Hector F. Espitia-Navarro

Bioinformatics PhD Candidate

48 Finch Trl NE, Atlanta, GA 30308, USA

 (+1) 404-519-5452 | hspitia@gatech.edu | hspitia | hspitia

Education & Training

Georgia Institute of Technology   *Atlanta, USA*
PhD in Bioinformatics | GPA: 3.8/4.0   *2015 – Present*

Universidad del Valle   *Cali, Colombia*
MSc in Engineering | GPA: 4.5/5.0   *2012 – 2015*
Computer Science and Software Engineering emphasis | Minor: bioinformatics

Universidad del Valle   *Cali, Colombia*
BSc in Computer Science | GPA: 4.4/5.0   *2005 – 2010*

Courses and Workshops

Cold Spring Harbor Laboratory   *New York, USA*
Advanced Sequencing Technologies & Applications   *2013*

Skills

Computational	*Proficient*: R, Bash, C++, Python, Perl, awk	*Familiar*: HTML, CSS, PHP, Java, LaTeX	*Learning*: Julia Software development, Linux environments, Git version control, machine learning, MySQL, PostgreSQL, Shiny applications, Jupyter notebooks
Bioinformatics	Bacterial sequence typing, WGS, metagenomics, RNA-seq, variant calling (including RAD-seq and GBS), pipeline development, SAMtools/BEDtools/VCFtools, FastQC
Languages	English: full professional proficiency Spanish: native language

Experience

Research

Georgia Institute of Technology | School of Biological Sciences | Jordan Lab   *Atlanta, USA*
Graduate Research Assistant   *2015 - Present*
Advisor: Dr. I. King Jordan

Efficient Alignment-free Software Applications for NGS-based Molecular Epidemiology

- Developed an alignment-free software based on *k*-mer frequencies for sequence typing directly from raw sequencing reads.
- Extended the *k*-mer approach to develop an alignment-free application for gene detection from raw sequencing reads.
- Developed a Web-accessible interface for sequence typing and phenotypic characterization of WGS samples. In colaboration with the Applied Bioinformatics Laboratory Inc., IHRC Inc., and Weems Design Studio.
- Developing an alignment-free software for *nifH* gene-based taxonomy assignment of amplicon sequencing samples.

Languages used: C++, R, Bash, Perl, Python, awk

Colombian Sugarcane Research Center (CENICAÑA) and Universidad del Valle *Cali, Colombia*
GRADUATE RESEARCH ASSISTANT *2012 - 2015*
Advisor: Dr. John Riascos; Co-advisor: Dr. Irene Tischer

General Bioinformatics Reasearch

- Established a bioinformatics platform for the sugarcane applied research.
- Developed pipelines for sugarcane RNA-seq transcriptome assembly, and identification of sugarcane variants related to agronomic traits of interest from RAD-seq and GBS data.

Development of a Data Mining Model for Identifying Sugarcane Candidate Genes Related to Drought and Flooding Tolerance

- Developed a framework for identifying sugarcane cDNAs related to water-stress tolerance, using RNA-seq expression analysis, co-expression network analysis, and machine learning.
- Deployed and maintained a Galaxy server instance for custom RNA-seq analysis pipelines.

Languages used: R, Bash, Perl, Python, awk, HTML, CSS, Javascript

Universidad del Valle | Bioinformatics and Biocomputing Research Group *Cali, Colombia*
UNDERGRADUATE RESEARCH ASSISTANT *2009 - 2010*
Advisor: Dr. Pedro Moreno; Co-advisor: Dr. Irene Tischer

Development of an Application for Multifractal Analysis on DNA and Protein Sequences using the Sand-box Method

- Developed a software for multifractal analysis on complete genomes.

Languages/tools used: C++, Python, UML

SOFTWARE DEVELOPMENT

Universidad del Valle | Informatics and Telecommunications Office *Cali, Colombia*
JUNIOR SOFTWARE DEVELOPMENT ENGINEER *2011 - 2012*

- Analyzed, designed and implemented software for an integrated Web-based management system for the Universidad del Valle.

Languages/tools used: UML, PHP (Symfony framework), HTML

TEACHING

Universidad del Valle *Palmira, Colombia*
ADJUNCT PROFESSOR *2010 - 2011*

- Teaching and administrating classes in engineering undergraduate programs.

Courses taught github: Algorithms and Programming, Systems General Theory, and Introduction to Computer Technology

Graduate Coursework

- Programming for Bioinformatics
- Computational Genomics
- Genomics and Applied Bioinformatics
- Environmental Microbial Genomics

- Human Genetics
- Applied Human Computational Genomics
- Simulation of Biology
- Machine Learning in Biomedicine

Honors & Awards

2015 **Overseas Doctorate Scholarship**. *Department of Science, Technology and Innovation (Colciencias), Government of Colombia.* *Bogotá, Colombia*

2010 **Honored BSc thesis**. *Universidad del Valle.* *Cali, Colombia*

Publications

Espitia-Navarro, **H. F.**, Rishishwar, L., Mayer, L. W., & Jordan, I. K. (2019). Rapid bacterial typing in the post-genomic era. In B. Budowle, S. Schutzer, & S. Morse (Eds.), *Microbial forensics* (3rd edition). Academic Press.

Espitia Navarro, **H. F.** (2015). *Desarrollo de un modelo de minería de datos para la identificación de genes candidatos asociados con tolerancia a estrés hídrico en caña de azúcar (Spanish) [Development of a data mining model for the identification of candidate genes associated with water stress in sugarcane]* (MSc Thesis, Universidad del Valle, Cali, Colombia).

Riascos-Arcos, J. J., **Espitia Navarro**, **H. F.**, & López Gerena, J. (2015). Evaluación de las herramientas de secuenciación masiva (NGS) para identificar genes asociados con tolerancia al estrés hídrico en caña de azúcar (Spanish) [Assessment of the new Next Generation Sequencing (NGS) tools to identify genes associated with tolerance to water deficit in sugarcane]. *Acta Agronómica*, *64*(4), 355–362. doi:10.15446/acag.v64n4.47772

Manuscrips in Preparation

Espitia-Navarro, **H. F.**, Chande, A. T., Nagar, S. D., Smith, H., Jordan, I. K., & Rishishwar, L. (2019). *Ultrafast sequence typing and gene detection with k-mers*.

Pattent Appplications

Espitia, **H.**, Chande, A. T., Jordan, I. K., & Rishishwar, L. (2017). *A method of sequence typing with in silico aptamers from a next generation sequencing platform*. Google Patents. Patent application US15/726,00. Retrieved from https://patents.google.com/patent/US20190108308A1

Posters & Presentations

Espitia-Navarro, **H. F.**, Chande, A. T., Nagar, S. D., Smith, H., Jordan, I. K., & Rishishwar, L. (2018). Ultrafast sequence typing with *in silico* DNA aptamers. Poster presented at the *ASM Microbe 2018*, June 7–11, Atlanta, GA, USA.

Espitia-Navarro, **H. F.**, Chande, A. T., Nagar, S. D., Smith, H., Jordan, I. K., & Rishishwar, L. (2017). Ultrafast sequence typing with *in silico* DNA aptamers. Oral presentation at the *IV Colombian Congress of Bioinformatics and VIII Iberoamerican Conference of Computational Biology*, September 13–15, Cali, Colombia.

Espitia-Navarro, **H. F.**, Chande, A. T., Smith, H., Jordan, I. K., & Rishishwar, L. (2017). Ultrafast sequence typing with *in silico* DNA aptamers. Poster presented at the *11th International Bioinformatics Conference*, November 3–4, Atlanta, GA, USA.

Espitia, **H. F.**, Tischer, I., & Riascos, J. J. (2015). Transcriptomic Studies to Identify Genes Related to Drought and Flooding Tolerance in Sugarcane. Oral presentation at the *11th Germplasm & Breeding, 8th Molecular Biology Workshop*, June 1–5, Saint-Gilles, Réunion Island, France.

Franco, C. M., **Espitia**, **H. F.**, Izquierdo, P., López, J., Salazar, F. A., & Riascos, J. J. (2015). Comparison of SSR and SNP Genetic Diversity Analyses in a Population of 220 Sugarcane Genotypes. Poster presented at the *11th Germplasm & Breeding, 8th Molecular Biology Workshop*, June 1–5, Saint-Gilles, Réunion Island, France.

Riascos, J. J., **Espitia**, **H. F.**, & Tischer, I. (2015). An Integrative Approach to Select Candidate Genes Derived from Sugarcane Transcriptomic Data. Poster presented at the *Plant & Animal Genome XXIII*, January 10–14, San Diego, CA, USA.

Riascos, J. J., **Espitia**, **H. F.**, & López, J. (2013). Identification and Characterization of Transcripts Associated with Tolerance to Drought and Flooding Stress in Sugarcane. Poster presented at the *Plant & Animal Genome XXI*, January 11–15, San Diego, CA, USA.