Strain	Habitat of Origin	Isolation Strategy	GenBank or EMBL Accession						
			SSU rRNA and/or ITS	glnA	MLSA	catA	nasA	Pept.	Genome
QLW-P1DMWA-1	Pond-1	sDAM	AJ879783.1	FN823098.2	CP000655	ND	LT622191	LT622210	CP000655
QLW-P1DNSYA-2	Pond-1	sDAM	AJ879800.1	HE646491.1	HE646484.1 - HE646494.1	ND	NS	LT622211	NS
QLW-P1FAT50D-2	Pond-1	sFAM	AJ879799.1	HE646502.1	HE646495.1 - HE646505.1	ND	NS	LT622212	NS
P1-Kol5	Pond-1	DP	HE646564.1	HE646480.1	HE646473.1 - HE646483.1	ND	NS	LT622213	NS
P1-Kol6	Pond-1	DP	HE646566.1	HE646524.1	HE646517.1 - HE646527.1	ND	ND	LT622214	NS
P1-Kol7	Pond-1	DP	HE646567.1	HE646535.1	HE646528.1 - HE646538.1	ND	LT622192	LT622215	NS
P1-Kol8	Pond-1	DP	HE646568.1	HE646546.1	HE646539.1 - HE646549.1	ND	ND	LT622216	LVLD00000000
P1-05-13	Pond-1	sFAM	HE646561.1	HE646447.1	HE646440.1 - HE646450.1	ND	LT622193	LT622217	NS
P1-05-86	Pond-1	sFAM	HE646569.1	HE646557.1	HE646550.1 - HE646560.1	ND	ND	LT622218	NS
P1-4-10KL	Pond-1	tDAM	HE646565.1	HE646513.1	HE646506.1 - HE646516.1	ND	LT622194	LT622219	LVJO00000000
P1-4-10GR	Pond-1	tDAM	HE646562.1	HE646458.1	HE646451.1 - HE646461.1	ND	NS	LT622220	NS
P1-3-8A	Pond-1	tDAM	HE646563.1	HE646469.1	HE646462.1 - HE646472.1	ND	NS	LT622221	NS
QLW-P2DMWB-1	Pond-2	sDAM	FN429654.1	FN823099.2	HE647093.1 - HE647101.1	ND	LT622195	LT622222	NS
MWH-Tro5-3-9	Pond Trog-5	tDAM	FN429711.1	FN823185.2	HE646931.1 - HE646939.1	LT622179	LT622196	LT622223	NS
MWH-Tro5-4-4	Pond Trog-5	tDAM	HE647119.1	HE647104.1	HE646940.1 - HE646948.1	NS	NS	LT622224	NS
MWH-Tro6-1-1	Pond Trog-6	tDAM	HE647120.1	HE647105.1	HE646949.1 - HE646957.1	LT622180	LT622197	ND	NS
MWH-Tro6-4-2	Pond Trog-6	tDAM	FN429712.1	FN823186.2	HE646967.1 - HE646975.1	NS	NS	ND	NS
MWH-Tro6-1-6	Pond Trog-6	tDAM	HE647121.1	HE647106.1	HE646958.1 - HE646966.1	LT622181	LT622198	LT622225	NS
MWH-Tro7-1-10	Pond Trog-7	tDAM	HE647123.1	HE647108.1	HE646985.1 - HE646993.1	NS	NS	LT622226	NS
MWH-Tro7-2-5	Pond Trog-7	tDAM	HE647122.1	HE647107.1	HE646994.1 - HE647002.1	NS	NS	LT622227	NS
MWH-Tro7-2-8KL	Pond Trog-7	tDAM	HE647125.1	HE647110.1	HE647012.1 - HE647020.1	NS	NS	LT622228	NS
MWH-Tro7-2-8GR	Pond Trog-7	tDAM	HE647124.1	HE647109.1	HE647003.1 - HE647011.1	LT622182	LT622199	LT622229	NS
MWH-Tro7-1-4	Pond Trog-7	tDAM	FN429713.1	FN823187.2	HE646976.1 - HE646984.1	LT622183	LT622200	LT622230	LVLF00000000
MWH-Tro8-2-1	Pond Trog-8	tDAM	HE647126.1	HE647111.1	HE647021.1 - HE647029.1	NS	LT622201	LT622231	NS
MWH-Tro8-2-5	Pond Trog-8	tDAM	HE647127.1	HE647112.1	HE647030.1 - HE647038.1	NS	LT622202	LT622232	NS
MWH-Tro8-3-6	Pond Trog-8	tDAM	HE647128.1	HE647113.1	HE647048.1 - HE647056.1	NS	LT622203	LT622233	NS
MWH-Tro8-4-1	Pond Trog-8	tDAM	HE647129.1	HE647114.1	HE647057.1 - HE647065.1	LT622184	LT622204	LT622234	NS
MWH-Tro8-4-4	Pond Trog-8	tDAM	HE647130.1	HE647115.1	HE647066.1 - HE647074.1	NS	LT622205	LT622235	NS
MWH-Tro8-2-9	Pond Trog-8	tDAM	FN429710.1	FN823183.2	HE647039.1 - HE647047.1	LT622185	LT622206	LT622236	LVLG00000000
Tro8F10W22	Pond Trog-8	sFAM	CP015018	LN908882.1	CP015018	LT622186	LT622207	ND	CP015018
MWH-RechtKolB	Pond Rechteckteich	sFAM	FN429668.1	FN823145.2	HE646913.1 - HE646921.1	LT622187	ND	ND	CP015016
MWH-RechtKolC	Pond Rechteckteich	sFAM	HE647118.1	HE647103.1	HE646922.1 - HE646930.1	NS	ND	ND	NS
MWH-RechtKol4	Pond Rechteckteich	sFAM	HE647117.1	HE647102.1	HE646904.1 - HE646912.1	LT622188	ND	ND	CP015017
MWH-Recht1	Pond Rechteckteich	tDAM	FN429706.1	FN823146.2	HE646895.1 - HE646903.1	NS	ND	ND	LVLE00000000
MWH-GrInsel-1	Pond Großer Inselteich	tDAM	FN556009.1	FN823190.2	HE646886.1 - HE646894.1	LT622189	ND	ND	NS
MWH-Unterer-4A-1	Pond Unterer Teich	tDAM	FN429705.1	FN823182.2	HE647075.1 - HE647083.1	LT622190	LT622208	LT622237	NS
MWH-Unterer-4A-2	Pond Unterer Teich	tDAM	HE647131.1	HE647116.1	HE647084.1 - HE647092.1	NS	LT622209	ND	NS
QLW-P1FAT50C-4	Pond-1	sFAM	AJ879778.1	FN823104.2	CP015922	CP015922	CP015922	ND	CP015922
QLW-P1DATA-2	Pond-1	sDAM	AJ879801.1	FN823102.2	LZMQ00000000	ND	ND	ND	LZMQ00000000
MWH-Tro8-2-5gr	Pond-1	sFAM	LZMR00000000	LZMR00000000	LZMR00000000	ND	LZMR00000000	ND	LZMR00000000
MWH-Adler-W8	Adlerlacke	sFAM	FN429735.1	FN823208.2	LZFI00000000	LZFI00000000	LZFI00000000	ND	LZFI00000000

sDAM, standard Dilution Acclimatization Method; sFAM, standard Filtration Acclimatization Method; tDAM, targeted isolation by DAM; DP, direct plating of dilutions on NSY agar plates

ND, not detected by PCR; NS, detected by PCR but not sequenced

Table S1. List of the 37 cultivated *P. asymbioticus* and the four *Polynucleobacter* sp. strains including accession numbers.

Habitat	**Site**	**Geographic Coordinates**	**Altitude (m.a.s.l.)**	**Max. Depth (m)**	**pH**	**Conductivity (µS cm-1)**	**PnecC (% of bact.)**[*]	***P. asymbioticus* (% of bact.)**
Pond-1	Loibersbacher Höhe	N 47.740° E 13.302°	1286	1.2	5.4	11.3	32	12[†]
Pond-2	Loibersbacher Höhe	N 47.739° E 13.302°	1297	0.6	5.5	10.0	38	9[†]
Pond Unterer Teich	Rauriser Urwald	N 47.078° E 12.989°	1676	ND	5.8	7.7	35	ND
Pond Großer Inselteich	Rauriser Urwald	N 47.075° E 12.991°	1704	ND	5.0	10.1	27	ND
Pond Rechteckteich	Rauriser Urwald	N 47.076° E 12.995°	1731	1.2	5.2	6.7	59	ND
Pond Trog-8	Trög	N 47.249° E 13.266°	1958	2.0	4.9	7.2	39	ND
Pond Trog-7	Trög	N 47.250° E 13.267°	1962	1.7	5.7	3.7	33	ND
Pond Trog-5	Trög	N 47.249° E 13.270°	1965	ND	4.8	7.6	48	ND
Pond Trog-6	Trög	N 47.250° E 13.272°	2001	ND	5.1	5.1	51	ND
Adlerlacke	-	N 47.373° E 12.093°	1770	ND	4.8	9.2	43	ND

ND = not determined

values for pH, conductivity, PnecC and *P. asymbioticus* are the means of multiple measurements

[*] determined by FISH probe PnecC-16S-445 (Hahn MW, Pockl M, Wu QL. 2005. Appl Environ Microbiol 71:4539-4547.)

[†] Hahn MW, Scheuerl T, Jezberova J, Koll U, Jezbera J, Simek K, Vannini C, Petroni G, Wu QL. 2012. PLoS One 7:e32772.

Table S2. Characterization of the nine habitats of origin of the *P. asymbioticus* strains and the habitat of origin of one *Polynucleobacter* sp. strain.

Strain	Sequencing	Yield (Mbp)	read count	%Q30	Mean Q
P1-4-10KL	MiSeq 2x150bp	771	5,142,282	97.96	37.76
P1-Kol8	MiSeq 2x150bp	181	1,204,432	95.87	37.09
MWH-Tro-7-1-4	MiSeq 2x150bp	316	2,109,486	97.66	37.67
MWH-Tro-8-2-9	MiSeq 2x150bp	434	2,893,930	96.93	37.45
Tro8F10W22	MiSeq 2x300bp	222	738,492	80.41	32.03
MWH-Recht1	MiSeq 2x150bp	367	2,449,702	98.01	37.78
MWH-RechtKolB	MiSeq 2x150bp	482	3,210,870	97.49	37.62
MWH-RechtKol4	MiSeq 2x150bp	209	1,390,006	96.53	37.27

- "Yield (Mbp)": number of bases called in mega bases.
- All reads are passed filter, i.e. reads have passed the default Illumina filter procedure (chastity filter).
- "%Q30": represents the percentage of bases with a quality score of at least 30 (inferred base call accuracy of 99.9%). The Q-score is a prediction of the probability of a wrong base call.

Table S3. Sequencing statistics for eight *P. asymbioticus* genomes.

Strain	hash length	number of nodes	N50 (bp)	max. node length (bp)	total length (bp)	total reads	used reads	avgerage coverage
P1-4-10KL	135	51	699,341	818,821	2,160,771	5,142,282	4,749,862	330
P1-Kol8	97	67	229,702	461,334	2,159,551	1,099,684	1,067,096	74
MWH-Tro-7-1-4	127	46	267,647	639,815	2,221,922	2,109,486	1,962,288	132
MWH-Tro-8-2-9	131	57	287,834	461,381	2,287,817	2,893,930	2,631,366	173
Tro8F10W22	97	45	282,233	531,708	2,197,670	718,464	702,193	96
MWH-Recht1	133	45	363,948	500,247	2,128,350	2,449,702	2,274,965	160
MWH-RechtKolB	137	34	295,536	555,775	2,352,775	3,210,870	2,924,163	186
MWH-RechtKol4	97	38	286,500	574,860	2,379,812	1,291,174	1,255,593	79

The table refers to the Velvet assembly output. Some gaps between contigs have been closed subsequently.

Table S4. Statistics after the assembly with Velvet.

GI variant	genome	GI position	length (kbp)	CDS count	IMG/ER scaffold ID	start coordinate	end coordinate
CSC	QLW-P1DMWA-1	2	62	65	2639765941	281483	343853
	P1-4-10KL	2	56	65	2639765857	287436	343853
	P1-Kol8[†]	2	75	90	2639765884 - 2639765887	158746	18019
	MWH-Tro7-1-4	2	90	99	2636567154	311481	401453
	MWH-Tro8-2-9	2	92	100	2639765932	282668	374712
	Tro8F10W22	2	79	89	2636562151	280050	359460
	MWH-Recht1[†]	2	59	62	2639765903 - 2639765907	289927	18897
	MWH-RechtKolB	2	282	109	2636567168	323497	605182
	MWH-RechtKol4	2	281	109	2636493768	323497	604711
ACD	MWH-Tro7-1-4[†]	5	77	86	2636567154 - 2636567156	632362	15000
	MWH-Tro8-2-9	5	76	81	2639765932	609604	685229
	Tro8F10W22	5	77	84	2636562151	592616	669579
	MWH-RechtKolB	21	48	52	2636567168	2007660	2055751
	MWH-RechtKol4	21	48	52	2636493768	2034104	2082195
NIT	QLW-P1DMWA-1	12	17	16	2639765941	988990	1005961
	P1-4-10KL	12	17	16	2639765857	988990	1005961
	MWH-Tro7-1-4	12	17	16	2636567156	1120719	1137690
	MWH-Tro8-2-9	9	19	19	2639765932	963412	982591
	Tro8F10W22	12	17	16	2636562151	1079595	1096566
LIP	MWH-RechtKol4	10	28	26	2636493768	1015955	1044433
COD	MWH-Tro7-1-4	19	15	13	2636567164	1787751	1802581
	Tro8F10W22	19	15	13	2636562151	1793404	1808234
	MWH-RechtKolB	12	22	16	2636567168	1139510	1161737
	MWH-RechtKol4	12	22	16	2636493768	1158343	1180570
GGR	QLW-P1DMWA-1	14	42	8	2639765941	1102705	1144985
	P1-4-10KL[†]	14	44	33	2639765857 - 2639765882	1102705	20039
	P1-Kol8[†]	14	42	18	2639765889 - 2639765899	749742	23699
	MWH-Tro7-1-4[†]	14	27	16	2636567156 - 2636567164	542792	12404
	Tro8F10W22	14	94	54	2636562151	1182944	1277293
	MWH-Recht1[†]	14	30	17	2639765907 - 2639765916	806835	12562
SUL	MWH-RechtKolB	17	10	11	2636567168	1699391	1709669
	MWH-RechtKol4	17	10	11	2636493768	1725835	1736113
HMR	QLW-P1DMWA-1	20	26	29	2639765941	1737019	1762904
	P1-4-10KL	20	26	29	2639765882	1738918	1764803
	P1-Kol8	20	26	29	2639765902	1738138	1764014
	MWH-Tro8-2-9	20	48	39	2639765938	1841582	1889998
	MWH-Recht1	20	26	29	2639765927	1731909	1757794

CSC, cell surface composition; ACD, aromatic compound degradation; NIT, nitrate assimilation; LIP, lipid metabolism; COD, carbon-monoxide dehydrogenation; GGR, giant gene region; SUL, sulfate transport; HMR, heavy metal resistance

[†] scaffolds could not be assembled in the region of the respective GI

Table S5. Characterization of eight selected GI variants.

Strain	Pept.[*]	GT1	GT2	NAT	NDP	UDP	assigned CSC type
QLW-P1DMWA-1	1	🟥	🟥	🟥	🟥	🟥	A1
QLW-P1DNSYA-2	1	🟥	🟥	🟥	🟥	🟥	A1
QLW-P1FAT50D-2	1	🟥	🟥	🟥	🟥	🟥	A1
P1-Kol5	1	🟥	🟥	🟥	🟥	🟥	A1
P1-Kol6	5	🟥	🟥	🟥	🟥	🟩	B5
P1-Kol7	1	🟥	🟥	🟥	🟥	🟥	A1
P1-Kol8	5	🟥	🟥	🟥	🟥	🟩	B5
P1-05-13	1	🟥	🟥	🟥	🟥	🟥	A1
P1-05-86	5	🟥	🟥	🟥	🟥	🟩	B5
P1-4-10KL	1	🟥	🟥	🟥	🟥	🟥	A1
P1-4-10GR	1	🟥	🟥	🟥	🟥	🟥	A1
P1-3-8A	1	🟥	🟥	🟥	🟥	🟥	A1
QLW-P2DMWB-1	1	🟥	🟥	🟥	🟥	🟥	A1
MWH-Tro5-3-9	6	🟥	🟥	🟩	🟥	🟩	C6
MWH-Tro5-4-4	6	🟥	🟥	🟩	🟥	🟩	C6
MWH-Tro6-1-1	🟥	🟥	🟥	🟥	🟩	🟥	D-
MWH-Tro6-4-2	🟥	🟥	🟥	🟥	🟩	🟥	D-
MWH-Tro6-1-6	1	🟥	🟥	🟥	🟩	🟥	E1
MWH-Tro7-1-10	3	🟥	🟥	🟩	🟩	🟥	F3
MWH-Tro7-2-5	3	🟥	🟥	🟩	🟩	🟥	F3
MWH-Tro7-2-8KL	3	🟥	🟥	🟩	🟩	🟥	F3
MWH-Tro7-2-8GR	3	🟥	🟥	🟩	🟩	🟥	F3
MWH-Tro7-1-4	4	🟥	🟥	🟩	🟩	🟥	F4
MWH-Tro8-2-1	6	🟥	🟥	🟩	🟥	🟩	C6
MWH-Tro8-2-5	6	🟥	🟥	🟩	🟥	🟩	C6
MWH-Tro8-3-6	6	🟥	🟥	🟩	🟥	🟩	C6
MWH-Tro8-4-1	6	🟥	🟥	🟩	🟥	🟩	C6
MWH-Tro8-4-4	6	🟥	🟥	🟩	🟥	🟩	C6
MWH-Tro8-2-9	2	🟥	🟥	🟩	🟥	🟩	C2
Tro8F10W22	🟥	🟥	🟩	🟥	🟩	🟥	G-
MWH-RechtKolB	🟥	🟩	🟥	🟥	🟥	🟥	H-
MWH-RechtKolC	🟥	🟩	🟥	🟥	🟥	🟥	H-
MWH-RechtKol4	🟥	🟩	🟥	🟥	🟥	🟥	H-
MWH-Recht1	🟥	🟥	🟥	🟥	🟩	🟥	D-
MWH-GrInsel-1	🟥	🟩	🟥	🟥	🟥	🟥	H-
MWH-Unterer-4A-1	1	🟩	🟥	🟥	🟥	🟥	I1
MWH-Unterer-4A-2	🟥	🟩	🟥	🟥	🟥	🟥	H-

[*] numbers are depicting the sequence type of the sequenced PCR product

Pept., Peptidoglycan/LPS O-acetylase OafA/YrhL, contains acyltransferase and SGNH-hydrolase domains

GT1, Glycosyltransferase involved in LPS biosynthesis, GR25 family

GT2, glycosyl transferase, family 25

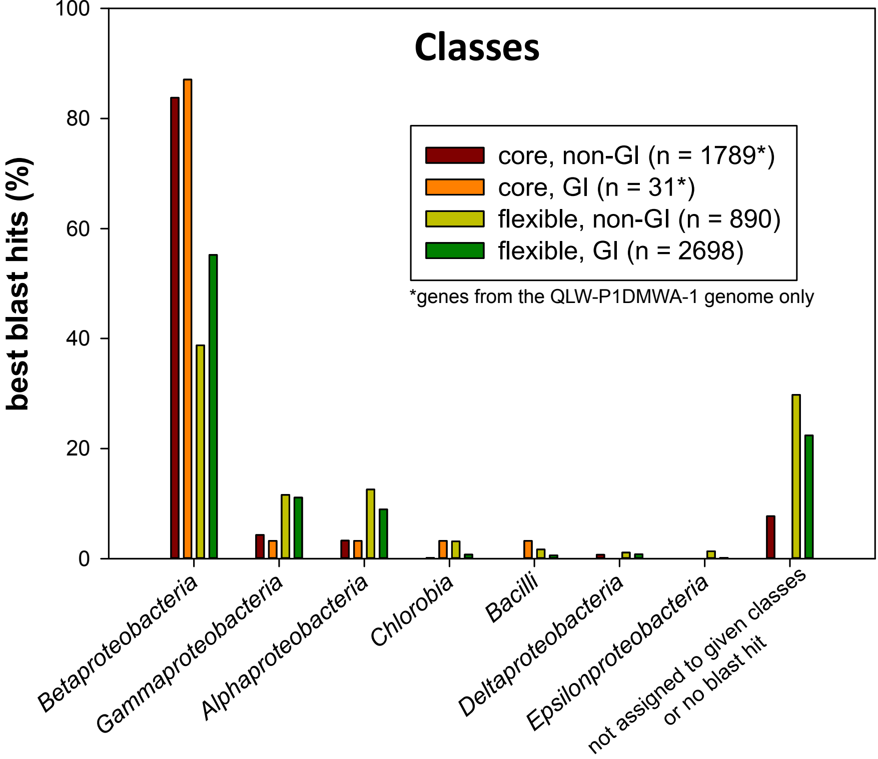
NAT, N-acetylglucosaminyldiphosphoundecaprenol N-acetyl-beta-D-mannosaminyltransferase

NDP, NDP-sugar epimerase, incl. UDP-GlcNAc-inverting 4,6-dehydratase FlaA1 and capsular polysacc. biosynth. prot. EpsC

UDP, UDP-N-acetylmuramyl pentapeptide phosphotransferase

🟩 detected by PCR 🟥 not detected by PCR

Table S6. Presence/absence pattern for six genes contained in CSC and Peptidoglycan/LPS O-acetylase sequence type.



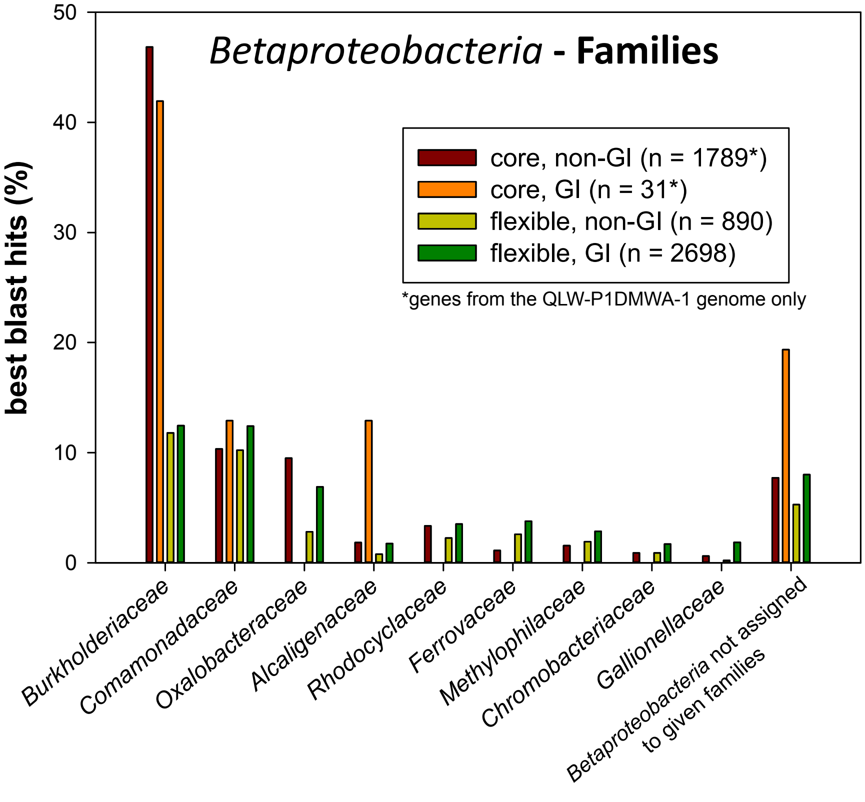


Figure S1. Best blast hit analysis of four groups of genes (core non-GI, core GI, flexible non-GI and flexible GI). Taxonomic assignment is shown on the rank of classes and for *Betaproteobacteria* on the rank of families.

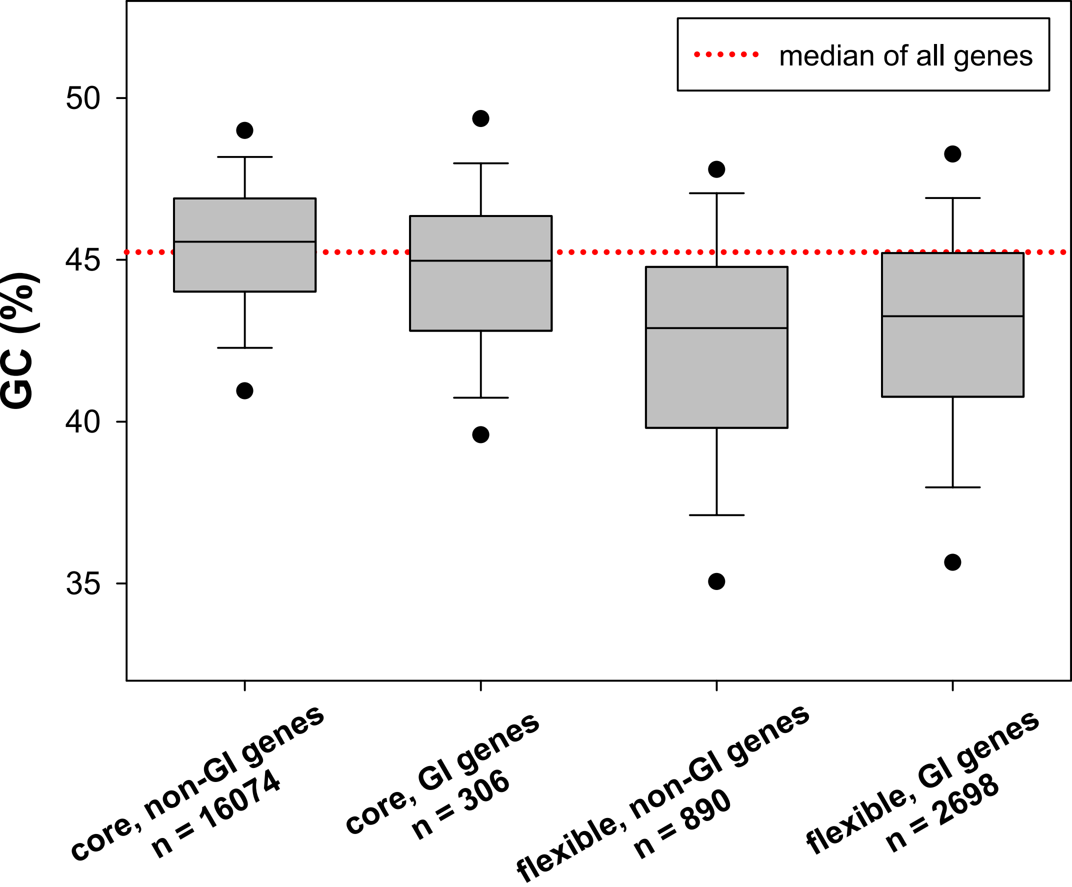


Figure S2. GC% distribution of the genes in the four groups.

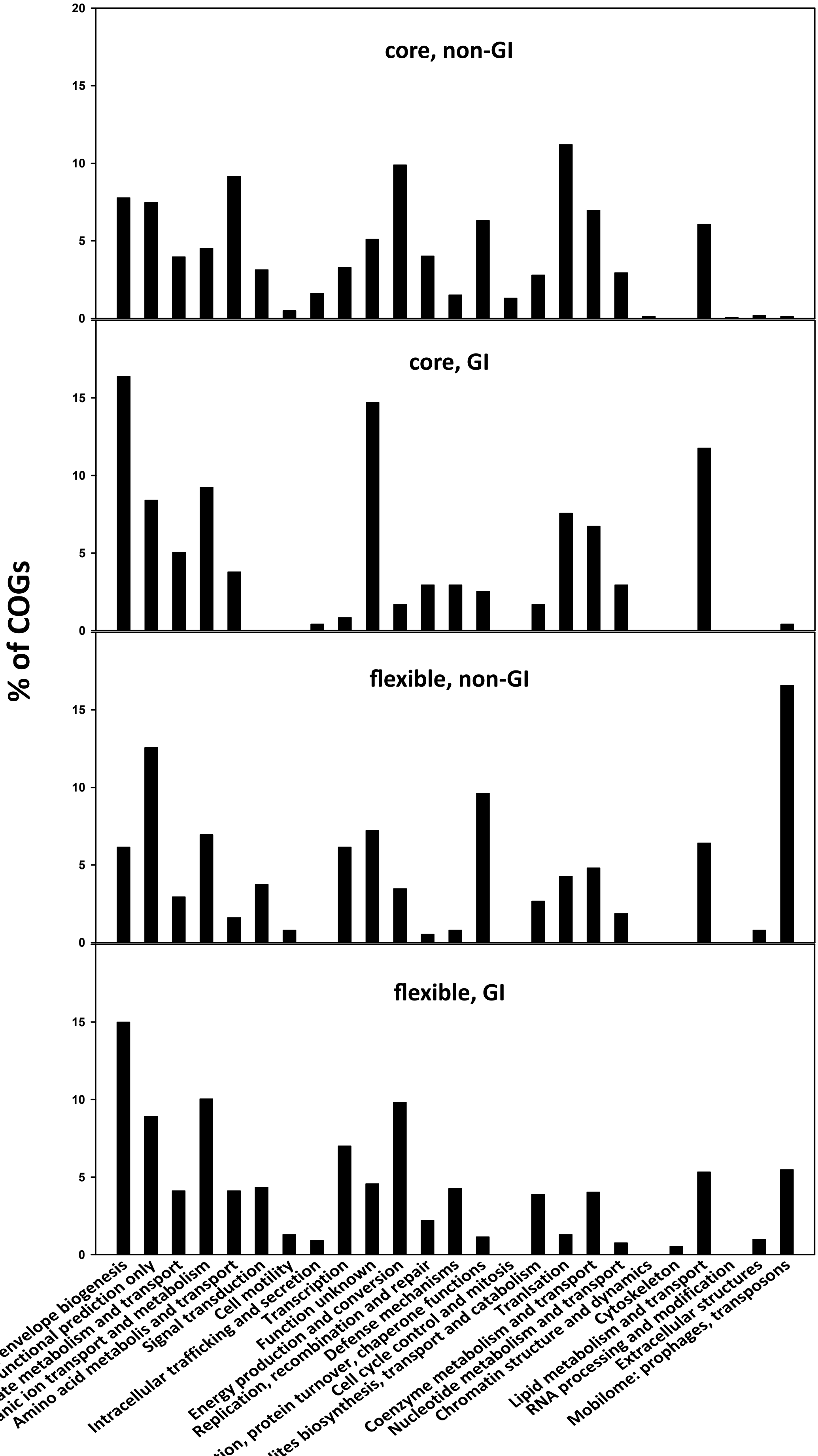


Figure S3. COGs distribution of the genes in the four groups.

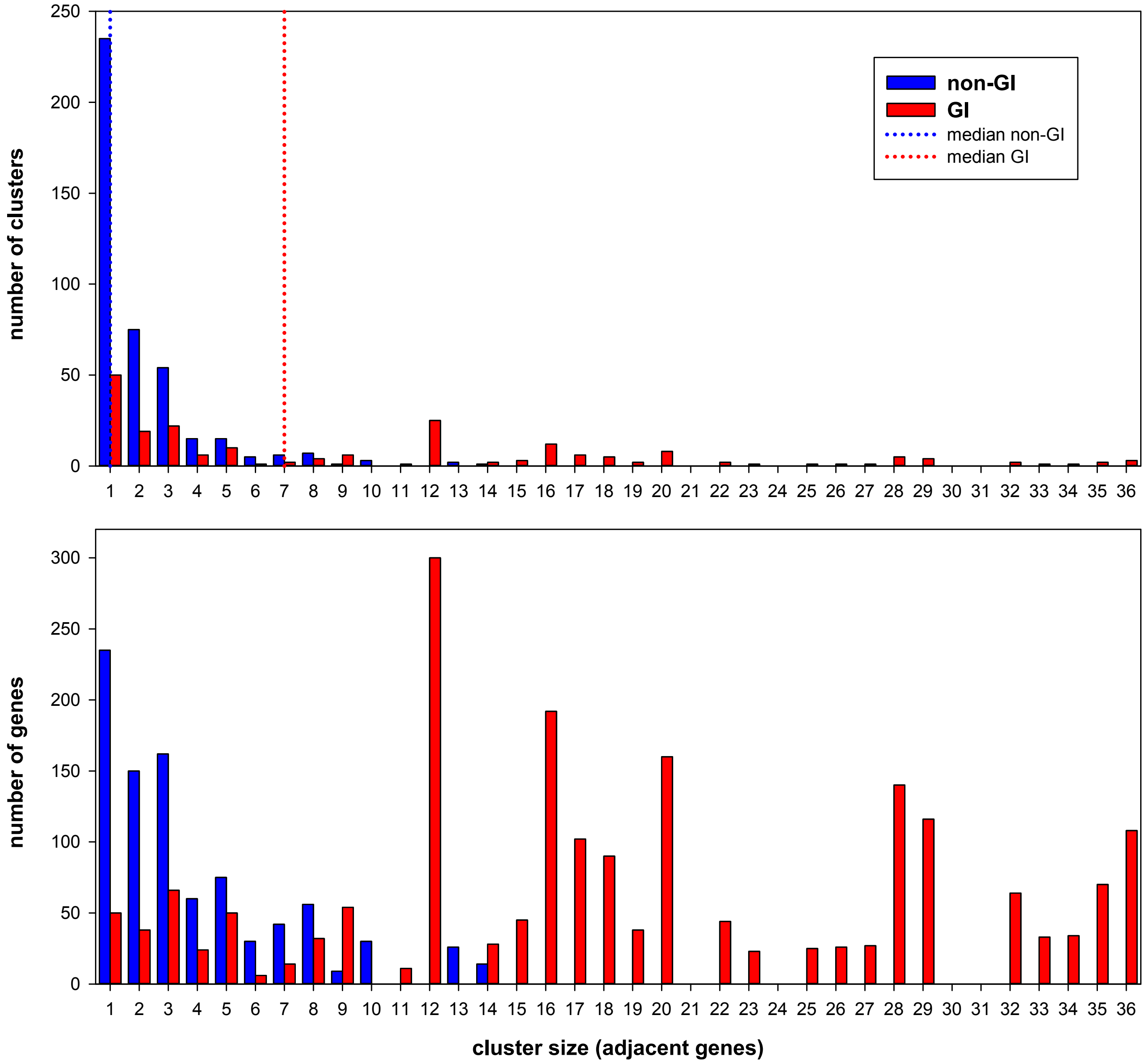


Figure S4. Clustering of flexible GI and non-GI genes.