Methods

Animal welfare approval was not needed for this study since all data came from animals raised in a commercial setting by The Maschhoffs, LLC (Carlyle, IL, USA). All pigs were harvested in commercial facilities under the supervision of USDA Food Safety and Inspection Service.

Animals and sample collection

Data were collected from crossbred individuals that were obtained from 28 funding Duroc sires and 747 commercial F_1 sows composed of Yorkshire × Landrace or Landrace × Yorkshire. The pigs were weaned at 18.64 ± 1.09 days old and were moved to nursery-finishing facility. Pigs were kept in 334 single-sire single-sex pens with 20 pigs per pen. The test period began the day that pigs were moved to the nursery-finishing facility. During the nursery, growth and finishing period all pigs were fed a standard pelleted feed based on sex and live weight. Details of diet and their nutritional values are provided (see additional File 1). The pigs received a standard vaccination and medication routine (see additional File 2). End of test (**Off-test**) was reached when the average weight of pigs of each pen reached 138 kg. The average age at off_test was 196.4 ± 7.80 days. Fecal samples for 16S rRNA sequencing were collected as follows. Rectal swabs were collected from all pigs at three stages: weaning (**Wean**), 15 weeks post weaning (**Mid-test**; average 118.2 ± 1.18 days), and off_test. Four pigs from each pen were selected as detailed by [23] and their rectal swabs were used for subsequent microbial sequencing. There were 1,205, 1,295 and 1,273 samples at weaning, Mid-test and Off-test respectively. Distribution of samples across families, time points and sex are provided (see additional file 3).

Illumina amplicon sequencing

DNA extraction, purification, illumina library preparation and sequencing were done as described by [24]. Briefly, total DNA (gDNA) was extracted from each rectal swab by mechanical disruption in phenol: chloroform. The DNA was purified using a QIAquick 96 PCR purification kit (Qiagen, MD, USA). Purification was performed per the manufacturer's instruction with the following minor modifications: (i) sodium acetate (3 M, pH 5.5) was added to Buffer PM to a final concentration of 185 mM to ensure optimal binding of genomic DNA to the silica membrane; (ii) crude DNA was combined with 4 volumes of Buffer PM (rather than 3 volumes); and (iii) DNA was eluted in 100 µL Buffer EB (rather than 80µL). All sequencing was performed at DNA Sequencing Innovation Laboratory at the Center of Genome Sciences and Systems Biology at Washington University in St. Louis. Phased, bi-directional amplification of the v4 region (515-806) of the 16S rRNA gene was employed to generate indexed libraries for Illumina sequencing as described in [25]. Sequencing was performed on an Illumina MiSeq instrument (Illumina, Inc. San Diego, USA), generating 250 bp paired-end reads.

16S rRNA gene sequencing and quality control of data

Pairs of 16S rRNA gene sequences were first merged into a single sequence using FLASH v1.2.11 [26] with a required overlap of at least 100 and less than 250 base pairs in order to provide confident overlap. Sequences with a mean quality score below Q35 were then filtered out using PRINSEQ v0.20.4 [27]. Sequences were oriented in the forward direction and any primer sequences were matched and trimmed off. Mismatch was allowed up to 1 base pair. Sequences were subsequently demultiplexed using QIIME v1.9 [28]. Sequences with greater than 97% nucleotide sequence were clustered into operational taxonomic units (**OTU**) using QIIME with the following settings: `max_accepts = 50,`

`max_rejects = 8, percent_subsample = 0.1 and -suppress_step4`. A modified version of GreenGenes [29,30] was used as reference database. Input sequences that had 10% of the reads with no hit to the reference database were then clustered de novo with UCLUST [29] to generate new reference OTU to which the remaining 90% of reads were assigned. The most abundant sequence in each cluster was used as representative sequence for the OTU. Sparse OTU were then filtered out by requiring a minimum total observation count of 1,200 for an OTU to be retained, the resulting OTU table was rarefied to 10,000 counts per sample and after data processing and quality control 1,755 OTU were retained for further analysis.

Genotyping

All pigs were genotyped with the PorcineSNP60 v2 BeadChip (Illumina, Inc., San Diego, CA). Quality control procedures were applied by removing the SNPs that had call rate less than 0.90 and minor allele frequency less than 0.05. After quality control the number of SNPs remaining for further analyses was 42,529.

Phenotypic data

Phenotypic data collection was done as described by Wilson et al [23]. Meat quality traits (intramuscular fat content (**IMF**), Minolta a* (**MINA**), Minolta b* (**MINB**), minolta L* (**MINL**), ultimate pH (**PH**), subjective color score (**SCOL**), subjective marbling score (**SMARB**), subjective firmness score (**SFIRM**), shearing force (**SSF**)) and carcass composition traits (Belly weight (**BEL**), ham weight (**HAM**), loin weight (**LOIN**), fat depth (**FD**), loin depth (**LD**) and carcass average daily gain (**CADG**)) were used for the current analysis. All the traits were measured as described by Khanal et al [31]. A summary of traits used in current analysis is reported in Table 1.

Statistical analysis

The data were analyzed using ASReml v4 [32]. Univariate analyses were conducted to estimate heritabilities, microbiabilities and variance components for each trait. Single trait models were fitted as:

$$y_{ijklmn} = \mu + dl_i + cg_j + sex_k + animal_l + pen_{m(j)} + e_{ijklmn} \quad (1)$$

where μ was the overall mean, dl_i was the ith fixed effect of dam line (2 levels), cg_j was the jth fixed effect of the contemporary group (6 levels), sex_k was the kth fixed effect of sex (2 levels), $animal_l$ was the random animal genetic effect, $pen_{m(j)}$ was the random effect of pen nested within contemporary group and e_{ijklmn} was the random residual. Pen and residuals were assumed normally distributed with mean zero and with variances $\mathbf{I}\sigma^2_{pen}$ and $\mathbf{I}\sigma^2_e$, respectively, where **I** was an identity matrix. The random effect of animal was assumed normally distributed with mean 0 and variance $\mathbf{G}\sigma^2_a$ where **G** was a realized genomic relationship matrix obtained according to VanRaden [33] as:

$$\mathbf{G} = \frac{(\mathbf{M}-\mathbf{P})(\mathbf{M}-\mathbf{P})'}{2\sum_{j=1}^{m} p_j(1-p_j)}$$

where **M** is a matrix of marker alleles with *m* columns (*m* = total number of markers) and *n* rows (*n* = total number of genotyped individuals), and **P** is a matrix containing the frequency of the second allele (p_j), expressed as $2p_j$. $\mathbf{M_{ij}}$ was -1 if the genotype of individual i for SNP j was homozygous for the first allele, 0 if heterozygous, or 1 if the genotype was homozygous for the second allele. Narrow sense heritabilities were estimated as $h^2 = \frac{\sigma^2_a}{\sigma^2_P}$, with $\sigma^2_P = \sigma^2_a + \sigma^2_{pen} + \sigma^2_e$.

We added the microbiome information to model (1) in order to estimate the changes in heritability due to the incorporation of microbiome information at each stage of sample collection. Model (2) was then:

$$y_{ijklmno} = \mu + dl_i + cg_j + sex_k + animal_l + microbiome_m + pen_{n(j)} + e_{ijklmno} \quad (2)$$

Where *dl*, *cg*, *sex*, *animal*, *pen* and *e* were as previously described and $microbiome_m$ was the random effect of the animal microbiome. The effect of the microbiome was assumed normally distributed with mean 0 and variance $\mathbf{O}\,\sigma_m^2$ in which **O** was a microbial correlation matrix among individuals and σ_m^2 was the microbiome variance. The matrix **O** was created following Camarinha-Silva et al [20]. Briefly, **O** was obtained as $\mathbf{O} = \frac{1}{q}\mathbf{X}\mathbf{X}^{\mathbf{T}}$, with matrix **X** of dimension of $n \times q$, where n is the number of animals and q is the number of OTU. **X** was constructed from **S**, a matrix of equivalent dimensions $n \times q$. Each element of the **S** matrix, S_{ij}, was the relative abundance of OTU j in animal i (plus 0.001). The elements of X were calculated as:

$$X_{ij} = \frac{log\,(S_{ij}) - \overline{logS_{.j}}}{sd(logS_{.j})}$$

where $S_{.j}$ was the vector of the j[th] column of **S**. The **O** matrix was created for each stage (Wean, Mid-test and Off-test) separately and fitted in each model separately with different **O** matrix. The contribution of the microbiome to the overall variance (microbiability) was calculated as: $m^2 = \frac{\sigma_m^2}{\sigma_P^2}$ [16]. The total variance σ_P^2 was in this case obtained as $\sigma_P^2 = \sigma_a^2 + \sigma_m^2 + \sigma_{pen}^2 + \sigma_e^2$.

Bivariate analyses were subsequently conducted to estimate genomic and microbial correlations among traits. Bivariate models were of form:

$$\begin{bmatrix} \boldsymbol{y_1} \\ \boldsymbol{y_2} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X_1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{X_2} \end{bmatrix} \begin{bmatrix} \boldsymbol{b_1} \\ \boldsymbol{b_2} \end{bmatrix} + \begin{bmatrix} \boldsymbol{Z_1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Z_2} \end{bmatrix} \begin{bmatrix} \boldsymbol{a_1} \\ \boldsymbol{a_2} \end{bmatrix} + \begin{bmatrix} \boldsymbol{K_1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{K_2} \end{bmatrix} \begin{bmatrix} \boldsymbol{o_1} \\ \boldsymbol{o_2} \end{bmatrix} + \begin{bmatrix} \boldsymbol{W_1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{W_2} \end{bmatrix} \begin{bmatrix} \boldsymbol{p_1} \\ \boldsymbol{p_2} \end{bmatrix} + \begin{bmatrix} \boldsymbol{e_1} \\ \boldsymbol{e_2} \end{bmatrix} \quad (3)$$

where $\boldsymbol{y_1}$ and $\boldsymbol{y_2}$ were the vectors of phenotypic measurements for trait 1 and trait 2 respectively; $\boldsymbol{X_1}$ and $\boldsymbol{X_2}$ were the incidence matrices relating the fixed effects to vector $\boldsymbol{y_1}$ and vector $\boldsymbol{y_2}$ respectively; $\boldsymbol{b_1}$ and $\boldsymbol{b_2}$ were the vector of fixed effect for trait 1 and trait 2 respectively; $\boldsymbol{Z_1}$ and $\boldsymbol{Z_2}$ were the incidence matrices relating the phenotypic observations to the vector of random animal effects for trait 1 and trait 2 respectively; $\boldsymbol{a_1}$ and $\boldsymbol{a_2}$ were the vectors of random animal effect for trait 1 and trait 2 respectively; $\boldsymbol{K_1}$ and $\boldsymbol{K_2}$ were the incidence matrices relating the phenotypic observations to the vector of random microbiome effect for trait 1 and trait 2 respectively; $\boldsymbol{o_1}$ and $\boldsymbol{o_2}$ were the vectors of random microbiome effect for trait 1 and trait 2 respectively; $\boldsymbol{W_1}$ and $\boldsymbol{W_2}$ were the incidence matrices relating the phenotypic observations to the vector of random pen effects for trait 1 and trait 2 respectively; $\boldsymbol{p_1}$ and $\boldsymbol{p_2}$ were the vector of random pen effect for trait 1 and trait 2 respectively; and $\boldsymbol{e_1}$ and $\boldsymbol{e_2}$ were the vectors of random residuals for trait 1 and trait 2 respectively. The fixed effects and random effects were the same as fitted in the univariate analyses.

The additive effects were again assumed normally distributed with means 0 and variance $\text{Var}\begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \mathbf{C} \otimes \mathbf{G}$; where $\mathbf{C} = \begin{bmatrix} \sigma_{a1}^2 & \sigma_{a12} \\ \sigma_{a21} & \sigma_{a2}^2 \end{bmatrix}$. The elements of the covariance matrix **C** were defined as: σ_{a1}^2, the genetic variance for trait 1, σ_{a2}^2 , the genetic variance for trait 2, $\sigma_{a12} = \sigma_{a21}$, the additive genetic covariance between trait 1 and trait 2. Similar assumptions were made for the microbiome effect for which the covariance structure was assumed $\text{Var}\begin{bmatrix} o_1 \\ o_2 \end{bmatrix} = \mathbf{Q} \otimes \mathbf{O}$; with $\mathbf{Q} = \begin{bmatrix} \sigma_{m1}^2 & \sigma_{m12} \\ \sigma_{m21} & \sigma_{m2}^2 \end{bmatrix}$. The elements **Q** were: σ_{m1}^2 , the

microbiome variance for trait 1, σ^2_{m2} , the microbiome variance for trait 2 and $\sigma_{m12} = \sigma_{m21}$ the microbiome covariance between trait 1 and trait 2. The pen (co)variance structure was $\text{Var}\begin{bmatrix}p_1\\p_2\end{bmatrix} = \mathbf{W} \otimes \mathbf{I}$; with $\mathbf{W} = \begin{bmatrix}\sigma^2_{p1} & 0\\0 & \sigma^2_{p2}\end{bmatrix}$ and **I** an identity matrix. The **W** matrix elements were: σ^2_{p1}, and σ^2_{p2} being the pen variance for trait 1 and 2, respectively. Pen effects were assumed uncorrelated among traits. The residual variance was given by $\text{Var}\begin{bmatrix}e_1\\e_2\end{bmatrix} = \mathbf{R} \otimes \mathbf{I}$; where $\mathbf{R} = \begin{bmatrix}\sigma^2_{e1} & \sigma^2_{e12}\\\sigma^2_{e21} & \sigma^2_{e2}\end{bmatrix}$ and **I** was an identity matrix. The components of **R** were defined as: σ^2_{e1} was the residual variance for trait 1, σ^2_{e2} was the residual variance for trait 2, $\sigma^2_{e21} = \sigma^2_{e12}$ was the residual covariance among the two traits. Preliminary analyses (data not shown), showed that correlations were not estimable for the traits with estimated microbiome variance less than 3%. Microbial correlations were therefore estimated among traits for which microbiome explained at least 3% of total phenotypic variance. In all cases, microbial correlations were not estimated at weaning since microbiome accounted for less than 3% of total variance for all traits.

Diversity analysis and its correlation with traits

The diversity analysis performed in this paper was aimed at investigating the distribution of alpha diversities at Wean, Mid-test and Off-test. The R package "vegan" [34] was used to calculate alpha diversity at each stage. The diversity was measured using Shannon index, and was computed as: $-\sum_{i=1}^{n} p_i \, ln(p_i)$ where p_i was the proportional abundance of i^{th} OTU. To estimate the correlation among different traits and alpha diversity at weaning (**alpha_w**), week 15 (**alpha_mid**) and end of test (**alpha_off**), bivariate analyses were conducted using ASReml v.4 [32] by removing the effect of microbiome from model (3) and fitting diversity as the dependent variable.