

Supplementary Material for ‘phyLoSTM: A Novel Deep Learning model on Disease Prediction from Longitudinal Microbiome data’

Contents

1	Supplementary Tables	4
2	Supplementary Figures	27
3	References	35

List of Supplementary Tables

1	Results evaluating performance of our model by changing network parameters on the simulated data. The boldfaced attributes represent the parameter values for which the model performs the best.	4
2	Table detailing the clusters in the DIABIMMUNE three country cohort [1] based on the phyla containing maximum number of OTUs. The right handside represents the species in each cluster. The numbering provided to each species provides a unique identifier to each OTU which is further used in Heatmaps as labels for the x and y axis, in Supplementary Figure 1, 2 and 3 to illustrate the correlations between the OTUs.	5
3	Association of environmental variables to various allergy outcomes in the DIABIMMUNE three country cohort.	18
4	Distribution of samples into cases and controls, in each allergy outcome in the DIABIMMUNE three country cohort.	19

5	AUC values tabulated for neural network models on test set of DIABIMMUNE three country cohort. The results are reported on considering model performance without (w/o) weight balancing in the neural network and with weight balancing in the neural network. Note the consistent increase in AUC as shown in column 3, 5, 7 and 9, using weight balancing.	20
6	Association of environmental variables to preterm birth outcome in the DiGiulio case-control study.	21
7	AUC as obtained by various feature extraction techniques combined with LSTM module to identify temporal dependency in DIABIMMUNE three country cohort with overall allergy as outcome. Alongwith conventional ML models, Multi dimensional feature selection [2] is also used for comparison against taxoNN for feature extraction.	22
8	Comparison of AUC values obtained on the test set of OTUs from metagenomic dataset and 16s dataset of the DIABIMMUNE three country cohort on the 'Overall Allergy' outcome.	23
9	AUC values tabulated for various machine learning methods upon class based stratification for DIABIMMUNE three country and DiGuilio preterm studies. . . .	24
10	Comparison of AUCs as obtained by phyLoSTM methodology incorporating temporal data across 6 time points in the DIABIMMUNE cohort for the overall allergy outcome with conventional ML models taking data from individual time points as inputs.	25
11	Comparison of AUCs as obtained by phyLoSTM methodology incorporating temporal data across 6 time points in the DiGiulio preterm study with conventional ML models taking data from individual time points as inputs.	26

List of Supplementary Figures

1	Heatmaps for the Spearman rank of the OTUs in the cluster, Phylum Proteobacteria, (a) before ordering and (b) after the ordering based on correlation of the OTUs in the DIABIMMUNE three country cohort	28
2	Heatmaps for the Spearman rank of the OTUs in the cluster, Phylum Firmicutes, (a) before ordering and (b) after the ordering based on correlation of the OTUs in the DIABIMMUNE three country cohort	29

3	Heatmaps for the Spearman rank of the OTUs in the cluster, Phylum Bacteroidetes, (a) before ordering and (b) after the ordering based on correlation of the OTUs in the DIABIMMUNE three country cohort	30
4	Comparison of AUC values as observed at each individual 6 time points using <i>taxoNN_{corr}</i> approach with the final higher AUC value of 0.897 using proposed phyLoSTM framework on the simulated study.	31
5	Relative abundance of OTUs in each phylum of the DiGiulio study at four time points measured every three months. For each box plot the upper whisker extends from the hinge to the largest value no further than 1.5 * IQR from the hinge (where IQR is the inter-quartile range, or distance between the first and third quartiles). The lower whisker extends from the hinge to the smallest value at most 1.5 * IQR of the hinge. Data beyond the end of the whiskers are called "outlying" points and are plotted individually.	32
6	95% confidence intervals obtained for the mean AUC values for 10 times 10-fold cross validation on the training set for the DiGiulio study on the preterm delivery output [3]	33
7	Relative abundance of OTUs in each phylum of the DIABIMMUNE three country cohort at six time points measured half-yearly upto age of three of the infant. For each box plot the upper whisker extends from the hinge to the largest value no further than 1.5 * IQR from the hinge (where IQR is the inter-quartile range, or distance between the first and third quartiles). The lower whisker extends from the hinge to the smallest value at most 1.5 * IQR of the hinge.	34

1 Supplementary Tables

Supplementary Table 1: Results evaluating performance of our model by changing network parameters on the simulated data. The boldfaced attributes represent the parameter values for which the model performs the best.

Performance Analysis		
Parameter	Values	AUC
Stride Size	1	0.878
	2	0.856
	3	0.844
	4	0.821
	5	0.806
Window Size	3	0.822
	4	0.844
	5	0.871
	6	0.849
	7	0.838
No. of filters	16	0.839
	32	0.871
	64	0.848
No. of Causal OTUs	8	0.787
	16	0.841
	32	0.875

Supplementary Table 2: Table detailing the clusters in the DIABIMMUNE three country cohort [1] based on the phyla containing maximum number of OTUs. The right handside represents the species in each cluster. The numbering provided to each species provides a unique identifier to each OTU which is further used in Heatmaps as labels for the x and y axis, in Supplementary Figure 1, 2 and 3 to illustrate the correlations between the OTUs.

OTUs in DIABIMMUNE Three country cohort		
Cluster	Phylum	Species
Cluster 1	p_Proteobacteria	1. s__Brevundimonas_diminuta 2.s__Brevundimonas_unclassified 3.s__Bartonella_unclassified 4.s__Agrobacterium_tumefaciens 5.s__Agrobacterium_unclassified 6.s__Shinella_zoogloeoides 7.s__Paracoccus_unclassified 8.s__Gluconobacter_frateurii 9.s__Sphingobium_unclassified 10.s__Sphingobium_yanoikuyae 11.s__Bordetella_bronchiseptica_parapertussis 12.s__Bordetella_pertussis 13.s__Burkholderia_unclassified 14.s__Cupriavidus_unclassified 15.s__Lautropia_mirabilis 16.s__Ralstonia_pickettii 17.s__Ralstonia_unclassified 18.s__Burkholderiales_bacterium_1_1_47 19.s__Comamonas_unclassified 20.s__Variovorax_unclassified 21.s__Oxalobacter_formigenes 22.s__Parasutterella_excrementihominis 23.s__Sutterella_parvirubra 24.s__Sutterella_wadsworthensis 25.s__Eikenella_corrodens 26.s__Neisseria_elongata 27.s__Neisseria_flavescens 28.s__Neisseria_meningitidis 29.s__Neisseria_subflava 30.s__Neisseria_unclassified 31.s__Bilophila_unclassified 32.s__Bilophila_wadsworthia

33.s__Desulfovibrio_desulfuricans
34.s__Desulfovibrio_piger
35.s__Campylobacter_conciscus
36.s__Campylobacter_curvus
37.s__Campylobacter_gracilis
38.s__Campylobacter_jejuni
39.s__Campylobacter_showae
40.s__Campylobacter_upsaliensis
41.s__Campylobacter_ureolyticus
42.s__Aeromonas_caviae
43.s__Aeromonas_unclassified
44.s__Aeromonas_veronii
45.s__Succinatimonas_hippeii
46.s__Spiribacter_unclassified
47.s__Buchnera_aphidicola
48.s__Citrobacter_freundii
49.s__Citrobacter_koseri
50.s__Citrobacter_unclassified
51.s__Citrobacter_youngae
52.s__Cronobacter_sakazakii
53.s__Enterobacteriaceae_bacterium_9_2_54FAA
54.s__Enterobacter_aerogenes
55.s__Enterobacter_cloacae
56.s__Enterobacter_hormaechei
57.s__Escherichia_albertii
58.s__Escherichia_coli
59.s__Escherichia_hermannii
60.s__Escherichia_unclassified
61.s__Hafnia_alvei
62.s__Klebsiella_oxytoca
63.s__Klebsiella_pneumoniae
64.s__Klebsiella_sp_MS_92_3
65.s__Klebsiella_unclassified
66.s__Morganella_morganii
67.s__Pantoea_agglomerans
68.s__Pantoea_unclassified
69.s__Proteus_mirabilis
70.s__Proteus_penneri
71.s__Proteus_unclassified
72.s__Providencia_unclassified
73.s__Raoultella_ornithinolytica
74.s__Salmonella_enterica

		<p>75.s__Salmonella_unclassified 76.s__Serratia_liquefaciens 77.s__Shigella_flexneri 78.s__Shigella_sonnei 79.s__Yersinia_enterocolitica 80.s__Yersinia_unclassified 81.s__Halomonas_unclassified 82.s__Actinobacillus_unclassified 83.s__Aggregatibacter_aphrophilus 84.s__Aggregatibacter_segnis 85.s__Aggregatibacter_unclassified 86.s__Haemophilus_haemolyticus 87.s__Haemophilus_influenzae 88.s__Haemophilus_paraahaemolyticus 89.s__Haemophilus_parainfluenzae 90.s__Haemophilus_paraphroahaemolyticus 91.s__Haemophilus_pittmaniae 92.s__Haemophilus_sputorum 93.s__Acinetobacter_baumannii 94.s__Acinetobacter_guillouiae 95.s__Acinetobacter_johnsonii 96.s__Acinetobacter_junii 97.s__Acinetobacter_pittii_calcoaceticus_nosocomialis 98.s__Acinetobacter_unclassified 99.s__Acinetobacter_ursingii 100.s__Enhydrobacter_aerosaccus 101.s__Cellvibrio_unclassified 102.s__Pseudomonas_alcaligenes 103.s__Pseudomonas_fragi 104.s__Pseudomonas_unclassified 105.s__Stenotrophomonas_maltophilia 106.s__Stenotrophomonas_unclassified</p>
Cluster 2	p_Firmicutes	<p>107. s__Gemella_haemolysans 108.s__Gemella_sanguinis 109.s__Gemella_unclassified 110.s__Macrococcus_caseolyticus 111.s__Staphylococcus_aureus 112.s__Staphylococcus_caprae_capitis 113.s__Staphylococcus_epidermidis 114.s__Staphylococcus_haemolyticus 115.s__Staphylococcus_hominis 116.s__Staphylococcus_lugdunensis</p>

117.s__Staphylococcus_saprophyticus
118.s__Abiotrophia_defectiva
119.s__Carnobacterium_maltaromaticum
120.s__Granulicatella_adiacens
121.s__Granulicatella_elegans
122.s__Granulicatella_unclassified
123.s__Enterococcus_avium
124.s__Enterococcus_casseliflavus
125.s__Enterococcus_cecorum
126.s__Enterococcus_dispar
127.s__Enterococcus_durans
128.s__Enterococcus_faecalis
129.s__Enterococcus_faecium
130.s__Enterococcus_gallinarum
131.s__Enterococcus_gilvus
132.s__Enterococcus_hirae
133.s__Enterococcus_italicus
134.s__Enterococcus_malodoratus
135.s__Enterococcus_raffinosis
136.s__Lactobacillus_acidophilus
137.s__Lactobacillus_amylovorus
138.s__Lactobacillus_brevis
139.s__Lactobacillus_buchneri
140.s__Lactobacillus_casei_paracasei
141.s__Lactobacillus_coryniformis
142.s__Lactobacillus_curvatus
143.s__Lactobacillus_delbrueckii
144.s__Lactobacillus_fermentum
145.s__Lactobacillus_gasseri
146.s__Lactobacillus_helveticus
147.s__Lactobacillus_hilgardii
148.s__Lactobacillus_johnsonii
149.s__Lactobacillus_kefiranofaciens
150.s__Lactobacillus_mucosae
151.s__Lactobacillus_oris
152.s__Lactobacillus_otakiensis
153.s__Lactobacillus_parafarraginis
154.s__Lactobacillus_plantarum
155.s__Lactobacillus_reuteri
156.s__Lactobacillus_rhamnosus
157.s__Lactobacillus_rossiae
158.s__Lactobacillus_ruminis

159.s__Lactobacillus_sakei
160.s__Lactobacillus_salivarius
161.s__Lactobacillus_vaginalis
162.s__Lactobacillus_versmoldensis
163.s__Lactobacillus_zeae
164.s__Pediococcus_acidilactici
165.s__Pediococcus_lolii
166.s__Pediococcus_pentosaceus
167.s__Pediococcus_unclassified
168.s__Leuconostoc_citreum
169.s__Leuconostoc_lactis
170.s__Leuconostoc_mesenteroides
171.s__Leuconostoc_pseudomesenteroides
172.s__Leuconostoc_unclassified
173.s__Weissella_cibaria
174.s__Weissella_confusa
175.s__Weissella_paramesenteroides
176.s__Weissella_unclassified
177.s__Lactococcus_garvieae
178.s__Lactococcus_lactis
179.s__Lactococcus_raffinolactis
180.s__Streptococcus_agalactiae
181.s__Streptococcus_anginosus
182.s__Streptococcus_australis
183.s__Streptococcus_constellatus
184.s__Streptococcus_dysgalactiae
185.s__Streptococcus_gallolyticus
186.s__Streptococcus_gordonii
187.s__Streptococcus_infantarius
188.s__Streptococcus_infantis
189.s__Streptococcus_intermedius
190.s__Streptococcus_lutetiensis
191.s__Streptococcus_macedonicus
192.s__Streptococcus_mitis_oralis_pneumoniae
193.s__Streptococcus_mutans
194.s__Streptococcus_parasanguinis
195.s__Streptococcus_parauberis
196.s__Streptococcus_pasteurianus
197.s__Streptococcus_peroris
198.s__Streptococcus_pseudopneumoniae
199.s__Streptococcus_salivarius
200.s__Streptococcus_sanguinis

201.s__Streptococcus_sobrinus
202.s__Streptococcus_sp_C150
203.s__Streptococcus_sp_GMD4S
204.s__Streptococcus_thermophilus
205.s__Streptococcus_tigurinus
206.s__Streptococcus_vestibularis
207.s__Butyricicoccus_pullicaecorum
208.s__Clostridiaceae_bacterium_JC118
209.s__Clostridium_asparagiforme
210.s__Clostridium_bolteae
211.s__Clostridium_butyricum
212.s__Clostridium_celatum
213.s__Clostridium_citroniae
214.s__Clostridium_clostridioforme
215.s__Clostridium_hathewayi
216.s__Clostridium_hylemonae
217.s__Clostridium_leptum
218.s__Clostridium_methylpentosum
219.s__Clostridium_nexile
220.s__Clostridium_perfringens
221.s__Clostridium_phytofermentans
222.s__Clostridium_scindens
223.s__Clostridium_sp_7_2_43FAA
224.s__Clostridium_sp_ATCC_BAA_442
225.s__Clostridium_sp_KLE_1755
226.s__Clostridium_sp_L2_50
227.s__Clostridium_sp_MSTE9
228.s__Clostridium_sp_SS2_1
229.s__Clostridium_sporogenes
230.s__Clostridium_symbiosum
231.s__Clostridium_tyrobutyricum
232.s__Anaerococcus_hydrogenalis
233.s__Anaerococcus_obesiensis
234.s__Anaerococcus_prevotii
235.s__Anaerococcus_tetradus
236.s__Anaerococcus_vaginalis
237.s__Finegoldia_magna
238.s__Parvimonas_unclassified
239.s__Peptoniphilus_harei
240.s__Peptoniphilus_lacrimalis
241.s__Clostridiales_bacterium_1_7_47FAA
242.s__Flavonifractor_plautii

243.s_Pseudoflavonifractor_capillosus
244.s_Anaerofustis_stercorihominis
245.s_Eubacterium_brachy
246.s_Eubacterium_eligens
247.s_Eubacterium_hallii
248.s_Eubacterium_limosum
249.s_Eubacterium_ramulus
250.s_Eubacterium_rectale
251.s_Eubacterium_siraeum
252.s_Eubacterium_sp_3_1_31
253.s_Eubacterium_ventriosum
254.s_Pseudoramibacter_alactolyticus
255.s_Anaerostipes_caccae
256.s_Anaerostipes_hadrus
257.s_Anaerostipes_sp_3_2_56FAA
258.s_Anaerostipes_unclassified
259.s_Blautia_hansenii
260.s_Blautia_hydrogenotrophica
261.s_Blautia_producta
262.s_Ruminococcus_gnavus
263.s_Ruminococcus_obeum
264.s_Ruminococcus_torques
265.s_Butyrvibrio_crossotus
266.s_Butyrvibrio_unclassified
267.s_Cellulosilyticum_lentocellum
268.s_Coprococcus_catus
269.s_Coprococcus_comes
270.s_Coprococcus_eutactus
271.s_Coprococcus_sp_ART55_1
272.s_Dorea_formicigenerans
273.s_Dorea_longicatena
274.s_Dorea_unclassified
275.s_Lachnospiraceae_bacterium_1_1_57FAA
276.s_Lachnospiraceae_bacterium_1_4_56FAA
277.s_Lachnospiraceae_bacterium_2_1_58FAA
278.s_Lachnospiraceae_bacterium_3_1_46FAA
279.s_Lachnospiraceae_bacterium_3_1_57FAA_CT1
280.s_Lachnospiraceae_bacterium_4_1_37FAA
281.s_Lachnospiraceae_bacterium_5_1_57FAA
282.s_Lachnospiraceae_bacterium_5_1_63FAA
283.s_Lachnospiraceae_bacterium_6_1_63FAA
284.s_Lachnospiraceae_bacterium_7_1_58FAA

285.s_Lachnospiraceae_bacterium_8_1_57FAA
286.s_Lachnospiraceae_bacterium_9_1_43BFAA
287.s_Lachnospiraceae_bacterium_ICM7
288.s_Lachnospiraceae_bacterium_oral_taxon_082
289.s_Roseburia_hominis
290.s_Roseburia_intestinalis
291.s_Roseburia_inulinivorans
292.s_Roseburia_unclassified
293.s_Oscillibacter_sp_KLE_1728
294.s_Oscillibacter_sp_KLE_1745
295.s_Oscillibacter_unclassified
296.s_Clostridium_bartlettii
297.s_Clostridium_difficile
298.s_Clostridium_hiranonis
299.s_Clostridium_sordellii
300.s_Peptostreptococcaceae_noname_unclassified
301.s_Peptostreptococcus_anaerobius
302.s_Peptostreptococcus_stomatis
303.s_Peptostreptococcus_unclassified
304.s_Anaerotruncus_colihominis
305.s_Anaerotruncus_unclassified
306.s_Faecalibacterium_prausnitzii
307.s_Ruminococcaceae_bacterium_D16
308.s_Ruminococcus_albus
309.s_Ruminococcus_bromii
310.s_Ruminococcus_callidus
311.s_Ruminococcus_champanellensis
312.s_Ruminococcus_flavefaciens
313.s_Ruminococcus_lactaris
314.s_Ruminococcus_sp_5_1_39BFAA
315.s_Subdoligranulum_sp_4_3_54A2FAA
316.s_Subdoligranulum_unclassified
317.s_Subdoligranulum_variabile
318.s_Catenibacterium_mitsuokai
319.s_Coprobacillus_sp_29_1
320.s_Coprobacillus_sp_D6
321.s_Coprobacillus_unclassified
322.s_Clostridium_innocuum
323.s_Clostridium_amosum
324.s_Clostridium_spiroforme
325.s_Erysipelotrichaceae_bacterium_21_3
326.s_Erysipelotrichaceae_bacterium_2_2_44A

		<p>327.s__Erysipelotrichaceae_bacterium_3_1_53 328.s__Erysipelotrichaceae_bacterium_5_2_54FAA 329.s__Erysipelotrichaceae_bacterium_6_1_45 330.s__Eubacterium_biforme 331.s__Eubacterium_cylindroides 332.s__Eubacterium_dolichum 333.s__Holdemania_filiformis 334.s__Holdemania_sp_AP2 335.s__Holdemania_unclassified 336.s__Solobacterium_moorei 337.s__Turcibacter_sanguinis 338.s__Turcibacter_unclassified 339.s__Acidaminococcus_fermentans 340.s__Acidaminococcus_intestini 341.s__Acidaminococcus_sp_HPA0509 342.s__Acidaminococcus_unclassified 343.s__Phascolarctobacterium_succinatutens 344.s__Anaeroglobus_geminatus 345.s__Dialister_invisus 346.s__Dialister_micraerophilus 347.s__Dialister_succinatiphilus 348.s__Megamonas_funiformis 349.s__Megamonas_hypermegale 350.s__Megamonas_rupellensis 351.s__Megamonas_unclassified 352.s__Megasphaera_elsdenii 353.s__Megasphaera_micronuciformis 354.s__Megasphaera_unclassified 355.s__Mitsuokella_multacida 356.s__Mitsuokella_unclassified 357.s__Selenomonas_artemidis 358.s__Selenomonas_noxia 359.s__Veillonella_atypica 360.s__Veillonella_dispar 361.s__Veillonella_parvula 362.s__Veillonella_ratti 363.s__Veillonella_unclassified</p>
Cluster 3	p_Bacterioidetes	<p>364. s__Bacteroides_barnesiae 365.s__Bacteroides_caccae 366.s__Bacteroides_cellulosilyticus 367.s__Bacteroides_clarus</p>

368.s__Bacteroides_coprocola
369.s__Bacteroides_coprophilus
370.s__Bacteroides_dorei
371.s__Bacteroides_eggerthii
372.s__Bacteroides_faecis
373.s__Bacteroides_finegoldii
374.s__Bacteroides_fragilis
375.s__Bacteroides_intestinalis
376.s__Bacteroides_massiliensis
377.s__Bacteroides_nordii
378.s__Bacteroides_oleiciplenus
379.s__Bacteroides_ovatus
380.s__Bacteroides_plebeius
381.s__Bacteroides_salanitronis
382.s__Bacteroides_salyersiae
383.s__Bacteroides_sp_1_1_6
384.s__Bacteroides_sp_2_1_22
385.s__Bacteroides_sp_2_1_33B
386.s__Bacteroides_sp_2_1_56FAA
387.s__Bacteroides_sp_3_1_19
388.s__Bacteroides_sp_3_2_5
389.s__Bacteroides_sp_4_3_47FAA
390.s__Bacteroides_stercoris
391.s__Bacteroides_thetaiotaomicron
392.s__Bacteroides_uniformis
393.s__Bacteroides_vulgatus
394.s__Bacteroides_xylanisolvans
395.s__Bacteroidales_bacterium_ph8
396.s__Barnesiella_intestinihominis
397.s__Butyricimonas_synergistica
398.s__Coprobacter_fastidiosus
399.s__Dysgonomonas_gadei
400.s__Dysgonomonas_mossii
401.s__Dysgonomonas_unclassified
402.s__Odoribacter_splanchnicus
403.s__Odoribacter_unclassified
404.s__Parabacteroides_distasonis
405.s__Parabacteroides_goldsteinii
406.s__Parabacteroides_johnsonii
407.s__Parabacteroides_merdae
408.s__Parabacteroides_sp_D13
409.s__Parabacteroides_unclassified

		410.s__Porphyromonas_asaccharolytica 411.s__Porphyromonas_bennonis 412.s__Porphyromonas_endodontalis 413.s__Porphyromonas_somerae 414.s__Porphyromonas_uenonis 415.s__Tannerella_forsythia 416.s__Alloprevotella_tannerae 417.s__Alloprevotella_unclassified 418.s__Paraprevotella_clara 419.s__Paraprevotella_unclassified 420.s__Paraprevotella_xylaniphila 421.s__Prevotella_amnii 422.s__Prevotella_bergensis 423.s__Prevotella_bivia 424.s__Prevotella_buccae 425.s__Prevotella_buccalis 426.s__Prevotella_copri 427.s__Prevotella_disiens 428.s__Prevotella_histicola 429.s__Prevotella_loescheii 430.s__Prevotella_melaninogenica 431.s__Prevotella_micans 432.s__Prevotella_nanceiensis 433.s__Prevotella_nigrescens 434.s__Prevotella_oralis 435.s__Prevotella_oris 436.s__Prevotella_oulorum 437.s__Prevotella_stercorea 438.s__Prevotella_timonensis 439.s__Alistipes_finegoldii 440.s__Alistipes_indistinctus 441.s__Alistipes_onderdonkii 442.s__Alistipes_putredinis 443.s__Alistipes_senegalensis 444.s__Alistipes_shahii 445.s__Alistipes_sp_AP11 446.s__Alistipes_sp_HGB5 447.s__Alistipes_unclassified 448.s__Capnocytophaga_unclassified 449.s__Cellulophaga_unclassified 450.s__Pedobacter_unclassified
Cluster 4	1. p_Acidobacteria	452. s__Granulicella_unclassified

2. p_Chlorobi	453. s__Chlorobium_phaeobacteroides
3. p_Deinococcus_Thermus	454. s__Deinococcus_unclassified
4. p_Synergistetes	455. s__Pyramidobacter_piscolens
5. p_Tenericutes	456. s__Candidatus_Phytoplasma_unclassified
6. p_Verrucomicrobia	457. s__Akkermansia_muciniphila
7. p_Fusobacteria	458. s__Cetobacterium_somerae
	459. s__Fusobacterium_gonidiaformans
	460. s__Fusobacterium_mortiferum
	461. s__Fusobacterium_nucleatum
	462. s__Fusobacterium_periodonticum
	463. s__Fusobacterium_ulcerans
	464. s__Fusobacterium_varium
	465. s__Leptotrichia_unclassified
	466. s__Leptotrichia_wadei
	467. s__Pyramidobacter_piscolens
8.	468. s__candidate_division_TM7_single_cell_isolate_TM7b
p_Candidatus_Saccharibacteria	469. s__candidate_division_TM7_single_cell_isolate_TM7c
9. p_Actinobacteria	470. s__Actinobaculum_massiliense
	471.s__Actinobaculum_schaalii
	472.s__Actinobaculum_unclassified
	473.s__Actinomyces_europaeus
	474.s__Actinomyces_graevenitzii
	475.s__Actinomyces_johnsonii
	476.s__Actinomyces_naeslundii
	477.s__Actinomyces_neuii
	478.s__Actinomyces_odontolyticus
	479.s__Actinomyces_sp_HPA0247
	480.s__Actinomyces_turicensis
	481.s__Actinomyces_urogenitalis
	482.s__Actinomyces_viscosus
	483.s__Varibaculum_cambriense
	484.s__Brevibacterium_unclassified
	485.s__Corynebacterium_accolens
	486.s__Corynebacterium_durum
	487.s__Corynebacterium_kroppenstedtii
	488.s__Corynebacterium_propinquum
	489.s__Corynebacterium_pseudodiphtheriticum
	490.s__Gordonia_otitidis
	491.s__Rothia_mucilaginosa
	492.s__Rothia_unclassified
	493.s__Propionibacterium_acnes

494.s__Propionibacterium_avidum
495.s__Propionibacterium_freudenreichii
496.s__Propionibacterium_granulosum
497.s__Propionimicrobium_lymphophilum
498.s__Alloscardovia_omnicolens
499.s__Bifidobacterium_adolescentis
500.s__Bifidobacterium_angulatum
501.s__Bifidobacterium_animalis
502.s__Bifidobacterium_bifidum
503.s__Bifidobacterium_breve
504.s__Bifidobacterium_catenulatum
505.s__Bifidobacterium_dentium
506.s__Bifidobacterium_longum
507.s__Bifidobacterium_pseudocatenulatum
508.s__Bifidobacterium_pseudolongum
509.s__Gardnerella_vaginalis
510.s__Parascardovia_denticolens
511.s__Scardovia_unclassified
512.s__Scardovia_wiggisiae
513.s__Adlercreutzia_equolifaciens
514.s__Atopobium_minutum
515.s__Atopobium_parvulum
516.s__Atopobium_rimae
517.s__Atopobium_vaginae
518.s__Collinsella_aerofaciens
519.s__Collinsella_intestinalis
520.s__Collinsella_stercoris
521.s__Collinsella_tanakaei
522.s__Collinsella_unclassified
523.s__Coriobacteriaceae_bacterium_BV3Ac1
524.s__Coriobacteriaceae_bacterium_phI
525.s__Cryptobacterium_curtum
526.s__Eggerthella_lenta
527.s__Eggerthella_sp_1_3_56FAA
528.s__Eggerthella_unclassified
529.s__Enterorhabdus_caecimuris
530.s__Gordonibacter_pamelaeae
531.s__Olsenella_uli
532.s__Olsenella_unclassified
533.s__Slackia_exigua
534.s__Slackia_piriformis

Supplementary Table 3: Association of environmental variables to various allergy outcomes in the DIABIMMUNE three country cohort.

(a)

Variables		Milk Allergy		p-value
		Cases	Controls	
Mean age in days (sd)		413 (286)	442 (241)	0.32
Sex (N (%))	Male	30 (58.8%)	77 (53.4%)	0.51
	Female	21 (41.8%)	67 (46.5%)	
Delivery Type (N (%))	Caesarean	3 (5.8%)	14 (9.7%)	0.40
	Vaginal	48 (94.1%)	130 (90.2%)	
Ethnicity (N (%))	Estonia	21 (41.1%)	49 (34%)	<0.001
	Finland	27 (52.9%)	44 (30.5%)	
	Russia	3 (5.8%)	51 (34.5%)	
Breast Feeding (N (%))	Yes	9 (17.6%)	93 (64.5%)	0.02
	No	42 (82.3%)	51 (35.4%)	

(b)

Variables		Egg Allergy		p-value
		Cases	Controls	
Mean age in days (sd)		471 (318)	426 (236)	0.75
Sex (N (%))	Male	19 (51.3%)	70 (44.3%)	0.76
	Female	18 (48.6%)	88 (55.7%)	
Delivery Type (N (%))	Caesarean	2 (5.4%)	15 (9.4%)	0.63
	Vaginal	35 (94.5%)	143 (90.5%)	
Ethnicity (N (%))	Estonia	10 (27%)	60 (37.9%)	0.01
	Finland	21 (56.7%)	50 (31.6%)	
	Russia	6 (16.2%)	48 (30.3%)	
Breast Feeding (N (%))	Yes	31 (83.7%)	104 (65.8%)	0.05
	No	6 (16.2%)	54 (34.1%)	

(c)

Variables		Peanut Allergy		p-value
		Cases	Controls	
Mean age in days (sd)		383 (226)	437 (254)	0.41
Sex (N (%))	Male	5 (55.5%)	102 (54.8%)	1
	Female	4 (44.4%)	84 (45.1%)	
Delivery Type (N (%))	Caesarean	0 (0%)	17 (9.1%)	0.34
	Vaginal	9 (100%)	169 (90.8%)	
Ethnicity (N (%))	Estonia	4 (44.4%)	66 (35.4%)	0.153
	Finland	5 (55.5%)	66 (35.4%)	
	Russia	0 (0%)	54 (29%)	
Breast Feeding (N (%))	Yes	5 (55.5%)	130 (69.8%)	0.58
	No	4 (44.4%)	56 (30.1%)	

(d)

Variables		Overall Allergy		p-value
		Cases	Controls	
Mean age in days (sd)		431 (301)	437 (226)	0.46
Sex (N (%))	Male	38 (57.4%)	69 (53.4%)	0.69
	Female	28 (42.4%)	60 (46.5%)	
Delivery Type (N (%))	Caesarean	3 (4.5%)	14 (10.8%)	0.22
	Vaginal	63 (95.4%)	115 (89.1%)	
Ethnicity (N (%))	Estonia	24 (36.3%)	46 (35.6%)	<0.001
	Finland	34 (51.5%)	37 (28.6%)	
	Russia	8 (12.1%)	46 (35.6%)	
Breast Feeding (N (%))	Yes	54 (81.8%)	81 (62.7%)	0.01
	No	12 (18.1%)	48 (37.2%)	

Supplementary Table 4: Distribution of samples into cases and controls, in each allergy outcome in the DIABIMMUNE three country cohort.

Time point	Egg Allergy		Milk Allergy		Peanut Allergy		Combined Allergy	
	Cases	Controls	Cases	Controls	Cases	Controls	Cases	Controls
T1 (6M)	28	79	41	66	6	101	51	56
T2 (12M)	31	154	33	151	7	178	50	135
T3 (18M)	35	162	46	151	12	185	61	136
T4 (24M)	31	120	40	111	5	146	51	100
T5 (30M)	13	37	15	35	3	48	20	30
T6 (36M)	10	32	13	29	0	43	18	24

Supplementary Table 5: AUC values tabulated for neural network models on test set of DIABIM-MUNE three country cohort. The results are reported on considering model performance without (w/o) weight balancing in the neural network and with weight balancing in the neural network. Note the consistent increase in AUC as shown in column 3, 5, 7 and 9, using weight balancing.

Methods	Overall allergy		Milk Allergy		Egg Allergy		Peanut Allergy	
	AUC w/o weight balancing	AUC Test with weight balancing	AUC w/o weight balancing	AUC Test with weight balancing	AUC w/o weight balancing	AUC Test with weight balancing	AUC w/o weight balancing	AUC Test with weight balancing
CNN-LSTM	0.744	0.762	0.735	0.741	0.662	0.673	0.559	0.585
LSTM	0.717	0.734	0.709	0.715	0.649	0.653	0.557	0.569
taxoNN_corr	0.695	0.711	0.687	0.691	0.635	0.641	0.553	0.561

Supplementary Table 6: Association of environmental variables to preterm birth outcome in the DiGiulio case-control study.

Variables		Preterm birth		p-value
		Cases (N=11)	Controls (N=29)	
Gestational Day Of Collection (Median [IQR])		88 [75, 127.5]	90 [78, 159]	0.59
Race (N (%))	American Indian	2 (18.2%)	1 (3.4%)	0.25
	Asian	1 (9.1%)	2 (6.9%)	
	Black	1 (9.1%)	1 (3.4%)	
	Caucasian	4 (36.4%)	18 (62.1%)	
	Other	1 (9.1%)	5 (17.2%)	
	Pacific Islander	2 (18.2%)	1 (3.4%)	
Ethnicity (N (%))	Hispanic	2 (18.2%)	11 (37.9%)	0.25
	Non-Hispanic	9 (81.8%)	18 (62.1%)	

Supplementary Table 7: AUC as obtained by various feature extraction techniques combined with LSTM module to identify temporal dependency in DIABIMMUNE three country cohort with overall allergy as outcome. Alongwith conventional ML models, Multi dimensional feature selection [2] is also used for comparison against taxoNN for feature extraction.

Methods	AUC
LSTM+taxoNN	0.762
LSTM+MDFS [2]	0.705
LSTM+ Random Forest	0.702
LSTM+SVM	0.695
LSTM+Lasso	0.682
LSTM+CNN_basic	0.591

Supplementary Table 8: Comparison of AUC values obtained on the test set of OTUs from metagenomic dataset and 16s dataset of the DIABIMMUNE three country cohort on the 'Overall Allergy' outcome.

Methods	AUC in Metagenomic dataset	AUC in 16s dataset
phyLoSTM	0.762	0.755
LSTM	0.734	0.729
taxoNN_corr	0.711	0.713
Random Forest	0.639	0.640
Support Vector Machines	0.611	0.615
Ridge Regression	0.607	0.611
Lasso Regression	0.603	0.600
Gaussian Bayes Classifier	0.601	0.599
Naïve Bayes Classifier	0.609	0.603
CNN_basic	0.569	0.552

Supplementary Table 9: AUC values tabulated for various machine learning methods upon class based stratification for DIABIMMUNE three country and DiGiulio preterm studies.

Methods	AUC in DIABIMMUNE three country cohort	AUC in DiGiulio Preterm study
phyLoSTM	0.667	0.673
LSTM	0.659	0.671
taxoNN_corr	0.634	0.662
Random Forest	0.639	0.659
Support Vector Machines	0.611	0.641
Ridge Regression	0.607	0.634
Lasso Regression	0.603	0.630
Gaussian Bayes Classifier	0.601	0.631
Naïve Bayes Classifier	0.609	0.624
CNN_basic	0.569	0.604

Supplementary Table 10: Comparison of AUCs as obtained by phyLoSTM methodology incorporating temporal data across 6 time points in the DIABIMMUNE cohort for the overall allergy outcome with conventional ML models taking data from individual time points as inputs.

Methods	AUC across time points					
	Timepoint1	Timepoint2	Timepoint3	Timepoint4	Timepoint5	Timepoint6
phyLoSTM	0.762					
LSTM	0.734					
taxoNN_corr	0.711	0.707	0.700	0.691	0.684	0.680
Random Forest	0.639	0.641	0.631	0.629	0.619	0.611
Support Vector Machines	0.611	0.619	0.610	0.607	0.597	0.589
Ridge Regression	0.607	0.600	0.599	0.591	0.582	0.575
Lasso Regression	0.603	0.601	0.597	0.593	0.583	0.571
Gaussian Bayes Classifier	0.601	0.600	0.599	0.591	0.585	0.572
Naïve Bayes Classifier	0.609	0.611	0.607	0.598	0.583	0.577
CNN_basic	0.569	0.561	0.565	0.560	0.551	0.550

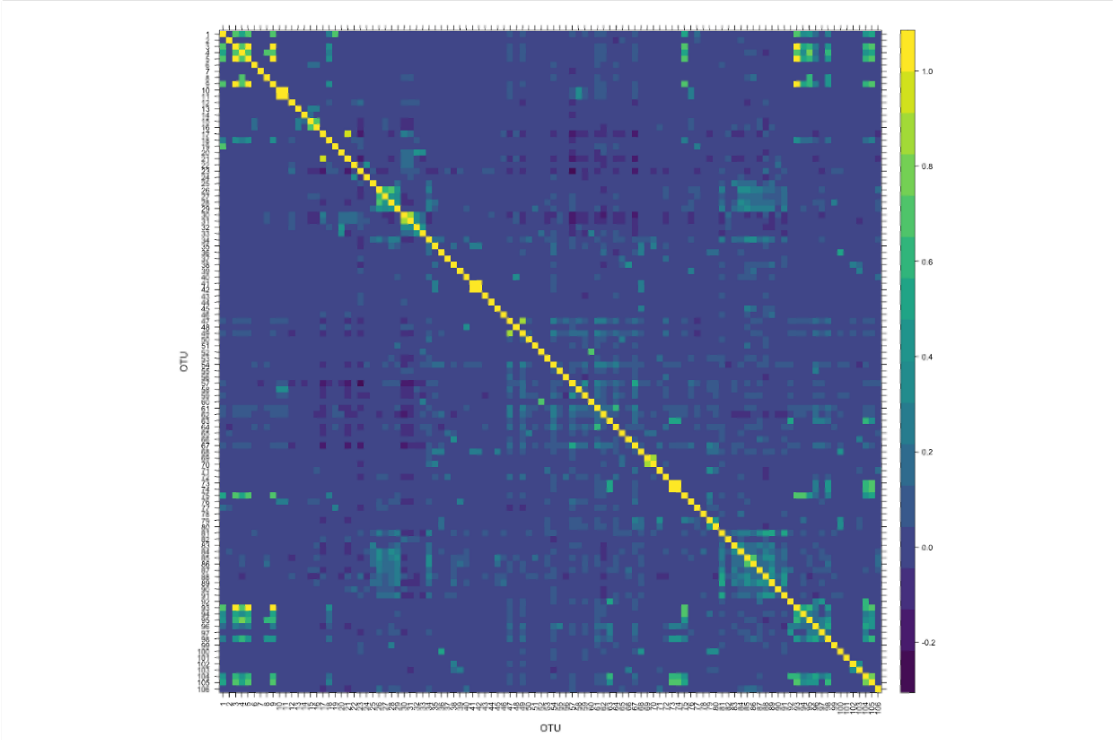
Supplementary Table 11: Comparison of AUCs as obtained by phyLoSTM methodology incorporating temporal data across 6 time points in the DiGiulio preterm study with conventional ML models taking data from individual time points as inputs.

	AUC across time points			
Methods	Timepoint1	Timepoint2	Timepoint3	Timepoint4
phyLoSTM	0.713			
LSTM	0.692			
taxoNN_corr	0.678	0.672	0.668	0.658
Random Forest	0.659	0.650	0.644	0.640
Support Vector Machines	0.641	0.633	0.628	0.620
Ridge Regression	0.634	0.631	0.625	0.620
Lasso Regression	0.630	0.625	0.622	0.616
Gaussian Bayes Classifier	0.631	0.622	0.619	0.615
Naïve Bayes Classifier	0.624	0.613	0.610	0.606
CNN_basic	0.604	0.597	0.589	0.580

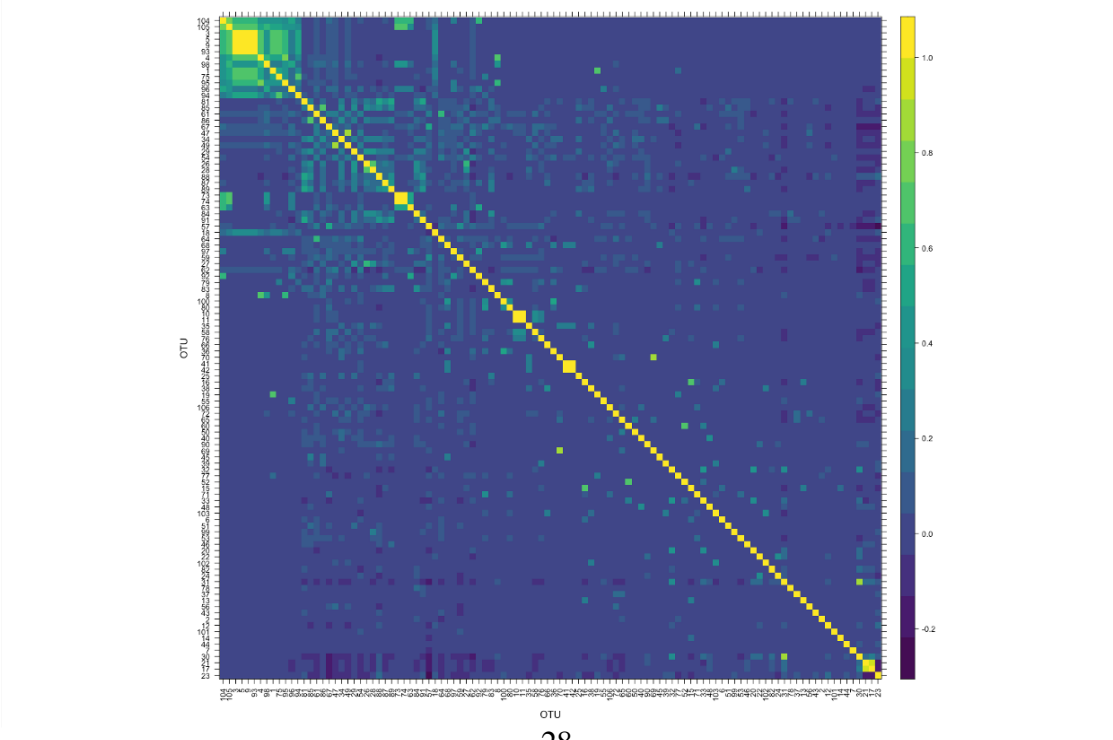
2 Supplementary Figures

Supplementary Figure 1: Heatmaps for the Spearman rank of the OTUs in the cluster, Phylum Proteobacteria, (a) before ordering and (b) after the ordering based on correlation of the OTUs in the DIABIMMUNE three country cohort

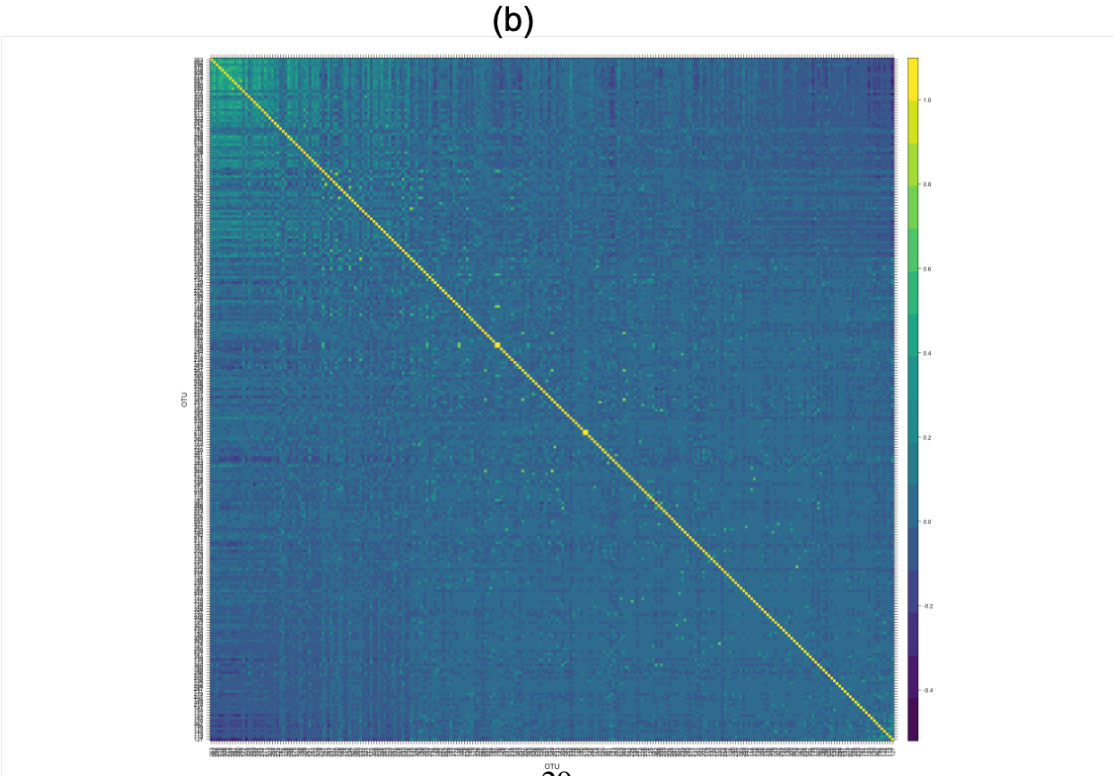
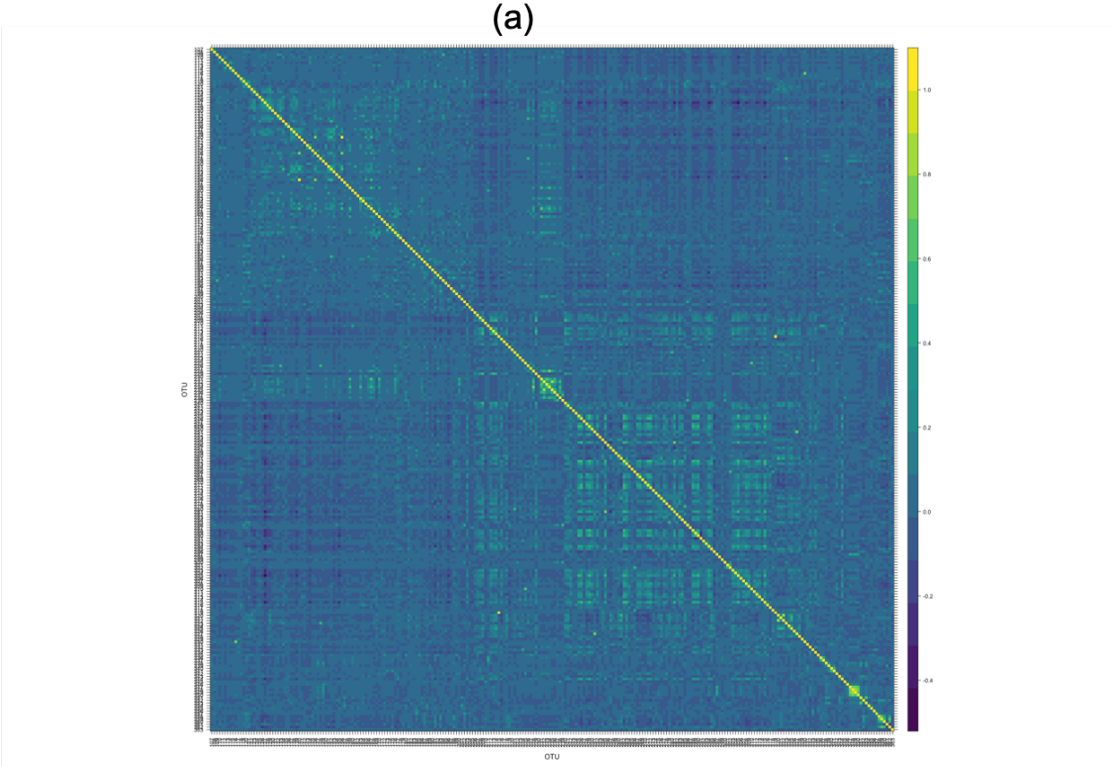
(a)



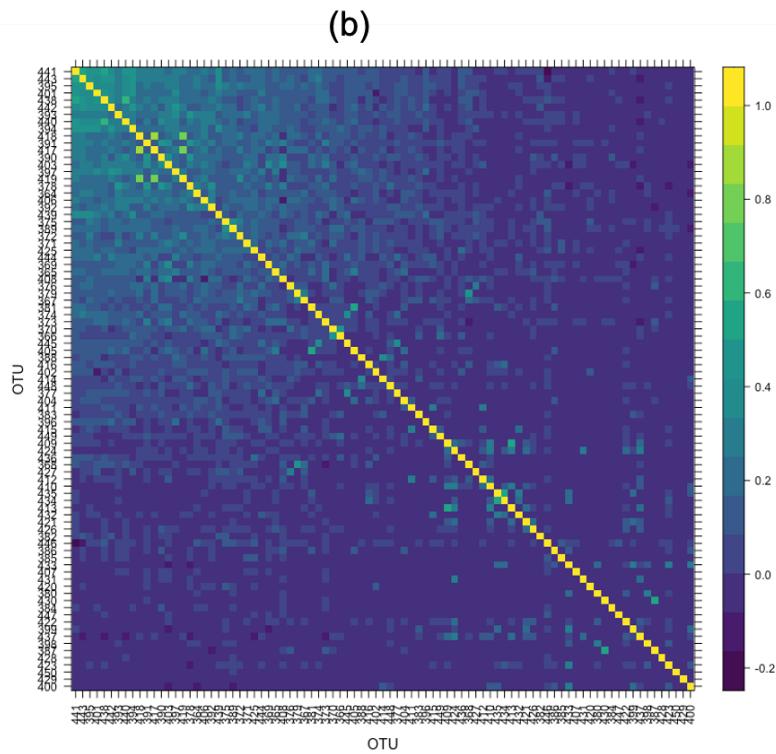
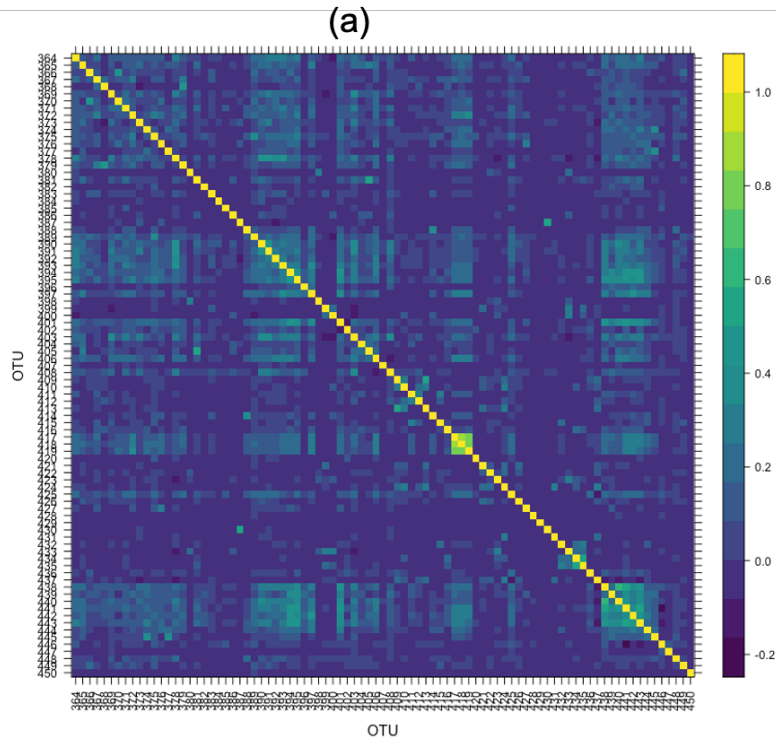
(b)



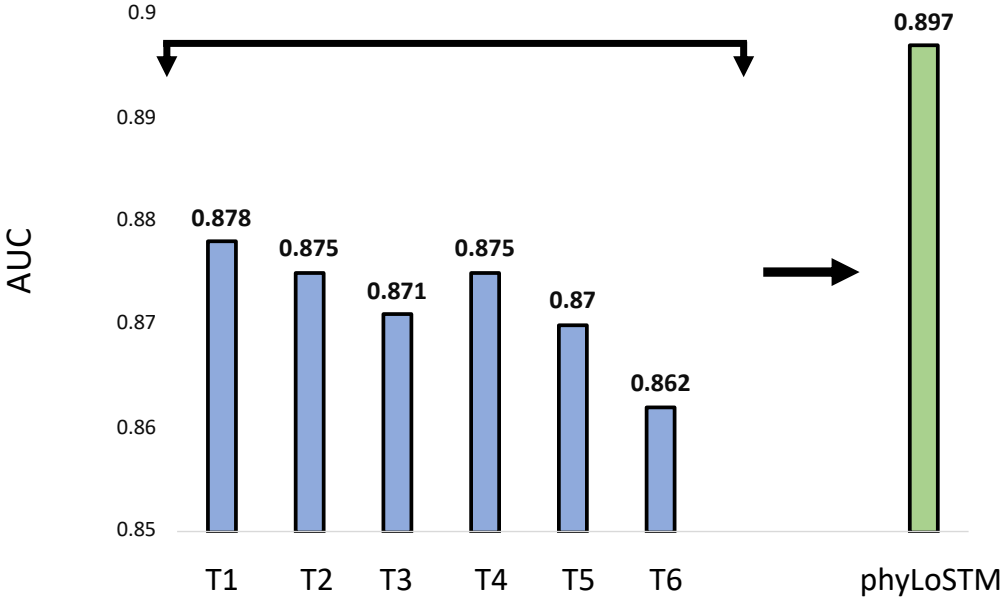
Supplementary Figure 2: Heatmaps for the Spearman rank of the OTUs in the cluster, Phylum Firmicutes, (a) before ordering and (b) after the ordering based on correlation of the OTUs in the DIABIMMUNE three country cohort



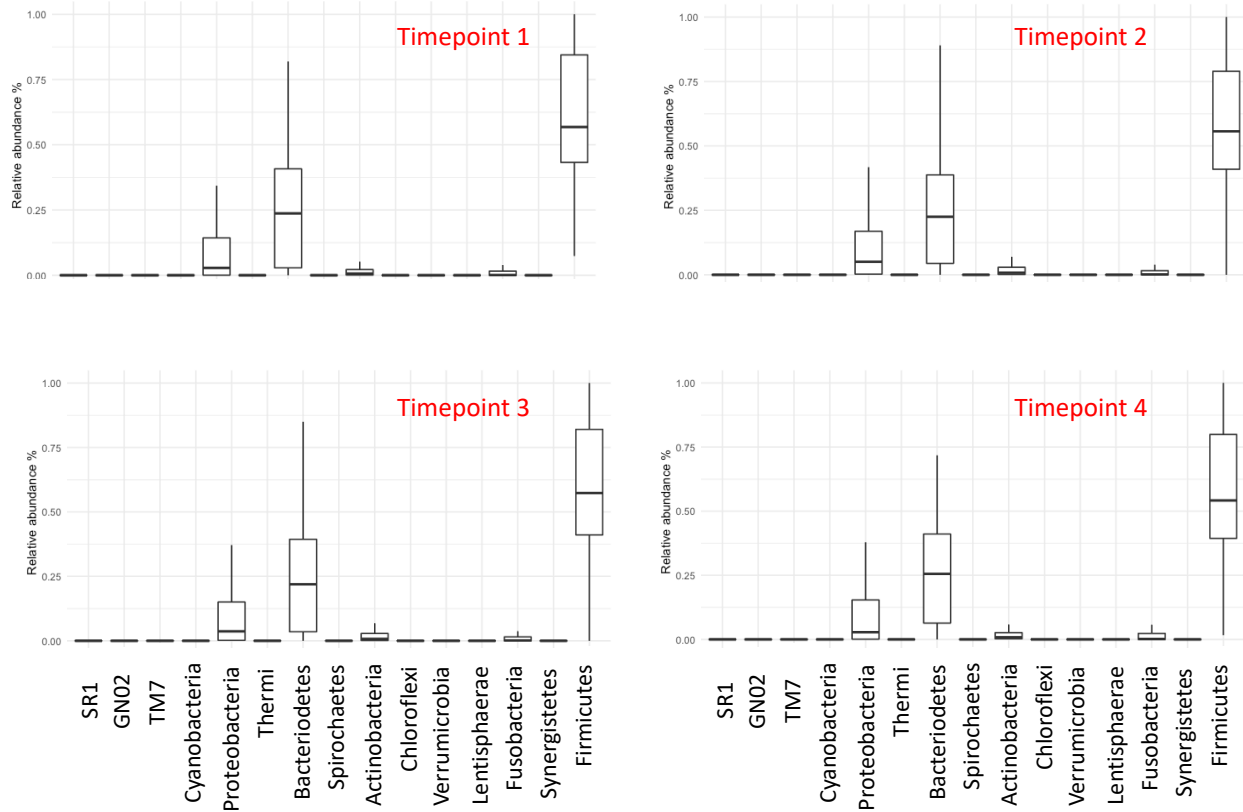
Supplementary Figure 3: Heatmaps for the Spearman rank of the OTUs in the cluster, Phylum Bacteroidetes, (a) before ordering and (b) after the ordering based on correlation of the OTUs in the DIABIMMUNE three country cohort



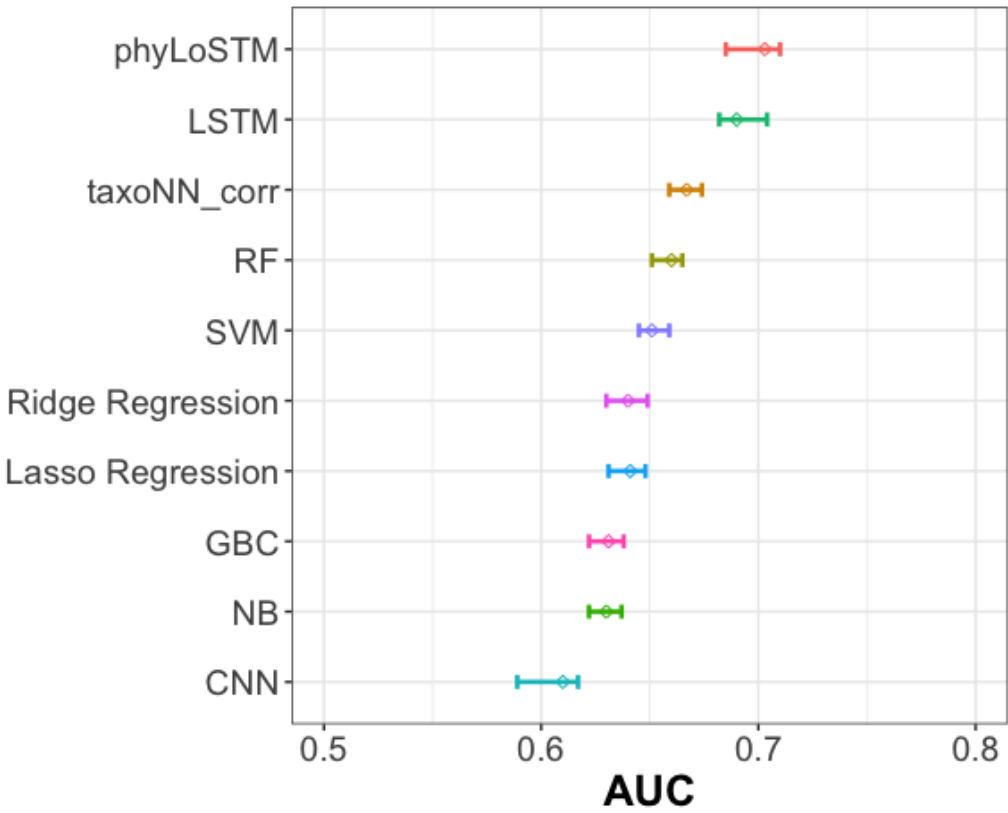
Supplementary Figure 4: Comparison of AUC values as observed at each individual 6 time points using *taxoNN_{corr}* approach with the final higher AUC value of 0.897 using proposed phyLoSTM framework on the simulated study.

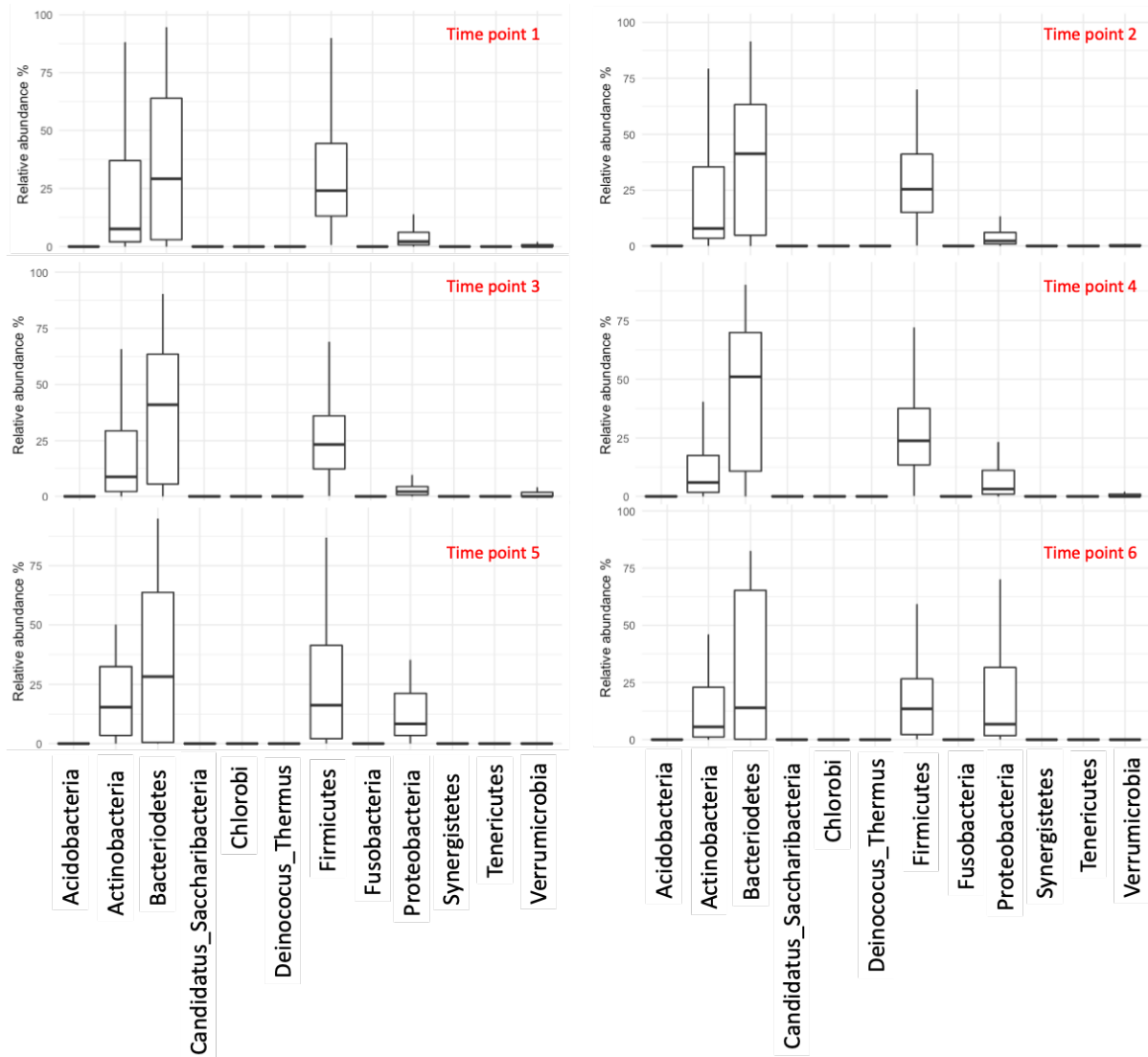


Supplementary Figure 5: Relative abundance of OTUs in each phylum of the DiGiulio study at four time points measured every three months. For each box plot the upper whisker extends from the hinge to the largest value no further than $1.5 * IQR$ from the hinge (where IQR is the inter-quartile range, or distance between the first and third quartiles). The lower whisker extends from the hinge to the smallest value at most $1.5 * IQR$ of the hinge. Data beyond the end of the whiskers are called "outlying" points and are plotted individually.



Supplementary Figure 6: 95% confidence intervals obtained for the mean AUC values for 10 times 10-fold cross validation on the training set for the DiGiulio study on the preterm delivery output [3]





Supplementary Figure 7: Relative abundance of OTUs in each phylum of the DIABIMMUNE three country cohort at six time points measured half-yearly up to age of three of the infant. For each box plot the upper whisker extends from the hinge to the largest value no further than $1.5 * IQR$ from the hinge (where IQR is the inter-quartile range, or distance between the first and third quartiles). The lower whisker extends from the hinge to the smallest value at most $1.5 * IQR$ of the hinge.

3 References

- [1] Vatanen, T. *et al.* Variation in microbiome lps immunogenicity contributes to autoimmunity in humans. *Cell* **165**, 842–853 (2016).
- [2] Mnich, K. & Rudnicki, W. R. All-relevant feature selection using multidimensional filters with exhaustive search. *Information Sciences* **524**, 277–297 (2020).
- [3] DiGiulio, D. B. *et al.* Temporal and spatial variation of the human microbiota during pregnancy. *Proceedings of the National Academy of Sciences* **112**, 11060–11065 (2015).