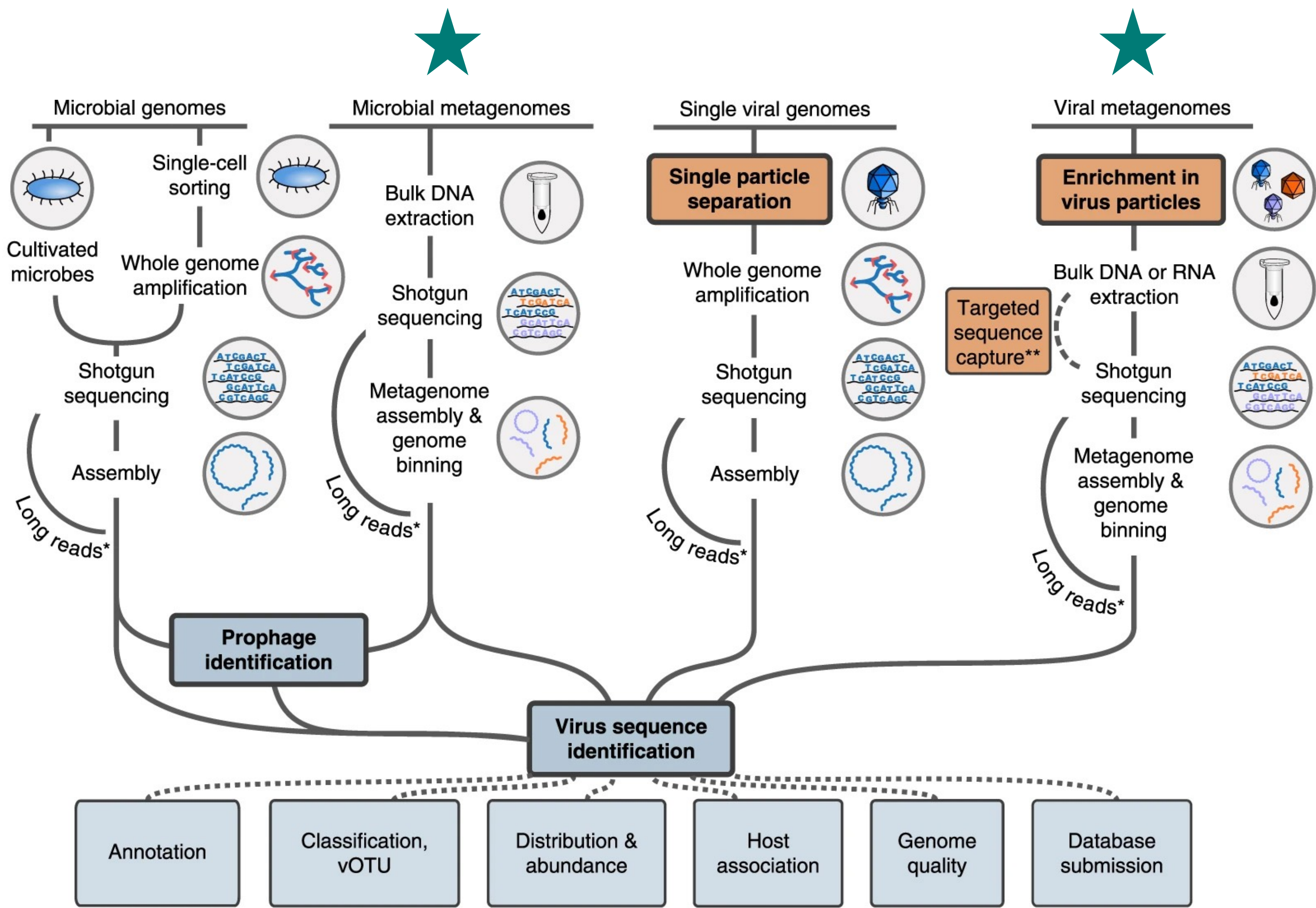


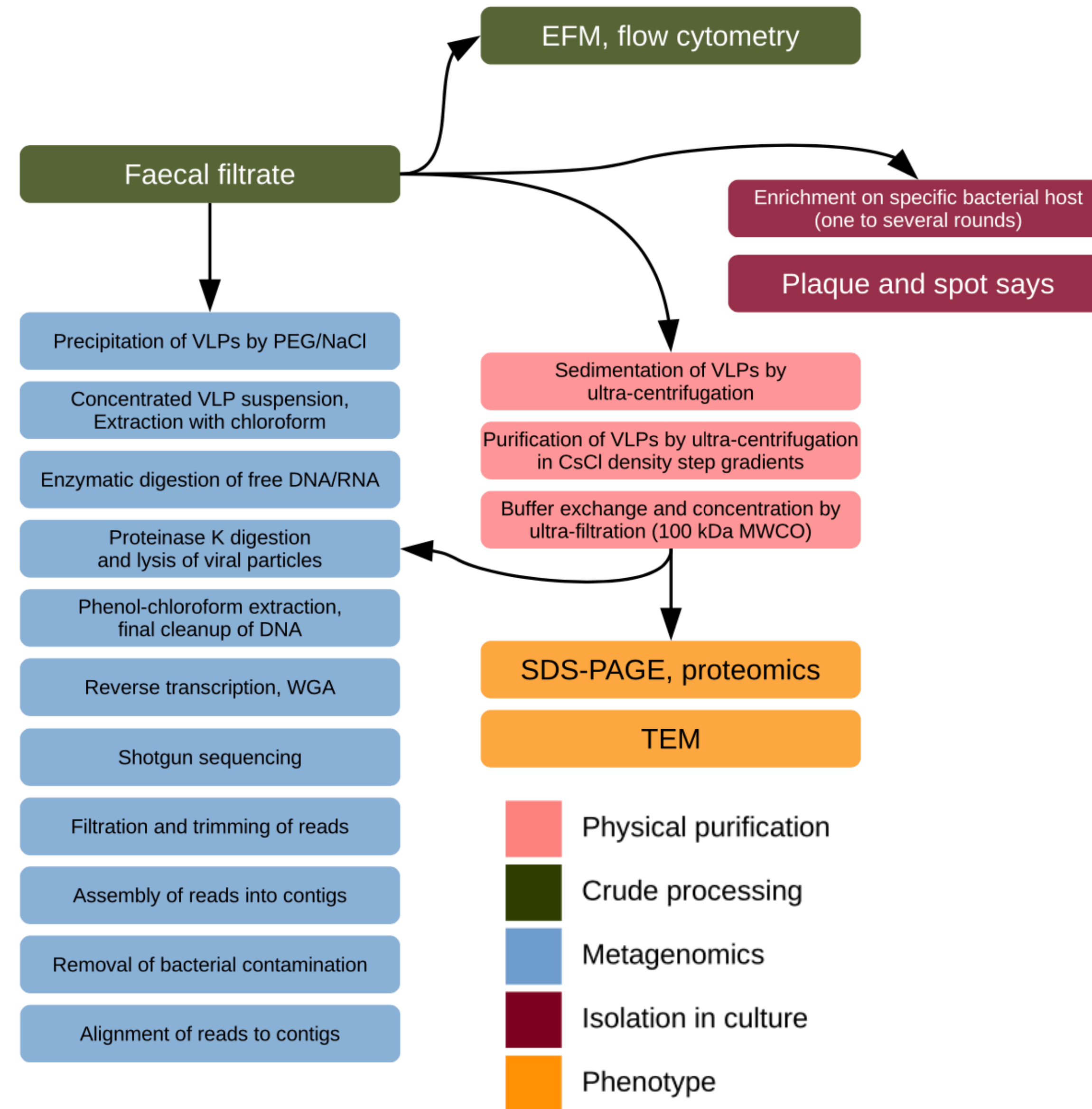
2do Workshop: Bioinformática en Genómica Viral

Día 3: Metagenómica Viral

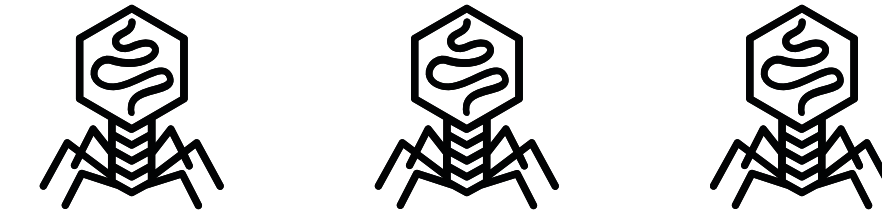
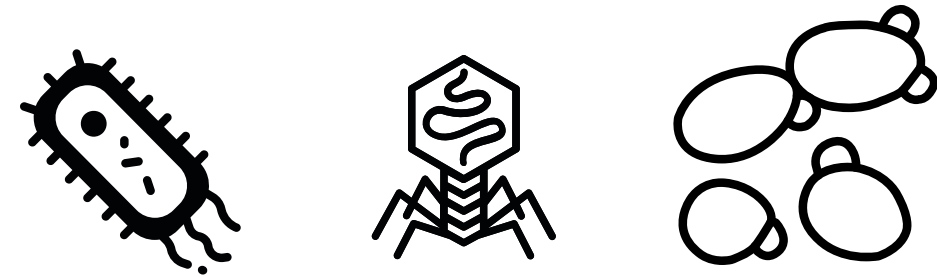
Multiples maneras de identificar genomas de virus no-cultivables





Métodos para enriquecer y analizar el viroma



Bulk metagenomics vs VLP

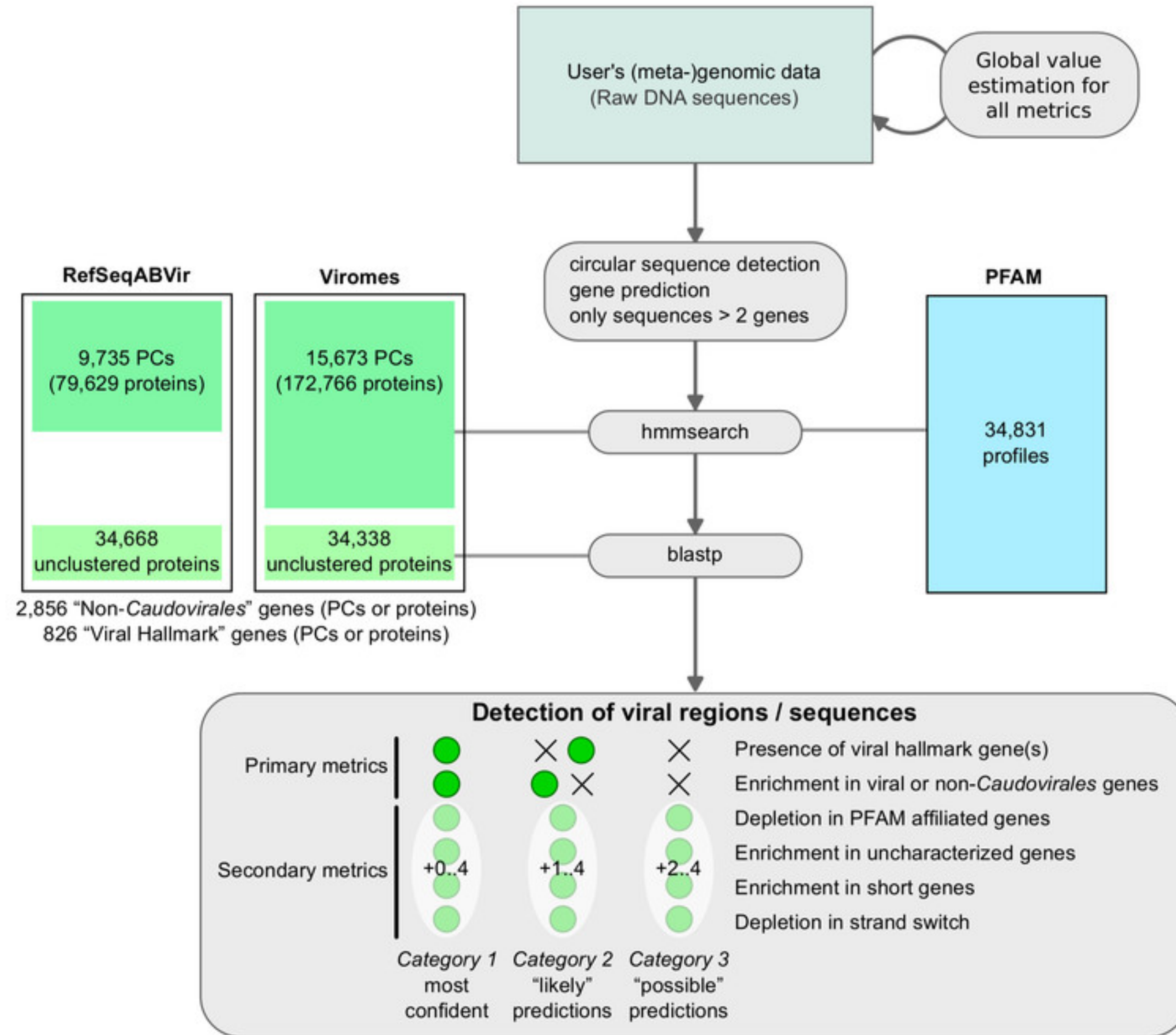


- 
- Muestreo Integral: Nos permite relacionar las diferentes comunidades microbianas dentro del ambiente.
 - Identificación más precisa de profagos y sus hospederos.
 - Menor Especificidad Viral: difícil detectar virus de baja abundancia.
 - Análisis de datos complejo
 - Mayor Ruido: potencial enmascaramiento de señales virales.

- 
- Método más sensible para detectar virus ya que las partículas virales está enriquecidas.
 - Reducción del Ruido de Fondo: Elimina la mayoría del material genético no viral.
 - Dependiendo del protocolo, es posible identificar virus de RNA
 - No hay captura de profagos
 - Alto costo en preparación de librerías

Aproximaciones computacionales para predecir secuencias virales.

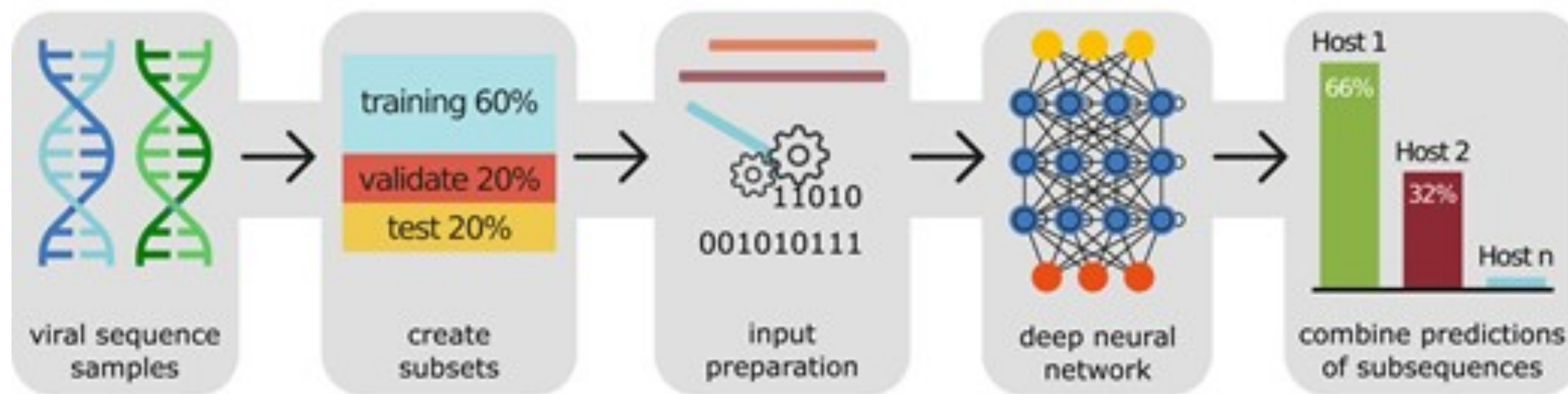
Basados en homologia



- Se usa alineamientos locales y modelos ocultos de Markov (HMM) contra bases de datos curadas.
- Uso de ventanas móviles (sliding window) para identificación de regiones ricas en genes virales.

Aproximaciones computacionales para predecir secuencias virales.

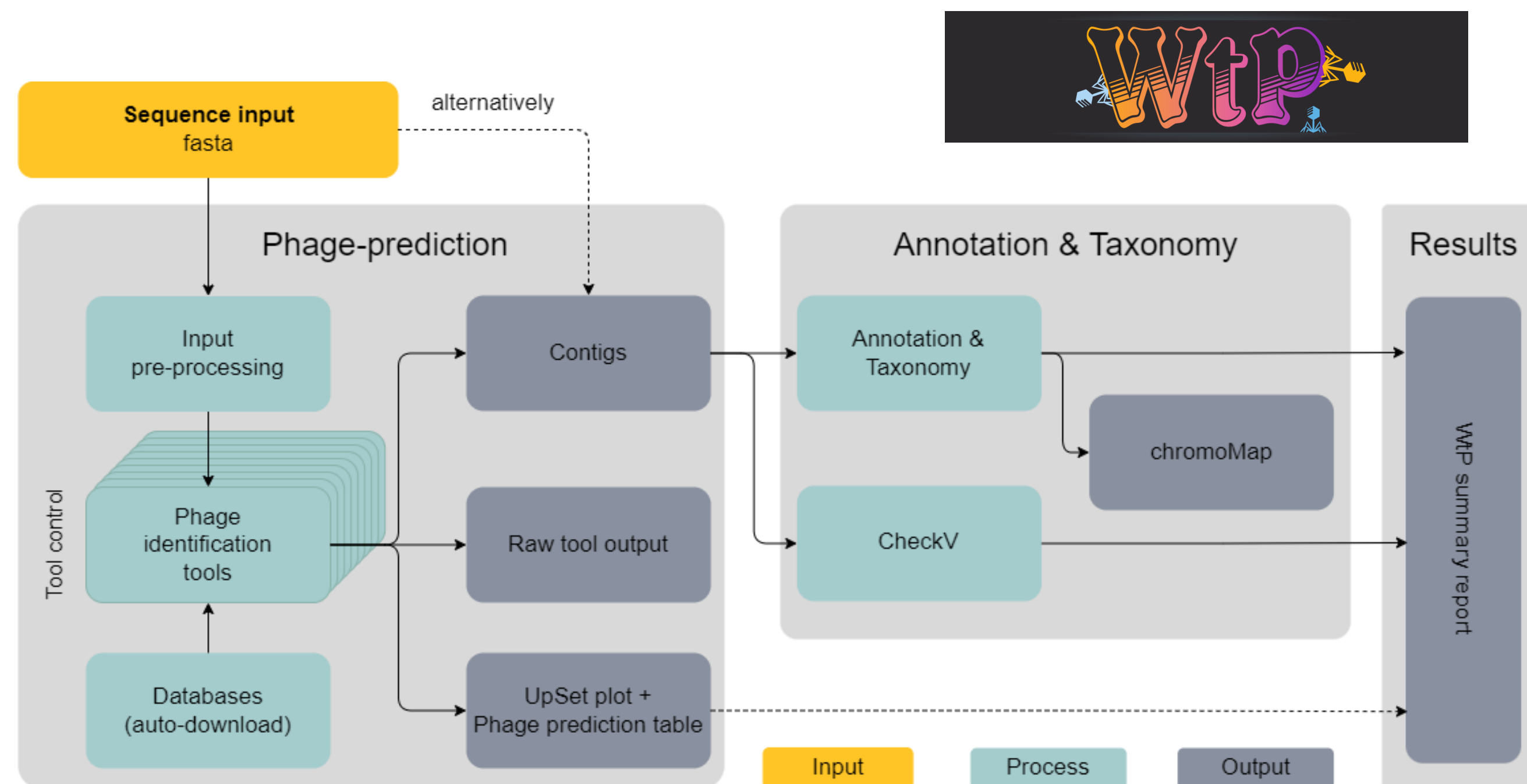
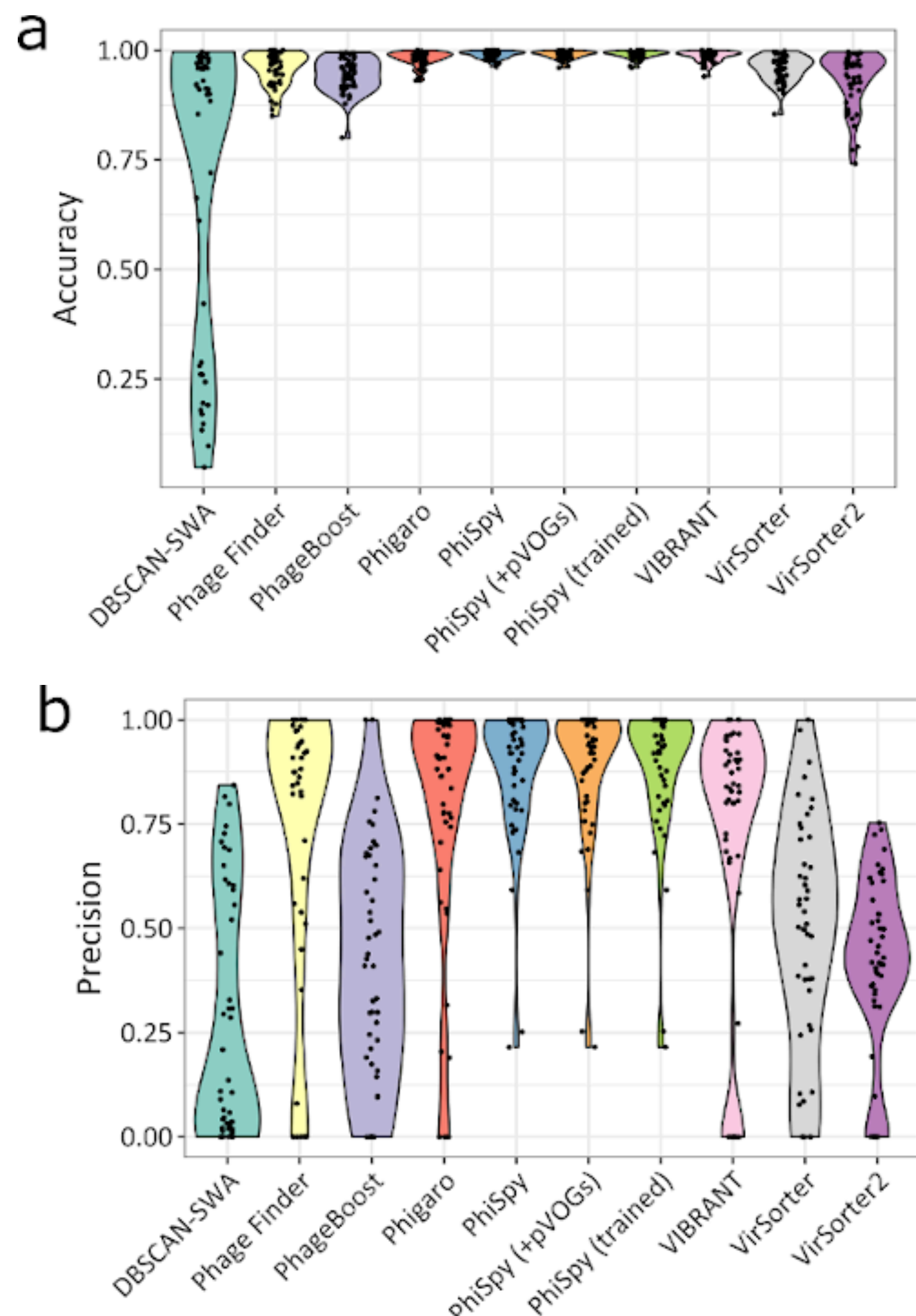
Basados en aprendizaje de máquina



- No hay búsqueda directa en bases de datos, sin embargo es necesario entrenar los modelos con las secuencias previamente identificadas.
- **Datos de Entrada:** Las secuencias de ADN se convierten en un formato que la red neuronal puede entender. Esto puede incluir:
 - k-mers: Sub secuencias de longitud fija
 - Composición de nucleótidos: Porcentaje de A, T, C, G en la secuencia.
 - Marcadores genéticos: Presencia de genes específicos de fagos.
- **Entrenamiento:** La red se entrena con un conjunto de datos etiquetados, donde algunas secuencias son fagos conocidos y otras no.
 - Durante el entrenamiento, la red ajusta sus parámetros (pesos) para distinguir entre secuencias virales y no virales.
- **Predicción:** Una vez entrenada, la red neuronal puede analizar nuevas secuencias y predecir si son fagos o no, basándose en los patrones que aprendió.

Que herramienta usar ?

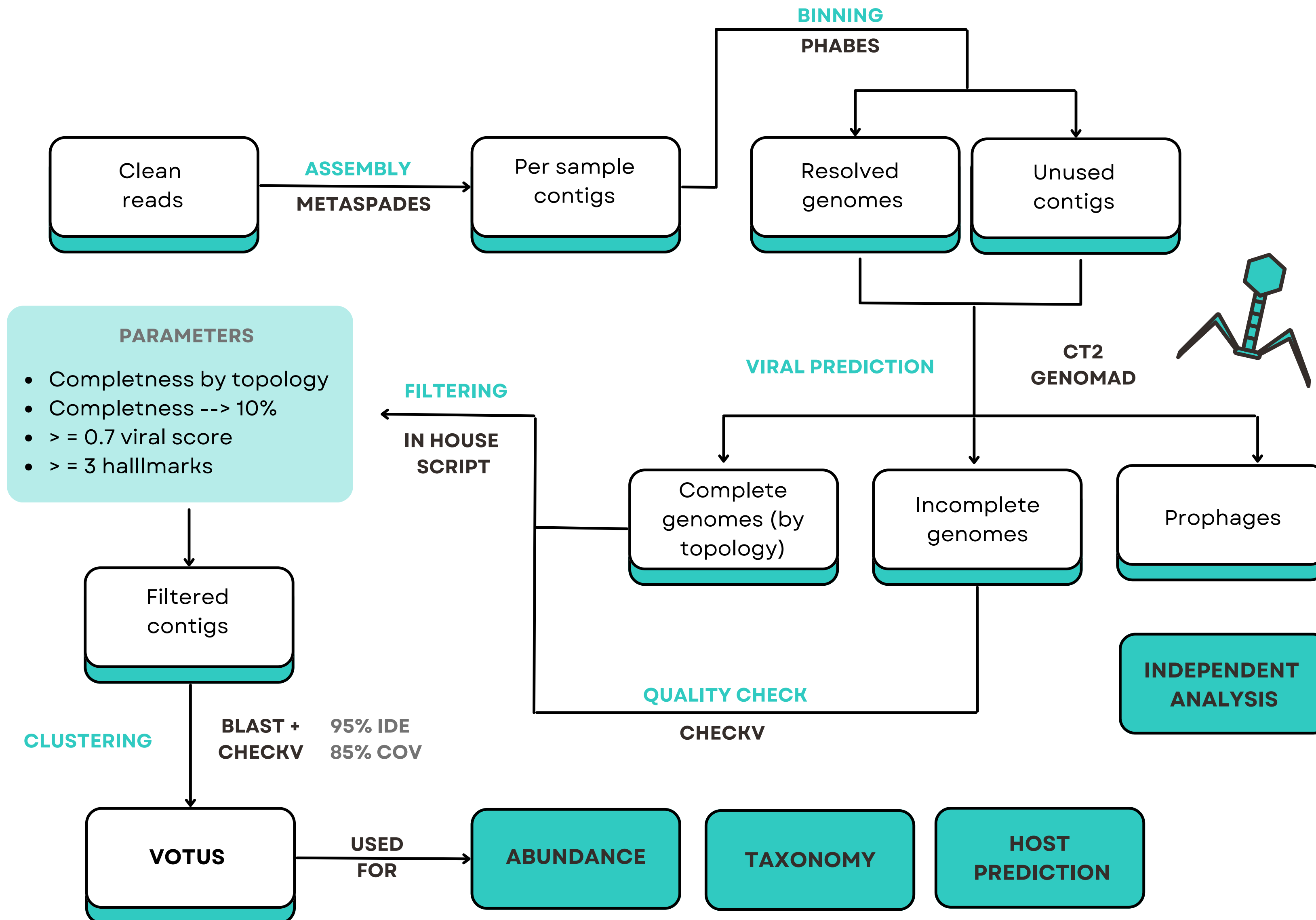
Flujos de trabajo con multiples herramientas como opción



- Dada la variación en precisión de las herramientas, varios flujos de trabajo usan multiples herramientas.
- Los contigs virales son escogidos dependiendo el numero de herramientas que los identifican.

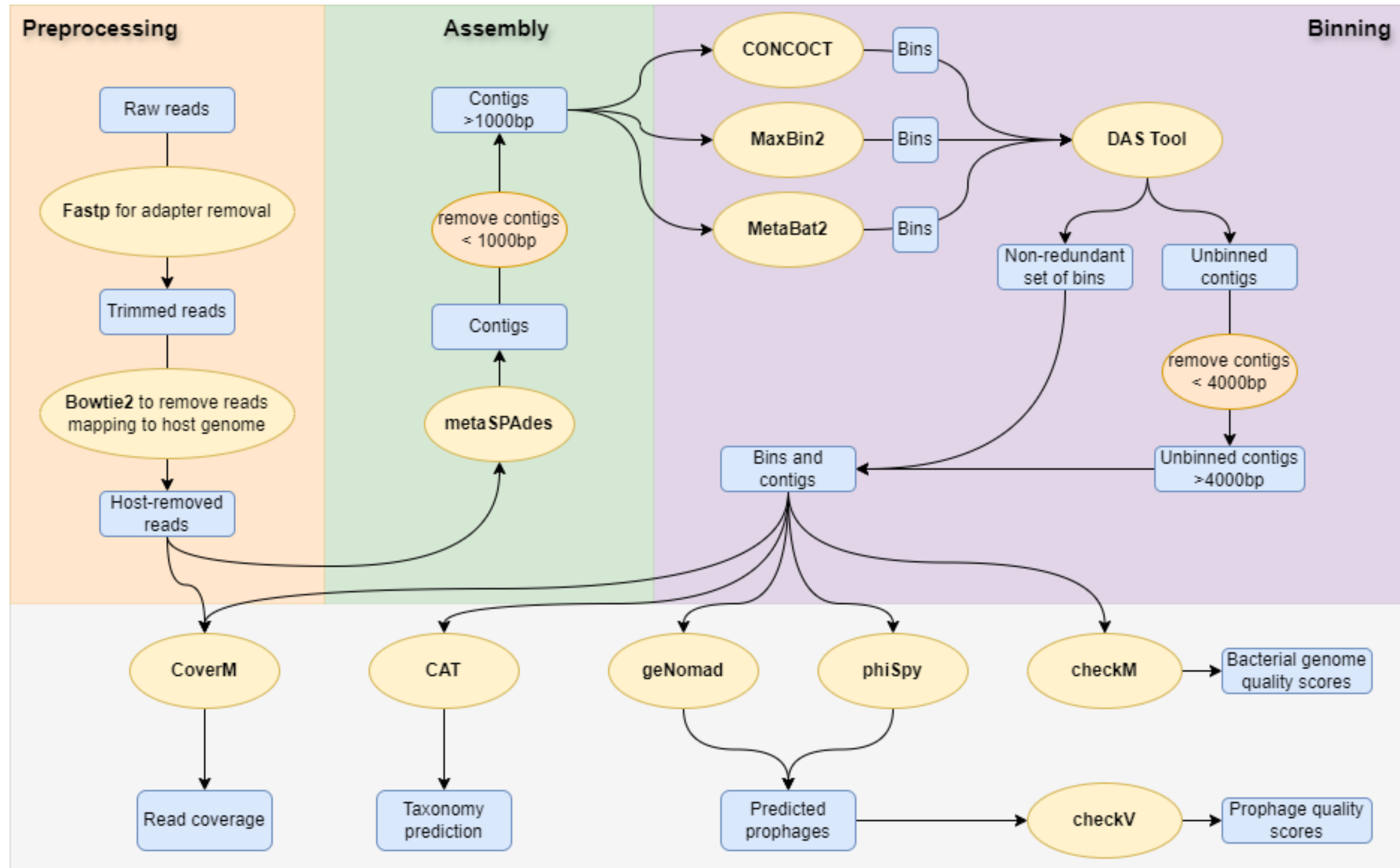
Workflow detallado análisis virales - No existe un goal standard

VLP análisis



Workflow detallado análisis virales - No existe un goal standard

Bulk metagenomic análisis



Hands on

*[https://github.com/BCVI/2do-Workshop-
Genomica-Viral/tree/main](https://github.com/BCVI/2do-Workshop-Genomica-Viral/tree/main)*