

MOLECULAR ECOLOGY RESOURCES

Supplemental Information for:

In-Depth Comparative Analysis of Illumina® MiSeq Run Metrics: Development of a Wet-lab Quality Assessment Tool

George John Kastanis, Luis V. Santana-Quintero, Maria Sanchez-Leon, Sara Lomonaco,
Eric W. Brown and Marc W. Allard

Table of Contents:

Table S1	Page 2
Table S2	Page 3
Figure S1	Page 4
Figure S2	Page 5
Figure S3	Page 6

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32

MOLECULAR ECOLOGY

RESOURCES

33
34
35
36
37

Table S1. Eigenvalues of the PCA components. The first five principal components, indicated in bold, contain 86.41% of the variability of the 486 MiSeq run data set.

Principal Component	Eigenvalue	Percentage of Variance	Cumulative
1	5.164	34.43%	34.43%
2	4.244	28.29%	62.72%
3	1.52	10.13%	72.86%
4	1.213	8.09%	80.95%
5	0.82	5.47%	86.41%
6	0.522	3.48%	89.89%
7	0.45	3.00%	92.89%
8	0.353	2.35%	95.25%
9	0.3	2.00%	97.25%
10	0.173	1.15%	98.40%
11	0.118	0.78%	99.19%
12	0.063	0.42%	99.61%
13	0.045	0.30%	99.90%
14	0.009	0.06%	99.97%
15	0.005	0.03%	100.00%

MOLECULAR ECOLOGY RESOURCES

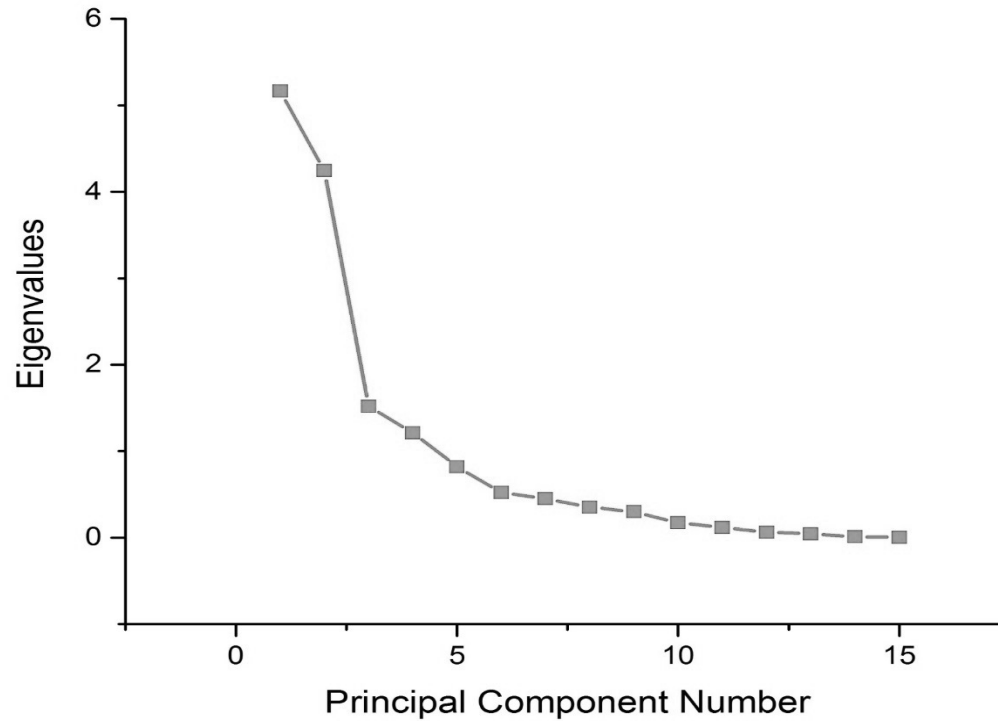
38
39
40 **Table S2. PCA coefficient table of the extracted Eigenvalues of the 15 selected MiSeq run metrics.** The top three PCA coefficients
41 for each principal component are in bold.
42

	Coefficients of				
	PC1	PC2	PC3	PC4	PC5
Clusters PF	0.168	0.287	0.272	-0.283	0.066
%\geqQ30 (Overall)	0.188	0.400	-0.092	0.145	0.130
%\geqQ30 (R1)	0.186	0.378	-0.177	0.209	-0.002
%\geqQ30 (IR1)	0.200	0.225	0.383	-0.132	0.010
%\geqQ30 (IR2)	0.254	0.204	0.233	-0.332	0.124
%\geqQ30 (R2)	0.178	0.382	-0.045	0.092	0.214
Phasing (R1)	-0.162	-0.158	0.018	0.120	0.939
Prephasing (R1)	-0.166	-0.220	0.383	-0.437	0.088
Phasing (R2)	-0.155	0.103	0.403	0.598	-0.014
Prephasing (R2)	-0.177	0.008	0.585	0.295	-0.143
Total Reads	0.328	-0.300	0.020	0.143	0.027
Reads PF	0.386	-0.205	0.120	0.054	0.024
Total Yield	0.383	-0.217	0.108	0.053	0.005
Q30 Yield	0.412	-0.128	0.063	0.051	0.063
Cluster Density	0.289	-0.311	-0.039	0.202	-0.044

43
44
45
46

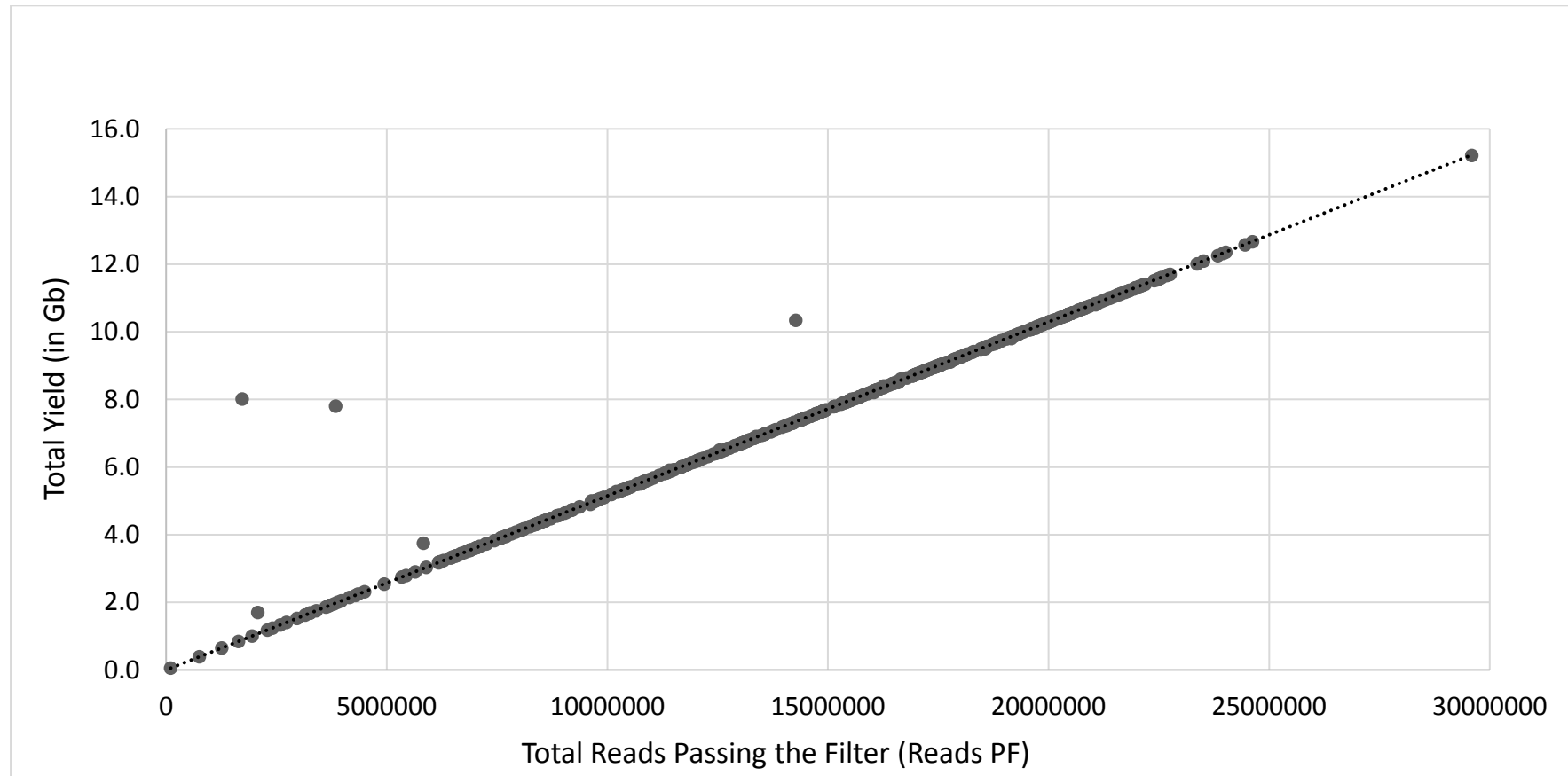
MOLECULAR ECOLOGY RESOURCES

47
48
49 **Figure S1. Scree plot of the 15 observed MiSeq run metrics.** The Scree plot generated shows its characteristic “elbow” in the vicinity
50 of the fifth component out of the 15. This signifies that the majority of the data can be expressed using only the first five components of
51 the data set.
52



MOLECULAR ECOLOGY RESOURCES

54
55
56 **Figure S2. Graph of the Total Yield (in Gb) versus the number of Total Reads Passing the Filter (Reads PF).** A single point
57 represents a MiSeq run performed and data points are shown for all five MiSeq instruments. R^2 value of the graph is 0.9742. Runs away
58 from the straight line depicts MiSeq runs that presented issues properly reporting into the Illumina® SAV software.
59



60

MOLECULAR ECOLOGY RESOURCES

Figure S3. Cluster Density range depicted with K-means. MiSeq runs outside the manufacturer CD range (1000 K/mm²-1200 K/mm²) shown in red and runs within manufacturer recommended CD range shown in blue.

