

Eukaryotes, read depth

Michel de Lange

01/11/2021

Accumulation Curves within Samples.

We want to establish whether the actual read depth was sufficient. Are we likely to have found more phyla if we had had more reads? In order to answer this question, I have done a set of simulations. Within each simulation, I have subsampled each Sample until the total number of observed phyla in the sample had been found. I then plotted one accumulation curve as an example, and I collected statistics of all accumulation curve: The number of OTUs that needed to be read before we observed all phyla in the sample. There is one such number for each of 1000 simulations for each sample. I then report the maximum of these 1000 numbers for each sample, and the 95% quantile. For example, if the 95% quantile for one sample was 5000, then we can be 95% sure that we would have found all observed phyla by reading 5000 OTUs. If this number is substantially below the actual number of sequences read, we can be confident that we have read a sufficient number of sequences.

I will only try another approach, but I haven't done this yet: Suppose we add an extra phylum to each sample, as one we have not observed, and suppose it is (?half as?) as abundant as the least abundant phylum in this sample. How likely are we to have observed this extra phylum within the actual number of OTUs read?

Table 1: Longtable

Sample	n phyla	Sample Abundance	p95	p100
18S.1.1.C.PCM.MM	2	1458	27	66
18S.1.1.D.PCM.ME	10	18812	4005	7563
18S.1.1.F.PCM.ME	3	9020	784	1617
18S.1.1.G.PCM.ME	6	5340	4158	5020
18S.1.1.H.PCM.ME	4	35026	6796	19288
18S.1.10.A.PCM.MM	1	378	1	1
18S.1.10.B.PCM.MM	2	1363	6	14
18S.1.10.C.PCM.MM	2	1689	13	32
18S.1.10.D.PCM.ME	2	166	14	31
18S.1.10.E.PCM.ME	3	1448	429	832
18S.1.10.F.PCM.ME	5	20748	13630	19211
18S.1.10.G.PCM.ME	6	14319	5664	9538
18S.1.11.A.PCM.MM	2	4697	11	20
18S.1.11.B.PCM.MM	1	2	1	1
18S.1.11.C.PCM.MM	6	19591	4123	8768
18S.1.11.D.PCM.ME	7	46082	725	1509
18S.1.11.E.PCM.ME	8	7457	149	366
18S.1.11.G.PCM.ME	8	15771	3655	7239
18S.1.2.B.PCM.MM	2	4369	2044	3546
18S.1.2.C.PCM.MM	6	21167	692	1642

(Continued on Next Page...)

Table 1: Longtable (*continued*)

Sample	n phyla	Sample Abundance	p95	p100
18S.1.2.D.PCM.ME	2	5487	7	19
18S.1.2.E.PCM.ME	4	12105	1173	2623
18S.1.2.F.PCM.ME	2	432	47	112
18S.1.2.G.PCM.ME	7	11566	155	328
18S.1.2.H.PCM.ME	7	3696	906	1583
18S.1.3.B.PCM.MM	2	16	5	9
18S.1.3.C.PCM.MM	1	1423	1	1
18S.1.3.D.PCM.ME	5	6208	801	2009
18S.1.3.E.PCM.ME	5	10298	1152	2407
18S.1.3.H.PCM.ME	3	503	95	182
18S.1.4.B.PCM.MM	6	6076	50	119
18S.1.4.D.PCM.ME	4	2362	71	146
18S.1.4.E.PCM.ME	3	1163	433	831
18S.1.4.F.PCM.ME	5	20475	15550	20372
18S.1.4.G.PCM.ME	1	3487	1	1
18S.1.4.H.PCM.ME	4	5947	4740	5827
18S.1.5.B.PCM.MM	3	20482	4736	8964
18S.1.5.C.PCM.MM	1	2	1	1
18S.1.5.D.PCM.ME	2	568	442	555
18S.1.5.E.PCM.ME	2	9029	593	1529
18S.1.5.F.PCM.ME	8	7140	3966	5612
18S.1.5.H.PCM.ME	4	15322	9946	13981
18S.1.6.B.PCM.MM	4	5890	20	67
18S.1.6.C.PCM.MM	1	78	1	1
18S.1.6.D.PCM.ME	12	40399	11381	19871
18S.1.6.E.PCM.ME	5	18220	7307	12197
18S.1.6.F.PCM.ME	3	869	115	317
18S.1.6.G.PCM.ME	2	1575	51	138
18S.1.6.H.PCM.ME	6	5954	293	557
18S.1.7.A.PCM.MM	1	15620	1	1
18S.1.7.B.PCM.MM	3	86804	424	858
18S.1.7.C.PCM.MM	5	953	98	233
18S.1.7.D.PCM.ME	1	10	1	1
18S.1.7.E.PCM.ME	2	2037	76	140
18S.1.7.F.PCM.ME	3	2433	229	444
18S.1.7.G.PCM.ME	4	4294	114	276
18S.1.7.H.PCM.ME	7	4688	837	1836
18S.1.8.A.PCM.MM	2	16744	7	15
18S.1.8.B.PCM.MM	3	2379	23	61
18S.1.8.D.PCM.ME	5	4203	29	59
18S.1.8.E.PCM.ME	5	7962	99	227
18S.1.8.F.PCM.ME	7	20267	566	1514
18S.1.8.G.PCM.ME	1	53	1	1
18S.1.8.H.PCM.ME	9	10536	1477	2955
18S.1.9.A.PCM.MM	1	820	1	1

(Continued on Next Page...)

Table 1: Longtable (*continued*)

Sample	n phyla	Sample Abundance	p95	p100
18S.1.9.B.PCM.MM	2	2402	6	13
18S.1.9.D.PCM.ME	4	11771	5380	8837
18S.1.9.E.PCM.ME	4	6820	5139	6628
18S.1.9.F.PCM.ME	1	602	1	1
18S.1.9.G.PCM.ME	7	19521	11712	17241
18S.1.9.H.PCM.ME	5	28268	517	1148
18S.2.1.A.PCM.ME	2	4934	333	613
18S.2.1.B.PCM.ME	6	2600	189	375
18S.2.1.C.PCM.LT	8	31078	2343	5840
18S.2.1.D.PCM.LT	8	9690	81	197
18S.2.1.E.PCM.LT	6	9182	211	567
18S.2.1.F.PCM.LT	5	30442	141	306
18S.2.1.G.PCM.LT	5	8660	82	183
18S.2.1.H.PCM.LT	10	14341	2376	4844
18S.2.10.A.PCM.ME	11	16069	1889	3325
18S.2.10.B.PCM.LT	10	12684	6698	10374
18S.2.10.C.PCM.LT	1	2	1	1
18S.2.10.D.PCM.LT	11	179903	10044	24426
18S.2.10.E.PCM.LT	1	2	1	1
18S.2.10.F.PCM.LT	6	13568	1949	5492
18S.2.10.G.PCM.LT	8	10778	5531	9015
18S.2.11.B.PCM.LT	10	13970	2090	3786
18S.2.11.C.PCM.LT	12	14834	933	1928
18S.2.11.D.PCM.LT	8	5289	340	848
18S.2.11.E.PCM.LT	4	20014	81	208
18S.2.11.F.PCM.LT	7	6663	2216	4131
18S.2.2.A.PCM.ME	3	2354	68	177
18S.2.2.B.PCM.ME	3	1691	37	80
18S.2.2.C.PCM.LT	4	4167	45	137
18S.2.2.D.PCM.LT	6	7093	5526	6970
18S.2.2.E.PCM.LT	10	20629	3500	8821
18S.2.2.F.PCM.LT	8	14258	280	590
18S.2.2.G.PCM.LT	8	19146	1488	3559
18S.2.2.H.PCM.ME	1	2454	1	1
18S.2.3.A.PCM.ME	2	110403	1288	2638
18S.2.3.C.PCM.LT	1	2	1	1
18S.2.3.D.PCM.LT	11	5319	4213	4981
18S.2.3.E.PCM.LT	4	4400	47	128
18S.2.3.F.PCM.LT	12	14402	427	877
18S.2.3.G.PCM.LT	3	5265	41	103
18S.2.4.A.PCM.ME	6	26469	540	1070
18S.2.4.B.PCM.ME	4	2813	103	253
18S.2.4.C.PCM.LT	5	5776	79	162
18S.2.4.D.PCM.LT	10	86257	30831	53971
18S.2.4.E.PCM.LT	5	14329	362	844

(Continued on Next Page...)

Table 1: Longtable (*continued*)

Sample	n phyla	Sample Abundance	p95	p100
18S.2.4.F.PCM.LT	11	71747	31562	47239
18S.2.4.G.PCM.LT	4	4290	28	69
18S.2.5.A.PCM.ME	6	6611	178	344
18S.2.5.B.PCM.ME	8	2582	278	664
18S.2.5.C.PCM.LT	8	12004	909	1812
18S.2.5.D.PCM.LT	10	42924	5705	14945
18S.2.5.F.PCM.LT	6	2101	1341	1890
18S.2.5.G.PCM.LT	7	27119	12446	23387
18S.2.6.A.PCM.ME	12	1882	525	1184
18S.2.6.B.PCM.ME	10	20668	708	1285
18S.2.6.C.PCM.LT	7	27162	6455	12804
18S.2.6.D.PCM.LT	9	31444	2002	4264
18S.2.6.E.PCM.LT	6	30281	7782	18801
18S.2.6.F.PCM.LT	5	9004	91	188
18S.2.7.A.PCM.ME	11	52727	42147	50990
18S.2.7.C.PCM.LT	11	23378	14107	21586
18S.2.7.D.PCM.LT	8	88274	17101	32693
18S.2.7.E.PCM.LT	6	10969	261	500
18S.2.7.F.PCM.LT	5	13291	205	410
18S.2.8.A.PCM.ME	7	7936	4032	6662
18S.2.8.B.PCM.LT	7	13802	870	2306
18S.2.8.C.PCM.LT	13	44312	14025	28044
18S.2.8.D.PCM.LT	11	19798	1696	5223
18S.2.8.E.PCM.LT	9	20792	937	2213
18S.2.8.F.PCM.LT	6	2549	69	153
18S.2.8.G.PCM.LT	7	23341	923	1943
18S.2.9.A.PCM.ME	3	41336	993	2248
18S.2.9.C.PCM.LT	7	26464	1427	2957
18S.2.9.D.PCM.LT	6	7431	406	821
18S.2.9.E.PCM.LT	7	12351	991	1731
18S.2.9.F.PCM.LT	7	12783	1782	3994
18S.2.9.G.PCM.LT	5	63942	8070	16670

- ‘n phyla’ is the number of phyla observed in the sample.
- ‘Sample abundance’ is the total abundance for the sample.
- p95 is the total abundance required so that we found all phyla in the sample 95% of the time.
- p100 is the highest abundance required to find all phyla in the sample over all 1000 simulations.

If p95 is well below the Sample Abundance, then this would give us confidence that the total read depth was sufficient.

Appendix.

