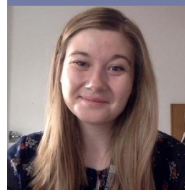


In This Edition

2 Pages

- **CASE STUDY:** Using Nimbus for Oxford Nanopore Sequence Basecalling and Genome Assemblies
- **TECHNICAL GUIDE:** Running Singularity CPU Applications using Slurm on Nimbus
- HPC News and Updates
- Tip of the Month



Sarah Cameron (PhD Researcher, Department of Life Sciences)
Email: sc3445@bath.ac.uk

Decoding the *Bordetella pertussis* pangenome: Using Nimbus for Oxford Nanopore Sequence Basecalling and Genome Assemblies

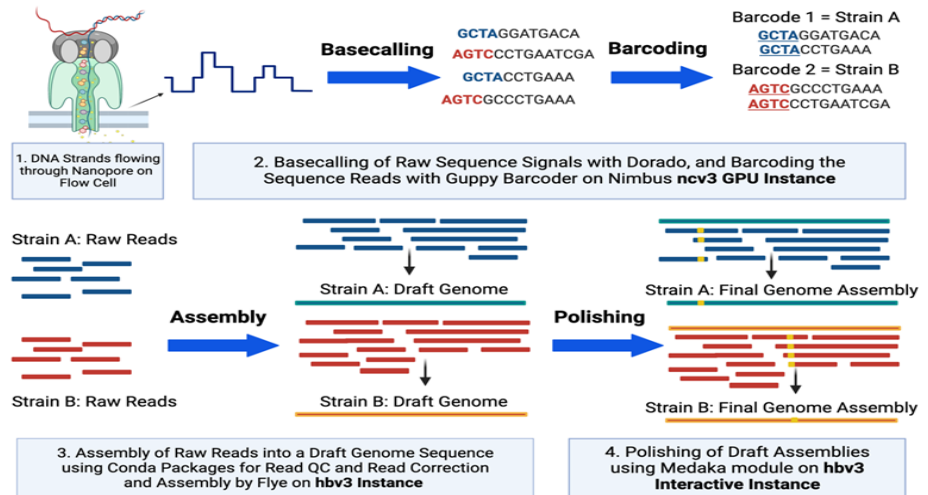


Figure 1: Basecalling and Assembly of Genomes on Nimbus (Image created in Biorender.com)

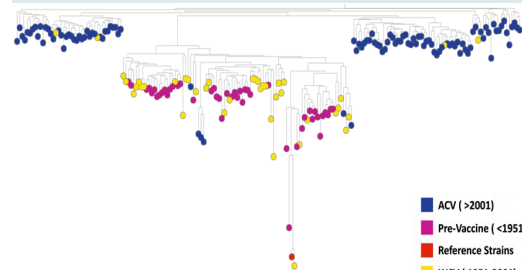


Figure 2: Pangenome Neighbour-Joining Tree of the *B. pertussis* pangenome with node colours representing the vaccine era the strain was isolated from (created by Microreact.org).

Sarah is a PhD student within the Department of Life Sciences. Her research focuses on understanding the genomic variation and evolutionary response of *Bordetella pertussis*, the causative agent of whooping cough, to vaccination pressures.

The human bacterial pathogen *Bordetella pertussis*, causative agent of whooping cough, is frequently regarded as a 'monomorphic' organism due to its low single nucleotide polymorphism (SNP) mutation rate and lack of new gene acquisition [1]. However, since the advent of long-read whole genome sequencing, the ability to synthesize a DNA sequence representative of its full genome has arisen and the true genomic variation of this organism is being only now appreciated. Complete genome sequences reveal a repertoire of genome arrangement mechanisms offering alternative routes of adaptation by deletion, duplication, or inversion of genes [2]. Two types of vaccine have been used in the UK childhood vaccination programme to prevent whooping cough. To investigate *B. pertussis* evolution across eras of different vaccine use, Sarah sequenced the full genomes of 200 strains isolated across 100-years in the UK (1920-2020). These isolates covered the pre-vaccine (1920-1951), whole cell vaccine (WCV, 1951-2001) and acellular vaccine (ACV, 2001-2020) periods.

"Nimbus aids Sarah's research by parallelizing basecalling on multiple GPU cards, unveiling insights into the species' evolutionary response to vaccination pressure"

from 200 strains in batches of 24. As the DNA passes through each Nanopore an electrical current is recorded and later the signal can be translated to individual letters of DNA: A, T, C and Gs (Figure 1). The raw files (.POD5) from a single run are sizable (150-200GB) and were transferred to Nimbus via the XDrive. The first slurm script processed the sequencing files, basecalling the signal into DNA letters using trained machine learning models provided by ONT and removing barcodes to trace sequences back to their source strains. GPU-optimized basecalling with Dorado and pre-processing with the POD5 python package expedited the process on Nimbus, which offered multiple GPUs despite storage limitations in campaign. Completion time averaged 5 hours

per sequencing run, costing ~ £15/100 GB of raw reads on a spot instance as compared to £60/100 for paygo instance. To manage the large .fastq files outputted by basecalling, Sarah processed the remaining pipeline stages on Nimbus. Using hbv3 instances for their increased memory, a second Slurm script employed Canu for read correction, alongside quality control steps (Porechop, Chopper, FastQC). The ONT conda package Flye facilitated draft assembly, while Medaka was utilized for final polishing, installed as a standalone module on Nimbus hbv3 and ncv3 instances by the research computing team.

Overall, Sarah successfully produced 200 high-quality genome assemblies of *B. pertussis* across a century, paving the way for insightful pangenome analysis. This analysis delved into the evolutionary adaptation of the species to vaccine-induced pressures, particularly through gene deletions. Figure 2 shows a pangenome neighbour-joining phylogenetic tree with the nodes coloured by vaccine era the strain was isolated from. The cluster of older isolates (pre-vaccine & WCV) led to the identification of genes that ACV strains likely deleted.

References

1. Mooi, F. R. (2010). "Bordetella pertussis and vaccination: the persistence of a genetically monomorphic pathogen." *Infect Genet Evol* 10(1): 36-49.
2. Weigand, M. R., et al. (2017). "The History of Bordetella pertussis Genome Evolution Includes Structural Rearrangement." *J Bacteriol* 199(8).

Running Singularity Containerized Applications (CPU) Using Slurm Workload Manager on Nimbus

To maintain frontend node availability for user login sessions, it is essential to submit all workloads as batch jobs on Nimbus. While in the previous article (issue 1), we discussed about singularity containerization, this article emphasizes the submission of batch jobs using slurm with code encapsulated within a container image. For demonstration, we will illustrate how a shared memory program (OpenMP) and a distributed memory (MPI) program are executed using Slurm scripts on Nimbus. The subsequent article in the next issue will cover running GPU applications inside the containers.

Example Slurm Scripts to Run Singularity Container Applications

Example-1: Executing an OpenMP program from inside the container (**ex_01.sif**) which prints Hello World on provided number of threads on a single node.

```
# The slurm job script to run on 1 node
[user@nimbus-1-login-1~]$ cat example1.sh
#!/bin/bash
#SBATCH --job-name=test1_container
#SBATCH --output=test1_%A_%a.out
#SBATCH --error=test1_%A_%a.err
#SBATCH --account=CA-CS1HGN-001
#SBATCH --qos=paygo-fsv2-4
#SBATCH --nodes=1
#SBATCH --ntasks-per-node=4
#SBATCH --partition=paygo-fsv2-4
#SBATCH --time=01:00:00

module load Singularity/4.0.0
export OMP_NUM_THREADS=4

singularity exec /your/working/dir/
ex_01.sif /data/a.out >>test1.txt

# Submit the job
$ sbatch example1.sh

# Check the output file
$ cat test1.txt
Hello World... from thread = 0
Hello World... from thread = 3
Hello World... from thread = 2
Hello World... from thread = 1
```

Modify the scripts according to your nimbus account. The container images for both examples are available in Nimbus directory: **/apps/container_images/examples**

Example-2: Running an MPI program from inside the container (**ex_02.sif**) in a batch script using MPI library from outside the container across multiple nodes.

```
# The slurm job script to run an MPI application on 2 compute nodes.
[user@nimbus-1-login-1~]$ cat example2.sh
#!/bin/bash
#SBATCH --job-name=test2_container
#SBATCH --output=test2_%A_%a.out
#SBATCH --error=test2_%A_%a.err
#SBATCH --account=CA-CS1HGN-001
#SBATCH --qos=spot-hc-44
#SBATCH --nodes=2
#SBATCH --ntasks-per-node=4
#SBATCH --partition=spot-hc-44
#SBATCH --time=01:00:00

module load Singularity/4.0.0
module load mpi/openmpi

mpirun -np 8 singularity exec /your/working/
dir/ex_02.sif /data/hello_world_MPI.bin
>>test2.txt

# Submit the job
$ sbatch example1.sh

# Check the output file
$ cat test2.txt
Hello world from processor nimbus-1-spot-hc-
44-pg0-1, rank 0 out of 8 processors
.
Hello world from processor nimbus-1-spot-hc-
44-pg0-2, rank 7 out of 8 processors
```

HPC News and Updates

Nimbus Onboarding Training Announcement

Join us for the Nimbus HPC User Onboarding Training on **April 10th, 2024**. This half-day morning session is ideal for new Nimbus users or those curious about its benefits. Register using the link below which also has the agenda to secure your spot:

<https://forms.office.com/e/WcYUjELG2m>

Tip of the Month

The slurm command "**sacct**" can help in listing the historical jobs with their ids, status etc. within a specified period of time on Nimbus.

```
[user@login1 ~]$ sacct -u username --
starttime=2023-10-01 --endtime=2024-02-01 --
format=JobID,State,Submit,Start,End,Elapsed
```

Acknowledgements

The Research Computing team would like to thank all contributors for the current issue of **HPCBytes**.

- If you would like to contribute a case study or article to be featured in **HPCBytes**, please get in touch with the Research Computing team.
- If you would like to hear more, please subscribe to the Research Computing mailing list here: <https://forms.office.com/e/rF8rLWbakA>

Contact us

Research Computing Team
Digital, Data and Technology
Email: it-hpc@bath.ac.uk