

## **CUFSAA Resource Request**

### **Seeking Funds to Upgrade the Computer Facilities in the Department of Plant Sciences, University of Colombo for the Bioinformatics Special Degree**

#### **1.0 Requester's Information**

**1.1 Full Name :** Dr. C. M. Nanayakkara

**1.2 Department :** Plant Sciences

**1.3 Title :** Head (Department of Plant Sciences, University of Colombo)

**1.4 Phone No :** +94 11 258 5038

**1.5 Email:** chandi@pts.cmb.ac.lk

#### **2.0 Resource Requested**

Two high-performance computers

#### **3.0 Significance**

##### **3.1 Explain how the requested resource benefit students and/or the department**

In 2006, the Department of Plant Sciences, University of Colombo introduced a special degree program in Bioinformatics for the first time in Sri Lanka, with the intention of providing an educational background that blends with Biology and Computer Science. The field of Bioinformatics is one of the promising research areas worldwide. Over the last few years, this field has grown enormously due to the generation of voluminous and incremental datasets, especially from genomic, transcriptomic, proteomic and metabolomics-based research data. Dealing with such complex, big datasets require more sophisticated and efficient algorithms and computing infrastructures.

##### **3.2 Would the requested resource benefit students? Yes No**

###### **3.2.1. If YES, explain how**

The requested resource could no doubt benefit students and young academic staff members of the Department to carry quality research and build collaborations with international research laboratories. Furthermore, all the students attached to the Department especially Bioinformatics special degree

students will get an opportunity to practice and apply related theories learned in the classroom. Currently, the Department hosts 5-8 Bioinformatics special degree students every year. Each special degree student is expected to conduct a bioinformatics related research projects as a significant part of their final year course work. However, lack of high-end computer facilities has limited the type of projects that can be conducted and the level of quality at which they can be conducted. Please see Section 4.1 for more details.

## **4.0 Current Status**

### **4.1 Current status of affairs in the absence of the requested resource**

Currently, inadequate computational power to handle complex biological datasets has become a limiting factor for conducting high-quality research in this field, hindering the progress of students and scientists. The Department has 03 desktop computers with 8 GB RAM to conduct research in Bioinformatics, and most often these are also shared with other special degree students in the Department. These computers do not have the capacity to store, manage and process a large amount of biological data generated from high-throughput omics technologies; this computational power is only sufficient to perform the first steps in the Bioinformatics data analysis pipelines. Many of the more complex steps in the pipeline require a high level of RAM capacity.

Most of the algorithms/software we use in research such as, genome/transcriptome mapping and analysis tools (e.g. Bedtools, Samtools, Bowtie2, Tophat, Cufflinks, FastQC), protein structure modeling tools (e.g. ITASSER), phylogenetic analysis tools (e.g. MEGA), drug designing tools are freely available. However, they demand high level of computing power due to the massive amount of information that has to be processed simultaneously. As an alternative, students often use their own laptops to perform analyses. However, most personal laptops do not have the required capacity to handle big datasets; most of the student laptops have either 4GB or 8GB memory (RAM) with i5 or maximum i7 processors. Most often, students have to run programs overnight, and frequent laptop crashes have been experienced due to overloaded hard-drives. Hence, students are compelled to limit their analyses to low memory consuming algorithms/software or perform for a small number of samples. Although the students show motivation and enthusiasm to conduct cutting-edge research, lack of computer infrastructures greatly has hindered generation of high quality, publishable results from their undergraduate research projects. On the other hand, not all students have personal laptops, and the only option they have is to utilize computer facilities available at the Department.

Furthermore, almost all the theoretical aspects relating to the field including Image Processing, Graphic Designing, and Neural Computing are conducted by experts from the School of Computing (UCSC), University of Colombo. Unfortunately, our students lack computer facilities in the Department to practice these programs.

**4.2 Is there any other alternative if CUFSA could not accommodate the requested resource?**

There is no any other alternative as many funding agencies do not provide funds to purchase high-performance computers for long-term use. In most cases funding is provided to purchase computers to conduct a specific project and it is required that the purchased computers are returned to the funding agency once the project is completed (e.g. funding obtained from the National Science Foundation).

**4.3 Explain the impact on students and/or on the department if CUFSA could not accommodate the requested resource**

This resource is requested for both teaching and research purposes. It will limit students getting involved in high quality research and also obtaining hands-on experience in novel computational tools that may require high computational power.

**5.0 Rationale - Please explain the rationale of your request**

At present, the Department accommodates 08 undergraduate students for the B.Sc. in Bioinformatics program every year. As a partial fulfillment of their degree program, all the special degree students are required to conduct a research project in the respective field. The Department has qualified and experienced scientists to conduct and supervise Bioinformatics related research projects. Hence, the Department wishes to set up a high-performance computer facility to enhance applied bioinformatics skills among undergraduate students and initiate genome-scale research projects. Some of the ongoing undergraduate research projects for 2019 are as follows,

- Analysis of bacterial community profiling using metagenomic approaches
- Analysis of effector proteins to elucidate infection strategies of fungal plant pathogens
- Genome-wide identification and analysis of Late Embryogenesis Abundant (LEA) genes in Banana (*Musa acuminata*): an insight into motif and gene structure arrangement
- Uncovering gene regulatory networks controlling plant embryogenesis

- Transcriptome analysis of *Arabidopsis POL2* mutants – an insight to gene functions at a global scale
- Protein-protein interaction network module analysis for finding functionally important genes in root development of *Oryza sativa*.

Analyses to be performed through the proposed high-performance computer facility:

- Mapping/assembly genomes and transcriptomes
- Downstream Bioinformatics analyses of high-throughput sequencing data such as differential gene expression analysis, Single nucleotide polymorphism (SNP) detection, construction and analysis of biological networks i.e. gene regulatory networks, protein-protein interaction networks
- Metagenomic analysis
- Phylogenetic analysis
- Statistical analyses of large datasets i.e. >30GB, >50GB
- Protein structure modeling and prediction
- Design computational models for biological environmental mechanisms

Open-source software packages to be installed/run in the proposed computer facility:

- R and Bioconductor
- Bedtools
- Samtools
- Bowtie2 and other sequence alignment tools including de novo sequence alignment algorithms
- Qiime2
- FASTQC
- Phylogenetic analysis tools such as MEGA for large scale datasets
- Matlab
- Protein structure modeling tools such as ITASSER
- Python3
- Ofc Linux OS
- Drug designing tools

**6.0 Other Funding Sources - Besides CUFSAA what are the other potential funding sources for your request?**

University of Colombo small research grant (competitive research grant)

**6.1 Have you pursued other funding resources?** Yes (University of Colombo small research grant)

**6.2 If not why?**

**6.3 If yes what was the outcome?** Unsuccessful

**7.0 How soon do you need the requested resource? and why?**

Since the new batch of the Bioinformatics special degree has just started their research projects, if the requested resource could be delivered immediately, that will benefit the current batch of students as well.

**8.0 Propose specifications such as the brand, the model, the cost and the source for purchasing.**

The Department wishes to purchase **two** high-performance computers to meet the current requirement.

Specifications and the tentative itemized price for **one machine** are given in the following table. (all the values are in LKR)

<b>SPECIFICATIONS FOR BIOINFORMATICS RESEARCH PC</b>		
Processor	Intel Core i9 7900X 3.3-4.3Ghz 10-core/20-thread Processor	180000
Motherboard	MSI X299 SLI Plus	60000
Memory	Corsair Vengeance LPX 32GB (2x16GB) DDR4 3200MHz Kit	80000
	Corsair Vengeance LPX 32GB (2x16GB) DDR4 3200MHz Kit	80000
Storage	Samsung 860 Evo 250GB M.2 Solid State Drive	23500
	Toshiba DT01ACA200 2TB 7200 RPM SATA 6Gbps	15500
Casing	Corsair Carbide-series 100R	11500
Power Supply	Corsair RM750X 750W 80+ Gold Modular	28500
	<b>SUM TOTAL</b>	<b>479000</b>

## **8.1 Alternative specifications**

These are the minimum specifications required for a high performance computer, in order to conduct bioinformatics analyses.

## **9.0 Provided CUFSAA fulfills your request, explain how you would report the progress on the impact of the resource.**

We could provide progress reports in 6 and 12 months after successful installation of the resource. The reports will include the number and types of projects conducted, their outcomes, data processed, bioinformatics tools used, etc. In addition, **CUFSAA** can be informed of any abstracts/full research articles published.

## **10.0 Are you willing to provide an annual progress report on the use and impact of the resource to CUFSAA for two consecutive years?**

Yes. In addition, we will acknowledge CUFSAA for providing necessary facilities in abstracts/full research articles that will result due to this facility.