Supplementary materials

Accurate quantification of bacterial 16S rRNA gene copy number in metagenomic DNAs accounting variable DNA integrity levels

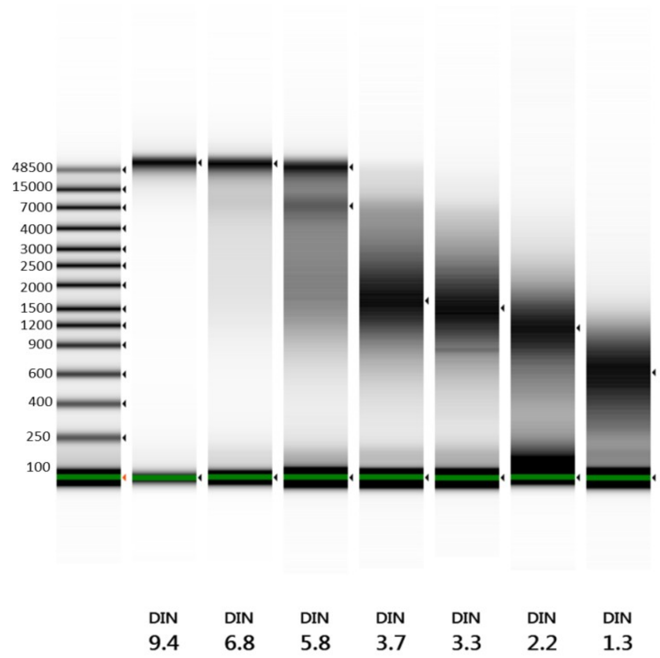


Fig. S1. Qualitative analysis of one of the 5 independent replicates of the ZymoBIOMICS/humDNA mock, at different degradation levels by Agilent 2200 TapeStation. First lane represents the original ZymoBIOMICS/humDNA showing the highest DNA quality (DIN 9.4); moving from left to right lanes, a degradation series of the ZymoBIOMICS/humDNA mock obtained by setting different sonication parameters, is shown.

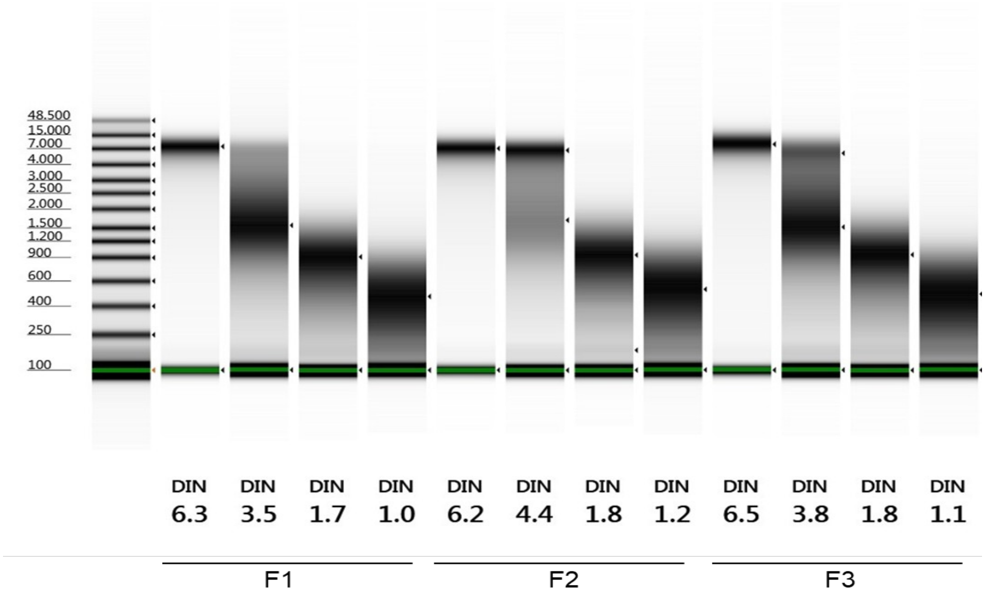


Fig. S2. Qualitative analysis of the three metagenomic DNAs (F1-F2-F3), at different degradation level, by Agilent 2200 TapeStation. For each sample, first lane represents the original sample (DIN 6.5±0.6). Moving from left to right lanes, 3 degradation levels, obtained by setting different sonication parameters, are shown and corresponding to DIN4 (DIN 4.5±1), DIN2 (DIN 2.05±0.35) and DIN1 (DIN 1.35±0.35).

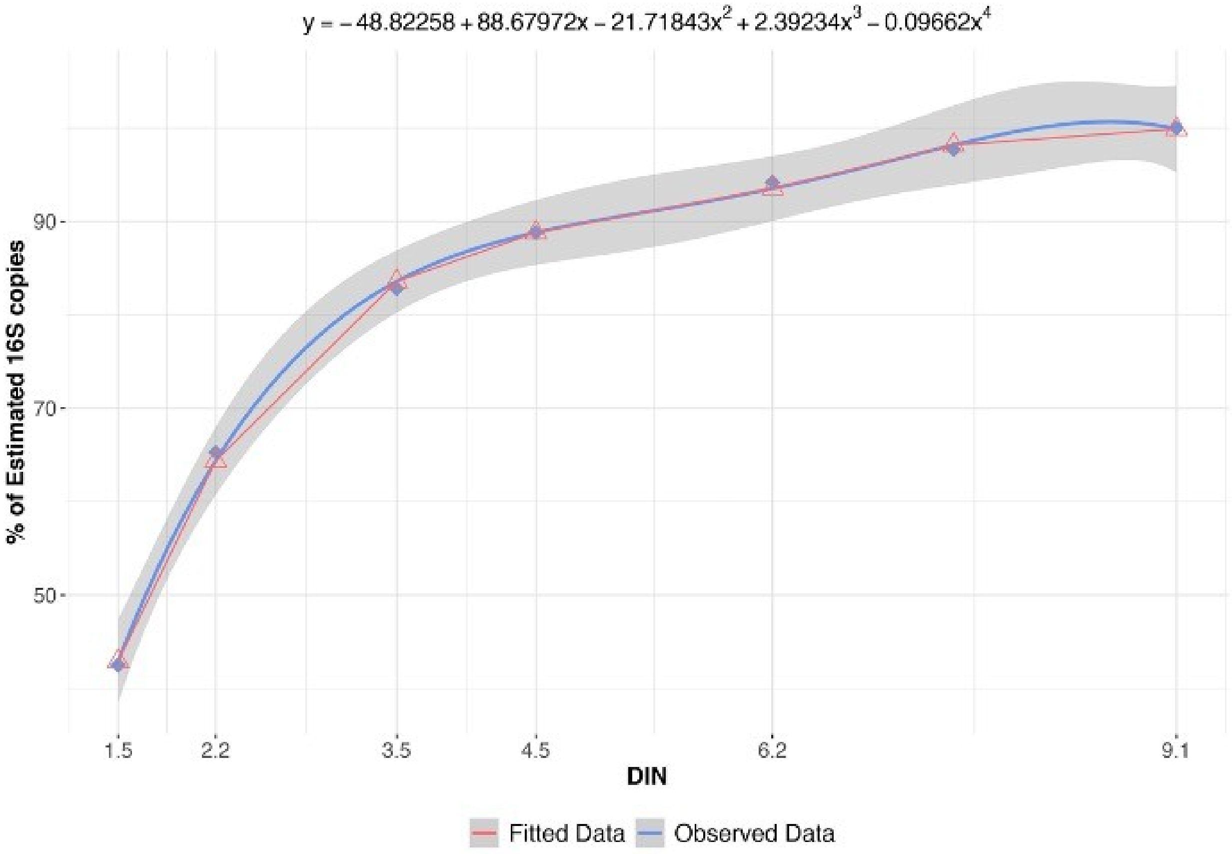
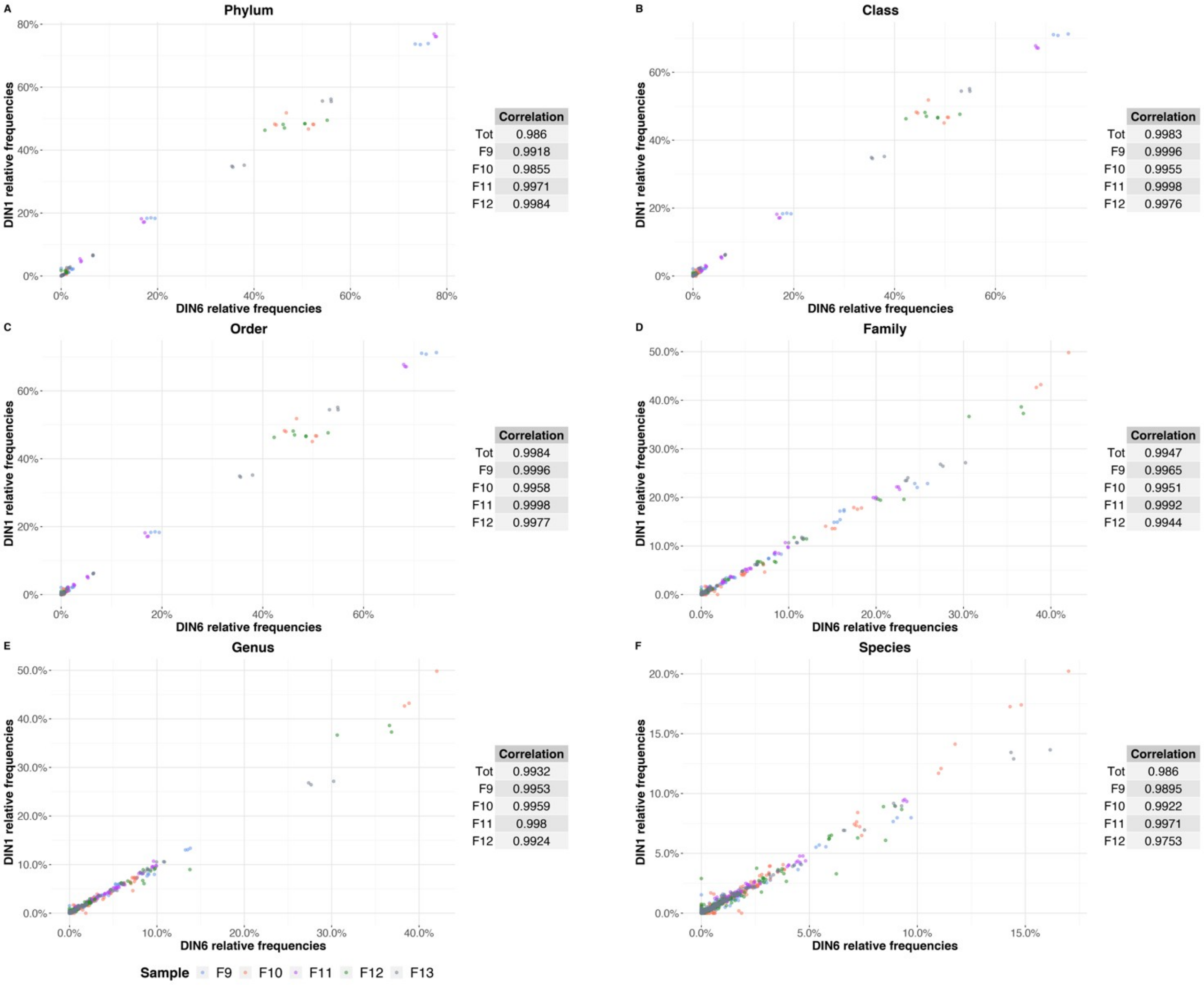


Fig. S3. Graphical representation of the inferred polynomial regression model used to predict the percentage of 16S rRNA copies underestimation as a function of the measured DIN. Scatter plot representing the percentage of 16S rRNA copies underestimation as a function of the measured DIN in ZymoBIOMICS/gDNA DNA mixture. In particular, blue points and red triangles represent the observed and the inferred percentage of 16S rRNA copies underestimation, respectively. The blue and the red line represents the regression interpolation between the observed and estimated values, respectively.



A – Phylum

	Correlation
Tot	0.986
F9	0.9918
F10	0.9855
F11	0.9971
F12	0.9984

B – Class

	Correlation
Tot	0.9983
F9	0.9996
F10	0.9955
F11	0.9998
F12	0.9976

C – Order

	Correlation
Tot	0.9984
F9	0.9996
F10	0.9958
F11	0.9998
F12	0.9977

D – Family

	Correlation
Tot	0.9947
F9	0.9965
F10	0.9951
F11	0.9992
F12	0.9944

E – Genus

	Correlation
Tot	0.9932
F9	0.9953
F10	0.9959
F11	0.998
F12	0.9924

F – Species

	Correlation
Tot	0.986
F9	0.9895
F10	0.9922
F11	0.9971
F12	0.9753

Fig. S4. Representation of the observed relative abundances in DIN6 and DIN1 samples and Pearson correlation. Scatter plot representing the observed the taxa relative abundances in DIN6 and DIN1 samples and summarized at 6 taxonomic levels (i.e. phylum, class, order, family, genus and species). Points are colored according to the samples. Beside the plot the observed Pearson correlation between DIN6 and DIN1 relative frequencies is reported for all data and per sample.

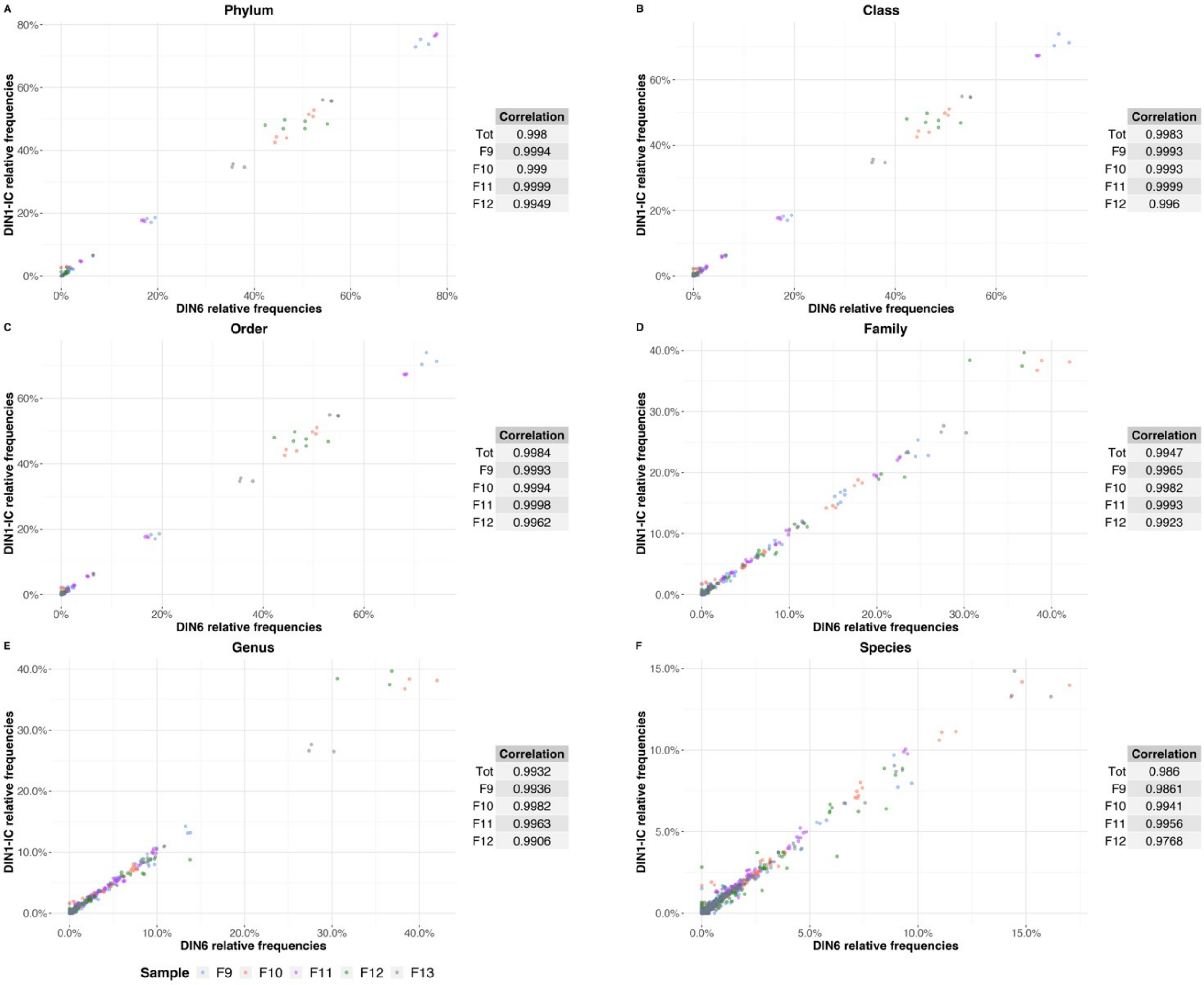
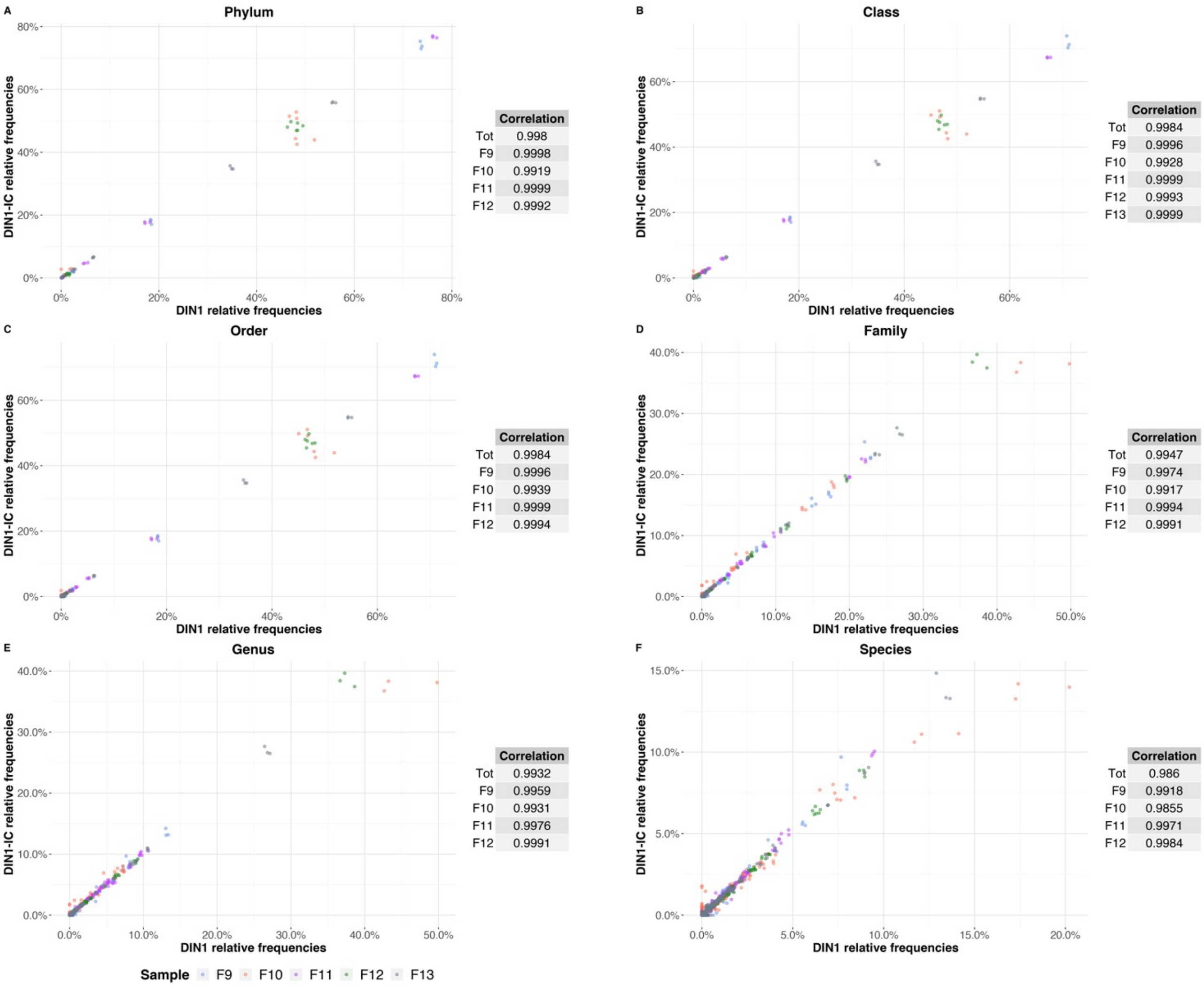


Fig. S5. Representation of the observed relative abundances in DIN6 and DIN1-IC samples and Pearson correlation. Scatter plot representing the observed the taxa relative abundances in DIN6 and DIN1-IC samples and summarized at 6 taxonomic levels (i.e. phylum, class, order, family, genus and species). Points are colored according to the samples. Beside the plot the observed Pearson correlation between DIN6 and DIN1-IC relative frequencies is reported for all data and per sample.



A

	Correlation
Tot	0.998
F9	0.9998
F10	0.9919
F11	0.9999
F12	0.9992

B

	Correlation
Tot	0.9984
F9	0.9996
F10	0.9928
F11	0.9999
F12	0.9993
F13	0.9999

C

	Correlation
Tot	0.9984
F9	0.9996
F10	0.9939
F11	0.9999
F12	0.9994

D

	Correlation
Tot	0.9947
F9	0.9974
F10	0.9917
F11	0.9994
F12	0.9991

E

	Correlation
Tot	0.9932
F9	0.9959
F10	0.9931
F11	0.9976
F12	0.9991

F

	Correlation
Tot	0.986
F9	0.9918
F10	0.9855
F11	0.9971
F12	0.9984

Fig. S6. Representation of the observed relative abundances in DIN1 and DIN1-IC samples and Pearson correlation. Scatter plot representing the observed the taxa relative abundances in DIN1 and DIN1-IC samples and summarized at 6 taxonomic levels (i.e. phylum, class, order, family, genus and species). Points are colored according to the samples. Beside the plot the observed Pearson correlation between DIN1 and DIN1-IC relative frequencies is reported for all data and per sample.

DIN	Total 16S copies/ng	SD	Underestimation *vs* DIN9 (%)	SD
9.1±0.1	1,387,126	± 56974	/	/
7.5±0.3	1,355,263	± 44705	3.1	± 1.8
6.2±0.4	1,306,880	± 60842	5.5	± 2.2
4.5±0.3	1,231,914	± 59374	11.1	± 2.8
3.5±0.3	1,148,583	± 64630	17.2	± 4.7
2.2±0.3	906,182	± 40182	34.7	± 2.9
1.5±0.3	589,273	± 61146	57.5	± 4.4

Table S1. 16S rRNA copy number quantification in ZymoBIOMICS/humDNA mock with different DINs. For each degradation level, the 16S rRNA copies/ng DNA quantified by ddPCR, the underestimation calculated as percentage (%) of 16S un-detected copies compared to highest quality mock (DIN 9.1±0.1) are reported. Values are reported as mean of duplicates of 5 independent replicates, for each DIN value ± SD.

DIN	Sample	Total 16S copies/ng DNA	SD	Total 16S copies/ng Mean	SD Mean	Underestimation vs DIN 6 (%)	SD	Underestimation (%) Mean	SD Mean
6.5±0.6	F1	1223124	57803	1144734	203336	0	0	0	0
	F2	1361889	245888			0	0		
	F3	1282938	105955			0	0		
	F4	1331529	85439			0	0		
	F5	1321086	141547			0	0		
	F6	1151349	201438			0	0		
	F7	1086539	155931			0	0		
	F8	1329989	69962			0	0		
	F9	701284	55346			0	0		
	F10	900000	65309			0	0		
	F11	1165279	123914			0	0		
	F12	1132581	153796			0	0		
	F13	893960	42420			0	0		
4.5±1	F1	1091539	103597	996223	182890	5.8	3.2	11.1	7.1
	F2	1107400	78279			3.9	2.0		
	F3	1049361	82566			11.2	5.7		
	F4	1131488	39292			11.4	3.0		
	F5	1258400	106302			8.5	6.8		
	F6	984413	90103			14.5	7.8		
	F7	1086539	155931			2.4	1.8		
	F8	1179676	69962			20.6	6.8		
	F9	677229	85770			3.4	12.2		
	F10	647391	226664			28.1	25.2		
	F11	957450	52688			17.8	4.5		
	F12	872321	71610			23.0	6.3		
	F13	907696	13624			-2.0	1.5		
2.05±0.35	F1	876115	79581	733039	105534	22.7	6.5	35.2	7.6
	F2	763376	102681			35.5	8.7		
	F3	715735	50826			44.2	4.0		
	F4	843066	82614			36.7	6.2		
	F5	886205	32302			32.9	2.4		
	F6	716005	61184			48.8	4.4		
	F7	662184	42253			39.1	3.9		
	F8	703366	108543			45.1	9.2		
	F9	546509	21479			22.1	3.1		
	F10	572561	45815			36.4	5.1		
	F11	799559	90359			31.4	7.8		
	F12	672387	48567			41.0	4.3		
	F13	772435	111698			11.0	12.5		
1.35±0.35	F1	504369	56572	497605	82871	60.5	5.9	56.9	6.8
	F2	489871	35727			58.6	3.0		
	F3	504795	62743			60.7	4.9		
	F4	595220	55114			55.3	4.1		
	F5	599077	46067			54.7	3.5		
	F6	589600	50994			70.3	6.2		
	F7	365542	11535			66.4	1.1		
	F8	379608	45943			66.2	3.9		
	F9	438346	50134			37.5	7.1		
	F10	385363	25225			57.2	2.8		
	F11	547997	71955			53.0	6.2		

	F12	559273	75166		51.0	6.6	
	F13	509811	90310		43.0	10	

Table S2. 16S rRNA copy number quantification in 13 metagenomic DNA samples with different degradation level. For each DIN value, the 16S copies/ng DNA in 13 metagenomic DNAs (F1-F13) quantified by ddPCR. the underestimation calculated as percentage (%) of 16S undetected copies compared to highest quality DNA (DIN 6.5±0.6) are reported. Values are reported as mean of at least three replicate experiments ± SD.

DIN	Sample	Total 16S copies	SD	*A. muciniphila* 16S copies	SD	Ratio 16S *A. muciniphila*/total 16S (%)
6.3±0.4	F9	619827	131915	10537	2273.2	1.7
	F10	795695	9102	7641	1067.3	1.0
	F11	1082260	210534	6462	249.5	0.6
	F12	1094413	130301	2240	127.3	0.2
	F13	908216	33515	1087	129.7	0.1
1.2±0.2	F9	269587	65960	4543	363.5	1.7
	F10	386099	29420	3084	440.6	0.8
	F11	348731	88900	2296	119.1	0.7
	F12	520198	60541	1033	109.3	0.2
	F13	487064	35054	508	28.6	0.1
1.2±0.2 input corrected	F9	618696	86434	11980	890.2	1.9
	F10	788188	56716	7446	283.2	0.9
	F11	915742	52889	7113	775.9	0.8
	F12	1192981	71564	2568	68.1	0.2
	F13	896312	13282	1080	175.5	0.1

Table S3. Total and *A. muciniphila* 16S copy number quantification in 5 metagenomic DNA samples at DIN6. DIN1 and input corrected DIN1 values. For each DIN value, the total and the *A. muciniphila* 16S copies quantified by ddPCR are reported as mean of at least duplicate experiments ± SD. for each sample. The *A. muciniphila* 16S copies relative abundance was calculated as the percentage ratio (%) between *A. muciniphila* 16S copies and total 16S copies in each sample.

rank	Sbj	DIN6	DIN1	DIN1-IC	Common DIN6 - DIN1	Only in DIN6	Only in DIN1	Common DIN6 - DIN1-IC	Only in DIN6	Only in DIN1-IC	Common DIN1 - DIN1-IC	Only in DIN1	Only in DIN1-IC	Common to all	% Common
phylum	**F9**	6	6	6	6	0	0	6	0	0	6	0	0	6	100,0%
	F13	4	4	4	4	0	0	4	0	0	4	0	0	4	100,0%
	F12	4	4	4	4	0	0	4	0	0	4	0	0	4	100,0%
	F11	4	4	4	4	0	0	4	0	0	4	0	0	4	100,0%
	F10	5	5	5	5	0	0	5	0	0	5	0	0	5	100,0%
class	**F9**	7	7	7	7	0	0	7	0	0	7	0	0	7	100,0%
	F13	5	5	5	5	0	0	5	0	0	5	0	0	5	100,0%
	F12	3	4	3	2	1	2	2	1	1	3	1	0	2	40,0%
	F11	7	7	7	7	0	0	7	0	0	7	0	0	7	100,0%
	F10	6	5	7	4	2	1	5	1	2	5	0	2	4	50,0%
order	**F9**	7	7	7	7	0	0	7	0	0	7	0	0	7	100,0%
	F13	4	5	4	4	0	1	4	0	0	4	1	0	4	80,0%
	F12	3	4	3	2	1	2	2	1	1	3	1	0	2	40,0%
	F11	8	8	8	8	0	0	8	0	0	8	0	0	8	100,0%
	F10	6	5	7	4	2	1	5	1	2	5	0	2	4	50,0%
family	**F9**	14	16	15	14	0	2	14	0	1	15	1	0	14	87,5%
	F13	10	11	10	10	0	1	10	0	0	10	1	0	10	90,9%

	F12	9	8	8	7	2	1	7	2	1	8	0	0	7	70,0%
	F11	15	15	16	15	0	0	15	0	1	15	0	1	15	93,7%
	F10	13	12	14	11	2	1	12	1	2	12	0	2	11	73,3%
genus	**F9**	26	28	27	26	0	2	26	0	1	27	1	0	26	92,9%
	F13	20	20	19	20	0	0	19	1	0	19	1	0	19	95,0%
	F12	16	17	17	14	2	3	14	2	3	17	0	0	14	73,7%
	F11	24	24	24	24	0	0	24	0	0	24	0	0	24	100,0%
	F10	20	19	21	18	2	1	19	1	2	19	0	2	18	81,8%
species	**F9**	36	37	36	36	0	1	36	0	0	36	1	0	36	97,3%
	F13	31	32	30	31	0	1	30	1	0	30	2	0	30	93,7%
	F12	31	29	29	26	5	3	27	4	2	28	1	1	26	76,5%
	F11	38	36	36	36	2	0	36	2	0	36	0	0	36	94,7%
	F10	35	31	36	30	5	1	33	2	3	30	1	6	29	76,3%

Table S4. Summary of common and uncommon observed taxa in metabarcoding analysis between the tested DNA qualities. For each taxonomic rank (i.e. phylum, class, order, family, genus) and for each subject (i.e. Sbj) are listed the number of observed taxa per DNA qualities (i.e. DIN6, DIN1 and DIN1-IC). Following, the observed taxa were pairwise compared among the DNA qualities and the number of common (e.g. Common to DIN6 DIN1) and uncommon taxa (e.g. Only in DIN6, Only in DIN1) are shown. Finally, The number of taxa observed in all samples (i.e. Common to all) and the relative abundance referred to the total number of observed taxa (i.e. % Common) are displayed.

rank	Sbj	taxon	DIN6	DIN1	DIN1-IC
class	**F12**	*Actinobacteria*	0,43%	0,89%	0,91%
	F12	*Coriobacteriia*	0,24%	0,67%	0,42%
	F10	*Actinobacteria*	0,41%	0,14%	0,70%
	F10	*Coriobacteriia*	0,37%	1,16%	2,13%
order	**F13**	*Eggerthellales*	0,20%	0,61%	0,41%
	F12	*Coriobacteriales*	0,18%	0,51%	0,29%
	F12	*Bifidobacteriales*	0,40%	0,84%	0,86%
	F10	*Coriobacteriales*	0,34%	1,10%	1,91%
	F10	*Bifidobacteriales*	0,27%	0,14%	0,51%
family	**F9**	*Odoribacteraceae*	0,46%	0,51%	0,54%
	F9	*Oscillospiraceae*	0,32%	0,64%	0,48%
	F13	*Eggerthellaceae*	0,20%	0,61%	0,41%
	F12	*Bifidobacteriaceae*	0,40%	0,84%	0,86%
	F11	*Atopobiaceae*	0,45%	0,48%	0,52%
	F10	*Bifidobacteriaceae*	0,27%	0,14%	0,51%
	F10	*Coriobacteriaceae*	0,34%	1,10%	1,91%
genus	**F9**	*Oscillibacter*	0,32%	0,64%	0,48%
	F9	*Anaerofilum*	0,49%	0,60%	0,56%
	F12	*Paraprevotella*	0,36%	0,56%	0,54%
	F12	*Bifidobacterium*	0,40%	0,84%	0,86%
	F12	*Eisenbergiella*	0,41%	0,65%	0,59%
	F10	*Collinsella*	0,34%	1,10%	1,78%
	F10	*Bifidobacterium*	0,27%	0,14%	0,51%
species	**F9**	*Alistipes sp. NML05A004*	0,40%	0,63%	0,40%
	F13	*Bacteroides thetaiotaomicron*	0,39%	0,50%	0,41%
	F12	*Parabacteroides chinchillae*	0,35%	0,52%	0,47%
	F12	*Bifidobacterium adolescentis*	0,35%	0,70%	0,64%
	F12	*Paraprevotella clara*	0,36%	0,56%	0,54%
	F10	*Collinsella aerofaciens*	0,34%	1,10%	1,78%
	F10	*Bifidobacterium pseudocatenulatum*	0,27%	0,14%	0,51%
	F10	*Alistipes sp. RMA 9912*	0,38%	0,44%	0,64%

Table S5. Inspection of uncommon taxa relative abundances. The relative abundances of uncommon taxa, differing among the three tested DNA qualities, are shown. The relative abundances are listed per each sample and are expressed as the average value of the three replicates.

Rank	Sbj	Correlation DIN6 - DIN1	p-value	Correlation DIN6 - DIN1-IC	p-value	Correlation DIN1 - DIN1-IC	p-value
phylum	**F12**	0,99692	0	0,99491	0	0,99921	0
	F9	0,99956	0	0,99941	0	0,99978	0
	F10	0,99473	0	0,999	0	0,99194	0
	F11	0,99978	0	0,99988	0	0,99992	0
	F13	0,99924	0	0,99906	0	0,99988	0
class	**F12**	0,99757	0	0,99603	0	0,99939	0
	F9	0,99961	0	0,99929	0	0,99959	0
	F10	0,9955	0	0,99928	0	0,99346	0
	F11	0,99979	0	0,99986	0	0,99992	0
	F13	0,99942	0	0,99927	0	0,9999	0
order	**F12**	0,99772	0	0,99625	0	0,99943	0
	F9	0,99962	0	0,99933	0	0,99962	0
	F10	0,99576	0	0,99939	0	0,99388	0
	F11	0,9998	0	0,99985	0	0,99992	0
	F13	0,9995	0	0,99933	0	0,9999	0
family	**F12**	0,99444	0	0,99234	0	0,9991	0
	F9	0,99653	0	0,99651	0	0,99743	0
	F10	0,99507	0	0,99825	0	0,99169	0
	F11	0,9992	0	0,99932	0	0,99938	0
	F13	0,99879	0	0,99837	0	0,99956	0
genus	**F12**	0,9924	0	0,99062	0	0,99907	0
	F9	0,99528	0	0,99363	0	0,99594	0
	F10	0,99593	0	0,99825	0	0,99314	0
	F11	0,99804	0	0,99634	0	0,99765	0
	F13	0,99849	0	0,99768	0	0,99942	0
species	**F12**	0,97528	0	0,97676	0	0,99843	0
	F9	0,98948	0	0,98614	0	0,99181	0
	F10	0,99217	0	0,99411	0	0,98551	0
	F11	0,99714	0	0,99557	0	0,99713	0
	F13	0,99602	0	0,99506	0	0,99751	0

Table S6. Pearson correlation among the observed relative abundances. The observed relative taxon frequencies were pairwise compared among the three DNA qualities (i.e. DIN6, DIN1 and DIN1-IC) by inferring the Pearson correlation. The resulting data are listed according to taxonomic rank (i.e. phylum, class, order, family, genus) and for each subject (i.e. Sbj). For each measured correlation values is also listed the relevant p-value.