

**Additional file 1. The available lancelet PGBD5 genomic sequences are complicated by tandem duplications within presumed protein coding sequences.** The apparently complete lancelet PGBD5 gene spanning exons 1-7 is annotated as Protein and Transcript ID 79338 in the *B. floridae* v1.0 genome assembly and was retrieved from the (-) strand of scaffold 84 (<http://genome.jgi-psf.org/pages/search-for-genes.jsf?organism=Brafl1>). The 7 lancelet coding exons are shown in red font; for illustrative purposes, exon 1 begins with the AUG initiator and exon 7 ends with UAG terminator because the transcription start site and polyadenylation site cannot be reliably predicted. Introns as well as DNA flanking the 5' and 3' ends of the ORF are shown in black. For clarity, each set of imperfect tandem repeats is highlighted alternately in yellow and cyan. Note that the XM\_002604443 mRNA model encodes a more divergent protein sequence than the one shown in Figure 1 because the imperfect tandem repeats differ in the corresponding genomic sequence on the UCSC Genome Browser (chrUn: 288364515-288370215) which has 7 x 60 bp repeats instead of 2 overlapping exon 3, and 6 x 108 bp repeats instead of 3 flanking and including exon 5.

```

>scaffold84
GACCTTTGACATGTAGTTCTAAACTGTGTGGCAAGTGAGAGGCAAGAAGAAATGTTTTGCTGTAATAAATTTTACAGTGTTCGAAAGGTATATAGAGCC 847700
TCACATTGACTTAAATCTGCAGAATATCCTGTAGTGTACCTTAGCATAGTCAGCATCTCAGGAAGTGCCTCCGCTCAATCCCTGCAGCATCAGCACC 847600
ATGGATGATGAGGAGGACAGAGACGGCCGGCCGCTCCAACATCTCTGAGACCTCCGTCCAACCCGGCGGATCAGGAGGGCTCCAGATTCCCGAGC 847500 ex1
ACATGTTCCGGCGGGTCGGCAAGCTCGGACGACGCCAGCCGCTCCGAGATGAGCCAATCGGATGATGATGAGGATTCTGAGGAGATGGGTGATGC 847400
AGATGAGGATGGAGGCGCAGTCCCGATGATGGAGGGTGGACCGCATCACTCACTCCGACCCCGCTGCTTCAGCGAGCCGAAACGGTCCGGCCG 847300
GTCGAGACGTCACGCCGGACATGGCCCGATCGACTTCTCCACCTGCTGGTCCCGAGCCCGTCTGAGCGCCATCGTGCAGCAGACCAATCGCTATG 847200
CACACAAGTACCAGCTGAAGAACGGCCGGGACGAGGAGTGGCCGAATGTGGACCTCCGCGGAGATGAAGGTATGATTGAAAAGACTCATTAGACTTTAA 847100
ACTGTATCTCTACATTAATGCGTTGTCTGACCCCTGCTCTTCCCTCATGACCTCAATGAAAAGTAGCTACGGGCGGATTGAGCTTTCATGAAAAAAC 847000
AATGGTTCAAACAATTAATGCTCCAGGTTGTGTCCTGATGCTTATTTGTGAGCTGTGCTTTTAAATTAATTTGTCCTCCACTCATATGTTTGGGACC 846900
ATCATGATGTTGTTAATTTTCCATGTTAATCTTTCCCAATTCATTTTAAACTTAGGAAAGCACATGTACCGTAGTTACATGTCAGGAAGTGAAGATAT 846800
TTTGGACAGATGGGGTGAATTTTTTCCGTCACACAGGCAAAATTTCTGTCCAGGACGGAGTGACAGGTCCTGGAGATGGTCTGTCAGCAGCAGATA 846700
TTTAAGGGGGCTGAACGTGAATTAAGAAAGCCACTACACATACAGTAGTACATTATGCTCAGATATAGCCCAAAGTACAAAAATACCACAAATTTAT 846600
TATGGGTGAAATCTTGTCCATTTCCGACGGCGTTCATGGCCATCTTCATGTACACCAGCTGGTGCACGGCGTCTTCCGAGGACATCTCCGCCCTGTGG 846500 ex2
GAGGCTGACGAGATAACGTACAACCCGGTTCATGGCGAGCATCTGACGAGCGCCGCTCCAGCGCATCATGGCTACTTCCAGTGGTGTCTCCAGCG 846400
GCCAGCAGCAGGGCGACAGTCTCGCAAGATCCGCCCTTCCCTCGAGCCGTCGAACTCCTCCCGAATGAGTTCAAGACTGCCAGGTACTCTGATAC 846300
TGCAATTCCTGATTTTTTCAATGATTTTTATGTTGAAATGTTTTCGTTTTTCATGTTGACGACTGTAAACGAGAATTATCATTACTGATAACAATAAT 846200
TTGCTTTGGTGACAGTAAGAAATGATTGTTCCATGAACATTTAGTTTCAAAAAGTTAAATTTCTTTTGTAGCATATCTTATTACCTGGATGTCTAAT 846100
CTTCATTGACGAATAACAAGGTATTTTGCATCTTCGCTTTAGACCCAGTATCCACGAGCCTCCGCTGGAGGAGGAGCCGCTGATGTGGTTCGCGCCA 846000 ex3
CTGGACAGAAACGGCGCAAGAAGAACGGTTCACCTCTGGGTGCGACAGTGCACAGTACAGGGTTCATCTCCAGGTGAGCTGTCTGTGTGTTCCCT
CCATGTTTACAGTGCACAGTACAGGGTTCATCTCCAGGTGAGCTGTCTGTGTGTTTCCCTCCATGTTTATGATGCCATAAACCTATCTAGGGGTGGGTACC 845800
TAAGTACAAAGTAAAATGGTCTCATTCTCTGATTTGGCCGTCCTCATTGGAATATACACTTTCATGCCAAAAATTCATTAAAGCAGTATATCTGGAGA 845700
TGTTTTGCTTTCATCGCAAGCAGCCGTTAATGTTACATGTAGGAAATAAAAAAATGGCCACCTAAATCCAGCGCAATCCAAAATGCTGGAGTGGCTC 845600
CAGGAGCCTGTTTGAAGTGAATCCCCCTAGTAAATAAAAAAATAAACATGTATCTGCTTCCTAGATCCACATCCATGTGAAGGAGAAGAGCGCA 845500 ex4
CAGCCTGAGCCACACGAGAAAGTTACTCCGTCGCCCGCAGCCTCTCTCCGACCTGATTGGCTCGGGACGGAAACCATTTGTCTGTCATCCCGGCC
AACACCAGCCTGCGTCTGTTCCGCGAGCTCGAGTCCCAGCGGTGTACGGCTCGGGGTGTCACGGACGCCAGGCACACCGACTGCGAGGGACTCCCTG 845300
CCAGCGTCTCGCAAGCCTCCGACATCCCTGCGCAAGTATGACAGCATCGTCCGACCAAGGGCAGGTTTGTGTTGTTGTTGTTGTTGTTTACTAG 845200
ATGGTATCCCGCCAGCGTGTGTCGCAAGCCTCCGGCATCCCTGCGCAAGTATGACAGCGTCCGCGCAAGGGCAGGTTGTTGTTGTTGTTGTTGTTG 845100 ex5
TTTACTAGATGGTATCCCAACAGCGTCTCGCAAGCCTCCGGCATCCCTGCGGAAATGACAGCATCGTCCGCAACCAAGGGCAGGTTTGTGTTGTTG
TTTGTGTTGTTTACTTGTGTTAAGATCTCCAGTAAAGTACTCATAAAAAATACATCATGCTCTTAGTAGTCTGCTTGAGACAATTTAACACATGACTCTACA
GCTTACTGAAGCTTCACTATTATATGCAAACTAGCAATAACAACAGTACAGTAAAGTGTACAGAGAGAGCAGAGGCTGAAATGTTAGTTGATGTGACA 844800
CCTGGTAGTTTTTGGCTACCATGTACATAGTGTCTACTTTGTCCATTTCTTGATACCTTGTTCCAATGCAGATATCCGCTGTGTTGTGGAACCTCGCA 844700
CGCAACCCCTCAGGTTTGTCTGTTCAATTTGTTTCAAACAACAAGTTACTAAGGTTAAACCCCTGTAACAACCTGTAATAAAGTACAAGCTGCATATC 844600
CAGTCAGGATTAATTCATCTTCTCACACAGAAATGGACCATTTCGATAAAGTATGGTGGATTCTTTAACATGCTCTAGGTGGCTCTCTTTAAACATG 844500
GGGCTCCAGATTTATGTCCTATCAAAGCTAGGCATCTCGATTTTACCTGACCCAAAGTATGAAAGTAGGTGTAAGTTCATTTAGAAAGGCACGACTTTG 844400
GGTCTGTGGAGTTTAAACCCAGAACAGTATGATCCAAAGTCAATATCAAACATTCGCGGCATCATAGTTCAAGTCAATCAAGTCAAGTCAAGTCAAGT 844300
AATCACACTCTAGTTGCCACCACCTAGGAGCAGGTCAATTAACCCAGCCAGCAAGAGACTGTGTAATACAGGCTAACAATGTCAACAACCTCTACAA 844200
CAAGACCCTTTGCCCCCTCCCCAGGTTGCTGACGAACCGTCTCGCCCGCGGGCAGGGGAAGCTGGTCTGTAAGAAGAACCGGAGGGCTGACGA 844100 ex6
GGAACGGCGGTGCTCGAGGCTGTGAGGCCTCCAGGTGGGTGCTTCACTTTGTTGTTGTTGTTGTTGTTGTTGTTGTTTATTGCATACACAGTAA
AATCCCCTCATGTTGTAACACAATGTGGTCCCAAATGAATATAGGCATGTGATTATGATACCTGTAATAATGCATGTAGATTCCGAGGATTAAATTTCC 843900
GTGGTAGCGGAAAAAGGACTTTTTGCGGTGGTATTAAGTTCGCAGTAGCACCTGAGGCACTGFACTCTTACTGTTACTGCCATGGAAAAATGTTCCG 843800
GGTGGTTTTAAGGTTGCGGTGAAGGACCACCGCAAAAACCGCAACATTAACCACCGCAACGTTTCTGCATTTACAGTATATCTTGACAATAGTGA 843700
TGACAATAACAATGTACATCTACTTCTTAAGAATCTTCTCTGTTTCAGAACCATCAAGTACATCGACAAGTACAACGAGAAGTACTCTGGTA 843600 ex7
CATCATCTCCACAAGCCAGCAAGAGTGGCAGCAGCTGTTCTGGACGCTCCTCAGCTGGCCATGAACAACGCTACATCATGTAACAAGCTGCCCCG
GCGCAGCGCACTCCCGTCTCAGCCGACGACTTCCGGGCGAGAGTGTGCGGCCATGGCCGGGCGTACTACCAGGAGCCCCCAGCAGGGGGAGG 843400
AGTAGCCCCCAGGGGAGGAGTGTGCGTCCATGGCCAGGCGTACTACCAGGAGCCCCCAGGGGGAGGAGTGTGCGGGCCATGGCCA 843300
GGCCCTACTACCAGGAGCCCCCAGCAAGGGGAGGGTGTGCGGCCATGGCTGGGCCGTACTACCAGGAGCCCCCAGGGGGAGGAGTGTGCG 843200
GC6C

```