

# **master's project notes**

# Contents

<b>1</b>	<b>Evaluation of NumPy vs CuPy in npstructures and BioNumpy</b>	<b>3</b>
----------	---	----------

# 1 Evaluation of NumPy vs CuPy in npstructures and BioNumpy

To initiate exploration of what CuPy does well in comparison to NumPy, I am carefully benchmarking each step of the k-mer counting pipeline, going from an input fasta file of 20 million reads to the finished counter object, in order to fully understand which parts of the npstructures and BioNumpy code is lacking relative performance using both NumPy and CuPy as the backend array module.

This pipeline can be abstracted into x parts: 1) reading and formatting a chunk of reads from the input fasta file, resulting in a npstructures.RaggedArray object containing a single read per row, where each base is represented as a single unsigned 8-bit integer.