**ANIMAL: An International Journal of Animal Bioscience**

**Supplementary file:**

**Fecal microbiome sequences in relation to the egg laying performance of hens using amplicon-based metagenomic association analysis**

A. A. Elokil, M. Magdy, S. Melak, H. Ishfaq, A. Bhuiyan, L. Cui, M. Jamil, S. Zhao and S. Li

**Table S1** Statistical table for data pre-processing and quality control in high and low egg laying performance of hens

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Sample ID** | **Raw PE** | **Raw Tags** | **Clean Tags** | **Effective Tags** | **Effective**  **Rate (%)** | **Effective**  **Bases (nt)** | **Avg. Length (nt)** | **Q20 content (%)** | **Q30 content (%)** | **GC**  **Content (%)** |
| **H1** | 148349 | 122638 | 117956 | 100696 | 67.88 | 46611525 | 463 | 96.93 | 92.87 | 51.6 |
| **H2** | 117520 | 96746 | 93172 | 89885 | 76.48 | 41528018 | 462 | 96.94 | 92.86 | 51.18 |
| **H3** | 168216 | 138511 | 133304 | 113477 | 76.46 | 52590427 | 463 | 96.84 | 92.73 | 51.1 |
| **H4** | 160507 | 133496 | 128476 | 116468 | 72.56 | 54065798 | 464 | 96.86 | 92.71 | 51.19 |
| **H5** | 153776 | 127443 | 122925 | 106822 | 69.47 | 49339870 | 462 | 96.96 | 92.93 | 51.1 |
| **L1** | 135724 | 111277 | 107175 | 96101 | 70.81 | 44100335 | 459 | 96.99 | 92.95 | 51.98 |
| **L2** | 138579 | 114990 | 110722 | 95882 | 69.19 | 44407593 | 463 | 96.87 | 92.73 | 51.35 |
| **L3** | 144418 | 119718 | 115175 | 101454 | 70.25 | 47112510 | 464 | 96.88 | 92.8 | 51.4 |
| **L4** | 158150 | 131811 | 127248 | 107483 | 67.96 | 49433621 | 460 | 97.03 | 93.03 | 51.55 |
| **L5** | 133767 | 109220 | 105659 | 89256 | 66.72 | 40607651 | 455 | 97.08 | 93.11 | 52.36 |

H: hens of high laying performance; L: hens of low laying performance; Samples per group n =5

Column abbreviations in order:

(1) Sample ID: Sample name

(2) Raw Reads: Pair-end (PE) raw reads without primers and barcodes

(3) Raw Tags: Merged raw Tags

(4) Clean Tags: Tags with low quality sequences removal

(5) Effective Tags: Tags with chimeric sequences removal

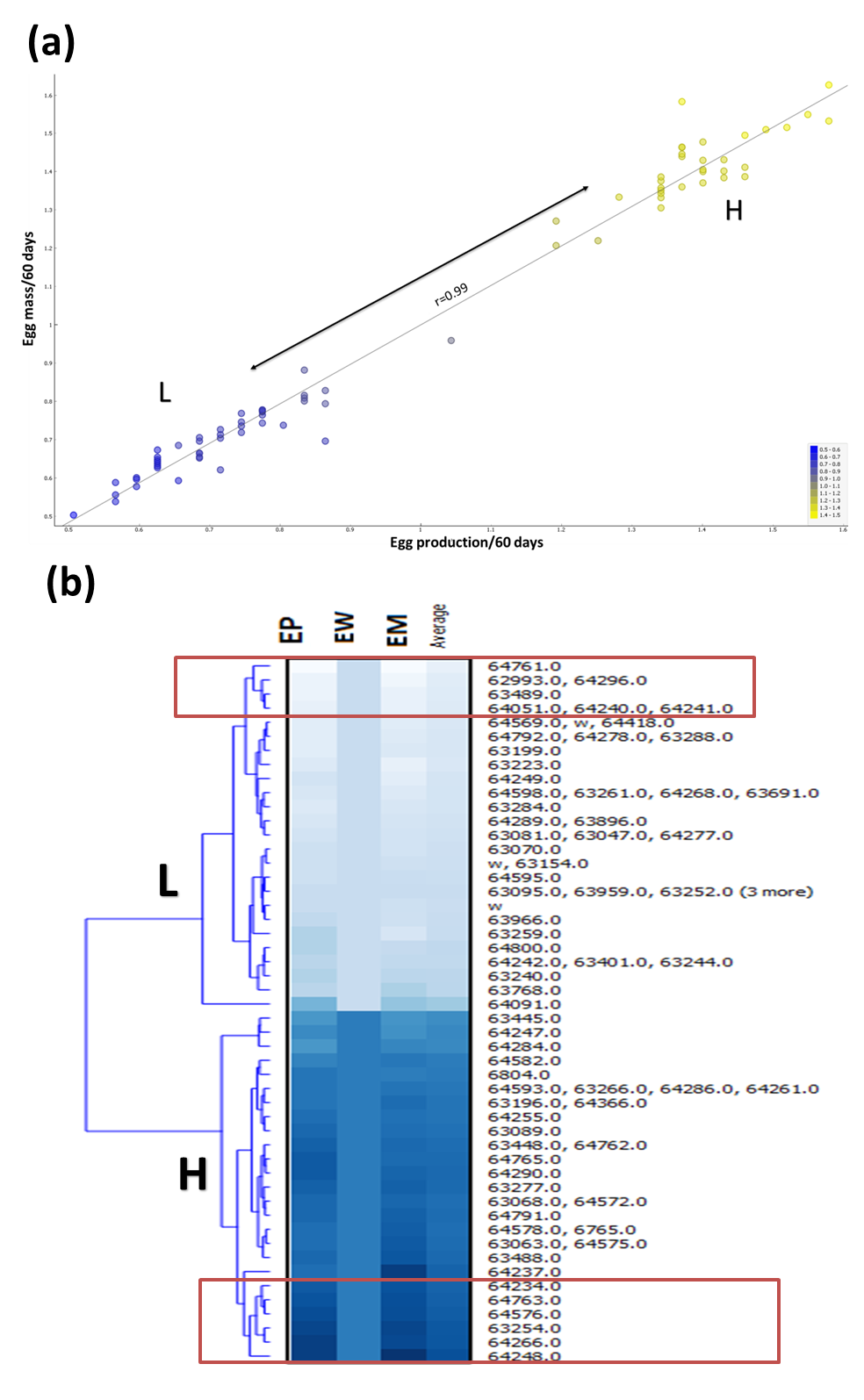
(6) Effective Rate(%): The proportion of Effective Tags to Raw PE-Reads

(7) Effective Bases (nt): Bases of Effective Tags

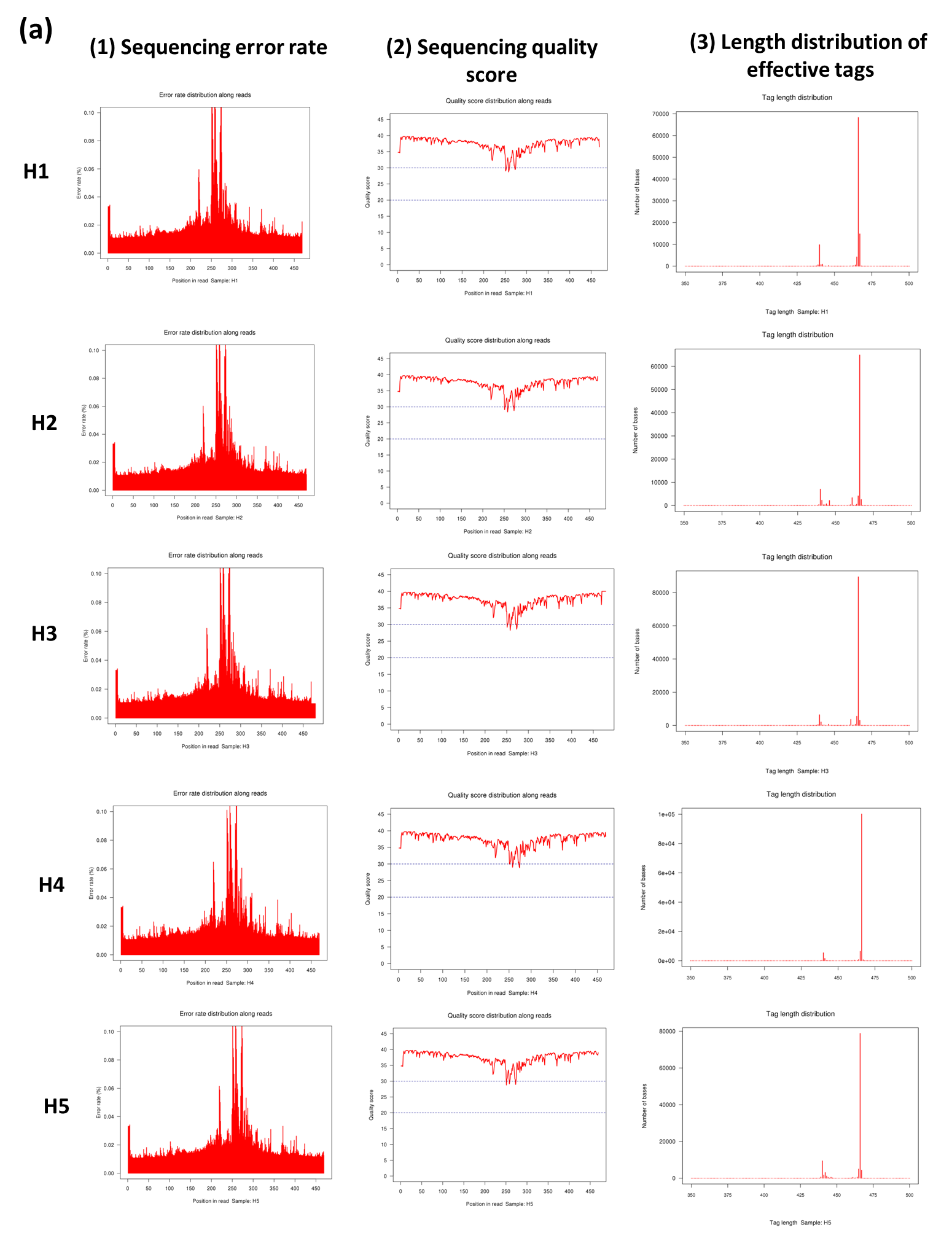
(8) Avg Length (nt): Average length of Effective Tags

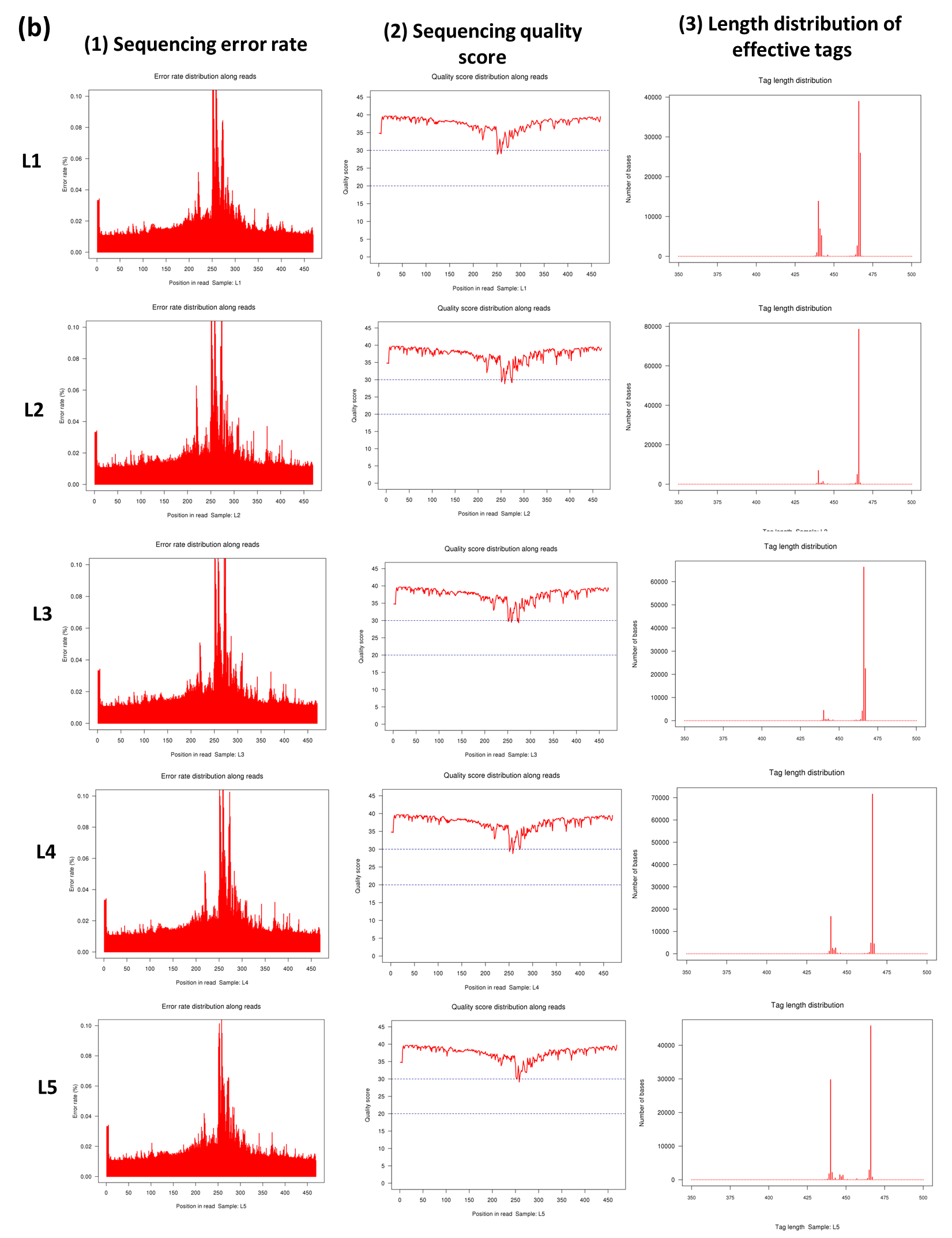
(9, 10) Q20 (%)/Q30 (%): base quality scores of Effective Tags

(11) GC (%): GC content of Effective Tags.

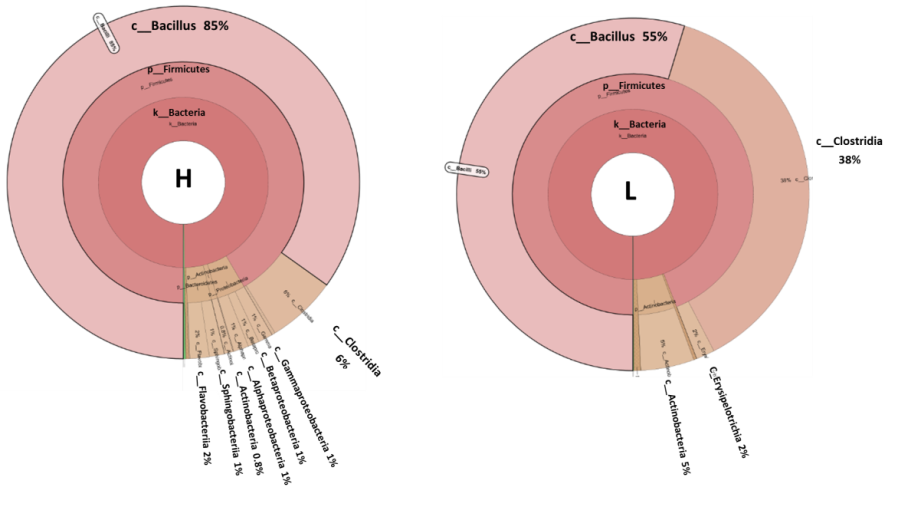
****

**Figure S1**Data processing for egg performance to select H and L groups (n=30 for each group) from a total population of 192 laying hens. Phenotype data of egg production, egg weight, egg mass, age at first egg was standardized from the value of sample/average traits value for each individual number hen (numbered by wing). (a) scatter Plot (egg mass/egg production), color by average performance. (b) The heatmap of individual hen performance based on average egg performance, all values were estimated from mosaic display of individual wing number of hen (hens per group n=30). H: hens of high laying performance; L: hens of low laying performance.

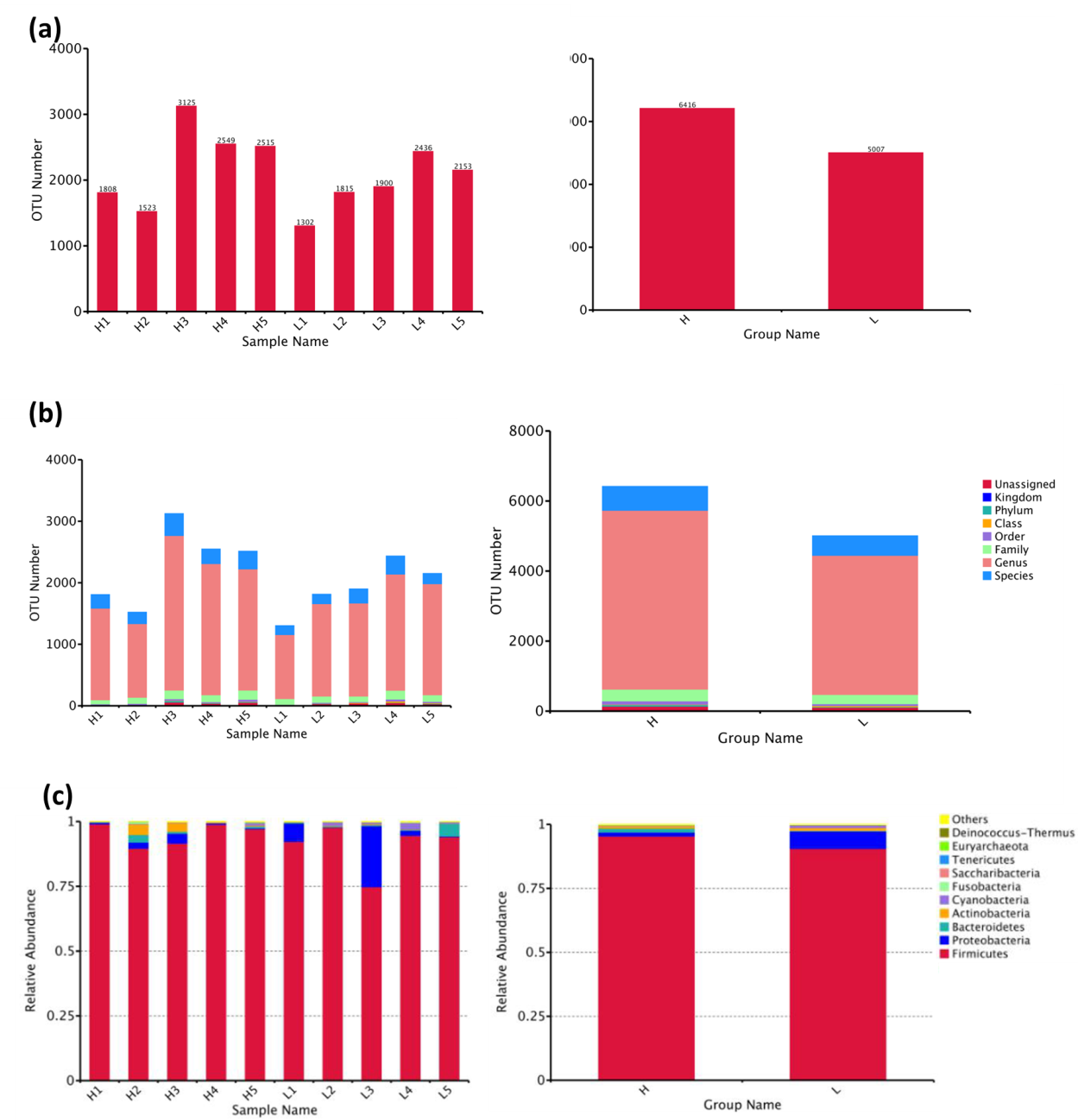
****

****

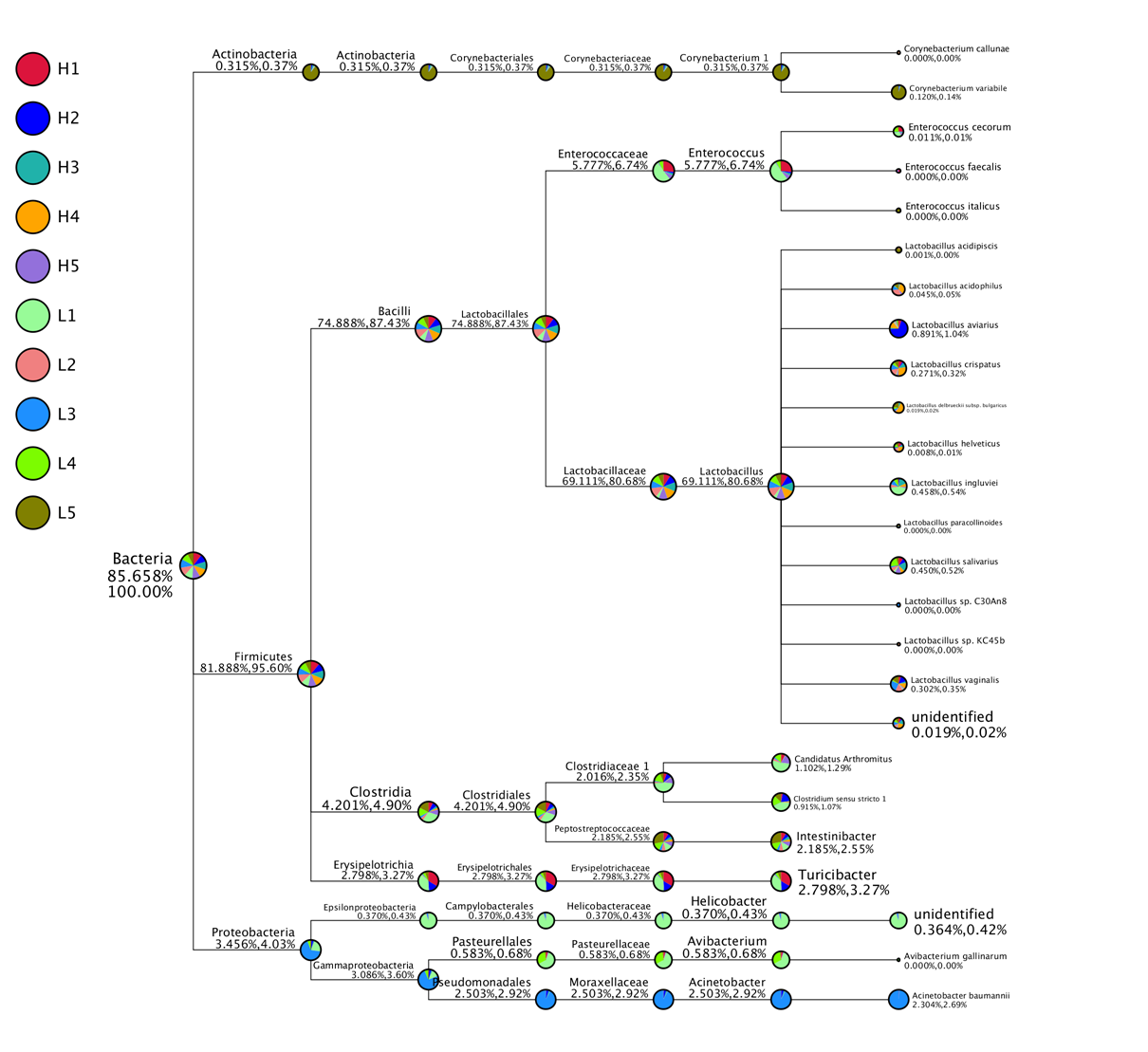
**Figure S2** Data quality control, a and b shown sequencing quality score, length distribution of effective tags, and sequencing error rate of samples in H&L groups of laying hens, respectively. The sequencing quality score of a given base, Q, is defined by the following equation: Q = - 10log10(e), where e is the estimated probability of the base call being wrong. Higher Q scores indicate a smaller probability of error. For example, a quality score of 20 represents an error rate of 1 in 100, with a corresponding call accuracy of 99%, (Samples per group n=5). H: hens of high laying performance; L: hens of low laying performance.



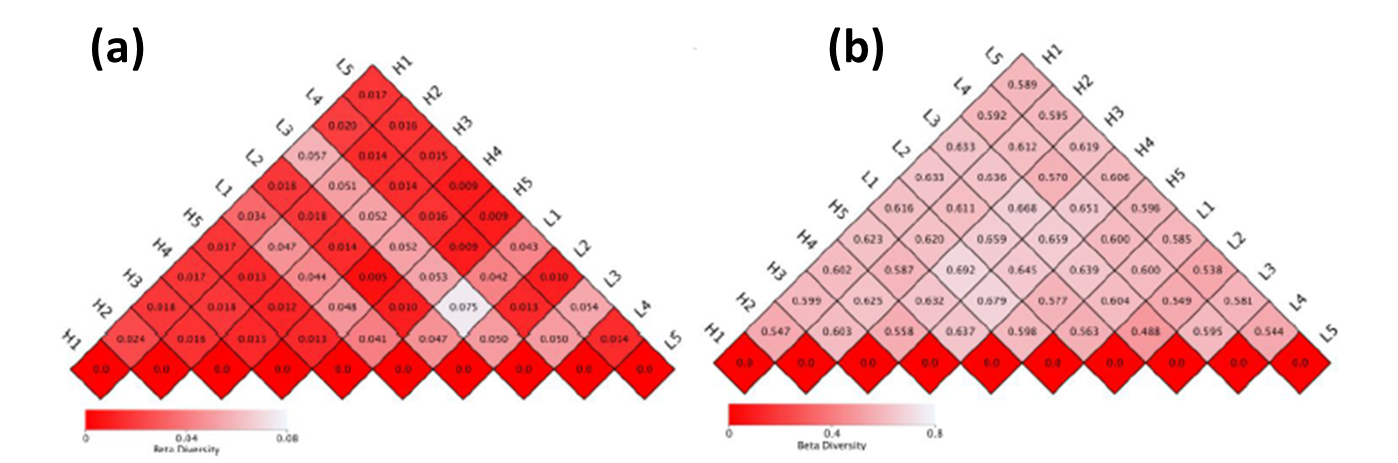
**Figure S3**Krona taxonomy visualization for distribution of the fecal microbiota composite for H&L groups of laying hens, graphs generated by using KRONA program, (Samples per group n=5). H: hens of high laying performance; L: hens of low laying performance.

****

**Figure S4** Statistics differences among samples and between groups of high and low egg performance of laying hens; (a) Operational taxonomic units diversity analysis, (b) Species annotation in different classification levels (Kingdom, Phylum, Class, Order, Family, Genus, Species), (c) Species relative abundance layout, (Samples per group n=5). H: hens of high laying performance; L: hens of low laying performance.

****

**Figure S5** The classification tree for multiple samples based on particular concerned species by default of top 10 genera for each sample for H&L groups of laying hens. The number above (after the taxonomic ranks) represents the relative abundance of the whole corresponding taxon, while the second number represents the relative abundance of the selected of corresponding taxon, (Samples per group n=5). H: hens of high laying performance; L: hens of low laying performance.



**Figure S6** Heatmap of beta diversity indices based on variation coefficient of two samples of high and low egg performance of laying hens, a and b are unweighted unifrac distance and weighted unifrac distance, respectively. The smaller value reveals less variation between the two samples in species diversity exists. The number in the square shows the variation of each pair of samples. The highest value stands for weighted unifrac distance, while the lowest value stands for unweighted unifrac distance, (Samples per group n=5). H: hens of high laying performance; L: hens of low laying performance.