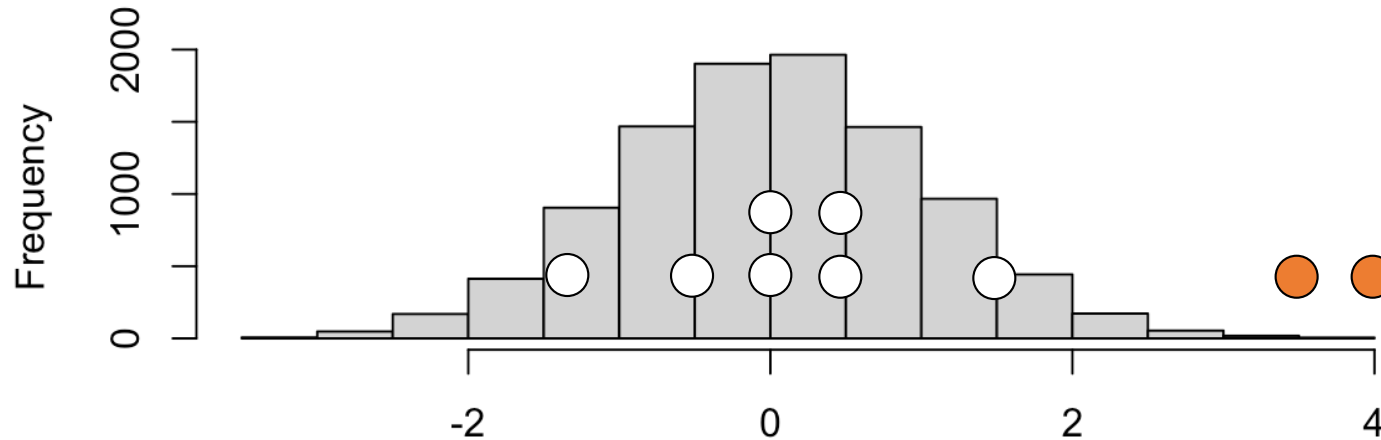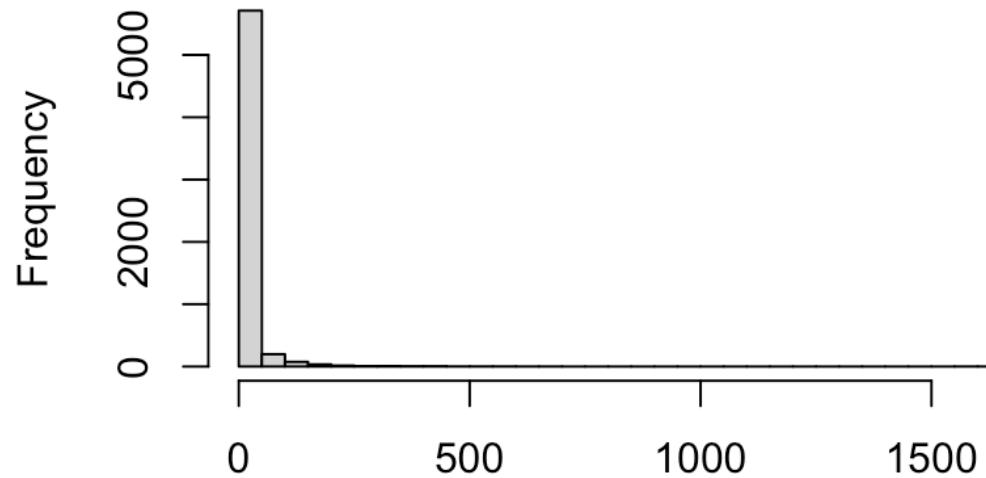# Statistical analyses:
Data assumptions

- Parametric statistical tests assume normally distributed data:

- This allows the test to 'infer' missing data for groups with low N
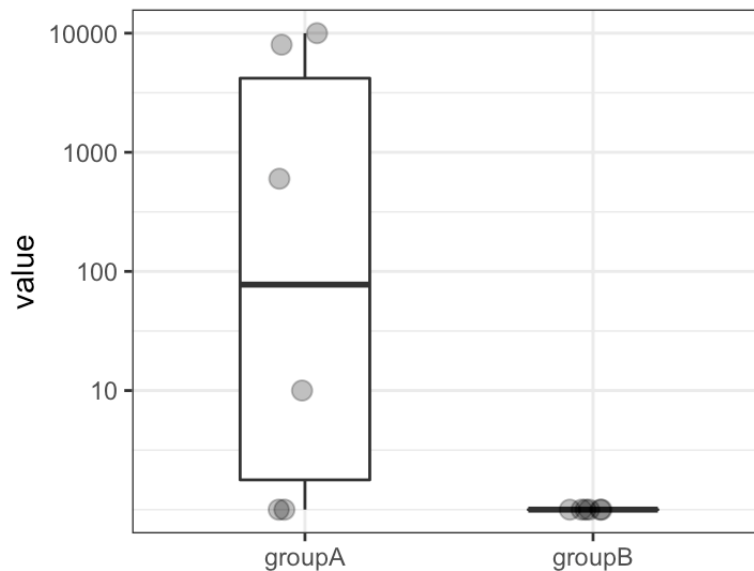
Statistical analyses:
Data assumptions

- Parametric statistical tests assume normally distributed data
- Microbiome data are not normally distributed:

Statistical analyses:
Data assumptions

- Parametric statistical tests assume normally distributed data
- Microbiome data are not normally distributed:

T-test P = 0.161
(parametric)

Mann-Whitney P = 0.028
(non-parametric)

Non-parametric tests with dichotomous variables:

- Goal: Identify taxa that are in differential abundance between two groups of samples
- Method: make no assumptions about data distribution using non-parametric statistics
  - non-parametric statistical test for each taxon, successively
    - ex. Mann-Whitney U test/Wilcoxon rank-sum test
  - then perform multiple comparisons corrections
    - ex. Benjamini-Hochberg false discovery rate calculation

Non-parametric tests with dichotomous variables: Results

| Row.names	wilx_P	wilx_stat	log2fc	numnonzero	rank	BH_Q	Kingdom	Phylum	Class	Order	Family	Genus
0f4cbad3b65eeea78a91f7bf9e73c5e4	0.001018407	0.0	-Inf	26	16	0.0952019	Bacteria	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	Anaerostipes
6598140bdeb7355cf46e0b33c25e9ff0	0.005458783	50.0	Inf	7	46	0.0952019	Bacteria	Firmicutes	Bacilli	Lactobacillales	Lactobacillaceae	Lactiplantibacillus
66d9c2e13cb8020d78bdf39cae8f3440	0.005458783	50.0	Inf	9	25	0.0952019	Bacteria	Firmicutes	Bacilli	Lactobacillales	Lactobacillaceae	Pediococcus
884cb13b27b7dd2a3750d189c988f647	0.005458783	50.0	Inf	6	19	0.0952019	Bacteria	Proteobacteria	Gammaproteobacteria	Enterobacterales	Enterobacteriaceae	Klebsiella
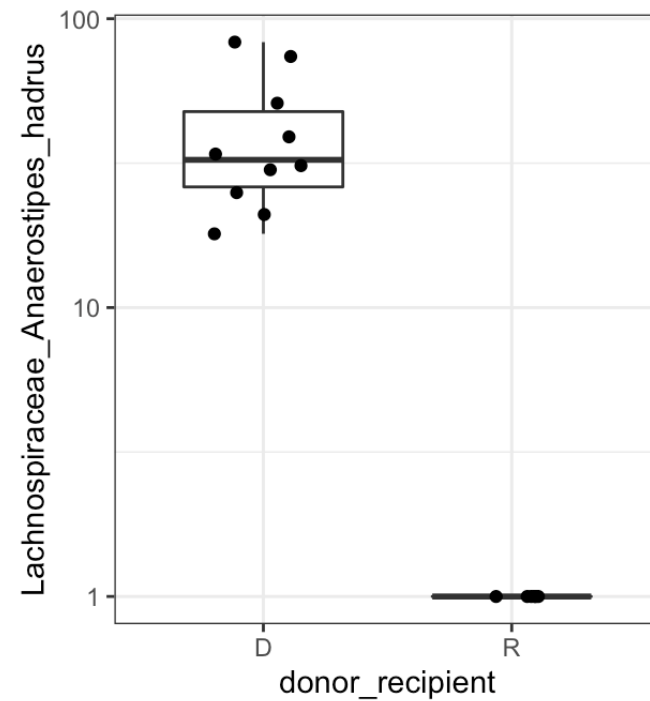
Controlling for confounding variables in group comparisons: linear mixed effects models

- Goal: Identify taxa that are in differential abundance between two groups of samples and control for confounding variables
- Method: linear mixed effects models
  - Requires linear variables (ideally normally distributed) → convert data to 'linear'-esque space using transformations
    - Ex. limma, arcsin square root, log with pseudocount
  - Perform linear mixed effects with covariates as fixed/random effects
  - Perform multiple comparisons correction
    - ex. Benjamini-Hochberg false discovery rate calculation (FDR 'Q' value)

Controlling for confounding variables in group comparisons: linear mixed effects models

Row.names	P	t	Kingdom	Phylum	Class	Order	Family	Genus	Species	BHq
0f4cbad3b65eeea78a91f7bf9e73c5e4	1.010065e-10	-19.788603	Bacteria	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	Anaerostipes	hadrus	1.494897e-08
e5ef806843f7664da2a1b26dc23e13c1	3.522567e-04	-4.791362	Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	Phocaeicola	<NA>	2.478389e-02
055c4b7300665006464d7f8e7a0214c957	8.783672e-04	-4.372881	Bacteria	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	Fusicatenibacter	saccharivorans	2.478389e-02
c867a71e863b24c0a5d422dd6d72e02d	8.938139e-04	-4.369494	Bacteria	Firmicutes	Negativicutes	Acidaminococcales	Acidaminococcaceae	Phascolarctobacterium	faecium	2.478389e-02
4d51427c6465d3c97af11af0edd132be	1.073987e-03	-4.275329	Bacteria	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	Agathobacter	<NA>	2.478389e-02
c322e6afe2a271d465ca4bd5971f739a	1.118331e-03	-4.160847	Bacteria	Firmicutes	Clostridia	Clostridiales	Ruminococcaceae	Faecalibacterium	<NA>	2.478389e-02
dd0234a1d48f74a011f58a58b206e6ad	1.172211e-03	-4.208625	Bacteria	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	Blautia	<NA>	2.478389e-02
6598140bdeb7355cf46e0b33c25e9ff0	1.370366e-03	4.052329	Bacteria	Firmicutes	Bacilli	Lactobacillales	Lactobacillaceae	Lactiplantibacillus	<NA>	2.535177e-02
66d9c2e13cb8020d78bdf39cae8f3440	1.644626e-03	3.955422	Bacteria	Firmicutes	Bacilli	Lactobacillales	Lactobacillaceae	Pediococcus	<NA>	2.704496e-02
db5b82bfeecda9de9a33e7f4db90ee7f	1.974811e-03	3.858670	Bacteria	Firmicutes	Negativicutes	Veillonellales	Veillonellaceae	Megasphaera	micronuciformis	2.922720e-02
d5d0e236ef6bc5291a873c4a842795ac	2.400135e-03	-3.825696	Bacteria	Firmicutes	Clostridia	Clostridiales	Peptostreptococcaceae	Intestinibacter	bartlettii	3.229272e-02
884cb13b27b7dd2a3750d189c988f647	2.904909e-03	3.655812	Bacteria	Proteobacteria	Gammaproteobacteria	Enterobacterales	Enterobacteriaceae	Klebsiella	<NA>	3.433790e-02
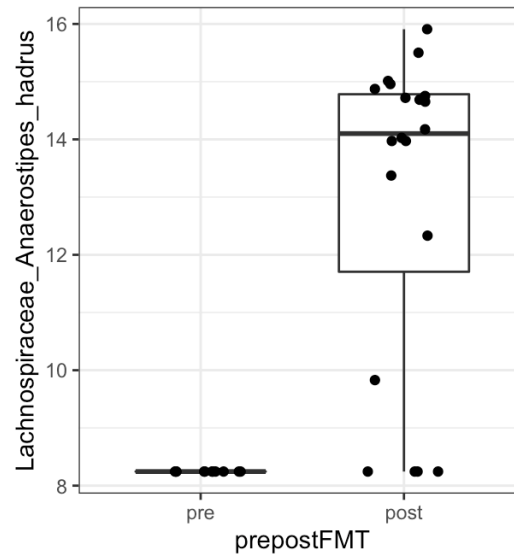
Controlling for intra-individual variance in longitudinal comparisons: linear mixed effects models

- Goal: Identify taxa that are in differential abundance between groups of samples in a longitudinal study

- Method: Control for intra-individual covariance using linear mixed effects models
  - Requires linear variables (ideally normally distributed), convert to 'linear'-esque space using transformations
    - Ex. limma, arcsin square root, log with pseudocount
  - Perform linear mixed effects with subject ID as a 'random effect'
  - Perform multiple comparisons correction
    - ex. Benjamini-Hochberg false discovery rate calculation (FDR 'Q' value)

Controlling for intra-individual variance in longitudinal comparisons: linear mixed effects models

Row.names	P	t	Kingdom	Phylum	Class	Order	Family	Genus	Species	famgenspec	BHq
9bb453680381a926dfca5f9e44c697fa	8.519397e-09	8.625376	Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	Bacteroides	ovatus	Bacteroidaceae_Bacteroides_ovatus	9.626918e-07
dd0234a1d48f74a011f58a58b206e6ad	1.864890e-08	8.222640	Bacteria	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	Blautia	<NA>	Lachnospiraceae_Blautia_NA	1.053663e-06
0f4cbad3b65eeea78a91f7bf9e73c5e4	6.152660e-08	7.739555	Bacteria	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	Anaerostipes	hadrus	Lachnospiraceae_Anaerostipes_hadrus	2.317502e-06
055c4b730066500064d7f8e7a0214c957	3.315587e-07	6.984980	Bacteria	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	Fusicatenibacter	saccharivorans	Lachnospiraceae_Fusicatenibacter_saccharivorans	9.366534e-06
e5ef806843f7664da2a1b26dc23e13c1	9.808455e-07	6.500447	Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	Phocaeicola	<NA>	Bacteroidaceae_Phocaeicola_NA	2.216711e-05
6598140bdeb7355cf46e0b33c25e9ff0	1.245175e-06	-6.400116	Bacteria	Firmicutes	Bacilli	Lactobacillales	Lactobacillaceae	Lactiplantibacillus	<NA>	Lactobacillaceae_Lactiplantibacillus_NA	2.345080e-05
80f715cf372b40bf5f783d6f3dec9210	3.089105e-06	6.041145	Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	Bacteroides	<NA>	Bacteroidaceae_Bacteroides_NA	4.986698e-05
db5b82bfeecda9de9a33e7f4db90ee7f	5.941851e-06	-5.760539	Bacteria	Firmicutes	Negativicutes	Veillonellales	Veillonellaceae	Megasphaera	micronuciformis	Veillonellaceae_Megasphaera_micronuciformis	8.392864e-05
a0531d77346b0efcc8bfa411fcebe945	8.687391e-06	5.622630	Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Rikenellaceae	Alistipes	putredinis	Rikenellaceae_Alistipes_putredinis	1.016041e-04
7a48aa7f3e7e5985addac38ece2de88f	8.991517e-06	5.609260	Bacteria	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	Blautia	obeum	Lachnospiraceae_Blautia_obeum	1.016041e-04
4d51427c6465d3c97af11af0edd132be	4.257686e-05	5.031517	Bacteria	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	Agathobacter	<NA>	Lachnospiraceae_Agathobacter_NA	4.269406e-04
d9104b547cf822e787d4e09e8c5be6cb	4.885754e-05	4.932664	Bacteria	Actinobacteria	Coriobacteriia	Coriobacteriales	Coriobacteriaceae	Collinsella	aerofaciens	Coriobacteriaceae_Collinsella_aerofaciens	4.269406e-04
884cb13b27b7dd2a3750d189c988f647	4.911705e-05	-4.925839	Bacteria	Proteobacteria	Gammaproteobacteria	Enterobacterales	Enterobacteriaceae	Klebsiella	<NA>	Enterobacteriaceae_Klebsiella_NA	4.269406e-04