Dear Dr. Xiang,

I hope this email finds you well.

The reason for emailing is to seek your consent in me using your published data on -

Z. Y. Du, A. J. Harris, Q. J. Xiang, Phylogenomics, co-evolution of ecological niche and morphology, and historical biogeography of buckeyes, horsechestnuts, and their relatives (Hippocastaneae, Sapindaceae) and the value of RAD-Seq for deep evolutionary inferences back to the Late Cretaceous. Mol Phylogenet Evol 145, 106726 (2020).

I am a PhD student from the Royal Botanic Garden Edinburgh and am working on a project on using nuclear data to tell plant species apart. Specifically, I am meta-analysing different nuclear DNA data sets (RAD, target-capture, whole-genome sequencing etc), where the sequence data contain multiple individuals from multiple congeneric species.

With the available data, I am assessing

• the proportion of plant species in a genus that resolves as monophyletic with nuclear data

• the proportion of named species that have diagnostic DNA sequence variants (fixed species-specific substitutions), and how frequent these are

• the amount of sequence data from the nuclear genome that is optimal for telling species apart (e.g. on average, how many loci are needed to give maximal species resolution in each data set).

I think there are some basic questions to be answered here that will be very useful for the design of future DNA based plant identification approaches, as well as providing insights into the genomic nature of differences between plant species.

I think there is the potential for the resulting paper to be interesting if we get it to fruition.

I would appreciate much hearing back… even better to have any ideas that will strengthen this study.

I hope that all makes sense.

Best wishes,

Wu