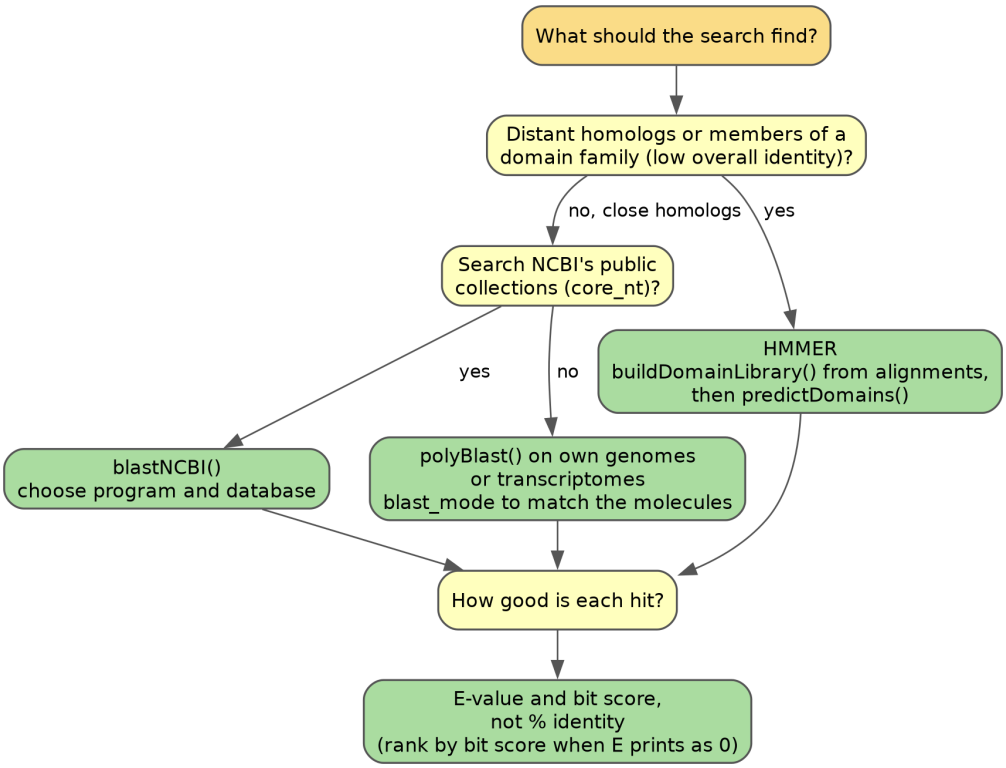
homology: summary

Integrated Bioanalytical Chemistry · thebustalab.github.io/integrated_bioanalytics



A homology search starts with two choices. **BLAST or HMMER**: BLAST is the fast first pass for close homologs and rapid annotation, while HMMER profiles built from alignments are more sensitive for distant homologs and shared domains. **Where to search**: `blastNCBI()` sends one sequence to NCBI's servers, and `polyBlast()` searches our own genomes or transcriptomes with many queries at once, with a `blast_mode` that matches the molecule types. Hits are then judged by E-value and bit score rather than percent identity, and ranked by bit score once E-values become so small that R prints them as zero.

Function	What it does	Arguments students vary
`blastNCBI()`	remote BLAST of one sequence on NCBI's servers, returning a tibble of hits	`query_sequence`, `program` (`"blastn"`, `"megablast"`), `database` (`"core_nt"`)
`polyBlast()`	local BLAST+ search of one or more queries against several sequence collections, writing a hit list and one file per hit (returns nothing)	`named_subjects_list`, `query_in_path`, `blast_mode`, `e_value_cutoff` (default 1), `queries_in_output`, plus the paths `sequences_of_interest_directory_path`, `monolist_out_path`, `blast_module_directory_path`
`buildDomainLibrary()`	trains profile HMMs from multiple sequence alignments into a domain library	`alignment_in_paths`, `domain_library_out_path`
`predictDomains()`	scores sequences against a domain library to find likely family members	`fasta_in_path`, `domain_library_in_path`

Query	Subjects	`blast_mode`
nucleotide	nucleotide	`"nnblastn"` (classic blastn)
protein	nucleotide, translated	`"ntblastp"` (like tblastn)
protein	protein	`"pnblastp"` (like blastp)