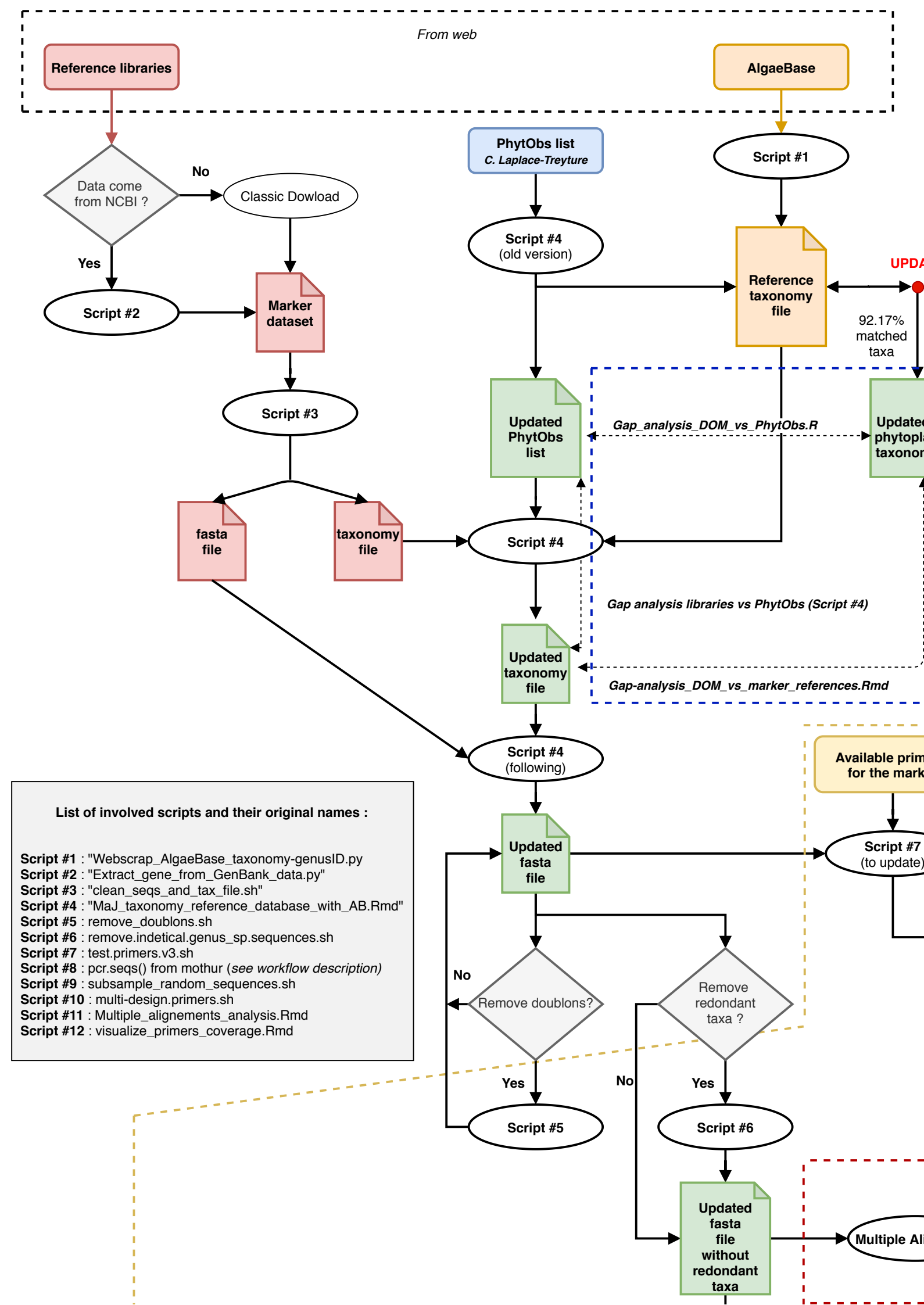
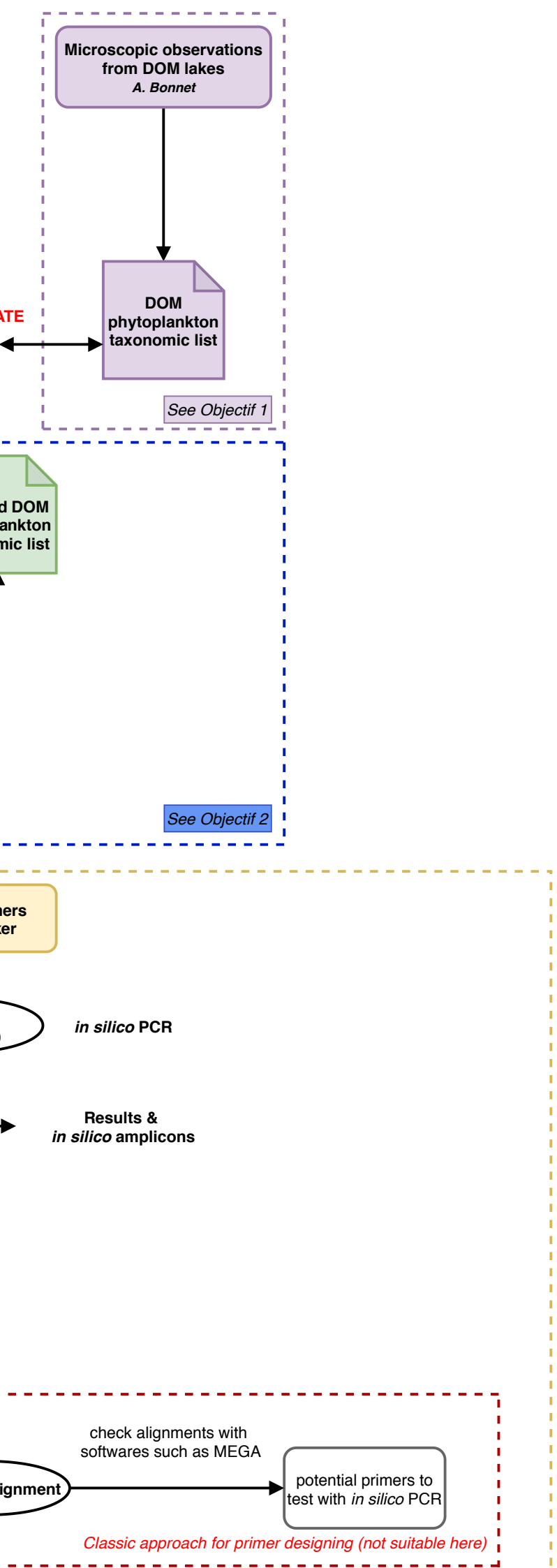
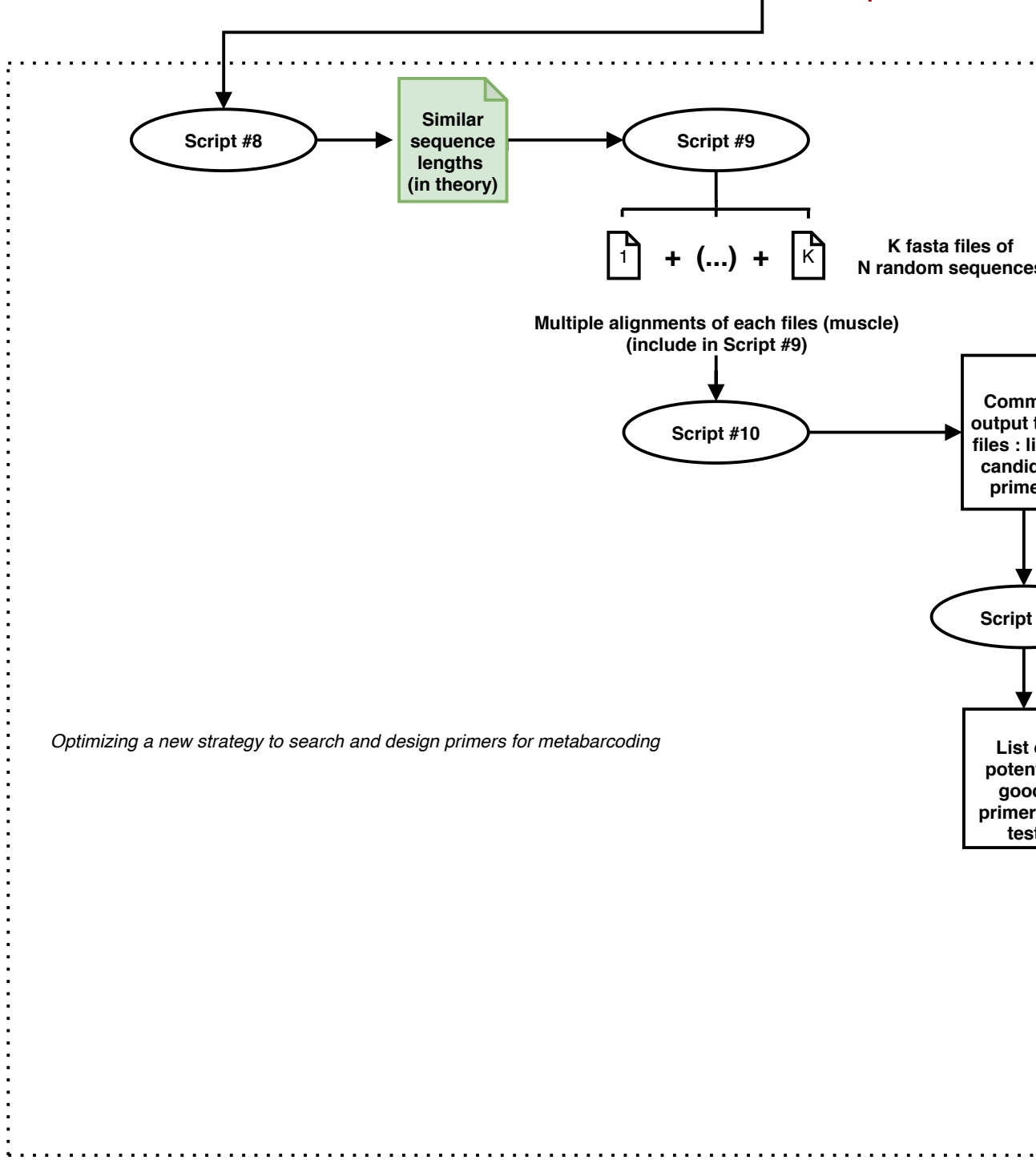
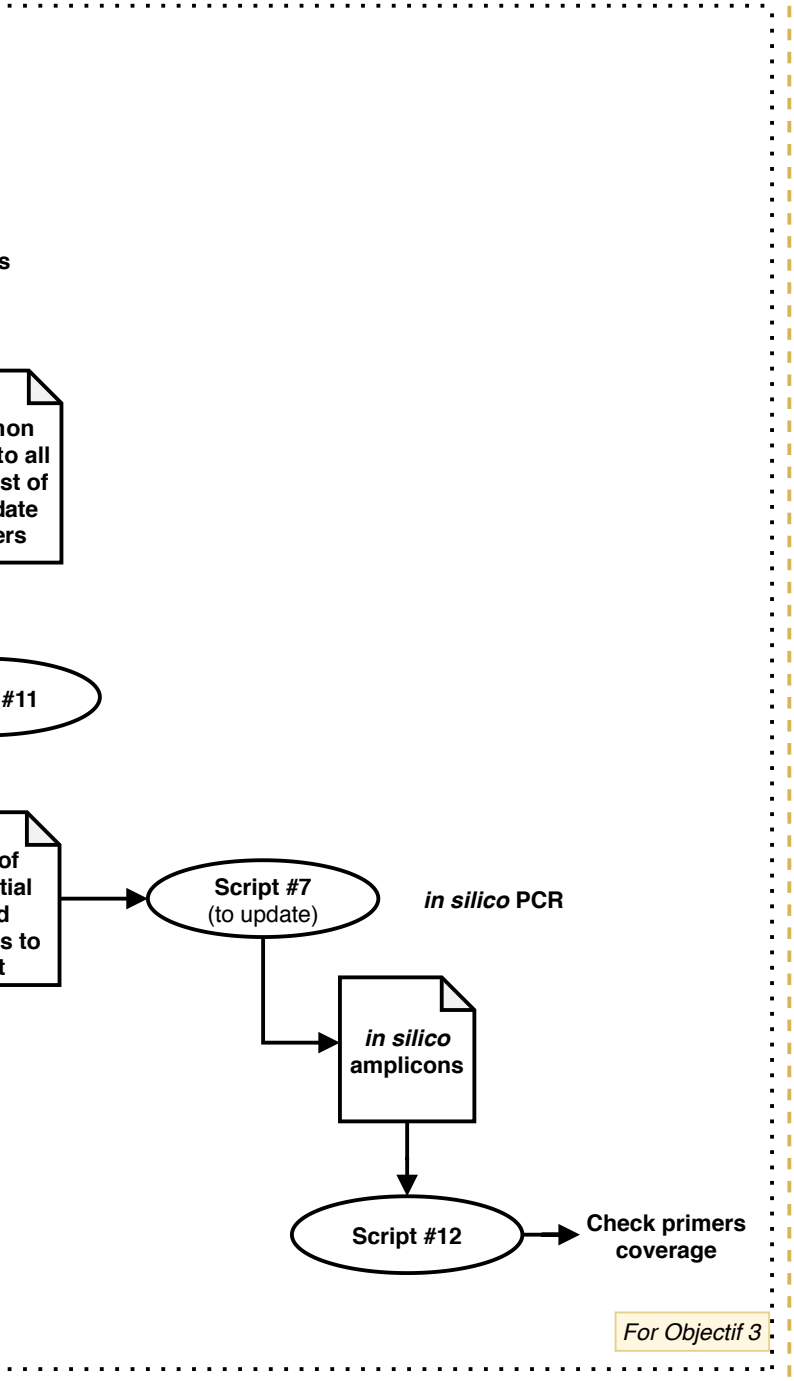

From web
Reference libraries
AlgaeBase
Data come from NCBI ?
No
Yes
Classic Dowload
Script #2
Marker dataset
Script #3
fasta file
taxonomy file
PhytObs list
C. Laplace-Treyture
Script #4 (old version)
Script #1
Reference taxonomy file
UPDA
92.17% matched taxa
Updated PhytObs list
Gap_analysis_DOM_vs_PhytObs.R
Updated phytopl taxonor
Script #4
Gap analysis libraries vs PhytObs (Script #4)
Updated taxonomy file
Gap-analysis_DOM_vs_marker_references.Rmd
Script #4 (following)
Available prim for the mark
Updated fasta file
Script #7 (to update)
No
Remove doublons?
Remove redondant taxa ?
Yes
No
Yes
Script #5
Script #6
Updated fasta file without redondant taxa
Multiple Al
List of involved scripts and their original names :
Script #1 : "Webscrap_AlgaeBase_taxonomy-genusID.py
Script #2 : "Extract_gene_from_GenBank_data.py"
Script #3 : "clean_seqs_and_tax_file.sh"
Script #4 : "MaJ_taxonomy_reference_database_with_AB.Rmd"
Script #5 : remove_doublons.sh
Script #6 : remove.indetical.genus_sp.sequences.sh
Script #7 : test.primers.v3.sh
Script #8 : pcr.seqs() from mothur (see workflow description)
Script #9 : subsample_random_sequences.sh
Script #10 : multi-design.primers.sh
Script #11 : Multiple_alignements_analysis.Rmd
Script #12 : visualize_primers_coverage.Rmd


Microscopic observations from DOM lakes
A. Bonnet
DOM phytoplankton taxonomic list
See Objectif 1
See Objectif 2
in silico PCR
Results & in silico amplicons
check alignments with softwares such as MEGA
potential primers to test with in silico PCR
Classic approach for primer designing (not suitable here)


Script #8
Similar sequence lengths (in theory)
Script #9
1
+ (...) +
K
K fasta files of N random sequences
Multiple alignments of each files (muscle) (include in Script #9)
Script #10
Comm output files : li candic prime
Script
List potent good primer test


s
non
to all
st of
date
ers
#11
of
tial
d
s to
t
Script #7
(to update)
in silico PCR
in silico
amplicons
Script #12
Check primers
coverage
For Objectif 3