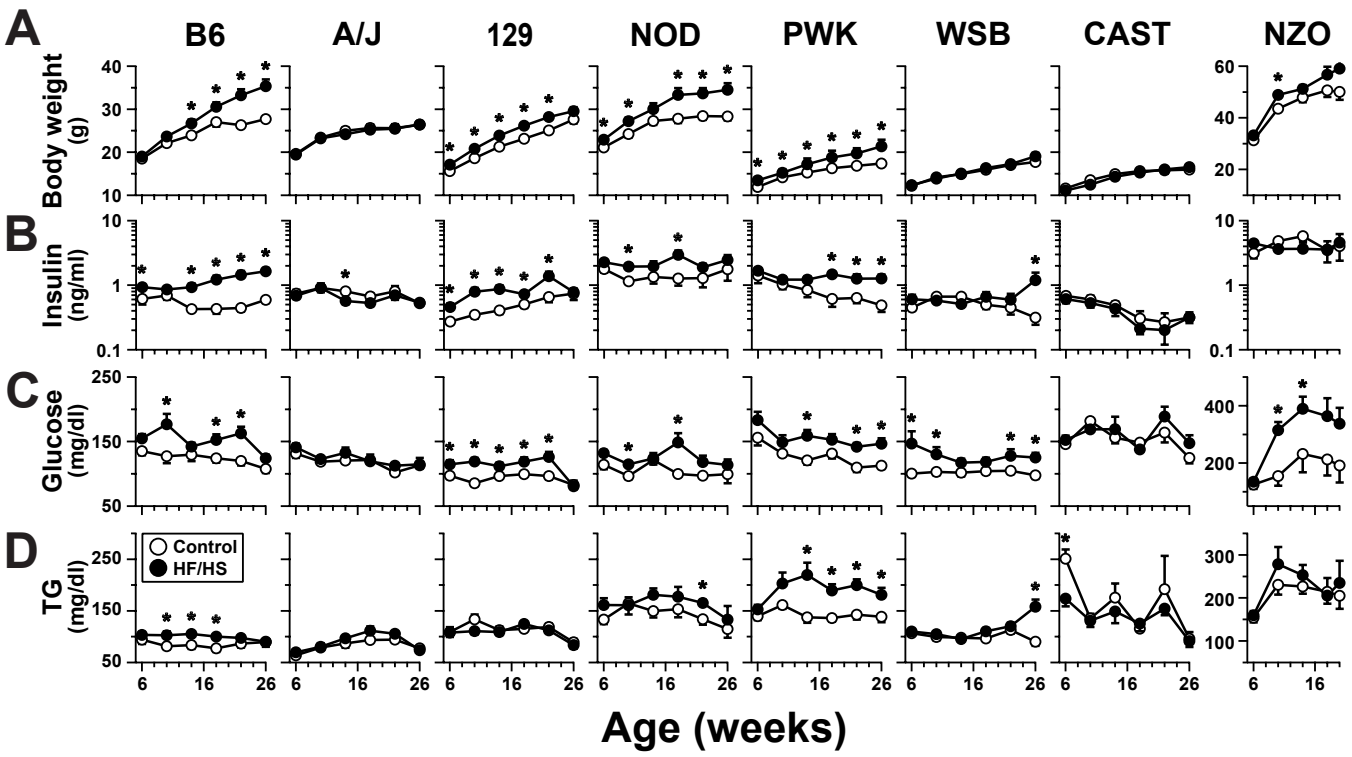

Figure S1. Segregation of metabolic syndrome among the founder strains of the Collaborative Cross (CC). Related to Figure 1. (A) Body weight, (B) fasting plasma insulin, (C) glucose and (D) triglycerides (TG) were determined at various ages for CC founder mice fed either a high-fat/high-sucrose (HF/HS) or a control diet for 22 weeks. Note differences in Y-axis scale for NZO mice. *$p < 0.05$. data are mean ± SEM, n ≥ 9 mice/genotype/diet.

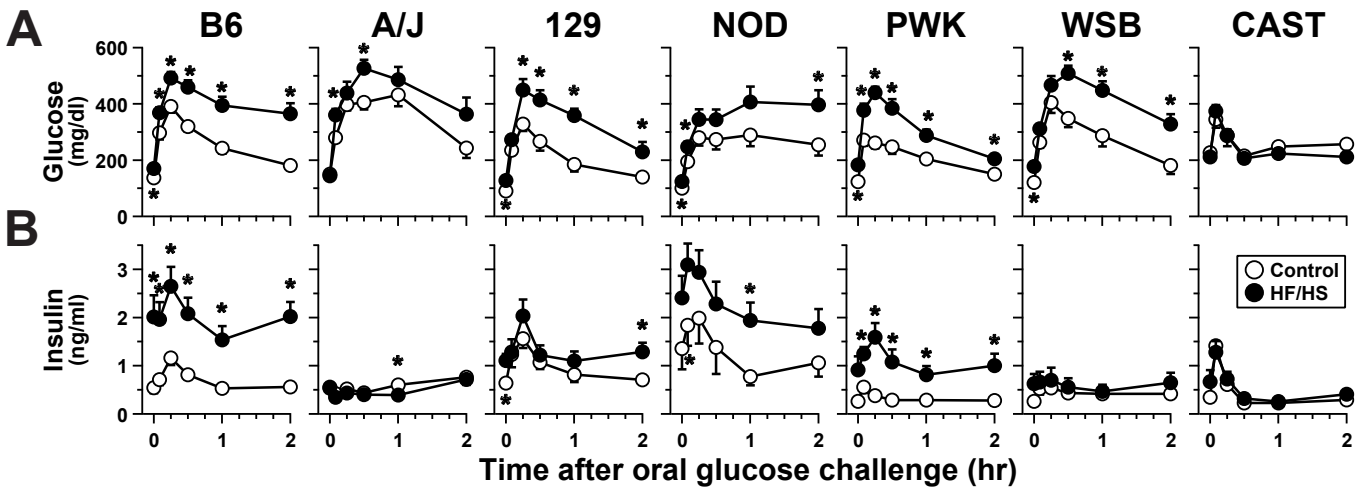

Figure S2. Diet-induced glucose tolerance and insulin sensitivity differ among CC founder mice. Related to Figure 1. Male mice were maintained on either a control or high-fat/high-sucrose (HF/HS) diet. At 22 weeks of age, mice were given glucose bolus (2 g/kg body weight) via oral gavage following a 4-hour fast. Blood was collected via retro-orbital bleed at 0, 5, 15, 30, 60, and 120 minutes following the glucose bolus, and used to determine plasma (A) glucose and (B) insulin levels. NZO mice did not survive 22 weeks of age on HF/HS diet. *$p < 0.05$. Data are mean ± SEM, n ≥ 9 mice/genotype/diet.

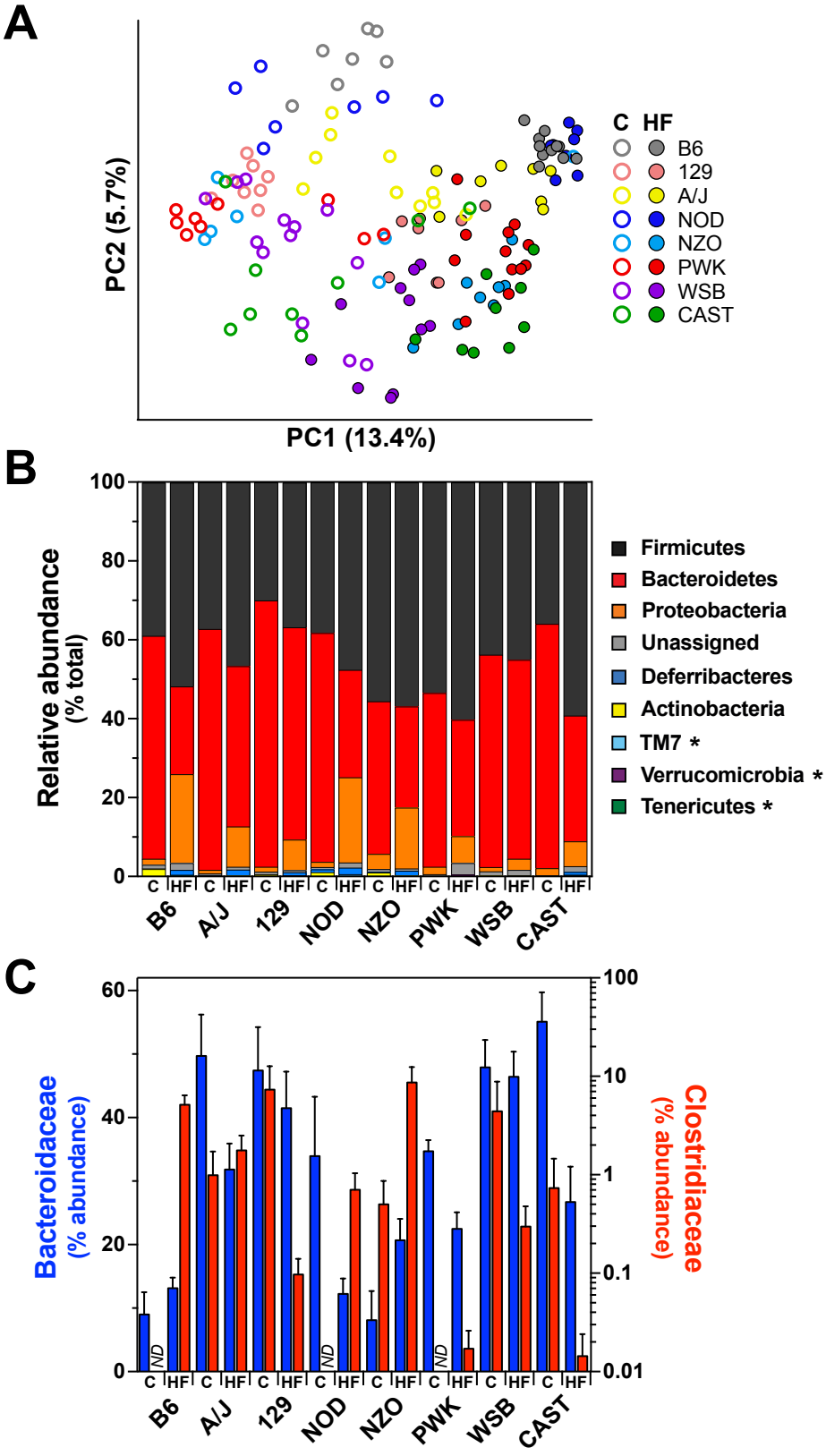

Figure S3. Host genotype and diet affect microbial composition. Related to Figure 2. (A) Principal Coordinate Analysis (PCoA) of unweighted UniFrac distances for the cecal microbiota of the founder mice. Open symbols, control diet; filled symbols HF/HS diet. (B) Relative abundance of 8 major microbiota phyla identified in cecal contents from CC founder mice maintained on control (*abv. C*) or HF/HS (*abv. HF*) diet for 22 weeks. Phyla ordered by mean abundance; * denotes mean phyla abundance < 1%. (C) Relative abundance of Bacteroidaceae and Clostridicaceae in CC founder strains. Family not detected marked as ND. Data are mean ± SEM, n ≥ 9 mice/genotype/diet.

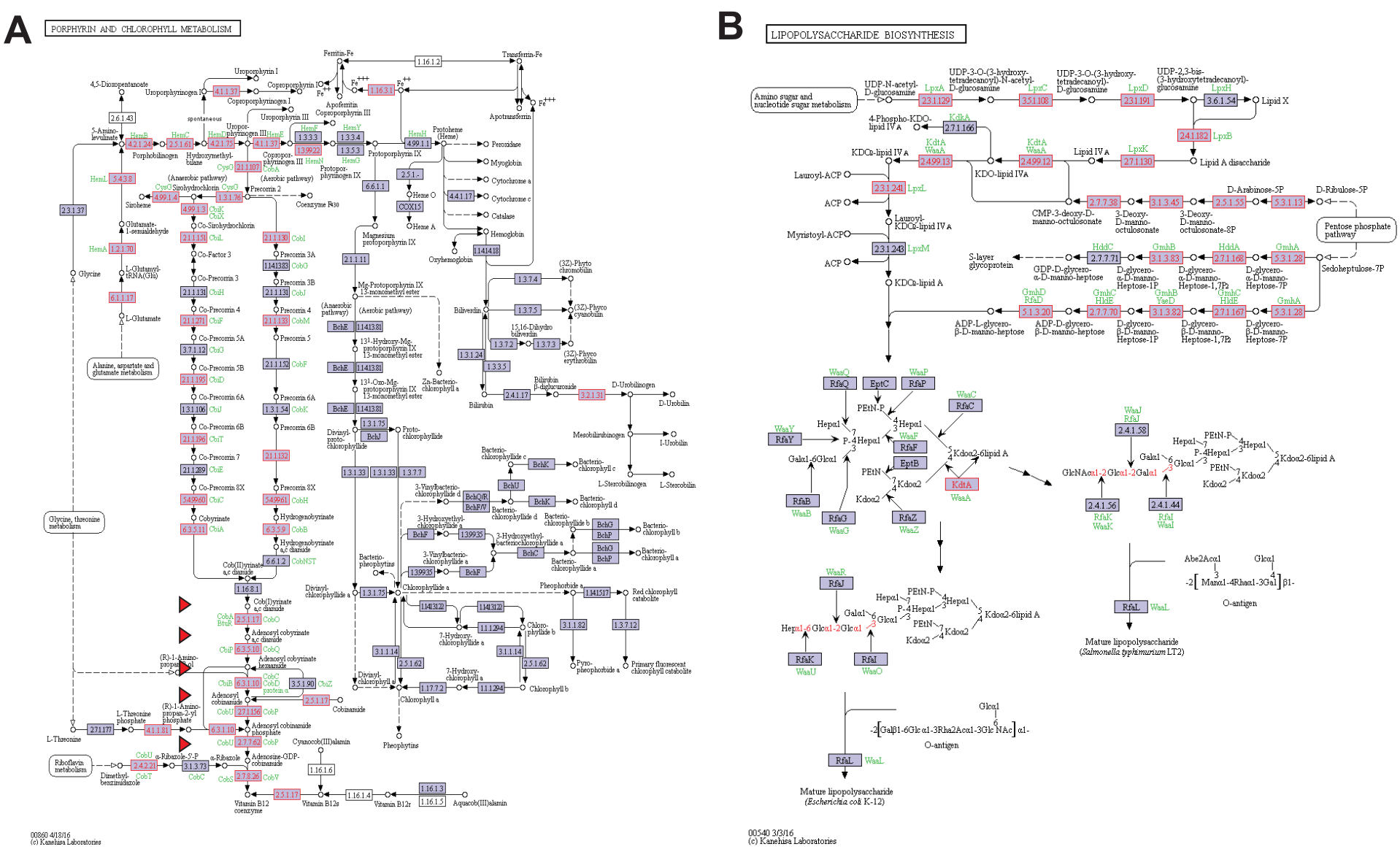

Figure S4. Microbial pathways enriched in CAST-derived microbiota. Related to Figure 4. (A) Vitamin B12 biosynthesis is functionally enriched in CAST-derived microbiota. KEGG pathway for "Porphyrin and Chlorophyll Metabolism" (map00860). Fifty-six genes within the pathway were more abundant in $B6_{CAST}$ than $B6_{B6}$ microbiota. ECs higher in the $B6_{CAST}$ microbiota compared to $B6_{B6}$ microbiota colored in red. The KO annotations for the 56 genes in the Porphyrin and Chlorophyll Metabolism pathway were input to the Reconstruct Module of KEGG Mapper. Red triangles indicate members of the module for vitamin B12 (cobalamin) biosynthesis, which is nearly complete; the missing block in this module corresponds to EC number 3.1.3.73, which is not boxed in red. (B) Lipopolysaccharide biosynthesis is functionally enriched in CAST-derived microbiota. KEGG pathway for "Lipopolysaccharide Biosynthesis" (map00540). Forty-six genes within the pathway were more abundant in $B6_{CAST}$ than $B6_{B6}$ microbiota. ECs higher in the $B6_{CAST}$ microbiota compared to $B6_{B6}$ microbiota colored in red.

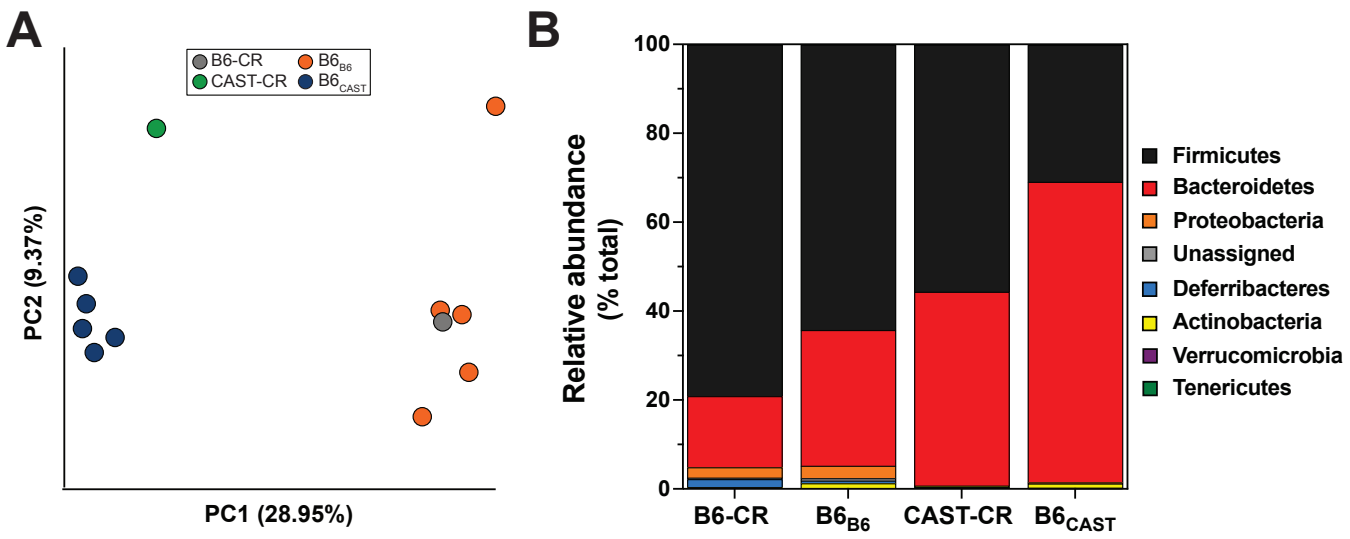

Figure S5. Microbiota composition of B6-CR, CAST-CR, $B6_{B6}$ and $B6_{CAST}$ mice used for insulin secretion studies. Related to Figure 6. (A) Principal Coordinate Analysis (PCoA) of unweighted Unifrac distances, and (B) relative abundance of microbial phyla. Data are mean ± SEM, n = 5 for $B6_{B6}$ and $B6_{CAST}$ mice and n = 1 for CR mice.

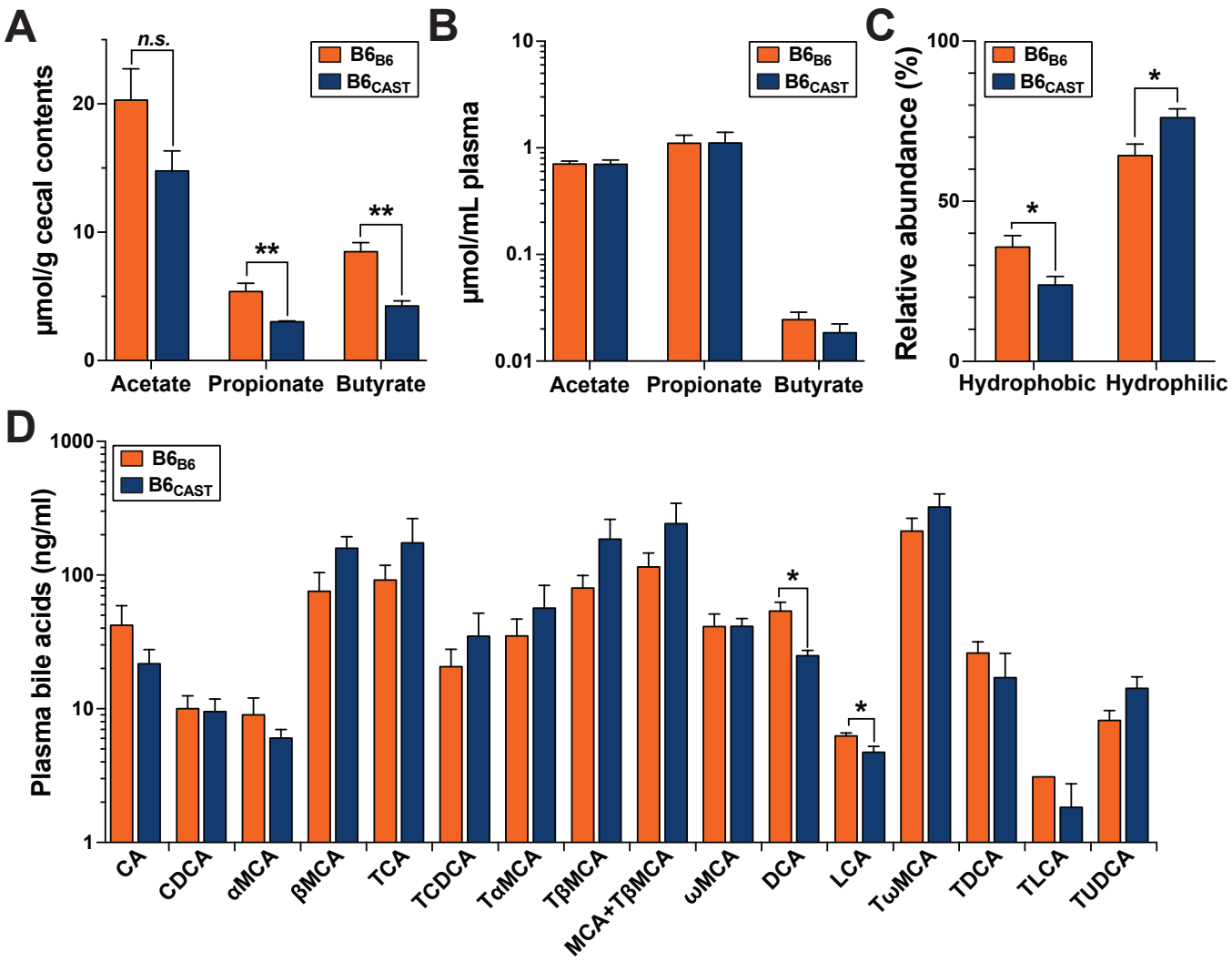

Figure S6. SCFA and BA measurements in $B6_{B6}$ and $B6_{CAST}$ mice. Related to Figure 6. (A) Cecal and (B) plasma SCFA concentrations from $B6_{B6}$ and $B6_{CAST}$ animals used for insulin secretion studies as determined by GC-MS. (C) Relative abundance of hydrophobic and hydrophilic BA species, and (D) abundance of major plasma BA species determined by UPLC-MS/MS. *p < 0.05, **p < 0.01. Data are mean ± SEM, n = 5.

Table S1. Related to Figures 1-5. Composition of diets used in this study

Low glycemic control diet (TD.08810)

Component	g/Kg
Casein	210.0
L-Cystine	3.0
INACTIVE Hi-Maize 220 (Resistant Starch)	500.0
Maltodextrin	100.0
Sucrose	39.1
Anhydrous Milkfat	20.0
Soybean Oil	20.0
Cellulose	35.0
Mineral Mix, AIN-93G-MX (94046)	35.0
Vitamin Mix, AIN-93-VX (94047)	15.0
Choline Bitartrate	2.8
TBHQ, antioxidant	0.0
Yellow Food Color	0.1

High-fat high-sucrose diet (TD.08811)

Component	g/Kg
Casein	195.0
L-Cystine	3.0
Sucose	340.0
Corn Starch	56.9
Maltodextrin	60.0
Anhydrous Milkfat	210.0
Soybean Oil	20.0
Cellulose	50.0
Mineral Mix, AIN-93G-MX (94046)	43.0
Vitamin Mix, AIN-93-VX (94047)	19.0
Choline Bitartrate	3.0
TBHQ, antioxidant	0.0
Green Food Color	0.1

Table S2. Related to Figure 2. Correlation coefficients and p-values in Collaborative Cross Founders.

Table S3. Related to Figure 4. (A) Percent of quality trimmed reads that mapped the mouse genome database. (B) Percent of quality trimmed reads that mapped the mouse metagenome database.

Table S4. Related to Figure 4. Differential abundance analysis of metagenomic data.

Table S5. Related to Figures 4 and S4. Functional enrichment within genes more abundant in B6 or CAST microbiota. (A)Enzyme commission numbers. (B) Evolutionary genealogy of genes: non-supervised orthologous groups (eggNOG). (C) KEGG orthologous (KO) terms. (D) Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways.

Supplemental Experimental Procedures

Fasting plasma measurements. Following a 4h fast, blood was collected via retro-orbital bleed in EDTA-coated eppendorf tubes. Blood samples were centrifuged and plasma was collected and stored at -80°C until further analysis. Plasma glucose was quantified using the Thermo-Fisher Infinity Glucose Oxidase reagent (Pittsburgh, PA), insulin was quantified using the Millipore-Linco Sensitive Rat Insulin RIA (Billerica, MA), and triglycerides levels were quantified Thermo-Fisher Infinity Triglycerides reagents (Pittsburgh, PA).

Oral Glucose Tolerance Test (oGTT). Mice were fasted for four hours prior to testing and were challenged with an oral dose of 2 g/kg body weight glucose at time 0. Blood was collected via retro-orbital bleed at 0, 5, 15, 30, 60 and 120 minutes post glucose challenge. Blood samples were centrifuged and plasma collected and stored at -80°C until further analysis.

Triglyceride measurement. Liver triglycerides (TG) were quantified following the Bligh and Dyer extraction method (Bligh & Dyer 1959). Briefly, ~30 mg frozen liver tissue was homogenized using a 40X dilution with 1X PBS. Total lipids were extracted from the liver homogenate in methanol-chloroform (2:1). The organic extract with dried and reconstituted in 10% Triton X-100 in isopropanol. Triglyceride content was determined by colorimetric assay from Wako (Richmond, VA) according to the manufacturer's instructions and expressed in µg of triglycerides per milligram of protein.

Microbiome Sample Processing. Genomic DNA was extracted from feces and cecum using a bead-beating protocol (Turnbaugh et al. 2009). Briefly, mouse fecal pellets (~50 mg) or cecal contents were re-suspended in a solution containing 500 µl of extraction buffer [200 mM Tris (pH 8.0), 200 mM NaCL, 20 mM EDTA], 210 µl of 20% SDS, 500 µl phenol:chloroform:isoamyl alcohol (pH 7.9, 25:24:1) and 500 µl of 0.1-mm diameter zirconia/silica beads. Samples were mechanically disrupted using a bead beater (BioSpec Products, Barlesville, OK; maximum setting for 3 min at room temperature), followed by centrifugation, recovery of the aqueous phase, and precipitation with isopropanol. NucleoSpin Gel and PCR Clean-up Kit (Macherey-Nagel, Bethlehem, PA) was used to remove contaminants. Isolated DNA was eluted in 5 mM Tris/HCl (pH 8.5) and was stored at -20°C until further use.
Collaborative cross founders: Amplicons of ~330 bp, spanning variable region 2 (V2) of the bacterial 16S rRNA gene, were generated by using modified primers 27F and 338R that incorporated sample specific barcodes (Muegge et al. 2011). A final library for sequencing was created by combining equimolar ratios of amplicons from the individual samples. The 16S rRNA amplicon mixture was subjected to 454 pyrosequencing conducted on a Roche GS Junior (Roche, Indianapolis, IN) with the Lib-L kit and Titanium chemistry.
Transplant: Amplification of 16S rRNA genes (V4) was done from DNA by PCR using unique 8-bp barcodes on the forward and reverse primers and fused with Illumina sequencing adapters (Kozich et al. 2013). Each sample was amplified in duplicate in a reaction volume of 25µl using KAPA HiFi HotStart DNA polymerase (KAPA Biosystems, Wilmington, MA), 10µM of each primer and ~25ng of genomic DNA. PCR was carried out under the following conditions: initial denaturation for 3 min at 95°C, followed by 25 cycles of denaturation for 30 s at 95°C, annealing for 30 s at 55°C and elongation for 30 s at 72°C, and a final elongation step for 5 min at 72°C. PCR products were purified with the NucleoSpin Gel and PCR Clean-up kit (Macherey-Nagel, Bethlehem, PA) and then quantified using Qubit dsDNA HS Assay kit (Invitrogen, Oregon, USA). Samples were pooled and sequenced on the Illumina MiSeq 2x250bp platform.

Microbiota Analysis in QIIME. Demultiplexing of 16S rRNA gene sequences, quality control and operational taxonomic unit (OTU) binning were performed using Quantitative Insights Into Microbial Ecology (QIIME)(Caporaso, Kuczynski, et al. 2010) version 1.9.1. Quality filtered reads were trimmed of Illumina adaptor and barcode sequences. Sequences were then clustered in OTUs using an open-reference OTU picking protocol based on 97% identity using UCLUST (Edgar 2010) against the Greengenes reference database (McDonald et al. 2012). Representative sequences (most abundant sequence in OTUs) were picked, aligned to GreenGenes Core reference alignment (DeSantis et al. 2006) using PyNAST (Caporaso, Bittinger, et al. 2010). Taxonomic assignments were associated with OTUs based on the taxonomy associated with the Greengenes reference sequence defining each OTU. UniFrac distances between samples were calculated using the Greengenes reference tree (Lozupone & Knight 2005). Greengenes reference sequences, trees and taxonomy data used in the analysis can be found at: http://greengenes.secondgenome.com/downloads/database/13_5

The resulting biom-formatted OTU table was filtered to remove singletons. CC founder cecal samples sequenced by 454

pyrosequencing were rarefied to an even sampling depth of 900 reads, and 5 samples were removed from the dataset as assigned reads fell below the rarefaction point of 900 reads/sample. Donor and recipient samples from microbiota transplants were rarefied to an even sampling depth of 10,000 reads/sample. The relative abundance of each taxon was calculated by dividing the sequences pertaining to a specific taxon by the total number of sequences for that sample. OTUs representing less than 0.1% were removed for relative abundance assessments and correlation analyses. Assessments of alpha-diversity and beta-diversity were also conducted on the rarefied OTU table in QIIME. Principal coordinate analysis (PCoA) was performed in QIIME using UniFrac distances calculated from the Greengenes reference tree. Permutation Multivariate Analysis of Variance (PERMANOVA) was used to compare strength of sample groups (diet, genotype) for founder PCoA using the compare_categories.py command in QIIME. Linear discriminant analysis (LDA) effect size (LEfSe) was used to identify taxa that discriminated between the fecal microbiota of transplant recipient mice using standard parameters ($p < 0.05$, LDA score 2.0)(Segata et al. 2011). For correlation analyses, only microbial families with at least one non-zero measurement for each strain on at least one diet were included. Correlations between microbiota and phenotypes and association testing were performed in R. Correlation coefficients and adjusted *p*-values are reported in **Table S2**.

Metagenomic analysis. Raw reads were pre-processed using the fastx toolkit (version 0.0.13) (Hannon Lab): raw reads were demultiplexed using fastx_barcode_splitter (specifying -bol -partial 2 and -mismatches 2), barcodes were trimmed using fastx_trimmer, (specifying –f 9 and –Q 33), and quality trimmed using fastq_quality_trimmer, (specifying -t 20 -l 30 and -Q 33). In order to filter out host contaminating reads in the metagenome samples, we identified paired and unpaired reads in our demultiplexed and trimmed files, and mapped them independently to the mouse genome assembly (Ensembl release 84, GRCm38.dna.toplevel) using Bowtie2 (v. 2.2.7) (Langmead & Salzberg 2012) with default settings (**Table S3**). From this output, we then identified reads that did not map to the mouse genome using samtools view (version 1.3) (Li et al. 2009), specifying –f 4 only for unpaired reads, or both –f4 and –f 8 for paired reads in addition to default settings, and regenerated .fastq files containing only reads that did not map to the mouse genome (custom perl scripts).

In order to examine gene-level abundance differences among our samples, we utilized the mouse gut metagenome gene sequences available from the Mouse Gut Metagenome Project (downloaded from gigadb.org: http://gigadb.org/dataset/100114) (Xiao et al. 2015). "Decontaminated" paired and unpaired reads were independently mapped against genes in the mouse gut metagenome assembly with Bowtie2 using default settings (v. 2.2.7) (Langmead & Salzberg 2012). A table of raw read counts was generated using htseq-count command (v. 0.6.0) (Anders et al. 2014), specifying a 'mock'.gff file containing "gene" entries, whose lengths were lengths of genes, for example:

S-Fe10_GL0000040 mock gene 1 1870 . + . gene_id "S-Fe10_GL0000040";

The resulting raw read count table was filtered to exclude low abundance genes, defined here as genes with average raw read counts of less than 10 across all 12 samples (total number of 73,905 genes), and then input into DESeq2 (version 1.10.1) (Love et al. 2014), for library size normalization (default settings). To allow for comparison of individual gene abundances, counts were further normalized by gene length to give "reads per kilobase gene" (**Table S4**).

In order to examine the similarity of the B6-derived microbiota DESeq2 was also used to identify genes differentially abundant between the B6 and CAST-derived microbiota using default settings, and found a very large number of genes differentially abundant (29,283 in B6 > CAST microbiota; 10,742 in CAST > B6 microbiota, with FDR < 0.05, **Table S4**), indicating dramatic genic diversity between B6 and CAST-derived microbiota.

Functional (KEGG Orthology (KO)and eggNOG annotations) and taxonomic annotations corresponding to the detected genes from the aforementioned mouse gut metagenome were downloaded (gigadb.org: http://gigadb.org/dataset/100114)(Xiao et al. 2015). This was further expanded to include enzyme commission numbers (ECs) and KEGG pathway information, by mapping these functions from KEGG to the individual genes by way of KO annotations. Enrichment of these functional groups in genes of increased abundance in B6 or CAST-derived microbiota compared to background (all genes detected) was examined using a Fisher's exact test (*p*-value < 0.05).

Bile acid analysis. ~100 mg feces were homogenized in 500 µl 50:50 water:methanol. Next 500 µl of alkaline acetonitrile (5% ammonium hydroxide in acetonitrile) was added to the homogenate, which was then heated for 20 minutes at 75°C. 500 µl of the mixture was centrifuged at 11,000 RPM for 10 minutes and 250 µl of the supernatant was collected and evaporated under N2 gas. Samples were reconstituted in 50 µl 50:50 water:methanol and ^{2}H4-CDCA was added to the samples for a final concentration of 2 µg/ml. For serum samples, 1 ml ice-cold acetonitrile was added to 50 µl serum and spiked with the internal standard for a final concentration of 2 µg/ml ^{2}H4-CDCA. The mixture was vortexed, centrifuged at 15,000 x g for 10 minutes, and the supernatant was aspirated and evaporated under vacuum. The LC-MS/MS conditions used were as described (Zhang & Klaassen 2010).

Measurement of SCFAs. Flash-frozen cecal contents (100mg) or plasma (50 µl) were mixed with 20 µl internal standards (acetic-d4 acid, Sigma-Aldrich #233315; propionic-3,3,3-d3 acid, CDN isotopes #D-80; and butyric-d7 acid, CDN isotopes #D-171) and acidified with 20 µl 33% HCl. Two rounds of extraction using 1 ml diethyl ether were carried out by mixing for 10 minutes at room temperature following by centrifugation at 1932 x g for 10 minutes at 4°C. Extracts (60 µl) were then incubated at room temperature for 2 hours with 2 µl N-tert-Butyldimethylsilyl-N-methyltrifluoroacetamide (MTBSTFA, Sigma-Aldrich #394882). Derivatized samples (1 µl) were injected onto an Agilent 7890B/5977A GC/MSD instrument with a DB1-ms column. A linear temperature gradient was

used, wherein the initial temperature of 80°C was held for 1 minute, then increased to 280°C at a rate of 15°C per minute prior to a final hold at 280°C for 5 minutes. The source temperature was set to 200°C and emission current to 300mA. The injector and transfer line temperatures were set to 250°C. Quantitation was performed using selected ion monitoring acquisition mode and metabolites were compared to relevant labeled internal standards using Agilent Mass Hunter v Acquisition B.07.02.1938. The m/z of monitored ions are as follows: 117 (acetic acid), 120 (acetic-d4 acid), 131 (propionic acid), 134 (propionic-3,3,3-d3 acid), 145 (butyric acid), and 152 (butyric-d7 acid). Concentrations were normalized to g of cecal contents or ml plasma.

Islet isolation, *ex vivo* insulin secretion and RNA isolation. Intact pancreatic islets were isolated from mice using a collagenase digestion procedure (Rabaglia et al. 2005). Briefly, islets were carefully hand-picked under a stereo microscope to remove contaminating acinar tissue. For insulin secretion assays, single islets were placed in a well of a 96-well microtiter plate and used to determine the amount of insulin secreted in response to low (1.7 mM) or high (16.7 mM) glucose, KCl (40 mM, plus 1.7 mM glucose), or the incretin hormone GLP-1 (100 nM, plus either 8.3 or 16.7 mM glucose). From each mouse, 7 islets were used per secretory condition, and 5 mice were surveyed per strain (B6, A/J, WSB, CAST), or transplant group ($B6_{B6}$, $B6_{CAST}$). Insulin secretion was monitored over a 45 min period. Insulin levels in the medium as well as that remaining within the islets was determined by ELISA.
Islets used for RNA isolation were washed twice with phosphate buffered saline (PBS) and centrifuged at 1500 rpm, 3 min RT. The PBS supernatant was removed and 350 µl RLT buffer (Qiagen, Hilden, Germany) was added. Islets were homogenized by hand for 1 min with a plastic micropestel (USA Scientific) and stored at -80°C until RNA purification. Total RNA was purified using the RNeasy Mini Kit (Qiagen, Hilden, Germany) following manufacturer's directions with on-column TURBO DNase treatment (Invitrogen, Carlsbad, CA).

Quantitative Real-Time PCR. SuperScript II Reverse Transcriptase with oligo(dT) primer (all from Invitrogen, Carlsbad, CA) was used to synthesize 20 µl cDNA templates from 100 ng purified RNA. cDNA was diluted 2X before use and qRT-PCR reactions were prepared in a 10µl volume using SsoAdvanced Universal SYBR Green Supermix (Bio-Rad, Hercules, CA, USA) and 400 nM specific primers targeting the gene of interest (FXR-F [5'-CCAACCTGGGTTTCTACCC-3']; FXR-R [5'-CACACAGCTCATCCCCTTT-3']). Reactions were run on a CFX96 Real-Time PCR System (Bio-Rad, Hercules, CA, USA). Relative gene expression was calculated by the ΔΔCt method using β-actin as an internal control.

Statistical Analysis. The data are expressed as mean ± SEM and analyzed using GraphPad Prism 6.0 (GraphPad Software, La Jolla, CA). Multiple groups were analyzed by one-way or two-way ANOVA followed by Bonferroni's multiple comparisons test. Significant differences between two groups were evaluated by two-tailed unpaired Student's t-test or Mann-Whitney U test for samples that were not normally distributed. Pearson's correlations between microbiota and phenotypes and association testing were performed in R. The level of significance was set at $p < 0.05$; $*p < 0.05$, $**p < 0.01$, $***p < 0.001$, $****p < 0.0001$.

Supplemental References

Anders, S., Pyl, P.T. & Huber, W., 2014. HTSeq - A Python framework to work with high-throughput sequencing data. *bioRxiv*, p.002824.

Bligh, E.G. & Dyer, W.J., 1959. A rapid method of total lipid extraction and purification. *Canadian journal of biochemistry and physiology*, 37(8), pp.911–917.

Caporaso, J.G., Bittinger, K., et al., 2010. PyNAST: a flexible tool for aligning sequences to a template alignment. *Bioinformatics*, 26(2), pp.266–267.

Caporaso, J.G., Kuczynski, J., et al., 2010. QIIME allows analysis of high-throughput community sequencing data. *Nature Methods*, 7(5), pp.335–336.

DeSantis, T.Z. et al., 2006. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Applied and environmental microbiology*, 72(7), pp.5069–5072.

Edgar, R.C., 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, 26(19), pp.2460–2461.

Hannon Lab, FASTX Toolkit. Available at: http://hannonlab.cshl.edu/fastx_toolkit/.

Kozich, J.J. et al., 2013. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Applied and environmental microbiology*, 79(17), pp.5112–5120.

Langmead, B. & Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9(4), pp.357–359.

Li, H. et al., 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16), pp.2078–2079.

Love, M.I., Huber, W. & Anders, S., 2014. *Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2*, Cold Spring Harbor Labs Journals.

Lozupone, C. & Knight, R., 2005. UniFrac: a new phylogenetic method for comparing microbial communities. *Applied and environmental microbiology*, 71(12), pp.8228–8235.

McDonald, D. et al., 2012. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *The ISME Journal*, 6(3), pp.610–618.

Muegge, B.D. et al., 2011. Diet drives convergence in gut microbiome functions across mammalian phylogeny and within humans. *Science (New York, N.Y.)*, 332(6032), pp.970–974.

Rabaglia, M.E. et al., 2005. Alpha-Ketoisocaproate-induced hypersecretion of insulin by islets from diabetes-susceptible mice. *American journal of physiology. Endocrinology and metabolism*, 289(2), pp.E218–24.

Segata, N. et al., 2011. Metagenomic biomarker discovery and explanation. *Genome biology*, 12(6), p.R60.

Turnbaugh, P.J. et al., 2009. A core gut microbiome in obese and lean twins. *Nature*, 457(7228), pp.480–484.

Xiao, L. et al., 2015. A catalog of the mouse gut metagenome. *Nature Biotechnology*, 33(10), pp.1103–1108.

Zhang, Y. & Klaassen, C.D., 2010. Effects of feeding bile acids and a bile acid sequestrant on hepatic bile acid composition in mice. *Journal of lipid research*, 51(11), pp.3230–3242.