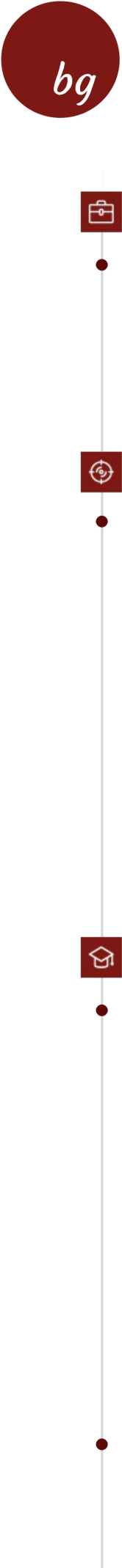
**Bishwa K. Giri**

**Data Scientist, Python Developer**

**Contact :** [kirannbishwa01@gmail.com](mailto:kirannbishwa01@gmail.com) , 336-3929572



**Professional Summary**

* 6+ years of experience with genomics and transcriptomics data analyses - sequence alignment, variant calling, haplotype phasing and population genetics, genotype to phenotype association.
* **Experienced in wet lab methods like next generation sequence reads library preparation, qPCR, PCR, primer design, allele testing, etc.**
* **5+ years of experience in programming and development of several full-length application and pipelines for big data analyses, transformation and statistical tests.**
* Experienced in working on Windows, UNIX, and Linux operating systems.
* **Experienced in remotely running data analyses on a HPC (high performance computation) cluster using ssh protocol.**
* **Advanced knowledge of big data analyses in bio-informatics** – regularly analyzed several gigabytes (+500 GBs) of sequence reads using front end user (command line based) high-end tools like BWA, GATK pipeline, Picard, phaser, rnaSTAR, g2gtools, samtools, bowtie, tophat, bcftools, vcftools, emase, IGB, IGV, GenomeBrowser. QCtools, BLAST, BOWTIE, MUSCLE etc.
* **Regularly anlayzed genomic, transcriptomics databases** – SAM/BAM files, VCF, GTF, GFF, fasta, fastq, RefSeq, HapRefPanel.
* **Skilled in the development of application required in bio-informatics domian –** Independently developed and contributed to several full length open source bioinformatics application and projects for analyzing big databases from genomics and RNAseq technologies at UNC Greensboro. Also, forked and optimized other public repositories involved in bio-informatics data analyses.

**Applications hosted on GitHub @** [**https://github.com/everestial**](https://github.com/everestial).

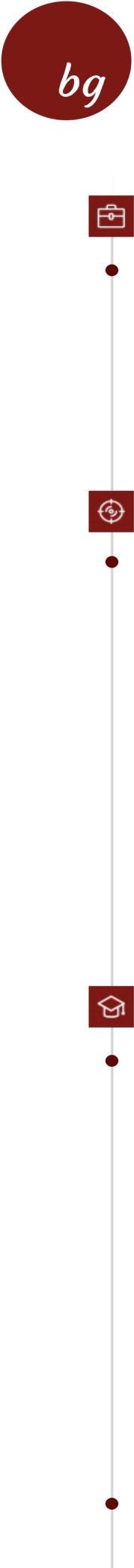
* + - * + **Developed data processing pipeline** [***aseCADG***](https://github.com/everestial/ASE-CADG) **for analyzing ASE (allele specific expression) from RNAseq reads of the F1 hybrids.**
        + **Developed haplotype phasing application using first order markov chains -**[***phase-Extender***](https://github.com/everestial/phase-Extender) ***and*** [***phase-Stitcher***](https://github.com/everestial/pHASE-Stitcher) ***that forms a part of***  [***phaseIT***](https://github.com/everestial/PhaseIT/tree/master) ***pipeline.***
        + **Developed python application** [***VCF-Simplify***](https://github.com/everestial/VCF-Simplify)**for simplifying and building VCF files for easy VCF data manipulation.**
* Experience with online bio-informatics tools, resources and databases like NCBI, Ensemble, Plaza, KEGG, cyverse, iplant, uniprot, OMIM, 1000 genomes, HapMap etc.
* **Solid knowledge on theoretical and applied aspects of both frequentist and bayesian statistics.**
* **Experienced in data filtration, wrangling, transformation and structured flow** using python and R and associated packages - pandas, dask, numpy, cyvcf2, tidyr, dplyr, shinyr, SQLite, Rcmdr, sci-kit learn.
* **Experienced in data visualization using python and R libraries** - matplotlib, ggplot2, plotly, plot, etc.
* **Experienced in application and data analyses workflow optimization** via multiprocessing, debugging and troubleshooting.
* **Experienced in cyber security domain -** used tools like splunk, tanium, elasticsearch, threatconnect, NDEX, etc. for data extraction and generating threat intelligence.
* **Experienced in jira, jenkins, confluence** for knowledge sharing, issues, application and development tracking.

**Skills**

* **Office Productivity :** Microsoft Office Package (Word, Excel, PowerPoint), LlibreOffice, Google docs.
* **Operating Systems :** Windows 98, 2000 Server, XP, Vista, 7 and UNIX/Linux, ubuntu.
* **Programming, Scripting and Database management**
* bash, sed, awk, ssh.
* **Advanced skills on Python and R programming.**
* Rstudio, Rcmdr, bioconductor, JupyterLab.
* **Python Packages:**  pandas, dask, pyspark, numpy, Sqlite, cyvcf2, matplotlib, Django, TkInter, keras, multiprocessing, scrapy, etc.
* **R packages:** ggplot2, plotly, tidyr, dplyr, shinyr, enrichGO etc.
* **BioConductor libraries:** DESeq2, edgeR, karyoploteR, clusterprofiler, DOSE, etc.
* Experience with SQL, NoSQL, JSON database.
* Experience working with larger than memory database using dask.
* Good knowledge of data architecture used in numpy, pandas and dask like arrays, dataframe, hdf5, etc.
* Working knowledge of GO programming language.
* Familiar with google cloud genomics data analyses and data processing pipeline setup.
* **Security and log management tools knowledge:**

Tanium, Splunk, Threatconnect, ElasticSearch, NDEX (network data explorer).

* **Statistics:**
* Advance knowledge of both frequentist and bayesian statistics.
* Experienced in statistical data analyses using regression, ANOVA, markov chain, MCMC (markov chain monte carlo) models, hidden markov model, etc.
* **HPC (high perfomance computation) :** Experienced working in HPC system with job submission, data query and database management.
* **IDE for application DevOps:** Experience with version control/integration/tracking tools like Git, Subversion, PyCharm, Docker, Anaconda navigator, Spyder, JupyterLab.
* Excellent interpersonal and time management skills and motivated by challenges and ability to perform well under pressure and multi-task effectively.
* Excellent communication skills in both verbal and professional writing. Fluent in languages : english, hindi, nepali.
* **Highly skilled in interdisciplinary framework with integrative knowledge in biology, molecular biology, bioinformatics, statistics, big database transformation and management, scripting and application development.**
* **Skilled in cyber security domain – analyze machine data logs to identify cyber and information threat. Generate network visualization for tracking network/host/ip relationships to provide threat intelligence.**
* Ability to work in a team to achieve team goals.
* Willing to learn more and open to the ideas of application and methods developments for advancement of STEM, interdisciplinary projects , open source projects and research.
* Willing to actively engage in the use of new generation of technologies like machine learning, deep learning, cloud computing etc.

 **Experience**

**Data Engineer for Wells Fargo via. Apex Systems (Sept, 2018-Current)**

* Extracted machine data from Splunk indexes to monitor, research and provide cyber threat intelligence.
* Used tanium, ThreatConnect, NDEX to monitor and further refine security data logs to reduce false positives.
* Involved in the development of MCMC (markov chain monte carlo) statistical models to generate quantitative risk assessment.
* Involved in the development of network visualization of network/hosts/ip to provide threat intelligence monitoring.

**Graduate Assistant (7+ years) for UNC Greensboro, Biology Department (2011-2018)**

* Taught General Biology Lab courses (Bio 105, Bio 111 & Bio 112) and upper level Bio 392 (Genetics) Lab.
* Trained undergraduate researchers in several biology, bio-informatics and data analyses research projects.

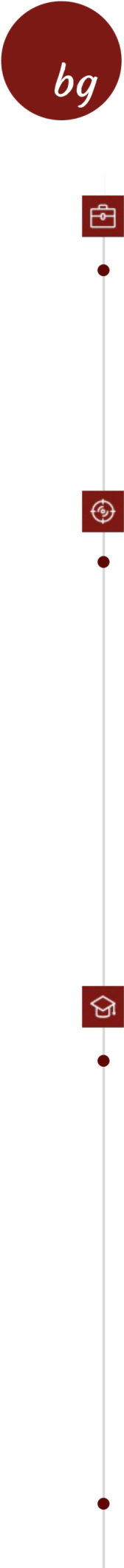
**Researcher and data analyst (7+ years) in the lab of Dr. David Remington, UNC Greensboro**

* Researched evolution of life history tradeoffs in [***Arabidopsis lyrata***](https://plants.usda.gov/core/profile?symbol=arly2).
* Maintained research stock genotypes of *A. lyrata* in greenhouse and growth chamber.
* Actively involved in managing plant growth, plant care, phenotyping, phenotype alteration using auxin inhibition assays, test of apical dominance using radioactive H3 isotope.
* Actively involved in genomic DNA extraction, RNA extraction, next generation sequence reads library preparation.
* Generated hypothesis and predictions about genomics and RNAseq databases to associate gene to phenotype (lifehistory traits) in [***Arabidopsis lyrata***](https://plants.usda.gov/core/profile?symbol=arly2).
* Developed statistical methods for test of hypothesis and interpretation of the results.
* Analyzed +500 GBs of next generation sequence data (genome resequence and RNAseq reads) using softwares like BWA, rnaSTAR, phaser, GATK, Picard, etc.
* Regularly prepared data analyses jobs and submitted it to the HPC cluster (Henry on NC state university) using ssh protocol.
* Created databases of aligned BAM files, variant call files (VCF), quality metrics, statistics of the observed variant from the aligned sequence reads.
* Developed algorithm involving first order markov chains for classifying haplotype blocks for haplotype phase extension.
* Developed ASE (allele specific expression) pipeline [***aseCADG***](https://github.com/everestial/ASE-CADG) [for the test of ASE (allele specific expression) in F1 hybrids.](https://github.com/everestial/ASE-CADG)
* Developed full length python based haplotype phasing application – [***phase-Stitcher***](https://github.com/everestial/pHASE-Stitcher)and[***phase-Extender***](https://github.com/everestial/phase-Extender)and established haplotype phasing pipeline [***phaseIT***](https://github.com/everestial/PhaseIT/tree/master).
* Created databases of genome wide phased haplotype for all the samples using the applications [***phase-Stitcher***](https://github.com/everestial/pHASE-Stitcher)and[***phase-Extender***](https://github.com/everestial/phase-Extender).
* Identify problem trends pro-actively.
* Maintain information confidentiality and security.
* Analyze, troubleshoot and report any issues on the file/application server systems.
* Analyze the output files and verify that it confirms technical requirements.
* Involved in further curation of the output database, data management, query, and statistical analyses of the refined dataset to test the hypothesis.
* Developed statistical models for count data analyses to resolve ASE (allele specific expression) patterns in RNAseq data from F1 hybrids and established the pipeline [***aseCADG***](https://github.com/everestial/ASE-CADG) .

**Undergraduate Researcher, Institute of Forestry, Pokhara, Nepal**

* Wildlife habitat mapping and ecological modeling using ArcGIS, EdrasImagine.
* Collected database on species distribution using field survey, catch and release technique.
* Generated hypothesis and tested data using statistical tools SPSS, R.

**Thesis: “**Habitat suitability Mapping and Species Identification of Chiroptera: A case Study from Kaski District, Nepal.” Thesis submitted to Tribhuvan University, Institute of Forestry, Pokhara Campus, Pokhara.

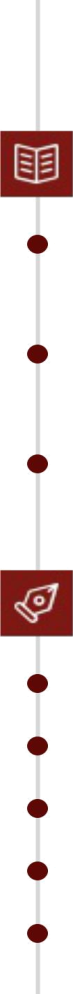
 **Education**

**PhD (Environmental Health Science with focus on Bio-Informatics), 2011- 2019**

University of North Carolina at Greenboro, NC

**B.Sc Forestry, 2005 – 2009**

Institute of Forestry, Pokhara, Nepal

**Projects/ Publications (on track)**

* Haplotype phasing in heterogenous genome and F1 hybrids using [***phase-Extender***](https://github.com/everestial/phase-Extender) ***and*** [***phase-Stitcher***](https://github.com/everestial/pHASE-Stitcher) ***.***
* Analyses of allele specific expression in *A. lyrata* F1 hybrids using [***aseCADG***](https://github.com/everestial/ASE-CADG) - pipeline for competitive alignment of sequence reads on a diploid genome.
* Auxin inhibition affects lateral shoot development and alters life history traits in *A. lyrata.*

**Personal Sites:**

**GitHub :** [*https://github.com/everestial*](https://github.com/everestial)

**LinkedIn :** [*https://www.linkedin.com/in/bishwa-kiran-84628728/*](https://www.linkedin.com/in/bishwa-kiran-84628728/)

**Wordpress:** [*https://wordpress.com/post/everestialblog.wordpress.com/21*](https://wordpress.com/post/everestialblog.wordpress.com/21)

**StackOverflow:**

[*https://stackoverflow.com/users/story/6346698*](https://stackoverflow.com/users/story/6346698)