

Supplement Table 3 Preprocessing statistics and quality control of raw data.

Sample Name	Raw PE	Combined	Qualified	Nochime	Base (nt)	AvgLen (nt)	Q20	Q30	GC%	Effective%
E11	90,693	83,851	82,127	75,343	31,620,447	420	98.1	94.12	51.28	83.07
E18	94,082	87,650	86,063	80,542	33,579,594	417	98.05	94.09	52.05	85.61
E19	101,661	92,017	89,299	85,307	36,421,597	427	98.01	93.88	54.73	83.91
E20	91,545	81,651	78,836	74,659	31,777,527	426	97.95	93.79	54.17	81.55
E28	96,769	89,544	87,608	82,673	35,158,329	425	97.98	93.76	52.42	85.43
E3	107,367	100,471	97,863	91,504	39,113,449	427	97.91	93.71	54.07	85.23
E5	105,345	97,691	95,504	64,863	27,613,407	426	97.93	93.69	53.81	61.57
E6	94,646	88,902	86,774	80,948	34,622,531	428	97.9	93.67	54.85	85.53
E7	98,520	91,219	89,091	82,039	34,812,668	424	97.97	93.8	54.21	83.27
E8	93,450	86,419	84,812	80,618	33,216,736	412	98.15	94.35	53.04	86.27
P13	108,303	100,274	97,115	92,195	39,403,706	427	98.04	93.96	54.78	85.13
P14	97,195	89,917	87,740	84,080	35,982,347	428	98	93.86	56.02	86.51
P16	109,018	95,745	90,882	86,091	36,703,066	426	97.53	92.84	53	78.97
P17	97,802	83,198	79,708	73,967	31,507,846	426	98.14	94.14	54.51	75.63
P18	101,244	93,624	90,982	87,555	37,446,833	428	98.01	93.89	54.86	86.48
P19	109,382	89,521	84,283	80,600	34,399,954	427	98.04	93.92	55.76	73.69
P21	91,237	72,410	67,575	64,255	27,417,444	427	98.02	93.94	55.5	70.43
P27	98,535	85,374	81,879	77,898	33,270,778	427	98.03	93.89	53.07	79.06
P4	96,882	90,983	88,896	85,551	36,617,810	428	97.97	93.8	55.78	88.3
P6	95,008	85,188	82,342	75,868	32,384,696	427	97.88	93.6	53.19	79.85

Note; E8-28: CDL group; P4-P27: CCA group. All sequences data were uploaded on Mendeley

(<https://data.mendeley.com/datasets/y4fm2gbdzf/1>).

Raw PE (Pair End), original data derived from the indicated primer pair; Combined, original data were assembled by overlapped sequence; Qualified, low-grade quality and short-length reads were filtered out to generate qualified data; Nochime, chimera of qualified sequences were filtered out to generate effective sequences; Base, the base number of total data; AvgLen, average length of qualified tags; Q20, the percentage of bases with sequencing error rate <1%; Q30, the percentage of bases with sequencing error rate <.1%; GC%, GC content; Effective%, the percentage of effective tags."