Simultaneous compression of multiple error-corrected short-read sets for faster data transmission and better de novo assemblies

Tao Tang[1], Gyorgy Hutvagner[2], Wenjian Wang[3], and Jinyan Li[1,*]

[1]Advanced Analytics Institute, University of Technology Sydney, Broadway, NSW 2007, Australia
[2]School of Biomedical Engineering, University of Technology Sydney, Broadway, NSW 2007, Australia
[3]School of Computer and Information Technology, Shanxi University, Shanxi Road, Shanxi 030006, China
[*]Corresponding Author

January 13, 2022

1 Details of datasets

Table S1: Details of dataset-33 (33 files, 102.39GB)

Name	Read number	Retrieved from
SRR017498.fastq	2197691	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR017/SRR017498/SRR017498.fastq.gz
SRR017499.fastq	2185062	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR017/SRR017499/SRR017499.fastq.gz
SRR017500.fastq	2067941	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR017/SRR017500/SRR017500.fastq.gz
SRR017501.fastq	9304954	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR017/SRR017501/SRR017501.fastq.gz
SRR350414.fastq	13873841	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR350/SRR350414/SRR350414.fastq.gz
SRR350416.fastq	13800627	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR350/SRR350416/SRR350416.fastq.gz
SRR350417.fastq	13971754	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR350/SRR350417/SRR350417.fastq.gz
SRR364390.fastq	12429076	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR364/SRR364390/SRR364390.fastq.gz
SRR364391.fastq	13524457	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR364/SRR364391/SRR364391.fastq.gz
SRR364392.fastq	13327151	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR364/SRR364392/SRR364392.fastq.gz
SRR364393.fastq	14013116	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR364/SRR364393/SRR364393.fastq.gz
SRR364394.fastq	13776778	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR364/SRR364394/SRR364394.fastq.gz
SRR391480.fastq	23074156	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR391/SRR391480/SRR391480.fastq.gz
SRR391482.fastq	24012306	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR391/SRR391482/SRR391482.fastq.gz
SRR391483.fastq	27413485	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR391/SRR391483/SRR391483.fastq.gz
SRR391484.fastq	26501216	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR391/SRR391484/SRR391484.fastq.gz
SRR404183.fastq	4259218	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR404/SRR404183/SRR404183.fastq.gz
SRR404184.fastq	7816147	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR404/SRR404184/SRR404184.fastq.gz
SRR404185.fastq	4946871	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR404/SRR404185/SRR404185.fastq.gz
SRR404186.fastq	6803145	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR404/SRR404186/SRR404186.fastq.gz
SRR404187.fastq	17124858	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR404/SRR404187/SRR404187.fastq.gz
SRR404188.fastq	14145375	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR404/SRR404188/SRR404188.fastq.gz
SRR449375.fastq	13031496	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR449/SRR449375/SRR449375.fastq.gz
SRR449376.fastq	13197078	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR449/SRR449376/SRR449376.fastq.gz
SRR449377.fastq	13479636	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR449/SRR449377/SRR449377.fastq.gz
SRR449378.fastq	10366702	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR449/SRR449378/SRR449378.fastq.gz
SRR449379.fastq	13147649	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR449/SRR449379/SRR449379.fastq.gz
SRR513465.fastq	13437700	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR513/SRR513465/SRR513465.fastq.gz
SRR547561.fastq	13242851	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR547/SRR547561/SRR547561.fastq.gz
SRR547562.fastq	12911326	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR547/SRR547562/SRR547562.fastq.gz
SRR547563.fastq	15882034	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR547/SRR547563/SRR547563.fastq.gz

Name	Read number	Retrieved from
SRR547564.fastq	15651448	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR547/SRR547564/SRR547564.fastq.gz
SRR547565.fastq	13524315	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR547/SRR547565/SRR547565.fastq.gz

Table S2: Details of dataset-47 (47 files, 208.97GB)

Name	Read number	Retrieved from
ERR532390_1.fastq	17018488	http://ftp.sra.ebi.ac.uk/vol1/fastq/ERR532/ERR532390/ERR532390_1.fastq.gz
ERR532394_1.fastq	30746428	http://ftp.sra.ebi.ac.uk/vol1/fastq/ERR532/ERR532390/ERR532390_1.fastq.gz
SRR062634_1.fastq	24476109	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR062/SRR062634/SRR062634_1.fastq.gz
SRR062635_1.fastq	24479905	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR062/SRR062635/SRR062635_1.fastq.gz
SRR062641_1.fastq	24958448	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR062/SRR062641/SRR062641_1.fastq.gz
SRR062644_1.fastq	24051140	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR062/SRR062644/SRR062644_1.fastq.gz
SRR062646_1.fastq	24136249	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR062/SRR062646/SRR062646_1.fastq.gz
SRR062655_1.fastq	22997965	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR062/SRR062655/SRR062655_1.fastq.gz
SRR067591.fastq	2387280	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR067/SRR067591/SRR067591.fastq.gz
SRR354180.fastq	6152228	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR354/SRR354180/SRR354180.fastq.gz
SRR354181.fastq	4907117	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR354/SRR354181/SRR354181.fastq.gz
SRR354182.fastq	3578549	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR354/SRR354182/SRR354182.fastq.gz
SRR354183.fastq	5575465	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR354/SRR354183/SRR354183.fastq.gz
SRR361240.fastq	4651469	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR361/SRR361240/SRR361240.fastq.gz
SRR361241.fastq	5137558	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR361/SRR361241/SRR361241.fastq.gz
SRR361242.fastq	5385821	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR361/SRR361242/SRR361242.fastq.gz
SRR361243.fastq	5437878	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR361/SRR361243/SRR361243.fastq.gz
SRR361244.fastq	4951700	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR361/SRR361244/SRR361244.fastq.gz
SRR361245.fastq	3998188	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR361/SRR361245/SRR361245.fastq.gz
SRR387477.fastq	3759745	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR387/SRR387477/SRR387477.fastq.gz
SRR387478.fastq	3890863	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR387/SRR387478/SRR387478.fastq.gz
SRR387479.fastq	3050218	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR387/SRR387479/SRR387479.fastq.gz
SRR387480.fastq	4470576	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR387/SRR387480/SRR387480.fastq.gz
SRR387481.fastq	570358	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR387/SRR387481/SRR387481.fastq.gz
SRR387482.fastq	3171526	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR387/SRR387482/SRR387482.fastq.gz
SRR401414.fastq	11745032	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR401/SRR401414/SRR401414.fastq.gz
SRR401415.fastq	20578433	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR401/SRR401415/SRR401415.fastq.gz
SRR401416.fastq	20515752	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR401/SRR401416/SRR401416.fastq.gz
SRR401417.fastq	14646834	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR401/SRR401417/SRR401417.fastq.gz
SRR445612_1.fastq	22691950	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR445/SRR445612/SRR445612_1.fastq.gz
SRR445635_1.fastq	19586556	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR445/SRR445635/SRR445635_1.fastq.gz
SRR445718.fastq	32943665	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR445/SRR445718/SRR445718.fastq.gz
SRR445719.fastq	30690825	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR445/SRR445719/SRR445719.fastq.gz
SRR445724.fastq	50896528	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR445/SRR445724/SRR445724.fastq.gz
SRR490961.fastq	49127668	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR490/SRR490961/SRR490961.fastq.gz
SRR490962.fastq	45583212	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR490/SRR490962/SRR490962.fastq.gz
SRR490976.fastq	33261944	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR490/SRR490976/SRR490976.fastq.gz
SRR504462.fastq	21246896	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR504/SRR504462/SRR504462.fastq.gz

SRR504463.fastq	26425413	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR504/SRR504463/SRR504463.fastq.gz
SRR504464.fastq	26186672	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR504/SRR504464/SRR504464.fastq.gz
SRR504465.fastq	26350142	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR504/SRR504465/SRR504465.fastq.gz
SRR504466.fastq	24871839	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR504/SRR504466/SRR504466.fastq.gz
SRR504468.fastq	22474516	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR504/SRR504468/SRR504468.fastq.gz
SRR518633.fastq	21896197	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR518/SRR518633/SRR518633.fastq.gz
SRR518634.fastq	24974345	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR518/SRR518634/SRR518634.fastq.gz
SRR518635.fastq	23529206	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR518/SRR518635/SRR518635.fastq.gz
SRR518636.fastq	15501764	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR518/SRR518636/SRR518636.fastq.gz

Table S3: Details of dataset-50 (50 files, 71.04GB)

Name	Read number	Retrieved from
SRR020684_1.fastq	10375605	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR020/SRR020684/SRR020684_1.fastq.gz
SRR020687_1.fastq	10394255	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR020/SRR020687/SRR020687_1.fastq.gz
SRR020688_1.fastq	8837322	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR020/SRR020688/SRR020688_1.fastq.gz
SRR020690_1.fastq	9534829	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR020/SRR020688/SRR020690_1.fastq.gz
SRR020692_1.fastq	9260242	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR020/SRR020692/SRR020692_1.fastq.gz
SRR020696_1.fastq	8961001	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR020/SRR020696/SRR020696_1.fastq.gz
SRR020697_1.fastq	9827620	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR020/SRR020697/SRR020697_1.fastq.gz
SRR020698_1.fastq	9658118	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR020/SRR020698/SRR020698_1.fastq.gz
SRR020699_1.fastq	9540494	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR020/SRR020699/SRR020699_1.fastq.gz
SRR033506_1.fastq	9668964	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR033/SRR033506/SRR033506_1.fastq.gz
SRR033508_1.fastq	9800692	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR033/SRR033508/SRR033508_1.fastq.gz
SRR033513_1.fastq	10103429	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR033/SRR033513/SRR033513_1.fastq.gz
SRR033516_1.fastq	10490546	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR033/SRR033516/SRR033516_1.fastq.gz
SRR033540_1.fastq	10022304	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR033/SRR033540/SRR033540_1.fastq.gz
SRR035578.fastq	17211229	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR035/SRR035578/SRR035578.fastq.gz
SRR035580.fastq	18183656	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR035/SRR035580/SRR035580.fastq.gz
SRR035581.fastq	17419085	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR035/SRR035581/SRR035581.fastq.gz
SRR035582.fastq	17661304	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR035/SRR035582/SRR035582.fastq.gz
SRR035583.fastq	17342832	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR035/SRR035583/SRR035583.fastq.gz
SRR035585.fastq	18387207	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR035/SRR035585/SRR035585.fastq.gz
SRR035586.fastq	19275633	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR035/SRR035578/SRR0355786.fastq.gz
SRR042754_1.fastq	12211192	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR042/SRR042754/SRR042754_1.fastq.gz
SRR042756_1.fastq	11931416	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR042/SRR042756/SRR042756_1.fastq.gz
SRR042758_1.fastq	11149360	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR042/SRR042758/SRR042758_1.fastq.gz
SRR042760_1.fastq	11350648	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR042/SRR042760/SRR042760_1.fastq.gz
SRR042762_1.fastq	11648481	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR042/SRR042762/SRR042762_1.fastq.gz
SRR042768_1.fastq	12045980	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR042/SRR042768/SRR042768_1.fastq.gz
SRR306333.fastq	1467267	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR306/SRR306333/SRR306333.fastq.gz
SRR306334.fastq	1324182	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR306/SRR306334/SRR306334.fastq.gz

SRR306335.fastq	1265603	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR306/SRR306335/SRR306335.fastq.gz
SRR306336.fastq	1695994	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR306/SRR306336/SRR306336.fastq.gz
SRR306337.fastq	1506514	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR306/SRR306337/SRR306337.fastq.gz
SRR306338.fastq	1171861	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR306/SRR306333/SRR306338.fastq.gz
SRR306339.fastq	820255	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR306/SRR306339/SRR306339.fastq.gz
SRR306340.fastq	791082	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR306/SRR306340/SRR306334-.fastq.gz
SRR306341.fastq	1102312	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR306/SRR306341/SRR306341.fastq.gz
SRR306342.fastq	985592	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR306/SRR306342/SRR306342.fastq.gz
SRR404352_1.fastq	9218083	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR404/SRR404352/SRR404352_1.fastq.gz
SRR404353_1.fastq	12132806	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR404/SRR404353/SRR404353_1.fastq.gz
SRR404354_1.fastq	10353791	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR404/SRR404354/SRR404354_1.fastq.gz
SRR404355_1.fastq	13407653	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR404/SRR404355/SRR404355_1.fastq.gz
SRR445237_1.fastq	4165017	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR445/SRR445237/SRR445237_1.fastq.gz
SRR445239_1.fastq	2763227	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR445/SRR445239/SRR445239_1.fastq.gz
SRR447746.fastq	12077910	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR447/SRR447746/SRR447746.fastq.gz
SRR447747.fastq	12077910	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR447/SRR447747/SRR447747.fastq.gz
SRR448375.fastq	8703586	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR448/SRR448375/SRR448375.fastq.gz
SRR448376.fastq	18597630	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR448/SRR448376/SRR448376.fastq.gz
SRR448378.fastq	2151255	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR448/SRR448378/SRR448378.fastq.gz
SRR448379.fastq	18619182	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR448/SRR448379/SRR448379.fastq.gz
SRR448380.fastq	20000000	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR448/SRR448380/SRR448380.fastq.gz

Table S4: Details of dataset-100 (100 files, 197.97GB)

Name	Read number	Retrieved from
SRR013495.fastq	3971263	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR013/SRR013495/SRR013495.fastq.gz
SRR013496.fastq	5115896	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR013/SRR013496/SRR013496.fastq.gz
SRR013497.fastq	6504569	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR013/SRR013497/SRR013497.fastq.gz
SRR013499.fastq	5884129	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR013/SRR013499/SRR013499.fastq.gz
SRR013500.fastq	9159886	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR013/SRR013500/SRR013500.fastq.gz
SRR013501.fastq	5684218	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR013/SRR013501/SRR013501.fastq.gz
SRR013528.fastq	4028371	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR013/SRR013528/SRR013528.fastq.gz
SRR013529.fastq	4559407	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR013/SRR013529/SRR013529.fastq.gz
SRR013531.fastq	4781540	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR013/SRR013531/SRR013531.fastq.gz
SRR013532.fastq	6475821	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR013/SRR013532/SRR013532.fastq.gz
SRR013533.fastq	6257147	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR013/SRR013533/SRR013533.fastq.gz
SRR014480.fastq	9548334	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR014/SRR014480/SRR014480.fastq.gz
SRR014482.fastq	6934214	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR014/SRR014482/SRR014482.fastq.gz
SRR014483.fastq	10843095	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR014/SRR014483/SRR014483.fastq.gz
SRR014485.fastq	17748216	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR014/SRR014485/SRR014485.fastq.gz
SRR016657.fastq	5297302	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR016/SRR016657/SRR016657.fastq.gz
SRR016658.fastq	5649145	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR016/SRR016658/SRR016658.fastq.gz

SRR016659.fastq	5858682	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR016/SRR016659/SRR016659.fastq.gz
SRR016660.fastq	5520965	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR016/SRR016660/SRR016660.fastq.gz
SRR020725.fastq	13714225	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR020/SRR020725/SRR020725.fastq.gz
SRR020730.fastq	9860543	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR020/SRR020730/SRR020730.fastq.gz
SRR026735.fastq	32979341	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR020/SRR020735/SRR020735.fastq.gz
SRR026736.fastq	30280269	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR020/SRR020736/SRR020736.fastq.gz
SRR026737.fastq	30140163	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR020/SRR020737/SRR020737.fastq.gz
SRR026739.fastq	40085387	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR020/SRR020739/SRR020739.fastq.gz
SRR026740.fastq	31755713	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR020/SRR020740/SRR020740.fastq.gz
SRR027721.fastq	10653262	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR027/SRR027721/SRR027721.fastq.gz
SRR027729.fastq	8253275	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR027/SRR027729/SRR027729.fastq.gz
SRR028687.fastq	4652065	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR028/SRR028687/SRR028687.fastq.gz
SRR028688.fastq	4560037	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR028/SRR028688/SRR028688.fastq.gz
SRR028689.fastq	4455967	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR028/SRR028689/SRR028689.fastq.gz
SRR029570.fastq	8563815	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR029/SRR029570/SRR029570.fastq.gz
SRR029571.fastq	7531808	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR029/SRR029571/SRR029571.fastq.gz
SRR066593.fastq	21006018	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR066/SRR066593/SRR066593.fastq.gz
SRR080711.fastq	14017947	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR080/SRR080711/SRR080711.fastq.gz
SRR080715.fastq	24197156	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR080/SRR080715/SRR080715.fastq.gz
SRR080716.fastq	22894550	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR080/SRR080716/SRR080716.fastq.gz
SRR080717.fastq	20673618	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR080/SRR080717/SRR080717.fastq.gz
SRR080718.fastq	19001046	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR080/SRR080718/SRR080718.fastq.gz
SRR080720.fastq	14795589	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR080/SRR080720/SRR080720.fastq.gz
SRR084762.fastq	25184235	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR084/SRR084762/SRR084762.fastq.gz
SRR084763.fastq	29554060	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR084/SRR084763/SRR084763.fastq.gz
SRR084765.fastq	24060050	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR084/SRR084765/SRR084765.fastq.gz
SRR084766.fastq	29958406	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR084/SRR084766/SRR084766.fastq.gz
SRR084767.fastq	10133748	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR084/SRR084767/SRR084767.fastq.gz
SRR084768.fastq	16672771	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR084/SRR084768/SRR084768.fastq.gz
SRR084769.fastq	40134954	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR084/SRR084769/SRR084769.fastq.gz
SRR084770.fastq	32869693	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR084/SRR084770/SRR084770.fastq.gz
SRR088652.fastq	21660781	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR088/SRR088652/SRR088652.fastq.gz
SRR088653.fastq	7079884	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR088/SRR088653/SRR088653.fastq.gz
SRR088654.fastq	24464213	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR088/SRR088654/SRR088654.fastq.gz
SRR090691.fastq	17476965	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR090/SRR090691/SRR090691.fastq.gz
SRR090692.fastq	14647692	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR090/SRR090692/SRR090692.fastq.gz
SRR090693.fastq	13287496	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR090/SRR090693/SRR090693.fastq.gz
SRR090695.fastq	13870449	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR090/SRR090695/SRR090695.fastq.gz
SRR090697.fastq	10980890	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR090/SRR090697/SRR090697.fastq.gz
SRR094511.fastq	1441633	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR094/SRR094511/SRR094511.fastq.gz
SRR094512.fastq	7657150	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR094/SRR094512/SRR094512.fastq.gz
SRR094513.fastq	7736706	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR094/SRR094513/SRR094513.fastq.gz

SRR094514.fastq	7583385	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR094/SRR094514/SRR094514.fastq.gz
SRR094516.fastq	773118	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR094/SRR094516/SRR094516.fastq.gz
SRR094517.fastq	5752685	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR094/SRR094517/SRR094517.fastq.gz
SRR094525.fastq	7444615	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR094/SRR094525/SRR094525.fastq.gz
SRR095649.fastq	9294528	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR095/SRR095649/SRR095649.fastq.gz
SRR095650.fastq	8779699	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR095/SRR095650/SRR095650.fastq.gz
SRR095652.fastq	20858642	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR095/SRR09552/SRR095652.fastq.gz
SRR095654.fastq	4846577	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR095/SRR09554/SRR095654.fastq.gz
SRR095655.fastq	5403390	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR095/SRR09554/SRR095654.fastq.gz
SRR095656.fastq	18082551	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR095/SRR09556/SRR095656.fastq.gz
SRR095669.fastq	14148223	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR095/SRR09569/SRR095669.fastq.gz
SRR096532.fastq	32542377	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR096/SRR096532/SRR096532.fastq.gz
SRR096533.fastq	28304102	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR096/SRR096533/SRR096533.fastq.gz
SRR096534.fastq	20135423	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR096/SRR096534/SRR096534.fastq.gz
SRR096535.fastq	16893565	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR096/SRR096535/SRR096535.fastq.gz
SRR096536.fastq	26572941	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR096/SRR096536/SRR096536.fastq.gz
SRR096538.fastq	15101283	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR096/SRR096538/SRR096538.fastq.gz
SRR096539.fastq	14559143	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR096/SRR096539/SRR096539.fastq.gz
SRR096540.fastq	13960496	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR096/SRR096540/SRR096540.fastq.gz
SRR097543.fastq	22199559	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR097/SRR097543/SRR097543.fastq.gz
SRR097544.fastq	18732099	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR097/SRR097544/SRR097544.fastq.gz
SRR097545.fastq	30471715	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR097/SRR097545/SRR097545.fastq.gz
SRR097547.fastq	32581224	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR097/SRR097547/SRR097547.fastq.gz
SRR097548.fastq	29890600	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR097/SRR097548/SRR097548.fastq.gz
SRR097549.fastq	6201090	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR097/SRR097549/SRR097549.fastq.gz
SRR097550.fastq	33301051	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR097/SRR097550/SRR097550.fastq.gz
SRR097556.fastq	25748009	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR097/SRR097556/SRR097556.fastq.gz
SRR097557.fastq	7897159	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR097/SRR097557/SRR097557.fastq.gz
SRR097558.fastq	34114251	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR097/SRR097558/SRR097558.fastq.gz
SRR097559.fastq	5145574	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR097/SRR097559/SRR097559.fastq.gz
SRR097560.fastq	18143915	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR097/SRR097560/SRR097560.fastq.gz
SRR097561.fastq	6980830	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR097/SRR097561/SRR097561.fastq.gz
SRR112679.fastq	30758792	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR112/SRR112679/SRR112679.fastq.gz
SRR112681.fastq	28366417	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR112/SRR112681/SRR112681.fastq.gz
SRR112684.fastq	22888982	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR112/SRR112684/SRR112684.fastq.gz
SRR112685.fastq	25292030	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR112/SRR112685/SRR112685.fastq.gz
SRR112687.fastq	25199157	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR112/SRR112687/SRR112687.fastq.gz
SRR121525.fastq	19855297	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR121/SRR121525/SRR121525.fastq.gz
SRR121527.fastq	24432741	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR121/SRR121527/SRR121527.fastq.gz
SRR121528.fastq	30911782	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR121/SRR121528/SRR121528.fastq.gz
SRR121530.fastq	31046209	http://ftp.sra.ebi.ac.uk/vol1/fastq/SRR121/SRR121530/SRR121530.fastq.gz

2 Pseudo code

Algorithm 1 Path graph construction

1: **procedure** CONSTRUCT(Vertex set V, Distance matrix D)
2:    $n \leftarrow$ `vertex number`
3:    $W \leftarrow$ `Weight Matrix`, $W[i][j] = e^{-D[i][j]}$
4:    $w \leftarrow$ `elements array`, $w[i] \leftarrow$ `the id of highest value in the` i`th row of` W
5:    $p \leftarrow$ `vertex array.`
6:    $W[a][b] \leftarrow$ `maximum value in` W
7:    $W[a][b] \leftarrow -1, W[b][a] \leftarrow -1$
8:    $p[1] \leftarrow a, p[2] \leftarrow b$
9:    **for** $i = 2; i < n$ **do** $i++$
10:       $p[i + 1] \leftarrow w[p[i]]$
11:       $W[p[i]][p[i + 1]] \leftarrow -1$
12:       $W[p[i + 1]][p[i]] \leftarrow -1$
13:    **return** p

---

3 N50 score of files corrected by Karect with different parameter values

In the graphs in this section, empty bar indicates that the N50 score of corresponding value was not computed correctly.
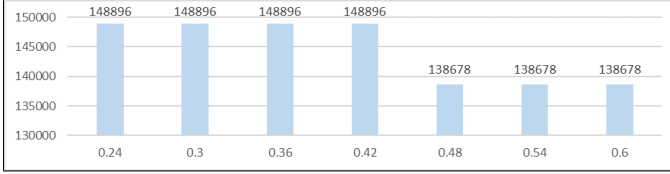
Figure 1: N50 score of **SRR354181** with different **aggressive** values of Karect
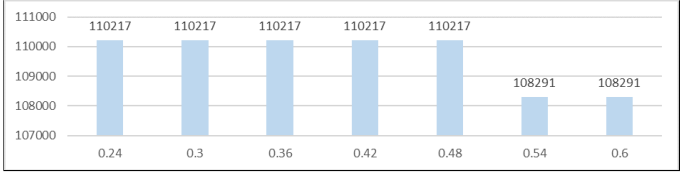
Figure 2: N50 score of **SRR354182** with different **aggressive** values of Karect
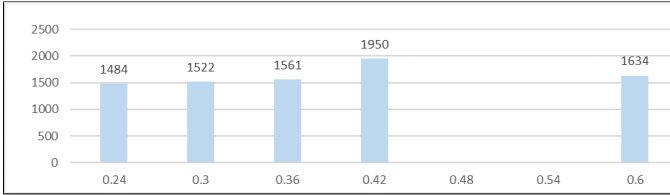
Figure 3: N50 score of **SRR361240** with different **aggressive** values of Karect
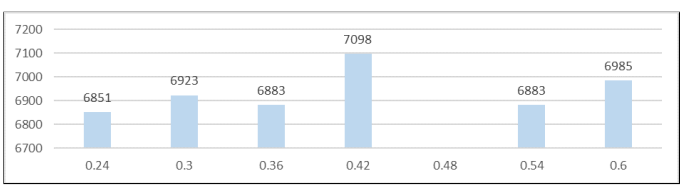
Figure 4: N50 score of **SRR361241** with different **aggressive** values of Karect
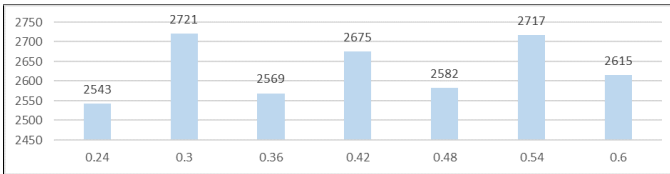
Figure 5: N50 score of **SRR361243** with different **aggressive** values of Karect
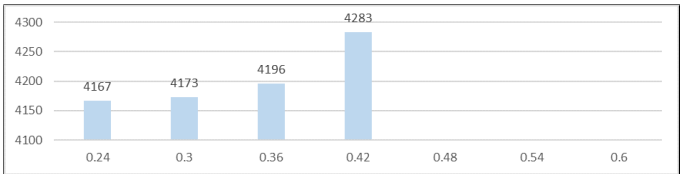
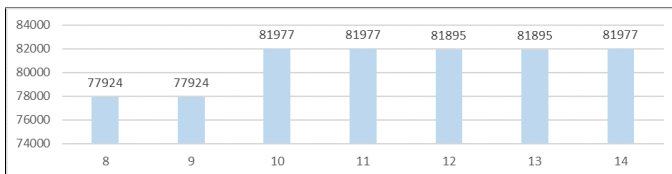Figure 6: N50 score of **SRR361245** with different **aggressive** values of Karect
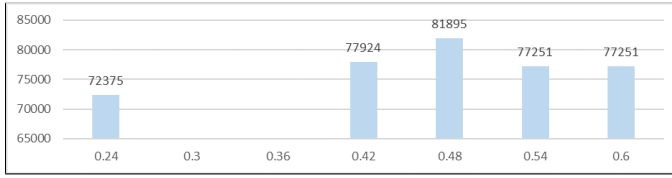
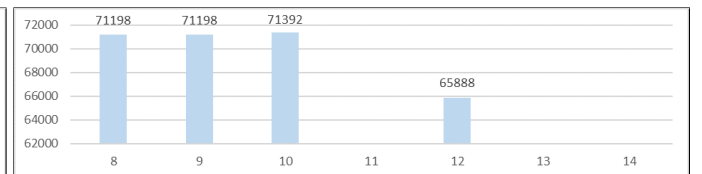Figure 7: N50 score of **SRR387479** with different **aggressive** values of Karect
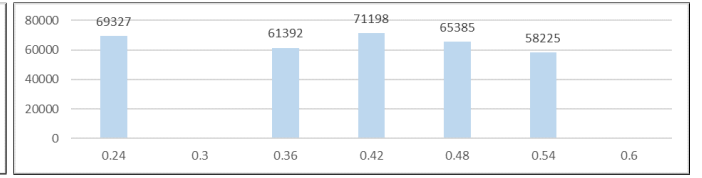
Figure 8: N50 score of **SRR387480** with different **aggressive** values of Karect
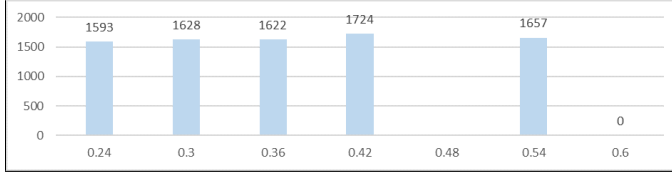
Figure 9: N50 score of **SRR401414** with different **aggressive** values of Karect
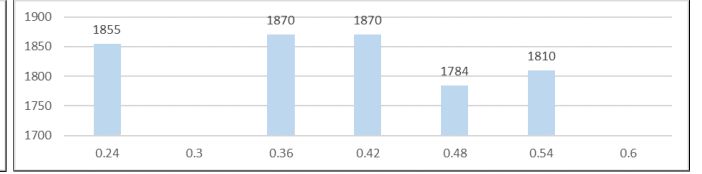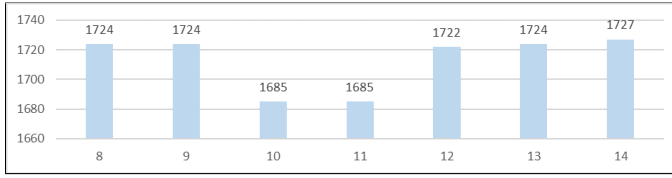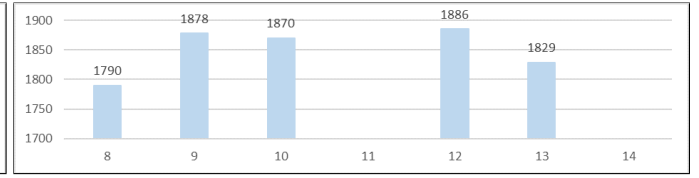

Figure 10: N50 score of **SRR401415** with different **aggressive** values of Karect
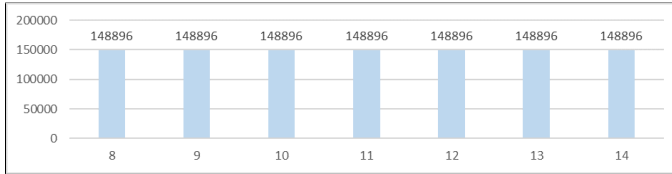
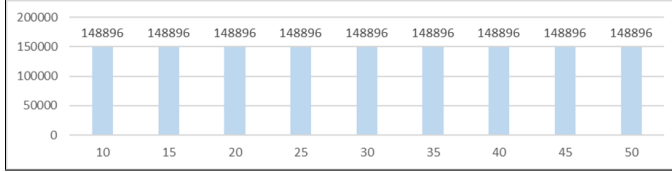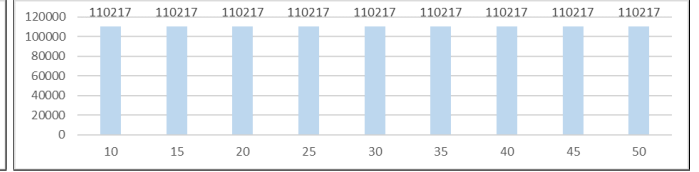Figure 11: N50 score of **SRR354181** with different **kmer** values of Karect
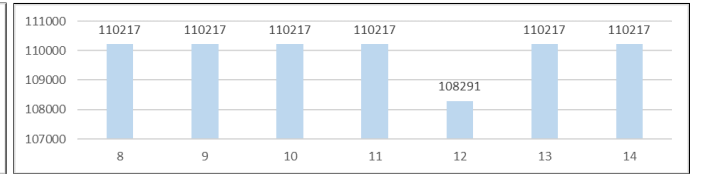
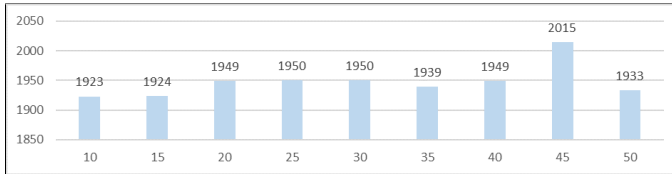Figure 12: N50 score of **SRR354182** with different **kmer** values of Karect
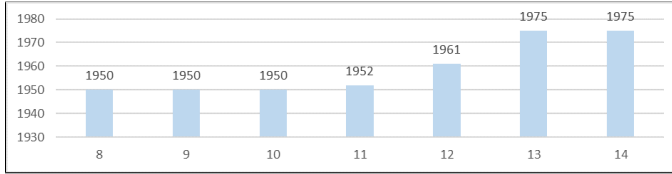
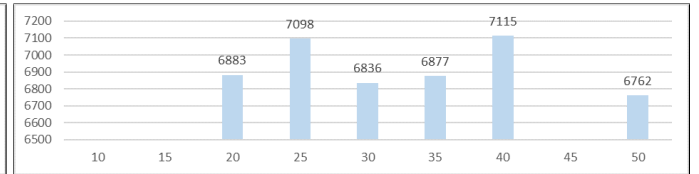Figure 13: N50 score of **SRR361240** with different **kmer** values of Karect
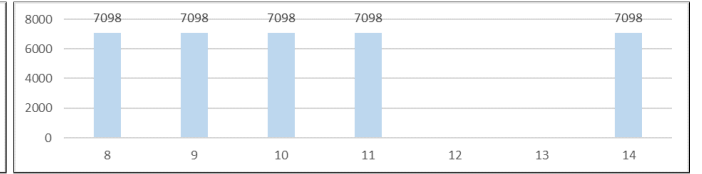
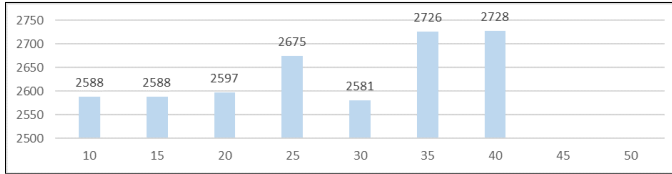Figure 14: N50 score of **SRR361241** with different **kmer** values of Karect
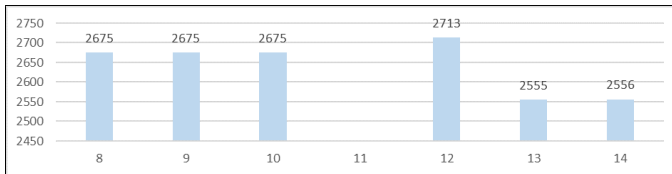
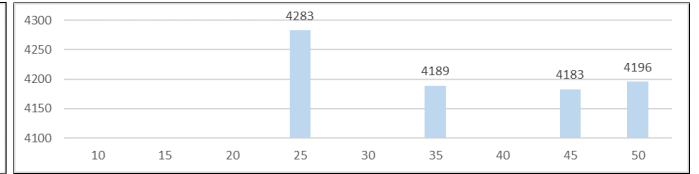Figure 15: N50 score of **SRR361243** with different **kmer** values of Karect
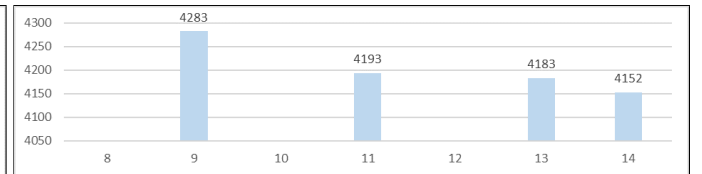
Figure 16: N50 score of **SRR361245** with different **kmer** values of Karect

Figure 17: N50 score of **SRR387479** with different **kmer** values of Karect

Figure 18: N50 score of **SRR387480** with different **kmer** values of Karect

Figure 19: N50 score of **SRR401414** with different **kmer** values of Karect

Figure 20: N50 score of **SRR401415** with different **kmer** values of Karect

Figure 21: N50 score of **SRR354181** with different **kmer** values of Karect

Figure 22: N50 score of **SRR354182** with different **kmer** values of Karect

Figure 23: N50 score of **SRR361240** with different **kmer** values of Karect

Figure 24: N50 score of **SRR361241** with different **kmer** values of Karect

Figure 25: N50 score of **SRR361243** with different **kmer** values of Karect

Figure 26: N50 score of **SRR361245** with different **kmer** values of Karect
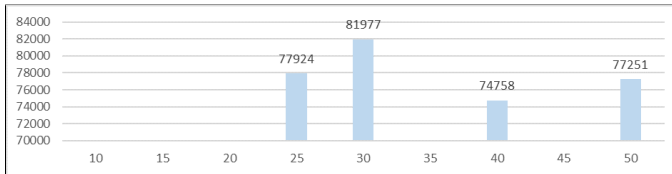

Figure 27: N50 score of **SRR387479** with different **kmer** values of Karect
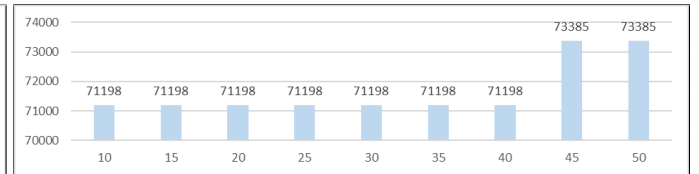

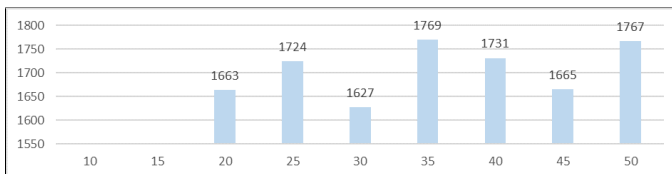Figure 28: N50 score of **SRR387480** with different **kmer** values of Karect

Figure 29: N50 score of **SRR401414** with different **kmer** values of Karect
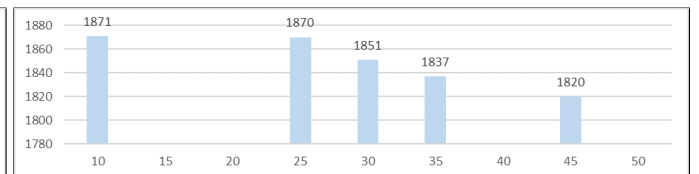

Figure 30: N50 score of **SRR401415** with different **kmer** values of Karect