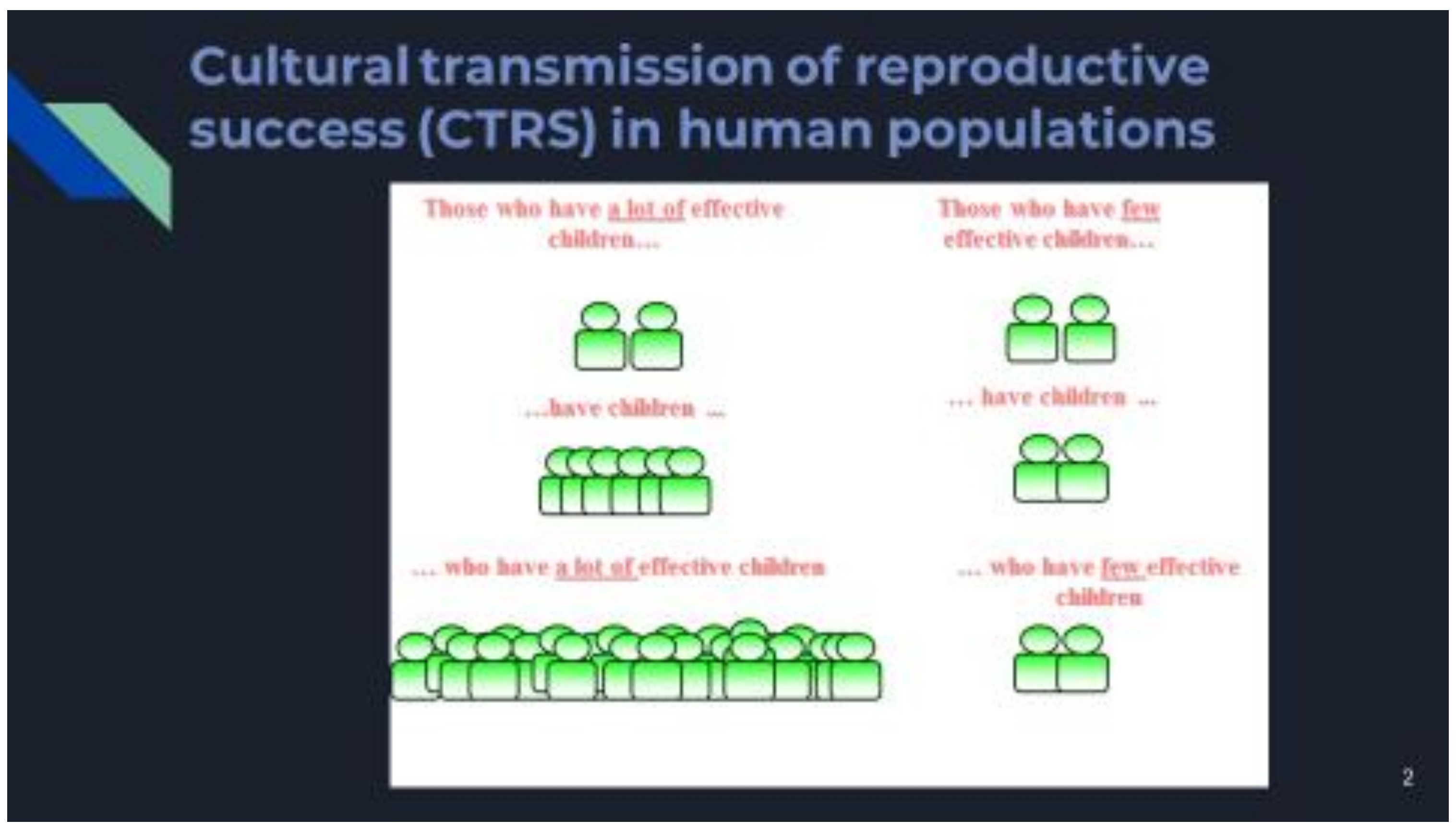
Inferring cultural transmission of reproductive success through machine learning methods

Master2 internship of **Ferdinand Petit**

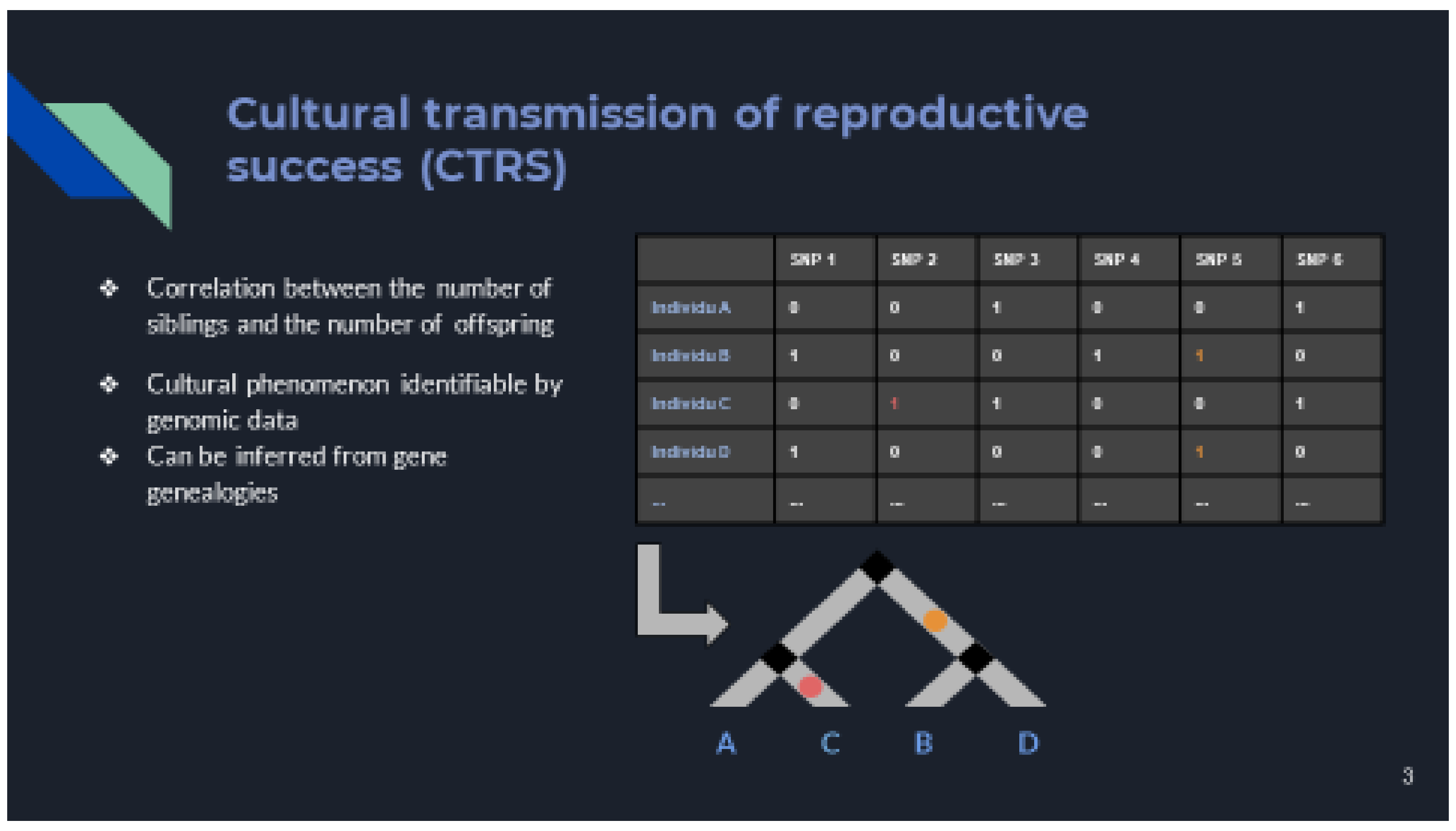
Supervisor: **AUSTERLITZ Frédéric**, Directeur de Recherche CNRS, Laboratoire Eco-anthropologie, UMR 7206 CNRS/MNHN/Univ de Paris. Co-supervisors: **GUEZ Jérémy**. PhD student UMR 7206 Eco-anthropologie, **Jay Flora**, Chargée de Recherche CNRS, Laboratoire Interdisciplinaire des Sciences du Numérique, Université Paris-Saclay.

data intelligence institute of Paris

Work supported by diiP, IdEx Université de Paris, ANR-18-IDEX-0001



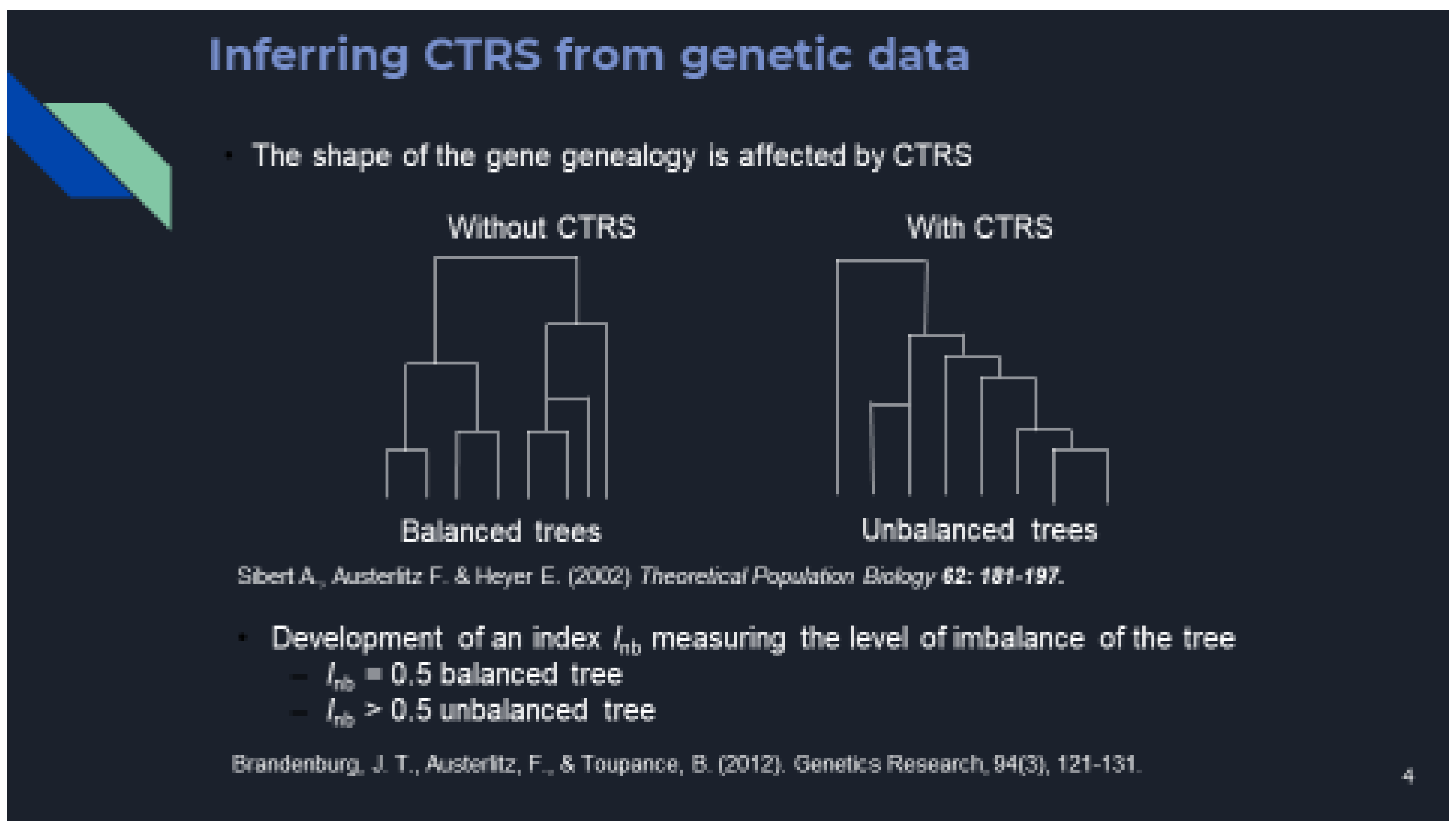
Cultural transmission of reproductive success (CTRS) in human populations



Cultural transmission of reproductive success (CTRS)

- ❖ Correlation between the number of siblings and the number of offspring
- ❖ Cultural phenomenon identifiable by genomic data
- ❖ Can be inferred from gene genealogies

	SNP 1	SNP 2	SNP 3	SNP 4	SNP 5	SNP 6
Individu A	0	0	1	0	0	1
Individu B	1	0	0	1	1	0
Individu C	0	1	1	0	0	1
Individu D	1	0	0	0	1	0
--	--	--	--	--	--	--



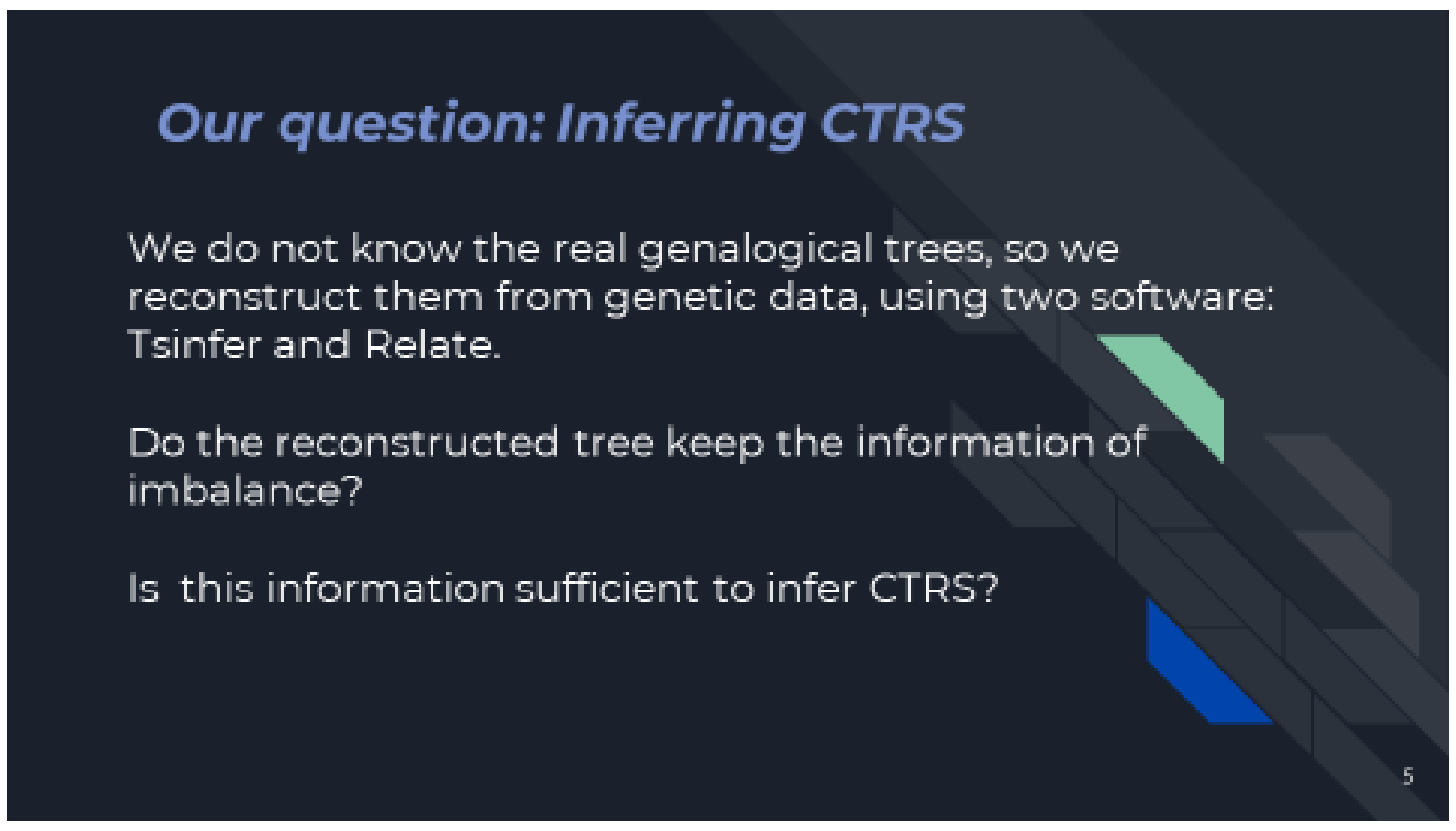
Inferring CTRS from genetic data

- The shape of the gene genealogy is affected by CTRS

Sibert A., Austerlitz F. & Heyer E. (2002) *Theoretical Population Biology* **62: 181-197.**

- Development of an index I_{nb} measuring the level of imbalance of the tree
  - $I_{nb} = 0.5$ balanced tree
  - $I_{nb} > 0.5$ unbalanced tree

Brandenburg, J. T., Austerlitz, F., & Toupance, B. (2012). Genetics Research, 94(3), 121-131.

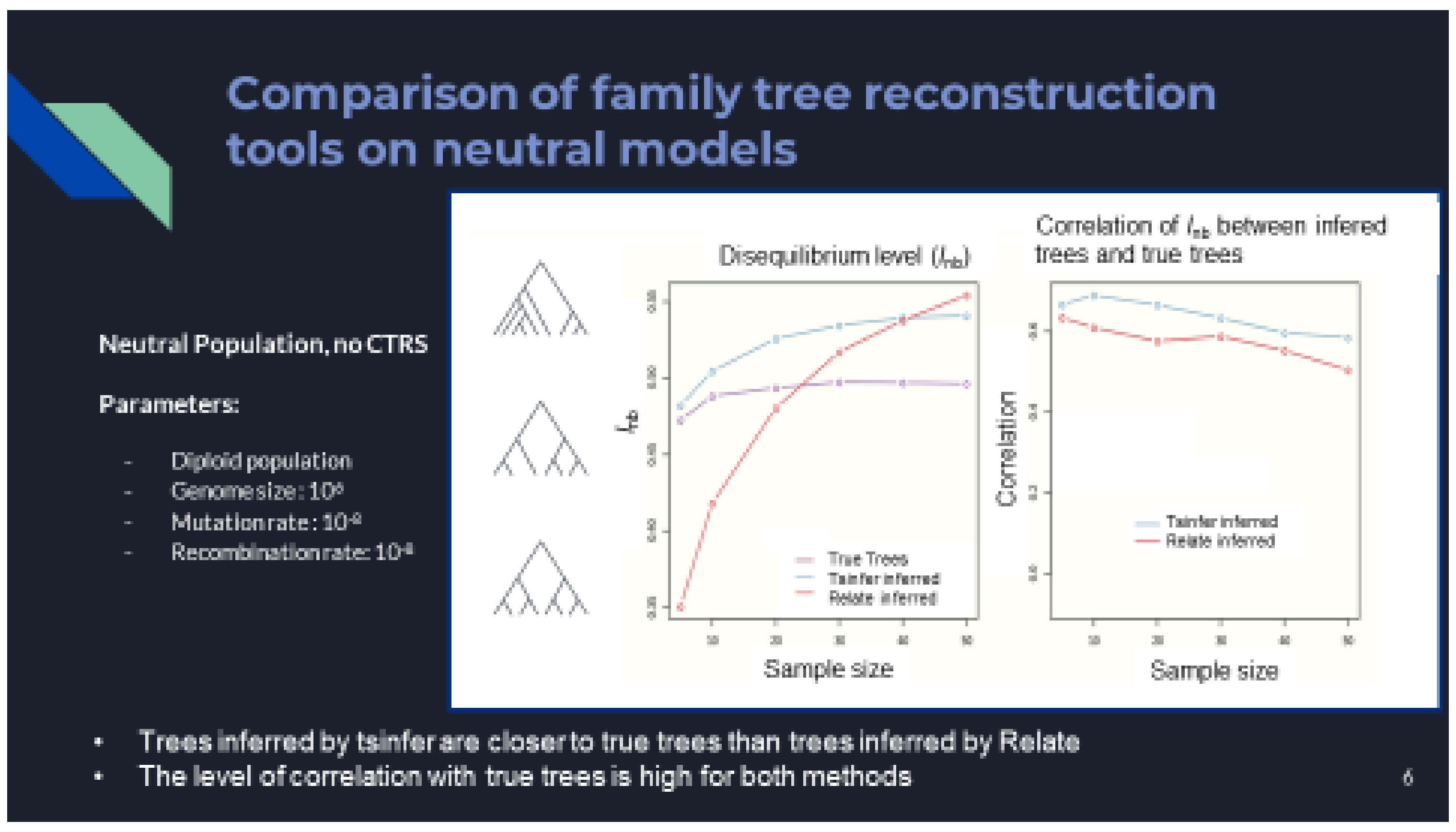


Our question: Inferring CTRS

We do not know the real genalogical trees, so we reconstruct them from genetic data, using two software: Tsinfer and Relate.

Do the reconstructed tree keep the information of imbalance?

Is this information sufficient to infer CTRS?



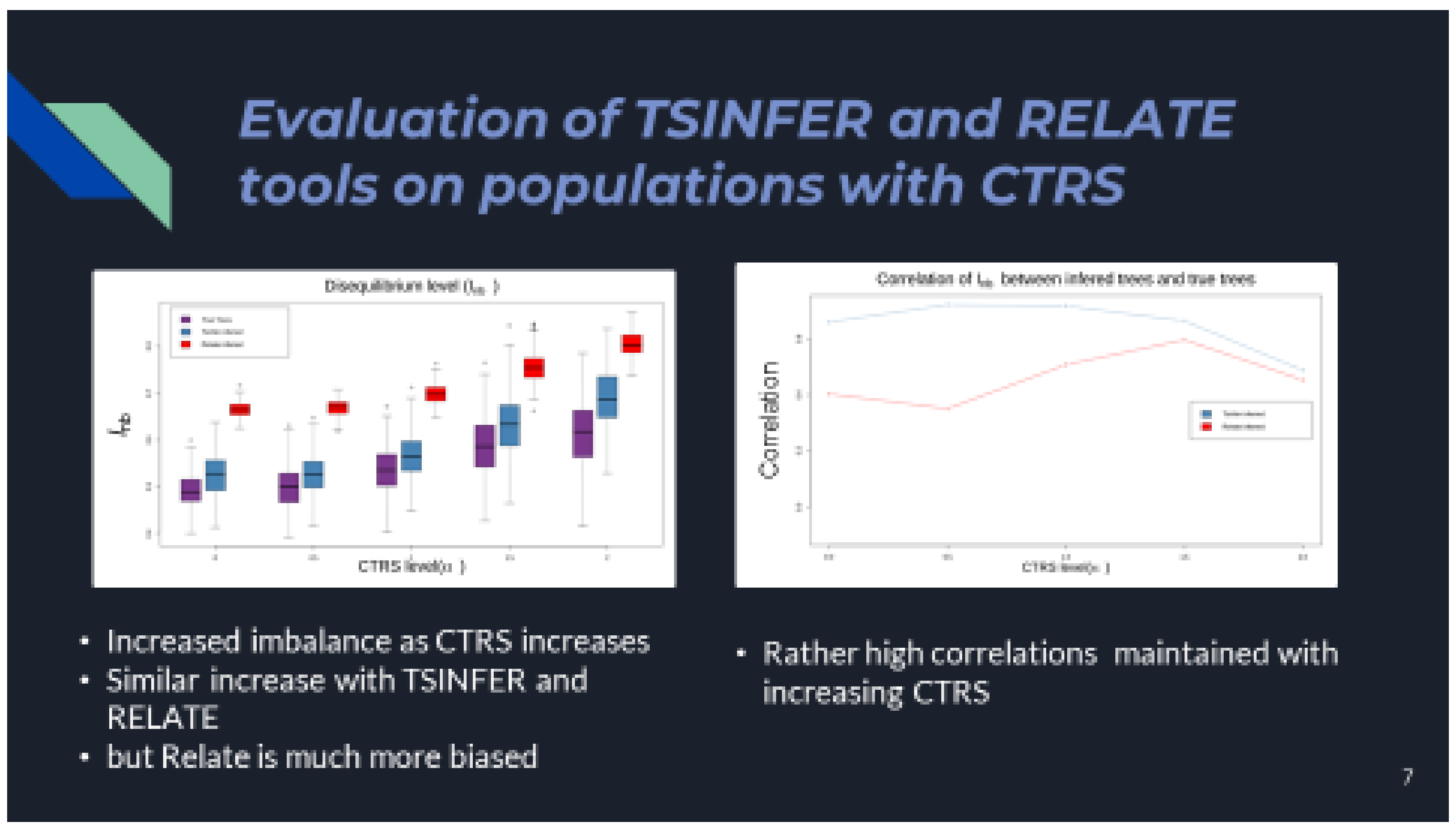
Comparison of family tree reconstruction tools on neutral models

Neutral Population, no CTRS

Parameters:

- Diploid population
- Genome size: 10^{8}
- Mutation rate: 10^{-8}
- Recombination rate: 10^{-8}

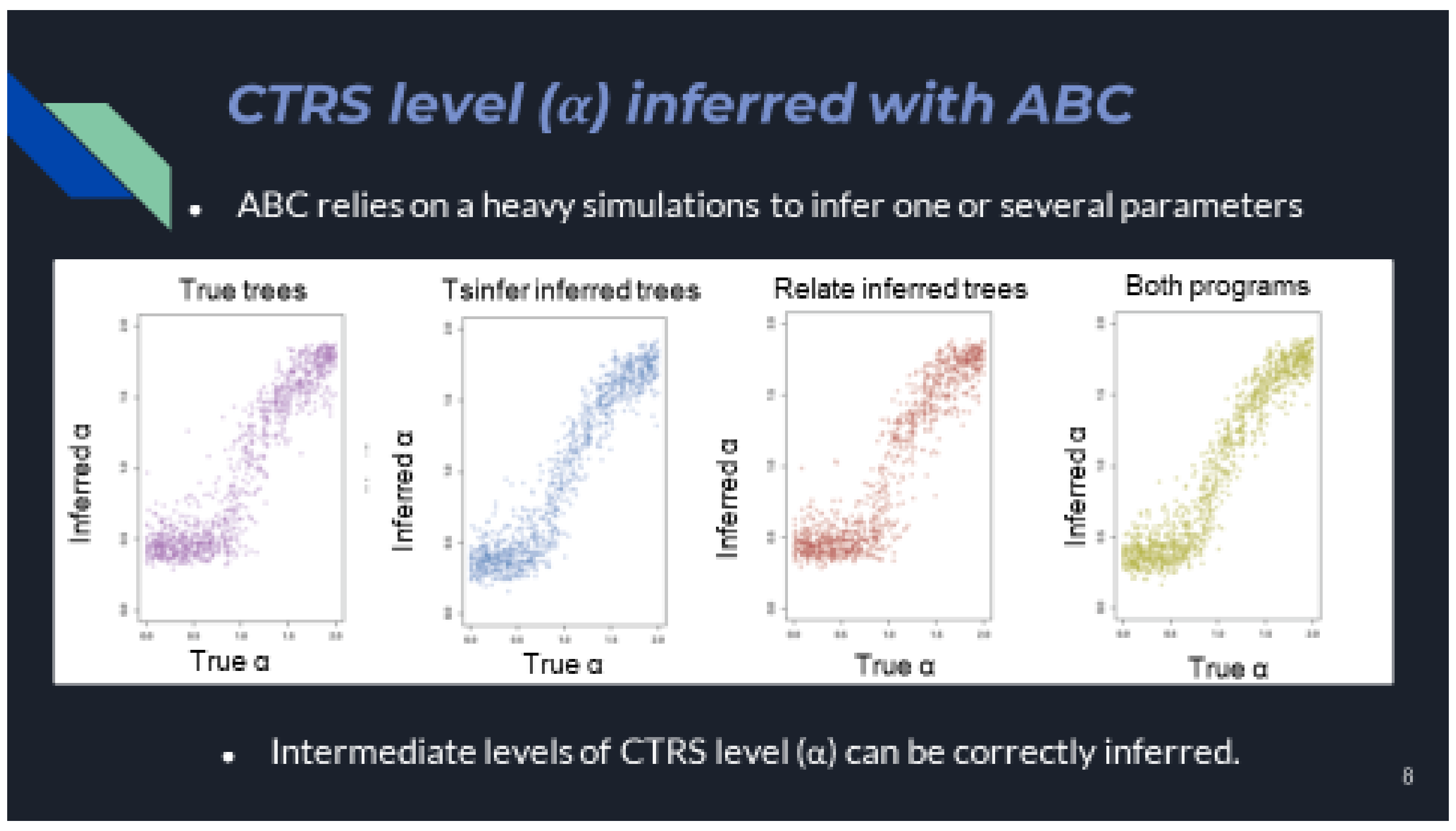
- Trees inferred by tsinfer are closer to true trees than trees inferred by Relate
- The level of correlation with true trees is high for both methods



Evaluation of TSINFER and RELATE tools on populations with CTRS

- Increased imbalance as CTRS increases
- Similar increase with TSINFER and RELATE
- but Relate is much more biased

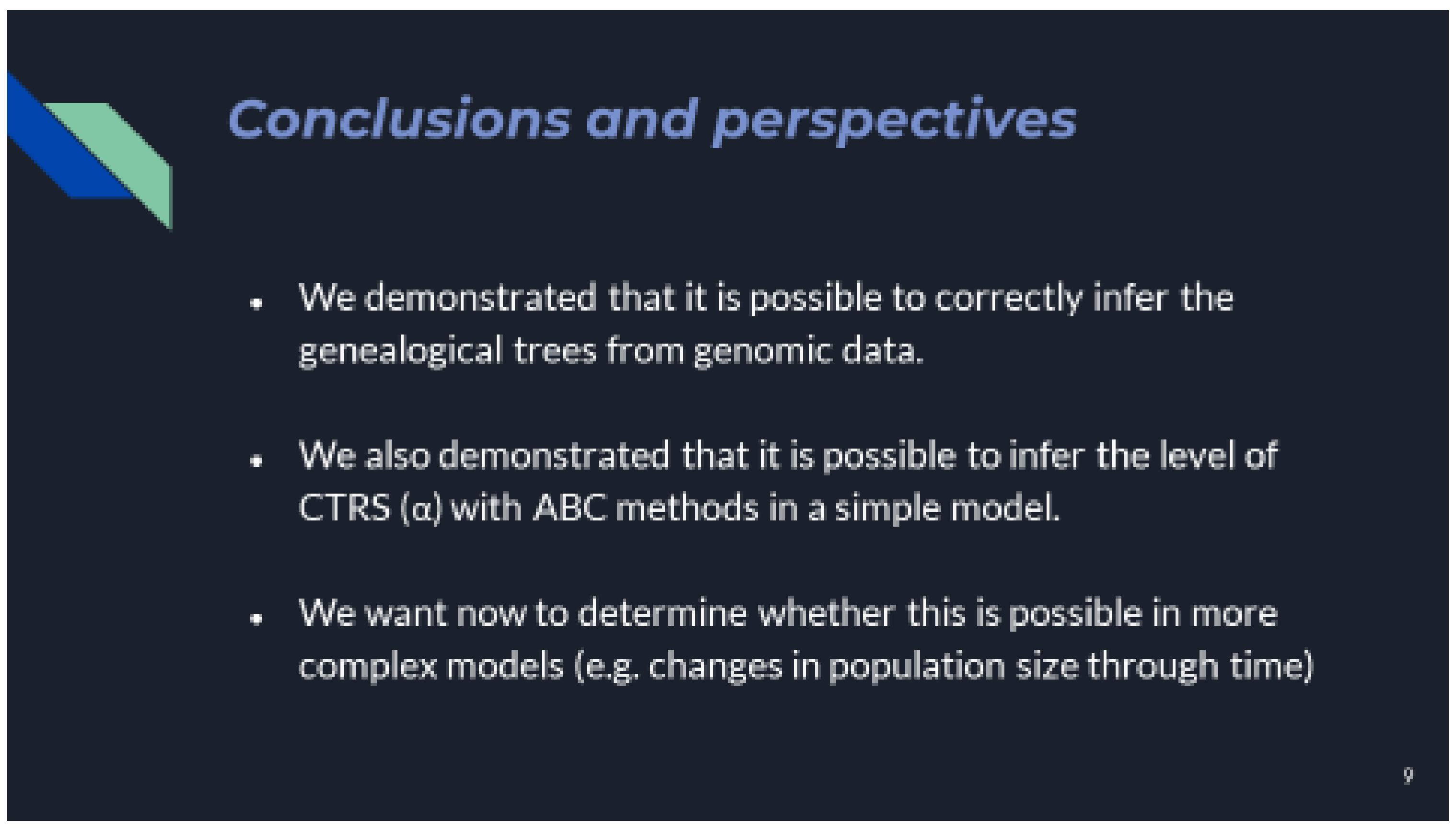
- Rather high correlations maintained with increasing CTRS



CTRS level (α) inferred with ABC

- ABC relies on a heavy simulations to infer one or several parameters

- Intermediate levels of CTRS level (α) can be correctly inferred.



Conclusions and perspectives

- We demonstrated that it is possible to correctly infer the genealogical trees from genomic data.
- We also demonstrated that it is possible to infer the level of CTRS (α) with ABC methods in a simple model.
- We want now to determine whether this is possible in more complex models (e.g. changes in population size through time)