IDC School of Design
अभिकल्प विद्यालय
IIT Bombay

Project 1 (P1)

Data visualisation on Comparative Genomics

Guide:
Prof. Venkatesh Rajmanickam

By:
Niharika Kumawat
18U130020

Approval

The B.Des Design Project – 1 titled “Data Visualisation in Comparative Genomics” by Niharika Kumawat, Roll Number 18U130020 is approved, in partial fulfilment of the Bachelor in Design Degree at the IDC School of Design, Indian Institute of Technology Bombay.

Project Guide:

Chairperson:

Internal Examiner:

External Examiner:

Declaration

I hereby declare that this document contains my original ideas and exploration. I have adequately cited and referenced the original sources wherever they have been used as a part of this project. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/ source in my submission. I understand that any violation of the above will because for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Niharika Kumawat
18U130020

IDC School of Design, IIT Bombay
November 2021

Acknowledgements

I am grateful to my guide Prof. Venkatesh Rajmanickam for his guidance and valuable inputs which helped me shape my project towards its completion.

My sincere gratitude to my panel, Prof. Jayesh Pillai, Prof. Vivek Kant and Prof. Swati Pal for their valuable feedback and comments over the various stage presentations.

Lastly, I'm thankful to all my professors, family and friends for their constant support and encouragement throughout the course of this project.

Contents

Abstract

This project aims to deliver a data visualisation around the complex topic of “Comparative Genomics”. The aim is to simplify a complex topic as such into a form which can be ingested by the general public with general insights and induce curiosity in them. The outcome is a web interface combining narrative and two interactive visualisations comparing the details of 30 widely known species of animals. A lot of static and dynamic tools exist to visualise genomic data but most of them are often very complex for any non-field expert user to use. This project aims to offer an entry point to the general audience to get introduced to the field of Comparative Genomics and its basics.

Two levels of comparison are undertaken in this project, one is general comparison of basic details with the first visualisation and second is comprehensive sequence comparison between the chosen species. The user can explore on their, details about the species of their interest from the given species by interacting with the tool. Genomics and Comparative Genomics being a growing field of research are important topics of interest, knowledge of which is important and can help individuals discover new opportunities in various fields.

1 Introduction

1.1 Given brief

Public Understanding of Science through visualization
The project will entail data visualization design of a science phenomenon of student’s choice. There has to be a strong 'data' component which entails analysis, synthesis, and transformation of data in an original manner that hopefully leads to newer understandings. One can do a visualization of pre-synthesized data also, but the focus is not on data transformation, but designing visual forms for communication.

For this project, I decided on “Comparative Genomics” as the topic of my data visualisation after exploring a few fields that I found interesting.

1.2 Comparative Genomics

Genome is the sum total of an organism's DNA i.e. the complete set of genetic instructions. Each genome contains all of the information needed to build that organism and allow it to grow and develop. Thus, comparative genomics is basically comparing these genetic details among different species using which we can pinpoint regions of similarity and difference between various species at the fundamental level.

1.3 Motivation

Being based on basic genetics and evolution, I felt it’s a fundamental topic, pre-requisites of which, are introduced in school but the topic itself is not explored. Thus, being simplified, it has a potential to induce curiosity in the audience to explore more things in the broader scope of Evolution, Genetics and Genomics.

1.4 Scope and Goal

For this project, the aim was to simplify the complex topic of Comparative Genomics. The aim is not towards providing expertise in this field but to give an introductory idea into Comparative Genomics, it's why's and how's for a general audience.

The goal is to connect what people already know about genes and evolution with the growing field of Comparative Genomics using datasets of common animals that we see around us.

1.5 Target audience

This visualisation is not curriculum designed thus it's aimed at supplementary learning for kids or even for adults. The visualisation is aimed for a general public understanding thus the demographics aren't restricted to a defined age group or occupational field but following are the basic pre-requisites to understand the visualisation:

Level of literacy: Secondary school or above
Basics of cells, DNA and genes are required for the audience to understand and interpret this visualisation. These topics are introduced at Secondary School level biology in most schools.

Expertise in domain: Novice
As the aim is to simplify the topic, the audience is expected to be novices which may or may not have any prior knowledge in the field of Comparative Genomics.

Interest in domain: Moderate to High
Certain topics like evolution and genetics are topics, basics of which are already introduced and taught early on. Comparative Genomics on the other hand is a growing field which can be considered an extension (with regards to this visualisation) of those basics topics that people already know. Thus, this visualisation is aimed towards those self-motivated kids and adults who want to explore more in this area, knowledge of which might not be necessary but imbibes curiosity in them to learn and know more about organisms around them.

2 Literature review

2.1 The visual display of quantitative information (Book by Edward R Tufte) [1]

Written by the Edward Tufte, considered as one of the pioneers in information design and data visualisation, this book begins from the history of data visualization covering everything from theory and principles of building data graphics. Following are some of the highlights from my reading:

- Directing attention towards substantive content rather than towards methodology and technique of coding
- Importance of connecting to the semantics and content of data
- Advertising the data not the technique
- Physical realization of the data
- Using visualisation such that the insights are both understood and retained
- Number of information carrying dimensions depicted should not exceed the number of dimensions in the data
- Focus on showing data variation instead of design variation
- Avoiding bulky grids and unintentional optical art caused by pattern fills
- Maximizing the data ink ratio within reason
- Erasing redundant ink within reason

Fig 1: A redrawing of Jon Snow’s Cholera map plotting the location of deaths from cholera in Central London, Sept 1845 which helped in identifying the cause of the disease through water pumps.

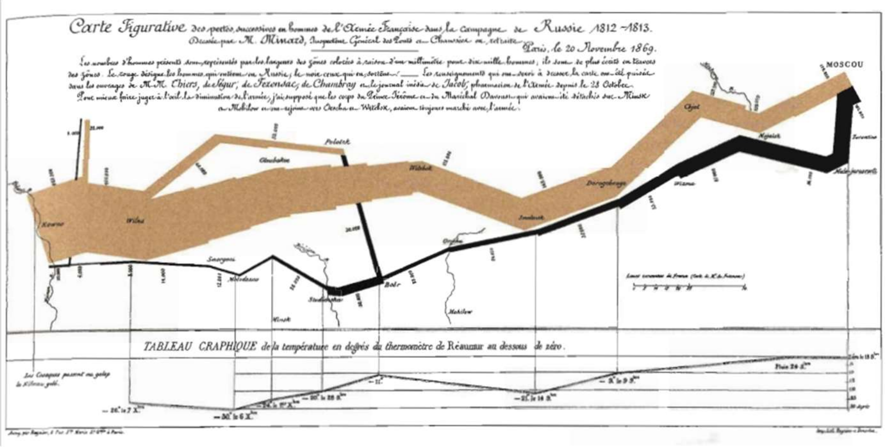


Fig 2: A classic visualisation of Charles Joseph Minard, showing the terrible fate of Napoleon's army in Russia.

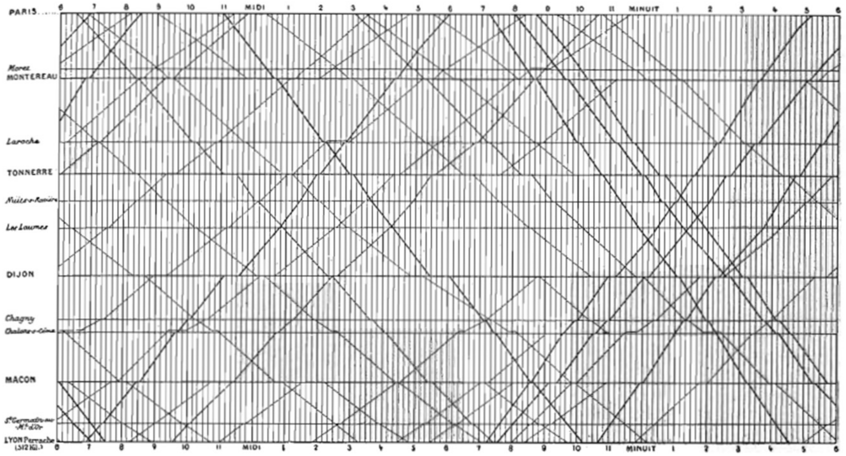


Fig 3: E.J. Marey's graphical train schedule from Paris to Lyon in the 1880s.

2.2 **Comparative Genomics** (Paper by *V de Crécy-Lagard and A Hanson)* [2]

“Comparative genomics can be simply defined as the comparison of biological information derived from whole-genome sequences. Comparative genomics therefore began in 1995, when the first two whole organism genomes (for the bacteria Haemophilus influenzae RD and Mycoplasma genitalium G37) were published. Very soon thereafter came bioinformatics tools to compare the genome sequences themselves, and the RNAs, proteins, and gene annotations that can be derived from them. These tools are constantly evolving to deal with the exponential proliferation of sequenced genomes driven by advances in sequencing technology, and to become more comprehensive and user-friendly.”

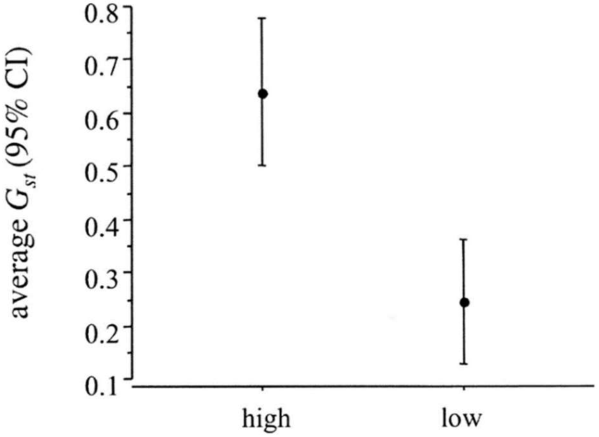
“The dramatic fall in sequencing costs has made it possible to compare the genomes of dozens of closely related species or strains at the DNA level to identify single-nucleotide polymorphisms (SNPs), indels, or large insertions and deletions. The availability of such information is changing epidemiology and evolutionary biology in major ways. For instance, analysing Yersinia pestis strains has now shown how the Black Death agent evolved in China and spread around the world, and comparing modern Helicobacter pylori isolates has allowed tracking of early human migrations. Similarly, it has been possible to track the source strains and their dissemination paths in recent Vibrio cholerae and Staphylococcus aureus outbreaks. For eukaryotes, analysis of diverse yeast genomes has solved the mystery of the origin of lager-brewing yeast. Thus, although these types of analyses have only just started, they have had immediate benefits (e.g., in understanding the spread of infectious disease). In future, they promise to help answer long-standing biological questions such as which molecular determinants characterize species or individuals and how populations evolve.”

2.3 **Designing effective visualisation in biological sciences**

(Workshop by Martin Krzywinski) [3]

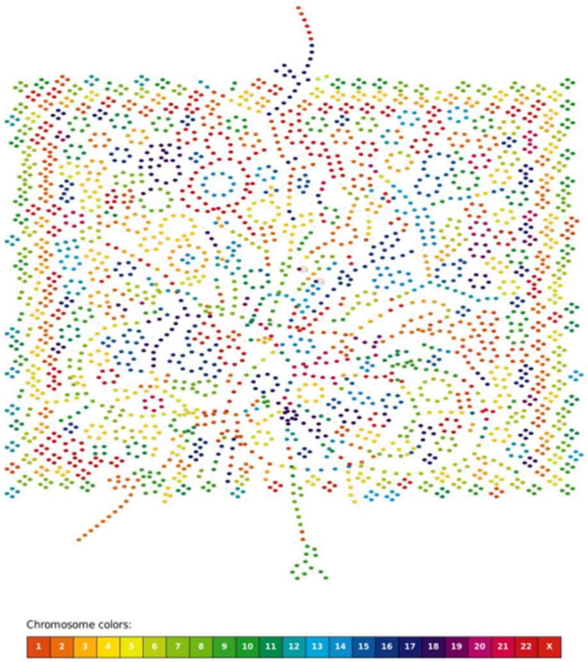
Martin Krzywinski is known for his work in bioinformatics, and the interface of science and art, and is also the one behind the Circos webtool developed specifically for genomic visualisations. In the workshop material he writes about the basics of designing an informative as well as information rich visualization in biological sciences. Few highlights from the workshop slides:

- Legibility in terms of resolution of data points, parsability limits and basics of colour perception & contrast.
- Clarity in terms of reducing redundancy, clarity of the message, reducing excess ink,
- Lastly, attractiveness in terms of grids and layouts



Informative, not information rich

Fig 4: Mean and 95% confidence interval of average Gst based on direct sequencing of fragments within high (N = 8) and low (N = 6) differentiation regions in natural population of house mice



Information rich, not informative

Fig 5: The breaks point graph G(M,R,D,Q,H,C) (obverse edges are not shown) of six mammalian genomes. Alekseyev, M.A. and P.A. Pevzner, Breakpoint graphs and ancestral genome reconstructions. Genome Res, 2009.

3 Secondary Research

3.1 Science articles

Comparative Genomics | https://dx.doi.org/10.1371%2Fjournal.pbio.0000058 [4]

Comparative Genomics | Learn Science at Scitable (nature.com) [5]

Eukaryotic Genome Complexity | Learn Science at Scitable (nature.com)[6]

Read various articles and research papers on Comparative Genomics, the process, resolution at multiple depths, tools used as well as it's relevance. This visualisation is centred around the following findings:

- Major principle of comparative genomics is that common features of two organisms will often be encoded within the DNA that is conserved between the species. [4]
- Comparison of whole genome sequences provides a highly detailed view of how organisms are related to each other at the genetic level. [5]
- Comparative genomics also provides a powerful tool for studying evolutionary changes among organisms, helping to identify genes that are conserved or common among species, as well as genes that give each organism its unique characteristics. [5]
- How big is it? That is usually the first question asked about an organism's genome. A simple comparison of the general features of genomes such as genome size, number of genes, and chromosome number presents an entry point into comparative genomic analysis. A very early lesson learned in the "genomic era" is that genome size does not correlate with evolutionary status, nor is the number of genes proportionate to genome size. [6]
- From the sequence comparison. We have learnt that the similarity between the genomes of two species is largely dependent upon the evolutionary distance between them. The more distantly related two organisms are, the less sequence similarity or shared genomic features will be detected between them. Very similar genomes separated by about 5 million years of evolution (such as human and

chimpanzee) are particularly useful for finding the sequence differences that may account for subtle differences in biological form. [4]

- Dramatic results have emerged from the rapidly developing field of comparative genomics. Comparison of the fruit fly genome with the human genome reveals that about sixty percent of genes are conserved. That is, the two organisms appear to share a core set of genes. Researchers have also found that two-thirds of human genes known to be involved in cancer have counterparts in the fruit fly. [5]
- In addition to its implications for human health, comparative genomics may benefit the broader animal world and ecological studies as well. As sequencing technology grows easier and less expensive, it will find wide applications in agriculture, biotechnology, and zoology as a tool to tease apart the often-subtle differences among animal and plant species. Such efforts might also lead to the rearrangement of our understanding of some branches of the evolutionary "tree of life," as well as point to new strategies for conserving rare and endangered species. [5]

3.2 Existing Visualisations and tools

Circos [7]

“Circos is a software package for visualizing data and information. It visualizes data in a circular layout — this makes Circos ideal for exploring relationships between objects or positions. There are other reasons why a circular layout is advantageous, not the least being the fact that it is attractive.

Circos is ideal for creating publication-quality infographics and illustrations with a high data-to-ink ratio, richly layered data and pleasant symmetries. You have fine control each element in the figure to tailor its focus points and detail to your audience.

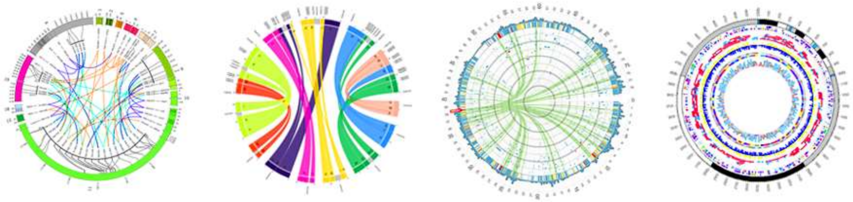


Fig 6: Images created with Circos, illustrating links, ribbons, tiles and a variety of 2D data tracks. If it's round, Circos can probably do it

Circos is flexible. Although originally designed for visualizing genomic data, it can create figures from data in any field—from genomics to visualizing migration to mathematical art.”

Cinteny [8]

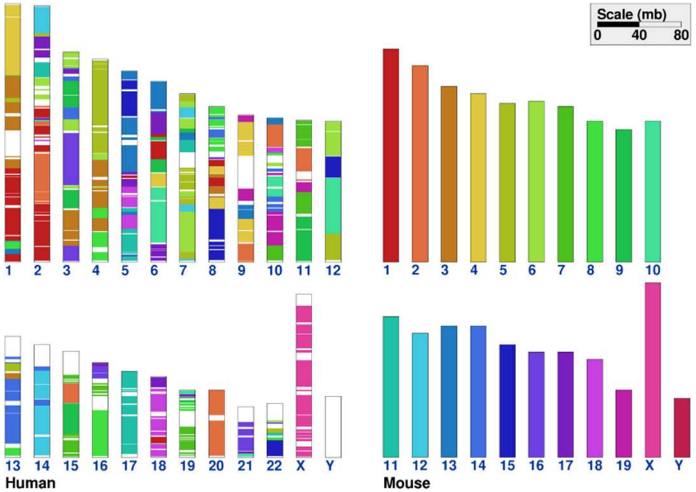


Fig 7: Comparison of Human vs Mouse conserved sequence similarity

Web server that lets you compare two species by either complete genome sequence analysis or comparing single chromosomes.

Genome Browser [9]

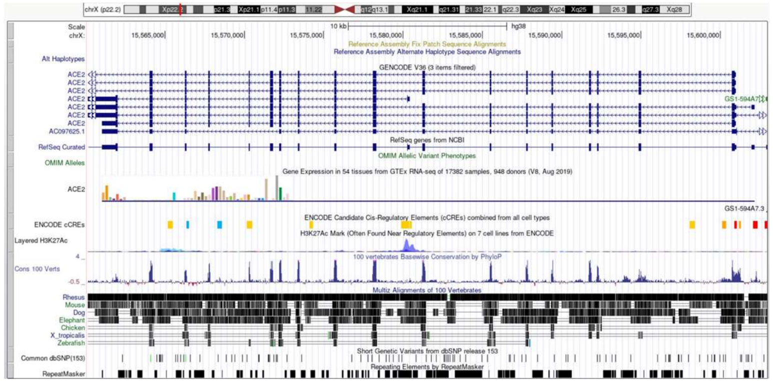


Fig 8: Overview and comparison of a small part of Chromosome X of various genomes to Chromosome X of Reference Human genome.

A very complex online and downloadable detailed plotting tool with multiple custom filters and options to compare various parts of different genomes or even entire genomes.

“It is an interactive website offering access to genome sequence data from a variety of vertebrate and invertebrate species and major model organisms, integrated with a large collection of aligned annotations. The Browser is a graphical viewer optimized to support fast interactive performance and is an open-source, web-based tool suite built on top of a MySQL database for rapid visualization, examination, and querying of the data at many levels. The Genome Browser Database, browsing tools, downloadable data files, and documentation can all be found on the UCSC Genome Bioinformatics website.”

Most of the visualisation and tools I explored are either static or very complex or both. My goals with my ideation will be to make the visualisation interactive as well as a simplified for the audience.

3.3 NCERT

The prerequisites to understand this topic is basic understanding of cell and its components i.e. Nucleus, DNA, Chromosomes, Genes and their structure & basic functions
So, I went through the science NCERT books to understand if and when are these topics introduced in the curriculum and based on that my I defined the pre-requisites for my target audience as above secondary school as these topics are introduced in 9th and 10th grades.

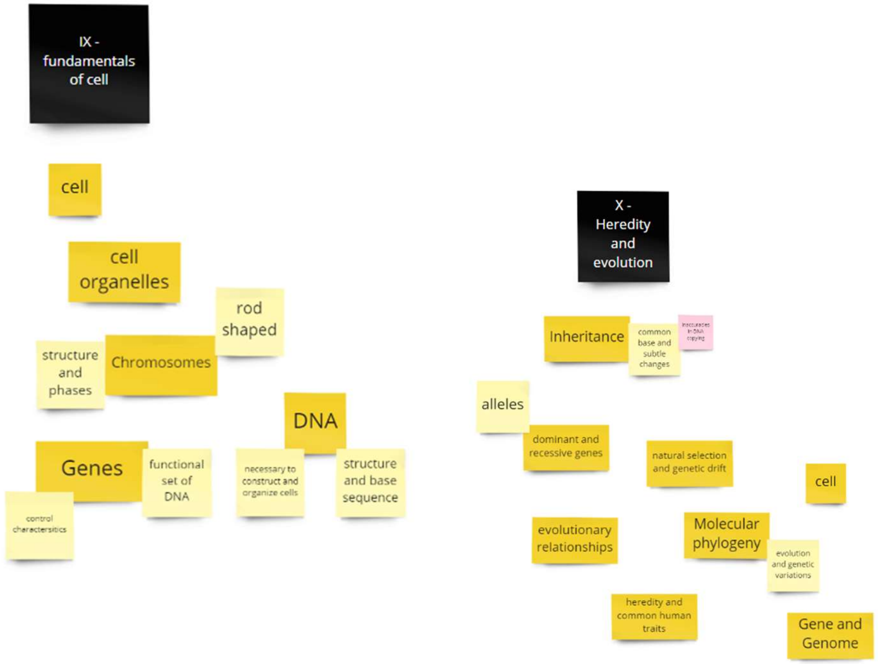


Fig 9: Topics introduced in 9th and 10th grade NCERT textbooks.

4 Primary research

Semi-structured interview with a Secondary School biology teacher

Questions asked to:

- Understand the pre-requisites
- Understand how these topics are approached
- Know about his approach of introducing out of curriculum content
- Explore ways to introduce and simplify complex topics
- Understand the difficulties, relevance and feedback to this topic

Insights

- Cell and genetics are basic pre requisites which introduced at two different points of time
- He mentioned introducing topics out of syllabus which he founds interesting by connecting them to the topics already in their books when students get bored during lectures
- He starts with basics to introduce the topic and also asks the students to look up new research being done in the field.
- Lastly, he made a point about how it's hard for people to connect cell basics like DNA and genes to broader ideas of evolution and gene expression as these topics are taught disconnectedly. So comparative genetics could be a good topic to form that connecting link while also being relevant and a growing research field.

5 Refined scope

Based on these insights, I refined my scope of visualization focused on comparative genomics by tying up the ends with narrative on basic genomics and in the end it's relevance in medicine and other fields.

Basic Genomics (Narrative)

Basics of DNA and genes.

Comparative Genomics (Data Visualisation)

- General
- Sequence

Genomics and health (Narrative)

Relevance of Comparative Genomics

6 Visualisation specifics

6.1 Visualisation typology

It will be a web-based interactive visualisation weaved in with basic narrative for easy understanding.

Structural form: Network visualisation (network visualizations explore the relationship between datasets i.e. the genomes in this case)

Intent: Intent can be narrative, instructive, Explorative or Simulative. This visualisation will be a combination of both narrative and explorative intent which will allow the audience to understand the basics from the text and explore various insights from their interaction with the visualisation.

Information type: Quantitative (Number of genes, Similarity in genomes)

Communication method: Interactive Visualisation

Based on the intent and Scott Berinato's model for types of data visualiastion, this visualiastion is a combination of both Idea illustration and Visual Discovery involving trend spotting, sense making and deep analysis.

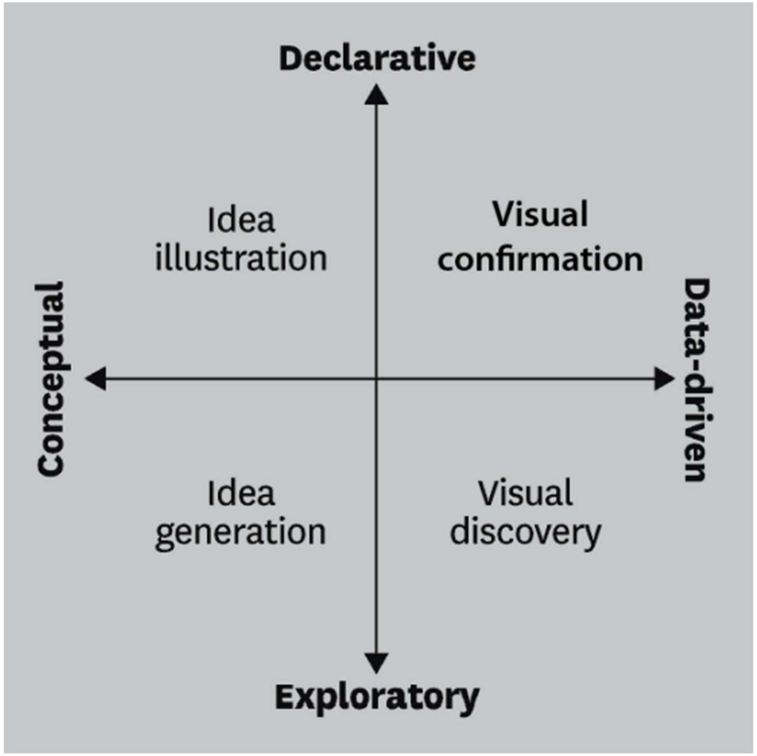


Fig 10: Visualizations That Really Work by Scott *Berinato*

6.2 Dataset, Sources and Tools

List of animals (30 widely known animals)

- Human
- Chimpanzee
- Gorilla
- Orangutan
- Rhesus Macaque
- Mouse
- Rat
- Squirrel
- Rabbit
- Pig
- Sheep
- Goat
- Cow
- Dolphin
- Alpaca
- Horse
- Panda
- Dog
- Cat
- Elephant
- opossum
- Pigeon
- Turkey
- Chicken
- Mallard Duck
- American alligator
- Zebrafish
- Atlantic cod
- Lamprey
- Mosquito

Data Source

- **General comparison –** Ensembl genome browser 104 (Basic Genome details)
- **Sequence comparison -** Table Browser (ucsc.edu) (Genome assembly data of the selected species)

Tools

- **Data processing –** Cinteny - Server for Synteny and Genome Rearrangement Analysis (cchmc.org)
- **Interactive Data Visualisation –** D3.js script library

7 Ideation

General Comparison

Design 1

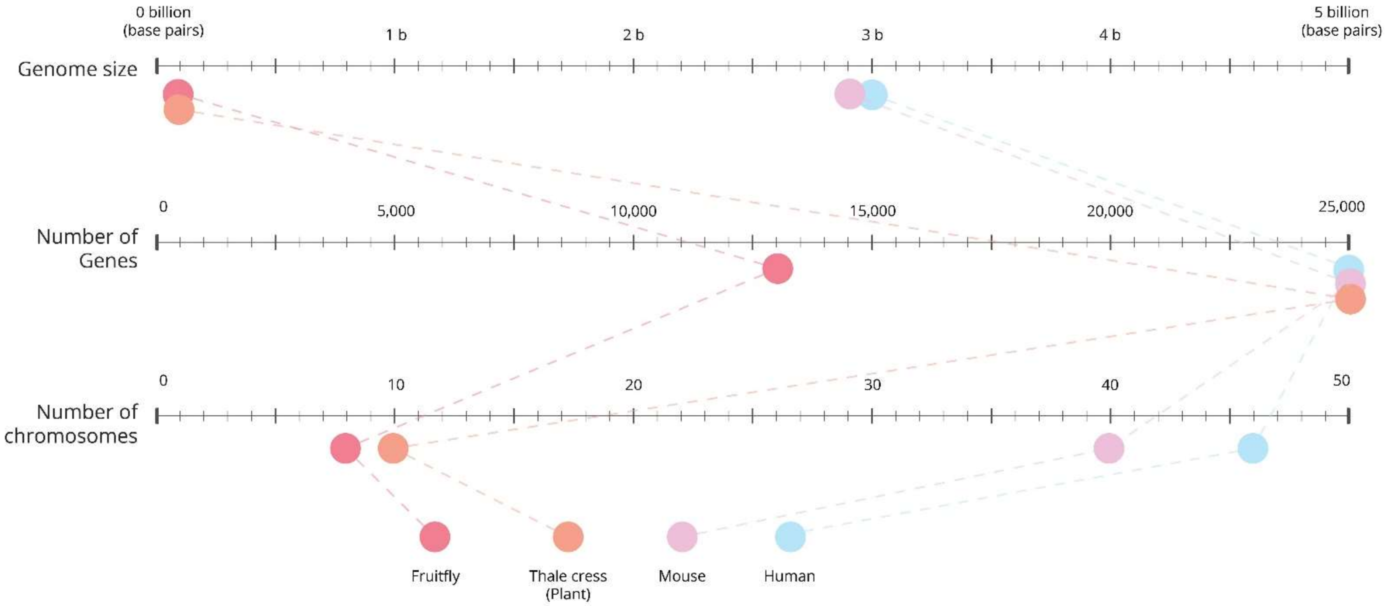


Fig 11: Idea exploration for visualisation

Above figure is an ideation for the general comparison where the gemome size and number of genes are compared on absolute scale. Each species is represented by a color coded circle on each of the scales based on the species genome size and number of genes. On Hover, one species is highlighted and one can see the finer details and datapoints.

Design 2

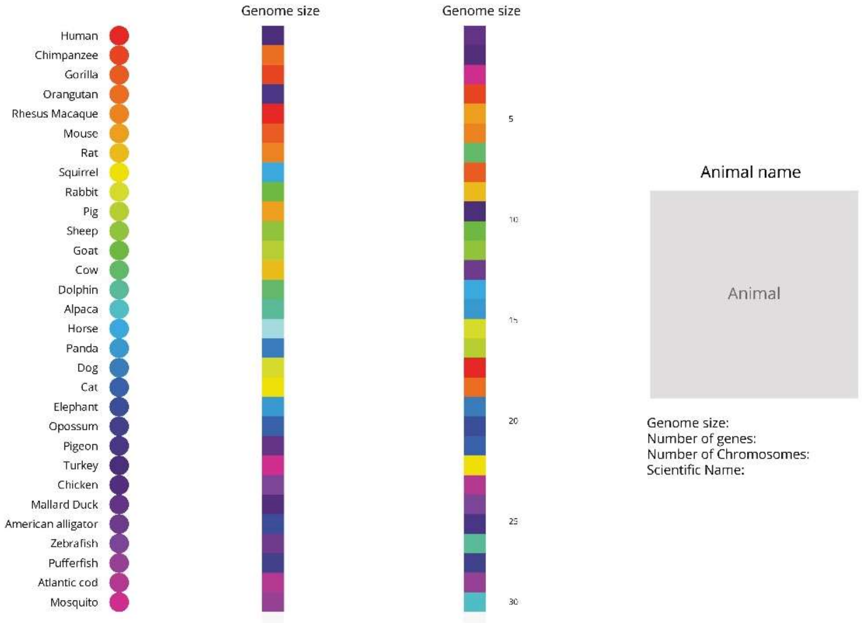


Fig 12: Idea exploration for visualisation

In this idea each species is represented by a colour and coded based on their evolutionary distance from our reference species “Human”. This means according to the coding; chimpanzee is the closest ancestor to humans while Mosquito is the farthest. The next column arranges the species based on their increasing rank of Genome size, i.e. Rank 1 on top will have the largest genome while Rank 30 at the bottom will have the smallest genome from the selected species. Similarly, the third column represents rankings based on number of genes. On Hover over a particular species, it’ll be highlighted and an overview will be given of the species on the right.

Design 3

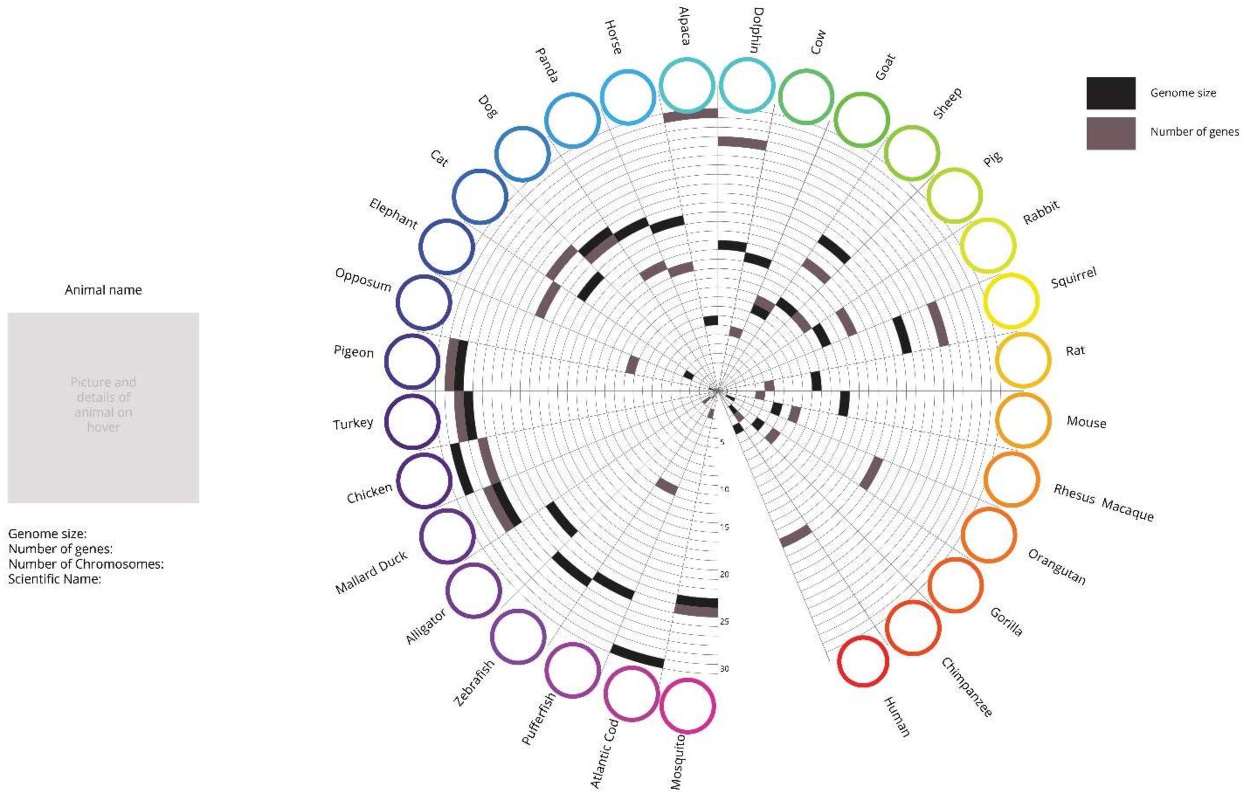


Fig 13: Idea exploration for visualisation

In this exploration, all the species are arranged around the circle in the order (anticlockwise) of their evolutionary distance from Humans, closest being Chimpanzee and farthest being Mosquito. The inner scale consists of 30 concentric circles divided into 30 arcs for each animal.

Rank of each species based on the first the genome size and second the number of genes is marked by two different colour fills in the circle sections. On Hover over a particular species, it'll be highlighted and an overview will be given of the species on the left.

Sequence Comparison

Design 1

This diagram is an attempt to simplify the circular visualisation/ circos diagram used for genomic circular visulisations. Inner segments represent the genomes of different species. Regions of similarity between the selected species are connected by chords at the positions where the sequences are coonserved. This let’s you compare the genetic makeup of one species in terms of another specie/s based on the sequence similarity. The outer segment of the reference species denotes the percentage similarity to each chosen species.

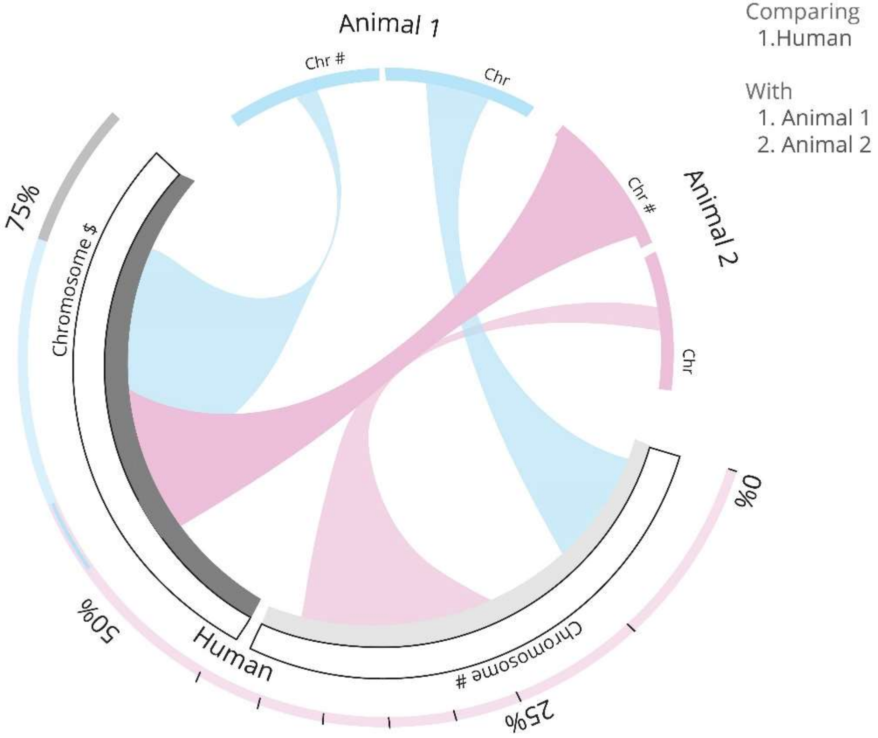


Fig 14: Idea exploration for visualisation

Design 2

Fig 15: Idea exploration for visualisation

In this exploration again, all the species are colour coded based on their evolutionary distance from Humans. The user can choose from the drop-down on the left, two species they want to compare in terms of the sequence similarity. The inner bars in the visualisation represent the genome of each species divided into different chromosomes. The outer solid bars represent the total length of similar segment found in the other species. On hover over any section of the genome, that chromosome and it’s connections with all the chromosomes of the other species is highlighted.

Design 3

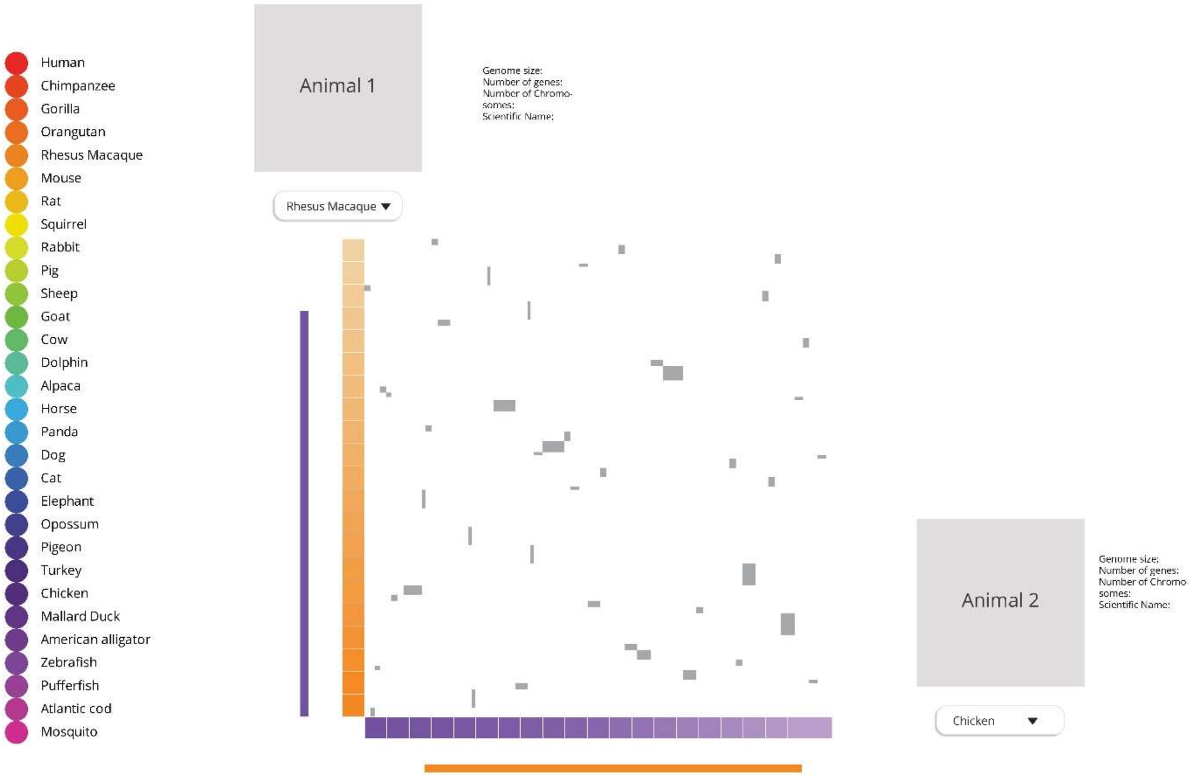


Fig 16: Idea exploration for visualisation

In this exploration again, all the species are colour coded based on their evolutionary distance from Humans. The user can choose from the two drop-downs on the axes, two species they want to compare in terms of the sequence similarity. The X and Y axes represent the genome of the two selected species. Each rectangle in the quadrant is a region of similarity, position based on the position of region on the genome and dimensions based on the region length on each genome. On hover over any section of the genome, that chromosome and it’s connections with all the chromosomes of the other species is highlighted.

8 Final Design

General Comparison

I tried to understand the insights that I want to convey with my visualisation and then choose the design which conveys it in the most simplified and easy to decode way. For the general comparison, the central idea is that Size of Genome or the Number of genes doesn’t correlate with each other or with the evolutionary status of the species. So, for a selected set of species, we can observe how the same species gets ranked on different parameters keeping in mind the evolutionary status of the species. So, I chose Design 2 as the Final design for this visualisation.

This visualisation will be bounded by the narrative we’ve curated for this visualisation. Before this visualisation, basics of genes, DNA and genome will be discussed to set the context.

Sequence Comparison

For this visualisation the goal is to highlight how even seemingly distant species have large regions of similarity between their genomes and how this relationship increases as we compare species closer to each other (in terms of evolutionary distance). So, with this tool we can observe the similarity between two chosen species and how it varies as we chose distant species. For this visualisation, I finalized Design 2 based on its suitability with the goals.

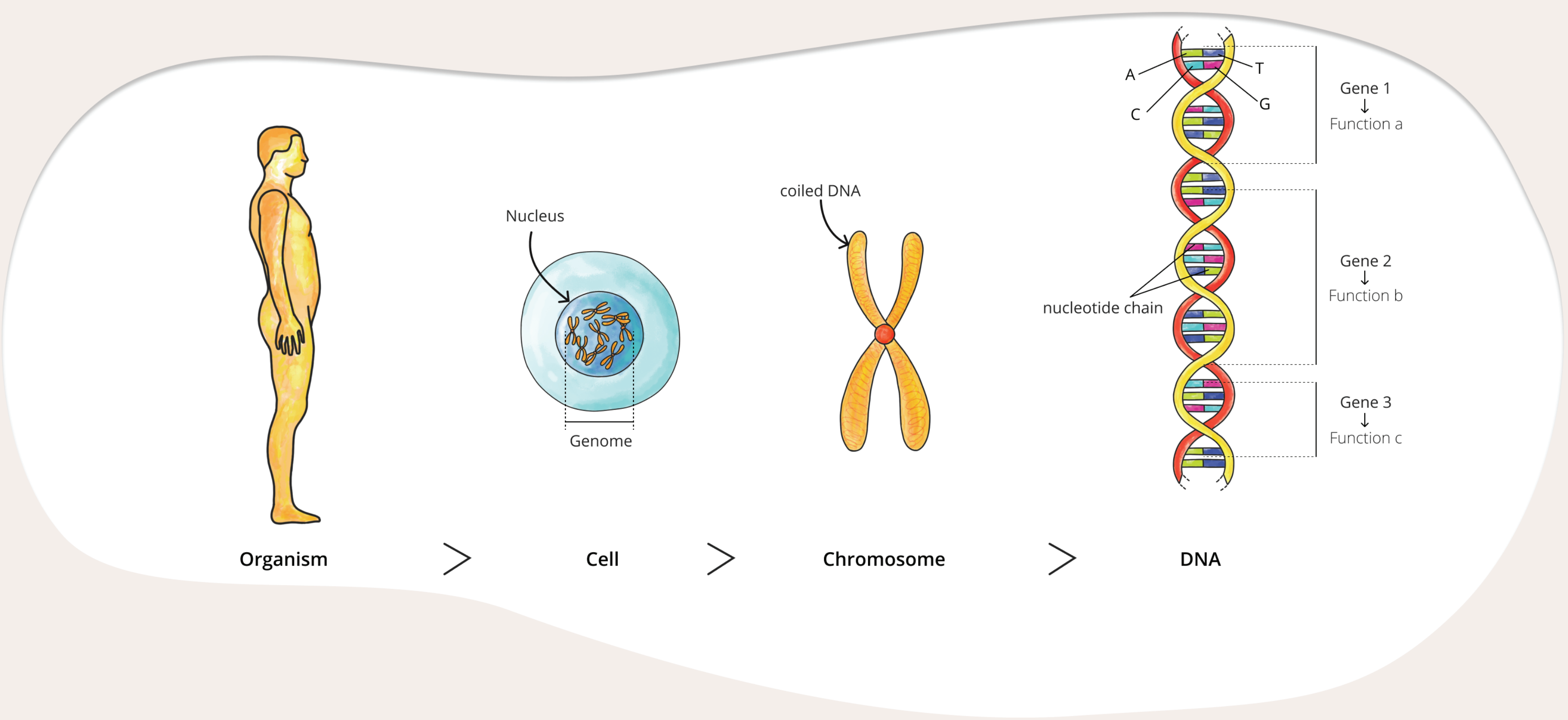
This is the second visualisation of this data story bounded by the narrative and closes with discussing the relevance of Comparative Genomics as a field of research currently and in the future to come.

Illustration style

Fig 17: Illustration style

Comparative Genomics

We are aware of the diversity of species around us ranging from mammals to insects to fishes to plants. Every living thing on this planet is made of cells which consists of DNA molecules. Within DNA, are written instructions that are sufficient to make an organism and the means by which organisms pass information to their offspring. Genes are the working subunits of DNA. Each gene contains a particular set of instructions, usually coding for a particular protein or for a particular function. On the other hand, Genome is termed as the complete set of the genetic information or instructions which is encoded in the DNA. In short, genome is a long, long molecule comprised of DNA chains and thus governs an organism's development, bodily functions and physical characteristics. Amazingly enough, this information is coded only by four nucleotides: Adenosine (A), Cytosine (C), Guanine (G) and Thymine (T). Understanding the order of these nucleotides in the DNA molecules is the fundamental approach of sequencing.

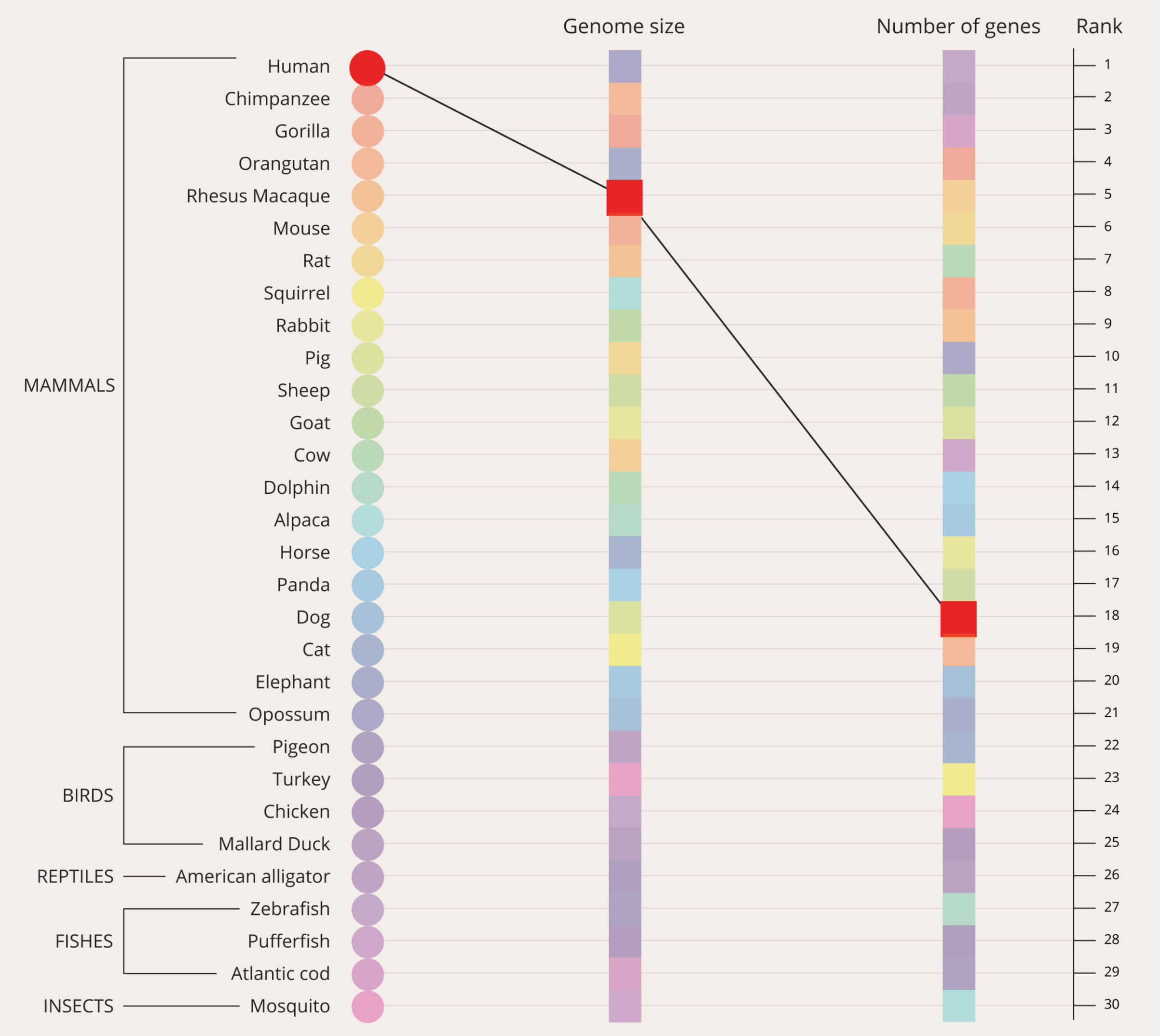


By looking at a crowd of people, one can realize, we all look very different. However, through sequencing recently, we've discovered that we humans are all actually 99.9% identical or even more than that. It's quite surprising how much diversity can be created from a very small number of changes in the genome. So, the human genome from generation to generation has barely changed at all and most of these changes are small, random changes. Comparative genomics is a field of research in which genomes of different species are compared on various levels with the help of which researchers can understand what, at the molecular level, distinguishes different life forms from each other.

We can compare our genomes to chimpanzees which are our closest relatives diverging from us approximately 6 million years ago. We can even compare our genomes to that of much more distant species like fruit flies, mosquitoes or even bacteria. When we do that, we find out that we share a surprising amount of sequence even with species as remote from us as bacteria. But thinking about it sort of makes sense as every living thing on the planet uses DNA as it's basic code and every living thing has at least a certain set of basic functions common between them. For example, every living thing on the planet has to copy its DNA in order to make copy of cells and bacteria uses a very similar mechanism to copy DNA as humans, so we find the genes in bacteria in those cases, similar to human genes.

How big is it?

This is usually the first question asked when studying an organism's genome. A simple comparison of the general features of genomes such as genome size, number of genes, and chromosome number presents an entry point into comparative genomic analysis. But, a very early lesson learned in the "genomic era" is that genome size does not correlate with evolutionary status, nor is the number of genes proportionate to genome size. The following visualisation let's you explore for yourself where different species stand in terms of evolution and the corresponding genome size and number of genes:

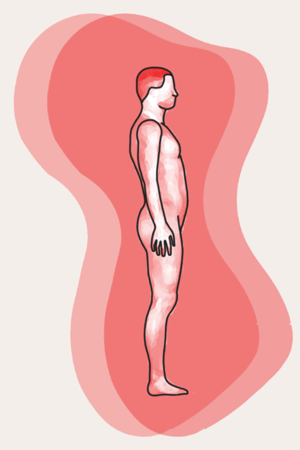


On the very left is the evolutionary tree for the selected species. Each species is coded by a unique color on a VIBGYOR gradient. Thus colors closer to each other on the gradient represents species with close evolutionary distance.

Middle column rearranges the species based on decreasing size of genome with rank 1 on top with the largest genome size. Last column rearranges the species based on decreasing number of genes with rank 1 on top with highest number of genes.

To compare the different variables for one species, hover over it's circle and it'll be highlighted and it's details will be shown below:

HUMAN



Genome Size: 3.09 gb

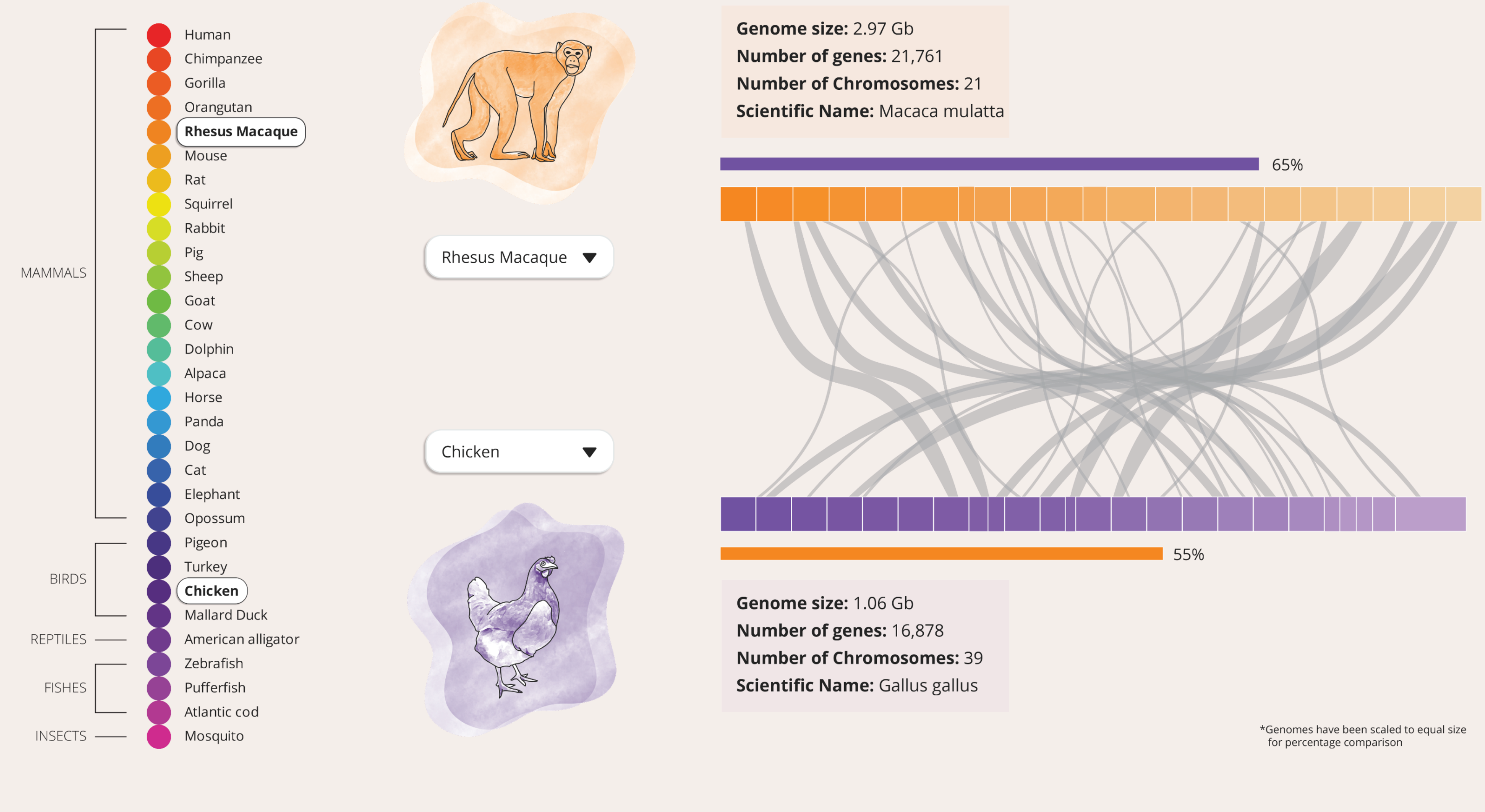
Number of genes: 20,442

Number of chromosomes: 23

Scientific name: Homo sapiens

How similar are we?

From sequencing, we have learnt that the similarity between the genomes of two species is largely dependent upon the evolutionary distance between them. The more distantly related two organisms are, the less sequence similarity or shared genomic features will be detected between them. The next visualisation is a tool to compare the genome similarities between any two selected species and their evolutionary relationship as compared to other species:



Genome size: 2.97 Gb
Number of genes: 21,761
Number of Chromosomes: 21
Scientific Name: Macaca mulatta

Genome size: 1.06 Gb
Number of genes: 16,878
Number of Chromosomes: 39
Scientific Name: Gallus gallus

*Genomes have been scaled to equal size for percentage comparison

There are many applications of Comparative Genomics today and the list is growing as we're getting better and better at sequencing. Comparing genomes helps us pinpoint genes that are essential to life and further helps to understand what genes relate to various biological systems, which in turn may translate into innovative approaches for treating human disease and improving human health.

Comparative genomics is increasingly being used in the areas ranging from human development and behaviour to metabolism and susceptibility to disease. Comparative genomics has also been significant in identification of drug targets of many infectious diseases. For example, comparative analyses of fungal genomes have led to the identification of many significant targets for new antifungal medicines. Comparative analysis of genomes of individuals with genetic disease against healthy individuals may also reveal ways of eliminating that disease.

As genome sequencing technology grows easier and less expensive, it's expected to find wide applications in agriculture, biotechnology, and zoology as a tool to tease apart the subtle differences among animal and plant species. Such efforts might point to new strategies for conserving rare and endangered species and various other applications.

Prototype link

Please visit the following link for the prototype –

https://homepages.iitb.ac.in/~18u130020/

9 Conclusion

Comparative Genomics is a very expansive topic and a large amount of time was spent in understanding the content and designing the narrative for the visualisation. This visualisation doesn’t aim for expertise in the field but just to imbibe the introductory idea of Comparative Genomics and how similar diverse species are when it comes to the genetic level. Efforts were made to incorporate the basic central idea around Comparative Genomics and shape into this narrative visualisation. I wanted the outcome to be a fully coded web interface but due to time constraints and limited expertise, I ended up using Figma for the most parts of it. I would like to work more on it later to make a fully functional coded outcome. This visualisation may be explored, played around with and encourage students and adults to look up things in detail for areas that interest them.

References

[1] Tufte, E. (1983). The Visual Display of Quantitative Information

[2] Crécy-Lagard, V., Hanson, A. (2013). Comparative Genomics in Brenner's Encyclopaedia of Genetics (Second Edition)

[3] Krzywinski, M. (July 2011). Designing effective visualizations in the biological sciences, PSA ANNUAL MEETING 2011 GENOMICS WORKSHOP mkweb.bcgsc.ca/psa/designing.effective.visualizations.biology.pdf

[4] Hardison, RC (November 2017). Comparative Genomics doi: 10.1371/journal.pbio.0000058

[5] Touchman, J. (2010) Comparative Genomics. Nature Education Knowledge 3(10):13

[6] Pray, L. (2008) Eukaryotic genome complexity. Nature Education 1(1):96

[7] Krzywinski, M. et al. Circos: an Information Aesthetic for Comparative Genomics. Genome Res (2009) 19:1639-1645 http://circos.ca/

[8] Sinha, UA., Meller, J. (2007). Cinteny: flexible analysis and visualization of synteny and genome rearrangements in multiple organisms. BMC Bioinformatics, 8:82 Cinteny - Server for Synteny and Genome Rearrangement Analysis (cchmc.org)

[9] Kent, WJ., Sugnet, CW., Furey, TS., Roskin, KM., Pringle, TH., Zahler, AM., Haussler, D. The human genome browser at UCSC. Genome Res. 2002 Jun;12(6):996-1006. https://cirm.ucsc.edu/index.html