

# High-throughput workflow for RNAseq data treatment linking laboratory data server and remote parallel calculation platform

*Marie GARAVILLON-TOURNAYRE, David LOPEZ, Marion LANDI, Corentin HOCHART,  
Jean-Stéphane VENISSE, Boris FUMANAL, Aurélie GOUSSET-DUPONT, and Philippe LABEL*



philippe.label@clermont.inra.fr  33+(0)473 407 922



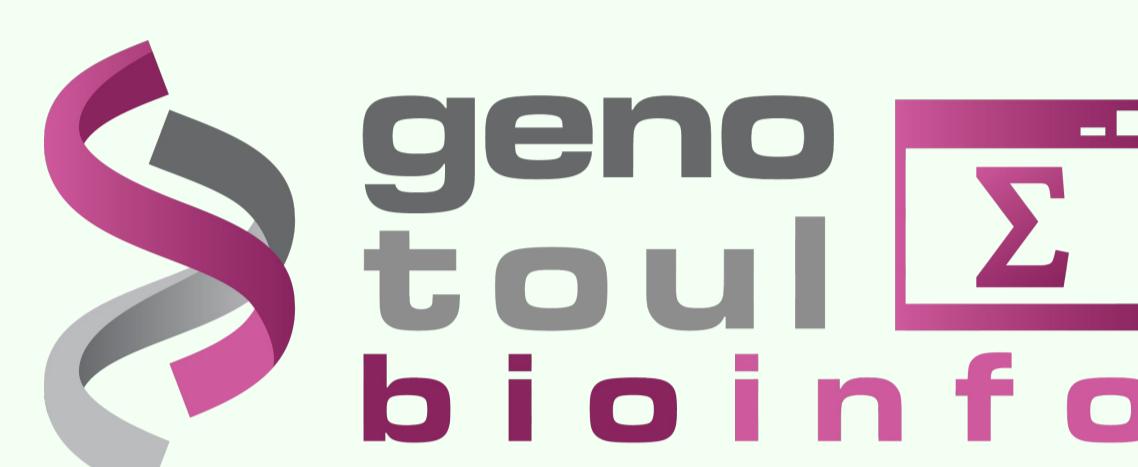
33+(0)473 407 922

# **PIAF UMB-547**

A large, three-dimensional red question mark is positioned in the center-right of the frame. To its right, a small white humanoid figure with its hand to its chin appears to be in a state of thought or contemplation. The background features a light pink rectangular area with a red dashed border. Above this area, the word "PROBLEM" is written in large, bold, red capital letters, oriented diagonally from the bottom-left towards the top-right. The overall composition suggests a theme of inquiry or seeking a solution to a problem.

1. RNAseq softwares using graphical user interfaces are slow and data volume limited
  2. We want to maximize the use of calculation resources and speed-up results

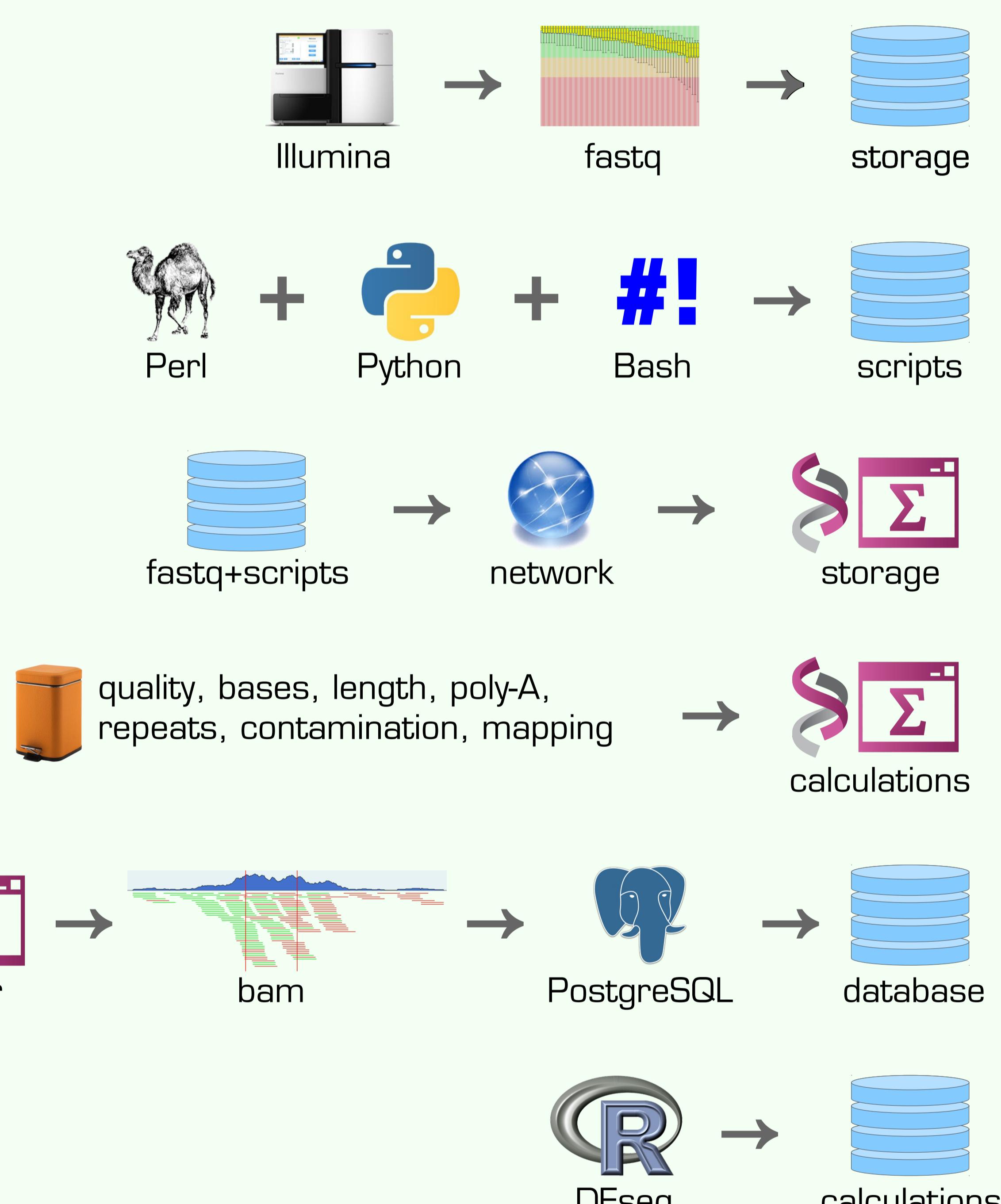
A 3D white stick figure stands in the foreground, holding a large green checkmark with both hands. The figure is positioned in front of a light blue rectangular background. Diagonally across the background, the word "SOLUTION" is written in large, bold, green, sans-serif capital letters. The letters are partially cut off at the top right corner of the image. A dashed green line forms a parallelogram around the word "SOLUTION".



# *remote cluster*

4856 cpu - 34.6Tb

# Tasks automatization through scripted workflow



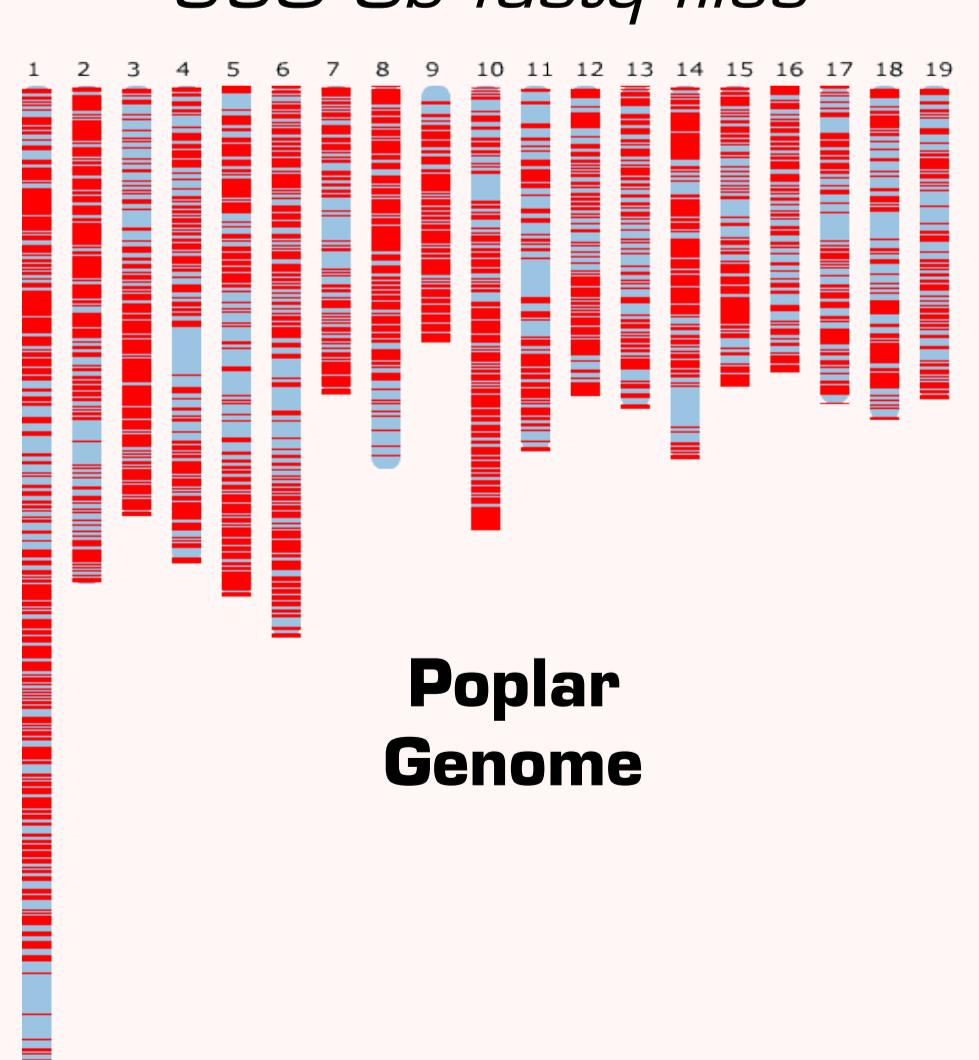
1. **sequences stored and secured**
  2. **scripts automatically prepared locally**
  3. **transfert of sequences and scripts**
  4. **calculations massively parallelized**
  5. **mapping results download**
  6. **biostatistics performed locally**

A large, bold red stamp with the word "EXAMPLE" in white capital letters is angled across the top left corner of the page. Below the stamp is a photograph of a black poplar tree (Populus nigra) with dense green foliage, standing upright on a stack of four light-colored wooden boards. The boards are arranged in a stepped, overlapping fashion, with the tree centered on the top board.



# INPUT

1,4 Gseq  
paired-end 2 x 100 bp  
48 samples, 6 genotypes,  
2 conditions, 4 repetitions  
*350 Gb fasta files*



## **OUTPUT**

7% sequences filtered out  
83% mapped on reference  
*66 Gb bam files*  
*129 Mb count files*

