

# Speeding up NGS analysis through local and remote computing resources

B. Wolf, P. Kuonen, *University of Applied Sciences Western Switzerland*  
T. Dandekar, *University of Würzburg, Germany*  
David Atlan, *Phenosystems, Belgium*

## Introduction

The explosion of NGS data which requires increasingly fast computers to keep up with the analysis pushed smaller laboratories to their limits. We previously presented a way for GensearchNGS users to distribute sequence alignment over multiple computers. This possibility has now been expanded to combine multiple computers in the same network with cloud computing resources. For our prototype, the user can by request add Amazon AWS EC2 cloud instances to the alignment process. The cloud resources are dynamically created and destroyed on demand, transparently to the user. It is also possible to combine local alignment on the computer starting the alignment, distributed alignment on multiple computers in the same network and cloud computing in any possible configuration. Even completely offloading the alignment is now possible. This flexibility allows especially smaller laboratories to adapt the software configuration to their needs.

## Distribution setup

The sequence alignment distribution uses a stream based approach. Currently the custom GensearchNGS alignment algorithm is used, but any alignment algorithm could be adapted to use this setup.

The distributed alignment was tested using several configurations. The first configuration used a single laptop to perform the sequence alignment of 5 million 100bp paired end reads. The second configuration added a desktop computer to align the data in combination with the laptop. Configuration 3 added one instance of an Amazon AWS EC2 instance to the computers used in configuration 2. Configuration 4 used the laptop as the base station, which did not perform any alignment but only read the raw data and save the aligned reads into a file. The alignment was done using 2 Amazon AWS EC2 instances. This last configuration is interesting as the client machine does not

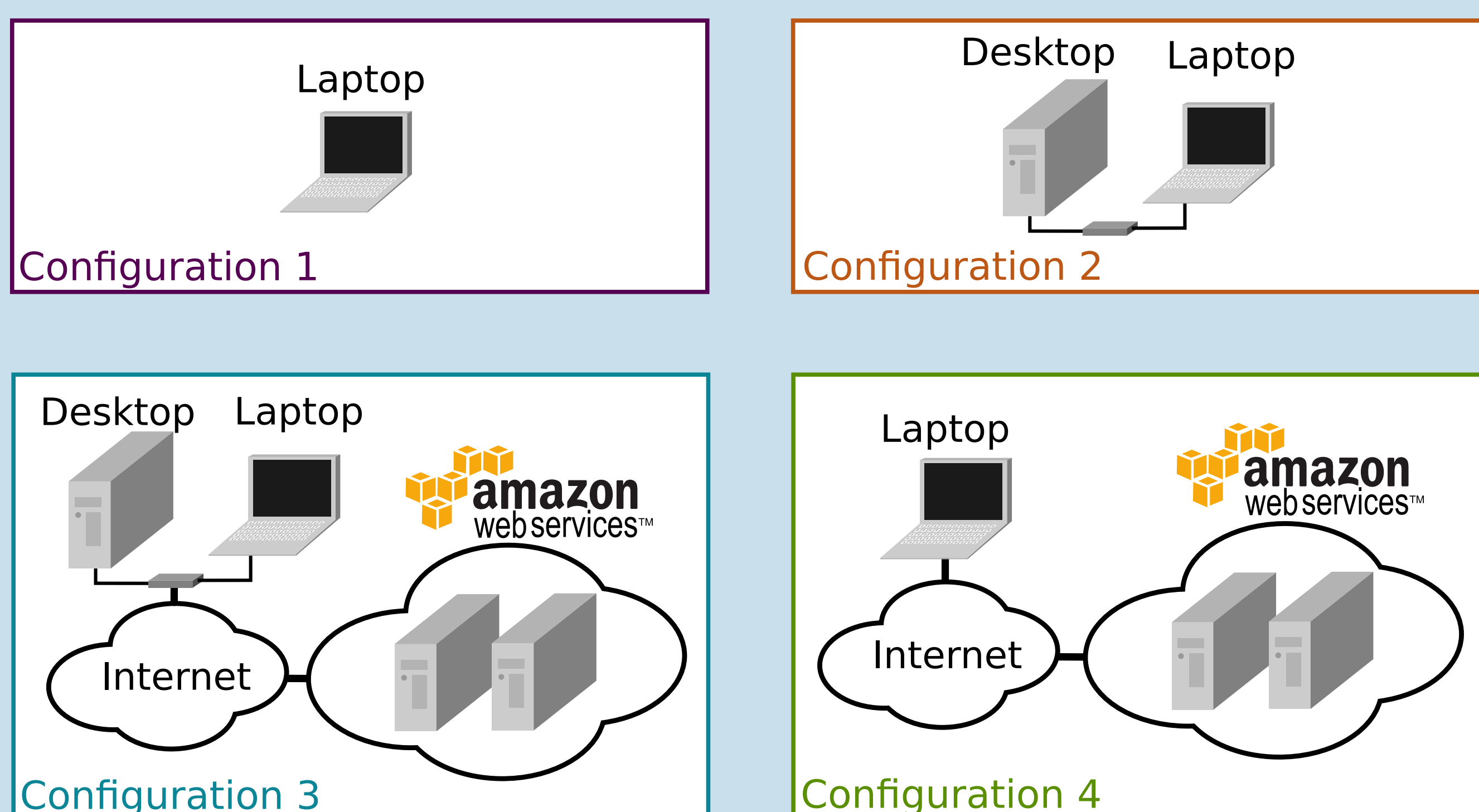


Figure 2: The 4 different configurations used for benchmarking

## Results

The alignment time was measured on all described configurations. Figure 3 shows the time required to perform the alignment.

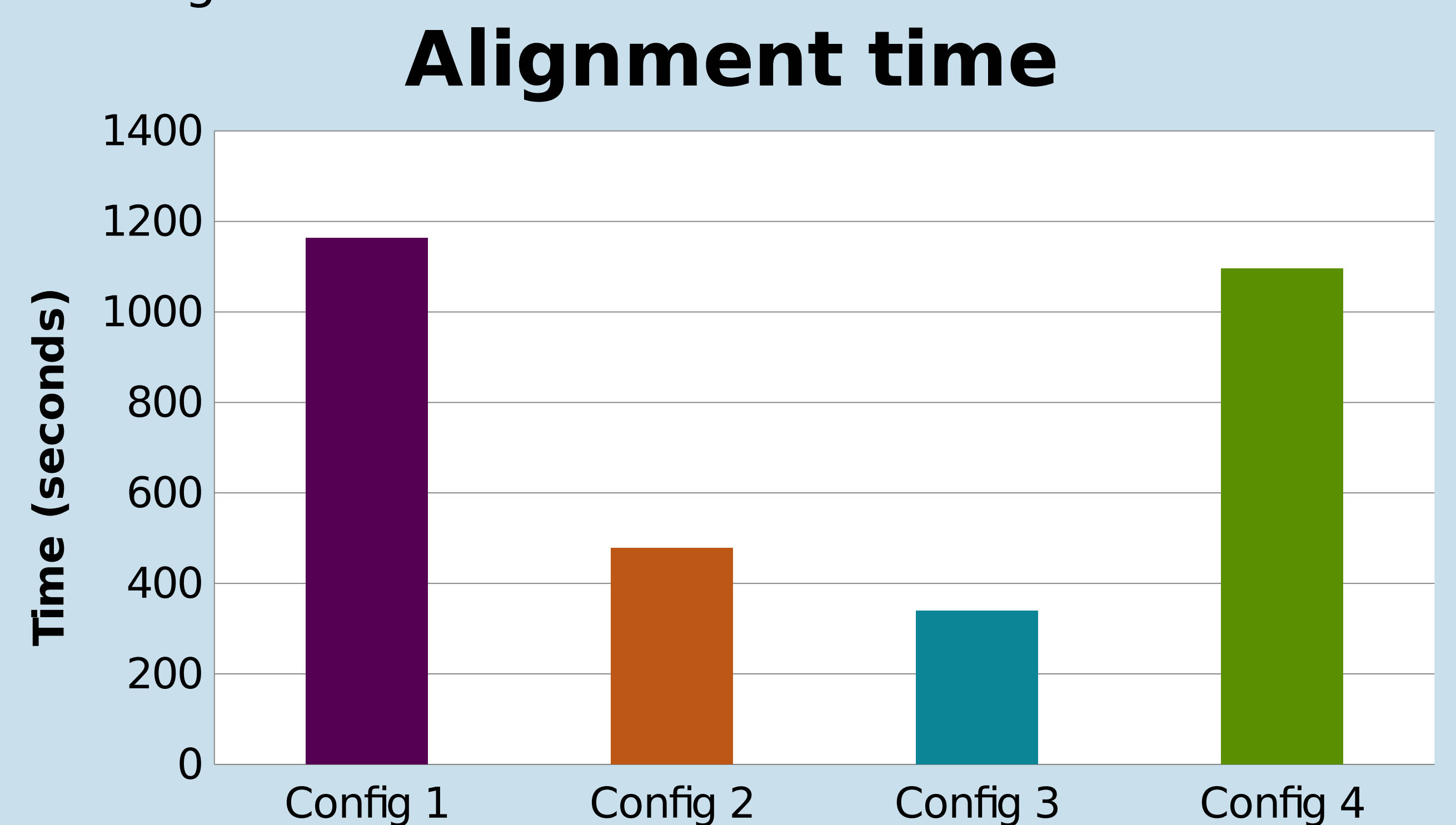


Figure 3: Alignment time on the different configurations

We can see a nice speedup when using the different configurations, making this type of distribution viable. In the future we will compare our approach with existing distributed aligners, noting that none of them has the ability to run on all 4 of the described configurations. The issue of security in the cloud for sequence alignment is also a topic to which needs investigating. One of the solutions is to use a private cloud, which is supported by our implementation.

## Conclusion

We introduced of cloud computing and the flexibility to distribute the alignment workload over different configurations, 4 of them that have been presented on this poster. Some of those configurations are unique to our distribution method. The addition of cloud computing greatly increases the flexibility of the computing infrastructure for NGS data analysis, benefiting especially small laboratories. In conclusion we can note the following highlights of this poster:

- Flexible alignment distribution in various configurations, some unique to our distribution method
- Infrastructure cost reduction for NGS data analysis
- Rapid parallel Genome analysis with high potential for individualized medicine
- Secure data analysis through the use of private clouds

Contact: Beat wolf, [beat.wolf@hefr.ch](mailto:beat.wolf@hefr.ch)

Website GensearchNGS:  
<http://www.phenosystems.com/>



PhenoSystems  
**GensearchNGS**  
MOLECULAR GENETICS MADE EASY

