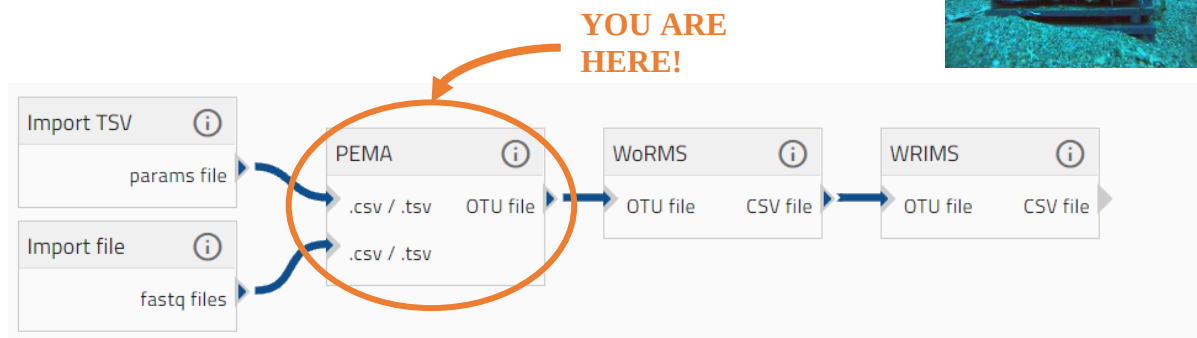
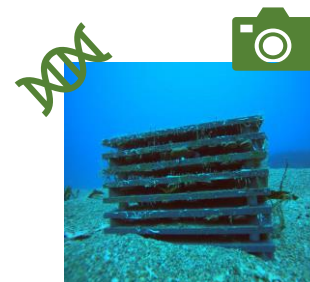


PEMA runner: fact sheet



Aim	This service runs the P.E.M.A. pipeline to process the DNA sequences from the ARMS samples selected using the services “ARMS filter and select/ARMS choose and parameterize” and downloaded using the service “PEMA sequences retriever”. This service can take some hours to run depending on the number of DNA sequences/ARMS samples that need to be analysed.
Input	A <i>Parameter.tsv</i> file provided by the user and the <i>fastq</i> files downloaded from ENA. The number of <i>fastq</i> files used as input for this service varies based on the number of DNA sequences that the user has chosen/selected within the previous workflow steps.
Requirements	The user must specify only some parameters in the <i>Parameter.tsv</i> file depending on the genetic marker chosen for the analysis. This file must be uploaded by the user and an example file can be found at: https://github.com/hariszaf/pema/blob/master/analysis_directory/parameters.ts .
Analytical tools	This service will always run the last released version of P.E.M.A. PEMA uses a series of analytical tools, databases, and Big Data Script language for the sequence analyses including: BigDataScript programming language - https://pcingola.github.io/bds/manual/site/index.html FASTQC - https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ Trimmomatic - http://www.usadellab.org/cms/?page=trimmomatic Cutadapt - https://cutadapt.readthedocs.io/en/stable/ BayesHammer - included in SPAdes - http://cab.spbu.ru/software/spades/ PANDAsseq - https://github.com/neufeld/pandaseq OBITools - https://pythonhosted.org/OBITools/welcome.html BLAST Command Line Applications - https://www.ncbi.nlm.nih.gov/books/NBK52640/ VSEARCH-2.9.1 - https://github.com/torognes/vsearch/releases/tag/v2.9.1

	SWARM - https://github.com/torognes/swarm CROP - https://github.com/tingchenlab/CROP CREST - https://github.com/lanzen/CREST RDPClassifier - https://github.com/rdpstaff/classifier (RPDtools are required in order to execute RDPClassifier) SILVA db - https://www.arb-silva.de/no_cache/download/archive/current/Exports/ MIDORI db - http://reference-midori.info/index.html "phat" algorithm, from the "gappa" package - https://github.com/lczech/gappa/wiki/Subcommand:-phat MAFFT - https://mafft.cbrc.jp/alignment/software/ RAxM-ng - https://github.com/amkozlov/raxml-ng PaPaRa - https://cme.h-its.org/exelixis/web/software/papara/index.html EPA-ng - https://github.com/Pbdas/epa-ng
Output	Three output files will be produced by the service: • A <i>final_table.tsv</i> which is the OTU table. The columns correspond to the sequence/samples processed, whereas the first column is an ID column and the final one contains the species taxonomic information. The specific columns for each sequence/sample contain integers values with the number of reads assigned to each sample and taxon and the title of these columns is the ENA accession number; • A <i>rdpClassifierInput.fasta</i> file. This file can be useful if the user wishes to subsequently analyse the DNA sequences using BOLD. This output file will be produced only when COI sequences are processed in PEMA. • A <i>PEMA_outputs.zip</i> folder with all outputs produced by the pipeline and that the user can download and inspect.
Citations	Haris Zafeiropoulos, Ha Quoc Viet, Katerina Vasileiadou, Antonis Potirakis, Christos Arvanitidis, Pantelis Topalis, Christina Pavloudi, Evangelos Pafilis, PEMA: a flexible Pipeline for Environmental DNA Metabarcoding Analysis of the 16S/18S ribosomal RNA, ITS, and COI marker genes, GigaScience, Volume 9, Issue 3, March 2020, giaa022, https://doi.org/10.1093/gigascience/giaa022
Metadata	https://metadatalogue.lifewatch.eu/srv/eng/catalog.search#/metadata/e6289cf6-dde7-4584-89e6-bd7e907a4b5f
Further readings	PEMA article - https://doi.org/10.1093/gigascience/giaa022 PEMA GitHub - https://github.com/hariszaf/pema