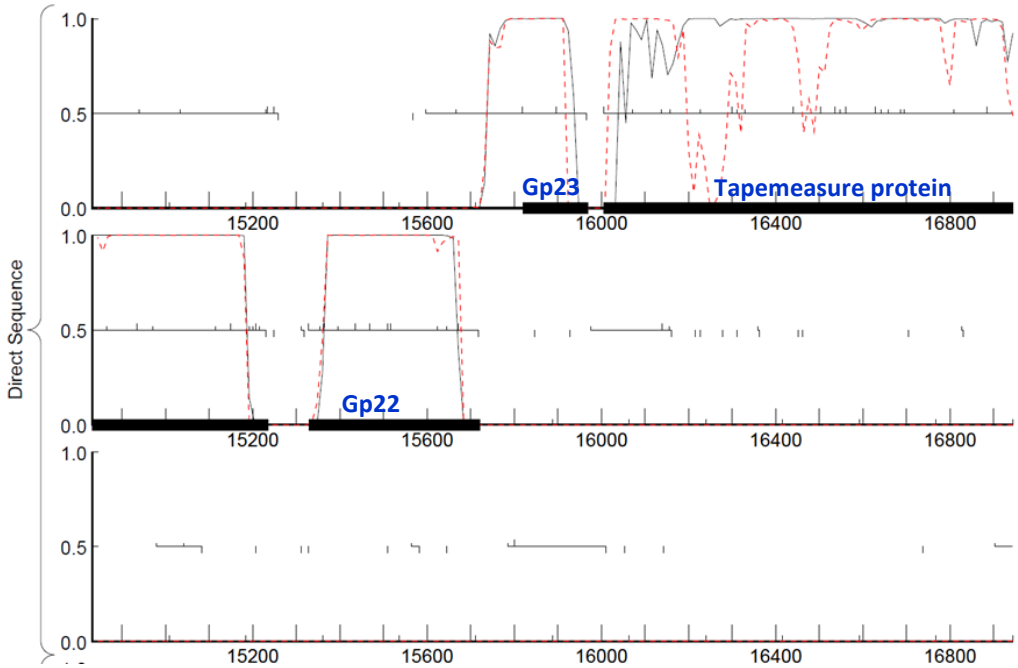
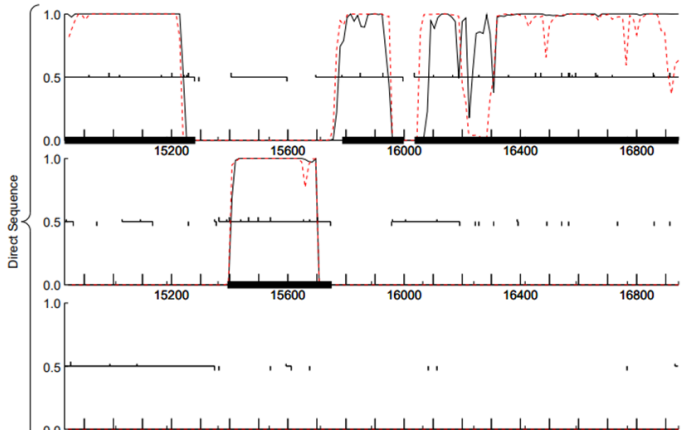
**Region of interest:**
Microdon 15329 - 15969
SEA_MICRODON_22 [15329-15721]
SEA_MICRODON_23 [15820-15969]

**Rationale:**
There is a large, uncalled region of coding potential in the +3 frame upstream of SEA_MICRODON_23 (downstream of SEA_MICRODON_22). These genes are flanked by the major tail protein gene (upstream of SEA_MICRODON_22) and the tapemeasure protein gene (downstream of SEA_MICRODON_23). This is the same syntenic region where tail assembly chaperones involving a programmed frameshift have been routinely called in Actinobacteriophages. Calling the SEA_MICRODON_22 gene on its preceding start codon would generate a substantial overlap (55 bp) with SEA_MICRODON_23.




This same behaviour (non-called coding potential upstream of Gp23) is observed in other members of the BH cluster, and calling the longer version of the SEA_MICRODON_22 homolog results also in substantial overlap.

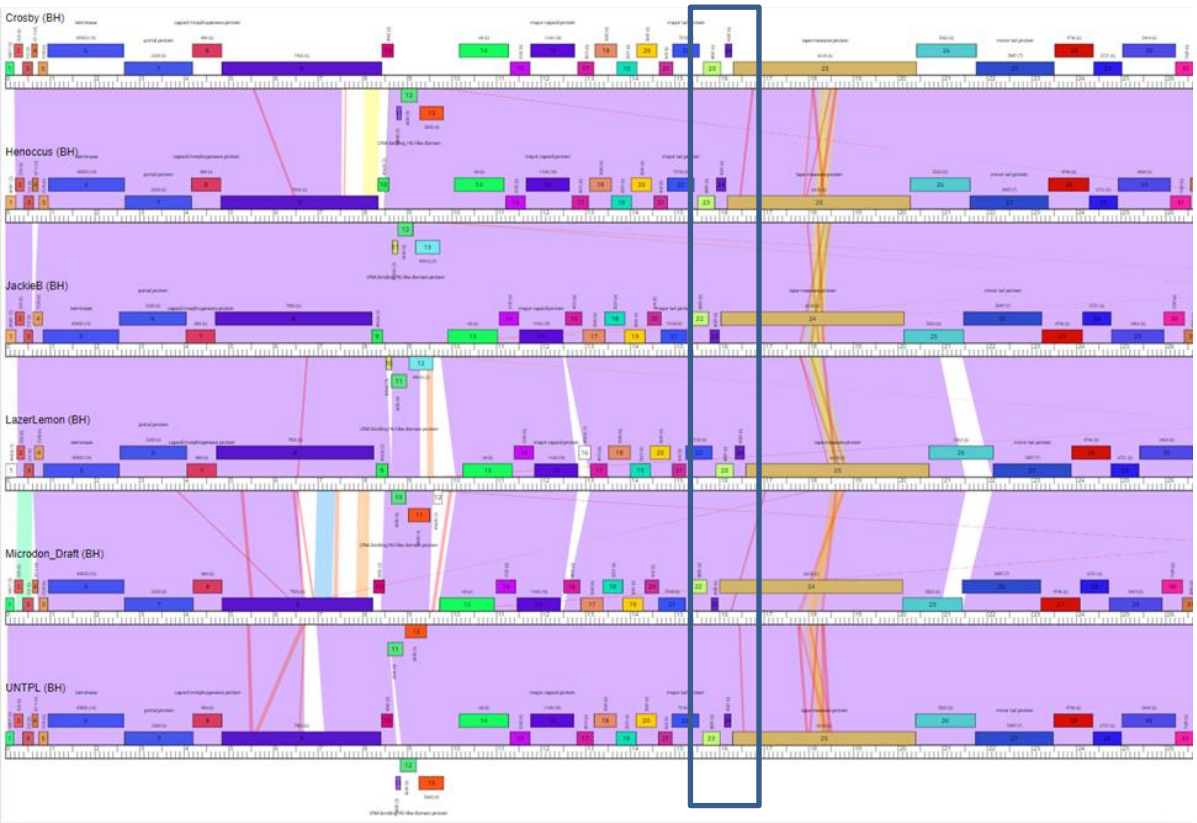


**DNA sequence**

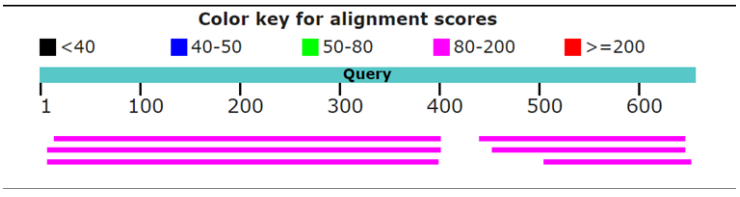
**ATG**AGCGGTTTCGACGCAGGCACGGCCGTGGAGCCGATGGACTACGACTTCACCACCGTCCCCGGCGGCGTGGGCAA
GGGCACCGTTCCCGAGCCCTCCACGAAGGAGATGCAGGTCTTCCAGCGCGAGTTCGCCAAGGTCATGCGCAAGGGCC
AGAAGCTCGAAGTCTCCGACGAGGACGCGATGAAGATGTCCGAGAAGGAGTTCGACAAGCTCCAGGCGGACGCGCAG
GCCATCGGTGAGGAGCTGGACGAGCTGATCGGCACCCTGTGCAAGGGCTCCCCGTCCACCGAGGAAGTGGCGACCCT
GCCCTTCCGCGTGAAGACGGCCTTCTCCAAGTGGCTCATGGAGCAGTTTGCCCCGGAAGGCGGGACCTCCGGTACGA
GGCAG**TAA**CCGGAGGCGAGAACCAGCGAGCGATCTACTACCTGGTACGGCGGAAGCTGGGCTACAGCCCGGCCGAGT
GGGACGACCTCCCATGGCATCACCAGCGG**ATG**TTCCTGGACGGGCTCCACACGGAGCTGAAGGACCCCGACGACGAC
TCGCAGGGGACCGGTCCGGAGACCGACGGCATGGCCAACGAGGACGCGTTCGACAAGGCCCTGAGCGGCGCCAAGAG
GCGTCGCGCGGAGCGCGCCACA**TAG**

**Genetic surroundings**

Gp15 → major capsid protein
Gp21 → major tail protein
Gp24 → tapemeasure protein




Gp23 protein gene homologs have been called at different lengths (longer genes) on similar phages (BH cluster).




BLAST with Microdon gp23

**Translation map**

```
  1   E  R  F  R  R  R  H  G  R  G  A  D  G  L  R  L  H  H  R  P
  1  *  A  V  S  T  Q  A  R  P  W  S  R  W  T  T  T  S  P  P  S
  1 M  S  G  F  D  A  G  T  A  V  E  P  M  D  Y  D  F  T  T  V
  1 ATGAGCGGTTTCGACGCAGGCACGGCCGTGGAGCCGATGGACTACGACTTCACCACCGTC

 21   R  R  R  G  Q  G  H  R  S  R  A  L  H  E  G  D  A  G  L  P
 21  P  A  A  W  A  R  A  P  F  P  S  P  P  R  R  R  C  R  S  S
 21 P  G  G  V  G  K  G  T  V  P  E  P  S  T  K  E  M  Q  V  F
 61 CCCGGCGGCGTGGGCAAGGGCACCGTTCCCGAGCCCTCCACGAAGGAGATGCAGGTCTTC

 41   A  R  V  R  Q  G  H  A  Q  G  P  E  A  R  S  L  R  R  G  R
 41  S  A  S  S  P  R  S  C  A  R  A  R  S  S  K  S  P  T  R  T
 41 Q  R  E  F  A  K  V  M  R  K  G  Q  K  L  E  V  S  D  E  D
121 CAGCGCGAGTTCGCCAAGGTCATGCGCAAGGGCCAGAAGCTCGAAGTCTCCGACGAGGAC

 61   D  E  D  V  R  E  G  V  R  Q  A  P  G  G  R  A  G  H  R  *
 61  R  *  R  C  P  R  R  S  S  T  S  S  R  R  T  R  R  P  S  V
 61 A  M  K  M  S  E  K  E  F  D  K  L  Q  A  D  A  Q  A  I  G
181 GCGATGAAGATGTCCGAGAAGGAGTTCGACAAGCTCCAGGCGGACGCGCAGGCCATCGGT

 81   G  A  G  R  A  D  R  H  P  V  Q  G  L  P  V  H  R  G  S  G
 81  R  S  W  T  S  *  S  A  P  C  A  R  A  P  R  P  P  R  K  W
 81 E  E  L  D  E  L  I  G  T  L  C  K  G  S  P  S  T  E  E  V
241 GAGGAGCTGGACGAGCTGATCGGCACCCTGTGCAAGGGCTCCCCGTCCACCGAGGAAGTG

101   D  P  A  L  P  R  E  D  G  L  L  Q  V  A  H  G  A  V  C  P
101  R  P  C  P  S  A  *  R  R  P  S  P  S  G  S  W  S  S  L  P
101 A  T  L  P  F  R  V  K  T  A  F  S  K  W  L  M  E  Q  F  A
301 GCGACCCTGCCCTTCCGCGTGAAGACGGCCTTCTCCAAGTGGCTCATGGAGCAGTTTGCC

121   G  R  R  D  L  R  Y  E  A  V  T  G  G  E  N  Q  R  A  I  Y
121  R  K  A  G  P  P  V  R  G  S  N  R  R  R  E  P  A  S  D  L
121 P  E  G  G  T  S  G  T  R  Q  *  P  E  A  R  T  S  E  R  S
361 CCGGAAGGCGGGACCTCCGGTACGAGGCAGTAACCGGAGGCGAGAACCAGCGAGCGATCT

141   Y  L  V  R  R  K  L  G  Y  S  P  A  E  W  D  D  L  P  W  H
141  L  P  G  T  A  E  A  G  L  Q  P  G  R  V  G  R  P  P  M  A
141 T  T  W  Y  G  G  S  W  A  T  A  R  P  S  G  T  T  S  H  G
421 ACTACCTGGTACGGCGGAAGCTGGGCTACAGCCCGGCCGAGTGGGACGACCTCCCATGGC

161   H  Q  R  M  F  L  D  G  L  H  T  E  L  K  D  P  D  D  D  S
161  S  P  A  D  V  P  G  R  A  P  H  G  A  E  G  P  R  R  R  L
161 I  T  S  G  C  S  W  T  G  S  T  R  S  *  R  T  P  T  T  T
481 ATCACCAGCGGATGTTCCTGGACGGGCTCCACACGGAGCTGAAGGACCCCGACGACGACT

181   Q  G  T  G  P  E  T  D  G  M  A  N  E  D  A  F  D  K  A  L
181  A  G  D  R  S  G  D  R  R  H  G  Q  R  G  R  V  R  Q  G  P
181 R  R  G  P  V  R  R  P  T  A  W  P  T  R  T  R  S  T  R  P
541 CGCAGGGGACCGGTCCGGAGACCGACGGCATGGCCAACGAGGACGCGTTCGACAAGGCCC

201   S  G  A  K  R  R  R  A  E  R  A  T  *
201  E  R  R  Q  E  A  S  R  G  A  R  H  I
201 *  A  A  P  R  G  V  A  R  S  A  P  H
601 TGAGCGGCGCCAAGAGGCGTCGCGCGGAGCGCGCCACATAG
```

A M K M S E K E F
→ called gp22 (SEA_MICRODON_22), on frame +1

Q G T G P E T D G M
→ called gp23 (SEA_MICRODON_23), on frame +3

G R R D L R Y E A V T
→ potential region of frameshift, on frame +3 (overlaps gp22)

Y L V R R K L G Y S P A E W D D L P W H H Q R M F L D
→ HHPred hit (P=87.52) to KOG0722 Molecular chaperone (DnaJ superfamily)

**Molecular chaperone hit**

The region upstream of SEA_MICRODON_23 contains a valid 82 amino acid (246 bp) protein sequence in the +3 frame, extending up to 48 amino acids (144 bp) into the coding sequence of SEA_MICRODON_22 (frame +1). A segment of this SEA_MICRODON_23 upstream sequence, before the overlap with SEA_MICRODON_22, maps to a 'molecular chaperone DnaJ' domain using HHpred (P=87.52). Refinement of the HHpred search using MSA of this conserved stretch of sequence from all 6 BH phages returns a 'molecular chaperone DnaJ' hit with P=90.7.

```
HHPred hit (P=90.7) to KOG0722 Molecular chaperone (DnaJ superfamily) using MSA

>Crosby
LVRRKLGYSPAEWDDLPWHHQRMFLD
>Henoccus
LVRRKLHYSPADWDALPWHHQRMFLE
>JackieB
LVRRKLHYSPADWDALPWHHQRMFLE
>LazerLemon
LVRRKLGYGPDDWDALPWHHQRMFLD
>Microdon
LVRRKLGYSPAEWDDLPWHHQRMFLD
>UNTPL
LVRRKLGYSPAEWDDLPWHHQRMFLD
```

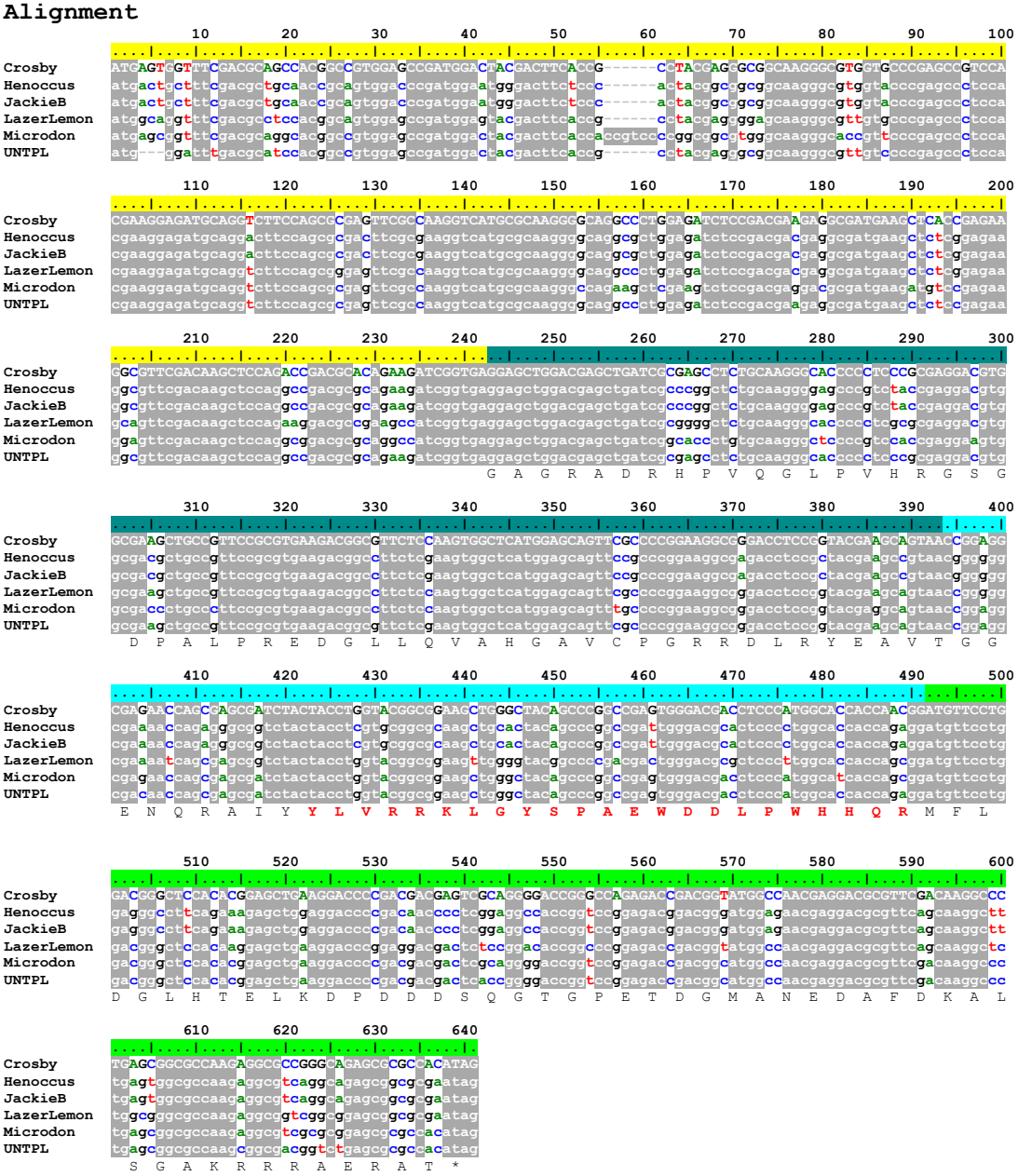
Template alignment

**KOG0722 Molecular chaperone (DnaJ superfamily) [Posttranslational modification, protein turnover, chaperones]**

Probability: 90.7 E-value: 0.08 Score: 32.87 Aligned Cols: 26 Identities: 31% Similarity: 0.426

```
Q ss_pred                          CCccccCCChHhhhhCccHhhhhccC
Q Seq0(translate          1        LVRRKLGYSPAEWDDLPWHHQRMFLD  26 (26)
Q Consensus               1        lvrrkl~y~p~dwd~lpwhhqrmfld  26 (26)
                                   |.|+.||+|...|++|+-+.+-+|++
T Consensus             255        l~r~~l~~s~~~~~~~~~~~~~~~~~  280 (329)
T KOG0722               255        LIRRYLGVSQMEFDQKTDEDIDDLFE  280 (329)
T ss_pred                          HHHHHhCCCHHHHhcCCHHHHHHHHH
```

## Alignment




A MSA of the DNA encompassing SEA_MICRODON_22 and SEA_MICRODON_22 with DNA from other BH cluster phages shows no substantial evidence of more non-synonymous substitutions in the putative frameshifted region (including that encoding the putative chaperone domain) than in the 'proper' gp23 region. This suggests that the region may be actively translated.

## Conclusion

Multiple lines of evidence (synteny, HHpred, GeneMark coding potential, MSAs) point towards warranting a larger gene call for SEA_MICRODON_22 and its homologs. The fact that the extended SEA_MICRODON_22 region matches a chaperone domain, together with the syntenic location of these genes, strongly suggests that these are tail chaperones. Multiple attempts to locate possible slippage sequences have been performed using known

references (PMC5009743, PMC3711429) have not been successful, but given the evidence we suggest that the longer version of the SEA_MICRODON_22 should be called, and that these genes be annotated as tail chaperones.