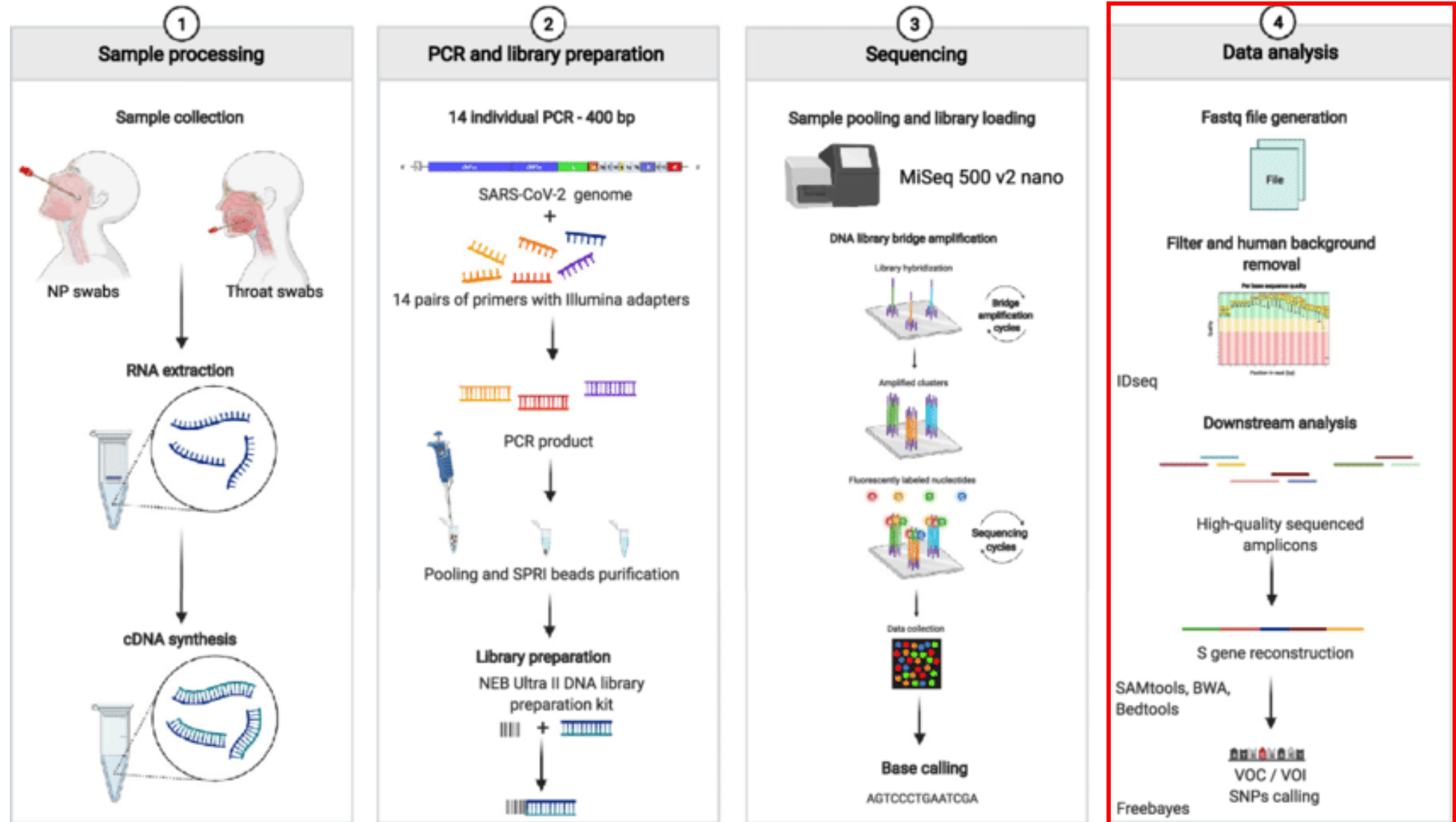
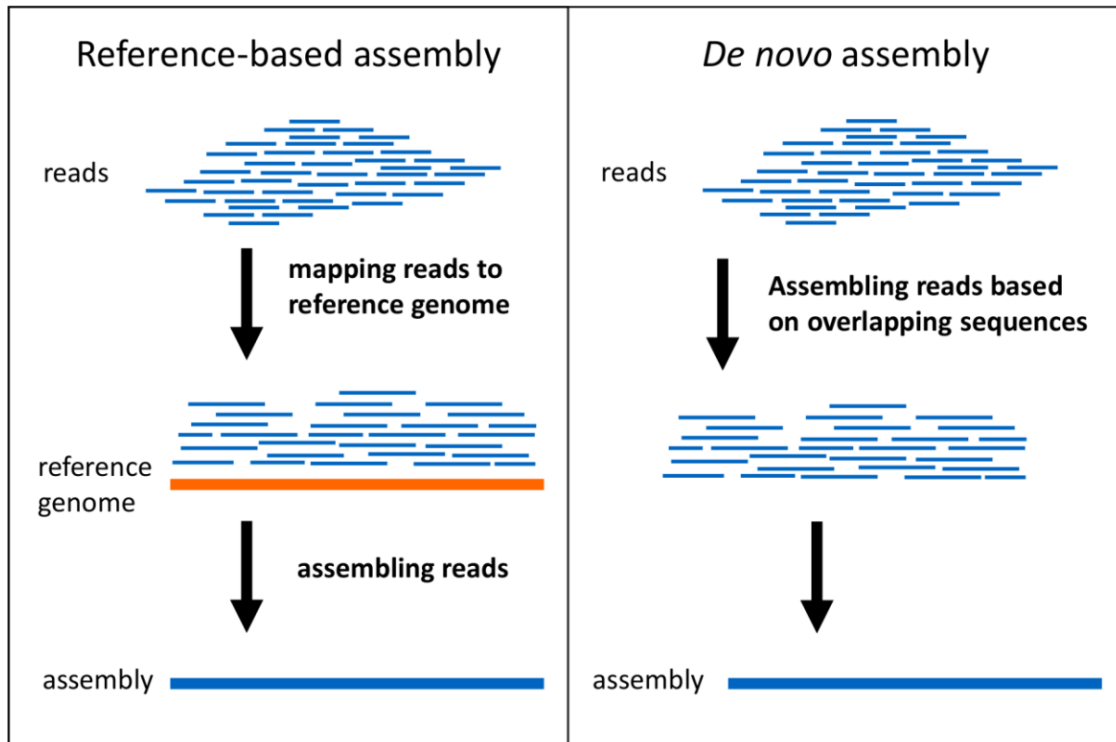


# PROCESAMIENTO DE NGS DATA Y ENSAMBLAJE GENÓMICO CON REFERENCIA

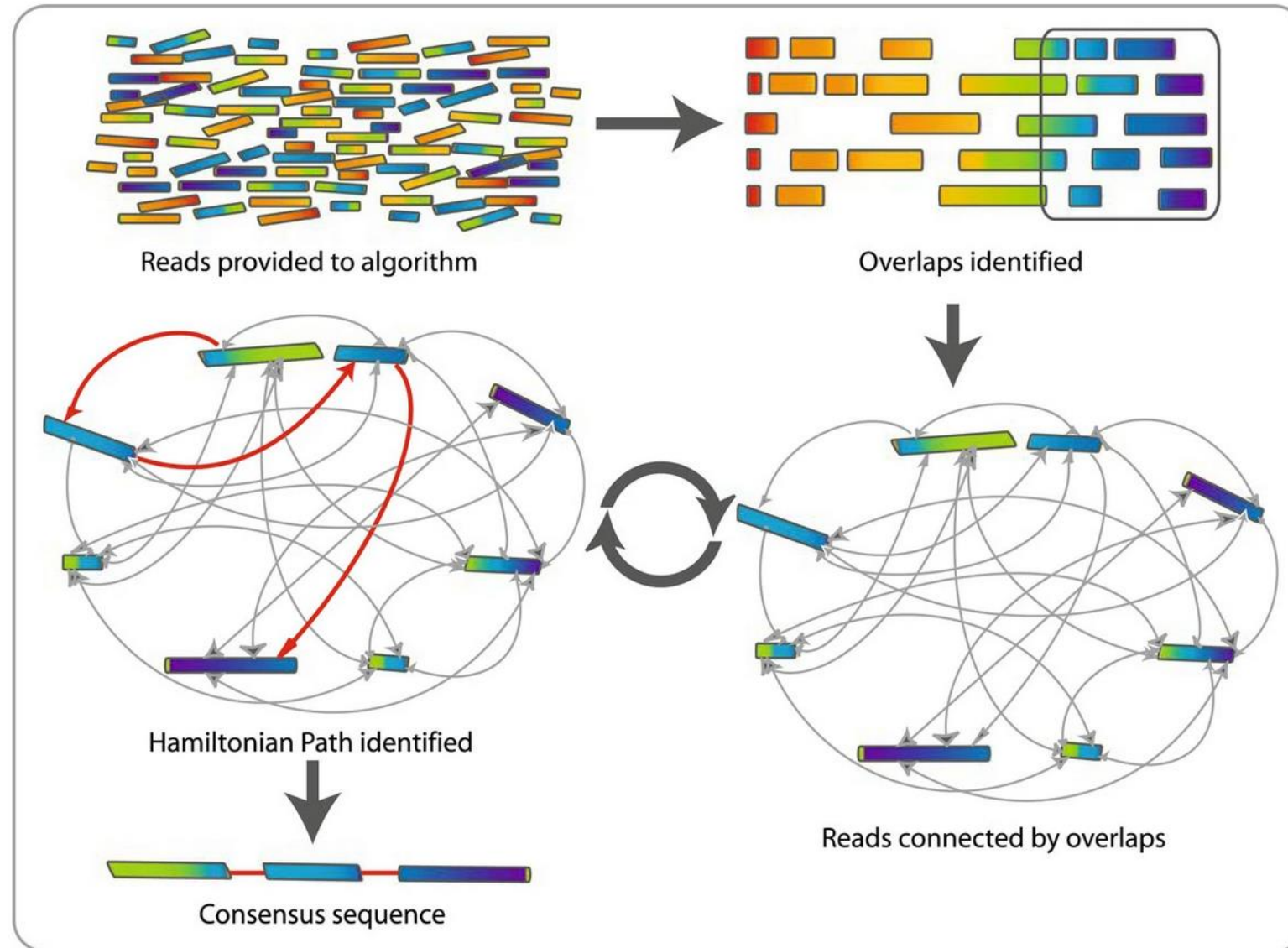
Murilo Cassiano, M.Sc.



# Tipos de ensamblaje genómico



# ¿Cómo recuperar la secuencia original?



# ¿Cómo recuperar la secuencia original?





# Sabemos cuál es la secuencia original

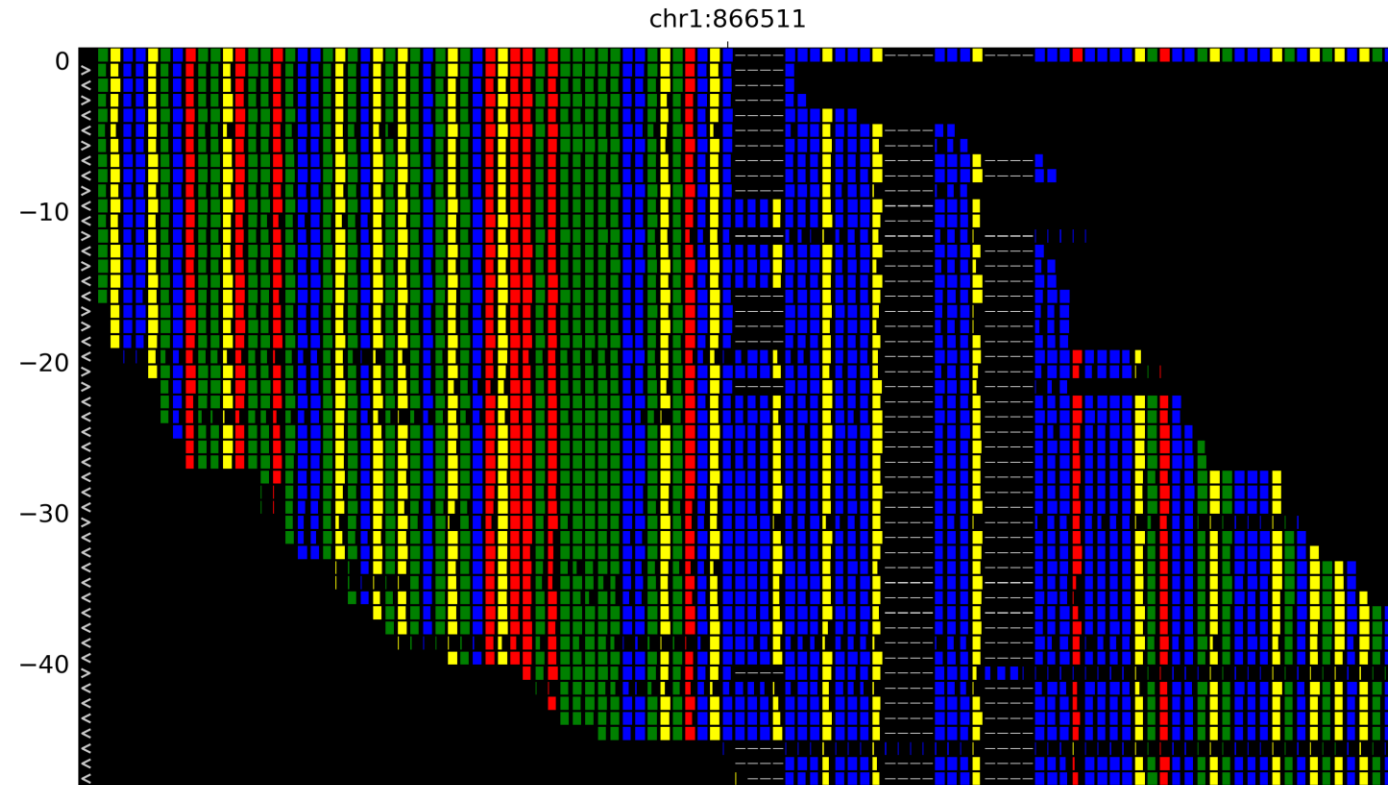
## Consensus

### Target Sequence

...ACTACGTGCTATGCAATCAGATTGCATGTCGGATCGATTTAGCTTAGCTCTGCTAGTC...

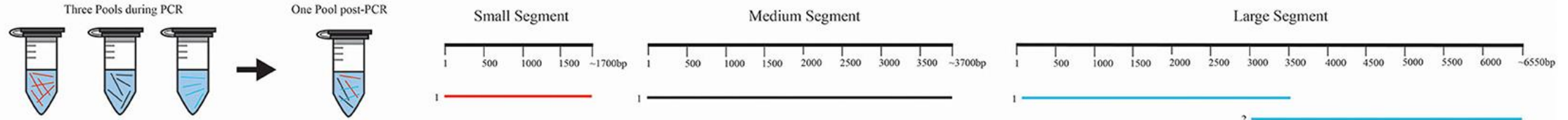
### Mapped Sequences

```
ACTACGCGCTAGGCAAACAG      AGATCGATATAGCGTAGCTC
ACGTACTATGCAAACAGATT      TCG-TTTAGTTTAGTTCTGC
TGGTATCCAAACAGTTTGCA      TTTTAGCTAGCTATGCTAG
TTTGCAACGAGATAGCATGT      AAGCT-AGCTCTGTTAGTC
AC      GCAAACAGA-TGCATATCGG      CTTAGCTATGCTAGTC
A-TAC      AACAGAGTGCATGACGGATC      AGCTCTGCTCGTC
ACTACGTG      AGATTACGTGTCGGTTCGAT      TCTGGTAGTC
-CTCCGTGCTA      TTGCAT-TCGCATCGATTTA      GCTTGTC
ACTACGGGCTTTGC      CATGGCGGA-CGATGTAGCT      AGTC
AATACGTCCTATGCAA      GTCGGATAGATTT-GCTTAG
ACTACGT-CTATGCAAACAG      GGATCGATCTAGCTTAGCTC
ACGTGCTAAGCAAACAGATT      TCGAGTTAG-TTAACCTCTGC
TGTATGCAA-CAGTTTGCA      ATTTGGCTTAGCTCCGCTAG
TATTCAAACAGATTGCATGT      TAGTTT-GCTGTGCTAGTC
GCAAACAGAGTGCAGGTCGG      CTTGGCTTTGCTGGTC
```

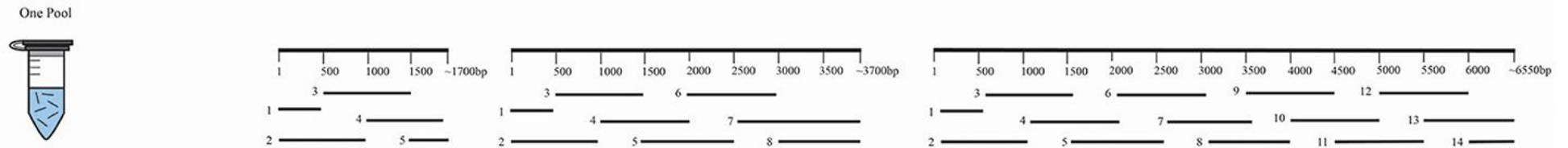


# Podemos enriquecer nuestra muestra

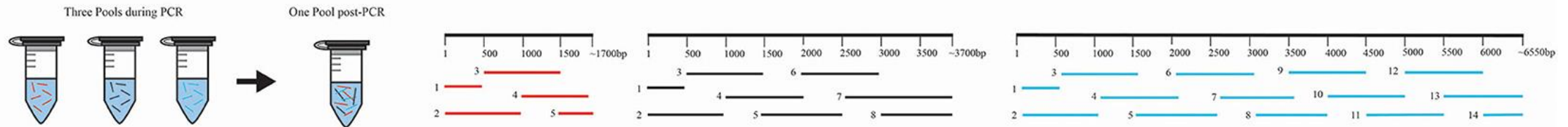
## A Whole Segment PCR



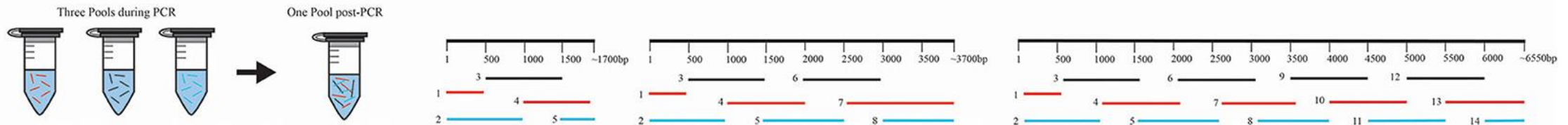
## B Whole Genome Multiplex Tiling



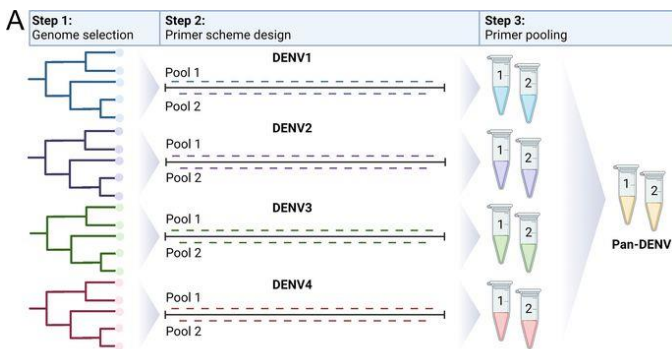
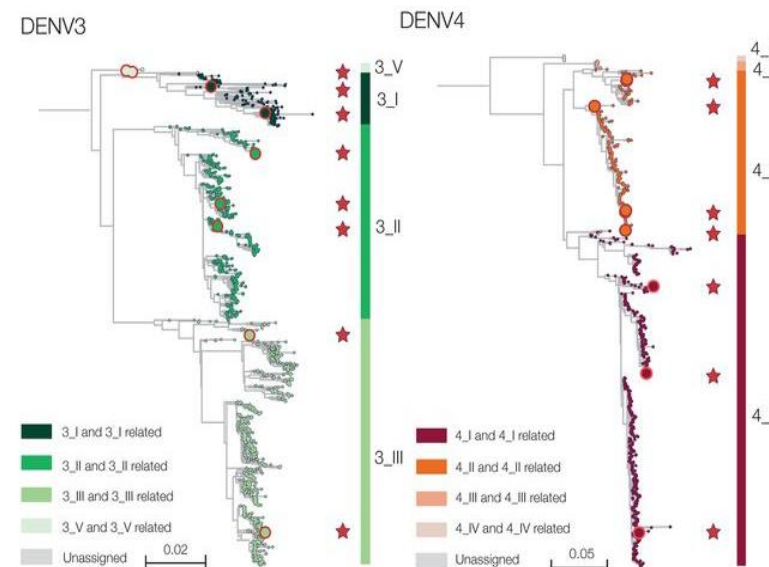
## C Segment-Specific Tiling



## D Disjointed Tiling

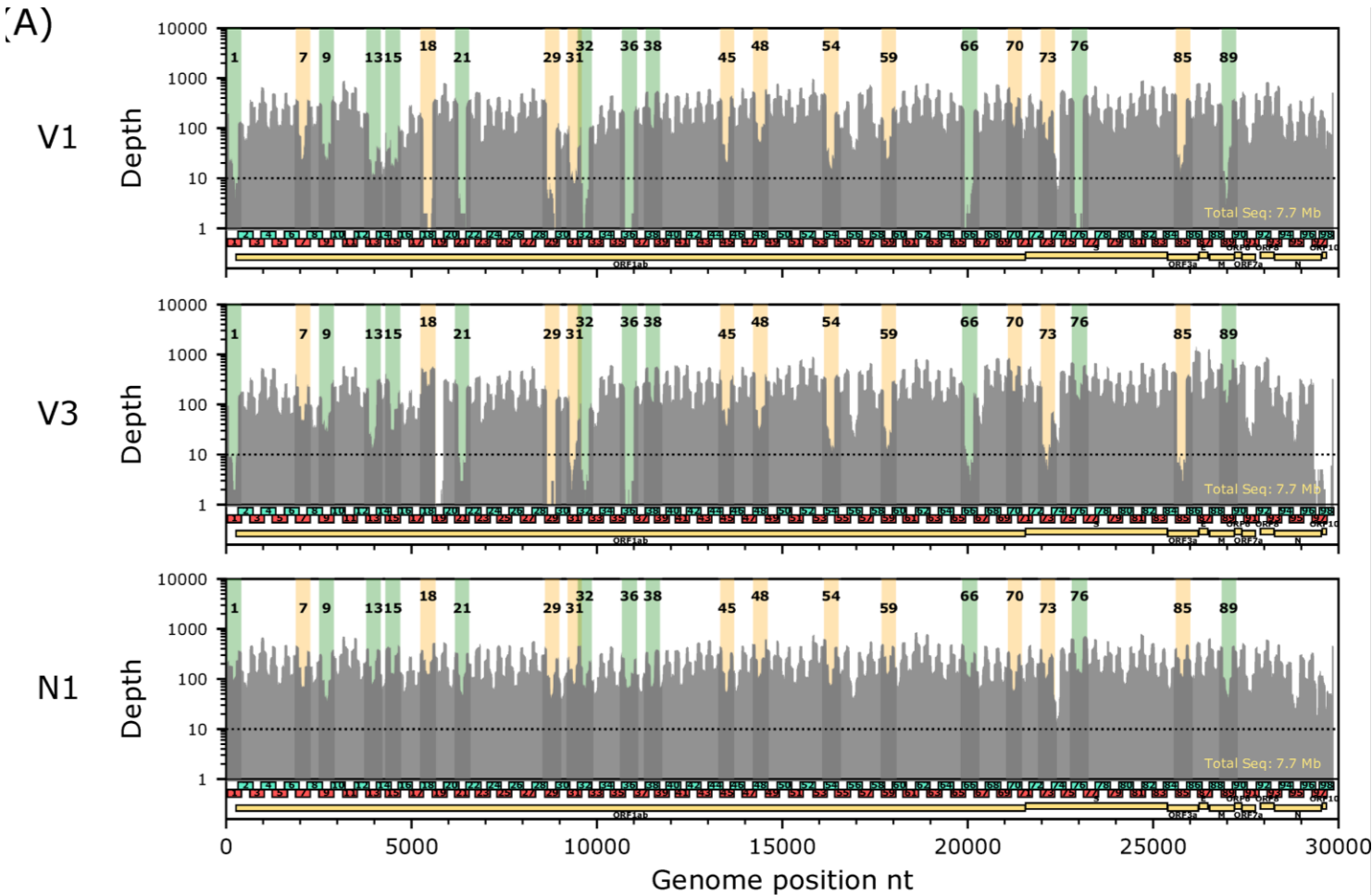


<https://doi.org/10.1101/2023.10.13.23296997>





# Amplicon-based también tiene limitaciones



A

Pool 1

7\_LEFT vs 9\_RIGHT  
ATCAGAGGCTGCTCGTTTGTGA > 3'  
3' < CACAACACCGCTACGACAGTA

21\_LEFT vs 73\_LEFT  
TGGCTATTGATTATAAACACTACACCC > 3'  
3' < TGTGGGTTTTTACCTAGTAATGTTTAAAC

15\_LEFT vs 15\_RIGHT  
ACAGTGCTTAAAAAGTGTAAGTGCC > 3'  
3' < TCACGGTCGATGTCAAAGACAA

21\_LEFT vs 29\_RIGHT  
TGGCTATTGATTATAAACACTACACCC > 3'  
3' < TGTGTGGTAGTTTTGAATATCTCATGTGA

45\_LEFT vs 89\_LEFT  
TACCTACAACCTGTGCTAATGACCC > 3'  
3' < TTACTGGTGTACCTTGCGCATG

85\_LEFT vs 13\_RIGHT  
ACTAGCACTCTCCAAGGGTGT > 3'  
3' < TCCCACAAAATTGACGACACCA

1\_RIGHT vs 31\_LEFT  
CATCTTTAAGATGTTGACGTGCCTC > 3'  
3' < CGGCACGGATGTCATGAGTCTT

59\_LEFT vs 89\_RIGHT  
CTCACGCATGATGTTTCATCTGCA > 3'  
3' < ACAAAGTAGAGCAACTGAAAGTCCA

Pool 2

18\_LEFT and 76\_RIGHT  
TGGAAATACCCACAAGTTAATGGTTTAAAC > 3'  
3' < TACCAAAATGTCCGTGTCACA

36\_LEFT vs 70\_RIGHT  
TTAGCTTGGTTGTACGCTGCTG > 3'  
3' < GACGACAATACAGAAATTTCTTCCAGT

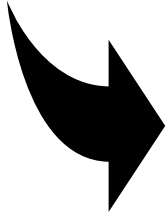
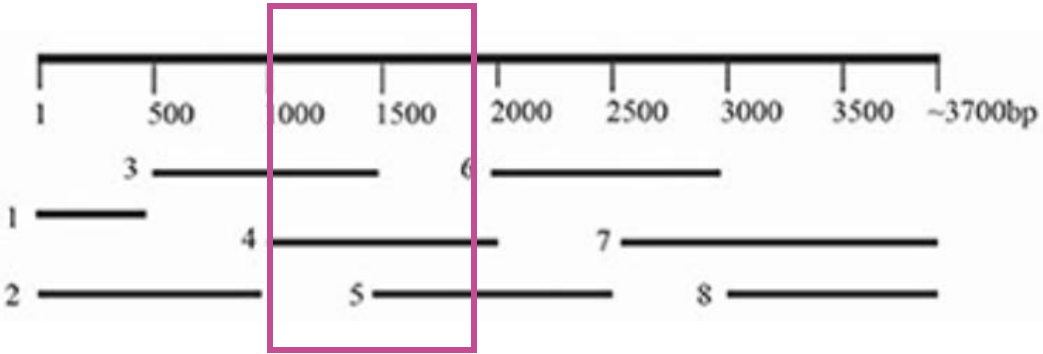
36\_LEFT vs 70\_RIGHT  
TTAGCTTGGTTGTACGCTGCTG > 3'  
3' < GACGACAATACAGAAATTTCTTCCAGT

66\_LEFT vs 54\_RIGHT  
GGGTGTGGACATTGCTGCTAAT > 3'  
3' < ACACGATTACCTGTTCAAAACCAA

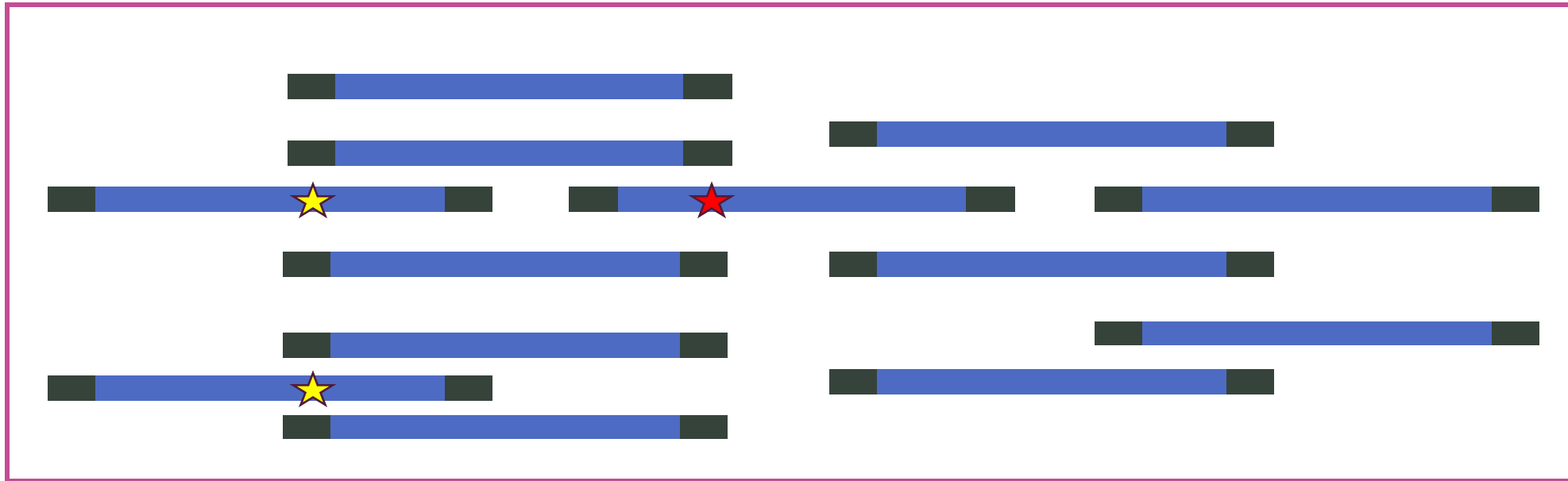
36\_LEFT vs 48\_RIGHT  
TTAGCTTGGTTGTACGCTGCTG > 3'  
3' < CGTGCGACGAAGACCATTAGAT

32\_RIGHT vs 38\_RIGHT  
AGCACATCACTACGCACTTTAGA > 3'  
3' < GCGATGAAATCTGACTGAGAACCAC

# Identificar variaciones genéticas incorrectas



**Potencial de interpretación errónea:** Las secuencias de primers pueden ser confundidas con variantes reales, llevando a resultados incorrectos.



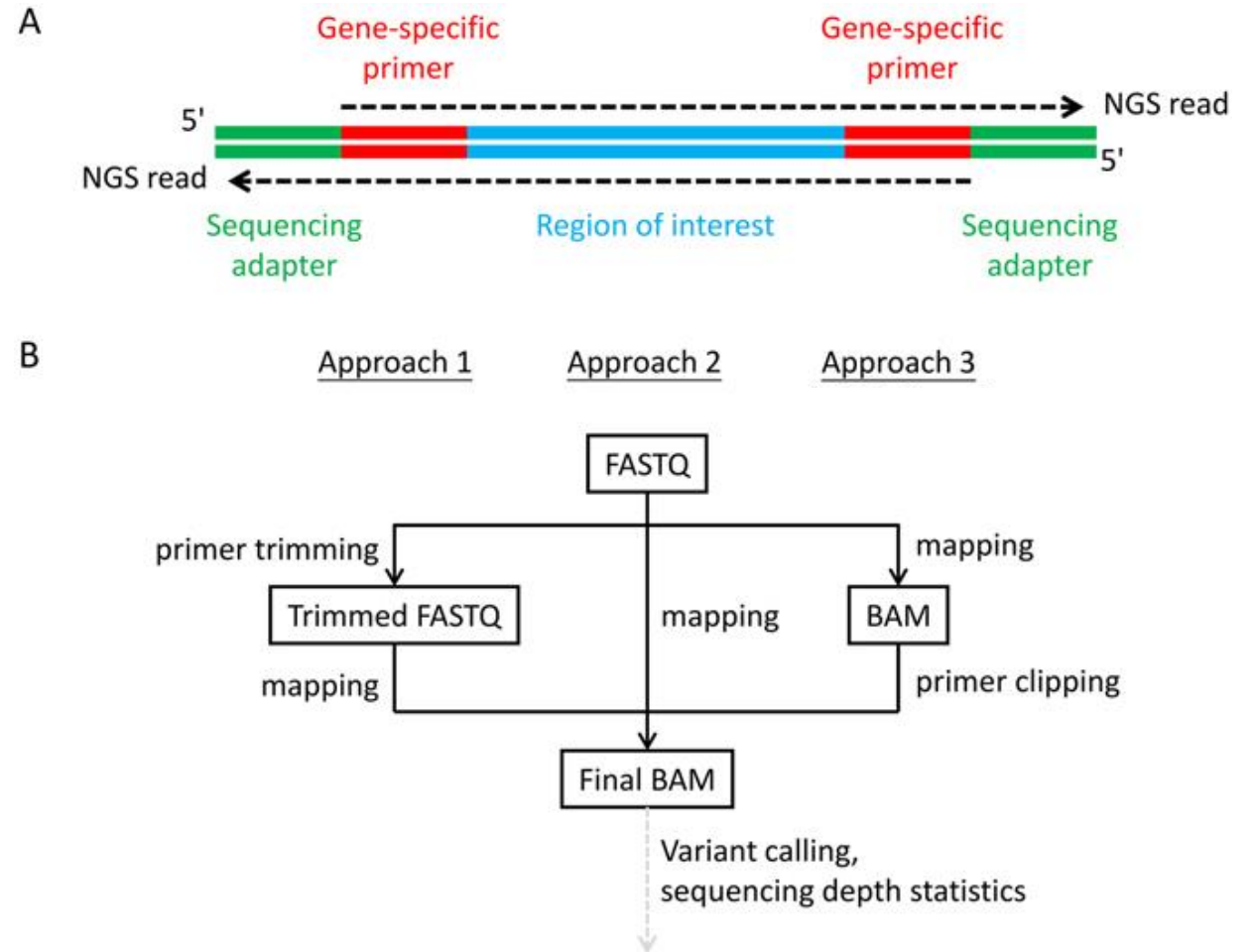
# 1. Sacar a los primers antes de mapear

## *Pros:*

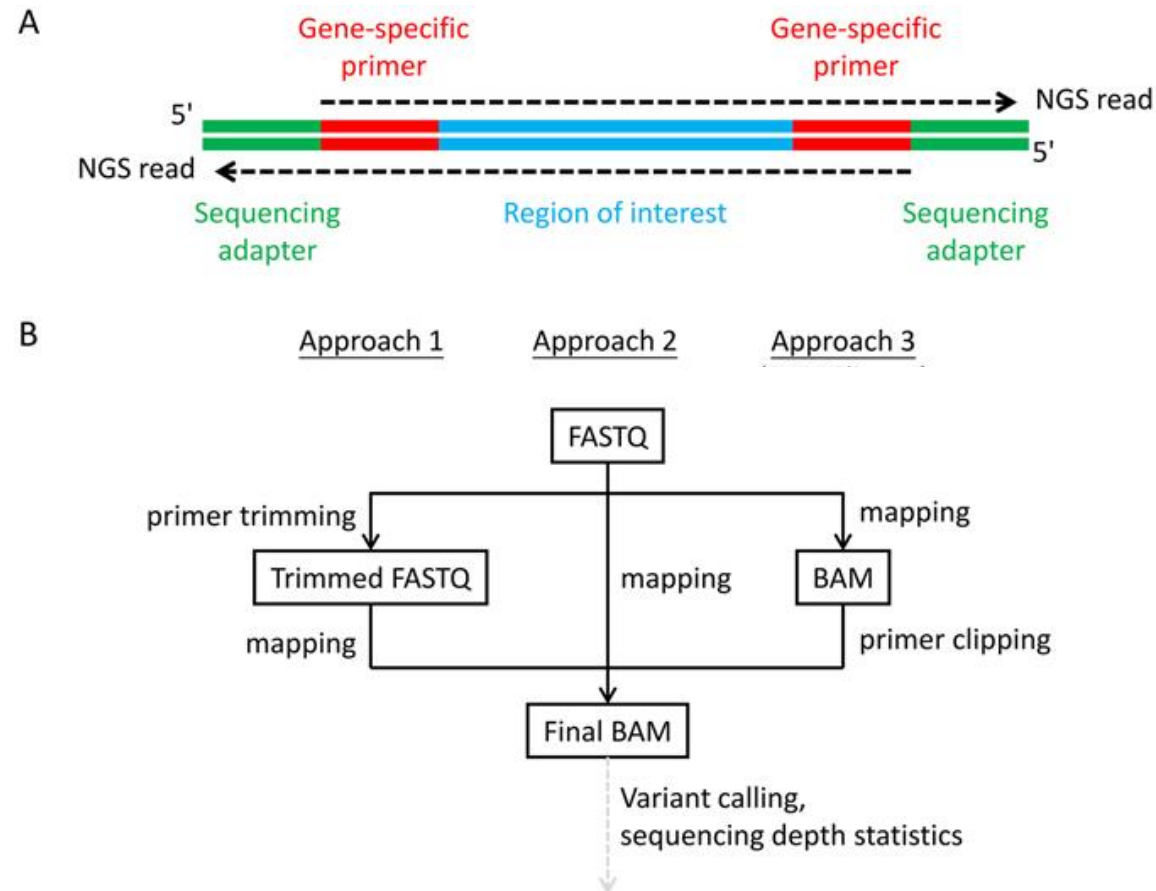
- **Menos falsos positivos:** reduces la posibilidad de identificar variaciones genéticas incorrectas basadas en los mismos.
- **Reducción del tamaño de los datos:** Menos datos para procesar puede resultar en un proceso más rápido.

## *Contras:*

- **Requiere herramientas adicionales:** Necesitarás herramientas específicas para eliminar eficientemente los primers.
  - **Pérdida de información:** Si no se hace correctamente, podrías perder partes importantes de la secuencia.
- Mapear puede quedar más difícil.**



## 2. No sacar a los primers



### Pros:

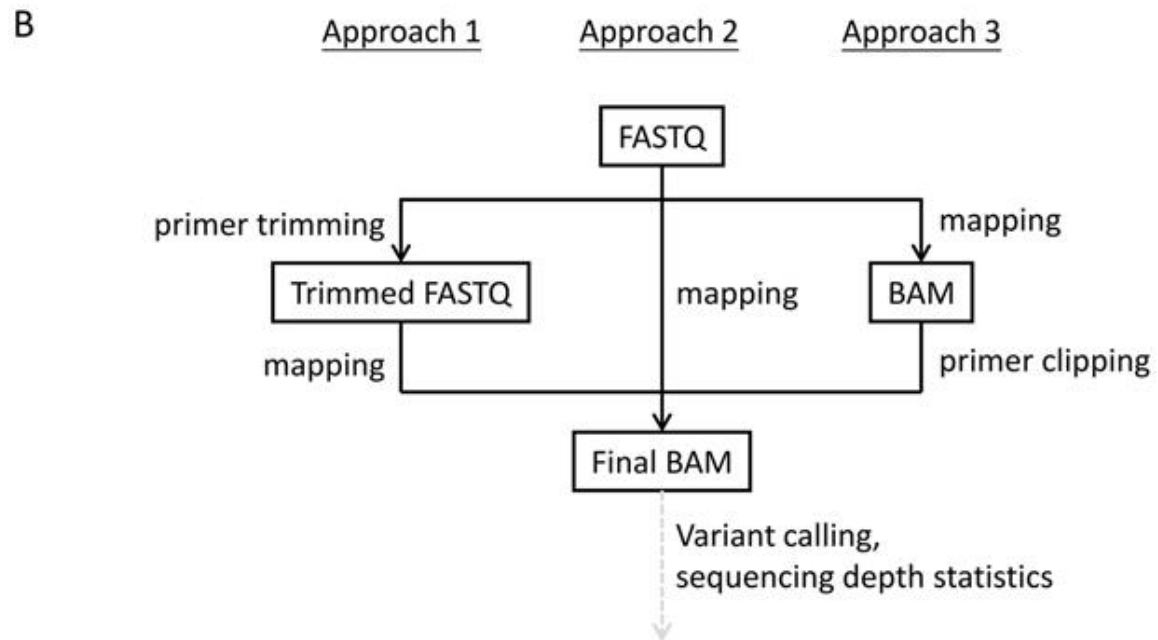
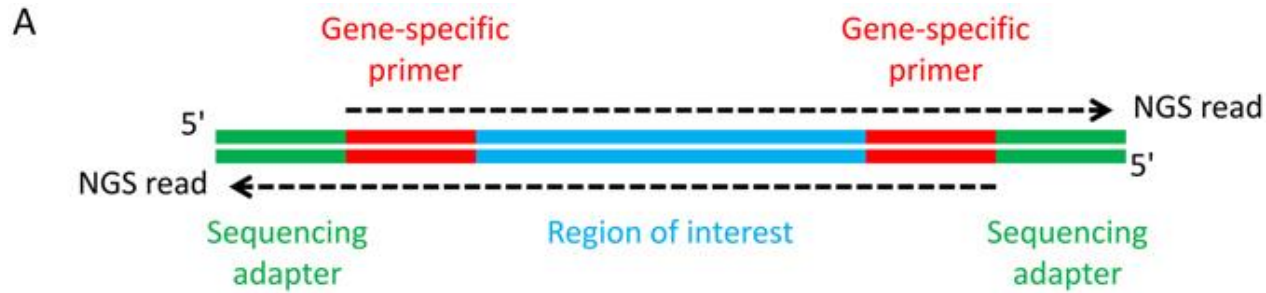
- **Más sencillo:** No necesitas pasos adicionales antes de mapear.
- **Preservación de toda la secuencia:** No hay riesgo de perder información al recortar los primers.

### Contras:

- **Limitado:** Solo logra funcionar bien si sabes cuales son los únicos hotspots de mutación de las secuencias que estudias. Hay que tener altísima cobertura.
- **Potencial de interpretación errónea:** Las secuencias de primers pueden ser confundidas con variantes reales, llevando a resultados incorrectos.



### 3. Sacar a los primers después de mapear:



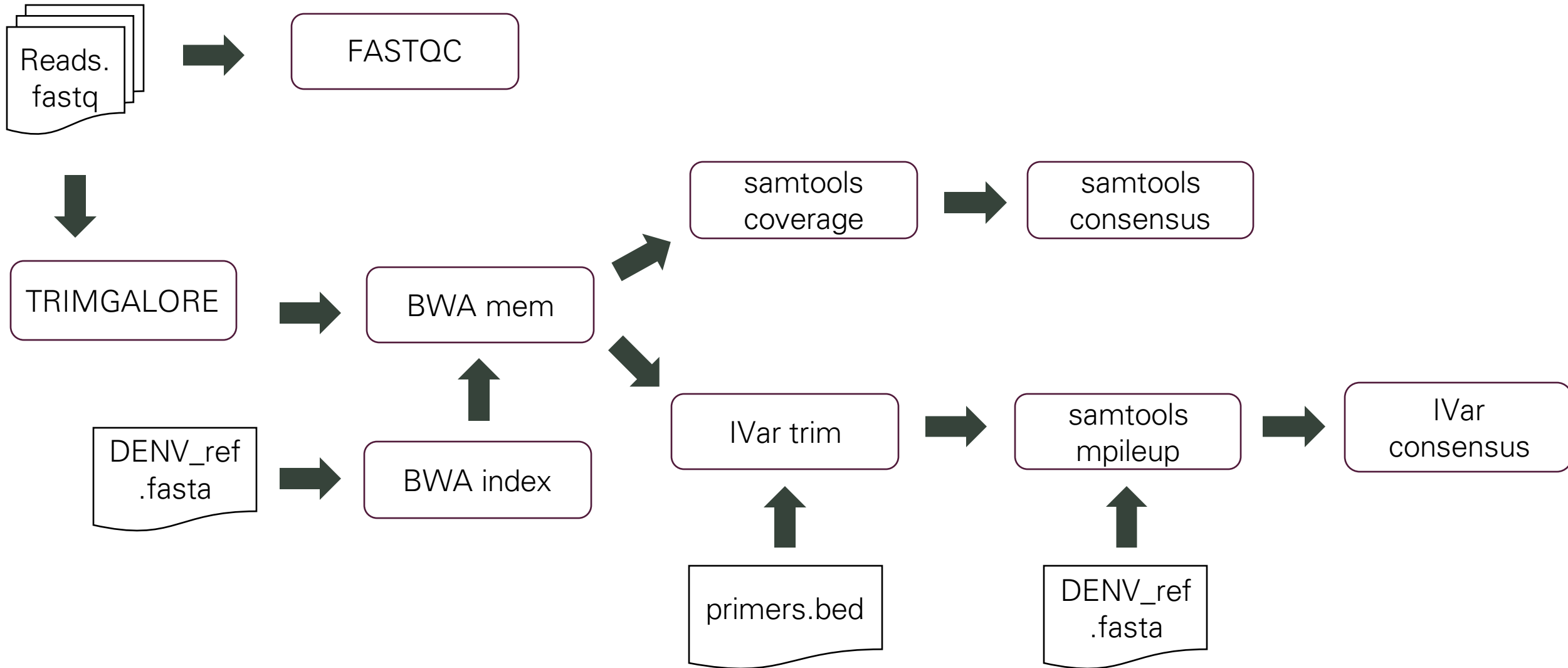
*Pros:*

- **Flexibilidad:** Puedes decidir eliminarlos después de observar cómo afectan las alineaciones.
- **Mapeo inicial completo:** Al mantener los primers, puedes asegurarte de que toda la secuencia se mapee correctamente antes de decidir eliminarlos.

*Contras:*

- **Proceso más largo:** Tendrás que mapear y luego procesar los datos nuevamente para eliminar los primers.
- **Potencial de interpretación errónea:** Al igual que no sacar los primers, existe el riesgo de confundir las secuencias de primers con variantes reales.

# Flujo de trabajo sugerido



index		1 item	19 minutes ago
	DENV_REF.fasta	42.8 KiB	10/28/23 at 11:42 PM
primers		1 item	21 minutes ago
	primers.bed	8.2 KiB	10/28/23 at 11:58 PM
raw		2 items	26 minutes ago
	SRR24053947_R2.fastq.gz	8.2 MiB	10/28/23 at 12:34 AM
	SRR24053947_R1.fastq.gz	8.2 MiB	10/28/23 at 12:34 AM