Georgia Tech

Bio Latino America

Biotechnology for public health and economic development in the Americas

Georgia Institute of Technology

MESSAGE FROM THE VICE PROVOST

Georgia Tech, consistently ranked among the very best technological universities in the world, is uniquely positioned to lead technological changes that are fundamental to the advancement of the human condition across the globe. Over the past two decades, Georgia Tech has grown into one of the most globalized public technological universities, with partnerships in more than 30 countries. Global engagement strategy remains a top priority for Georgia Tech, and it includes a renewed commitment to strengthen our links with Latin America. A unique opportunity exists to explore new partnerships between Georgia Tech and Latin American institutions, particularly in the space of the biosciences and biotechnology, for a variety of reasons:

- Latin American students, who have traditionally been a part of Tech's student body, continue to apply in increasing numbers, bringing their energy, perspectives and valued leadership skills to the Tech community.
- The alumni of Latin America, a strong, diverse and influential constituency, have been and continue to serve as a resource and advisory source to their country and their alma mater.
- Latin American governments increasingly recognize the need to develop knowledge-based economies that rely on research and education to stimulate innovation and growth.
- The region is home to numerous outstanding universities that increasingly emphasize research and development in addition to traditional educational activities.
- The marriage of biological and information sciences has ushered in a new age where virtually every aspect the human condition will be touched by biotechnology.

The 21st century has been hailed as the 'Age of Biology'. As was the case with physics in the early 20th century, our current understanding of biological systems is increasing rapidly along with our ability to engineer them. These advances are driven in large part by developments in information technology, and accordingly information-driven biotechnology will play an increasingly important role in shaping the human condition – in Latin America and beyond.

The Georgia Tech **Bio Latino America Initiative** is a faculty led effort to apply innovations in biotechnology to challenges in public health and economic development across the Americas. This effort leverages Georgia Tech's existing strengths in the applied biosciences – genomics and bioinformatics in particular – to address complex, regional-specific problems through collaborative research, development and education.

The Georgia Tech administration is very enthusiastic about this line of work and impressed with the progress that has been made thus far. But we are also eager to grow this program via deeper engagement with alumni, institutional and government partners in the region. We have collected this brief portfolio of the collaborative biotechnology work that we are currently engaged in with Latin American partners in order to showcase our capabilities in this area, and with the hopes of stimulating the formation of new partnerships in the region.

Yves Berthelot
Vice Provost for International Initiatives
President, GT Lorraine
Georgia Institute of Technology

THE BIO LATINO AMERICA INITIATIVE

OUR VISION

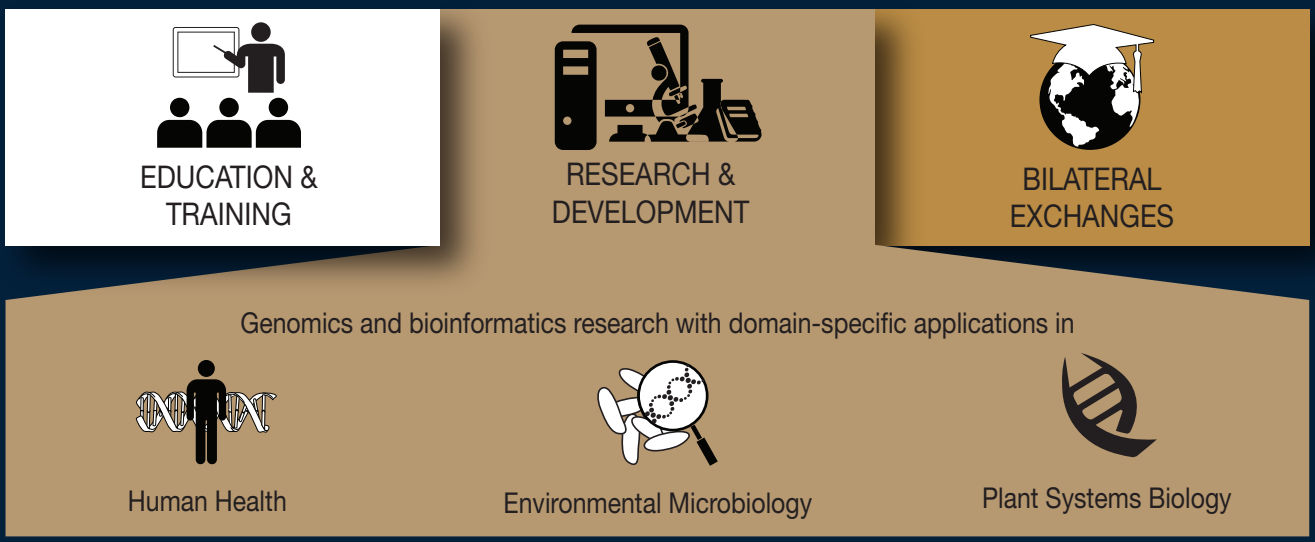
Georgia Tech will define the technological research university of the 21st century. To this end, we aim to expand Georgia Tech's global influence with a special emphasis on Latin America.

OUR MISSION

Georgia Tech scientists solve complex problems through innovative research, development & education. The Bio Latino America initiative applies biotechnology to address challenges in public health & economic development throughout the Americas.

COLLABORATIVE ACTIVITIES

The Bio Latino America mission is executed via collaborative efforts in three core areas: Education & Training, Research & Development, and Bilateral Exchanges



OUR PARTNERS IN LATIN AMERICA

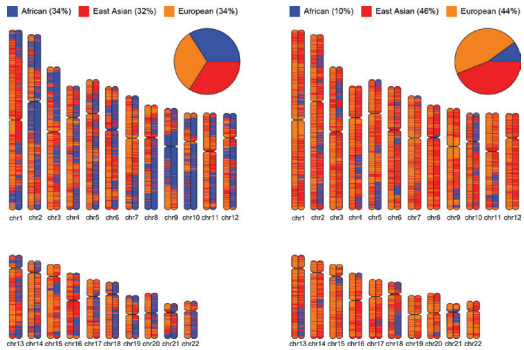
UNIVERSIDAD NACIONAL DE COLOMBIA
Universidad del Valle
BIOS
INDICASAT AIP
COLCIENCIAS COLOMBIA
Universidad Tecnológica del Chocó "Diego Luis Córdoba"
Universidad de los Andes
SENACYT
PABI
INCAUCA S.A.
REGENERAR
Corpoica
CENTRO DE INGENIERÍA GÉNETICA Y BIOTECNOLOGÍA

RESEARCH & DEVELOPMENT

Georgia Tech faculty and graduate students engage in collaborative research efforts with partner institutions in Latin America. Our collaborative research combines genomic and bioinformatic approaches with domain-specific applications to human health, environmental microbiology and plant systems biology.

Ancestry, Admixture and Selection in Latino Genomes

The genome sequences of Latin Americans are characterized by three-way admixture patterns with ancestral contributions from Africa, the Americas and Europe. These genome sequences can be considered to be novel in terms of human evolution since they contain allelic combinations that have never previously existed in the same genetic background. This project explores the implications of the creation of such novel admixed Latin American genomes for natural selection and human health. The working hypothesis is based on the axiom that specific allelic variants evolved separately in ancestral human populations based on their regional-specific utility, i.e. their relationship to health and fitness in a particular environment. Then these pre-evolved ancestral population-specific alleles may have been selected in modern admixed Latino populations based on their utility in the new environment. In order to test this hypothesis, researchers from Georgia Tech are collaborating with faculty and students from the School of Medicine at Universidad Libre in Cali, Colombia as well as computational biologists from BIOS in Manizales to analyze complete genome sequences of 94 Colombian individuals.



A Population Genomic Health Profile for Colombia

Researchers from Georgia Tech are collaborating with scientists from Universidad Libre and Regenerar in Cali, Colombia to survey Colombian genome sequence diversity. A collection of complete genome sequences of 94 individuals from the city of Medellin is being characterized to this end. The allelic diversity of this sample will be measured and compared to world-wide levels of allelic variation in order to identify alleles that are enriched (or depleted) in the Colombian population. These Colombian-characteristic alleles will be evaluated with respect to their relationship to human health and disease in order to create a genomic health profile for the country. This work will result in a database of Colombian genome sequence diversity, associated with its phenotypic consequences, which will be hosted and distributed by BIOS – the Colombian National Center for Bioinformatics and Computational Biology. The population genomic profile derived from this research effort will serve as a baseline for future studies of genomic and personalized medicine in Colombia.

ChocoGen: Patrimonio Genético del Pacífico Colombiano

ChocoGen is a collaborative research project aimed at the discovery and characterization of the genetic heritage of the people of Chocó. The population of Chocó, a state located on the pacific coast of Colombia, has a uniquely African genetic heritage with admixture from Europe and the Americas. Researchers from the ChocoGen project are using analysis of genomic sequences sampled from volunteers from the population of Chocó to characterize their genetic ancestry, the quantity and nature of admixture between ancestral populations and the possible relationship between ancestry, admixture and determinants of health and disease. The ChocoGen project is a partnership among the Universidad Tecnológica del Chocó and BIOS in Colombia, Georgia Tech in the USA and the PanAmerican Bioinformatics Institute.

Clinical Metagenomics for Tropical Diseases

Infectious tropical diseases caused by microbial pathogens represent a sustained threat to public health in Latin America. Clinical metagenomic approaches, i.e. culture-free sequencing of patient samples, hold great promise for the diagnosis and treatment of such diseases owing to the fact that they are unbiased and highly sensitive. The main goal of this project is to establish an accurate diagnostic method for the identification and characterization of pathogenic agents that cause febrile illnesses in pediatric patients treated at the Clinica Universitaria Rafael Uribe Uribe (CURUU) in Cali, Colombia. To do this, we are developing metagenomic strategies that utilize deep sequencing of sterile sites from infected patients coupled with novel bioinformatic analysis methods to characterize the sequences of infectious agents that are causing disease. This project is a collaboration between researchers from Georgia Tech and Universidad Libre.

RESEARCH & DEVELOPMENT

Discovery of Native Biofertilizers for Sugar Cane

Biofertilizers are microorganisms that promote plant growth by increasing the supply or availability of primary nutrients. Nitrogen-fixers are biofertilizers that function by converting (unusable) molecular Nitrogen into a form that can be readily assimilated by plants. The goal of this project is to develop regional-specific cultures of Nitrogen-fixing bacteria that will allow for an increased yield of sugarcane and reduced use of urea based fertilizer, thus providing for more economical and sustainable cropping practices. To do this, we are using a combination of metagenomic and selective media-based techniques to characterize the native microbiome associated with sugarcane cultivated in the Cauca region of Colombia. Once identified in this way, these native microbes will be cultured and developed as biofertilizers for locally grown sugarcane crops. This project represents a public-private collaboration between Georgia Tech, Regenerar in Cali and Incauca, the largest sugar cane company in Colombia.

Metagenomic Approaches to Functional Biodiversity

Georgia Tech researchers are working with international collaborators from INDICASAT to conduct metagenomics, reverse genomics and targeted culture-based surveys of Panamanian marine microbial biosystems in an effort to identify enzymatic activities that drive the production of bioactive metabolites. Panamá is a country with an abundance of coral reefs that contain a rich and diverse coral fauna, and coral-associated microbes encode a wide variety of enzymatic activities involved in the production of bioactive metabolites. The INDICASAT team and collaborators have shown that Panamanian endemic coral-associated bacteria are rich sources of antimicrobial and anti-parasitic compounds. Building on this prior work, a combination of community metagenomics and reverse genomics approaches is being employed to 1) characterize the diversity of coral-associated bacteria from Panamá, and 2) evaluate the genomic determinants of bioactive secondary metabolite biosynthesis in these organisms.

```
    } else {
        print color 'bold red';
        print "unrecognized answer.  Exiting the program.\n";
        print color 'reset';
        exit;
    }
}

$mmx .= "g";

##############################################
($sec,$min,$hour,undef) = localtime();

# Perform Quality Control using PRINSEQ
print color 'reset';
print color 'bold blue';
printf ("\n\n[%02d:%02d:%02d] Started QC\n\n", $hour, $min, $sec);
# QC for single reads file
if($sep == 0){
    print "prinseq-lite.pl -fastq $r -min_len 50 -trim_qual_left 24 -trim_qual_right 24 -trim_qual_window 5 -trim_qual_step 1 -min_qual_mean 28 -out_good $out.good.$$ -o
    print color 'reset';
    `prinseq-lite.pl -fastq $r -min_len 50 -trim_qual_left 24 -trim_qual_right 24 -trim_qual_window 5 -trim_qual_step 1 -min_qual_mean 28 -out_good $out.good.$$ -out_bad
    $r = "$out.good.$$.fastq";
} else {
# QC for paired end reads files
    print "prinseq-lite.pl -fastq $r -fastq2 $r2 -derep 14 -min_len 50 -trim_qual_left 24 -trim_qual_right 24 -trim_qual_window 5 -trim_qual_step 1 -min_qual_mean 28 -ou
    print color 'reset';
    `prinseq-lite.pl -fastq $r -fastq2 $r2 -derep 14 -min_len 50 -trim_qual_left 24 -trim_qual_right 24 -trim_qual_window 5 -trim_qual_step 1 -min_qual_mean 28 -good
    $r = "$out.good.".$$."_1.fastq";
    $r2 = "$out.good.".$$."_2.fastq";
}
print color 'bold blue';
($sec,$min,$hour,undef) = localtime();
printf ("\n\n[%02d:%02d:%02d] Finished QC\n\n", $hour, $min, $sec);

# Check if the reference fla indices exist for bwa, index if they do not exist
if(!(-e "$ref.amb" && -e "$ref.ann" && -e "$ref.bwt" && -e "$ref.pac" && -e "$ref.sa")){
    ($sec,$min,$hour,undef) = localtime();
    printf ("\n\n[%02d:%02d:%02d] Indexing the reference genome (will take some time)\n\n", $hour, $min, $sec);
    print "bwa index $ref\n\n";
    print color 'reset';
    `bwa index $ref`;
    print color 'bold blue';
    ($sec,$min,$hour,undef) = localtime();
    printf ("\n\n[%02d:%02d:%02d] Finished indexing the reference genome\n\n", $hour, $min, $sec);
}
```

Bioinformatics Algorithms and Methods

Researchers from Georgia Tech are partnering with international collaborators from BIOS in Colombia and INDICASAT in Panamá to develop a suite of algorithms and methods for the analysis of microbial genomic, transcriptomic and proteomic datasets. All of the bioinformatic algorithms and methods developed are tailored to collaborative experimental studies being conducted at field research sites in Colombia and Panamá. Work is proceeding along two lines: 1) analytical pipeline development whereby existing methods will be assembled using best practices into integrated bioinformatics workflows that can be readily accessed by working researchers, and 2) de novo code development for the creation of novel bioinformatics methods from first principles. The entire collection of tools will deployed at BIOS, the Colombian National Center for Bioinformatics and Computational Biology, which will serve as a single point of entry for investigators working on the analysis of microbial diversity in Latin America.

High Performance and Cloud Computing (HPCC) for Bioinformatics

Faculty and students from the Georgia Tech Bioinformatics graduate program are teaming with the Colombian National Center of Bioinformatics and Computational Biology (BIOS) to create a distributed high-performance bioinformatics analysis platform for microbial genomic and metagenomic datasets. HPCC resources are currently not widely accessible at partner institutions in Latin America. Thus, the creation of such resources, along with the educational and training activities that will be deployed to facilitate their use, is designed to extend participation in bioinformatics and microbial genomics research to laboratories across Latin America. BIOS will serve as a single port of entry for international access to the HPCC analytical platform to be developed. Two classes of methods are being developed for the utilization of HPCC resources. First, bioinformatic and genome analysis code will be adopted and optimized to distributed and/or multi-core computational environments to ensure optimal usage of HPCC resources. Second, a cloud computing portal, built on wrapper code used to create an intuitive user interface, is being developed in order to create an HPCC platform that can be readily accessed by microbial diversity researchers across the Americas.

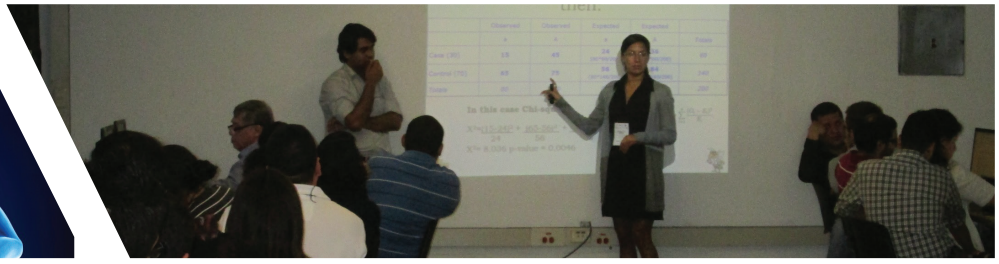
EDUCATION & TRAINING

Georgia Tech faculty and graduate students offer short courses, workshops and seminars, covering various themes in the biosciences, throughout Latin America.

II Grand Session of Epidemiogy, Colombia 2014

ICGEB: Advanced Techniques for Clinical & Personalized Human Genomics, Colombia 2014

XLVII National Congress of Biological Sciences, Colombia 2012



II International Bioinformatics Course, Colombia 2012

I International Basic Bioinformatics Course, Colombia 2011

BILATERAL EXCHANGES

Georgia Tech faculty laboratories host international visitors who receive state-of-the-art training in numerous bioscience research areas, such as biotechnology, bioinformatics, clinical genomics, metagenomics, microbiology and systems biology. Georgia Tech students and faculty also participate in on site research, development and capacity building at partner institutions in Latin America.

Host Lab: Prof. King Jordan, School of Biology, Georgia Tech

Yina Alejandra Cifuentes Triana,
MS student
Universidad Nacional de Colombia
Bogotá, Colombia

Juan Esteban Gallo,
PhD student
Universidad del Rosario
Bogotá, Colombia

Host Lab: Kostas Konstantinidis, School of Civil & Environmental Engineering, Georgia Tech

Henry Fabian Tobar Tossee,
PhD student
Universidad del Valle
Cali, Colombia

Luis Miguel Rodriguez,
MS student
Universidad de Los Andes
Bogotá, Colombia

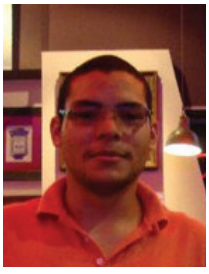
Host Lab: Prof. Joel Kostka, School of Biology, Georgia Tech

Lina C. Valderrama-Aguirre,
MSc
Universidad del Valle
Cali, Colombia

Luz Karime Medina,
MSc
Pontificia Universidad Javeriana
Bogotá, Colombia

Host Lab: Prof. Fredrik Vannberg, School of Biology, Georgia Tech

Ninfa Andrea Fernandez Joaqui,
MS student
Universidad Libre
Cali, Colombia



Jock Randolph Chichaco,
MD, PhD student
INDICASAT-AIP
Clayton, Panamá

GEORGIA TECH BIO LATINO AMERICA LEADERSHIP

Yves Berthelot
Vice Provost for International Initiatives
President, GT Lorraine
yves.berthelot@provost.gatech.edu

Marta Garcia
Associate Vice President,
International Development
marta.garcia@gatech.edu

Sara Araujo
Managing Director of Development,
Latin America
saraisa39@yahoo.com

King Jordan
Associate Professor, School of Biology
Faculty Director, Bio Latino America
king.jordan@biology.gatech.edu

GEORGIA TECH PARTICIPATING LABORATORIES AND UNITS

Mark Borodovsky
Bioinformatics Lab
Computational Sciences & Engineering
http://exon.gatech.edu/

King Jordan
Bioinformatics & Genomics Lab
School of Biology
http://jordan.biology.gatech.edu/

Joel Kostka
Environmental Microbiology Lab
School of Biology
http://www.joelkostka.net/

Greg Gibson
Center for Integrative Genomics
School of Biology
http://www.cig.gatech.edu/

Kostas Konstantinidis
Metagenomics Lab
School of Civil & Environmental Engineering
http://enve-omics.gatech.edu/

Fredrik Vannberg
Clinical Genomics Lab
School of Biology
http://vannberg.biology.gatech.edu/

PanAmerican Bioinformatics Institute
http://panambioinfo.org/

UniValle-GeorgiaTech Genome Research Center
http://univalle.gatech.edu/

ChocoGen
http://www.chocogen.com/