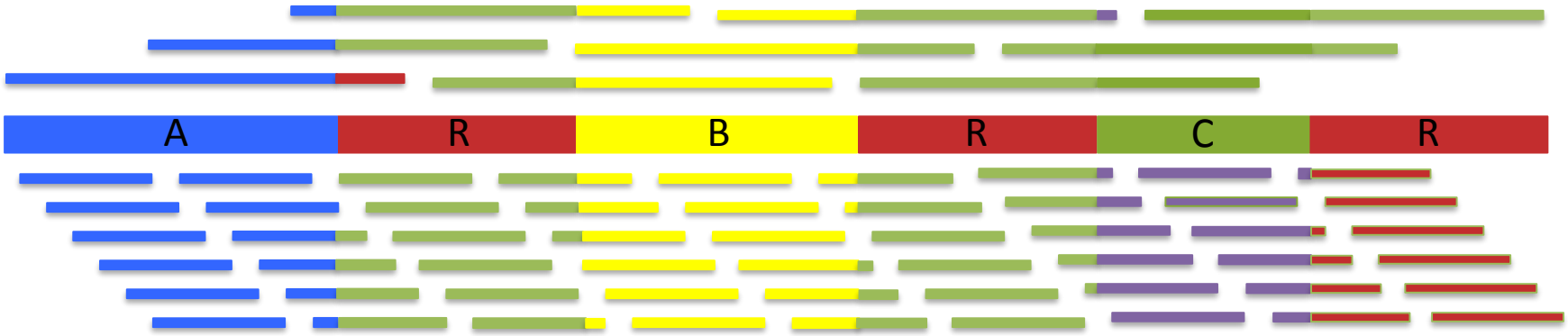
Preparing extra-long DNA fragments for RS II and Sequel applications

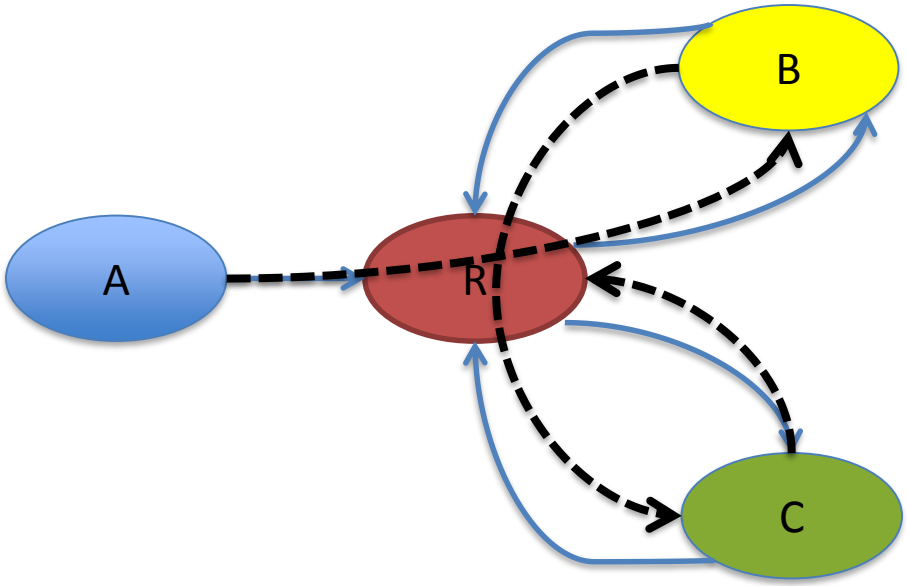
Sara Goodwin. Ph.D.

Pacific Biosciences user group meeting

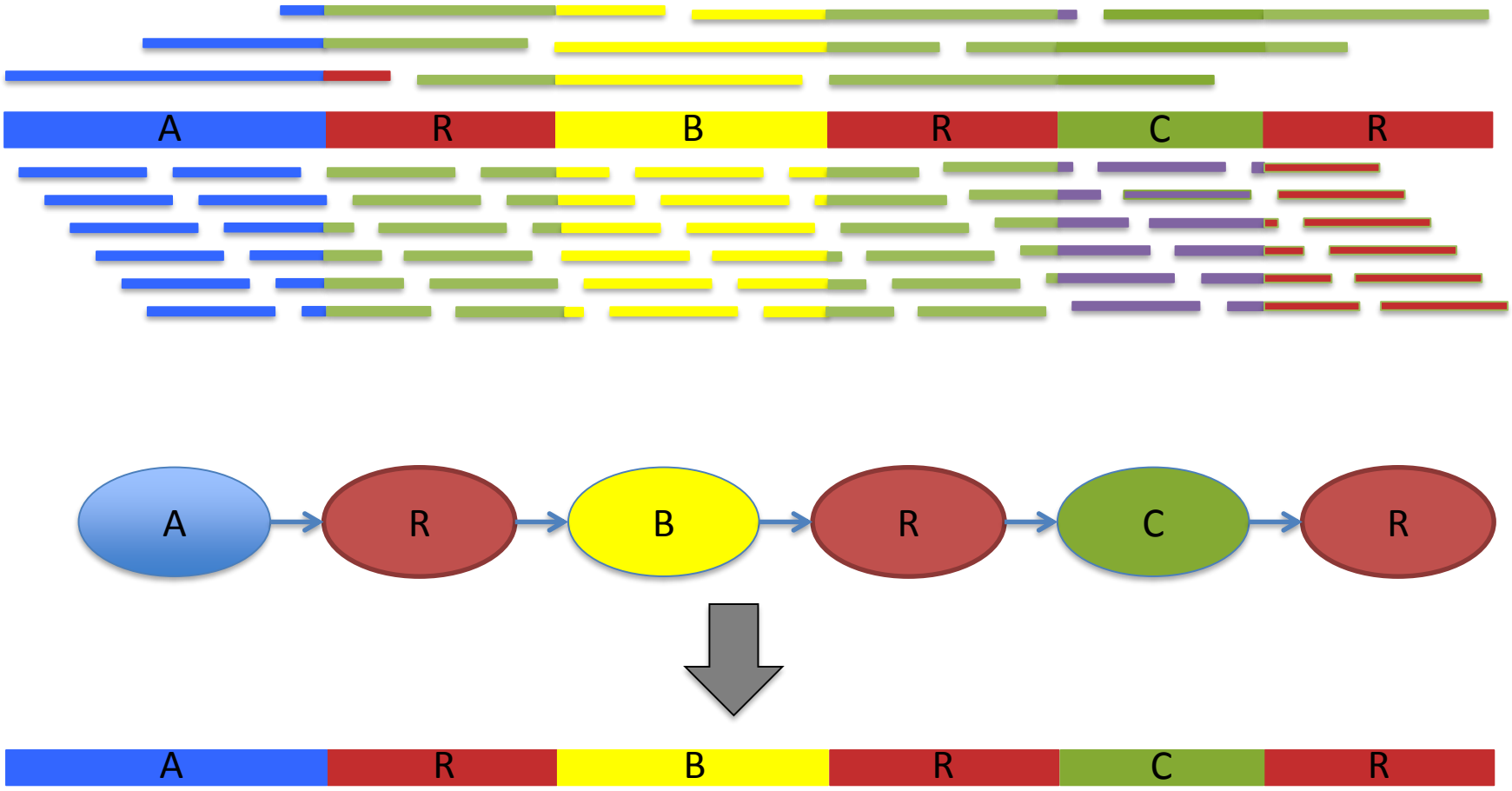
6/27/17

Assembly Complexity





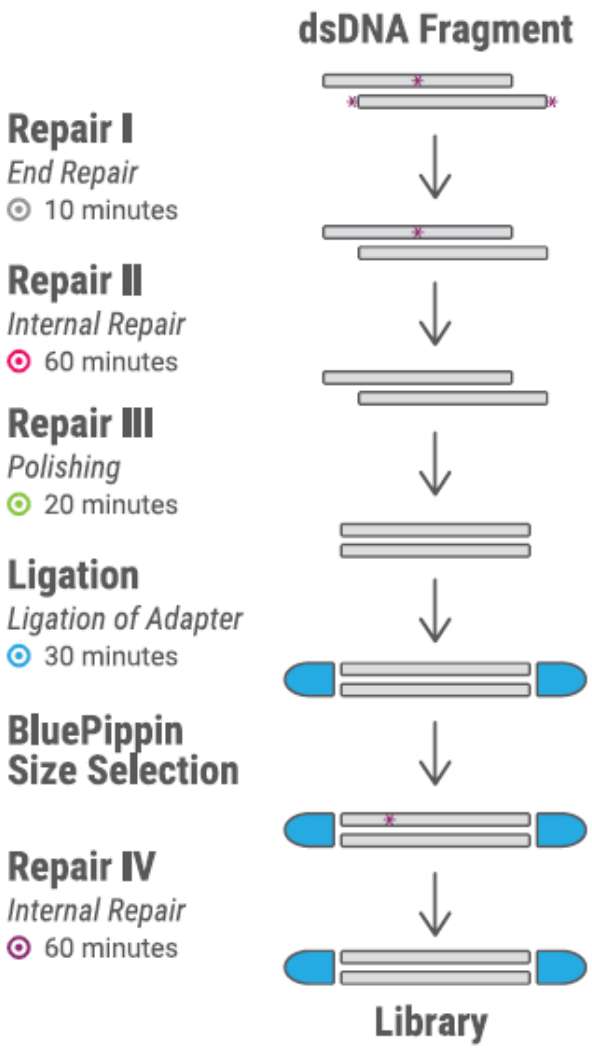
Assembly Complexity

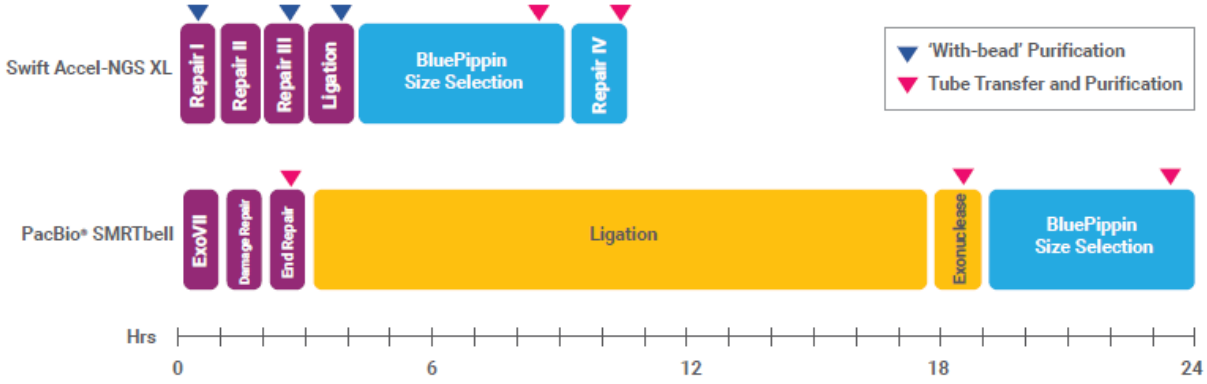


The advantages of SMRT sequencing
Roberts, RJ, Carneiro, MO, Schatz, MC (2013) *Genome Biology.* 14:405

Essential steps for long library generation

Long Read Library prep with Swift Accel NGS XL





- Excluding size selection, library prep is 5 hours
- Not compatible with small fragments
- All steps prior to size selection occur “on-bead”

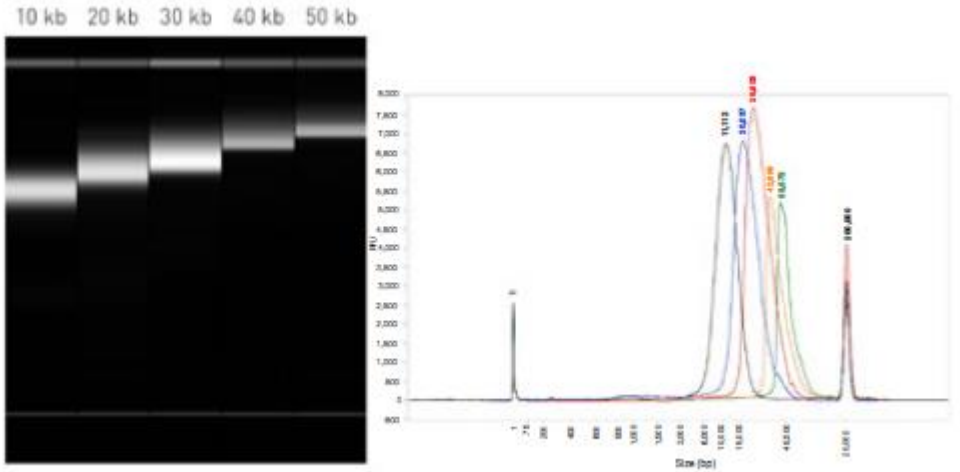
Library preparation yield

Prep	Size	Starting	Pre Size Selection	Post Size Selection	Recovery
Standard Maize	20kb	10ug	2.5ug	500ng	5.0%
Swift Maize	30kb	4.5ug	1.1ug	440ng	9.5%
Standard Grape	30kb	25ug	4.4ug	880ng	3.5%
Swift Grape	30kb	5ug	1.7ug	214ng	4.2%

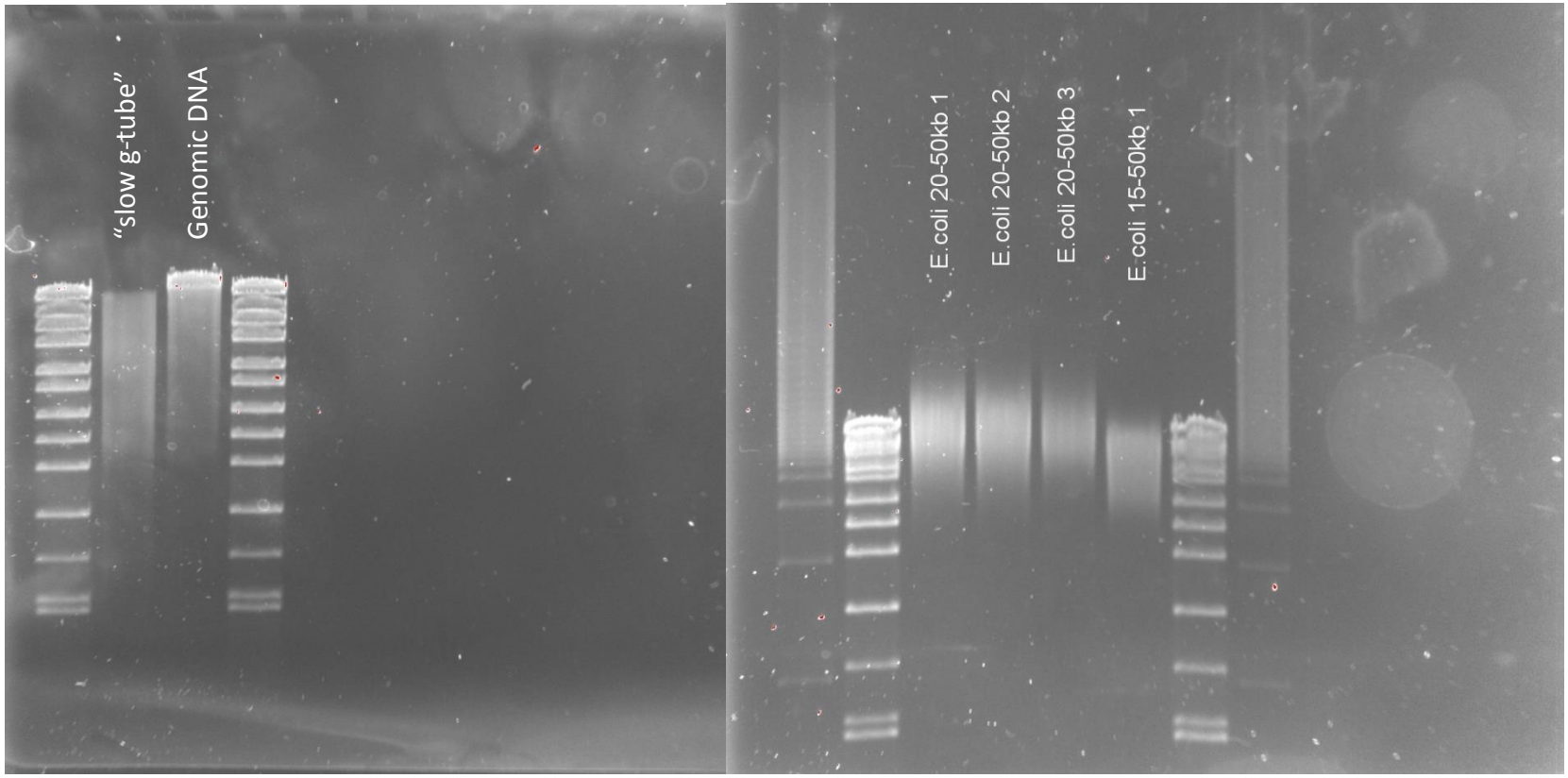
Sequencing performance

Prep	Subreads per cell	Yield	Mean	N50	Reads >10kb	Bases in Reads >10kb	Reads >20kb	Bases in reads >20kb	Proportion of bases >20kb
Standard Maize	623383	868Mb	6152	10076	38694	500Mb	1555	38Mb	4.3%*
Swift Maize	53540	670Mb	12596	19265	28222	550Mb	11824	317Mb	47.3%
Standard Grape	26708	300Mb	6386	11152	5221	91Mb	1522	39Mb	13.5%
Swift Grape	97095	941Mb	9698	12786	41455	638Mb	6050	148Mb	15.7%

Shearing with Megarupter



PFGE of various fragmentation strategies



Library preparation

Sample	Starting	Fragmentation	Post ER	Post Exo	Size selection	Final	Recovery
Standard	10ug	20kb	4.6ug	2ug	10-50kb	0.5ug	5.0%
Mega 30	20ug	30kb	12.1ug	5.6ug	15-50kb	1ug	5.0%
Mega 50	30ug	50kb	22ug	12ug	20-50kb	1.5ug	6.8%

Must have at least 5ug prior to size selection for 50kb prep

Megarupter sequencing results RSII

Experiment	Loading	reads	Bases	mean	N50	reads>10 kb	bases >10kb	reads>20 kb	bases >20kb	reads>30 kb	bases >30kb
slow gtube (ecoli)	300pM	79412	0.93Gb	11697.93	17087	41223	0.74Gb	12788	0.32Gb	1851	0.06
mega 50 (e coli)	300pM	24986	0.30Gb	11875.54	18869	12318	0.24Gb	5079	0.13Gb	1107	0.04
mega 30 (e coli)	300pM	69476	0.72Gb	10324.20	16158	29867	0.54Gb	9603	0.24Gb	1433	0.05
Standard (Maize)	150pM	181900	1.69Gb	9292.23	12618	79336	1.19Gb	9493	0.24Gb	1385	0.05

experiment	% of reads >10kb	%of bases over 10kb	% of reads >20kb	%of bases over 20kb	% of reads >30kb	%of bases over 30kb
slow gtube ecoli	52%	80%	16%	35%	2%	7%
mega 50 (e coli)	49%	81%	20%	45%	4%	13%
mega 30 (e coli)	43%	75%	14%	34%	2%	7%
standard (Maize)	44%	70%	5%	14%	1%	3%

Megarupter sequencing results Sequel

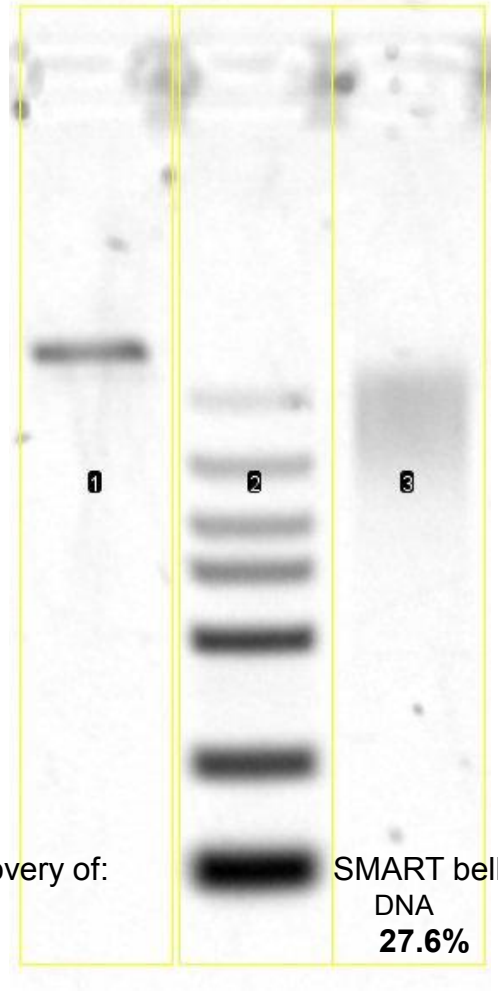
experiment	loading	reads	bases	mean	N50	reads>10kb	bases >10kb	reads>20kb	bases >20kb	reads>30kb	bases >30kb
Human 20kb	10pM	331312	2.60	7847	12309	114535	1.71	12778	0.32	1636	0.06
Human 20kb	15pM	466467	3.61	7737	11920	157048	2.30	15644	0.39	1922	0.07
Human 20kb	20pM	628679	4.61	7326	11621	191771	2.81	20034	0.50	2433	0.09
Human 20kb	40pM	686344	4.88	7112	11504	199638	2.93	20493	0.51	2747	0.10
Maize 30kb mega	10pM	167241	1.60	9548	15711	68657	1.19	17122	0.43	2201	0.08
Maize 30kb mega	15pM	255887	2.36	9219	15105	100555	1.71	23078	0.58	3157	0.11
Maize 30kb mega	20pM	282556	2.54	9001	14745	108616	1.82	22809	0.56	2787	0.10
Maize 30kb mega	40pM	481594	4.12	8563	14264	171853	2.86	36723	0.91	4542	0.16
E coli 50kb sequel	40pM	196924	1.97	10024	16012	80959	1.47	26216	0.68	5054	0.18
E coli 50kb sequel	40pM	177945	1.74	9775	16111	69513	1.28	23863	0.62	4645	0.16
E coli 50kb sequel	60pM	207223	2.17	10484	17455	88328	1.67	33460	0.88	6953	0.24
E coli 50kb sequel	80pM	302294	2.67	8832	16186	103545	1.93	37116	0.98	7992	0.28

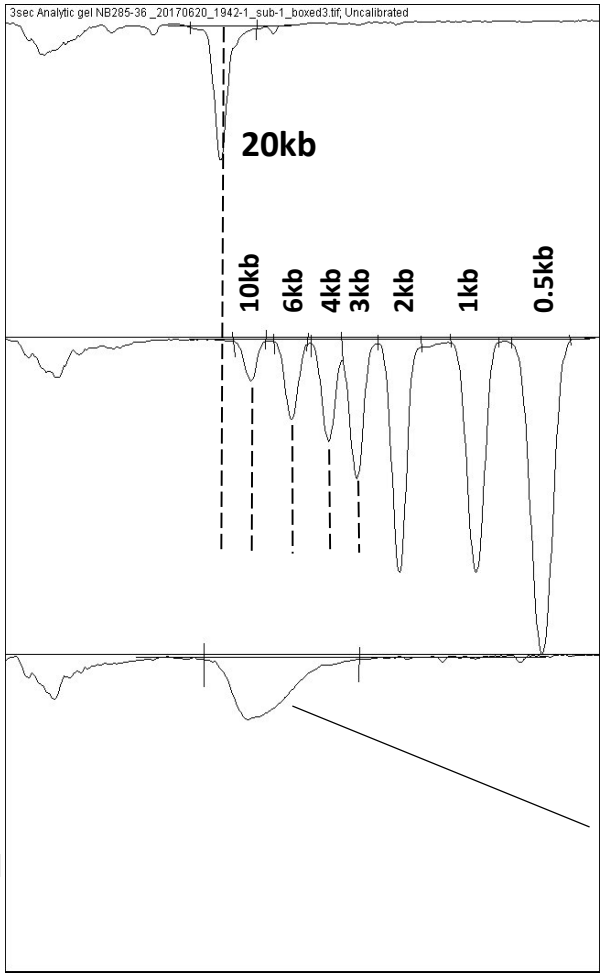
Experiment	% of reads >10kb	%of bases over 10kb	% of reads >20kb	%of bases over 20kb	% of reads >30kb	%of bases over 30kb
Human 20kb	34.57%	65.63%	3.86%	12.25%	0.49%	2.25%
Human 20kb	33.67%	63.69%	3.35%	10.75%	0.41%	1.91%
Human 20kb	30.50%	61.10%	3.19%	10.80%	0.39%	1.91%
Human 20kb	29.09%	60.03%	2.99%	10.52%	0.40%	2.06%
Maize 30kb mega	35.68%	69.43%	7.63%	22.06%	0.94%	3.77%
Maize 30kb mega	38.44%	71.42%	8.07%	22.18%	0.99%	3.79%
Maize 30kb mega	39.30%	72.54%	9.02%	24.44%	1.23%	4.62%
Maize 30kb mega	41.05%	74.29%	10.24%	26.64%	1.32%	4.74%
E coli 50kb sequel	42.62%	76.96%	16.15%	40.48%	3.36%	11.13%
E coli 50kb sequel	41.11%	74.27%	13.31%	34.52%	2.57%	8.89%
E coli 50kb sequel	39.06%	73.34%	13.41%	35.73%	2.61%	9.24%
E coli 50kb sequel	34.25%	72.34%	12.28%	36.66%	2.64%	10.38%

New Size Selection Strategies

SMART belled DNA (0.29ug) Recovered from 1.39ug

Ln	20kb DNA Std 5	Rev.Mass Ladder Std 6	SMART belled DNA sample 7





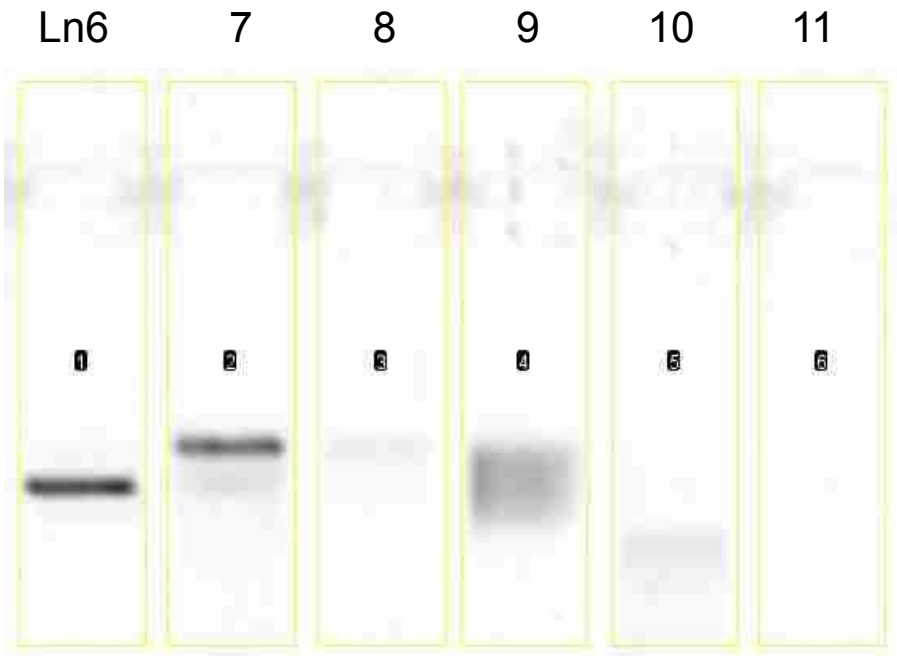
Recovered SMART belled DNA

Recovery of: SMART belled DNA **27.6%**

· SMART belled DNA recovery following size selection procedure.

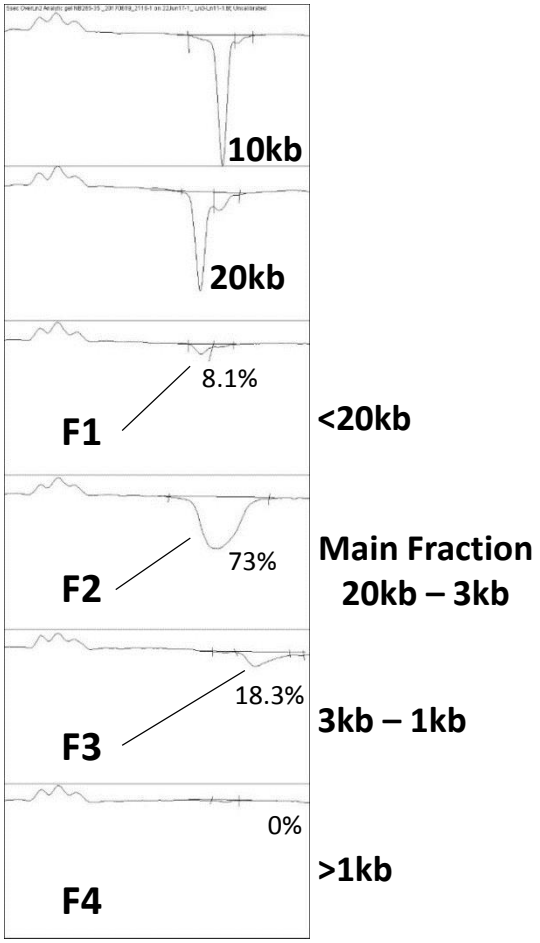
Recovery of “10kb” sheared E coli Genomic DNA, fractionated

Recovery in Ln6 (10kb DNA Std), Ln7 (20kb DNA Std),
Ln8 (Fraction1 ->20k), Ln9(Fraction2 – Main (20kb-3kb)),
Ln10(Fraction3-(3kb-1kb), Ln11(Fraction4 - <1kb)



Recovery of:

10kb Std	20kb Std	F1 (**>20kb**)	F2 (Main)	F3 (**3kb-1kb**)	F4 (**<1kb**)
39%	39%	8.1%	**73%**	18.3%	0%

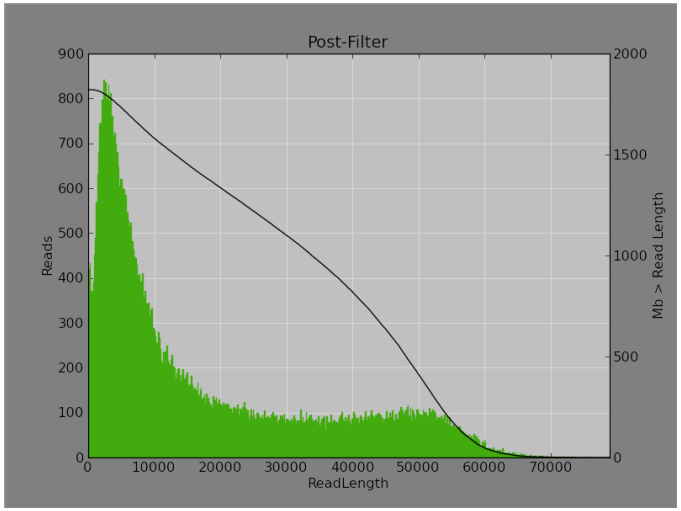


· Multiple fractions recovered simultaneously – (2-3hrs)

Library and Sequencing performance

Sample	Starting	Fragmentation	Post Exo	Size selection	Final	Recovery
Standard	10ug	20kb	2ug	10-50kb	0.5ug	5%
10 kb e coli	10ug	10kb	2.2ug	>500bp	~0.3ug	3%

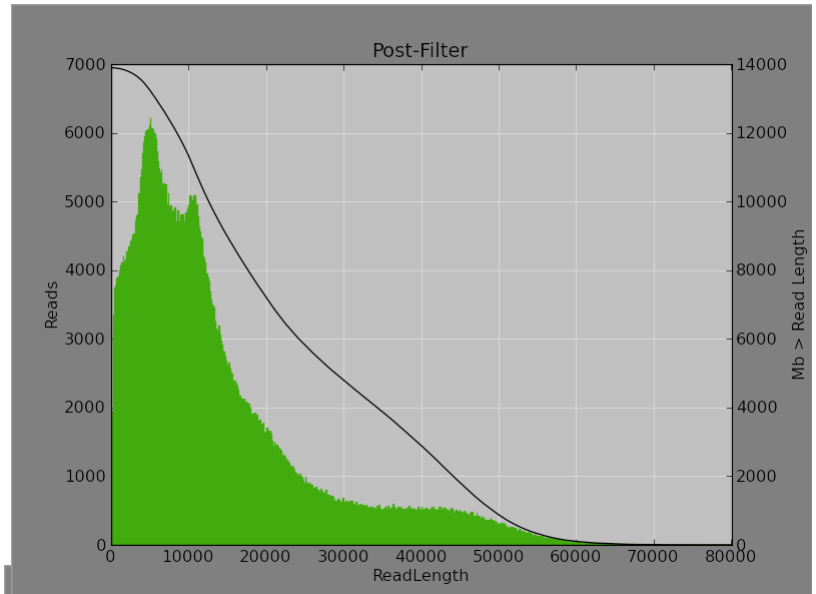
Experiment	Loading	reads	bases	mean	N50	reads>10 kb	bases >10kb
Standard (Maize)	150pM	181900	1.69	9292.23	12618	79336	1.19
10 kb e coli	100pM	515957	1.81	3499	5000	15119	0.20

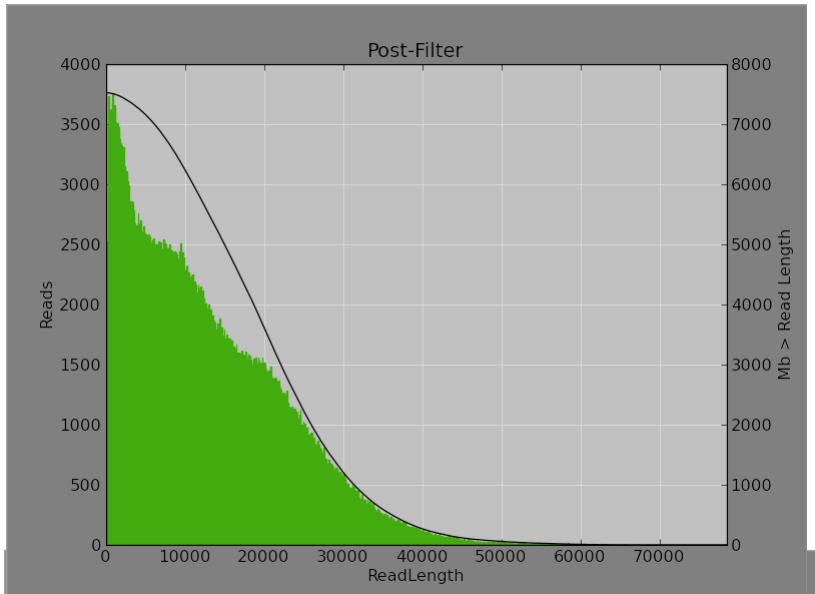


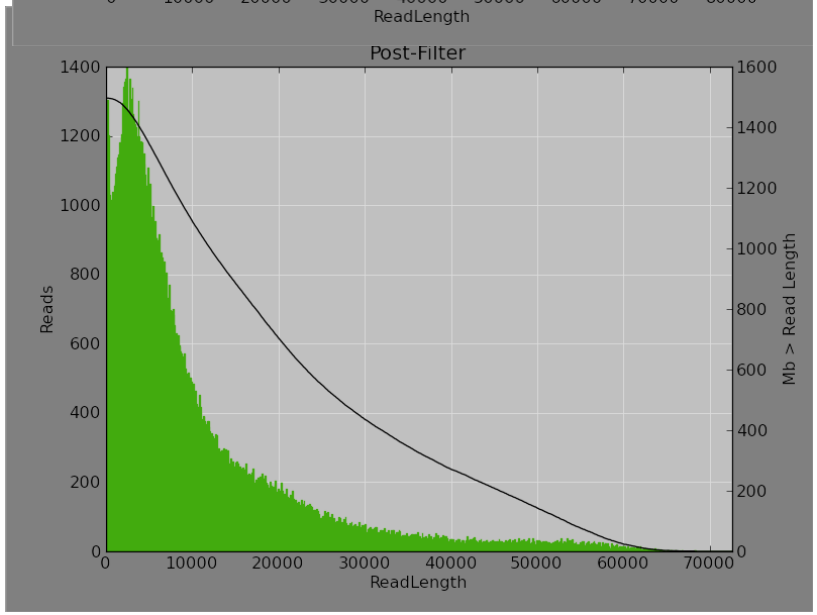
Acknowledgments

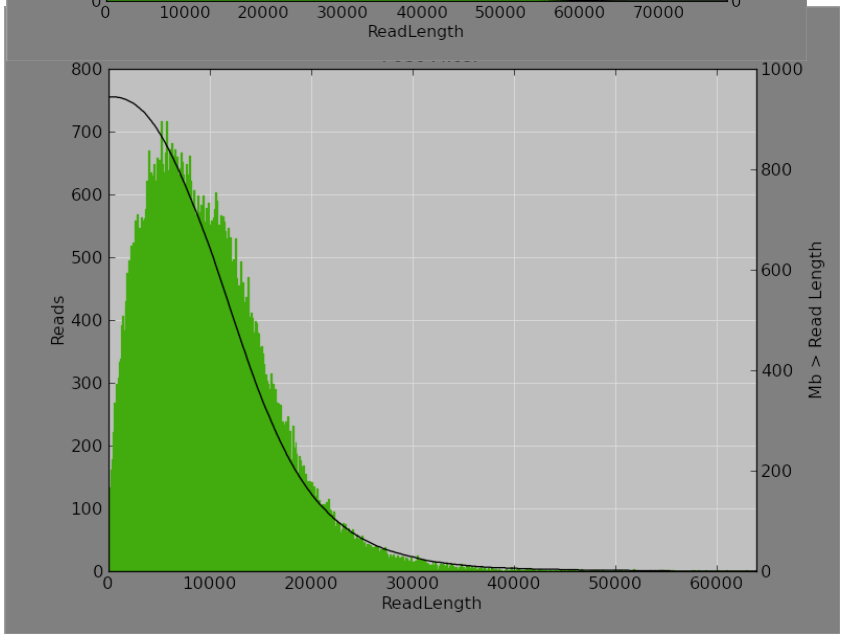
W. Richard McCombie, Senem Mavruk, Robert Wappel

Subread distribution Swift

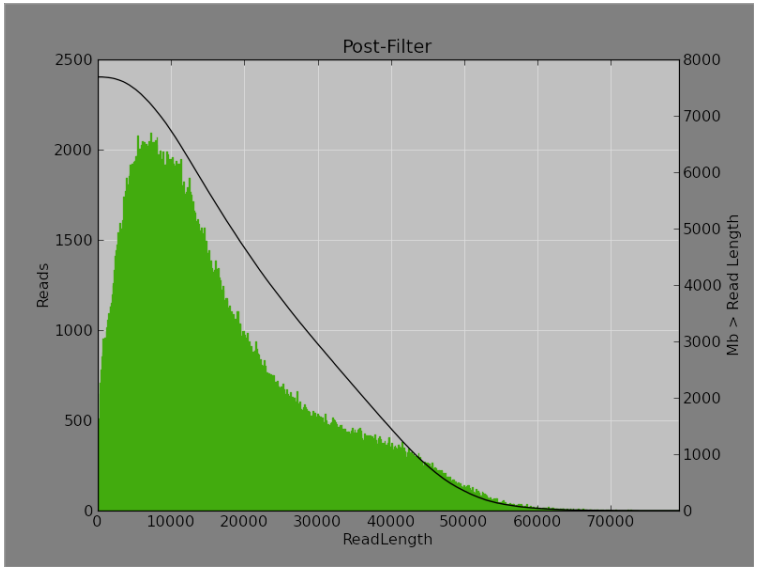


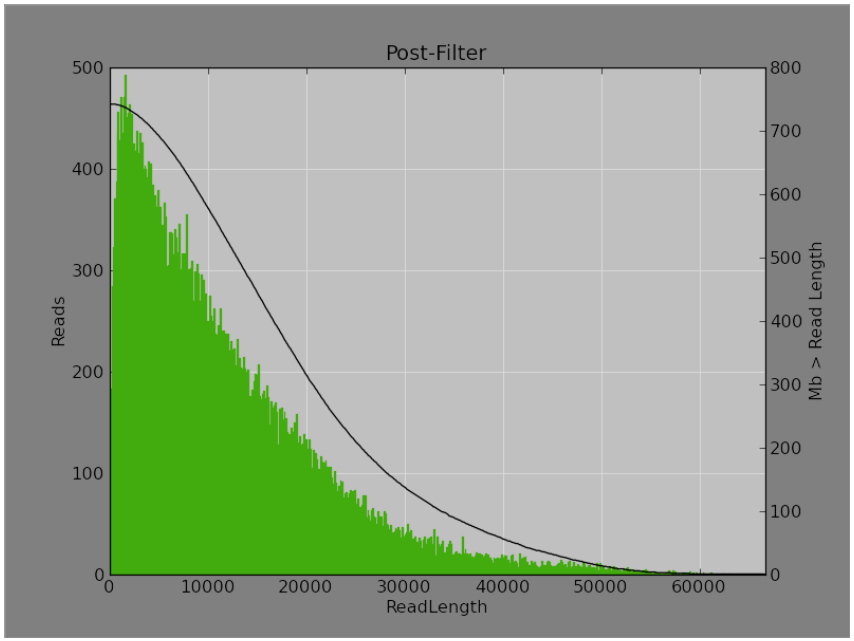


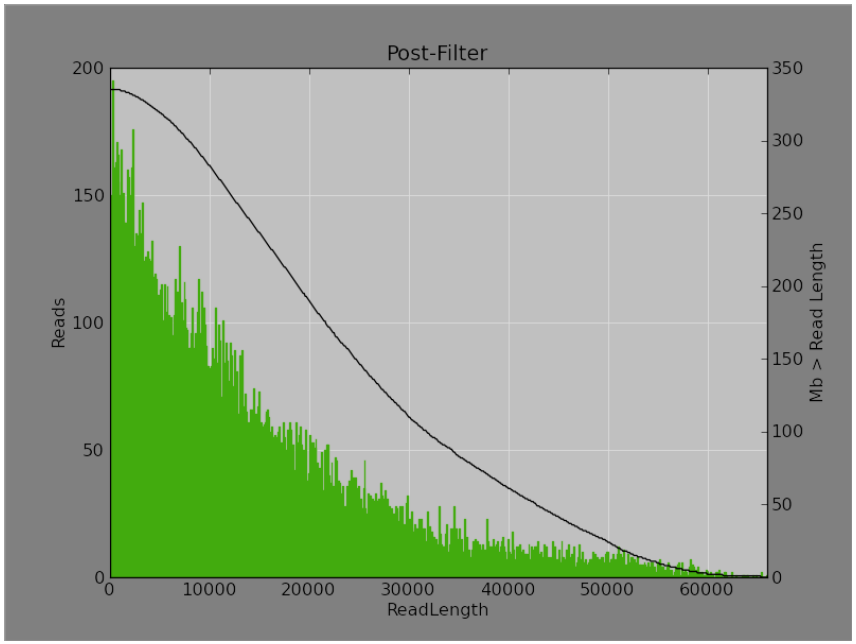




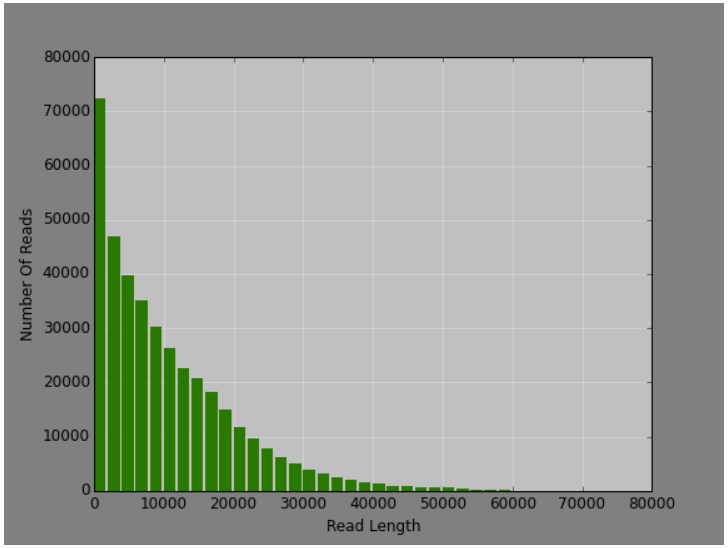
RS II

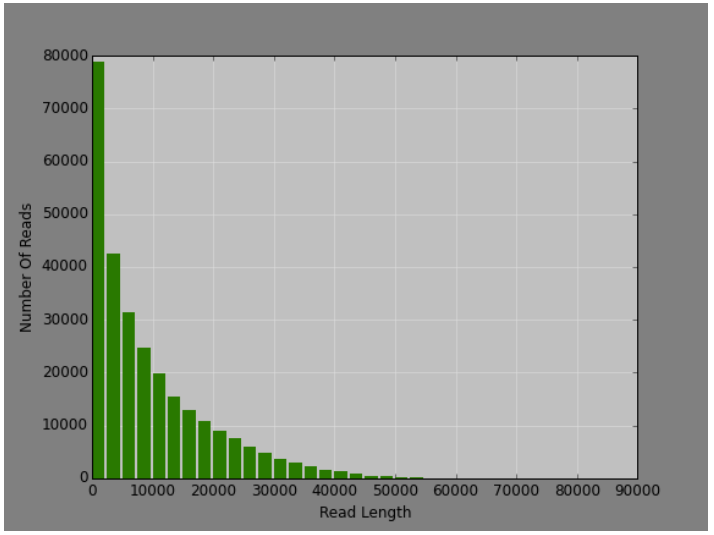

Post-Filter
Reads
Mb > Read Length
ReadLength
0
500
1000
1500
2000
2500
1000
2000
3000
4000
5000
6000
7000
8000
10000
20000
30000
40000
50000
60000
70000


Post-Filter
Reads
Mb > Read Length
ReadLength
0
100
200
300
400
500
100
200
300
400
500
600
700
800
10000
20000
30000
40000
50000
60000


Post-Filter
Reads
Mb > Read Length
ReadLength
0
50
100
150
200
50
100
150
200
250
300
350
10000
20000
30000
40000
50000
60000

Sequel


Number Of Reads
Read Length
0
10000
20000
30000
40000
50000
60000
70000
80000


Number Of Reads
Read Length
0
10000
20000
30000
40000
50000
60000
70000
80000
90000