SUPPLEMENTARY DOCUMENTS

Metaproteomics reveals functional differences in intestinal microbiota development of preterm infants

Zwittink RD[1], van Zoeren-Grobben D[2], Martin R[3], van Lingen RA[2], Groot Jebbink LJ[2], Boeren S[4], Renes IB[3], van Elburg RM[3,5], Belzer C[1,*], Knol J[1,3,*]

1. Laboratory of Microbiology, Wageningen University, Wageningen, the Netherlands
2. Princess Amalia Dpt of Paediatrics, Dpt of Neonatology, Isala, Zwolle, The Netherlands
3. Nutricia Research, Utrecht, The Netherlands
4. Laboratory of Biochemistry, Wageningen University, Wageningen, The Netherlands
5. Emma Children's Hospital, AMC, Amsterdam, The Netherlands

* These authors contributed equally

Running title: Microbiota development in preterm infants

Corresponding Author:

Clara Belzer

Laboratory of Microbiology, Wageningen University; Stippeneng 4, 6708WE, Wageningen, The Netherlands T: +31 483751 E-mail: clara.belzer@wur.nl

SUPPLEMENTARY DOCUMENTS

Content:

Table S1: **Sampling scheme for metaproteomics and 16S-rRNA sequencing**. Red dots: samples for metaproteomics, Black dots: samples for 16s-rRNA gene sequencing, A-J: individual infants.

Table S2: **Contents of the in-house generated protein database.**

Table S3: **Correlation between 16S- and protein-based taxonomic classification.** P-values below 0.05 are considered significant (Spearman correlation with Monte Carlo permutation 10.000x).

Table S4: **Relative abundance of *Bifidobacterium*-derived beta-galactosidase and ABC transporters for oligosaccharides during postnatal weeks 3-6.** Relative abundances were calculated using iBAQ intensities.

Figure S1: **Weighted unifrac distance between (A) and within (B) infants.**
A: Unifrac distance was determined with samples collected from all infants at time points meconium, week 1 and week 2 to show variation between infants at these time points.
B: Unifrac distance was determined with samples collected from each infant during the first two postnatal weeks to show variation within infants.
Mec: meconium, A-J: individual infants

Figure S2: **Microbiota composition profiles during the first two postnatal weeks.** Taxonomy is based on 16S-rRNA gene sequencing**.** Taxonomic profiles are shown for each available time point (1-14) per infants (A-J).

Figure S3: **Principal response curve analysis summarising the differences in protein-based taxonomic profiles (A), 16S-based taxonomic profiles (B) and protein-based KEGG Orthology identifier profiles between EP25-26, EP27 and VP infants throughout postnatal weeks 3-6.** The response curve of VP infants is set as the reference to which the other infant groups are compared.

Figure S4: **Proportion of aerobes/fac. anaerobes and anaerobes according to 16S-rRNA gene sequencing (A) and metaproteomics (B) data.** Per time point, average relative abundance per gestational age group is shown.

Figure S5: **Taxonomic profiles per functional category.** Taxonomic profiles were made for the most abundant KEGG Brite level B functional categories. Per time point, average relative abundances for each GA group are shown. Relative abundances were calculated using iBAQ intensities. Mec: meconium

Figure S6: **Protein-based (A) and 16S-based (B) taxonomic profiles and protein-based functional profiles (C) for all samples collected during postnatal weeks 3-6.**

Table S1: **Sampling scheme for metaproteomics and 16S-rRNA sequencing**. Red dots: samples for metaproteomics, Black dots: samples for 16s-rRNA gene sequencing, A-J: individual infants.

Infant / Time point	**1**	**2**	**3**	**4**	**5**	**6**	**Wk 1**	**8**	**9**	**10**	**11**	**12**	**13**	**Wk 2**	**Wk 3**	**Wk 4**	**Wk 6**
A				●🔴		●	●	🔴	●	●	●	●		●🔴	🔴	●🔴	●🔴
B		●		●				🔴	●	●	●	●	●🔴	●	●🔴	●	●🔴
C			●	●🔴			●🔴	●	●		●	●	●	●🔴	●🔴	●🔴	●🔴
D			●	●	●	●🔴		●	●		●	●	●	●🔴	●🔴	●🔴	●🔴
E		●🔴		●	●		●🔴	●			●	●	●🔴	●	●🔴	●🔴	●🔴
F			●🔴	●			●		●		●	●	●	●🔴	●🔴	🔴	●🔴
G		🔴	●	●		●	●🔴		●	●				🔴	●🔴	●🔴	●🔴
H	●🔴	●🔴	●	●🔴	●🔴		●🔴		●🔴	●	●🔴	●🔴	●🔴	●🔴	●🔴	●🔴	●🔴
I			●🔴	●	●	●		●🔴	●	●	●🔴			🔴	●🔴	●🔴	●🔴
J			●🔴		●	●	●🔴	●	●	●	●	●	●🔴		●🔴	●🔴	●🔴

Table S2: **Contents of the in-house generated protein database**

Kingdom	Species	Sequences	Taxonomy Uniprot ; reference proteome
Bacteria	Bifidobacterium	1723	206672 - Bifidobacterium longum (strain NCC 2705) ; Reference proteome
		1834	484020 - Bifidobacterium bifidum BGN4 ; Reference proteome
		1838	1254439 - Bifidobacterium thermophilum RBL67 ; Reference proteome
		1525	442563 - Bifidobacterium animalis subsp. lactis (strain AD011) ; Reference proteome
		2121	401473 - Bifidobacterium dentium (strain ATCC 27534 / DSM 20436 / JCM 1195 / Bd1) ; Reference proteome
		1629	367928 - Bifidobacterium adolescentis (strain ATCC 15703 / DSM 20083 / NCTC 11814 / E194a) ; Reference proteome
		1913	1385940 - Bifidobacterium breve JCM 7019
		1659	1263060 - Bifidobacterium pseudocatenulatum CAG:263
	Propionibacterium	2294	267747 - Propionibacterium acnes (strain KPA171202 / DSM 16379) ; reference proteome (Reference proteome)
	Xanthomonadales	4365	522373 - Stenotrophomonas maltophilia (strain K279a) ; Reference proteome
	Corynebacterium	2265	257309 - Corynebacterium diphtheriae (strain ATCC 700971 / NCTC 13129 / Biotype gravis) ; Reference proteome
		2508	548476 - Corynebacterium aurimucosum (strain ATCC 700975 / DSM 44827 / CN-1) ; Reference proteome
		2018	645127 - Corynebacterium kroppenstedtii (strain DSM 44385 / CCUG 35717) ; Reference proteome
	Bacteroides	3543	1263037 - Bacteroides caccae CAG:21 ; Reference proteome
		4782	226186 - Bacteroides thetaiotaomicron (strain ATCC 29148 / DSM 2079 / NCTC 10582 / E50 / VPI-5482) ; Reference proteome
		4234	272559 - Bacteroides fragilis (strain ATCC 25285 / NCTC 9343) ; Reference proteome
		4465	357276 - Bacteroides dorei
	Lactobacillus	3038	220668 - Lactobacillus plantarum (strain ATCC BAA-793 / NCIMB 8826 / WCFS1); Reference proteome
		2688	321967 - Lactobacillus casei (strain ATCC 334) ; Reference proteome
		1865	557436 - Lactobacillus reuteri (strain DSM 20016) ; reference proteome
		2168	387344 - Lactobacillus brevis (strain ATCC 367 / JCM 1170) ; reference proteome
		1851	1069534 - Lactobacillus ruminis (strain ATCC 27782 / RF3) ; reference proteome
		1703	362948 - Lactobacillus salivarius (strain UCC118) ; Reference proteome
		1859	272621 - Lactobacillus acidophilus (strain ATCC 700396 / NCK56 / N2 / NCFM) ; Reference proteome
		2886	1088720 - Lactobacillus rhamnosus ATCC 8530
		2749	1446494 - Lactobacillus paracasei N1115

	Staphylococcus	2889	93061 - Staphylococcus aureus (strain NCTC 8325) ; Reference proteome
		2463	176279 - Staphylococcus epidermidis (strain ATCC 35984 / Reference proteome62A) ; Reference proteome
	Enterococcus	3101	226185 - Enterococcus faecalis (strain ATCC 700802 / V583) ; Reference proteome
		2671	333849 - Enterococcus faecium DO ; Reference proteome
	Streptococcus	1946	862971 - Streptococcus anginosus C238
		1840	1069533 - Streptococcus infantarius (strain CJ18)
		2105	208435 - Streptococcus agalactiae serotype V (strain ATCC BAA-611 / 2603 V/R) ; Reference proteome
		2030	171101 - Streptococcus pneumoniae (strain ATCC BAA-255 / R6) ; Reference proteome
	Clostridium	2659	195102 - Clostridium perfringens (strain 13 / Type A); Reference proteome
		3752	272563 - Peptoclostridium difficile (strain 630) ; Reference proteome
	Enterobacter	5044	716541 - Enterobacter cloacae subsp. cloacae (strain ATCC 13047 / DSM 30054 / NBRC 13535 / NCDC 279-56) ; Reference proteome
	Escherichia	4305	83333 - Escherichia coli (strain K12) ; Reference proteome
	Shigella	3691	300267 - Shigella dysenteriae serotype 1 (strain Sd197)
	Klebsiella	4765	272620 - Klebsiella pneumoniae subsp. pneumoniae (strain ATCC 700721 / MGH 78578) ; Reference proteome
	Proteus	3606	529507 - Proteus mirabilis (strain HI4320) ; Reference proteome
	Veillonella	1843	479436 - Veillonella parvula (strain ATCC 10790 / DSM 2008 / JCM 12972 / Te3) ; Reference proteome
	Pseudomonas	5563	208964 - Pseudomonas aeruginosa (strain ATCC 15692 / PAO1 / 1C / PRS 101 / LMG 12228) ; Reference proteome
	Citrobacter	4991	290338 - Citrobacter koseri (strain ATCC BAA-895 / CDC 4225-83 / SGSC4696) ; Reference proteome
		4651	1333848 - Citrobacter freundii CFNIH1
	Bacillus	5219	226900 - Bacillus cereus (strain ATCC 14579 / DSM 31) ; Reference proteome
		5066	1367477 - Bacillus infantis NRRL B-14911 ; Reference proteome
	Paenibacillus	6210	324057 - Paenibacillus sp. (strain JDR-2) ; Reference proteome
	Faecalibacterium	2755	718252 - Faecalibacterium prausnitzii L2-6 ; Reference proteome
	Ruminococcus	2797	657313 - Ruminococcus torques L2-14 ; Reference proteome
		1811	657321 - Ruminococcus bromii L2-63
	Acinetobacter	3465	470 - Acinetobacter baumannii ; Reference proteome
		3598	871585 - Acinetobacter calcoaceticus (strain PHEA-2)
	Rhodanobacter	3773	666685 - Rhodanobacter denitrificans ; Reference proteome
	Eubacterium	4506	903814 - Eubacterium limosum (strain KIST612) ; Reference proteome

		3545	515619 - Eubacterium rectale (strain ATCC 33656 / VPI 0990) ; Reference proteome
		2309	1263078 - Eubacterium hallii CAG:12
	Anaerostipes	3141	1262699 - Anaerostipes sp. CAG:276 ; Reference proteome
	Roseburia	3629	718255 - Roseburia intestinalis XB6B4 ; Reference proteome
		2653	1263105 - Roseburia inulinivorans CAG:15 ; Reference proteome
		3351	585394 - Roseburia hominis (strain DSM 16839 / NCIMB 14029 / A2-183) ; Reference proteome
	Blautia	3087	1263061 - Blautia hydrogenotrophica CAG:147 ; Reference proteome
	Erysipelotrichaceae	3917	1262981 - Erysipelotrichaceae bacterium CAG:64 ; Reference proteome
	Fusobacterium	2046	190304 - Fusobacterium nucleatum subsp. nucleatum (strain ATCC 25586 / CIP 101130 / JCM 8532 / LMG 13131) ; Reference proteome
	Succinatimonas	1970	1262974 - Succinatimonas sp. CAG:777 ; Reference proteome
	Haemophilus	1707	71421 - Haemophilus influenzae (strain ATCC 51907 / DSM 11121 / KW20 / Rd) ; Reference proteome
		1694	233412 - Haemophilus ducreyi (strain 35000HP / ATCC 700724) ; Reference proteome
	Anaerococcus	1691	525919 - Anaerococcus prevotii (strain ATCC 9321 / DSM 20548 / JCM 6508 / PC1) ; Reference proteome
	Gardnerella	1365	525284 - Gardnerella vaginalis (strain ATCC 14019 / 317) ; Reference proteome
	Finegoldia	1631	334413 - Finegoldia magna (strain ATCC 29328) ; Reference proteome
	Rothia	1991	680646 - Rothia mucilaginosa (strain DY-18) ; Reference proteome
		2212	762948 - Rothia dentocariosa (strain ATCC 17931 / CDC X599 / XDIA) ; Reference proteome
	Parabacteriodes	3830	435591 - Parabacteroides distasonis (strain ATCC 8503 / DSM 20701 / NCTC 11152) ; Reference proteome
		2964	1263094 - Parabacteroides merdae CAG:48 ; Reference proteome
	Bradyrhizobium	8253	224911 - Bradyrhizobium diazoefficiens (strain JCM 10833 / IAM 13628 / NBRC 14792 / USDA 110); Reference proteome
	Achromobacter	6445	762376 - Achromobacter xylosoxidans (strain A8) ; Reference proteome
	Subdoligranulum	1341	1262970 - Subdoligranulum sp. CAG:314 ; Reference proteome
	Blastococcus	4793	1146883 - Blastococcus saxobsidens (strain DD2) ; Reference proteome
	Micrococcus	2207	465515 - Micrococcus luteus (strain ATCC 4698 / DSM 20030 / JCM 1464 / NBRC 3333 / NCIMB 9278 / NCTC 2665 / VKM Ac-2230) ; Reference proteome
	Dyella	4342	1379159 - Dyella jiangningensis ; Reference proteome
	Microbacterium	3671	979556 - Microbacterium testaceum (strain StLB037) ; Reference proteome
	Paracoccus	3363	1367847 - Paracoccus aminophilus JCM 7686 ; Reference proteome
		4403	318586 - Paracoccus denitrificans (strain Pd 1222) ; Reference proteome

	Burkholdia	4872	13373 - Burkholderia mallei
		6902	216591 - Burkholderia cenocepacia (strain ATCC BAA-245 / DSM 16553 / LMG 16656 / NCTC 13227 / J2315 / CF5610) ; Reference proteome
	Methylobacterium	6566	426117 - Methylobacterium sp. (strain 4-46) ; Reference proteome
	Bartonella	1466	283166 - Bartonella henselae (strain ATCC 49882 / Houston 1) ; Reference proteome
Fungi	Candida	9044	237561 - Candida albicans (strain SC5314 / ATCC MYA-2876) ; Reference proteome
		6226	294747 - Candida tropicalis (strain ATCC MYA-3404 / T1) ; Reference proteome
		5780	578454 - Candida parapsilosis (strain CDC 317 / ATCC MYA-4646) ; Reference proteome
		5209	284593 - Candida glabrata (strain ATCC 2001 / CBS 138 / JCM 3761 / NBRC 0622 / NRRL Y-65) ; Reference proteome
Animalia	Homo sapiens	67493	9606 - Homo sapiens ; Reference proteome
	Bos taurus	23898	9913 - Bos taurus ; Reference proteome

Table S3: **Correlation between 16S- and protein-based taxonomic classification**. P-values below 0.05 are considered significant (Spearman correlation with Monte Carlo permutation 10.000x).

	Bifidobacterium	***Enterobacter/Klebsiella***	***Enterococcus***	***Streptococcus***	***Veilonella***	***Clostridium***	***Staphylococcus***
Spearman correlation	0.794*	0.739*	0.572*	0.686*	0.029	0.495*	0.226
p-value	7.537E-07	1.063E-05	0.002	7.809E-05	0.884	0.009	0.256
	Corynebacterium	***Propionibacterium***	***Lactobacillus***	***Escherichia***	***Finegoldia***	***Anaerococcus***	***Haemophilus***
Spearman correlation	0.206	0.114	0.052	-0.210	0.317	0.097	-0.099
p-value	0.304	0.573	0.798	0.293	0.107	0.629	0.622

* Significant correlation

Table S4: **Relative abundance of *Bifidobacterium*-derived beta-galactosidase and ABC transporters for oligosaccharides during postnatal weeks 3-6.** Relative abundances were calculated using iBAQ intensities.

Bifidobacterium derived beta-galactosidase

GA group	Mean	SD	25-26 vs 27	25-26 vs 30	27 vs 30
EP 25-26	1.38E-04	1.57E-04	0.258	0.029*	0.028*
EP 27	5.13E-05	9.70E-05			
VP	8.93E-03	1.40E-02			

Bifidobacterium derived ABC transporters for oligosaccharides

GA group	Mean	SD	25-26 vs 27	25-26 vs 30	27 vs 30
EP 25-26	2.60E-03	3.26E-03	0.171	0.00007**	0.00002**
EP 27	5.91E-04	8.85E-04			
VP	1.70E-02	1.01E-02			

* Independent sample t-test p<0.05

** Independent sample t-test P<0.001

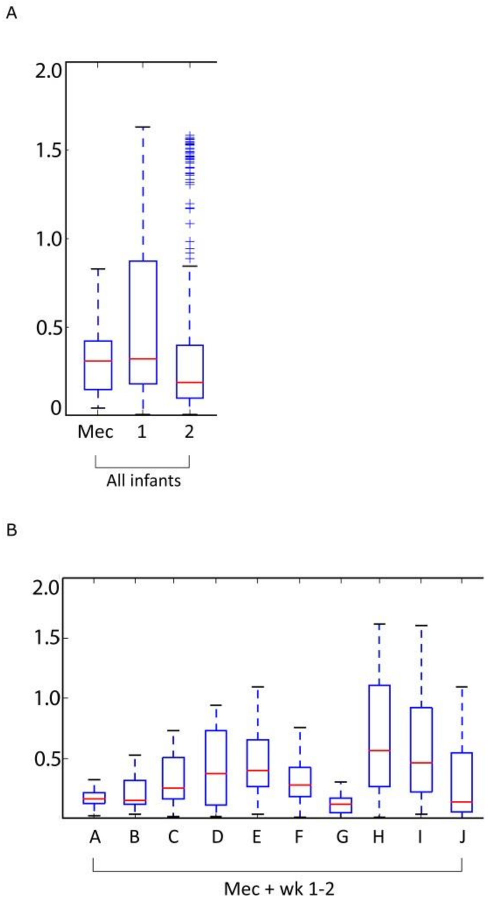

Figure S1: **Weighted unifrac distance between (A) and within (B) infants.**

A: Unifrac distance was determined with samples collected from all infants at time points meconium, week 1 and week 2 to show variation between infants at these time points.

B: Unifrac distance was determined with samples collected from each infant during the first two postnatal weeks to show variation within infants.

Mec: meconium, A-J: individual infants

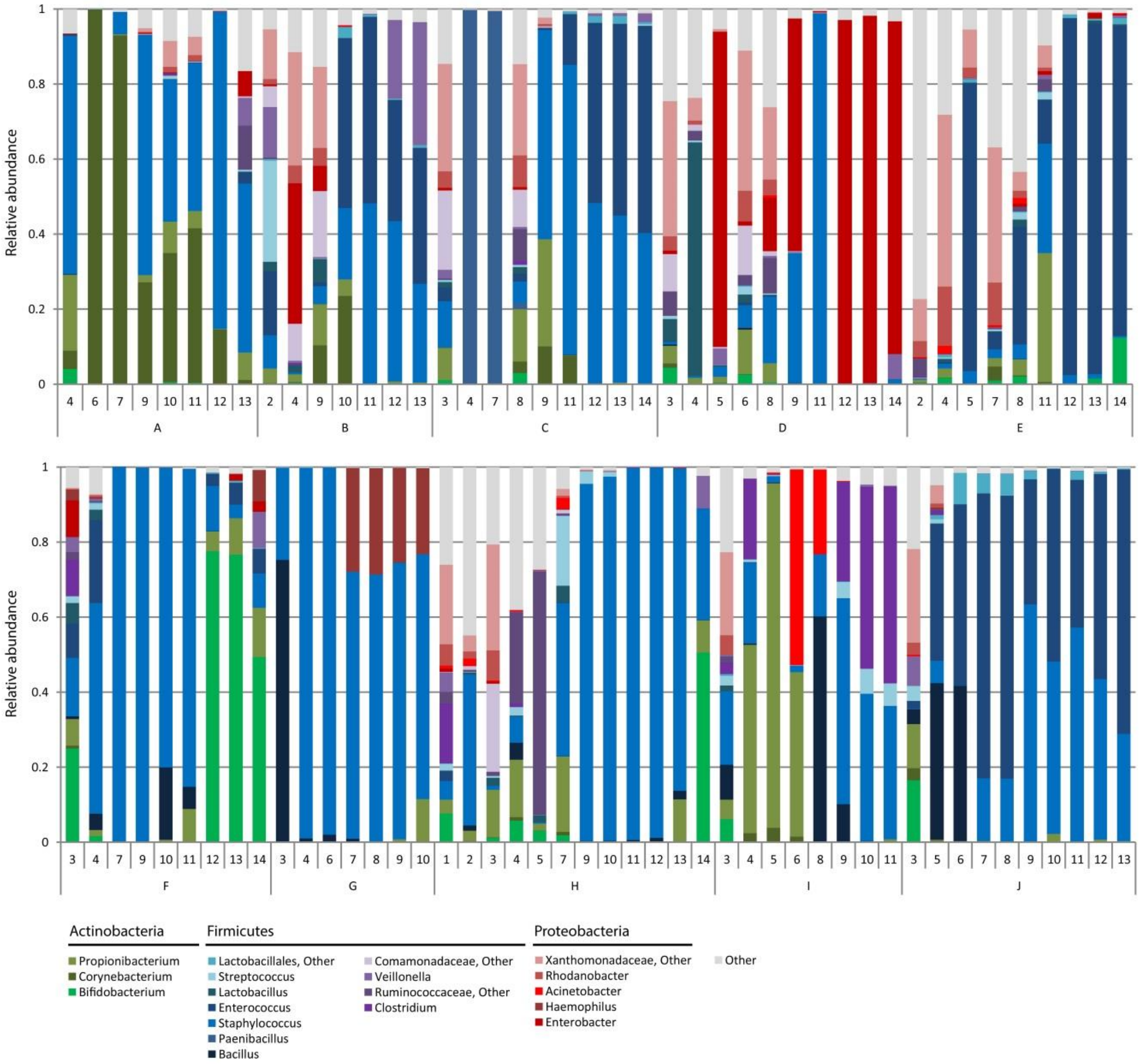

Figure S2: **Microbiota composition profiles during the first two postnatal weeks.** Taxonomy is based on 16S-rRNA gene sequencing**.** Taxonomic profiles are shown for each available time point (1-14) per infants (A-J).

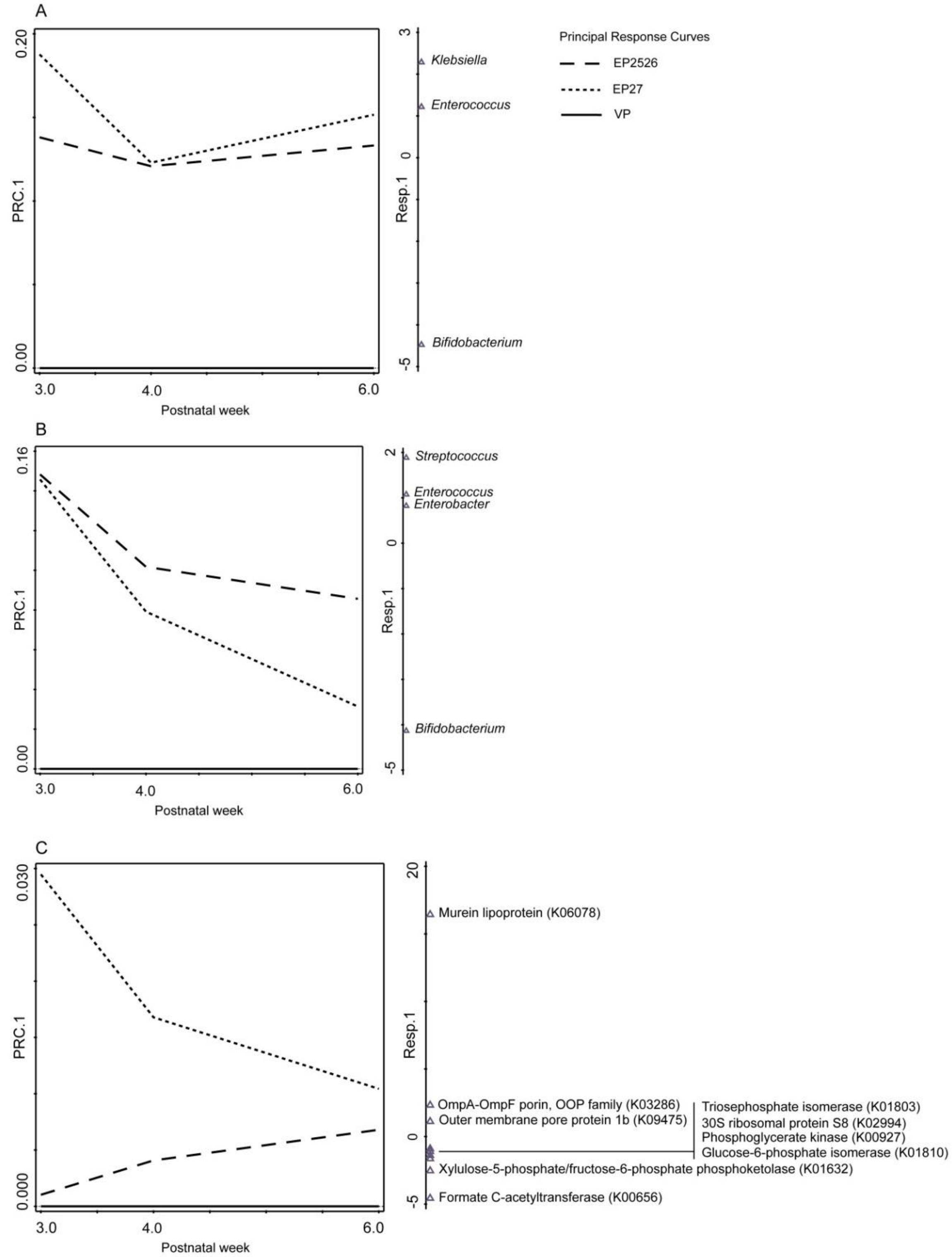

Figure S3: **Principal response curve analysis summarising the differences in protein-based taxonomic profiles (A), 16S-based taxonomic profiles (B) and protein-based KEGG Orthology identifier profiles between EP25-26, EP27 and VP infants throughout postnatal weeks 3-6.** Genera and proteins with a score lower than -0.85 or higher than 0.85 are shown on Resp1.

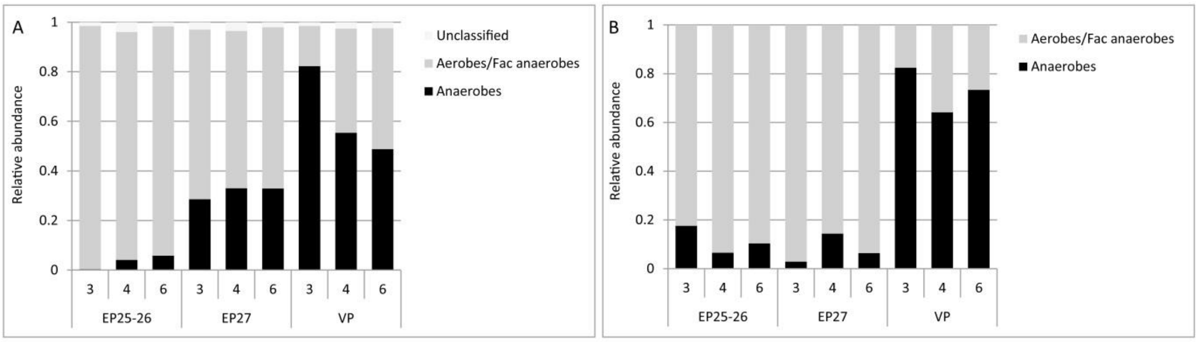

Figure S4: **Proportion of aerobes/fac. anaerobes and anaerobes according to 16S-rRNA gene sequencing (A) and metaproteomics (B) data.** Per time point, average relative abundance per gestational age group is shown.

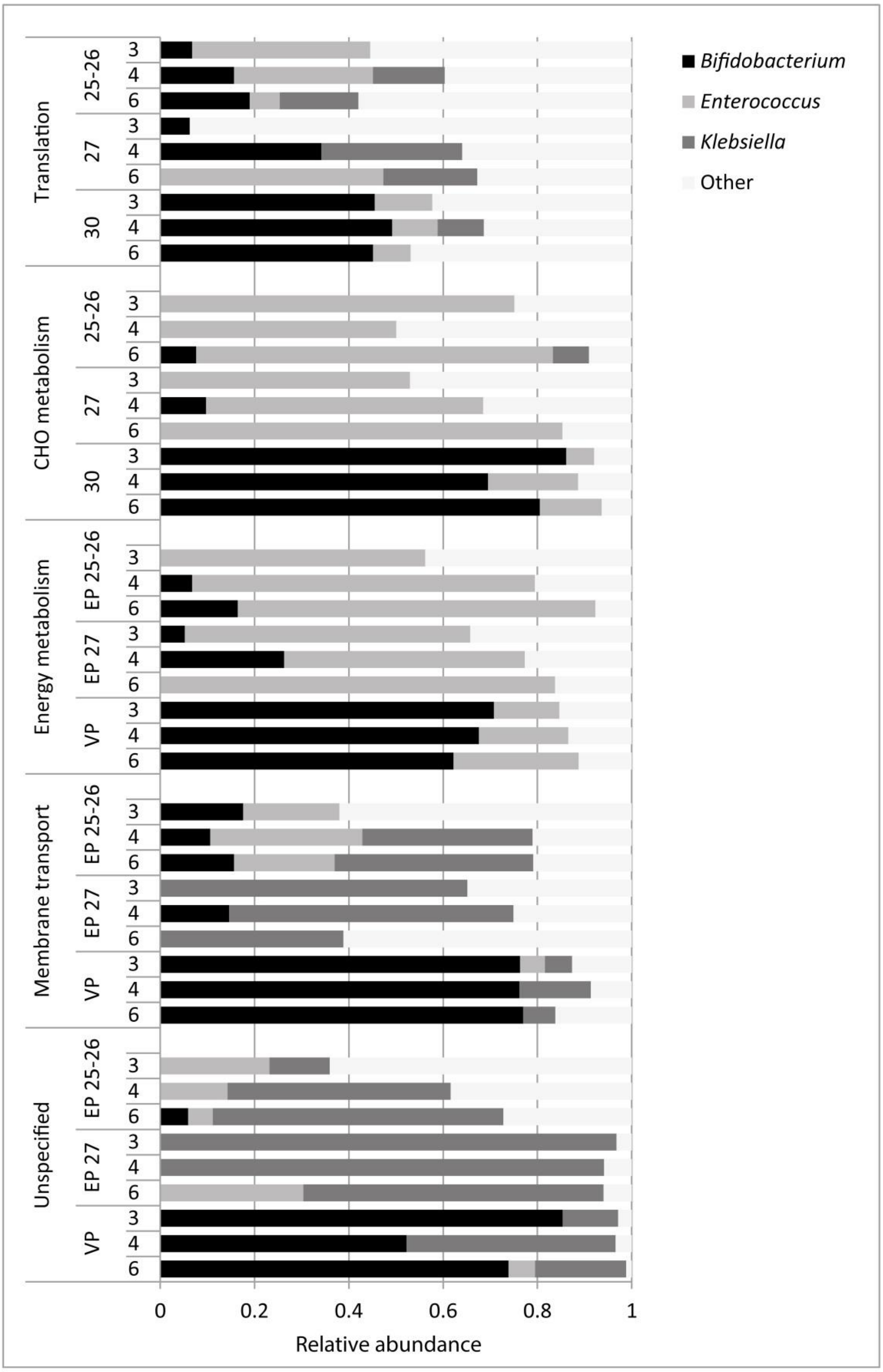

Figure S5: **Taxonomic profiles per functional category.** Taxonomic profiles were made for the most abundant KEGG Brite level B functional categories. Per time point, average relative abundances for each GA group are shown. Relative abundances were calculated using iBAQ intensities.

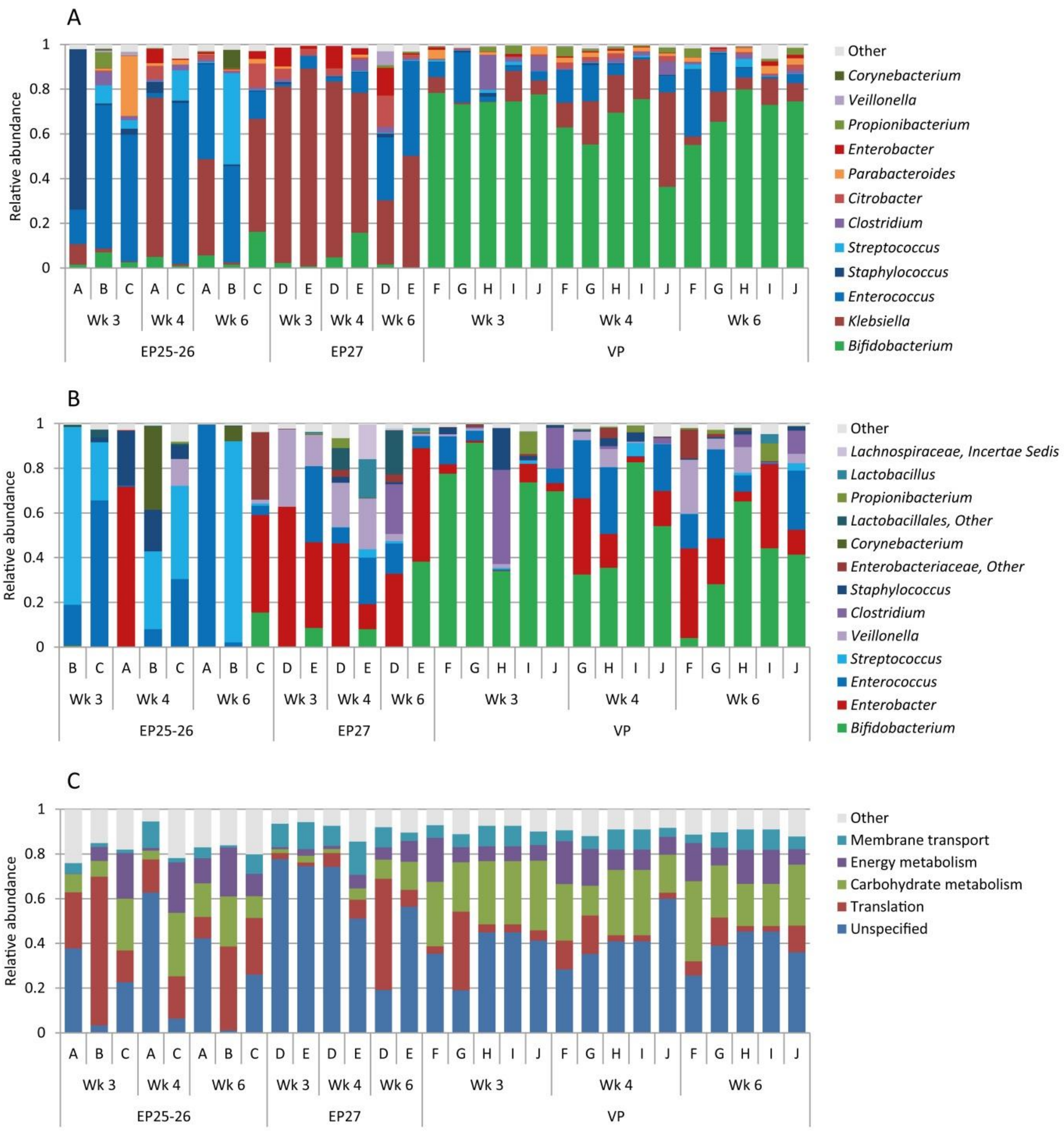

Figure S6: **Protein-based (A) and 16S-based (B) taxonomic profiles and protein-based functional profiles (C) for all samples collected during postnatal weeks 3-6.**