Evolution is commonly measured using comparative phylogenetic analysis. Comparisons of orthologous characters and sequences from different species are used to infer organismal evolution. Analyses of duplicated genes can be used to root phylogenetic trees and infer ancestral groups. The expansion of gene families through gene and genome duplications allowed more complex regulatory and developmental pathways to evolve in multicellular eukaryotes. In prokaryotes and single celled eukaryotes, the acquisition of foreign genes by horizontal gene transfer is the main mechanism for gene family expansion; it allows genomes to evolve new traits quickly and facilitates the assembly of new metabolic pathways. Additionally, prokaryotic organisms with short generation times will accumulate genetic adaptations at a much faster rate than organisms with longer generation times (e.g., humans). In multicellular animals where somatic cells and gametes are separate, acquisition of foreign genes is rare, leading to high levels of similarity in gene content. However, multicellular eukaryotes have evolved in close association with prokaryotic symbionts that impact development, physiology and ecology of the association. To understand the evolution of the complex human systems we must consider the genomes of the associated microbiota, known as the microbiome. We must therefore consider the human as a holobiont, a complex ecosystem, whose evolutionary fitness is determined by the host, the symbionts, and their interactions.

**Orthologous, Paralogous and Xenologous Genes**

Sequences or structures that evolved from a single ancestral structure or sequence are homologous. To classify the different types of homology, Fitch (1970) introduced the terms orthology and paralogy. Orthologous structures or sequences in two organisms are homologues that evolved from the same feature in their most recent common ancestor; however, orthologues do not necessarily retain their ancestral function. Since the evolution of orthologues reflects organismal evolution, molecular systematics has been concerned traditionally with comparing orthologous sequences. By contrast, paralogues are homologues whose evolution reflects gene duplication events. For example, the β chain of haemoglobin is a parologue of both the haemoglobin α chain and myoglobin, because they each evolved from the same ancestral globin gene through repeated gene duplication events. Only the deepest split in a phylogenetic tree relating homologous proteins determines orthology versus paralogy (Fitch, 1970). If the deepest split between two genes corresponds to a speciation event, those genes are orthologues. If the split corresponds to a gene duplication event, then those genes would be considered paralogues. To clearly distinguish whether two genes are orthologues or paralogues, a rooted phylogeny is necessary.
Using genes that encode resistance to antibiotics as a model, the term xenology was coined for homologues that were acquired by an organism through horizontal gene transfer (Fitch, 1970). Synology denotes homologues that arose from the fusion of complete genomes (Gogarten, 1994), such as bacterial genes brought into the eukaryotic cell through the mitochondrial endosymbiont.

See also Homologous, Orthologous and Paralogous Genes.

**Gene number and genome organization**

In the human genome, protein-coding genes tend to exist in nonrandom clusters that are separated by large stretches of non protein-coding DNA, referred as gene-poor ‘deserts’. Regulatory elements also appear to be clumped into regulatory rich and poor regions (Zhang et al., 2007). See also Clustering of Highly Expressed Genes in the Human Genome, Gene Clustering in Eukaryotes, Genome Organization of Vertebrates, and Isochores. Up to 1% of the total human genome comprises exons – the regions of genes that encode proteins. Introns, the regions in genes that are spliced out during the creation of messenger ribonucleic acid (mRNA), make up about 24% of the genome. The number of protein coding genes is estimated to be approximately 22,000 (Pertea and Salzberg, 2010). More than 90% of multi exon coded proteins undergo alternative splicing, allowing more than a single protein to be translated from a region of exons. This further complicates estimates of protein-coding gene number, although the function of the vast majority of alternatively spliced transcripts remains unknown, and some may reflect transcriptional noise rather than a distinct function (Pertea 2012).

Additionally, individual humans can differ slightly in genome content with variation related primarily to deletions or regions of segmental duplication. Comparison of two different human genomes, one from Africa and one from Asia with the reference genome at NCBI showed 5Mb of unique DNA in each of the new genomes. Estimates suggest a human pan-genome would include up to 40Mb more DNA (or < 0.01%) than the reference genome (Pertea and Salzberg, 2010). This is a very small difference compared to the estimated 90% difference in size between the pan-genome and a single reference genome of the bacterium *E. coli*. Comparison of 60 *E. coli* genomes suggests that less than 10% of the genes in the *E. coli* pan-genome are present in all of the 61 *E. coli* genomes analyzed. (Lukjancenko et al., 2010)

In addition to protein-coding mRNAs, a new class of transcripts, collectively called non-coding RNAs (ncRNA), has been identified. These transcripts do not code for proteins and they originate from intergenic regions, introns, or from sequences antisense to known transcripts. Currently their function is not well understood (Johnson et al., 2005). These ncRNA encoding regions expand the traditional definition of a gene to include a myriad of non-protein coding sequences and hint towards complex patterns of expression and regulation during development of different cell types. The growing list of functional RNAs increases the number of estimated genes from 22,000 to 30,000–40,000 genes (Pertrea 2012). Non-coding RNAs (ncRNAs) include small interfering RNA (siRNA) and microRNA (miRNA) which play a central role in RNA interference by binding to specific mRNA molecules to increase or decrease the amount of protein translated. Other ncRNAs that have
a role in gene silencing include the PIWI-interacting RNAs (piRNAs) which bind PIWI proteins during spermatogenesis and are thought to be involved in silencing transposons in the genome. Promoter-associated RNAs (PARs), transcription initiation RNAs (tiRNAs), X-inactivation RNAs (xiRNAs), and various other classes of ncRNAs are also suggested to have functional roles (Pertrea 2012). The long non-coding RNAs (lncRNAs), defined as ncRNAs longer than 200 bp, undergo splicing with similar frequency to protein-coding mRNAs and are probably the least well-understood transcripts. Ponjavic et al. (2007) analyzed a set of more than 3000 long noncoding RNAs (lncRNAs), and found the substitution pattern and indel distribution in comparison of mouse and human homologues suggest that these macro RNAs are under purifying selection. Although the function of many of these nonprotein-coding RNAs is still to be determined, they may be key regulators of epigenetic gene regulation in mammalian cells (Pertrea 2012).

Based on the differential expression, localization and patterns of conservation in ncRNAs, it is likely that the portion of the human genome that is functional has previously been underestimated. Analysis of substitution rates suggests that 6.5-10% of the genome appears to be under selective constraint (Meader et al., 2010). At the other extreme of functional DNA estimates, the EnCODE project combined data from a variety of analyses to map RNA transcribed regions, protein-coding regions, transcription-factor binding sites, chromatin structure and DNA methylation sites. The study encompassed 1,640 genome wide data sets from 147 different human cell types. The non-translated regions that may have regulatory functions included elements such as enhancers, promoters and regions that contribute to the structure of chromatin. The sum of these data were interpreted to suggest that up to 80% of the genome contains elements that participate in at least one of these functions (Zhang et al., 2007).

The ratio of constrained (and therefore likely functional) non-protein coding bases to coding bases in Drosophila is 2 while in humans it is between 5 and 8. Much of the apparent differences in complexity between species may be due to a varying amount of noncoding regulatory sequence, regulating a fairly stable core of protein-coding genes (Meader et al., 2010). This is compatible with the notion that much of the organismal complexity and interspecific differences of mammals, are encoded in the non-protein-coding functional complement rather than in protein-coding sequence (Ponting and Hardison, 2011).

**Duplicated and repetitive DNA**

Repeated sequences derived from transposable elements comprise 43–45% of the genome (Li et al., 2001) and include long interspersed nuclear elements (LINEs or L1 elements), short interspersed nuclear elements (SINEs), DNA transposons and long-terminal repeat (LTR) elements. SINEs include Alu repeats (a distinct class of retrotransposon-amplified repeat DNA that arose with the evolution of primates) and comprise about 10% of the genome (Li et al., 2001). Transposable elements can have a significant role in gene duplication through the formation of pseudogenes that lack introns (Kazazian, 2004). L1 elements can mobilize transcribed DNA and are involved in exon reshuffling. Many known proteins incorporate truncated L1 or Alu elements in their transcripts through alternative splicing events (Li et al., 2001). Only 35–40 subfamilies of transposable elements remain
actively mobile in the human genome (Mills et al., 2007). Comparison of human
and chimpanzee genomes indicates that since the two species diverged, human endogenous
retrovirus K (HERV-K) and L1 elements are active in both species, while Alu elements show
approximately 3-fold higher activity in humans (The Chimpanzee Sequencing and Analysis
Consortium, 2005). See also Centromeric Sequences and Sequence Structures, Long
Interspersed Nuclear Elements (LINEs), Retroviral Repeat Sequences, Transposable
Elements: Evolution, and Transposons

Several ancient genome duplications occurred in the evolution of the vertebrate, plant, and
fungal lineages (Van De Peer, 2009). It is difficult to decide if these whole genome
duplication resulted from an autochthonous autopolyploidization or as a consequence of a
between-species hybridization (an allopolyploidization) The latter process is particularly
important in plant evolution and breeding. Following these whole-genome duplications,
many duplicated genes undergo pseudogenization - a few duplicates acquire new functions
following sub or neo-functionalization (Van de Peer, 2009)

Duplicated segments in the human genome are generally enriched in protein coding genes
(Zhang et al., 2005), and hence they have the potential to evolve novel transcripts, either as
whole-gene duplications or through the creation of mosaic genes. For example, 11 new
transcripts have been identified in the 10% of chromosome 22 that originated through
segmental duplication (Bailey et al., 2002). One region of chromosome 16 contains a newly
evolved, unique family of repeats. This gene family, named ‘morpheus’, consists of highly
similar genes evolving so rapidly they show no sequence similarity to known genes from
other organisms, and seem to be under positive selection (Johnson et al., 2001). See also
Chromosome 16, Chromosome 22, and Segmental Duplications and Genetic Disease.

Gene duplications can be either DNA or RNA mediated. RNA mediated duplication results in
genes that have lost introns and regulatory regions of the original gene; consequently the
rate with which duplicated genes turn into pseudogenes is much higher for RNA than for
DNA mediated duplications. However, because the former occur at a much higher rate, about
half of the functional duplicated copies in mammals were determined to have originated
through RNA intermediates (Jun et al, 2009). Using comparative genome analyses, Cicarelli
et al. identified 22 primate-specific gene duplications that are maintained as a single copy in
other metazoan genomes (Cicarelli et al., 2005). Eighty-two per cent of these duplications
are part of genome regions that underwent recent segmental duplications.

Recent variations in the number of paralogues in the lineage leading to humans and within
the human populations are considered to reveal genomic regions under selective pressures
(Gokcumen et al., 2011; Han et al., 2011). In asymmetric evolution after duplication (AED),
one duplicate evolves or degrades faster than the other and often becomes functionally or
conditionally specialized. In a study on asymmetrically duplicated genes, confirmed
duplicated genes sets identified across 13 vertebrate genomes were enriched in functional
categories related to neuron differentiation and response to external stimuli (Prosdocimi et
al., 2012).
Human Orthologues in Other Genomes

Many proteins evolved early in the metazoan lineage and have orthologues in invertebrate genomes, in fact only 7% of the protein motifs in humans are vertebrate specific and it appears most of the protein complexity is due to shuffling of existing domains. The initial human genome sequence contained detectable homologues to 61% of proteins found in Drosophila, 43% of proteins found in Caenorhabditis elegans, and 46% of proteins found in Saccharomyces cerevisiae. Thirteen hundred and eight groups of proteins, encompassing 3129 human proteins, were found to contain at least one orthologue in each of the four species (human, fruitfly, nematode and yeast). These groups of proteins represent basic housekeeping functions in the cell, including respiration, transcription, translation and membrane functions. Of these groups, 564 contained only one orthologue (and no paralogues) from human, fruitfly, nematode and yeast (Lander et al., 2001) representing genes that had not undergone duplication or modification. This is a small percentage of the complete gene complement and indicates the extensive occurrence of gene duplication in the evolution of lineages. The large number of duplicated genes poses a challenge for identification of orthologues among eukaryotic genomes. Consequently, the numbers of orthologous gene sets vary with detection method. An analysis using four different methods for orthologue detection found 7, 663 orthologs shared between humans and C. elegans (or ~38% of C. elegans proteins) and illustrates the complexity of finding orthologs between two species (Shaye and Greenwald, 2011). In analysis of the chimpanzee genome, 13,454 pairs of human and chimpanzee genes were designated as orthologues with high-quality alignments, while addition of rat and mouse sequences reduced the number of unambiguously orthologous genes to 7043 (The Chimpanzee Sequencing and Analysis Consortium, 2005). See also Alignment: Statistical Significance, Sequence Similarity, and Similarity Search.

Orthologous regions between the genomes are not limited to coding regions of a genome. The conserved noncoding portions of the genome (so-called 'dark matter') have been analyzed by comparative genome analyses (Johnson et al., 2005). Multiple stretches of the human genome are identified as being extraordinarily conserved across large evolutionary distances (called 'ultra conserved elements', or UCEs). The 481 regions of the human genome that are over 200 bp in length are 100% identical between human, rat and mouse genomes and many of them are also highly conserved in chicken, dog and fish (Bejerano et al., 2004). Most of these UCEs lie outside of exons, are under stronger purifying selection than non-synonymous sites in protein-coding genes (Katzman et al., 2007) and still await functional assignment and explanation for such remarkable sequence conservation. The 0.14% of the human genome consisting of regions of less striking conservation, but still of high similarity, are found in human and four other vertebrate genomes (mouse, rat, chicken, and Fugu rubripes)(Siepel et al., 2005). These highly conserved elements (HCEs) are longer than UCEs and only 42% of them overlap with known exons. The reason for conservation of HCEs is unknown, as in the case of UCEs, but roles of control in gene expression and post-transcriptional regulation are suggested based on individual examples.

In addition to studying conserved regions, which may provide a hint towards functionality, it is also interesting to look into the fastest evolving (compared to other vertebrates) regions of human genome. 34 498 genomic regions that are ≥96% identical in chimpanzee, mouse and rat genome, but show changes in the human genome, were examined (Pollard et al., 2006).
Only approximately 20% of these regions overlap with exons and 202 show evidence for accelerated evolution in the human genome. Many of the 202 human-accelerated regions are located either in introns of the genes related to transcription and DNA binding or adjacent to such genes.

**Xenologues in the Human Genome**

Initial analyses of the draft human genome were interpreted to suggest that the human genome contained 113–223 genes that probably originated from horizontal gene transfer from bacteria directly into the human lineage (Lander et al., 2001). Given a close association between a prokaryotic symbiont and a eukaryotic host, gene transfer into the nucleus of the eukaryotic host, even in case of multicellular animals, is possible (e.g. Kondo et al., 2002). (A well studied example for bacteria-to-eukaryote transfer is the many mitochondrial genes that now reside in the nucleus.) However, few, if any, of the postulated bacteria to human transfers have upheld closer scrutiny (Andersson et al., 2001; Salzberg et al., 2001). A reanalysis by Salzberg et al. (2001) has shown that the number suggested initially was affected by a species-sampling effect (i.e. by the number of nonvertebrate genomes that were included in the analyses). Differential gene loss might also produce similar results (Andersson et al., 2001). In addition, the direction of potential horizontal gene transfer remains unclear. Thus, the existence of putatively transferred genes directly from bacteria to the human lineage remains unconfirmed and requires additional analyses with more genomic data. See also [Bacterial DNA in the Human Genome](#), and [Homologous, Orthologous and Paralogous Genes](#)

**The Human Microbiome**

Although no recent xenologues were confirmed to be present in the human genome, humans are home to a complex coevolved microbial community. The small generation times, lack of nucleus, and unicellular life cycle of bacteria make them conducive to relatively rapid evolution compared to humans. As eukaryotes evolved over time and developed more complex body plans, prokaryotes adapted to inhabit these newly developing niches. The limitations on the effective population size in these developing niches, imposed by host number, cell number, cell space, and population bottle necks during host transmission resulted in selective sweeps and specialization in colonizing different eukaryotic tissues (Toft et al., 2010). Currently each of us has about 100 trillion bacterial cells found in various locations from the skin to the lining of the alimentary canal and the urogenital tract. The heaviest colonization on the human body occurs in the gut or large intestine where densities approach $10^{11}-10^{12}$ cells per gram of colon contents (Walter and Ley, 2011). Humans are born sterile and are colonized during development with organisms from the environment, initially during passage through the birth canal and through subsequent contact with the primary caregiver. Factors such as breast-feeding and vaginal birth increase the similarity between maternal and infant microbiome until the age of 2.5, when the microbiota of the children becomes more unique, more stable, and more like that of an adult (Parfrey and Knight, 2012). The collective number of different species associated with human intestine is ~1000-1500, while the number of species associated with any single individual is ~160 species, suggesting distinct and adaptable symbiotic populations relative to the environmental parameters specific to each individual. At a larger taxonomic scale microbiota cluster with respect to the host diet (herbivores, omnivores, carnivores, (Fraune and Bosch,
Orthologues, Paralogues and Horizontal Gene Transfer in the Human Holobiont – Page: 7

2010); however, within primates the composition of the gut microbiota tracks the evolutionary history of the host organism (Ochman et al., 2010), revealing a tight co-evolutionary relationship.

Figure: To the left of the body, underlined in red, are the number of human cells which make up the average human body, to the right underlined in purple are the number of prokaryotic cells associated with different locations of the body. The units are bacterial cells per ml, and thus the cumulative amount of prokaryotic cells in each organ is much greater. Additionally, beneath the body is a chart comparing the number of human cells and genes in the human genome (in red) to the number of cells per ml of bacteria in the large intestine, one of the most heavily colonized areas of the human body, and the number of non-redundant prokaryotic genes isolated from the large intestine (in purple).

Human genomes are 99.9% similar between individuals, however the genetic material of the microbiota between even closely related individuals is 70-90% different (Parfrey and Knight 2012). The microbiota contain approximately 150 times more non-redundant genes than in the human genome, suggesting functional flexibility as an important role of the microbiota (Qin et al, 2010). Metagenomic analysis of fecal samples collected from 124 individuals were pooled and revealed 3.3 million non-redundant genes across all samples, 8% of these were genes shared between at least 50% of subjects, while 72% were rare genes present in less than 20% of subjects (Qin et al, 2010). Despite differences in composition between individuals, microbiomes appear largely functionally equivalent (Walter and Ley 2012). Thus, the genetic information present in humans is a composite of Homo sapiens genes and genes present in the genomes of the trillions of microbes that colonize our adult bodies (Turnbaugh et al, 2006). When functional categories of genes were compared between the gut microbiota and the human genome using odds ratios, the gut microbiota showed a significant enrichment in genes involved in metabolism, which were underrepresented in the human genome (Gill et al., 2006). Metabolic specialization encourages high species diversity and niche partitioning related to substrate preference (Spor et al., 2011). ‘Our’ microbial genomes (the microbiome) encode metabolic capacities that we have not had to evolve in our nuclear genome.

See also: Metagenomics and Microbial Communities, Endosymbionts
**Possible roles of the Microbiota**

Human bacterial symbionts contribute to the absorption of carbohydrates, lipids and micronutrients, as well as metabolism of xenobiotics and toxins (Gill et al., 2006). It is difficult to gauge the extent of the impact of microbiota in human physiology is, however, as fecal transplants of microbiota from healthy subjects have been used to alleviate chronic *Clostridium difficile* infections in patients where antibiotics are ineffective; in 95% of cases colitis caused by dysbiosis was alleviated after transplantation (Gough et al., 2011). There is a well-demonstrated correlation between states of dysbiosis and disease in humans, diseases such as Crohn’s, IBD, allergies, celiac’s disease, gastric cancer, autism, obesity, anorexia, type II diabetes, type I diabetes, multiple sclerosis, and rheumatoid arthritis (Clemente et al., 2012); however, it is not known if correlation can be attributed to causality as many of the diseases where a dysbiosis is noted are autoimmune disorders, and the tight link between the microbiota and immune function is difficult to tease apart. Much has been learned about the relationship between endogenous microbiota and the host organisms by using mouse gnotobiotic models (i.e., animals that harbor only a defined set of microorganisms). Studies using gnotobiotic mice have shown that microbes are involved in the development innate immunity through mucosal fortification, and additionally play a definitive role in development of the adaptive immune systems. Such studies illustrate a profound difference in physiology, especially relating to host defense, between animals with and without microbiota, implying co-evolution between the host and its symbionts for the purpose of collaboration against infectious agents (Lee and Mazmanian, 2010). Supporting this hypothesis, studies of specific members of the microbiota, such as *Lactobacillus* have shown such organisms have a protective effect against many forms of intestinal dysbiosis by inducing protective modifications to both the mucin and the epithelial barrier, secreting antimicrobial substances, and replenishing suppressed beneficial microbiota (Mattar et al., 2001). Additionally ‘parasitic’ organisms such as helminthes may modulate the immune system and elicit a protective effect against certain types of dysbiosis, alleviating symptoms of arthritis, multiple sclerosis, type I diabetes, and Crohn’s disease (Rook et al., 2012). These examples suggest long term co-evolution toward the currently established and delicately balanced relationship between the host and the microbiota, which is essential to maintain homeostasis (Rook et al., 2012).

**Horizontal Transfers in the Human Microbiome**

The field of horizontal gene transfer (HGT) within the human microbiome has exploded recently with the Human Microbiome project funded by NIH. In one study a total of 13,514 high confidence HGT genes were identified in the genomes of 308 human microbes (Liu et al., 2012). Most of the genes were involved in either catalysis or metabolism, again highlighting the important role of the microbiota in metabolic functions. In another study a screen of 2,235 human associated bacterial genomes from different body sites showed a network of 10,770 unique, recently transferred genes; in most of which the HGT occurred between isolates from ecologically similar, but geographically and phylogenetically distinct environments (Smillie et al., 2011). Bacteria involved in transfers often share similar body sites, oxygen tolerance or ability to cause disease, indicating an important role for ecology (environment) in driving these networks of gene sharing. A classic example of HGT is the transfer of antibiotic resistance genes; such genes have a selective advantage in the gut.
environment and can be transferred from the outside environment to the gut microbiota via food sources. In a study by Lester (2006) volunteers were fed a strain of vancomycin resistant *Enterococcus faecium* isolated from a chicken; subsequently vancomycin resistance was transferred to the human gut *E. faecium* in these volunteers, providing evidence for food as a reservoir for possible HGTs. An example of HGT impacting the nutrients available to the holobiont was recently discovered in the human microbiome of Japanese individuals. Genes for porphyranases, alginases and agarases, enzymes which facilitate the breakdown of carbohydrates in algal cell walls, were transferred from marine algal parasites, to the gut organism *Bacteroidetes plebeius* (Hehemann et al., 2010). These HGTs allow the gut bacterium to utilize seaweed as a carbon source, and confers a secondary benefit to the human host, who can now utilize metabolites released by the bacterium after the food source is broken down. In these examples HGTs increase the fitness primarily of the microbes associated with the human gut, and have a secondary benefit on the human host.

### The Hologenome Theory of Evolution

The preceding observations support the hologenome theory of evolution (Fraune and Bosch, 2010, Rosenberg et al., 2009): the unit of selection is the holobiont, in this case the human host and the microbial symbionts. The microbial symbionts profoundly affect the fitness of the host organism; in turn the evolutionary trajectory of the microbiota is impacted by the health and well being of the human host. Selection acting at either level, the microbiota or the human host, will act on the collective set of genes, the hologenome, in such a way that genes maintained and expressed by any organism present in the holobiont will have an effect on the holobiont as a whole. The holobiont can adapt to changing environmental conditions through acquiring new symbionts, or symbionts already present may acquire new genes and properties through HGT. The interactions between host and symbiont (including commensals and parasites) have a long evolutionary history. Disturbance of these long established interactions may have surprising consequences for human health and well-being (Rook, 2012).

### References:


Rook GAW (2012) A Darwinian View of the Hygiene or "Old Friends" Hypothesis. Microbe 7 (4) 173-180


**Keywords:**
Orthologous
Paralogous
Symbiosis
Holobiont
Hologenome
Co-evolution
Gene duplication
Horizontal Gene Transfer

**Glossary Terms**

constrained bases: nucleotide positions in a gene or genome that have experienced fewer substitution events than expected for an non-functional part of the genome

dysbiosis: microbial imbalance, often occurring in the digestive tract and on the skin.

microbiome: The sum of the genomes of all the microbial symbionts carried by a host.

ncRNA: non-coding RNAs are RNAs that do not encode a protein.

neofunctionalization: one of the possible processes following a gene duplication in which one of the paralogs acquires a new function

pan-genome: the pan-genome of a taxonomic unit (often a species) describes the set of all genes present in at least one member of that taxonomic unit.

positive selection: describes natural selection leading to fixation of a mutation because it provided a benefit to the organism. Also known as Darwinian selection.

pseudogenization: the process of a protein coding gene losing its ability to code a protein.

selective sweeps: the fixation in a population of a gene that provides a selective advantage. The gene that provides the increased fitness can carry a neighboring part of the genome with it to fixation.

subfunctionalization: one of the possible processes following a gene duplication in which the two paralogs each retain some non-overlapping functions of the ancestral

**Key concepts**

Orthologous structures or sequences in two organisms are homologues that evolved from the same feature in their last common ancestor; orthologues reflect organismal evolution.

Paralogues are homologues whose evolution reflects gene duplication events.
Genomes can evolve by acquiring genes through horizontal gene transfer, or from the fusion of complete genomes through symbiosis.

Individual humans can differ slightly in genome content with variation related primarily to deletions or regions of segmental duplication.

Apparent differences in complexity between species may be due to a varying amount of non-coding regulatory sequence, regulating a fairly stable core of protein-coding genes.

Repeated sequences derived from transposable elements comprise a large portion of the genome and can have a significant role in gene duplication through the formation of pseudogenes that lack introns.

Duplicated segments in the human genome are generally enriched in protein coding genes and have the potential to evolve novel transcripts, either as whole-gene duplications or through the creation of mosaic genes.

Variations in the number of paralogues in humans reveal genomic regions under selective pressures.

Orthologous regions among genomes are found in both protein coding exons and non-coding regions of the genome. Rapidly evolving regions of the human genome include intergenic regions that may be important for gene regulation.

Humans can be viewed as a holobiont, a complex ecosystem whose evolutionary fitness is determined by interactions of the host and the microbiota.

The microbiome, the sum of microbial genomes carried in our symbionts, encode metabolic capacities that we have not had to evolve in our nuclear genome.

**Further Reading**


