



Running CAW with AWS Batch

Hopefully my last talk about CAW ever...

Maxime Garcia

@gau

@MaxUlysse

<https://maxulysse.github.io/>

2018-01-31

SciLifeLab NGI / BarnTumörBanken

SciLifeLab

NATIONAL ATCA GENOMICS INFRASTRUCTURE

NBIS
NATIONAL BIOINFORMATICS INFRASTRUCTURE SWEDEN

Karolinska Institutet

KTH
KONSTALLEG
ROYAL INSTITUTE OF TECHNOLOGY

Stockholms
universitet

UPPSALA
UNIVERSITET



Definition

A screenshot of a web browser displaying Google search results for the query "CAW". The search bar at the top contains "CAW". Below the search bar, the Google logo is visible, followed by the word "CAW" and a magnifying glass icon. A horizontal navigation bar includes "All" (which is underlined), Images, Videos, News, Maps, More, Settings, and Tools. A message indicates "About 64 400 000 results (0,55 seconds)". On the left, a sidebar titled "Dictionary" shows the word "caw" with its phonetic transcription "/kɔ:/". It defines "caw" as a noun, describing it as the harsh cry of a rook, crow, or similar bird. It also defines "caw" as a verb, meaning to utter a caw. An example sentence is provided: "rooks cawed in the dark trees". At the bottom of the sidebar, there is a link "Translations, word origin, and more definitions" with a downward arrow icon. A "Feedback" link is located at the bottom right of the sidebar.

What is CAW?



🌐 <http://opensource.scilifelab.se/projects/caw/>

- Nextflow pipeline

What is CAW?



🌐 <http://opensource.scilifelab.se/projects/caw/>

- Nextflow pipeline
- Developed at NGI



What is CAW?



🌐 <http://opensource.scilifelab.se/projects/caw/>

- Nextflow pipeline
- Developed at NGI
- In collaboration with NBIS



What is CAW?



🌐 <http://opensource.scilifelab.se/projects/caw/>

- Nextflow pipeline
- Developed at NGI
- In collaboration with NBIS
- Support of The Swedish Pediatric Tumor Biobank



What does CAW do?



What does CAW do?



- WGS analysis (Tumor/Normal pair or Germline)

What does CAW do?



- WGS analysis (Tumor/Normal pair or Germline)
- Handles both GRCh37 and GRCh38

What does CAW do?



- WGS analysis (Tumor/Normal pair or Germline)
- Handles both GRCh37 and GRCh38
- Based on GATK best practices for processing FASTQ files

What does CAW do?



- WGS analysis (Tumor/Normal pair or Germline)
- Handles both GRCh37 and GRCh38
- Based on GATK best practices for processing FASTQ files
- SNPs, SNVs and indels

What does CAW do?



- WGS analysis (Tumor/Normal pair or Germline)
- Handles both GRCh37 and GRCh38
- Based on GATK best practices for processing FASTQ files
- SNPs, SNVs and indels
 - MuTect1, MuTect2, Strelka, and GATK HaplotypeCaller

What does CAW do?



- WGS analysis (Tumor/Normal pair or Germline)
- Handles both GRCh37 and GRCh38
- Based on GATK best practices for processing FASTQ files
- SNPs, SNVs and indels
 - MuTect1, MuTect2, Strelka, and GATK HaplotypeCaller
- Structural variants

What does CAW do?



- WGS analysis (Tumor/Normal pair or Germline)
- Handles both GRCh37 and GRCh38
- Based on GATK best practices for processing FASTQ files
- SNPs, SNVs and indels
 - MuTect1, MuTect2, Strelka, and GATK HaplotypeCaller
- Structural variants
 - Manta

What does CAW do?



- WGS analysis (Tumor/Normal pair or Germline)
- Handles both GRCh37 and GRCh38
- Based on GATK best practices for processing FASTQ files
- SNPs, SNVs and indels
 - MuTect1, MuTect2, Strelka, and GATK HaplotypeCaller
- Structural variants
 - Manta
- Heterogeneity, ploidy and CNVs

What does CAW do?



- WGS analysis (Tumor/Normal pair or Germline)
- Handles both GRCh37 and GRCh38
- Based on GATK best practices for processing FASTQ files
- SNPs, SNVs and indels
 - MuTect1, MuTect2, Strelka, and GATK HaplotypeCaller
- Structural variants
 - Manta
- Heterogeneity, ploidy and CNVs
 - ASCAT

What does CAW do?



- WGS analysis (Tumor/Normal pair or Germline)
- Handles both GRCh37 and GRCh38
- Based on GATK best practices for processing FASTQ files
- SNPs, SNVs and indels
 - MuTect1, MuTect2, Strelka, and GATK HaplotypeCaller
- Structural variants
 - Manta
- Heterogeneity, ploidy and CNVs
 - ASCAT
- Containers (portable, reproducible)

What does CAW do?



- WGS analysis (Tumor/Normal pair or Germline)
- Handles both GRCh37 and GRCh38
- Based on GATK best practices for processing FASTQ files
- SNPs, SNVs and indels
 - MuTect1, MuTect2, Strelka, and GATK HaplotypeCaller
- Structural variants
 - Manta
- Heterogeneity, ploidy and CNVs
 - ASCAT
- Containers (portable, reproducible)
 - Docker or Singularity

Where to use CAW?

- Any POSIX compatible system

Where to use CAW?



- Rackham

Where to use CAW?



- Rackham
- Bianca

Where to use CAW?



- Rackham
- Bianca
- Irma

Where to use CAW?



- Rackham
- Bianca
- Irma

Where to use CAW?



- AWS Batch

Where to use CAW?



CAW with AWS Batch



CAW with AWS Batch



- A single command line

```
nextflow run main.nf -profile awsbatch -w s3://caw-test-results/work \
--genome smallGRCh37 --sample s3://caw-test-data/tsv/tiny-s3.tsv \
--outDir s3://caw-test-results/Results
```

CAW with AWS Batch



- A single command line

```
nextflow run main.nf -profile awsbatch -w s3://caw-test-results/work \
--genome smallGRCh37 --sample s3://caw-test-data/tsv/tiny-s3.tsv \
--outDir s3://caw-test-results/Results
```

🌐 <https://maxulysse.github.io/2017/11/16/Running-CAW-with-AWS-Batch/>

Going further



- Run a full size test sample

Going further



- Run a full size test sample
- Gather reports

Going further



- Run a full size test sample
- Gather reports
- Get a pricing

Going further



- Run a full size test sample
- Gather reports
- Get a pricing
- Wait for Amazon to finally come to Stockholm

The List of People Involved

Sebastian DiLorenzo	Markus Mayrhofer
Jesper Eisfeldt	Monica Nistèr
Phil Ewels	Björn Nystedt
Maxime Garcia	Pall Olason
Szilveszter Juhos	Markus Ringnér
Max Käller	Pelin Sahlén
Malin Larsson	Johanna Sandgren
Marcel Martin	Teresita Díaz De Ståhl

Where to find us?

- We are on the SciLifeLab Slack 
#cancer-pipeline

Where to find us?

- We are on the SciLifeLab Slack 
`#cancer-pipeline`
- We have a gitter channel
<https://gitter.im/SciLifeLab/CAW>

Where to find us?

- We are on the SciLifeLab Slack 
`#cancer-pipeline`
- We have a gitter channel
<https://gitter.im/SciLifeLab/CAW>
- Our code is hosted on Github 
<https://github.com/SciLifeLab/CAW>

Any questions?

