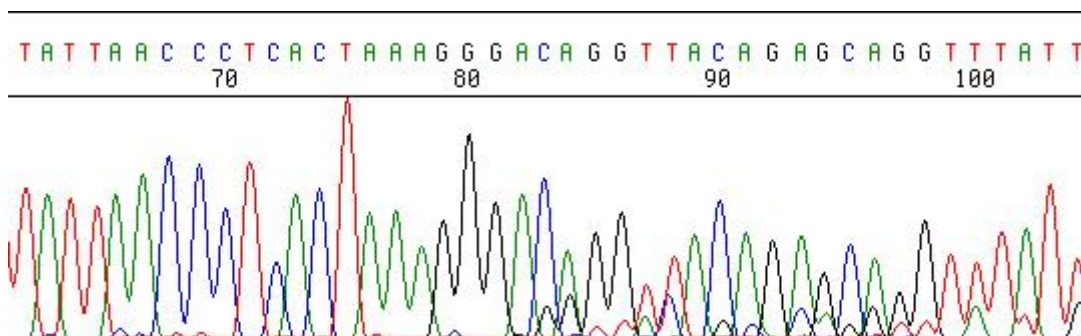


# Documentación proceso Automatización entrega secuencias sanger

v.1.0



## Contenido

1	Introducción .....	3
2	Pasos a realizar por la Unidad de Genómica .....	3
2.1	Actualización del fichero de configuración de la carrera .....	3
2.2	Copia de ficheros a recurso compartido .....	4
3	Recepción del correo con el enlace.....	4
4	Requisitos e instalación del script de automatización .....	5
4.1	Descripción del script .....	5
4.2	Requisitos .....	6
4.3	Instalación .....	6
4.4	Logs.....	6

## 1 INTRODUCCIÓN

---

La Unidad de Genómica lleva a cabo el servicio de Secuenciación Sanger del Instituto de Salud Carlos III. A través de este servicio los investigadores pueden solicitar la secuenciación de sus muestras a partir de un formulario. Una vez concluido el servicio la Unidad de genómica pone a disposición del investigador los resultados de la secuenciación utilizando un recurso compartido situado en Cibeles.

Esta situación presenta una serie de problemas:

- Obligatoriedad de realizar la copia de ficheros/resultados a la carpeta del investigador solicitante correspondiente. Una misma carrera de secuenciación contiene secuencias de varios usuarios, por lo que la separación de cada grupo de ficheros a cada investigador tiene que hacerse de manera manual.
- Los permisos a las carpetas de los distintos laboratorios que se utilizan en Cibeles se mantienen de manera difícil en el tiempo debido a la gran movilidad del personal en el centro.
- Mediante este mecanismo no hay una notificación al usuario de que tiene disponibles las secuencias, siendo éste un requisito solicitado para cumplir la norma ISO por la que está acreditada la Unidad de Genómica para el ensayo de secuenciación Sanger.

Para solucionar estos problemas se ha desarrollado un proceso de automatización que permite la compartición personalizada de los resultados de cada carrera a cada investigador solicitante. Además se realiza una notificación por mail a la persona o personas solicitantes con un link a la descarga de los archivos necesarios.

Por último se especifica que las muestras serán borradas automáticamente al cabo de 7 días, siendo los datos almacenados para su backup y archivo por la Unidad de Genómica con un proceso ajeno a este procedimiento.

## 2 PASOS A REALIZAR POR LA UNIDAD DE GENÓMICA

---

### 2.1 ACTUALIZACIÓN DEL FICHERO DE CONFIGURACIÓN DE LA CARRERA

Este fichero contiene todas las muestras de la secuenciación que han de ser asignadas a los investigadores. Para realizar esta separación se ha de rellenar en el campo “Comment” la dirección de correo del investigador para poder enviarle un email con el enlace de la carpeta donde se puede descargar las muestras.

Ejemplo:

Well	Sample Name	Comment	Results Group 1	Instrument Protocol 1	Analysis Protocol 1
A01	1992PBS_R	jperez@isciii.es	Secuenciacion	LongSeq50_POP7_Z_00518	3730P7BD3_1_seq

Rellenando todas las filas de la columna “Comment” asignando cada una de ellas con el correo del investigador se podrá asignar todas las muestras pertenecientes a un investigador y permitir con ello que pueda acceder a ellas.

En el caso de que una muestra o conjunto de muestras hayan de compartirse con más de 1 investigador, se separaran los correos electrónicos con 2 puntos “:”.

Ejemplo:

Well	Sample Name	Comment	Results Group 1	Instrument Protocol 1	Analysis Protocol 1
A11	44587SAR_R	<b>igarcia@isciii.es:pruiz@isciii.es</b>	Secuenciacion	LongSeq50_POP7_Z	37OP7B9_1_seq

## 2.2 COPIA DE FICHEROS A RECURSO COMPARTIDO

Una vez finalizada la carrera y disponibles las secuencias a compartir, se copiarán al recurso compartido sanger\_seq al que tiene acceso la Unidad de genómica:

- Fichero “.txt” generado para la configuración de carrera con la información del solicitante. Ej. GN18-190INFA.txt
- Carpeta con las secuencias obtenidas de la carrera. Ej. GN18-190INFA

## 3 RECEPCIÓN DEL CORREO CON EL ENLACE

El proceso de automatización detectará cuando los resultados de una nueva carrera estén disponibles y separará los resultados por investigador generando una carpeta compartida donde sólo tendrán acceso los usuarios establecidos en el archivo de configuración de la carrera.

Por último enviará un correo electrónico como el que se muestra a continuación informando de que los resultados están disponibles, junto con el link necesario para acceder a las secuencias.

Sanger sequencing run GN18-190AINFA has finished.

Your samples has been successfully sequenced!

You can retrieve your samples from:

\\barbarroja\20181030 GN18-190AINFA m.jimenez smonzon

The samples will be available for 7 days, afterwards the folder will be deleted.

If you have any question/issue please contact us:

Contacto	Email
Bioinformatics Unit	<a href="mailto:bioinformatica@isciii.es">bioinformatica@isciii.es</a>
Genomics Unit	<a href="mailto:azaballos@isciii.es">azaballos@isciii.es</a>

**BU-ISCIII**

Bioinformatics Unit , Instituto de Salud Carlos III  
Majadahonda  
Madrid

#### NOTA:

Para una misma carrera, el usuario podrá recibir más de un correo, dependiendo de las muestras que ha compartido con otros usuarios.

Es decir, si en el fichero configuración de la carrera fuese como en este ejemplo:

Well	Sample Name	Comment	Results Group 1	Instrument Protocol 1	Analysis Protocol 1
A10	44587SAR_R	<a href="mailto:paco@isciii.es">paco@isciii.es</a> : <a href="mailto:pedro@isciii.es">pedro@isciii.es</a>	Secuenciacion	LongSeq50_POP7_Z	37OP7B9_1_seq
C01	45787SAR_R	<a href="mailto:paco@isciii.es">paco@isciii.es</a> : <a href="mailto:eva@isciii.es">eva@isciii.es</a>	Secuenciacion	LongSeq50_POP7_Z	37OP7B9_1_seq
D05	44857SAR_R	<a href="mailto:paco@isciii.es">paco@isciii.es</a>	Secuenciacion	LongSeq50_POP7_Z	37OP7B9_1_seq

El usuario Paco recibiría:

- Un correo con el enlace a las muestras compartidas con Pedro
- Otro correo con el enlace de las muestras compartidas con Eva
- Otro correo con el enlace de las muestras que son sólo de él.

## 4 REQUISITOS E INSTALACIÓN DEL SCRIPT DE AUTOMATIZACIÓN

### 4.1 DESCRIPCIÓN DEL SCRIPT

Para el desarrollo de este proceso de automatización se ha desarrollado un script en bash que realiza los siguientes pasos:

- Separación de las secuencias por investigador.

- Compartición de las carpetas mediante CIFS/SAMBA.
- Envío de correo electrónico por carpeta compartida.

La ejecución del script es controlado por un proceso crontab, configurado para visualizar la carpeta compartida sanger\_seq de la Unidad de genómica, y esperar a que una nueva tanda de secuencias estén disponibles. Cuando esto ocurre el proceso crontab lanza el script de compartición de secuencias.

## 4.2 REQUISITOS

Para su funcionamiento el script necesita funcionar en un servidor con S.O Linux, el programa está testado y desarrollado en centos 6.10 por lo que se recomienda su utilización.

Además el programa tiene las siguientes dependencias:

- Instalación y configuración del servidor samba.
- Integración del sistema Linux con el directorio activo del ISCIII usando samba, winbind y kerberos. (ejemplo: panoramix, servidor de ficheros en Bioinfo01)
- Conectividad: puertos de correo y samba abiertos para permitir envío de correo y conexión de los usuarios a los recursos compartidos de samba.

## 4.3 INSTALACIÓN

1. Añadir nuevo recurso compartido (sanger\_seq) con usuarios de Unidad de genómica con permisos de acceso y escritura, en el fichero de configuración de samba /etc/samba/smbdir.conf.
2. Crear carpeta /etc/samba/shares donde se irán creando las configuraciones de las distintas carpetas compartidas.
3. Crear carpeta sanger\_seq\_users donde se irán creando las distintas carpetas compartidas para los usuarios.
4. Cambiar rutas necesarias en script sanger\_script\_contrab.sh
5. Configurar crontab que lance el script sanger\_script\_crontab.sh para que se ejecute cada (tiempo por determinar).

## 4.4 LOGS

Se generan tres logs diferentes:

1. Crontab log: información sobre las operaciones que realiza el script.
2. Run\_processed: nombre de las carreras de secuenciación sanger procesadas.
3. Samba\_folders: información sobre las carpetas que se van compartiendo indicando carrera, investigador/es, fecha y número de ficheros.

## VERSION CONTROL

Version	Author	Date	Reasons for the Document update
V.1.0	Luis Chapado	20/11/2018	Primera guía de usuario y de instalación del proceso de automatización de entrega de resultados sanger.
V. 1.1	Sara Monzón	26/11/2018	Revisión del texto. Introducción añadida e información de requisitos de instalación,