Towards pocket-sized genomic analyses: cross-platform benchmark of multi-organism genomic data indexing and alignment

Treccani Mirko, Veschetti Laura, Malerba Giovanni

Department of Neurosciences, Biomedicine and Movement Sciences

University of Verona

www.sites.google.com/view/gmlab

mirko.treccani@univr.it

BITS 2022 Verona June 27 - 29

Background

Current socio-economic situation: the UN 2030 Agenda
Raspberry Pi Foundation Sustain 2025 project
Community interest for Raspberry Pi in bioinformatic
GMLab newborn project: BioVRPi

Aims

Development of a low-cost, stable and tested bioinformatic environment
Cross-platform benchmarking multi-organism genomic analyses

Platforms (64-bits)

Raspberry Pi 4 (RaspiOS): 8 GB RAM - USB HDD storage
Laptop (MacOS): Intel Core i5 - 16 GB RAM - SSD storage
Desktop (Ubuntu): Intel Core i7 - 32 GB RAM - HDD storage

Organisms

SARS-CoV-2 (virus): GCF_009858895.2 - 29.9 Kbp
Escherichia coli (bacterium): GCF_000005845.2 - 4.6 Mbp
Caenorhabditis elegans (nematode): GCF_000002985.6 - 100.3 Mbp

Materials and Methods

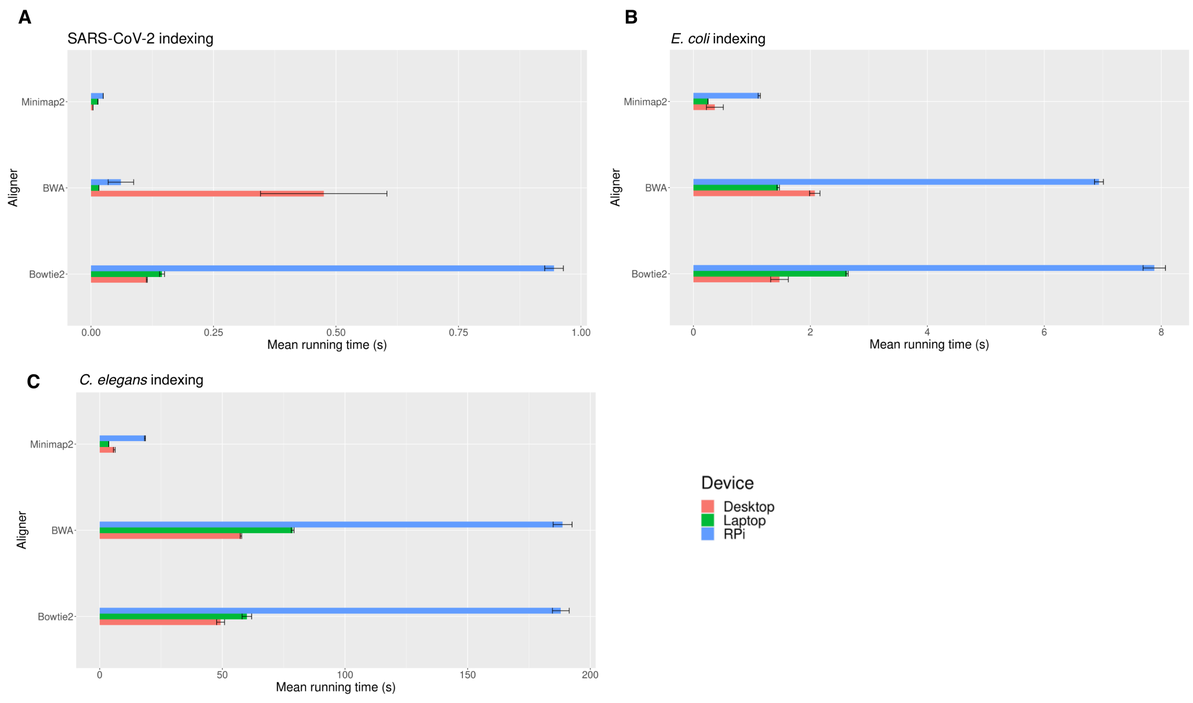
In silico reads generation
- ART Illumina v2.5.8: HiSeqX v2.5 TruSeq profile

Indexing and alignment
- BWA v0.7.17-r1188
- Bowtie2 v2.4.5
- Minimap2 v2.17

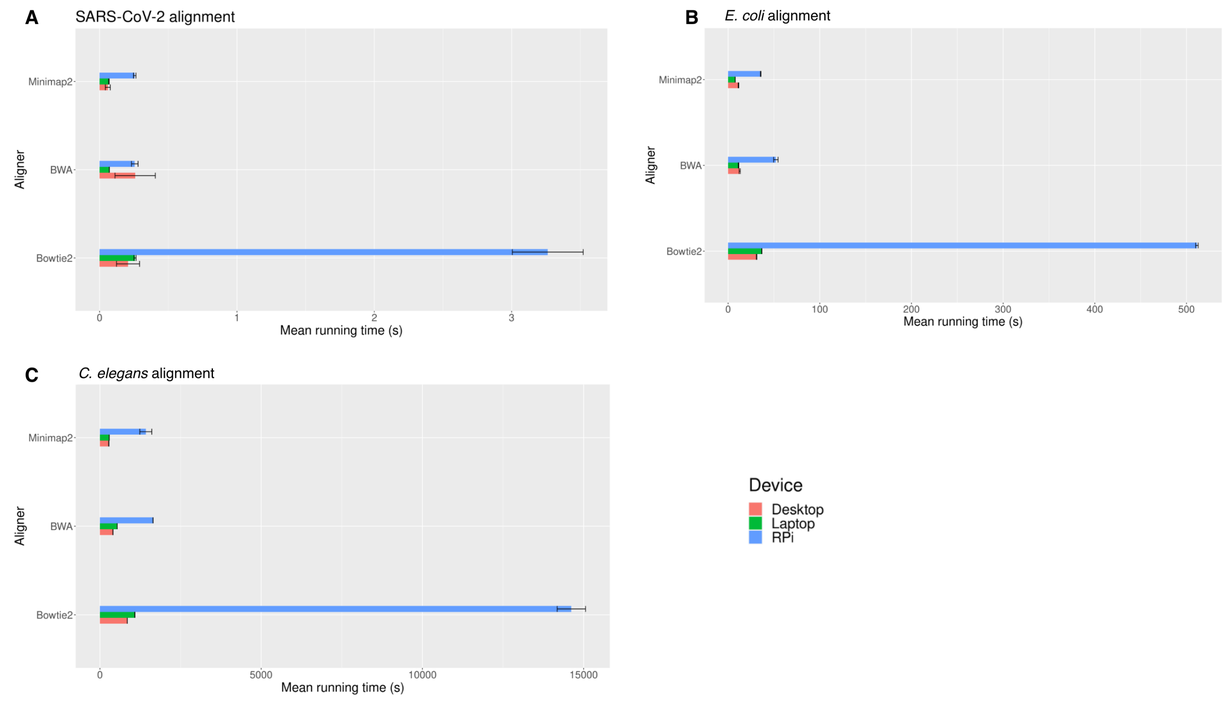
Benchmarking
- Hyperfine v1.13.0: 3 warmup simulations and 10 runs

Results

Indexing



Alignment



Organism	BWA Mean time (standard deviation) [s]			Bowtie2 Mean time (standard deviation) [s]			Minimap2 Mean time (standard deviation) [s]		
	RPi	Laptop	Desktop	RPi	Laptop	Desktop	RPi	Laptop	Desktop
SARS-CoV2	0.061 (0.026)	0.016 (0.001)	0.475 (0.129)	0.945 (0.019)	0.145* (0.005)	0.114 (0.001)	0.025 (0.0001)	0.014 (0.0003)	0.004 (0.0001)
Escherichia coli	6.935 (0.076)	1.450 (0.021)	2.076 (0.089)	7.881 (0.191)	2.627 (0.020)	1.472 (0.151)	1.126 (0.019)	0.249 (0.002)	0.367 (0.145)
Caenorhabditis elegans	188.713 (3.847)	78.674 (0.537)	57.600 (0.310)	187.987 (3.429)	60.000 (1.930)	49.290 (1.603)	18.448 (0.185)	3.690 (0.015)	5.963 (0.353)

Organism	BWA Mean time (standard deviation) [s]			Bowtie2 Mean time (standard deviation) [s]			Minimap2 Mean time (standard deviation) [s]		
	RPi	Laptop	Desktop	RPi	Laptop	Desktop	RPi	Laptop	Desktop
SARS-CoV2	0.256 (0.024)	0.071 (0.001)	0.259 (0.147)	3.263 (0.079)	0.258 (0.008)	0.207 (0.084)	0.256 (0.009)	0.068 (0.002)	0.060 (0.018)
Escherichia coli	52.087 (2.419)	11.400 (0.298)	12.667 (0.640)	511.404 (1.401)	36.747 (0.182)	31.066 (0.246)	35.519 (0.329)	7.718 (0.061)	11.342 (0.303)
Caenorhabditis elegans	1646.538 (2.752)	536.830 (0.921)	403.359 (3.823)	14619.361 (440.325)	1082.906 (5.437)	845.148 (2.281)	1422.040 (185.270)	288.117 (2.084)	274.682 (4.442)

Take-Home Messages

1. Raspberry Pi is a low-cost and environmental-friendly alternative for genomic analyses
2. Storage devices may have influenced RPi performances and needs to be taken into account
3. RPi devices turned out to be efficient for micro-organisms with room for improvement
4. Bowtie2 turned out to be the more time and resources consuming aligner
5. Tools and parameters optimization for Raspberry Pi ARM architecture needs to be investigated and will lead to a greater scalability for more complex organisms

Find out more