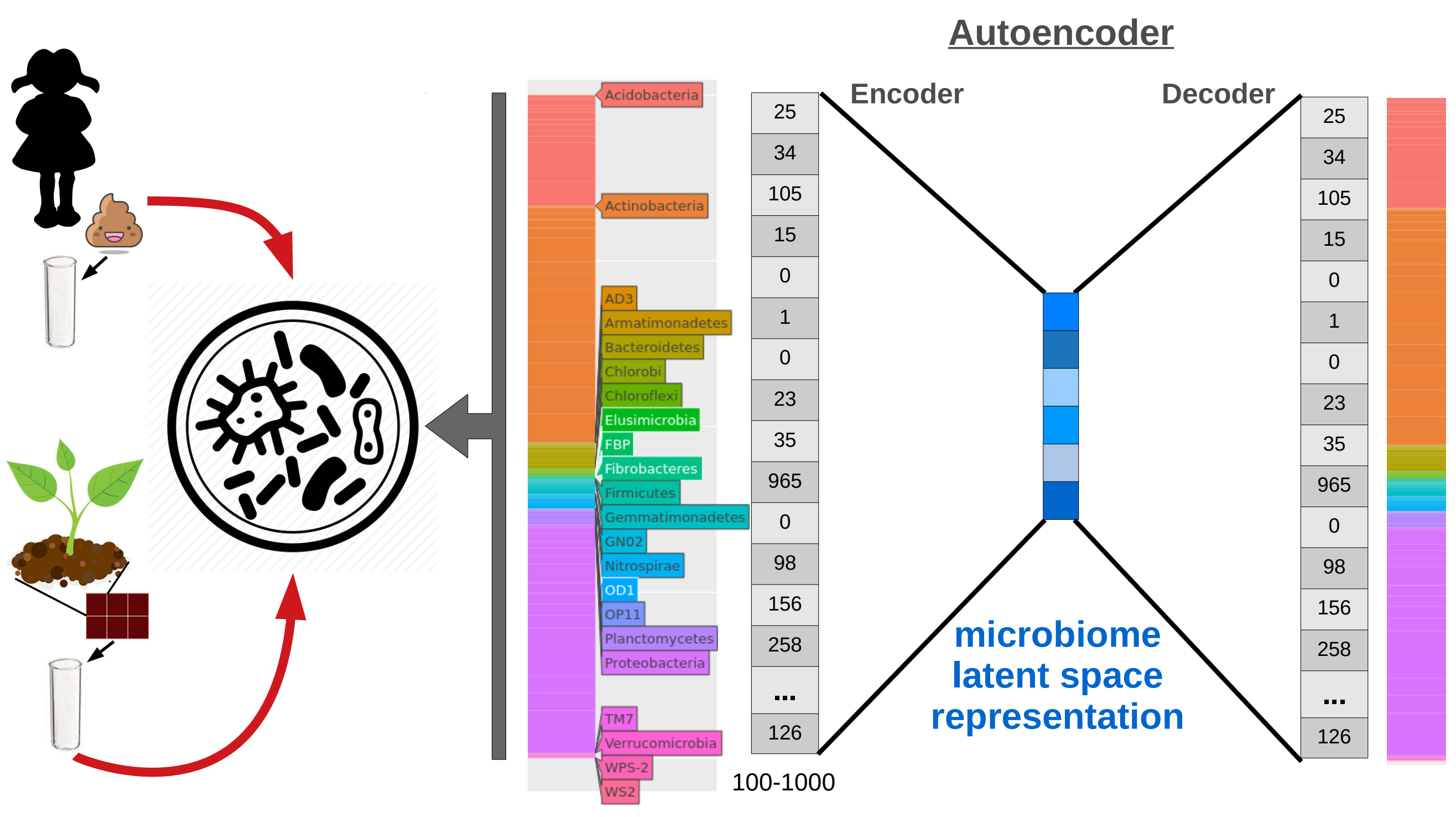# Condensed Microbiome Representation using Transfer and Deep Learning to Promote Microbial Composition Prediction
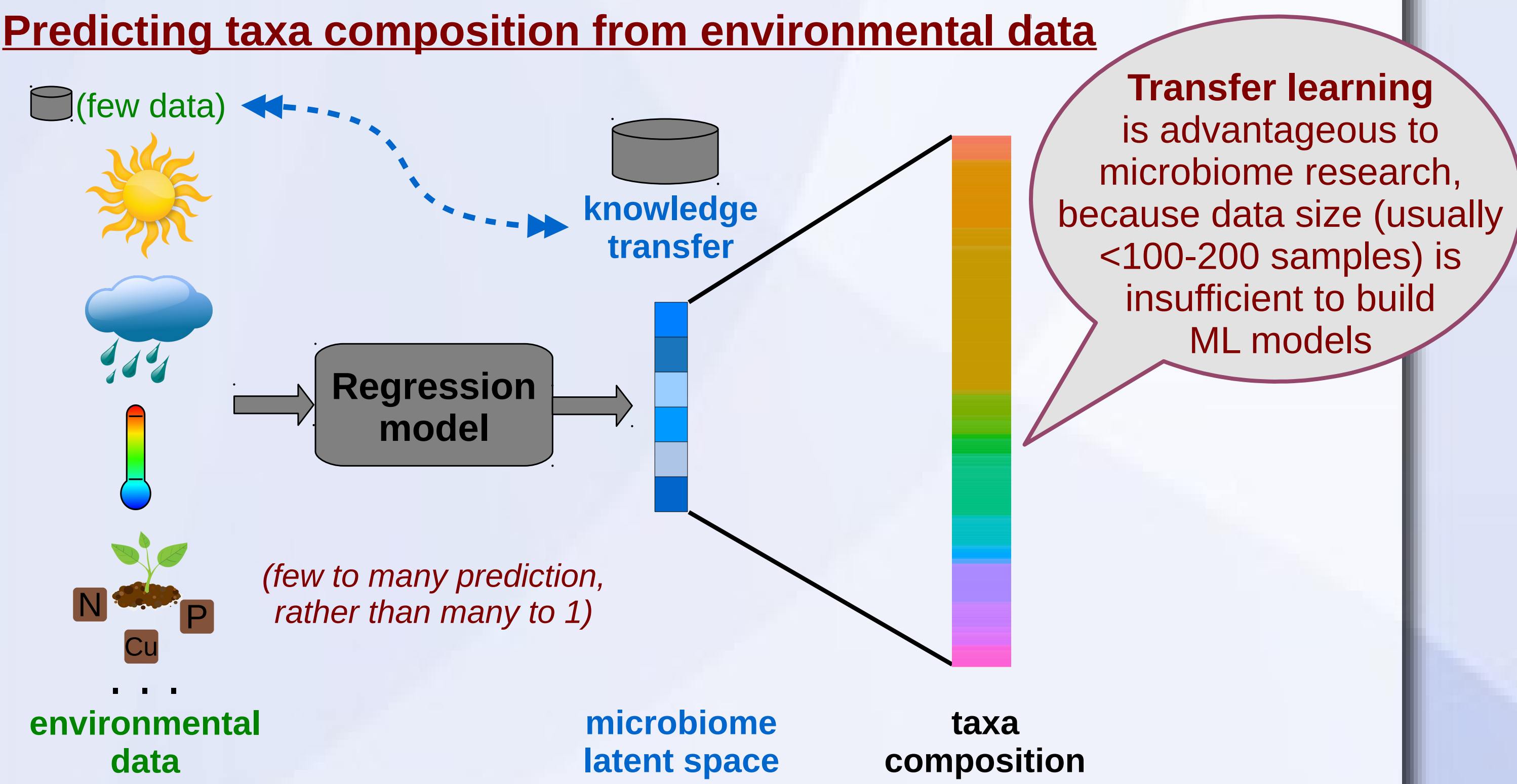
Sara Cabello, Beatriz García-Jiménez and Mark D. Wilkinson

- <u>Motivation</u>: Reducing the dimensionality of microbiome datasets to apply knowledge discovery approaches
- <u>Methods</u>: We selected a Deep Learning architecture - an autoencoder - to condense a long vector of OTU abundances (hundreds), that describes a microbiome sample, into an encoded representation (< 10 values)
- <u>Results</u>: We transfer knowledge from a published dataset [Walters et al., 2018] of around 5000 maize root microbiome samples into our autoencoder model, reducing from 717 taxa to a latent space of 6 rational numbers representing the microbial composition. Additionally, we predict microbial composition of maize root microbiome using a few environmental variables (plant age, temperature or precipitation)

Keywords: Microbiome, Deep Learning, Transfer Learning, Autoencoder, Dimensionality reduction
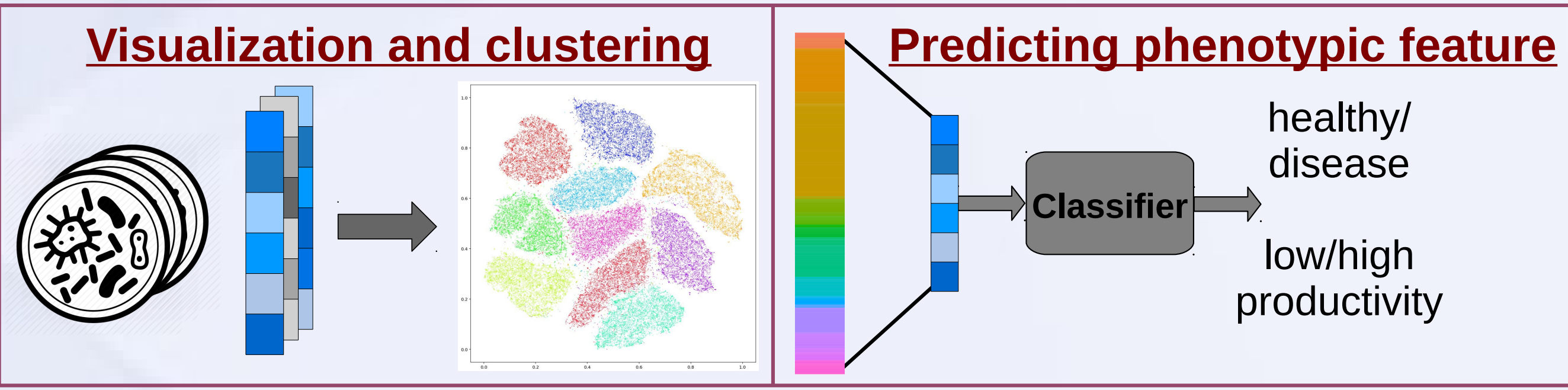
Method

Autoencoder

Encoder | Decoder

microbiome latent space representation

100-1000

Applications

Predicting taxa composition from environmental data

(few data) → knowledge transfer

Regression model

Transfer learning is advantageous to microbiome research, because data size (usually <100-200 samples) is insufficient to build ML models

(few to many prediction, rather than many to 1)

environmental data → microbiome latent space → taxa composition

Case study: maize rhizosphere

- Walters et al., 2018: doi:10.1073/pnas.1800918115
- Study the influence of microbiome in agronomically important plants (maize: *Zea mays L. subsp.mays*)
- 4,855 samples & 717 OTUs
- Mapping variables:
  - elevation, T, precipitation, plant age, maize line and variety

Visualization and clustering

Predicting phenotypic feature

Classifier → healthy/disease, low/high productivity

Results

Latent space performance

Avg. performance	MSE	SMAPE	Pearson corr.
All OTUs (717)	0.0007	65.61%	0.57
OTUs corr.>0.5 (458)	0.0009	53.48%	0.77

GENUS COMPOSITION: Maize_Line = Popcorn

Original data — Relative abundance

Transformed data — Relative abundance

Good / Bad sample predictions

Samples

GENUS:
g__ , g__A17, g__Actinoplanes, g__Aeromicrobium, g__Amycolatopsis, g__Bradyrhizobium, g__Candidatus Solibacter, g__Cellulomonas, g__DA101, g__Devosia, g__Fimbriimonas, g__Flavisolibacter, g__Fluviicola, g__Gemmatimonas, g__Hyphomicrobium, g__Knoellia, g__Luteolibacter, g__Nitrobacter, g__Nocardioides, g__Pedobacter, g__Pedomicrobium, g__Phenylobacterium, g__Phycicoccus, g__Pilimelia, g__Pseudomonas, g__Rhodoplanes, g__Saccharothrix, g__Skermanella, g__Steroidobacter
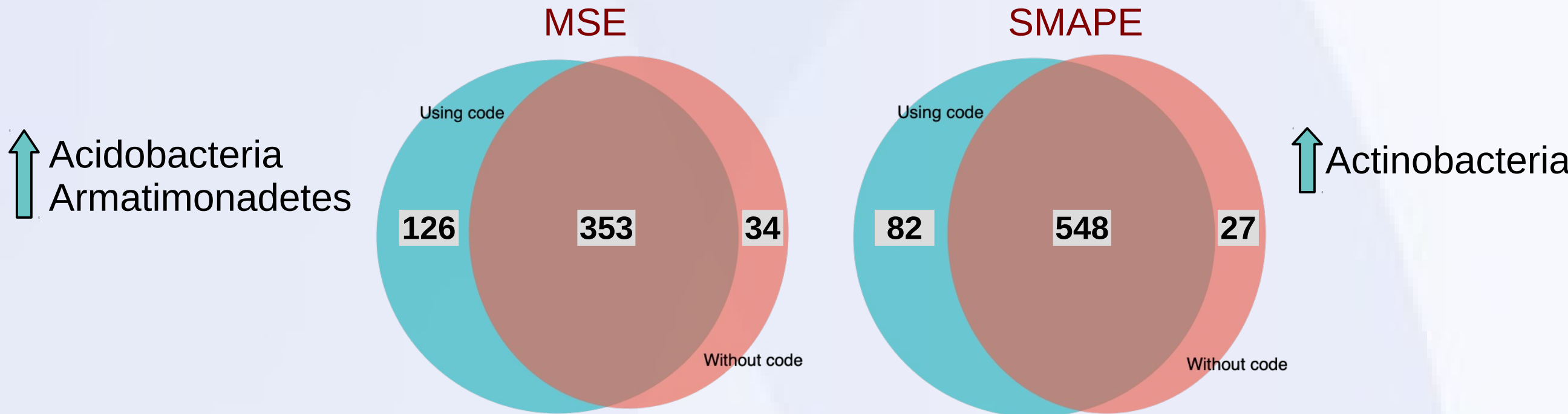
Application

Predicting taxa composition from environmental data

Challenge: Obtaining 717 abundance values from only a few environmental variables

- **Performance**

Predictor	MSE	SMAPE
Default	0.0027	71.71%
Without code	0.0022	71.68%
Latent space (code)	0.0018	68.69%

- **OTUs predicted better than default**

MSE — Using code | Without code: 126 | 353 | 34

↑ Acidobacteria Armatimonadetes

SMAPE — Using code | Without code: 82 | 548 | 27

↑ Actinobacteria

- **Relevant features** (agree with [Walters et al., 2018]):
  - plant age, precipitation

Note: MSE: Mean Square Error.   SMAPE: Symmetric Mean Absolute Percentage Error

Contributions:
- Novel dimensionality reduction approach to define latent space in 16S microbiome. It could be applied to any environment (gut, ocean, urban soil, etc.) where enough samples available
- Ability to undertake challenging tasks in microbiome data analysis, such as to predict the microbial composition of hundreds of taxa based on a small number of features, rather than the common prediction of a phenotypic feature
- The knowledge encoded within our microbiome autoencoder model can be reused, via Transfer Learning, into novel but related studies with fewer samples

beatriz.garcia@upm.es
http://wilkinsonlab.info/microbiomes

CBGP — CENTRO DE BIOTECNOLOGÍA Y GENÓMICA DE PLANTAS — UPM-INIA
EXCELENCIA SEVERO OCHOA
POLITÉCNICA — "Ingeniamos el futuro"
CAMPUS DE EXCELENCIA INTERNACIONAL
INIA — Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria