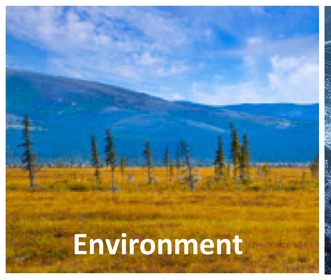
Development of GPU accelerated Smith-Waterman kernel for metagenomics workflow

Muaaz G. Awan, NERSC

Microbiomes are Critical to Energy and the Environment



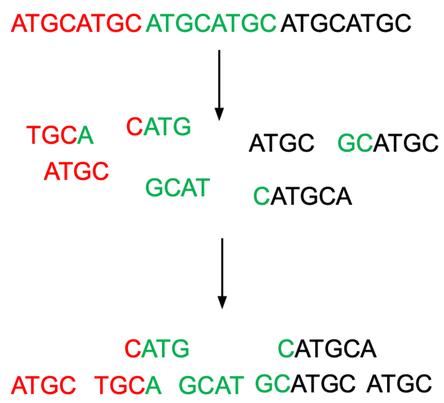
- **Microbes:** single cell organisms, such as bacteria and viruses
- **Microbiomes:** communities of 1000s of microbial species, less than 1% individually culturable in a lab (and thus sequenced)
- **Metagenomics:** genome sequencing on these communities (growing exponentially)

Genome Assembly Primer

Genome of an organism is sampled using sequencing machines in the form of reads.

Objective is to assemble the pieces of genome back into their correct places *denovo.*

Like solving a puzzle, without a reference!

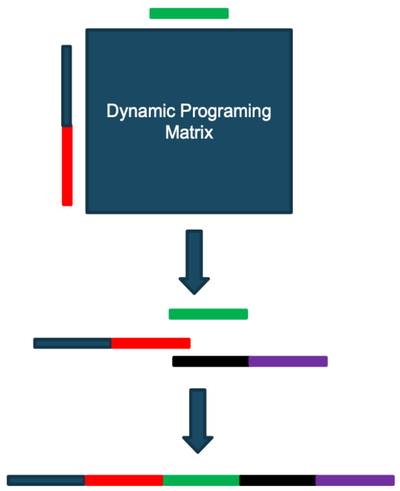


Smith-Waterman, core of assembly.

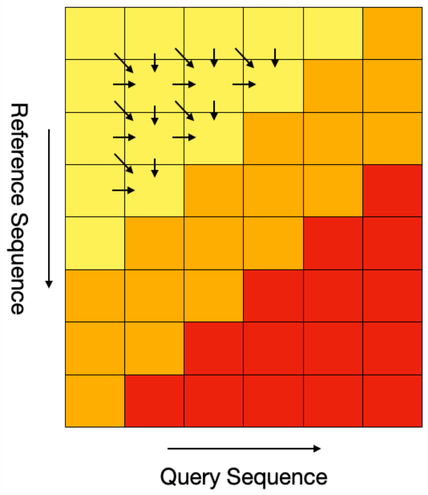
At the core of most assembly frameworks Smith-Waterman algorithm aligns sequences.

This alignment information is used to stitch together different overlapping parts of the genome.

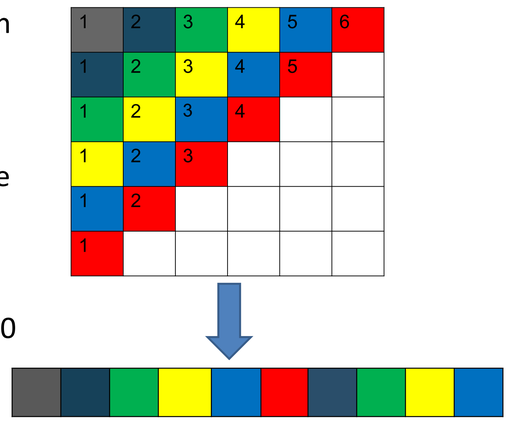
This information can be used to assemble a complete genome.



Cell dependencies and memory access pattern

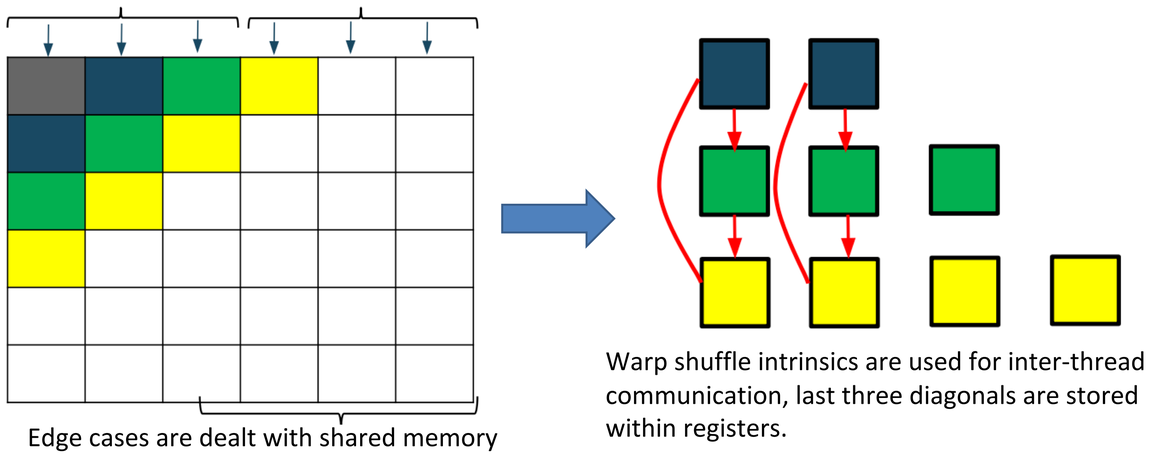


- Each cell depends on its top, top diagonal and cell on the left.
- This restricts parallelism along the minor-diagonal.
- This results in each thread accessing a memory location 200 bytes apart



Register shuffle for inter-thread communication

Inter-warp communications go through shared memory

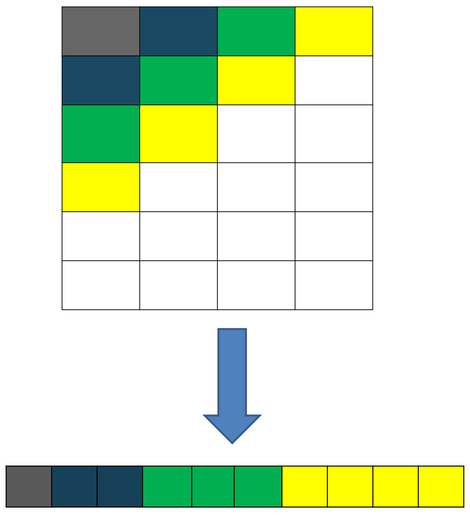


Warp shuffle intrinsics are used for inter-thread communication, last three diagonals are stored within registers.

Diagonal Major Indexing

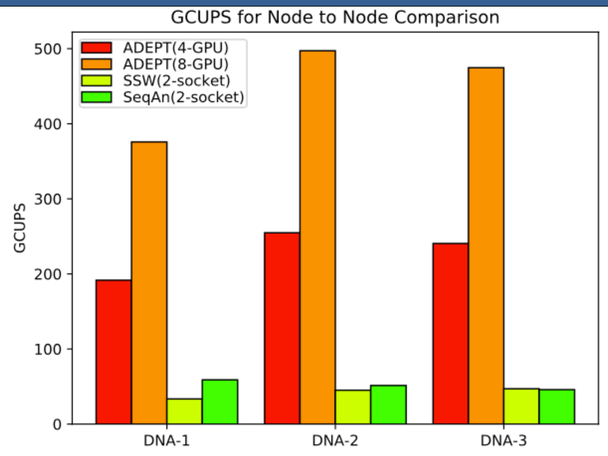
For traceback phase, complete matrix needs to be stored.

Matrix is re-organized such that anti-diagonal elements occupy contiguous memory locations.



Performance

- Cori GPU (V100) Nodes were used for GPU runs.
- Cori Haswell Nodes were used for CPU runs.
- GCUPs = $Total\ Cells / Time$



- DNA-1 -> 32.2 mil. alns.
- DNA-2 -> 8.8 mi. alns.
- DNA-3 -> 16.1 mil. Alns.

Portability

- CUDA Kernel was ported to HIP using hipify-perl.
- Shuffle intrinsics and error checks had to be hand converted.
- No compromise on performance was observed.
- rocm 3.3.0 was used for these runs