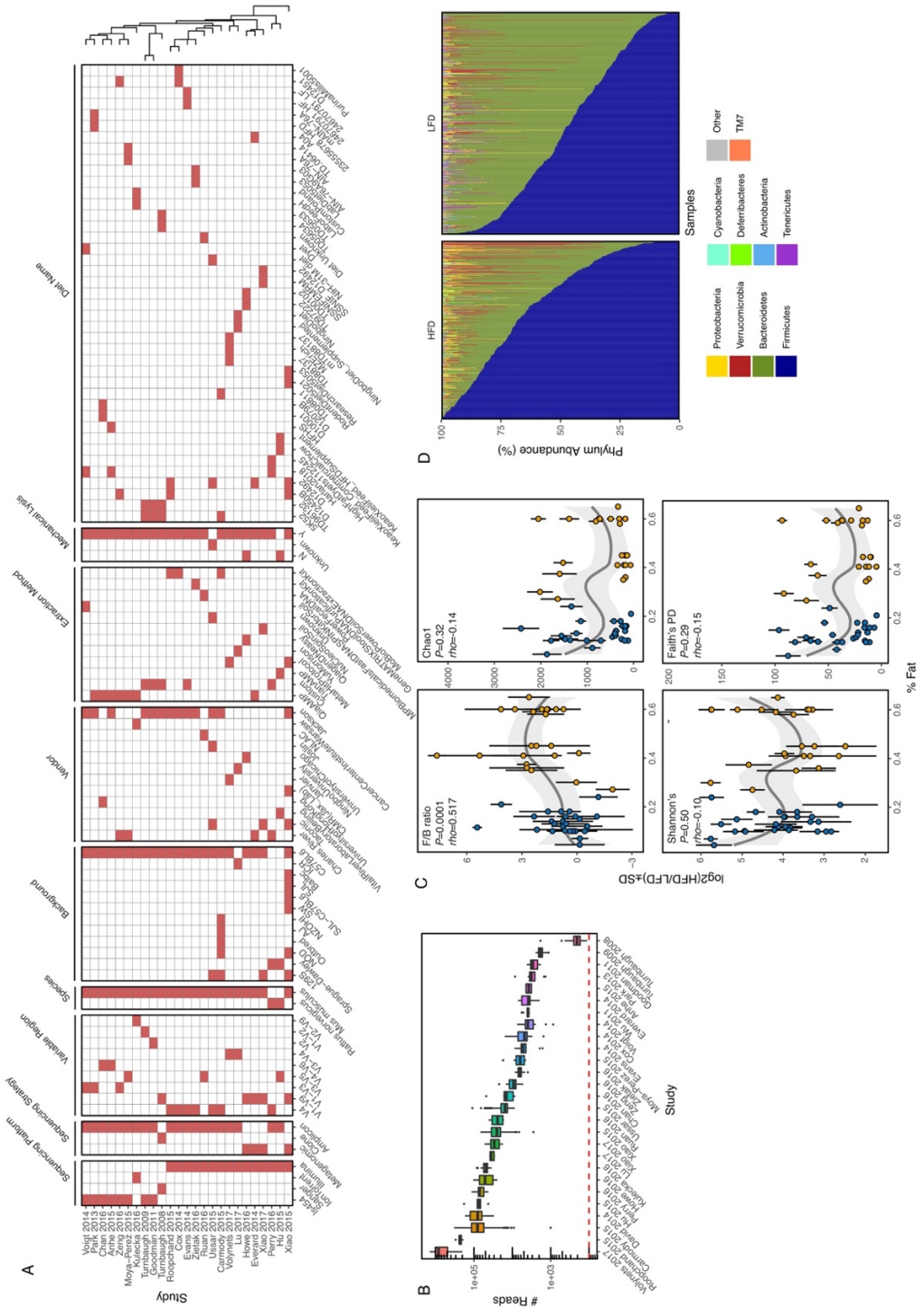


**Figure S1. Study data and correlations relating to Figure 2. (A)** Study metadata with studies ordered by UPGMA clustering. **(B)** Read depth distribution among studies. The dashed red line denotes the subsampling depth for pooled analysis. **(C)** Correlations of diversity and dietary fat content. Each point represents a study's HFD or LFD group (as colored) with the x axis representing the fat content. The trend line represents a LOESS regression ± standard error. Dietary fat is positively correlated with the ratio of Firmicutes to Bacteroidetes (P=0.0001, *rho*=0.517). **(D)** Phylum-level bar plot contrasting HFD and LFD across all studies.

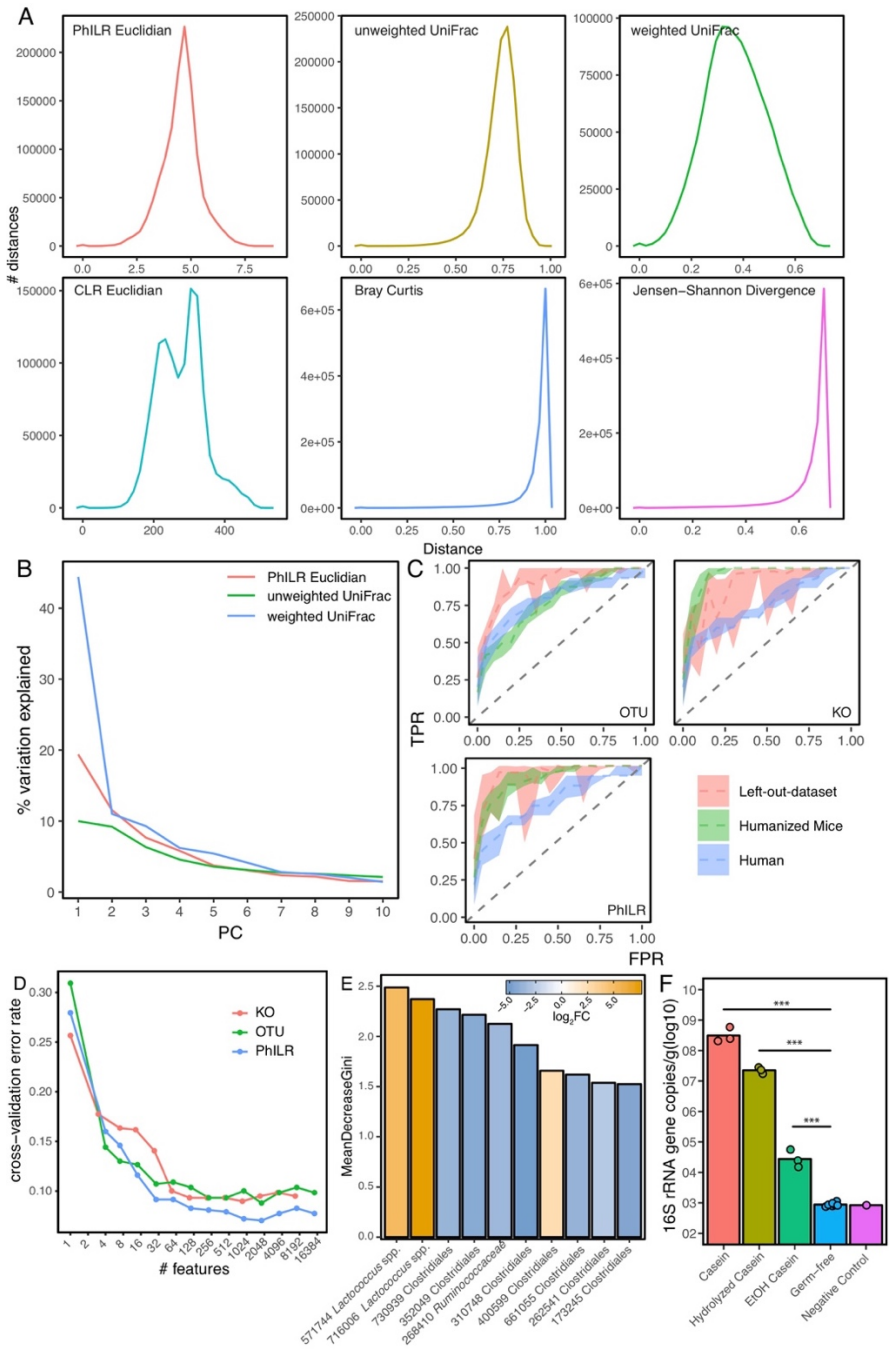


**Figure S2. Ordination and feature selection data relating to Figures 3 and 4**. **(A)** Distribution of intersample distances. A high degree of saturation is noted in Bray Curtis (non-overlapping=1) and Jensen-Shannon Divergence (non-overlapping=ln(2)=0.693). CLR euclidian was pre-emptively removed due to possible effects of zero-replacement for logarithmic normalization. **(B)** Scree plot showing variation explained by axes for PCoAs in **Figure 3. (C)** Leave-one-dataset-out (LODO) receiver operator curves demonstrate that predictive accuracy can be obtained across studies. The dashed line represents the median and the top and bottom of the ribbon represent the 3rd and 1st quartiles respectively. **(D)** 10-fold cross validation for feature selection identifies a minimum core set of informative features. **(E)** The most important features (as ranked by mean decrease in GINI coefficient) from a random forest classifier demonstrates an over-representation of *Lactococcus* spp. **(F)** Bacterial DNA content in various casein preparations. P<0.001 ANOVA with TukeyHSD.

**Table S1 relating to Figure 1. Descriptions of studies meeting inclusion criteria.**

| Study Identifier | PubmedID | Type | # samples | # LFD | # HFD | Platform | HFD[1] | LFD[1] | Variable Regions Covered | Strain specified in paper | Vendor |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Anhe 2015 | 25080446 | Murine | 4 | 2 | 2 | Is454 | High-Fat High Sugar Diet (NR) | Harlan 2018 (NC) | 6-8 | C57BL/6J | Jackson |
| Carmody 2015 | 25532804 | Murine | 334 | 184 | 150 | Illumina | TD08811 (C) | 5K52 (NC) | 4 | 129S1SvImJ, A/J, C57BL/6J, NZO/HlLtJ, NOD/LtJ, Outbred | Jackson |
| Chan 2016 | 27821063 | Murine | 9 | 4 | 5 | Is454 | D12079B (C) | D10001 (C) | 4-6 | Apoe-/- (C57BL/6) | University of Hong Kong |
| Cox 2014 | 25126780 | Murine | 54 | 24 | 30 | Illumina | D12451 (C) | PurinaMills5001 (NC) | 4 | C57BL/6J, Swiss Webster colonized with C57BL/6J | Jackson (C57BL/6J), Taconic (Swiss Webster) |
| Evans 2014 | 24670791 | Murine | 11 | 5 | 6 | Illumina | Custom HFD (C*) | Custom LFD (C*) | 4 | C57BL/6J | Jackson |
| Everard 2014 | 24694712 | Murine | 16 | 9 | 7 | Illumina | D12492 (C) | A04 (NC) | Metagenome | C57BL/6J | Charles River Belgium |
| Goodman 2011 | 21436049 | Murine | 20 | 10 | 10 | Is454 | TD96132 (C) | 5K52 (NC) | 2 | C57BL/6J | Jackson |
| Howe 2016 | 26473721 | Murine | 6 | 3 | 3 | Illumina | TD97222 (C) | TD00102 (C) | Metagenome | C57BL/6 | University of Chicago |
| Hu 2015 | 26129950 | Murine | 6 | 3 | 3 | Illumina | Keao Xieli Feed HFD Supplemented (NR) | Keao Xieli Feed Commercial Chow (NR) | 3-5 | Sprague Dawley | Vital River Laboratory Beijing |
| Kulecka 2016 | 27559357 | Murine | 32 | 16 | 16 | IonTorrent | Custom HFD (NR) | Labo Feed H (NC) | 2-9 | C57BL/6W | Cancer Center Institute Warsaw |
| Lu 2017 | 28567037 | Murine | 6 | 3 | 3 | Illumina | Custom HFD (NR) | Custom LFD (NR) | 3-4 | ICR | Ningbo University |
| Moya-Perez 2015 | 26161548 | Murine | 16 | 8 | 8 | Is454 | TD.06414 (C) | AIN-76A (C) | 3-5 | C57BL/6 | Charles River France |
| Park 2013 | 23555678 | Murine | 18 | 9 | 9 | Is454 | Custom HFD (C) | Modified AIN-76A (C) | 1-3 | C57BL6/J | Jackson |
| Perry 2016 | 27279214 | Murine | 15 | 8 | 7 | Illumina | High Fat Diet Dyets 112245 (C**) | Harlan 2018 (NC) | 4 | Sprague Dawley | Charles River |
| Roopchand 2015 | 25845659 | Murine | 20 | 10 | 10 | Illumina | D12492 (C) | D12450B (C) | 4 | C57BL6/J | Jackson |
| Ruan 2016 | 27892926 | Murine | 11 | 5 | 6 | Illumina | Custom HFD (NR) | Custom LFD (NR) | 3-5 | C57BL/6J | NLAC |
| Turnbaugh 2008 | 18407065 | Murine | 30 | 12 | 18 | Sanger | TD96132 (C), TD05634 (C) | 5K52 (NC), TD05633 (C) | 1-9 | C57BL/6J | Jackson |
| Turnbaugh 2009 | 20368178 | Murine | 26 | 14 | 12 | Is454 | TD96132 (C) | 5K52 (NC) | 1-2 | C57BL6/J | Jackson |
| Ussar 2015 | 26299453 | Murine | 109 | 41 | 68 | Illumina | D12492 (C) | NIH-31M diet (NC) | 4 | 129S1/SvImJ, 129Jax/Jos, 129S6/SvEvTac, 129Tac/Jos, C57BL/6J, B6Jax/Jos | Jackson, Taconic, Joslin |
| Voigt 2014 | 24848969 | Murine | 18 | 10 | 8 | Is454 | High Fat Diet from Dyet (NC) | Harlan 2018 (NC) | 1-3 | C57BL/6J | Jackson |
| Volynets 2017 | 28356436 | Murine | 72 | 60 | 12 | Illumina | Modified TD88137 (C) | M-Z Erich (C***) | 3-4 | C57BL/6J | Janvier |
| Xiao 2015 | 26414350 | Murine | 146 | 111 | 35 | Illumina | D12492 (C) | RodentDiet5021 (NC), D12450B (C), ResearchDiet5053 (NC) | Metagenome | Sv129, C57BL/6, C57BL/6J, BALB/c, SJL, 129S, SJL-C57BL/6, NODLtJ, Swiss Webster | Taconic Denmark, Jackson, Taconic, CMR (Jax Lab), Taconic USA CMR |
| Xiao 2017 | 28390422 | Murine | 35 | 16 | 19 | Illumina | SSNIF_D12492 (C) | SSNIFEMRM (NC) | Metagenome | 129S6/SvEvTac, C57BL/6JBomTac | Taconic Denmark |
| Zeng 2016 | 27362974 | Murine | 12 | 6 | 6 | Is454 | D12451 (C) | D12450B (C) | 1-3 | C57BL/6 | Charles River |
| Zietak 2016 | 27304513 | Murine | 47 | 23 | 24 | Illumina | AIN-76A9G03 (C) | LabDiet5053 (NC) | 4 | C57BL/6J | Jackson |
| David 2014 | 24336217 | Human | 19 | 10 | 9 | Illumina | NA | NA | 4 | NA | NA |
| Wu 2011 | 21885731 | Human | 10 | 5 | 5 | Is454 | NA | NA | 1-2 | NA | NA |

[1]C = Casein containing diet; NC = no casein reported in dietary formulation; NR = casein content of diet not reported

*Suspect casein in HFD and LFD given similarity to other diets reported previously

**Suspect casein in HFD and LFD given all HFD from Dyets vendor report use of casein

***Suspect casein given protein source for diets is highly suggestive of casein content

**Table S2 relating to Figure 2 and 3. ADONIS tests for entire dataset.**

| | | With Lactococcus OTUs | | Without Lactococcus OTUs | |
|---|---|---|---|---|---|
| **Metric** | **Term[1]** | **$R^2$** | ***P*-value** | **$R^2$** | ***P*-value** |
| *PhILR Euclidean* | | | | | |
| | Diet_Classification | 0.0335 | 0.001 | 0.0323 | 0.001 |
| | StudyID | 0.4865 | 0.001 | 0.4874 | 0.001 |
| | Residuals | 0.4800 | | 0.4804 | |
| *weighted UniFrac* | | | | | |
| | Diet_Classification | 0.1155 | 0.001 | 0.1108 | 0.001 |
| | StudyID | 0.2910 | 0.001 | 0.2840 | 0.001 |
| | Residuals | 0.5937 | | 0.6052 | |
| *unweighted UniFrac* | | | | | |
| | Diet_Classification | 0.0491 | 0.001 | 0.0441 | 0.001 |
| | StudyID | 0.2672 | 0.001 | 0.2637 | 0.001 |
| | Residuals | 0.6837 | | 0.6922 | |

[1]Formula=(Distance Matrix ~ Diet_Classification+StudyID, strata=StudyID)

**Table S3 relating to Figure 4. Random Forest Data Sets**

| Murine Training Set | | Murine Test Set | | External Murine Samples | | Humanized Mice | | Humans | |
|---|---|---|---|---|---|---|---|---|---|
| StudyID | # Samples | StudyID | # Samples | StudyID | # Samples | StudyID | # Samples | StudyID | # Samples |
| Anhe 2015 | 3 | Anhe 2015 | 1 | Evans 2014 | 11 | Goodman 2011 | 20 | David 2014 | 19 |
| Carmody 2015 | 216 | Carmody 2015 | 118 | Everard 2014 | 16 | Turnbaugh 2009 | 26 | Wu 2011 | 10 |
| Chan 2016 | 6 | Chan 2016 | 3 | Xiao 2015 | 146 | | | | |
| Cox 2014 | 37 | Cox 2014 | 17 | | | | | | |
| Howe 2016 | 5 | Howe 2016 | 1 | | | | | | |
| Hu 2015 | 5 | Hu 2015 | 1 | | | | | | |
| Kulecka 2016 | 23 | Kulecka 2016 | 9 | | | | | | |
| Lu 2017 | 4 | Lu 2017 | 2 | | | | | | |
| Moya-Perez 2015 | 9 | Moya-Perez 2015 | 6 | | | | | | |
| Park 2013 | 12 | Park 2013 | 6 | | | | | | |
| Perry 2016 | 10 | Perry 2016 | 5 | | | | | | |
| Roopchand 2015 | 13 | Roopchand 2015 | 7 | | | | | | |
| Ruan 2016 | 9 | Ruan 2016 | 2 | | | | | | |
| Turnbaugh 2008 | 23 | Turnbaugh 2008 | 7 | | | | | | |
| Ussar 2015 | 73 | Ussar 2015 | 36 | | | | | | |
| Voigt 2014 | 11 | Voigt 2014 | 7 | | | | | | |
| Volynets 2017 | 43 | Volynets 2017 | 29 | | | | | | |
| Xiao 2017 | 26 | Xiao 2017 | 9 | | | | | | |
| Zeng 2016 | 8 | Zeng 2016 | 4 | | | | | | |
| Zietak 2016 | 33 | Zietak 2016 | 14 | | | | | | |
| Total | 569 | | 284 | | 173 | | 46 | | 29 |

**Table S5 relating to Figure 4. Receiver operator curve (ROC) areas under the curve (AUC).**

| | AUROC | | | | AUROC without *Lactococcus* | | | |
|---|---|---|---|---|---|---|---|---|
| **Data Set** | **F/B ratio** | **OTU** | **KO** | **PHILR** | **F/B ratio** | **OTU** | **KO** | **PHILR** |
| Murine Training Set | 0.7670 | 1.0000 | 1.0000 | 1.0000 | 0.7607 | 1.0000 | 1.0000 | 1.0000 |
| Murine Test Set | 0.8005 | 0.9697 | 0.9716 | 0.9849 | 0.7957 | 0.9669 | 0.9784 | 0.9816 |
| External Murine Sample | 0.7878 | 0.9301 | 0.8073 | 0.8592 | 0.7832 | 0.9119 | 0.8245 | 0.8923 |
| Humanized Mice | 0.9508 | 0.7888 | 1.0000 | 0.8210 | 0.9394 | 0.6089 | 0.9754 | 0.7472 |
| Human | 0.4333 | 0.5881 | 0.8595 | 0.7571 | 0.4333 | 0.6167 | 0.7262 | 0.7762 |