



Programas de Internacionalización Study Abroad 2021

Course title: Narrowing the interrogation of DNA. Target Enrichment in Plant Phylogenomics

Teaching period: July 5 to July 9, 2021

Teaching hours: 22,5 hours

Academic coordinators: Esteban Salmerón Sánchez y Fabián Martínez Hernández

Knowledge area: Botany

1. INTRODUCTION

Target enrichment is a cost-effective and efficient method for researchers to capture specific regions of interest. Reducing the amount of DNA being interrogated allows researchers to perform experiments with more statistically relevant numbers of samples. Investigators can reliably sequence large numbers of genes (e.g. over 100 genes). It also provides greater sequencing depth, which increases the ability of Next-Generation Sequencing (NGS) to detect variants from limited sample amounts, identifying, for instance, disease-associated mutations or single nucleotide variants (SNPs). Thus, hybridization-based enrichment is a useful strategy for analyzing specific genetic variants. Several strategies are employed for sequence capture, in this course we will discuss hybrid capture enrichment with genome skimming (also known as HybSeq). It uses probes, also known as baits, that hybridize specifically with targeted regions of interest. Pools of thousands of long custom biotinylated oligonucleotides can be used to selectively enrich for specific regions of interest in an NGS library prior to sequencing.

This course will put special focus on plant systems. High-throughput sequencing provides an avenue for bridging the micro- and macroevolutionary gap in phylogenetics by scaling up the number of loci and individuals within populations and across species that can be sequenced at a reasonable cost. It has gained popularity in recent years because, unlike whole-genome sequencing, this technique can work with both fresh and old museum material and generally demands less complex bioinformatics. Reduced representation techniques such as restriction site-associated DNA sequencing or genotype-by-sequencing work well for population-level analysis and do not require a reference genome. Yet, their outputs are SNPs or small reads from anonymous loci, which make assessment of gene orthology challenging in the case of deep divergences. HybSeq is promising for non-model organisms because it generates thousands of DNA sequences from low or single-copy nuclear genes, combining exon capture with genome skimming of intronic and intergenic regions, flanking the targeted exon regions. HybSeq also generates highly repetitive DNA from organellar genomes as a by-product; the latter is important to detect reticulate evolution and introgression.

2. OBJECTIVES

1. To provide an overview of the library preparation and target enrichment protocols required for next-generation sequencing (NGS).
2. To learn bioinformatics skills to analyze the raw data generated with this technique, with special focus on plant systems.



Programas de Internacionalización Study Abroad 2021

3. CONTENT	
Modules	Description
MODULE A <i>BEST DNA PRACTICES</i>	<ul style="list-style-type: none"> • Introduction to DNA extractions for NGS • DNA quantification and DNA fragmentation • Sampling strategies
MODULE B <i>LIBRARY PREPARATION</i>	<ul style="list-style-type: none"> • Library Preparation Protocols • Best practices and tips
MODULE C <i>HYB-SEQ</i>	<ul style="list-style-type: none"> • Hybridization protocol • Pooling samples • Sequencing preparation protocol
MODULE D <i>PROBE DESIGN</i>	<ul style="list-style-type: none"> • Probe design • Selection of genomic and transcriptomic resources
MODULE E <i>DATA ANALYSES</i>	<ul style="list-style-type: none"> • Processing raw data and quality check • Downstream analyses: from DNA matrices to SNPs discovery

4. METHODOLOGY
<p>In this course, we will go through different wet lab protocols to prepare samples for targeted sequencing. We will discuss best practices for species sampling, genomic DNA extractions and pooling strategies, with a special focus on plant systems (Modules A-C).</p> <p>We will also have training on probe design using genomic and transcriptomic data publicly available. Bioinformatic pipelines (e.g. MarkerMiner, MrBait) will be used to detect intron/exon boundaries, and to identify low or single-copy nuclear genes needed to develop the gene target probes. We will discuss probes design (Module D) and selection of target loci. Finally, bioinformatics pipelines will be used to check the quality of the data sequenced and to process raw data for further analyses (Module E). Students will learn to extract sequences from raw data, built DNA matrices and conduct phylogenetic analyses in an automated way using loops and bash scripts. Moreover, they will learn how to extract SNPs from HybSeq data and to investigate the possibilities of its output in population genetics. We will use open-source softwares and codes to explore and analyze different datasets provided by the instructor. Students could also bring and use their own datasets. All these pipelines can be run in either Unix or Windows environments.</p>

5. PROFESSIONAL VISITS AND COMPLEMENTARY ACADEMIC ACTIVITIES



Programas de Internacionalización Study Abroad 2021

We will invite Dr. Sonia Herrando-Moraria (Instituto Botánico de Barcelona –IBB/CISC) for a talk in Module E.

6. ASSESSMENT

Students will be evaluated upon their participation during the course (50%). They will also be requested to submit a small project based on HybSeq technique (50%).

7. LECTURERS

Tamara Villaverde, *PhD in Environmental Sciences.*

Tamara Villaverde is a postdoctoral researcher at the University of Almeria with an UAL-FEDER funded project (EUGEANDA) focused on the evaluation of the genetic diversity of endangered species and their invasive relatives using genomic data. Her main research interests are related with the evolution, systematics and biogeography of plants. She has a broad teaching experience in BSc and MSc degrees.

Sonia Herrando-Moraira, *PhD Candidate.*

Sonia Herrando-Moraira is coursing the last year of doctoral dissertation at the Botanical Institute of Barcelona (IBB-CSIC). She is working on plant biodiversity, mainly in the sunflower family, from a systematic and ecological perspectives. She has a special interest in bioinformatic analyses of high throughput molecular data and ecological niche modeling tools. As a teaching experience, she collaborated in the module “Bioinformatic tools in Phylogenetic and Phylogegnomic studies” within the master’s degree in Plant Biology, Genomics and Biotechnology at the Autonomous University of Barcelona.

Organized by:

Vice-Rectorate for Internationalization
UNIVERSIDAD DE ALMERÍA
Tel. +34 950 01 5816
E-mail: sabroad@ual.es