Supplemental Figures for Linford A. et al. Crystal structure and putative substrate identification for the *Entamoeba histolytica* low molecular weight tyrosine phosphatase.



Supplemental Fig. 1. Schematic representation of epitope-tagged putative substrate genes subcloned into pGIR209. Selected putative substrate genes identified via mass spectrometry (Table 1) were subcloned into pGIR209 with epitope tags E, S, and KT3 added. 5' actin and 3' actin, 5' and 3' flanking regions of the *Entamoeba histolytica* actin gene; *Neo*, neomycin resistance gene; *luc*, luciferase gene; 5' *hgl* and 3' *hgl*, 5' and 3' flanking regions of the *E. histolytica hgl* gene; tetO, tetracycline operator; Putative Substrate, genes for putative substrate proteins of the LMW-PTP identified by mass spectrometry. N-tag, N-terminally added E-S-KT3 epitope tag; C-tag, C-terminally added E-S-KT3 epitope tag. As a control for transfection, drug selection, and expression of an epitope-tagged protein in amebae, URE3-BP was also cloned and expressed. The pGIR209 diagram is adapted from Ramakrishnan G, Vines RR, Mann BJ, Petri WA Jr. A tetracycline-inducible gene expression system in *Entamoeba histolytica*. Mol Biochem Parasitol 1997; 84(1):93-100, and Katz U, Bracha R, Nuchamowitz Y, Milstein O, Mirelman D. Comparison between constitutive and inducible plasmid vectors used for gene expression in *Entamoeba histolytica*. Mol Biochem Parasitol 2003; 128(2):229-33.

Ent his 1	1		54
Ent his 2	1		54
Ent diepar	1		54
Ent dispar	1		54
Cia lam P15	1		54
Giard intee	1		54
Dictvo diec	1		63
Klu lactie	1		59
Sacch cerev	1		62
Sabir pombo	1		50
Sorg bicol	97	MANTAL PLATE AND	168
Zon move 1	97	MAREAS IADARARARARAFIA UF VEDONI CASTAREAVIA ILVSA ROJUSATI - IDSAUTI GINEGUNAJSAN-	160
Zea mays 1	1	MAREASIAGARSAREARFFAYERYENGAGAICOSFAREAYFALLYSKAGESARE - IDSAGIIGIRGANADSAN-	72
Deven ant	0 E	MAREAS IAGARSAREAR FFAVER VEDENICKS FAREAVER IN SKRUDSKET - IDSAGTIGTEGENAS SKRU	12
Hord mild	00	margasarareda - garff supervised on tensor and supervised to the supervised to the supervised to the supervised on tensor and supervised on the supervised	1 1 0
Apph thal	1	MAREAS IPAPALA - SARPIAVLEV CLORICKSPAALAVERNIVSKRUKSKE - IDSAUTIGINEGKNAJSKE -	£0
Arab thai	1	MATPPTTQ-SSETKPISVLFVCLGNICKSPAAEGVFROIVKRGLDSKFN-IDSAGTIDIHEGRMADPKM-	68
Vit Vinir Dem tricke	1	MASGESIDIQSREESULEVCLORICKSPAALGVETDAVRKRGLDSRER-IDSAGTINTHEGRLADERM-	67
Pop trieno	1	MASS-TEIRFFAVLEVELGENICKSPAAEGVIDIVERGEDSER- DOAGTINIEGENPADSE	63
Ricin comm	1	MASSTETETRYFSULFVCLGNTCKSPAAEGVFROTVKNSLDSKFN-TDSAGTINTHEGRPADFKM-	55
Nost punct	1	MTIDRGQKTSWPYKLLFVCLGNICRSPSAENIMNHLTEQAGLSDAII-CDSAGTSSHVGSPPDKKM-	66
Lyngbya sp	1	MPYQLLFVCLGNICRSPSAENIMNHQIELAHLGRSIL-CDSAGTSSYHIGSPPDARM-	56
Cyano sp	1		56
Arthro max	1	MSIRLEFVCIGNICRSPSAENIMNHLITEEANLSDRIV-CDSAGTSGTHIGSAPDPRM-	56
Prochio mar	1	MIRKVLFICLGNICRSPAAEAIFLHHLREIDLINEFI-VDSAGTGGWHVGKKADSRM-	56
Bdello bact	1	MTRONNSPRIIOMKSORLEFVEIGNICRSPTAEAVAAHLIRORDEPWV-VDSAGTSGAHDGEMADPRS-	69
Bact caccae	1	IDSAGILSYHOGELPDSRM-	55
Bact theta	1	IDSAGILSYHQGELPDSRM-	55
Bact fragil	1	IDSAGILAYHQGELPDSRM-	56
Parab merd	1	MMEEKKGEYKILFVCLGNICRSPSAEAVMKKLVODAGLDGRIK-IDSAGIIGYHAGEKADPRM-	62
Vibrio chol	1	IDSAGTIGYHQGNPPDARS-	55
Pseudo aeru	1	MRVLFVCLGNICRSPTAEGVFRRKVEEAGLASRIH-VDSAGTAGWHAGKAPDQRT-	54
Homo isof b	1	MAEQATKS <mark>VLFVCLGNICRS</mark> PIAEAVFRKLVTDQNISENWV-IDSGAVSDWNVGRSPDPRA-	60
Homo isof c	1	MAEQATKS <mark>VLFVCLGNICRS</mark> PIAEAVF <mark>RKLVTDQNISE</mark> NWR-V <mark>DSA</mark> ATSG <mark>Y</mark> EIGNPPDYRG-	60
Mus musc 1	1	MAEVGSKS <mark>VLFVCLGNICRSPIAEAVFRKLVTD</mark> EK <mark>VSD</mark> NWA-I <mark>DSS</mark> AVSDWNVGRPPDPRA-	60
Mus musc 2	1	MAEVGSKS <mark>VLFVCLGNICRSPIAEAVFRKLVTD</mark> EKVSDNWR-IDSAATSTYEVGNPPDYRG-	60
Bos taurus	1	MAEQVTKS <mark>VLFVCLGNICRS</mark> PIAEAVF <mark>RKLVTDQNISDNW</mark> R-I <mark>DSAAT</mark> S <mark>TY</mark> ELGNPP <mark>DCR</mark> G-	60
Xenop laev	1	MAQQGNGKS <mark>VLFVCLGNICRSPIAEAVFQKLVTD</mark> AG <mark>ISKEW</mark> S-I <mark>DSA</mark> ATSDWNVGSSP <mark>DSR</mark> A-	61
Gal gallus	1	MAAGEVKS <mark>VLFVCLGNICRS</mark> PI <mark>AEAVFRKLVTD</mark> EKVENKWR-IDSAA <mark>T</mark> STYEIGNPPDYRG-	60
Dros melan	1	DSAAIGGWHVGNRADPRA-	55
Danio rerio	1	MAASSGKS <mark>VLFVCLGNICRS</mark> PIAEAVFRKMATDSGVVDKWV-IDSGATSDWNTGSTP <mark>DAR</mark> G-	60
Ent his 1	100	EQ-YKQKIFKMVDFCTTIKTTEVPOTTGGEKGFHRVIDILEDACENLIIKLEEGKLIN	157
Ent his 1 Ent his 2	100 100	EQ-YKQKIFKMVDFCTTIKTTEVPDDYYGGEKGFHRVIDILEDACENLIIKLEEGKLIN EQ-YKQKIFKMVDFCTTIKTTEVPDDYYGGEKGFHRVIDILEDACENLIIKLEEGKLIN	157 157
Ent his 1 Ent his 2 Ent dispar	100 100 100	EQ-YKQKIFKMVDFCTTIKTTEVPDFYMGGEKGFHRVIDILEDACENLIIKLEEGKLIN EQ-YKQKIFKMVDFCTTIKTTEVPDFYMGGEIGFHRVIDILEDACENLIIKLEEGKLIN EO-YKQKIFKMVDFCTTIKTTEVPDFYMGGEIGFHRVIDILEDACENLIIKLDEGKLIN	157 157 157
Ent his 1 Ent his 2 Ent dispar Ent invad	100 100 100 100	EQ-YKQKIFKMVDFCTTIKTTEVPDFYMGGEKGFHRVIDILEDACENLIIKLEEGKLIN EQ-YKQKIFKMVDFCTTIKTTEVPDFYMGGEKGFHRVIDILEDACENLIIKLEEGKLIN EQ-YKQKIFKMVDFCTTIKTTEVPDFYMGGEIGFHRVIDILEDACENLIIKLEEGKLIN KE-YONKIFKMVDFCTTIKTTEVPDFYMGGLEGFHRVIDILEDQCENLIDQLENAGDE	157 157 157 156
Ent his 1 Ent his 2 Ent dispar Ent invad Gia lam P15	100 100 100 100 100	EQ-YKQKIFKMVDFCTTIKTTEVPDFY GGEKGFHRVIDILEDACENLIIKIEEGKLIN -EQ-YKQKIFKMVDFCTTIKTEVPDFY GGEIGFHRVIDILEDACENLIIKIDEGKLTN -KE-YQDKIFKMVDFCTTIKTE-VPDFY KE-YQDKIHKMVEYCKISKDSEVPDFY KE-YQDKIHKMVEYCKISKDSE-VPDFY CGR-PELVYSKVEYCRSIDRINSGKAL	157 157 157 156 154
Ent his 1 Ent his 2 Ent dispar Ent invad Gia lam P15 Giard intes	100 100 100 100 100 100	EQ-YKQKIFKMVDFCTTIKTTEVPDFYMGGEKGFHRVIDILEDACENLIIKLEEGKLIN EQ-YKQKIFKMVDFCTTIKTTEVPDFYMGGEKGFHRVIDILEDACENLIIKLEEGKLIN EQ-YKQKIFKMVDFCTTIKTEVPDFYMGGEIGFHRVIDILEDACENLIIKLDEGKLIN KE-YQQKIFKMVEYCKKSKDEVPDFYMGGEIGFHRVIDILEDACENLIIKLDEGKLIN KR-DQSKVRMFRDVDFSGK-GDVPDFYMGGR-FELVYSNVERTGRSIDDILENSGKL KR-DDSKINFRDVDPSGK-GDVPDFYMGGR-FELVYSNVERTGRSIDDILKSGKL	157 157 157 156 154 154
Ent his 1 Ent his 2 Ent dispar Ent invad Gia lam P15 Giard intes Dictvo disc	100 100 100 100 100 100	EQ-YKQKIFKMVDFCTTIKTTEVPDFYMGGEKGFHRVIDILEDACENLIIKLEEGKLIN EQ-YKQKIFKMVDFCTTIKTTEVPDFYMGGEKGFHRVIDILEDACENLIIKLEEGKLIN EQ-YKQKIFKMVDFCTTIKTTEVPDFYMGGEIGFHRVIDILEDACENLIIKLEEGKLIN KE-YQDKIFKMVDFCTTIKTEVPDFYMGGLFGFKVIDILEDCGKNLLDQLENAGDE KR-DQSKVRMFRDYDPEGK-GDVPDFYMCGRFELVYSVWERTCRSIDDIKNGKAL KR-DQSKIRMFRDYDPEGK-GDVPDFYMGGR-FELVYSVWERTCRSIDDIKNGKAL KR-DRSKIRMFRDYDPEGK-GDVPDFYMGGR-FELVYSVWERTCRSIDDIKNGKAL DN-HINTIKRGFWTH-KNN-VEDFYMGGNSNFMICFNWHCINVTKKIENAM	157 157 157 156 154 154 179
Ent his 1 Ent his 2 Ent dispar Ent invad Gia lam P15 Giard intes Dictyo disc Klu lactis	100 100 100 100 100 100 126	EQ-YKQKIFKMVDFCTTIKTTEVPDFYMGGEKGFHRVIDILEDACENLIIKLEEGKLIN EQ-YKQKIFKMVDFCTTIKTTEVPDFYMGGERGFHRVIDILEDACENLIIKLEEGKLIN EQ-YKQKIFKMVDFCTTIKTEVPDFYMGGERGFHRVIDILEDACENLIIKLEGKLITN KE-YQDKIHKMVEYCKKSKDEEVPDFYMGGERGFHRVIDILEDACENLIIKLEGKGKIN KR-DQSKVRMFPDYDFSGK-GDVPDFYMGGERGFHRVIDILEDACENLIIKLKSGKAL KR-DQSKIKFNDYDFSGK-GDVPDFYMGGR-FELVYSNVERTGRSIDDNIKSGKAL D-N-HIATIKRLGENTHI-KKINVEDFYMGGNEGFEVPKOVWFSKEPUFELES KR-DQSK-IKKENNTDNKFOD-IVDDFYMGGNEGFEVPKOVWFSKEPUFELES	157 157 157 156 154 154 179 157
Ent his 1 Ent his 2 Ent dispar Ent invad Gia lam P15 Giard intes Dictyo disc Klu lactis Sacch cerev	100 100 100 100 100 126 105	EQ-YKQKIFKMVDFCTTIKTTEVPDFYYGGEKGFHRVIDILEDACENLIIKLEEGKLIN EQ-YKQKIFKMVDFCTTIKTTEVPDFYYGGEKGFHRVIDILEDACENLIIKLEEGKLIN EQ-YKQKIFKMVDFCTTIKTTEVPDFYYGGEIGFHRVIDILEDACENLIIKLGEGKITN KE-YQDKIFKMVEYCKKSKDSEVPDFYYGGLKGFCKVIDMLEDGCKNLDQLENAGDE KR-DQSKVRMFRDYDPEGK-GDVPDFYYCGRFELYYSNVERTCRSILDRIKSGKAL KR-DRSKIRMFRDYDFEGK-GDVPDFYYCGRFELYYSNVERTCRSILDRIKSGKAL KB-DRSKIRMFRDYDFEGK-GDVPDFYYGGN-FELYYSNVERTCRSILDRIKSGKAL DN-HIATIKMIGEYHTH-KINVEDFYYGGNSNFNICFMNVHGLVNFLKEIESLN KE-PKAQLSLFGKNNTNOKFQD-IVDBFYGGNDGFEYHFKQVWFSGPTEKEN	157 157 156 154 154 179 157 161
Ent his 1 Ent his 2 Ent dispar Ent invad Giard intes Dictyo disc Klu lactis Sacch cerev Schiz nombe	100 100 100 100 100 126 105 108	EQ-YKQKIFKMVDFCTTIKTTEVPDFYKGGEKGFHRVIDILEDACENLIIKLEEGKLIN EQ-YKQKIFKMVDFCTTIKTTEVPDFYKGGEKGFHRVIDILEDACENLIIKLEEGKLIN -EQ-YKQKIFKMVDFCTTIKTEVPDFYKGGEXGFHRVIDILEDACENLIIKLEGKLIN KE-YQDKIHKMVEYCKKSKDSEVPDFYKGGEXGFHRVIDILEDACENLIIKLEKAGGE KR-DQSKVRMFRDYDFCGK-GDVPDFYKGGEXGFHRVIDILEDGCKNLDQLENAGGE KR-DQSKVRMFRDYDFGGK-GDVPDFYKCGRFELVYSVERTCRSILDRIKNGKAL D-N-HIATIKKLGYHTH-KKINVEDFYKGGNREFFLVSVERTCRSILDRIKNGKAL KR-QSAKVLIFGDNTNDGVQTIEDFWKGGNEGFEVFKQUTYFSCPLELES EG-SKAKVLIFGDNTNDGVQTIEDFWKGGSDGFCNVCYUTVFSKQFLKKEL D-G-GSAKVLIFGDNTNDGVQTIEDFWKGGSDGFCNVCYUTVFSKQFLKKEL	157 157 156 154 154 179 157 161
Ent his 1 Ent his 2 Ent dispar Ent invad Gia lam P15 Giard intes Dictyo disc Klu lactis Sacch cerev Schiz pombe Sarg bicol	100 100 100 100 100 126 105 108 105	EQ-YKQKIFKMVDFCTTIKTTEVPDYYGGEKGFHRVIDILEDACENLIIKLEEGKLIN EQ-YKQKIFKMVDFCTTIKTTEVPDYYGGEIGFHRVIDILEDACENLIIKLEEGKLIN EQ-YKQKIFKMVDFCTTIKTTEVPDYYGGEIGFHRVIDILEDACENLIIKLDEGKLIN KR-DQSKVRMFDVDFSCK-GDVPDYYGGEIGFHRVIDILEDACENLIIKLBAGDE KR-DGSKIKMFDVDFSCK-GDVPDYYGGR-FELVYSWERTCRSIDDIKNGKAL C-R-DSKIKMFDVDFSCH-GD-VPDYYGGR-FELVYSWERTCRSIDDIKNGKAL D-N-HIATIKELGEYNHH-KKINVEDYYGGRNEGFEYNFKQVYFSCPILES KR-DSKISLFGKNNTDNKFQD-IVDDYYGGRNEGFEYNFKQVYFSCPILES KR-DSKVKMFFDNTFSCVFKUVDDYYGGNEGFEXNFKQVYFSCPILES G-GSRKVLFGENNTROGVQTIEDJWGGNEGFEXNFKQVYFSCPILKSIA G-SRKVVLFGEYNSFOVSKIVDDYYGGDGFGCVIDUFDACFLIDSIA 	157 157 156 154 154 179 157 161 156 279
Ent his 1 Ent his 2 Ent dispar Ent invad Gia 1am P15 Giard intes Dictyo disc Klu lactis Sacch cerew Schiz pombe Sorg bicol Zaa mare 1	100 100 100 100 126 105 108 105 223 215	EQ-YKQKIFKMVDFCTTIKTTEVPDFNGGEKGFHRVIDILEDACENLIIKIEEGKLIN EQ-YKQKIFKMVDFCTTIKTTEVPDFNGGEIGFHRVIDILEDACENLIIKIEEGKLIN -EQ-YKQKIFKMVDFCTTIKTEVPDFNGGEIGFHRVIDILEDACENLIIKIDGKLIN KFYQDKIFKMVDFCTTIKTE-VPDFNGGEIGFHRVIDILEDACENLIIKIDGKLIN KR-DGKVENFDDFCTIKTE-VPDFNGGEIGFHRVIDILEDACESILDRIKGKAL KR-DGKIFMFDDFCDFGK-GDVPDFNGGR-FELVYSIVERTGRSILDRIKGKAL KR-DGKIFMFDDFCGK-GDVPDFNCGRFELVYSIVERTGRSILDRIKGKAL KR-DRSKIFMFDDFCG-GD-VPDFNGGNEGFEVFKQVNFSCFLELES ES-SKAKVLFGDNNTNDGTVGTIEDIWGDIDFENFKQVTFSCFLELES ES-SKAKVLFGDNNTNDGTVGTIEDIWGJSDGFCVIDILEDACESILDRISISA CS-SKAKVLFGCNSTSSEC-VPDFNGGSDGFCVIDILEDACESILDSIVSISA CN-APKVKLMCSYCKRTESEVPDFNGGSDGFCVIDILEDACESILDSIVSISA	157 157 156 154 154 179 157 161 156 279 274
Ent his 1 Ent his 2 Ent dispar Ent invad Gia 1am P15 Giard intes Dictyo disc Klu lactis Sacch cerev Schiz pombe Sorg bicol Zea mays 1 Zea mays 2	100 100 100 100 100 126 105 108 105 223 215	EQ-YKQKIFKMVDFCTTIKTEVPDYYGGEKGFHRVIDILEDACENLIIKLEEGKLIN EQ-YKQKIFKMVDFCTTIKTEVPDYYGGEIGEHRVIDILEDACENLIIKLEEGKLIN EQ-YKQKIFKMVDFCTTIKTEVPDYYGGEIGEHRVIDILEDACENLIIKLEGGKLT KR-DQSKVRMFDVDFCTTIKTE-VPDYYGGEIGEHRVIDILEDACENLIIKLEGGKLT KR-DQSKVRMFDVDFCGK-GDVPDYYGGEIGEHRVIDILEDACENLIIKLEGGKLT KR-DGSKVRMFDVDFGG-GDVPDYYGGR-FELVSNVERTGRSIDDIKNGKAL D-N-HIATIKRLGENTH-KKINVEDYYGGNFGFEVVIDILEDACESIDDIKNGKAL KR-DGSKVLFGCNNTDNKFQD-IVDDYYGONSNFNICFNIVHECLVNFLKEIESLN KR-DGSKVLFGCNNTDNKFQD-IVDDYYGONSGFEVFEVEVFFEQUELES EG-SKARVCLFGDNNTNDGFVQTILEDIW GGEGGFEVFUDILEDACESIDDIVASISG CS-APKKVKLMGSYCKGHTSEVPDYYGGPGGFEKVLDILEDACESIDDIVASISG DS-APKKVKLMGSYCKGHTSEVPDYYGGPGGFEKVLDILEDACESIDDSIVASISG D-S-APKKVKLMGSYCKGHTSEVPDYYGGPGGFEKVLDILEDACESIDDSIVASISG	157 157 156 154 154 179 157 161 156 279 274
Ent his 1 Ent his 2 Ent dispar Ent dispar Ent invad Gia 1am P15 Giard intes Dictyo disc Klu lactis Sacch cerev Schiz pombe Sorg bicol Zea mays 1 Zea mays 2 Oruza est	100 100 100 100 100 126 105 108 105 223 215 127 210	EQ-YKQKIFRMVDFCTTIKTTEVPDFYGGEKGFHRVIDILEDACENLIIKLEEGKLIN EQ-YKQKIFRMVDFCTTIKTTEVPDFYGGEIGFHRVIDILEDACENLIIKLEEGKLIN EQ-YKQKIFRMVDFCTTIKTTEVPDFYGGEIGFHRVIDILEDACENLIIKLEEGKLIN KR-DQSKVRMFDVDFCTIKTTE-VPDFYGGLGFHRVIDILEDACENLIIKLERAGDE KR-DQSK-VRMFDVDFSG-GDVPDFYGGR-GEVYSWERTGRSIDNIRAGGE KR-DSSKIRMFDVDFSG-GD-VPDFYGGR-FELVYSWERTGRSIDNIKAGKAL D-N-HIATIKRLGEYHTH-KKIN-VEDFYGGN-FELVYSWERTGRSIDNIKAGKAL ER-DSKISJFCKNNTDNFFQD-IVDDFYGGNSGFEVFRQVWYFSQFLELS C-G-SRAKVLFGEVASFQVSKIVDDFYGSDGFGCCYIQIVDFSQNFLKSIA ES-RAKVLFGEVASFQVSKIVDDFYGSDGFGCCYIQIVDFSQNFLKSIA EN-APDKVKLMGSYCKRHTSSEVPDFYGGOGGFEKVIDILEDACESLDSIVANAASISG -DS-APKKVKLMGSYCKRHTSSEVPDFYGGOGGFEKVIDILEDACESLDSIVANAASISG D-S-SAPKKVKLMGSYCKRHTSEVPDFYGGOGGFEKVIDILEDACESLDSIVANAASISG	157 157 156 154 154 179 157 161 156 279 274 186 268
Ent his 1 Ent his 2 Ent dispar Ent invad Gia 1am P15 Giard intes Dictyo disc Klu lactis Sacch cerev Schiz pombe Sorg bicol Zea mays 1 Zea mays 2 Oryza sat Hord wuld	100 100 100 100 100 126 105 223 215 127 210 203	EQ-YKQKIFKMVDFCTTIKTTEVPDFYGGEKGFHRVIDILEDACENLIIKIEEGKLIN EQ-YKQKIFKMVDFCTTIKTTEVPDFYGGETGFHRVIDILEDACENLIIKIEEGKLIN EQ-YKQKIFKMVDFCTTIKTEVPDFYGGETGFHRVIDILEDACENLIIKIEGKLIN KE-YQDKIHKMVETCHKSKDSEVPDFYGGETGFHRVIDILEDACENLIIKIEKAGKAL KR-DQSKVRMFRDYDFGGK-GDVPDFYGGR-FELVSNVERTCRSILDRIKNGKAL KR-DQSKVRMFRDYDFGGK-GDVPDFYGGR-FELVSNVERTCRSILDRIKNGKAL N-HIATIKRLGEYHTH-KKINVEDFYGGNRGFEYHDILEDCFPLEES KR-QCSKA-VLFGGNNTDNKTQOTIEDIWGDIGDFEYHFKQUTFFSCPLELES EG-SKAKVLFGGNNTDNKTQOTIEDIWGGSDGFGDCYIQLVDFSQNLKSIA D-S-APKKVKLMCSYCKHTESEVPDFYGGSQGFEKVIDILEDACESILDSIVASISG -D-S-APKKVKLMCSYCKHTESEVPDFYGGPQGFEKVIDILEDACESILDSIVANNASISG -D-S-S-PKKVKLMCSYCKHTESEVPDFYGGPQGFEKVIDILEDACESILDSIVANNASISG -D-S-SPKKVKLMCSYCKHTESEVPDFYGGPQGFEKVIDILEDACESILDSIVANNASISG -D-S-SPKKVKLMCSYCKHTESEVPDFYGGPQGFEKVIDILEDACESILDSIVANNASISG	157 157 156 154 154 179 157 161 156 279 274 186 268 262
Ent his 1 Ent his 2 Ent dispar Ent invad Gia lam P15 Giard intes Dictyo disc Klu lactis Sacch cerew Schiz pombe Sorg bicol Zea mays 1 Zea mays 2 Oryza sat Hord vulg arab thal	100 100 100 100 126 105 223 215 127 210 203	EQ-YKQKIFRMVDFCTTIKTEVPDYYGGEKGFHRVIDILEDACENLIIKLEEGKLIN EQ-YKQKIFRMVDFCTTIKTEVPDYYGGEGEGHRVIDILEDACENLIIKLEEGKLIN EQ-YKQKIFRMVDFCTTIKTE-VPDYYGGEIGHRVIDILEDACENLIIKLEGGKLT KR-DQSKVRMFDVDFCTTIKTE-VPDYYGGEIGHRVIDILEDACENLIIKLEGGKLT KR-DGSKINMFDVDFCTTIKTE-VPDYYGGEIGHRVIDILEDACESILDILNGGKL KR-DSKINMFDVDFCG-GD-VPDYYGGNCGFVIDURETGRSIDNIKNGKAL DN-HIATIKRLGFNTH-KKINVEDYYGGNCGFENFKQIVYSETGRSIDNIKNGKAL DN-HIATIKRLGFNTH-KKINVEDYYGGNCGFENFKQIVYSEGFLELES KR-DSKINMFGVDFGVGITEDIWGMSNFNICFNIVHDCLVNFLKEI-ESLN KG-SKAKVULFGDNTNGFUQTIEDIWGGNCGFENFKQIVYSEGFLELES GG-SKAKVULFGUNTDNGFVQTIEDIWGGNCGFENVIDILEDACESLLDSIVANNASISG -DS-APNKVKLMCSYCKQHTESE-VPDYYGGPQGFEKVIDILEDACESLLDSIVANNASISG -DS-APNKVKLMCSYCKQHTESE-VPDYYGGPQGFEKVIDILEDACESLLDSIVANNASISG -DS-APNKVKLMCSYCKQHTESE-VPDYYGGPQGFEKVIDILEDACESLLDSIVANNASYFA S-SNKKVKLMCSYCKRHTESE-VPDYYGGPQGFEKVIDILEDACESLLDSIVANNASYFA S-CMKVKLMCSYCKRHTESE-VPDYYGGPQGFEKVIDILEDACESLLDSIVANNASYFA	157 157 156 154 154 157 161 156 279 274 186 268 268 268
Ent his 1 Ent his 2 Ent dispar Ent invad Gia 1am P15 Giard intes Dictyo disc Klu lactis Sacch cerew Schiz pombe Sorg bicol Zea mays 1 Zea mays 2 Oryza sat Hord vulg Arab thal Vit winif	100 100 100 100 100 105 105 223 215 127 210 203 123	EQ-YKQKIFKMVDFCTTIKTTEVPDFYGGEKGFHRVIDILEDACENLIIKIEEGKLIN EQ-YKQKIFKMVDFCTTIKTTEVPDFYGGEIGHRVIDILEDACENLIIKIEEGKLIN EQ-YKQKIFKMVDFCTTIKTEVPDFYGGEIGHRVIDILEDACENLIIKIEGKLIN KE-YQDKIHKMVEYCKKSKDSEVPDFYGGEIGHRVIDILEDACENLIIKIEGKKLI KE-YQDKIHKMVEYCKSKDSEVPDFYGGR-FELYYSIVERTCRSILDRIKGGKAL KR-DRSKVKLHCSYKRNTDSEVO-IVDDFYGGR-FELYYSIVERTCRSILDRIKGGKAL KE-YQDKISKKNTDNKFQDIPDFGKGDVPDFYGGREGFEVFEVFEVFECFELES ES-SKAEVLFGDNNTNDGVQTIEDWGDIGDEFYFRQITFSQFLKKEL CS-SKAEVLFGDNNTNDGVQTIEDWGSIGGFEVIDLEDACESLLDSIVASISG D-S-ARKVKLHCSYCKRHTESEVPDFYGGPGFEKVLDLEDACESLLDSIVASISG -D-S-SARK-VKLHCSYCKRHTESEVPDFYGGPGFEKVLDLEDACESLLDSIVANNASISG -D-S-SARK-VKLHCSYCKRHTESEVPDFYGGPGFEKVLDLEDACESLLDSIVANNASISG -D-S-SARK-VKLHCSYCKRHTESEVPDFYGGPGFEKVLDLEDACESLLDSIVANNASISG -D-S-SARK-VKLHCSYCKRHTESEVPDFYGGPGFEKVLDLEDACESLLDSIVANNASISG -D-S-SARK-VKLHCSYCKRHTESEVPDFYGGPGFEKVLDLEDACESLLDSIVANNASISG -D-S-SARK-VKLHCSYCKRHTESE-VPDFYGGPGFEKVLDLEDACESLLDSIVANNASISG -D-S-SARK-VKLHCSYCKRHTESE-VPDFYGGPGFEKVLDLEDACESLLDSIVANNASISG -D-S-SARK-VKLHCSYCKRHTESE-VPDFYGGPGFEKVLDLEDACESLLDSIVANNASISG -D-S-SARK-VKLHCSYCKRHTESE-VPDFYGGPGFEKVLDLEDACESLLDSIVANNASISG -D-S-SARK-VKLHCSYCKRHTESE-VPDFYGGPGFEKVLDLEDACESLLDSIVANNASISG -D-S-SARK-VKLHCSYCKRHTESE-VPDFYGGPGFEKVLDLEDACESLLDSIVANNASISG -D-S-SARK-VKLHCSYCKRHTESE-VPDFYGGPGFEKVLDLEDACESLLDSIVANNASISG -D-S-SARK-VKLHCSYCKRHTESE-VPDFYGGPGFEKVLDLEDACESLLDSIVANNASISG -D-S-SARK-VKLHCSYCKRHTESE-VPDFYGGPGFEKVLDLEDACESLLDSIVANNASISG -D-S-DRK-VKLHCSYCKRHTESE-VPDFYGGPGFEKVLDLEDACESLLDSIVANNASISG -D-S-DRK-VKLHCSYCKRHTESE-VPDFYGGPGFEKVLDLEDACESLLDSIVANNASISG -D-S-DRK-VKLHCSYCKRHTESE-VPDFYGGPGFEKVLDLEDACESLLDSIVANNASISG -D-S-DRK-VKLHCSYCKRHTESE-VPDFYGGPGFEKVLDLEDACESLLDSIVANNASISG -D-S-DRK-VKLHCSYCKRHTESE-VFDFYGGPGFEKVLDLEDACESLLDSIVANNASISG -D-S-DRK-VKLHCSYCKRHTESE-VFDFYGGPGFEKVLDLEDACESLLDSIVANNASISG -D-S-DRK-VKLHCSYCK	157 157 156 154 154 179 157 161 156 279 274 186 268 262 172
Ent his 1 Ent his 2 Ent dispar Ent invad Gia lam P15 Giard intes Dictyo disc Klu lactis Sacch cerev Schiz pombe Sorg bicol Zea mays 1 Zea mays 2 Oryza sat Hord vulg Arab thal Vit vinif	100 100 100 100 100 126 105 223 215 127 210 203 123 123	EQ-YKQKIFRMVDFCTTIKTEVPDYYGGEKGFHRVIDILEDACENLIIKLEEGKLIN EQ-YKQKIFRMVDFCTTIKTEVPDYYGGEIGEHRVIDILEDACENLIIKLEEGKLIN EQ-YKQKIFRMVDFCTTIKTEVPDYYGGEIGEHRVIDILEDACENLIIKLEGGKIT KR-DQSKVRMFDVDFCTTIKTE-VPDYYGGEIGEHRVIDILEDACENLIIKLQGGKIT KR-DQSKVRMFDVDFCTTIKTE-VPDYYGGEIGEHRVIDILEDACENLIIKLEGGKIT KR-DQSKVRMFDVDFGC-GD-VPDYYGGR-GEVYSVERTCRSIDDIKNGKAL D-N-HIATIKRLGENTH-KKINVEDYYGGNFGEFVYSVERTCRSIDDIKNGKAL CG-SKAKVCLFGDMNTNDGFVQTIIEDIW GONGEFYNFKQVTFSEQVIELES S-GFKVKLMCSYCKHTESE-VPDYYGGPGFEKVLDILEDACESLLDSIVANNASISG -DS-APNKVKLMCSYCKHTESE-VPDYYGGPGFEKVLDILEDACESLLDSIVANNASISG -DS-APNKVKLMCSYCKHRTESE-VPDYYGGPGFEKVLDLLEDACESLLDSIVANNASISG -DS-GFKKVKLMCSYCKRHTESE-VPDYYGGPGFEKVLDLLEDACESLLDSIVANNASISG -DS-GFKK-VKLMCSYCKRHTESE-VPDYYGGPGFEKVLDLLEDACESLLDSIVANNASISG -DS-GFKK-VKLMCSYCKRHTESE-VPDYYGGPGFEKVLDLLEDACESLLDSIVANNASISG -DS-GFKK-VKLMCSYCKRHTESE-VPDYYGGPGFEKVLDLLEDACESLLDSIVANNASISG -D-S-GFKK-VKLMCSYCKRHTESE-VPDYYGGPGFEKVLDLLEDACESLLDSIVANNASISG -D-S-CFKK-VKLMCSYCKRHTESE-VPDYYGGPGFEKVLDLLEDACESLLDSIVANNASISG -D-S-CFKK-VKLMCSYCKRHTESE-VPDYYGGPGFEKVLDLLEDACESLLDSIVANNASISG -D-S-CFKK-VKLMCSYCKRHTESE-VPDYYGGPGFEKVLDLLEDACESLLDSIVANNASISG -D-S-CFKK-VKLMCSYCKRHTESE-VPDYYGGPGFEKVLDLLEDACESLLDSIVANNASISG -D-S-CFKK-VKLMCSYCKRHTESE-VPDYYGGPGFEKVLDLLEDACESLLDSIVANNASISG -D-S-CFKK-VKLMCSYCKRHTESE-VPDYYGGPGFEKVLDLLEDACESLLDSIVANNASISG -D-S-CFKK-VKLMCSYCKRHTESE-VPDYYGGPGFEKVLDLLEDACESLLDSIVANNASISG -D-S-CFKK-VKLMCSYCKRHTESE-VPDYYGGPGFEKVLDLLEDACESLLDSIVANNASISG -D-S-CFKK-VKLMCSYCKRHTESE-VPDYYGGPGFEKVLDLLEDACESLLDSIVANNASISG -D-S-CFKK-VKLMCSYCKRHTESE-VPDYYGGPGFEKVLDLLEDACESLLDSIVANNASISG -D-S-CFKK-VKLMCSYCKRHTESE-VPDYYGGPGFEKVLDLLEDACESLLDSIVANNASISG -D-S-CFKK-VKLMCSYCKRHTESE-VPDYYGGPGFEKVLDLLEDACESLLDSIVANNASISG -D-S-CFKK-VKLMCSYCKRHTESE-VPDYYGGPGFEKVLDLEDACESLLDSIVANNASISG -D-S-CFKK-VKLMCSYCKRHTESE-VPDYYGGPGFEKVLDLEDACESLLDSIVANNASISG -D-S-CFKK-VKLD	157 157 156 154 154 157 161 156 279 274 186 268 262 177 182
Ent his 1 Ent his 2 Ent dispar Ent dispar Ent dispar International dispar Signal disparation Sacch cerev Schiz pombe Sorg bicol Zea mays 1 Zea mays 1 Zea mays 1 Zea mays 2 Oryza sat Hord vulg Arab thal Vit vinif Pop tricho	100 100 100 100 126 105 223 215 127 210 203 123 122 118	EQ-YKQKIFKMVDFCTTIKTTEVPDFT GGERGFHRVIDILEDACENLIIKIEEGKLIN EQ-YKQKIFKMVDFCTTIKTTEVPDFY GGERGFHRVIDILEDACENLIIKIEEGKLIN EQ-YKQKIFKMVDFCTTIKTEVPDFY GGERGFHRVIDILEDACENLIIKIDEGKLIN EGOEKVKINDFCTTIKTEVPDFY 	157 157 157 156 154 154 159 157 161 156 279 274 186 268 262 177 182 173
Ent his 1 Ent his 2 Ent dispar Ent invad Gia lam P15 Giard intes Dictyo disc Klu lactis Sacch cerev Schiz pombe Sorg bicol Zea mays 1 Zea mays 2 Oryza sat Hord vulg Arab thal Vit vinif Pop tricho Ricin comm	1000 1000 1000 1000 105 2233 2155 2210 2033 1227 1233 1222 1188 1200	EQ-YKQKIFKMVDFCTTIKTTEVPDYYGGEKGFHRVIDILEDACENLIIKIEEGKLIN EQ-YKQKIFKMVDFCTTIKTTEVPDYYGGEIGFHRVIDILEDACENLIIKIEEGKLIN EQ-YKQKIFKMVDFCTTIKTTEVPDYYGGEIGFHRVIDILEDACENLIIKIEGGKLT KE-YQDKIHKMVEYCKKSKDSEVPDYYGGEIGFHRVIDILEDACENLIIKIEGGKLT KR-DQSKVRMFRDYDFGCK-GDVPDYYGGR-FELVYSVERTCRSIDDIKNGKAL D-N-HIATIKRLGEYHTH-KKINVEDYYGGNGFEYNDUEDCVFEKEISLD KEQ-SKARVLFGGNNTNNFQOVQTILEDW GONGFEYNDUECVFFKEL EG-SKARVLFGGNNTNNFQOVQTILEDW GONGGFEYNDIDLEDACESLDSIVANASISG -D-S-APKKVKLMCSYCKHTESEVPDYYGGPGFEKVLDLEDACESLDSIVANASISG -D-S-APKKVKLMCSYCKHTESEVPDYYGGPGFEKVLDLEDACESLDSIVANASISG -D-S-S-PKKVKLMCSYCKHTESEVPDYYGGPGFEKVLDLEDACESLDSIVANASISG -D-S-S-PKKVKLMCSYCKHTESEVPDYYGGPGFEKVLDLLEDACESLDSIVANASISG -D-S-S-PKKVKLMCSYCKHTESEVPDYYGGPGFEKVLDLLEDACESLDSIVANASISG -D-S-S-PKKVKLMCSYCKKHTESEVPDYYGGPGFEKVLDLLEDACESLDSIVANASISG -D-S-S-PKKVKLMCSYCKKHTESE-VPDYYGGPGFEKVLDLLEDACESLDSIVANASISG -D-S-S-PKKVKLMCSYCKKHTESE-VPDYYGGPGFEKVLDLLEDACESLDSIVANASISG -D-S-SAPKKVKLMCSYCKKHTESE-VPDYYGGPGFEKVLDLLEDACESLDSIVANASISG -D-S-SAPKKVKLMCSYCKKHTESE-VPDYYGGPGFEKVLDLLEDACESLDSIVANASISG -D-S-APKKVKLMCSYCKKHTESE-VPDYYGGPGFEKVLDLLEDACESLDSIVANASISG -D-S-APKKVKLMCSYCKKHTESE-VPDYYGGPGFEKVLDLLEDACESLDSIVANASISG -D-S-APKKVKLMCSYCKKHTESE-VPDYYGGPGFEKVLDLLEDACESLDSIVANASISG -D-S-APKKVKLMCSYCKKHTESE-VPDYYGGPGFEKVLDLLEDACESLDSIVANASISG -D-S-APKKVKLMCSYCKKHTESE-VPDYYGGPGFEKVLDLLEDACESLDSIVANASISG -D-S-APKKVKLMCSYCKKHTESE-VPDYYGGPGFEKVLDLLEDACESLDSIVANASISG -D-S-APKKVKLMCSYCKKHTESE-VPDYYGGPGFEKVLDLLEDACESLDSIVANASISG -D-D-ADKKVKLMCSYCKKHTESE-VPDYYGGPGFEKVLDLLEDACESLDSIVANASISG -D-D-ADKKVKLMCSYCKKHTESE-VPDYYGGPGFEKVLDLLEDACESLDSIVANASISG -D-D-ADKKVKLMCSYCKKHTESE-VPDYYGGPGFEKVLDLLEDACESLDSIVANASISG -D-D-ADKKVKLMCSYCKKHTESE-VPDYYGGPGFEKVLDLLEDACESLDSIVANASISG -D-D-ADKKVKLMCSYCKKHTESE-VPDYYGGPGFEKVLDLLEDACESLDSIVANASISG -D-D-ADKKVKLMCSYCKKHTESE-VPDYYGGPGFE	157 157 157 154 154 154 154 154 157 161 156 279 274 186 268 262 177 182 173 172
Ent his 1 Ent his 2 Ent dispar Ent dispar Ent dispar Sidard intes Dictyo disc Klu lactis Sacch cerev Schiz pombe Sorg bicol Zea mays 1 Zea mays 1 Zea mays 2 Oryza sat Hord vulg Arab thal Vit vinif Pop tricho Ricin comm Nost punct	100 100 100 100 105 223 215 127 210 203 123 122 118 120 114	EQ-YKQKIFKMVDFCTTIKTTEVPDFT GGERGFHRVIDILEDACENLIIKIEEGKLIN EQ-YKQKIFKMVDFCTTIKTTEVPDFT GGERGFHRVIDILEDACENLIIKIEEGKLIN 	157 157 157 156 154 154 154 157 161 156 279 274 186 262 262 177 182 173 172
Ent his 1 Ent his 2 Ent dispar Ent invad Gia 1am P15 Giard intes Dictyo disc Klu lactis Sacch cerev Schiz pombe Sorg bicol Zea mays 1 Zea mays 1 Zea mays 2 Oryza sat Hord vulg Arab thal Vit vinif Pop tricho Ricin comm Nost punct Lyngbya sp	100 100 100 100 126 105 123 215 223 210 203 127 210 203 122 118 120 114	EQ-YKQKIFKMVDFCTTIKTTEVPDYYGGEKGFHRVIDILEDACENLIIKIEEGKLIN EQ-YKQKIFKMVDFCTTIKTTEVPDYYGGEIGFHRVIDILEDACENLIIKIEEGKLIN EQ-YKQKIFKMVDFCTTIKTEVPDYYGGEIGFHRVIDILEDACENLIIKIEGKLIN KR-YQDKIFKMVDFCTTIKTE-VPDYYGGEIGFHRVIDILEDACENLIIKIEKGKLI KR-DQSKVRMFRDYDFCTIKTE-VPDYYGGR-FELVYSWERTCRSILDRIKNGKAL KR-DQSKVRMFRDYDFGGK-GDVPDYYGGR-FELVYSWERTCRSILDRIKNGKAL N-HIATIKRLGYNTDNKTQOJVQTILEDWGONGFEVNDILEDCFNEUGLINFLKEL S-4PKQLSIFGNNTDNKTQOJVQTILEDWGGSDGFGDCYIQVDFSQNFLKSIA S-APKVKLMCSYCKNTDNKTQ-VPDYYGGSDGFEKVIDILEDACESLLDSIVASISG S-APKKVKLMCSYCKHTESEVPDYYGGPGGFEKVIDILEDACESLLDSIVANASISG S-APKKVKLMCSYCKHTESEVPDYYGGPGGFEKVIDILEDACESLLDSIVANASISG S-APKKVKLMCSYCKHTESEVPDYYGGPGGFEKVIDILEDACESLLDSIVANASISG S-APKKVKLMCSYCKKHTESEVPDYYGGPGGFEKVIDILEDACESLLDSIVANASISG S-GFKKVKLMCSYCKKHTESEVPDYYGGPGGFEKVIDILEDACESLLDSIVANASISG S-APKKVKLMCSYCKKHTESEVPDYYGGPGGFEKVIDILEDACESLLDSIVANASISG S-GFKKVKLMCSYCKKHTESEVPDYYGGPGGFEKVIDILEDACESLLDSIVANASISG S-GFKKVKLMCSYCKKHTESEVPDYYGGPGGFEKVIDILEDACESLLDSIVANASISG S-GFKKVKLMCSYCKKHTESEVPDYYGGPGGFEKVIDILEDACESLLDSIVANASISG S-GFKKVKLMCSYCKKHTESEVPDYYGGPGGFEKVIDILEDACESLLDSIVANASISG S-GFKKVKLMCSYCKKHTESEVPDYYGGPGGFEKVIDILEDACESLLDSIVANASISG S-GFKKVKLMCSYCKKHTESEVPDYYGGPGGFEKVIDILEDACESLLDSIVANASISG 	157 157 156 154 154 154 156 279 274 186 268 268 268 268 262 2177 182 177 182 173 172
Ent his 1 Ent his 2 Ent dispar Ent dispar Ent invad Gia lam P15 Giard intes Dictyo disc Klu lactis Sacch cerew Schiz pombe Sorg bicol Zea mays 1 Zea mays 1 Zea mays 2 Oryza sat Hord vulg Arab thal Vit vinif Pop tricho Ricin comm Nost punct Lyngbya sp Cyyano sp	100 100 100 100 126 105 223 215 223 127 210 203 123 122 118 120 114 103 102	EQ-YKQKIFRMVDFCTTIKTTEVPDYYGGEKGFHRVIDILEDACENLIIKLEEGKLIN EQ-YKQKIFRMVDFCTTIKTTEVPDYY-GGEIGFHRVIDILEDACENLIIKLEEGKLIN EQ-YKQKIFRMVDFCTTIKTTEVPDYYGGEIGFHRVIDILEDACENLIIKLEGGKLIT KR-DQSKVRMFDVDFCTTIKTE-VPDYYGGEIGFHRVIDILEDACENLIIKLEGGKLIT KR-DQSKVRMFDVDFCTTIKTE-VPDYYGGEIGFHRVIDILEDACENLIIKLEGGKLIT KR-DSKIRMFDVDFCTTIKTE-VPDYYGGNEGFEYNSVERTCRSIDDIKNGKAL D-N-HIATIKELGEYNTH-KKINVEDYYGGNEGFEYNSVERTCRSIDDIKNGKAL KR-DSKIRMFNVDYFCTGYTIE KR-DDSKIKMFCVDYGVILEDIWGGNEGFEYNSVERTCRSIDDIKNGKAL S-SRAKVLIFCDNNTNNGVQTIEDIWGGNEGFEYNSVERTCRSIDDISISG S-SRAKVLIFCDNNTNNGVQTIEDIWGGNEGFEYNSVERTCRSIDDISISG S-APKKVKLMCSYCKRTESE-VPDYYGGPGCFEKVLDILEDACESLDSIVANNASISG S-APKKVKLMCSYCKRTESE-VPDYYGGPGCFEKVLDILEDACESLDSIVANNASISG S-GFKKVKLMCSYCKRTESE-VPDYYGGPGCFEKVLDILEDACESLDSIVANNASISG S-GFKKVKLMCSYCKRTESE-VPDYYGGPGCFEKVLDILEDACESLDSIVANNASISG S-GFKKVKLMCSYCKRTESE-VPDYYGGPGCFEKVLDILEDACESLDSIVANNASISG S-GFKKVKLMCSYCKRTESE-VPDYYGGPGCFEKVLDILEDACESLDSIVANNASISG S-GFKKVKLMCSYCKRTESE-VPDYYGGPGCFEKVLDILEDACESLDSIVANNASISG S-GFKKVKLMCSYCKRTESE-VPDYYGGPGCFEKVLDILEDACESLDSIVANNASISG S-GFKKVKLMCSYCKRTESE-VPDYYGGPGCFEKVLDILEDACESLDSIVANNASISG S-GFKKVKLMCSYCKRTESE-VPDYYGGPGCFEKVLDILEDACESLDSIVANNASISG S-GFKKVKLMCSYCKRTESE-VPDYYGGPGCFEKVLDILEDACESLDSIVANNASISG S-GFKKVKLMCSYCKRTESE-VPDYYGGPGCFEKVLDILEDACESLDSIVANNASISG 	157 157 156 154 154 159 157 161 156 279 274 186 268 262 177 182 173 172 169 162
Ent his 1 Ent his 2 Ent dispar Ent dispar Ent invad Gia 1am P15 Giard intes Dictyo disc Klu lactis Sacch cerew Schiz pombe Sorg bicol Zea mays 1 Zea mays 1 Zea mays 1 Zea mays 2 Oryza sat Hord vulg Arab thal Vit vinif Pop tricho Ricin comm Nost punct Lyngbya sp Cyano sp Arthro max	100 100 100 100 126 105 128 223 225 127 210 203 123 122 118 120 114 103 102	EQ-YKQKIFKMVDFCTTIKTEVPDYKGGEKGFHRVIDILEDACENLIIKIEEGKLIN EQ-YKQKIFKMVDFCTTIKTEVPDYKGGEIGHRVIDILEDACENLIIKIEEGKLIN EQ-YKQKIFKMVDFCTTIKTEVPDYKGGEIGHRVIDILEDACENLIIKIEGKLIN KR-YQCKIFKMVDFCTTIKTEVPDYKGGEIGHRVIDILEDACENLIIKIEKGKLIN KR-DQSKVKMFRDIDFCTIKTE-VPDYKGGEIGHRVIDILEDACENLIIKIEKGKLIN KR-DQSKVKMFRDIDFCTIKTEVPDYKGGEIGHRVIDILEDACENLIIKIEKGKLIN KR-DQSKVKMFRDIDFCTIKTEVPDYKGGR-FELVSIVSWERTCRSILDRIKNGKAL KR-DQSKVKMFRDIDFGK-GDVPDYKGGN-FELVSIVSWERTCRSILDRIKNGKAL KR-DQSKVKMFRDIDFGVOTILEDWGNBGFEVHDLEDCVNTLKEL-SLN KE-FAQLSIFGKNITDNKFQDIVDIVDUYGGNGGFEVHDLEDACESLDSIVASISG QG-SARVLFGDNNTNDGVQTILEDWGGIQGFEKVLDLEDACESLDSIVASISG D-S-ARKVKLMCSYCKRHTESEVPDYKGGDQGFEKVLDLEDACESLDSIVASISG -DS-ARKVKLMCSYCKRHTESEVPDYKGGDQGFEKVLDLEDACESLDSIVANASISG -DS-ARKVKLMCSYCKRHTESEVPDYKGGDQGFEKVLDLEDACESLDSIVANASISG -DS-GPKKVKLMCSYCKRHTESE-VPDYKGGDQGFEKVLDLEDACESLDSIVANASISG -DS-GPKKVKLMCSYCKRHTESE-VPDYKGGDQGFEKVLDLEDACESLDSIVANANSISG -DS-GPKKVKLMCSYCKRHTESE-VPDYKGGDQGFEKVLDLEDACESLDSIVANANSISG -DD-ARKKVKLMCSYCKRHTESE-VPDYKGGDQGFEKVLDLEDACESLDSIVANASISG -DD-ARKKVKLMCSYCKRHTESE-VPDYKGGDQGFEKVLDLEDACESLDSIVANASISG -DD-ARKKVKLMCSYCKRHTESE-VPDYKGGDQGFEKVLDLEDACESLDSIVANASISG -DD-ARKKVKLMCSYCKRHDETE-VPDYKGGDQGFEKVLDLEDACESLDSIVANASISG -DD-ARKKVKLMCSYCKRHDETE-VPDYKGGDQGFEKVLDLEDACESLDSIVANASISI -PD-ARKKVKLMCSYCKRHDETE-VPDYKGGDQGFEKVLDLEDACESLDSIVANASISI -PD-ARKKVKLMCSYCKRHDETE-VPDYKGGPQGFEKVLDLEDACESLDSIVANASISI -PD-ARKK-VKLMCSYCKRHDETE-VPDYKGGPQGFEKVLDLEDACESLDSIVANASISG -PD-ARKK-VKLMCSYCKRHDETE-VPDYKGGPQGFEKVLDLEDACESLDSIVANASISG -PD-ARKK-VKLMCSYCKRHDETE-VPDYKGGPQGFEKVLDLEDACESLDSIVANESLDSIVAN 	157 157 156 154 154 154 161 156 279 274 182 262 262 277 182 173 172 169 162 169 162 161
Ent his 1 Ent his 2 Ent dispar Ent invad Gia lam P15 Giard intes Dictyo disc Klu lactis Sacch cerev Schiz pombe Sorg bicol Zea mays 1 Zea mays 2 Oryza sat Hord vulg Arab thal Vit vinif Pop tricho Ricin comm Nost punct Lyngbya sp Cyano sp Arthro max Prochlo mar	100 100 100 100 126 105 223 215 127 210 203 122 118 120 114 103 102 103	EQ-YKQKIFKMVDFCTTIKTTEVPDYYGGEKGFHRVIDILEDACENLIIKIEEGKLIN EQ-YKQKIFKMVDFCTTIKTTEVPDYYGGEKGFHRVIDILEDACENLIIKIEEGKLIN EQ-YKQKIFKMVDFCTTIKTTE-VPDYYGGEKGFHRVIDILEDACENLIIKIEGKLIN KR-DQSKVRMFPDDFCTTIKTE-VPDYYGGEKGFHRVIDILEDACENLIIKIEGKLIN KR-DQSKVRMFPDDFCTKIKTE-VPDYYGGEKGFHRVIDILEDACENLIIKIEGKLIN KK-BSKIRMFDDFCTKIKTE-VPDYYGGEKGFHRVIDILEDACENLIIKIEGKLIN K	157 157 156 154 159 157 161 156 279 182 274 186 268 262 177 182 173 172 169 162 161 160 156
Ent his 1 Ent his 2 Ent dispar Ent dispar Ent dispar International dispar Signal disparation Sacch cerev Schiz pombe Sorg bicol Zea mays 1 Zea mays 1 Zea mays 2 Oryza sat Hord vulg Arab thal Vit vinif Pop tricho Ricin comm Nost punct Lyngbya sp Cyano sp Arthro max Prochlo mar Bdello bact	100 100 100 100 126 105 223 215 227 210 203 123 122 118 120 114 103 102 103 115	EQ-YKQKIFKMVDFCTTIKTTEVPDYIGGEKGFHRVIDILEDACENLIIKIEEGKLIN EQ-YKQKIFKMVDFCTTIKTTEVPDYIGGEIGHRVIDILEDACENLIIKIEEGKLIN EQ-YKQKIFKMVDFCTTIKTTEVPDYIGGEIGHRVIDILEDACENLIIKIEEGKLIN KR-DQKVFKMVDFCTTIKTEVPDYIGGEIGHRVIDILEDACENLIIKIQGKLIN KR-DQKVFKMVDFCTTIKTE-VPDYIGGEIGHRVIDILEDACENLIIKIQGKLIN KR-DQKVFKKWTDNKGO-VPDYIGGR-FELVISIVERTGRSILDRIKAGKAL KR-DQKVIKKWNTDNKGO-VPDYIGGNEGFEVFDELEES EG-SKAKVLFGWNTDNKGO-VDDYIGGNEGFEVFDELEES EAPKVKLMCSYCKNTESEVPDYIGGNEGFEVFDLEDACESLDSIVANASISG -DS-APKKVKLMCSYCKRTESEVPDYIGGNEGFEVIDLEDACESLDSIVANASISG -DS-APKKVKLMCSYCKRTESEVPDYIGGNEGFEVIDLEDACESLDSIVANASISG -DS-APKKVKLMCSYCKRTESEVPDYIGGNEGFEVIDLEDACESLDSIVANASISG -DS-APKKVKLMCSYCKRTESEVPDYIGGNEGFEVIDLEDACESLDSIVANASISG -DS-APKKVKLMCSYCKRTESEVPDYIGGNEGFEVIDLEDACESLDSIVANASISG -DS-APKKVKLMCSYCKRTESEVPDYIGGNEGFEVIDLEDACESLDSIVANASISG -DS-APKKVKLMCSYCKRTESEVPDYIGGNEGFEVIDLEDACESLDSIVANASISG -DS-APKKVKLMCSYCKRTESEVPDYIGGNEGFEVIDLEDACESLDSIVANASISG -DS-APKKVKLMCSYCKRTESEVPDYIGGNEGFEVIDLEDACESLDSIVANASISG -DS-APKKVKLMCSYCKRTESE-VPDYIGGNEGFEVIDLEDACESLDSIVANASISG -DS-APKKVKLMCSYCKRTESE-VPDYIGGNEGFEVIDLEDACESLDSIVANASISG -DD-AHKKVKLMCSYCKRTESE-VPDYIGGNEGFEVIDLEDACESLDSIVANASISG -DD-AHKKVKLMCSYCKRTESE-VPDYIGGNEGFEVIDLEDACESLDSIVANASHLDS -PD-ADKKVKLMCSYCKRTESE-VPDYIGGNEGFEVIDLEDACESLDSIVANASHLDS -PD-AHKKVKLMCSYCKRTESE-VPDYIGGNEGFEVIDLEDACESLDSIVASHLDS -RD-AHKKVKLMCSYCKKHDETEVPDYIGGNEGFEVIDLEDACESLDSIVASHLDS -RD-AHKKVKLMCSYCKKHDEFE-VPDYIGGNEGFEVIDLEDACESLDSIVASHLDS -RD-AHKKVKLMCSYCKKHDEFE-VPDYIGENEGFEVIDLEDACESLDSIVASHLDS -RD-AHKKVKLMCSYCKKHDEFE-VPDYIGENEGFEVIDLEDACESLDSIVASHLDS -RD-AHKKVKLMCSYCKKHDEFE-VPDYIGENEGFEVIDLEDACESLDSIVASHLDS -RD-AHKKVKLMCSYCKKHDEFE-VPDYIGENEGFEVIDLEDACESLDSIVASHLDS -RD-AHKKVKLMCSYCKKHDEFE-VPDYIGENEGFEVIDLEDACESLDSIVASHLDS -RD-AHKKVKLMCSYCKKHDEFE-	157 157 156 154 154 159 157 161 156 279 278 268 262 268 262 177 182 163 172 169 162 161 156 170 156 169 162 161 156 162
Ent his 1 Ent his 2 Ent dispar Ent invad Gia lam P15 Giard intes Dictyo disc Klu lactis Sacch cerev Schiz pombe Sorg bicol Zea mays 1 Zea mays 2 Oryza sat Hord vulg Arab thal Vit vinif Pop tricho Ricin comm Nost punct Lyngbya sp Cyano sp Arthro max Prochlo mar Bact caccae	100 100 100 100 126 105 223 215 127 210 203 123 123 122 118 120 114 103 102 103 115 103	EQ-YKQKIFKMVDFCTTIKTTEVPDYYGGEKGFHRVIDILEDACENLIIKIEEGKLIN EQ-YKQKIFKMVDFCTTIKTTEVPDYYGGEIGFHRVIDILEDACENLIIKIEEGKLIN EQ-YKQKIFKMVDFCTTIKTTEVPDYYGGEIGFHRVIDILEDACENLIIKIEGKLIN KF-YQDKIFKMVDFCTTIKTTE-VPDYYGGEIGFHRVIDILEDACENLIIKIEGKLIN KF-YQDKIHHMVEYCKKISKDSEVPDYYGGEIGFHRVIDILEDACENLIIKIEKGKLI KR-DQSKVRMFRDYDFCTKIKTE-VPDYYGGR-FELYSNVERTCRSILDNIKNGKAL N-HIATIKRLGEYNTIKTE-VPDYYGGR-FELYSNVERTCRSILDNIKNGKAL N-HIATIKRLGEYNTDNKTO'UTILEDWGNEGFEYNDILEDCENLDSIKNGKAL 	157 157 154 154 154 154 159 157 161 156 279 274 186 262 274 182 262 177 182 173 172 169 162 161 156 156 156 156 156 157 157 157 157 157 156 154 154 154 154 154 154 154 154 155 155
Ent his 1 Ent his 2 Ent dispar Ent dispar Ent diapar Siard intes Dictyo disc Klu lactis Sacch cerev Schiz pombe Sorg bicol Zea mays 1 Zea mays 1 Zea mays 2 Oryza sat Hord vulg Arab thal Vit vinif Pop tricho Ricin comm Nost punct Lyngbya sp Cyano sp Arthro max Prochlo mar Bdello bact Bact caccae Bact theta	100 100 100 100 126 105 223 215 220 203 123 122 118 120 114 103 103 103 103	EQ-YKQKIFRMVDFCTTIKTEVPDY GGECKFHRVIDILEDACENLIKLEKGKLIN EQ-YKQKIFRMVDFCTTIKTEVPDY GGECKFHRVIDILEDACENLIKLEKGKLIN 	157 157 154 154 154 179 157 161 279 274 156 268 262 279 274 156 268 262 177 182 268 262 177 182 268 173 172 169 161 160 156 161 160 156 161 164 175 175 155 155 155 155 155 155 155 155
Ent his 1 Ent his 2 Ent dispar Ent invad Gia lam P15 Giard intes Dictyo disc Klu lactis Sacch cerev Schiz pombe Sorg bicol Zea mays 1 Zea mays 2 Oryza sat Hord vulg Arab thal Vit vinif Pop tricho Ricin comm Nost punct Lyngbya sp Cyano sp Arthro max Prochlo mar Bdello bact Bact caccae Bact theta Bact fragil	100 100 100 100 105 105 223 215 127 210 203 123 123 123 123 123 123 123 123 123 12	EQ-YKQKIFRMVDFCTTIKTEVPDYYGGECKGFHRVIDILEDACENLIIKIEEGKLIN EQ-YKQKIFRMVDFCTTIKTEVPDYYGGEIGHRVIDILEDACENLIIKIEEGKLIN EQ-YKQKIFRMVDFCTTIKTEVPDYYGGEIGHRVIDILEDACENLIIKIEGKLIN KR-YQKIFRMVDFCTTIKTEVPDYYGGEIGHRVIDILEDACENLIIKIEKGKLIN KR-DQSKVRMFRDJDFCTIKTE-VPDYYGGEIGHRVIDILEDACENLIIKIEKGKLI KR-DQSKVRMFRDJDFCTIKTE-VPDYYGGEIGHRVIDILEDACENLIIKIEKGKLI KR-DQSKVRMFRDJDFCGK-GDVPDYYGGR-FELVSNVRTCRSILDNIKNGKAL D-N-HIATIKKLGYTHI-KKIN-VPDYYGMNGFFYNFDCVNFIKGPIKKEL S-RAKVLFGDNNTNGGVQTIEDIWGGSDGFGCVIDILEDACESLIDSIVASISG 	157 157 154 154 179 157 161 279 274 186 268 268 268 262 177 182 268 173 172 169 160 156 171 156 156 157
Ent his 1 Ent his 2 Ent dispar Ent invad Gia lam P15 Giard intes Dictyo disc Klu lactis Sacch cerew Schiz pombe Sorg bicol Zea mays 1 Zea mays 2 Oryza sat Hord vulg Arab thal Vit vinif Pop tricho Ricin comm Nost punct Lyngbya sp Cyano sp Arthro max Prochlo mar Bdello bact Bact theta Bact fragil Parab merd	100 100 100 100 105 108 223 215 127 223 123 123 123 123 123 123 123 123 123	E-Q-YKQKIFRMVDFCTTIKTEVPDY GGECKFHRVIDILEDACENLIKIEEGKLIN E-Q-YKQKIFRMVDFCTTIKTEVPDY GGECKFHRVIDILEDACENLIKIEEGKLIN E-Q-YKQKIFRMVDFCTTIKTEVPDY GGECKFHRVIDILEDACENLIKIDEGKLIN E-YQCKIFRMVDFCTTIKTEVPDY 	157 157 154 154 154 179 157 279 278 268 268 268 268 268 268 268 268 268 26
Ent his 1 Ent his 2 Ent dispar Ent dispar Ent dispar International and the second Scale and the second Scale and second Scale	100 100 100 100 105 108 105 223 215 127 210 203 123 123 123 123 123 120 114 103 103 103 103 103 103 104 109 102	EQ-YKQKIFRMVDFCTTIKTEVPDYIGGEKGFHRVIDILEDACENLIIKIEEGKLIN EQ-YKQKIFRMVDFCTTIKTEVPDYIGGEKGFHRVIDILEDACENLIIKIEEGKLIN EQ-YKQKIFRMVDFCTTIKTEVPDYIGGEKGFHRVIDILEDACENLIIKIEGKLIN KR-YQCKIFRMVDFCTTIKTEVPDYIGGEKGFHRVIDILEDACENLIIKIEGKLIN KR-DQSKVRHFDVDFCTTIKTEVPDYIGGEKGFHRVIDILEDACENLIIKIEGKLIN KR-DQSKVRHFDVDFCTTIKTE-VPDYIGGEKGFHRVIDILEDACESILDRIKGKAL KR-DQSKVRHFDVDFCTTIKTE-VPDYIGGEKGFHRVIDILEDACESILDRIKGKAL KR-DQSKVRHFDVDFCTG-GD-VPDYICGR-FELVYSVERTCRSILDRIKGKAL KR-DQSKVKHCSYKKNTDNKCOVOTIEDWIGGNGGFEKVLDLEDACESILDRISGKAL KE-FAQLSIFKKNTDNKCOVOTIEDWIGGNGGFEKVLDLEDACESILDSIVASISG CSARKVLFGDNNTNGGVQTIEDWIGGNGGFEKVLDLEDACESILDSIVASISG D-S-ARKVKLMCSYKRHTESEVPDYIGGPGFEKVLDLEDACESILDSIVANNASISG -DS-ARKVKLMCSYKRHTESEVPDYIGGPGFEKVLDLEDACESILDSIVANNASISG -DS-ARKVKLMCSYKRHTESEVPDYIGGPGFEKVLDLEDACESILDSIVANNASISG -DS-GFKKVKLMCSYKRHTESEVPDYIGGPGFEKVLDLEDACESILDSIVANNASISG -DS-GFKKVKLMCSYKRHTESEVPDYIGGPGFEKVLDLEDACESILDSIVANNASISG -DS-GFKKVKLMCSYKRHTESEVPDYIGGPGFEKVLDLEDACESILDSIVANNASISG -DS-GFKKVKLMCSYKRHTESEVPDYIGGPGFEKVLDLEDACESILDSIVANNASISG -DD-ARKKVKLMCSYKRHTESEVPDYIGGPGFEKVLDLEDACESILDSIVANNASISG -DD-ARKKVKLMCSYKRHTESEVPDYIGGPGFEKVLDLEDACESILDSINAENHISA -PD-ARKKVKLMCSYKRHTESEVPDYIGGPGFEKVLDLEDACESILDSINAENHISA -PD-ARKKVKLMCSYKRHTESE-VPDYIGGPGFEKVLDLEDACESILDSINAENHISA -PD-ARKKVKLMCSYKRHTETE-VPDYIGGPGFEKVLDLEDACESILDSINAENHISA -PD-ARKKVKLMCSYKRHTETE-VPDYIGGPGFEKVLDLEDACESILDSIVANSEF -RQ-YONKVKLMCSYKRHTETE-VPDYIGGPGFEKVLDLEDACESILDSINAENHISA -PD-ARKKVKLMCSYKRHTETE-VPDYIGGPGFEKVLDLEDACESILDSINAENHISA -PD-ARKKVKLMCSYKRHTETE-VPDYIGGPGFEKVLDLEDACESILDSINAENHISA -PD-ARKKVKLMCSYKRHTETE-VPDYIGGPGFEKVLDLEDACESILDSINAENHISA -PQ-YONKVKLMCSYKRHTETE-VPDYIGGPGFEKVLDLEDACESILDSINAENHISA 	157 157 154 154 179 157 156 279 274 156 268 268 262 279 274 173 172 169 161 166 156 156 156 156 156 156
Ent his 1 Ent his 2 Ent dispar Ent invad Gia lam P15 Giard intes Dictyo disc Klu lactis Sacch cerev Schiz pombe Sorg bicol Zea mays 1 Zea mays 1 Zea mays 2 Oryza sat Hord vulg Arab thal Vit vinif Pop tricho Ricin comm Nost punct Lyngbya sp Cyano sp Arthro max Prochlo mar Bdello bact Bact caccae Bact theta Bact theta Bact fragil Parab merd Vibrio chol	100 100 100 100 1266 105 127 210 203 123 122 118 120 103 102 103 103 103 103 103 103 103 103 103 103	EQ-YKQKIFRMVDFCTTIKTEVPDYYGGEKGFHRVIDILEDACENLIIKIEEGKLIN EQ-YKQKIFRMVDFCTTIKTEVPDYYGGEIGFHRVIDILEDACENLIIKIEEGKLIN EQ-YKQKIFRMVDFCTTIKTEVPDYYGGEIGFHRVIDILEDACENLIIKIEGKLIN KF-YQNKIFRMVDFCTTIKTEVPDYYGGEIGFHRVIDILEDACENLIIKIEGKLIN KK	157 157 156 154 154 154 159 161 156 279 182 262 274 182 262 274 182 262 2177 182 161 156 160 156 157 169 166 157 155 156
Ent his 1 Ent his 2 Ent dispar Ent dispar Ent dispar Ent dispar Side and the second Soluty of	1000 1000 1000 1005 1055 2233 1227 2100 2033 1223 1231 1231 1231 1231 1231	EQ-YKQKIFRMVDFCTTIKTEVPDYIGGEKGFHRVIDILEDACENLIKIEEGKLIN EQ-YKQKIFRMVDFCTTIKTEVPDYIGGEIGHRVIDILEDACENLIKIEEGKLIN EQ-YKQKIFRMVDFCTTIKTEVPDYIGGEIGHRVIDILEDACENLIKIEEGKLIN 	$157\\157\\157\\154\\154\\179\\279\\274\\186\\262\\177\\161\\156\\171\\173\\172\\162\\161\\156\\156\\156\\155\\158$
Ent his 1 Ent his 2 Ent dispar Ent invad Gia lam P15 Giard intes Sidto disc Klu lactis Sacch cerev Schiz pombe Sorg bicol Zea mays 1 Zea mays 2 Oryza sat Hord vulg Arab thal Vit vinif Pop tricho Ricin comm Nost punct Lyngbya sp Cyano sp Arthro max Prochlo mar Bdello bact Bact theta Bact theta Bact theta Bact fragil Parab merd Vibrio chol Pseudo aeru Homo isof b	1000 1000 1000 1000 105 2233 2155 2233 1232 1122 1188 1200 103 1033 1033 1033 1033 1033 1031 1031 1032 1032	EQ-YKQKIFRMVDFCTTIKTEVPDYYGGEKGFHRVIDILEDACENLIIKIEEGKLIN EQ-YKQKIFRMVDFCTTIKTEVPDYYGGEIGFHRVIDILEDACENLIIKIEEGKLIN EQ-YKQKIFRMVDFCTTIKTEVPDYYGGEIGFHRVIDILEDACENLIIKIEGKLIN KR-YQKIFRMVDFCTTIKTEVPDYYGGEIGFHRVIDILEDACENLIIKIEKGKLIN KR-YQKIFRMVDFCTTIKTE-VPDYYGGEIGFHRVIDILEDACENLIIKIEKGKAL KR-PQKVKMFRDJDFCTGKG-GDVPDYYGGEIGFHRVIDILEDACENLIIKIEKGKAL KR-PQKVKMFRDJDFCGK-GDVPDYYGGR-FELVYSWERTCRSILDRIKNGKAL D-N-HIATIKRLGEYKINTONKYOJVQTILEDWYGMSGFKVHDILEDCKELLDRISNGKAL 	$157\\157\\157\\154\\154\\159\\279\\161\\156\\262\\279\\162\\177\\182\\262\\177\\182\\262\\177\\182\\161\\156\\156\\157\\154\\158\\158\\158$
Ent his 1 Ent his 2 Ent dispar Ent dispar Ent dispar Ent dispar Schill and P15 Giard intes Scient disparts Sacch cerev Schiz pombe Sorg bicol Zea mays 1 Zea mays 1 Zea mays 2 Oryza sat Hord vulg Arab thal Vit vinif Pop tricho Ricin comm Nost punct Lyngbya sp Cyano sp Arthro max Prochlo mar Bact caccae Bact theta Bact fragil Parab merd Vibrio chol Pseudo aeru Homo isof c Homo isof c	1000 1000 1000 1000 1050 2233 2152 127 2100 2033 1222 1133 1022 1033 1031 1021 1031 103	EQ-YKQKIFRMVDFCTTIKTEVPDY GGERGHRVIDILEDACENLIKIEEGKLIN EQ-YKQKIFRMVDFCTTIKTEVPDY GGERGHRVIDILEDACENLIKIEEGKLIN EQ-YKQKIFRMVDFCTTIKTEVPDY GGERGHRVIDILEDACENLIKIDEGKLIN E	$157\\ 157\\ 157\\ 154\\ 154\\ 179\\ 161\\ 1279\\ 274\\ 186\\ 2262\\ 177\\ 172\\ 161\\ 157\\ 161\\ 157\\ 162\\ 161\\ 156\\ 157\\ 155\\ 154\\ 158\\ 158\\ 158\\ 158\\ 158\\ 158\\ 158\\ 158$
Ent his 1 Ent his 2 Ent dispar Ent invad Gia lam P15 Giard intes Dictyo disc Klu lactis Sacch cerev Schiz pombe Sorg bicol Zea mays 1 Zea mays 2 Oryza sat Hord vulg Arab thal Vit vinif Pop tricho Ricin comm Nost punct Lyngbya sp Cyano sp Arthro max Prochlo mar Bdello bact Bact caccae Bact theta Bact fragil Parab merd Vibrio chol Pseudo aeru Homo isof b Homo isof c Mus musc 1	1000 1000 1000 1000 105 108 105 2233 1227 2100 2102 103 1033 1033 1033 1033 1033	EQ-YKQKIFRMVDFCTTIKTEVPDYTGGEKGFHRVIDILEDACENLIIKIEEGKLIN EQ-YKQKIFRMVDFCTTIKTEVPDYTGGEIGHRVIDILEDACENLIIKIEEGKLIN EQ-YKQKIFRMVDFCTTIKTEVPDYTGGEIGHRVIDILEDACENLIIKIEGKLIN 	$157\\157\\157\\156\\154\\154\\268\\262\\262\\262\\177\\182\\169\\162\\160\\156\\156\\156\\156\\156\\156\\156\\156\\156\\156$
Ent his 1 Ent his 2 Ent dispar Ent invad Gia lam P15 Giard intes Dictyo disc Klu lactis Sacch cerew Schiz pombe Sorg bicol Zea mays 1 Zea mays 2 Oryza sat Hord vulg Arab thal Vit vinif Pop tricho Ricin comm Nost punct Lyngbya sp Cyano sp Arthro max Prochlo mar Bdello bact Bact fragil Parab merd Vibrio chol Pseudo aeru Homo isof b Homo isof c Mus musc 1 Mus musc 2 Bos taurus	1000 1000 1000 1000 1020 1025 1027 2100 2033 1232 1022 1033 1022 1033 1033	EQ-YKQKIFRMVDFCTTIKTEVPDY GGERGHRVIDILEDACENLIKIEEGKLIN EQ-YKQKIFRMVDFCTTIKTEVPDY GGERGHRVIDILEDACENLIKIEGGKLT 	$157\\157\\157\\156\\154\\154\\179\\157\\161\\279\\274\\162\\262\\262\\262\\262\\177\\172\\169\\165\\156\\156\\157\\156\\156\\156\\156\\156\\156\\156\\156\\156\\156$
Ent his 1 Ent his 2 Ent dispar Ent invad Gia lam P15 Giard intes Dictyo disc Klu lactis Sacch cerev Schiz pombe Sorg bicol Zea mays 1 Zea mays 2 Oryza sat Hord vulg Arab thal Vit vinif Pop tricho Ricin comm Nost punct Lyngbya sp Cyano sp Arthro max Prochlo mar Bedllo bact Bact fragil Parab merd Vibrio chol Pseudo aeru Homo isof c Mus musc 1 Mus musc 1 Mus musc 2 Bos taurus Xenop laev	1000 1000 1000 1055 2235 1277 2203 1225 1270 2033 1022 1023 1033 1033 1033 1033 103	EQ-YKQKIFRMVDFCTTIKTEVPDYIGGEKGFHRVIDILEDACENLIKIEEGKLIN EQ-YKQKIFRMVDFCTTIKTEVPDYIGGEIGHRVIDILEDACENLIKIEEGKLIN EQ-YKQKIFRMVDFCTTIKTEVPDYIGGEIGHRVIDILEDACENLIKIEGKKLIN 	$\begin{array}{r} 157\\ 157\\ 158\\ 154\\ 154\\ 161\\ 274\\ 161\\ 274\\ 162\\ 262\\ 274\\ 182\\ 262\\ 274\\ 172\\ 162\\ 177\\ 162\\ 156\\ 156\\ 156\\ 156\\ 156\\ 158\\ 158\\ 158\\ 158\\ 158\\ 158\\ 158\\ 158$
Ent his 1 Ent his 2 Ent dispar Ent invad Gia lam P15 Giard intes Dictyo disc Klu lactis Sacch cerev Schiz pombe Sorg bicol Zea mays 1 Zea mays 2 Oryza sat Hord vulg Arab thal Vit vinif Pop tricho Ricin comm Nost punct Lyngbya sp Cyano sp Arthro max Prochlo mar Bdello bact Bact theta Bact theta Bact theta Bact theta Homo isof b Homo isof c Mus musc 1 Mus musc 2 Bos taurus Xenop laev Gal gallus	1000 1000 1000 1000 1005 1005 2233 2155 2233 2122 2100 2033 1222 1018 1220 1124 1033 1032 1051 1051 1055 1055 1055 1077 1055	E-Q-YKQKIFRMVDFCTTIKTEVPDY GGERGHRVIDILEDACENLIKIEEGKLIN E-Q-YKQKIFRMVDFCTTIKTEVPDY GGERGHRVIDILEDACENLIKIEGGKLIT EGKLIN 	$\begin{array}{r} 157\\ 1557\\ 1554\\ 179\\ 1554\\ 179\\ 156\\ 279\\ 186\\ 262\\ 177\\ 182\\ 173\\ 172\\ 162\\ 177\\ 162\\ 156\\ 157\\ 155\\ 154\\ 158\\ 158\\ 158\\ 158\\ 158\\ 158\\ 158\\ 158$
Ent his 1 Ent his 2 Ent dispar Ent dispar Ent dispar Ent dispar Schiz pombe Schiz pombe Arthro max Prochlo mar Bdello bact Bact theta Bact fragil Parab merd Vibrio chol Pseudo aeru Homo isof b Homo isof b Homo isof b Homo isof b Homo isof c Mus musc 1 Mus musc 1 Schiz gallus Dros melan	100 100 100 100 105 125 223 215 223 123 123 123 123 123 123 123 123 123	EQ-YKQKIFRMVDFCTTIKTEVPDYIGGEKGFHRVIDILEDACENLIKIEEGKLIN EQ-YKQKIFRMVDFCTTIKTEVPDYIGGEIGHRVIDILEDACENLIKIEEGKLIN 	$\begin{array}{r} 157\\ 1557\\ 1554\\ 1759\\ 154\\ 1757\\ 156\\ 154\\ 1757\\ 161\\ 156\\ 274\\ 262\\ 262\\ 262\\ 177\\ 1757\\ 169\\ 155\\ 156\\ 156\\ 156\\ 156\\ 156\\ 158\\ 158\\ 158\\ 158\\ 158\\ 158\\ 158\\ 158$

Ent his 1	55	RK-VGKSR-GYQVDSISRPVVSSDFKNFDYIFAMDNDNYYELLD-RCP	99
Ent his 2	55	<mark>RK-VGKSR-GYQVDSISRPVVSSDFKNFDYIFVMDNDNYYELL</mark> D- <mark>RC</mark> <mark>P</mark>	99
Ent dispar	55	RK-VGKSR-GYQVDSISRPVVSSDFKNFDYIFAMDNDNYYELLD-RCP	99
Ent invad	55	RA-AGLGR-GYTFDHISRQVKQSDFSNFDIIVAMDDSNYSDLMY-NCP	99
Gia lam P15	55	QKAVADNL-GISFSKKSRHWRKSDYDDFDLILAMDKSNFQDIMR-EVG	99
Giard intes	55	<mark>QKAI</mark> TDKL-GISFSKK <mark>SR</mark> HWR <mark>NSDY</mark> DGFDLILAMDKS <mark>NFQDI</mark> <mark>I</mark> K-EIG	99
Dictyo disc	64	QSCNQNMGR-AISEES-LKHFKSIPLHRARQFTDEDFSKFDYIFAMDESNLSNIKKVLKHST-TK	125
Klu lactis	60	AS-TCRKN-GVPISHSAQQIKSKHFKEFDYIICMDESNLRNLKR-IQP	104
Sacch cerev	63	VS- <mark>ICK</mark> QH- <mark>GVKIN</mark> HKGKQIKTKHFDEYDYIIGMDESNINNLKK-IQP	107
Schiz pombe	60	LE-VLKKN-GIHTKHLARKLSTSDFKNFDYIFAMDSSNLRNINR-VKP	104
Sorg bicol	169	RA-ASKKR-GIEVTSISRPIRPSDFRDFDLILAMDRQNYEDILNSFERWRRKE-PLP	222
Zea mays 1	161	<mark>R</mark> A-AS <mark>KKR-GIEVTSI</mark> <mark>SRPI</mark> KP <mark>SDFRDFDLILAMD</mark> RQ <mark>NYEDI</mark> LNSFERWRRK <mark>E</mark> -PL <mark>P</mark>	214
Zea mays 2	73	<mark>R</mark> A-AS <mark>KKR-GIEVTSISRPI</mark> KP <mark>SDFRDFDLILAMDRQNYEDI</mark> LNSFERWRRK <mark>E</mark> -PL <mark>P</mark>	126
Oryza sat	156	RS-ASKKR-GIEVTSISRPVKPSDFRDFDLILAMDRQNYEDLLSSFERWRHRE-PLP	209
Hord vulg	149	IS-TSKKR-GIEVTSISRPIKPSDFRIFDLILAMDRONYEDILSSFDRWOHKE-TLP	202
Arab thal	69	RS-AAKRR-GIEITSLSRPIKASDFREFDLILAMDDONKEDILKAYNVWKARG-NFP	122
Vit vinif	68	RA-ASKRR-GIEITSISRPIQPSDFRDFDLILAMDKQNRDDIMEAFDRWRFRE-TLP	121
Pop tricho	64	<mark>R</mark> A-AS <mark>KRR-GIEITSISRPI</mark> RP <mark>SDFRDFDIILAMDNONREHI</mark> MEAFNKWKFK <mark>E</mark> -TL <mark>P</mark>	117
Ricin comm	66	RA-ASKRR-GIEITSISRPIRPSDFRDFDIIILAMDKONREDILEAFNRWKVRD-SLP	119
Nost punct	67	SA-AAA <mark>TKLGIKL</mark> RGQ <mark>AR</mark> QFQK <mark>SDFQDFDLILAMDQENYENI</mark> LT-LD-RT	113
Lyngbya sp	57	TK-AAS <mark>SR-G</mark> ILLKGQ <mark>AR</mark> QFGREDFENFDLILAMDYDNYESIVR-LD-SA	102
Cyano sp	57	NA-AASG <mark>R-G</mark> IPLKGK <mark>AR</mark> QFQP <mark>SDFENFDLILAMDRENYRDI</mark> RR-L <mark>P</mark> P	101
Arthro max	57	TM-AAKKR-GIVLKGSARQFTASDFDNFDLILAMDQDNYQQILY-LD-RG	102
Prochlo mar	57	RS-AALSR-GIVIESRARQISLNDFNNFDLILTMDQSNLDDVNS-LAK	101
Bdello bact	70	IL-HGERR-GYDLTSISRAIRESDYYDFDWILAMDASNLEHLRQ-RCP	114
Bact caccae	56	RA-HAARR-GYQLVHRSRPVRTEDFYNFDLIIGMDDRNIEDLKD-KAPSP	102
Bact theta	56	<mark>R</mark> A-HAAR <mark>R-GYQL</mark> VHR <mark>SRPVRTEDFYNFDLI</mark> IG <mark>MDDRN</mark> IDDLK <mark>E-K</mark> APST	102
Bact fragil	57	RA-HAARR-GYELVHRSRPVRTEDFYNFDLIIGMDDRNMDDLKE-KAPSP	103
Parab merd	64	RS-HAARR-GYKLDSVSRPVCTEDFFDFDLLIGMDNRNIDDLKR-KAP	108
Vibrio chol	56	<mark>K</mark> A-A <mark>GEKR-GY</mark> SFSG <mark>I</mark> K <mark>ARKI</mark> RDE <mark>DF</mark> VKFDWILAADQENLAELKA- <mark>RC</mark> <mark>P</mark>	101
Pseudo aeru	55	<mark>R</mark> L-AA <mark>QRR-GYDLSAL</mark> RG <mark>RQV</mark> GIE <mark>DF</mark> SRHDL <mark>ILAMD</mark> LS <mark>NLADL</mark> ETL <mark>R</mark> AG	101
Homo isof b	61	VS-CL <mark>RN</mark> H-GIHTAHKARQITKEDFATFDYILCMDESNLRDLN-RKS	104
Homo isof c	61	QS-CMKRH-GIPMSHVARQITKEDFATFDYILCMDESNLRDLN-RKS	104
Mus musc 1	61	VS-CLRNH-GISTAHKARQITKEDFATFDYILCMDESNLRDLN-RKS	104
Mus musc 2	61	QN-CMRKH-GIHMQHIARQITKEDFATFDYILCMDESNLRDLN-RKS	104
Bos taurus	61	QA-CMRKH-GIPMSHVARQVTKEDFVTFDYILCMDESNLRDLN-RKS	104
Xenop laev	62	LK-CLKSH-SIETSHRR-RGS	106
Gal gallus	61	QT-CMKKH-GITMNHIRCVTKDDFQTFDYILCMDESNLRDL	104
Dros melan	56	IS-TLOKH-GLKCTHIVRQIRKQDFSEFDYIFGMDEDNMSELRR-LAP	100
Danio rerio	61	LA-CLRKH-GIETDHRARQVTKDDFMSFDYILCMDESNLRDLNK-KAS	105

Supplemental Fig. 2. Alignment of the protein sequence of the *E*. *histolytica* LMW-PTPs with other selected LMW-PTP sequences. The *E. histolytica* LMW-PTP protein XP_656359 sequence (Ent his 1) was aligned with other LMW-PTPs. As compared with XP_656359: Identical amino acids, conservative substitutions, active site, DPYY loop. Supplemental Figs 3 and 4 show phylogenetic trees for all selected LMW-PTPs without and with the PTP1B outgroup. Supplemental Data Set 2 lists tabulates species, LMW-PTP accession numbers, and percent identity and similarity to the *E. histolytica*

LMW-PTP protein XP_656359 for each LMW-PTP and PTP1B homologous sequence for each organism.

2



Supplemental Fig. 3. A phylogenetic tree comparing the E. histolytica LMW-PTP protein XP 656359 and selected LMW-PTP proteins from different species. Species abbreviations, LMW-PTP accession numbers, and percent identity and similarity to the E. histolytica LMW-PTP protein XP 656359 for each LMW-PTP sequence are found in Supplemental Data Set 2, and LMW-PTP protein alignments are shown in Supplemental Fig. 2. The phylogenetic tree was created with "one-click" phylogenic analysis using online software [www.phylogeny.fr] by inputting the protein sequences for the selected LMW-PTPs. Selected LMW-PTPs are color-coded in terms of the general phylogenetic groups into which they fall: cyanobacteria, multicellular animals, unicellular eukaryotes, bacteria, and plants. The scale bar at the bottom of the figure indicates the branch length representing the amount of genetic change

(the number of nucleotide substitutions divided by the length of the protein sequence)

[http://epidemic.bio.ed.ac.uk/how_to_read_a_phylogeny]. The numbers in blue (between 0 and 1) at each node indicate the measure of support for the node (the node represents a putative ancestor), with 1 being maximal support. The phylogeny.fr web site uses the approximate likelihood-ratio test (aLRT) for branches to calculate this figure [Anisimova M, Gascuel O. Approximate likelihood-ratio test for branches: a fast, accurate, and powerful alternative. Syst Biol 2006; 55(4):539-52]. LMW-PTP sequences from cyanobacteria are shown in blue, those from multicellular animals are shown in violet, unicellular eukaryotic LMW-PTP sequences are red, bacterial LMW-PTPs are gray, and plant LMW-PTPs are green.



Supplemental Figure 4. A phylogenetic tree comparing the E. histolytica LMW-PTP protein and selected LMW-PTP proteins from different species, and including as an outgroup proteins homologous to PTP1B from those same **species.** Species abbreviations are in Supplemental Data Set 2, and LMW-PTP protein alignments are shown in Supplemental Fig. 2. The phylogenetic tree was created using online software (www.phylogeny.fr) by inputting the protein sequences for the selected LMW-PTP and PTP1B homologues. The human PTP1B (isoform 1) sequence was input into BLAST [http://blast.ncbi.nlm.nih.gov/Blast.cgi] to identify the E. histolytica homologue; BLAST was then used to identify PTP1B homologues in the same organisms chosen for LMW-PTP comparison using the E. histolytica homologue. Sequences are labeled as LMW-PTP or PTP1B following the species abbreviation, and are color-coded in terms of the general phylogenetic groups into which they fall: blue for cyanobacteria, violet for multicellular animals, red for unicellular eukaryotes, gray for bacteria, and green for plants. The scale bar at the bottom of the figure indicates the branch length representing the amount of genetic change (the number of nucleotide substitutions divided by the length of the protein sequence). The numbers in blue (between 0 and 1) at each node indicate the measure of support for the node (the node represents a putative ancestor), with 1 being maximal support. Accession numbers, and percent identity and similarity to the E. histolytica LMW-PTP protein XP_656359 for each LMW-PTP and PTP1B sequence are found in Supplemental Data Set 2. Supplemental Fig. 3 shows the tree without the PTP1B outgroup, and more information is found in its legend.

Protein sequence of hypothetical protein 328t.00002 (gi56465624)

MNAFILLFIALATAEDIYSPFKRSCVIQ NRINTITAQTKAIHNKIRADEQQERNLN DSLYYLYRDLKVATVAEDRARLELEIEA IQKKLISVRIHKVNMLRDIRKETDKITA PFRDSIARSNKIERHVGLEDEGNFVDER LKTQATIKKVTSQLANKYATLAAKIAAQ SNVAKÄAKKFNNDAAKVQAFVAKKAEAA YTKVHSKIVKAISAAVENGVARNNVETV AKTAIEQIVDKLMMKVAAQPSVEAADKV AQNIVKSQLKQIKKAVKVTPKKVVKKPV

Peptide Sequence
(K) YATVAEDR (A)
(K) VTSQLANK (Y)
(K) VAAQPSVEAADK (V)
(K) TAIEQIVDK(L)
(R) NNVE TVAK (T)
(R) LELEIEAIQK (K)
(K) KVTSQLANK (Y)
(K) KFNNDAAK (V)
(R) INTITAQTK (A)
(K) IAAQSNVAK (A)
(R) HVGLEDEGNFVDER (L)
(K) AISAAVENGVAR (N)

Peptide Sequence

(K) ENFAGECLER (I)

(K) GNVLFETcK(E)

(R) IKDMDDAYK (H)

(K) MAEOISEGIK(S)

(K) TLDLGNYHLK (F)

(R) VFNDQmmER(I)
(K) VLVScDVFGGHYADER(V)

Peptide sequence

(K) DFEVVEK (L)

(K) AANILVSDDGVcK(L)

(K) AFNDFIALccVK(N)

(K) DTLHGLAYLHK (M)

(R) ERPSALELLK (C)

(R) IDGDTTDVEK(E)
(R) ITDLIEKENEVIAEFGSR(K)

(K) LADFGVSTK (M)

(K) MFNIVESSVGDVLK(E)

(K) VVIAYGSAYGYTQK(M)
(K) YPEATVIGTmAALNNIK(Y)

(R) AIFLIPNRPPPTLSTPEEVSK (A)

(K) HYFDcIFGPFK (N)

(K) FQPNSNELQTcFEWGK (K)

(R) FVQcFGSFGWSGEGVK (N)

(K) GLDMIE TOMGFPSDELK (A)

(K) GNVLFETcKENFAGEcLER(I)

Protein sequence of the conserved hypothetical protein (gi56465028)

MEELINKNNQLIKEVVETLPNIQQSIDDVN NNYQSYLESIKNEVFELKENETYSEAIKRG EGIFDRMDETERRRKLVIEKSQETINQIEI MKNKISK<mark>IIEECNNDEETFGKT</mark>RNELIDKI ISIEEMIMKDKLFRRTEETDTITSEFNKK EEWKKHYSDYLERKKRKEEEEEKERKKQEE ERLQIMKGMNSMEEMRQIEEWTNKKVGNVL FESDVDDWNKNTSVFDQKIKKKEHIMIVIE DEEGNKFGGYVNSKIDEVESCITDQKSFLF SLESKGRMKRMMKFDIKEPQCAFYLNSQTD DCLFEFGFGHDICVFKENNKTQSYCYQRSF EYNGISNALCGKQLPERFTQKRIMVIEMK

Protein sequence of the putative ARP2/3 complex 34 kda subunit (gi56471818)

MAPNLLLESYNRIVYARVLEKLTTEGGIKK GLKEKLADFNQSIWYIERTKEKMTISVSTP CWKQLKENRSLEFLQKVYGSDLVEADSNSE YNLILTIPEKHDKPEEFAMNAAKLITNNLI GPAVVLADEVKNNKADEKLVQIDYRPGESY WLKPNGDRLTVIFSIKFDDKDDAVFGRVFI NEFSKSAAGCPSCDVVTRKNAPPPSELKTV EGLAEDNCYISFLLEKRHLNNPQKTLEVLM SCRNYINFHIKCSKA

Protein sequence of the hypothetical protein 503.t00001 (gi56464309)

MSGIGDAFGRKFYQIKTHVGAGQKTMDSDV QYARNKLSESYKKFKNILDVIKKLAPTVHA TNLMQVEVLTSLGDCVNTSPETKSDIDSI ISTFQKIDEGVNTYETRIESDIIVPLKTYM EQFKVMEKRFEICHNRRVDMDRYHDSVLSI SKKPPGKQSGLGEAQNKYNVARDLYNVLRN EIIADVEKLTSSAEEVVSPICGTLIVSYTD YLNHLNNYWGEASEVSSNFRICALDPAPII TPGESSMVIDANVFSKKSSDVMGGTYEPSE FENSGSSTTTSTTTTTTTVSTSSGKRPPP PPPSRKEQVRCEVEYQAQEQGELSFKEGDI ITVLKKEGDWWLGELKDNKDTSHTIMFLHF KFTFELKDYFISFSSLFIFQNKEL

Protein sequence of the putative ribosephosphate pyrophosphokinase (gi56473321)

MSVTTSLSAEQWQHDHFIVHTSKAGYLAQS ISAVLGYPIIPIIHK<mark>KFSDGECYYRI</mark>ATPE RTSLVGRDVVVVSSIVDDQELLEVVRIGGE LAELGTKRRIFVIPYLMYSTMERAVHPGEV VTCKSTVRMLCGIPSSGLGNLFMLMDLHTS GIIHYFEGTVQAMELYAESTLEQAIAENID FSEPVIFGSADLGRPLWVETFANHFGVGIA FIRKSRSFEDTHVIGEPIGDVKGKHVVIYD DLTRSAGSLIKACNAYDLNGAIKVTAVVTH LSLIDDDVVQKVIDSRIDRVIATNSVRSQ MDAVKNSKKFIICDISPVFAKQIRACLNNN ND Peptide Sequence (K) EGDWWLGELK (D) (K) IDEGVNTYETR (I) (R) IESDIIVPLK (T) (W) COST CTOOWL (W)

Peptide Sequence

(K) GmNSMEEMR (O)

(K) NTSVFDOK(I)

Peptide Sequence

(R) SLEFLQK (V)

(K) TLEVLmScR(N)

(R) VFINEFSK(S)

(K) FDDKDDAVFGR (V)

(K) LADFNOSIWYIER (T)

(K) SAAGcPScDVVTR (K)

(K) LLTNMLIGPAVVLADEVK (N)

(K) VGNVLFESDVDDWNK (N)

(K) IIEEcNNDEETFGK(T)

(R) RTEETDTITSEFNK (K)

(R) SFEYNGISNALcGK (Q)

(K) KVGNVLFESDVDDWNK(N)

(K) IDEGVNIVETR(I) (K) IDEGVNIVETR(I) (R) IESDIIVPLK(T) (K) QSGLGEAQNK(Y) (K) SDIDSIISTFQK(I) (K) TmDSDVQYAK(N) (R) YHDSVLSISK(K)

Peptide Sequence (R) AVHPGEVVTcK (S) (K) FSDGEcYYR (I) (K) KFIIcDISPVFAK (Q) (R) VIATNSNVR (S)

Protein sequence of type A flavoprotein (gi56474182)

MKALEVVKDLYWVGVFDKELRVFDIIMTTP YGTSYNSFLIKSEKGNVLFETCKENFAGEC LERIEDVIGKEGKLDYIVLNHTEPDHSGSL VHILEKYPEATVIGTMAALNNIKYIGHIKE NTKTLNSGKIKTLDLGNYHLKFLIQPFLHW PDTMMTVIEEMKVLVSCDVFGGHYADERVF NDQMMERIKDMDAYKHYFDCIFGFENYV IKGLDMIETQMGFPSDELKAICCSHGPVLR THIKENIERYRQMAQFIALKNKVVIAYGSA YGYTQKMAEQISEGIKSTGVEVKMFNIVES SVGDVLKEFEDAKGLLLGTPTLVNDTIPPI MQIACSLNPTIHCNRFVQCFGSFGWSGEGV KNLSARIVQLKVHQPVEPLSIKFQPNSNEL QTCFEWGKKFAEALKA

Protein sequence of the putative serine/threonine protein kinase (gi56466736)

MTTPTKLQVSQMQKINWVAECSKDFEVVEK LGEGSYGSVFHAIHRPTGTHAAVKTVRIDG DTTDVEKE IVLMKQCTDYIVRFGHFFVD KDLWIVMEICGAGSVCDIMNILNSTLDERE IQIILKDTLHGLAYLHKMHMIHRDIKAANI LVSDDGVCKLADFGVSTKMGGTNKLQKTLI GTPYWMAPEVILESGHNEKADIWSVGITAI EMFDGRPHSELHAMKAIFLIPNRPPTLS TPEEVSKAFNDFTALCVKNFREPSALEL LKCDFMKPNGNRKRITDLIEKENEVIAEFG SRKKALEQGEDDDEGDDTWIKK

Supplemental Fig. 5. Sequences of unique peptides and their locations in the protein sequences of putative substrate proteins identified by mass spectrometry and selected for further evaluation. Locations of the unique peptides are shown in yellow in the protein sequence, and the list of unique peptides identified for each protein is located in the table to the right of each protein sequence.

Hypothetical protein 328.t00002 (gi56465624; XM_644144)

- 1 MNAFILLFIALATAEDI<mark>X</mark>SPFKRSCVIQNRINTITAQTKAIHNKIRADEQ 51 OERNLNDSLYYLYRDKK<mark>X</mark>ATVAEDRARLELEIEAIOKKLHSVRIHKVNML
- 101 RDIRKETDKITAPFRDSIARSNKIERHVGLEDEGNFVDERLKTQATIKKV
- 151 TSQLANKYATLAAKIAAQSNVAKAAKKFNNDAAKVQAFVAKKAEAAY
- 201 HSKIVKAISAAVENGVARNNVETVAKTAIEQIVDKLMMKVAAQPSVEAAD
- 251 KVAQNIVKSQLKQIKKAVKVTPKKVVKKPVKKAVAKKVVKKAAPKPAAKK
- 301 IAKK

Phosphorylati	ion site	es predicted	1:	Tyr: 3
Name	Pos	Context	Score	Pred
		V		
328 +00002	10	AFDTVCDFV	0 0/0	*V*

	v		
18	AEDI <mark>Y</mark> SPFK	0.948	*Y*
68	rdlk <mark>y</mark> atva	0.970	*Y*
197	AEAA <mark>Y</mark> TKVH	0.829	*Y*
	18 68 197	18 AEDI <mark>X</mark> SPFK 68 RDLK <mark>X</mark> ATVA 197 AEAA <mark>X</mark> TKVH	v 18 AEDI¥SPFK 0.948 68 RDLK¥ATVA 0.970 197 AEAA¥TKVH 0.829

Type A flavoprotein (gi56474182; XM_651854)

		-				
			v			
Name		Pos	Context	Score	Pred	
Phosp	horylati	on site	s predicte	ed:	Tyr:	4
401	ADADINA					
401	AFALKA					~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
351	GSFGWSG	EGVKNLS	ARIVOLKVHC	PVEPLSIF	FOPNSI	NELOTCFEWGKKF
301	SVGDVLK	EFEDAKG	LLLGTPTLVN	IDTIPPIM	DIACSL	NPTIHCNRFVQCF
251	RQWAQPI	ALKNKVV	IAYGSAYGYI	QKMAEQIS	SEGIKS	TGVEVKMFNIVES
200	CIFGPFK	N <mark>X</mark> VIKGI	DMIETQMGFE	PSDELKAI	CCSHGP	VLRTHIKENIERY
131	EDIMITV.	LEEPICVL	WSCDVE GGH	ADERVENT	QUUER	INDEDDA
151	DDTMMTV	TEEMENT	VSCDVFCCU	ADERVENI	OMMER	
101	TVIGTMA	ALNNIKY	IGHIKENTKI	LNSGKIKI	LDLGN	YHLKFLIQPFLHW
51	TCKENFA	GECLERI	EDVIGKEGKI	D <mark>Y</mark> IVLNH	CEPDHS	GSLVHILEKYPEA
1	MKALEVV	KDLYWVG	VFDKELRVFD	DIIMTTPYC	STSYNS	FLLKSEKGNVLFE

		v		
Flavopr	76	GKLD <mark>Y</mark> IVLN	0.821	*Y*
Flavopr	174	FGGH <mark>Y</mark> ADER	0.795	*Y*
Flavopr	195	MDDA <mark>y</mark> khyf	0.513	*Y*
Flavopr	209	PFKN <mark>Y</mark> VIKG	0.574	*Y*
		~		

Supplemental Fig. 6. Protein sequences and predicted tyrosine phosphorylation sites in selected putative substrate proteins. Tyrosine phosphorylation sites were predicted using the NetPhos 2.0 server at

Serine/threonine protein kinase, putative (gi56466736; XM_645070)

MTTPTKLQVSQMQKINWVAECSKDFEVVEKLGEGS<mark>Y</mark>GSVFHAIHRPTGTH 1 51 AAVKTVRIDGDTTDVEKEIVLMKQCDTD**Y**IVRFFGHFFVDKDLWIVMEIC 101 GAGSVCDIMNILNSTLDEREIQIILKDTLHGLAYLHKMHMIHRDIKAANI LVSDDGVCKLADFGVSTKMGGTNKLQKTLIGTPYWMAPEVILESGHNEKA 151 201 DIWSVGITAIEMFDGRPPHSELHAMRAIFLIPNRPPPTLSTPEEVSKAFN 251 DFIALCCVKNFRERPSALELLKCDFMKPNGNRKRITDLIEKENEVIAEFG 301 SRKKALEQGEDDDEGDDEQDDDDTEEDDTEEEEDDTEEDDSDDGDDTM 351 VIRK Phosphorylation sites predicted: Tyr: 2

Name	Pos	Context	Score	Prec
ProtKina	36	GEGS <mark>Y</mark> GSVF	0.816	*Y*
ProtKina	79	CDTD <mark>y</mark> ivrf	0.809	*Y*

Conserved hypothetical protein (gi56465028, XP_001914335)

1	MEELINKNNQLIKEVVETLPNIQQSIDDVNNNYQSYLESIKNEVFELKEN
51	ET <mark>¥</mark> SEAIKRGEGIFDRMDETERRRKLVIEKSQETINQIEIMKNKISKIIE
101	ECNNDEETFGKTRNELIDKIISIEEMIMKDKLFRRTEETDTITSEFNKKK
151	EEWKKH <mark>Y</mark> SDYLERKKRKEEEEEKERKKQEEERLQIMKGMNSMEEMRQIEE
201	WTNKKVGNVLFESDVDDWNKNTSVFDQKIKNKEHIMIVIEDEEGNKFGG <mark>Y</mark>
251	VNSKIDEVESCITDQKSFLFSLESKGRMKRMMKFDIKEPQCAFYLNSQTD
301	DCLFEFGFGHDICVFKENNKTQSYCYQRSFEYNGISNALCGKQLPERFTQ
351	KRIMVIEMK
Phosp	phorylation sites predicted: Tyr: 3

Name	Pos	Context	Score	Pred
		v		
Cons hyp	53	ENET <mark>Y</mark> SEAI	0.927	*Y*
Cons hyp	157	WKKH <mark>Y</mark> SDYL	0.562	*Y*
Cons hyp	250	KFGG <mark>Y</mark> VNSK	0.953	*Y*

Ribose-phosphate pyrophosphokinase, putative (gi56473321, XP_656141)

Rib-phos

Rib-phos

Rib-phos

1	MSVTTSLSAEQWQHDHFIVHTSKAGYLAQSISAVLGYPIIPIIHK	KFSDG
51	EC Y YRIAIPERTSLVGRDVVVVSSIVDDQELLEVVRIGGELAELG	「KRRI
01	FVIPYLMYSTMERAVHPGEVVTCKSTVRMLCGIPSSGLGNLFMLM	DLHTS
51	GIIHYFEGTVQAMEL <mark>Y</mark> AESTLEQAIAENIDFSEPVIFGSADLGRP	LWVET
01	FANHFGVGIAFIRKSRSFEDTHVIGEPIGDVKGKHVVIYDDLTRS	AGSLI
51	KACNA <mark>Y</mark> LDNGAIKVTAVVTHLSLIDDDVVQKVIDSRIDRVIATNS	NVRSQ
01	MDAVKNSKKFIICDISPVFAKQIRACLNNNND	
hosp	horylation sites predicted: Tyr: 3	
ame	Pos Context Score Pred	

DGECYYRIA 0.784 *Y*

166 AMEL<mark>Y</mark>AEST 0.956 *Y* 256 ACNA<mark>Y</mark>LDNG 0.947 *Y*

Calcium-binding protein, putative, URE3-sequence specific DNA binding protein (URE3-
BP) (AAG18423).

1	MQPPVAN	FCLWNL	2PIQGSWMGAAC	TYOWPPS	20 KNIWWF.PLL	N.L.T.b.
51	TRIYQWF	MGVDRDI	RSGTLEINELMM	GQFPGGI	RLSPQTALRM	MRIF
101	NGHISFY	EFMAMYI	KFMELAYNLFVM	NDRNRSG	TLEPHEILPAI	LQQL
151	NQRTSLL	LHRLFA	RGMAFCDLNCWI	AICAFAA	QTRSAYQMIFN	4NPY
201	KPFNPME	FGKFLD	VVTSLLE			
Phosp	horylati	on site	es predicted	:	Tyr: 2	
Phosp Name	horylati	on site Pos	es predicted Context	: Score	Tyr: 2 Pred	
Phosp Name URE3-	borylati	on site Pos	es predicted Context V PLDO Y TRIY	Score	Tyr: 2 Pred	
Phosp Name URE3- URE3-	BP BP	on site Pos 50 197	es predicted Context 	: Score 0.804 0.698	Tyr: 2 Pred * <u>Y</u> * * <u>Y</u> *	

http://www.cbs.dtu.dk/services/NetPhos/. URE3-BP was included as a control. Protein sequences are shown with predicted phosphorylated tyrosines in bold and highlighted, with their surrounding residues in a table below. The closer the score is to a value of one, the more likely the phosphorylation of that tyrosine residue.

ARP2/3 complex 34 kDa subunit, putative (gi56471818, DS571335)

- 1 MAPNLLLESYNRIVYARVLEKLTTEGGIKKGLKEKLADFNQSIWYIERTK 51 EKMTISVSTPCWKQLKENRSLEFLQKVYGSDLVEADSNSE<mark>Y</mark>NLILTIPEK
- 101 HDKPEEFAMNAAKLLTNMLIGPAVVLADEVKNNKADEKLVOIDYRPGES
- 151 WI.KPNGDRI.TVIESIKEDDKDDAVEGRVEINEESKSAAGCPSCDVVTRKN
- 201 APPPSELKTVEGLAEDNC¥ISFLLEKRHLNNPQKTLEVLMSCRN¥INFHI
- 251 KCSKAFLHIRMRNKVSHLQLVLNRAKPEREVEKKTASGRTFKK

Phosphorylation sites predicted: Tyr: 4

d

Hypothetical protein 503.t00001 (gi56464309, XP_656019)

1	MSGIGDAFGRKFYQIKTHVGAGQKTMDSDVQ Y AKNKLSES <mark>Y</mark> KKFKNILDV
51	IKKLAPTVHATNLMQVEVLTSLGDCVVNTSPETKSDIDSIISTFQKIDEG
101	VNTYETRIESDIIVPLKTYMEQFKVMEKRFEICHNRRVDMDR <mark>Y</mark> HDSVLSI
151	SKKPPGKQSGLGEAQNKYNVARDLYNYLRNEIIADVEKLTSSAEEVVSPI
201	CGTLIVSYTD Y LNHLNNYWGEASEVSSNFRICALDPAPIITPGESSMVID
251	ANVESKKSSDVMDGTYEPSEFENSGSSTTTSTTTTTTVSTSSGKRPPPP

301 PPSRKEQVRCEYE

351 SHTIMFLHFKFTFELKDYFISFSSLFIFQNKEL

Phosphorylation sites predicted: Tyr: 6

Name	Pos	Context	Score	Pred
		v		
501.t	32	SDVQ <mark>Y</mark> AKNK	0.957	*Y*
501.t	41	LSES <mark>Y</mark> KKFK	0.519	*Y*
501.t	143	dmdr <mark>y</mark> hdsv	0.971	*Y*
501.t	211	SYTD <mark>Y</mark> LNHL	0.927	*Y*
501.t	266	MDGT <mark>Y</mark> EPSE	0.607	*Y*
501.t	314	CEYE <mark>Y</mark> QAQE	0.575	*Y*
		~		

Supplemental Table 1. Crystallographic X-ray data and refinement statistics. Crystallized recombinant *E. histolytica* LMW-PTP representing different states along the catalytic pathway were subjected to X-ray diffraction in order to solve their structures: the enzyme with no ligand bound (apo structure), with the substrate mimic HEPES bound, and with the phosphate analog sulfate bound. Citation for MolProbity: Davis IW, Leaver-Fay A, Chen VB, Block JN, Kapral GJ, Wang X, et al. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. Nucleic Acids Res 2007; 35(Web Server issue):W375-W383.

	Apo structure	HEPES (form 1)	HEPES (form 2)	Phosphate analog
PDB ID	3ILY	3IDO	3JS5	3JVI
Crystal parameters				
Space group	P4 ₃	$P2_{1}2_{1}2_{1}$	$P2_{1}2_{1}2_{1}$	P4 ₃ 2 ₁ 2
Cell dimensions	45.7, 45.7, 132.3	60.6, 70.4, 79.2	38.6, 60.3, 72.56	45.0, 45.0, 133.3
$a=b=c$ (Å), $\alpha=\beta=\gamma$ (°)	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90
Data set				
X-ray Source	FRE+	ALS 5.0.3	MicroMax 007 HF	FRE+ Superbright
	Superbright			
Wavelength (Å)	1.54	0.9765	1.5418	1.5418
Resolution (Å)	50-2.2	50-2.2	50-1.95	50-1.80
	(2.28-2.20)	(2.28-2.20)	(2.02-1.95)	(1.86-1.80)
Rmerge	0.063 (0.316)	0.140 (0.492)	0.066 (0.198)	0.059 (0.231)
I/sigma (I)	23.6 (4.3)	14.0 (2.4)	19.6 (4.7)	17.7 (4.1)
Completeness	98.3% (85.4%)	98.6% (89.5%)	97.9% (84.9%)	94.6% (98.2%)
# reflections, unique	13,504 (1154)	17,494 (1539)	12,697 (1069)	23,136
Multiplicity	4.5 (3.4)	8.1 (5.7)	4.4 (2.2)	4.3 (2.5)
Refinement statistics				
R _{work}	0.211 (0.245)	0.206 (0.281)	0.169 (0.215)	0.206 (0.260)
R _{free}	0.264 (0.325)	0.254 (0.315)	0.212 (0.244)	0.249 (0.361)
RMSD bond lengths (Å)	0.016	0.010	0.009	0.012
RMSD bond angles (Å)	1.529	1.211	1.087	1.273
Ramachandran:				
Preferred	96.2%	98.7%	98.1%	97.4%
allowed	99.3%	100%	100%	99.4%
disallowed	0.7%	0%	0%	0.6%
MolProbity clash score	7.82 (95 th)	4.94 (98 th)	4.00 (98 th)	$10.22(72^{nd})$
Molprobity score	21.8 (75 th)	1.56 (98 th)	1.19 (99 th)	2.01 (62 nd)