Continuous pre- and post-transplant exposure to a disease-associated gut microbiome promotes hyper-acute graft-versus-host disease in wild-type mice

Kate L Bowerman[1], Antiopi Varelias[2,3], Nancy Lachner[1], Rachel D Kuns[2], Geoffrey R Hill[2,3,4,5], Philip Hugenholtz[1]

[1]Australian Centre for Ecogenomics, School of Chemistry and Molecular Biosciences, The University of Queensland, St Lucia, Queensland, Australia.

[2]QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia.

[3]Faculty of Medicine, The University of Queensland, St Lucia, Queensland, Australia.

[4]Clinical Research Division, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA.

[5]Division of Medical Oncology, University of Washington, Seattle, Washington, USA.

Supplementary figures

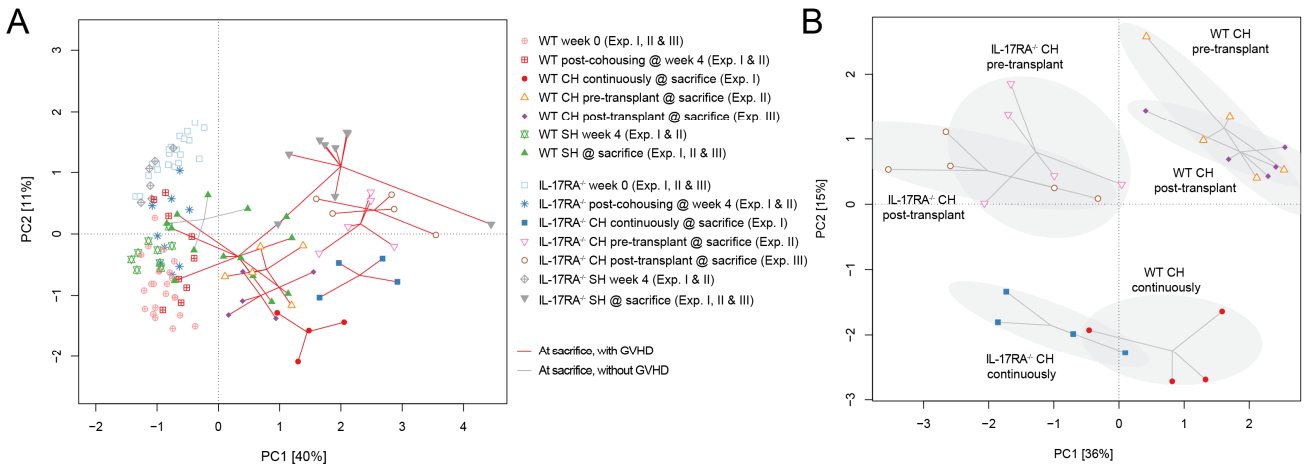

Figure S1. Genome level ordination analysis.

PCA plots based on read counts per genome generated from read mapping to recovered metagenome-assembled genomes plus reference genomes available from NCBI. (A) PCA of all mice at each time point; pre-cohousing, post-cohousing and time of sacrifice. Points representing pre-transplant groups are unconnected. Post-transplant groups sacrificed due to the development of GVHD are connected with a red line. Post-transplant groups that did not develop GVHD (a subset of separately housed WT mice) are connected with a grey line. (B) Independent PCA of all cohoused post-transplant mice. Centroids connect mice from each individual experimental setup. CH: cohoused, SH: separately housed.

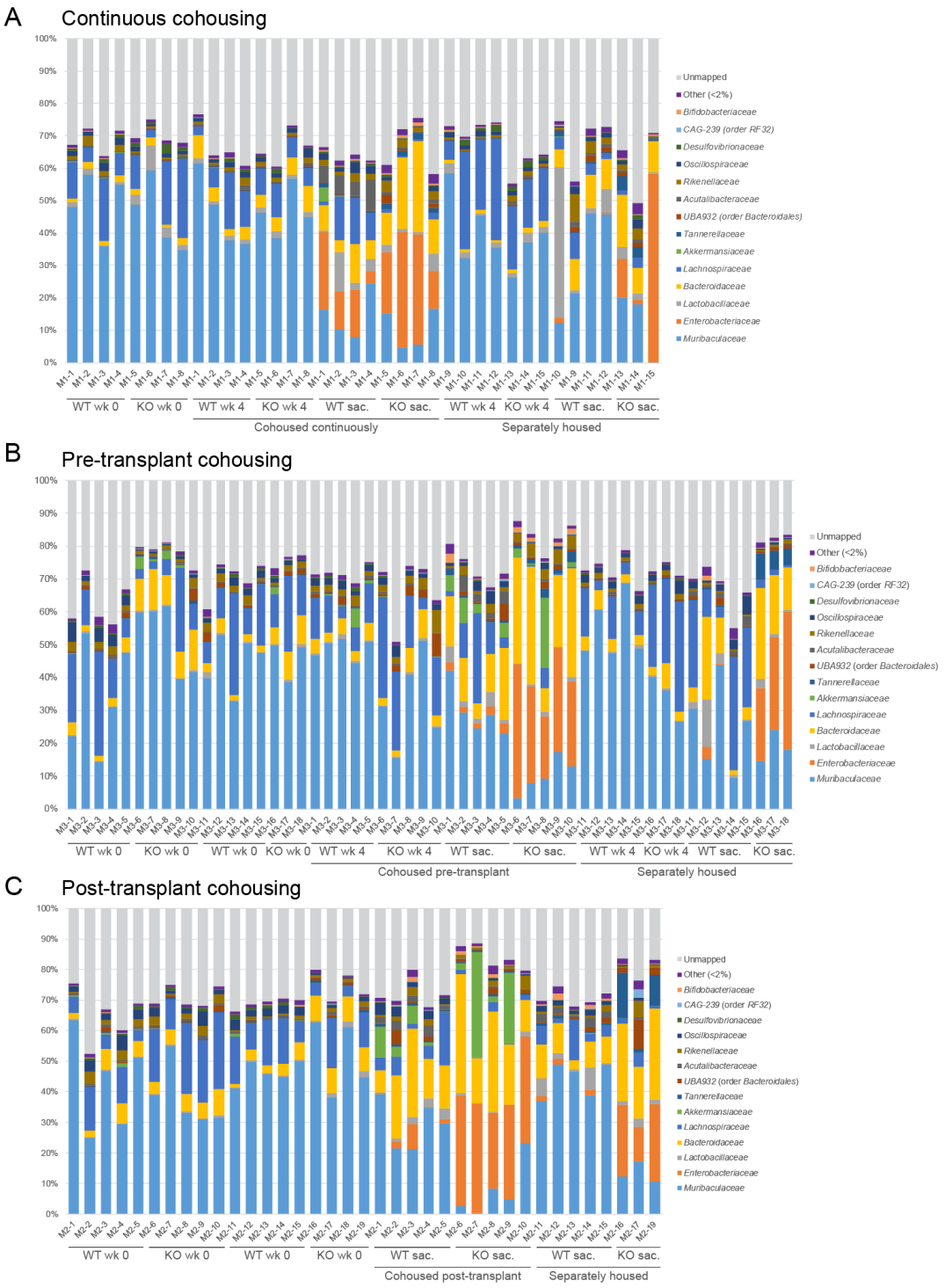

Figure S2. Family level relative abundance

Family level relative abundance based on read counts per genome generated from read mapping to recovered metagenome-assembled genomes plus reference genomes available from NCBI. Each bar represents an individual mouse at each time point: pre-cohousing or pre-transplant as applicable (week 0), post-cohousing (week 4), at time of sacrifice (sac.). (A) Mice cohoused continuously. (B) Mice cohoused pre-transplant only. (C) Mice cohoused post-transplant only. Families with average relative abundance <2% are grouped under Other. Unmapped indicates percentage of reads that did not map to the genome database.

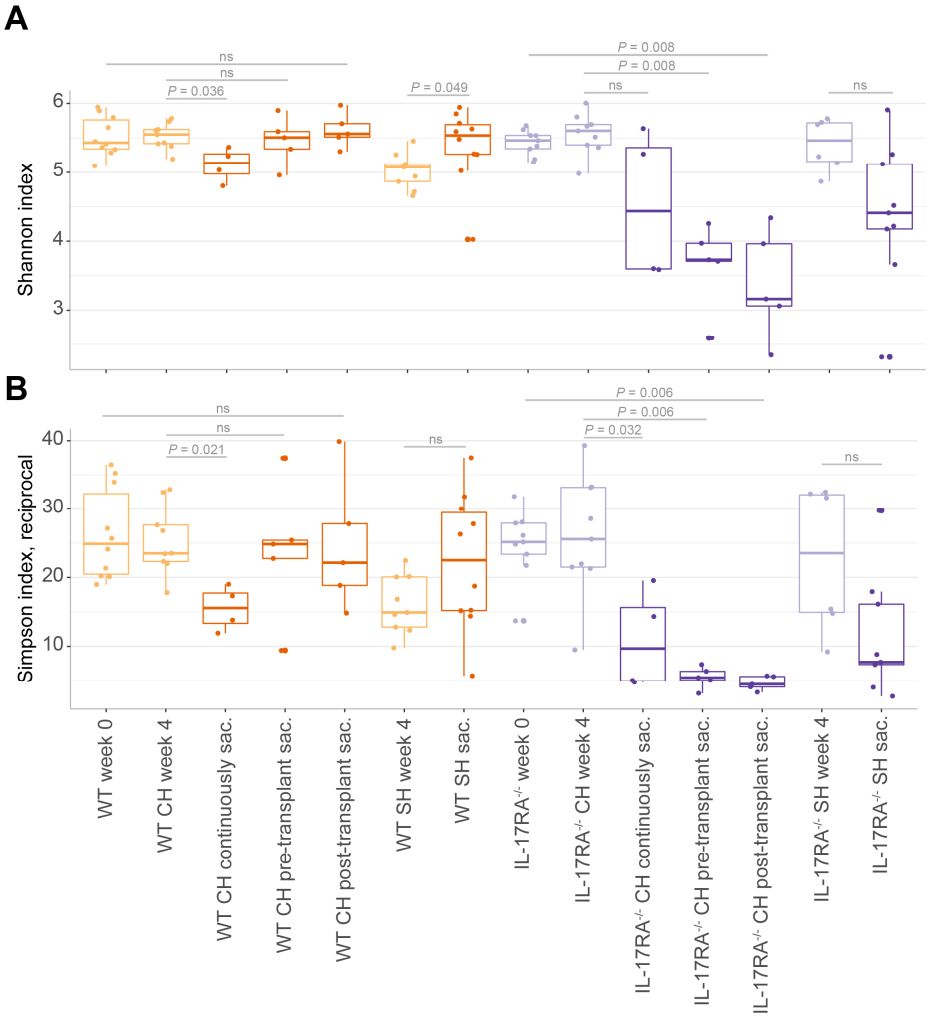

Figure S3. Alpha diversity

Shannon (A) and Simpson (B) indices within each group were calculated using raw read counts from genome mapping normalized for library size using DESeq2 size factor adjustment. Boxes denote first and third quartiles, whiskers extend 1.5x the inter-quartile range. P values are derived from pairwise Wilcoxon rank sum analysis with Benjamini-Hochberg adjustment for multiple comparisons. Statistical results are displayed for pre- vs post-transplant comparisons only.

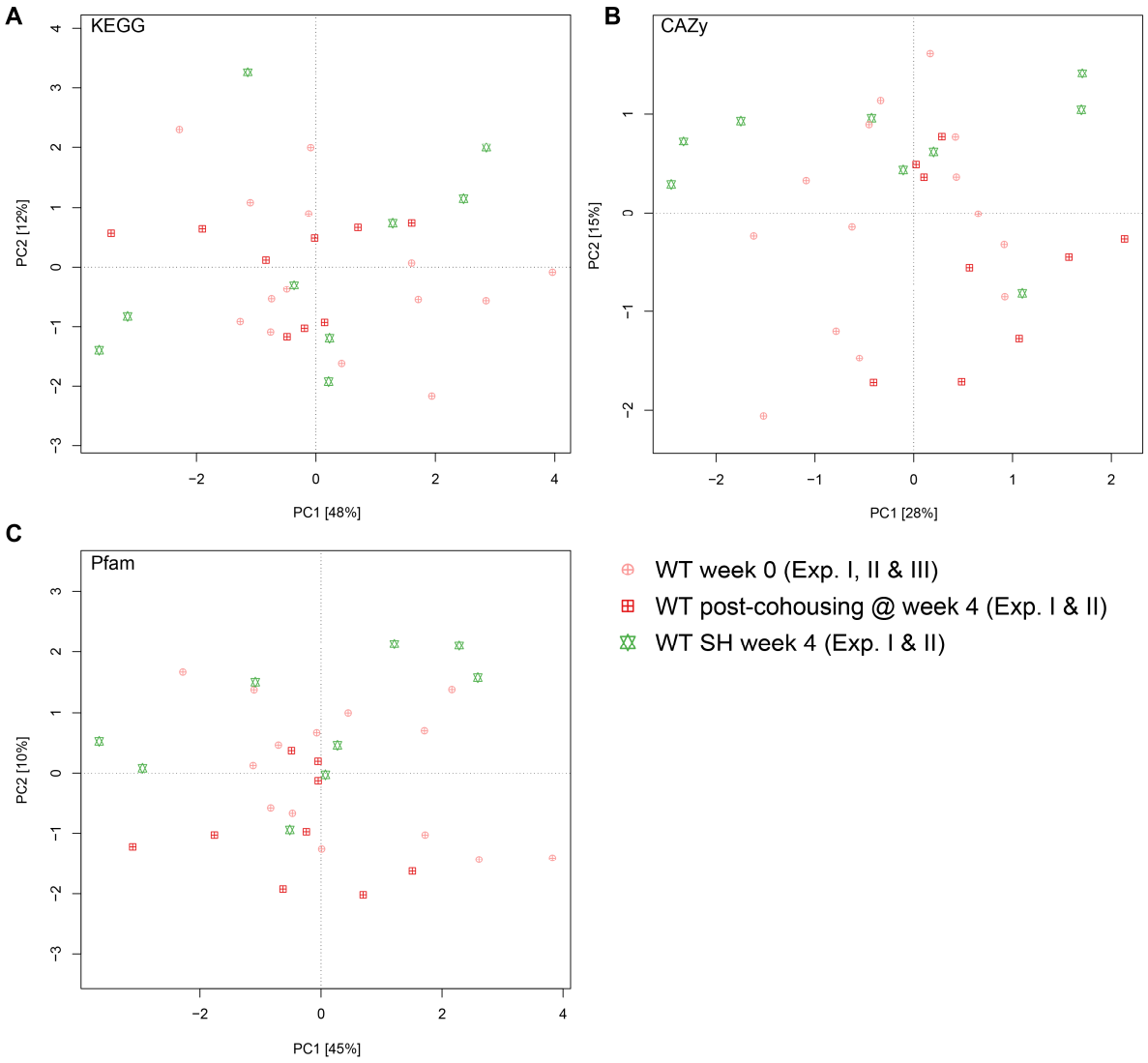

Figure S4. Functional level ordination analysis.

Comparison of pre-transplant WT mice at the functional level using mice from Experiments I, II and III. Week 0 mice include pre-cohousing WT mice from Experiments I and II and pre-transplant WT mice from Experiment III. Principal component analysis undertaken using read counts from annotation of raw reads using functional databases KEGG (A), CAZy (B) and Pfam (C).

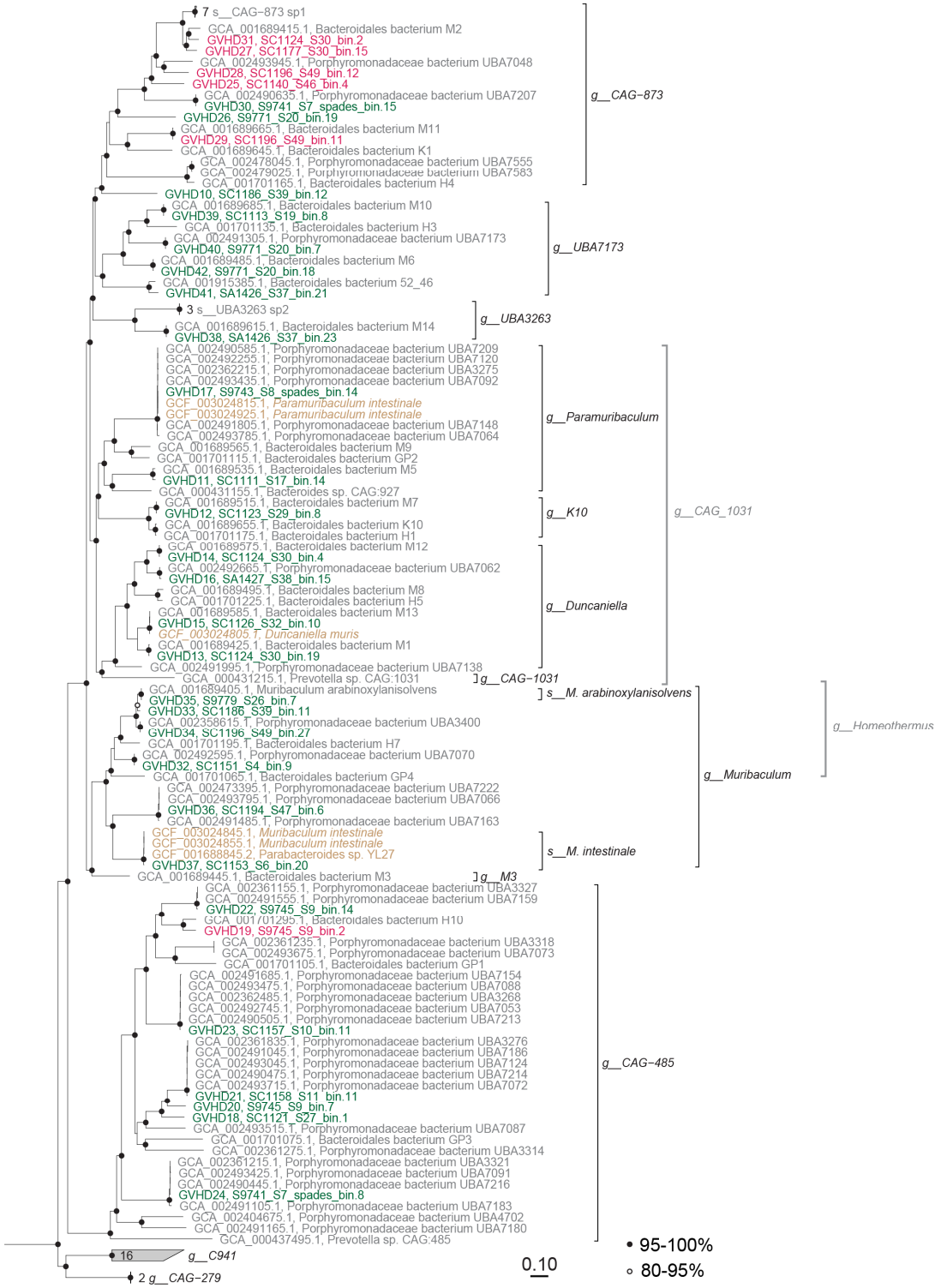

Figure S5. *Muribaculaceae* genome tree

Maximum-likelihood phylogenetic tree constructed using 120 concatenated single copy marker genes. Genomes included were publicly available from NCBI (grey) and MAGs recovered from this study (green). MAGs in pink are those enriched or depleted during cohousing. Genomes in brown are derived from cultured isolates. Taxonomy in black is consistent with release 04-R89 of the Genome Taxonomy Database. Taxonomy in grey is consistent with release 03-R86.

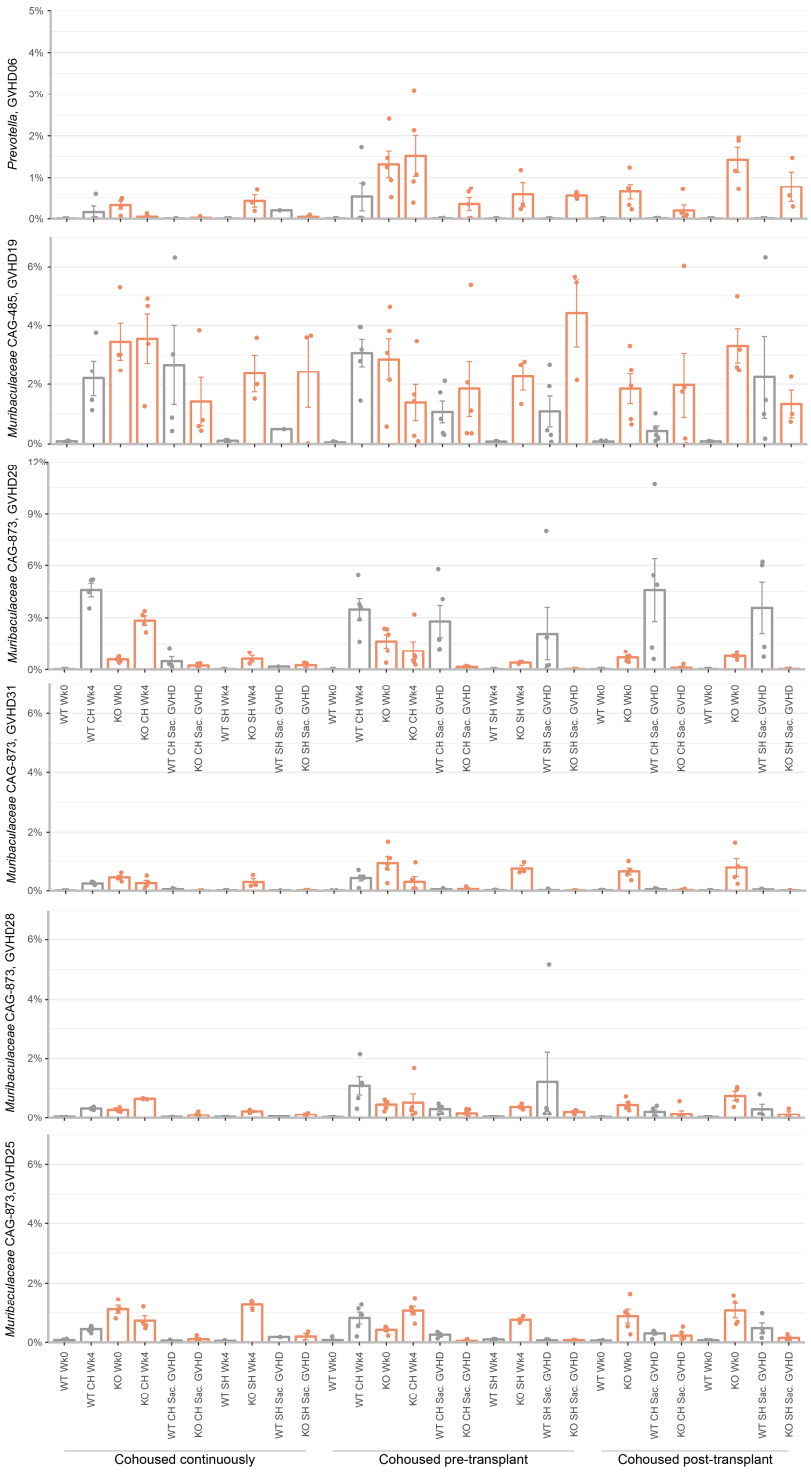

Figure S6. Genome relative abundance

Relative abundance of select species displaying consistently increased abundance in WT mice during cohousing pre-transplant (Experiments I & II). Relative abundance values are based on read mapping to genome database including recovered MAGs and publicly available genomes from NCBI.

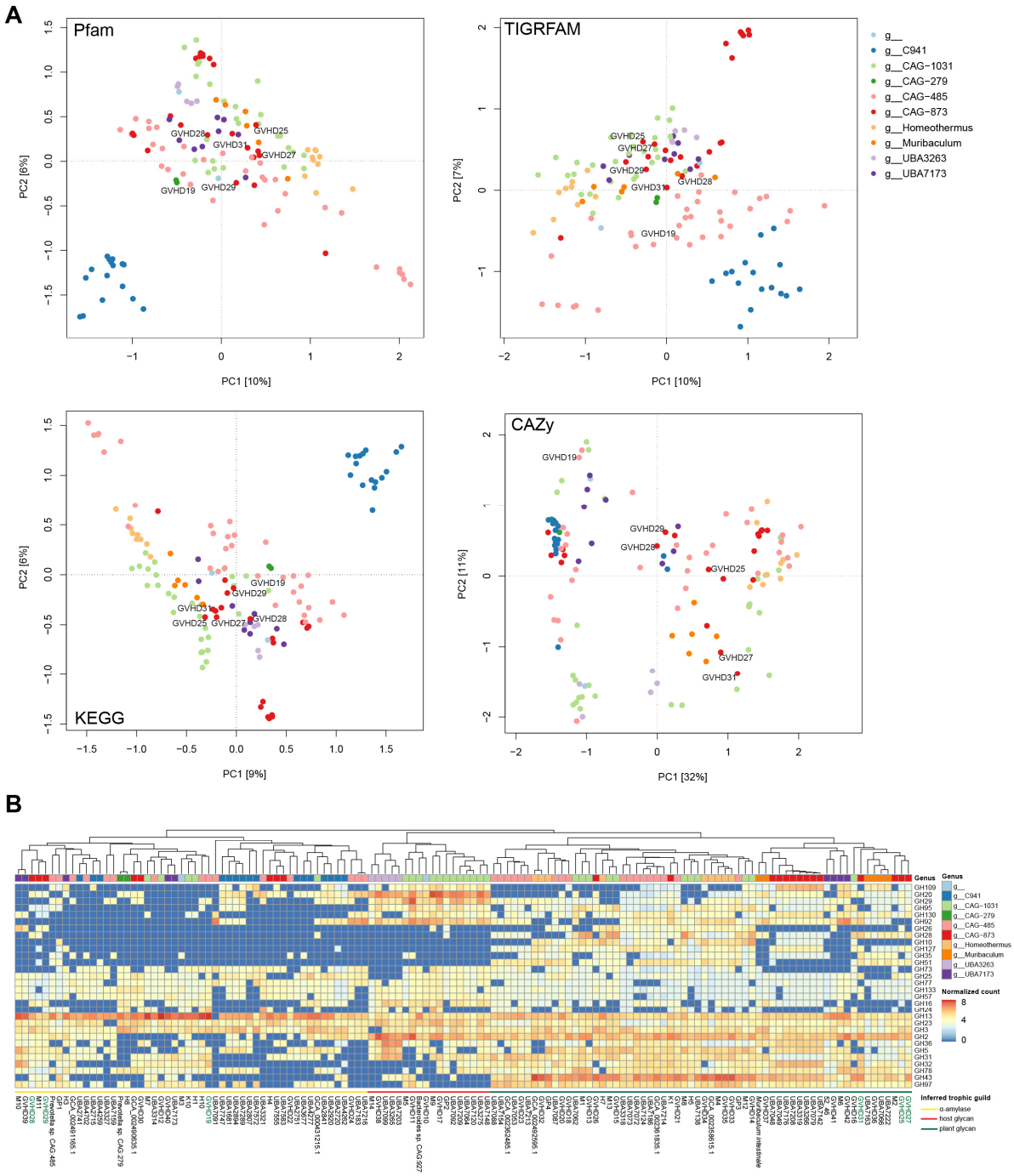

Figure S7. Functional annotation of *Muribaculaceae*.

Comparison of *Muribaculaceae* genomes enriched in WT mice during cohousing with other genomes from family. Proteins from recovered genomes and publicly available genomes from the family *Muribaculaceae* were annotated with functional domains from Pfam, TIGRFAM, KEGG and CAZy databases. (A) PCA plots for each database produced from domain counts per genome. Enriched and depleted MAGs are annotated in each plot. (B) Heatmap of enzyme counts within top 30 glycoside hydrolase categories for recovered MAGs and public genomes. Green coloured genomes are those enriched or depleted during cohousing pre-transplant. Coloured bar indicates inferred trophic guild based on CAZy profile as per Ormerod et al 2016. Genera were defined using GTDB release 03-R86.

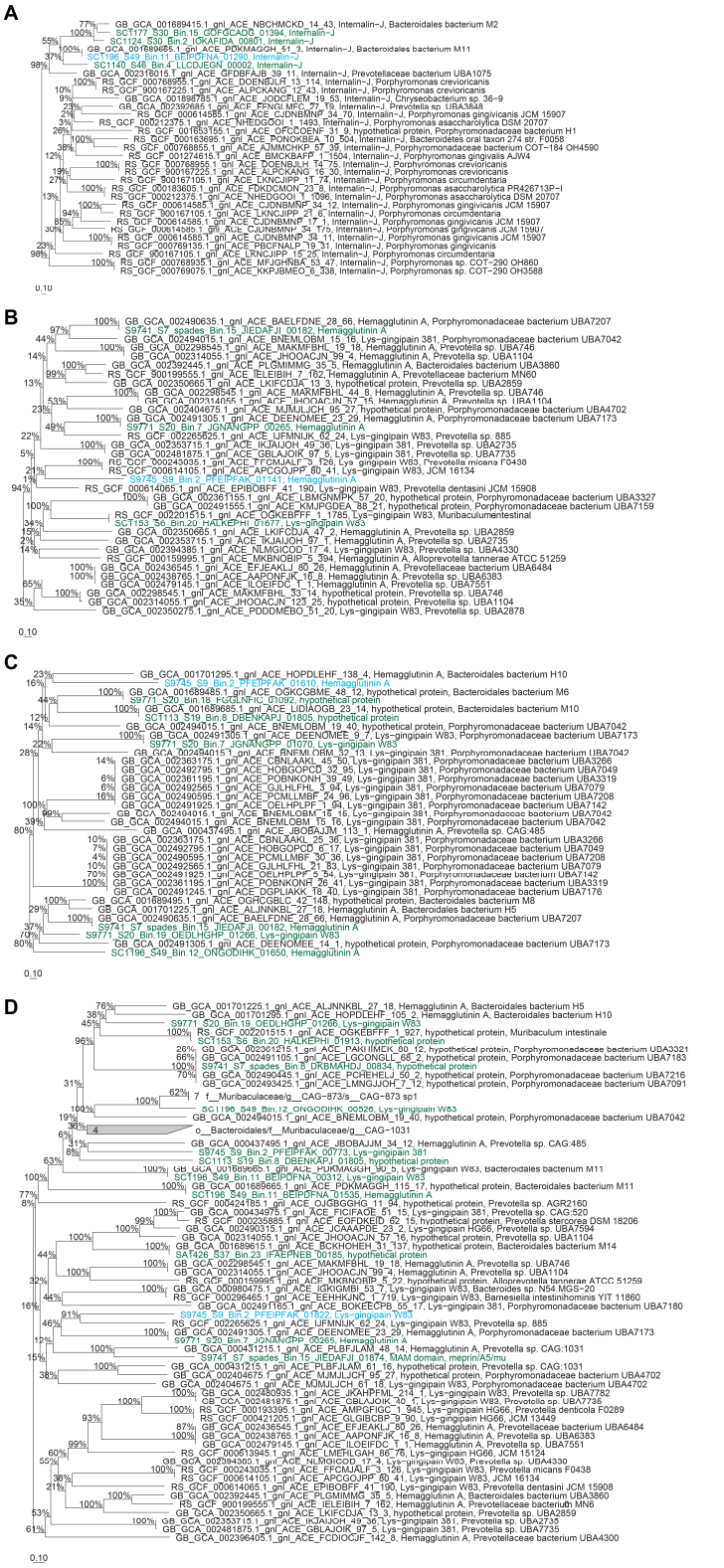

Figure S8. Gene trees

Gene trees constructed from BLAST search of recovered MAGs and genomes from NCBI (GTDB release 03-RS86). Seed protein for each search marked in blue, search result proteins from MAGs marked in green. (A) Putative internalin-J protein from GVHD29. (B-D) Putative hemagglutinin A proteins from GVHD19.

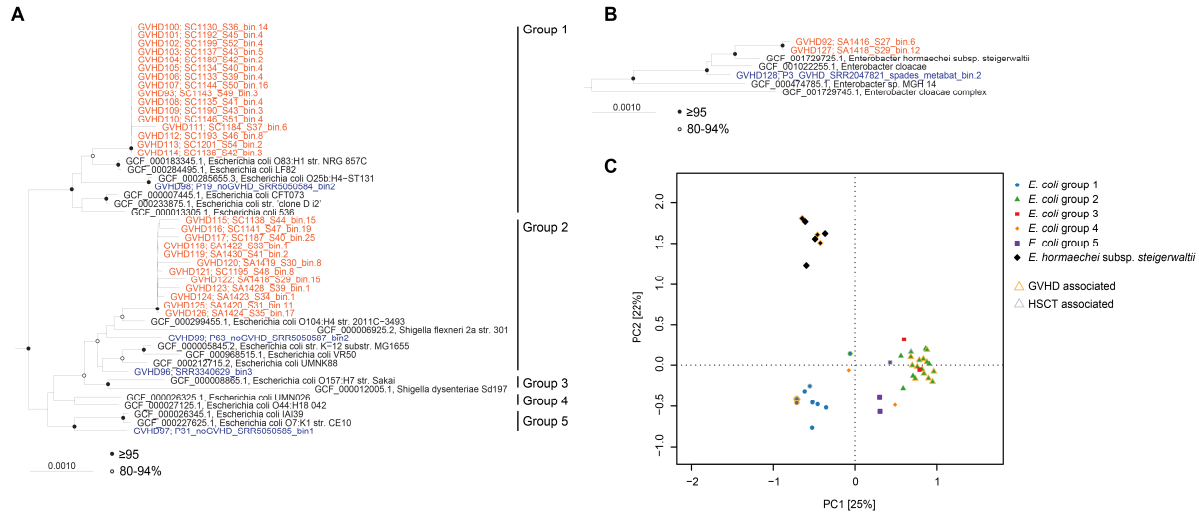

Figure S9. *Escherichia coli* and *Enterobacter hormaechei* subsp. *steigerwaltii* strain analysis

(A&B) Maximum-likelihood phylogenetic trees constructed using 120 concatenated single copy marker genes for E. coli (A) and *E. hormaechei* subsp. *steigerwaltii* (B). Genomes included were publicly available from NCBI (black), MAGs recovered from this study (orange) and from datasets available from NCBI SRA (blue). (C) PCA of virulence factor gene profiles of genomes from A & B.

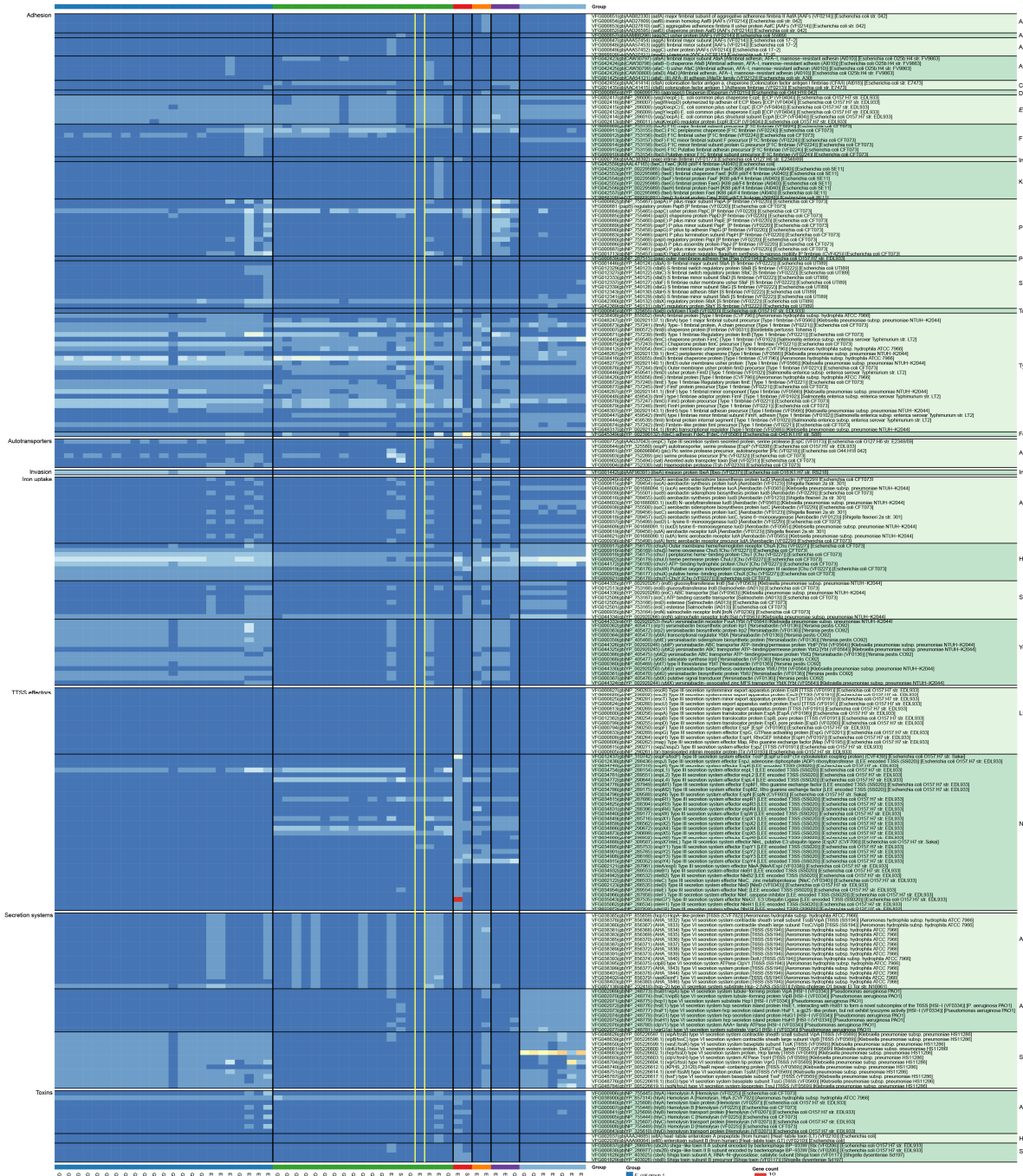

Figure S10. Virulence factor profile of *Escherichia coli* and *Enterobacter hormaechei* subsp. *steigerwaltii* strains and reference genomes.

Virulence profiles determined using BLAST against the VFDB. Genomes included are those from **Fig. S9A & B** in the same order as presented in the respective trees. Annotation groups are as per the VFDB. Yellow vertical lines mark *E. coli* K-12.