

---

# **RAPIFILT MANUAL DE USUARIO**

# **RAPIFILT USER MANUAL**

---

Versión 1.0 / Version 1.0

# Contenido / Contents

1.	ESPAÑOL . . . . .	2
1.1.	Descripción . . . . .	2
1.2.	Instalación en Linux y MAC . . . . .	2
1.3.	Instalación usando el paquete BIOCONDA . . . . .	2
1.4.	Ejecución . . . . .	2
1.5.	Resultados . . . . .	3
1.6.	Opciones de configuración . . . . .	3
1.7.	Preguntas Frecuentes . . . . .	3
2.	ENGLISH . . . . .	5
2.1.	Description . . . . .	5
2.2.	Linux and MAC Installation . . . . .	5
2.3.	CONDA installation . . . . .	5
2.4.	Running . . . . .	5
2.5.	Results . . . . .	6
2.6.	Filter options . . . . .	6
2.7.	FAQ . . . . .	6

# 1. ESPAÑOL

## 1.1. Descripción

RAPId FILTER (RAPIFILT) es un software escrito en C para el control de calidad de secuencias de DNA. Este permite remover bases con baja calidad y eliminar secuencias cortas.

Base	A	G	T	C	A	T	G	A	T	C	A	T
Calidad	(	#	:	;	?	@	#	%	:	;	?	@
Valor	7	2	25	26	30	31	2	4	25	26	30	31
	a)								b)			

Figura 1. a) Secuencia origen. b) Salida generada por RAPIFILT (Q>20).

## 1.2. Instalación en Linux y MAC

Descargue los códigos ([aquí](#)) y ejecute:

```
$> unzip RAPIFILT_v1.1.zip
$> cd RAPIFILT_v1.1
$> make
$> export PATH=$PATH:$(pwd)
```

## 1.3. Instalación usando el paquete BIOCONDA

RAPIFILT está disponible en el gestor de paquetes BIOCONDA. Para usar esta opción es necesario instalar [Miniconda](#) y ejecutar el siguiente comando.

```
$> conda install -c bioconda rapifilt
```

## 1.4. Ejecución

Ejecute RAPIFILT directamente desde la terminal, use como argumentos el nombre del archivo, la opción de filtrado y nombre del archivo de salida.

Ejemplo:

```
$> rapifilt -fastq bmini.fq -l 20 -r 20 -o bmini_salida
```

## 1.5. Resultados

RAPIFILT genera los siguientes archivos:

- `bmini_salida.fastq`: Archivo con las secuencias filtradas.
- `bmini_salida_bad.fastq`: Secuencias eliminadas.
- `bmini_salida_stat.txt`: Estadística por base.

## 1.6. Opciones de configuración

Use las siguientes opciones para configurar RAPIFILT:

- `-h` Muestra la ayuda
- `-v` Información del programa y la versión
- `-f` Salida en formato FASTA
- `-l` Valor de corte al inicio de la secuencia
- `-r` Valor de corte al final de la secuencia
- `-w` Tamaño de la ventana para verificar los niveles de calidad
- `-m` Longitud mínima de la secuencia
- `-mx` Longitud máxima de la secuencia
- `-fastq` Entrada en formato FASTQ
- `-sff` Entrada en formato 454 tipo SFF
- `-i` Entrada en formato Illumina (Forward, Reverse)
- `-o` Nombre del archivo de salida
- `-tb` Remueve las primeras n bases sin importar la calidad
- `-te` Remueve las últimas n bases sin importar la calidad
- `-bin` Tamaño usado para computar la estadística por base

## 1.7. Preguntas Frecuentes

### ▪ Formato de entrada

RAPIFILT soporta secuencias en formato FASTQ y SFF, el archivo puede estar comprimido en formato gz. Para Illumina (paired-end) las secuencias deben estar en archivos independientes. Archivos interlineados no son directamente soportados pero pueden ser previamente procesados con nuestro programa [unmerge](#).

### ▪ Sistema Operativo requerido

RAPIFILT está diseñado en C y ha sido probado bajo ambientes Linux y MAC. Para Windows recomendamos la versión disponible en el gestor de paquetes BIOCONDA.

- **Códigos y archivos de prueba**

Los códigos originales y secuencias de prueba pueden ser descargados desde: [RAPIFILT-GitHub](#)

- **Información adicional**

Por favor diriga sus preguntas a los siguientes correos:

[bernardo.benavides@udea.edu.co](mailto:bernardo.benavides@udea.edu.co)

[felipe.cabarcas@udea.edu.co](mailto:felipe.cabarcas@udea.edu.co)

## 2. ENGLISH

### 2.1. Description

RAPId FILTER (RAPIFILT) is a C program to filter DNA sequences. It trims low-quality bases and removes short sequences.

Base	A	G	T	C	A	T	G	A	T	C	A	T
Quality	(	#	:	;	?	@	#	%	:	;	?	@
Value	7	2	25	26	30	31	2	4	25	26	30	31
	a)								b)			

Figure 1. a)Original sequence. b)RAPIFILT output sequence (Q>20).

### 2.2. Linux and MAC Installation

Download RAPIFILT source codes ([here](#)) and type:

```
$> unzip RAPIFILT_v1.1.zip
$> cd RAPIFILT_v1.1
$> make
$> export PATH=$PATH:$(pwd)
```

### 2.3. CONDA installation

RAPIFILT is available for the BIOCONDA repository. You should install [Miniconda](#) and execute the following command.

```
$> conda install -c bioconda rapifilt
```

### 2.4. Running

RAPIFILT can be executed from the console with the user's arguments.  
Example:

```
$> rapifilt -fastq bmini.fq -l 20 -r 20 -o bmini_out
```

## 2.5. Results

RAPIFILT generates the following files:

- `bmini_out.fastq`: Filtered sequences.
- `bmini_out_bad.fastq`: Removed sequences.
- `bmini_out_stat.txt`: Statistics per base.

## 2.6. Filter options

RAPIFILT supports the following options:

- `-h` Help message
- `-v` Program and version information
- `-f` Enable fasta output (default fastq)
- `-l` Set left-cut value for quality scores (int default 0)
- `-r` Set right-cut value for quality scores (int default 0)
- `-w` Set windows size to check on the quality scores (int default 1)
- `-m` Filter sequence shorter than `min_len` (int default 1)
- `-mx` Filter sequence larger than `max_len` (int default 5000)
- `-fastq` single fastq input (file.fastq) the file can be gz compressed
- `-sff` 454 input (file.sff)
- `-i` Illumina inputs(file1.fastq file2.fastq) the files can be gz compressed
- `-o` Desired fastq output file. If not specified to stdout
- `-tb` Remove `n` bases from the begins of sequencing fragments (int default 0)
- `-te` Remove `n` bases from the ends of sequencing fragments (int default 0)
- `-bin` Bin size used to compute statistic per base (int default 1)

## 2.7. FAQ

### ▪ Input sequences

RAPIFILT receives unpaired reads in FASTQ or SFF format; the file can be gz compressed. Illumina paired-end (separated forward and reverse files) is also supported. Interlaced forward and reverse are not directly supported, but our [unmerge](#) program can preprocess them.

### ▪ What operating system is required?

RAPIFILT is a C program, and it has been tested on Linux and MAC environments. For computers with Windows OS, we recommended the BIOCONDA version.

- **Source codes and test files**

Source codes and test files can be download from: [RAPIFILT-GitHub](#)

- **Additional information**

Please contact us:

[bernardo.benavides@udea.edu.co](mailto:bernardo.benavides@udea.edu.co)

[felipe.cabarcas@udea.edu.co](mailto:felipe.cabarcas@udea.edu.co)