

Supplementary information

for

Gut microbiome composition is predictive of incident type 2 diabetes in a population cohort of 5 572 Finnish adults

Matti O. Ruuskanen, PhD^{1*}, Pande P. Erawijantari, DSc^{1*}, Aki S. Havulinna, DSc (tech.)^{2,3},
Yang Liu, MSc^{4,5}, Guillaume Méric, PhD^{4,6}, Jaakko Tuomilehto, MD, PhD^{2,7,8}, Michael Inouye,
PhD^{4,9}, Pekka Jousilahti, MD, PhD², Veikko Salomaa, MD, PhD², Mohit Jain, PhD^{10,11}, Rob
Knight, PhD^{12,13,14,15}, Leo Lahti, PhD^{1†}, Teemu J. Niiranen, MD, PhD^{2,16†}

*These authors contributed equally to this work

†These authors contributed equally to this work

¹Department of Computing, University of Turku, Turku, Finland

²Department of Public Health and Welfare, Finnish Institute for Health and Welfare, Helsinki, Finland

³Institute for Molecular Medicine Finland, FIMM - HiLIFE, Helsinki, Finland

⁴Cambridge Baker Systems Genomics Initiative, Baker Heart and Diabetes Institute, Melbourne, Victoria, Australia

⁵Department of Clinical Pathology, Melbourne Medical School, The University of Melbourne, Melbourne, Victoria,
Australia

⁶Department of Infectious Diseases, Central Clinical School, Monash University, Melbourne, Victoria, Australia

⁷Department of Public Health, University of Helsinki, Helsinki, Finland

⁸Saudi Diabetes Research Group, King Abdulaziz University, Jeddah, Saudi Arabia

⁹Department of Public Health and Primary Care, Cambridge University, Cambridge, United Kingdom

¹⁰Department of Medicine, University of California, San Diego, La Jolla, California, USA

¹¹Department of Pharmacology, University of California, San Diego, La Jolla, California, USA

¹²Jacobs School of Engineering, University of California, San Diego, La Jolla, California, USA

¹³Center for Microbiome Innovation, University of California San Diego, La Jolla, California, USA

¹⁴Department of Pediatrics, School of Medicine, University of California San Diego, La Jolla, California, USA

¹⁵Department of Computer Science & Engineering, University of California San Diego, La Jolla, California, USA

¹⁶Division of Medicine, Turku University Hospital, Turku, Finland

Table S1. Results of multivariable adjusted Cox regression models for incident type 2 diabetes.

Predictor	Eastern Finland				Western Finland		
	Coefficient	Hazard ratio	P-value	Adjusted p-value	Coefficient	Hazard ratio	P-value
PC1	-0.246	0.782 (95% CI, 0.693-0.883)	1×10^{-4}	0.0099	-0.067	0.935 (95% CI, 0.785-1.114)	0.4542
Cluster 1	0.182	1.2 (95% CI, 1.082-1.33)	5×10^{-4}	-	0.168	1.183 (95% CI, 1.018-1.375)	0.0284
Cluster 2	0.123	1.13 (95% CI, 1.008-1.267)	0.0358	-	0.062	1.064 (95% CI, 0.898-1.261)	0.4742
Cluster 3	0.211	1.235 (95% CI, 1.106-1.381)	2×10^{-4}	-	0.08	1.083 (95% CI, 0.917-1.278)	0.3465
Cluster 4	0.217	1.242 (95% CI, 1.115-1.384)	1×10^{-4}	-	0.099	1.104 (95% CI, 0.939-1.299)	0.2306
Cluster 5	0.19	1.209 (95% CI, 1.092-1.339)	3×10^{-4}	-	0.163	1.177 (95% CI, 1.016-1.364)	0.0297
<i>Blautia</i> sp. KLE 1732	0.198	1.219 (95% CI, 1.09-1.364)	5×10^{-4}	0.0153	0.123	1.13 (95% CI, 0.958-1.334)	0.1469
<i>Clostridium phoceensis</i>	0.197	1.218 (95% CI, 1.09-1.361)	5×10^{-4}	0.0153	0.014	1.014 (95% CI, 0.858-1.198)	0.8702
<i>Coprococcus comes</i>	0.194	1.214 (95% CI, 1.084-1.36)	8×10^{-4}	0.0153	-0.044	0.957 (95% CI, 0.807-1.135)	0.6148
<i>[Clostridium] citroniae</i>	0.192	1.212 (95% CI, 1.086-1.352)	6×10^{-4}	0.0153	0.192	1.212 (95% CI, 1.035-1.42)	0.0172
<i>Eggerthella lenta</i>	0.187	1.205 (95% CI, 1.08-1.344)	8×10^{-4}	0.0153	0.071	1.074 (95% CI, 0.909-1.269)	0.4022
<i>Blautia obeum</i>	0.18	1.197 (95% CI, 1.071-1.338)	0.0015	0.0221	0.112	1.118 (95% CI, 0.952-1.313)	0.1747
<i>[Clostridium] glycyrrhizinilyticum</i>	0.176	1.193 (95% CI, 1.08-1.318)	5×10^{-4}	0.0153	0.103	1.108 (95% CI, 0.942-1.304)	0.2163
<i>Anaerostipes hadrus</i>	0.171	1.186 (95% CI, 1.061-1.327)	0.0028	0.0279	-0.065	0.937 (95% CI, 0.795-1.105)	0.4416
<i>[Clostridium] symbiosum</i>	0.169	1.184 (95% CI, 1.061-1.321)	0.0025	0.0277	0.11	1.117 (95% CI, 0.949-1.314)	0.1832
<i>Bacteroides vulgatus</i>	0.166	1.18 (95% CI, 1.047-1.331)	0.0067	0.0461	0.121	1.128 (95% CI, 0.945-1.347)	0.1821
<i>Oscillibacter</i> sp. KLE 1745	0.165	1.179 (95% CI, 1.054-1.319)	0.0039	0.0325	0.091	1.096 (95% CI, 0.926-1.296)	0.2875
<i>[Clostridium] boltea</i>	0.164	1.178 (95% CI, 1.068-1.3)	0.0011	0.0172	0.185	1.204 (95% CI, 1.043-1.39)	0.0115
<i>[Ruminococcus] gnavus</i>	0.162	1.176 (95% CI, 1.06-1.304)	0.0023	0.0277	0.155	1.168 (95% CI, 1.006-1.355)	0.0414
<i>Eubacterium ramulus</i>	0.157	1.17 (95% CI, 1.048-1.306)	0.0053	0.0389	0.042	1.043 (95% CI, 0.885-1.229)	0.6161
<i>Tyzzerella nexilis</i>	0.15	1.162 (95% CI, 1.046-1.292)	0.0052	0.0389	0.158	1.171 (95% CI, 1.011-1.355)	0.0347
<i>Alistipes putredinis</i>	-0.159	0.853 (95% CI, 0.767-0.949)	0.0035	0.0325	-0.054	0.948 (95% CI, 0.805-1.116)	0.5192
<i>Sutterella wadsworthensis</i>	-0.17	0.844 (95% CI, 0.752-0.947)	0.004	0.0325	-0.023	0.977 (95% CI, 0.823-1.16)	0.7939
<i>Alistipes indistinctus</i>	-0.177	0.838 (95% CI, 0.747-0.94)	0.0025	0.0277	0.023	1.023 (95% CI, 0.862-1.213)	0.7961

- Significant p-values (< 0.05) for models using data from Western Finland are indicated in bold font.

Table S2. Results of multivariable adjusted Cox regression models for incident type 2 diabetes after excluding participants diagnosed with type 2 diabetes in the first two years of the follow-up period.

Predictor	Eastern Finland				Western Finland		
	Coefficient	Hazard ratio	P-value	Adjusted p-value	Coefficient	Hazard ratio	P-value
PC1	0.258	1.295 (95% CI, 1.144-1.465)	0	0.0058	-0.065	0.937 (95% CI, 0.786-1.118)	0.4717
Cluster 1	0.183	1.201 (95% CI, 1.082-1.334)	6×10^{-4}	NA	0.172	1.188 (95% CI, 1.023-1.38)	0.0241
Cluster 2	0.212	1.237 (95% CI, 1.102-1.387)	3×10^{-4}	NA	0.092	1.097 (95% CI, 0.924-1.301)	0.2897
Cluster 3	-0.212	0.809 (95% CI, 0.727-0.9)	1×10^{-4}	NA	-0.126	0.881 (95% CI, 0.753-1.031)	0.1152
Cluster 4	0.249	1.283 (95% CI, 1.145-1.439)	0	NA	0.011	1.011 (95% CI, 0.851-1.202)	0.898
Cluster 5	0.208	1.232 (95% CI, 1.103-1.375)	2×10^{-4}	NA	0.123	1.131 (95% CI, 0.963-1.328)	0.1336
Cluster 6	0.197	1.218 (95% CI, 1.098-1.351)	2×10^{-4}	NA	0.166	1.181 (95% CI, 1.019-1.368)	0.0272
<i>Blautia sp. KLE 1732</i>	0.213	1.238 (95% CI, 1.105-1.386)	2×10^{-4}	0.0128	0.134	1.144 (95% CI, 0.969-1.35)	0.1124
<i>Coprococcus comes</i>	0.211	1.235 (95% CI, 1.101-1.386)	3×10^{-4}	0.0128	-0.03	0.971 (95% CI, 0.817-1.152)	0.7324
<i>Clostridium phoceensis</i>	0.198	1.219 (95% CI, 1.089-1.365)	6×10^{-4}	0.0128	0.008	1.008 (95% CI, 0.853-1.192)	0.9238
<i>[Clostridium] citroniae</i>	0.196	1.217 (95% CI, 1.089-1.36)	5×10^{-4}	0.0128	0.197	1.218 (95% CI, 1.039-1.427)	0.015
<i>Blautia obeum</i>	0.192	1.211 (95% CI, 1.083-1.356)	8×10^{-4}	0.0146	0.124	1.132 (95% CI, 0.963-1.33)	0.1323
<i>[Clostridium] glycyrrhizinilyticum</i>	0.18	1.197 (95% CI, 1.082-1.325)	5×10^{-4}	0.0128	0.112	1.118 (95% CI, 0.951-1.315)	0.1775
<i>Eubacterium ramulus</i>	0.18	1.198 (95% CI, 1.072-1.339)	0.0015	0.0178	0.057	1.059 (95% CI, 0.899-1.248)	0.4919
<i>Eggerthella lenta</i>	0.18	1.197 (95% CI, 1.071-1.338)	0.0015	0.0178	0.078	1.081 (95% CI, 0.915-1.278)	0.3596
<i>Roseburia intestinalis</i>	0.177	1.193 (95% CI, 1.063-1.34)	0.0028	0.0239	-0.035	0.965 (95% CI, 0.813-1.147)	0.6885
<i>Anaerostipes hadrus</i>	0.173	1.189 (95% CI, 1.061-1.332)	0.0029	0.0239	-0.057	0.945 (95% CI, 0.8-1.115)	0.4997
<i>[Clostridium] symbiosum</i>	0.171	1.187 (95% CI, 1.062-1.326)	0.0024	0.0238	0.119	1.127 (95% CI, 0.958-1.325)	0.1489
<i>Bacteroides vulgatus</i>	0.171	1.187 (95% CI, 1.05-1.341)	0.0059	0.0389	0.112	1.118 (95% CI, 0.936-1.335)	0.2172
<i>[Ruminococcus] gnavus</i>	0.166	1.181 (95% CI, 1.063-1.312)	0.002	0.0213	0.161	1.175 (95% CI, 1.013-1.363)	0.0327
<i>Tyzzerella nexilis</i>	0.164	1.178 (95% CI, 1.059-1.31)	0.0025	0.0238	0.163	1.177 (95% CI, 1.018-1.361)	0.0281
<i>[Clostridium] bolteae</i>	0.162	1.176 (95% CI, 1.064-1.299)	0.0014	0.0178	0.187	1.206 (95% CI, 1.044-1.392)	0.0108
<i>Dorea sp. 5-2</i>	0.16	1.173 (95% CI, 1.047-1.315)	0.0061	0.0389	0.043	1.044 (95% CI, 0.881-1.237)	0.6181
<i>Alistipes putredinis</i>	-0.155	0.856 (95% CI, 0.768-0.954)	0.005	0.0367	-0.063	0.939 (95% CI, 0.798-1.106)	0.4512
<i>Sutterella wadsworthensis</i>	-0.16	0.852 (95% CI, 0.758-0.959)	0.0077	0.0458	-0.014	0.986 (95% CI, 0.83-1.171)	0.872

- Significant p-values (< 0.05) for models using data from Western Finland are indicated in bold font.

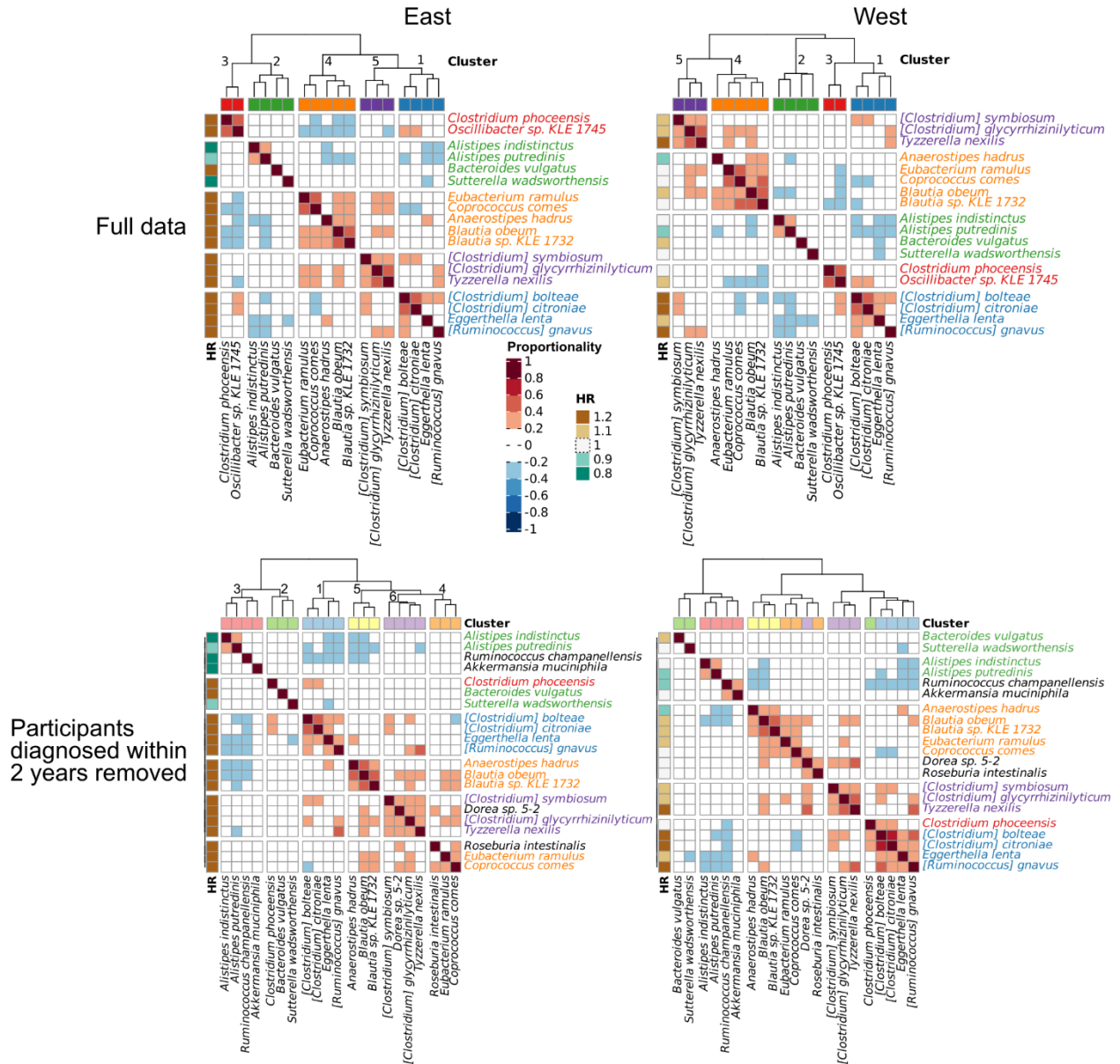


Figure S1. Effect of excluding participants diagnosed with type 2 diabetes in the first two years of the follow-up period on taxa filtering, proportional abundance, and clustering. The 21 taxa in the lower panels are annotated both with the five clusters of the full data (taxa names colors), and a new clustering conducted separately in the subset Eastern Finland data (color bar on top of the heatmaps on the lower row panels). The two annotations can also be identified by their different color schemes.

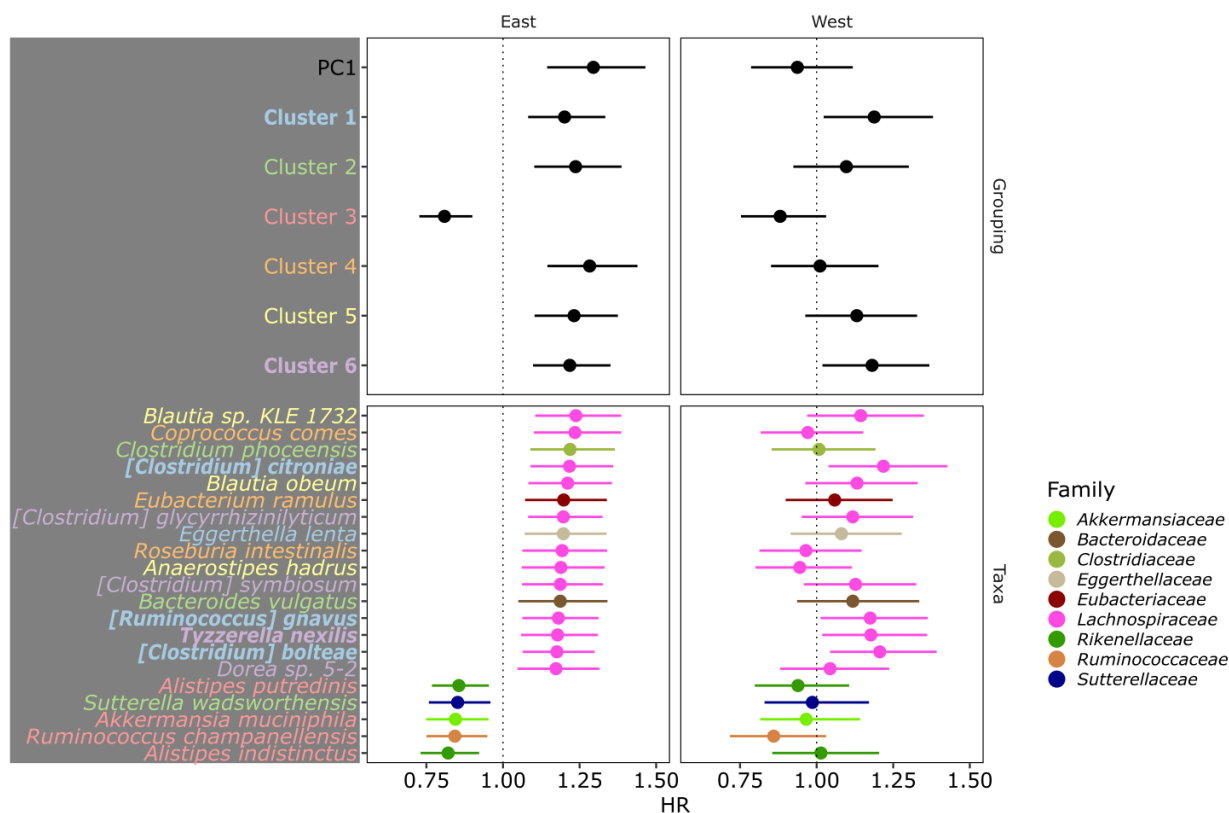


Figure S2. Hazard ratios between models in Eastern and Western Finland data after excluding participants diagnosed with type 2 diabetes in the first two years of the follow-up period.

Features with significant associations in the validation (Western Finland) data are indicated in bold and the taxa colors show their membership in a cluster. The information in this figure can be found in numeric format in **Table S2**.

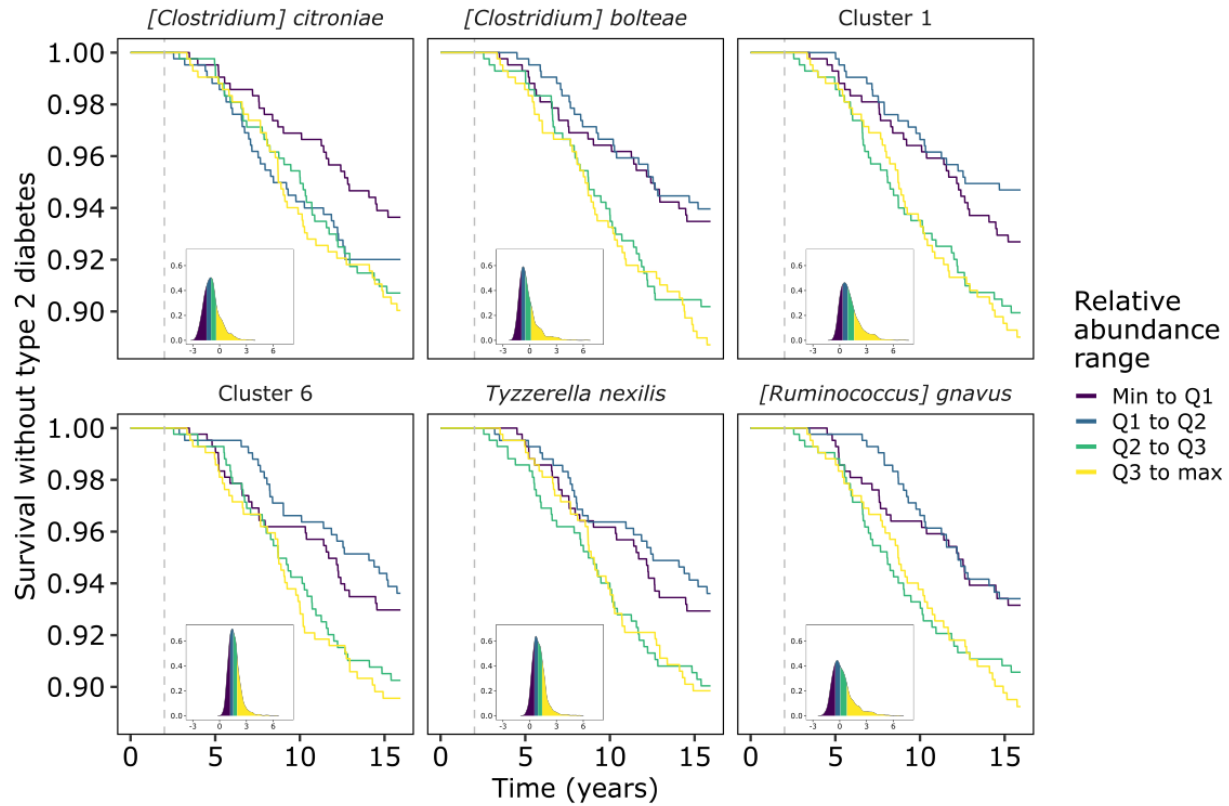


Figure S3. Kaplan-Meier curves for features with significant effect sizes in both data sets after excluding participants diagnosed with type 2 diabetes in the first two years of the follow-up. Diabetes-free survival times of participants in Western Finland is shown. Curves are separated by ranges between quartiles of relative abundance of each feature. Distribution of the participants with the same relative abundance ranges are included as inlays for each of the features. The cut-off for the removal of participants diagnosed in the first two years of the follow-up period is indicated with a dashed gray line.