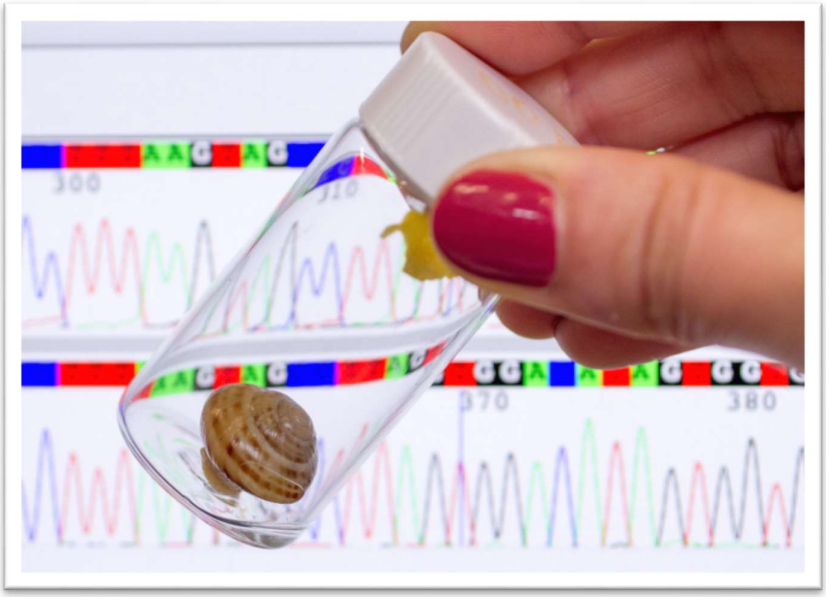
July 2019

Lab Protocol II

Sanger Sequencing



ANNOTATED DNA SEQUENCES = GENETIC REFERENCE

The nucleotide sequence of a known region of the genome is a source of information about the specimen. The sequences of different specimens thus make it possible to characterize genetically the species, for one or more regions of the genome. The genetic inventory of biodiversity therefore serves as a reference point for Swiss species and has a variety of applications.

Markers & Primers The markers selected to characterize the species must be informative regions of the DNA - low intraspecific variability and high interspecific variability. A good marker must have adjacent sites preserved to facilitate the creation of primers that, depending on the planned application, provide more or less complete coverage of the species in the group of organisms studied. The length of the DNA sequence must be short enough to be used with current amplification and sequencing methods.

PCR & Sequencing DNA sequencing is the process of determining the sequence of nucleic acids that compose the target region – selected marker. With the PCR, the selected specific region is amplified exponentially to generate thousands of additional copies of this particular DNA fragment.

Reference sequences

DNA sequences annotated, stored and gathered for easy public access.

Files

DNA sequencing produces two types of files: individual chromatograms - ab1; edited consensus sequences - fst.

Data

Information relating to the acquisition of the sequences is essential, as well as the sequence files themselves - chromatograms and consensus sequences.

1	2	3
CHROMATOGRAMS SwissBOL - Files	CONSENSUS SEQUENCES SwissBOL - Files	INFORMATION SwissBOL - Matrix

sofia.wyler@unine.ch

BEST PRACTICES

Manipulation Appropriate and sterile instruments must be used when preparing PCR and handling DNA extracts. The use of gloves is mandatory during all handling and precautions must be taken to avoid cross-contamination. When preparing PCR, it is essential to use sterilized auxiliaries (containers, tubes, eppendorfs, falcons, tips) and the use of filter tips is clearly an advantage. In addition, it is strongly recommended to use pipettes only used to work with low DNA concentrations, as well as handling in pre-PCR rooms. The vast majority of current PCR methods are based on a thermal cycle and the use of thermostable DNA polymerase. Once amplified, DNA is much less prone to contamination and vigilance can be released for the post-PCR processes of verification and purification of the amplification product, and then sequencing.

Preparation A file name that can be understood from the beginning makes it easier to validate the data. Consensus sequences must be submitted according to one of two alternatives: 1. one file per sequence (file name: VOUCHERID_SEQMARKER.fst or EXTID_SEQMARKER.fst); 2. one file per marker gathering all sequences for this project (file name: SEQMARKER.fst; sequence title: VOUCHERID or EXTID). The chromatogram files must be named according to the following rule: VOUCHERID_TRAMARKER_TRASEQ_PRIMER.ab1 or EXTID_TRAMARKER_TRASEQPRIMER.ab1).

Information To facilitate the maintenance of links between the various objects, it is recommended to keep the unique identifier assigned to the specimen - BARCODE TAG (unique specimen identifier on label provided by GBIF.ch); INFOSPECIES ID (identifier of the occurrence assigned by a data centre); MUSEUM ID (catalogue number provided by the museum institution); FIELD ID (arbitrary identifier temporarily assigned during the sampling). Information about the primers used to amplify and sequence DNA ensures data quality.

IDENTIFICATION	VOUCHERID	1	Internal identifier used to ensure links between genetic information - BARCODE TAG (GBIFCHID) or INFOSPECIES ID (occurrence ID) or MUSEUM ID (catalog Number)
	PROJECTCODE	1	Code attributed to the project to which sequence aquisition is associated
EXTRACTION (EXT)	DNA_EXTID	1	Identifier attributed by the lab that assures temporarily the link between genetic information
	DNA_EXTINS	1	Institution /Company where DNA was extracted
SEQUENCE CONSENSUS (SEQ)	SEQ_GENOME	1	Organelle source of the sequence
	SEQ_MARKER	1	Fragment or region of the genome targetted
	SEQ_FILENAME	1	Name of the fasta file containing the consensus sequence
	SEQ_REPOSITORYURL	4	Internet site where the fasta file containing the consensus sequence was deposited
	SEQ_NCBI	2	Accession number given by GenBank to the sequence at the moment of deposition
	SEQ_BOLDPROCESSID	2	Identifier assigned by BOLD to the specimen record
	SEQ_BIBLIOREF	2	Bibliographic reference of the article where the sequence was published for the first time
	SEQ_BIBLIOREFDOI	4	ID associated to the article where the sequence was published for the first time
IDENTIFICATION	VOUCHERID	1	Internal identifier used to ensure links between genetic information - BARCODE TAG (GBIFCHID) or INFOSPECIES ID (occurrence ID) or MUSEUM ID (catalog Number)
CHROMATOGRAMMES (TRA)	TRA_FILENAME	1	Name of the chromatogram file corresponding to the individual sequence
	TRA_MARKER	1	Fragment or region of the genome targetted
	TRA_INS	3	Institution / Company where DNA sequencing was performed
	TRA_STAFF	4	Person who executed the sequencing - first and last names written in full
	TRA_YEAR	4	Year or date (ISO) of Sequecing YYYY-MM-DD
	TRA_SEQPRIMER	1	Name of the primer used to sequence
	TRA_PCRPRIMER1	1	Name of the Forward primer used in the amplification reaction
	TRA_PCRPRIMER2	1	Name of the Reverse primer used in the amplification reaction
IDENTIFICATION	PROJECTCODE	1	Code attributed to the project to which sequence aquisition is associated
PRIMERS (PRI)	PRI_MARKER	1	Fragment or region of the genome targetted
	PRI_NAME	1	Name given to the primer by the authors who designed it
	PRI_SEQ	1	DNA sequence of the primer (5'-3')
	PRI_DIR	1	Direction of the primer
	PRI_BIBLIOREF	2	Bibliographic reference of the article where the primer was published for the first time
	PRU_BIBLIOREFDOI	4	ID of the article where the primer was published for the first time

Prioritisation of information : 1 – mandatory ; 2 – basic ; 3 – recommended ; 4 – optional.

sofia.wyler@unine.ch