



# GASAL2: Alineación de Secuencias Acelerada por GPU

Presentamos GASAL2, una biblioteca GPU para la alineación de secuencias de ADN y ARN que supera a las bibliotecas CPU y GPU existentes. Dada la complejidad computacional de los algoritmos de alineación de secuencias, se han propuesto varias soluciones aceleradas. NVBIO es la única biblioteca GPU disponible para datos NGS de alto rendimiento, pero con rendimiento limitado. GASAL2 ofrece una alternativa superior.



por **Maiev ShadowSong**

# Introducción al Problema: Complejidad de la Alineación de Secuencias

Muchas aplicaciones de procesamiento de datos de secuenciación NGS dependen en gran medida de algoritmos de alineación de secuencias para identificar similitudes entre fragmentos de ADN. Programas como BWA-MEM y Bowtie2, ensambladores de ADN como PCAP y PHRAP, y algoritmos de corrección de errores de lectura como Coral y ECHO, utilizan repetidamente estos algoritmos.

La alineación de secuencias es el proceso de editar dos o más secuencias utilizando brechas y sustituciones para que coincidan estrechamente. Se realiza mediante programación dinámica, con algoritmos de alineación global (Needleman-Wunsch) y local (Smith-Waterman), ambos mejorados por Gotoh para usar penalizaciones de brecha afín.



# Objetivo del Estudio: Superar las Limitaciones Actuales



## Alineación de ADN/ARN

Desarrollar una biblioteca acelerada por GPU para la alineación de secuencias de ADN/ARN, capaz de realizar alineación global, local y semi-global.



## Rendimiento Superior

Superar a las implementaciones SIMD optimizadas para CPU (SeqAn, Parasail) y a la biblioteca GPU de NVIDIA (NVBIO).



## Integración Sencilla

Proporcionar APIs de alto rendimiento que puedan integrarse fácilmente en herramientas bioinformáticas existentes.

# Flujo de Trabajo de GASAL2



## Etapa 1: Empaquetamiento de Datos

Las secuencias se transfieren a la memoria de la GPU y se empaquetan en enteros de 32 bits. Este proceso es hasta 750 veces más rápido que en NVBIO, eliminando el tiempo de empaquetamiento.



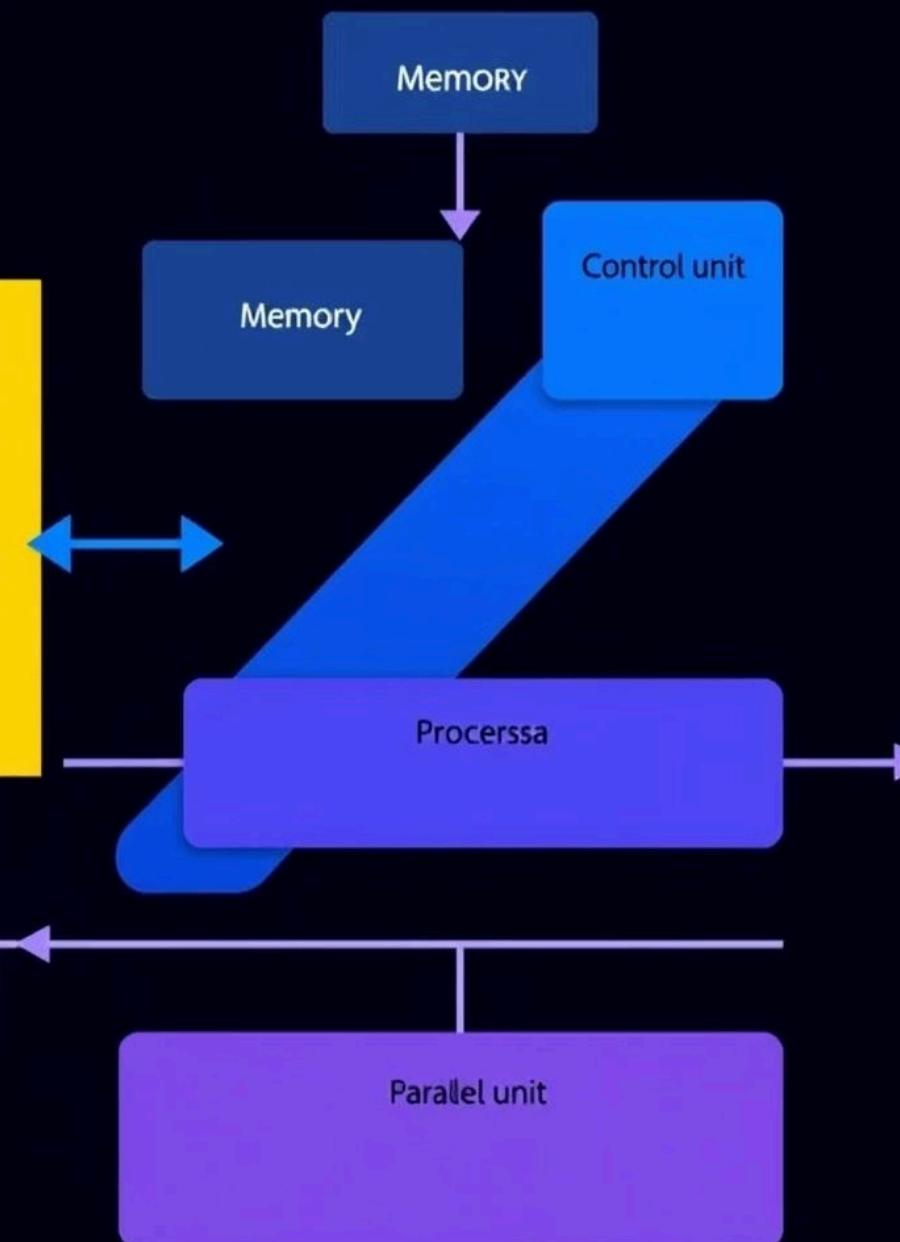
## Etapa 2 (Opcional): Complemento Inverso

Cualquier número de secuencias puede ser invertido y/o complementado en la GPU, aprovechando el alto paralelismo de la tarea.



## Etapa 3: Alineación

El kernel de alineación realiza la alineación por pares utilizando un esquema de puntuación de brecha afín. Cada hilo de la GPU alinea un par de secuencias de forma independiente.



# Arquitectura del Modelo: Optimización de GPU

## Paralelización Inter-Secuencia

GASAL2 asigna un par de secuencias a cada hilo de la GPU, permitiendo que todas las alineaciones se ejecuten en paralelo sin dependencias de datos.

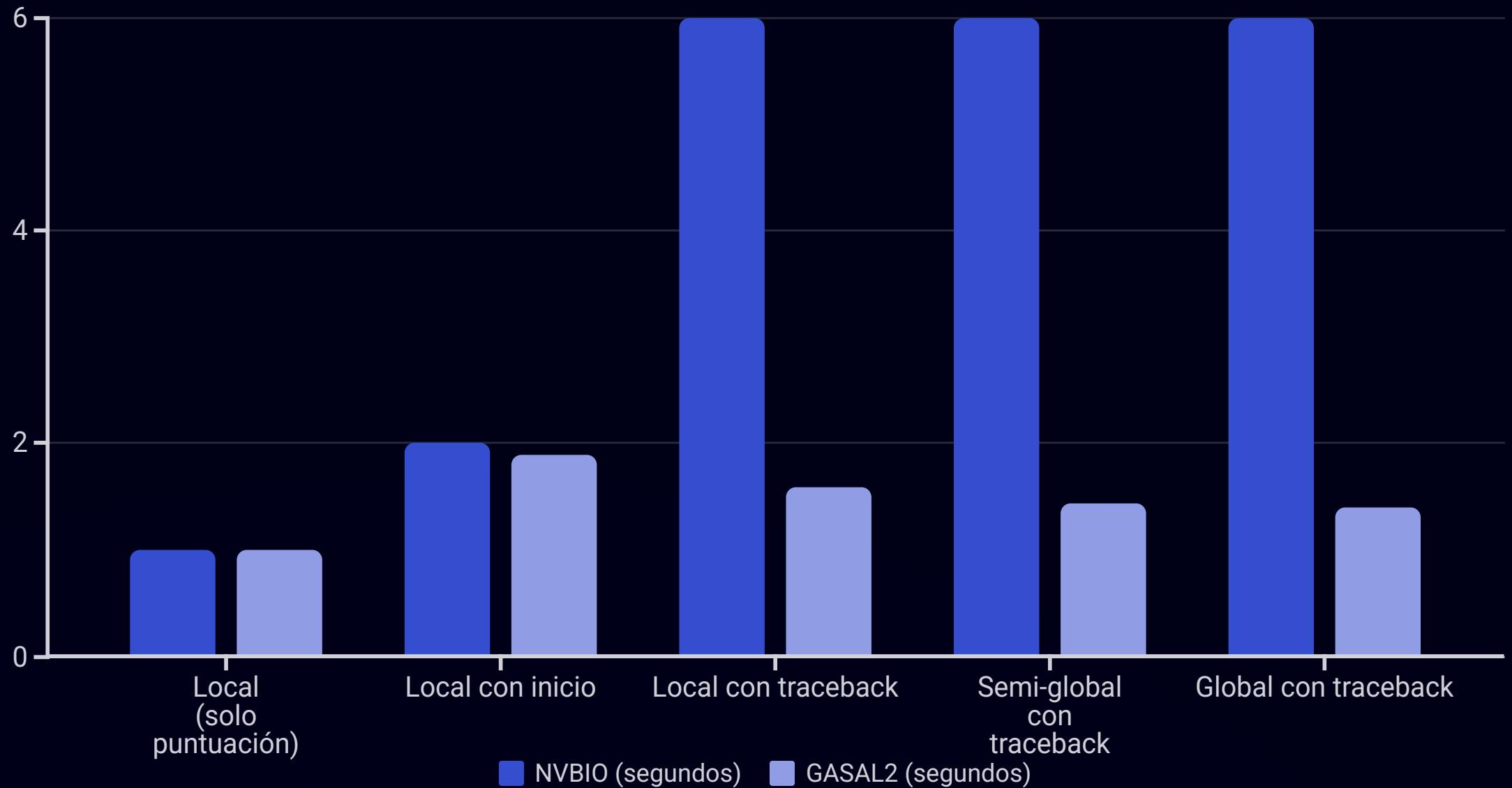
## Procesamiento de Tiles 8x8

Las secuencias se empaquetan en palabras de 32 bits (8 bases), lo que permite calcular un tile de 8x8 de la matriz de programación dinámica con una sola lectura de memoria, reduciendo las solicitudes de memoria.

## Manejo de Memoria Eficiente

En lugar de almacenar toda la matriz, solo se guarda una columna de 8 elementos y una fila completa, reduciendo los requisitos de memoria de  $O(n^2)$  a  $O(n)$ .

# Resultados Principales: Rendimiento Superior



Los resultados muestran que los tiempos de ejecución del kernel de alineación de NVBIO y GASAL2 son casi idénticos para "solo puntuación" y "con inicio". Sin embargo, con la computación de traceback, los kernels de GPU de GASAL2 son aproximadamente 4 veces más rápidos que NVBIO. Esto se debe a la eficiencia del enfoque de GASAL2 para reducir las celdas de DP necesarias para el inicio.

# Comparativas: GASAL2 vs. Bibliotecas Existentes

21x

Más rápido que Parasail en GPU (GTX 1080 Ti) para longitudes de consulta de hasta 300 bases.

13x

Más rápido que NVBIO para longitudes de consulta de hasta 100 bases.

750x

Más rápido en el empaquetamiento de secuencias en GPU que NVBIO.

GASAL2 supera a las implementaciones SIMD optimizadas para CPU como SeqAn y Parasail, así como a la biblioteca basada en GPU de NVIDIA, NVBIO. La capacidad de GASAL2 para realizar el empaquetamiento de secuencias en la GPU es única y contribuye significativamente a su rendimiento superior.

# Caso de Uso: Aceleración de BWA-MEM

## BWA-MEM Original

El hilo de la CPU permanece inactivo mientras la extensión se realiza en la GPU, lo que lleva a una subutilización.



## GASAL2 (Acelerado)

Funciones de alineación asíncronas permiten superponer la ejecución de CPU y GPU, eliminando el tiempo de extensión.

## GASAL (Modificado)

Mejoras en el manejo de bases 'N' y asignación de memoria, pero aún con tiempos de extensión más lentos.

GASAL2 acelera BWA-MEM, un mapeador de lectura de ADN. La alineación local se acelera 20 veces, lo que resulta en una mejora general de la aplicación de 1.3 veces en comparación con la CPU con hasta 12 hilos. La ejecución asíncrona de GASAL2 permite una superposición completa de la ejecución de CPU y GPU, optimizando el uso de recursos.

# Ventajas Clave de GASAL2 y Conclusiones



## Alto Rendimiento

Hasta 21 veces más rápido que Parasail y 13 veces más rápido que NVBIO.



## Ejecución Asíncrona

Permite la superposición completa de la ejecución de CPU y GPU, eliminando el tiempo de inactividad.



## Versatilidad

Soporta alineación local, global y semi-global, con y sin traceback, y empaquetamiento en GPU.



## Fácil Integración

APIs de alto rendimiento que facilitan su uso en diversas herramientas bioinformáticas.

GASAL2 es una biblioteca de alineación de secuencias de ADN y ARN acelerada por GPU de alto rendimiento. Sus capacidades únicas, como el empaquetamiento de secuencias en GPU y la ejecución asíncrona, la distinguen de las bibliotecas anteriores. Los resultados experimentales confirman su superioridad en velocidad y eficiencia, convirtiéndola en una opción ideal para el procesamiento de datos NGS de alto rendimiento.