I Meeting of Systematics, Biogeography and Evolution (SBE)

@ManoloFPerez

sites.google.com/site/manolofperez

Using deep learning for unifying genomic data and traits in species delimitation

MF Perez; I Sanmartín; BC Faircloth; LAC Bertollo; MB Cioffi

REAL JARDÍN BOTÁNICO

LSU Louisiana State University

ufscar

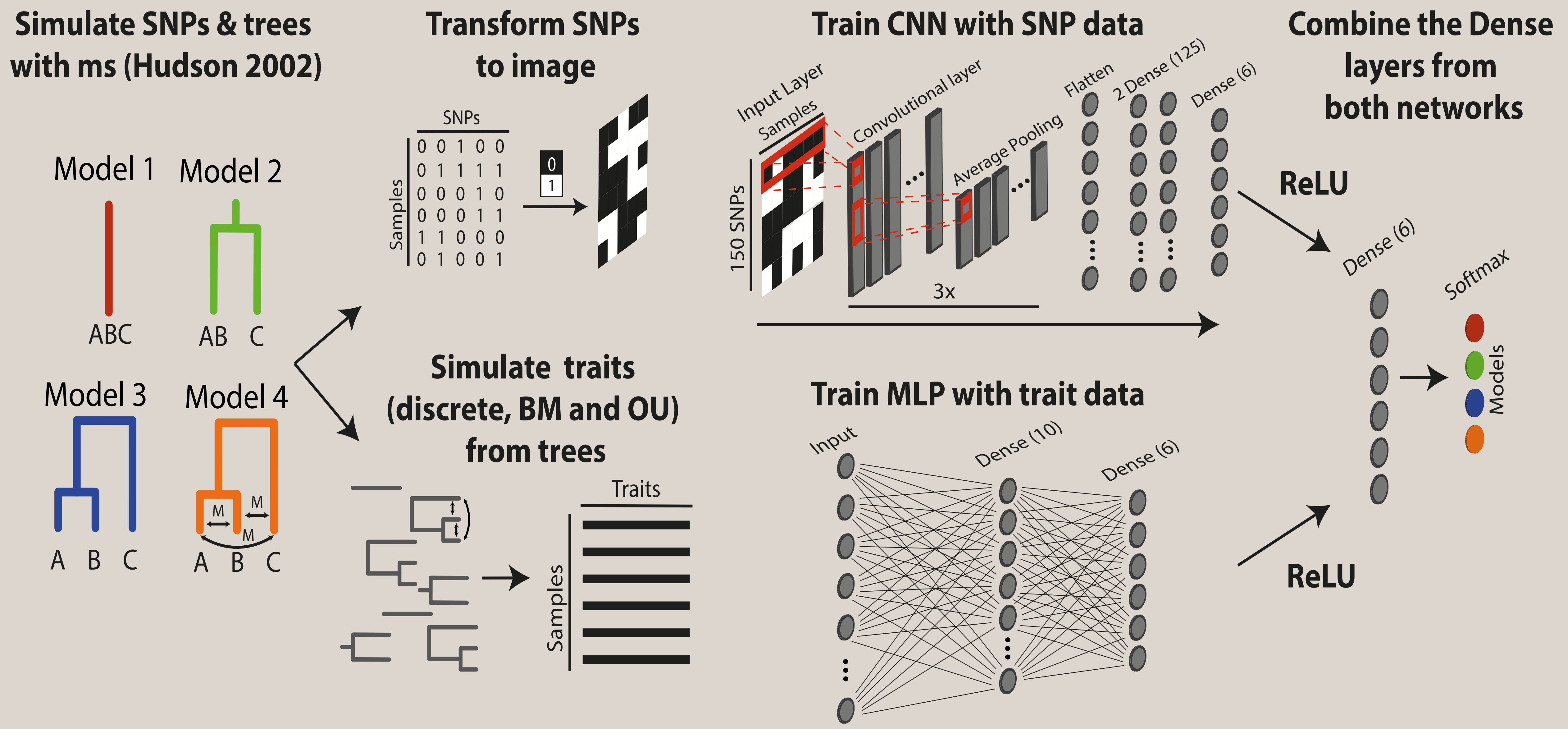
Introduction

Different **species concepts - distinct** strategies to **identify species boundaries** (de Queiroz 2007). It is important to adopt a **multidisciplinary approach**, by assessing **different sources of evidence** (Carstens et al. 2013).

Most approaches consist in analyzing **genomic and phenotypical/geographical** information **separately,** followed by **visual/qualitative comparison**. Methods that actually **integrate** different data are **limited** to up to a **few hundreds of loci** and **simple models** of evolution (Solís-Lemus et al. 2015).

We present a method based on **simulated data and deep learning**, that **combines** both **genomic and trait** information in a unified framework.

Methods



Simulate SNPs & trees with ms (Hudson 2002)

Transform SNPs to image

SNPs

Samples

0	0	1	0	0
0	1	1	1	1
0	0	0	1	0
0	0	0	1	1
1	1	0	0	0
0	1	0	0	1

Simulate traits (discrete, BM and OU) from trees

Train CNN with SNP data

Train MLP with trait data

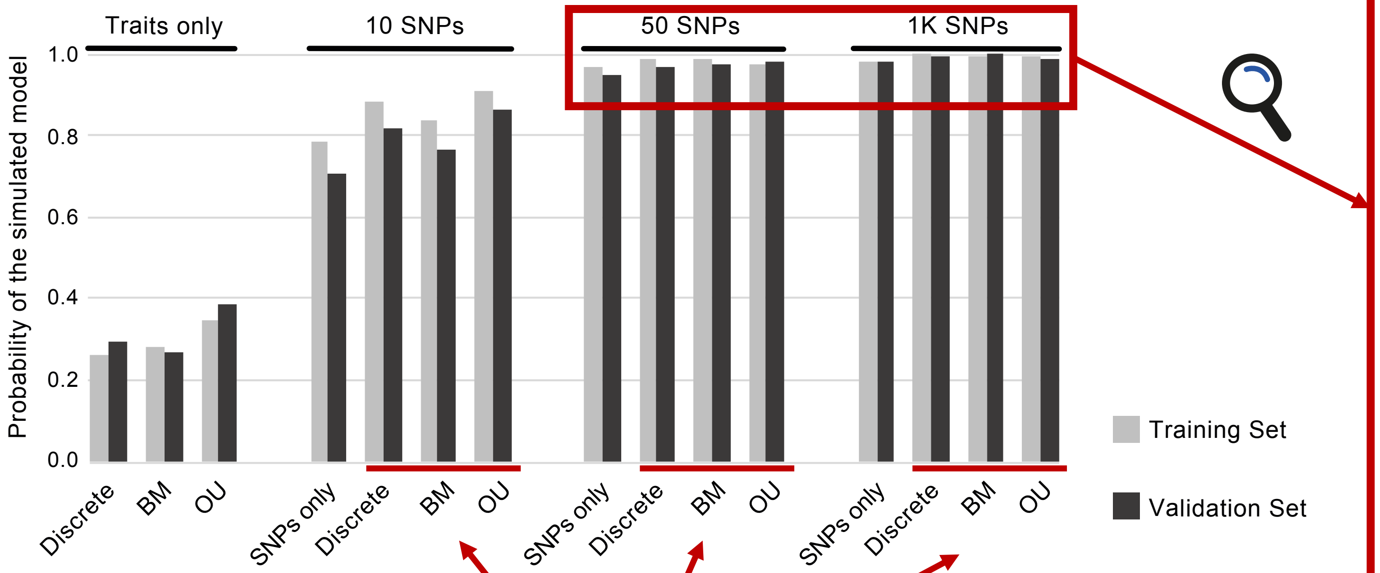
Combine the Dense layers from both networks

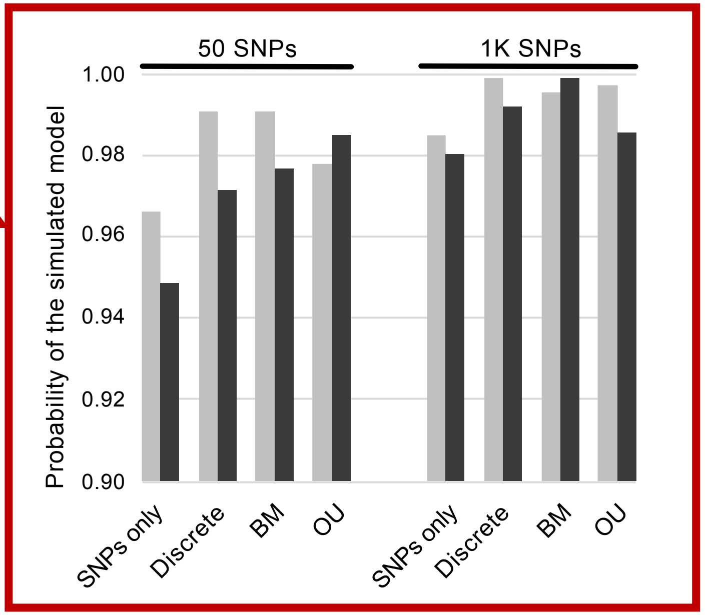
Results

Probabilities are low when using only traits.

Increasing SNPs also raised the probability of recovering the right model.

- little improvement with > 50 SNPs.





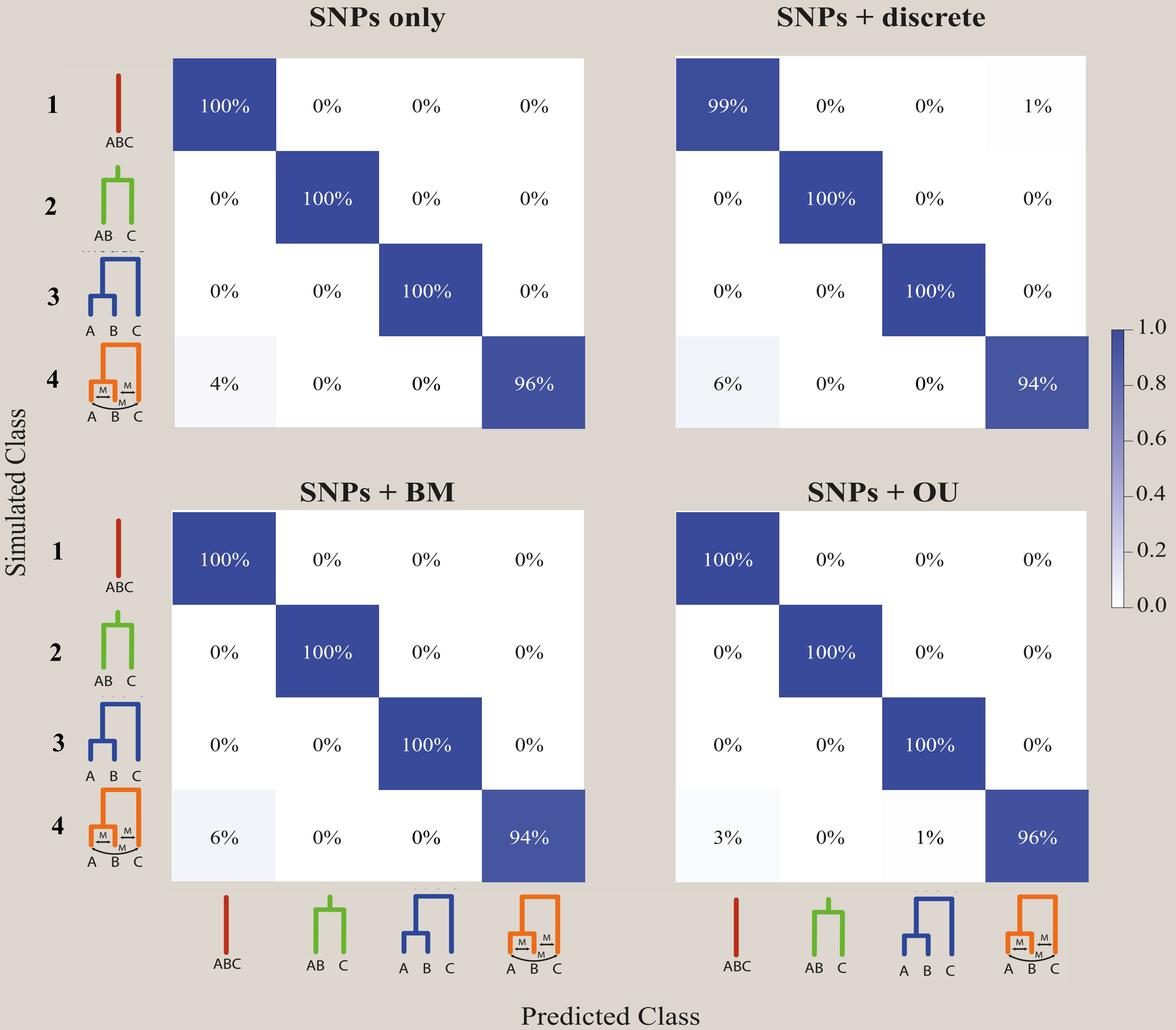
Using both genomic and trait data recovered slightly better results than using only SNPs.

Conclusions

The **accuracy** of our approach was **very high** (confusion matrix with the **test set**). **Confusion** of model 4 (**migration**) with model 1 (**one species**).

Incorporating traits resulted in **similar accuracy** to using only SNPs.

Traits incorporate information **complimentary** to genomic data that might be **useful for species delimitation.**



SNPs only

Simulated Class	ABC	AB C	A B C	A B C (M)
1 ABC	100%	0%	0%	0%
2 AB C	0%	100%	0%	0%
3 A B C	0%	0%	100%	0%
4 A B C (M)	4%	0%	0%	96%

SNPs + discrete

Simulated Class	ABC	AB C	A B C	A B C (M)
1 ABC	99%	0%	0%	1%
2 AB C	0%	100%	0%	0%
3 A B C	0%	0%	100%	0%
4 A B C (M)	6%	0%	0%	94%

SNPs + BM

Simulated Class	ABC	AB C	A B C	A B C (M)
1 ABC	100%	0%	0%	0%
2 AB C	0%	100%	0%	0%
3 A B C	0%	0%	100%	0%
4 A B C (M)	6%	0%	0%	94%

SNPs + OU

Simulated Class	ABC	AB C	A B C	A B C (M)
1 ABC	100%	0%	0%	0%
2 AB C	0%	100%	0%	0%
3 A B C	0%	0%	100%	0%
4 A B C (M)	3%	0%	1%	96%