

**Guía para el Desarrollo del caso**  
**Maestría en Biología Computacional**  
**Titulación 2026-01**

**Caso: Resiliencia Andina: Integrando Genética Molecular en un Sistema de Agricultura Regenerativa para la Papa en la Región Andina**

El desarrollo de este caso debe estar alineado con *genómica real de papa* y con las técnicas moleculares actuales para fitomejoramiento.

Este documento es una guía de los procesos que los maestrantes pueden considerar para el desarrollo del caso y el reporte. Los datos, programas, paquetes o plataformas mencionadas no son de uso obligatorio, sino que dependerá del maestrante hacer la selección de los datos y el flujo de trabajo adecuado para aplicarlo de acuerdo al enfoque y alcance de su investigación.

**1. Especie de estudio**

***Solanum tuberosum* L. (papa cultivada)**

Datos taxonómicos, genómicos, fenotípicos, etc.

Familia:

Características del Genoma:

Tamaño del genoma:

**Especies silvestres relevantes (parientes)**

Datos tomados del **Potato Wild Relatives (PWR)**:

- *Solanum chacoense*
- *Solanum commersonii*
- *Solanum demissum*
- *Solanum phureja*
- *Solanum stenotomum*

Características de las especies silvestres:

- Resistencia a heladas
- Resistencia a sequía
- Tolerancia a patógenos
- Eficiencia en uso de nutrientes

**Bases de datos**

<b>Base</b>	<b>Qué aporta</b>
<b>NCBI GenBank</b>	Genomas, transcriptomas
<b>Ensembl Plants</b>	Anotaciones oficiales
<b>Sol Genomics Network (SGN)</b>	SNPs, QTLs, genes
<b>CIP (International Potato Center)</b>	Accesiones andinas
<b>EBI ENA / SRA</b>	Datos RNA-seq reales

Ejemplo real de datasets:

- Genoma referencia:  
*Solanum tuberosum* DM v6.1 (PGSC)

- RNA-seq estrés frío:  
SRA: **SRP074937**
- Población diversa:  
**CIP 300 genotypes panel**

## 2. Diseño experimental

### Poblaciones a muestrear

Grupo	N
Variedades modernas comerciales	40
Variedades nativas andinas	40
Especies silvestres	20
<b>Total</b>	<b>100 genotipos</b>

### Posibles fenotipos a estudiar (datos de campo de bibliografía)

Bajo sistema regenerativo:

- Rendimiento (kg/ha)
- Índice de vigor
- Tolerancia a estrés hídrico o térmico
- Eficiencia de uso de nitrógeno
- Biomasa radicular
- Índice de micorrización

### Datos moleculares a generar

#### Genómica

- **GBS o WGS ligero**
- Illumina NovaSeq
- ~10x cobertura por genotipo

#### Transcriptómica

RNA-seq en:

- Raíz
- Hoja

Condiciones:

- Estrés frío
- Estrés hídrico
- Control

## 3. Pipeline experimental (resumen visual)

Campo (datos bibliográficos)

↓

→ Fenotipado → ADN/RNA

↓

Secuenciación

↓

Datos crudos

↓

Análisis bioinformático  
↓  
Genes candidatos / SNPs  
↓  
Selección genómica

#### **4. Pipeline bioinformático completo**

##### **4.1 Control de calidad**

###### **Datos**

- FASTQ (DNA-seq / RNA-seq)

###### **Herramientas**

- FastQC
- MultiQC
- Trimmomatic

###### **Resultado**

- Reads filtradas y limpias

##### **4.2 Alineamiento**

###### **DNA-seq**

- BWA-MEM
- SAMtools

###### **RNA-seq**

- STAR
- HISAT2

###### **Resultado**

- Archivos BAM alineados al genoma DM v6.1

##### **4.3 Llamado de variantes (SNPs)**

- GATK HaplotypeCaller
- FreeBayes

###### **Resultado:**

- Archivo VCF con ~2–5 millones SNPs

##### **4.4 Análisis poblacional**

###### **Herramientas**

- PLINK
- ADMIXTURE
- PCA (R)

###### **Resultados**

- Estructura genética
- Diversidad genética
- Diferenciación entre grupos

##### **4.5 GWAS (asociación genotipo–fenotipo)**

###### **Herramientas**

- GAPIT (R)
- Tassel

## **Resultado**

- SNPs asociados a:
  - Helada
  - Sequía
  - Eficiencia nutricional

## **4.6 Transcriptómica diferencial**

### **Herramientas**

- featureCounts
- DESeq2 / edgeR

### **Resultado**

- Genes diferencialmente expresados (DEGs)
- Mapas de expresión bajo estrés

## **4.7 Integración multi-ómica**

Cruce de:

- SNPs GWAS
- DEGs
- QTLs SGN

Resultado:

### **Lista priorizada de genes candidatos adaptativos**

Ejemplos reales:

- CBF/DREB (respuesta al frío)
- P5CS (prolina, estrés hídrico)
- NRT1/NRT2 (nitrógeno)
- WRKY (estrés general)

## **4.8 Selección genómica**

### **Herramientas**

- rrBLUP
- BGLR

### **Resultado**

- Modelos predictivos
- Valor genómico estimado (GEBV)

## **5. Resultados esperados**

### **A nivel biológico**

- Las variedades nativas poseen alelos únicos asociados a:
  - Mayor biomasa radicular
  - Mejor tolerancia al frío
  - Menor dependencia de nitrógeno

### **A nivel genómico**

- Identificación de:
  - 50–200 SNPs adaptativos
  - 20–40 genes candidatos clave

### **A nivel aplicado**

- Se propone:

- Cruzamientos dirigidos
- Marcadores moleculares
- Selección genómica temprana

### Impacto real

Desarrollo de variedades de papa **productivas + resilientes + adaptadas a agricultura regenerativa.**

### 6. Análisis de resultados

Para el análisis de los resultados será necesario hacer una discusión en función del enfoque del estudio y los resultados obtenidos. El análisis tendrá como objeto relacionar los resultados obtenidos con datos previos respecto al genoma/transcriptoma de la papa en un contexto de conservación de diversidad genómica.

### 7. Reporte científico

Los maestrantes pueden utilizar los mismos puntos desarrollados en esta guía para redactar su reporte científico. Deberán subir al aula virtual el reporte escrito e incluir en éste el enlace de una carpeta del drive en la cual depositarán todos los archivos generados en la resolución del caso.

### 8. Rúbrica de evaluación

El caso se evaluará sobre 50 puntos de acuerdo a los siguientes criterios:

<b>Criterio</b>	<b>Excelente</b>	<b>Bueno</b>	<b>Regular</b>	<b>Insuficiente</b>
Caracterización de la especie	Usa $\geq 5$ bases de datos públicas con citas correctas; identifica $\geq 10$ genes relevantes <b>(10 puntos)</b>	Usa 3-4 bases de datos; identifica 5-9 genes <b>(8 puntos)</b>	Usa 1-2 bases; identifica $< 5$ genes <b>(6 puntos)</b>	No usa datos públicos o los usa incorrectamente <b>(4 puntos)</b>
Diseño experimental	Propuesta realista, con tamaño muestral justificado, controles y réplicas <b>(5 puntos)</b>	Diseño adecuado pero con alguna limitación en controles o réplicas <b>(4 puntos)</b>	Diseño vago o poco reproducible. <b>(3 puntos)</b>	Diseño inviable o ausente <b>(2 puntos)</b>
Pipeline bioinformático	Flujo completo, programas actualizados, integra $\geq 2$ tipos de datos ómicos.	Flujo adecuado pero falta integración o detalle <b>(8 puntos)</b>	Flujo incompleto o con herramientas obsoletas <b>(6 puntos)</b>	Pipeline ausente o erróneo <b>(4 puntos)</b>

	<b>(10 puntos)</b>			
Resultados esperados y análisis	Concretos, cuantificables, vinculados claramente al problema, se analiza con claridad el impacto biológico <b>(15 puntos)</b>	Generalizados pero vinculados al problema, se analiza con proyección biológica <b>(12 puntos)</b>	Vagamente relacionados, se analiza vagamente el impacto biológico <b>(10 puntos)</b>	No se vinculan al problema, no se analiza con enfoque biológico <b>(8 puntos)</b>
Originalidad y factibilidad	Propuesta innovadora, con recursos reales (ej. SRA, CRISPR) <b>(5 puntos)</b>	Propuesta convencional pero factible <b>(4 puntos)</b>	Propuesta poco innovadora o con recursos limitados <b>(3 puntos)</b>	Propuesta inviable o poco fundamentada <b>(2 puntos)</b>
Presentación, estructura y referencias	Informe bien organizado, con gráficos claros, cronograma realista, referencias que justifican el estudio <b>(5 puntos)</b>	Estructura clara pero con deficiencias menores, referencias adecuadas <b>(4 puntos)</b>	Estructura confusa o incompleta, referencias incompletas <b>(3 puntos)</b>	Desorganizado, sin formato solicitado, sin referencias <b>(2 puntos)</b>

## 9. Recursos Adicionales para Estudiantes

- Bases de datos clave:
  - SolGenomics (solgenomics.net)
  - NCBI SRA (Proyectos de papa andina)
  - Ensembl Plants (S. tuberosum)
  - PGRDEU (recursos fitogenéticos)
- Herramientas bioinformáticas sugeridas:
  - Galaxy para flujos reproducibles
  - Docker para entornos controlados
  - R (paquetes: ggplot2, bioconductor)