



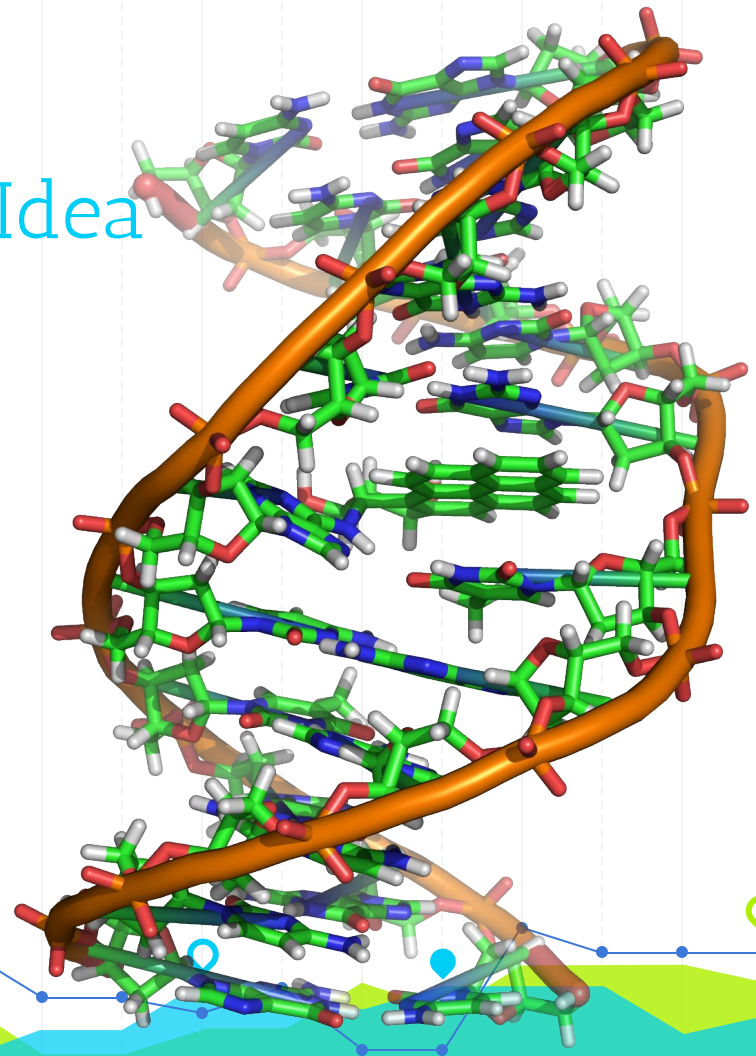
# Nitrogenase Finder

Rebecca Gettys, Liv Kelley, Erica Lee



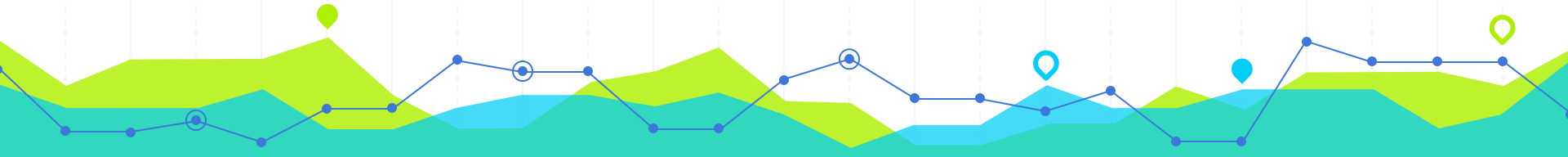
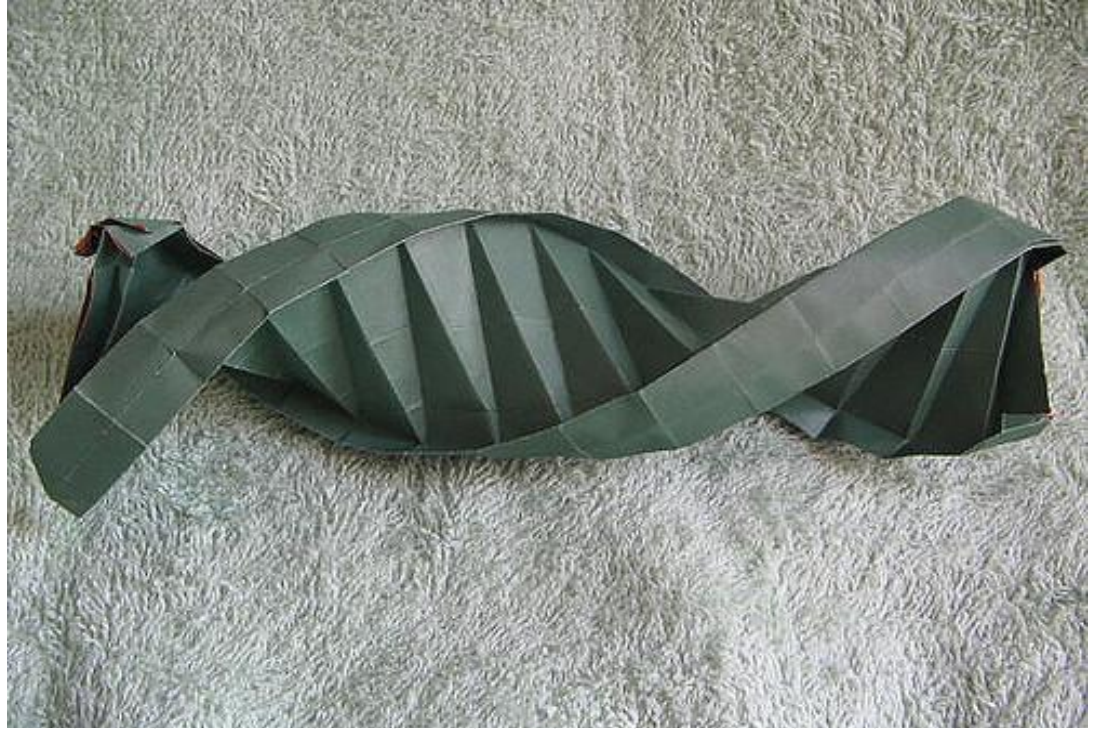
# Background for the Big Idea

- ◉ Biological explanations
- ◉ Jean's summer work
- ◉ Recreating something like "BLAST" for understanding
- ◉ Won't be a full version
- ◉ Should still be useful

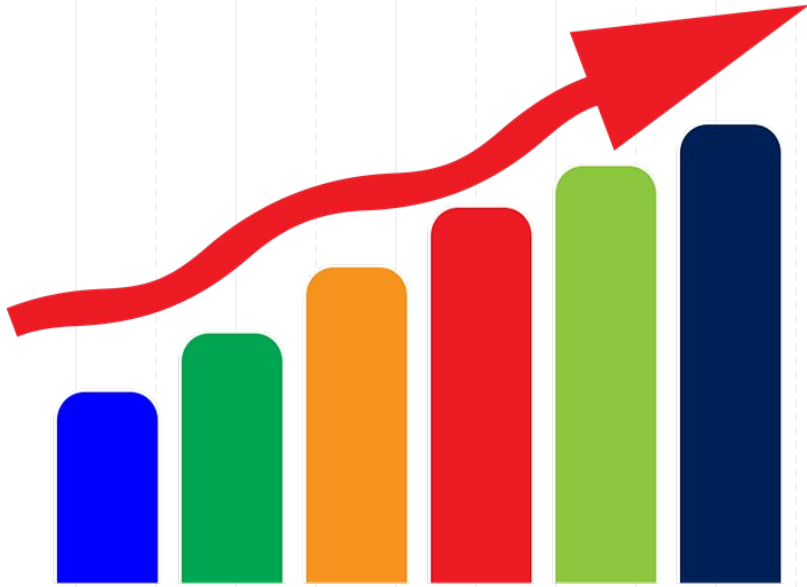


# The Big Idea

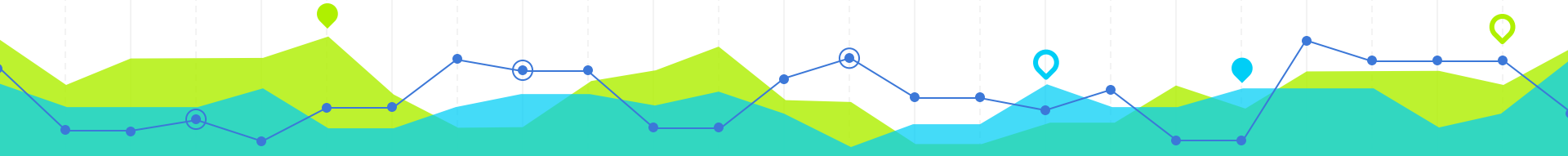
- ◉ Identify genes responsible for nitrogen fixation within bacterial genomes.
- ◉ Find location.
- ◉ Accurately visualize the DNA.
- ◉ Visualize the percentage similarity.



# Progress

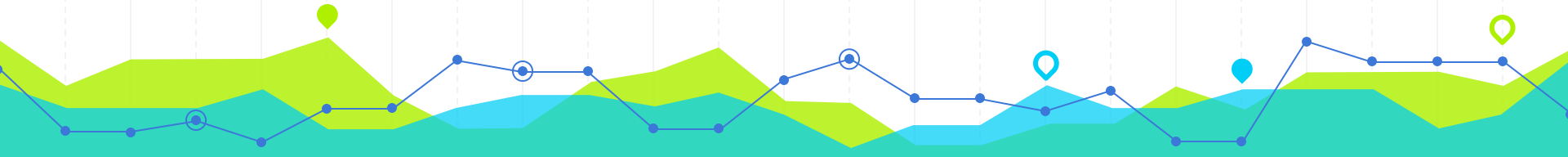
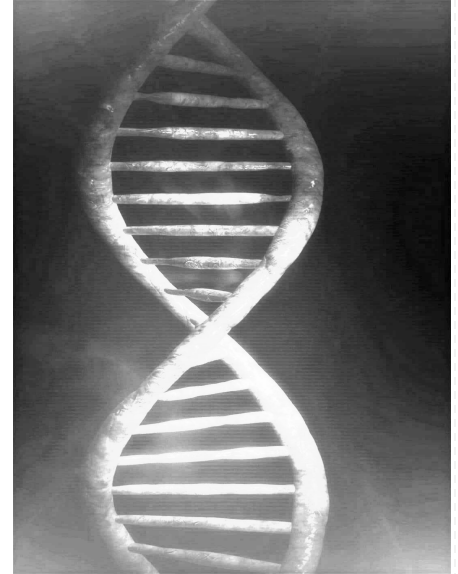


- Nitrogenase finder vs. gene finder
- Simple visualization vs. real data
  - Length smaller segments cover
  - Capacity for a reverse complement feature

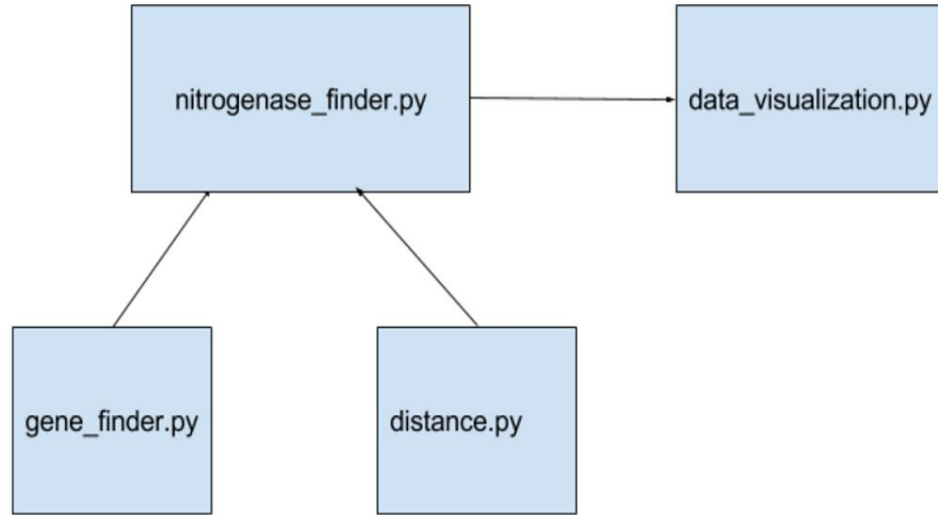


# Implementation and Design Details

- ◉ Levenshtein algorithm
- ◉ Object oriented visualization
- ◉ A few different steps.
  - ◉ Getting the bacterial Genome
  - ◉ Plugging it into nitrogenase finder (find the genes and output some data about them)
  - ◉ Properly running the visualization

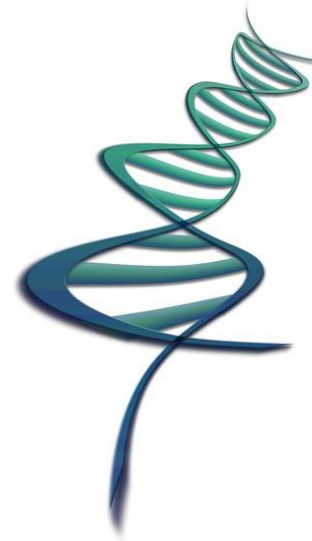


# Overall Structure



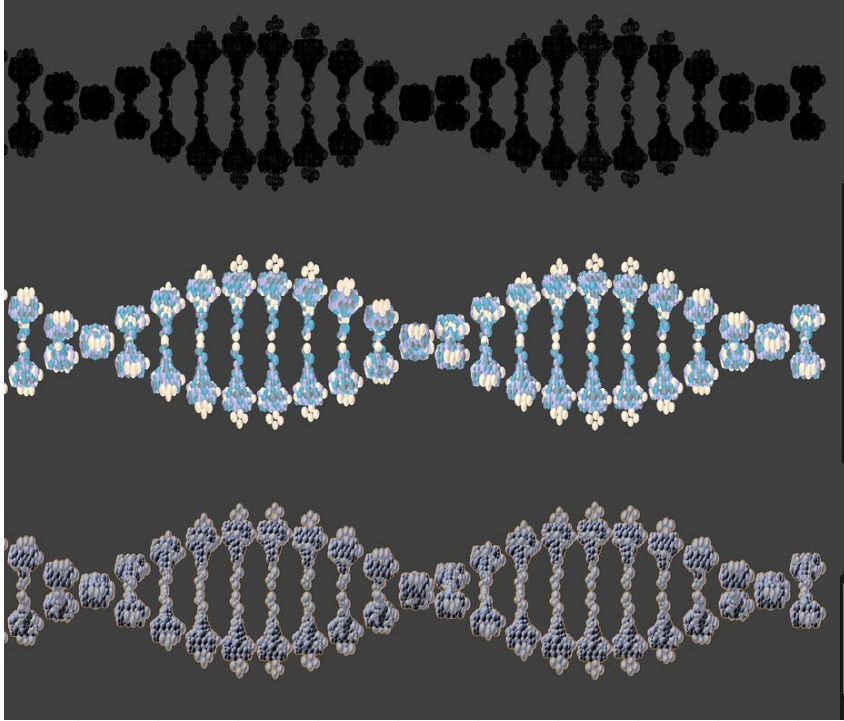
# Gene\_Finder/Distance/Nitrogenase Finder Specific

- ◉ Nitrogenase finder takes the genomes and uses levenshtein distances to processes them into:
  - ◉ Possible ORFs (matches) - length, start, end locations
  - ◉ Levenshtein Distances/percent similarity
  - ◉ Includes forward/reverse DNA sequence
  - ◉ Length of the nitrogenase





# For Gene\_Finder

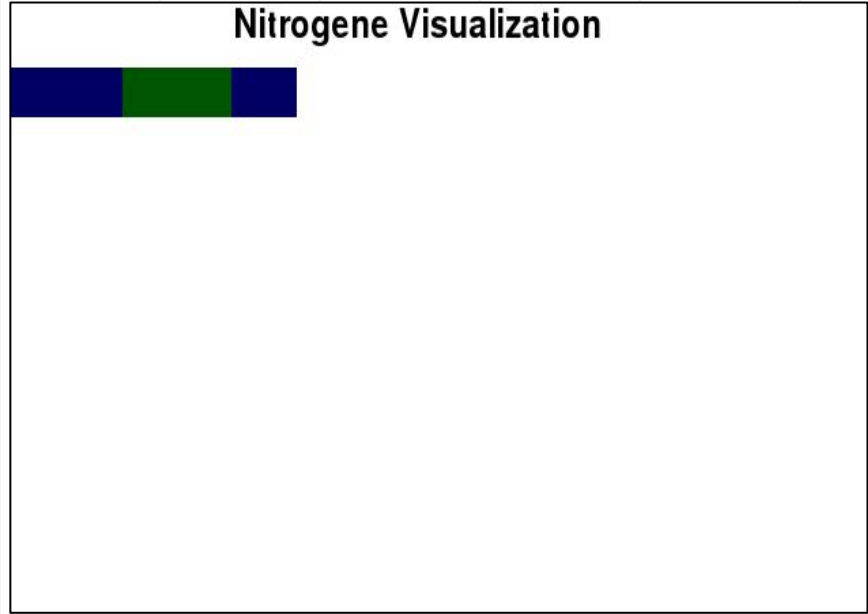


- Working, more optimization would be nice
- Possibility for adding object orientation to increase speed
- Also still haven't switched to pypy - will make it \*much\* faster
- No longer pursuing match-within-a-match feature
- Nitrogenase\_finder is in great shape



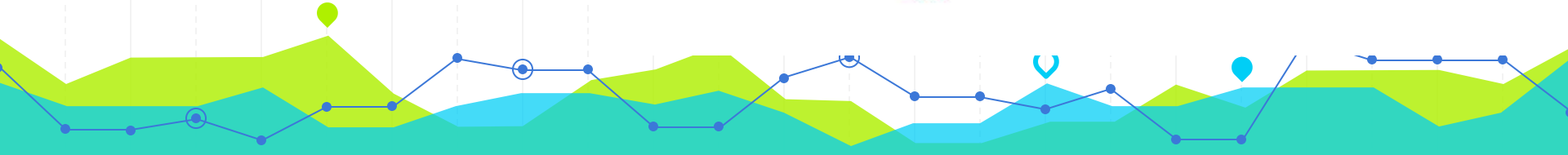
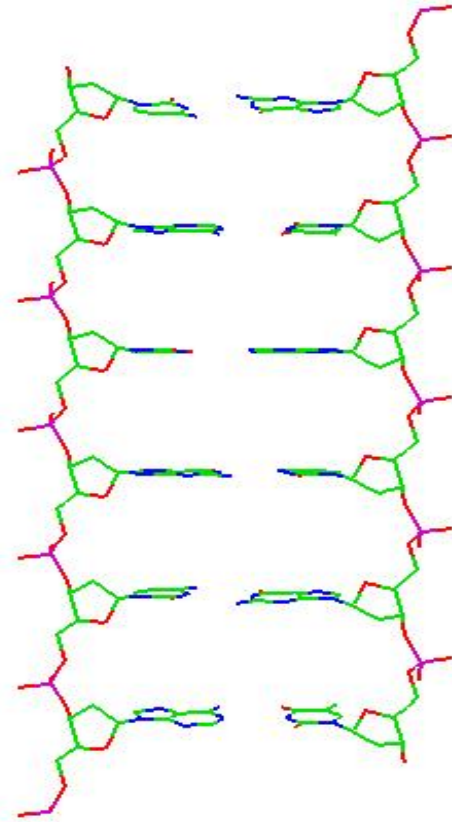
# Visualization Specifics

- Fixing bugs
- Pickle processing
- Input data
- Command line threshold
- For loops drawing rectangles
- If statements determining color
- For loops for scrolling



# For Visualization

- Could run faster
- Is there anything unnecessary left in the code?
- Is it getting accurate data



# Overall Project Status

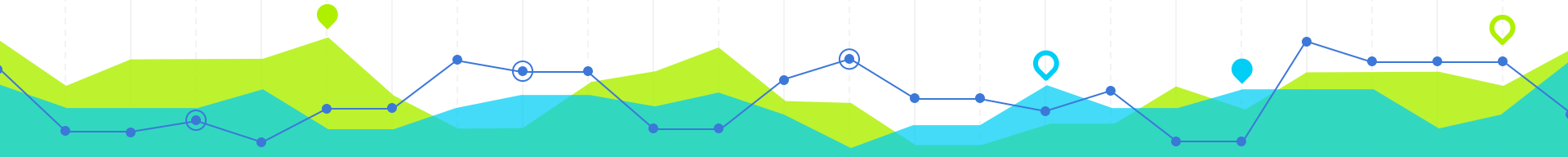
- ◉ Code spits out the right data
- ◉ Could include more features
- ◉ Bugs
- ◉ Code running time is problematic sometimes
- ◉ Crashing issues



# Future Implementation and Design



- Adding a key for the different colors
  - Other visualization features
- Making a better user interface
- Object orientation gene\_finder
  - Speeding up code



# Thank You!



## Picture Sources

- 1) <https://pixabay.com/en/dna-science-biology-medical-304162/>
- 2) <https://en.wikipedia.org/wiki/DNA>
- 3) <https://www.flickr.com/photos/georigami/140085167>
- 4) <https://en.wikipedia.org/wiki/DNA>
- 5) <https://pixabay.com/en/ascending-graph-bar-graphs-progress-1173935/>
- 6) <https://pixabay.com/en/deoxyribonucleic-acid-graphic-dna-64273/>
- 7) <https://www.flickr.com/photos/bombaglot/3272357111>
- 8) <https://www.flickr.com/photos/bombaglot/3272357111>

