Table 1

|  |  |  |  |
| --- | --- | --- | --- |
| **No.** | **Strains** | **Access. No.** | **Length (amino acids)** |
| 1 | *B. safensis* strain CFA06 | KEP29487.1 | 616 |
| 2 | *B. safensis* strain JG-B5T | RAU57979.1 | 616 |
| 3 | *B. safensis* strain RIT372 | KIZ56126.1 | 616 |
| 4 | *B. safensis* | WP\_041109021.1 | 616 |
| 5 | *B. safensis* strain LK12 | KMN80921.1 | 616 |
| 6 | *B. safensis* strain LK32 | KML49660.1 | 616 |
| 7 | *B. safensis* strain LK21 | KML10526.1 | 616 |
| 8 | *B. safensis* strain Ingolstadt | OYN66870.1 | 616 |
| 9 | *B. safensis* strain JPL\_MERTA8-2 | KKD42423.1 | 616 |
| 10 | *B. safensis* strain SCAL1 | KMK71141.1 | 616 |

Table 2

|  |  |  |  |
| --- | --- | --- | --- |
| **No.** | **Program** | **Function** | **Source** |
| 1 | BioEdit ver. 7.2.5 (Hall, 1999) | Editing of biological sequences and alignment. | [www.mbio.ncsu](http://www.mbio.ncsu).edu/BioEdit/bioedit.html |
| 2 | BLAST (Altschul et al., 1990) | Comparing nucleotide or protein sequences to sequence databases and calculating the statistical significance of matches. | <https://blast.ncbi.nlm.nih.gov/Blast.cgi> |
| 3 | ClustalX ver. 2.1 (Larkin et al., 2007) | Performing multiple sequence alignment (MSA) to find conserved/unconserved regions in the alignment. | www.clustal.org |
| 4 | Clustal Omega (Sievers et al., 2011) | Performing MSA. Numbers indicating positions of bases/residues in the alignment are shown in this version. | [www.ebi.ac.uk/Tools/msa/clustalo/](http://www.ebi.ac.uk/Tools/msa/clustalo/) |
| 5 | MEGA ver. 7.0.18 (Kumar et al., 2016) | Inferring phylogenetic trees. | www.megasoftware.net |
| 6 | SnapGene ver. 2.3.2 | Planning, visualizing, and simulating restriction-ligation cloning, PCR, and gel electrophoresis. | SnapGene software (from GSL Biotech; available at [snapgene.com](https://www.snapgene.com/)) |
| 7 | GC Content Calculator (McGowan, 2015) | Calculating %GC of a given nucleotide sequence. | www.jamiemcg. github.io/bioinf/index.html |
| 8 | EMBOSS Transeq (Rice et al., 2000) | Translating nucleotide sequences to their corresponding peptide sequences. | [www.ebi.ac.uk/Tools/st/emboss\_ transeq/](http://www.ebi.ac.uk/Tools/st/emboss_%20transeq/) |
| 9 | InterProScan (Zdobnov and Apweiler, 2001) | Scanning a sequence for matches against the InterPro protein signature databases. | [www.ebi.ac.uk/interpro/search/sequence-search](http://www.ebi.ac.uk/interpro/search/sequence-search) |
| 10 | ScanProsite (De Castro et al., 2006) | Scanning protein sequences for the occurrence of patterns, profiles and motifs against the PROSITE database. | [www.prosite.expasy.org/scanprosite/](http://www.prosite.expasy.org/scanprosite/) |
| 11 | Compute pI/Mw (Gasteiger et al., 2005) | Predicting protein isoelectric point (pI) and molecular weight (MW). | [web.expasy.org/compute\_pi/](https://web.expasy.org/compute_pi/) |
| 12 | ProtParam (Gasteiger et al., 2005) | Computing amino acid composition. | web.expasy.org/protparam/ |
| 13 | MyDomains (Hulo et al., 2007) | Creating custom domain figures. | [www.prosite.expasy.org/mydomains/](http://www.prosite.expasy.org/mydomains/) |
| 14 | PredictProtein (Yachdav et al., 2014) | Predicting protein secondary structures composition. | www.predictprotein.org |
| 15 | Sequence Annotated by Structure (SAS) (Milburn et al., 1998) | Predicting secondary structures positions in a given protein sequence. | www.ebi.ac.uk/thornton-srv/databases/sas/ |
| 16 | SWISS-MODEL (Waterhouse et al., 2018) | Modeling of 3D protein structures based on homology. | www.swissmodel.expasy.org/interactive |

Table 3

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Clear zones around the bacterial colonies in Congo-red dyed media |  | | |  | | |  | | |
| Isolate | DSI 2 | | | DSI 3 | | | DSI 5 | | |
| Repetition | I | II | III | I | II | III | I | II | III |
| Clear zone diameter (A) (cm) | 1.6 | 1.7 | 1.9 | 0.9 | 0.95 | 0.9 | 0.9 | 1.4 | 0.8 |
| Colony diameter (B) (cm) | 0.7 | 0.75 | 0.7 | 0.9 | 0.8 | 0.8 | 0.8 | 0.8 | 0.75 |
| Ratio A/B | 2.29 | 2.27 | 2.71 | 1.00 | 1.19 | 1.13 | 1.13 | 1.75 | 1.07 |
| Mean | 2.42 | | | 1.11 | | | 1.32 | | |

Table 4

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| DSI 2 | *B. safensis* SCAL1 | *Bacillus* sp.  WP8 | *B. safensis* U17-1 | *B. safensis* U41 | *B. pumilus* BY-1 | *B. xiamenensis* VV3 |
| Nucleotide Sequence | | | | | |
| 99.3% | 99.19% | 97.73% | 97.73% | 95.30% | 87.26% |
| Amino Acid Sequence | | | | | |
| 99.35% | 99.2% | 98.2% | 98.2% | 97.2% | 91.4% |

Amino acid sequences of endoglucanase from: *B. safensis* SCAL1 (KMK71141.1); *Bacillus* sp. WP8 ([WP\_039178590.1](https://www.ncbi.nlm.nih.gov/protein/WP_039178590.1?report=genbank&log$=prottop&blast_rank=7&RID=HVANE2WB015)); *B. safensis* U17-1 ([APT54136.1](https://www.ncbi.nlm.nih.gov/protein/1129027050)); *B. safensis* U41 ([APT49588.1](https://www.ncbi.nlm.nih.gov/protein/1129022501)); *B. pumilus* BY-1 ([AEX68682.1](https://www.ncbi.nlm.nih.gov/protein/372100184)); dan *B. xiamenensis* VV3 ([AOZ87246.1](https://www.ncbi.nlm.nih.gov/protein/1095430120)).

Table 5

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **No** | **Amino Acid (AA)** | **AA Group** | **AA Percentage in** | | | |
| **EG\* DSI 2** | **EG BS** | **EG BA** | **EG BL** |
| 1 | Arginine (Arg/R) | Charged (side chains may form salt bridge) | 3.4 | 3.0 | 2.6 | 1.7 |
| 2 | Lysine (Lys/K) | 5.0 | 8.8 | 3.3 | 6.0 |
| 3 | Aspartic acid (Asp/D) | 6.5 | 7.2 | 8.9 | 5.2 |
| 4 | Glutamic acid (Glu/E) | 6.2 | 4.0 | 8.3 | 2.2 |
| 5 | Histidine (His/H) | 2.4 | 1.4 | 2.6 | 1.7 |
| Total | | | 23.5 | 24.4 | 25.7 | 16.8 |
| 6 | Glutamine (Gln/Q) | Polar / hydrophilic (form hydrogen bond) | 4.5 | 4.2 | 4.0 | 2.6 |
| 7 | Asparagine (Asn/N) | 5.7 | 7.6 | 6.9 | 9.9 |
| 8 | Glysine (Gly/G) | 7.1 | 9.0 | 8.3 | 9.1 |
| 9 | Serine (Ser/S) | 6.2 | 7.2 | 7.3 | 12.1 |
| 10 | Threonine (Thr/T) | 6.2 | 6.8 | 3.6 | 8.2 |
| 11 | Tyrosine (Tyr/Y) | 6.0 | 4.0 | 4.3 | 6.0 |
| 12 | Cysteine (Cys/C) | 0.3 | 0.6 | 0 | 0 |
| Total | | | 36 | 39.4 | 34.4 | 47.9 |
| 13 | Alanine (Ala/A) | Nonpolar / hydrophobic (commonly present inside protein core) | 10.6 | 7.0 | 7.9 | 6.5 |
| 14 | Isoleusine (Ile/I) | 3.6 | 6.2 | 6.9 | 6.0 |
| 15 | Leusine (Leu/L) | 7.1 | 7.0 | 5.3 | 3.0 |
| 16 | Methionine (Met/M) | 1.8 | 1.6 | 2.3 | 0.9 |
| 17 | Phenilalanine (Phe/F) | 4.5 | 3.0 | 3.3 | 3.4 |
| 18 | Valine (Val/V) | 5.5 | 5.0 | 6.9 | 6.5 |
| 19 | Proline (Pro/P) | 4.2 | 3.4 | 3.6 | 2.6 |
| 20 | Tryptophan (Trp/W) | 3.1 | 2.8 | 3.6 | 6.5 |
| Total | | | 40.4 | 36 | 39.8 | 35.4 |

\*EG DSI 2 = thermostable endoglucanase from *Bacillus safensis* DSI 2

EG BS = thermostable endoglucanase from *B. subtilis* I15 (NCBI No. ACK38261.1)

EG BA = thermolabile endoglucanase from *B. agaradhaerens* (PDB No. 7A3H)

EG BL = thermolabile endoglucanase from *B. licheniformis* (PDB No. 2JEN)