All the read datasets used in this thesis are shown above in table 3-1. The reproducing of the GAGE-B papers results was initially started with all datasets in table 3-1. The datasets were first assembled according to the recipe on their site with an assembler named Velvet. Out of the 8 assemblers used by the GAGE-B researchers, Velvet was chosen first because it was the most easy-to-install assembler available in regards to dependencies and student permission on the faculty-computers. After a while, realizing the amount of time and effort needed to reproduce the results fully, the focus shifted from all datasets and assemblers, to one set of MiSeq and HiSeq data assembled using all the 8 assemblers instead. The dataset that was chosen for this was the specie Vibrio cholera. The reason behind this choice was that Vibrio cholerae consisted of the smallest set of MiSeq and HiSeq data, with a total of 3.5 GB compared to a total of 4.5-8 GB for the other species. The choice was based on the fact that assembling the data with multiple assemblers, interpreting the results and comparing them to the original GAGE-B results, as described in the paper, were all quite time-consuming tasks.

Originally, the thought behind doing the assemblies when precompiled results were available was that the results were going to be used later on in the Galaxy tool. The tool could assess the results and see if they correlated with the GAGE-B conclusion. But unfortunately, it did not go as planned since not all assemblies were carried out successfully. The task was still carried out partially, but not all results are equally relevant or informative.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | GAGE-B Hiseq ctg | Soap Hiseq ctg | GAGE-B Miseq ctg | Soap Miseq ctg | GAGE-B Hiseq scf | Soap Hiseq scf | GAGE-B Miseq scf | Soap Miseq scf |
| Contig ≥ 200 | 139 | 462 | 244 | 439 | - | - | - | - |
| Scaffold ≥ 500 | - | - | - | - | 75 | 77 | 165 | 166 |
| N50 | 135 118 | 21 660 | 71 357 | 29 551 | 200 529 | 200 760 | 91 942 | 92 055 |
| NA50 | 112 904 | 21 657 | 71 357 | 29 551 | 181 115 | 181 222 | 91 942 | 92 054 |
| # misassembly | 15 | 2 | 12 | 2 | 15 | 6 | 14 | 6 |
| # local misassembly | 50 | 0 | 44 | 0 | 76 | 81 | 77 | 111 |
| Unaligned | 1 | 1 | 2 | 2 | 1 | 1 | 2 | 2 |
| Genome fraction | 97.295 | 96.488 | 96.940 | 96.230 | 97.305 | 97.517 | 97.050 | 97.068 |
| # genes | 3479 | 3274 | 3442 | 3336 | 3480 | 3485 | 3443 | 3432 |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | GAGE-B Hiseq ctg | SPAdes Hiseq ctg | GAGE-B Miseq ctg | SPAdes Miseq ctg | GAGE-B Hiseq scf | SPAdes Hiseq scf | GAGE-B Miseq scf | SPAdes Miseq scf |
| Contig ≥ 200 | 205 | 158 | 1 475 | 1 480 | - | - | - | - |
| Scaffold ≥ 500 | - | - | - | - | 106 | 106 | 145 | 147 |
| N50 | 83 518 | 137 662 | 262 160 | 262 160 | 98 274 | 225 904 | 262 160 | 262 160 |
| NA50 | 83 518 | 137 656 | 262 160 | 262 160 | 95 858 | 214 751 | 262 160 | 262 160 |
| # misassembly | 4 | 4 | 5 | 5 | 21 | 7 | 5 | 5 |
| # local misassembly | 2 | 9 | 4 | 4 | 17 | 11 | 4 | 4 |
| Unaligned | 1 | 3 | 57 | 57 | 1 | 3 | 57 | 57 |
| Genome fraction | 97.439 | 98.611 | 98.643 | 98.752 | 98.209 | 98.753 | 98.648 | 98.752 |
| # genes | 3 483 | 3 571 | 3 598 | 3597 | 3 544 | 3 586 | 3599 | 3598 |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | GAGE-B Hiseq ctg | Velvet Hiseq ctg | GAGE-B Miseq ctg | Velvet Miseq ctg | GAGE-B Hiseq scf | Velvet Hiseq scf | GAGE-B Miseq scf | Velvet Miseq scf |
| Contig ≥ 200 | 261 | 246 | 201 | 179 | - | - | - | - |
| Scaffold ≥ 500 | - | - | - | - | 85 | 104 | 138 | 133 |
| N50 | 40 877 | 46 346 | 92 036 | 105 176 | 172 545 | 163 386 | 109 996 | 105 176 |
| NA50 | 40 877 | 42 805 | 67 096 | 105 176 | 171 505 | 163 055 | 75 901 | 105 176 |
| # misassembly | 4 | 9 | 14 | 5 | 10 | 10 | 22 | 6 |
| # local misassembly | 8 | 9 | 2 | 3 | 129 | 123 | 13 | 6 |
| Unaligned | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Genome fraction | 97.038 | 96.34 | 97.563 | 96.234 | 97.14 | 96.251 | 97.598 | 96.242 |
| # genes | 3 386 | 3 392 | 3 491 | 3 460 | 3 400 | 3 379 | 3 492 | 3 459 |