International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-8 Issue-6, March 2020

Assessment the Quality of Genome Assemblies by using QUAST Tool for Metagenomics

Swapnaja More, Ajit More.

Abstract*: Number of assembly algorithms have emerged out but due to constraints of genome sequencing techniques no one is perfect. Various methods for assembler's comparison have been developed, but none is yet a recognized standard. The problem of evaluating assemblies of formerly unsequenced species has not been considered, because mostly existing methods for comparing assemblies are only applicable to new assemblies of finished genomes. For comparing and evaluating genome assemblies we have used QUAST (Quality Assessment Tool). This tool is used to assess the quality of leading assembly software by evaluating quality metrics. Assemblies with a reference genome, as well as without a reference can be evaluated by QUAST tool. For genome assembly evaluation based on alignment of contigs to a reference, it is a modern tool. In this study we demonstrate QUAST performance by comparing several leading genome assemblers on three metagenomic datasets.

***Keywords:* genome assembly, contig, misassembly, metagenomics,etc**

I. INTRODUCTION

Modern DNA sequencing technologies are not succeeded till the date to interpret a complete sequence of a chromosome. Alternatively, they generate large numbers of reads of thousands of consecutive bases, sampled from different regions of the genome. Genome assembly software can combine overlapping clones i.e. reads that form a physical map of the genome into larger regions called contigs. In present scenario sequencing technologies and software has many limitations that obstructs reconstruction of full chromosomes. These hurdles are inclusive of errors in reads and huge number repeats in the genome(Ghurye, Cepeda-Espinoza et al.; 2016). To get rid of all of these challenges, resulting in many differences in the generated contigs, different assembly programs use different innovative approaches. Lot of work has been done in present day on methods to compare different assemblers.

Several important findings have been made in recent years by directly extracting data from the environment. Metagenomics, defined as the sequencing and analysis of genomic data taken directly from the environment, is a new and rapidly developing field that makes it practicable to study uncultured organisms(Fernando Meyer 2019). The data collected in metagenomic projects are usually highly fragmented and belong to many organisms unlike organism specific genome projects in which a single genome is considered. NGS technologies enables extraction of short reads without cloning organisms even from low abundant species. In fact, the data produced in these experiments supposed to be huge, confusing, and having fragments from many of species which are highly varying in homology and abundance. As per view of Boisvert, Raymond et al. (2012) all these challenges led to a new computational problem of metagenome assembly followed by a diversity of methods which needs a standard benchmark procedure. Presents assembly evaluation viewpoint are mostly not design to evaluate metagenomes(Zhang 2018). For closely related reference genome no one uses contig alignments.

A QUAST is used by us to evaluate a complete range of metrics as the number of metrics not big that it would merge as hard to interpret all of them. The visualizations and interface are easy to use, representative and informative to the researchers for assessing the quality of assemblies of new species even without a reference genome by applying QUAST(Alexey Gurevich, Vladislav Saveliev et al. 2013). Fortunately, QUAST is quite fast and it could be run simultaneously on multi-core processors.

II. MATERIALS AND METHODS

A. Reference-based evaluation

There are experimented and reviewed metagenomic datasets consist of known species content or simulated reads (Sébastien Boisvert1 2012). To evaluate the assembly methods based on alignments with reference genomes by implementing QUAST these metagenomic datasets are used. Four crucial steps applied for pipeline:

- QUAST is implementing with all input assemblies against the combined references of genome. QUAST can report not only one but all ambiguous alignments which are required from closely related species of metagenomic datasets.
- All contigs are partitioned into groups, each of which contains sequences mapped to a particular reference genome. The contigs mapped to all possible genomes go into respective matching group and unaligned contigs are kept in separate group.
- QUAST is initiated for each input reference separately and provided with a corresponding group of contigs. Unaligned contigs are grouped together and processed without any input reference.

Manuscript received on February 10, 2020.
Revised Manuscript received on February 20, 2020.
Manuscript published on March 30, 2020.
* Correspondence Author
Swapnaja More*, Department of Computer Science, Y.M. College, Bharati Vidyapeeth Deemed University, Pune, Maharashtra, India.
Email: swapnaja2325@gmail.com.
Dr. Ajit More, Department of Computer Applications and System Studies, IMED, Bharati Vidyapeeth Deemed University, Pune, Maharashtra, India. Email : ajit.more@bharatividyapeeth.edu

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

- Subsequently, the outcomes of all QUAST runs are grouped collectively summary reports and visualizations. A user can outlook separate result such as summery of the results for the complete dataset and detailed QUAST result for each run
- Misassembly indicated by no. of interspecies translocations where the border sequences align to obvious references
- Likely incorporate interspecies translocation with an unknown genome through the wide variety of contigs that consist of both massively aligned and unaligned fragments(Sheng-Yong Niu, Jinyu Yang et al. 2017).

B. De novo evaluation

For each detected species, one strain with the great score is persisting inside the assembly.

TABLE-I: GENOME DATABASE AND APPLIED GENOME ASSEMBLIES.

Genome Database	Assemblies
CAMI(Le Bras, Collin et al. 2016)	Gold Assembly IDBA-UD SOAPdenovo2 SPAdes
MetaHIT(Blanco-MÃguez, Alberto Gutiérrez-Jácome et al., 2017)	M_Gold Assembly M_IDBA-UD M_Ray M_SOAPdenovo2 SPAdes
HMP(Group, Peterson et al. 2009)	H_IDBA-UD H_Ray H_SOAPdenovo2 H_SPAdes

When running QUAST without provided reference genomes (CAMI and HMP datasets), BLAST alignment (Ye, McGinnis et al. 2006) and references downloading from NCBI were the most time-consuming and time may significantly vary depending on internet connection bandwidth. Almost all processes excluding reference downloading are parallelized, with each assembly processed in a separate thread, so QUAST works faster on computers with more CPUs.

HTML report including charts and interactive summary report is generated along with key statistics for all references and assemblies. (Calle and Luz 2019)

Summary plots are classified into three groups:

- Misassembly plots: relocations, inversions, translocations and interspecies translocations are types as distribution of misassemblies. They exist in two views such as across all assemblies per reference and across all references per assembly.
- Metric-level plots: for all assemblies against all references one per metric plot is generated. Starting from the best genome and ordered by the mean value between all assemblies.
- Krona charts: it is available only in the de novo evaluation mode, one for the whole dataset and one per assembly, In krona chart representation of taxonomic profile by round charts. (Ondov, Brian Bergman et al. 2015).

The interactive summary HTML reports generate cluster of tables and plots for all statistical data, references and assemblies. Each row of the table shows a value for the merged reference and can be enlarged to visualize each reference values separately. (See below Figures)

III. RESEARCH METHODOLOGY

A. QUAST Report with CAMI Dataset

The ability to retrieve and organize fragment of genomic DNA from any natural context has opened a window into the broad universe of uncultivated microbes. The simulated dataset named Critical Assessment of Metagenome Interpretation (CAMI) is a new community designed aiming for an independent, comprehensive and considerate evaluation of methods. In computational metagenomics, current edge challenges for the CAMI results provide a blueprint for software selection to discover answer for specific research questions.

We assembled the CAMI dataset using IDBA-UD (Yu Peng 2012), SPAdes (Bankevich, Nurk et al. 2012), Ray Meta (Sébastien Boisvert1 2012), and SOAPdenovo2 (Xie, Wu et al. 2014). Along with these genome assemblies CAMI provided supplementary gold assembly. (Reddy, Thomas et al. 2015). QUAST was run in de novo evaluation mode and references detected by algorithm were compared with the original ones used for the dataset simulation. QUAST was run on 19 genome sample from the CAMI project. Statistics are based on contigs size greater than equal to 500 bp.

Misassemblies report

	Gold_Assembly	IDBA_UD	SOAPdenovo2	SPAdes
# misassemblies	344	1208	60	874
# relocations	207	213	28	186
# translocations	59	472	26	242
# inversions	16	5	0	0
# interspecies translocations	62	518	6	446
# misassembled contigs	168	1085	58	781
Misassembled contigs length	7844380	8899419	112977	8242135
# possibly misassembled contigs	103	122	36	108
# possible misassemblies	365	237	42	216
# local misassemblies	217	159	57	148
# unaligned mis. contigs	6	16	4	14
# mismatches	246989	260796	182834	348226
# indels	6934	5660	9864	6485
# indels (<= 5 bp)	5547	4908	5592	5645
# indels (> 5 bp)	1387	752	4272	840
Indels length	28776	16701	75123	20084

Fig.A1. Misassemblies report for CAMI dataset

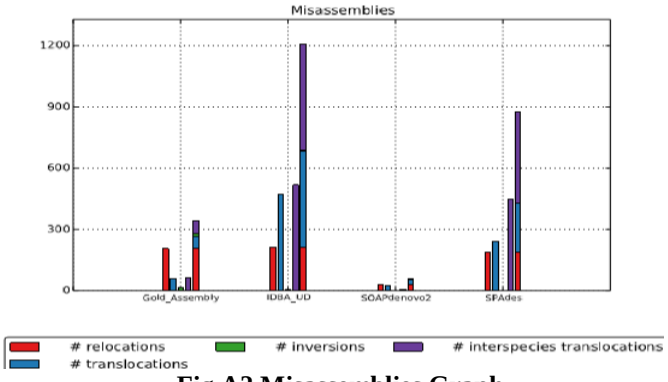

Fig.A2 Misassemblies Graph

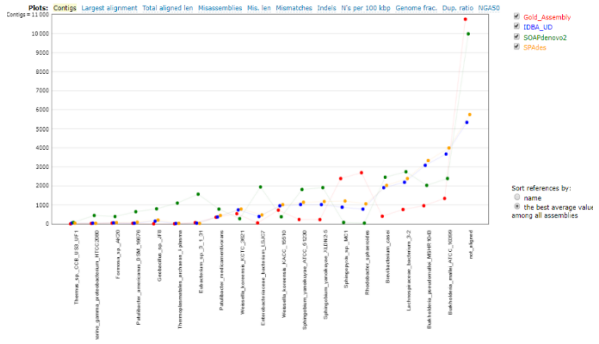

Fig.A3 Best contig average values among all assemblies

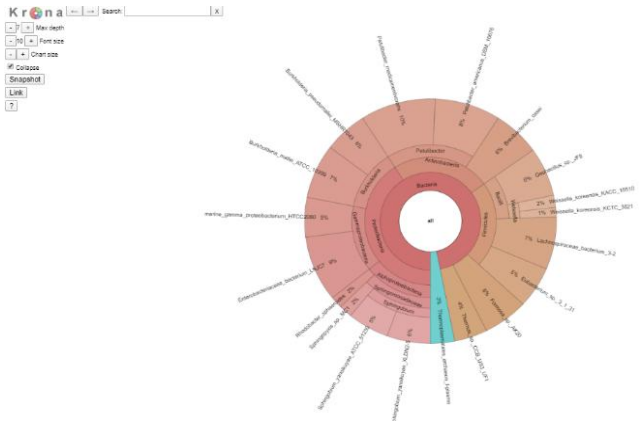

Fig.A4 Summery taxonomy chart with Chrona for CAMI

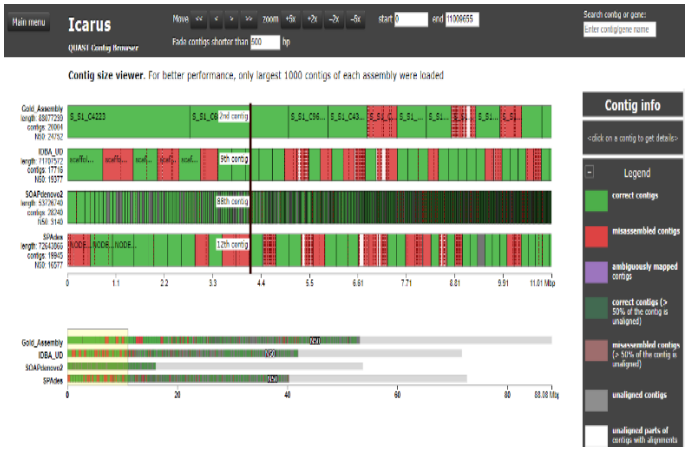

Fig. A5 Contig size from Icarus contig browser

Krona charts: Gold_Assembly IDBA_UD SOAPdenovo2 SPAdes Summary

Combined reference | 74 830 919 bp | 19 references | 886 fragments

Show heatmap

Worst Median Best

Genome statistics	Gold_Assembly	IDBA_UD	SOAPdenovo2	SPAdes
Genome fraction (%)	81.174	71.744	51.749	72.842
Duplication ratio	1.156	1.011	1.003	1.008
Largest alignment	2 780 101	800 397	202 541	505 955
Total aligned length	56 690 477	47 534 974	35 246 571	48 133 790
NGA50	...	...	...	...
LGA50	...	...	...	...
Misassemblies				
# misassemblies	344	1208	60	874
Misassembled contigs length	7 844 380	8 899 419	112 977	8 242 135
Mismatches				
# mismatches per 100 kbp	406.79	485.93	472.15	639
# indels per 100 kbp	11.42	10.55	25.47	11.9
# N's per 100 kbp	0.14	0	44.16	9.43
Statistics without reference				
# contigs	20 004	17 716	28 240	19 945
Largest contig	2 780 101	800 397	202 548	514 709
Total length	88 077 239	71 707 572	53 726 740	72 643 866
Total length (>= 1000 bp)	80 954 951	66 269 212	43 342 137	66 084 435
Total length (>= 10000 bp)	57 307 002	43 171 375	13 399 664	42 077 060
Total length (>= 50000 bp)	33 989 240	23 949 778	1 795 083	21 248 231

Fig.A6 QUAST Report for CAMI dataset.

All our assemblies were evaluated against downloaded 19 references. Supplementary Fig.A6 shows that "Gold Assembly", provided by the CAMI team, has the best results in majority of metrics. A high number of misassemblies for some references possibly indicate the presence of other organisms, closely related to the downloaded genomes. None of the assemblers may be called the best or the worst with regards to the majority of metrics. Largest contig is assembled with IDBA-UD assembly (800 397 bp). SPAdes have a slightly larger total length than IDBA-UD (72 643 866 bp versus 71 707 572 bp), and a significantly fewer number of misassemblies (1208 versus 874). SOAPdenovo2 has a very low number of misassemblies (only 60) but its genome fraction is smaller than IDBA-UD and SPAdes, and it has a very low value for total contigs length (1 795 083 bp).

B. QUAST Report with MetaHIT Dataset

To understand the role of the human intestinal microbiota in health and disease the MetaHIT is emerged. It is Illumina based metagenomic sequencing, characterization and assembly. Paired end analysis used for scaffolding of the fasta file of concatenated contig sequences. QUAST was ran on 13 genome sample from the MetaHIT real dataset. (Qin et al 2010).

Misassemblies report

	M_Ray	M_IDBA_UD	M_SOAPdenovo2	M_SPAdes
# misassemblies	274	547	167	592
# relocations	182	260	71	272
# translocations	16	115	42	123
# inversions	4	15	6	9
# interspecies translocations	72	157	48	188
# misassembled contigs	211	430	161	447
Misassembled contigs length	3736106	6884894	408857	7802409
# possibly misassembled contigs	129	352	77	510
# possible misassemblies	236	538	97	740
# local misassemblies	1125	244	5537	267
# unaligned mis. contigs	54	107	208	142
# mismatches	109985	223393	115494	249937
# indels	4115	11063	3613	13104
# indels (<= 5 bp)	3651	9572	3269	10681
# indels (> 5 bp)	464	1491	344	2423
Indels length	12366	35451	9439	48993

Fig.B1. Misassemblies report for MetaHIT dataset

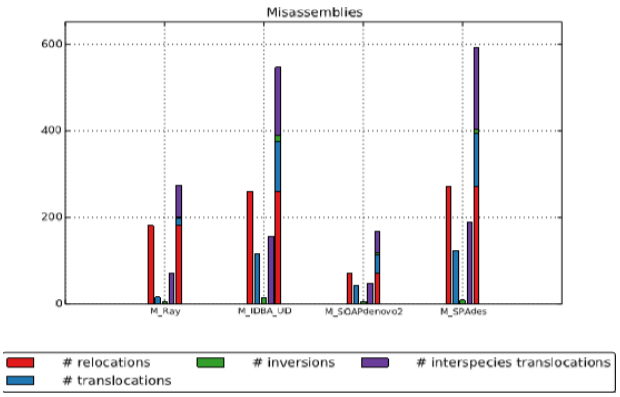

Fig B2. Misassemblies Graph

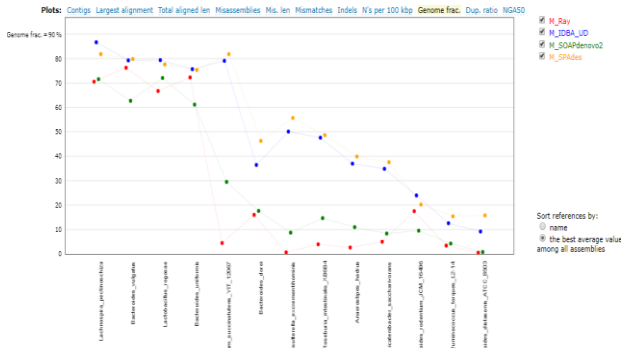

Fig. B3 Genome fraction values among all assemblies for MetaHIT

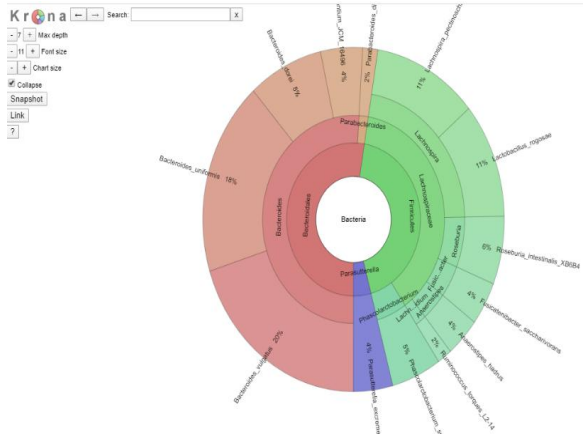

Fig.B4 Summery taxonomy chart with Chrona for MetaHIT

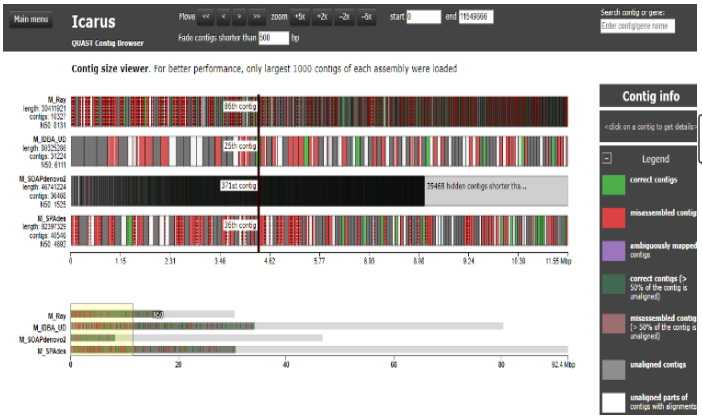

Fig.B5 Contig size from Icarus contig browser

Krona charts: M_Ray M_IDBA_UD M_SOAPdenovo2 M_SPAdes Summary

Combined reference | 50 139 093 bp | 13 references | 449 fragments

Worst Median Best Show heatmap

Genome statistics	M_Ray	M_IDBA_UD	M_SOAPdenovo2	M_SPAdes
Genome fraction (%)	25.972	46.189	28.071	47.505
Duplication ratio	1.06	1.022	1.049	1.028
Largest alignment	63 626	127 263	17 217	131 110
Total aligned length	11 015 623	20 317 649	11 701 370	21 029 369
NGA50	...	...	...	...
LGA50	...	...	...	...
Misassemblies				
# misassemblies	274	547	167	592
Misassembled contigs length	3 736 106	6 884 894	408 857	7 802 409
Mismatches				
# mismatches per 100 kbp	845.63	965.75	820.58	1054.2
# indels per 100 kbp	31.64	47.83	25.67	55.27
# N's per 100 kbp	2087.27	238.48	3730.51	1425.14
Statistics without reference				
# contigs	10 327	31 224	36 468	40 546
Largest contig	99 107	305 144	40 707	189 063
Total length	30 411 921	80 325 286	46 741 224	92 397 329
Total length (>= 1000 bp)	27 080 646	69 223 529	30 720 336	77 823 828
Total length (>= 10000 bp)	13 755 677	34 930 908	2 800 864	33 477 263
Total length (>= 50000 bp)	2 346 322	16 008 349	0	11 409 912

Fig.B6 Quast report for MetaHIT dataset.

Above fig.B6 shows summary HTML report for the MetaHIT dataset for four assemblies. IDBA-UD and SPAdes assembled more genomes than Ray and SOAPdenovo2. At the same time, IDBA-UD and SPAdes demonstrated their best results on different organisms.

C. QUAST Report with HMP Dataset

The HMP used 16S rRNA and whole metagenome shotgun (mWGS)(Bogaerts, Winand et al. 2019) complexity of the human microbiome. HMP(Group, Peterson et al. 2009) dataset having almost 2000 metagenomies and over 10 terabyte of DNA sequences. The design, implementation, and analysis of a big human microbiome is challenge.Essential goal of HMP's is to became a standard catalog of the microorganisms which generate from ordinary human hosts, having their ordinary structure of phylogeny, ecology, taxonomy, biogeography, metabolism, and function. QUAST was ran on the 28 reference genome samples from the HMP real dataset project without providing any references. However, all assemblies contain large fragments not aligned to the combined reference. A short version of the summary HTML report is shown in Figure E.

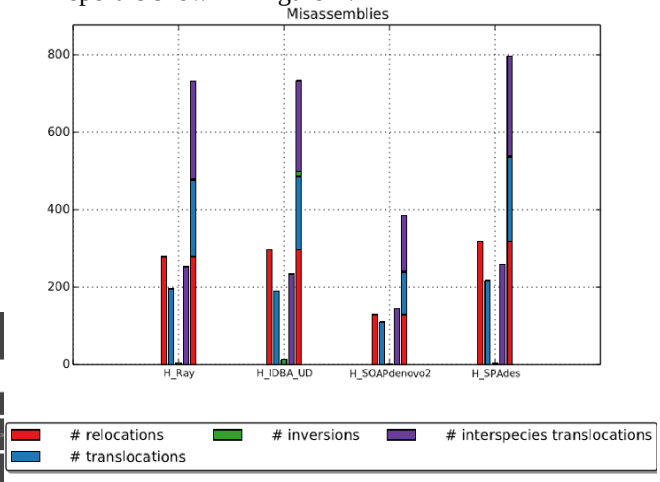

Fig.C1. Misassemblies Graph

Misassemblies report

	H_Ray	H_IDBA_UD	H_SOAPdenovo2	H_SPAdes
# misassemblies	732	733	385	797
# relocations	279	297	129	318
# translocations	196	189	110	217
# inversions	4	13	2	3
# interspecies translocations	253	234	144	259
# misassembled contigs	449	525	252	546
Misassembled contigs length	7715949	7353701	3415211	7913674
# possibly misassembled contigs	601	590	261	560
# possible misassemblies	1245	1103	571	1083
# local misassemblies	1181	1009	613	637
# unaligned mis. contigs	299	257	459	205
# mismatches	686020	781403	352377	816269
# indels	15898	18274	8997	18936
# indels (<= 5 bp)	14792	16385	7166	17495
# indels (> 5 bp)	1106	1889	1831	1441
Indels length	35033	59725	52939	47884

Fig.C2 Misassemblies report for HMP dataset

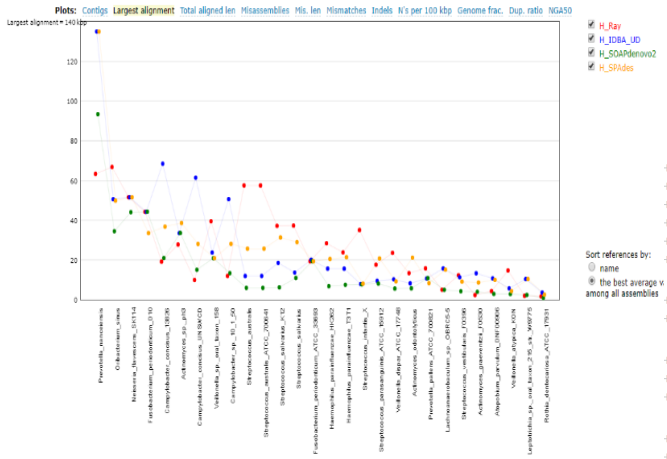

Fig.C3 Largest Alignmnet values among all assemblies

Krona charts: H_Ray H_IDBA_UD H_SOAPdenovo2 H_SPAdes Summary

Combined reference | 62 779 261 bp | 28 references | 756 fragments

Worst Median Best — Show heatmap

Genome statistics	H_Ray	H_IDBA_UD	H_SOAPdenovo2	H_SPAdes
Genome fraction (%)	35.007	43.695	20.844	44.27
Duplication ratio	1.229	1.1	1.069	1.103
Largest alignment	66 813	135 090	93 432	135 056
Total aligned length	21 923 189	26 454 307	11 996 203	26 338 223
NGA50	...	...	...	...
LGA50	...	...	...	...
Misassemblies				
# misassemblies	732	733	385	797
Misassembled contigs length	7 715 949	7 353 701	3 415 211	7 913 674
Mismatches				
# mismatches per 100 kbp	3123.97	2852.76	2693.03	2938.09
# indels per 100 kbp	72.4	66.71	68.76	68.16
# N's per 100 kbp	661.67	0.12	3523.3	142.89
Statistics without reference				
# contigs	20 766	55 710	36 865	49 424
Largest contig	442 828	509 970	560 918	386 771
Total length	72 065 155	99 459 279	64 684 975	92 249 098
Total length (>= 1000 bp)	65 266 007	77 350 395	48 925 980	72 172 611
Total length (>= 10000 bp)	38 598 919	26 853 750	20 575 482	28 960 702
Total length (>= 50000 bp)	14 105 535	14 013 926	10 168 529	14 017 133

Fig.C6 Quast report for HMP dataset.

Fig.C4 Summery taxonomy chart with Chrona for HMP

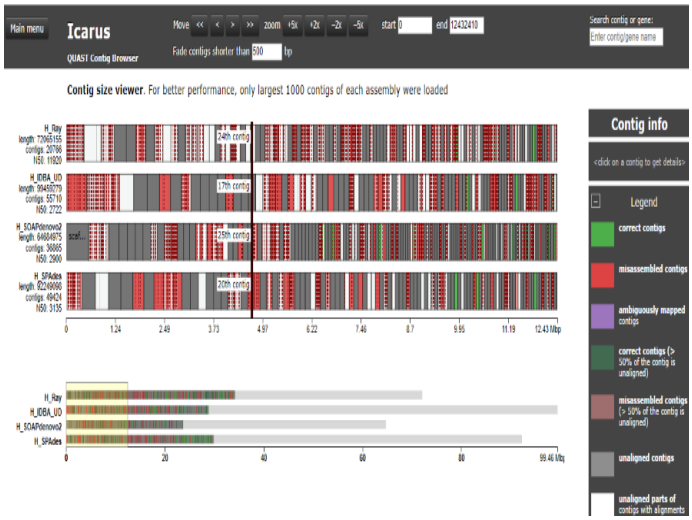

Fig.C5 Contig size from Icarus contig browser

IDBA-UD has the largest total length. IDBA-UD and SPAdes have a significantly higher genome fraction (43.6% and 44.2%) than Ray and, especially, SOAPdenovo2 (35% and 20% respectively). SOAPdenovo2 provides the most accurate assembly with a minimal number of misassemblies and mismatches (385), and has the largest contig (560918). Ray demonstrates the lowest number of contigs (20766).

IV. RESULTS AND DISCUSSION

In assembly approaches study, we compared several leading genome assemblers on two real datasets (MetaHIT and HMP) that do not have existing reference sequences and one simulated dataset (CAMI). De novo studies carried out with a large number of different assemblers along with different parameters. Recent survey report comes with assembly developments and comparisons, which reflect the importance of generating a high-quality, representative genome sequence (Bradnam et al., 2013; Powers et al., 2013). We used QUAST tool to estimate the results that analyze and compare large genomic de novo assemblies' verses reference sequences and determine applicable quality metrics. (fig. d)

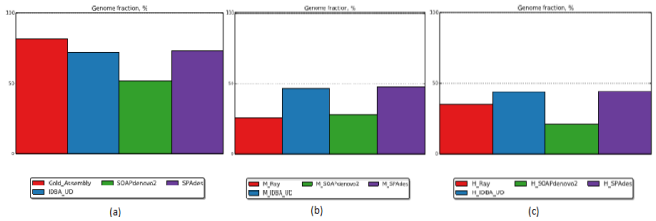

Fig d. Genome fraction count by various assemblers for (a) CAMI, (b) MetaHIT and (c) HMP datasets

This is noticeable from the above figure which shows the assemblers gives higher genome fraction with simulated CAMI dataset. Spades and IDBA-UD gives higher genome fraction count with HMP and MetaHIT database as compared to SOAPdenovo2 and Ray assemblers.

Whereas Ray perform well with HMP but not with MetaHIT and SOAPdenvo2 gives very little count for HMP as compared to CAMI and MeyaHIT databases.

V. CONCLUSION

In this work, we compared the capacity of modern-day genome assembly tools to assemble three metagenomic datasets i.e. CAMI, Metahit and HMP of various size range. These data is assembled using four commonly used leading assemblers in metagenomic studies: IDBA-UD (Peng *et al.*, 2012), SPAdes (Bankevich *et al.*, 2012), Ray Meta (Boisvert *et al.*, 2012), SOAPdenovo2 (Luo *et al.*, 2012)and one Gold_assembly, QUAST performance on all three datasets are demonstrated in above figures which shows comparison of different assembly performance. Datasets available for analysis were sequenced using Standard NGS technologies. Thus, the shown assembler performance produces collective summary HTML report, cluster of tables and plots for all statistics, references and assemblies. Deployment of standard pipeline is the major benefit of this work, for large scale genome assembly evaluation. It can help to reproduce this or similar standard in the future and can also compares some other genome assembly programs on any NGS dataset. None of the assemblers may be called the best or the worst with regards to the majority of metrics We trust that the present pipeline is also suitable for everyday quality control in progress research studies of genomes. Quast will helpful to researchers to evaluate different assembly algorithms and comparisons by using QUAST will help to select the one for their work.

ACKNOWLEDGMENT

This work has been supported by Department of Computer Science, Y.M. College, Pune and Department of Computer Applications and system Studies, IMED, Bharati Vidyapeeth, Pune for strong intramural support for experimental, computational work and for the critical reading of this manuscript.

REFERENCES

1. Alexey Gurevich, Vladislav Saveliev, et al. (2013). "QUAST: quality assessment tool for genome assemblies."
2. Ayling, M., M. D. Clark, et al. (2019). "New approaches for metagenome assembly with short reads." Briefings in Bioinformatics.
3. Bankevich, A., S. Nurk, et al. (2012). "SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing." J Comput Biol. 2012 May;19(5):455-77. doi: 10.1089/cmb.2012.0021. Epub 2012 Apr 16.
4. Blanco-MÃguez, Alberto Gutiérrez-Jácome, et al. (2017). "MAHMI database: a comprehensive MetaHit-based resource for the study of the mechanism of action of the human microbiota." Database: the journal of biological databases and curation 2017: baw157.
   Bogaerts, B., R. Winand, et al. (2019). "Validation of a Bioinformatics Workflow for Routine Analysis of Whole-Genome Sequencing Data and
5. Related Challenges for Pathogen Typing in a European National Reference Center: Neisseria meningitidis as a Proof-of-Concept." Frontiers in Microbiology 10(362).
6. Boisvert, S. b., F. d. r. Raymond, et al. (2012). "Ray Meta: scalable de novo metagenome assembly and profiling." Genome Biology 13(12): R122.
7. Bradnam K, et al. Assemblathon 2: evaluating de novo methods of genome assembly in three vertebrate species, GigaScience , 2013, vol. 2 pg. 10
8. Calle and M. Luz (2019). "Statistical Analysis of Metagenomics Data." Genomics & informatics 17(1): e6-e6.
9. Fernando Meyer, A. B., Peter Belmann, Stefan Janssen, Alice C. McHardy and David Koslicki (2019). "Microbiomes and Metagenomics" BMC Part of Springer Nature.
10. Ghurye, J. S., V. Cepeda-Espinoza, et al. (2016). "Metagenomic Assembly: Overview, Challenges and Applications." The Yale journal of biology and medicine 89(3): 353-362.
11. Powers J, et al. Efficient and accurate whole genome assembly and methylome profiling of E. coli,
12. BMC Genomics, 2013, vol. 14 pg. 675
13. Group, N. H. W., J. Peterson, et al. (2009). "The NIH Human Microbiome Project." Genome Research 19(12): 2317-2323.
14. Le Bras, Y., O. Collin, et al. (2016). "Colibâ€™read on galaxy: a tools suite dedicated to biological information extraction from raw NGS reads." GigaScience 5(1): 9.
15. Ondov, Brian Bergman, et al. (2015). Krona: Interactive Metagenomic Visualization in a Web Browser: 339-346.
16. Reddy, T. B., A. D. Thomas, et al. (2015). "The Genomes OnLine Database (GOLD) v.5: a metadata management system based on a four level (meta)genome project classification." Nucleic Acids Res. 2015 Jan;43(Database issue):D1099-106. doi: 10.1093/nar/gku950. Epub 2014 Oct 27.
17. Sébastien Boisvert1, Frédéric Raymond1,2, Élénie Godzaridis2, François Laviolette3 and Jacques Corbeil1,4 (2012). "Ray Meta: scalable de novo metagenome assembly and profiling ".
18. Sheng-Yong Niu, Jinyu Yang, et al. (2017). "Bioinformatics tools for quantitative and functional metagenome and metatranscriptome data analysis in microbes ".
19. Xie, Y., G. Wu, et al. (2014). "SOAPdenovo-Trans: de novo transcriptome assembly with short RNA-Seq reads." Bioinformatics. 2014 Jun 15;30(12):1660-6. doi: 10.1093/bioinformatics/btu077. Epub 2014 Feb 13.
20. Ye, J., S. McGinnis, et al. (2006). "BLAST: improvements for better sequence analysis." Nucleic Acids Research 34(suppl_2): W6-W9.
21. Yu Peng, H. C. M. L., S. M. Yiu and Francis Y. L. Chin (2012). "IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth ".
22. Zhang, Q. (2018). "Metagenome Assembly and Contig Assignment
23. " Book: Microbiome Analysis Series: Methods In Molecular Biology.

AUTHORS PROFILE

Swapnaja More
Qualification:
M.SC.(Bioinformatics), M.C.A. (Science)
Designation: Research Student
Experience: 4years of academics
Specialization: Bioinformatics
Research & Publications:

1. Published research paper in the JETIR International Journal with No :(ISSN-2349-5162)
Paper ID: JETIR1906V74.
Paper Title: Biometric cryptosystem using fuzzy vault for iris.

2. International Journal of Research and Analytical Reviews IJRAR (ISSN 2348-1269), Paper id: ijrar19k4459. Paper title: A review of bioinformatics tools for metagenomic data in the next-generation sequencing era.

3. Presented poster at NGBT 2019: Nextgen Genomics, Biology, Bioinformatics and Technologies Conference, Mumbai organized by SciGenom Research Foundation's (SGRF).

4. Published paper entitled "Explore the world of Bioinformatics with Data Mining", International Journal of Emerging Technologies and Innovative Research (www.jetir.org), ISSN:2349-5162, Vol.6, Issue 6, page no.39-43, June 2019, Available :http://www.jetir.org/papers/JETIR1908807.pdf

Dr. Ajit More, Designation: Program director
Qualification: M.C.M., M.C.A. Ph.D.
Specialization: Programming, Data mining and Business IntelligenceExperience: 25years
Research & Publications**:** 10 International, 20 National.