

“WGS-based bacterial typing and characterization are simplified and accelerated by the use of the Juno pipelines.”

“Juno”: a set of bioinformatics pipelines for the analysis of bacterial genomes for accredited diagnostics and public health laboratory surveillance

A. Hernández-Segura¹,
@alehsecura13
alejandra.hernandez.segura@rivm.nl

E. Hamer^{1,2} R. Wolthuis¹ T. Bosch¹ E. Franz³ M. van den Beld¹

¹ Infectious Disease Research, Diagnostics and Laboratory Surveillance (IDS), RIVM

³ Infectious Diseases, Epidemiology and Surveillance, RIVM

² Applied University of Arnhem and Nijmegen (HAN)

Introduction

- WGS-based techniques are increasingly used in diagnostics and surveillance of bacterial pathogens.
- Lack of bioinformatics expertise of end-users makes WGS difficult to use.
- Current web-based or commercial solutions have downsides:
 - Not easy to automate and/or make reproducible and traceable.
 - Expensive.
 - Incompatible with patient-privacy restrictions.

Objectives

- Facilitate bacterial identification and characterization through easy-to-use bioinformatics pipelines that can be installed locally.
- Validate performance of the pipelines and compatibility with ISO-standards.

Methods

- Using `snakemake` for workflow management and parallelization.¹
- Using `conda` for package/environment management.
- Using well-known and trusted bioinformatics tools:
 - Juno-assembly: `trimmomatic`,² `SPAdes`,³ `FastQC`, `bbtools`, `picard`, `checkM`⁴ and `QUAST`.⁵
 - Juno-typing: `kmerFinder`,⁶ `MLST`,⁷ `SeqSero2`,⁸ `SerotypeFinder`⁹ and `ShigEIFinder`.¹⁰
 - Juno-annotation: `circlator`¹¹ and `prokka`.¹²
 - Juno-resistance: `ResFinder` and `PointFinder`.¹³

Results

- Juno pipelines are:
 - Compatible with Linux-like systems (including WSL2 for Windows).
 - Compatible with HPC clusters.
 - In use for the diagnostic and surveillance at the RIVM.
 - Validated (or in process to validation) according to ISO-standards (ISO15189 and ISO16140-4).
- The use of even partially automated workflows has already accelerated our response to outbreaks/calamities. - The code for the Juno pipelines are available for download through GitHub

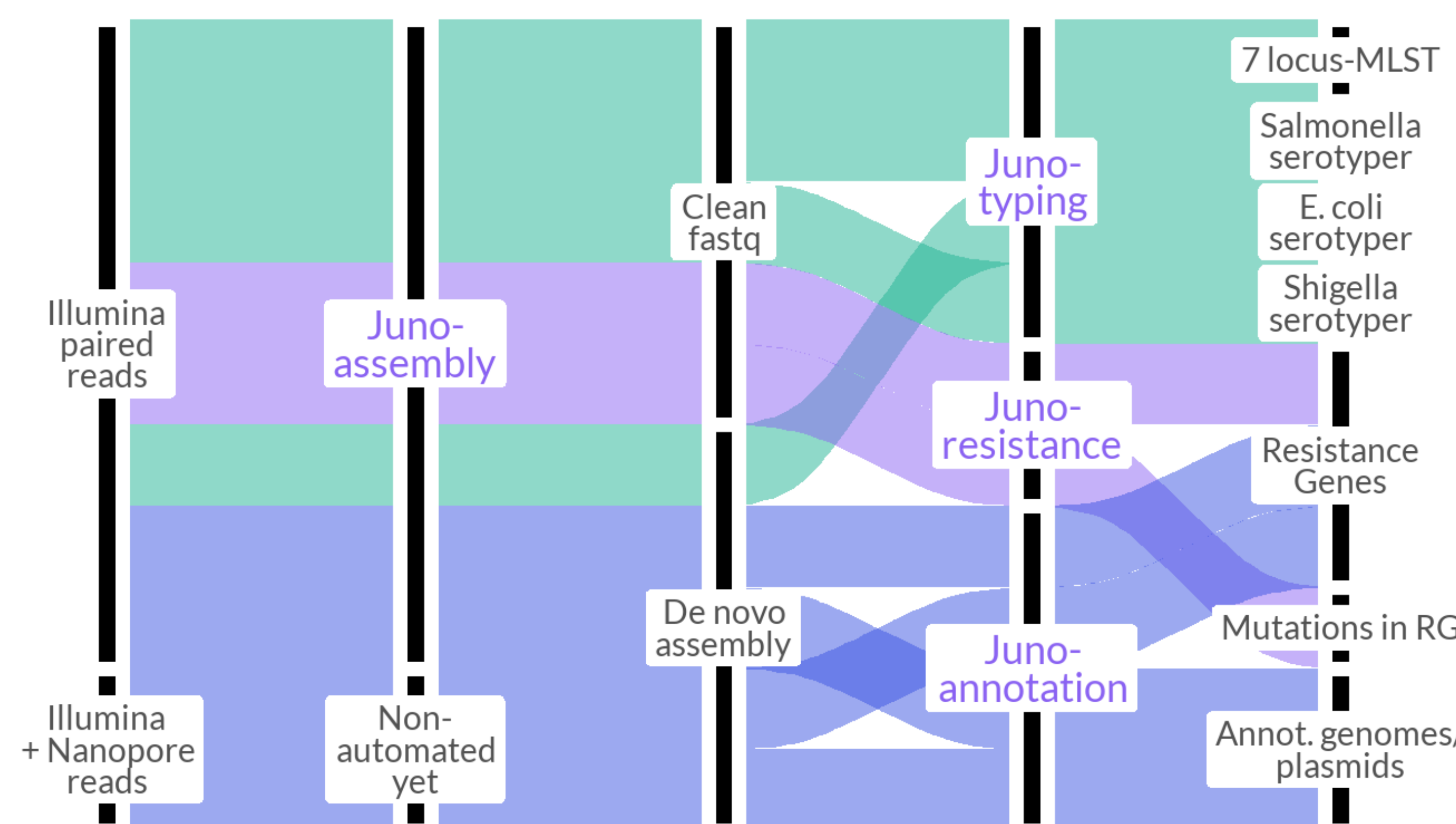


Figure 1: Overview of Juno pipelines. Only features that are fully developed or in progress are shown.

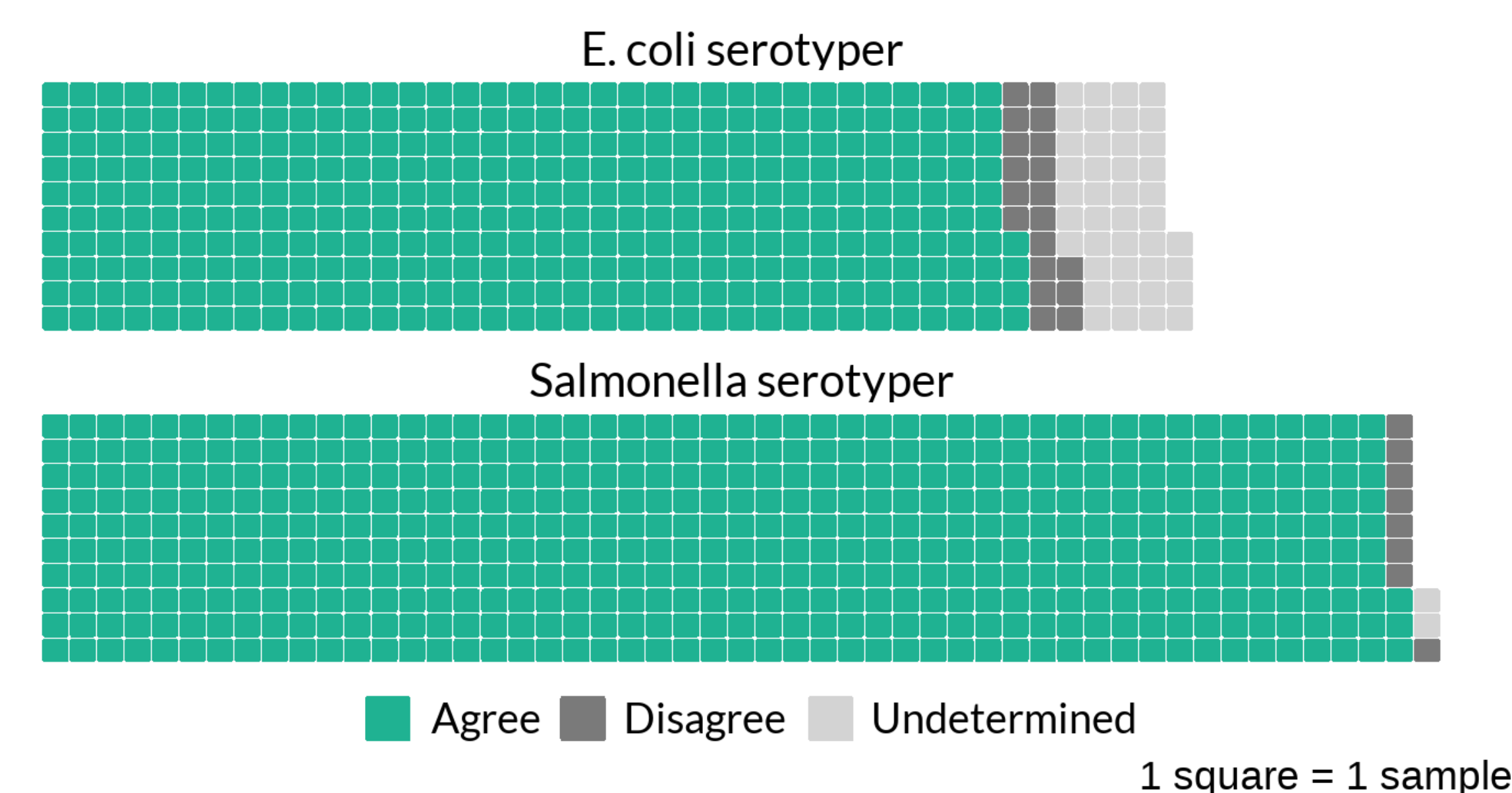


Figure 2: Validation of Juno-typing to meet ISO-standards. The plots show the results of the validation for the Salmonella and E. coli serotypers against the classical serotyping (Agree/Disagree). The Juno-typing pipelines could also render a serotype in rough samples which are difficult to serotype with traditional methods (Undetermined). Validation of the other pipelines is work in progress.

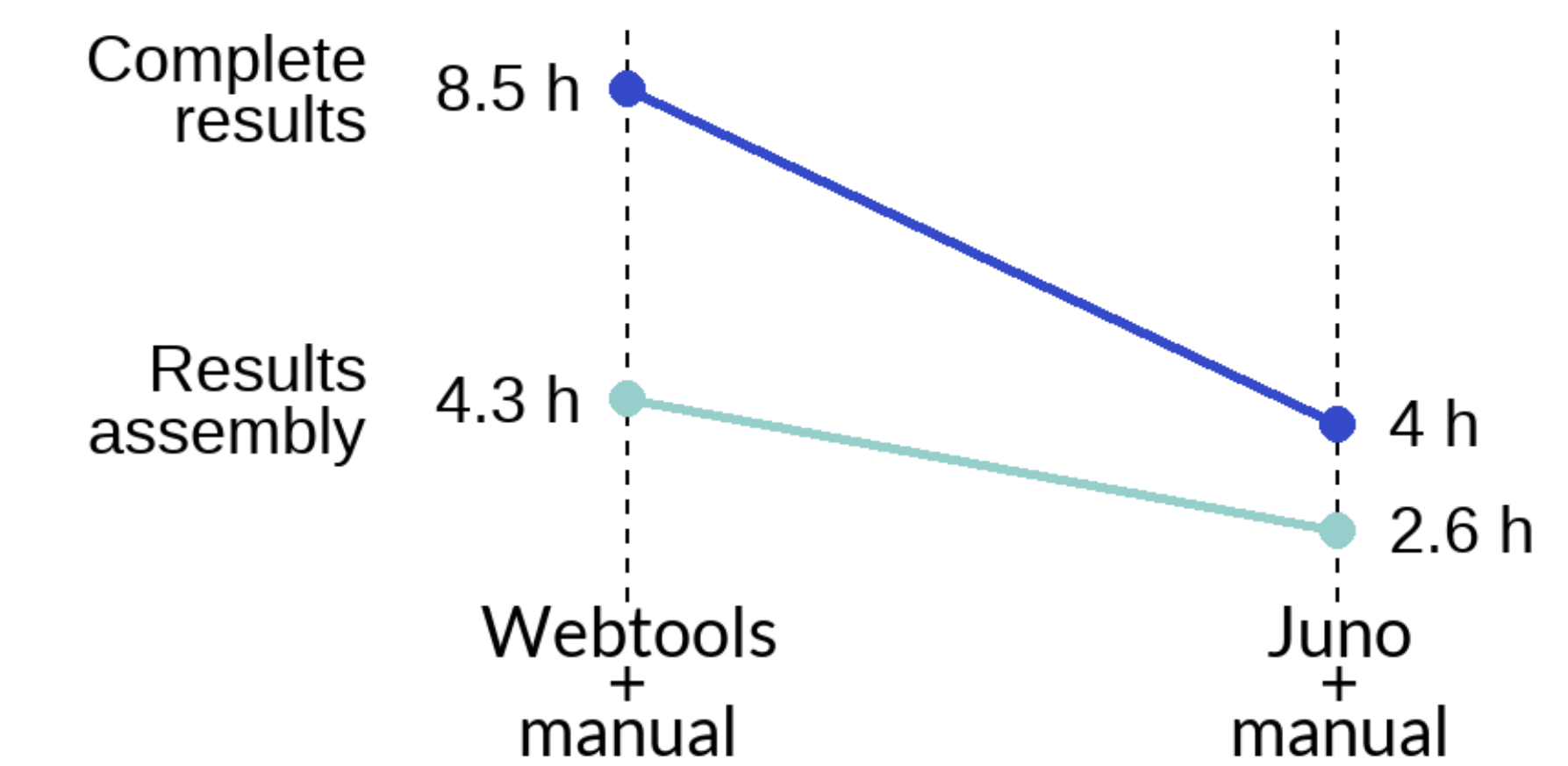


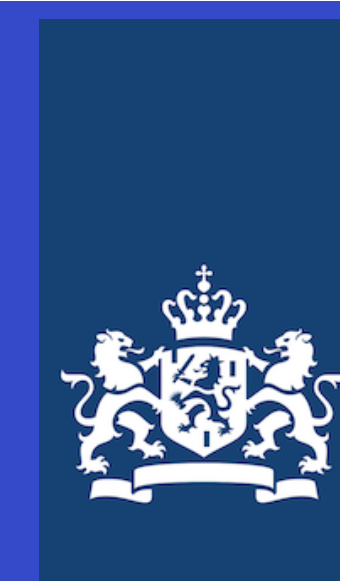
Figure 3: Time saved while having the automated Juno pipelines in place. A “calamity test” was performed to evaluate how fast the technicians could react in case of an outbreak emergency. 12 samples were analyzed with Webtools + manual work (no pipelines) or with the Juno pipelines. At the time of the test, only the Juno-assembly was fully automated. Juno-assembly takes ~30 min to analyze one full sequencing run (~40 samples) but the 2.6 h depicted in the figure take into account the waiting time in the queue of the HPC cluster (slowed down by the large amount of coronavirus samples). The Juno pipelines provide reports in the format used for downstream analyses, accelerating the whole process. In post-pandemic times and with all pipelines in place, we expect these times to be further reduced.

Conclusion

Efforts to make use of bioinformatics tools for diagnostics and surveillance are accelerating the response to outbreaks. Although the separate tools are available, having validated, logged and easy to use pipelines like Juno facilitate the validation, harmonization and accreditation of the process.

References

- Mölder, J.K.P., F. & Nahnsen, S. *F1000Research* 10, (2021).
- Bolger, A.M., Lohse, M. & Usadel, B. *Bioinformatics* 30, 2114–2120 (2014).
- Bankevich, A. et al. *J Comput Biol* 19, 455–477 (2012).
- Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P. & Tyson, G.W. *Genome Res* 25, 1043–1055 (2015).
- Gurevich, A., Saveliev, V., Vyahhi, N. & Tesler, G. *Bioinformatics* 29, 1072–1075 (2013).
- Larsen, M.V. et al. *J Clin Microbiol* 52, 1529–1539 (2014).
- Larsen, M.V. et al. *J Clin Microbiol* 50, 1355–1361 (2012).
- Zhang, S. et al. *Appl Environ Microbiol* 85, (2019).
- Joensen, K.G., Tetzschner, A.M., Iguchi, A., Aarestrup, F.M. & Scheut, F. *J Clin Microbiol* 53, 2410–2426 (2015).
- Zhang, X., Payne, M., Nguyen, T., Kaur, S. & Lan, R. *bioRxiv* (2021).doi:10.1101/2021.01.30.428723
- Hunt, M. et al. *Genome Biol* 16, 294 (2015).
- Seemann, T. *Bioinformatics* 30, 2068–2069 (2014).
- Bortolaia, V. et al. *Journal of Antimicrobial Chemotherapy* 75, 3491–3500 (2020).
- Thorne, B. (2019).at <https://github.com/brentthorne/posterdown>



Rijksoverheid

