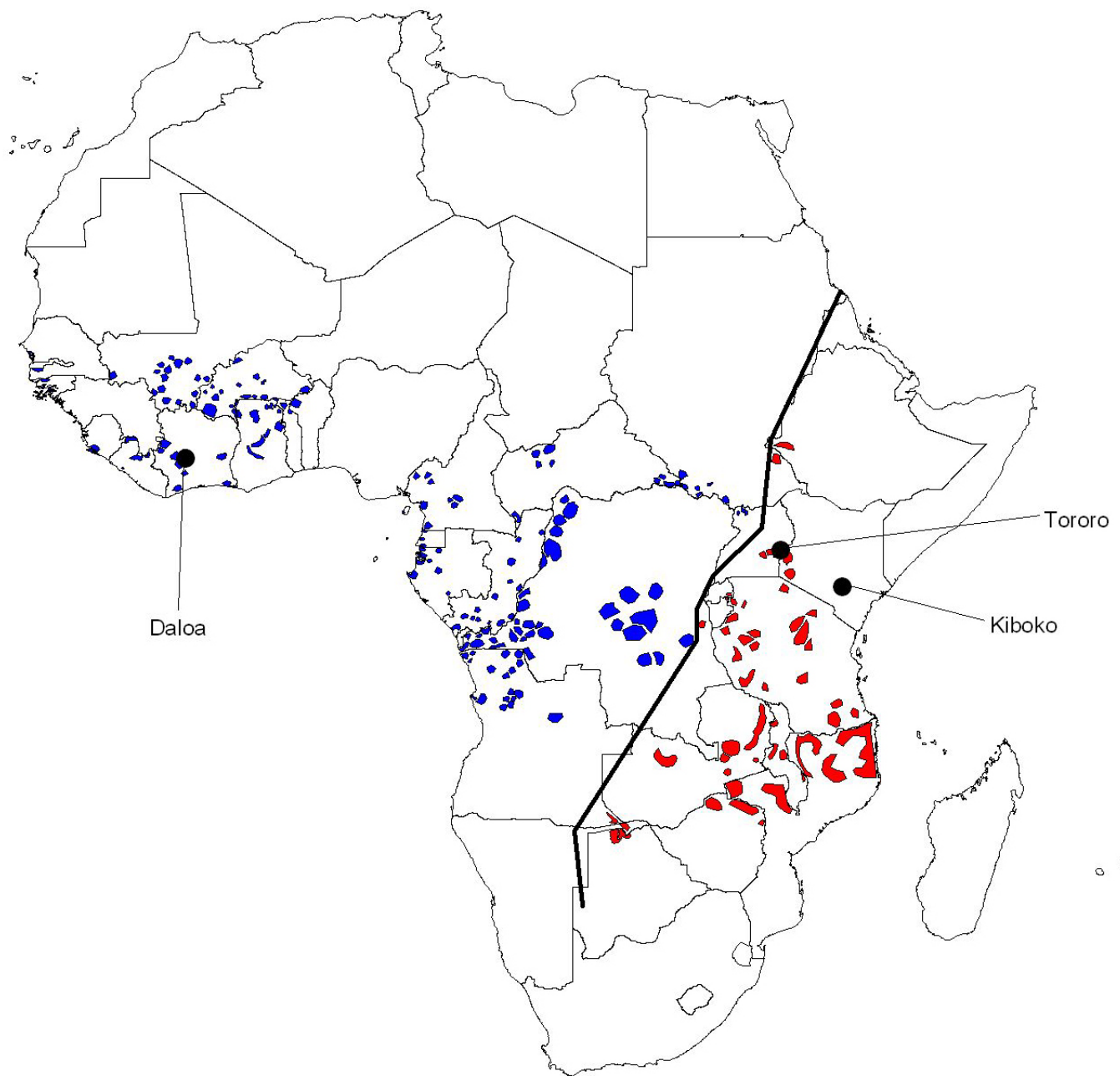


Additional file 1

1. Map showing location where isolates were collected.
2. A comparison of VSGs AnTat1.1b and AnTat 1.10
3. The following data are presented for each VSG :
 - a) Amino acid sequence with cysteine residues in red
 - b) Domain combination
 - c) N-terminal domain sequences of homologues identified using the whole VSG sequence to screen the *T. b. brucei* TREU927 and *T. b. gambiense* Daloa genomic sequences. The N-terminal domain was standardised by ending 5 residues before first cysteine in the C-terminal domain
 - d) Percentage identities measured using NCBI blast 2 sequences (<http://www.ncbi.nlm.nih.gov/blast/bl2seq/wblast2.cgi>)
 - e) Thre way alignment for the six VSGs used to determine the location of sequence divergence.

The screen was performed in 2006



Map of Africa showing the geographic origins of the isolates analysed in this paper. The expressed VSGs were characterised from field isolates from Tororo in Uganda; the *Trypanosoma brucei brucei* TREU927 genome reference strain came from Kiboko in Kenya and the *T. b. gambiense* genome strain came from Daloa in Cote d'Ivoire. Areas of endemic *T. b. rhodesiense* human trypanosomiasis are shown in red and *T. b. gambiense* human trypanosomiasis in blue. The range of *T. b. brucei* is far more extensive.

Location of differences between AnTat1.1 and AnTat1.10 in primary and tertiary structure of N-terminal domains



```
>Ant1.10
QTAGRPLADAVGKALCTYSKTAKRQANLAQALDRGITAAKKSQ
QAHELATIALAKLPHYREAAATILYAKNKRAEAEANIENWKGQ
KTKLVGEAMYSSGRIDELMLMLEGHRDQGSAGQTKTCLGAAGNG
NTVDEFVKTECDTEQDHNINADESIEQAARTLSQENRDPEAGG
GTNCKITGNLASDYDHPNDLSLLGGLLTIHNGGGFKATTTIKT
AAAANKLISALASKVNDIAANLKAHTESAPTTKQELKTLAAGK
ARSKLAAANDEYTSWEAGKPVNFDEHIKKVFGAEDGKDSAYAL
ALEGISIEVPQKPGTTESKQLYSMQPKDLMAALIGTIAEIQKAA
ATKAPCPKHKLTSA
```

```
>Ant1.1b
QTAGRPLADVVGKTLCTYSKTAKRQANLAQTLQRASSAAKQSR
QAQQLAALALAKLPDYKEAAATLLIYATHKIQDAQASIENWTGE
NTKLVGQAMYSSGRIDELMLLLEGHREDGANGQDKTCLGAAAGG
NTVNEFVKTECDTESGHNIEADNSNIGATDYDSHANELPLLGG
LTIHNAGGFKTGQSLQTAAPTNNKLSALKKNGAGVAAKLATVTA
ARPTNKQEFKTLASKAERAKLQAANEEYNNWPKGAKPADFDAH
IKKVFGAEDGKDSAYALALEGISIEVPQKPGTTESKQLYSMQPK
DLMAALIGTIAEIQKAAATKAPCPKHKLTSA
```

Score = 437 bits (1125), Expect = 4e-121
Identities = 254/366 (69%), Positives = 286/366 (78%), Gaps = 27/366 (7%)

Ant1.10	1	QTAGRPLADAVGKALCTYSKTAKRQANLAQALDRGITAAKKSQAHELATIALAKLPHY	60
		QTAGRPLAD V G K L C T Y S K T A K R Q A N L A Q L R + A A K + S + Q A + L A + A L A K L P Y	
Ant1.1b	1	QTAGRPLADVVGKTLCTYSKTAKRQANLAQTLQRASSAAKQSRQAQQLAALALAKLPDY	60
Ant1.10	61	REAAATILYAKNKRAEAEANIENWKGQKTKLVGEAMYSSGRIDELMLMLEGHRDQGSAG	120
		+EAAAT+LIYA +K +A+A+IENW G+ TKLVG+AMYSSGRIDELML+LEGHR+ + G	
Ant1.1b	61	KEAAATLLIYATHKIQDAQASIENWTGENTKLVGQAMYSSGRIDELMLLLEGHREDGANG	120
Ant1.10	121	QTKTCLGAAGNGNTVDEFVKTECDTEQDHNINADESIEQAARTLSQENRDPEAGGGTNC	180
		Q K T C L G A A G N T V + E F V K T E C D T E H N I A D S + I	
Ant1.1b	121	QDKTCLGAAAGGNTVNEFVKTECDTESGHNIEADNSNIG-----	159
Ant1.10	181	KITGNLASDYDHPNDLSLLGGLLTIHNGGGFKATTTIKTAAAANKLISALASKVNDIAA	240
		A+DYDSH N+L LLGGLLTIHN GGFK +++TAA NKLISAL +K +AA	
Ant1.1b	160	-----ATDYDSHANELPLLGGLLTIHNAGGFKTGQSLQTAAPTNNKLSALKKNGAGVAA	213
Ant1.10	241	NLKAHTESAPTTKQELKTLAAGKARSKLAANDEYTSWEAGKPVNFDEHIKKVFGAED	300
		L T + P T K Q E K T L L A + K R + K L A A N + E Y + W + G K P + F D H I K K V F G A E D	
Ant1.1b	214	KLATVTAARPTNKQEFKTLASKAERAKLQAANEEYNNWPKGAKPADFDAHIKKVFGAED	273
Ant1.10	301	GKDSAYALALEGISIEVPQKPGTTESKQLYSMQPKDLMAALIGTIAEIQKAAATKAPCPK	360
		GKDSAYALALEGISIEVPQKPGTTESKQLYSMQPKDLMAALIGTIAEIQKAAATKAPCPK	
Ant1.1b	274	GKDSAYALALEGISIEVPQKPGTTESKQLYSMQPKDLMAALIGTIAEIQKAAATKAPCPK	333
Ant1.10	361	HKLTSA 366	
		HKLTSA	
Ant1.1b	334	HKLTSA 339	

Ako1

```
MPTVASMQHHPKIKASNIFFRYLTLALTLTLLVALSLTRKGESAAAGD
GLNAAEFRALCHLAAIADLSADATTLPTIDPNEKQEIEMLNMSASQATWQAAFPAPDQPE
PDVQQAACNTGDKRIHCLKDYKFKQLKATLKTESAAEQTRGKHKLVPAIQETPEGRRVQR
QLHDLATEAADIISTYNSQHSQKQKLGALT'TVLKQALYGTGTAEKTGKYGDKWTATAT
RTTDCAQNMAGTSLRGDLAACLICINDSATNKQMCGN'TVGP'DNSNNWQASATTTNINGLVSK
CKTLVKPKLSAQ'TIRAALAI'FDAKLR'SHTGGGKDAI'VLGTPHTNGDCGGQANVACVDYTO
QLTPTNEGEQNKINWYQHLDQAASTLQELETAEIDRQHALRQLGDLKKRAYSLYNTLTKIA
NSPHTTKGTEIHAKEQNTKESKCKLKN'TTAAECPETDCDYDKTKNECKSKPGSETTATPT
GAGAAGT'TTEKCKDKKKDDCKDGC'KWEGETCKDSSILATKKFALS'VVSAAFVALLF
```

Domain combination: B1

GeneDB Blast results N-terminal domains

>Ako1

```
MPTVASMQHHPKIKASNIFFRYLTLALTLTLLVALSLTRKGESAAAGD
GLNAAEFRALCHLAAIADLSADATTLPTIDPNEKQEIEMLNMSASQATWQAAFPAPDQPE
PDVQQAACNTGDKRIHCLKDYKFKQLKATLKTESAAEQTRGKHKLVPAIQETPEGRRVQR
QLHDLATEAADIISTYNSQHSQKQKLGALT'TVLKQALYGTGTAEKTGKYGDKWTATAT
RTTDCAQNMAGTSLRGDLAACLICINDSATNKQMCGN'TVGP'DNSNNWQASATTTNINGLVSK
CKTLVKPKLSAQ'TIRAALAI'FDAKLR'SHTGGGKDAI'VLGTPHTNGDCGGQANVACVDYTO
QLTPTNEGEQNKINWYQHLDQAASTLQELETAEIDRQHALRQLGDLKKRAYSLYNTLTKIA
NSPHTTKGTEIHAKEQN
```

>gamb276c12

```
MPTVASMQHHPKIKASNIFFRYLTLALTLTLLVAISLTRKGESAAAGDGLNAAEFRALCHLAA
IADLSADG'TTLPAINPNEKQEI'EILNMSASQATWQAAFPDPDQPEPDVQACNTGDKRIH
CLKDYKFKQLKATLKTESAAEQTRGKHKLVPAIQETPEGRR'IQQLNDLAAEAADI'ISS
YESQHAQLKQKLGAL'TSVLKQALYGT'RTTQTGSTYADKWTATGSRTADCEAQKAGLSLR
GDLACL'CIDST'TRQMCNGIGITDSSSEWQTSASKINIDSLISKCKEITKPKLRAATIR
TALAI'FDAKLSHTQGSKDAI'VLGTPHTNGNCGQOASVACVDYTOQLTPTNEGEQNKINW
YQHLDQAASTLQELETAEIERQHALRQLGALKKRAYSLYNTLKIANS'PHTTKGTEIHAKE
QN
```

>contig11608 adjusted manually to maintain VSG reading frame
AEFRALCRLPPIDLLANATALPTIDPNEKQEI'EILNMSASQATWQAAFPAPDQADPEPE
PACKSEDQVVRCLDYQTFKQLKESLKSESA'AEKTGDKHKLVPAIQET
PEGRRVQRQLHDLAAESA'EVISSYDSQHSQKQKLGAA'LTSILKEALYAGATTATGENYD
DKWTAT'TTRAADCAAPQAGKSIRGDLACL'INDQ'TTSKQMC'NAVGT'DGGNDWQASATAT
NIDKLVGKCKTLVKPKL'TTRAIHKALSDFDAKLSHTGGAKDAI'VLGTPHTDGTGCGSTAN
KACVDYTEQLTPTTEGAQNKISWYQHLEHAART
LQELQTAQIERQDALRRIT'LKKRAYSLYNTLKIANS'PYTTKGTEIHAKEQN
Short at N-terminus due to end of contig – couldn't extend

NCBI blast 2 sequence scores for N-terminal domains excluding N-terminal signal sequences

Tororo v Tbg: 84%

Tororo v Tbb 927: 73% over available sequence

Tbg v Tbb927: 71% over available sequence

CLUSTAL W (1.83) multiple sequence alignment

```

Ako1      MPTVASMQHKKPIKASNIFFRYLTLALTLVLSLTRKGESAAAGDGLNAAEFRALCHLAA 60
TbgDal    MPTVASMQHKKPIKASNIFFRYLTLALTLVAISLTRKGESAAAGDGLNAAEFRALCHLAA 60
Tbb927    -----AEFRALCRLPP 11
                                         *****:*.

Ako1      IADLSADATTLPTIDPNEKQEIEMLNMSASQATWQAAFPAPDQPEPDVQOACNTGDKRIH 120
TbgDal    IADLSADGTTLPAINPNEKQEIIEILNMSASQATWQAAFPDPDQPEPDVQOACNTGDKRIH 120
Tbb927    IPDLLANATALPTIDPNEKQEIIEILNMSASQATWQAAFPAPDQADPEPEPACKSEDQVVR 71
*.** *:.*:***:****:*****:*****:***** *****:***: :***: * : **

Ako1      CLKDYKFKQKATLKTESAAEQTRGKHKLVPAIQETPEGRRVQRLHDLATEAADIIST 180
TbgDal    CLKDYKFKQKATLKTESAAEQTRGKHKLVPAIQETPEGRRIQKQLNDLAAEAADIISS 180
Tbb927    CLKDYQTFKQLKESLSESAEKTGDKHKLVPAIQETPEGRRVQRLHDLAAESAIEVISS 131
*****:.* ***** :***:*****:* .*****:*****:*****:*****:***:

Ako1      YNSQHSQKQKLGALTTLVTKQALYGTGTAECTGKYGDKWTATATRTTDCQNMAAGTSLR 240
TbgDal    YESQHAQLKQKLGALTSVLKQALYGTRTTQTGSTYADKWTATGSRTADCEAQKAGLSLR 240
Tbb927    YDSQHSQKQKLGAAALTSILKEALYGAGTTATGENYDDKWTATTTTRAADCAAPQAGKSIR 191
*:.***:***** *****:***:*****: * . . * ***** :***:*** ** *:*

Ako1      GDLACLINDSATNKQMCNGTVGPDNSNNWQASATTTNINGLVSKCKTLVKPKLSAQ TIR 300
TbgDal    GDLACLISDSTTRKQMCNGIGTDSSSEWQTSASKINIDSLISKKEITKPKLRAATIR 300
Tbb927    GDLACLINDQTTSKQMCNGAVGTDGGNDWQASATATNIDKLVGKCKTLVKPKLTTTRAIH 251
*****.*.:.** ***** :*. . . .:***:***: ** :*:.*** :.***** : :**

Ako1      AALAI FAKLRSHTGGGKDAIVLGPHTNGDCGGQANVACVDYDQQLTPTNEGEQNKINW 360
TbgDal    TALAI FAKLKSHTQGSKDAIVLGPHTNGNCGQASVACVDYDQQLTPTNEGEQNKINW 360
Tbb927    KALSDFDAKLKSHTGGAKDAIVLGPHTDGTGCGSTANKACVDYTEQLTPTTEGAQNKISW 311
** : *****:*** *.*****:*** ** *. *****:*****.* *****.*

Ako1      YQHLDQAASTLQELETAIEIDRQHALRQLGDLKKRAYS LYNTLKIANS PHTTKGTEIHAKE 420
TbgDal    YQHLDQAASTLQELETAIEIERQHALRQLGALKKRAYS LYNTLKIANS PHTTKGTEIHAKE 420
Tbb927    YQHLEHAARTLQELQTAQIERQDALRRITTLKKRAYS LYNTLKIANS PYTTKGTEIHAKE 371
*****:*** *****:***:***:***.***: : *****:*****:*****

Ako1      QN 422
TbgDal    QN 422
Tbb927    QN 373
**

```

Ako 2

```
MSAVATGFILAI AALKQAGAVSDGENAAIFKPLCAALQLAD
VKPTFEPPIQPKMPEPELELYRLNMSIAPKDWRAKFLKQSTAPPTAVDLPOGEKDAERQTR
WKTWTD TAVFLATGSNEKDLKSNYGLTTATDAQIEAIRATIHDIVETAHAIYTATATANE
TAPADDALLQKEIAQAVYGEETYGPAELTATKVLSGADAGYTAA CGGGGTHKPQNTVAGT
IICVCGCAEGQSHKPCCHKQATETAWQATNIPNRASWGNLRNVCPKATKTKITAHGLREL
VAAAKK IANTQGNVFIGSEGENTCNGNANGACVKITGGVASGNLKDDAIPWINKLEAVA
EQLTERGNYNSEQQRKTA AINELISRAKAVAKRAHYLFQFQKTAQAAAAGEATEGTVSKA
DCKQYTTNSTCPKNDCKWDSTTEDKGDHCKPKPGSEPTTAGTGDGDAKTDSKCSDKKKEE
DCKDGC KWDGKECKDSSILV NKQFAFSVVSAAF MALLF
```

Domain combination: B1

GeneDB Blast results N-terminal domains

>ako2

```
MSAVATGFILAI AALKQAGAVSDGENAAIFKPLCAALQLAD
VKPTFEPPIQPKMPEPELELYRLNMSIAPKDWRAKFLKQSTAPPTAVDLPOGEKDAERQTR
WKTWTD TAVFLATGSNEKDLKSNYGLTTATDAQIEAIRATIHDIVETAHAIYTATATANE
TAPADDALLQKEIAQAVYGEETYGPAELTATKVLSGADAGYTAA CGGGGTHKPQNTVAGT
IICVCGCAEGQSHKPCCHKQATETAWQATNIPNRASWGNLRNVCPKATKTKITAHGLREL
VAAAKK IANTQGNVFIGSEGENTCNGNANGACVKITGGVASGNLKDDAIPWINKLEAVA
EQLTERGNYNSEQQRKTA AINELISRAKAVAKRAHYLFQFQKTAQAAAAGEATEG
```

>Tb11.24.0011

```
MSAVAPGVILAI AALKQARGVTDGDNA AIFKPLCAALQLADV KPTFEPPIQANMPE
PLDLYRLNMSIAPKDWRAKFLNQGNKAASTPAEVPTDENDEELKARWKTWADTAVFLATD
NNEKDLKANYGLTATAGQIEAIRPTIHDITEAARAIYTADSDPGPAPDADELLQKEIAQ
AVYQGEQWGPEELTANKVLSGPD TGYTACGGGGTQKPQNTVAGT IICVCGTGNTQSKPK
CHKKQTTQTAWQATNIPNRASWGNLRNVCPKATKTKITAHGLREL VAAAKTIANTQGNV
FIGSEGENTCNGNAAGACVKITGGAPSGNLKDDAIPWINKLEAVADKLAARDNYNSEQRR
KKAADIADLISRAKAVTKRAHYLFQFQKTAKTAAAGEATEG
```

>Tbg843f08.plk

```
MSAVATGVILAI AAL TQARGVTDGDNA AIFKPLCATLQLADAKSTFEPPIQPKMPEPLDL
YRLNMSIAPKNWRRAKFLKQYNKATPTPAAADLPQGEKDAETTARWKTWADTAVFLATDNN
EKDLQANYGLTATAGQIEAIRPTIHDITEAARAIYTADSDPETTPEDDELLQK*IAEA
VYQGEQWGPEELTANKVLSGPD TGYTACGGSGTQKPQNTVAGT IICVCGCANTQTHKPC
HKKQTAETA WQSNIPDRSSWGNLR TICPKATKTKITAHGLREL VAAAKTIANTQGNV
FIGSEGENTCNGNANGACVKITGGAASGNLKDDAIPWINKLEAVADKLAARDNYNSEQRR
AAAIADLISRAKAVTKRAHYLFQFQKTAKTAAAGEATEG
```

NCBI blast 2 sequence scores for N-terminal domains excluding N-terminal signal sequences

Tororo v Tbg: 79%

Tororo v Tbb 927: 79%

Tbg v Tbb927: 87%

CLUSTAL W (1.83) multiple sequence alignment

```

Tbb927      MSVAVPGVILAI AALKQARGVTDGDNA AIFKPLCAALQLADV KPTFEPP IQANMPEPLDL 60
TbgDal      MSAVATGVILAI AAL TQARGVTDGDNA AIFKPLCATLQLADAKSTFEPP IQPKMPEPLDL 60
ako2        MSAVATGFILAI AALKQAGAVSDGENA AIFKPLCAALQLADV KPTFEPP IQPKMPEPLEL 60
          *****.*.*****.** .*:*:*****:*****.*.*****.:*****:*

Tbb927      YRLNMSIAPKDWRAKFLNQGNKAASTP--AEVPTDENDEELKARWKTWADTAVFLATDNN 118
TbgDal      YRLNMSIAPKNWRAKFLKQYNKATPTPAAADLPQGEKDAETTARWKTWADTAVFLATDNN 120
ako2        YRLNMSIAPKDWRAKFLKQS---TAPPTAVDLPQGEKDAERQTRWKTWTD TAVFLATGSN 117
          *****:*****:*      :..*  .::* .:* * :*****:*****..*

Tbb927      EKDLKANYGLATATAGQIEAIRPTIHDITEAARAIYTADSDP-GPAPDADELLOKEIAQA 177
TbgDal      EKDLQANYGLTTATAGQIEAIRPTIHDITEAARAIYTADSDP-ETTPEDDELLOK-IAEA 178
ako2        EKDLKSNYGLTTATDAQIEAIRATIHDIVETAHAIYTATATANETAPADDALLOKEIAQA 177
          ****:;****:*** .*****.*****.*:*****: . .:* * **** **:*

Tbb927      VYGEQEQWGPEELTANKVLSGPD TG YTTACGGGGTQKPQNTVAGTII CVCGTGNTQSKKPC 237
TbgDal      VYGEQEQWGPEELTANKVLSGPD TG YTTACGGSGTQKPQNTVAGTII CVCGCANTQTHKPC 238
ako2        VYGEETYGPAELTATKVLSGADAGYTAACGGGGTHKPQNTVAGTII CVCGCAEQSHKPC 237
          ***:* :** *****.*****.*:***:***.**:*****:***** .: *::***

Tbb927      HKKQTQTAWQATNIPNRASWGNLRNVC PKATKTKITAHGLRELVA AAKTIANTQGNNVF 297
TbgDal      HKKQTAETA WQGSNIPDRSSWGNLR TICPKATKTKITAHGLRELVA AAKTIANTQGNDVF 298
ako2        HKKQATETA WQATNIPNRASWGNLRNVC PKATKTKITAHGLRELVA AAKKIANTQGNDVF 297
          ****:;:****. :***:*:*****. :*****:*****.*****:***

Tbb927      IGSEGENTCNGNAAGACVKITGGAPSGNLKDDAIPWINKLEAVADKLAARDNYNSEQRRK 357
TbgDal      IGSEGENTCNGNANGACVKITGGAASGNL KDDAIPWINKLEAVADKLAARDNYNSEQRRK 358
ako2        IGSEGENTCNGNANGACVKITGGVASGNL KDDAIPWINKLEAVAEQLTERGNYNSEQQRK 357
          ***** ***** .*****:*****:;: * .*****:***

Tbb927      KAAIADLISRKAVTKRAHYLFQFQKTAKTAAAGEATEG 396
TbgDal      AAAIADLISRKAVTKRAHYLFQFQKTAKTAAAGEATEG 397
ako2        TAAINELISRKAVAKRAHYLFQFQKTAQAAAAGEATEG 396
          *** :*****:*****:*****:*****:*****
    
```

Bug1

>Bug1

```
MQALRRTLTISFVAALAWPQQASGSGNALKTAAWKEMCQTAKEL
RAVPTAAANKLSTRQTIAELEATALRLAAVAASQADVGDAATLGALAKELSTEASL
SSGYTAKTNKALTAASAAYVGGISEFLNLLTAGITGATNGCLAKQDGATLIGSRDE
LTDCPIDKELTVEETSSTIKTTFADGITTLKTDNVKDTGSNAALCVLFQTPDAPSK
LFHKAAEFLAAGGTLVNTATTTIKTEHGKKISKETNKVGTALIKQAHTQTHELLTEP
ENDPTSASKSIKALASDTNFTATVVRQLKLIGMAGSDKTLEDYAKSKIKTLLDDGKA
NFDTKWNALIDTKIYDGKSEKVKTEKASSITSTKELIQSVLYYQIDLKTKLKRKVEEL
EQKKNDCKEKTEFEVCNAIEGEQEENNTKDCHYDSKKDGKKCTLKKEVKEKLEKTNQET
GVKDGKNDRCVKHGTDKTKCEAENTAGQTPVCGFRKGDNEPEQDKEMCRNGSFLVNN
KFALSMVSAAFVALLF
```

Domain combination: A1

GeneDB Blast results N-terminal domains

Tbg:

Contig assembled manually based on:

>gamb1660g03.q1krevcomp

>gamb187e12.p1k

>gamb1588a05.p1k

>Bug1

```
MQALRRTLTISFVAALAWPQQASGSGNALKTAAWKEMCQTAKEL
RAVPTAAANKLSTRQTIAELEATALRLAAVAASQADVGDAATLGALAKELSTEASL
SSGYTAKTNKALTAASAAYVGGISEFLNLLTAGITGATNGCLAKQDGATLIGSRDE
LTDCPIDKELTVEETSSTIKTTFADGITTLKTDNVKDTGSNAALCVLFQTPDAPSK
LFHKAAEFLAAGGTLVNTATTTIKTEHGKKISKETNKVGTALIKQAHTQTHELLTEP
ENDPTSASKSIKALASDTNFTATVVRQLKLIGMAGSDKTLEDYAKSKIKTLLDDGKA
NFDTKWNALIDTKIYDGKSEKVKTEKASSITSTKELIQSVLYYQIDLKTKLKRKVEEL
E
```

>Tbb927chr5

```
MHI IARI IAF TYVAVLMCTQRSATTGAALKKAAWNKMCATEKELRTPATAEQKLTEAOG
TIEALENPALKMAAVAGAHADAGEALILAALAE LRSEAQT KLLNLKELRKTAL EAAVA
AAVYGGVSEFFTLMTAGQSGTINGCLADTDGSNVI SGKSGLGDCPIDQAI SLTDTSETLT
TTFTENGISSLKAADV KDDSTNANKCVL FQNAQAGNDR LFQHGAAFLAAGGTL SVSTAAT
TLKTGHGQNIKANSTATGAALIKQAHSATLELLKHATTEAEHV GKNKAKELASDSHFAAT
VVKKLKLIGMKGSDSTLEAYAKSIKSL LADGGENFGSKWQTLIDTPIYDGKSEEVKTEK
ASSITSTKELIQSVLYYQIDLKTKFKRAKEELE
```

>Tbg

```
MSRAKATTSFIFA AVLVC SQLGAATGNALKAAAWQDMCTTAE
QLTKIPALATKKLDQAQAVMADYRDAAMRLTAVAATRADSGDAIVLAILGSKLLEAAEET
AGNLKALRPLSIKAAAAAGTVYGGISDFLNL LSAHTGITCGCLNAKGDG DIMADTTELA
SCPIDNEIDLEKDTTGLETVVGDGKFKTLKTTDVKATSTSN TKCVLFQDASAVGNNKLFQ
KAATFLIAGGTLKVNSAGPTMTTAHGAALAEQSTAGGKLITQAHAAIKELLSQAETAPA
SAAKTKVKALAADS NFAKVVVNKLKLIGMEGSGNTLEHNAKVKIKNLLADGCKNFDSKWQ
ALIDTPIYDGKSEKIKTEKASSISSTTEIIQSVLYYQTKQQTDFLRLQKELE
```

NCBI blast 2 sequence scores for N-terminal domains excluding N-terminal signal sequences

Tororo v Tbg: 50%

Tororo v Tbb 927: 58%

Tbg v Tbb927: 50%

Bug2

MLATAETRQTARLLTLQTAVLAALVIPRSADAAAHHSSSGISG
FRAICELINLGAASCQDDQVGAESNDIKEAAAALINLTIANPAIITELEAKATPEEAIGTE
NSKAAQOCTGDNEWICKAAHSRLKQKKGLKTKQTLTELSTKPTLKESIEAVLNQLKTEIA
SIDNRSGAKQCEAKKALATAIGVQDTSKDNLALPEATTDRAATTCGKPSGTDKGTIAGKN
LLADAVCLCSSDAASSNAGDNCNLQGTGNLDFASAI SVSSSDLKLLGGCSAFKADTKL
TAHALEAALSASFNERFRNQGTNNKITNVIGHVDGTGAEC DGS DAGNKGACVYYGKSPGS
KALLPPAWREHITAAIGHLQHAEQERLRRHESIKRATTLNNTLAALILIAGDASSSQALR
EAAQTKQTNIGAQP GGKAKDCGQHKDNTTCTAHKNCKWEGESDTKGECKPKDGEGQTNP
AAGAGTDGKTNNTGSNSFAINKAPLLLAFLLF

Domain combination: B2

GeneDB Blast results N-terminal domains

Tbb 927 – best match CONTIG11866 32%

>Bug2

MLATAETRQTARLLTLQTAVLAALVIPRSADA
AAAHHSSSGISG
FRAICELINLGAASCQDDQVGAESNDIKEAAAALINLTIANPAIITELEAKATPEEAIGTE
NSKAAQOCTGDNEWICKAAHSRLKQKKGLKTKQTLTELSTKPTLKESIEAVLNQLKTEIA
SIDNRSGAKQCEAKKALATAIGVQDTSKDNLALPEATTDRAATTCGKPSGTDKGTIAGKN
LLADAVCLCSSDAASSNAGDNCNLQGTGNLDFASAI SVSSSDLKLLGGCSAFKADTKL
TAHALEAALSASFNERFRNQGTNNKITNVIGHVDGTGAEC DGS DAGNKGACVYYGKSPGS
KALLPPAWREHITAAIGHLQHAEQERLRRHESIKRATTLNNTLAALILIAGDASSSQALR
EAAQTKQTNIGAQP GG

>gamb856d03.q1k_1

MLATAETRQTARLLTLQTAVLAALVIPRSGSAA
AAHTSSGISGFRAICELINLGDASCQE
DQVGAESNDIKEAAAALINLTIANPAIITELEAKDTAEEAIGTENS KAAQOCSGDNEWICK
AAHRLKQKKGLKTKQTLTELSTKPTLKESIEAVLNQLKTEIASLDNRDGAKHCSEAKKA
LATAIGVQDTTKDNLNLPEATTDRAATTCGKPSGTHKGTIAGKNLLGDVAVCLSSDATSSN
AGENCNLQGTANLDFRS AISVTNSDLKLLGGCSAFKADSKLTAHALEAALSASFDSERF
RNQGTNNKITNVIGHVDGSGNECDGSNDGNKGACVYYGKSPGSKALLPPAWREHITAAIG
HLQHAEQERLRRHESIKRATTLNNTLAALILIAGDASSNQAALREAAQTKQTNIGAQP GG

NCBI blast 2 sequence scores for N-terminal domains
excluding n-terminal signal sequences

Tororo v Tbg: 93%

Tororo v Tbb 927: N/A

Tbg v Tbb927: N/A

Buteba 1

>Buteba 1

```
MSMSLSLTIRKRQPQNAITNAIATTTVFISIIILLAINSGATAPTTAVNAREFDLLCTLVRA
EENLEERQTASQAAKEVVALAAQIELILANLKHIERLAAAEPQDAPKESSSGDTPEACKA
PKSTVCTKAAQIYKRFRPDEKLALAFLAETTTGELRATFNVTLKQIATTATSHARYFGQNT
ETRQALAKIKKALYGSPEAKGDAIIDSADSTRSAACGKTDSNEANSAKKRAAAALICLCG
GDNTNTGNDACFTQTKADINYASKSGAVEKAWTEISQKCQANTRTTKVTTAQLKTAATEL
ATLIHQKRGDKAVAGLLGAAQINAGAVDCDGSDANGKGSCVILSTSAKKYKVETPDWLTA
MQEAINNLDDEQIEINNGQAEAHILALNSSLTTLLAQAVAPAKQAPTQAAVPPEKQTNP
QKDCTKNTKKKDCKEGDGCKWTNEDEETGSHCKAKNDGEGVKTENEGTTTTSATGNNSFV
IKTSPLLLAVLLF
```

Domain combination: B2

GeneDB Blast results N-terminal domains

Tbg:

Two non-overlapping hits with

>gamb333c09.plk

>gamb215b11.plk

cannot join or extend to form contigs

>But1

```
MSMSLSLTIRKRQPQNAITNAIATTTVFISIIILLAINSGATAPTT
AVNAREFDLLCTLVRA
EENLEERQTASQAAKEVVALAAQIELILANLKHIERLAAAEPQDAPKESSSGDTPEACKA
PKSTVCTKAAQIYKRFRPDEKLALAFLAETTTGELRATFNVTLKQIATTATSHARYFGQNT
ETRQALAKIKKALYGSPEAKGDAIIDSADSTRSAACGKTDSNEANSAKKRAAAALICLCG
GDNTNTGNDACFTQTKADINYASKSGAVEKAWTEISQKCQANTRTTKVTTAQLKTAATEL
ATLIHQKRGDKAVAGLLGAAQINAGAVDCDGSDANGKGSCVILSTSAKKYKVETPDWLTA
MQEAINNLDDEQIEINNGQAEAHILALNSSLTTLLAQAVAPAKQAPTQAAVPPEKQT
```

>Tb10.v4.0152

```
MSMSLSLTNRKQQLQNAITNAIATTTVLISIIILLAINSGAAAPTT
AVNAREFDLLCTLVRA
EENLEERQTAGQAAKEVVALAAKIELILANLKHIERLAAAEPQAAPKESTSDETPEACKA
AKSTVCNKAQIYKFRPDEKRALAFLAETTDKLSATVNVTLKQIADNKANHAAYFGEST
GKRPAFANVKKALYGDPAVTGDGIIDTGDNTRPKACGNTASAAAKSGKARAEAAALICLRG
SDSSNGGNAACFTGAALDTNYGSPSNDFATAWTQIKQCSQKQRQAKVTGEQLKTAASAL
NHLVHSHKGDKAVAGLLGYAEANTGAISCDGSTGNGKGSCVILSTSAKSYKVETPDWLTA
MQETINNLDGEKHEIEKGWKAEAHTLALNSSLTTLLALQAGSSKEAAPEKQASPEKQS
```

>Tbg

```
MSMSLSLTPRKQELRSAITNAITTTAVFINTIPLAINSDAAGPTK
AVNAREFELLCTLVQA
EDNLEARLSAGAESKSVVKLAEQLEIILPNLKHIERLAAAEPQAAPKESSSDETPEACKA
AKSTVCDKAAQIYKKLHPDEKLELAFLAETTSKL
gap
TRPTACGNTASAVANSAKKRAAAALICLCG
SDNTNTANAACFTGAAPNTNYGSPSNDFATAWTQIKQCSQKQKQAKVTGEQLKTAASPL
NHLVHSHKHGDKAVGLLGYAEANTGAISCDGSTGNGKGSCVILSTSAKSYKVETPDWVTA
MQETINNLDGEKHEIEKGWKAEAHTLALNSSLTTLLALPAGSSKEAAPEKQASPEKQS
```

NCBI blast 2 sequence scores for N-terminal domains
excluding N-terminal signal sequences

Tororo v Tbg: 67% (for available sequence)

Tororo v Tbb 927: 69%

Tbg v Tbb927: 87% (for available sequence)

Buteba 2

```
MAKISSKIEVSYIVLPLSLIFAVSGCHATQGDALSKENRDELCNLNKELAVL
PATYVKNLISQWDAEDLRASDELRLRIYLADNTPQKAAELAPLLAAHSAIVQRTATETAT
ETKEVAGLIAAAAAALNGQITEFFHIAANAYQAGGTGKGCVTGEDASKIVQGTKGLGDCG
ETLTVEPTAAAAALQTLTATGFTNLQGAKTTTQRADASADCRLFNTDGSNGVLGSQASAGG
LTFAGGYIKVAANADTITGSSLAAMNEGHEATHKAYHRLWRAYVTITKSKAATQAGYSRI
NLDQLTGTTEARRVFREAIQKTPKYDDAKDSDTVKQSMTAHYGPDAKFNSKTWELVDGHQ
VKKEAYETNGKGTEKLANIKDINELRKILTYYQTRQTSISNNAAQETKQCSEITAQEQKP
PETDTTCEKKGTGAECKDGCKVEGTGDSKKCVKDPTYKPKQAEGVKAEGKGTNTTGSNS
FVIKKSFLFLAFLLF
```

Domain combination: A2

GeneDB Blast results N-terminal domains

Tbb 927 best match Tb927.3.270 32% identity

>but2

```
MAKISSKIEVSYIVLPLSLIFAVSGCHA
TQGDALSKENRDELCNLNKELAVL
PATYVKNLISQWDAEDLRASDELRLRIYLADNTPQKAAELAPLLAAHSAIVQRTATETAT
ETKEVAGLIAAAAAALNGQITEFFHIAANAYQAGGTGKGCVTGEDASKIVQGTKGLGDCG
ETLTVEPTAAAAALQTLTATGFTNLQGAKTTTQRADASADCRLFNTDGSNGVLGSQASAGG
LTFAGGYIKVAANADTITGSSLAAMNEGHEATHKAYHRLWRAYVTITKSKAATQAGYSRI
NLDQLTGTTEARRVFREAIQKTPKYDDAKDSDTVKQSMTAHYGPDAKFNSKTWELVDGHQ
VKKEAYETNGKGTEKLANIKDINELRKILTYYQTRQTSISNNAAQETKQCSEITAQEQKP
P
```

>Tbg1

```
AKISSKIEVSYIVLPLSLIFAVSGCHA
TQGDAL*KENRNELCNLNKEIAVLPATYVKNLISQ
SETEEGKATDELRLRIYLADSTPKKAAELAPLLAAHSAIVQRTATETATETEEVAGLI
AAAGALNGQITEYFHIAANAYQSGSGKGCITADADSKIVQGPAGLTDCGLTTLSVDPVA
ATEITEVTKTVFPGLNARK
Gap
SQQSAGALVFAGGYITVAANTAEITATSLAAMNEGHATSHKAYHRLWKAYVTITKSKAAT
QATYSRINLADLTNTPEARRVFREAIQKTPKYDDAKDSDTVKQSMTAHXSPDVKFNSKTW
ELVDGHQVTKEGYETNGKGTEKLANIKDINELRKILTYYQTRQTSISNNAAQETKQCSEI
TAQGQKHQ
```

NCBI blast 2 sequence scores for N-terminal domains excluding n-terminal signal sequences

Tororo v Tbg: 84% over available sequence

Tororo v Tbb 927: N/A

Tbg v Tbb927: N/A

Buteba 3

```
MKLCIFCSVFI AVLQMV EIGASGAVTPGENAAEFAAVCSLV
QFAKAKIPPASEPQEIQKIINTLSVINFTLMDEQTRSTVEKNKEKTWEQIQSEHTGETKY
YGDHWDKWTRVAKLPPTDPEALAL KDWERHRANENLKQQIKSLLHEALNLVKEAATFKED
LKAASLERLQELALYGSV GKS AELKSADDRKGYCGKGT TDGTAAGVGATDSLYATILCLC
AGAATDNTV GKGCCLD CQSGANAAAWTAHSNGKDRAEFLASKCPAALIPEHPTTPELNHR
LAAFYKLALGTTSTTQT I KYTVGK VSGTGEGG CDGEVTGTSGRCAKYTAQQILQDDASLK
WRQSLKAAATLYDKQAQAKITTYRLADKLRLINATAAALYHTPLTQTQIQPMSRTDAATK
PTTNCSEHKSSKTTCENTGKCEWEANPSDKSKGECCKPKAGTETPAVGS GDGAAAGASATD
KKCSEKKTGECEKDGCKWEGTECKDSSILVNKKFALS VVCAAFVLLS
```

Domain combination: B1

GeneDB Blast results N-terminal domains

gambiense contig assembled by hand starting with

>gamb1420g06.q1k

>gamb1557e09.q1k

>gamb1465h01.q1k

>gamb655g05.plk

>but3

```
MKLCIFCSVFI AVLQMV EIGASGAVTPGEN
AAEFAAVCSLV
QFAKAKIPPASEPQEIQKIINTLSVINFTLMDEQTRSTVEKNKEKTWEQIQSEHTGETKY
YGDHWDKWTRVAKLPPTDPEALAL KDWERHRANENLKQQIKSLLHEALNLVKEAATFKED
LKAASLERLQELALYGSV GKS AELKSADDRKGYCGKGT TDGTAAGVGATDSLYATILCLC
AGAATDNTV GKGCCLD CQSGANAAAWTAHSNGKDRAEFLASKCPAALIPEHPTTPELNHR
LAAFYKLALGTTSTTQT I KYTVGK VSGTGEGG CDGEVTGTSGRCAKYTAQQILQDDASLK
WRQSLKAAATLYDKQAQAKITTYRLADKLRLINATAAALYHTPLTQTQIQPMSRTDAAT
```

>927chr9

```
MKLCIFRSVFI AVLQIFEIGASGAAPPGEN
AAEFAAVCSLVQFAKAKIPPASEPQEIQKI
INTLSAINFTLMDDLTRSTVEKNK*KT*EQIQSEHTGETLYFGDH*DKWTRVAKLPETDP
EALAL KDWERHRANNLKKQIKSLL*EALALDKEANAKEDLKVASITNLQEAALYGS PG
KSVELKTTDDRKG YFGKSTDGAAAGVGATDSLYATILCLCAGATGDEAGGKGCCLDCQS
GANGAAWTPPTSAGKARAEFLASKCPAALIPEHPTTTELNRHLRAVYKLALGTTSTNTQT I K
YTVGK VSGTGEDGCGGEVTGTIGRCVKYTAEQILEDEASLK*RQSLKAAAKLYDKQAQAK
ITMYRLADKLRLINATAAAL*HTTLT*QTI*PMSRTDAAT
```

>tbg

```
MKLCIFCSVFI AVLQMF EIGATGAATPGEN
AAEFAAVCSLVQFAKAKIPPASEPQEIQKIINTLSAINFTLMYEQTRST
VEKNKEKTWEQIQSEHTGETKY YGDNWDKRTRVAKLPETDPEALAL KDWERHRANNLKN
QIKSLLWEALSLKTEAASAKEDLKAANLERLQELALYGSAGKSAELKSADDRKGYCGKGT
TDGTAAGVGATDLLYATILCLCAGAATDNTAGKGCCLDCQSGPNAEAWTAHSDGKDRAEF
LASKCPAALIPEHPTTPELNHRLAAFYKLALGTTSTTQTVKYTVGK VSGTCEGGCDGEVT
GTTGRCAKYTAQKILQDDASLKWRQSLKAAATLYDKQAQAKITTYRLADKLRLINATATA
LYHTTLTQTQIQPMSRTDAAK
```

NCBI blast 2 sequence scores for N-terminal domains excluding n-terminal signal sequences

Tororo v Tbg: 92%

Tororo v Tbb 927: 84%

Tbg v Tbb927: 82%

CLUSTAL W (1.83) multiple sequence alignment

```
but3      MKLCIFCSVFIQVAVLQMVQVEIGASGAVTPGENAAEFAAVCSLVQFAKAKIPPASEPQEIQKI 60
tbgDal    MKLCIFCSVFIQVAVLQMVQVEIGATGAATPGENAAEFAAVCSLVQFAKAKIPPASEPQEIQKI 60
Tbb927    MKLCIFRSVFIQVAVLQIFEIGASGAAPPGENAAEFAAVCSLVQFAKAKIPPASEPQEIQKI 60
          *****
          *****:.*.***:*.***.*****

but3      INTLSVINFTLMDEQTRSTVEKNKEKTWEQIQSEHTGETKYYGDHWDKWTRVAKLPPTDP 120
tbgDal    INTLSAINFTLMYEQTRSTVEKNKEKTWEQIQSEHTGETKYYGDNWDKRTRVAKLPETDP 120
Tbb927    INTLSAINFTLMDDLTRSTVEKNKKT--EQIQSEHTGETLYFGDH-DKWTRVAKLPETDP 117
          *****
          ***** : *****:.* ***** *:*:* ** ***** **

but3      EALALKDWERHRANENLKQIQIKSLLHEALNLVKEAATFKEDLKAASLERLQELALYGSVG 180
tbgDal    EALALKDWERHRANNLNQIKSLLWEALSLSKTEAASAKEDLKAANLERLQELALYGSAG 180
Tbb927    EALALKDWERHRANNLNKQIQIKSLL-EALALDKEAANAKEDLKVASITNLQEAALYGSFG 176
          *****
          *****:***:***** ** * .***. *****.*.: .*** ***** *

but3      KSAELKSADDRKGYCGKGTTDGTAAGVGATDSLYATILCLCAGAATDNTVKGKCLDCQS 240
tbgDal    KSAELKSADDRKGYCGKGTTDGTAAGVGATDILLYATILCLCAGAATDNTAGKGCCLDCQS 240
Tbb927    KSVELKTTDDRKGYFGKSTDGAAAGVGATDSLYATILCLCAGATGDEAGGKGCCLDCQS 236
          **.*.***.:***** **.*.***:***** *****:*.*: *****

but3      GANAAAWTAHNSGKDRAEFLASKCPAALIPHPPTPELNHRLAAFYKLALGTTSTTQTIK 300
tbgDal    GPNAEAWTAHSDGKDRAEFLASKCPAALIPHPPTPELNHRLAAFYKLALGTTSTTQTVK 300
Tbb927    GANGAAWTPTSAGKARAFLASKCPAALIPHPPTTELNHRLRAVYKLALGTTSTNTQTIK 296
          *.*. ***. * ** *****.***** *.*.*****.*.***:*

but3      YTVGKVSQTGEGGCDGEVGTGSRCAKYTAQIQLQDDASLKWRQSLKAAATLYDKQAQAK 360
tbgDal    YTVGKVSQTCEGGCDGEVGTGTRCAKYTAQIQLQDDASLKWRQSLKAAATLYDKQAQAK 360
Tbb927    YTVGKVSQTGEDGCGGEVGTIGRCVKYTAEQILEDEASLK-RQSLKAAAKLYDKQAQAK 355
          *****
          *.*.***.***.***:***:***:***** *****.*****

but3      ITTYRLADKLRLINATAAALYHTPLTQTQIQPMSRTDAAT 400
tbgDal    ITTYRLADKLRLINATAALYHTTTLTQTQIQPMSRTDAAK 400
Tbb927    ITMYRLADKLRLINATAAALHTTTLTQ---IPMSRTDAAT 392
          ** *****:***: * ** *****.
```

Buteba 4

```
MPKATYALVVQTLTLLVSKV
HQVQPTAGDAIKKKYWTALCDIAVDADKIAAKALHNLQOPATPGAESLRNLLRALVYNLA
NNTTEPTSPGERMIWSHFEEASAAIDFYKGDKPAKLITAVRDAARANGAILDWINTQHKV
STSSHGCLSGGDSGEAPAAGSAGMATIDAKCVPNWTAVTAEAAATTAVGLQGLQGTLLGGQ
KGEDELANGGHHGCSNSNSANSVAVKLLNGGTGTSVAGQSPTLVAGIFKLTGTSMAMDDLNSV
ATKQTTHPYVYHIVQAVKTAGETLATEDEVSTIDRAKTSNSFKSVARVHLLGKATDSRRD
TEIPDKIQAVFGSPDKYTAIYKSNVDEMPLTNELTHDPNLKKLGDITDINDLMRLYFFFL
DRLKSHVKDLTAQLEEAKRQOSPksakekekeCHTKGQDKQEECEKLESQGCVFNKDGKD
GEKCTLKKEVETELEKATTSEEGKDGKATNSTGSNSFVINKAPLLLAFLLF
```

Domain combination: A2

GeneDB Blast results N-terminal domains

>but4

```
MPKATYALVVQTLTLLVSKVHQVQPTA
GDAIKKKYWTALCDIAVDADKIAAKALHNLQOPATPGAESLRNLLRALVYNLA
NNTTEPTSPGERMIWSHFEEASAAIDFYKGDKPAKLITAVRDAARANGAILDWINTQHKV
STSSHGCLSGGDSGEAPAAGSAGMATIDAKCVPNWTAVTAEAAATTAVGLQGLQGTLLGGQ
KGEDELANGGHHGCSNSNSANSVAVKLLNGGTGTSVAGQSPTLVAGIFKLTGTSMAMDDLNSV
ATKQTTHPYVYHIVQAVKTAGETLATEDEVSTIDRAKTSNSFKSVARVHLLGKATDSRRD
TEIPDKIQAVFGSPDKYTAIYKSNVDEMPLTNELTHDPNLKKLGDITDINDLMRLYFFFL
DRLKSHVKDLTAQLEEAKRQOSPksak
```

>Tb92703

```
MAKATYALVVQTLFLVSKVNQVQPTA
GDAIKKKY*KALCDIAVYADNIAAKALSSMQEPA
TPCRKALQNLRLRALVYNLANSTEQATPGEKMIWSHFAEQTGTTIEFYNGDNPAKLITAVR
DAARAGGAILDRIETQHTISTITHDCLSPNDAGGPPATGNTNLSMEAKCVSDWTAVKPK
KEATVAVSHVGLQGLSRNNAHSELSNGDHKCSNSNSADIAVKLLHGDTGTSLAANNPKLV
AEIFKLTDNGITTEGLANVAAKQQTNPYVYHIVQAVKASEETITLEAVSTLYKAKPSNSP
KSVARVHFLGTIHTDTTRDTEVPDKIQTAFGSPEEYTAIYSRNVDEMPLTKELTHDPNLK
KLSDITDINDLMRLYFFFSRLKSQVKDLTGQLEEAKRQOAPKSAE
```

>TbgChr3

```
MPKATYALVVQTLTLLVSQVQVQPTA
GNAIKKKY*KALCDIAVDADKIAAKALHNLQOPA
TPGAASLRNLLRALVYNLANNTEPTSPGERMIWSHFEEASAAIDFYKGDKPAKLITAVR
NAARTNGAILDWINTQHKVSTSSHGCLGADNTVTNLVDKDTAMASAAPNCVASWAEVTAN
NAGAATVSAQGFTEGLRVNLDKKTTHPYVYHIVQAVKTSEETITLEDVSTLDKAKTSNS
AGIFKLTDNGIATEGLRNVDENKKTTHPYVYHIVQAVKTSEETITLEDVSTLDKAKTSNS
FKGVARVHLLGKATDGSRDTEVPDKIQTAFGSPEEYTAIYSRNVDEMPLTKELTHDPNL
RKLGEITDINDLMRLYFFFSRLKSQVKDMTGQLEEAKRQOAPKSAK
```

NCBI blast 2 sequence scores for N-terminal domains
excluding N-terminal signal sequences

Tororo v Tbg: 74%

Tororo v Tbb 927: 67%

Tbg v Tbb927: 68%

CLUSTAL W (1.83) multiple sequence alignment

```

but4      MPKATYALVVQTLTLLVSKVHQVQPTAGDAIKKKYWTALCDIAVDADKIAAKALHNLQOPA 60
TbgDal    MPKATYALVVQTLTLLVSQVQVQPTAGNAIKKKY-KALCDIAVDADKIAAKALHNLQOPA 59
Tbb927    MAKATYALVVQTLFLVSKVNQVQPTAGDAIKKKY-KALCDIAVYADNIAAKALSSMQEPA 59
          *.*****:***:*:*****:***** .***** **:***** .:***

but4      TPGAESLRNLLRALVYNLANNTEPTSPGERMIWSHFEEASAAIDFYKGDKPAKLITAVR 120
TbgDal    TPGAASLRNLLRALVYNLANNTEPTSPGERMIWSHFEEASAAIDFYKGDKPAKLITAVR 119
Tbb927    TPCRKALQNLRLALVYNLANSTEQATPGEKMIWSHFAEQTGTTIEFYNGDNPACLITAVR 119
          **  :*:*****.** :*:***:*****:..:***:***:*****

but4      DAARANGAILDWINTQHKVSTSSHGCLSGGDSGEAPAAGSAGMATIDAKCVPNWTAVTAE 180
TbgDal    NAARTNGAILDWINTQHKVSTSSHGCLGADNTVTNLVDKDTAMASAAPNCVASWAEVTAN 179
Tbb927    DAARAGGAILDRIETQHTISTITHDCLSPNDAGGPPATGNTNLQSMEAKCVSDWTAVKPK 179
          :***:.***** *:***.*** :*.*** .:~ . .: : .:***.*** :*..

but4      AAATTAVGLQGLQGTLLGGQKGEGLANGGHGCNSNSANSVAVKLLNGGTGTSVAGQSPTLV 240
TbgDal    NAGAATVSAQFTGVLGTAVPHGTLANTDKKCNSSNGHTAIKLLNDATGTNVGPAEPKLV 239
Tbb927    KEATVAVSHVGLQELSRNNAHSELSNGDHKCNSSNADIYVAVKLLHGDTGTSVLAANNPKLV 239
          .:.:*. * : * * . . . * : * . : * * * * . . * : * * * . . * : * * *

but4      AGIFKLTGTSMAMDDLNVAT-KQTHPYVYHIVQAVKTAGETLATEDVSTIDRAKTSNS 299
TbgDal    AGIFKLTDNGIATEGLRNVDENKKTTHPYVYHIVQAVKTSEETITLEDVSTLTKAKTSNS 299
Tbb927    AEIFKLTDNGITTEGLANVAA-KQQTNPYVYHIVQAVKASEETITLAVSTLYKAKPSNS 298
          * *****.:.: :.* ** * : * :*****: : * : * * * : :*.***

but4      FKSVARVHLLGTKATDSRRDTEIPDKIQAVFGSPDKYTAIYKSNVDEMPLTNELTHDPNL 359
TbgDal    FKGVARVHLLGTKATDGSRDTEVPDKIQAFGSPEEYTAIYSRNVDEMPLTKELTHDPNL 359
Tbb927    PKSVARVHFLGTIHTDTTRDTEVPDKIQAFGSPEEYTAIYSRNVDEMPLTKELTHDPNL 358
          *.*****:*** ** *****:*****:*****:*****:*****:*****

but4      KKLGDITDINDLMRLYFFFLDRLKSHVKDLTAQLEEAKRQOSPksAK 406
TbgDal    RKLGEITDINDLMRLYFFFSRLKsQVKDMTGQLEEAKRQOPksAK 406
Tbb927    KKLSDITDINDLMRLYFFFSRLKsQVKDLTGQLEEAKRQOPksAE 405
          :*.:***** *****:***:*.*****:*****:
    
```

Buteba 6

MYALMLQLTAVLWLVLTQTQAANIVGGANRPDHAALCKFITMAAREVEVPEIPALPTDEY
NYIHMVNFSTAPEDWQNMFFEDKASKKAHADPKAAGATGRGFANWDQWQKTALSHLDA
TTADAKKQAGVVQLSDTEAALVKPQLAHLAALADELITELSGLQPGPGSTTASSAKQLL
TTAAYGEETVPKPNPNPEKVFGAALNGNRDAICEATSTGPQVATTLATLACVCHKGNAN
TVDPPVCTDQAKGSEWQDADPAATALTGNLAAIAKSCAKESKEKVTAELATAITDL
DKLIHKGATDAYLGSFVSTKCNSSGNGVCKVFASYNTKPSLVNKLWIPQLNDLVTK
LMDLERKKIRAEQLVERIKSLKKETTVVIEVAKATASSMEKFETKTNHQQTSEATSR
STHNTNTTCTQNKCKWDGTTEDKGCVADESKGATQTNAGTGAAGTTTEKCKDKKKDD
CKDGCWEGETCKDSSILATKKFALSUVSAAFAALLF

Domain combination: B1

GeneDB Blast results N-terminal domains

Tbg closest match >gamb1360f04.p1k_2

>But6

MYALMLQLTAVLWLVLTQTQA
ANIVGGANRPDHAALCKFITMAAREVEVPEIPALPTDEYNY
IHMVNFSTAPEDWQNMFFEDKASKKAHADPKAAGATGRGFANWDQWQKTALSHLDATTAD
AKKQAGVVQLSDTEAALVKPQLAHLAALADELITELSGLQPGPGSTTASSAKQLLTTAAYG
EETVPKPNPNPEKVFGAALNGNRDAICEATSTGPQVATTLATLACVCHKGNANTVDPPVCT
DQAKGSEWQDADPAATALTGNLAAIAKSCAKESKEKVTAELATAITDLDKLIHKGATD
AYLGSFVSTKCNSSGNGVCKVFASYNTKPSLVNKLWIPQLNDLVTKLMDLERKKIRAE
QLVERIKSLKKETTVVIEVAKATASSMEKFETKTNHQQTS

>Tb11.v4.0034

MYAIMLPMTIVLWLVTKTEG
TDIVGGANRQEHAALCKFVAMAPREVEIPTIPAMPEDDL
YIHMVNFSAATGSWQEMFYSDKAAKTHDNPKSAGQDGRGFEEWPRWTKIAAKKLEATT
GGQKNQAGIMELTEDQETLARDHLKHIRTRAHELAKELQSIQPPSDAPKDNTAKETIATA
VYGENAKPASNPDPKVFVTTGTRDSVCKAAGGNGNPATALEALTCVCFRGTNNVDPA
VCVQAADGGTWEAASTFATGVTAANLGLKAKSCFPDAGQITGTEILAAIANLNDLIHKDT
SNAYLGQFASSKCDGNKAQGVCLFSGFASKAQPIVEQLTWIPPLKNLALGLQTLSENKQ
RGLQIIEQIKLKKDAVVIVDSAKATAAALEKIKGGTQKQHTIG

NCBI blast 2 sequence scores for N-terminal domains
excluding n-terminal signal sequences

Tororo v Tbg: N/A

Tororo v Tbb 927: 46%

Tbg v Tbb927: N/A

Buw1

MSMKQGGAALVLLIMTIKRVNALVADESEAAHRAMCALTEAASATFTAPSGTEQADELNDK
IQANMSVADAHWQALFDSEAGSKPYDETTGDNRTIADSLGGQEKWDGWRKAFANIKALNI
GTATKGEYPKITNEVDRKIARQQLRPIAAQAIKLAAMKPLKAFKFLSDGKINKINENLRKAL
YGGDGLDQPTPTISKSFQDQD'TWNNLCQGSAKRKSIAAGDFFCICTGAAAATKQCSAAYSNT
EHSNQPSINTGWTALQKSCGKREQSKATVESIYAAVTQWRVALNQKSGSENSVWLGKSSG
TGTECAGTDGNTCVDYSGFFKANSADLATLPWLSALSEAAAKLRQAAEAAAAVKSLQTM
QALQGATVEIYSAASGVLAKEMTIAIPTAPPTQHELQPQTQKKQAEAEETECNTAGDNKA
CEKKTGCCTYDEAKDKGKRCCLSEEAKKVADGTAKDAKTNTTASNSFVIHKAPLLLLAFLLF

Domain combination: B2

GeneDB Blast results N-terminal domains

>Buw1

MSMKQGGAALVLLIMTIKRVNALVAD
ESEAAHRAMCALTEAASATFTAPSGTEQADELND
KIQAANMSVADAHWQALFDSEAGSKPYDETTGDNRTIADSLGGQEKWDGWRKAFANIKAL
NIGTATKGEYPKITNEVDRKIARQQLRPIAAQAIKLAAMKPLKAFKFLSDGKINKINENLR
KALYGGDGLDQPTPTISKSFQDQD'TWNNLCQGSAKRKSIAAGDFFCICTGAAAATKQCSAA
YSNTEHSNQPSINTGWTALQKSCGKREQSKATVESIYAAVTQWRVALNQKSGSENSVWLG
KSSGTGTECAGTDGNTCVDYSGFFKANSADLATLPWLSALSEAAAKLRQAAEAAAAVK
SLQTMQALQGATVEIYSAASGVLAKEMTIAIPTAPPTQHELQPQTQKKQAE

>Tb927_11

MSIQGGAALVALIMLATDRVKAVAD
ESEAAHRAMCALTEAASATFTTPAGSPEAEGLS
DKIQAANMSVADARWQAFQFDSEAGSKPYDETTGDNRTLADKLGKKEKWDGWRKAFANIKV
LNIGTKPQGEYPKINSEVDRKILARQQLRAVAQAATQLERTMKPLKAFKFLSDDKSNKINENLR
RKALYGGDGLDQPTPTISKSFQDQD'TWNNLCQGSAKRKSIAAGDFFCICTGADATTKQCSG
AYSNTAHDQSPTSINTGWTALQKSCGKREQSKATAGTIYAAVAQW*AALKQKASGSDNSVW
LGSSSTGTACAGTD'TNCTIDYSDFFKSAGADLATLPWLSALSDA
AQLKQAEAAAAVKSLQAIQALQGASVEIYSAASGVLAKEMPIAIPPTQ
RLQPQTQKKQAE

>gamb217f04

RQASTALRTLIMLT'TDRVKADAD
DTEAVHRAMCALTEASVTFTPPAAVAEPAGLNDKI
QAANMSVADARWQALFDSEVGSKPYDETTGDNRTLADKLGKKEKWDGWLQAVANIKALNI
GTKPEGEYPKINSEVDRKISRQQLRPIAAQATKLAIAMK*LKAFKFLSDDKINKINENLRKA
LYGGDGLDQPTPTISKSFQDQD'TWNNLCQGSAKRKSIAAGDFFCICTWQNAVTK*CSGAYS
HTAHNDQNNINTGWTALQKSCGKREQSKASVASIYAAIAQWRAALKQKNKGGSDNSVWLG
ASDNTGATCAGTD'TNCTVDYSDFFK
KTGGIDLAT*HRLKALTAVEQ*LKQAEETAAGVMSLQVQMQAVQASVEINSAAASG
VLSKEMPITVPTARSQSTEINYKHSKNK
Reading frame adjusted after DFFK

NCBI blast 2 sequence scores for N-terminal domains
excluding n-terminal signal sequences

Tororo v Tbg: 69%

Tororo v Tbb 927: 80%

Tbg v Tbb927: 78%

CLUSTAL W (1.83) multiple sequence alignment

```
Buw1      MSMKQGGAALVLLIMTIKRVNALVADESEAAHRAMCALTEAASATFTTAPSGTEQADELND 60
Tbb927    MSIQQGAAALVALIMLATDRVKAVADESEAAHRAMCALTEAASATFTTTPAGSPEAEGLSD 60
TbgDal    ---RQASTALRTLIMLT'TDRVKADADDTEAVHRAMCALTEAASVTFTTPAAVAEPAGLND 57
          :*..:**  ***  .          **;**.*****.***.*:.  :.  *.*

Buw1      KIQAANMSVADAHWQALFDSEAGSKPYDETTGDNRTIADSLGGQEKWDGWRKAFANIKAL 120
Tbb927    KIQAANMSVADARWQAFFDSEAGSKPYDETTGDNRTLADKLGKKEKWDGWRKAFANIKVL 120
TbgDal    KIQAANMSVADARWQALFDSEVGSKPYDETTGDNRTLADKLCGKEKWDGWLQAVANIKAL 117
          *****:***:****.*****:***.* *:***** :*.*****

Buw1      NIGTATKGEYPKITNEVDRKIQQLRPIAAQAIKLAAAMKPLKAFKAFKINENLR 180
Tbb927    NIGTKPQGEYPKINSEVDRKLRQQLRAVAAQATQLERTMKPLKAFKAFKINENLR 180
TbgDal    NIGTKPEGEYPKINSEVDRKISRQQLRPIAAQATKLAIAMK-LKAFKAFKINENLR 176
          **** .:*****.*****:*****:*****:* :** *****.* *****

Buw1      KALYGGDGLQPTPTISKDFGQDGTWNNLCQGSKRKSIAGDFFCICTGAAAATKQCSAA 240
Tbb927    KALYGGDGLSTPTLGKSFANGESWAGLCATKANRKSIAAGDFFCICTGADATTKQCSGA 240
TbgDal    KALYGGDGLSTPTLGKSFANGES-AGLCATKANRKSIAAGDFFCICTWQNAVTK-CSGA 234
          *****:*.**:.*** :*:* .** .*:*****.*** ** **.*

Buw1      YSNTEHSNQPSINTGWTALQKSCGKREQSKATVESIYAAVTQWRVALNQGK-SGSENSVW 299
Tbb927    YSNTAHDSQPSINTGWTALQKSCGKREQSKATAGTIYAAVAQW-AALKQKA-SGSDNSVW 298
TbgDal    YSHTAHNDQNNINTGWTALQKSCGKREQSKASVASIYAAIAQWRAALKQKNKGGSDNSVW 294
          **:* *.*. * *****:*****:.. :*****:** .**:* **.*:****

Buw1      LGSSSGTGTECAGTDGNTCVDYSGFFKANSAGDLATLPWLSALSEAAAKLRQAAEAAA 359
Tbb927    LGSSSTGTACAGTDNTCIDYSDFFKKSAGDLATLPWLSALSDDAAQLKQAEAAA 358
TbgDal    LGASDNTGATCAGTDNTCVDYSDFFKKTGGIDLATH-RLKALT-AVEQLKQAEETAAGV 352
          ** *.**:* ***** **:***.* ** .** ***** *.*:* *. :**:* **.*

Buw1      KSLQTMQALQGATVEIYSAAASGVLAKEMTIAIPTAPPTQHELOPOTQKKQAE 413
Tbb927    KSLQAQIQALQGASVEIYSAAASGVLAKEMPIAIPPTAPPTQORLOPOTQKKQAE 412
TbgDal    MSLQVQMVAQVQASVEINSAAASGVLKEMPITVPTARSQSTEINYKHSKNK-- 404
          ***.*:**:***:*** *****:***.*:*** . . .:: :.*::
```

Buw2

MTTPGKIVIAVYVITTAITYVSSANAALKD'TTWKSLCGLEEEELRKLPSVAATTLERQETSVN
AYKITGLK'TLLYGH'TLGD'SKDAAA'IASLGL'TLLKKADEEQSSKHTGTTT'AIKALTSSQELVG
RIHEGIEILAIATHDSVW'CLGN'TAGTADAAGERTGAGCKGKIHDLT'TTAANLGDDVIDSDGY
KTIQDITDGTGVADSSK'CPFT'AVAANHQ'TWGAGKANAKMIDGMLTFTTAEQVTRSGFTKKRA
AKDRLTDKLSVTVHADVMALKEAYRNTEIRAGSSALKAAAEDNQLTPTLRLLQRAPYNMPA
ADADAEAKNLVKT'KFSTT'GSTLEDLWKQIKDASVVDIMKETEEATTIEKIQOLETLRATIEL
YAASDKGKLAKL'KKELEQAQADKNGAKVSKTTEE'CN'CHKELGPCQKAGCKFDNSKNDGEKCF
PDPEAKTDDKKGREDGK'TTTTSTTGSNSVLINKAPILLAFLLF

Domain combination: A2

GeneDB Blast results N-terminal domains

Tbg closest is >gamb249a02.p1k at 31%

N-terminal domains

>Buw2

MTTPGKIVIAVYVITTAITYVSS
ANAALKD'TTWKSLCGLEEEELRKLPSVAATTLERQETS
VNAYKITGLK'TLLYGH'TLGD'SKDAAA'IASLGL'TLLKKADEEQSSKHTGTTT'AIKALTSSQ
ELVGR'IEHEGIEILAIATHDSVW'CLGN'TAGTADAAGERTGAGCKGKIHDLT'TTAANLGDDV
IDSDGYKTIQDITDGTGVADSSK'CPFT'AVAANHQ'TWGAGKANAKMIDGMLTFTTAEQVTR
SGFT'KKRAAKDRLTDKLSVTVHADVMALKEAYRNTEIRAGSSALKAAAEDNQLTPTLRLL
LQRAPYNMPAADADAEAKNLVKT'KFSTT'GSTLEDLWKQIKDASVVDIMKETEEATTIEKI
QOLETLRATIELYAASDKGKLAKL'KKELEQAQADKNGAKVS

>Tb927chr5

MTT'GKIVIAVYV'IAVATV'TTYVAG
TLAALKKNVKKPLCGLEEEELRMLPSVAATALKRQEAS
SNAYKITGFK'TLLYGH'TLDD'SKD'TAAIT'ALGQ'TLLKKADEEQNAKHATTATAIKALTFSQ
ELVGR'IEHEVIEILATATHDSVYFLGNSGATSNAAAERTTAGCKGKIHDLT'TTADNIDDDI
LDENGYKTI'DVENTNGVGD'SNKP'FTAQAASHQ'TWGAGTADAEMIDGMLTCTTAEQVAR
SGFT'KVGTANNRLTDALSITVHAD'TARPKSAYRETKIKGKAAIEQAAEDSOLTPTLKRL
LQRAPYNMPAADANSEADNLVKT'KFSTGTPLEKL*KEIKDASVVDIMKETEEPTTIEKIQ
QTE'LTRV'TIEFYAASDKEVLAKL'KKELDQAQA'AKNGGKVS
(Reading frames adjusted)

NCBI blast 2 sequence scores for N-terminal domains
excluding n-terminal signal sequences

Tororo v Tbg: N/A

Tororo v Tbb 927: 74%

Tbg v Tbb927: N/A

Do1

MIPATGSPAALAVPQQKQASRYLIVVVAAMTTKVATAALEEGANAPAF
HNLCTILAAADTDLAIGQETTTDKDKLLDTHQLNHTFSDQSWQNEFKTAKGPGNWRDTL
PEKYKGDQAKWQEMYPTWLKAAKEAETDEENTAVKKIGAEGLESEAQKLLLRPTVQQLAARA
HDIKNQLQSLAKLEQSDDAKGASKLLKEAAYGSATKTRENTAHTDIFASTCTSTATNCK
DDSSQKPPPTAAGVAMCLCAADGTKSKM CVNPQTALSNWNPQSNHVAEWAKLRVY CSEK
LGGDRPEEA INNALES LMANLHFNGNKLYLGAWQTGACRSGSGNAFCVSYTSDASKDASA
VNKQPWVEKTRRAVAALRNSRQATAEAKQLKRQLTTIVAELAAAAQAANKMNTKTINTEK
TAPTSTKDLNANCSKHKDSKTTCENTGKCEWKGKTETDGTCEVDESKVTTQTNVAKAGTD
SKTNTTGSNSFVIHTAPLLLAVLLF

Domain combination: B2

GeneDB Blast results N-terminal domains

>Do1

MIPATGSPAALAVPQQKQASRYLIVVVAAMTTKVATA
ALEEGANAPAF
HNLCTILAAADTDLAIGQETTTDKDKLLDTHQLNHTFSDQSWQNEFKTAKGPGNWRDTL
PEKYKGDQAKWQEMYPTWLKAAKEAETDEENTAVKKIGAEGLESEAQKLLLRPTVQQLAARA
HDIKNQLQSLAKLEQSDDAKGASKLLKEAAYGSATKTRENTAHTDIFASTCTSTATNCK
DDSSQKPPPTAAGVAMCLCAADGTKSKM CVNPQTALSNWNPQSNHVAEWAKLRVY CSEK
LGGDRPEEA INNALES LMANLHFNGNKLYLGAWQTGACRSGSGNAFCVSYTSDASKDASA
VNKQPWVEKTRRAVAALRNSRQATAEAKQLKRQLTTIVAELAAAAQAANKMNTKTINTEK
TAPTSTK

>Tb09.244.1110

MIALAFASVPKLRANAAVAAGEN
AALFRGLRITILSAAETKITLGSETTAEDPNIGAIQQL
NLTFSPKPTWKAVFKTKPGEGNWRDEIPDTHKENKQWKALYPAWLKAAQAEEDTEPNVLK
EVGAAKLSKHQKIYFGSEVAKVAAQAAKIAENLQELKTHEERSDAESAHLKLTAAFGGA
DKARTRLKHTDMFTSGCAADPQVCKDDTNTGKPPITVAGVIMCVCAADANEPKACINPQ
TALTTWANPSTATATEWPKISAYCTANSGQEINVDSIDAALDRVLATGRFDTKKLYIGAA
NSGACRSGSANGFCITTYTAASATDTSKLNSAPWLANVRKATRALRLSKEATEEAKRLKTO
IATLYAGLSGTAAAATTQIPEPPALDKGTTVARV

>gamb261a02.plk

MIALAFASVPKLRANAAVTAGEN
AALFRDLCTIVSGAKS
TATIGSETTTLDPTVGSIHQLNLTFS DPAWKAVFKTKPGKENWQEEIPDAHKENAQWKAL
YPTWLAAAQADEQDTELKVLKAVGADKLKPHQKAYFGSEVAKVAAQVSKIAETLQELKAE
ERSTDAESAQKLLRRTAAFGGPKTRSGLTHIDMFTSGCAANQAVCKDDDGTKGPPITVA
GVIMCVCAHDGNLGTACINPATALATWADPSAATQAEWTKISAYCTANSGQEINVDSIDD
ALGRVLATARFDSSKLYIGAANGGACRSGSANGFCITTYTAASATDTSKLTAAPWVANVRK
ATRALRLSKKATEDAKRLKAQLATLYAGLSGAAAAAARIPEPPALDKGTTVARLT

NCBI blast 2 sequence scores for N-terminal domains
excluding n-terminal signal sequences

Tororo v Tbg: 44%

Tororo v Tbb 927: 45%

Tbg v Tbb927: 77%

Do2

```
MSHARYVFLATFITAQMASFSAG
QAISTLANKHEYRELC1KL2LALARTGITLQDLSSPPTEDINLVMDINMSLSEVSWQNM3FKDKSKGEWHANA4AE
AKQEGKGYEQSWDKWLN5SRQLKEDAQTTEFKQLKLEDL6KPHQKRSLRRHVQAVAE7E8VAN
EASGFVGGESTEGQLNAAEATK9TLREAAYGAGDTTEAKVTAQQAFGAAMTDQARTATCTT
NAQGRAGKSVLATAACL10CL11KPQEGTQVDGACGMTLDGSGTWRDSAAAPTSADIQNLAKYC
AGDGSTNDPGARIKAAL12TALATSVVRGTTDAHLGAFKTGNCDGKNTGGICVELKNGARAE
DGGVTQLGWYEKLN13TLADKLIRRQQAQTRNKHAKEKIQKLVTTLKAFIKVTKGEAAPEIA
ADAPNAKKGAPG14PATSCSSYNTNAT15CHQNNCKWEENASDKSKGTCKPKTETETPPAGAGT
GAAGASAEDKKCSNKKKQEDCKDGCKWEGTECKD
SSILATKKFALS16VVSAAFVALLF
```

Domain combination: B1

GeneDB Blast results N-terminal domains

Tbb 927 present in reads

```
>tryp_XI-346e09.plk rev comp
reading frame adjusted by inclusion of n at 316
Cannot extend at 3' end
```

Tbg:

```
>gamb1337e11.plk_5 33% identity
```

```
>Do2
```

```
MSHARYVFLATFITAQMASFSAG
QAISTLANKHEYR
EL1KL2LALARTGITLQDLSSPPTEDINLVMDINMSLSEVSWQNM3FKDKSKGEWHANA4AE
AKQEGKGYEQSWDKWLN5SRQLKEDAQTTEFKQLKLEDL6KPHQKRSLRRHVQAVAE7E8VAN
EASGFVGGESTEGQLNAAEATK9TLREAAYGAGDTTEAKVTAQQAFGAAMTDQARTATCTT
NAQGRAGKSVLATAACL10CL11KPQEGTQVDGACGMTLDGSGTWRDSAAAPTSADIQNLAKYC
AGDGSTNDPGARIKAAL12TALATSVVRGTTDAHLGAFKTGNCDGKNTGGICVELKNGARAE
DGGVTQLGWYEKLN13TLADKLIRRQQAQTRNKHAKEKIQKLVTTLKAFIKVTKGEAAPEIA
ADAPNAKKGGA
```

```
>Tbb927
```

```
MPHVHYVFSATFISGEMVWHSTY
QTIRTLANKHKYRELCKILALARTDITLQDLSSPPHE
DINLVMDITMSLSDPSWQDMFVKDKSKGEWHANAADANQKGGYEDQWADWLAARKRIKK
LSKXTAFKGLNLAELKDYQKTNVRQH1VQAVAEQVATAASN2FVPATSEDAKLTSDDVTPTL
RQAAF3GDKAATEGGVTIQQAF
```

NCBI blast 2 sequence scores for N-terminal domains excluding N-terminal signal sequences

Tororo v Tbg: N/A

Tororo v Tbb 927: 64% over available sequence

Tbg v Tbb927: N/A

Kinu1

```
MDYKDCYGYLAALTFVVSLCSG
APLTDNDPVAEQISDLCCK
EKFYVETLIQNLKQRVQSTADSIKDMEKLKATWEKAAAATSDAAKKCLFSALEHKADAEL
LRELPKIEEATEAVTTAQIALSEHVGMLGATITLAKTKLDSGSSNHGNADTGSIRIALSR
TTATTDLCNEPATISDIKPGVSEIQPGKLFKLKLTTKPTDLHKHMFVDWLTIGGLKSCTAH
TSYDQNFFDGALSGCCOYSASGTAVATQASTKPPYATAAVTLFKNNDPEQQCEVTDLPAGGA
ADKHKKLQHSLCRALQLGNVKGHSRQLDGAALQSDHVVANTIRNCDPVFQKLTTATDGE
GTTELKKYIKEAYGSSASEFAEKFITNAEKLQMTLRLNDKIETKDLSAVSPGEQTAAALS
HIQGLHNKRELEAGKKSTSAAAVDPQKSEDCKGEKETCKNKKDGCEFKDGECLAKVTTA
AGTDGKTNTTGS
NSFVIHKTPLLLAVLLLLA
```

Domain combination: A2

GeneDB Blast results N-terminal domains

Tbg – contig cannot be assembled further in 3' direction

```
>gamb606g06.q1k
>gamb1287f01.q1k
```

```
>kinu1
```

```
MDYKDCYGYLAALTFVVSLCSG
APLTDNDPVAEQISDLCCK
EKFYVETLIQNLKQRVQSTADSIKDMEKLKATWEKAAAATSDAAKKCLFSALEHKADAEL
LRELPKIEEATEAVTTAQIALSEHVGMLGATITLAKTKLDSGSSNHGNADTGSIRIALSR
TTATTDLCNEPATISDIKPGVSEIQPGKLFKLKLTTKPTDLHKHMFVDWLTIGGLKSCTAH
TSYDQNFFDGALSGCCOYSASGTAVATQASTKPPYATAAVTLFKNNDPEQQCEVTDLPAGGA
ADKHKKLQHSLCRALQLGNVKGHSRQLDGAALQSDHVVANTIRNCDPVFQKLTTATDGE
GTTELKKYIKEAYGSSASEFAEKFITNAEKLQMTLRLNDKIETKDLSAVSPGEQTAAALS
HIQGLHNKRELEAGKKSTSAAAVDP
```

```
>Tb09.244.0250
```

```
MDYKDCYGYLAALTFVVSLCSG
APLTDNDPVAEQISDLCCKEKFYVETLIQNLKQRVQSTA
ASIKDMEKLKATWEKAAAATSDAAKKCLFSALEHKADAELLRELPKIEEATEAVTTAQIA
LSEHVGMLGATITLAKTKLDSGSSNHGNAGTGSIRIALSRTTATTDLCNEPATISDIKPG
VSEIQPKLFKLKLTTKPTDLHKHMFVDWLTIGGLTSCTAHTSYDQNFDAALSGCNYKTAA
TAEAKQATTKPTYGTTEVTLFKNNDPEQQCEVTDLPAGGAADKHKKLQHSLCRALQLGNV
NGRSLRQLNGAALQSDHVVANTIRNCDPVFQKLTTATDGEGTTELKKYIKEAYGSSASEF
AEKFITNAEKLQMTLRLNDKIETKDLSAVSPGEQTAAALSHIQGLHNKRELEAGKKSTSAAAVDP
```

```
>gamb1287f01
```

```
MDYKDCYGYLSALTFVASLCSG
APLTDNDPVAEQISDLCCKEKFYVETLIQKLKQRVQSTA
ASIKDMEKLKATWEKAAAATSDAAKKCLFSALEHKADAELLRELPKIEEATEAVTTAQIA
LSEHVGMLGATITLAKTKLDSGSSNHGNANTGSIRIALSRTTATTDLCNEPATISDIKHG
VSEIKADKLFKLKLTTKPTDLHKHMFVDWLTIGGLTSCTAHRSYDQNFDAALSGCQYSASG
KAVATQASTKPTYATAAVTLFKNNDPEQQCEVTDLPAGDSEDK
```

NCBI blast 2 sequence scores for N-terminal domains excluding N-terminal signal sequences

Tororo v Tbg: over 93% over available sequence

Tororo v Tbb 927: 95%

Tbg v Tbb927: 91% over available sequence

Maw-ero 1

```
MHQVLLAAIVLLSPMTKHASG
AIGAGANVAVFEALCGLISLSK
STITPPPPQVKTHLTELATLKKLNMSVSESSWRNLFHKNGAKDQYNDQVPQGVAKGDDWDA
KWSGWKEAEKALEKADTDPDLKAAGLTKPDTKLASAARHAI AALAESAETTAAGALEPPD
VEEEIGGSKLQ TALRTAAFGADVASEQQATIQA AFGGDATQTRSTACEASTAENKAQTVL
AAIACVCM DINSNAVNGACHHKHVAADWSGTSISL PNLQALAAAYCPPRPTKPITAAELR
AAIAAVENKITVFGTDGYLGAAVTS CDGQTANGVCVKFTGYKTNPQTSTANLPWIGK LKT
LAAKLEQREKANQKAQEITSIFKVDIKKVKFIVLEAKQKVEGQTTNTLSAAHAGGQSNVA
GKPC EEQTSNNTTCRADNNCKWDSKEESEGNFCKPKDGEEQTKTAGAGGDEAAAAKEEGKK
CSDKKKEEDCKSPDCKWEGTECKD
SSIIVTNQFALSMVSAAFVALLF
```

Domain combination: B1

GeneDB Blast results N-terminal domains

Tbb 927 best hit >Tb10.v4.0061 33% identity

```
>Mawero1
MHQVLLAAIVLLSPMTKHASGAIG
AGANVAVFEALCGLISLSK
STITPPPPQVKTHLTELATLKKLNMSVSESSWRNLFHKNGAKDQYNDQVPQGVAKGDDWDA
KWSGWKEAEKALEKADTDPDLKAAGLTKPDTKLASAARHAI AALAESAETTAAGALEPPD
VEEEIGGSKLQ TALRTAAFGADVASEQQATIQA AFGGDATQTRSTACEASTAENKAQTVL
AAIACVCM DINSNAVNGACHHKHVAADWSGTSISL PNLQALAAAYCPPRPTKPITAAELR
AAIAAVENKITVFGTDGYLGAAVTS CDGQTANGVCVKFTGYKTNPQTSTANLPWIGK LKT
LAAKLEQREKANQKAQEITSIFKVDIKKVKFIVLEAKQKVEGQTTNTLSAAHAGGQSN
```

```
>gamb1327e07
MISVIRALAVILSPIALVGTVDGTKS
TGANVAPFEALCGIIALAKSAVKVPAVTNSHIEQ
MNKVRKLNMSVTDPTWRIIFHKNGKPNEYNDNVPRGIVSGPDWAEQWKQLKEAEQKLENE
AEGTDFKNSHLNTATPDVKRIIRPRIA AIAARA AIAIKVKS IKEDPQTTEAL TDSKVQ TLLR
KAVFGDDLAD EQGATHQSLFGGTVSLSNRGTVCEASTQANKAQTAGAVIARLFLDSNSDIV
ADIFHHAATGGAHWNAGNPD LGQIQAH IARCRPAVSELLTYDKLIGLITAITTRISLWSG
DGYLGAVTSACSGTTATGSCVKFSGYKANPDQATSALQCLGHLRRLAEQLKDREDMNKKT
EEATRILK
```

NCBI blast 2 sequence scores for N-terminal domains
excluding n-terminal signal sequences

Tororo v Tbg: 42% over available sequence

Tororo v Tbb 927: N/A

Tbg v Tbb927: N/A

Mul 1

MKRKQLVLLTLVAAVAPASG
ALDDNAAEFRALCDVYNLYQQKAKMTEDSNDSMQPTLERLIKANLSTATETYFNNKDGAFCKEGKTDEQSIQKWREEA
NKMVSEEDTKTKAKKSTRMPPSEARSAANKRIMELHRDSENVLEEYKAAVAIAKAARQEAAAALTYAIFGDNNTDLTED
KIPSTTKQMCGNAAGGHADVGGIAHALTCLCSVTTGANNECAKGLTTTGNQGSQTATQTAENWKTIAEACSKRKSSMT
PTPELLEAITVGVEAHLGRQPNGAADGDNAYVLGKLGASSVCDMGTNKACINYKTQMTTGTKDVAWQRNLRKAAMELRK
AADLDRQAKLHKIHLQHNDEAAATILLTAELTSKAASTSTLSPSPRSNQQSAPGEEKCKPQNNSAEECPSEHCNYDTK
TKECKPKKTGSETPAAAGTGDASSGVDCKSHQTQQAACEAENKDVKGQKAVCGWIDYVEGTGKLPKPECRS
SSFLLNEKFALSVVSAFAALLF

Domain combination: B2

GeneDB Blast results N-terminal domains

Tbb 927 closest hit >Tb11.57.0037 36 % identity

Tbg closest hit >gamb761b01.q1k_0 30 % identity

N-terminal domains

>Mul1
MKRQQLVLLTLVAAVAPASGALDDNAAEFRALCDVYNLYQQKAKMTEDSNDSMQPTL
ERLIKANLSTATETYFNNKDGAFCKEGKTDEQSIQKWREEANKMVSEEDTKTKAKKST
RMPPSEARSAANKRIMELHRDSENVLEEYKAAVAIAKAARQEAAAALTYAIFGDNNTD
LTEDKIPSTTKQMCGNAAGGHADVGGIAHALTCLCSVTTGANNECAKGLTTTGNQGS
QTATQTAENWKTIAEACSKRKSSMTPTPELLEAITVGVEAHLGRQPNGAADGDNAYVL
GKLGASSVCDMGTNKACINYKTQMTTGTKDVAWQRNLRKAAMELRKAADLDRQAKLHK
IHLQHNDEAAATILLTAELTSKAASTSTLSPSPRSNQQSAP

NCBI blast 2 sequence scores for N-terminal domains
excluding n-terminal signal sequences

Tororo v Tbg: N/A

Tororo v Tbb 927: N/A

Tbg v Tbb927: N/A

Mul3

```
MNKKILILTVFALAAGKNKIQA
TDELNPAPDFAVLCRIVQHAK
AGFKQASEDEKSVASQLAATTAKLAIINDDDETRYLDKNNRTIYRITGDEPKIPKGAEEK
PLKAKLLQLERRAAELKNDYDNTFAEATRQIEQANDELAEAVYGAGAKFEDQGDSSKLIT
NARASSLFGATGTYNKNCGGTNGGGAASTSNVGITLVSDIYCLCIAGTASAKTCDQTTTA
LSHGTLFASTAGTGKEAFDALMAKCGSQPQOTSPSELHALLIAWQSKLGSFESTNTRDA
ARFIIGRADNPATGCTGAAKQHCVDYKSFLGATPSQDPRWVSKIKSAITKVQTAAEKSLR
LRHSLSLLEVVSQANLTYTEGLNSQRTQRHNAENNAPTQTTETDETCEKKGTDGCKDG
CKWDGEGDNKKCVKDLDYKPKQAEKGESKTGTTNTTGS
NSFIINKGPLLLTFLLF
```

Domain combination: B2

GeneDB Blast results N-terminal domains

Tbg contig assembled by hand

>gamb86e09.p1k

>gamb1296h01.q1k

Sequence adjusted to maintain frame by inclusion of n at 170
Cannot be extended

>Mul3

```
MNKKILILTVFALAAGKNKIQA
TDELNPAPDFAVLCRIVQHAK
AGFKQASEDEKSVASQLAATTAKLAIINDDDETRYLDKNNRTIYRITGDEPKIPKGAEEK
PLKAKLLQLERRAAELKNDYDNTFAEATRQIEQANDELAEAVYGAGAKFEDQGDSSKLIT
NARASSLFGATGTYNKNCGGTNGGGAASTSNVGITLVSDIYCLCIAGTASAKTCDQTTTA
LSHGTLFASTAGTGKEAFDALMAKCGSQPQOTSPSELHALLIAWQSKLGSFESTNTRDA
ARFIIGRADNPATGCTGAAKQHCVDYKSFLGATPSQDPRWVSKIKSAITKVQTAAEKSLR
LRHSLSLLEVVSQANLTYTEGLNSQRTQRHNAENNAPTQTT
```

>tbb927

```
MNKKILILTVFALAAGKNKIQA
TDELNPAPDFAVLCRIVQHAKAGFKQTSSEDEKSAASQL
AAKTAKLAIINDDDETR*LDKNNRTIYRITGKEPKIPTRAETKPLKAKLLEIERRAIELK
TDYDTKSSDSSRQIQQANDDLAEAVYGAGAKFDDQGDTSKLI TNGRASNLFKAGTYAKN
CGGTNGGGTATTSNVGITLVSDIYCLCIAGTATTKTCDQTTTASSHATLFASTAGSGKGS
SDALTAKCGSQPQOTSPSELHALLIAWQSKRGAFHFEFSTNTKDEARFIIG*ATNQATGCTG
ANKQHCVDYKYFLGGTSPQDALWVSKIKSAIGKVQTADEKSLRLRHNL SLLEAISDQANL
TYTESLNSQRTQRDNAENNVP TKQT
```

>tbg

```
KFDDQGDTSKLI TNGRAASLFGPAGTYAKNCGGTNGGGTATTSNVGITLVSDIYCLCIAG
TATTKTCDQTTTASSHETLFASTAGGGKASDALTAKCGSQPQOTSPSELHALLIAWQSK
RGAFHFEFSTNTKDEARFIIG-ATNQATGCTGANKQHCVDYKSFLGGTSPQDALWVSKIKSA
IGKVQTADEKSLRLRHNL SLLEAISDQANLTYTEGLNSQRTQRDNAENNVP TKQT
```

NCBI blast 2 sequence scores for N-terminal domains
excluding n-terminal signal sequences

Tororo v Tbg: 85% over available sequence

Tororo v Tbb 927: 83%

Tbg v Tbb927: 96% over available sequence

Mul 4

```
MTNAKRAFISFKTLFLITTTNLF TAE A
NTAPGDNTAQYKAICMLVNLANKCTKGGQATTTKTDTALVVGAINISLASNDFQQAIDTE
KDWATLPQDKKNKLGSEEDWQFWKESKRQLKKHQAWIKTFTGQPKTAAQRLATAEMTLR
ARKLYEQAQSASATPATSTAQLCKDALYGQGKTSSDGLAAGHATRANVCSDEAAGTNNKA
GTALFWDVLCCLCGGSSTHANTKQACGKGLEASETTRTDWTPNTNDQAAGDPLIQHCATFGG
AGHLNAVTLSEAWAALQTQIKTDSNTNAGKPLVLGAVDASGAFTGCTGNKVTNGGQCVQ
YKAKHFSNGETTIPWLSALRRAAEKAKSDEDNAKKAQQVEKELQLLNDTVIAILVGEKTT
TTAASAEESASKKQEEAEANCNKIEKDTDCKAKPKCAWDGTAKDPNKKCTLSEEAKKE
AAKEAAEKQTGTDGKTN'TTGS
NSFVIKKAPLLLAVLLF
```

Domain combination: B2

GeneDB Blast results N-terminal domains

Tbg contig assembled by hand

```
>gamb248h03.plk rev comp
>gamb381f10.plk
>gamb1357f09.plk rev comp
Reading frame adjusted by insertion of n at 1122
```

>Mul4

```
MTNAKRAFISFKTLFLITTTNLF TAE A
NTAPGDNTAQYKAICMLVNLANKCTKGGQATTT
KTDTALVVGAINISLASNDFQQAIDTEKDWATLPQDKKNKLGSEEDWQFWKESKRQLKK
HQAWIKTFTGQPKTAAQRLATAEMTLRARKLYEQAQSASATPATSTAQLCKDALYGQGKT
SSDGLAAGHATRANVCSDEAAGTNNKAGTALFWDVLCCLCGGSSTHANTKQACGKGLEASE
TRTDWTPNTNDQAAGDPLIQHCATFGGAGHLNAVTLSEAWAALQTQIKTDSNTNAGKPLV
LGAVDASGAFTGCTGNKVTNGGQCVQYKAKHFSNGETTIPWLSALRRAAEKAKSDEDNA
KKAQQVEKELQLLNDTVIAILVGEKTTTTTAASAEESASKKQEQE
```

>Tb927

```
MTNAKRAFIRFKTLFLLTAANLFTAE A
NTAAGDNTAQYKAICMLVNLATKCTKGGQEATT
KTDTALVVGAINISLAGSDFQQAIDTEKEWATLPQANKNKLGGSEEDWQFWKESKKQLKH
HQTWIKTFTAQPNATAQRLATAEMAKRACKLYAQAQTAEQTPAPTAAQLCKDSLYGHGAA
NSDGLKAGHSARENVCSEQAGGPNNKAGTALFWDLLCCLCAGDGQOETTKKACGKLEGSE
TRTAWAPDGNKEAGDPLIEHCAKFGGAGHLTAVTLEAAWAALQTQIKTDSSTNAGKALV
LGAVDSVSGGAFAGCTGNKGTNGGQCVQYKPKHFANGETTIPWLTALRRAAEKVKNDEDN
AKKAQQVEKELQLLNDTVIAILVGEKTTAT-AAPVEESASKKHQE
```

>Tbg

```
MTNAQRAFISFKTLFLITTTNLF TAE D
NTAPGDNTAQYKAICMLVNLVKKCTKGGQANTT
KTDTALVVGAINISLAGSDFQQAIDTEKELATLPQDNKNKLGGNKADWKFWKESKKQLKN
HQTWIKTFAAQPKTAARKRLATVEMAIRARNLYAEANQASATPAAYAAQLCKEALYGQGT
SSYGLKAGHATRENVCSDEAAGTSNKAGSALFWDVLCCLCGGSSTHANTKQACGKGL-ASE
TSTDWTPNSNDQTAGDPLIEHCAKFGGADHLTDVTLSEAWAALQSQIKTDSSTNAGKALV
LGAVDSIGGGAFAGCTGNKVTNGGQCVQYKAMHFANGETTISWLSALRRAAEKVKSDNEDN
AKKAQQVEKGLQPLKDTVIAILVGEKTTATAALVEESASKKQEQE
```

NCBI blast 2 sequence scores for N-terminal domains
excluding n-terminal signal sequences

Tororo v Tbg: 82%

Tororo v Tbb 927: 80%

Tbg v Tbb927: 79%

CLUSTAL W (1.83) multiple sequence alignment

```
Mul4      MTNAKRAFISFKTLFLITTTNLFTAANTAPGDNTAQYKAICMLVNLANKCTKGGQATTT 60
TbgDal    MTNAQRAFISFKTLFLITTTNLFTAEDNTAPGDNTAQYKAICMLVNLVKKCTKGGQANTT 60
Tbb927    MTNAKRAFIRFKTLFLLTAANLFTAANTAAAGDNTAQYKAICMLVNLATKCTKGGQEATT 60
          ****:*  *****:*:*:*****  ***.*****.*****.***** **

Mul4      KTD TALVVGAINISLASNDFQQAIDTEKDWATLPQDKKNKLGSEEDWQFWKESKRQLKK 120
TbgDal    KTD TALVVGAINISLAGSDFQQAIDTEKELATLPQDNKNKLGKNKADWKFWKESKKQLKN 120
Tbb927    KTD TALVVGALNISLAGSDFQQAIDTEKEWATLPQANKNKLGGSEEDWQFWKESKKQLKH 120
          *****:*  *****.*****: ***** :*****.: **:*****:***:

Mul4      HQAWIKTFTGQPKTAAQRLATAEMTLRARKLYEQAQSASATPATSTAQLCKDALYGQGKT 180
TbgDal    HQTWIKTFAAQPKTAAKRLATVEMAIRARNLYAEANQASATPAAYAAQLCKEALYGQGT 180
Tbb927    HQTWIKTFTAQPNATAQRLATAEMAKRAKLYAQAQTAEQTPAPTAAQLCKDSL YHGAA 180
          **:*****:.*:***:***.***: **:* ** :*: *. ***. :*****:***:* :

Mul4      SSDGLAAGHATRANVCSDEAAGTNNKAGTALFWDVLCCLCGGSSTHANTKQACGKGLSE 240
TbgDal    SSYGLKAGHATRENVCSDEAAGTSNKAGSALFWDVLCCLCGGSSTHANTKQACGKGL-ASE 239
Tbb927    NSDGLKAGHSARENVCSEQAGGPNNKAGTALFWDLLCCLCAGDGOQETTKKACGKGLEGE 240
          .* ** ***:*  *****:*.*.*****:*****:*****.*.. : .**:*  ***** .**

Mul4      TRTDWTPNTNDQAAGDPLIQHCATFGGAGHLNAVTLSEAWAALQTQIKTDSNTNAGKPLV 300
TbgDal    TSTDWTPNSNDQTAGDPLIEHCAKFGGADHLTDVTLSEAWAALQSQIKTDSSTNAGKALV 299
Tbb927    TRTAWAPDGNKKEAGDPLIEHCAKFGGAGHLTAVTLSEAWAALQTQIKTDSSTNAGKALV 300
          * * *:*  **: *****:***.***.*. *****:*****:*****.*****.**

Mul4      LGAVDASG-AAFTGCTGNKVTNGGQCVQYKAKHFSNGETTIPWLSALRRAAEKAKSDEDN 359
TbgDal    LGAVDSIGGAFAGCTGNKVTNGGQCVQYKAMHFANGETTISWLSALRRAAEKVKSDEDN 359
Tbb927    LGAVDSVSGGAFAGCTGNKGTNGGQCVQYKPKHFANGETTIPWLTALRRAAEKVNDEDN 360
          *****: . .**:*  ***** *****. **:*****.**:*****.*.*****

Mul4      AKKAQQVEKELQLLNDTVIAILVGEKTTTTTAASAEESASKKQOE 404
TbgDal    AKKAQQVEKGLQPLKDTVIAILVGEKTTAT-AALVEESASKKQOE 403
Tbb927    AKKAQQVEKELQLLNDTVIAILVGEKTTAT-AAPVEESASKKHQE 404
          ***** ** *:*  *****:* ** .*****:***
```