9.1: Genomes and their organization

Genomes are characterized by two complementary metrics, the number of base pairs of DNA and the number of genes present within this DNA. The number of base pairs is easier to measure, we can count them. This can, however, lead to a mistaken conclusion, namely that the number of base pairs of DNA within the genome of a particular species, organism, or even tissue within an organism is fixed and constant. In fact genomes are dynamic, something that we will return to shortly.

The genome of an organism (and generally the cells of which it is composed) consists of one or more DNA molecules. When we talk about genome size we are talking about the total number of base pairs present in all of these DNA molecules added together. The organism with one of the largest known genomes is the plant *Paris japonica*; its genome is estimated to be ~150,000 x 10^6 (millions of) base pairs\(^{255}\). In contrast the (haploid) human genome consists of ~3,200 x 10^6 base pairs of DNA. The relatively small genome size of birds (~1,450 x 10^6 base pairs) is thought to be due to the smaller genome size of their dinosaurian ancestors\(^{256}\). That said there are interesting organisms that suggest that in some cases, natural selection can act to dramatically increase or decrease genome size without changing gene number. For example, the carnivorous bladderwort *Utricularia gibba*, has a genome of ~80 x 10^6 base pairs and ~28,000 genes, significantly fewer base pairs of DNA, but apparently more genes than humans.

Very much smaller genomes are found in prokaryotes, typically their genomes are a few millions of base pairs in length. The smallest genomes occur in organisms that are obligate parasites and endosymbionts. For example the bacterium *Mycoplasma genitalium*, the cause of non-gonococcal urethritis, contains ~0.58 x 10^6 base pairs of DNA, which encodes ~500 distinct genes. An even smaller genome is found in the obligate endosymbiont *Carsonella ruddii*; it has 159,662 (~0.16 x 10^6) base pairs of DNA encoding "182 ORFs (open reading frames or genes), 164 (90%) overlap with at least one of the two adjacent ORFs"\(^{257}\). Eukaryotic mitochondria and chloroplasts, derived from endosymbionts, have very small genomes. Typically mitochondrial genomes are ~16,000 base pairs in length and contain ~40 genes, while
chloroplasts genomes are larger, ~120,000–170,000 base pairs in length, and encode ~100 genes. Most of the genes present in the original endosymbionts appear to have either been lost or transferred to the host cell’s nucleus. This illustrates a theme that we will return to, namely that genomes are not static. In fact, it is their dynamic nature that makes significant evolutionary change possible.

An interesting question is what is the minimal number of genes that an organism needs. Here we have to look at free living organisms, rather than parasites or endosymbionts, since they can rely on genes within their hosts. A common approach is to use mutagenesis to generate non-functioning (amorphic) versions of genes. One can then count the number of essential genes within a genome, that is, genes whose function is absolutely required for life. One complication is that different sets of genes may be essential in different environments, but we will ignore that for now. In one such lethal mutagenesis study Lewis et al found that 382 of the genes in *Mycoplasma genitalium* are essential; of these ~28% had no (as yet) known function.258

---

**References**


258 Essential genes of a minimal bacterium: [http://www.pnas.org/content/103/2/425.full](http://www.pnas.org/content/103/2/425.full)

---

**Contributors**

- Michael W. Klymkowsky (University of Colorado Boulder) and Melanie M. Cooper (Michigan State University) with significant contributions by Emina Begovic & some editorial assistance of Rebecca Klymkowsky.