Experiments on classifying unaligned reads (Section 3.1.2)

Speced seeds used in the experiments are shown in the following tables.

Table 1: Spaced seeds optimized for the hit number (obtained with IEDERA)

weight	seed
14	###-##-#-#---#-#-#--#--###
16	###-##-#-#---#-#-#--#--#####
18	####-##--##--#--#-#-#-##-####
20	#####-###---##--#--#-#-#-##-####
22	######--####----#--##-#-#--#######

Table 2: Spaced seeds optimized for the coverage (obtained with IEDERA)

weight	seed
14	###-###--#-#--#-##--###
16	###-###--#-#--#-##--#####
18	###-#--###-##--#-#--###-####
20	###-#-##-#-##--##--###-#-#####
22	####-##-#-#-##-###-#-#--##--#####

Table 3: Spaced seeds obtained by quadratic residues (QR-seeds)

weight	seed
14	##----#-###-##-###-#----##
16	#--#--#----###-#-#---####-##-##
18	#-#--##-#----###-####-###----#-##
20	#--#--##---#####-#-#--#-#-#####---##
22	#-##-#-##---#-----###-#---#####-###--#-#--#

Results of experiments are shown in teh following tables.

Table 4: Classification of *Mycobacterium vanbaalenii* (PYR-1, NC_008726) reads against *Mycobacterium smegmatis* (MC2 155, NC_018289) and *Mycobacterium gilvum* (PYR-GCK, NC_009338) genomes. Each entry contains a pair "Fraction of reads classified closer to *M.smegmatis* / Fraction (in %) of reads classified closer to *M.gilvum*".

weight	contig seed			hit-nb seed			cover seed			QR-seed		
	hit-nb	cov	Jac	hit-nb	cov	Jac	hit-nb	cov	Jac	hit-nb	cov	Jac
14	52/41	54/42	30/70	51/40	53/42	29/71	51/39	53/42	28/72	50/41	52/43	27/73
16	39/48	44/47	35/61	34/47	39/50	32/65	34/48	39/51	32/66	33/49	38/53	30/68
18	24/37	25/37	23/42	20/40	21/41	20/49	20/41	21/42	20/50	21/42	22/43	21/52
20	11/24	11/24	11/26	12/32	12/32	11/35	11/30	12/30	11/33	11/29	11/29	11/32
22	07/17	06/17	06/18	08/23	08/24	08/25	08/25	08/25	08/27	07/23	07/23	07/25

Table 5: Classification of *Bacillus thuringiensis* (serovar konkukian 97-27, NC_005957) reads against *Bacillus anthracis* (Ames, NC_003997) and *Bacillus cereus* (ATCC 14579, NC_004722) genomes. Each entry contains a pair "Fraction (in %) of reads classified closer to *B.anthracis* / Fraction of reads classified closer to *B.cereus*".

weight	contig seed			hit-nb seed			cover seed			QR-seed		
	hit-nb	cov	Jac	hit-nb	cov	Jac	hit-nb	cov	Jac	hit-nb	cov	Jac
14	83/14	78/17	87/13	83/13	80/15	88/12	83/13	79/15	88/12	83/13	80/15	88/12
16	81/11	80/12	87/11	82/11	82/11	88/11	82/11	82/11	88/11	82/11	81/12	88/11
18	79/09	79/09	85/09	80/09	80/09	85/09	80/09	80/09	85/09	80/09	80/10	85/09
20	77/08	77/08	83/08	79/08	80/08	84/08	79/08	79/08	84/08	79/08	79/08	84/08
22	76/08	76/08	82/08	79/08	79/08	84/08	79/08	79/08	84/08	79/08	79/08	84/08

Table 6: Classification of *Bacillus licheniformis* (ATCC 14580, NC_006270) reads against *Bacillus anthracis* (Ames, NC_003997) and *Bacillus pumilus* (SAFR-032, NC_009848) genomes. Each entry contains a pair "Fraction (in %) of reads classified closer to *B.anthracis* / Fraction of reads classified closer to *B.pumilus*".

weight	contig seed			hit-nb seed			cover seed			QR-seed		
	hit-nb	cov	Jac	hit-nb	cov	Jac	hit-nb	cov	Jac	hit-nb	cov	Jac
14	47/40	49/46	37/62	50/29	55/36	44/56	51/30	55/37	43/57	49/33	53/39	41/59
16	24/25	24/25	24/30	22/21	22/21	22/28	23/23	23/23	23/32	22/27	23/27	22/35
18	04/07	04/07	04/08	04/08	04/08	04/09	04/08	04/08	04/09	04/10	04/10	04/10
20	0.9/3.8	0.9/3.8	0.8/4.0	1.0/5.4	1.0/5.4	1.0/5.7	0.9/4.4	1.0/4.4	0.9/4.6	0.8/4.2	0.8/4.2	0.8/4.4
22	0.3/2.4	0.3/2.4	0.3/2.6	0.4/2.9	0.4/2.9	0.4/3.1	0.4/3.3	0.5/3.3	0.4/3.5	0.4/2.8	0.4/2.8	0.4/3.0