

Project: Speed Optimization of Augustus software

Goal: optimizing for speed and parallelizing the implementation of [AUGUSTUS](#)

About Augustus

- After sequencing the genome (DNA) of any species, molecular biologists are tasked with figuring out which sections of the genome encode genes (exons) and which sections are non-coding (introns). This task is called [gene prediction](#).
- AUGUSTUS is a gene-prediction software (part of the [BRAKER](#) pipeline) that can locate genes by looking at just the genome sequence (ab initio). If available, it can also incorporate "extrinsic hints" from protein / RNA sequencing of that species.
- AUGUSTUS usually takes a long time for gene prediction, even when running on High-Performance Computers (HPCs). Two reasons for this are poor parallelization and performance bottlenecks. Hence, the goal is to improve the implementation of AUGUSTUS to make it faster and more suitable for HPCs.

Physical requirements: The work will be done by connecting to HPCs remotely. So, access to a laptop/computer (basic) with internet connection (1+ Mbps) is needed.

Must be good at parallelization, multi-threading, concurrency, code optimization techniques, process synchronization, OS concepts, basics of molecular biology, etc

Tools and Languages

1. Linux OS and BASH
2. C/C++, Python, [Chapel](#)
3. Code profiling tools – [gprof](#) and [Intel vtune profiler](#)

Contact

1. MTech student, Rohit S <rohit2013107@iitgoa.ac.in>
2. Professor, Sharad Sinha <sharad@iitgoa.ac.in>