

## We accelerated the OpenCL backend of pyPaSWAS sequence aligner

```
:EDRATWKSNYFLKIIQLDDYPKCFIVGADNVGSKOMQIRMSLRGK-AVY
:EDRATWKSNYFLKIIQLDDYPKCFIVGADNVGSKOMQIRLSLRGK-AVY
:EDRATWKSNYFLKIIQLNDYPKCFIVGADNVGSKOMQIRLSLRGK-AIV
:ENKAAWKAQYFIKVVLELDFEFPKCFIVGADNVGSKOMONIRTSLRGL-AVY
:AG-SKRKKLFIEKATKLFTTYDKMIVAEADDFVGSQLOKIRKSIRGI-GAV
:AG-SKRKNVFIKATKLFTTYDKMIVAEADDFVGSQLOKIRKSIRGI-GAV
:LSKQKQKQMYIEKLSLQQYSKILLVHVDNVGNOMASVRKSLRGK-ATD
:KKIARWVDEVAELTEKIKTHKTIIANIEGFPADKLHEIRKKLRGK-ADJ
:RKIARWVDEVAELTEKIKTHKTIIANIEGFPADKLHEIRKKLRGK-ADJ
:RKVASWKKLEEVKELTELKNSNTILIGNLEGFPADKLHEIRKKLRGK-ATD
:KPIPEWKTMLRELEELFSKHRVVLADLTGTPFVVRVRKKLWKK-YPN
:RQYPARKVKIVSEATELLQKYPYVFLDLHGLSSRILHEYRRLRRY-GVJ
:EHIPQWKKDEIENIKELIQSHKVFQMVGIEGILATKMKIRRDLDV-AVI
:EHIPQWKKDEIENIKELIQSHKVFQMVRIEGILATKMKIRRDLDV-AVI
:--PEEYKVRAVEEIKRMISSKPVAIVSFRNVPAGOMKIRREFRGK-AEI
:PKVAEWKREEVKELTEKIKTHKTIIANIEGFPADKLHEIRKKLRGK-ADJ
:IAHVAEWKKEVQELHDLIKGYEVVGIANLADIPAROLOKMRQTLRDS-ALJ
:HKIAPWKIEEVNKLKELKNGQIVALVDMMEVPPAROLOEIRDKIR-GTMTI
:HKIAPWKIEEVNALKELKLSANVIALIDMMEVPPAROLOEIRDKIR-DQMTI
:AHVAEWKIEEVKTEKELIKSKPVAIVDMMDVPPAROLOEIRDKIR-DKVKI
:IAHVAEWKKEVEELANLIKSYVIALVDVSSMPAYPLSOMRRLIRENGGLI
```

Last year we accelerated the OpenCL-code in PaSWAS, which is open source software to do DNA/RNA/protein sequence alignment and trimming. It has users world-wide in universities, research groups and industry.

Below you'll find the benchmark results of our acceleration work. You can also test out yourself, as the code is [public](#). In the readme-file you can learn more about the idea of the software. Lots of background information is described in these two papers:

[Flexible, Fast and Accurate Sequence Alignment Profiling on GPGPU with PaSWAS](#) | [pyPaSWAS: Python-based multi-core CPU and GPU sequence alignment](#)

We chose PaSWAS because we really like bio-informatics and computational chemistry - the science is interesting, the problems are complex and the potential GPU-speedup is real. Other examples of such software we worked on are [GROMACS](#) and [TeraChem](#).

The efforts Stream HPC put into accelerating pyPaSWAS are very much appreciated. The speed-up has a direct effect on our ability to run even larger data sets. Also, Stream HPC shows that having computer science engineers on board life science projects is of great importance. Sven Warris - author of PyPaSWAS

### Benchmarks]

A number of different datasets were used for the query and database. The comparison is between the performance of the original code and our branch. The scoring matrix used is **BLOSUM80**, with all other parameters set to default. **NOTE: The results below are for the non-affine gap method.**

An example of how a benchmark is run, is: `time python pypaswas.py --loglevel=info -M BLOSUM80 data/canisLupusAnkyrinPRED.fasta ../canisLupusFamiliaris.faa_4284.faa --device_type=GPU --platform_name=AMD --framework=OpenCL --device=1`

You can change the variables as we put in the dataset description.

**Dataset 1] Query file:** canisLupusAnkyrin.fasta **Target file:** canisLupusAnkyrinPRED.fasta **Number of queries:** 8 **Number of targets:** 346

**OpenCL Device**OriginalStream SpeedupTimeTimeGCUPSTimeGCUPS  
Dual Intel Xeon CPU E5-2650 v3 @ 2.30GHz  
m48.244s0.03XXAMD Firepro s9300 (only 1 GPU used)0m16.648s0.090m4.841s0.475.2x  
AMD Radeon Instinct MI250m11.464s0.130m3.938s0.453.4x  
NVIDIA GeForce GTX9800m19.943s0.070m4.131s0.385.4x

**Dataset 2] Query file:** canisLupusAnkyrin.fasta**Target file:** canisLupusFamiliaris.faa\_4284.faa**Number of queries:** 8**Number of targets:** 4284

**OpenCL Device**OriginalStreamSpeedupTimeTimeGCUPSTimeGCUPS  
Dual Intel Xeon CPU E5-2650 v3 @ 2.30GHz  
m16.338s0.03XXAMD Firepro s9300 (only 1 GPU used)1m30.050s0.120m9.644s1.4111.7x  
AMD Radeon Instinct MI250m55.800s0.190m9.403s1.246.5x  
NVIDIA GeForce GTX9802m19.006s0.070m13.478s0.8111.5x

**Dataset 3] Query file:** canisLupusAnkyrinPRED.fasta**Target file:** canisLupusFamiliaris.faa\_4284.faa**Number of queries:** 346  
**Number of targets:** 4284

**OpenCL Device**originalStreamSpeedupTimeTimeGCUPSTimeGCUPS  
Dual Intel Xeon CPU E5-2650 v3 @ 2.30GHz  
m95m40.606s0.11XXAMD Firepro s9300 (only 1 GPU used)16m41.019s0.632m11.045s4.907.7x  
AMD Radeon Instinct MI2510m12.648s1.032m8.425s4.964.8x  
NVIDIA GeForce GTX98039m13.160s0.273m21.758s3.1511.6x

## Thanks

Big thanks go to Sven Warris, the author of the software and first author of the papers, who supported us with papers and background info to make this happen.

If you want to use the software for your own research, download a release at <https://github.com/swarris/pyPaSWAS/releases>