Outline

- What is metagenomics & why do we care?

- The opportunity of metagenomics & metagenomic prediction

- Future perspectives

- Challenges & ongoing research questions
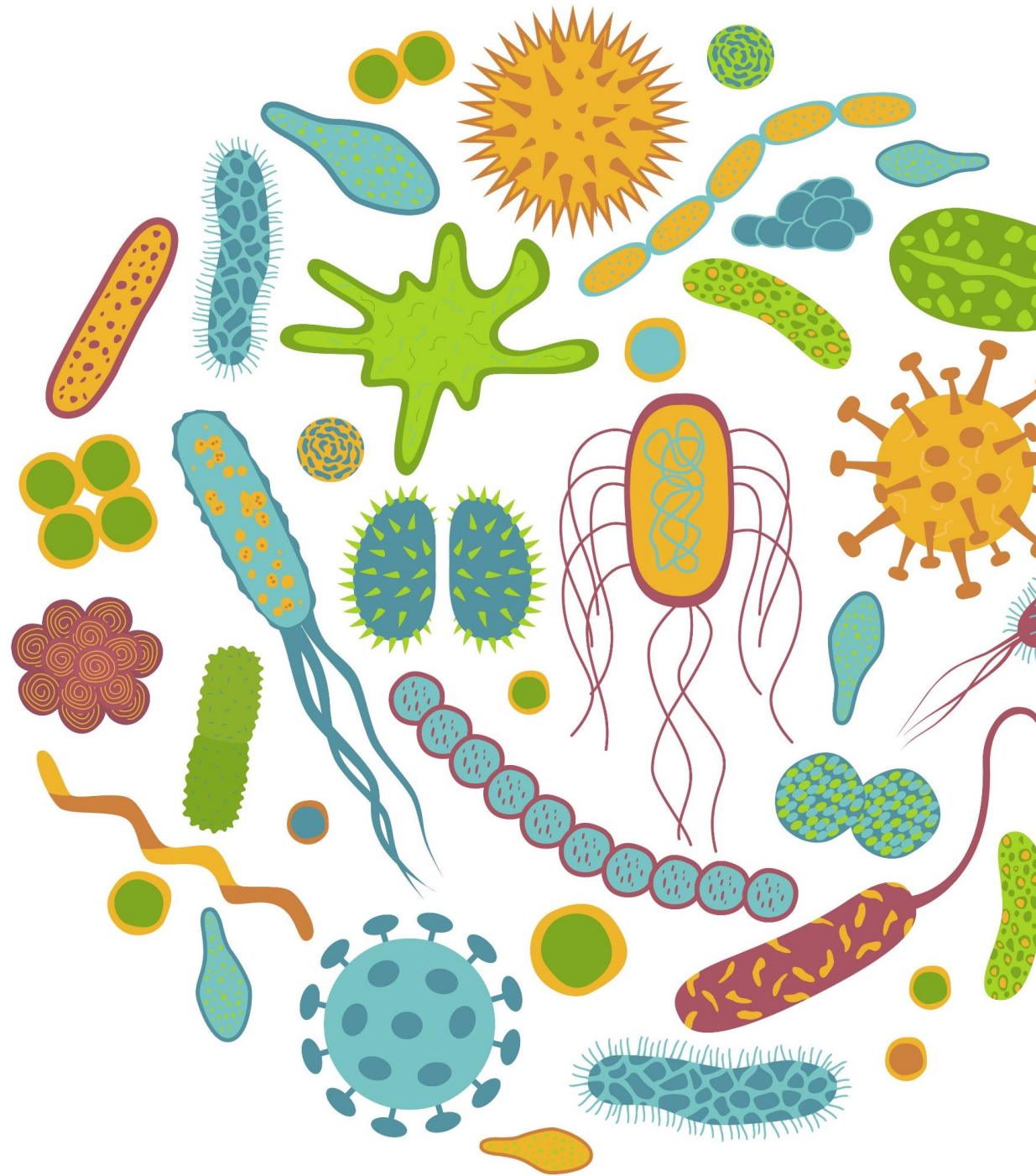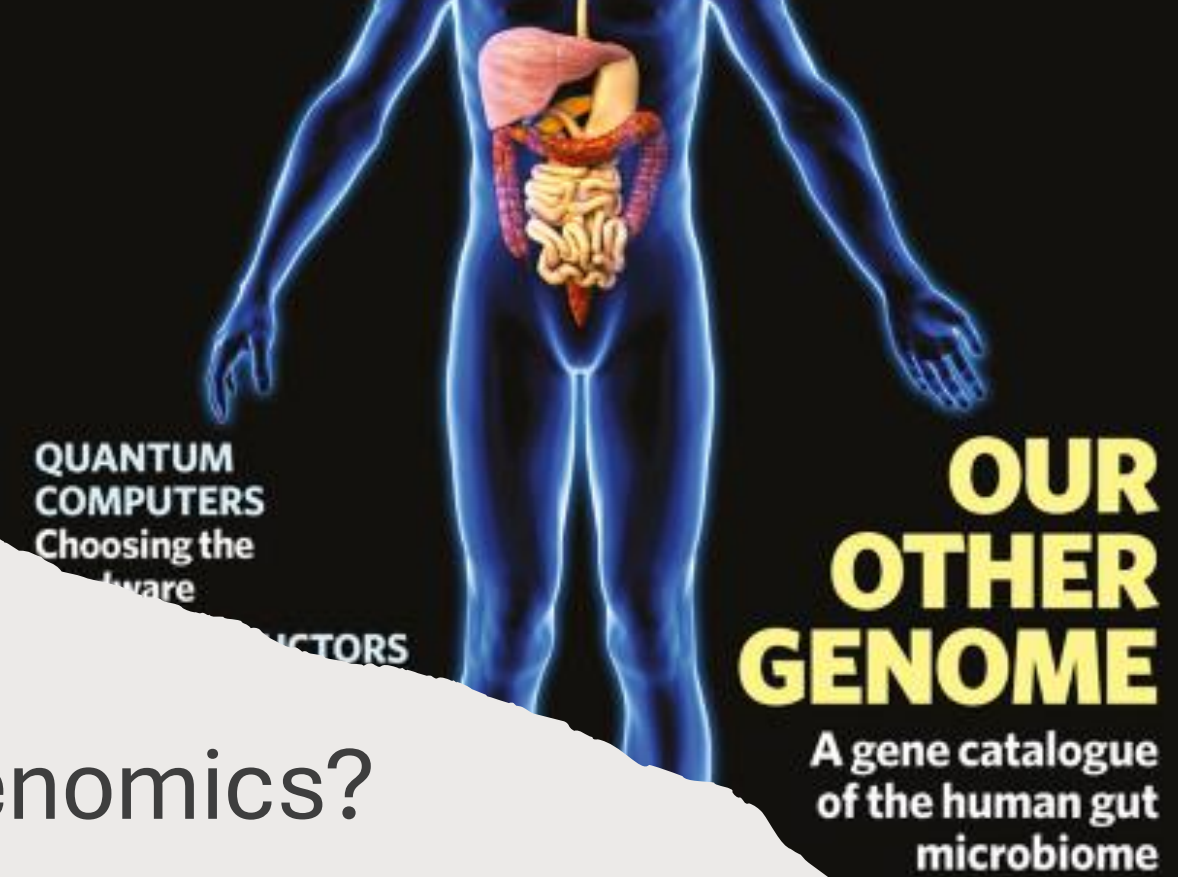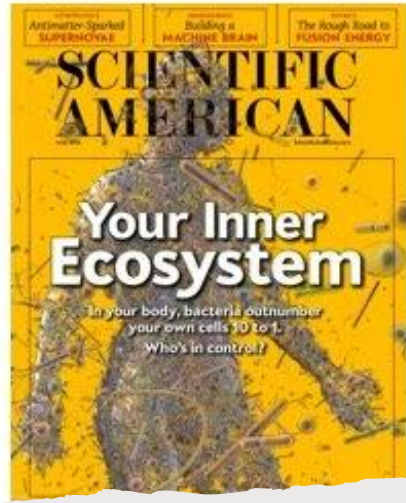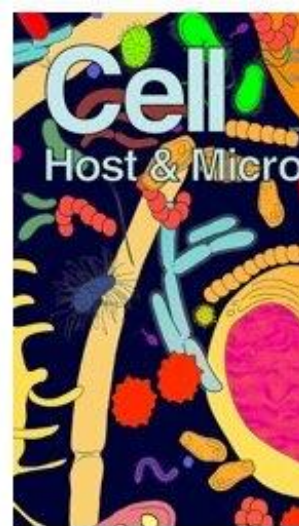
What is metagenomics?
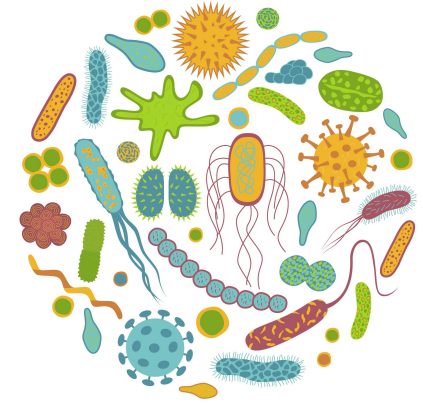
Definitions:

• *Microbiome* – a community of microbes

• *Metagenome* – the cumulative genomes of cells which make up the microbiome

• *Metagenomics* – is the study of the genomes of that microbial population

Why do we care about metagenomics?

Why do we care about metagenomics?

It influences so many things!
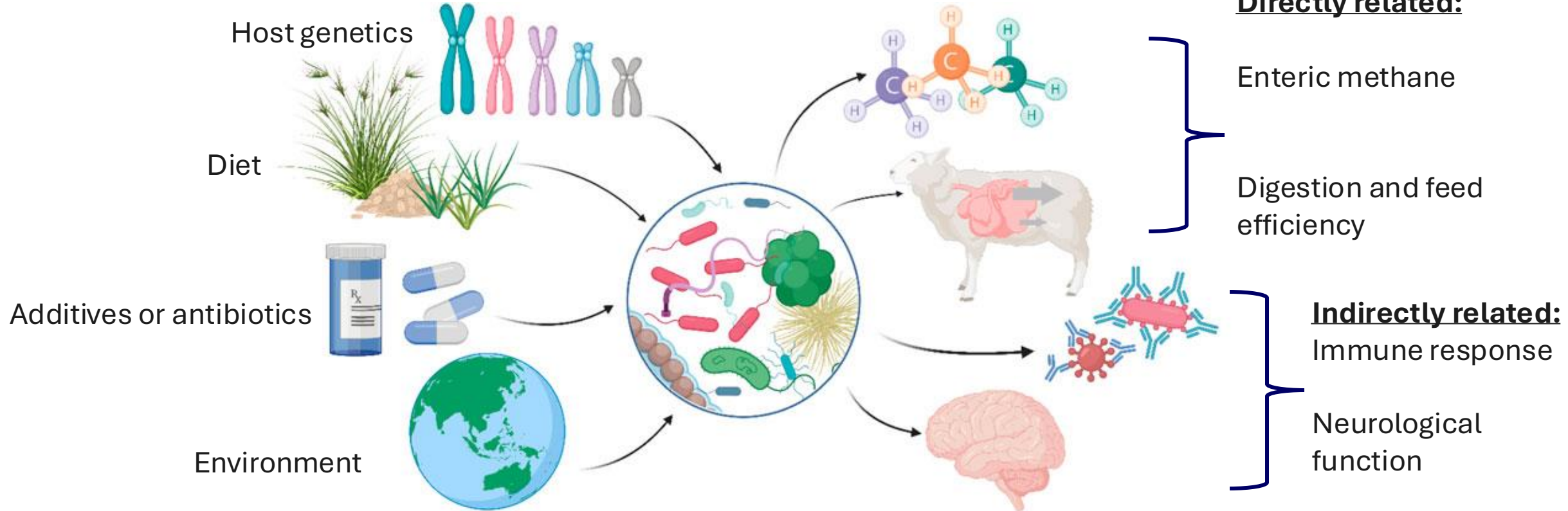
Figure: Ross & Hayes, Front Genet, 2022

Why do we care about metagenomics?

- Rumen metagenome: (summarized in Mizrahi and Jami, 2018)
  - Feed efficiency and RFI
  - Enteric methane emissions
  - Milk fat and yield

- Human gut metagenome: (summarized in Kho and Lal, 2018)
  - IBD
  - Celiac disease
  - Obesity
  - Infectious disease

- Saliva metagenome: (Alqedari et al, 2024)
  - Covid-19 severity

Why do we care about metagenomics?

- *Microbiability (m^2)* – proportion of phenotypic variance of the trait that is explained by between-animal differences in the microbial community
  - Difford, Lassen & Løvendahl, 2016

- Cattle CH_4 – 0.15 (Difford et al., 2018)
- Swine feed conversion ratio – 0.20 (Aliakbari et al., 2022)
- Swine back fat – 0.40 (He et al. 2022)
- Lamb live weight – 0.33 (Hess et al., 2023)
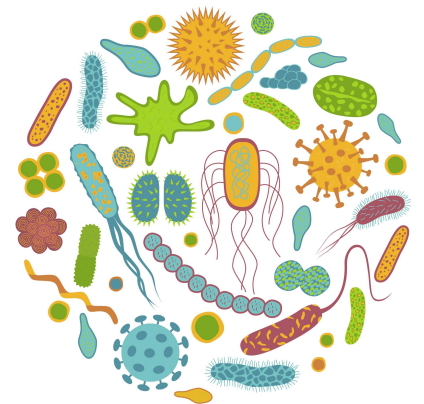
Why do we care about metagenomics?

- There is some component of host control of the microbiome that is independent of diet/environment

A heritable subset of the core rumen microbiome dictates dairy cow productivity and emissions

R. JOHN WALLACE (iD), GOOR SASSON, PHILIP C. GARNSWORTHY (iD), ILMA TAPIO (iD), EMMA GREGSON, PAOLO BANI (iD), PEKKA HUHTANEN (iD), ALI R. BAYAT (iD), FRANCESCO STROZZI, [...], AND ITZHAK MIZRAHI (iD)   +23 authors   Authors Info & Affiliations

Bovine host genome acts on rumen microbiome function linked to methane emissions

- Martínez-Álvaro, et al. Commun Biol, 2022
- 17% of rumen microbial genera had significant host genomic effects (h^2 = 0.13-0.61)
- 29 microbial genera host-genomically correlated with methane emissions (r = 0.59-0.93)

Bovine host genome acts on rumen microbiome function linked to methane emissions

- Martínez-Álvaro, et al. Commun Biol, 2022

Highlights the strength of a common host genomic control of specific microbial processes and CH$_4$
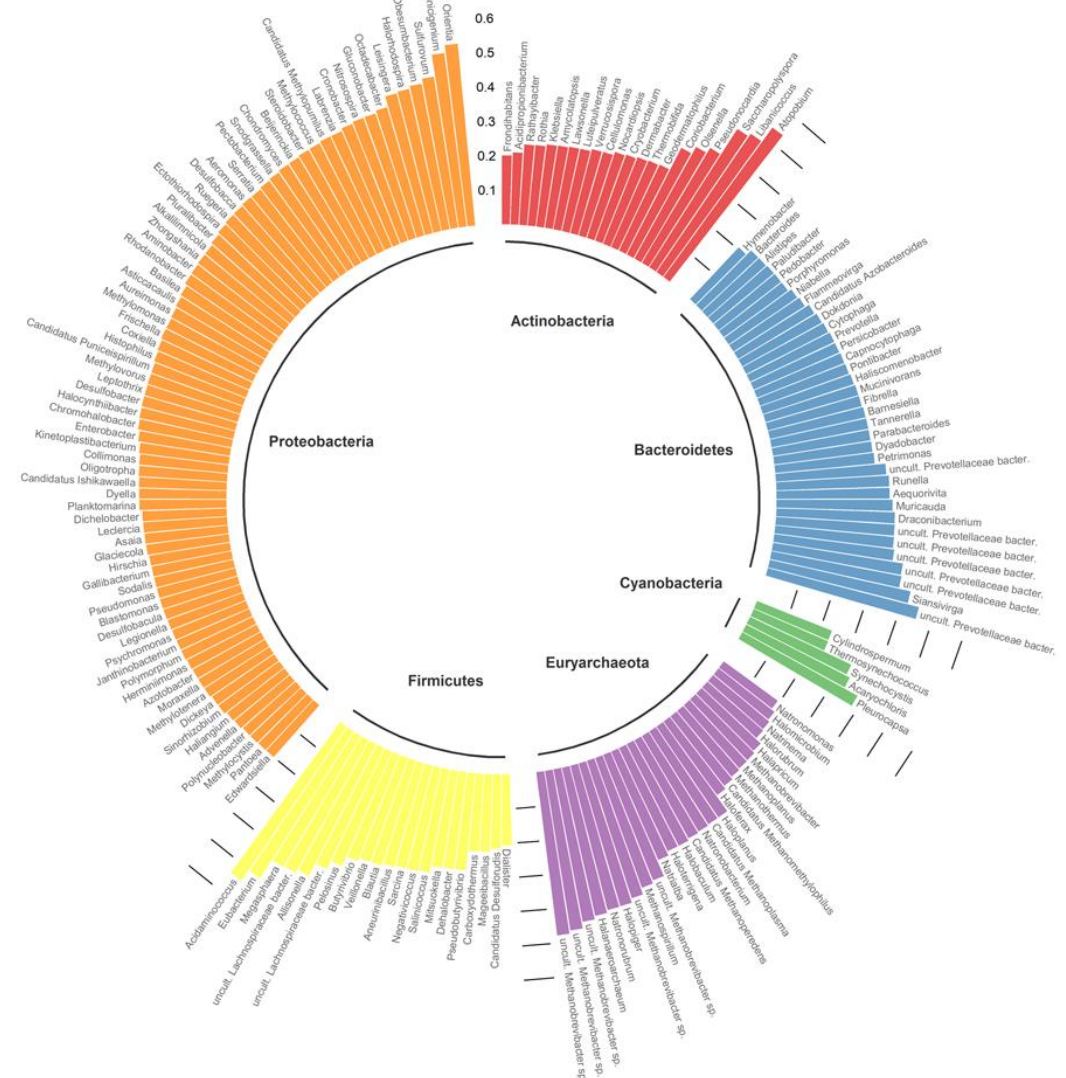
Bovine host genome acts on rumen microbiome function linked to methane emissions

- Selection based on:
  - 30 most informative microbial genes- mitigation potential of 17%
  - CH_4 phenotype using respiration chambers – 13% mitigation potential
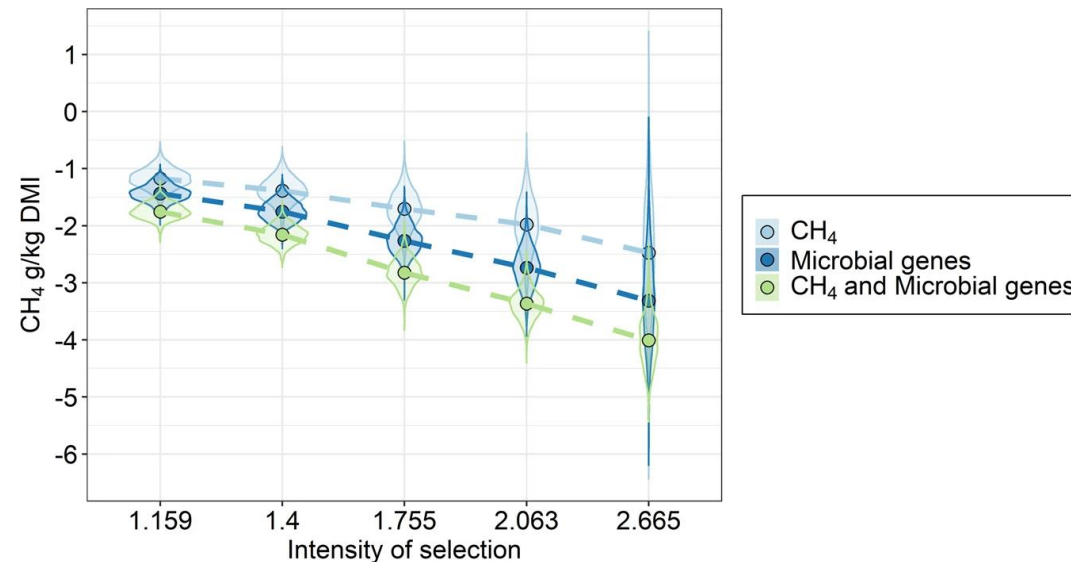
The host genome affects the comprehensive function of the microbiome in the cattle rumen

Martínez-Álvaro, et al. Commun Biol, 2022

Hypotheses

As the microbiome contributes a significant proportion of the variance of many traits, we may be able to use metagenomics to capture additional variance

Due to host-microbiome interactions, we may be able to positively select for more desirable microbiomes

B

Reference population

Phenotypic measurments

Microbiome samples

Microbiome relationship matrix

Prediction equation

$$y = Xb + Zu + e$$

$$\begin{bmatrix} u \\ e \end{bmatrix} \sim N \left[0, \begin{pmatrix} G\sigma_u^2 & 0 \\ 0 & R\sigma_e^2 \end{pmatrix} \right]$$

Accuracy

Predicted

Measured

Microbiome samples

Phenotypic measurments

Validation population

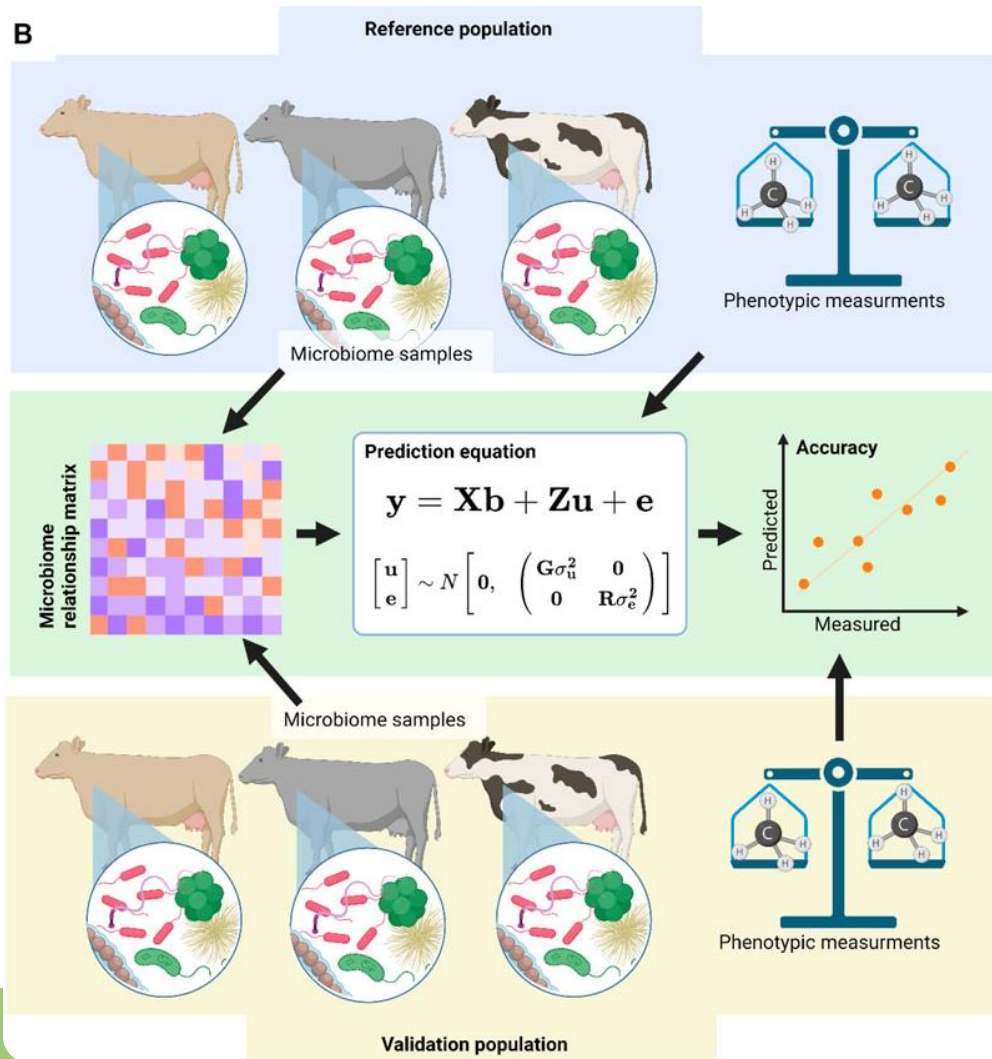# The opportunity of metagenomics

- **Metagenomic prediction**
  - Two types:
    - Phenotypic prediction
    - Proxy trait to predict genetic merit
  - Can be used in tandem with genotypes to inform prediction

Figure: Ross & Hayes, Front Genet, 2022

Metagenomic prediction

- First report of using BLUP for prediction, fitting metagenomic profiles from shotgun sequencing
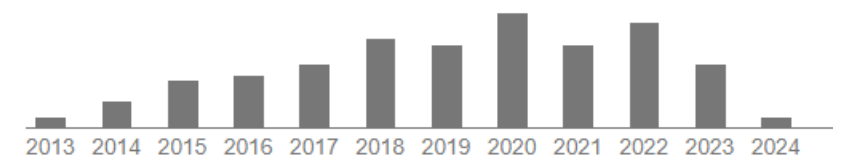- **2013**

Metagenomic Predictions: From Microbiome to Complex Health and Environmental Phenotypes in Humans and Cattle

Elizabeth M. Ross[1,2,3]*, Peter J. Moate[4], Leah C. Marett[4], Ben G. Cocks[1,2,3], Ben J. Hayes[1,2,3]

1 Biosciences Research Division, Department of Environment and Primary Industries, Bundoora, Victoria, Australia, 2 Dairy Futures Cooperative Research Centre, Bundoora, Victoria, Australia, 3 La Trobe University, Bundoora, Victoria, Australia, 4 Future Farming Systems Division, Department of Environment and Primary Industries, Ellinbank, Victoria, Australia

- Inspired by genomic BLUP, fitting a GRM
- Inspired by rapidly decreasing sequencing costs
- Inspired by difficulty measuring methane phenotypes and need for a proxy

- Prediction accuracy for methane = 0.466
- max n = 47 cattle

Cited by 144

2013 2014 2015 2016 2017 2018 2019 2020 2021 2022 2023 2024

Holobiont prediction

Model G   Model GM   Model GMO

Martinez Boggio et al., JDS, 2024

Holobiont prediction

$$ho^2 = (\sigma^2_{g_d} + \sigma^2_m + \sigma^2_{gxm})/\sigma^2_y$$

h² = 0.41
ho² = 0.58

h² = 0.29
ho² = 0.43

Martinez Boggio et al., JDS, 2024
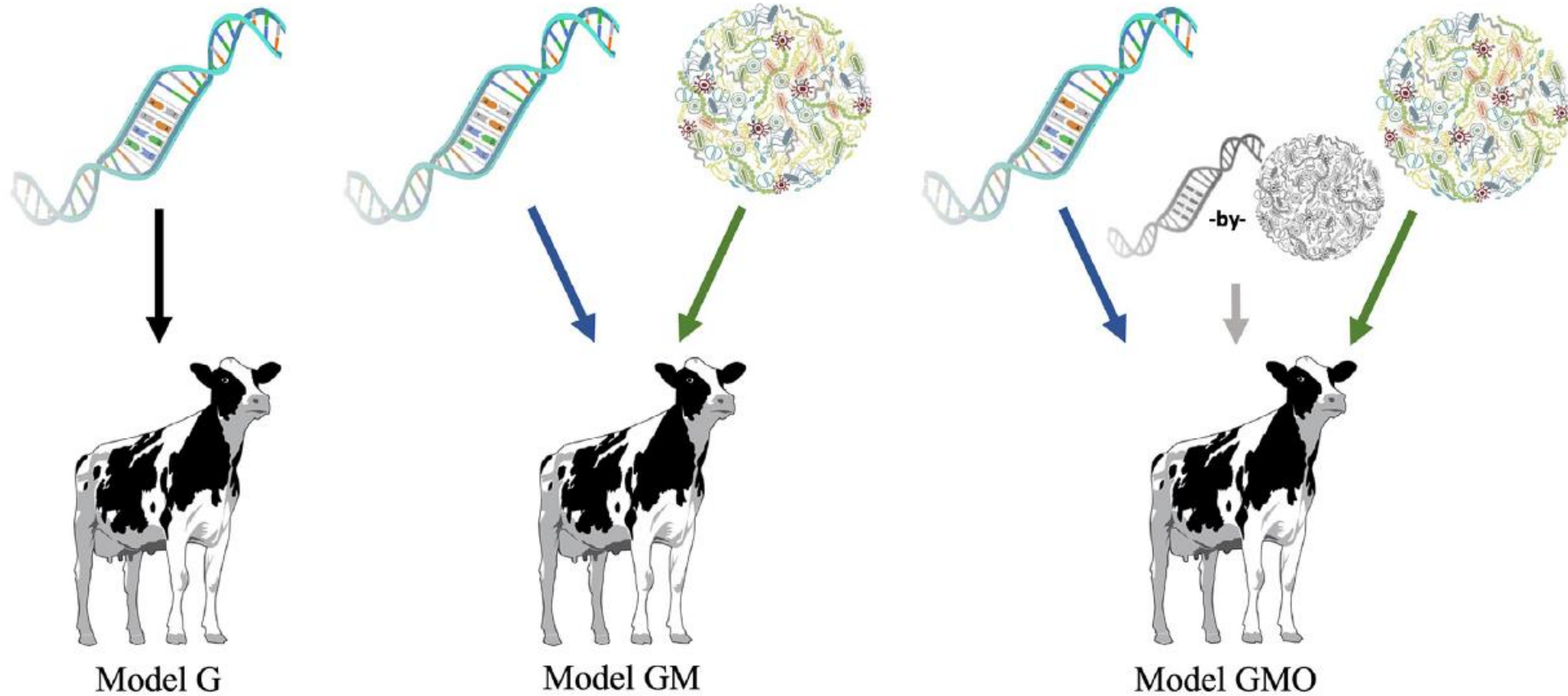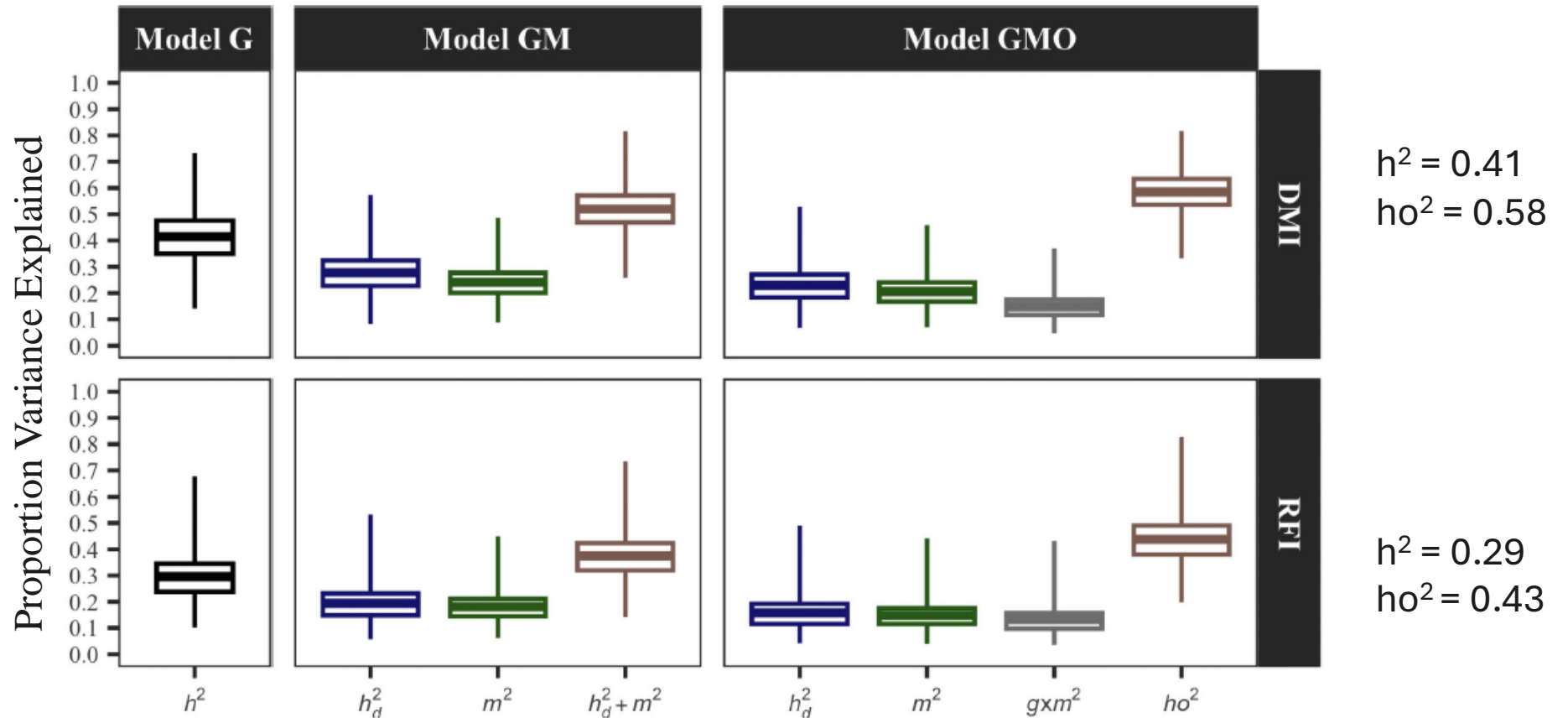
Holobiont prediction
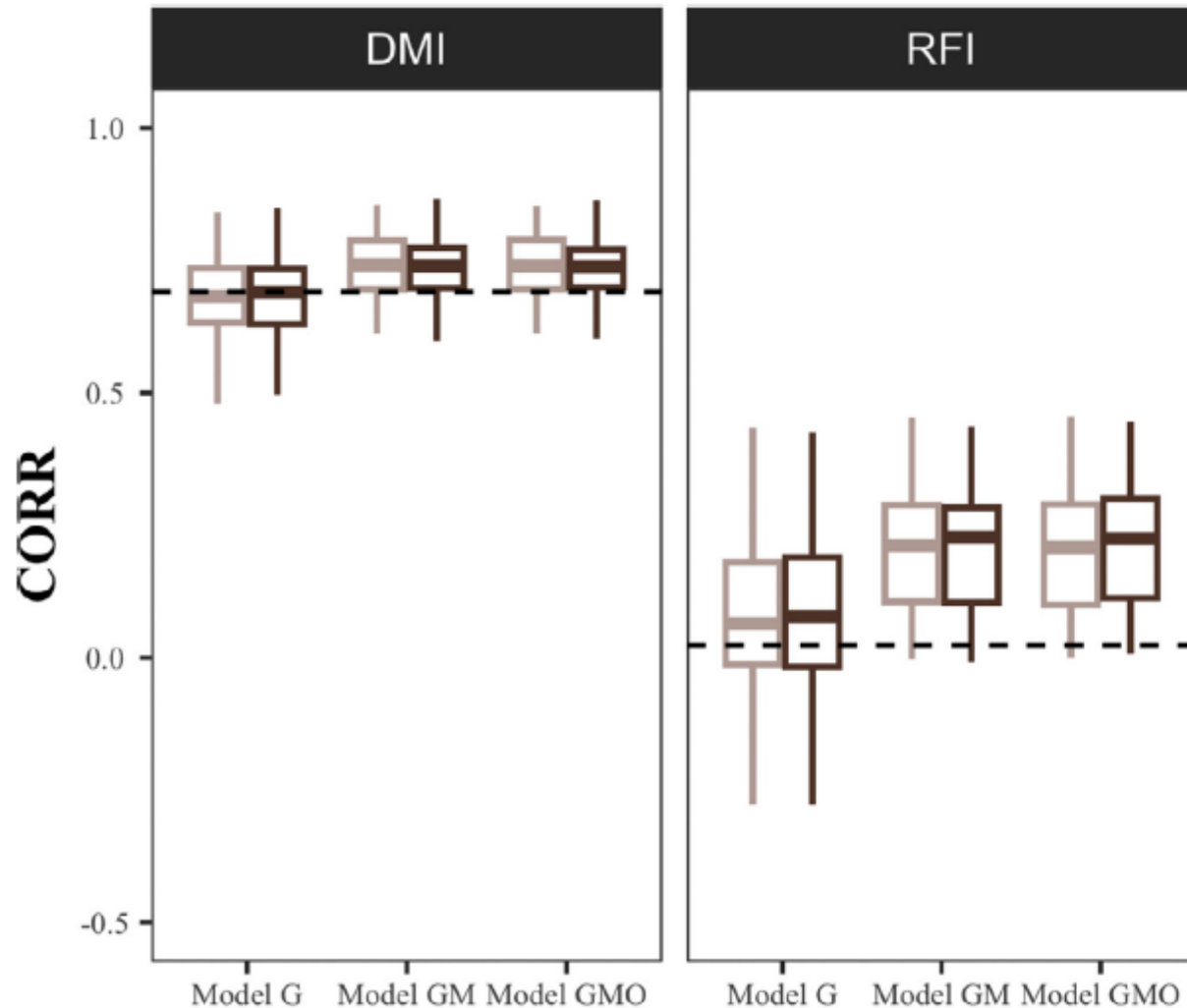
- The differences obtained between the h^2 and the h_d^2 strongly suggest that the microbiome mediates part of the host genetic effect

Metagenomic prediction

- Swine:
  - Prediction accuracy significantly increased by including microbiome data – more than 100% increase for back fat

- Bunnies:
  - Accurately sorted high vs low lifetime productivity females with 94% accuracy using just 53 amplicon seq variants

- Sheep:
  - Combining the metagenome profile with host genotype explained more than 70% of the variation in methane emissions and residual feed intake.
  - Increased prediction accuracy for growth and fleece weight

Christian Maltecca[1], Duc Lu[1], Constantino Schillebeeckx[2], Nathan P. McNulty[2], Clint Schwab[3], Caleb Shull[3] & Francesco Tiezzi[1]

RESEARCH ARTICLE   Open Access

Microbiome composition as a potential predictor of longevity in rabbits

Iliyass Biada[1], Noelia Ibáñez-Escriche[1*], Agustín Blasco[1], Cristina Casto-Rebollo[1] and Maria A. Santacreu[1]

RESEARCH ARTICLE   Open Access

Combining host and rumen metagenome profiling for selection in sheep: prediction of methane, feed efficiency, production, and health traits

Melanie K. Hess[1,2*], Larissa Zetouni[1,3], Andrew S. Hess[1,4], Juliana Budel[1,5], Ken G. Dodds[1], Hannah M. Henry[1], Rudiger Brauning[1], Alan F. McCulloch[1], Sharon M. Hickey[6], Patricia L. Johnson[1], Sara Elmes[7], Janine Wing[8], Brooke Bryson[9], Kevin Knowler[1], Dianne Hyndman[1], Hayley Baird[1], Kathryn M. McRae[1], Arjan Jonker[10], Peter H. Janssen[10], John C. McEwan[1] and Suzanne J. Rowe[1]

Metagenomic prediction

Steps:

1. Sample and phenotype collection
2. Sequencing the metagenome
3. Classification
4. Quantitative analysis
   a. Relationship matrix
   b. Prediction modeling

Figure: Ross & Hayes, Front Genet, 2022
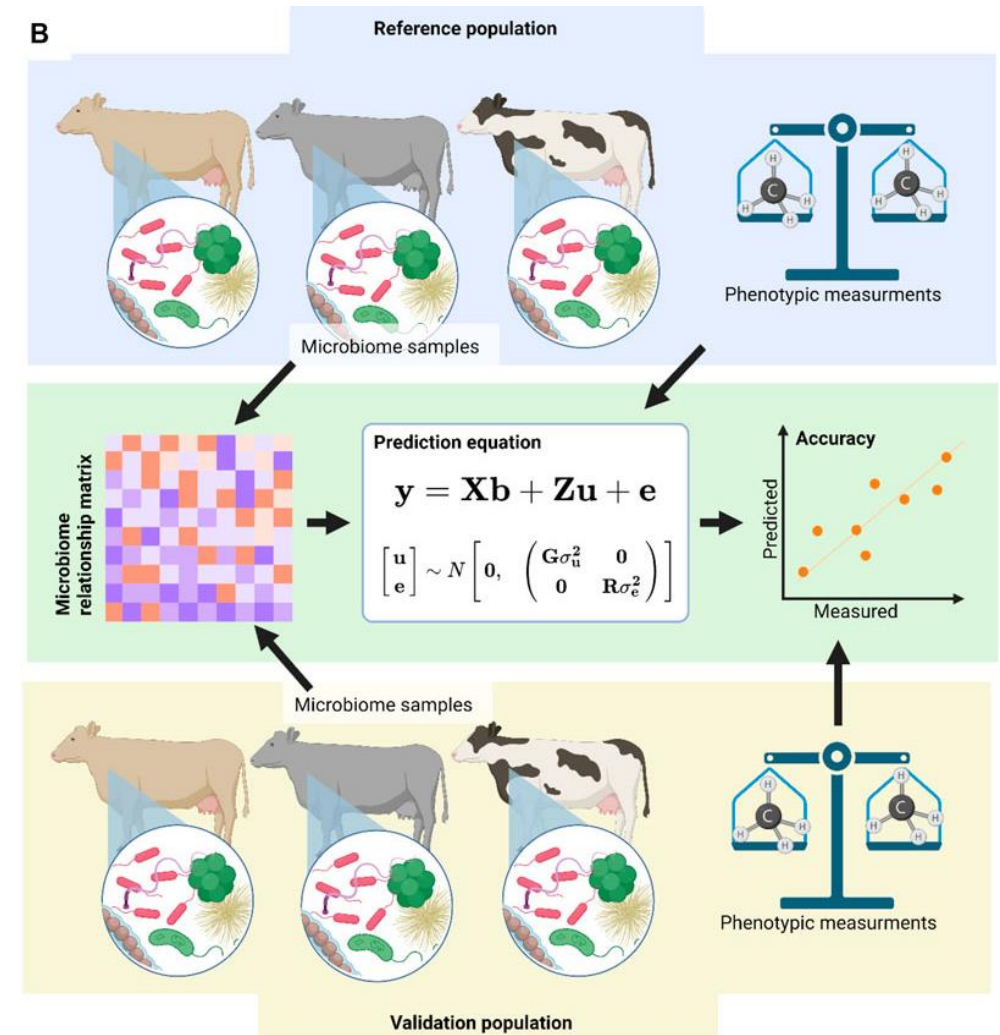
Metagenomic prediction

1. Sample and phenotype collection
   - Consider trait of interest and how that relates to the microbial community being sampled
     - Rumen, fecal, oral
     - Could be anything!

Photo credit: E. Ross, personal comm.

Metagenomic prediction

2. Sequencing the metagenome

Amplicon sequencing (16S, 18S, etc.)
- Pro: Most cost effective option
- Cons: Lower resolution, more difficult to characterize function or abundance

Short read sequencing
- Pro: Less biased approach
- Con: Single reads don't span entire genes

Long read sequencing
- Pro: Best taxonomic classification, highest alignment rate
- Con: More expensive

Figure: Lee, J Open Source Edu, 2019
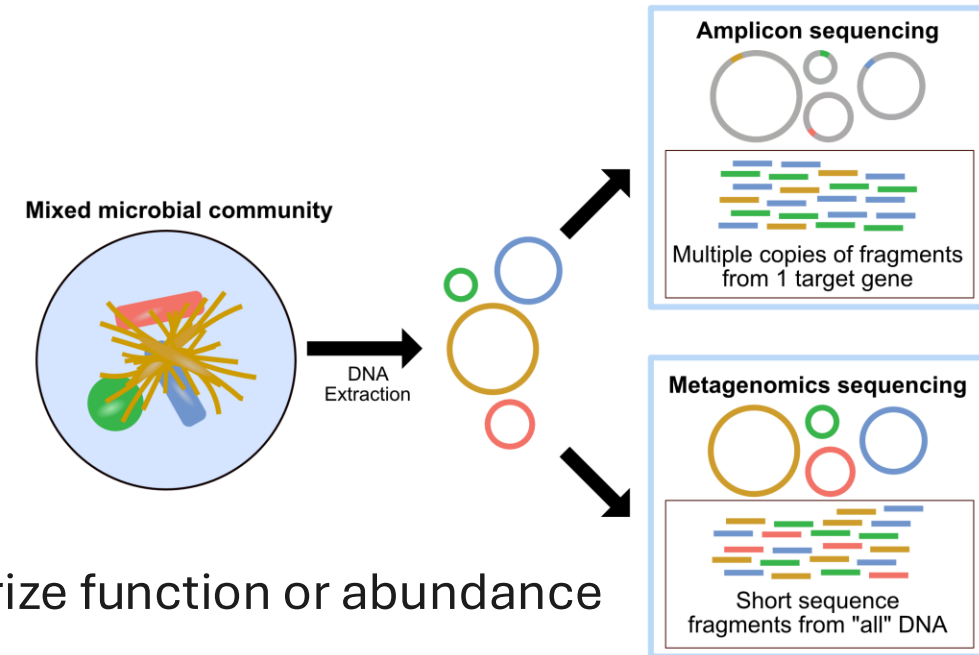
Metagenomic prediction

2. Sequencing the metagenome

- *Sequencing depth example:*
  - Short read: 3 million reads
    (Ross et al., 2012)

  - Long read: 1.5 million reads per sample
    (E. Ross, personal comm., 2024)
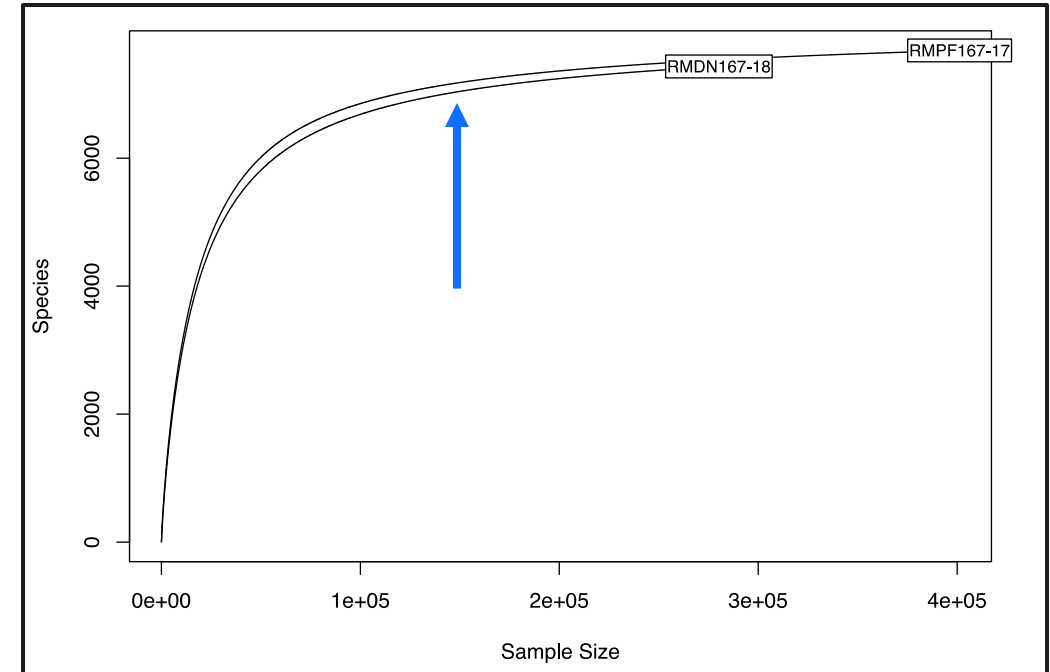
Metagenomic prediction

3. Classification – taxonomical or functional
   - **Taxonomical** – which species are present?
     - Direct alignment – assigns taxonomy using either public or assembled reference dataset
     - Reference free approaches – based on sequence similarity

Figure: modified Zhang et al, 2016

Metagenomic prediction

3. Classification – taxonomical or functional
   - **Taxonomical** – which species are present?
     - Genus or species
     - OTUs – Operational Taxonomic Units
       - Clusters sequence based on percent similarity - challenges with between study comparisons
     - ASVs – Amplicon Sequence Variants
       - Distinguishes single nucleotides – reproducible across studies

Figure: modified Zhang et al, 2016

Metagenomic prediction
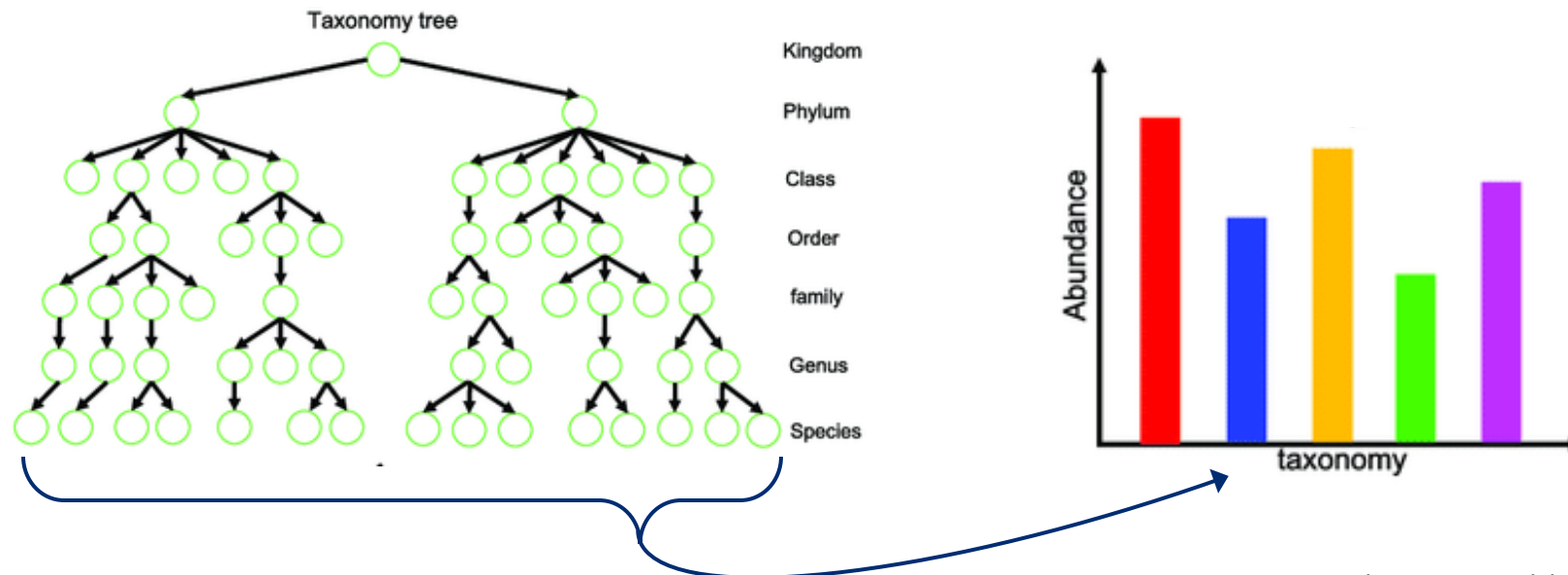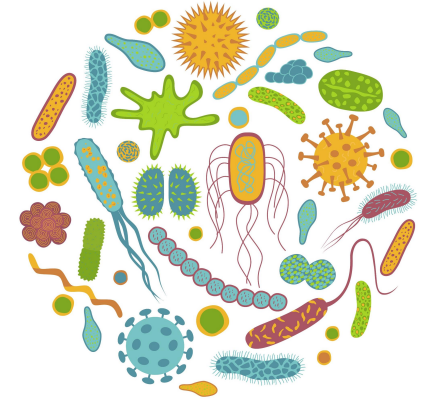
3. Classification – taxonomical or functional
   - **Functional** – what do the species do?
     - Gene content – direct alignment to reference dataset
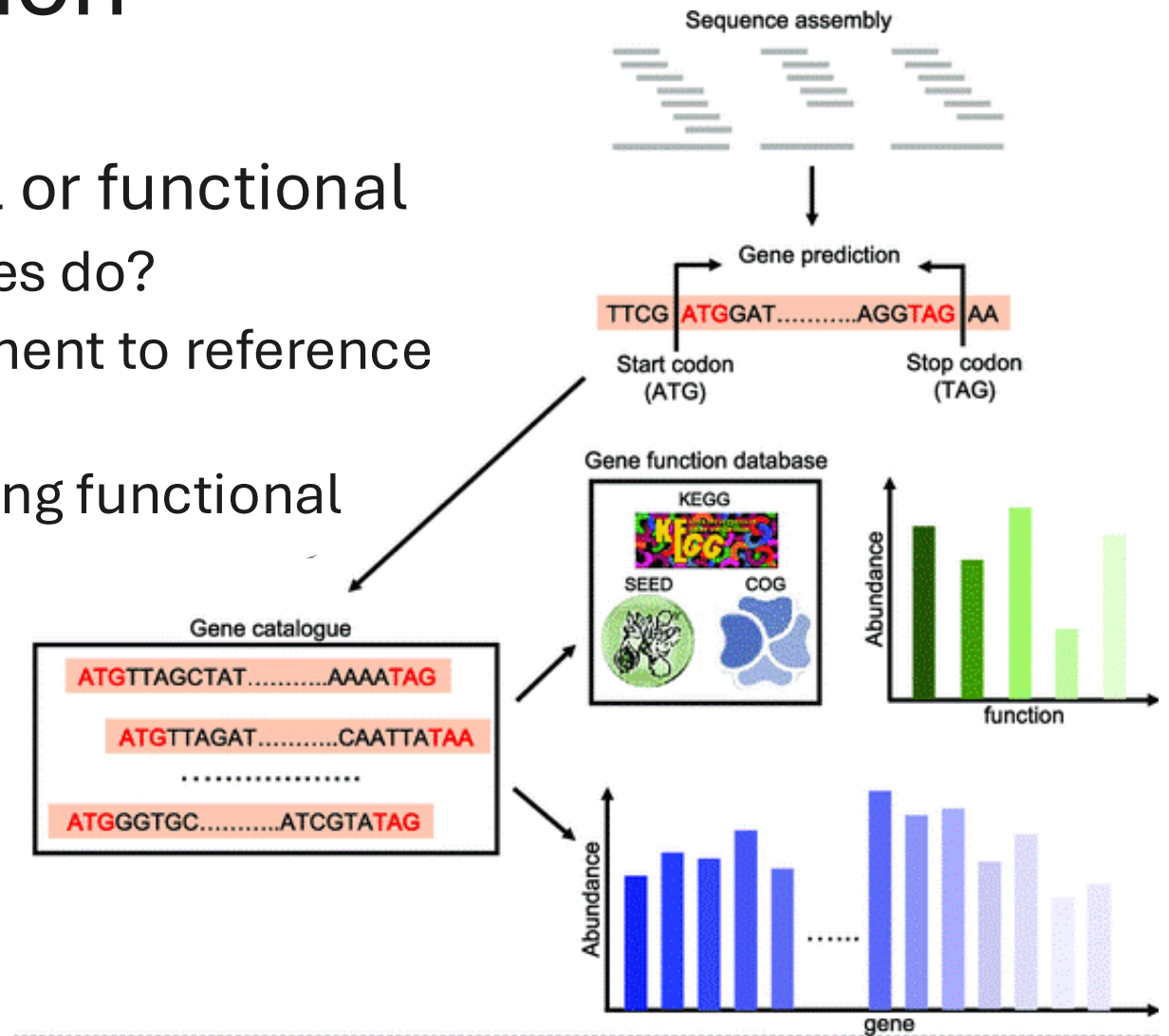   - Can be further classified using functional annotation

Figure: modified Zhang et al, 2016

Metagenomic prediction

	Taxa 1	Taxa 2	Taxa 3	Taxa 4	Taxa 5	Taxa 6	Taxa 7	Taxa 8	Taxa 9	Taxa 10	Taxa 11	Taxa 12
Animal 1	690	42	33	73	216	670	46	64	104	58	490	98
Animal 2	2475	786	1555	1124	2967	3086	555	1541	443	93	1174	2638
Animal 3	993	489	455	176	2506	1720	566	723	168	242	633	2583
Animal 4	343	455	70	467	267	196	5	307	219	10	88	278
Animal 5	762	37	1014	732	2145	1449	704	794	253	14	1618	1472
Animal 6	377	82	372	233	707	182	26	49	5	24	331	331
Animal 7	3705	855	1326	1403	3426	2620	460	1741	764	355	2224	127
Animal 8	559	39	5	1506	1961	2219	944	355	304	231	2121	2117
Animal 9	220	53	352	375	646	578	268	173	78	62	395	34
Animal 10	1762	1316	329	843	261	1466	341	1759	364	43	478	1042
Animal 11	327	27	102	252	350	128	77	398	5	34	523	121
Animal 12	2600	1623	369	1726	699	51	313	997	1042	398	1162	743
Animal 13	1835	817	1828	1294	2528	1813	71	2216	831	290	2169	914
Animal 14	406	81	70	88	618	485	195	126	22	47	271	90
Animal 15	2057	860	619	1321	2228	1147	477	1198	407	252	1443	1546
Animal 16	206	563	636	239	406	1304	375	884	132	79	77	1288
Animal 17	71	1552	1385	944	278	1536	266	2222	536	313	1509	809

Figure: Ross & Hayes, Front Genet, 2022

Metagenomic prediction

4. Quantitative Analysis
   a. Metagenomic relational matrices (MRM)
      - $M = \dfrac{XX'}{n}$ (Ross et al., 2013; Difford et al., 2018; Saborio-Montero et al., 2021; Boggio et al., 2023)
      - **X** is the matrix of natural log transformed bacterial and archaeal relative abundance

   <u>Alternative:</u>

   - Heritabilities of the microbiome were estimated by fitting the 3361 [microbes] as observed traits in 3361 univariate genomic models. (Martínez-Álvaro, et al., 2024)

   ***Nearly any approach requires a normalization or scaling step

Metagenomic prediction

4. Quantitative Analysis
   b. Prediction modeling
      - Phenotype vs proxy trait breeding value
      - Linear, Bayesian, machine learning
         - BLUP – *Ross et al., 2013*
         - BayesR-like – *Zhang et al, ISME, 2018*
         - Machine learning – *Maltecca et al, Sci Rep, 2019*

      - Fitting both the GRM and MRM simultaneously
         - Separate variance estimates
         - Microbiability and heritability

Feasibility of metagenomic predictions

- Value proposition of metagenomics:
  - Capturing trait variance for low-moderate heritability traits
  - Especially, difficult to measure traits

- Cost of:
  - Long read metagenomic sequence: $100 / sample
  - Feed efficiency trial: $750 / head
  - Year of missed cow production due to infertility: $900 (Prevatt et al, 2018)
  - BRD prevention, management, treatment, herd losses = $1 billion per year to industry (TAMU)

Feasibility of metagenomic predictions

Facilitating genetic selection for reduced enteric methane in northern Australia

- Simulated Case Study – E. Ross, unpublished

Assumptions:

1. Large effective population size: N = 1000

2. Heritability of methane: h^2 = 0.17 (Gonzalas-Recio et al., 2022)

3. Genetic correlation between CH_4:CO_2 ratio and microbiome predicted methane = 0.66 (Bilton et al., 2022)

4. Cost of metagenomic sequenced sample: $160 AUD

5. Cost of GreenFeed phenotype: $2750 AUD

Feasibility of metagenomic predictions

- The expected prediction accuracy was calculated using the equation from Daetwyler et al (2008):

  - $r = \sqrt{\dfrac{Nh^2}{Nh^2 + M_e}}$

    - Scaled by genetic corr. between direct and microbiome predicted methane

- **Scenario:** All animals have a metagenomic profile, 5% have GreenFeed phenotype

Feasibility of metagenomic predictions

- **Scenario:** All animals have a metagenomic profile, 5% have GreenFeed phenotype
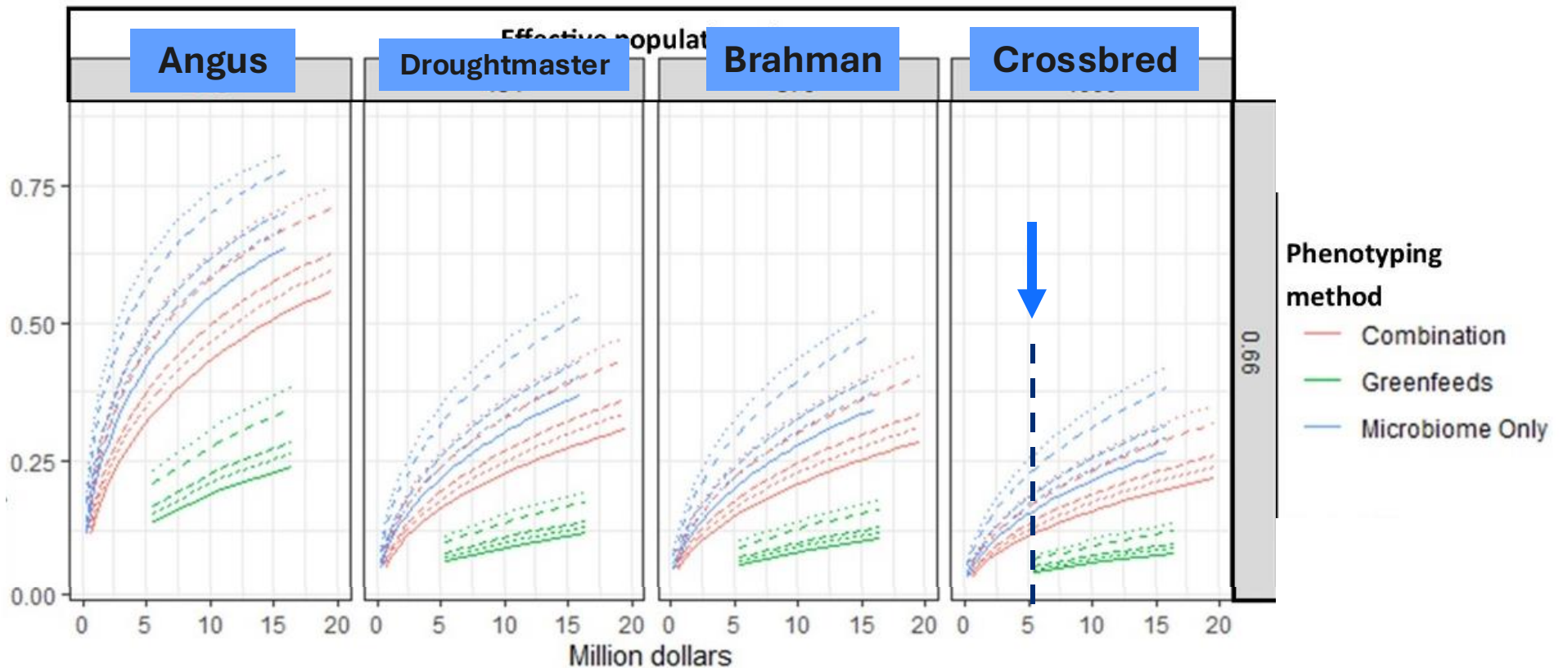
Future perspective: *Improving disease resistance*

BRD resistance	• Nasal microbiome
Pink eye resistance	• Nasal or ocular microbiome
Venereal disease resistance	• Vaginal and seminal microbiome

Future perspective: *Improved reproduction*

Improved conception rate
- Vaginal and seminal microbiome

Reduced early embryonic loss
- Vaginal and seminal microbiome

IVF success rate
- Follicular fluid and seminal microbiome

Challenges

- Required reference population sizes
  - Still a quantitative trait of low-moderate heritability
  - Still need large groups of animals with metagenomic profiles to be useful

  - $p > n$ – many datasets currently have more metagenomic classifications than there are sequenced animals

Challenges

- Data interpretation
  - Effect direction of metagenomic predictions is not necessarily known
    - Is the microbiome affecting the phenotype, or is the phenotype affecting the microbiome?
    - Enteric methane (microbiome effects phenotype)
    - Autism in humans (phenotype effects microbiome)

  - Timing of sampling and reporting of metadata to compare studies
    - Ex: reference free approaches, especially OTUs, are difficult to compare
    - Population and environmental differences
   - Dietary and developmental differences

Ongoing research questions

- Underlying biology

- Optimized modeling approaches

- Estimation of metagenomic relationships

- Optimized sequencing requirements

- Classification approaches

- Sufficient reference datasets for alignment

- Host-metagenome interactions

- Proof-of-concept for selection using metagenomic informed predictions
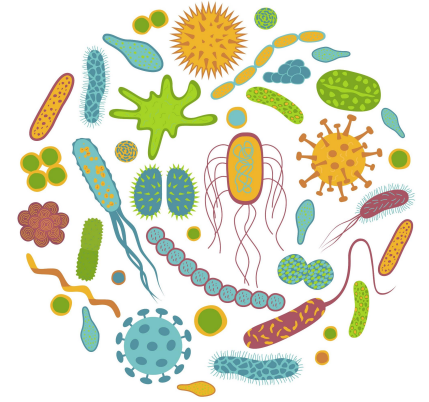
Summary

- The microbiome has potentially large effects on expression of traits of interest

- Between animal variation can be profiled using metagenomics

- Metagenomic prediction is an opportunity to increase prediction accuracy or use the metagenome as a proxy trait

- There is ample room for improvement – opportunities!
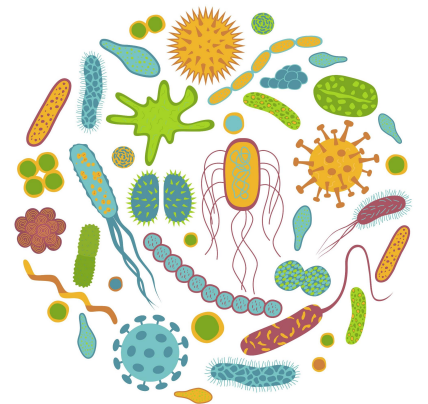
Acknowledgements

University of Queensland

Elizabeth Ross

Chian Ong

USMARC

Larry Kuehn

Jim Wells

Tara McDaneld

Tim Smith

Kristin Kuhn

Sungbong Shin

University of Nebraska – Lincoln

Samodha Fernando

THE UNIVERSITY OF QUEENSLAND
AUSTRALIA

QAAFI
Queensland Alliance for Agriculture and Food Innovation

USDA
Agricultural Research Service
U.S. DEPARTMENT OF AGRICULTURE

Questions?

Contact:

Bailey Engle   402-762-4264   bailey.engle@usda.gov

Mention of a trade name, proprietary product, or specific equipment does not constitute a guarantee or warranty by the USDA and does not imply approval to the exclusion of other products that may be suitable. USDA is an equal opportunity provider and employer.