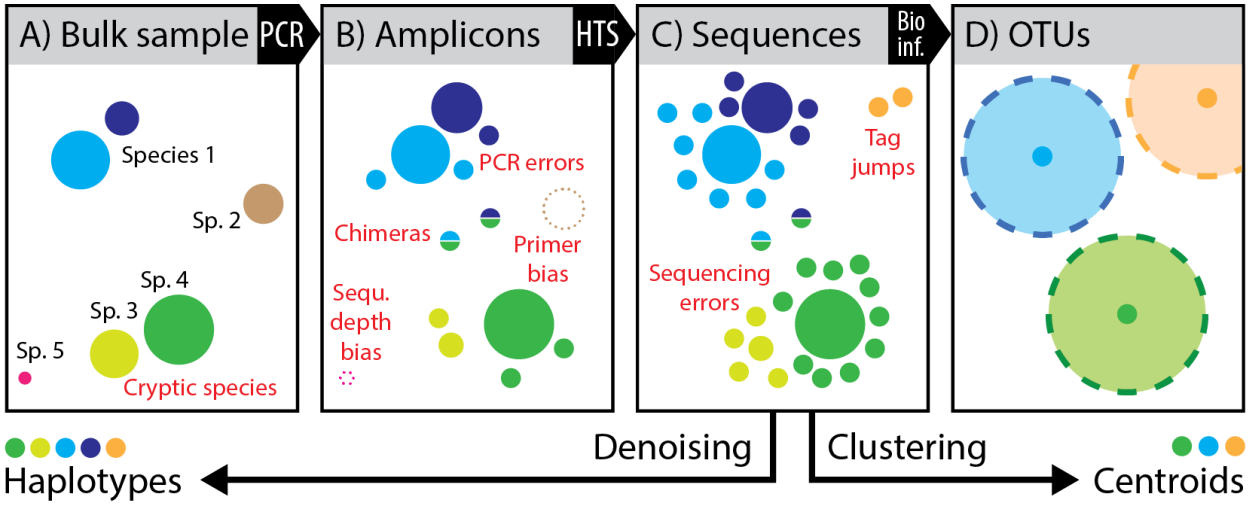


**Figure S1**: Overview of the metabarcoding process, with key biases potentially affecting sequence accuracy (shown in red). In the bulk sample (**A**) several species with different biomass (indicated by circle size) and distinct haplotypes (indicated by colour) are present. After tissue homogenisation and DNA extraction the COI marker is amplified using PCR (**B**), which can not only skew sequence abundance but also fail to amplify taxa due to primer bias (Elbrecht & Leese 2015) or insufficient sequencing depth in the case of underrepresented / rare taxa (Elbrecht et al. 2017). In the process of HTS (**C**) many new false sequence variants are generated due to sequencing errors (Schirmer et al. 2015), chimera formation (Edgar et al. 2011) and mixing of multiplexed samples (Schnell et al. 2015; Esling et al. 2015). The impact of these errors is usually reduced by strict quality filtering and clustering of similar sequences into operational taxonomic units (OTUs). Usually only the most abundant sequence in an OTU is considered and used to identify the respective species, which however means that information on genetic diversity is lost (Callahan et al. 2017, **D**). Thus, recently alternative denoising strategies have been developed to remove sequences affected by errors from the dataset and retain the actual haplotype sequences present in a sample (Eren et al. 2014; Amir et al. 2017; Callahan et al. 2016; Edgar 2016). Figure based on Figure S1 in Callahan *et al*. 2016.

References:

1 --- will add later
2
3
4
5
6
7
8