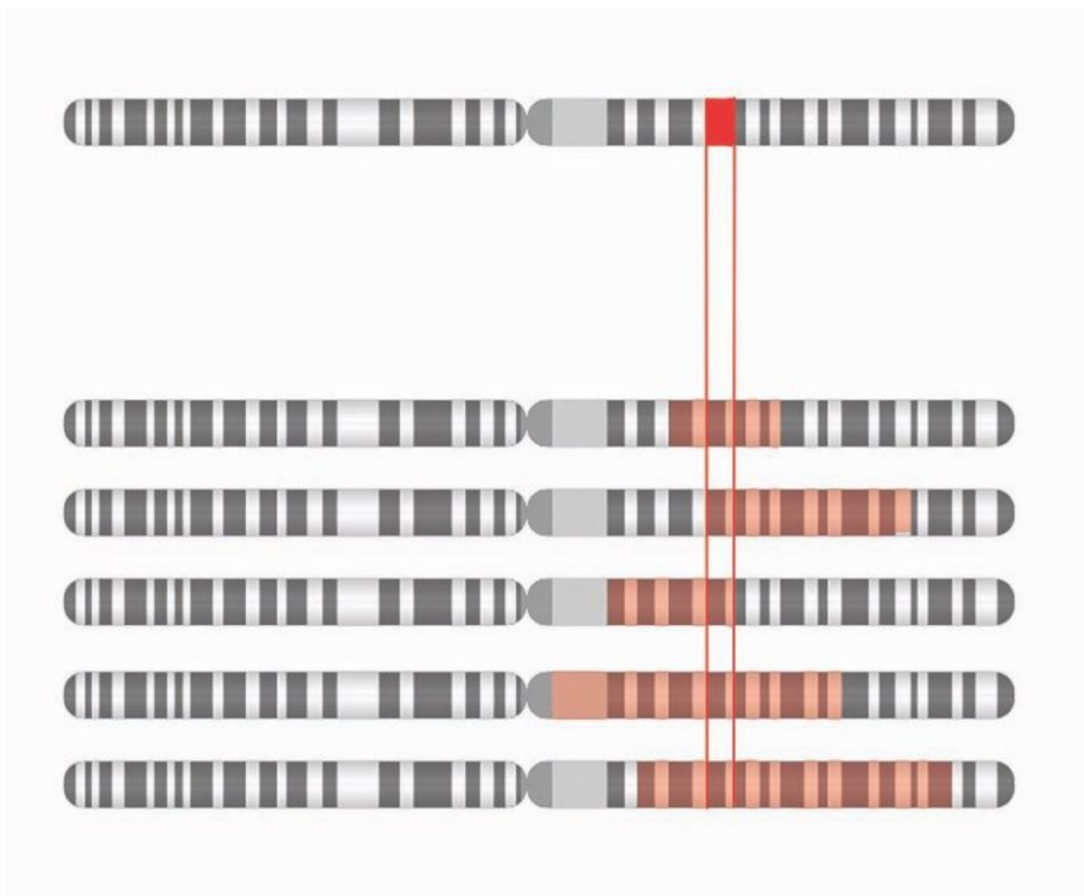


Comparative Genomics

[See online here](#)

Comparative genomics allows researchers to use a robust repertoire of techniques to compare the gene sequences of different organisms. Among these techniques are analytics, which can be used to not only draw structure/function comparisons between species but to also make predictions about unknowns. This knowledge is cutting edge in the development of new strategies to remedy human disease.



Comparative Genomics

Conserved DNA sequences allow researchers to gain understanding of the essential genes and proteins that organisms share. Knowing how genes relate to biological systems allow scientists to key in on their specific properties and thus exploit the differences between organisms. This is very important as it offers innovative approaches for combating pathogens. Understanding the evolutionary relationships between two organisms is essential in creating a coordinated attack against infection.

The Central Dogma of Molecular Biology describes how information flows from DNA to RNA and then to protein; the latter being the business end of the organism-level

transaction. As DNA sequencing rises in resolution and falls in cost, wider applications for genomics are found. Agriculture, biotechnology, pharmacology and global human standards of living all look to improve as a direct effect of comparative genomics.

Comparative genomics can be loosely defined as the large-scale comparison of genomes in order to understand the biology of individual genomes and to extract general principles that apply to groups of genomes. This definition might seem difficult to comprehend at first. To better understand this definition, one can dissect it. When we compare genomes to each other, we assume that biological sequences, structures, and functions are shared across different organisms.

Sequencing Genomes

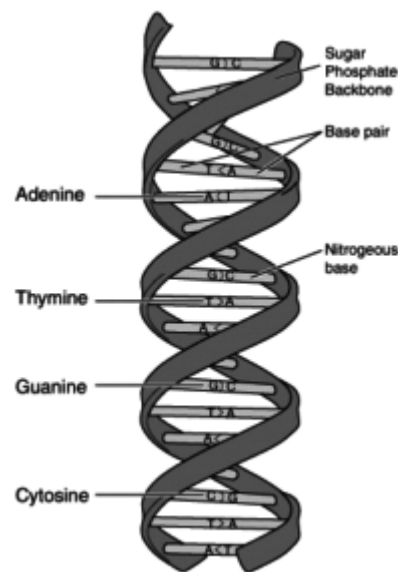


Image: "DNA structure and bases" by AutisticPsycho2. License: Public Domain

The genomes of virtually all living organisms are composed of **deoxyribonucleic acid (DNA)**. DNA is found in a double helix configuration constructed from the association of its four building blocks - adenine (A), thymine (T), cytosine (C) and guanine (G). A DNA chain consists of sequences that code for different proteins as well as regulatory sequences that turn the expression of genes off and on.

Once a given genome has been sequenced and uploaded to a database, comparisons of the general features, such as the number of genes and where those genes reside on chromosomes, can be made.

Polymerase Chain Reaction (PCR) is now widely used as a technique to amplify segments of chromosomal DNA or **complementary DNA** (DNA that is derived from reverse transcription of RNA). In order to perform PCR, the sequences of the target region must be known.

Oligonucleotide primers are designed as short single-stranded DNA fragments that are complementary to the target region sequences. A sample of DNA is denatured through light heat, which effectively separates its two strands. DNA polymerases can then attach to the 3' end of the designed primers and synthesize new strands of DNA.

Comparing Genomes



Image:
"Arabidopsis thaliana"
by Carl Axel
Magnus
Lindman -
<http://runeberg.org/nordflor/194.htm>
I. License:
Public
Domain

Genome size does not always correlate with advanced evolutionary standing, nor do the number of genes reveal the size of a genome. An example of this is demonstrated when comparing the genomes of two completely sequenced organisms - *Arabidopsis thaliana* and *Drosophila melanogaster*. *Arabidopsis* has 8 million less nucleotide base pairs than *Drosophila*, but it has 12,000 more genes. In fact, the small *Arabidopsis* plant has a similar number of genes as we do.

Chromosome-level comparisons gleam insight into the physical co-location between two organisms. This kind of information is provided by higher-resolution techniques applied to DNA sequences. The physical co-location or **synteny** is vital information that researchers use to gain clues about similarities in the groupings of genes. Synteny is generally conserved for genes with related functions, an observation that points to common evolutionary ancestry.

Human and rodent X-chromosomes are a brilliant illustration of synteny; scientists have marked the existence of reciprocal syntenic groups on chromosome 20 and chromosome 2, respectively. The arrangement of the genetic information is conserved along this entire block. Conversely, changes to the human and rodent genomes over millions of years is reflected in the rearrangement of certain genes.

Homologous genetic material, i.e., DNA that is similar between two organisms, allows scientists to compare blocks of genes. Many of the enzymes of intermediary metabolism among related organisms have conserved genes that are similarly arranged on chromosomes. Using the method of analytics, researchers identify similar regions of

chromosomes that correspond to regulatory sequences as well as functional loci of genes for the subsequently produced protein's function.

Benefits of Comparative Genomics



Image: “Under a high magnification of 15549x, this scanning electron micrograph (SEM) depicted some of the ultrastructural details seen in the cell-wall configuration of a number of Gram-positive *Mycobacterium tuberculosis* bacteria.” Photo Credit: Janice Carr; Content Providers(s): CDC/ Dr. Ray Butler; Janice Carr – This media comes from the Centers for Disease Control and Prevention’s Public Health Image Library (PHIL), with identification number #8438. License: Public Domain

Understanding the similarities and differences between the genes of diverse organisms allows us to study potential cures for diseases that have eluded us for hundreds of years. The knowledge of comparative genomics is at the forefront of rational drug design because many pharmaceuticals act to inhibit enzymes that are produced from genes.

***Mycobacterium tuberculosis* (MTB)** is the causative agent of [tuberculosis](#) in humans and their only known reservoir. MTB was the cause of the “White Plague” of the 17th and 18th centuries in Europe, a time period when approximately 100% of the European population was infected with the microorganism, and 25% of all adult deaths were attributed to the disease.

Today, **tuberculosis (TB)** is the leading bacterial cause of death and affects over one-third of the entire world population.

The primary route of transmission of TB is through **aerosols** when individuals carrying the disease [cough](#) active droplets containing the infectious bacteria. These active droplets can remain infective for hours.

MTB is a nonmotile rod-shaped bacterium that is an obligate aerobe. Complexes of MTB are always found in the oxygen-rich upper lobes of the [lungs](#), where they act as facultative intracellular parasite infecting macrophages. The aim of researchers is to understand various components of the mycobacterial cell wall, the metabolic pathways that lead to the biosynthesis of these constituents (in this case sterols) and how these pathways can be inhibited.

Isoprenoids are vital to several core bacteria cellular functions. Mycobacteria produce isoprenoids via a metabolic route that differs from the one used by humans – the **Methyl Erythritol Phosphate Pathway (MEP)**. Humans produce cholesterol via the mevalonate pathway (MEV), which synthesizes HMG-CoA through Acetyl-CoA.

Regardless of the route that is taken to generate the start and repeat unit, a common characteristic of the isoprenoids is that they are derived from two basic 5-carbon molecules: **isopentenyl diphosphate (IPP)** and its isomer **dimethylallyl diphosphate (DMAPP)**. In the MEP pathway, these 5-carbon building blocks are generated from pyruvate and glyceraldehyde-3-phosphate.

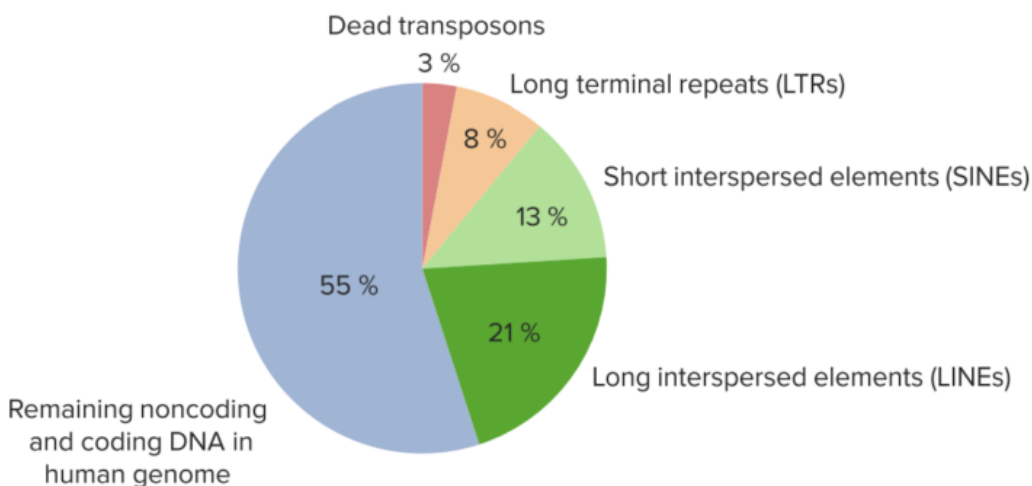
Subsequent reactions convert MEP to IPP and DMAPP for the generation of the three essential groups of compounds that arise from isoprenoids with respect to mycobacteria – sterols, ubiquinones, and dolichols. **Sterols** provide membrane stability as a primary function. **Ubiquinones** are electron carriers and the major component of bacterial aerobic respiratory chains.

Understanding the similarities and differences between the genomes of these organisms is at the forefront of the rational drug design for eradicating tuberculosis. Because these organisms possess different pathways for the synthesis of the same vital compound, one pathway can be chemically attacked without blocking any of the enzymes in the pathway of the other. In this case, the enzymes of the mycobacterial MEP pathway can be targeted without interfering with cholesterol biosynthesis in the patient.

Indeed, **fosmidomycin** is a phosphonic acid compound developed by researchers to inhibit **deoxyxylulose phosphate reductoisomerase (DXR)**—the enzyme that catalyzes the dedicated step of the MEP pathway MTB. As the understanding of the genome and proteome of MTB advances, further-developed chemical compounds will one day likely lead to the eradication of tuberculosis globally.

Transposable Elements

Transposable elements move from one location to another. They may be duplicated, or excised and inserted elsewhere.



LINEs

Long interspersed elements (6,000 bp)

- 21 % of the human genome
- Transpose themselves
- Retrotransposons: Contain reverse transcriptase

SINEs

Short interspersed elements (300 bp)

- Nested in LINEs
- Use LINE machinery
- Interrupt genes

LTRs

Long terminal repeats

- Reverse transcriptase

Dead transposons

- No longer have machinery to move

References

Shigi Y (1989) Inhibition of bacterial isoprenoid synthesis by fosmidomycin, a phosphonic acid-containing antibiotic. *J Antimicrob Chemother* 24: 131–145.

Brown AC, Parish T (2008) Dxr is essential in *Mycobacterium tuberculosis* and fosmidomycin resistance is due to a lack of uptake. *BMC Microbiol* 8: 78. doi:10.1186/1471-2180-8-78.

Nair SC, Brooks CF, Goodman CD, Strurm A, McFadden GI, et al. (2011) Apicoplast isoprenoid precursor synthesis and the molecular basis of fosmidomycin resistance in *Toxoplasma gondii*. *J Exp Med* 208: 1547–1559. doi:10.1084/jem.20110039.

Joseph Paul Calihan (2014) Ghost Killer – Targeting Sterol Biosynthesis in *Mycobacterium tuberculosis* (MTB). Texas Tech University

Jomaa H, Wiesner J, Sanderbrand S, Altincicek B, Weidemeyer C, et al. (1999) Inhibitors of the nonmevalonate pathway of isoprenoid biosynthesis as antimalarial drugs. *Science* 285: 1573– 1576.

Small, K. S., Brudno, M. *et al.* A haplome alignment and reference sequence of the highly polymorphic *Ciona savignyi* genome. *Genome Biol* 8, R41 (2007).

Stein, L. D., Bao, Z. *et al.* The genome sequence of *Caenorhabditis briggsae*: a platform for comparative genomics. *PLoS Biol* 1, E45 (2003).

Stein, L. D., Bao Z. *et al.* Genome sequence of the nematode *C. elegans*: a platform for investigating biology. The *C. elegans* Sequencing Consortium. *Science* 282, 2012–2018 (1998).

Legal Note: Unless otherwise stated, all rights reserved by Lecturio GmbH. For further legal regulations see our [legal information page](#).