Genomics session

Course: Organisation and utilisation of hologenomic datasets

Instructors: Dr. Jaelle Brealey, Dr. Melanie Parejo, Assoc. Prof. Morten Limborg

Date: 12 September 2022, Bilbao, Spain

Outline

9.10 - 9:40h	Intro lecture: From population genomics to hologenomes	Morten Limborg / Melanie Parejo
9:40:h- 10:00h	Pipeline presentation: *Host genome recovery from gut metagenomic samples*	Melanie Parejo
10.00 - 10:30	Coffee break ~30 min	
10:30 - 11:40	Hands on exercise: *mGWAS on salmon*	Jaelle Brealey / (Morten Limborg)
10:40 - 12:00	General discussion and questions	All

Why is the host genome a key component of the holobiont?

Heritability and **evolution**!

If host has no effect => only the **environment** that matters!!!

ARTICLE

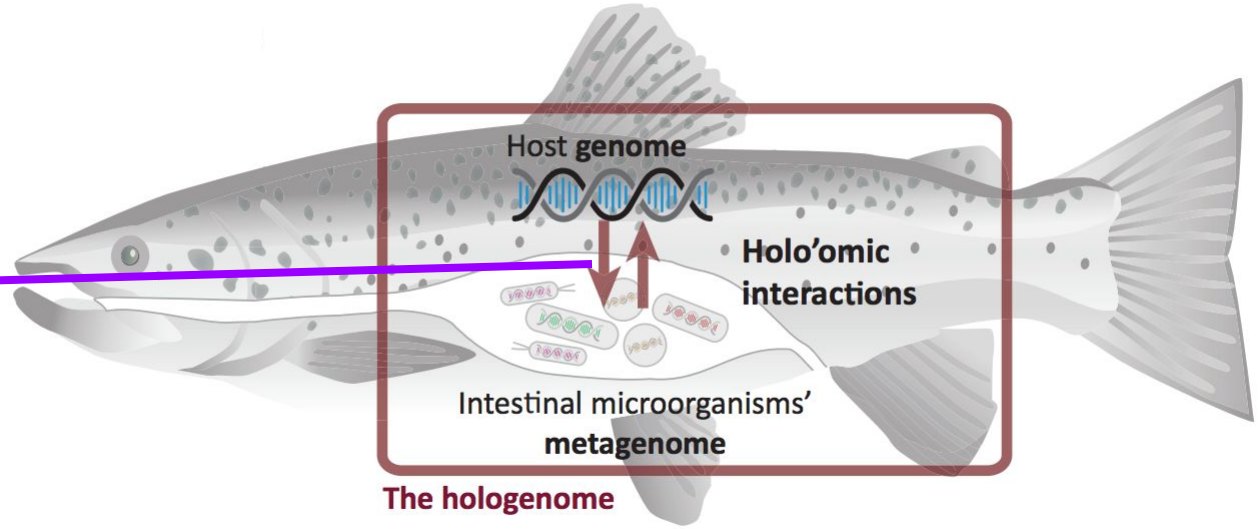
doi:10.1038/nature25973

Environment dominates over host genetics in shaping human gut microbiota

Daphna Rothschild[1,2]*, Omer Weissbrod[1,2]*, Elad Barkan[1,2]*, Alexander Kurilshikov[3], Tal Korem[1,2], David Zeevi[1,2], Paul I. Costea[1,2], Anastasia Godneva[1,2], Iris N. Kalka[1,2], Noam Bar[1,2], Smadar Shilo[1,2], Dar Lador[1,2], Arnau Vich Vila[3,4], Niv Zmora[5,6,7], Meirav Pevsner-Fischer[5], David Israeli[8], Noa Kosower[1,2], Gal Malka[1,2], Bat Chen Wolf[1,2], Tali Avnit-Sagi[1,2], Maya Lotan-Pompan[1,2], Adina Weinberger[1,2], Zamir Halpern[7,9], Shai Carmi[10], Jingyuan Fu[3,11], Cisca Wijmenga[3,12], Alexandra Zhernakova[3], Eran Elinav[5]§ & Eran Segal[1,2]§

Do we know if the host (genome) control the microbiome?

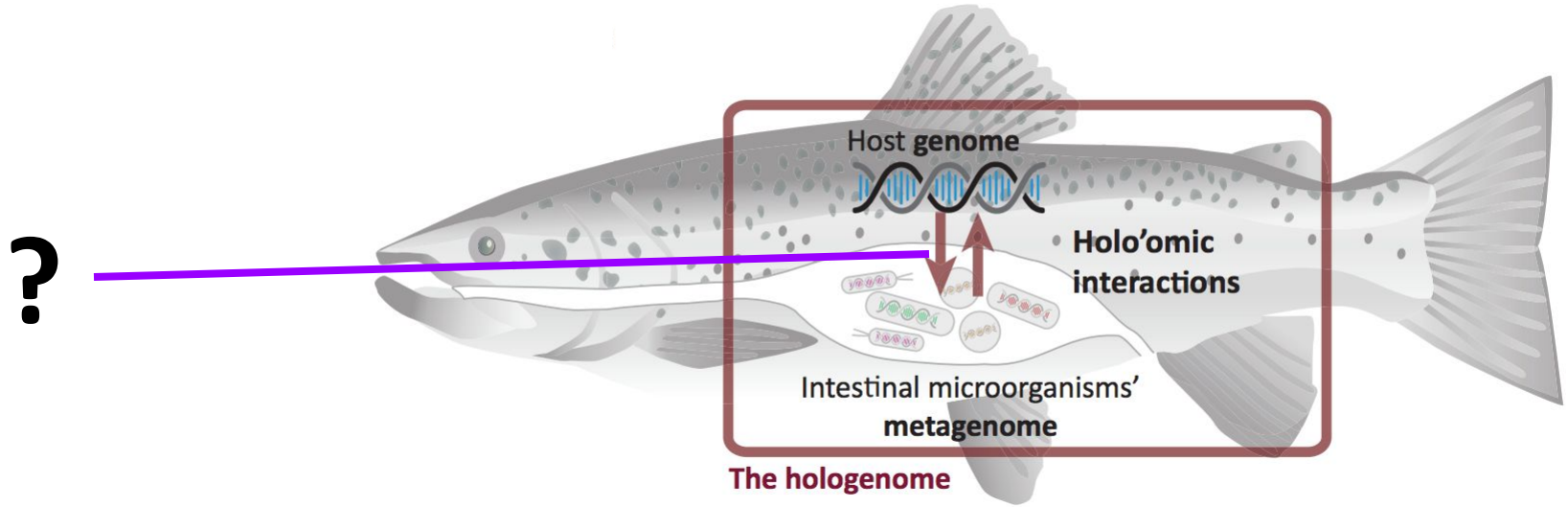
?



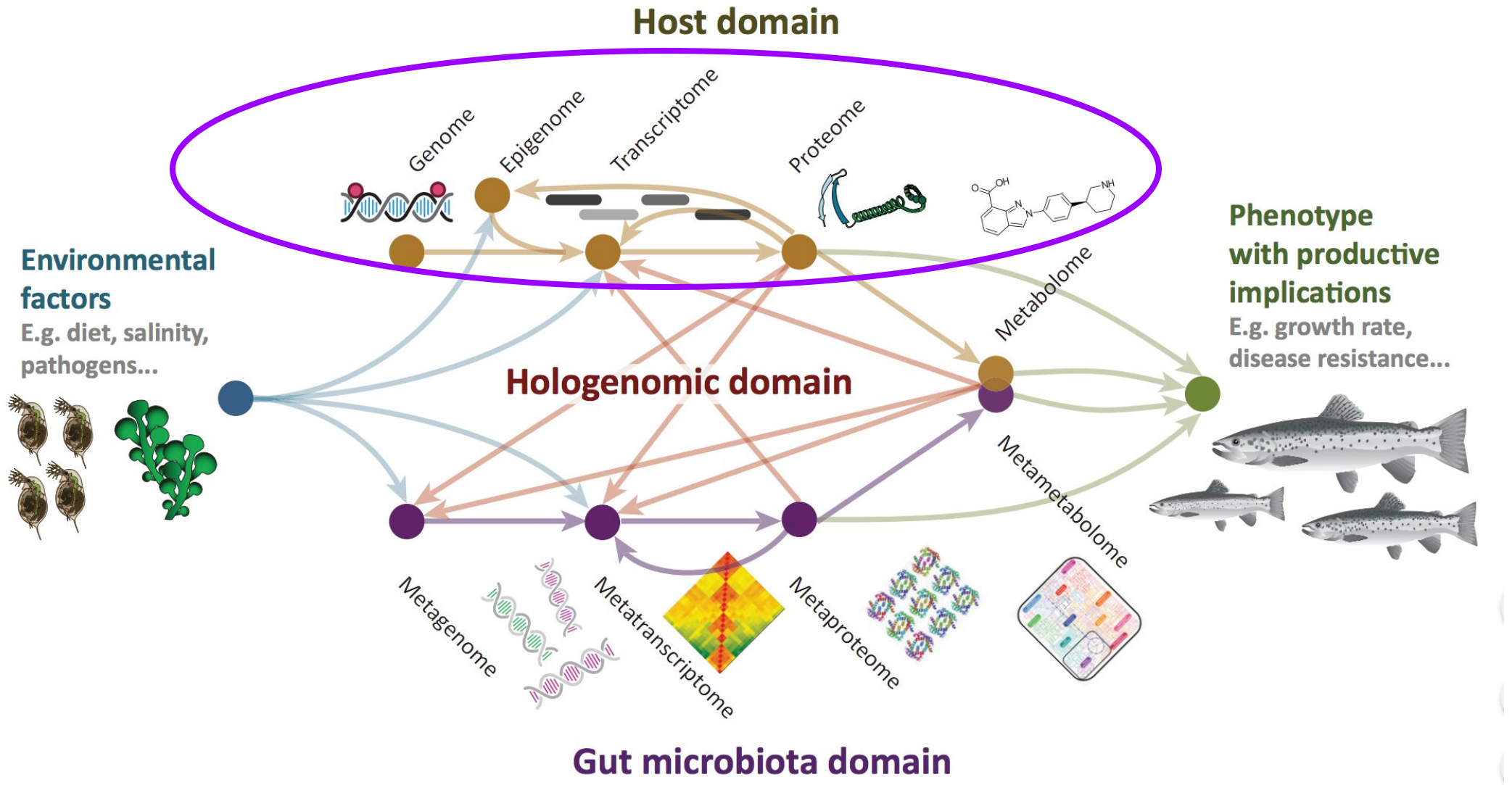
From where & how?

- Examples?
- How much do we know?
- Space & time?

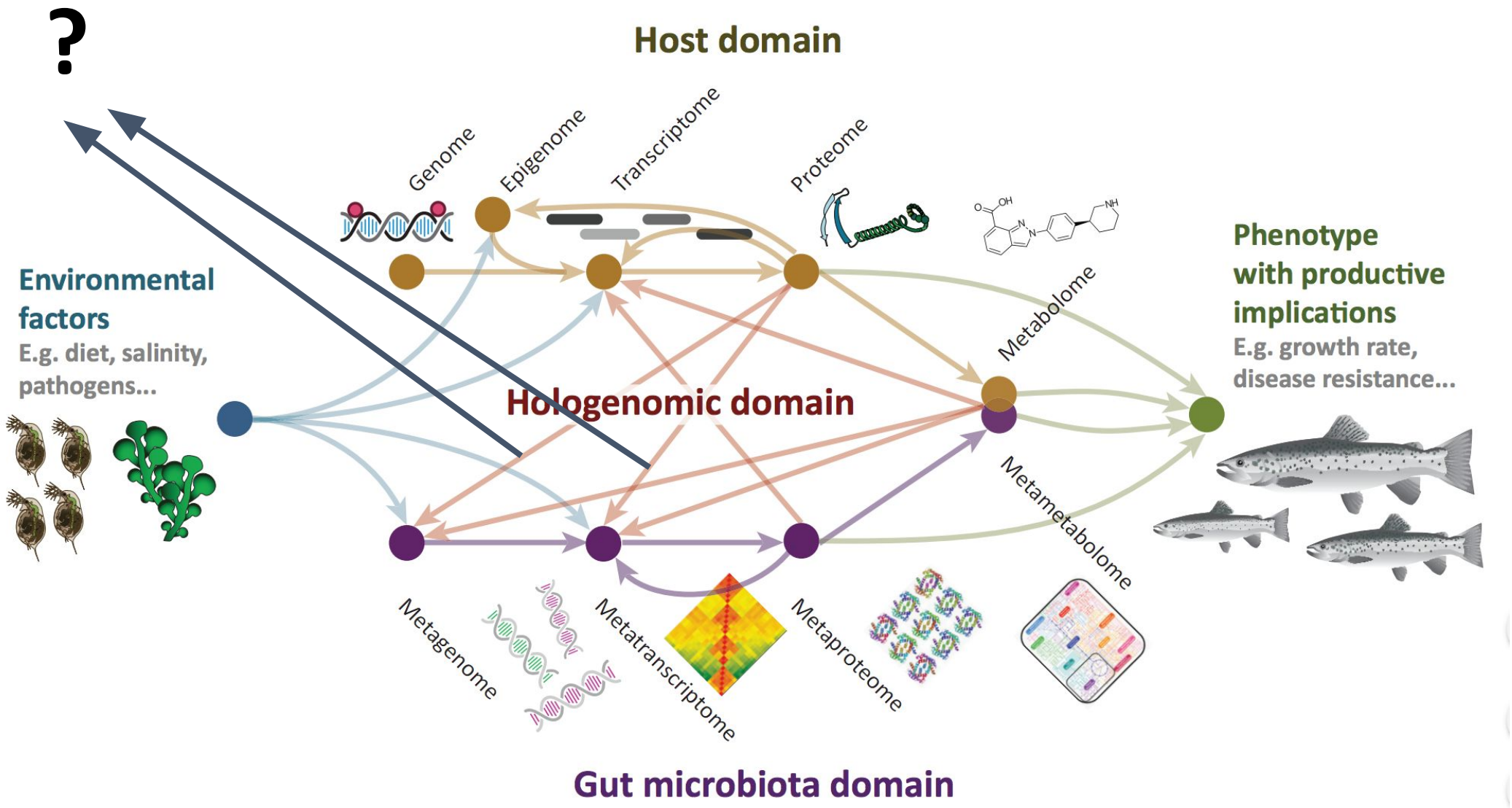
What are the mechanisms by which the host controls the microbiome?



What are the mechanisms by which the host controls the microbiome?



How can we test the role of the host?



How can we test the role of the host?

How can we test the role of the host?

EWAS

GWAS

Experimental studies

MGWAS

Comparative ‘omics

Other methods?

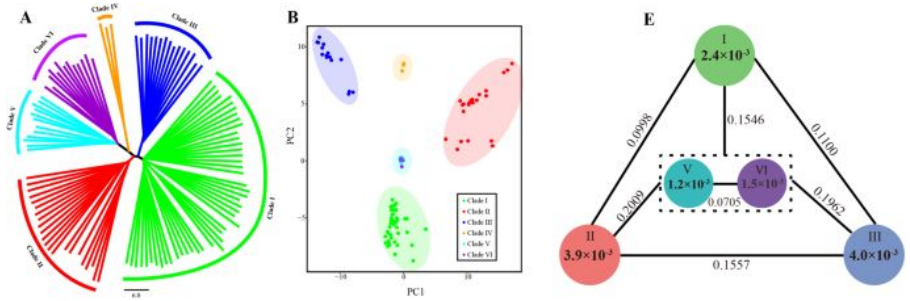
Before we start testing - we all need a basic recap to the fundamental field of **population genomics**

From population genomics to hologenomic hypothesis

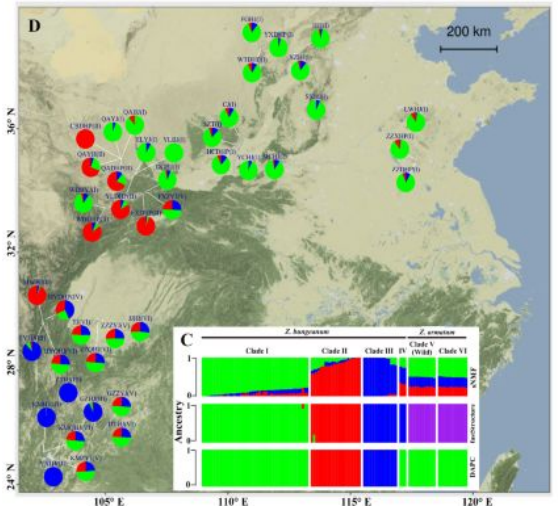
Index

- What is population genetics?
- Evolution and population genetics
- Why study population genetics/genomics?
- Redefining evolution 2.0 – the hologenome concept of evolution
- Hologenome theory of evolution
- Biology through the hologenomics lens
- Future perspectives

What is population genetics?



Population genetics is a field of biology that studies the genetic composition of biological populations, and the changes in genetic composition that result from the operation of various factors, including natural selection.



Important concepts: Natural selection, Mutation, Genetic drift, Gene flow, …

- What is the genetic composition of populations?
- What is the frequency of alleles and genotypes?
- What causes allele frequencies in populations to change? (= *Evolution*)
- Is fitness associated with certain alleles?

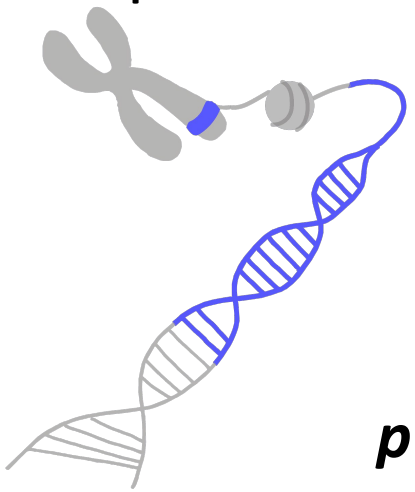
Feng et al. 2020

Evolution and population genetics

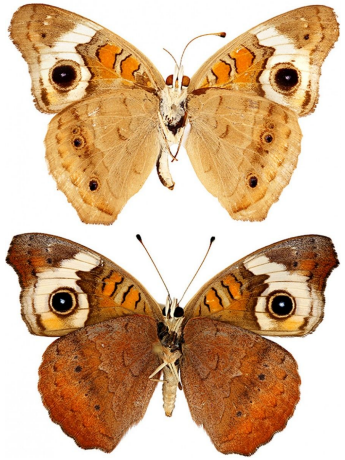
- Evolution is the change in allele frequency at a locus in a population over time

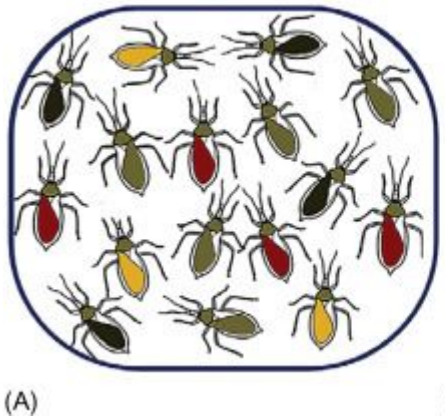
Evolutionary forces like natural selection act on different levels genes, individuals, groups, …

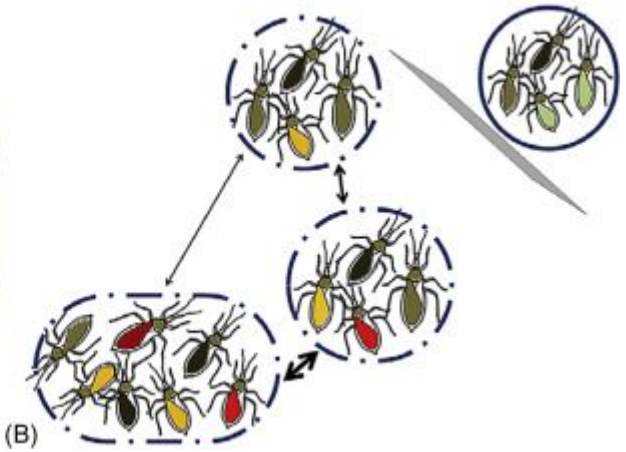
But the affects are seen in populations in the form of changes in allele frequencies study of population genetics



p + q

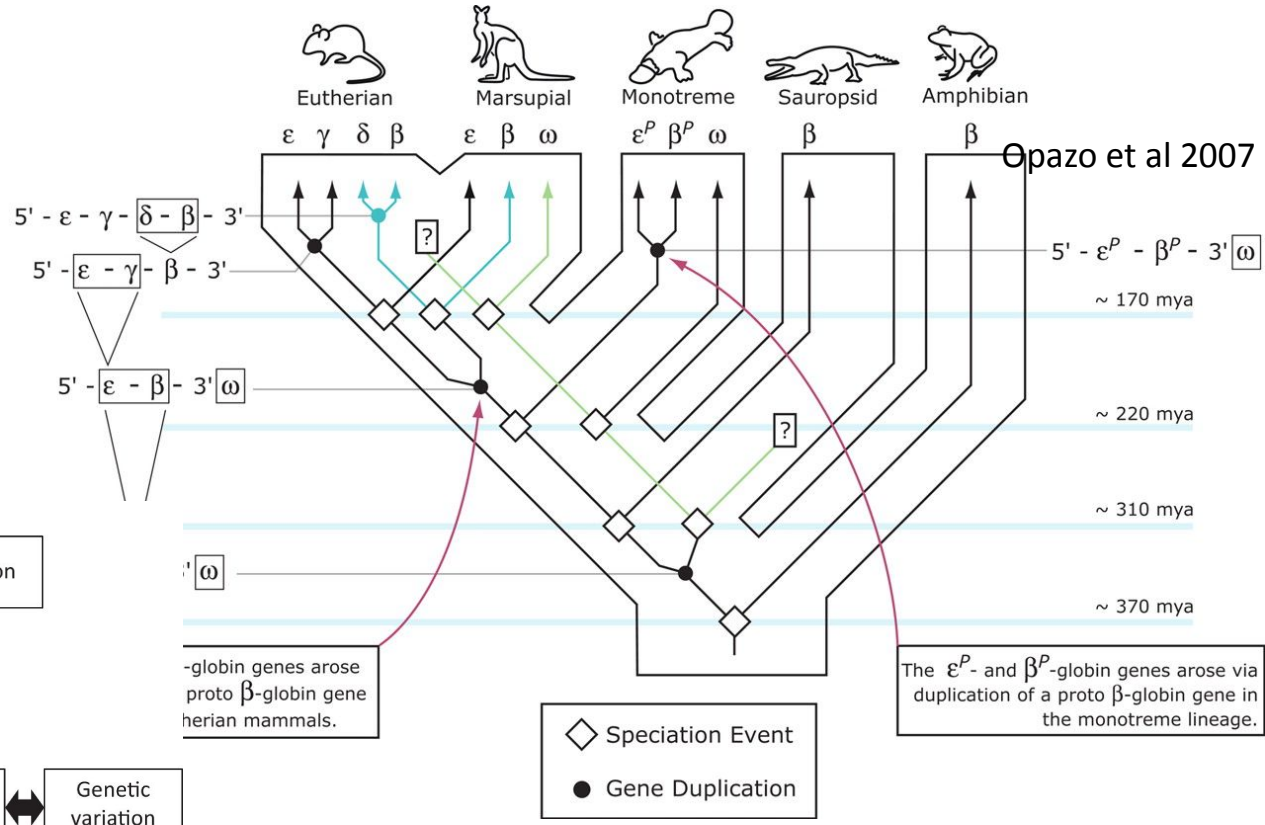




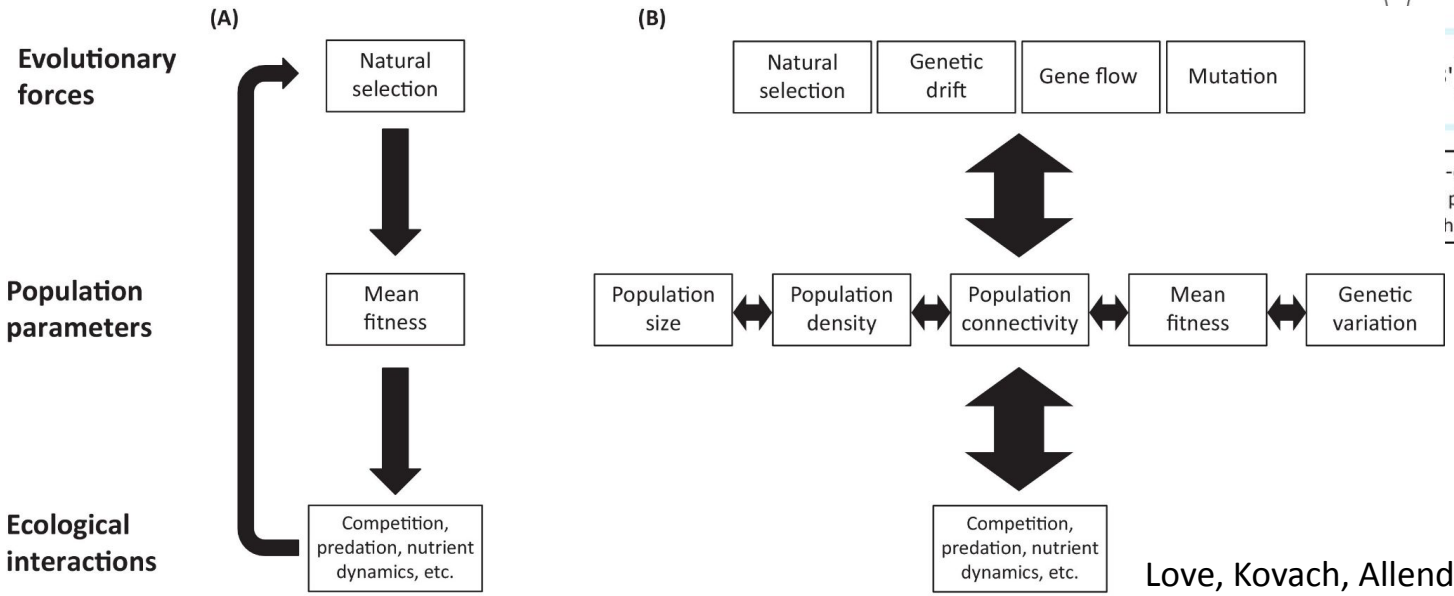


Why study population genetics/genomics?

1. Understand and refine theory
2. Understand the history of genes



Opazo et al 2007



Love, Kovach, Allendorf, 2017

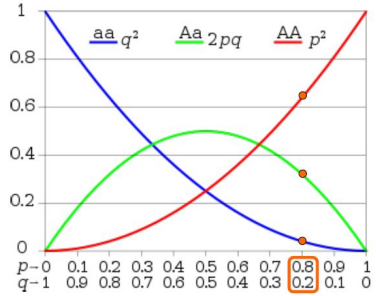
Trends in Ecology & Evolution

$p^2 + 2pq + q^2 = 1$

$p2 = (0.8)^2 = 0.64$

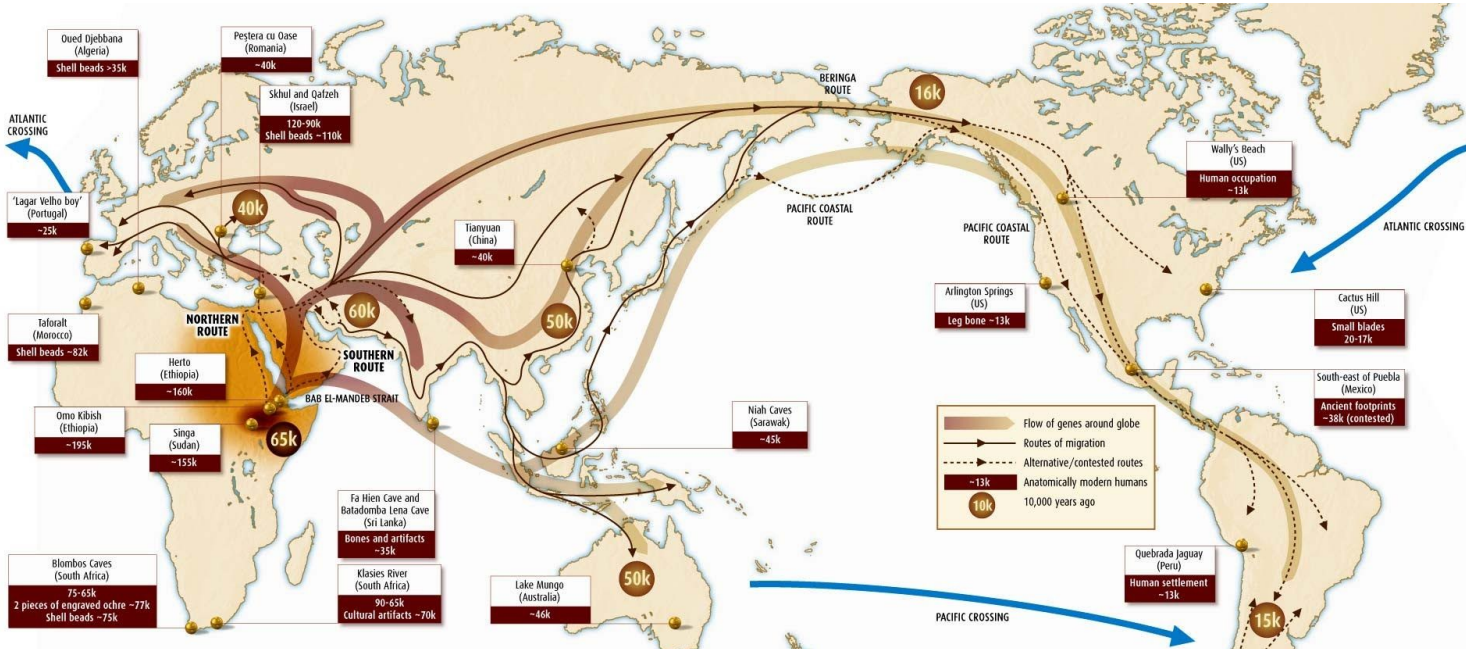
$2pq = 2*(0.8)*(0.2) = 0.32$

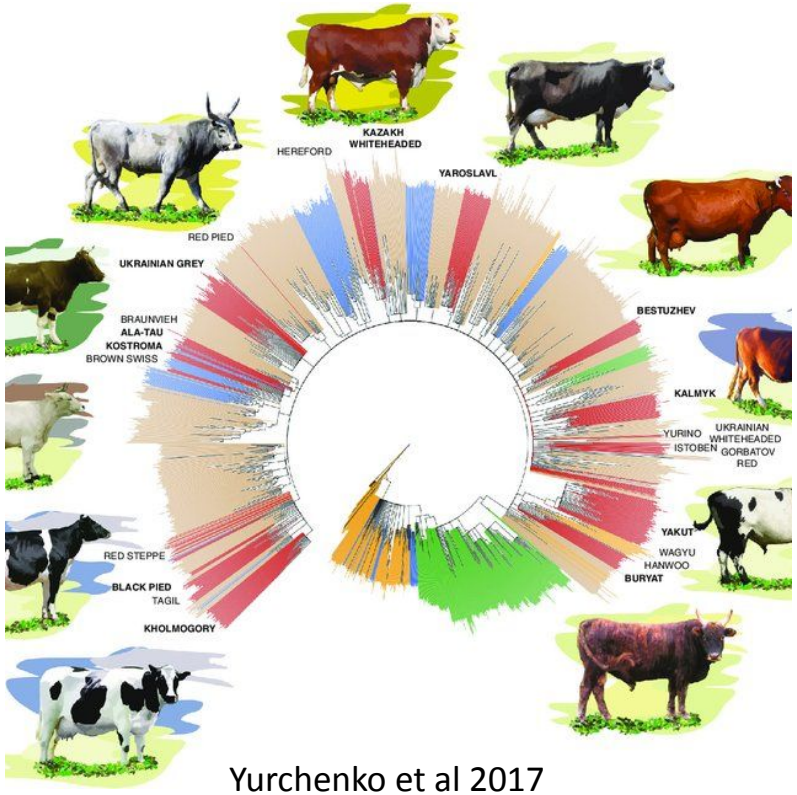
$q2 = (0.2)^2 = 0.04$



Why study population genetics/genomics?

3. Understand the history of populations

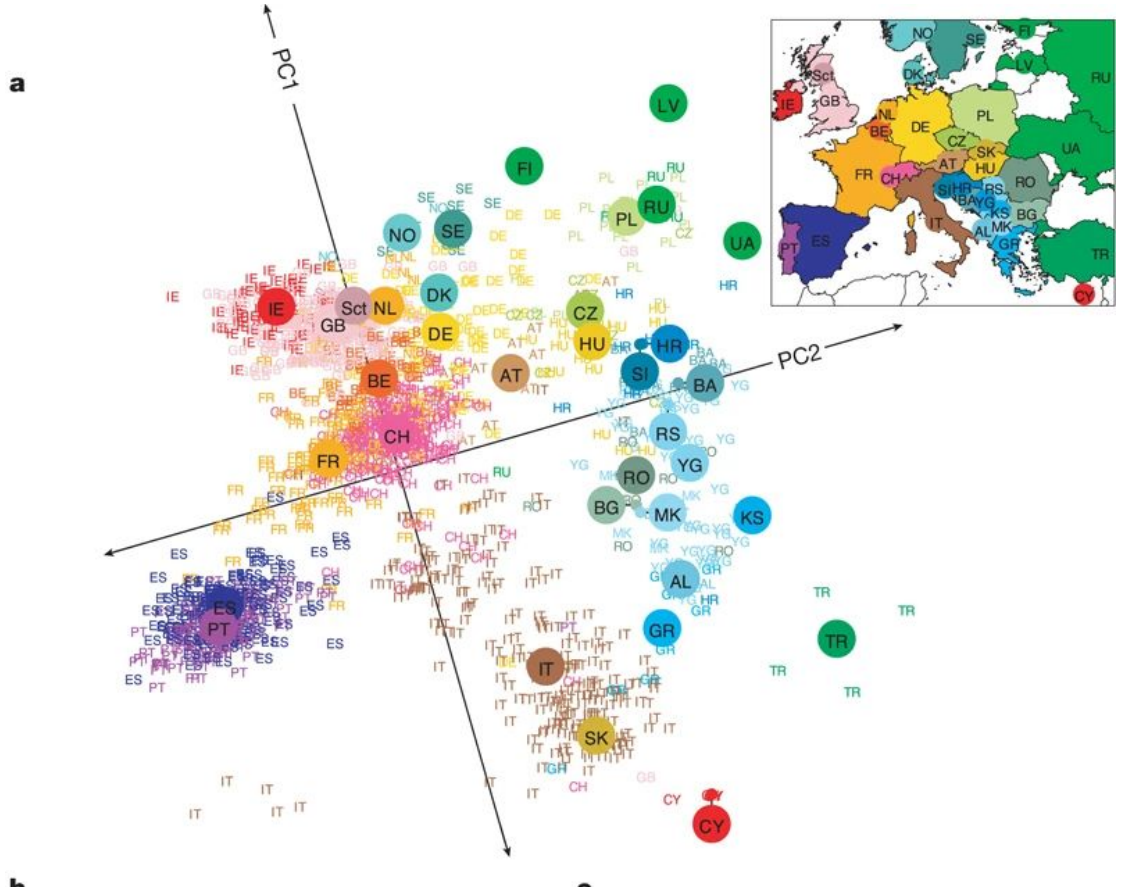




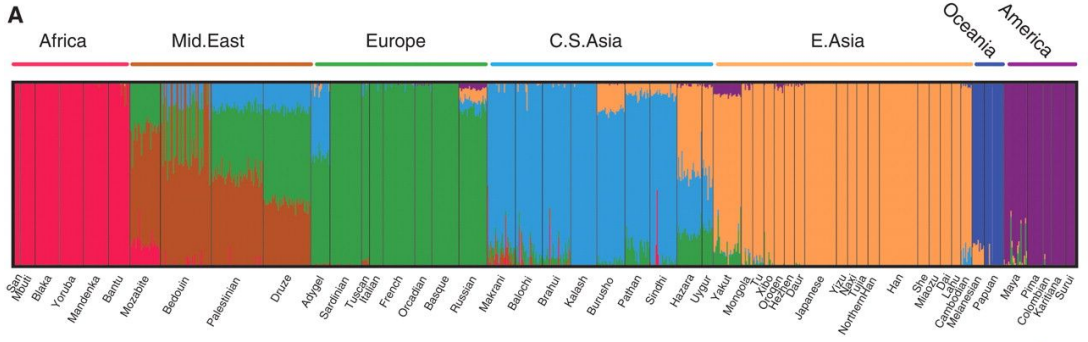
Yurchenko et al 2017

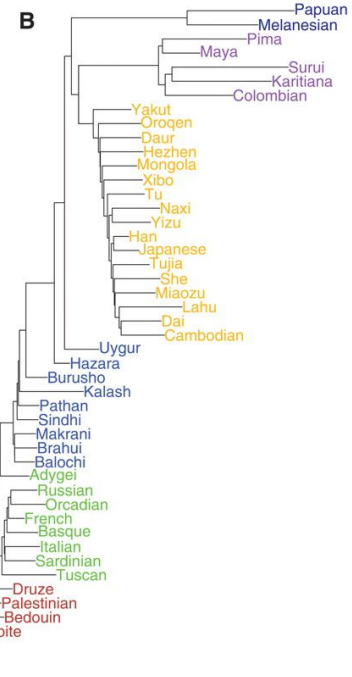
Why study population genetics/genomics?

4. Understand the relationship between individuals



November et al 2008



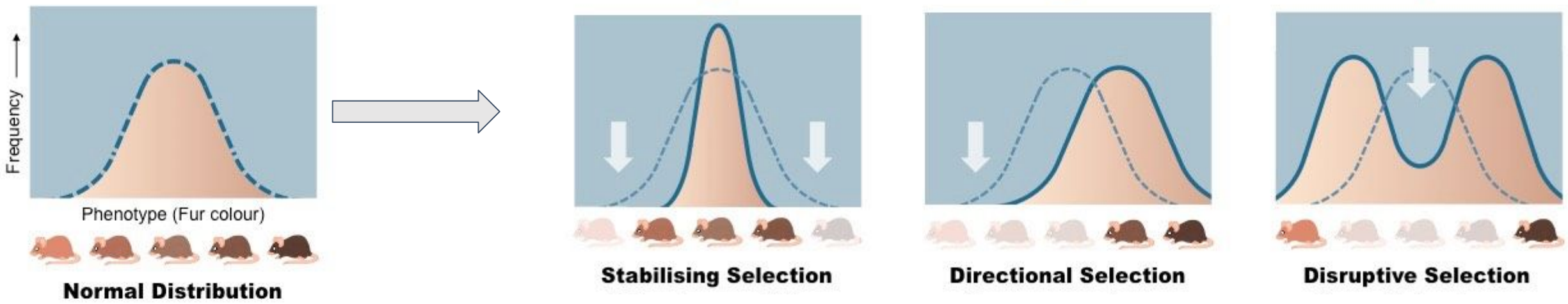


Worldwide Human Relationships Inferred from Genome-Wide Patterns of Variation

Jun Z. Li,[1,2]*† Devin M. Absher,[1,2]* Hua Tang,[1] Audrey M. Southwick,[1,2] Amanda M. Casto,[1] Sohini Ramachandran,[4] Howard M. Cann,[5] Gregory S. Barsh,[1,3] Marcus Feldman,[4]‡ Luigi L. Cavalli-Sforza,[1]‡ Richard M. Myers[1,2]‡

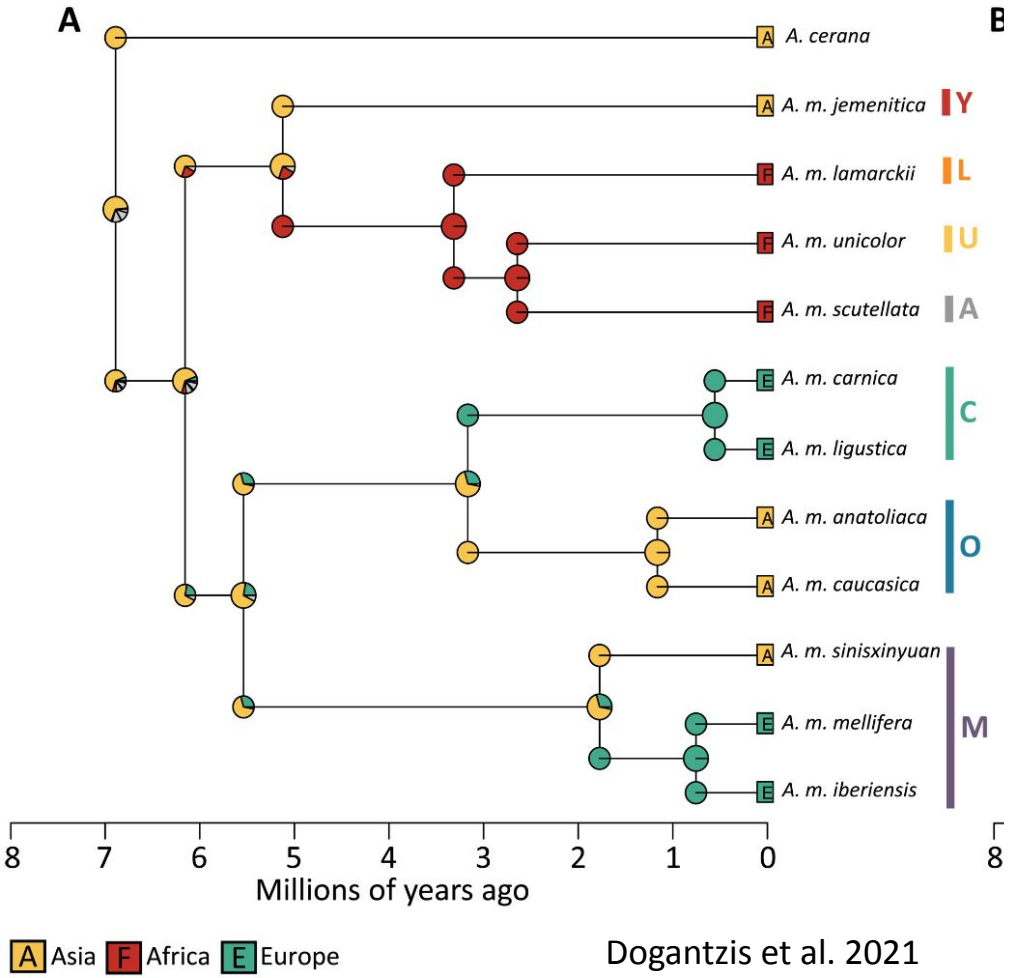
Why study population genetics/genomics?

5. Understand the evolutionary forces that shape life forms

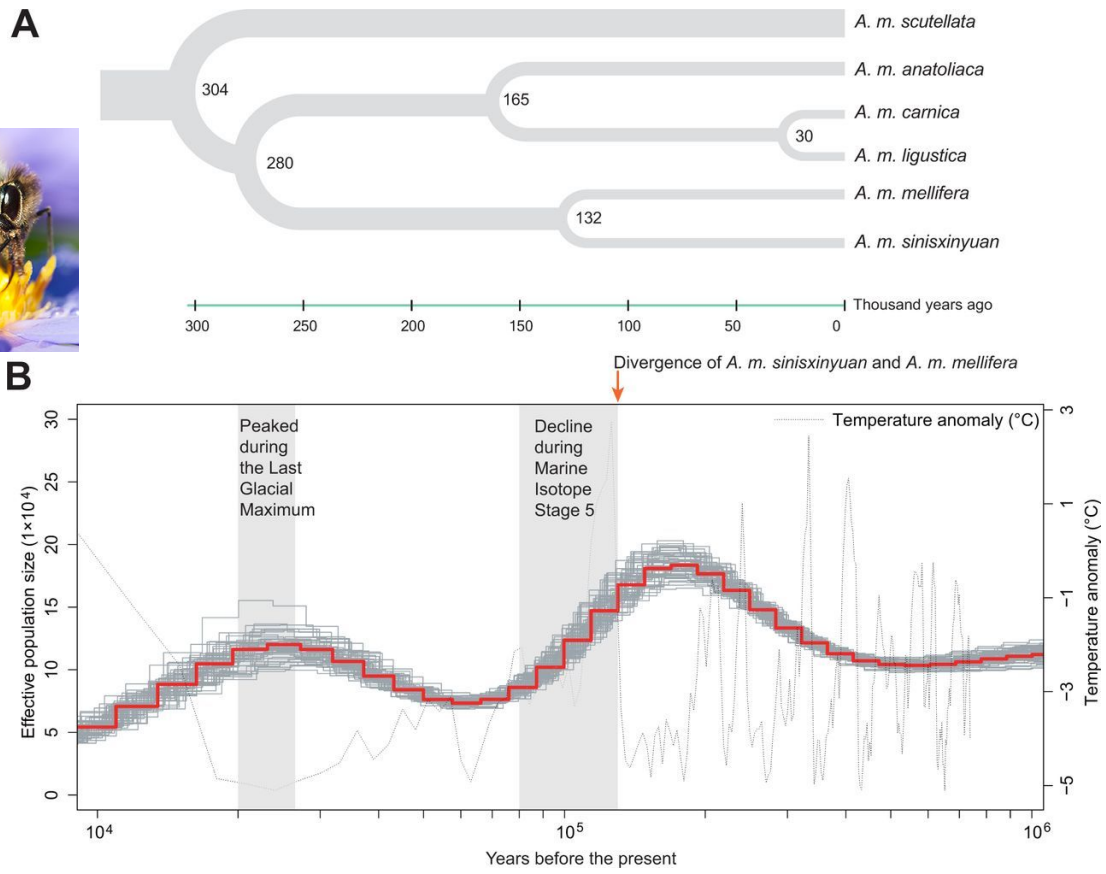


Why study population genetics/genomics?

6. Reconstruct the history and the timing of evolutionary events



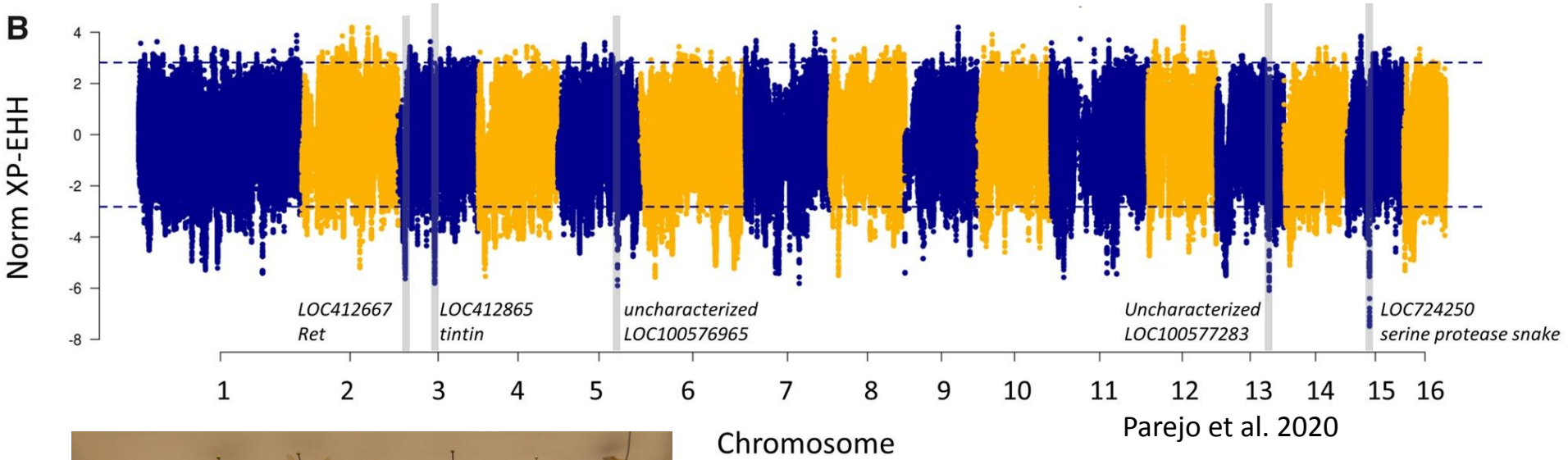
Dogantzis et al. 2021



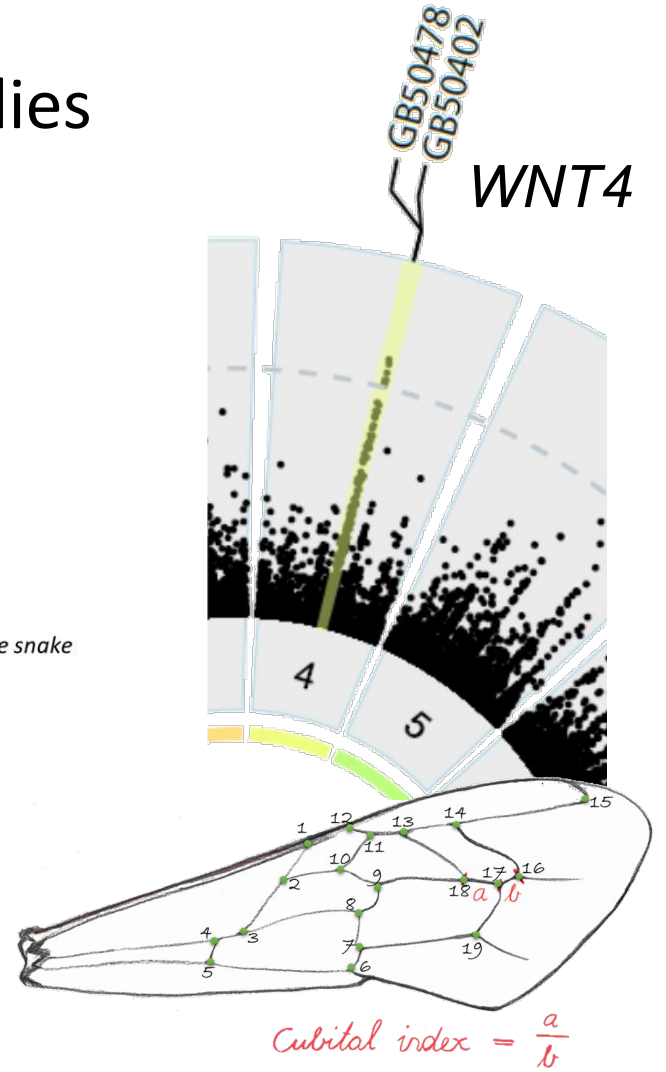
Chen et al.

Why study population genetics/genomics?

7. Selection signatures / Genome-wide association studies



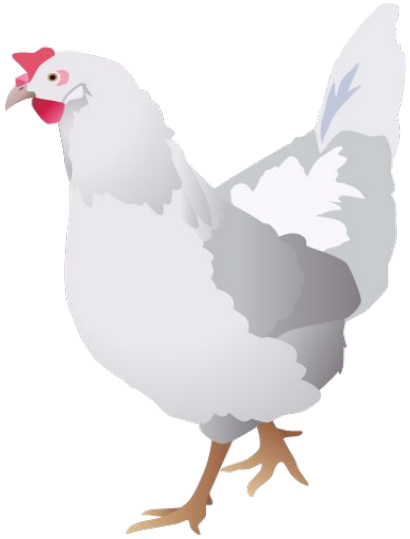
Parejo et al. 2020

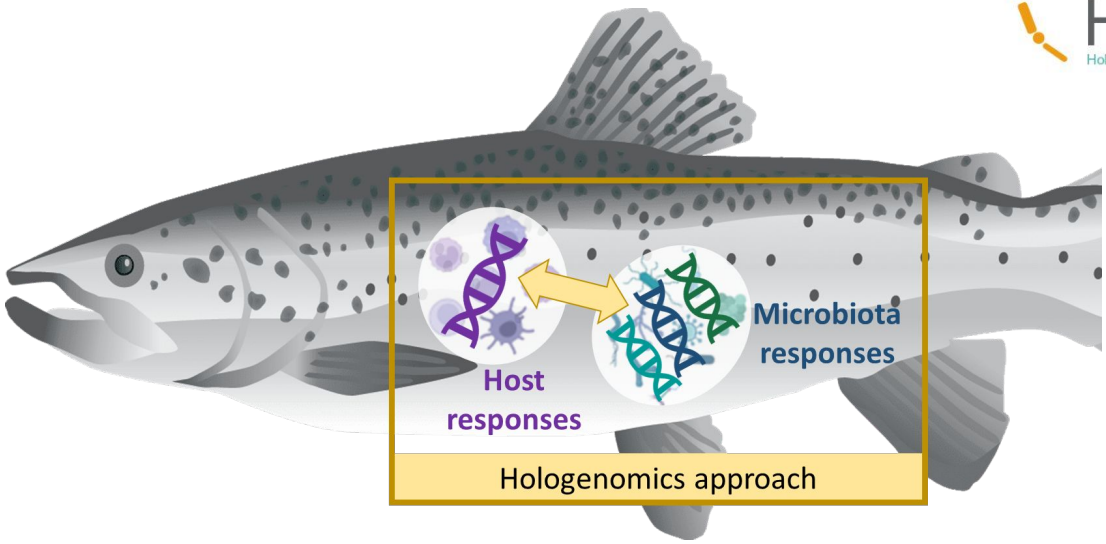


Parejo et al. 2017

Why study population genetics/genomics?

7. Selection signatures / Genome-wide association studies





AIM:
to find associations with production and immunity traits, and most importantly microbiota (hands-on exercise)

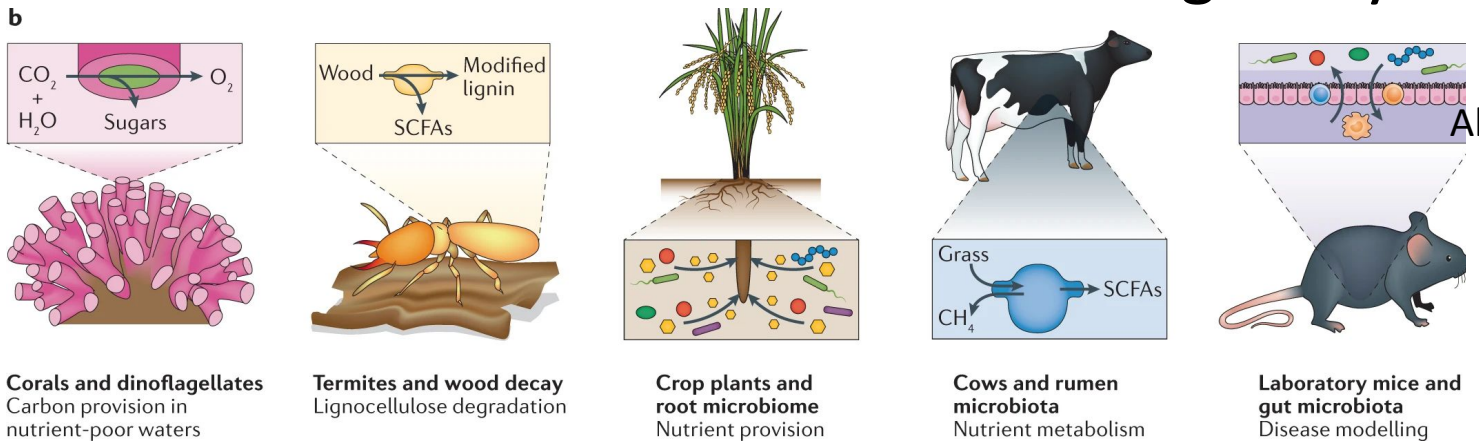
Why study population genetics/genomics?

1. Understand and refine theory
2. Understand the history of genes
3. Understand the history of populations
4. Understand the relationship between individuals
5. Understand the evolutionary forces that shape life forms
6. Reconstruct the history and the timing of evolutionary events
7. Selection signatures / Genome-wide association studies

Redefining evolution 2.0 – the hologenome concept of evolution

Redefining evolution 2.0 – the hologenome concept of evolution

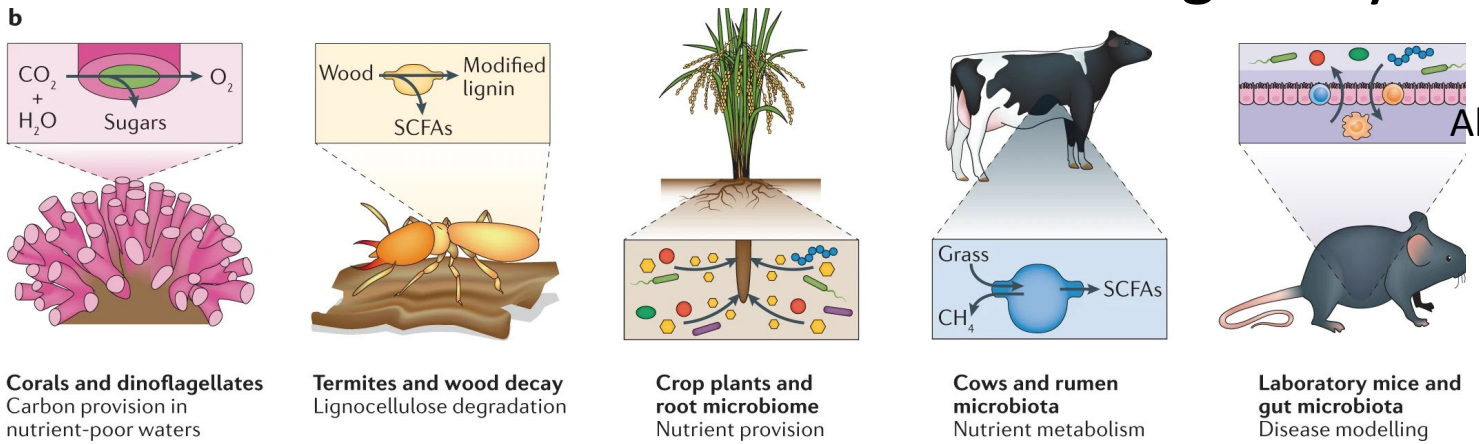
Pervasive host-microbiota interactions in biological systems



Alberdi et al. 2022

Redefining evolution 2.0 – the hologenome concept of evolution

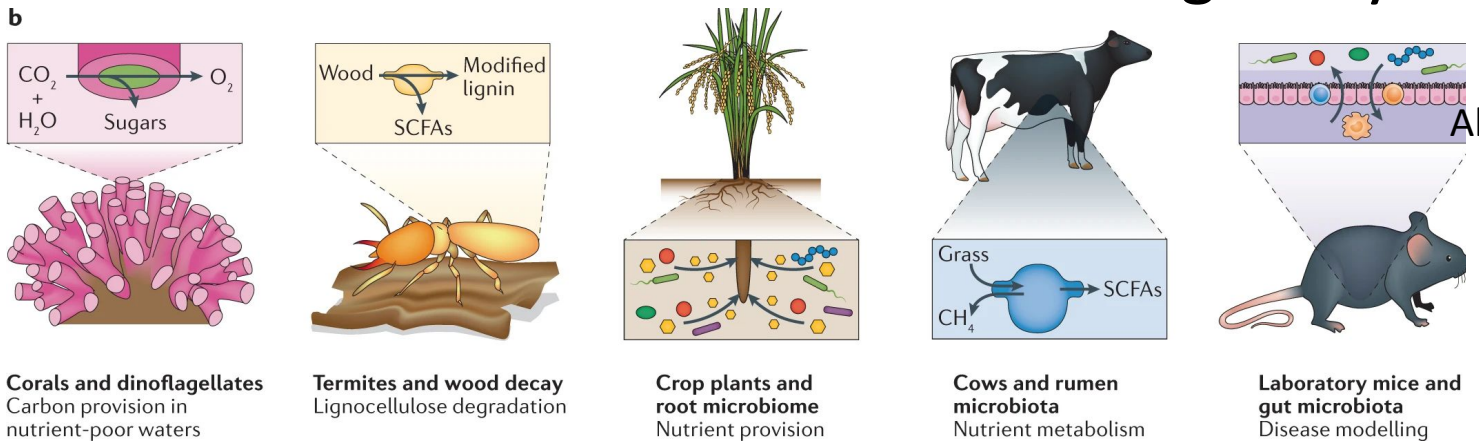
Pervasive host-microbiota interactions in biological systems



Alberdi et al. 2022

Redefining evolution 2.0 – the hologenome concept of evolution

Pervasive host-microbiota interactions in biological systems



Alberdi et al. 2022

 Holobiont and its hologenome as an independent level of selection in evolution

2008

REVIEW ARTICLE

Role of microorganisms in the evolution of animals and plants: the hologenome theory of evolution

Ilana Zilber-Rosenberg[1] & Eugene Rosenberg[2]

[1]Teaching at the Open University of Israel, Raanana, Israel; and [2]Department of Molecular Microbiology and Biotechnology, Tel Aviv University, Ramat Aviv, Israel

Rosenberg and Zilber-Rosenberg *Microbiome* (2018) 6:78
https://doi.org/10.1186/s40168-018-0457-9

Microbiome

REVIEW Open Access

The hologenome concept of evolution after 10 years

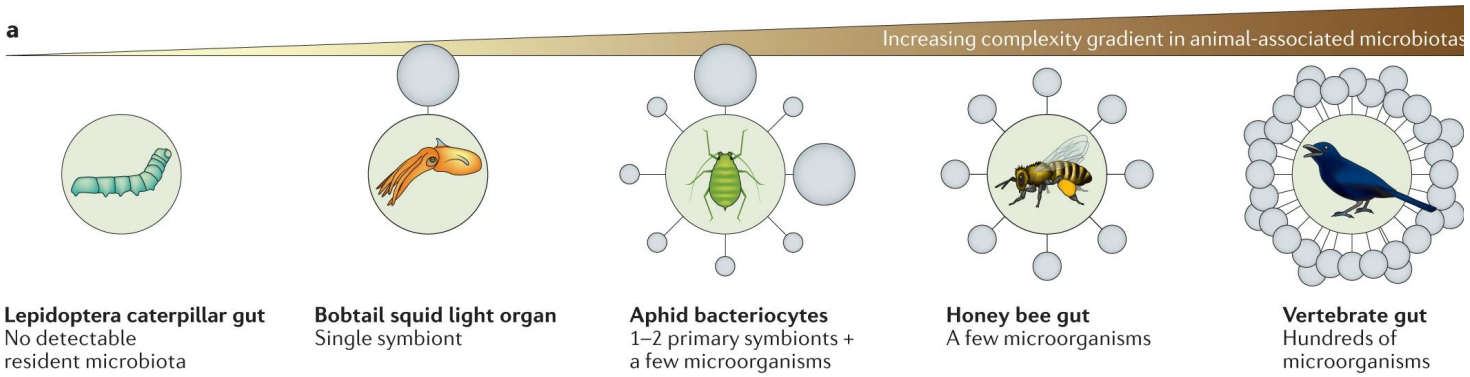
Eugene Rosenberg* and Ilana Zilber-Rosenberg

Hologenome theory of evolution
(Zilber-Rosenberg & Rosenberg, 2008 & 2018)

Four basic principles:

Hologenome theory of evolution
(Zilber-Rosenberg & Rosenberg, 2008 & 2018)

1. All animals and plants harbor abundant and diverse microbiota and are thus considered **holobionts**.
   - Human skin alone 1000 associated bacteria (Ying et al, 2015)
   - Human gut microbiome contains ~900 million unique protein coding genes = 400 x more bacterial than human genes (Yang et al 2009).
   - Core microbiome vs metabolic redundancy

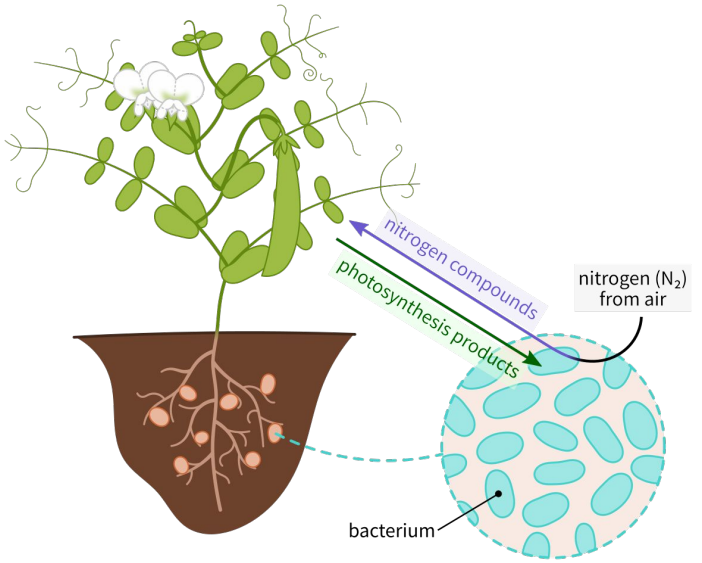


Alberdi et al. 2022

Hologenome theory of evolution
(Zilber-Rosenberg & Rosenberg, 2008 & 2018)

2. The host with its microbiome, the holobiont, functions generally as a distinct **biological entity** anatomically, metabolically, immunologically, during development and in evolution.

- «interactive fitness in holobionts» = beneficial interactions between microbiomes and their hosts lead to a better-adapted holobiont
  - Protection against pathogens, e.g. fecal transplants (Drekonja et al 2015
  - Provision of nutrients, e.g. nitrogen fixation in legumes



Hologenome theory of evolution
(Zilber-Rosenberg & Rosenberg, 2008 & 2018)

3. A significant fraction of the microbiome genome together with the host genome is **transmitted** from one generation to the next and thus can propagate unique properties of the holobiont.

- *Unit of selection in evolution*, = transmission between generations.
  - Breastfeeding
  - Coprophagy (eating mother's feces), e.g. Koala

Hologenome theory of evolution
(Zilber-Rosenberg & Rosenberg, 2008 & 2018)

4. **Genetic variation** in the hologenome can be brought about by changes in the host genome as well as by changes in the microbiome genome.

- Rapid changes in the microbiome genome could allow holobionts to adapt and survive under changing environmental conditions thus providing the time necessary for the host genome to adapt and evolve.
  - via i) amplification or reduction of the number of a specific microbial group;
  - ii) acquisition of novel microbes
  - iii) horizontal gene transfer

From population genomics to hologenomic hypothesis

From population genomics to hologenomic hypothesis

- Biology through the hologenomic lens
  - Does environment outweigh host genetics in determining the microbiota?
  - Do host–microbiota interactions shape host fitness?
  - Do microorganisms shape host evolution?
  - Does the microbiota provide an adaptive buffer to their hosts?
  - Do hosts shape microbial evolution?

Alberdi et al. 2022

PART 2

Pipeline Host Genome Recovery from Metagenomic Datasets

Pipeline Host Genome Recovery from Metagenomic Datasets

Sofia Marcos

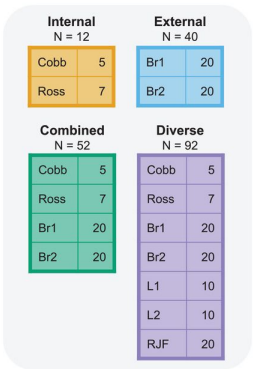
IDEA behind: Take advantage of the same DNA sample to get genomic information for hosts and microbiota.

- Cost-efficient

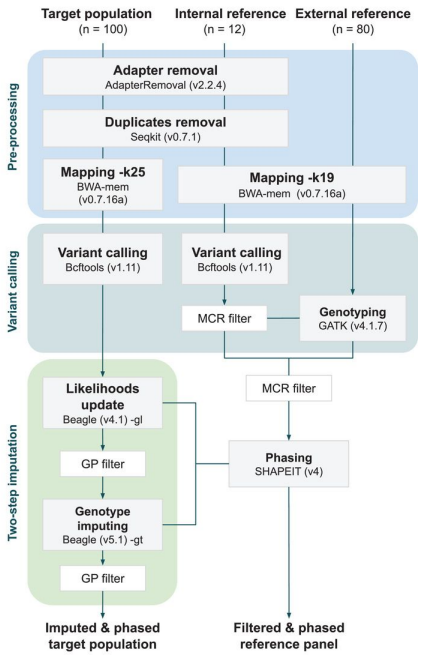
(a) Populations

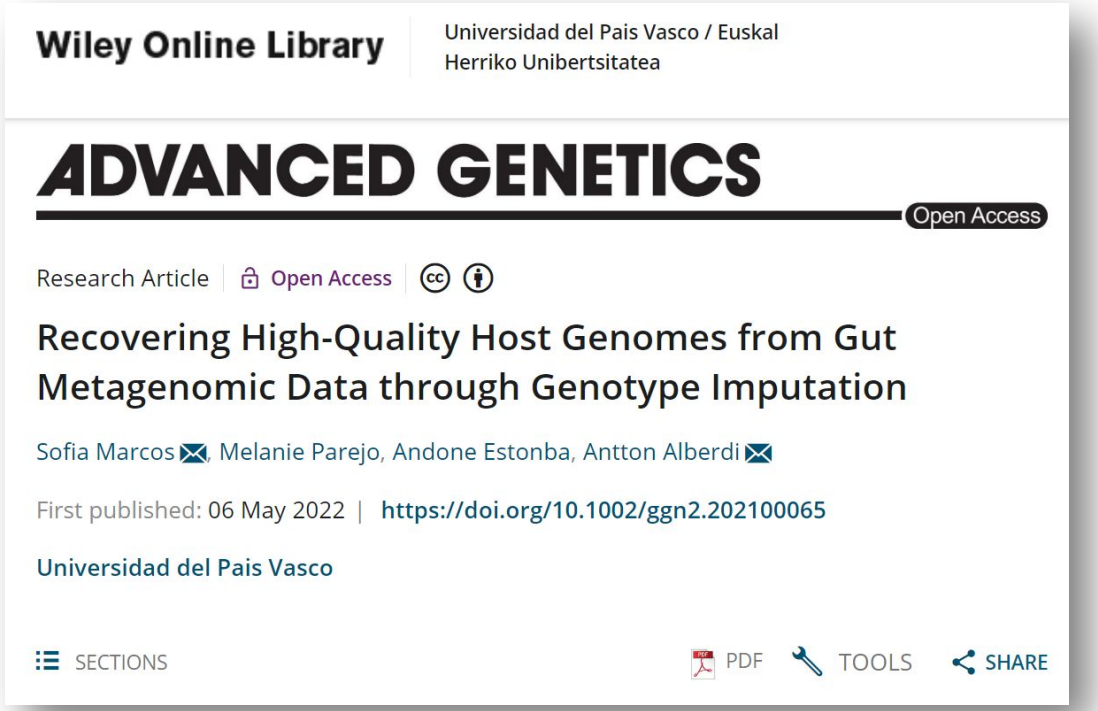
Populations	Target population	Internal reference	External reference
Tissue type	Caecum gut content	Ileum gut content	Blood
Host DNA (%)	~ 5	~ 95	-
Breeds	Cobb (N=47) Ross (N=53)	Cobb (N=5) Ross (N=7)	Br1 (N=20) Br2 (N=20) L1 (N=10) L2 (N=10) RJF (N=20)
Total N	100	12	80

(b) Reference panels



(c) Pipeline





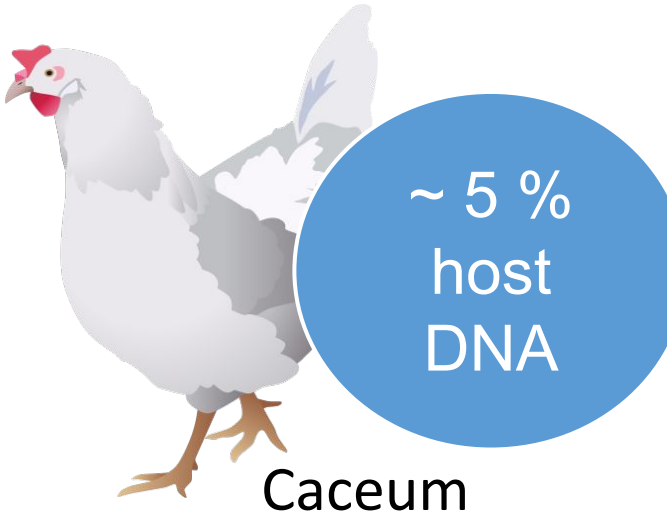
Pipeline Host Genome Recovery from Metagenomic Datasets

Problem:

Metagenomic datasets of host-associated microbial communities often contain variable amounts of host DNA, typically too little for accurate host genetic analyses

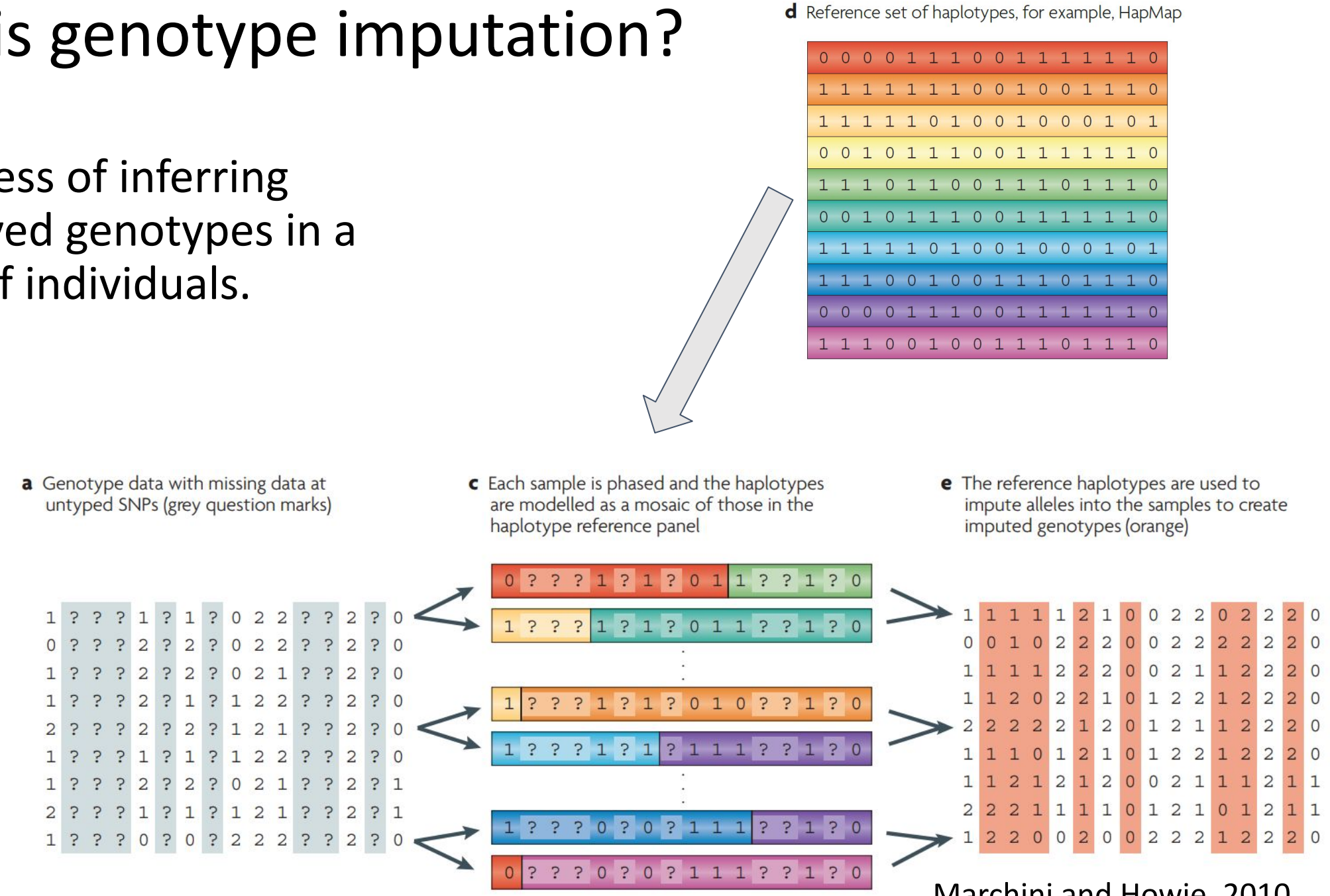
Solution

Genotype imputation can be employed to reconstruct host genotypes (if a reference panel is available).



What is genotype imputation?

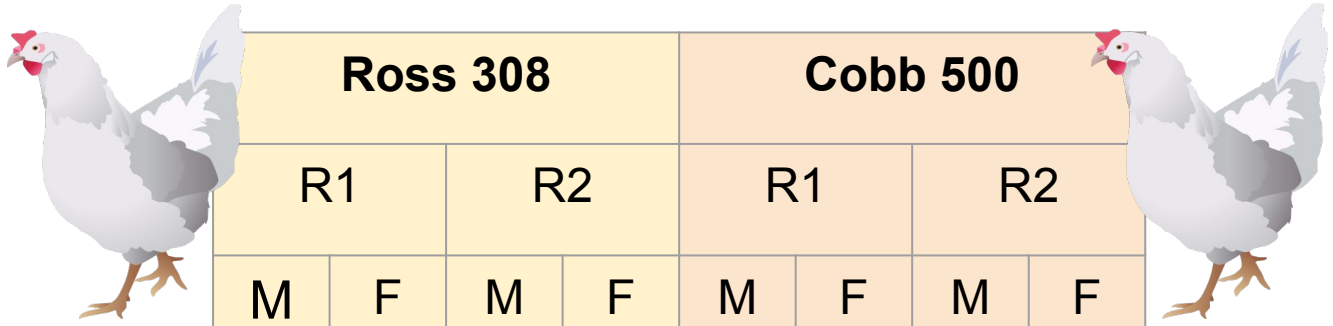
The process of inferring unobserved genotypes in a sample of individuals.



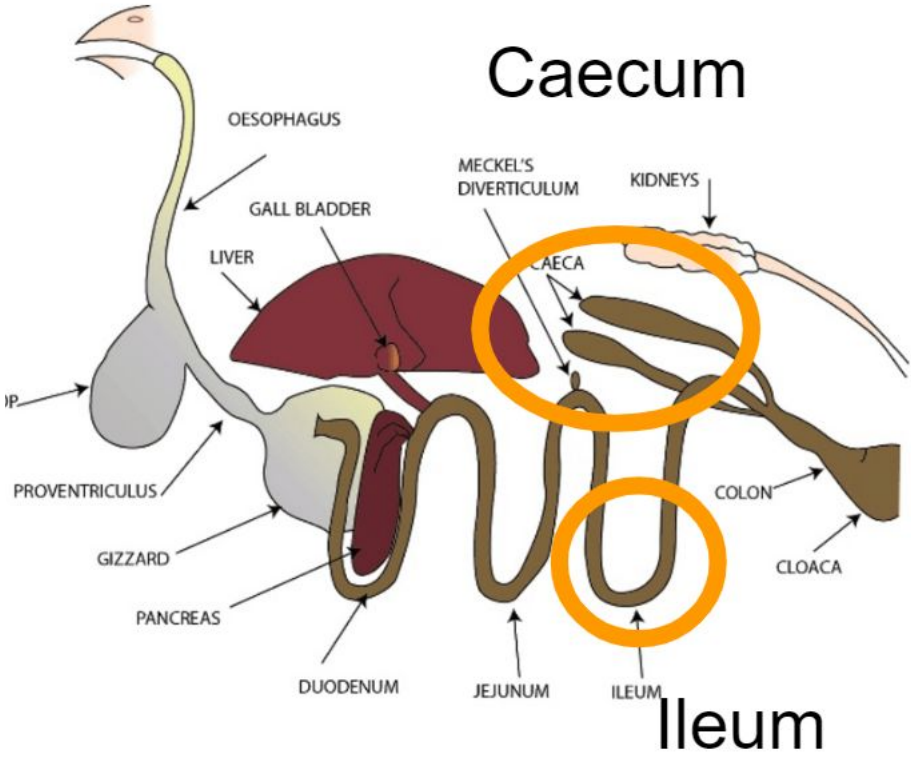
Marchini and Howie, 2010

Pipeline Host Genome Recovery from Metagenomic Datasets

- Holofood chicken genomics data
  - 2 broiler lines
    - Ross 308 - Aviagen
    - Cobb 500 - Cobb-Vantress
  - 2 origins for each line
  - 2 biological sexes



Ross 308				Cobb 500			
R1		R2		R1		R2	
M	F	M	F	M	F	M	F



Pipeline Host Genome Recovery from Metagenomic Datasets

- Populations and tested reference panels

(a) **Populations**

Populations	Target population	Internal reference	External reference
Tissue type	Caecum gut content	Ileum gut content	Blood
Host DNA (%)	~ 5	~ 95	-
Breeds	Cobb (N=47) Ross (N=53)	Cobb (N=5) Ross (N=7)	Br1 (N=20) Br2 (N=20) L1 (N=10) L2 (N=10) RJF (N=20)
Total N	100	12	80

Reference panels

Internal
N = 12

Cobb	5
Ross	7

External
N = 40

Br1	20
Br2	20

Combined
N = 52

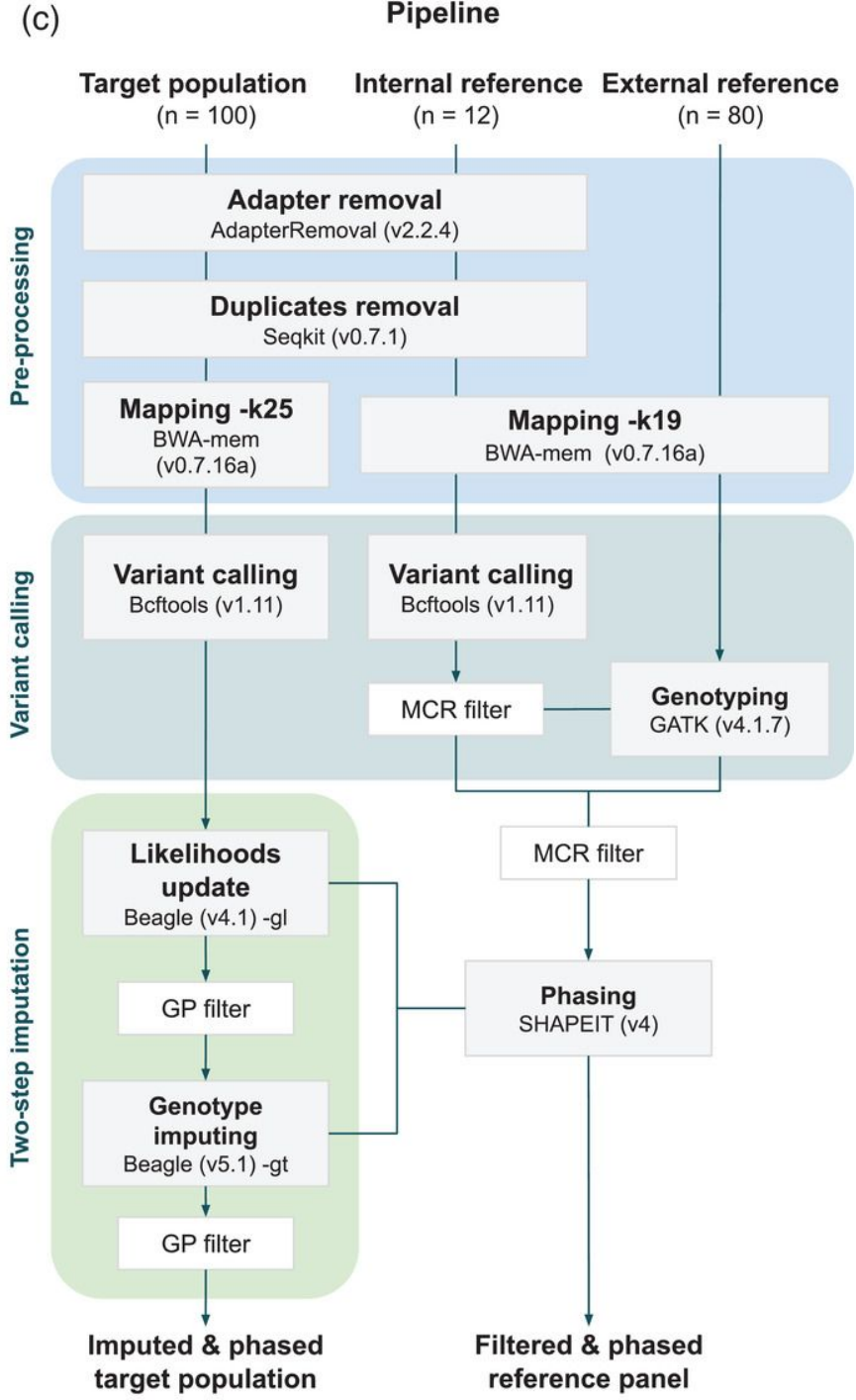
Cobb	5
Ross	7
Br1	20
Br2	20

Diverse
N = 92

Cobb	5
Ross	7
Br1	20
Br2	20
L1	10
L2	10
RJF	20

Pipeline Host Genome Recovery from Metagenomic Datasets

- Two-step imputation
  - Likelihood update (Beagle v4.1)
  - Filter on genotype probability
  - Impute genotypes (Beagle v5.1)
  - Filter again on genotype probability

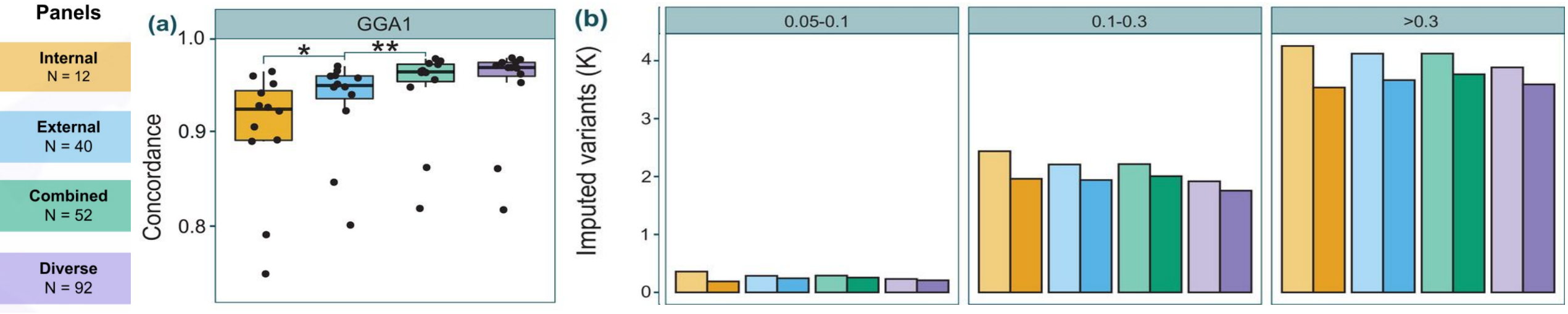


Pipeline Host Genome Recovery from Metagenomic Datasets

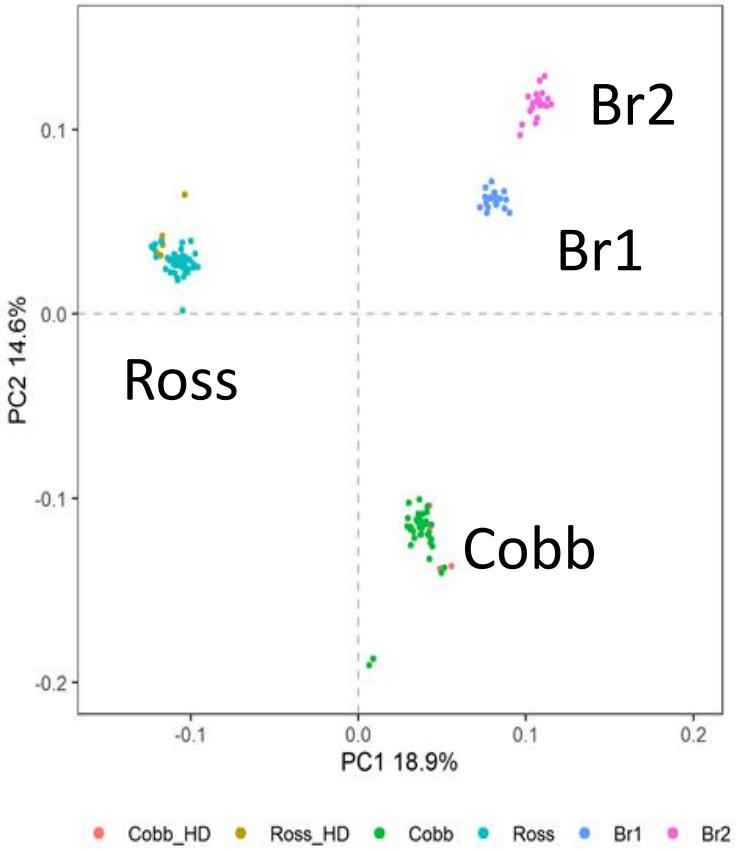
- Validation with 12 samples for which we have high and low sequence data.

Genotype concordance overall

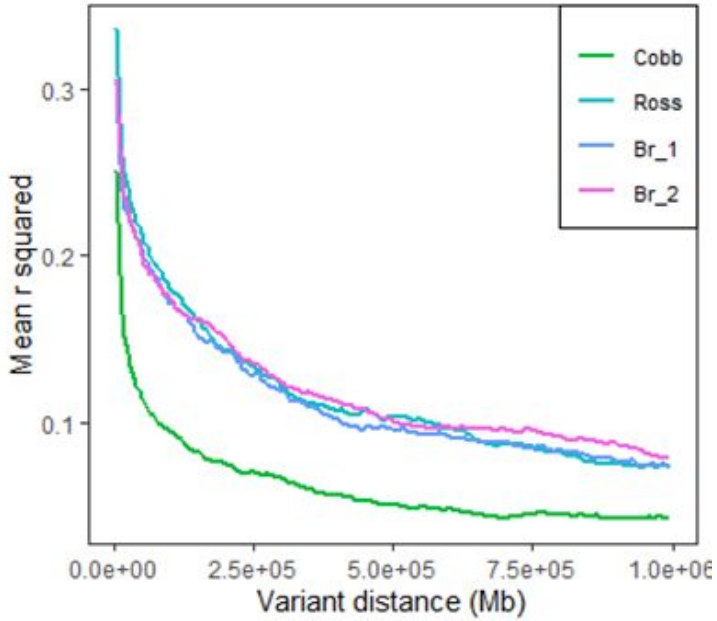
Number of imputed and correctly imputed variants In allele frequency bins

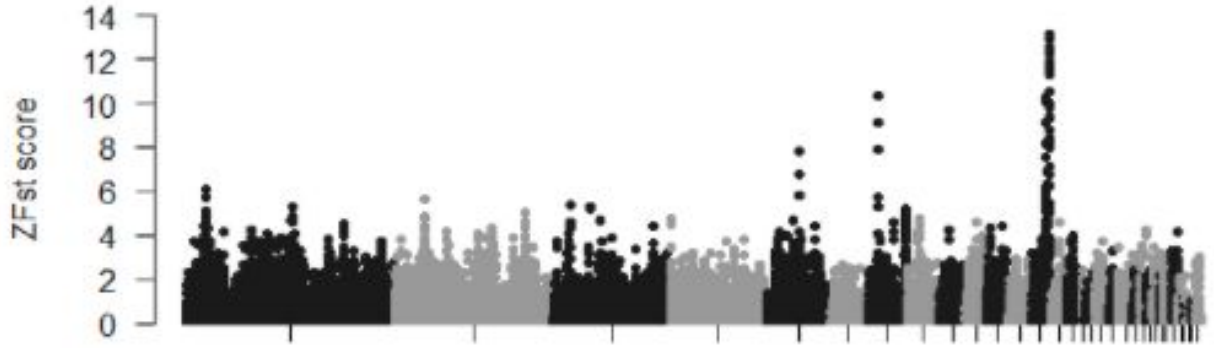


Downstream genetic analyses (ongoing work)



- PCA
- Population Structure
- Genetic Diversity
- Selection Signatures
- mGWAS
- ….





Applications

System	Pop. characteristics	N	Sample types	Host DNA	Metagenomic dataset	Ref. panel	Downstream analysis
Buffalo	River, swamp, and hybrid buffaloes	695	Gut, intestine, and rectum	<20%	[65]	[66]	Selection signatures
Cattle	Three crossbreeds and one pureline	282	Gut	*3%	[67]	[11]	Selection signatures
Pig	Various breeds	470	Fecal	*2%	[70]	[69]	Selection signatures
Chicken	Lohmann Brown and Silkie hens	90	Fecal	8%	[71]	Custom[57]	Selection signatures
Chicken	Red Junglefowl	51	Fecal	49%	[72]	Custom[57]	Implication on domestication
Rat	Sprague Dawley	49	Fecal	11%	[73]	Custom[74]	Host-microbiota association
Rat	SpragueDawley	84	Cecal	*51%	[75]	Custom[74]	Host-microbiota interactions
Mouse	Various breeds	184	Fecal	9%	[76]	Custom[77]	Differences between populations
Mouse	C57BL/6J	88	Fecal	<5%	[78]	Custom[77]	Differences between populations
Zebrafish	Single cohort	29	Fecal	*9%	[79]	Custom[80]	Population genetic inference
Honey bee	Eastern and Western honey bees	40	Gut	<10%	[81]	[82]	Differences between species

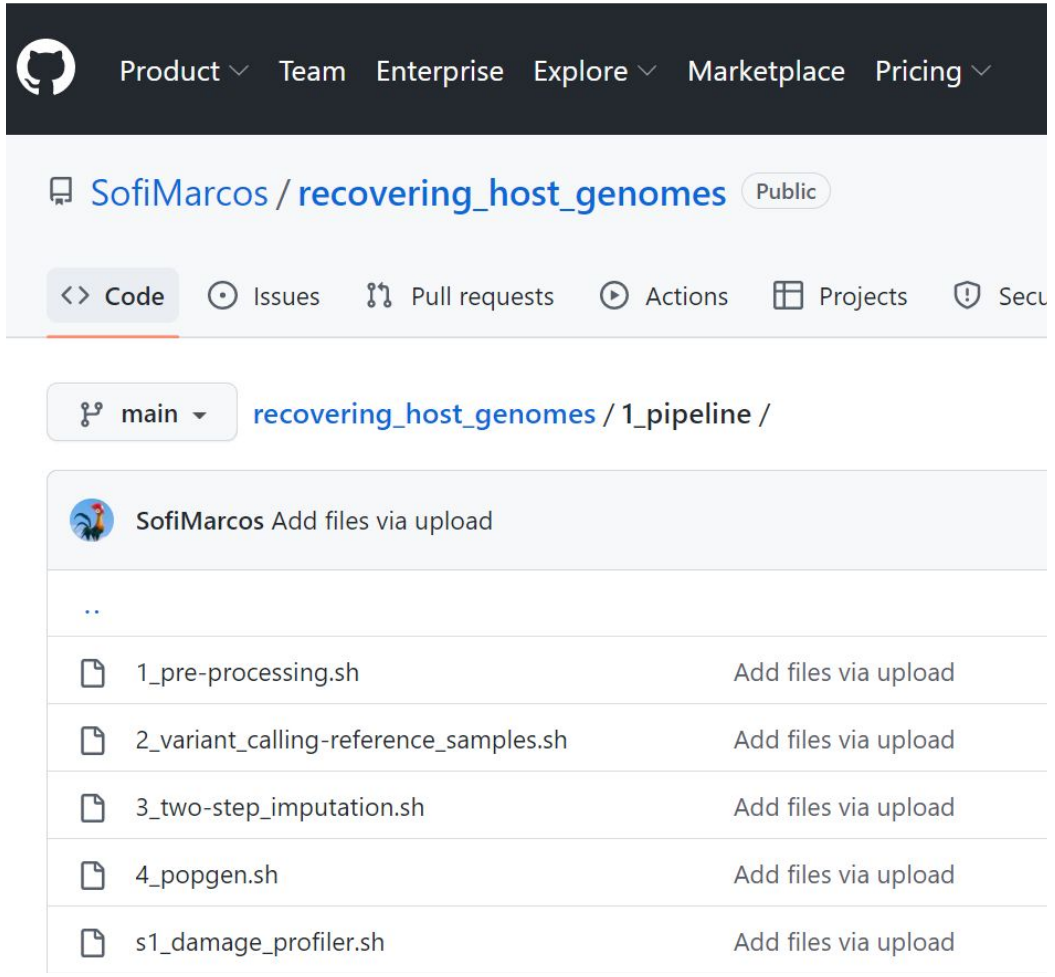
To consider for your studies

- Is there a reference genome available?
- Are there published sequences available to be used as a reference panel?
- How far/close related are the available references?
- What is the depth of coverage of your data? (low pass sequencing starts from 2x)
- Consider generating a custom internal reference panel, by deep-sequencing selected key individuals.
- Optional: test different imputation software

Pipeline Genome Recovery from Metagenomic Datasets

- Pipeline available in Github:

https://github.com/SofiMarcos/recovering_host_genomes/tree/main/1_pipeline



Poster ID 37

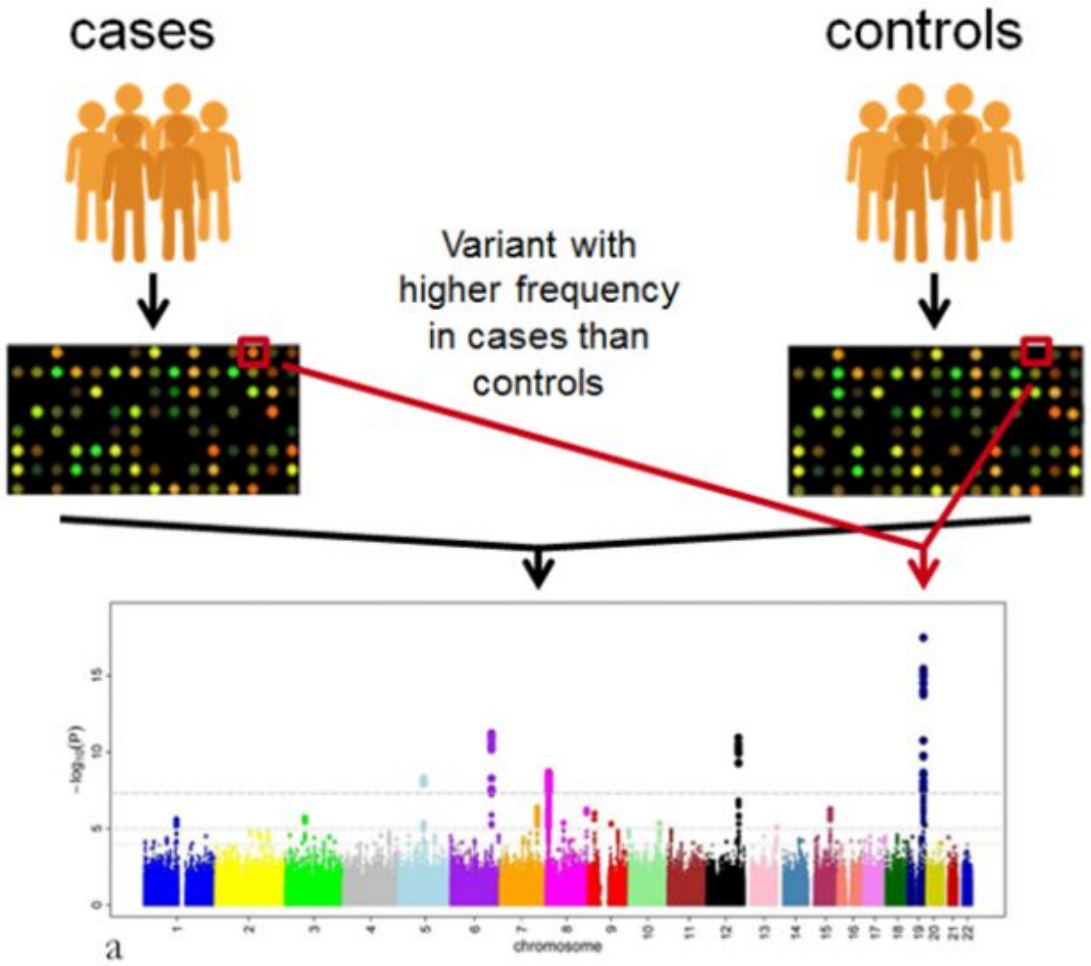
APPLIED HOLOGENOMICS CONFERENCE

Sofia Marcos

PART 3

- HANDS ON SESSION: microbiome-GWAS on salmon

GWAS



Trait (phenotype)

- disease cases vs controls
- height
- …

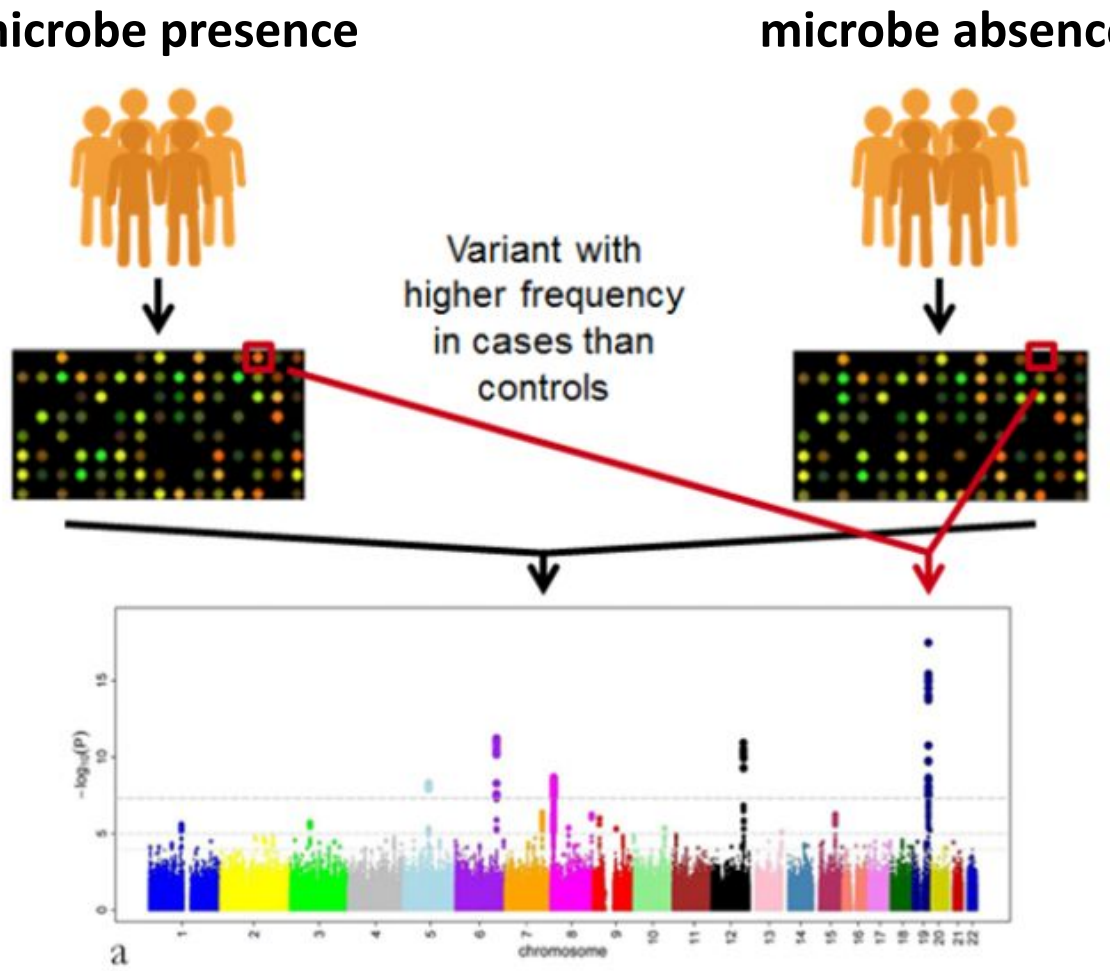
Genotyping information

- SNP array
- whole genome sequencing

Results

- SNPs associated with trait

microbiome (M)GWAS



Trait: microbiome composition
- abundance
- detection
- species, genus, family …

Genotyping information
- SNP array
- whole genome sequencing

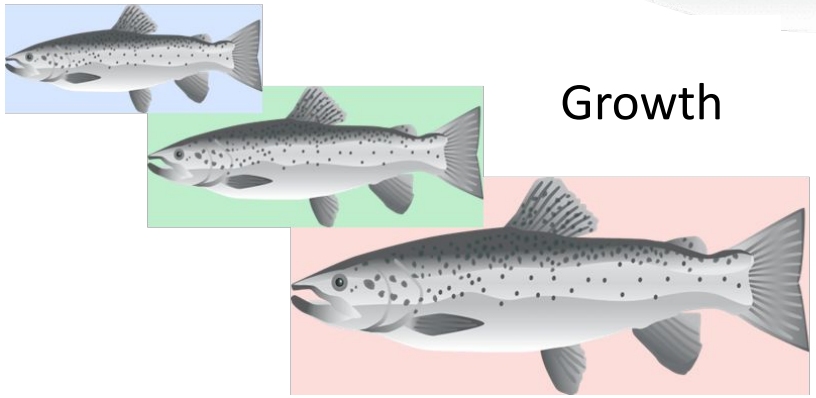
Results
- SNPs associated with trait

~~HoloFood~~ HoloFish dataset

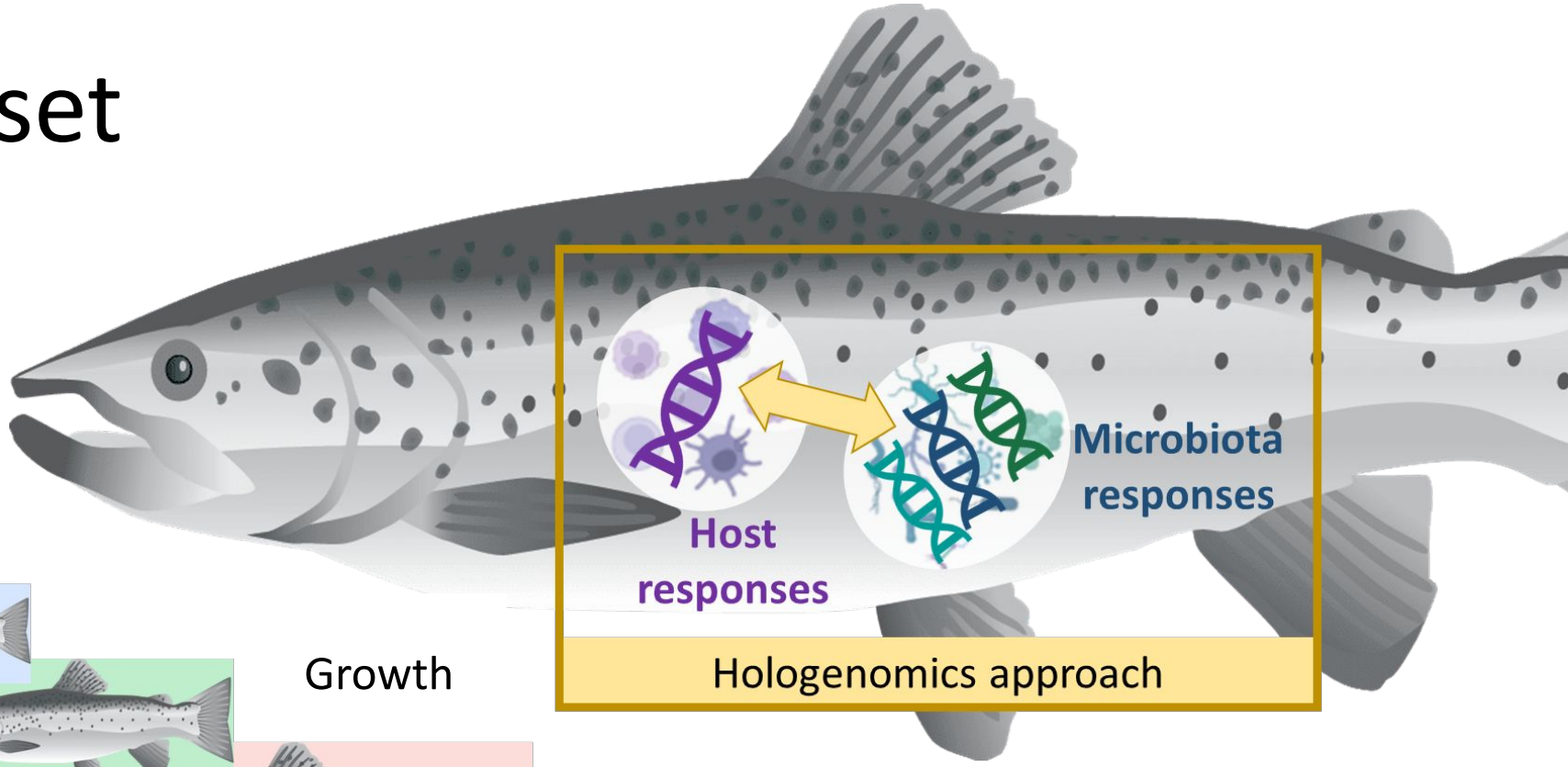
Dataset description

HoloFish dataset

Parasite infection

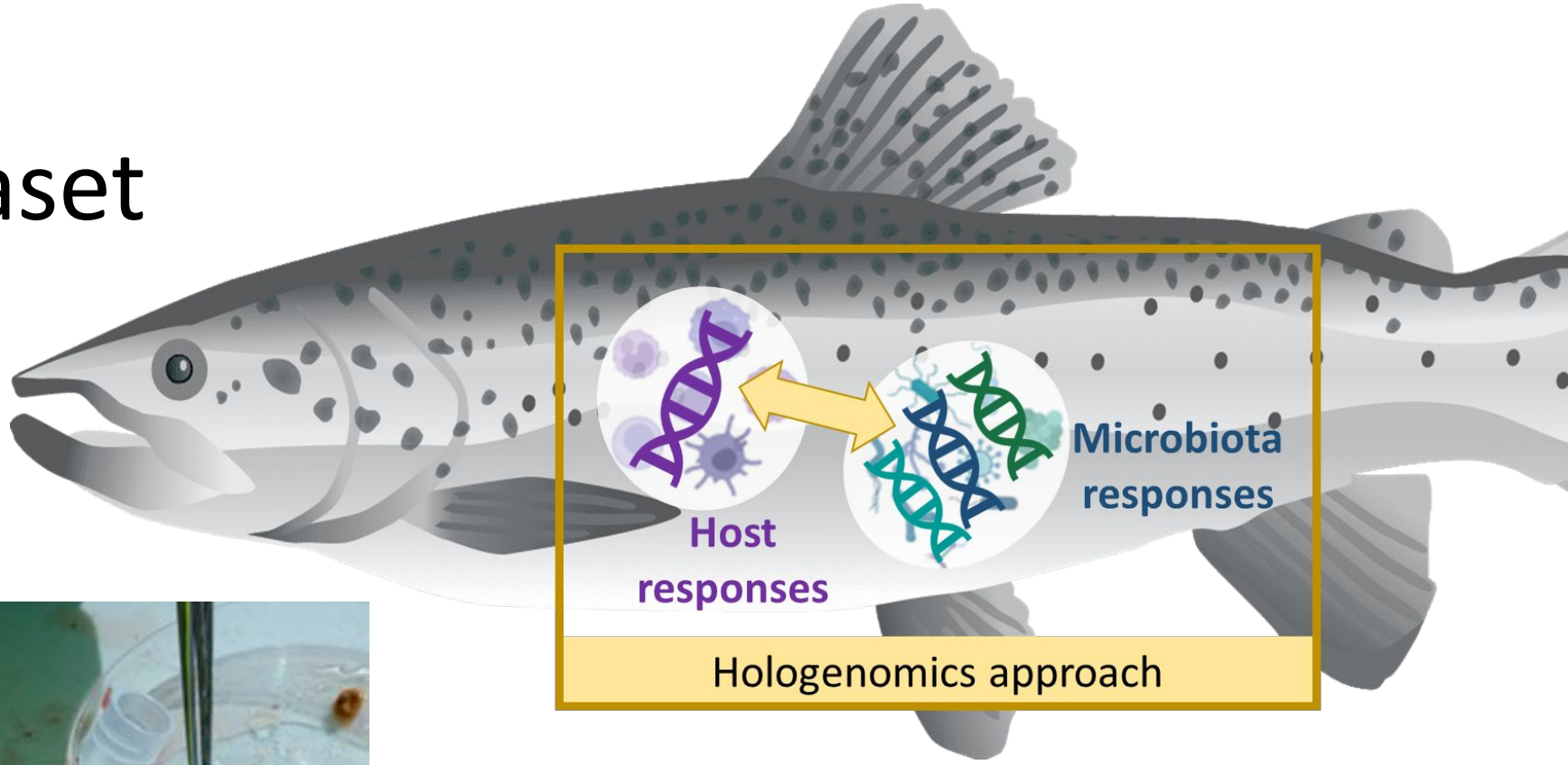


Growth



Dataset description

HoloFish dataset



microbiome: shotgun metagenomics

host genetics: genomics

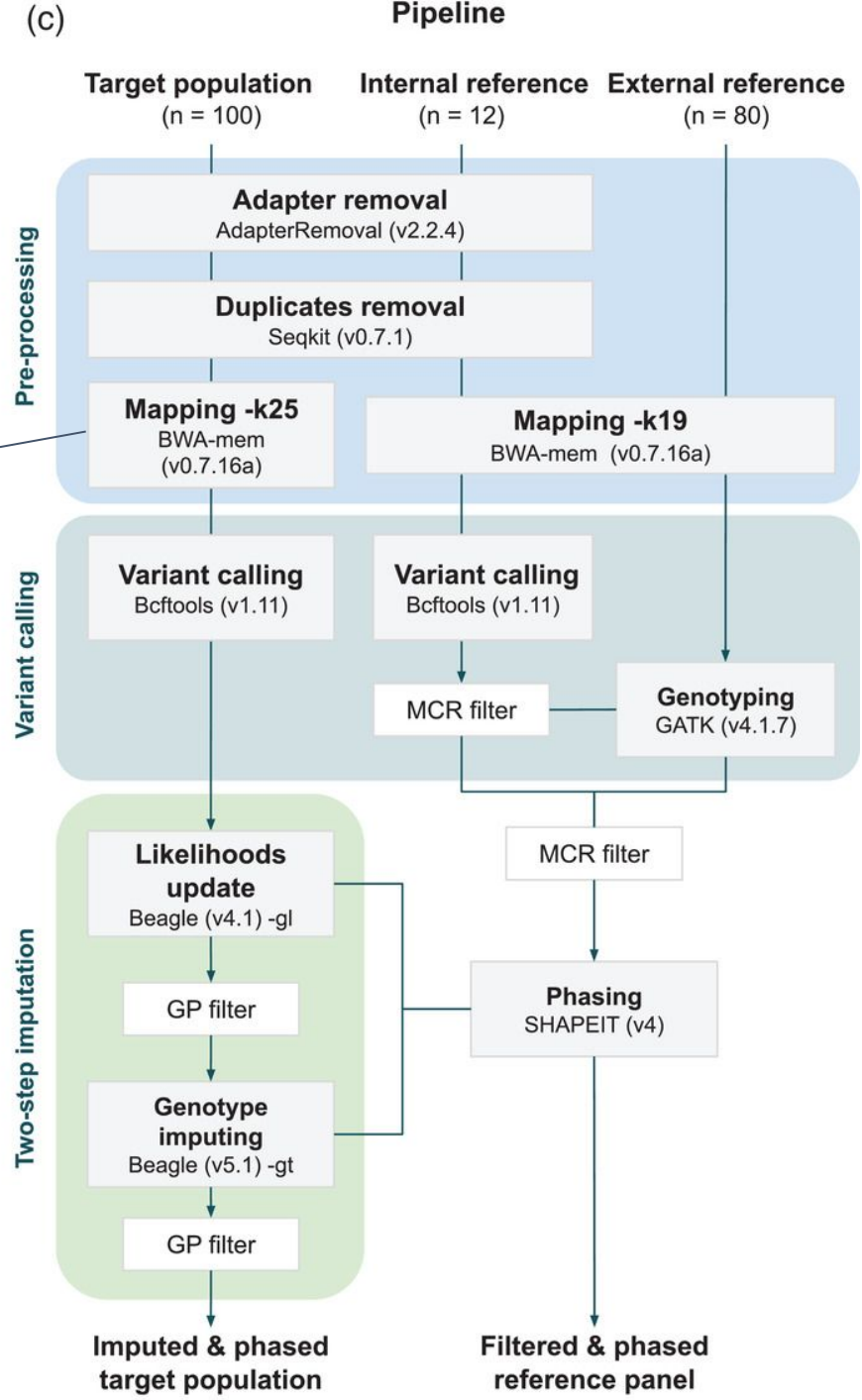
Processing - host genomics

Identify callable regions
GATK

Genotype likelihoods
ANGSD

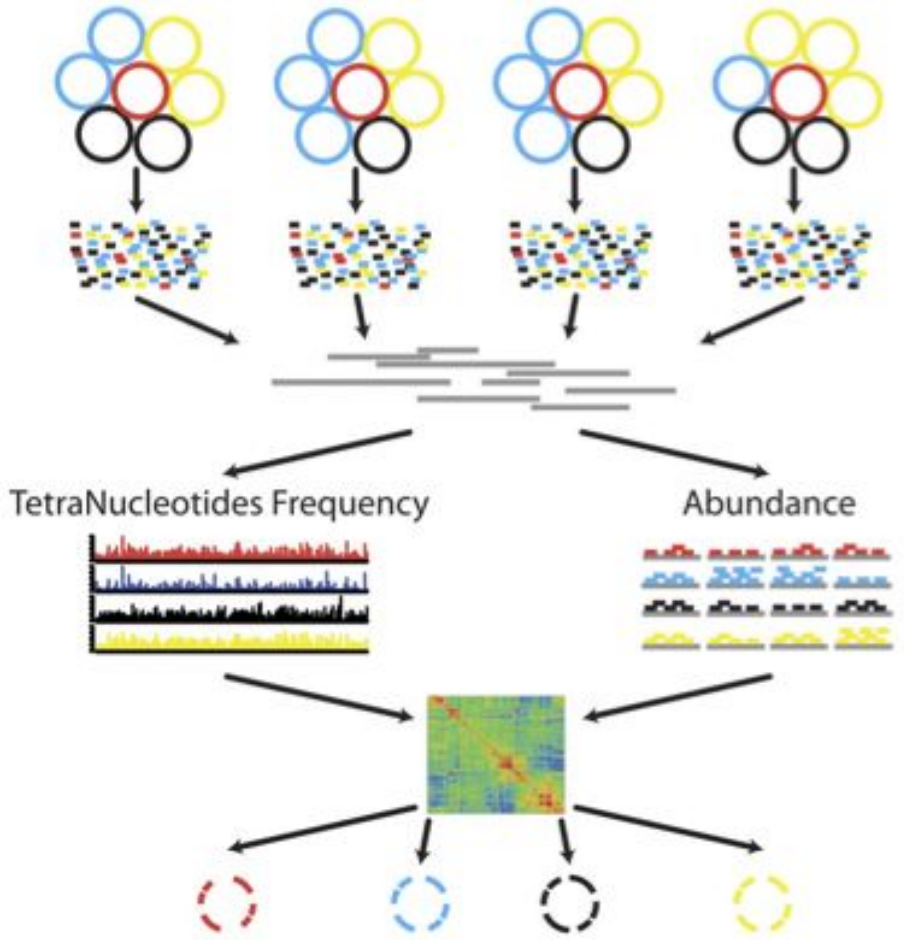
LD pruning
plink

Genotype probabilities
Beagle (v3.3.2)



Dataset description

Processing - metagenomics



- MEGAHIT
- metaSPADES

- MetaBAT
- maxbin
- concoct

- Anvio

MetaBAT workflow
(Kang, et al. PeerJ 2015)

Host variation data practical session

Hint

This practical session uses software available on the course-provided virtual machines. To follow this workshop at a later date, see the github repo for installation instructions.

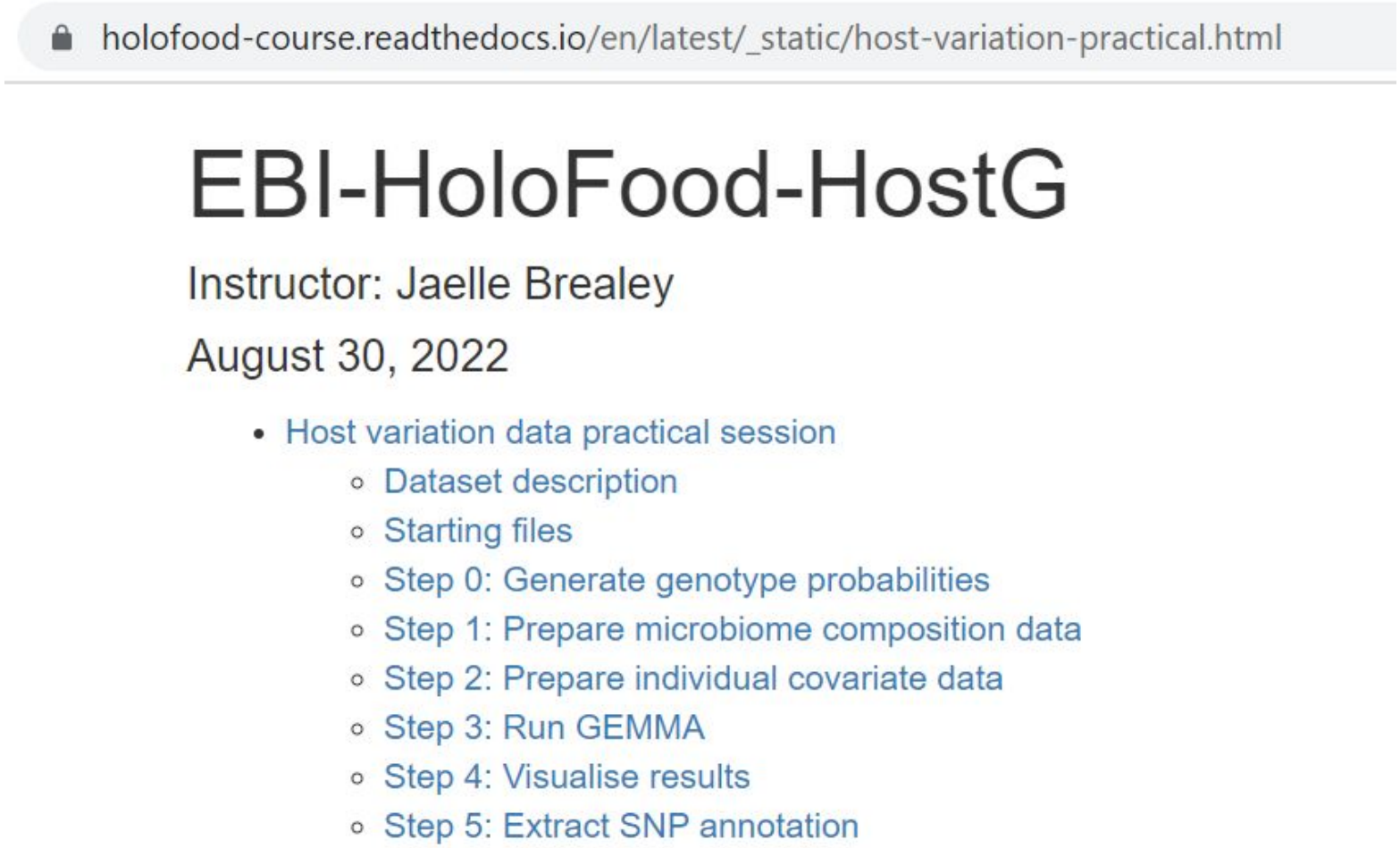
microbiome-GWAS using GEMMA

1. Prepare microbiome composition data
2. Prepare individual covariate data
3. Run GEMMA
4. Visualise GWAS results
5. Extract SNP annotation

Prerequisites

For this tutorial you will need to first load the conda environment by running:

```
conda activate mgwas-env
```



https://holofood-course.readthedocs.io/en/latest/_static/host-variation-practical.html