

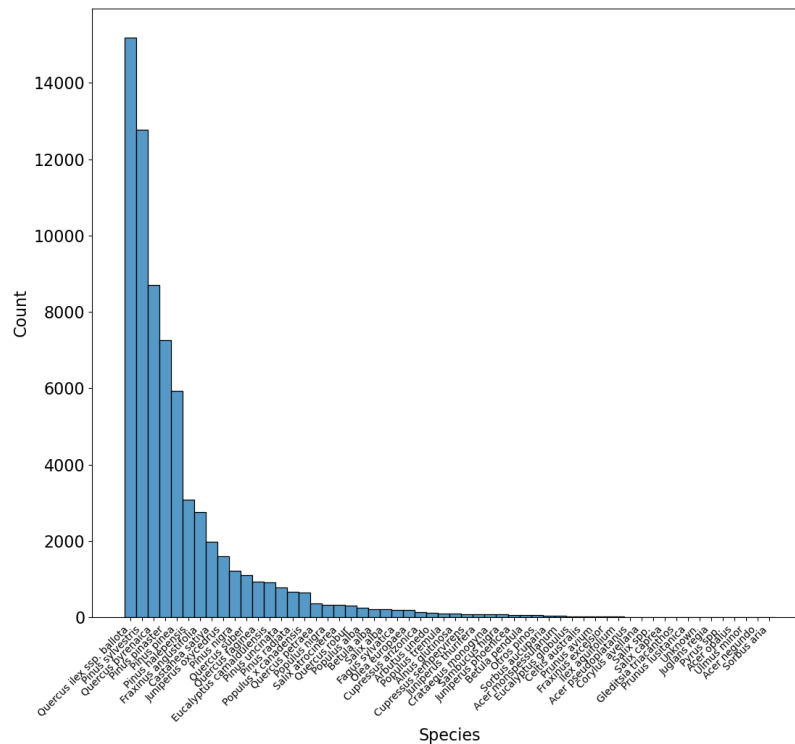
# Mining Field Data for Tree Species Recognition at Scale Supplementary Material

Dimitri Gominski<sup>1</sup>, Daniel Ortiz-Gonzalo<sup>1</sup>, Martin Brandt<sup>1</sup>,  
Maurice Mugabowindekwe<sup>1</sup>, and Rasmus Fensholt<sup>1</sup>

University of Copenhagen, Denmark  
{dg,gonzalo}@ign.ku.dk

## 1 Species Distribution

We indicate in Supplementary Figure 1 the complete distribution of tree species in the dataset (verified + unverified individuals). The distribution is typically long-tailed, with a few dominant species and numerous rare species.



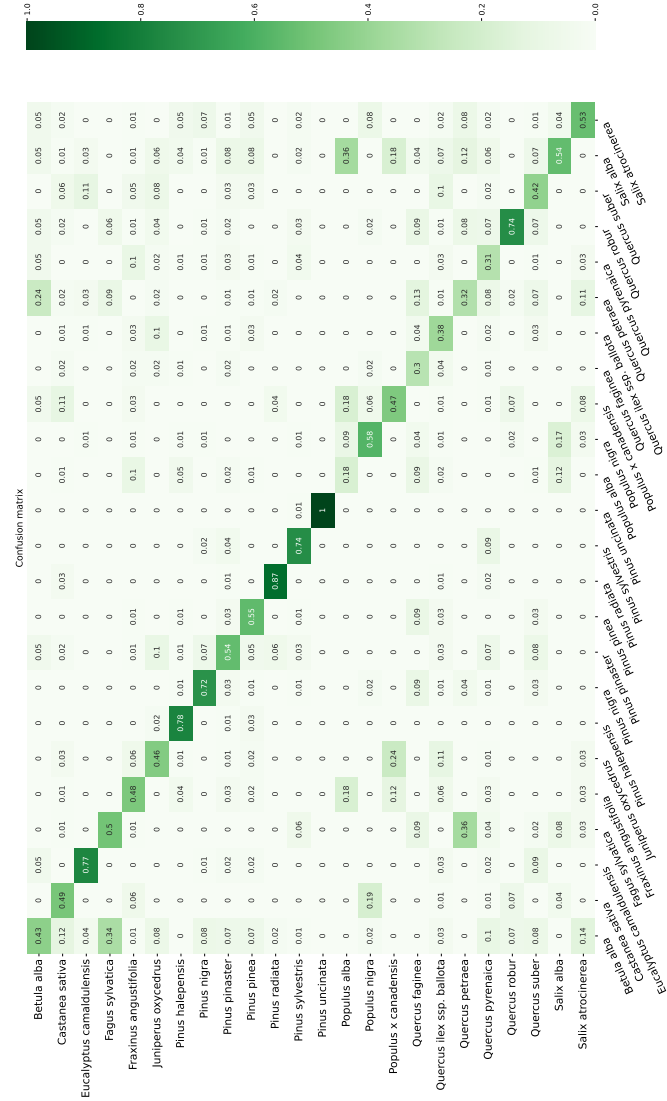
Supplementary Fig. 1: Full distribution of tree species.

## 2 Confusion Matrix

The confusion matrix of our best model (Supplementary Figure 2) trained on all labels + unlabeled examples with semi-supervised learning shows a relatively balanced performance accross species, with some rare species such as *Salix alba* having an accuracy  $> 50\%$ .

## 3 Acknowledgments

D.G. acknowledges support from the Novo Nordisk Foundation, grant number NNF21OC0069116. D.O-G. acknowledges support from the European Union's Horizon 2020 research and innovation programme under the MSCA grant agreement No 896512, project TREES4CLIMA.



Supplementary Fig. 2: Confusion matrix for our best model (all labels + 500k unlabeled images).