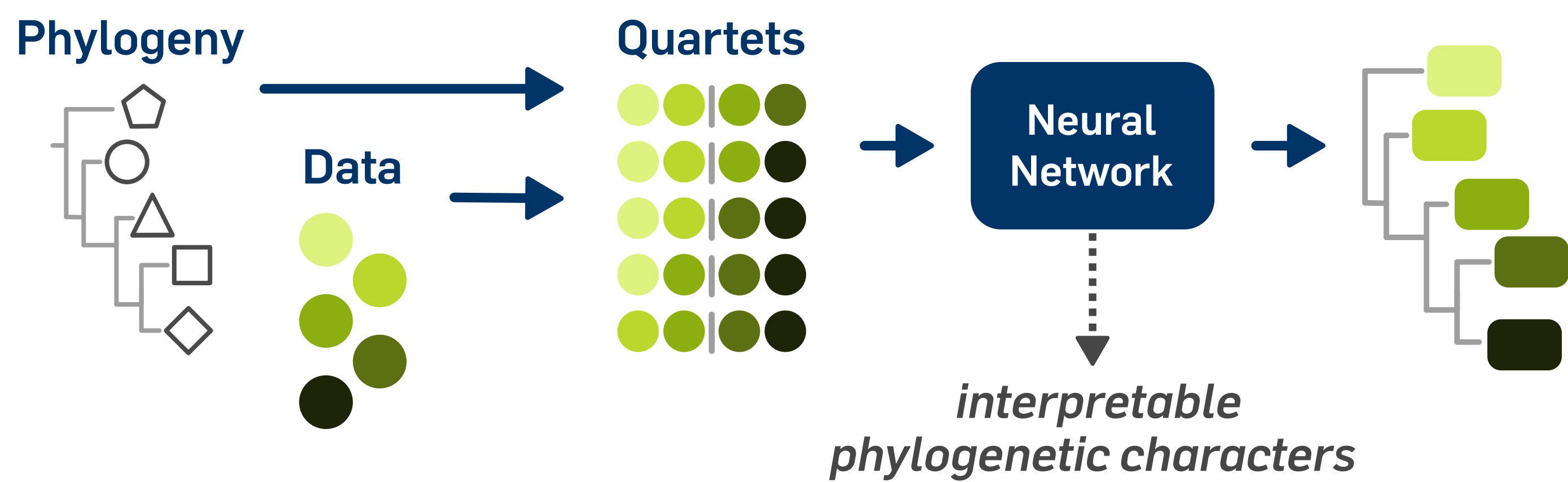


# A QUARTET-BASED APPROACH FOR INFERRING PHYLOGENETICALLY INFORMATIVE FEATURES FROM GENOMIC AND PHENOMIC DATA

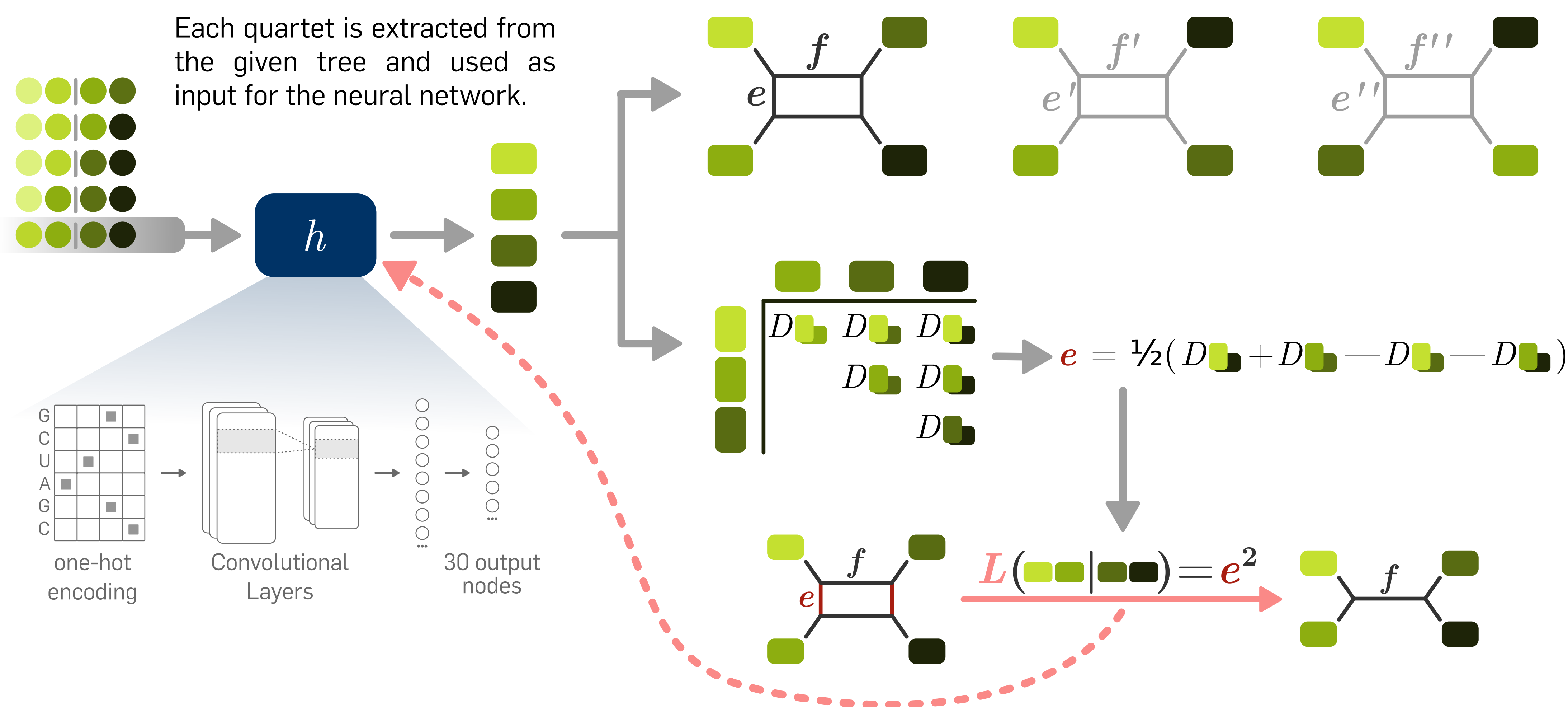
Vivian B. Brandenburg, Ben Luis Hack, Axel Mosig. Ruhr University Bochum

## IDEA



- Form quartets of input data from a given phylogeny
- Train Neural Network with quartets
- Network will focus on phylogenetic informative fraction of input
- Assign inferred features to interpretable phylogenetic characters

## METHODS



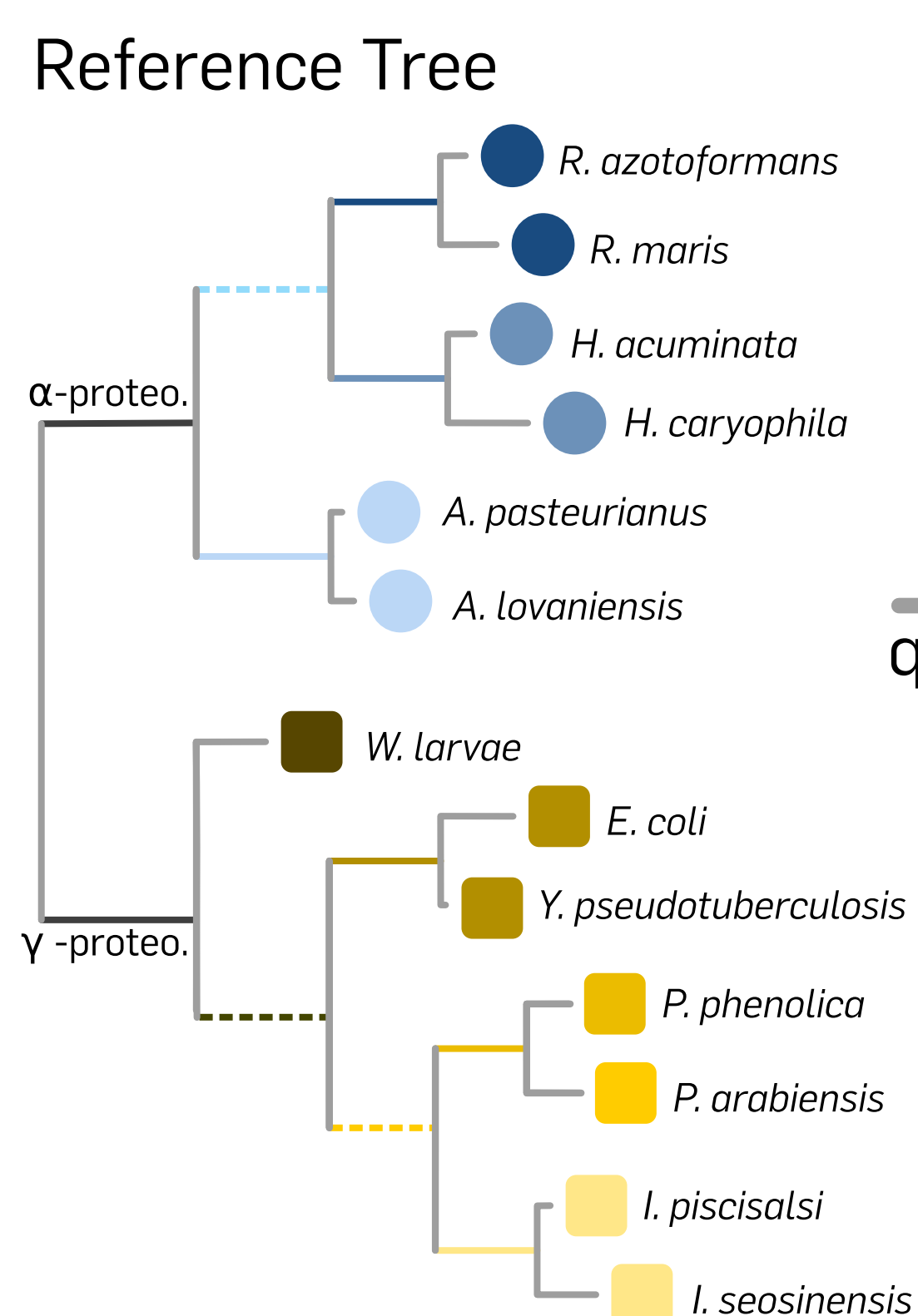
Each output quartet forms three splits, of which only one has no negative edges. The **Quartet Score** shows the fraction of correct output splits in all training data.

The euclidean distance between all output features is used to calculate **edge e**.

**Quartet Loss L** minimizes **edge e** to optimize towards the correct split.

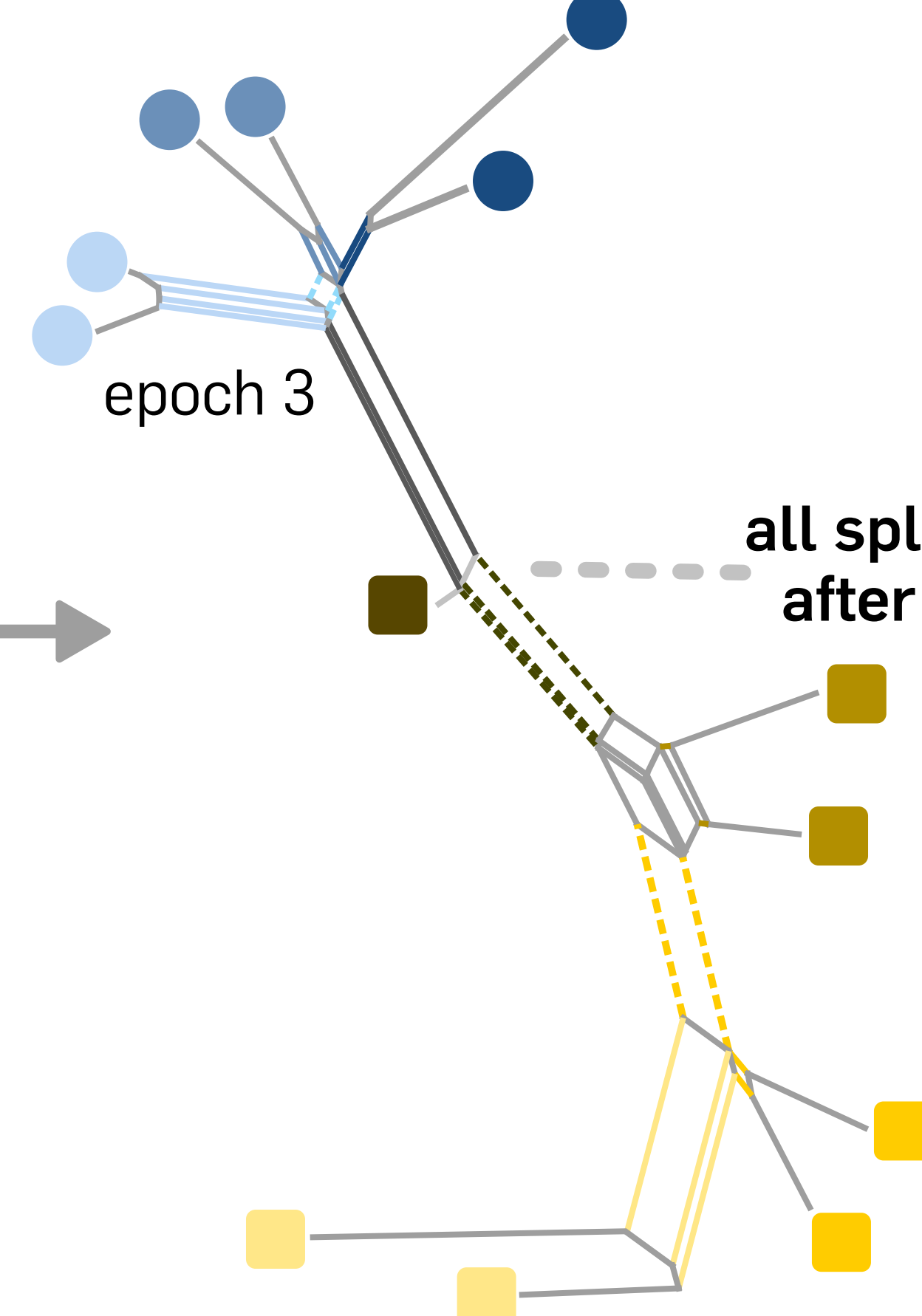
## PROOF OF CONCEPT

**Data:** nts 0-350 of 16S rRNA sequences

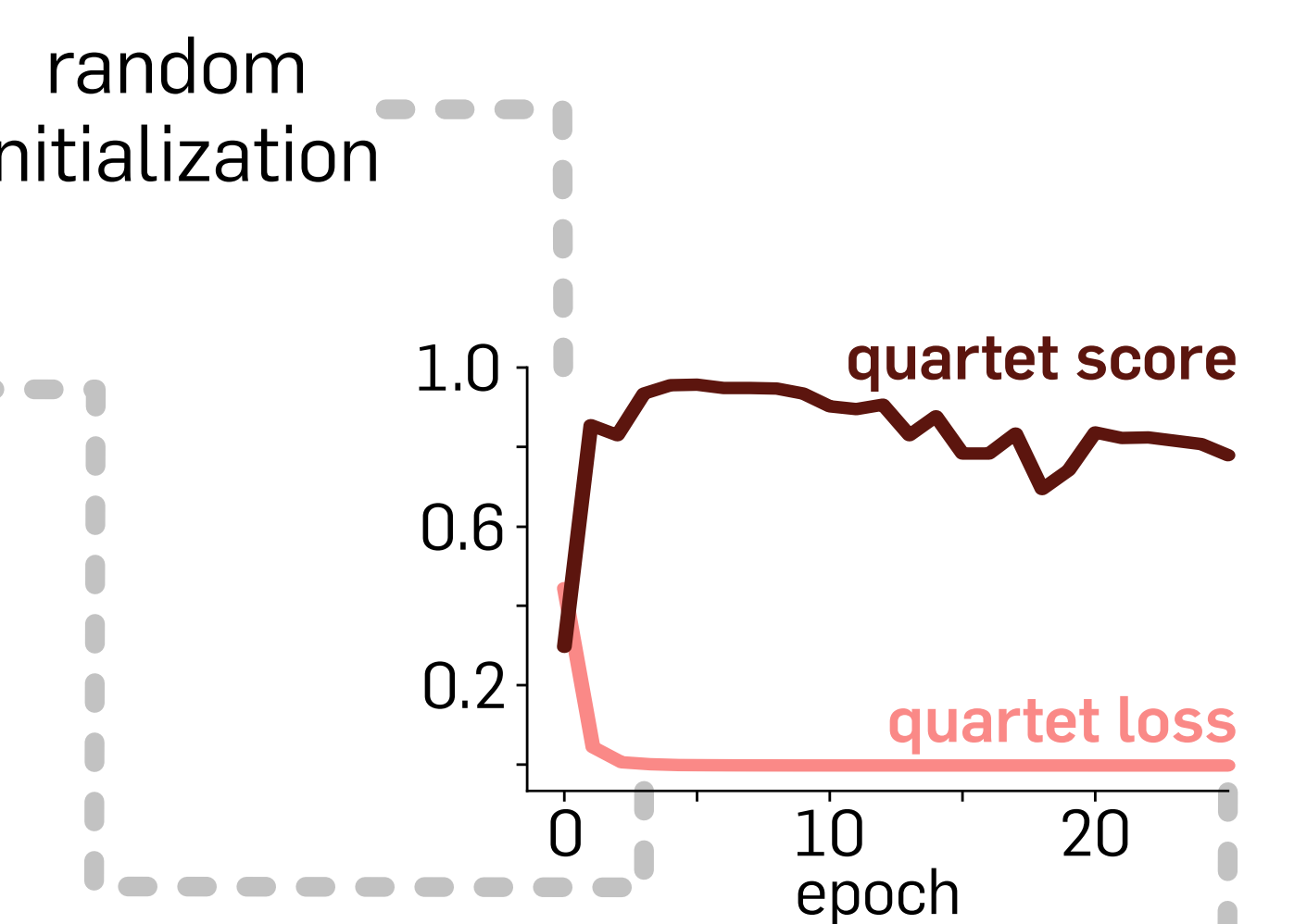


658 quartets

**Result**



The euclidean distance between all output feature vectors is calculated and visualized with *SplitsTree4*



## OUTLOOK

- Explore quartet loss function and network topology variations
- Assign phylogen. characters from inferred features, e.g by Class Activation Map
- Flexibility of input data type allows various extensions, eg:
  - ▶ Co-evolutionary patterns between RNAs and RNA-binding proteins
  - ▶ Image data for phenome evolution

## References

- Bandelt and Dress. Split decomposition. *Mol. Phylogenet. Evol.* 1992
- Bryant and Moulton. Neighbor-net. *Mol. Biol. Evol.* 2004
- Huson and Bryant. Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* 2006
- Zhou et al. Learning Deep Features for Discriminative Localization. *CVPR Proceedings* 2016

## Contact

Vivian.Brandenburg@rub.de

