**S3 Table.** BADGE settings with input range, default value und description. Changing BADGE settings can be easily done using any text editor. Open *BADGE.sh* with your text editor and change settings as desired. Save your changes before you start a new BADGE run. We recommend to save a copy of *BADGE.sh* somewhere else. If your version does not work anymore, just replace it by the original *BADGE.sh* script.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Setting name in BADGE | type | options | default value | description and notes |
| clean\_up | boolean | true/false  | true | if switched on (true), only the most important output will be kept – all other files are deleted |
| min\_DMG\_occurrence | real | 0 - 1 | 1 | minimum occurrence of a DMG in group A to be reported, a value of 0.5 would mean, that the DMGs to identify only have to be in 50% of all members of group A in order to be reported |
| special\_character | boolean | true/false | true | replaces incompatible special characters with ‘\_’ |
| num\_blast\_proc | real | > 0 | 4 | number of parallel blast processes to start |
| megablast\_perc\_identity\_cut | real | 0 - 100 | 95 | % identity threshold value – minimum identity of hit to be kept |
| megablast\_e\_value | real | > 0 | 1×10-15 | e-value threshold – maximum value of blast hit to be kept |
| megablast\_within\_group\_qscov | real | 0 - 1 | 0.95 | Query / subject coverage (length of query divided by length of subject and vice versa), threshold within a group – minimum value of blast hit to be kept (relevant for core determination – Step 3) |
| megablast\_between\_group\_qscov | real | 0 - 1 | 0.50 | Query / subject coverage threshold between groups – minimum value of blast hit to be kept (relevant for exclusion of ORFs – Step 1 and Step 2) |
| dc\_mode | boolean | true/false | false | if switched on (true), basic DMG identification is done using dc-megablast instead of megablast, more sensitive but slower |
| dc\_filter  | boolean | true/false | true | if switched on (true), DC-megablast filter will be applied to data – potential DMGs with a long (> 50 % qscov) but low identity (default between 70 – 90 % identity) hit in other group will be discarded |
| dc\_perc\_identity\_cut | real | 0 - 100 | 70 | c.f. megablast\_perc\_identity\_cut  |
| dc\_blast\_e\_value | real | > 0 | 10 | c.f. megablast\_e\_value |
| dc\_between\_group\_qscov | real | 0 - 1 | 0.50 | c.f. megablast\_between\_group\_qscov |
| blastn\_filter | boolean | true/false | true | if switched on (true), blastn filter will be applied to data – potential DMGs with a short (> 25 % qscov) but high identity (> 95 % identity) hit in other group will be discarded |
| blastn\_perc\_identity\_cut | real | 0 - 100 | 95 | c.f. megablast\_perc\_identity\_cut |
| blastn\_e\_value | real | > 0 | 10 | c.f. megablast\_e\_value |
| blastn\_between\_group\_qscov | real | 0 - 1  | 0.25 | c.f. megablast\_between\_group\_qscov |
| protein\_level | boolean | true/false | false | if switched on (true), ORFs will be translated and used to perform BADGE (Steps 1 and 3) using blastp – NOTE: if Protein-Level is active dc-megablast filter and blastn filter are disabled automatically |
| blastp\_perc\_identity\_cut | real | 0 - 100 | 50 | c.f. megablast\_perc\_identity\_cut |
| blastp\_e\_value | real | > 0 | 10 | c.f. megablast\_e\_value |
| blastp\_within\_group\_qscov | real | 0 - 1 | 0.50 | c.f. megablast\_within\_group\_qscov |
| blastp\_between\_group\_qscov | real | 0 - 1 | 0.50 | c.f. megablast\_between\_group\_qscov |
| fastatranslate\_geneticcode | real | 1 - 25 | 11 | genetic code to be used for translation – 11 corresponds to bacterial, archaeal and plant plastid code |
| fastatranslate\_frame | real | 1 - 3 | 1 | reading frame to be translated |
| protein\_level\_clean\_up | boolean | true/false | true | if switched on (true), translated orf files will be removed after BADGE is done |
| identify\_overlapping | boolean | true/false | false | if switched on (true), overlapping DMGs are labelled correspondingly |
| mut\_level\_nt | boolean | true/false | false | if switched on (true) BADGE will look for DMGs with ANY (even single nucleotide changes) differences to the opposite group |
| mut\_level\_aa | boolean | true/false | false | if switched on (true) BADGE will look for DMGs with ANY (even single amino acid changes) differences to the opposite group in protein\_level mode |