

Read mapping on De Bruijn graphs

Additional file

Limasset et. al.

May 24, 2016

Id	Species	nb reads	mean read length	coverage
SRR959239	<i>E.Coli</i>	5,372,832	98	114X
SRR065390	<i>C.Elegans</i>	67,617,092	100	68X
SRR1522085	<i>C.Elegans</i>	46,299,804	51	12X
SRR345593 and SRR345594	<i>Human</i>	3,040,306,840	100	95X

Table 1: Main characteristics of data sets used in this study. Presented read numbers are not the same than those in input of *GGMAP* because we only try to align read without 'N'

Reads set	Assembler	k	c	Mapping on assembly	Mapping on CDBG
				Mapping rate	Mapping rate
<i>E.coli</i>	Velvet	31	3	95.57%	97.17%
//	Minia	//	//	96.19%	//
<i>C.elegans</i>	Velvet	21	2	56.33%	89.15%
//	Minia	//	//	51.43%	//
//	Velvet	31	3	80.60%	92.34%
//	Minia	//	//	80.43%	//
Human	Minia	31	10	63.16%	85.51%

Table 2: Assembly and mapping approach comparison . Columns k and c are the assembly main parameters for Velvet and Minia (k is the length of k -mers and c the minimum coverage for a k -mers to be considered). The last column recalls the overall rate of reads mapped using *GMAP* on the CDBG obtained with the same read sets and k and c parameters. Because of limited computational resources, Velvet was not tested on the Human data-set. The '//' symbol indicates same value(s) between two consecutive lines.

Graph	Mapped set (nb reads)	Mapping Strategy	Nb reads mapped (on branching parts of DBG)	Nb reads fully mapped on unitigs	Overall nb mapped reads
<i>E.coli</i>	SRR959239 (5,128,790)	Greedy	687,074 (13.40%)	4,296,710 (83.78%)	4,983,784 (97.17%)
//	//	Complete	688,929 (13.43%)	4,295,814 (83.76%)	4,984,743 (97.19%)
<i>C.elegans_cpx</i>	SRR065390 (67,617,092)	Greedy	44,686,355 (66.09%)	15,592,918 (23.06%)	60,279,273 (89.15%)
<i>C.elegans_norm</i>	//	Greedy	13,994,715 (20.70%)	48,442,146 (71.64%)	62,436,861 (92.34%)
<i>C.elegans_norm</i>	SRR1522085 (22,509,110)	Greedy	3,523,416 (15.65%)	16,682,194 (74.11%)	20,205,610 (89.77%)
Human	SRR345593 and SRR345594 (2,967,536,821)	Greedy	1,004,182,363 (33.84%)	1,533,456,046 (51.67%)	2,537,638,409 (85.51%)

Table 3: Results of *BGREAT* on real read sets. We remind that the column “*Nb reads fully mapped on unitigs*” corresponds to the number of reads mapped by Bowtie2 with defaults parameters. We also remind that reads from SRR1522085 are mapped on the CDBG obtained using reads from read set SRR065390. For all other lines, the same read set was used both for constructing the CDBG and during the mapping.