

Project 3: Phylogenetic relationships

Since taxonomic identification is often difficult, especially in plants that are prone to hybridisation or propagating predominantly by vegetative means, such that flowers are not available, genetic barcoding is often a feasible alternative. The principle is to amplify marker genes, sequence them, and to construct a phylogenetic tree to assess the relationships. If the test sample clusters together with a known sequence from an identified species of a database, this can be used to authenticate otherwise recalcitrant accessions. Moreover, the genetic barcode works with any material that is available from the plant of interest, even for a seed (no matter, whether it is seed, leaf, or flower, the result from genetic barcoding is the same - it should be clear to you, why this is the case).

Barcoding genes should meet three criteria (CBOL, 2009): they should be universal, they should allow for good sequencing quality, and they should be endowed with discriminative power. In most cases, for plants, plastidic genes are used, because they can be amplified more easily as compared to nuclear genes (do you have an idea, why?), moreover, in contrast to nuclear markers, there is no heterozygosity, such that it is easy to determine the haplotype (do you have an idea, why there is no heterozygosity in plastid markers?). Depending on the scope of the barcoding study, conservative markers, such as *rbcl* (large subunit of RubisCo) or more variable markers (*trnH-psbA* intergenic spacer) can be used. After purification of the DNA, the barcode is amplified by genomic PCR, and then the amplicon is sequenced, the sequences are checked for quality, aligned, and compared to related sequences from the database to construct a phylogeny (for instance by the Neighbour-Joining algorithm).

Goal: For our strawberry project (see project description 1), the relationship between the accession should be addressed, to see the degree of variability in our data set, as well as the relationship between wild and cultivated strawberries and their American ancestors.

Approach: Since we are here in the realm of a single species, we need a barcoding marker with a high variability. Therefore, the *trnH-psbA* intergenic spacer is used. The sequences are aligned with homologues recovered from GenBank (using blast and taxonomy view), and a phylogenetic tree is constructed based on the Neighbour-Joining algorithm.

Background:

Information on molecular phylogeny:

lecture Molecular Phylogeny in the Plant Evolution course <http://www.botanik.kit.edu/botzell/579.php>

Protocols for barcoding

DNA extraction. Material is shock-frozen in liquid nitrogen, and stored at -80 °C. The frozen material is ground using a high-throughput disruptor (TissueLyser, Qiagen, Germany), and the DNA is extracted using the InnuPrep Plant DNA Kit (Macherey-Nagel) following the instructions of the producer. Around 50-100 mg of plant material is sufficient. The kit is based on affinity to a column, whereby the DNA binds to the matrix, such that contaminants can be eluted. Subsequently, the DNA can be eluted by elution buffer or nuclease-free water. Typical volumes are 100 µl of eluate. Quality is controlled by spectrophotometry via the Nano-Drop. The device is first initialised by pipetting 1.5 µl of nuclease-free water. Then the device is calibrated by pipetting 1.5 µl of the blank (the solvent for the DNA), then the DNA sample is measured (1.5 µl). The values give the concentration of DNA (in ng/µl) and its purity (determined as A_{260}/A_{280} measuring contamination by proteins and A_{260}/A_{230} measuring contamination by polysaccharides). The A_{260}/A_{280} should be ideally 1.8, higher values indicate RNA contamination, lower values indicate protein contamination. The A_{260}/A_{230} value should be between 1.8 and 2.2. The concentration should then be adjusted to 50 ng / µl for optimal use in genomic PCR.

PCR. Sample volume is 10 µl (30 µl for sequencing). Each sample with 1 µl buffer (10x), 1 µl BSA (10 mg / ml), 0.2 µl of each primer (10 pmol), 0.2 µl dNTPs (10 mmol each), and 0.1 µl Taq (New England Biolabs, Frankfurt) and ~0.5 µl template (50 ng/µl).

Table 1: PCR protocol for amplification of tnrH-psbA

step		temperature	Time
initial denaturation		95°C	2:00 min
denaturation	35x	95°C	0:30 min
annealing		60°C	0:30 min
elongation		68°C	1:00 min
final elongation		68°C	5:00 min
storage		12°C	-

Table 2: oligonucleotide primers for typical plant barcodes. Amplicon lengths are predicted for different *Ocimum* taxa investigated in the current study.

target	Sequence	Amplicon (bp)
rbclLa*	F: 5-ATGTCACCACAAACAGAGACTAAAGC-3 R: 5-CGTGGTGGACTTGATTTTAC-3	599

matK	F: 5-ACCCAGTCCATCTGGAAATCTGGTTC-3 R: 5-CGTACAGTACTTTTGTGTTTACGAG-3	878
trnL-F igs	F: 5-CGAAATCGGTAGACGCTACG-3 R: 5-ATTTGAACTGGTGACAGAG-3	803-1209
trnH-psbA igs**	F: 5-GTTATGCATGAACGTAATGCTC-3 R: 5-CGCGCATGGTGGATTCAATCC-3	420-446

*source for the oligonucleotide primers for rbcLa

Levin RA, Wagner WL, Hoch PC, et al. (2003) Family Level Relationships of Onagraceae Based on Chloroplast rbcL and ndhF Data. *Am J Bot* 90, 107-115

**source for the oligonucleotide primers for trnH-psbA

Sang T et al. (1997) Chloroplast DNA phylogeny, reticulate evolution, and biogeography of *Paeonia* (Paeoniaceae). *Am J Bot* 84, 1120.

Tate JA, Simpson BB (2003): Paraphyly of *Tarasa* (Malvaceae) and diverse origins of the polyploid species. *Systematic Botany* 28, 723–737.

Trouble shooting for recalcitrant samples (e.g. dried material from commercial samples).

Here, the DNA can be extracted using the Invisorb® MSpin Plant Mini Kit (Stratec Molecular) following the instructions of the producer. Further, for the PCR 10 mg/ml of bovine serum albumin were added.

sequence analysis. The sequences are processed like explained in the sequence tutorial. Basically, the quality of the sequences is assessed by visual inspection of the chromatograms, by BLAST search of the sequence and validation of the sequence identity, and by alignment of the two reads (forward and reverse complement) using the Alignment Explorer of the freeware MEGA (<http://www.megasoftware.net/>). In addition, the NCBI database can be searched by BLAST for homologues, which can be selected by help of the Taxonomy View tool to get a relevant collection of homologues from related taxa. The alignment can then be trimmed and used to construct and bootstrap a Neighbor-Joining (NJ) tree.