cases remain of genes typical of other kingdoms being anomalously distributed in only one (or, at most a few) metazoan lineage(s). Classic examples include the cellulose biosynthetic genes of ascidians, whereas cellulose is a common cell wall constituent in plants, fungi, slime molds, and bacteria, among animals, it is only known from the tunics of ascidians. This distribution gave rise to the idea that the biosynthetic pathway was acquired in ascidians by “horizontal transfer” (= LGT). However, if the genes were acquired by LGT,

**Fig. 2.7** A cluster of colony-forming *Salpingoeca rosetta* choanoflagellates. Triggered by the presence of bacteria, the single-celled choanoflagellate *Salpingoeca rosetta* divides and stays with its sisters to form a colony. Scanning electron microscope image courtesy of Nicole King/news.berkeley.ed



this happened a long time ago—much longer ago than is usually implied when LGT is invoked—because cellulose synthase genes from ascidians that diverged deep in the Cambrian are very similar. Rather than a recent LGT, this pushes the time of any “horizontal transfer” of cellulose synthase genes to very early in meta-zoan evolution. Similar scenarios apply to a number of other animal genes with anomalous distributions.

The *Hydra magnipapillata* genome was sequenced in 2010 by the Sanger whole genome shotgun method at the J. Craig Venter Institute (Chapman et al. 2010). Following the completion of sequencing, two assemblies of the genome were done. The assembly gives an estimated nonredundant genome size of 1.05 Gb. The *Hydra* genome is (A + T)-rich and contains ~20,000 bona fide protein-coding genes. All the essential signaling pathways controlling the formation of epithelia, muscle tissue, stem cells, nerve cells, and the innate immune system are present in the *Hydra* genome.

All these observations in choanoflagellates, sponges, and cnidarians together make not only a very strong point that the “molecular tool kit” leading to the evolution of higher animals and humans was already present at the beginning of multicellular life. They also show that animal life from the very beginning was multiorganismal and involved close association between simple animal hosts, bacteria, and in many cases also eukaryotic symbionts and viruses (for reference see, e.g., King 2010).

As we will see next, embedding this diversity of animal life within a clear phylogenetic framework is essential to understanding it from both molecular biology and evolutionary perspective.

## References

- Adamska M, Larroux C, Adamski M, Green K, Lovas E, Koop D, Richards GS, Zwafink C, Degan BM (2010) Structure and expression of conserved Wnt pathway components in the demosponge *Amphimedon queenslandica*. *Evol Dev* 12(5):494–518
- Adamska M, Degan BM, Green K, Zwafink C (2011) What sponges can tell us about the evolution of developmental processes. *Zoology (Jena)* 114(1):1–10, Review
- Alegado RA, King N (2014) Bacterial influences on animal origins. *Cold Spring Harb Perspect Biol* 6:a016162
- Alegado RA, Brown LW, Cao S, Dermenjian RK, Zuzow R, Fairclough SR, Clardy J, King N (2012) A bacterial sulfonolipid triggers multicellular development in the closest living relatives of animals. *ELife* 1:e00013
- Artamonova II, Mushegiand AR (2013) Genome sequence analysis indicates that the model eukaryote *Nematostella vectensis* harbors bacterial consort. *Appl Environ Microbiol* 79(22):6868–6873
- Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA, Horvath P (2007) CRISPR provides acquired resistance against viruses in prokaryotes. *Science* 315:1709–1712
- Chapman JA, Kirkness EF, Simakov O, Hampson SE, Mitros T, Weinmaier T, Rattei T, Balasubramanian PG, Borman J, Busam D, Disbennett K, Pfannkoch C, Sumin N, Sutton GG, Viswanathan LD, Walenz B, Goodstein DM, Hellsten U, Kawashima T, Prochnik SE, Putnam NH, Shu S, Blumberg B, Dana CE, Gee L, Kibler DF, Law L, Lindgens D, Martinez DE, Peng J, Wigge PA, Bertulat B, Guder C, Nakamura Y, Ozbek S, Watanabe H, Khalturin K, Hemmrich G, Franke A, Augustin R, Fraune S, Hayakawa E, Hayakawa S, Hirose M, Hwang JS, Ikeo K, Nishimiya-Fujisawa C, Ogura A, Takahashi T, Steinmetz PR, Zhang X, Aufschnaiter R, Eder MK, Gorny AK, Salvenmoser W, Heimberg AM, Wheeler BM, Peterson KJ, Bottger A, Tischler P, Wolf A, Gojobori T, Remington KA, Strausberg RL, Venter JC, Technau U, Hobmayer B, Bosch TC, Holstein TW, Fujisawa T, Bode HR, David CN, Rokhsar DS, Steele RE (2010) The dynamic genome of *Hydra*. *Nature* 464(7288):592–596
- Doolittle WF, Fraser P, Gerstein MB, Graveley BR, Henikoff S, Huttenhower C, Oshlack A, Ponting CP, Rinn JL, Schatz MC, Ule J, Weigel D, Weinstock GM (2013) Sixty years of genome biology. *Genome Biol* 14(4):113
- Doolittle WF, Zhaxybayeva O (2009) On the origin of prokaryotic species. *Genome Res* 19(5):744–756
- Fortunato SAV, Adamski M, Mendivil Ramos O, Leininger S, Liu J, Ferrier DEK, Adamska M (2014) Calcisponges have a *ParaHox* gene and dynamic expression of dispersed NK homeobox genes. *Nature* 514:620–623
- Fukami T, Wardle DA, Bellingham PJ, Mulder CP, Towns DR, Yeates GW, Bonner KI, Durrett MS, Grant-Hoffman MN, Williamson WM (2006) Above- and below-ground impacts of introduced predators in seabird-dominated island ecosystems. *Ecol Lett* 9:1299–1307
- Grissa I, Vergnaud G, Pourcel C (2007) The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC Bioinf* 23(8):172
- Hadfield MG (2011) Biofilms and marine invertebrate larvae: what bacteria produce that larvae use to choose settlement sites. *Ann Rev Mar Sci* 3:453–470
- Horvath P, Barrangou R (2010) CRISPR/Cas, the immune system of bacteria and archaea. *Science* 327:167–170
- Keeling PJ, McCutcheon JP, Doolittle WF (2015) Symbiosis becoming permanent: survival of the luckiest. *Proc Natl Acad Sci U S A* 112(33):10101–10103
- King N (2010) Nature and nurture in the evolution of cell biology. *Mol Biol Cell* 21:3801–3802
- Knoll AH (2003) *Life on a young planet*. Princeton University Press, Princeton
- Makarova KS, Haft DH, Barrangou R, Brouns SJJ, Charpentier E, Horvath P, Moineau S, Mojica FJM, Wolf YI, Yakunin AF, Oost J, Koonin EV (2011) Evolution and classification of the CRISPR-Cas systems. *Nat Rev Microbiol* 9:467–477



- Margulis L, Dorion S (2001) Marvellous microbes. *Resurgence* 206:10–12
- McFall-Ngai M, Hadfield M, Bosch T, Carey H, Domazet-Loso T, Douglas A, Dubilier N, Eberl G, Fukami T, Gilbert S, Hentschel U, King N, Kjelleberg S, Knoll A, Kremer N, Mazmanian S, Metcalf J, Nealon K, Pierce N, Rawls J, Reid A, Ruby E, Rumpho M, Sanders J, Tautz D, Wernegreen J (2013) Animals in a bacterial world, a new imperative for the life sciences. *Proc Natl Acad Sci U S A* 110(9):3229–3236
- Shinzato C, Shoguchi E, Kawashima T, Hamada M, Hisata K, Tanaka M et al (2011) Using the *Acropora digitifera* genome to understand coral responses to environmental change. *Nature* 476:320–323
- Shoguchi E, Shinzato C, Kawashima T, Gyoja F, Mungpakdee S, Koyanagi R, Takeuchi T, Hisata K, Tanaka M, Fujiwara M, Hamada M, Seidi A, Fujie M, Usami T, Goto H, Yamasaki S, Arakaki N, Suzuki Y, Sugano S, Toyoda A, Kuroki Y, Fujiyama A, Medina M, Coffroth MA, Bhattacharya D, Sato N (2013) Draft assembly of the symbiodinium minutum nuclear genome reveals dinoflagellate gene structure. *Curr Biol* 23:1399–1408
- Srivastava M, Simakov O, Chapman J, Fahey B, Gauthier MEA, Mitros T, Richards GS, Conaco C, Dacre M, Hellsten U, Larroux C, Putnam NH, Stanke M, Adamska M, Darling A, Degnan SM, Oakley TH, Plachetzki DC, Zhai Y, Adamski M, Calcino A, Cummins SF, Goodstein DM, Harris C, Jackson DJ, Leys SP, Shu S, Woodcroft BJ, Vervoort M, Kosik KS, Manning G, Degnan BM, Rokhsar DS (2010) The Amphimedon queenslandica genome and the evolution of animal complexity. *Nature* 466(7307):720–726
- Taylor LH, Latham SM, Woolhouse ME (2001) Risk factors for human disease emergence. *Philos Trans R Soc Lond B Biol Sci* 356(1411):983–989
- Williams GP, Babu S, Ravikumar S, Kathiresan K, Prathap SA, Chinnapparaj S, Marian MP, Alikhan SL (2007) Antimicrobial activity of tissue and associated bacteria from benthic sea anemone *Stichodactyla haddoni* against microbial pathogens. *J Environ Biol* 28:789–793
- Woese CR, Fox GE (1977) Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc Natl Acad Sci U S A* 74(11):5088–5090