# Abstract

An efficient determination of the genotype is important to many studies. Rather than sequencing the whole genome of an individual, genetic markers such as SNPs (single nucleotide polymorphism) are used for analysis. RAD (restriction site associated DNA) tags are DNA sequences, which are next to a restriction endonuclease recognition site. Restriction site associated DNA sequencing (RAD-seq) in combination with RAD analysis pipelines offer an efficient and robust way for the determination of genetic markers. In this thesis, an overview of existing RAD analysis pipelines is given and a comparison between the pipelines Stacks, PyRAD, ipyrad and dDocent is performed. The pipelines were compared by taking a look at the loci the pipelines recovered, phylogenetic and population structure results. Runtimes and memory footprint of the pipelines are also in the focus of this comparison.

Analyses for simulated and empirical data were performed. dDocent recovered more usable loci than the other pipelines and is also overall the fastest pipeline. But dDocent does not come with build-in functions for generating phylogenetic or population structure files. Stacks was also very fast except for one analysis. It also recovered the highest number of overall loci. PyRAD was the slowest pipeline for most of the data sets. But it recovered a relatively high number of loci. ipyrad performed much faster than PyRAD but it did not recover as many loci as PyRAD. The results of the pipelines overlapped strongly for the simulated data sets. But only the results of one of the empirical data sets did show a relatively high overlap. In general, Stacks had the largest memory footprint and was the only analysis pipeline that required more than the available 8 GB of memory on the test system.

**Key words**

RAD-seq; ddRAD-seq; Stacks; PyRAD; ipyrad; dDocent; rtd; Rainbow; species tree; population structure