



## SGIker Prestakuntza Eskaintza (Ikerkuntzarako Zerbitzu Orokorrak)

---

Ikastaroaren izenburua:

---

**NGS datuen analisi bioinformatikorako sarrera, nobo genomak mihiztatzeko eta populazioen genetikan aplikatzeko.**

<b>Data</b>	2021eko azaroaren 29tik abenduaren 3ra.
<b>Orduetgia</b>	9: 00-13: 00
<b>Iraupena</b>	20 ordu
<b>Tokia</b>	0.13 Mintegia, María Goyri Parke Zientifikoa, Bizkaiko Campusa

Hizlariak eta irakasleak:

---

**Luis Javier Chueca doktorea.**

Ikastaroaren helburuak:

---

1. 2. eta 3. belaunaldiko sekuentziak maneiatzea.
2. Ereduzkoak ez diren organismoen genomak mihiztatzea eta idaztea.
3. Bereizketa genetikoaren parametroak zenbatzea.
4. Populazio-egitura identifikatzea eta historia demografikoak eta ebolutiboak ondorioztatzea.



## Ikastaroaren edukiak:

---

Teoriko-praktikoa.

### UPV/EHUko Genomika Zerbitzuko (IKS) sekuentziazio masiboko azpiegituren eta ikastaroaren aurkezpena.

1.eguna:

Ikastaroaren aurkezpena (IKS)Sarrera. Genomak sekuentziatzeko partzuergo nagusiak (Earth Biogenome Project, Darwing Tree of Life, ERGA, etab.). Organismo ez-ereduen erreferentzia-genomak, azterketa-oinarri gisa. Unix sistema eragilearen oinarritzko sarrera (GNU/Linux). Galaxy Project-erako sarrera: web plataforma ireki bat ikerketa konputazional irisgarri, erreproduzigarri eta gardenerako. Hirugarren belaunaldiko sekuentziak genomak mihiztatzeko eta idazteko. Sekuentzia motak: PacBio, Oxford Nanopore Technologies (ONT). Kromosoma-mailako mihiztadura (Hi-C eta Omni-C datuak).

2.eguna:

Genomak mihiztatzea. Teoria eta kasu praktikoak (Flye, wtdbg2, HiFiasm). Mihiztatzeko aldamiok (SPPACE-Ir scaffolder, TBG-gapcloser). Errore-tasa baxuko illumina-irakurketen bidez mihiztatzea. Mihiztatzearen kalitatea ebaluatzea (QUAST eta BUSCO). Mihiztatzean egon daitekeen kutsadura identifikatzea (Blobtools).

3.eguna:

Transkriptuak Illumina (Trinity) datuetatik abiatuta mihiztatzea. DNA mitokondrial eta kloroplastikoa mihiztatzea eta idaztea (MitoZ, Mitos, Novoplasty). Genomen oharpena (Maker, Braker, GeMoMa, EVIDENCEModeler). Idatzoharren eskuzko zuzenketa. Apolo webgunerako sarrera: online editore kolaboratiboa eta berehalakoa ohar genomikoak editatzeko.

4.eguna:

Populazioen genetikarako datuak. DdRADseq, pool-seq eta whole-genome re-sequencing arteko desberdintasun nagusiak. Sekuentziak mapatzea eta lerrotzea erreferentziazko genomen aurrean. Indels eta Single Nucleotide Polymorphism (SNPs) (Samtools, GATK, FreeBayes) identifikatzea. VCF fitxategiak miatzea. SNP motak. Dibertsitate genetikoen estimazioa (Tajima's D, FST, DXY, \_and\_t = ut\_and\_gt; fA).

5.eguna:

Biztanleria-egituraren azterketa (APA eta ADMIXTURE). Proteinen gene kodifikatzaileetan hautespen positiboa identifikatzea MacDonald-Kreitman test bidez. Populazioen eta espezieen historia demografikoa. Dibergentzia ebolutiboko agertokien simulazioa « Site Frequency Spectrum » (SFS) eta « Approximate Bayesian Computation » delakoetatik abiatuta.

**\* Ikastaroaren edukiak eta/edo ordenak aldaketak izan ditzakete ikastaroaren garapenean.**



## Parte hartzailearen profila:

---

Postgraduatu eta ikertzaileak.

## Parte hartzaile kopurua (gutxienekoa/gehienezkoa):

---

4/20

### Kontaktua

- Irati Miguel doktorea eta Fernando Rendo doktorea.
- Genomikako Zerbitzua- Sekuentziazio eta Genotipatze Unitatea.
- Bioteknologiako Zentroa - Maria Goyri eraikina. UPV/EHUko parke zientifikoa. Bizkaiko campusa. Sarriena auzoa, z/g 48940 Leioa (Bizkaia).
- Tel.: 946012934/946018301.
- [irati.miguel@ehu.eus](mailto:irati.miguel@ehu.eus)  
[fernando.rendo@ehu.eus](mailto:fernando.rendo@ehu.eus)

### Prezioa

- UPV/EHUkoek: 200 €
- IEPkoek: 350 €
- Kanpokoek: 500 €

## Informazio gehigarria:

---

- Ikastaro teoriko-praktikoa izango da.
- Ikastaroa gaztelaniaz ematen da.
- Laborategiko segurtasun-arauak bete beharko dituzte.
- Ikastaroan bertaratze-ziurtagiria emango da. Ikastaroan parte hartu dutenek ziurtagiri hori jasoko dute, baldin eta ikastaroaren ordu kopuru guztiaren % 80 bete badute.
- **Laguntzaile bakoitzak bere ordenagailua ekarri behar du.**
- Gomendatutako sistema eragileak:  
Ubuntu LTS 18.04.6 edo geroagokoa.  
MacOS Katalina v10.15.7 edo geroagokoa.  
Windows 10.
- Komandoen lerroa eta programazioa ezagutzea gomendatzen da.



- Anima gaitetzen parte-hartzaileek honako tutorial hauek bisita ditzaten:  
<https://ryantutorials.net/linuxtutorial/>  
<https://slurm.schedmd.com/documentation.html/>  
<https://stackabuse.com/python-tutorial-for-absolute-beginners/>  
<https://www.statmethods.net/r-tutorial/index.html/>
- **Luis Javier Chueca** doktoreak ikastaroa ematen du ([luisjavier.chueca@ehu.eus](mailto:luisjavier.chueca@ehu.eus)). Doktoratu ondoko ikertzailea Euskal Herriko Unibertsitatean (UPV/EHU) eta LOEWE-Center for Translational Biodiversity Genomics (LOEWE-TBG) zentroan. Bere ikerketa-ildoak mendebaldeko paleartikoko espezieen egungo banaketa eta erlazio filogenetikoak ulertzea da, baita nitxoaren, espezializazioaren eta tokiko egokitzapenaren bilakaeraren oinarri genetikoak identifikatzea ere. Taxonomia tradizionala sekuentziazio-teknologiekin eta abangoardiako bioinformatikekin konbinatzen du, fauna maila genomikoan dibertsifikatzeko prozesuak sakon ikertzeko.



## Oferta formativa SGIker (Servicios Generales de Investigación)

---

Título del curso:

---

**Introducción al análisis bioinformático de datos NGS para ensamblaje de genomas de novo y aplicaciones a la genética de poblaciones.**

<b>Fechas</b>	Del 29 de noviembre al 3 de diciembre de 2021.
<b>Horario</b>	9:00-13:00
<b>Duración</b>	20 horas
<b>Lugar</b>	Seminario 0.13, Parque Científico María Goyri, Campus de Bizkaia

Ponentes y formadores:

---

**Dr. Luis Javier Chueca**

Objetivos que se pretenden alcanzar en el curso:

---

1. Manejo de secuencias de 2<sup>a</sup> y 3<sup>a</sup> generación.
2. Ensamblaje y anotación de genomas de organismos no modelo.
3. Estimación de diferentes parámetros de diferenciación genética.
4. Identificación de estructura poblacional e inferencia de historias demográficas y evolutivas.

Contenidos que se van a trabajar durante el curso:

---

De carácter teórico-práctico.

Día 1:

### **Presentación del curso y de las infraestructuras de secuenciación masiva del Servicio de Genómica de la UPV/EHU (SGIker).**

Introducción. Principales consorcios para secuenciación de genomas (Earth Biogenome Project, Darwin Tree of Life, ERGA, etc.). Genomas de referencia de organismos no-modelos como base de estudio. Introducción básica al sistema operativo UNIX (GNU/Linux). Introducción a Galaxy Project: una plataforma web abierta para investigación computacional accesible, reproducible y transparente.



Secuencias de tercera generación para el ensamblado y anotación de genomas. Tipos de secuencias: PacBio, Oxford Nanopore Technologies (ONT). Ensamblaje a nivel de cromosoma (datos Hi-C y Omni-C).

Día 2:

Ensamblado de genomas. Teoría y casos prácticos (Flye, wtdbg2, HiFiasm). Andamiaje de ensamblado (SPPACE-Ir scaffolder, TBG-gapcloser). Pulido de ensamblado mediante lecturas de illumina con baja tasa de error. Evaluación de la calidad del ensamblado (QUAST y BUSCO). Identificación de posible contaminación en el ensamblaje (Blobtools).

Día 3:

Ensamblaje de transcriptomas a partir de datos Illumina (Trinity). Ensamblaje y anotación de ADN mitocondrial y cloroplástico (MitoZ, Mitos, Novoplasty). Anotación de genomas (Maker, Braker, GeMoMa, EVIDENCEModeler). Corrección manual de anotaciones. Introducción a web Apollo: un editor online colaborativo e instantáneo para edición de anotación genómicas.

Día 4:

Datos para genética de poblaciones. Principales diferencias entre ddRADseq, pool-seq y whole-genome re-sequencing. Mapeo y alineación de secuencias frente a genomas de referencia. Identificación de Indels y Single Nucleotide Polymorphism (SNPs) (Samtools, GATK, FreeBayes). Exploración de archivos VCF. Tipos de SNPs. Estimación de diversidad genética (Tajima's D, FST, DXY, entre otros.)

Día 5:

Análisis de estructura poblacional (PCA y ADMIXTURE). Identificación de selección positiva en genes codificantes a proteínas mediante el uso de MacDonald-Kreitman test. Historia demográfica de poblaciones y especies (PSMC). Simulación de escenarios de divergencia evolutiva a partir del "Site Frequency Spectrum (SFS)" y "Approximate Bayesian Computación".

***\*El contenido y/u orden del curso puede verse sujeto a cambios durante el desarrollo del mismo.***

Perfil del participante:

---

Personas postgraduadas y personal investigador.

---



## Número de participantes (mínimo/máximo):

---

4 / 20

### Datos de contacto

- Dra. Irati Miguel y Dr. Fernando Rendo.
- Servicio de Genómica-Unidad de Secuenciación y Genotipado.
- Centro de Biotecnología Animal - Edificio Maria Goyri. Parque científico de la UPV/EHU. Campus de Bizkaia. Bº Sarriena, s/n 48940 Leioa (Bizkaia).
- Tfno.: 946012934 / 946018301.
- [irati.miguel@ehu.eus](mailto:irati.miguel@ehu.eus)  
[fernando.rendo@ehu.eus](mailto:fernando.rendo@ehu.eus)

### Precio

- Usuarios de la UPV/EHU: 200 €
- Usuarios de Organismos Públicos de Investigación: 350€
- Usuarios externos: 500 €

## Otra información adicional:

---

- El curso se imparte en castellano y es teórico-práctico.
- Se seguirán en todo momento las medidas de seguridad necesarias para realizar las prácticas.
- Se entregará certificado de asistencia. Los y las asistentes al curso recibirán dicho certificado siempre y cuando hayan completado al menos el 80% de la duración total del mismo.
- **Se requiere que cada asistente traiga su ordenador.**  
Sistemas operativos recomendados:  
Ubuntu 18.04.6 LTS o posterior  
MacOS Catalina v10.15.7 o posterior  
Windows 10
- Se recomienda un conocimiento básico de la línea de comandos y programación.  
Animámos a que los participantes visiten los siguientes tutoriales:  
<https://ryantutorials.net/linuxtutorial/>  
<https://slurm.schedmd.com/documentation.html/>  
<https://stackabuse.com/python-tutorial-for-absolute-beginners/>  
<https://www.statmethods.net/r-tutorial/index.html/>



- El curso lo imparte **Dr. Luis Javier Chueca (luisjavier.chueca@ehu.eus)**. Investigador postdoctoral en la Universidad del País Vasco (UPV/EHU) y el LOEWE-Center for Translational Biodiversity Genomics (LOEWE-TBG). Su línea de investigación se centra en comprender la distribución actual y las relaciones filogenéticas de especies del paleártico occidental, así como identificar la base genética de la evolución de nicho, especiación y adaptación local. Combina la taxonomía tradicional con tecnologías de secuenciación y bioinformáticas de vanguardia para investigar a fondo los procesos de diversificación de fauna a nivel genómico.

## SGIker training courses offered (Advanced Research Facilities)

---





## Course title:

---

**Introduction to the bioinformatic analysis of NGS data for de novo genome assembly and applications to population genetics.**

<b>Dates</b>	From November 29 to December 3, 2021.
<b>Hour</b>	9: 00-13: 00
<b>Duration</b>	20 hours
<b>Course venue</b>	Seminar 0.13, María Goyri Science Park, Bizkaia Campus

## Speakers and trainers:

---

**Doctor Luis Javier Chueca**

## Objectives to be fulfilled during the course:

---

1. Handling of 2nd and 3rd generation sequences.
2. Assembly and annotation of genomes of non-model organisms.
3. Estimation of different parameters of genetic differentiation
4. Identification of population structure and inference of demographic and evolutionary histories

## Content that is going to be worked on during the course:

---

### Theoretical-practical course

Day 1:

**Presentation of the course** and the massive sequencing infrastructures of the **Genomics Service of the UPV / EHU (SGIker)**.

Introduction. Main consortia for genome sequencing (Earth Biogenome Project, Darwin Tree of Life, ERGA, etc.). Reference genomes of non-model organisms as a basis for study. Basic introduction to the UNIX operating system (GNU / Linux). Introduction to Galaxy Project: an open web platform for accessible, reproducible and transparent computational research.

Third generation sequences for genome assembly and annotation. Sequence types: PacBio, Oxford Nanopore Technologies (ONT). Chromosome level assembly (Hi-C and Omni-C data).

Day 2:



Assembly of genomes. Theory and practical cases (Flye, wtdbg2, HiFiasm). Assembling scaffolding (SPPACE-Ir scaffolder, TBG-gapcloser). Assembly polishing using illumina readings with low error rate. Assembly quality evaluation (QUAST and BUSCO). Identification of possible contamination in the assembly (Blobtools).

Day 3:

Assembly of transcriptomes from Illumina data (Trinity). Assembly and annotation of mitochondrial and chloroplastic DNA (MitoZ, Mitos, Novoplasty). Genome annotation (Maker, Braker, GeMoMa, EvidenceModeler). Manual correction of annotations. Introduction to web Apollo: a collaborative and instant online editor for genomic annotation editing.

Day 4:

Data for population genetics. Main differences between ddRADseq, pool-seq and whole-genome re-sequencing. Sequence mapping and alignment against reference genomes. Identification of Indels and Single Nucleotide Polymorphism (SNPs) (Samtools, GATK, FreeBayes). VCF file browsing. Types of SNPs. Estimation of genetic diversity (Tajima's D, FST, DXY, , among others.)

5th day:

Population structure analysis (PCA and ADMIXTURE). Identification of positive selection in protein coding genes using the MacDonal-Kreitman test. Demographic history of populations and species (PSMC). Simulation of evolutionary divergence scenarios from the "Site Frequency Spectrum (SFS)" and "Approximate Bayesian Computation".

**\* The content and / or order of the course may be subject to change during its development.**

Participant profile:

---

Graduates and researchers.

Number of participants (minimum/maximum):

---

4/20

#### Contact

- Dra. Irati Miguel and Dr. Fernando Rendo.

#### Course fee

- UPV/EHU users: 200 €
- PRB users: 350 €



- Genomics Service-Sequencing and Genotyping Unit.
- Biotechnology Center - Maria Goyri Building. UPV / EHU Science Park. Bizkaia Campus.Bº Sarriena, s / n 48940 Leioa (Bizkaia).
- Phone: 946012934/946018301.
- [irati.miguel@ehu.eus](mailto:irati.miguel@ehu.eus)  
[fernando.rendo@ehu.eus](mailto:fernando.rendo@ehu.eus)

- External users: 500 €

### Other additional information:

- The course is taught in Spanish and is theoretical-practical.
- The security measures necessary to carry out the practices will be followed at all times.
- Certificate of attendance will be delivered. Those attending the course will receive said certificate as long as they have completed at least 80% of the total duration of the course.
- **Each attendee is required to bring their computer.**
- Recommended operating systems:  
Ubuntu 18.04.6 LTS or later.  
MacOS Catalina v10.15.7 or later.  
Windows 10.
- A basic knowledge of the command line and programming is recommended.
- We encourage participants to visit the following tutorials:  
<https://ryantutorials.net/linuxtutorial/>  
<https://slurm.schedmd.com/documentation.html/>  
<https://stackabuse.com/python-tutorial-for-absolute-beginners/>  
<https://www.statmethods.net/r-tutorial/index.html/>
- The course is taught by Dr. **Luis Javier Chueca** ([luisjavier.chueca@ehu.eus](mailto:luisjavier.chueca@ehu.eus)). Postdoctoral researcher at the University of the Basque Country (UPV / EHU) and the LOEWE-Center for Translational Biodiversity Genomics (LOEWE-TBG). His line of research focuses on understanding the current distribution and phylogenetic relationships of Western Palearctic species, as well as identifying the genetic basis for niche evolution, speciation and local adaptation. It combines traditional taxonomy with cutting-edge bioinformatics and sequencing technologies to thoroughly investigate fauna diversification processes at the genomic level.