
2 EXPERIMENTAL DATA

For this challenge, we will use human solutions to the Borderlands Science BLS game that uses genome fragments of human gut microbes. The data set is structured as follows:

- **steps:** A chronological list of player moves, detailing insertions of gaps ("") into DNA sequences. Each move is represented as a tuple (`sequence_index`, `position`):
 - *sequence_index*: The index of the sequence (0 to 18, 0-based indexing) where the insertion will occur.
 - *position*: The position (0 to 9, 0-based indexing) within the sequence where a gap (") will be inserted.
 - *Examples*:
 - * (1, 3) means insert a gap at position 4 in sequence 2.
 - * (0, 0) means insert a gap at position 1 in sequence 1.
- **score:** The player's final numerical alignment score computed by the gearbox score function available in the starter notebook
- **start:** The initial, unaligned DNA sequence configuration.
- **solution:** The final alignment of the DNA sequence that reflects all the insertions of the gaps.
- **moves:** A list of moves performed by the player, each formatted as {sequence}position}+;{timestamp}: Each sequence is labeled (A, B, C, D, E, F), representing different DNA fragment. Each column is numbered starting from 0. Examples:
 - **B1+;6653** → Insert a gap (") at **position 1** in sequence **B** (AGA—)
Result: A-GA—
 - **D0+;1254** → Insert a gap (") at **position 0** in sequence **D** (GGC—).
Result: -GGC—
- **accepted pairs** : Commonly substituted amino acid or nucleotide pairs in MSA, derived from evolutionary data and used for scoring alignments.

Note : both "steps" and "moves" represent the insertions made by the user, however, it will likely be more convenient for you to use steps.

3 OBJECTIVE

*Your goal is to implement a model that learns to make the optimal moves and returns the **optimal final solution** (i.e. the solution that yields the highest score). The judges must be able to run the notebooks and obtain the same results. A starter notebook can be found [here](#).*

4 EVALUATION

EXPLORATIONS (15 POINTS)

Your team will receive a score ranging from 1 to 15 points, primarily based on the following criteria:

- **Understanding of the Problem:** How well does your team identify the key components and potential challenges associated with the problem?
- **Strategy and Solution:** How clearly does your team articulate a strategy to solve the problem based on their understanding of the task?
- **Explanation and Justification:** How effectively does your team communicate their chosen approach, ensuring it is clear, complete, unambiguous, and concise?

RESULTS (25 POINTS)

Your team will receive a score ranging from 1 to 25 points, primarily based on the following criteria:

- **Performance on Blind Data:** How well does your approach optimize the game score on unseen data (not provided during training or testing)?
- **Quality of Solutions:** How effectively does your approach identify effective (good) solutions, even if they may not always be optimal?
- **Validation Strategy:** How well does your team implement a structured strategy to ensure that all the problem's requirements are met?

CONCLUSIONS (10 POINTS)

Your team will receive a score ranging from 1 to 10 points, primarily based on the following criteria:

- **Presentation Quality:** Does your presentation include all the necessary, relevant, and expected components?
- **Clarity and Innovation:** How well does your proposed approach articulate itself, ensuring it is innovative, clear, unambiguous, quantified, and precise?
- **Engagement and Communication:** How effectively do visual elements and language enhance audience understanding and engagement?

5 DISCLAIMER

The test set will be released at 11:00 a.m. on Sunday.

For submission, put your notebook, testing dataframe with your solutions inside a "Solutions" column, and slides on a GitHub repository and post it on Devpost.