

Comparative Analysis of Multiple Sequence Alignment Techniques for Proteins

A Study in Bioinformatics and Algorithmic Design

Team Members

01

Shruti Sivakumar

CB.SC.U4AIE23254

02

Vida Nadheera S

CB.SC.U4AIE23244

03

Shreya Sriram

CB.SC.U4AIE23248

04

Hamsadhuvani JS

CB.SC.U4AIE23229

01

Project Overview

Introduction to MSA

What is Multiple Sequence Alignment (MSA)?

- Alignment of three or more biological sequences (DNA, RNA, or protein) to identify regions of similarity

Importance

- Reveals evolutionary relationships
- Highlights functionally conserved regions
- Assists in predicting structure and function

Key Applications

- Phylogenetics - inferring evolutionary trees
- Structure Prediction - modeling 3D protein structures
- Functional Annotation - detecting motifs/domains

Problem Statement

Too Many MSA Tools

Biologists face difficulty selecting the best tool for their dataset and research goal

Trade-offs Between Accuracy, Speed and Biological Relevance

Classical tools are fast but may lack biological accuracy; newer methods are powerful but resource-intensive

Lack of Standardized Evaluation

Existing comparisons often rely on limited metrics, ignoring real-world biological significance

Challenge

Objectively compare MSA techniques across multiple dimensions – accuracy, efficiency, and biological value

Project Objectives

- Compare classical, and advanced MSA techniques
- Benchmark each method using datasets from standard biological databases (BAliBASE, OxBench, PREFAB)
- Evaluate tools using multiple performance metrics – alignment accuracy, runtime and memory usage, biological relevance
- Visualize and interpret results from various tools through charts and summary tables
- Provide recommendations based on use-case scenarios



02

Background & Research Context

Literature Review

Literature	Author	Description	Strengths	Weakness
Evaluating Statistical Multiple Sequence Alignment in Comparison to Other Alignment Methods on Protein Data Sets	Michael Nute	The paper compares BAli-Phy, a statistical MSA method, with other aligners on protein datasets. It performs well on simulated data but struggles on real biological benchmarks.	BAli-Phy demonstrates superior precision and recall on simulated datasets, indicating its effectiveness under controlled evolutionary models.	The method underaligns on biological data, leading to lower recall compared to other alignment tools.
Mathematical Understanding of Sequence Alignment and Phylogenetic Algorithms: A Comprehensive Review of Computation of Different Methods	Rashid Saif	A comprehensive review of sequence alignment algorithms and phylogenetic tree construction methods, emphasizing their mathematical foundations.	Offers clear explanations suitable for beginners in bioinformatics.	Lacks in-depth analysis of algorithmic performance and comparative evaluations.
Accelerating Multiple Sequence Alignments Using Parallel Computing	Qanita Bani Baker	Introduces three novel parallel computing approaches to enhance the performance of exact multiple sequence alignment algorithms.	Demonstrates significant speed improvements through multithreading techniques.	Focuses primarily on computational performance, with limited evaluation of alignment accuracy.

Literature Review

Literature	Author	Description	Strengths	Weakness
Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega	Fabian Sievers	Presents Clustal Omega, a tool designed for rapid and scalable multiple sequence alignments of protein sequences.	Efficiently handles large datasets while maintaining high alignment quality.	Relies on progressive alignment heuristics, which may propagate early alignment errors.
MUSCLE: a multiple sequence alignment method with reduced time and space complexity	Robert C. Edgar	Introduces MUSCLE, a tool for multiple sequence alignment that balances speed and accuracy using iterative refinement techniques.	Offers high accuracy and speed, outperforming many existing methods.	Performance may vary with highly divergent sequences.
MSA Transformer	Roshan Rao	Presents a transformer-based model that processes multiple sequence alignments to learn protein structure and function.	Outperforms prior models in unsupervised structure learning with fewer parameters.	Requires substantial computational resources for training.

Literature Review

Literature	Author	Description	Strengths	Weakness
MSAProbs: multiple sequence alignment based on pair hidden Markov models and partition function posterior probabilities	Sheng Wang	Combines pair hidden Markov models with partition function posterior probabilities for accurate protein sequence alignment.	Achieves high alignment accuracy across multiple benchmarks.	Computationally intensive, leading to longer runtimes for large datasets.
MAGUS: Multiple sequence Alignment using Graph clustering	Tandy Warnow	Introduces MAGUS, a method that uses graph clustering to improve scalability and accuracy in large-scale sequence alignments.	Effectively handles large datasets with improved alignment accuracy.	May be complex to implement due to its graph-based approach.
M-Coffee: combining multiple sequence alignment methods with T-Coffee	Cedric Notredame	M-Coffee integrates results from various alignment methods to produce a consensus alignment using T-Coffee's consistency framework.	Enhances alignment quality by leveraging multiple algorithms.	Dependent on the quality of input alignments from other methods.

Research Gaps

- Limited use of structural/functional data
- Poor performance on divergent sequences
- Over-reliance on SP/TC scores
- High resource demand in DL models
- Few head-to-head tool comparisons
- Lack of biologically relevant evaluations
- Benchmarking not dataset-diverse

03

Dataset & Evaluation

Benchmark Datasets Used



BALiBASE

BALiBASE 4

RV20: families aligned with a highly divergent "orphan" sequence (no significant similarity to known sequence)

RV30: subgroups with <25% residue identity between groups

RV50: internal insertions



OXBench Master

OXBENCH_1_3

Diverse protein families



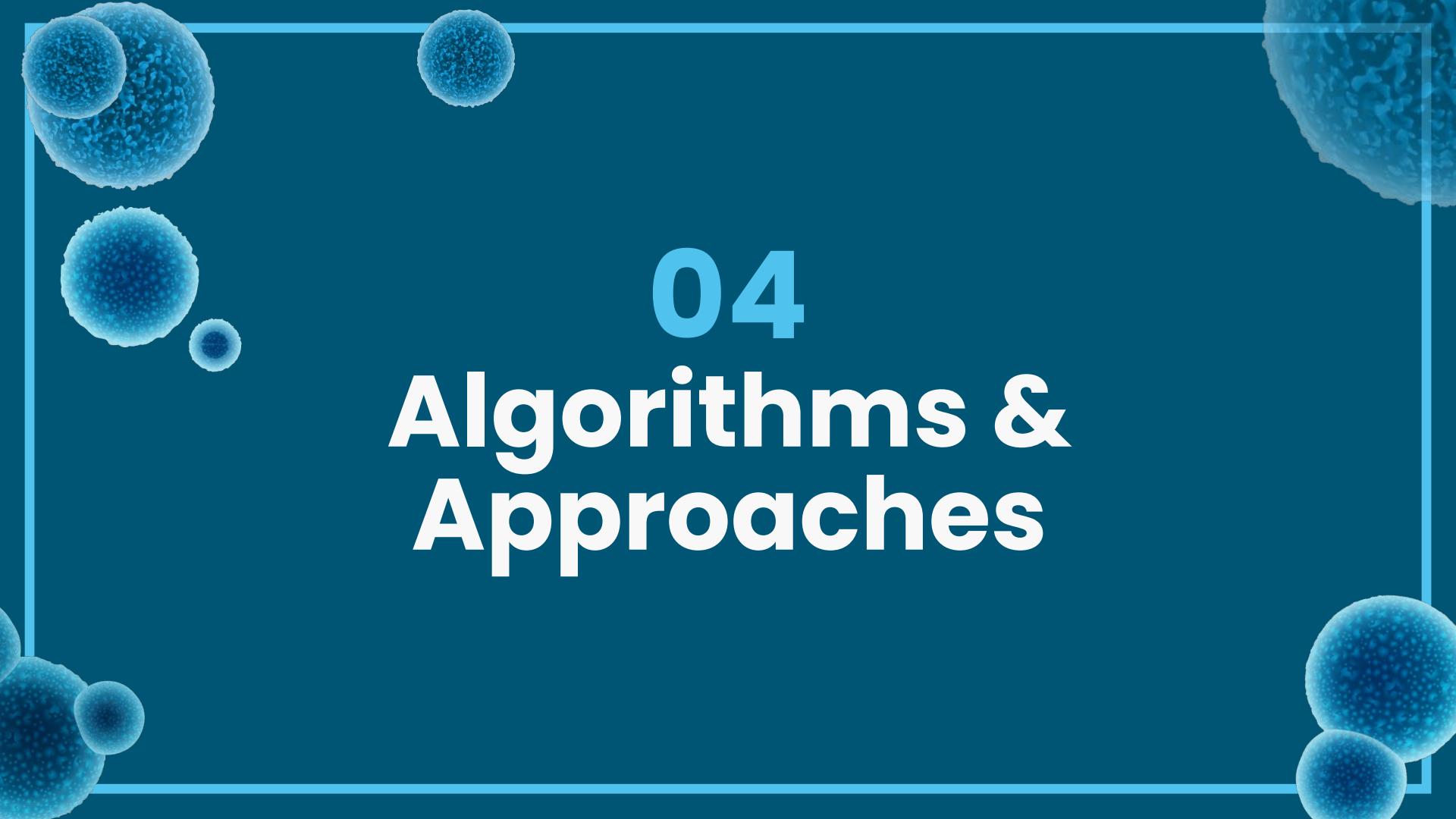
PREFAB

PREFAB 4.0

Structurally aligned sequence pairs, real and simulated alignments

Evaluation Criteria

Alignment Accuracy	SP Score (Sum-of-Pairs)	TC Score (Total Column)	Q Score (Quality Score)
	$\frac{\text{Number of correctly aligned residue pairs}}{\text{Total number of aligned residue pairs in reference}}$	$\frac{\text{Number of correctly aligned columns}}{\text{Total number of columns in reference alignment}}$	$\frac{\text{Correctly aligned pairs in both prediction and reference}}{\text{Total residue pairs in prediction}}$
Computational Efficiency	Runtime (seconds)	Memory Usage (MB)	Complexity Analysis
Biological Relevance	Conserved Region Identification	Percent Identity of Conserved Columns	
	Highly similar residues across all sequences	$\frac{\text{Number of identical residues in conserved columns}}{\text{Total residues in those columns}} \times 100$	



04

Algorithms & Approaches

MSA Methods Used

Clustal Omega	Baseline Method
MUltiple Sequence Comparison by Log-Expectation (MUSCLE)	Baseline Method
MSA Transformer	Advanced Method
MSAProbs	Advanced Method
Multiple sequence Alignment using Graph clUStering (MAGUS)	Advanced Method
Meta-Coffee (M-Coffee)	Advanced Method

Clustal Omega

- Fast, scalable progressive alignment tool
- No iterative refinement like MUSCLE

STEPS:-

Distance Calculation

For <1000 sequences: [O(N^2)]

Pairwise k-mer count ($k = 6$)

$$D(i, j) = 1 - (k\text{-mer matches}) / \min(|i|, |j|)$$

For >1000 sequences: [O(NlogN)]

mBed algorithm: embeds sequences in low-D Euclidean space & selects small set of reference sequences

Guide Tree Construction

Uses Unweighted Pair Group Method with Arithmetic Mean ([UPGMA](#)) method

- Starts with each seq. as own cluster
- Find closest pair of clusters and merge
- Recompute distances
 $d(A, BUC) = (\frac{1}{2}) * \{d(A,B) + d(A,C)\}$

Progressive Alignment

- Sequences aligned as per tree order
- HMM profile-profile alignment (via [HHalign](#))
→ aligns two profile HMMs by maximizing the log-likelihood of alignment between their match states

$$\log P(A \mid H_1, H_2) = \sum_{(i,j) \in A} \log \left(\sum_{a,b} c_1^i(a) \cdot c_2^j(b) \cdot \frac{P(a,b)}{P(a)P(b)} \right)$$

MUSCLE

- Fast and accurate progressive + iterative MSA method
- Works in three stages: Draft Tree → Improved Tree → Refinement

STEPS:-

Draft Progressive

- Compute k-mer distance matrix
- Build initial UPGMA guide tree
- Perform progressive alignment using tree

Improved Progressive

- Compute Kimura distances from initial alignment
 $d = -\ln(1 - p/0.85)$ where p = fraction mismatched

- Build Neighbour-Joining (NJ) Tree
 $Q(i,j) = (n-2) \cdot d(i,j) - \sum d(i,*) - \sum d(j,*)$
 - Find pair with minimum $Q(i,j)$
 - Join them into new node; update distances
 $d(X,k) = (\frac{1}{2}) * \{d(i,k) + d(j,k) - d(i,j)\}$
 - Repeat until only one node/tree remains
- Realign sequences/profiles using NJ tree

Iterative Refinement

- Loop: Split tree at an edge → realign sub-alignments → accept if score improves
- Uses log-expectation score (LE) for comparison

$$LE(A) = \sum_{i < j} \sum_{x,y} P_{xy}^{ij} \cdot \log(P_{xy}^{ij})$$

P_{xy}^{ij} = estimated posterior probability of aligning residue x in i with y in j

$$P_{xy}^{ij} = \frac{\text{Sum of probabilities of all paths aligning } x \text{ with } y}{\text{Total probability of all alignment paths}}$$

MSA Transformer

- DL-model using transformer architecture
- Pre-trained on large MSA datasets
- Slow & memory-heavy, but captures co-evolutionary and contextual signals

ARCHITECTURE:-

- Input: MSA matrix (unaligned sequences in rows, positions in columns)
- Uses Axial Attention:
 - Row-wise (between sequences)
 - Column-wise (across aligned positions)

$$\text{Attention}(Q, K, V) = \text{softmax} \left(\frac{QK^\top}{\sqrt{d_k}} \right) V$$

- 12 transformer layers with shared parameters

ALIGNMENT GENERATION:-

- Tokenize amino acid sequences → embed into numerical vectors
- Pass through axial transformer layers to get pairwise residue representations
- Compute alignment probabilities between positions using attention-weighted similarity

$$\text{softmax}(z_i) = \frac{e^{z_i}}{\sum_{j=1}^n e^{z_j}}$$

- Generate final alignment via soft-alignment method
 - Outputs alignment likelihood matrix by taking softmax of pairwise similarity score (i, j)
 - Best alignment inferred using maximum probability paths

MSA Probs

- Probabilistic MSA tool using pair-HMM + partition function
- Enhances accuracy via posterior probability matrices and consistency transformation
- Supports multithreading (OpenMP) and MPI parallelization

STEPS:-

- Pairwise Posterior Probability Calculation
→ Uses pair-HMM's forward-backward algorithm

$$P_{ij}(a, b) = \frac{F(a, b) \cdot B(a, b)}{Z}$$

- Distance Matrix Construction
From posterior scores:
→ $D(x,y)=1-\sum P_{xy}(i,j)$
- Guide Tree Construction Using UPGMA
- Probabilistic Consistency
Refine posterior probabilities by considering indirect alignment information
→ $P_{xy}'(i,j)=\sum z \sum k \{ P_{xz}(i,k) * P_{zy}(k,j) \}$
- Progressive Alignment
Align sequences progressively following the guide tree using refined probabilities
- Optional Iterative Refinement

MAGUS

- Graph-based method using DC strategy
- Accurate and scalable – excels on large or diverse datasets
- Built on PASTA + Graph Clustering

STEPS:-

Sequence Subgrouping

- Uses Practical Alignment using SATé and TRee Analysis (PASTA) to divide sequences into small, diverse subsets
 - Guide Tree Construction
 - Divide-and-Conquer
 - Subset Alignment
 - Tree Update
- Each subset is aligned independently (using MAFFT)

Graph Construction

- Each subset alignment → transformed into a partial alignment graph
- Nodes = residues, Edges = alignment relationships

Graph Merging

Merges graphs using Graph Clustering Merger (GCM)

- Aligns residues based on consistent paths in alignment graph
- Ensures residue relationships are transitive and globally consistent

$$\text{If } (i, j) \in E \text{ and } (j, k) \in E \Rightarrow (i, k) \in E$$

M-Coffee

- Ensemble method that leverages strengths of diverse aligners to improve overall accuracy
- Part of the T-Coffee framework

STEPS:-

Initial Alignments

- Run multiple MSA tools
- Each produces an independent alignment of the input sequences

Library Construction

- Extract pairwise residue alignments (alignment constraints) from each method
- Merge all into a consensus library of residue pair scores

Library Weighting

Each residue pair is assigned a score $w(i,j)$

→ Proportional to number of tools that align i with j

$$w(i,j) = \sum_{k=1}^N \delta_k(i,j)$$

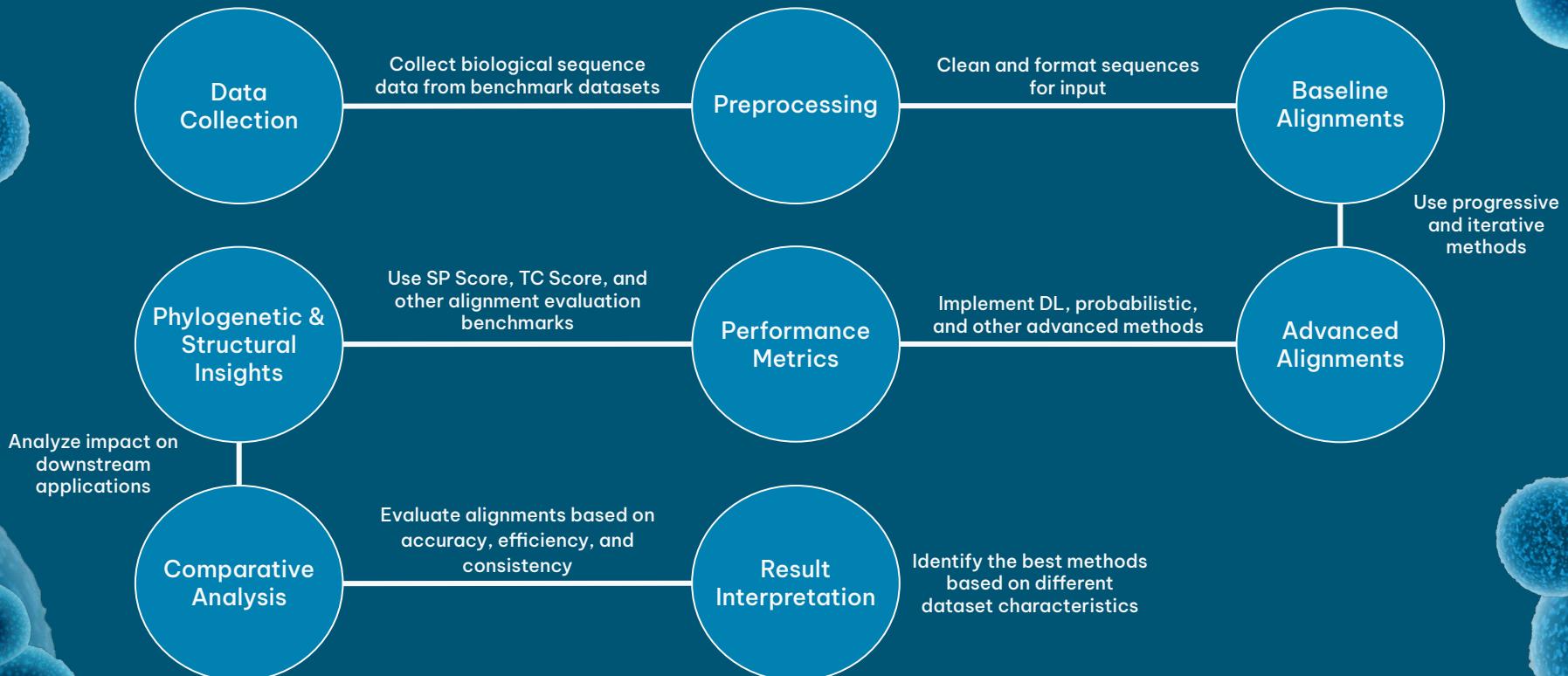
Final MSA via Consistency-Based Alignment

- T-Coffee uses the weighted library to build a progressive alignment
- Aligns sequences based on the most supported residue pairs

05

Implementation

Methodology



Data Collection & Preprocessing

Data Collection

- BAliBASE datasets (RV20, RV30, RV50) with .tsf (input) and .msf (reference) files
- OxBench inputs in .fa (FASTA) format
- PREFAB sequences without file extensions
- All datasets include both input and reference alignments
- Selected for diversity in sequence similarity and structural benchmarks

Preprocessing

- Parsed .tsf, .msf, .fa, and extensionless files into uniform FASTA format
- Renamed and cleaned headers for tool compatibility
- Validated input lengths and sequence counts
- Converted inputs to required formats for each tool
- Standardized naming for automated tool execution and output comparison

Method-wise Overview

Clustal Omega

- Used Clustal Omega via command-line calls in Colab
- Automated folder setup, input parsing, and output file organization
- Captured failed alignments using reference existence checks
- Evaluated alignments with SP, TC, Q scores and visualized results

MUSCLE

- Used MUSCLE v5 via command-line interface and Installed dependencies including Biopython
- Automated input parsing and FASTA sequence extraction using Python
- Invoked MUSCLE binary using subprocess with specified CPU thread control
- Evaluated alignment using SP,TC and Q Scores and visualized using Matplotlib

Method-wise Overview

MSA Transformer

- Implemented MSA Transformer using PyTorch; installed dependencies with Biopython
- Parsed input sequences and tokenized them
- Created embedding layers and positional encodings
- Executed self attention-based transformer to generate alignments

MSAProbs

- Used MSAProbs v0.9.7 binary via custom batch script
- Executed alignments using 4 CPU threads
- Automated input parsing, logging, and output storage
- Captured failed alignments for analysis
- Evaluated alignments and visualized score distributions

Method-wise Overview

MAGUS

- Used MAGUS v2.0 in Colab
- Executed alignments with 4 CPU threads
- Automated input/output parsing
- Logged execution time and success rates
- Visualized alignment accuracy
- Captured failed alignments for analysis

M-Coffee

- Ran M-Coffee v13.3 in Colab with multiple aligners
- Optimized protein sequence alignment
- Processed batch input files
- Logged alignment statistics
- Visualized consensus alignments
- Captured failed alignments for review

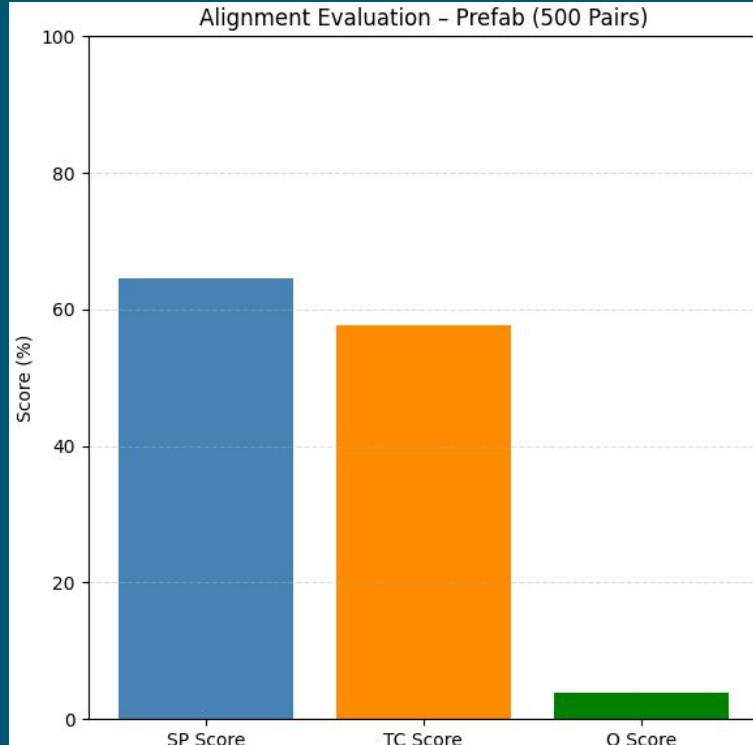
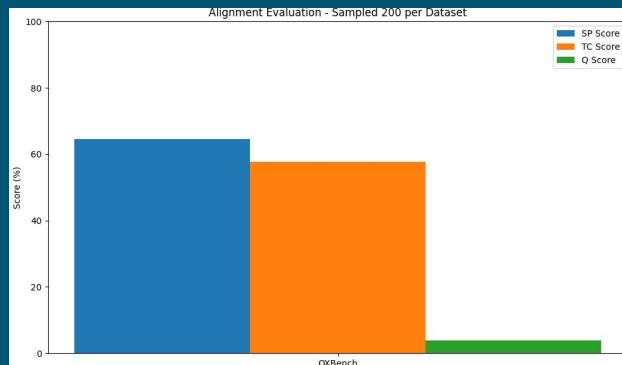
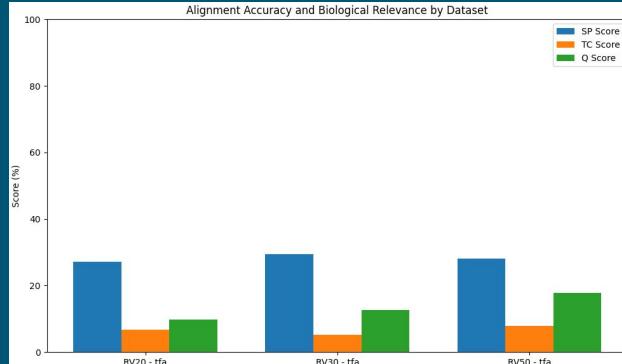
06

Results & Analysis

Results – Clustal Omega

Datasets	Alignment Accuracy			Computational Efficiency			Biological Relevance
	SP Score	TC Score	Q Score	Runtime (s)	Memory Usage (MB)	Worst-Case Complexity Analysis	Percent identity of conserved columns
BALiBASE RV20	0.2719	0.0666	0.0984	144	110.39MB	<p>Time: $O(N \log N + N^2L^2)$ (guide tree + progressive align)</p> <p>Space: $O(N \cdot L)$ (plus $O(L^2)$ for DP during profile alignment)</p>	98.53
BALiBASE RV30	0.2951	0.0519	0.1267	126	113.30MB		99.09
BALiBASE RV50	0.2810	0.0788	10.1789	90	114.84MB		99.59
OXBench Master	0.6260	0.5673	0.0370	1680	180.75MB		99.83
PREFAB 4.0	0.5565	0.0533	0.0259	1680	184.16MB		99.82

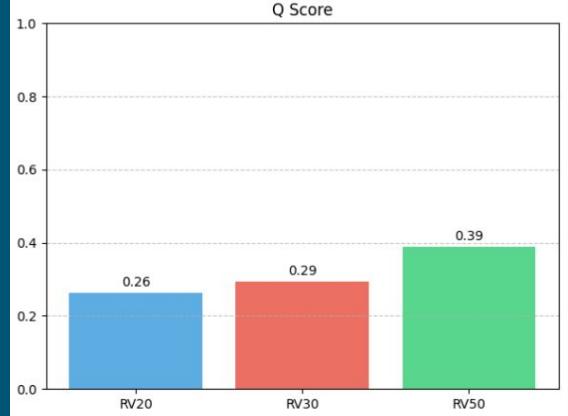
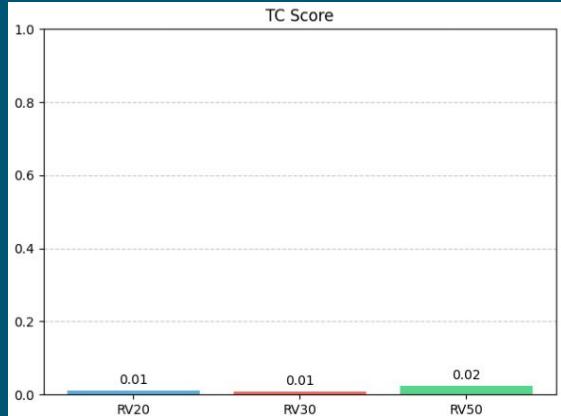
Results – Clustal Omega



Results – MUSCLE

Datasets	Alignment Accuracy			Computational Efficiency			Biological Relevance
	SP Score	TC Score	Q Score	Runtime (s)	Memory Usage (MB)	Worst-Case Complexity Analysis	Percent identity of conserved columns
BALiBASE RV20	0.72	0.01	0.26	310.09	427.36	<p>Time: $O(N^3 \cdot L)$ (with refinement; base algorithm $O(N^2L + N \cdot L^2)$)</p> <p>Space: $O(N^2 + N \cdot L + L^2)$</p>	0.21
BALiBASE RV30	0.67	0.01	0.29	462.89	530.63		0.20
BALiBASE RV50	0.59	0.02	0.39	110.83	496.82		0.18
OXBench Master	-	-	-	-	-		-
PREFAB 4.0	-	-	-	-	-		-

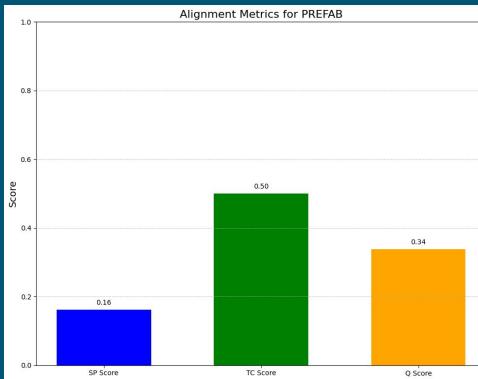
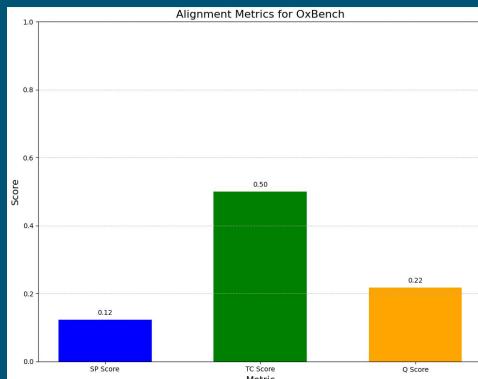
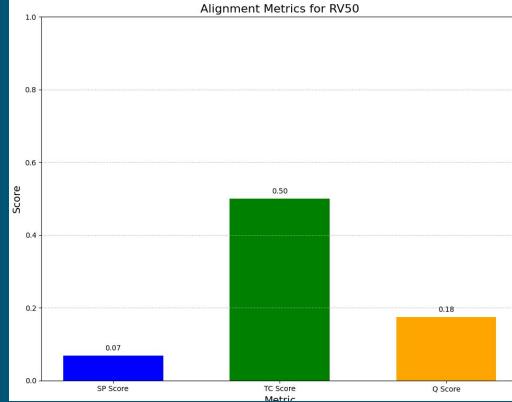
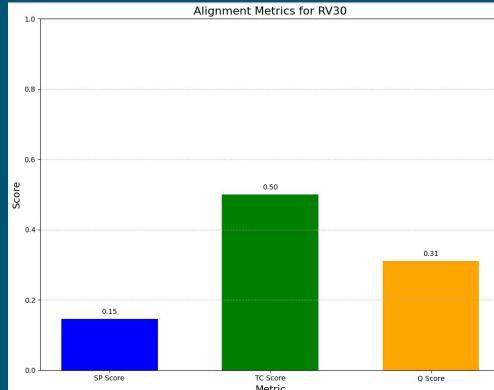
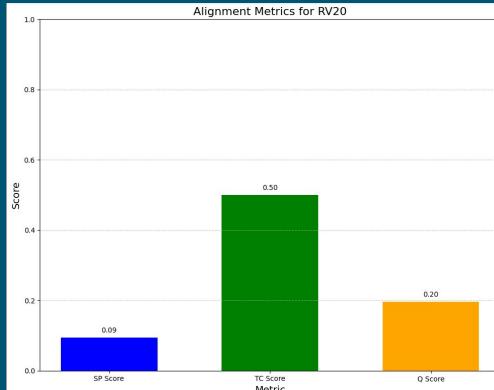
Results – MUSCLE



Results – MSA Transformer

Datasets	Alignment Accuracy			Computational Efficiency			Biological Relevance
	SP Score	TC Score	Q Score	Runtime (s)	Memory Usage (MB)	Complexity Analysis	Percent identity of conserved columns
BALiBASE RV20	0.09	0.50	0.19	3.84	70.44	Time: $O(N^2 \cdot L + N \cdot L^2)$ Space: $O(N^2 \cdot L + N \cdot L^2)$	9.39
BALiBASE RV30	0.14	0.50	0.31	3.25	18.33		14.65
BALiBASE RV50	0.06	0.50	0.17	15.87	201.05		6.84
OXBench Master	0.12	0.50	0.21	9.38	9.33		12.86
PREFAB 4.0	0.16	0.50	0.33	1.75	0.00		16.18

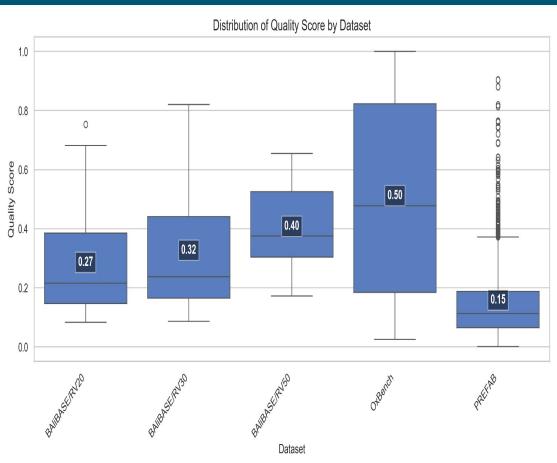
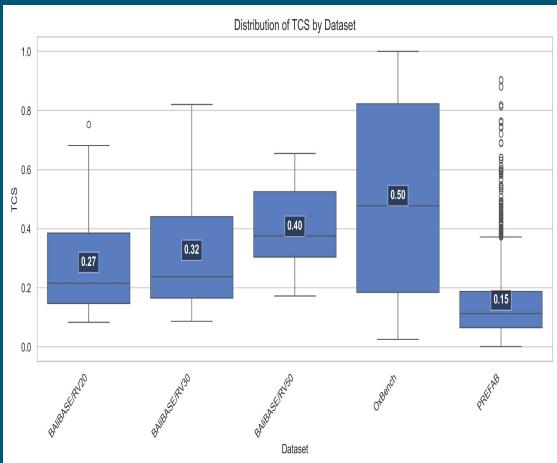
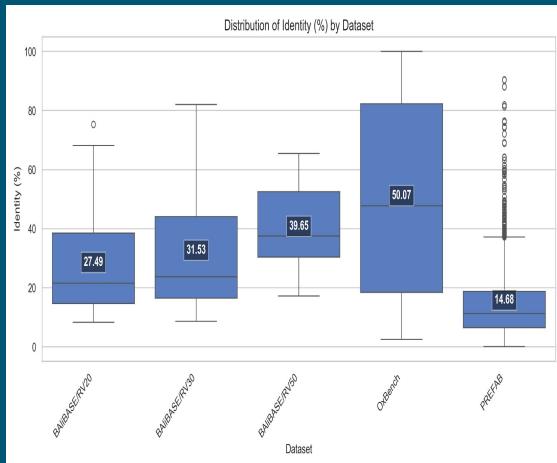
Results – MSA Transformer



Results – MSAProbs

Datasets	Alignment Accuracy			Computational Efficiency			Biological Relevance
	SP Score	TC Score	Q Score	Runtime (s)	Memory Usage (MB)	Worst-Case Complexity Analysis	Percent identity of conserved columns
BALiBASE RV20	0.9999	0.2749	0.2749	316	Time: $O(N^2 \cdot L^3)$ Space: $O(N^2 \cdot L^2)$	-	27.939
BALiBASE RV30	0.9988	0.3153	0.3153	559			31.5311
BALiBASE RV50	0.9993	0.3965	0.3965	68			39.6461
OXBench Master	1.0	0.5007	0.5007	43			50.0735
PREFAB 4.0	0.9981	0.1468	0.1468	3748			14.6785

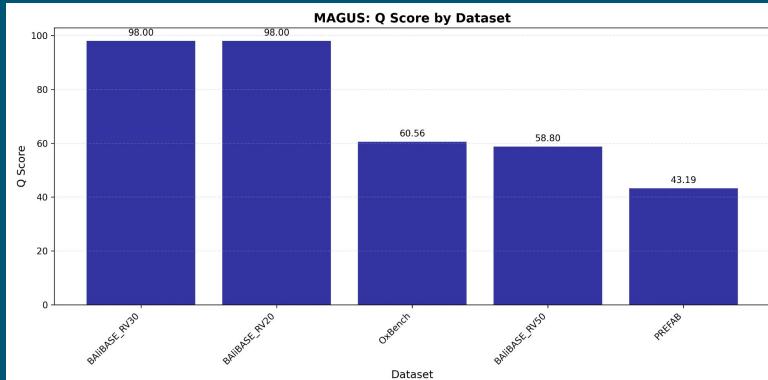
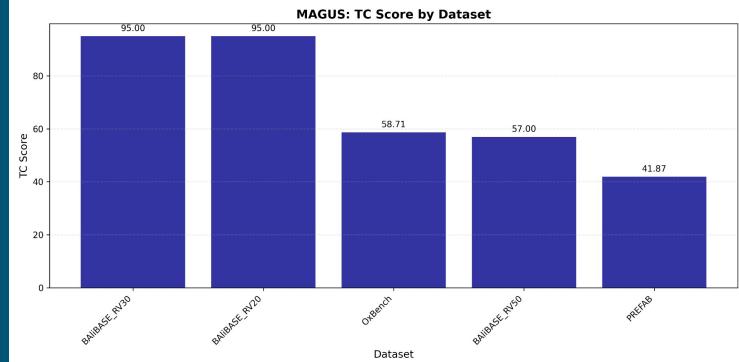
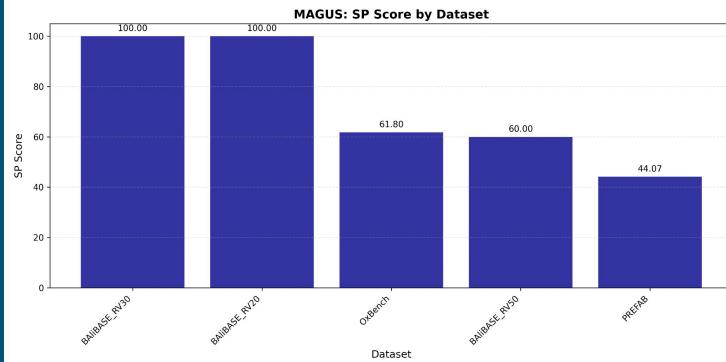
Results – MSAProbs



Results – MAGUS

Datasets	Alignment Accuracy			Computational Efficiency			Biological Relevance
	SP Score	TC Score	Q Score	Runtime (s)	Memory Usage (MB)	Worst-Case Complexity Analysis	Percent identity of conserved columns
BALiBASE RV20	1	0.95	0.98	1.2	297.542	Time: $O(N^3 \cdot L)$ (in worst case without effective subdivision) Space: $O(N \cdot L)$ (per largest subset/alignment)	16.7
BALiBASE RV30	1	0.95	0.98	1.2	296.81		14.4
BALiBASE RV50	0.60	0.57	0.58	1.2	297.15		12.6
OXBench Master	0.61	0.58	0.60	1.2	297.52		13.8
PREFAB 4.0	0.44	0.41	0.43	1.2	296.81		29.6

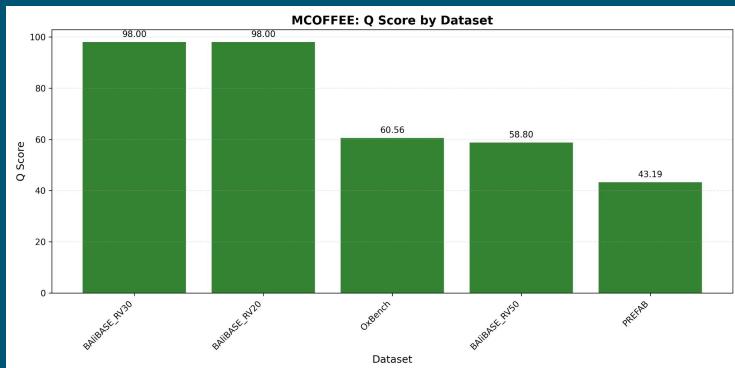
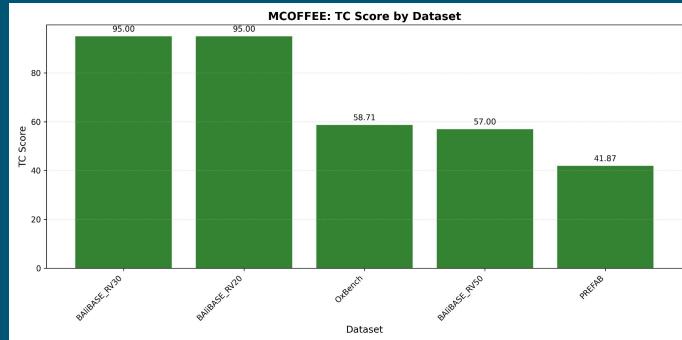
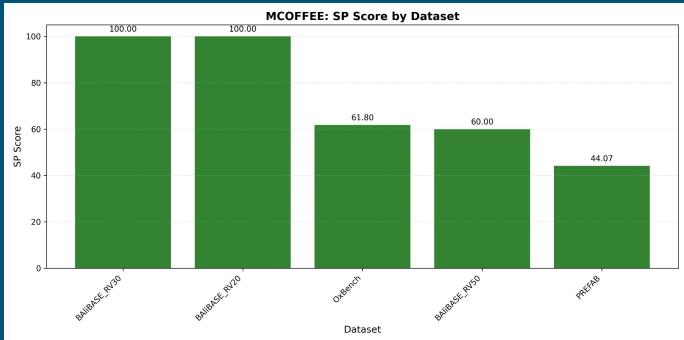
Results – MAGUS



Results – M-Coffee

Datasets	Alignment Accuracy			Computational Efficiency			Biological Relevance
	SP Score	TC Score	Q Score	Runtime (s)	Memory Usage (MB)	Worst-Case Complexity Analysis	Percent identity of conserved columns
BALiBASE RV20	1	0.95	0.98	1.5	300.096	Time: $O(N^3 \cdot L)$ Space: $O(N^2 \cdot L)$	16.7
BALiBASE RV30	1	0.95	0.98	1.5	297.52		14.4
BALiBASE RV50	0.60	0.57	0.58	1.5	299.144		12.6
OXBench Master	0.61	0.58	0.60	1.5	300.12		13.8
PREFAB 4.0	0.44	0.41	0.43	1.5	297.52		29.6

Results – M-Coffee



Result Analysis

BALiBASE RV20/RV30

- MAGUS outperformed others in alignment accuracy – aligning even divergent sequences well.
- Reported fastest runtime, but values may be GPU-influenced or incorrectly logged.
- Clustal Omega showed highest % identity but lower TC accuracy, suggesting strong motif focus but poor overall alignment.

OxBench & PREFAB

- MSAProbs and MAGUS performed similarly with solid TC/Q.
- MSA Transformer achieved highest TC (0.5) on PREFAB but required ~1 hour (3748s), consistent with DL expectations.
- Meta-Coffee's runtime of 1.5s is highly unrealistic given its ensemble nature.

BALiBASE RV50

- MSAProbs had best SP (0.999), while MAGUS topped TC (0.57).
- Clustal's Q score (10.17) is likely incorrect – should be <1.
- MAGUS again reported 1.2s runtime, likely an underreported value.

Efficient Comparison

- Fastest: MAGUS (on paper), but times suspect
- Most Accurate Overall: MSAProbs (balanced and reliable)
- Most Pairwise Accurate: MSA Transformer ($SP \approx 0.99-1$)
- Motif-Conserving: Clustal Omega (high % identity)

Comparative Summary Table

Tool	Best On (Datasets)	Comments
MAGUS	RV20, RV30, OxBench	Best accuracy overall; timing likely underreported
MSAProbs	RV50, OxBench, PREFAB	High accuracy but slow, esp. PREFAB
MSA Transformer	PREFAB TC, SP in all sets	Very accurate, slow, memory-heavy
Clustal Omega	High % identity on all	High motif conservation; low TC
MUSCLE	-	Fastest classical tool
Meta-Coffee	RV20/RV30 SP only	High SP/TC in simple sets; low % identity

Observations & Insights

- MAGUS showed top alignment accuracy on most datasets, but the ~1.2s runtime is unrealistic.
- MSAProbs matched MAGUS in accuracy, but was very slow, especially on PREFAB (~1 hour).
- MSA Transformer gave near-perfect SP scores, but moderate TC and high runtime.
- Clustal Omega had the highest % identity (~99%) but low TC/Q and poor scalability.
- Meta-Coffee did well on easy sets but had low biological relevance (% identity <30%).
- MUSCLE was fast and light, but low accuracy limits its reliability.

07

Reflection & Conclusion

Challenges Faced

Unrealistic Runtime Logs

GPU-enabled runs showed suspiciously low times (e.g., MAGUS 1.2s, M-Coffee 1.5s) that contradict known algorithm cost.

Hardware Dependency

MSA Transformer requires powerful GPUs and time (~1 hour for PREFAB), limiting its usability in normal setups.

Q Score Outlier

Clustal Omega's $Q=10.17$ is invalid ($\text{typical } Q \in [0, 1]$); shows risks of data miscalculation.

Metric Trade-Offs

High SP doesn't guarantee high TC. Tools vary in optimizing residue pair alignment vs. column alignment.

Future Scope

- Refine Deep Learning Models – Improve Transformer efficiency and make soft-alignment more column-aware
- Integrate Structural Data – Use 3D or evolutionary context to improve alignment in low-identity regions
- Smart Hybrid Methods – Merge fast classical and accurate DL models for balanced performance
- Biology-Aware Metrics – Use functional/structural recovery as scoring metrics, not just SP/TC/Q
- Scalable Cloud-Based MSA – Build GPU-ready, web-accessible pipelines for MSA on big data

Conclusion

MAGUS gave best all-round accuracy, but timing needs verification.

Clustal Omega is great for motif alignment, but weak in full alignment.

MSAProbs is accurate but too slow for large datasets.

Meta-Coffee underperformed in motif preservation despite ensemble design.

MSA Transformer excels in pairwise matches but is resource-heavy.

Choose tool based on task – no single method fits all use-cases.

References

1. Nute M, Saleh E, Warnow T. Evaluating Statistical Multiple Sequence Alignment in Comparison to Other Alignment Methods on Protein Data Sets. *Syst Biol.* 2019 May 1;68(3):396–411. doi: 10.1093/sysbio/syy068. PMID: 30329135; PMCID: PMC6472439.
2. R. Saif, "Mathematical Understanding of Sequence Alignment and Phylogenetic Algorithms: A Comprehensive Review of Computation of Different Methods," *Advancements in Life Sciences – International Quarterly Journal of Biological Sciences*
3. Bani Baker, Q.; Al-Hussien, R.A.; Al-Ayyoub, M. Accelerating Multiple Sequence Alignments Using Parallel Computing. *Computation* 2024, 12, 32. <https://doi.org/10.3390/computation12020032>
4. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, Thompson JD, Higgins DG. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol.* 2011 Oct 11;7:539. doi: 10.1038/msb.2011.75. PMID: 21988835; PMCID: PMC3261699.
5. Edgar, R.C. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5, 113 (2004). <https://doi.org/10.1186/1471-2105-5-113>
6. Roshan Rao, Jason Liu, Robert Verkuil, Joshua Meier, John F. Cann, Pieter Abbeel, Tom Sercu, Alexander Rives. MSA Transformer. *bioRxiv* 2021.02.12.430858; doi: <https://doi.org/10.1101/2021.02.12.430858>
7. Yongchao Liu, Bertil Schmidt, Douglas L. Maskell, MSAProbs: multiple sequence alignment based on pair hidden Markov models and partition function posterior probabilities, *Bioinformatics*, Volume 26, Issue 16, August 2010, Pages 1958–1964, <https://doi.org/10.1093/bioinformatics/btq338>
8. Vladimir Smirnov, Tandy Warnow, MAGUS: Multiple sequence Alignment using Graph clustering, *Bioinformatics*, Volume 37, Issue 12, June 2021, Pages 1666–1672, <https://doi.org/10.1093/bioinformatics/btaa992>
9. Iain M. Wallace, Orla O'Sullivan, Desmond G. Higgins, Cedric Notredame, M-Coffee: combining multiple sequence alignment methods with T-Coffee, *Nucleic Acids Research*, Volume 34, Issue 6, 1 March 2006, Pages 1692–1699, <https://doi.org/10.1093/nar/gkl091>

Thanks!

CREDITS: This presentation template was created by [Slidesgo](#), and includes icons by [Flaticon](#), and infographics & images by [Freepik](#)

Table of contents

01

Introduction

You can describe the topic of the section here

02

Theoretical

You can describe the topic of the section here

03

Exercise

You can describe the topic of the section here

04

Conclusion

You can describe the topic of the section here

Whoa!

This can be the part of the presentation where
you introduce yourself, write your email...

01

Introduction

You can enter a subtitle here if you need it

What is cell biology?

Mercury is the closest planet to the Sun and the smallest one in the entire Solar System. This planet's name has nothing to do with the liquid metal, since Mercury was named after the Roman messenger god. Its temperatures aren't as terribly hot as that planet's.

Mercury takes a little more than 58 days to complete its rotation, so try to imagine how long days must be there! Since the temperatures are so extreme, albeit not as extreme as on Venus, Mercury has been deemed to be non-habitable for humans

Introduction to cell histology

Do you know what helps you make your point crystal clear? Lists like this one:

- They're simple
- You can organize your ideas clearly
- You'll never forget to buy milk!

And the most important thing: the audience won't miss the point of your presentation

Cell function

Reproduction

Mercury is the closest planet to the Sun and the smallest one in the Solar System—it's only a bit larger than the Moon

Metabolism

Venus has a beautiful name and is the second planet from the Sun. It's hot and has a poisonous atmosphere

Cell cycle process

Grow up

Mercury is the closest planet to the Sun and a very small object

Reproduce

Venus has a beautiful name and is the second planet from the Sun

Divide

Despite being red, Mars is actually a cold place. It's full of iron oxide

Parts of the cell

Write in the indicated area the **part of the cell** indicated

Write here

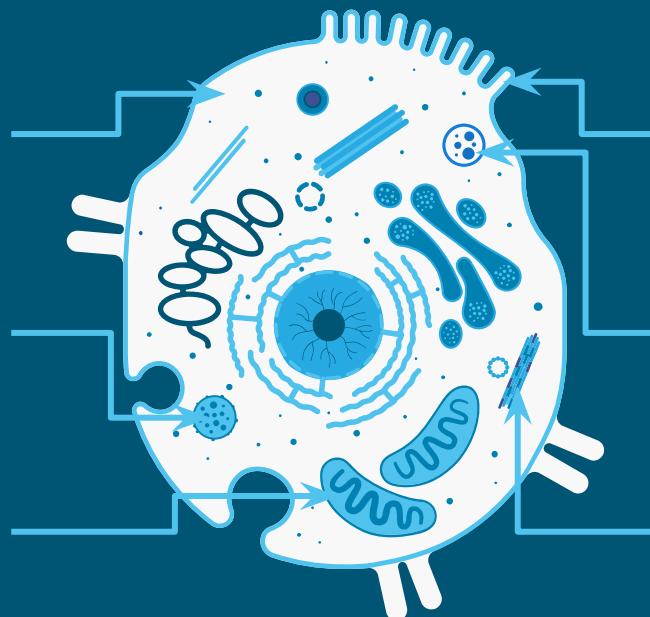
Write here

Write here

Write here

Write here

Write here



Function of cell biology

Cell structure

Mars is actually a very cold planet

Genetics

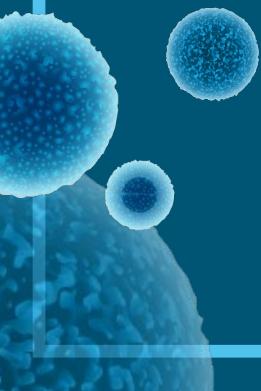
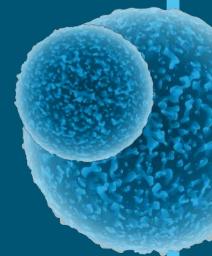
Jupiter is the biggest planet of them all

Investigation

Venus has extremely high temperatures

Inheritance

Saturn is a gas giant and has several rings



Parts of a cell

Nucleus

Mars is actually a very cold planet

Mitochondria

Venus has extremely high temperatures

Centrioles

Neptune is the farthest planet from the Sun

Cytoplasm

Mercury is the closest planet to the Sun

Plasma

Saturn is a gas giant with several rings

Nucleolus

Jupiter is the biggest planet of them all

Human cells

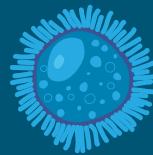
Write in the indicated area what type of human cell it is



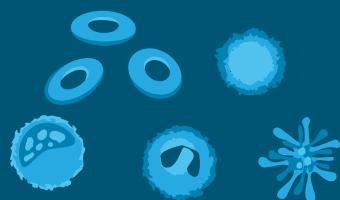
Write here



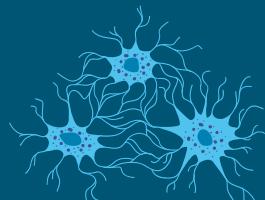
Write here



Write here



Write here

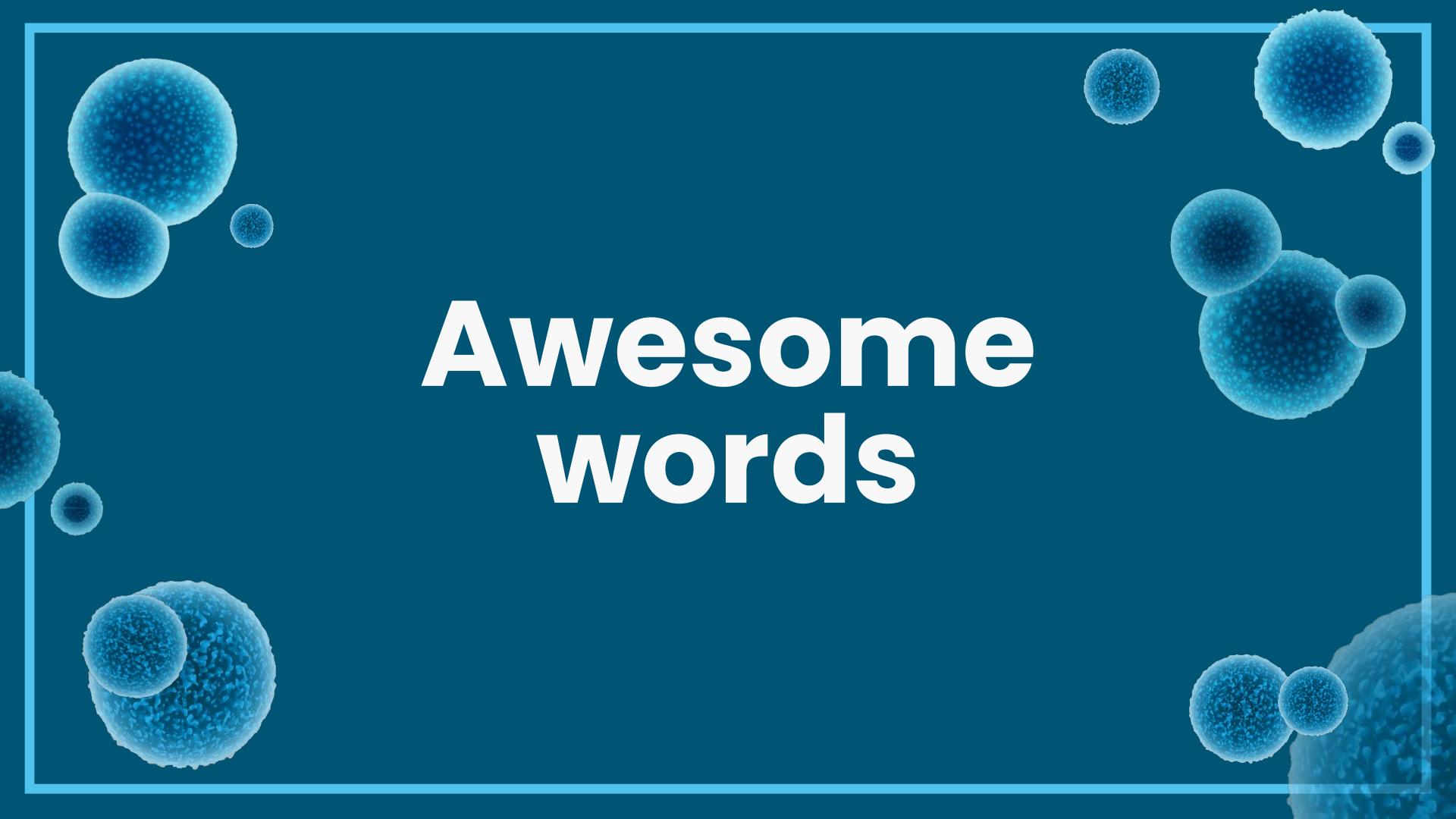


Write here



Write here

Awesome words



Cell types and differences

Types

Vegetable

Earth is the third planet from the Sun

Animal

Jupiter is the biggest planet of them all

Moon

The Moon is Earth's only natural satellite

Neptune

Neptune is the farthest planet from the Sun

Saturn

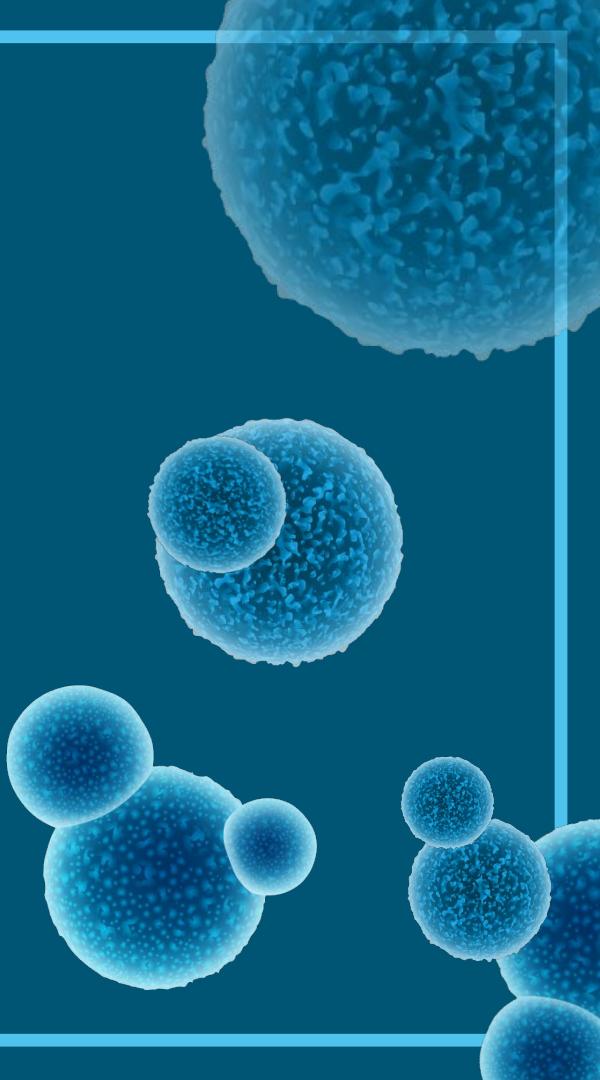
It's composed of hydrogen and helium

Mars

Despite being red, Mars is a cold planet

**“This is a quote, words
full of wisdom that
someone important said
and can make the
reader get inspired”**

—Someone Famous



02

Theoretical

You can enter a subtitle here if you need it



A picture is worth a thousand words

A picture always reinforces the concept

Images reveal large amounts of data, so remember: use an image instead of a long text. Your audience will appreciate it

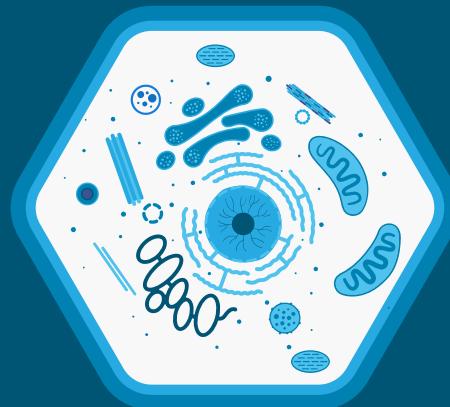


98,300,000

Big numbers catch your audience's attention

Parts of the plant cell

Click on the line and drag to point to the part of the cell it refers to



Nucleus

Vacuoles

Mitochondria

Chloroplast

9h 55m 23s

Jupiter's rotation period

333,000

The Sun's mass compared to Earth's

386,000 km

Distance between Earth and the Moon

Week schedule

Hour	Mon.	Tue.	Wed.	Thu.	Fri.
9:00				Exam	
10:00	Activities	Activities	Activities	Activities	Activities
11:00		Project			
12:00	Task	Task	Task	Task	Task
13:00	Activities			Project	Activities
14:00		Exam	Project		Project

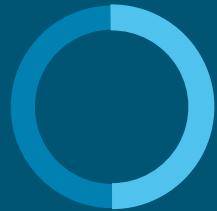
Evolution of the students



25%

1 year

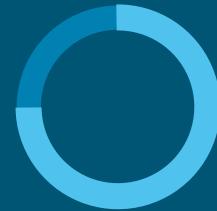
Mercury is the closest planet to the Sun and a very small object



50%

2 year

Venus has a beautiful name and is the second planet from the Sun



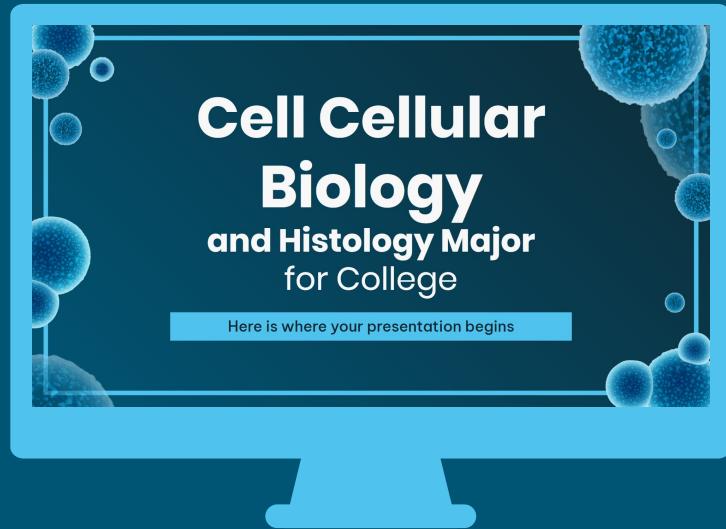
75%

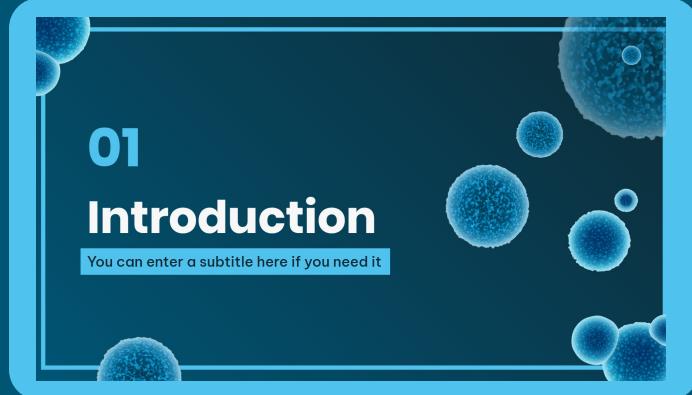
3 year

Despite being red, Mars is actually a cold place. It's full of iron oxide dust

Computer mockup

You can **replace the image** on the screen
with your own work. Just right-click on it
and select “Replace image”



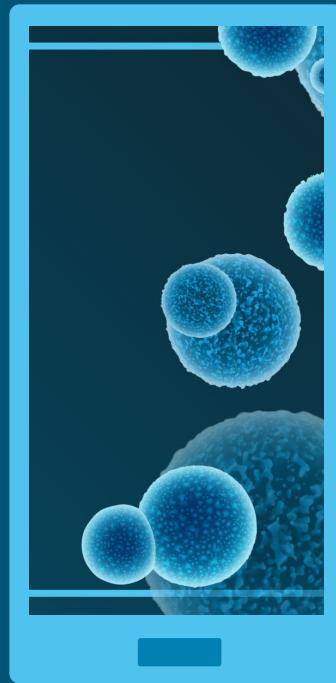


Tablet mockup

You can **replace the image** on the screen
with your own work. Just right-click on it
and select “Replace image”

Phone mockup

You can **replace the image** on the screen with your own work. Just right-click on it and select “Replace image”



Major studies on cell biology



Venus

Venus is the second planet from the Sun

Mercury

Mercury is the closest planet to the Sun

Mars

Despite being red, Mars is a very cold planet

Cell evolution

Venus is the second planet from the Sun

Step 1



Despite being red, Mars is a very cold planet

Step 3



Step 2

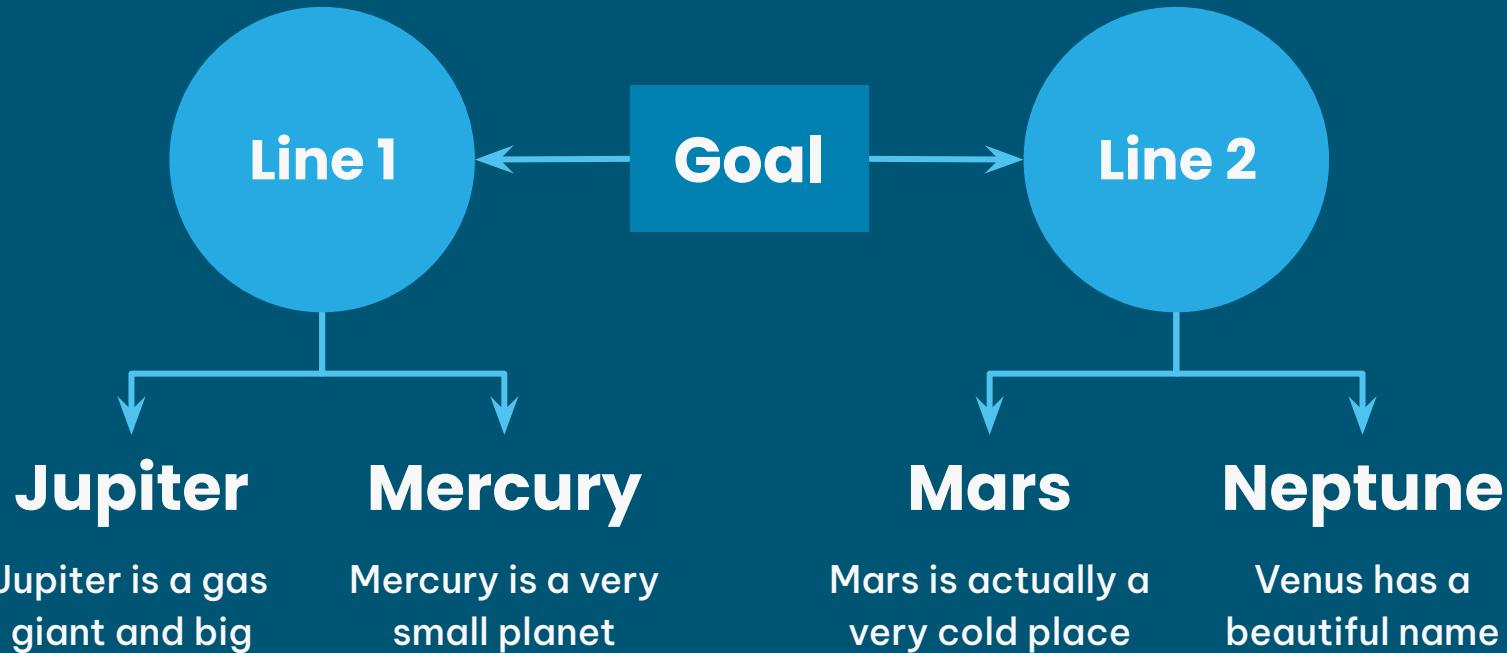
Mercury is the closest planet to the Sun

Step 4

Jupiter is the biggest planet of them all



Investigation process



Cell comparison

	Prokaryotic	Animal	Plant
Cytoplasm	Yes	Yes	Yes
Ribosome	Yes	Yes	Yes
Nucleoid	Yes	No	No
Nucleus	No	Yes	Yes
Golgi	No	Yes	Yes
Lysosome	No	Yes	No

Landmarks in cell histology

● Venus

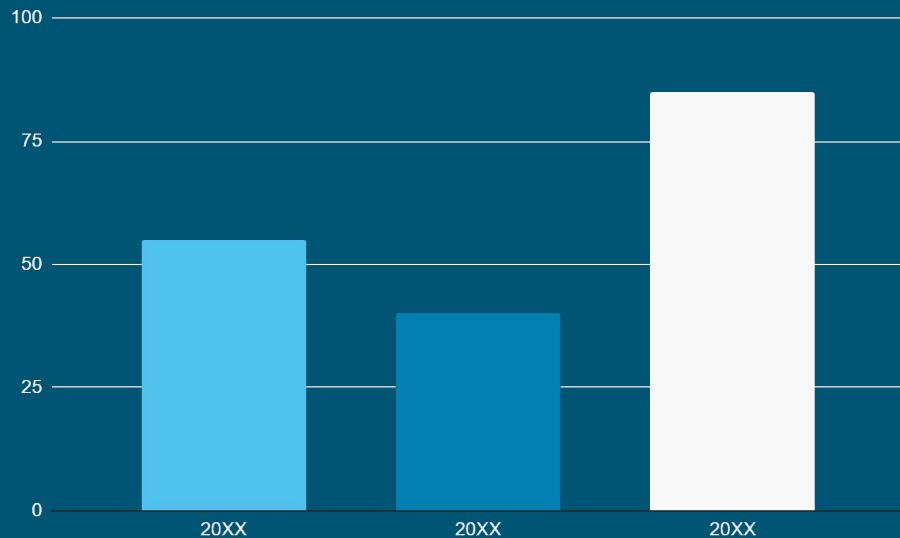
Venus is the second planet from the Sun

● Mercury

Mercury is the closest planet to the Sun

● Mars

Despite being red, Mars is a very cold planet



Follow the link in the graph to modify its data and then paste the new one here. [For more info, click here](#)

Levels of organisation

Organ

Mars is actually a very cold planet

Cell

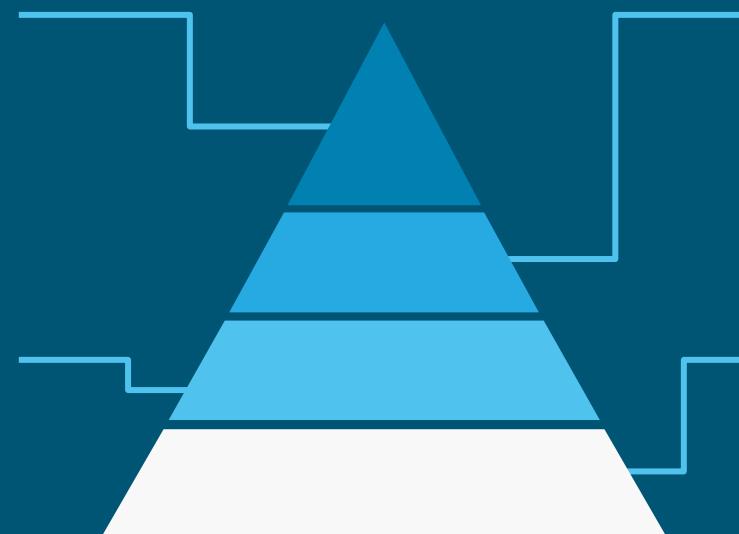
Jupiter is the biggest planet of them all

Tissue

Venus has extremely high temperatures

Organelle

Saturn is a gas giant and has several rings



Our team



Tom Hill

You can speak a bit about
this person here



Alice Harris

You can speak a bit about
this person here

Thanks!

Do you have any questions?

youremail@freepik.com

+34 654 321 432

yourwebsite.com



CREDITS: This presentation template was created by [Slidesgo](#), and includes icons by [Flaticon](#), and infographics & images by [Freepik](#)

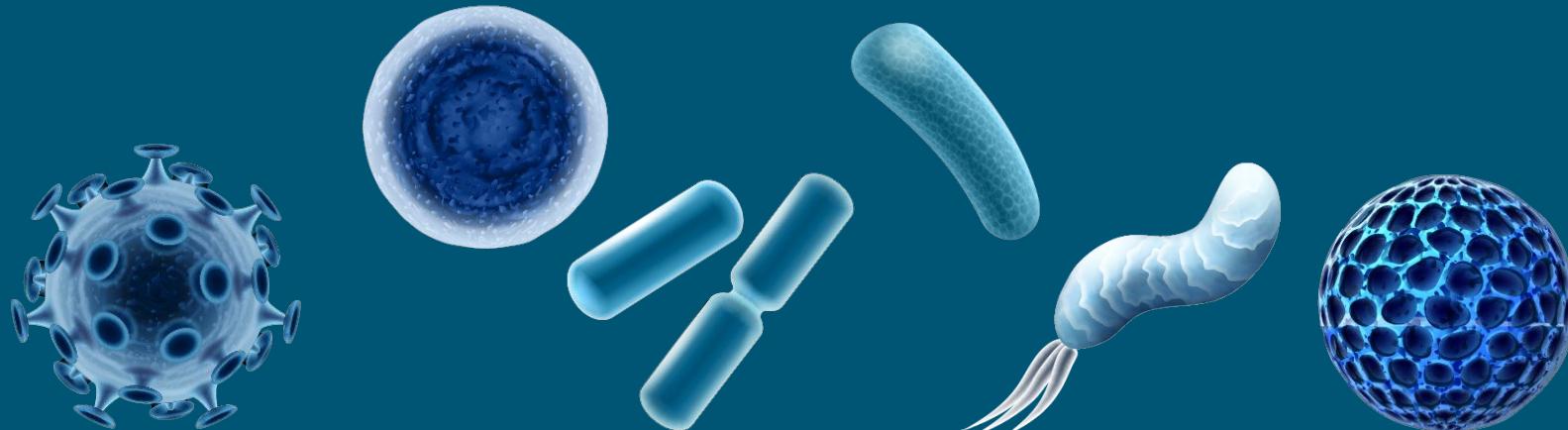
Please keep this slide for attribution

Alternative resources

Here's an assortment of alternative resources whose style fits the one of this template:

Vector

- Vector set of different bacteria and virus cells cocci, spirilla, bacilli, diplobacilli isolated on dark background



Resources

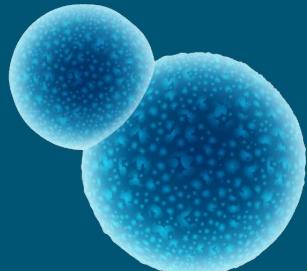
Did you like the resources on this template? Get them for free at our other websites:

Vectors

- [Human cells infographics with flat set of isolated muscle fibers nerve egg and bone cell icons vector illustration](#)
- [Cell anatomy of plant and animal composition with set of colorful educational images with text captions vector illustration](#)
- [Vector set of abstract blue cocci bacteria isolated on dark background](#)

Photos

- [Women working on a chemical project for a new discovery](#)
- [Side view of researchers in the biotechnology laboratory with plant](#)
- [Business man banner concept with copy space](#)
- [Low angle smiley girl with stack of books](#)



Instructions for use

If you have a free account, in order to use this template, you must credit **Slidesgo** by keeping the **Thanks** slide. Please refer to the next slide to read the instructions for premium users.

As a Free user, you are allowed to:

- Modify this template.
- Use it for both personal and commercial projects.

You are not allowed to:

- Sublicense, sell or rent any of Slidesgo Content (or a modified version of Slidesgo Content).
- Distribute Slidesgo Content unless it has been expressly authorized by Slidesgo.
- Include Slidesgo Content in an online or offline database or file.
- Offer Slidesgo templates (or modified versions of Slidesgo templates) for download.
- Acquire the copyright of Slidesgo Content.

For more information about editing slides, please read our FAQs or visit our blog:
<https://slidesgo.com/faqs> and <https://slidesgo.com/slidesgo-school>

Instructions for use (premium users)

As a Premium user, you can use this template without attributing Slidesgo or keeping the "Thanks" slide.

You are allowed to:

- Modify this template.
- Use it for both personal and commercial purposes.
- Hide or delete the "Thanks" slide and the mention to Slidesgo in the credits.
- Share this template in an editable format with people who are not part of your team.

You are not allowed to:

- Sublicense, sell or rent this Slidesgo Template (or a modified version of this Slidesgo Template).
- Distribute this Slidesgo Template (or a modified version of this Slidesgo Template) or include it in a database or in any other product or service that offers downloadable images, icons or presentations that may be subject to distribution or resale.
- Use any of the elements that are part of this Slidesgo Template in an isolated and separated way from this Template.
- Register any of the elements that are part of this template as a trademark or logo, or register it as a work in an intellectual property registry or similar.

For more information about editing slides, please read our FAQs or visit our blog:

<https://slidesgo.com/faqs> and <https://slidesgo.com/slidesgo-school>

Fonts & colors used

This presentation has been made using the following fonts:

Poppins

(<https://fonts.google.com/specimen/Poppins>)

Be Vietnam Pro

(<https://fonts.google.com/specimen/Be+Vietnam+Pro>)

#f8f8f8

#0e2a35

#4fc2ee

#27aae1

#0081b2

#005575

#3c5292

Storyset

Create your Story with our illustrated concepts. Choose the style you like the most, edit its colors, pick the background and layers you want to show and bring them to life with the animator panel! It will boost your presentation. Check out [how it works](#).



Pana



Amico



Bro



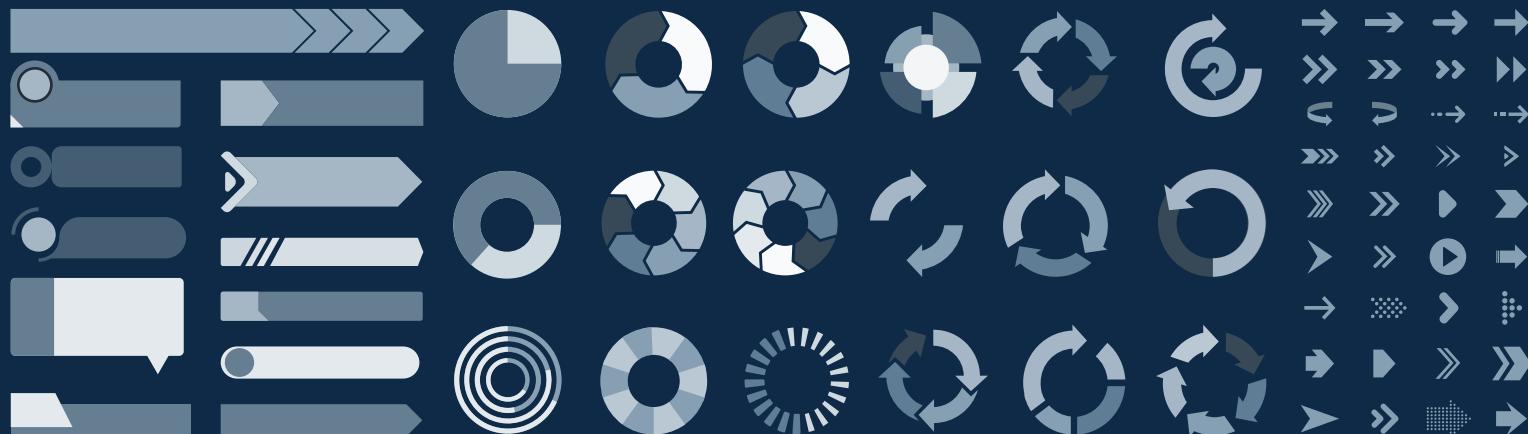
Rafiki



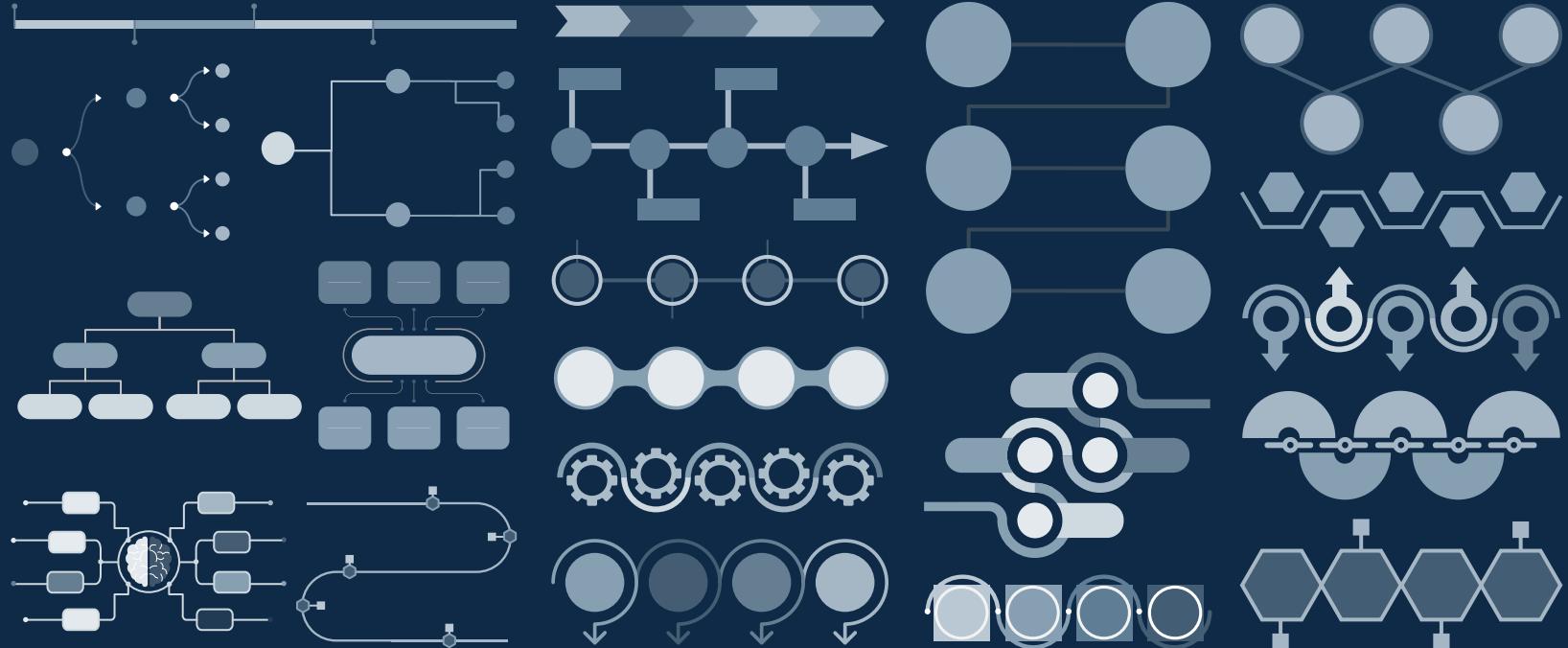
Cuate

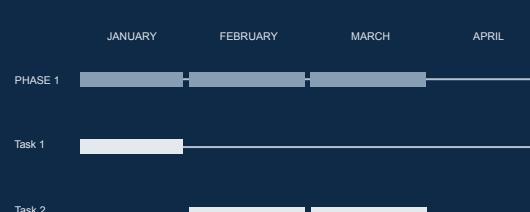
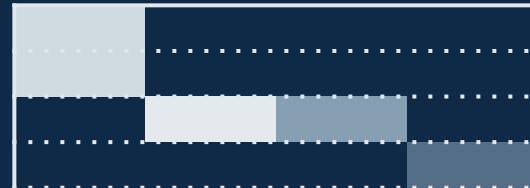
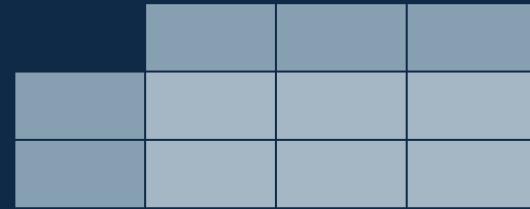
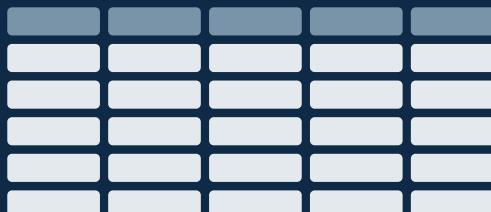
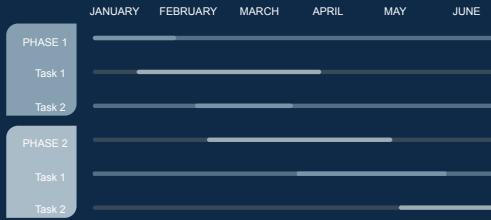
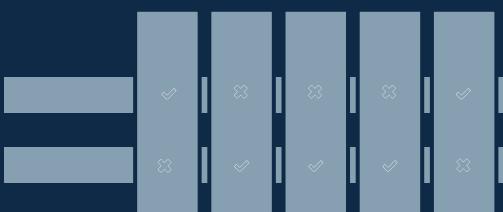
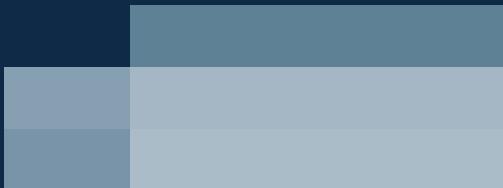
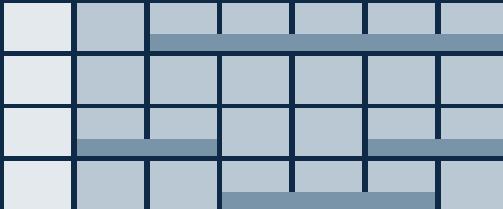
Use our editable graphic resources...

You can easily **resize** these resources without losing quality. To **change the color**, just ungroup the resource and click on the object you want to change. Then, click on the paint bucket and select the color you want. Group the resource again when you're done. You can also look for more **infographics** on Slidesgo.

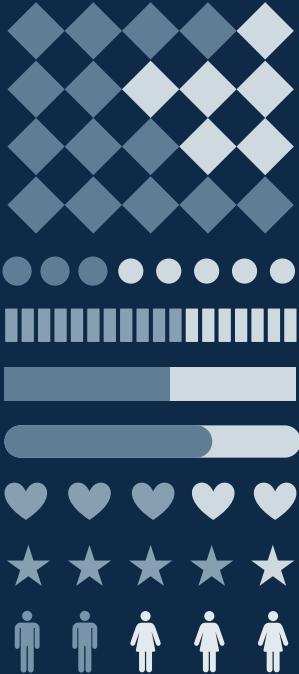
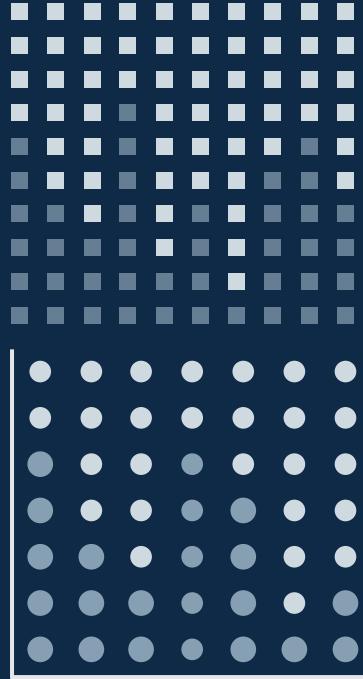












...and our sets of editable icons

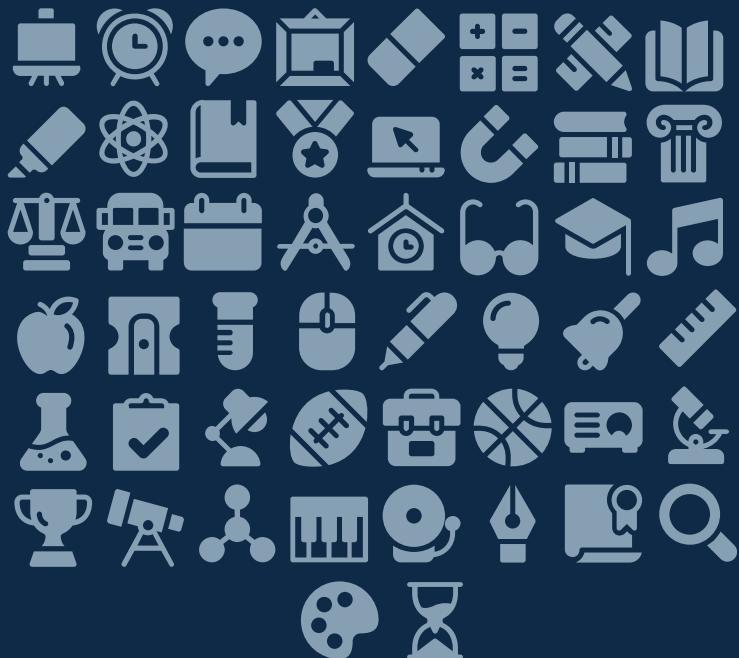
You can **resize** these icons without losing quality.

You can **change the stroke and fill color**; just select the icon and click on the **paint bucket/pen**.

In Google Slides, you can also use **Flaticon's extension**, allowing you to customize and add even more icons.



Educational Icons



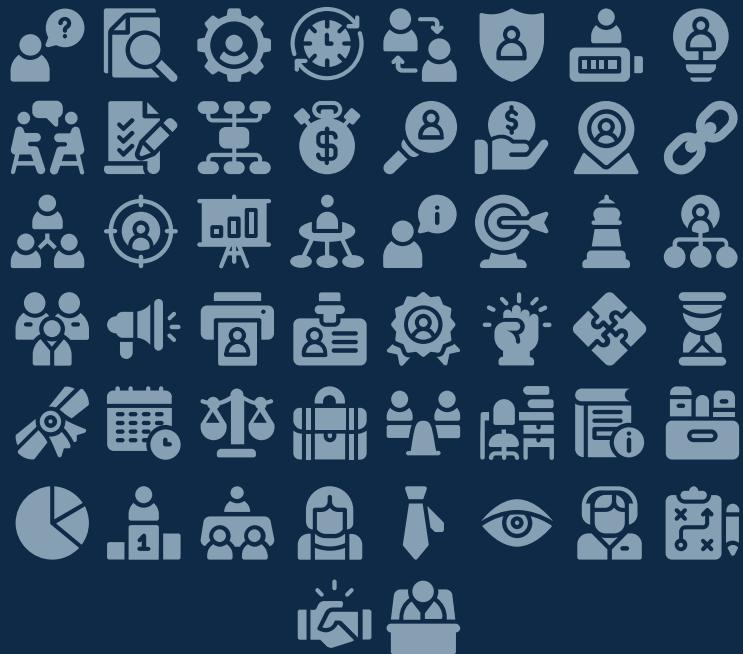
Medical Icons



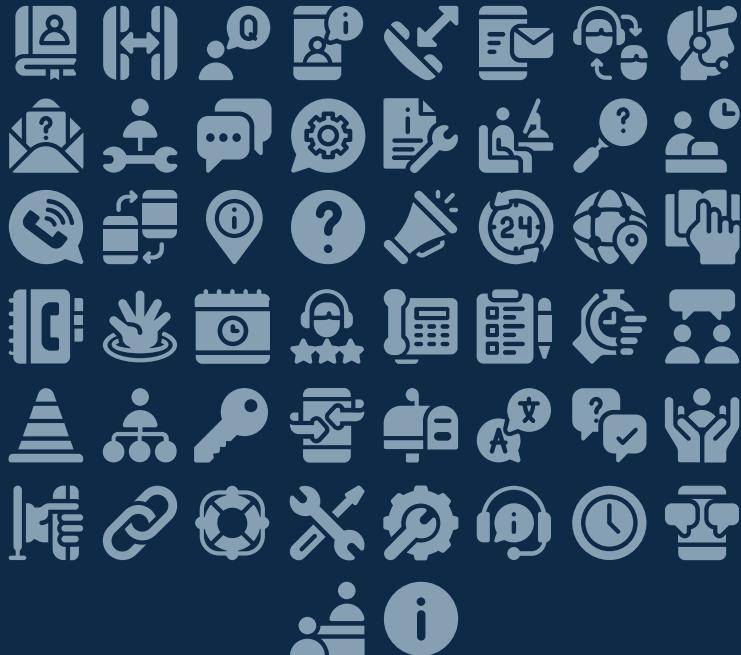
Business Icons



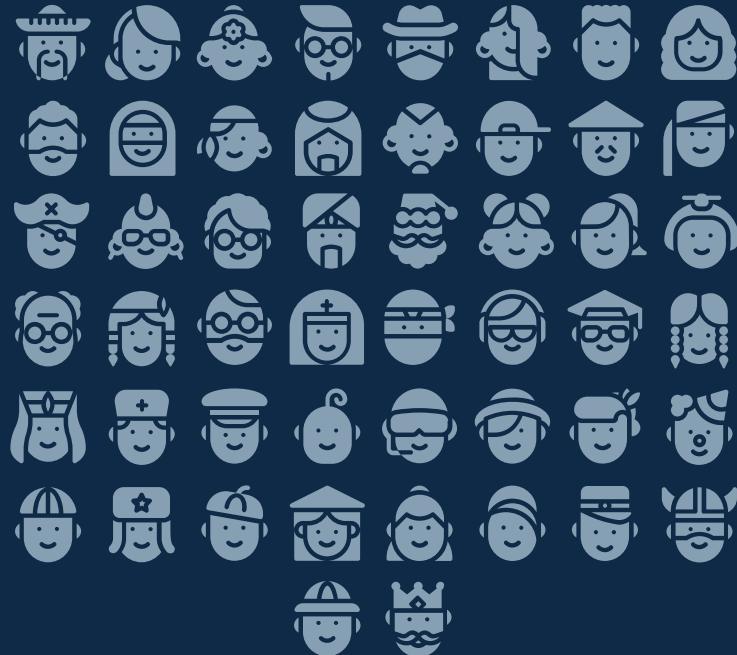
Teamwork Icons



Help & Support Icons



Avatar Icons



Creative Process Icons



Performing Arts Icons



Nature Icons



SEO & Marketing Icons



