

Title

Whoever¹, Sherine Awad², Whatever order^{3,*}

1 Author1 Dept/Program/Center, Institution Name, City, State, Country

2 Same as Titus Departments/Program/Center, Institution Name, City, State, Country

3 Author3 Dept/Program/Center, Institution Name, City, State, Country

*** E-mail: Corresponding author@institute.edu**

Abstract

Author Summary

Introduction

Materials and Methods

Polar (write correct name) datasets where downloaded from XX. The dataset represent XX brief description for the data.

Assembly Treatments -or call it protocol-

1. **Quality Filtering:** Low quality bases were trimmed and low quality reads were removed using trimmomatic. After quality trimming reads were either directly assembled, or first preprocessed with digital normalization and then assembled.
2. **Digital Normalization:** Digital normalization (diginorm) is a technique that down samples reads from highly abundant transcripts while retaining approximately the full sequence information content of the reads [?]
3. **Partitioning:**
4. **Reinflation:**

After Quality trimming, we either preprocess the data using digital normalization, partitioning, or reinflation. We the result assembly using four different assemblers; Velvet (cite), Idba (cite), Spades (cite), and Megahit (cite) and using the various assembly treatments.

Results

Evaluating Assembly Quality Using Various Assembly Treatments

Table 1 shows . Table 2 shows .

Time and Memory Comparisons

Table 3 shows the running time and memory utilizations for four assemblers and different treatments.

Table 1. Evaluating Assembly Using Different Treatments

Treatment/Quality Metric	Quality Filtering	Digital Normalization	Partition	Reinflation
(1) Velvet				
Genome Fraction	72.949	89.043	88.879	-
Unaligned Length	8,977,149	10,909,693	11,317,834	-
(2) Idba				
Genome Fraction	90.969	91.003	90.082	88.346
Unaligned Length	10,709,716	10,637,811	10,644,357	10,288,486
(3) Spades				
Genome Fraction	90.424	90.173	89.272	89.798
Unaligned Length	10,597,529	10,621,398	10,500,235	10,461,672
(4) Megahit				
Genome Fraction	89.961		88.769	
Unaligned Length	10,525,444		10,565,036	

Table 2. Evaluating Assembly Using Different Treatments

Treatment/Quality Metric	Quality Filtering	Digital Normalization	Partition	Reinflation
(1) Velvet				
No. of Unaligned Sequences				-
(2) Idba				
No. of Unaligned Sequences	2092523	2761871	3602590	
(3) Spades				
No. of Unaligned Sequences	3013782		3579651	
(4) Megahit				
No. of Unaligned Sequences				

SubSubsection 1.1

Assembly Validation

We estimated the percentage of unaligned sequences by each assembly treatment and using the five assemblers. We mapped the raw reads to each assembly. Then we extracted the unaligned sequences to each assembly. Table XX shows the percentages of unaligned sequences from raw reads to each assembly treatment using the five assemblers under study.

Discussion

Main points Explain the significance of each treatment

Table 3. Running Time and Memory Utilization

Treatment/Quality Metric	Quality Filtering	Digital Normalization	Partition	Reinflation
(1) Velvet				
Running Time	60:42:52	6:48:46	4:30:36	-
Memory Utilization in KB	1594851536	827412304	1156729920	-
(2) Idba				
Running Time	33:53:46	6:34:24	6:34:24	2:33:17
Memory Utilization in KB	129853424	104736448	93584624	393938608
(3) Spades				
Running Time	67:02:16	15:53:10	7:54:26	6:50:46
Memory Utilization in KB	400340512	127423856	129715072	434531888
(4) Megahit				
Running Time			0:00:27	
Memory Utilization in KB			7548668	

Acknowledgments

References

@articleBrown2012, Author = C Titus Brown and Adina Howe and Qingpeng Zhang and Alexis B Pyrkosz and Timothy H Brom, Journal = , Month = May, Number = , Title = A reference-free algorithm for computational normalization of shotgun sequencing data, Volume = , Year = 2012

Figure Legends

Tables

Supporting Information Legends