|  |
| --- |
| **Table S20.** *P. gigantea* ABC models |
| **Existing model** | **Suggested model / changes** |
| Model name | Protein ID | Location | Model name | Protein ID | Location |
| estExt\_Genewise1.C\_390060 | 87882 | scaffold\_39:94658-97723 (-) | estExt\_Genewise1Plus.C\_390061 | 104024 | scaffold\_39:94658-97792 (-) |
| estExt\_Genewise1.C\_810023 | 90887 | scaffold\_81:44869-49148 (-) | estExt\_Genemark1.C\_810016 | 128458 | scaffold\_81:44869-49150 (-) |
| fgenesh1\_pm.3\_#\_77 | 21134 | scaffold\_3:295746-296813 (-) | estExt\_Genemark1.C\_30113 | 124233 | scaffold\_3:295746-298376 (-) |
| fgenesh1\_pm.137\_#\_8 | 25601 | scaffold\_137:31582-34289 (+) | CE62642\_3357 | 194022 | scaffold\_137:31433-34608 (+) |
| estExt\_fgenesh1\_pm.C\_10057 | 27132 | scaffold\_1:171968-174644 (+) | Sequencing gap (N in the position 1891) of about 30 bp. As a result, I’d suggest that there is no intron (and no frame-shift) between exons 6 and 7.  |  |  |
| estExt\_fgenesh1\_pm.C\_3030004 and CE204994\_3653 | 32454 and 336374 | scaffold\_303:11013-12222 (-) and scaffold\_303:12521-13899 (-) | estExt\_Genewise1.C\_3030007 | 96729 | scaffold\_303:11013-13936 (-) |
| estExt\_Genewise1Plus.C\_30270 | 98346 | scaffold\_3:356376-357602 (-) | N-terminal part of the gene is missing from the database; the model can be extended upstream based on homology with the related *P. gigantea* proteins (proteins ID 124233 and 96729). | 534834 | scaffold\_3:356376-358055 |
| estExt\_Genewise1Plus.C\_1940030 | 111230 | scaffold\_194:39232-40470 (-) | CE123138\_2353 | 254518 | scaffold\_194:39105-42027 (-) |
| fgenesh1\_pg.113\_#\_1 | 14688 | scaffold\_113:148-1448 (-) | Only small fragment of the gene is present and not possible to reconstitute the entire protein sequence |  |  |
| fgenesh1\_pg.113\_#\_2 | 14689 | scaffold\_113:2091-7842 (-) | There are two sequence gaps within the gene, so it’s not possible to reconstitute the entire aa sequence.  | 534835 | scaffold\_113:2091-8447 (-) |
| fgenesh1\_pm.41\_#\_23 | 23474 | scaffold\_41:102552-108846 (+) | CE244673\_18999 | 376053 | scaffold\_41:101782-108976 (+) |
| fgenesh1\_pm.30\_#\_32 | 22983 | scaffold\_30:154474-160810 (-) | fre\_1\_e\_gw1.30.123.1 | 534837 | scaffold\_30:154474-160810 (-) |
| estExt\_fgenesh1\_pg.C\_1130003 | 36552 | scaffold\_113:14632-21227 (-) | There are two gaps within the sequence, thus it is not possible to reconstitute the entire protein sequence.  | 534838 | scaffold\_113:14632-21227 (-) |
| estExt\_fgenesh1\_pg.C\_2980004 | 38168 | scaffold\_298:10908-17662 (+) | fre\_1\_estExt\_Genemark1.C\_2980008 | 534839 | scaffold\_298:10770-17662 (+) |
| gm1.8277\_g | 121521 | scaffold\_175:36930-45224 (-) | fre\_1\_gm1.8277\_g | 534840 | scaffold\_175:36930-43613 (-) |
| estExt\_Genemark1.C\_570036 | 127559 | scaffold\_57:121648-128305 (+) | CE282348\_2765 | 413728 | scaffold\_57:121475-128597 (+) |
| CE33064\_4524 | 164444 | scaffold\_113:38450-44018 (+) | fre\_1\_estExt\_Genemark1.C\_1130009 | 534841 | scaffold\_113:37502-44008 (+) |
| MIX8954\_2964\_12 | 497319 | scaffold\_181:40693-47165 (+) | CE111560\_17494 | 242940 | scaffold\_181:39964-46863 (+) |
| fgenesh1\_pg.118\_#\_1 | 14797 | scaffold\_118:58-2352 (+) | More than half of the gene is missing, so it’s not possible to reconstitute the entire sequence of the deduced protein |  |  |
| fgenesh1\_pm.70\_#\_13 | 24318 | scaffold\_70:38980-41804 (+) | estExt\_Genemark1.C\_700014 | 127965 | scaffold\_70:38894-42042 (+) |
| estExt\_fgenesh1\_pm.C\_230030 | 28761 | scaffold\_23:117477-120357 (+) | fre\_1\_fgeneshTR\_pm.23\_\_\_29 | 534842 | scaffold\_23:117477-120256 (+) |
| CE60640\_8249 | 192020 | scaffold\_135:12700-14524 (+) | estExt\_Genemark1.C\_1350008 | 129467 | scaffold\_135:12058-14576 (+) |
| fgenesh1\_kg.274\_#\_2\_#\_Locus4143v1rpkm41.13 | 20542 | scaffold\_274:5209-7367 (+) | CE182818\_10251 | 314198 | scaffold\_274:5006-7492 (+) |
| estExt\_fgenesh1\_pg.C\_400020 | 34896 | scaffold\_40:97140-100665 (-) | There is a sequence gap of about 270 bp within the gene, thus it’s not possible to reconstitute the entire protein sequence |  |  |
| estExt\_Genewise1Plus.C\_340085 | 103495 | scaffold\_34:113777-115000 (-) | CE219763\_4564 | 351143 | scaffold\_34:113432-115436 (-) |

\*-amino acids VNIQAS (positions 100 – 105) should be replaced by residues KGQLLEI as in the protein model 25601