

Supporting Information

Selective MAP kinases proteolysis determined by activation loop flexibility

Lihi Gur-Arie¹, Maayan Eitan-Wexler², Nina Weinberger², Ilan Rosenshine¹, Oded Livnah²

¹Department Microbiology and Molecular Genetics, IMRIC, Faculty of Medicine, and

²Department of Biological Chemistry, Alexander Silverman Institute of Life Sciences,
The Wolfson Centre for Applied Structural Biology, The Edmond J. Safra Campus, The
Hebrew University of Jerusalem.

Hamiltonella_defensa (HD)	MR-RALKLNLTPLSFSS-----SSNSDTDQSLVKKSSVNWKNNHQLSFDGTONKNI
Erwinia_piriflorinigrans	MRPTSLQLTQPVLNPSSS-----DPASDIQSLVKISGVNWVWNNNQQLSFIGTDYKN
Shigella_boydii	MRPTSLNLALHQSS-----TSSSMSDADIESLVKTSSVQWIKNNNPQLRFQGTDHNI
Enterohemorrhagic E.coli (EHEC)	MRPTSLNLVHQSS-----RSSSMSDTDIESLVKASSVQWIKNNNPQLRFQGTDHNI
Enteropathogenic E.coli (EPEC)	MRPTSLNLVHQSS-----TSSSMSDTDIESLVKASSVQWIKNNPQLRFQGTDHNI
Escherichia_albertii	MRPTSLNLALHQSS-----TSSSMSDTDIESLVKASSVQWIKNNPQLRFQGTDHNI
Providencia_alcalifaciens	MRPTSLKLTLPSLPTSSS-----HSSSATDIDQYLVKMSSVSWIKNNNQQLHFGLGTDHKI
Izakielia_capsodis	MRPTFLNLTPSPSSLFSNSNSNSTTSDIDQSLVKMRSRVHVVKNNNQQLCFHGTNHKI
Salmonella_enterica_arizonae (SEA)	MRPTSLKLTLPSLHLPSSS-----NS-ATDIQSLVKRSGVWRWVKNNNQQLSFHGTDHKI
Pantoea_stewartii	MRPTSLNLVLP-----NSISATDVQSLVKMSSVLWIKNQNQQLCFHGTDHKI
Salmonella_enterica_Sandiego	MRPTSLKLTLPSLHLPSSS-----NSISAIIDQSLVKRSGVWRWVKNNNQQLSFHGTDHKI
Leclercia	MRPTSLNLTLPSLPLPSSS-----NSISATDIDQSLVKRSGVWRWVKNNNQQLSFHGTDHKI
Salmonella_enterica_Braenderup	MRPTSLNLTLPSLPLPSSS-----NSVSATDIDQSLVKMSGVWRWVKNNNQQLCFHGDLKI
Salmonella_enterica_salamae	MRPTSLNLTLPLPLPSSS-----NSVSATDIDQSLVKMSGVWRWVKNNNQQLCFHGDLKI
Citrobacter_rodentium (CR)	MRPTSLNLTLPSLPLPSSS-----NSISATDIDQSLVKMSGVWRWVKNNNQQLCFHGDLKI
Erwinia_tracheiphila	MRPTSLNLTLPSLPLPSSS-----NSISATDIDQSLVKISGVWRWVKNNNQQLCFHGDLKI
	** * : * :: * : * * : * * * * : : :
 Hamiltonella_defensa (HD)	 YQRLEALKDKESTETGKELLNCIESVSRLKSEKLIIRLDSTELGVTAAHCAENAENFRGT
Erwinia_piriflorinigrans	YRQLETALDKVKSTNTGRILLKCIALTSQLSKSKLAIFLNLSEELMVEPHCKTDAENFRGT
Shigella_boydii	YQQIEAALDKIGSTETGRVLLNAIESISRLKSETVVIHLNSSRLGVIAHRIIDAENHRGT
Enterohemorrhagic E.coli (EHEC)	YQQIEAALDKIGSTETGRVLLNAIESISRLKSETVVIHLNSSRLGVMAHRDIDAEENHRGT
Enteropathogenic E.coli (EPEC)	YQQIEAALDKIGSTETGRVLLNAIESISRLKSETVVIHLNSSRLGVMAHRDIDAEENHRGT
Escherichia_albertii	YQKLEAALDKIESTNTGRILLNSIELTSRLKSEKLVHILNSSELGVIAHCSTDAAENSRTG
Providencia_alcalifaciens	YQKLEAALDKIESTDTGRILLNCIELTSRLKSLLAIYNCTELMVVAHCDADAENSRTG
Izakielia_capsodis	YQKLEAALDKIESTDTGRILLNCIELTSRLKSEKLAELGVVAHCNTDAENARGT
Salmonella_enterica_arizonae (SEA)	YQKLEAALDKIESTNTGRILLNCIELTSRLKSEKLAELGVVAHCNTDAENARGT
Pantoea_stewartii	YQKLEAALDKIESTNTGRILLNCIELTSRLKSEKLAELGVVAHCNTDAENARGT
Salmonella_enterica_Sandiego	YQKLEAALDKIESTNTGRILLNCIELTSRLKSEKLAELGVVAHCNTDAENARGT
Leclercia	YQKLEAALDKIESTNTGRILLNCIELTSRLKSEKLAELGVVAHCNTDAENARGT
Salmonella_enterica_Braenderup	YQKLEAALDKIESTNTGRILLNCIELTSRLKSEKLAELGVVAHCNTDAENARGT
Salmonella_enterica_salamae	YQKLEAALDKIESTNTGRILLNCIELTSRLKSEKLAELGVVAHCNTDAENARGT
Citrobacter_rodentium (CR)	YQKLEAALDKIESTNTGRILLNCIELTSRLKSEKLAELGVVAHCNTDAENARGT
Erwinia_tracheiphila	YQKLEAALDKIESTNTGRILLNCIELTSRLKSEKLAELGVVAHCNTDAENARGT
	* : * * : * : * : * * : * : * . * * * * : * : * * :
 Hamiltonella_defensa (HD)	 GSYLCNCNLNVQEEESIGKGIEQTELDACTIVEHELLHVFHNLNGERLKAVFQPEVETYSP
Erwinia_piriflorinigrans	GSDFHCNFDAVE-YFFDQGMDLVEFQACVVPHELLHVFHNLNGERLRRVISQPEPQTDYP
Shigella_boydii	GSDFHCNLDNAVE-YPCGEGLSVDHFATIVEHELLHVFHNLNGERLKVESSRPESQKYS
Enterohemorrhagic E.coli (EHEC)	GSDFHCNLDNAVE-YPCGEGLSVDHFATIVEHELLHVFHNLNGERLKVESSRPESQKYS
Enteropathogenic E.coli (EPEC)	GSDFHCNLDNAVE-YPCGEGLSVDHFATIVEHELLHVFHNLNGERLKVESSRPESQKYS
Escherichia_albertii	GSDFHCNLDNAVE-YPCGEGLSVDHFATIVEHELLHVFHNLNGERLKVESSRPESQKYS
Providencia_alcalifaciens	GSDFHCNLDNAVE-YPCGEGLSVDHFATIVEHELLHVFHNLNGERLKVESSRPESQKYS
Izakielia_capsodis	GSDFHCNLDNAVE-YTNGQQGSLGDFAHIVEHELLHVFHNLNGERLKVESSRPESQKYS
Salmonella_enterica_arizonae (SEA)	GSNFHCNLDNAVE-YPCGGQGISLVDFHACTIVEHELLHVFHNLNGERLKVVESSQPEQTDCS
Pantoea_stewartii	GSDFHCNLDNAVE-YPCGGQGISLVDFHACTIVEHELLHVFHNLNGERLKVVESSQPEQTHY
Salmonella_enterica_Sandiego	GSDFHCNLDNAVE-YPCGGQGISLVDFHACTIVEHELLHVFHNLNGERLKVVESSQPEQTHY
Leclercia	GSDFHCNLDNAVE-YPCGGQGISLVDFHACTIVEHELLHVFHNLNGERLKVVESSQPEQTHY
Salmonella_enterica_Braenderup	GSDFHCNLDNAVE-YPCGGQGISLVDFHACTIVEHELLHVFHNLNGERLKVVESSQPEQTHY
Salmonella_enterica_salamae	GSDFHCNLDNAVE-YPCGGQGISLVDFHACTIVEHELLHVFHNLNGERLKVVESSQPEQTHY
Citrobacter_rodentium (CR)	GSDFHCNLDNAVE-YPCGGQGISLVDFHACTIVEHELLHVFHNLNGERLKVVESSQPEQTHY
Erwinia_tracheiphila	GSDFHCNLDNAVE-YPCGGQGISLVDFHACTIVEHELLHVFHNLNGERLKVVESSQPEQTHY
	* : * * : : * : : . : * * * * : * : * :
 Hamiltonella_defensa (HD)	 FLEEARTVGLGSFSEEVFSENKFREEDIGIPRVRSYAHESLILHDNTFTMFHENKQSHPLL
Erwinia_piriflorinigrans	LLLEEARTEVGLGAFSEEVLSNKFRKEIGAPRTSYFYDLYVNDNTITGMGYERKHLHPLL
Shigella_boydii	LLLEEARTEVGLGAFSEEVLSNKFRFREEIGMPRTSYFPHDSALIHDDNTVSQGFLQVRLHPLL
Enterohemorrhagic E.coli (EHEC)	LLLEEARTEVGLGAFSEEVLSNKFRFREEIGMPRTSYFPHDSALIHDDNTVSLGFQQVRLHPLL
Enteropathogenic E.coli (EPEC)	LLLEEARTEVGLGAFSEEVLSNKFRFREEIGMPRTSYFPHDSALIHDDNTVSLGFQQVRLHPLL
Escherichia_albertii	LLLEEARTEVGLGAFSEEVLSNKFRFREEIGMPRTSYFPHDSALIHDDNTVSLGFQQVRLHPLL
Providencia_alcalifaciens	LLLEEARTEVGLGAFSEEVLSNKFRFREEIGMPRTSYFPHDSALIHDDNTVSLGFQQVRLHPLL
Izakielia_capsodis	LLLEEARTEVGLGAFSEEVLSNKFRVEIGIPRTFYSHDLALIHDNTVTQGFQRKHLHPLL
Salmonella_enterica_arizonae (SEA)	FLEEARTVGLGSFSEEVLSNKFRFREEIGVPRTFYPRDPYLIHDNTVTQGLQRKHLHPLL
Pantoea_stewartii	LLLEEARTEVGLGAFSEEVLSNKFRFREEIGVPRTFYPHDSSIHDNTVTQGFQRKHLHPLL
Salmonella_enterica_Sandiego	FLEEARTVGLGSFSEEVLSNKFRFREEIGVPRTFYPHNSLIIHDNTVSQGFLQRKHLHPLL
Leclercia	LLLEEARTEVGLGAFSEEVLSNKFRFREEIGMPRTSYFPHDSSIHDNTVTQGFQRKHLHPLL
Salmonella_enterica_Braenderup	FLEEARTVGLGSFSEEVLSNKFRFREEIGMPRTSYFPHNSLIIHDNTVSQGFLQRKHLHPLL
Salmonella_enterica_salamae	LLLEEARTEVGLGAFSEEVLSNKFRFREEIGMPRTSYFPHDSSIHDNTVTQGFQRKHLHPLL
Citrobacter_rodentium (CR)	FLEEARTVGLGSFSEEVLSNKFRFREEIGMPRTSYFPHDSSIHDNTVTQGFQRKHLHPLL
Erwinia_tracheiphila	LLLEEARTEVGLGAFSEEVLSNKFRFREEIGMPRTSYFPHDSSIHDNTVTQGFQRKHLHPLL
	* : * * : * : * : * : * : * : * : * : :

Figure S1. NleD ortholog protein sequences. NleD protein sequences, as determined by PSI-Blast search results (NCBI). Sequences were aligned using Clustal Omega (Sievers et al., 2011). The HELLH Zn-metalloprotease motif is encircled in green. An asterisk (*) indicates a fully conserved residue. A colon (:) and a period indicates conservation of strongly similar and weakly similar properties respectively.



Figure S2: Sequence alignment of JNK2, p38 α , and ERK2 MAP kinases. The regions which were chosen for replacement are colored as detail below:Marked in yellow: activation loop; bolded: TXY motif- brown: α EF/ α F loop -green: L13- α G -blue: MKI. Magenta: part of L16

Table S1– Bacterial strains used in this study. The strain name in our collection is brackets.

Name	Description	Reference or source
E2348/69	EPEC wild-type, strain E2348/69	James Kaper
KB4298	E2348/69 $\Delta nleD::kan$	(Baruch et al., 2011)
BL21(DE3)	F-, <i>ompT</i> , <i>hsdSβ(rβ-mβ-)</i> , <i>dcm</i> , <i>gal</i> , (DE3) <i>tonA</i>	Novagen
KB4583	<i>Salmonella enterica</i> serovar Arizona (SEA) strain CDC346-86 SARC5	ATCC
ICC168 (4368)	<i>Citrobacter rodentium</i>	Gad Frankel
Rosetta	<i>E. coli</i> strain Rosetta	Novagen

Table S2: Plasmids used in this study

Name	Description	Reference or source	Name in our collection
JNK2	pET28a, bacterial expression vector encoding hexa-His tag JNK2	David Engelberg	pLG4470
p38 α	pET28a, bacterial expression vector encoding hexa-His tag p38 α	This study	
ERK2	pET15b, bacterial expression vector encoding hexa-His tag ERK2	David Engelberg	pLG4472
NleD	pET28a, bacterial expression vector encoding hexa-His tag truncated <i>nleD</i> EPEC 19-232	This study	pLG6728
ERK _{A-JNK}	pET15b, bacterial expression vector encoding ERK2 with JNK2 activation loop (residues 169-188 from JNK2)	This study	pLG6725
ERK _{A-p38}	pET15b, bacterial expression vector encoding ERK2 with p38 α activation loop (residues 167-185 from p38 α)	This study	pLG6726
p38 _{A-ERK}	pET28a, bacterial expression vector encoding p38 α with ERK2 activation loop (residues 164-188 from ERK2)	This study	pLG6727
ERK _{G-p38}	pET15b, bacterial expression vector encoding ERK2 with G-helix from p38 α (residues 223-229 from p38 α)	This study	
ERK _{G-JNK}	pET15b, bacterial expression vector encoding ERK2 with G-helix from JNK2 (residues 225-231 from JNK2)	This study	
ERK _{G,M,L-p38}	pET15b, bacterial expression vector encoding ERK2 with G-helix, MKI and L16 from p38 α (residues 223-229, 243-262, 327-333, respectively, from p38 α)	This study	
ERK _{G,M,L-JNK}	pET15b, bacterial expression vector encoding ERK2 with G-helix, MKI and L16 from JNK2 (residues from 225-231, 245-262, 342-348, respectively, from JNK2)	This study	
ERK _{F-p38}	pET15b, bacterial expression vector encoding ERK2 with α EF/ α F loop from p38 α (residues 196-202 from p38 α)	This study	
ERK _{G,F-p38}	pET15b, bacterial expression vector encoding ERK2 with α EF/ α F loop and G-helix from p38 α (residues	This study	

	196-202 and 223-229, respectively, from p38 α)		
ERK _{G,F,M,L-p38}	pET15b, bacterial expression vector encoding ERK2 containing ERtrp38+ ER α EF/ α Fp38	This study	
p38 α D316N	pET28a, bacterial expression vector encoding hexa-His tag p38 α with D316N	Lab collection	
pET28a	Bacterial expression vector	Novagen	pLG4483
NleD EPEC	pET28a, bacterial expression vector encoding <i>nleD</i> _{EPEC}	This study	pLG4480
pSA10	pKK177-3 derivative containing <i>lacI</i> ^q . bacterial expression vector.	(Schlosser-Silverman et al., 2000)	1060
NleD EPEC	pSA10, bacterial expression vector encoding <i>nleD</i> _{EPEC}	(Baruch et al., 2011)	pKB4345
NleD CR1	pSA10, bacterial expression vector encoding <i>nleD</i> _{1CR}	This study	pKB4505
NleD CR2	pSA10, bacterial expression vector encoding <i>nleD</i> _{2CR}	This study	pKB4506
NleD SEA	pSA10, bacterial expression vector encoding <i>nleD</i> _{SEA}	This study	pSK4611
NleD HD	pSA10, bacterial expression vector encoding <i>nleD</i> _{HD}	This study	pSK4702
GST-7Ala-GFP	pGEX2T, bacterial expression vector encoding GST – 7 alanine loop – GFP	This study	pLG5088
GST- JNK ₁₆₉₋₁₉₂ -GFP	pGEX2T, bacterial expression vector encoding GST –JNK2 _{D169 – R192} – GFP	This study	pKB4699
GST- JNK ₁₇₆₋₁₉₂ -GFP	pGEX2T, bacterial expression vector encoding GST –JNK2 _{A176 – R192} – GFP	This study	pLG5011
GST- JNK ₁₈₀₋₁₉₂ -GFP	pGEX2T, bacterial expression vector encoding GST –JNK2 _{F180 – R192} – GFP	This study	pLG5102
GST- JNK ₁₇₆₋₁₈₈ -GFP	pGEX2T, bacterial expression vector encoding GST –JNK2 _{A176 – T188} – GFP	This study	pLG5101
GST- JNK ₁₈₀₋₁₈₆ -GFP	pGEX2T, bacterial expression vector encoding GST –JNK2 _{F180 – V186} – GFP	This study	pLG5086
GST- JNK ₁₈₂₋₁₈₈ -GFP	pGEX2T, bacterial expression vector encoding GST –JNK2 _{M182 – T188} – GFP	This study	pLG5087
GST- ERKA-GFP	pGEX2T, bacterial expression vector encoding GST-ERK1 ₁₈₃₋₂₁₇ – GFP	This study	pLG5103
JNK2	pGEX2T, bacterial expression vector encoding hexa-His tag JNK2	This study	pLG4901

JNK2 P184E	pGEX2T, bacterial expression vector encoding hexa-His tag JNK2 mutant P184E	This study	pLG5186
p38β	pET28a, bacterial expression vector encoding hexa-His tag p38β	David Engelberg	pLG4495
ERK1	pET28a, bacterial expression vector encoding hexa-His tag ERK1	This study	pLG4774
JNK1	pSRα3, mammalian expression vector encoding HA-JNK1,	Eitan Shaulian	pLG4487
JNK2	pSRα3, mammalian expression vector encoding HA-JNK2	Eitan Shaulian	pLG4488
p38α	pcDNA3, mammalian expression vector encoding HA-p38α	David Engelberg	pLG4683
p38β	pcDNA3, mammalian expression vector encoding HA-p38β	David Engelberg	pLG4684
p38γ	pcDNA3, mammalian expression vector encoding HA-p38γ	David Engelberg	pLG4685
p38δ	pcDNA3, mammalian expression vector encoding HA-p38δ	David Engelberg	pLG4686
ERK1	pcEFL, mammalian expression vector encoding hexa-His ERK1	David Engelberg	pLG4675
ERK2	pcEFL, mammalian expression vector encoding HA-ERK2	David Engelberg	pLG4676
JNK2 D326A	pSRα3, mammalian expression vector encoding HA-JNK2 with D326A mutation	This study	pLG6680
JNK2 E329A	pSRα3, mammalian expression vector encoding HA-JNK2 with E329A mutation	This study	pLG6681
JNK2 E331A	pSRα3, mammalian expression vector encoding HA-JNK2 with E331A mutation	This study	pLG6682
JNK2 D326A/ E329A/ E331A	pSRα3, mammalian expression vector encoding HA-JNK2 with the mutations D326A, E329A and E331A	This study	pLG6683

Table S3 – The primers used in this study

Plasmid number	Primer sequence
pLG4480	CAGAATCATATGCGCCCTACGTCCCTC AGAATTCAAGCTAAAGCAATGGATGCAG
pLG4901	CggatccATGAGCGACAGTAAATGTG GGAATTCTCATCGACAGCCTCAAGG
pLG5011	CAGGGATCCCGCGTGCACTAACTTCATGATGACCCCTACGT GGTGACACGGTACTACCGGATGGTGAGCAAGGGCGAG CCAGAATTCTTACTTGTACAGCTCGTCCATG
pLG5086	CAGGGATCCTCATGATGACCCCTACGTGGTGACAATGGTGAGCAA GGGCGAG CCAGAATTCTTACTTGTACAGCTCGTCCATG
pLG5087	CAGGGATCCATGACCCCTACGTGGTGACAATGGTGAGCAA GGGCGAG CCAGAATTCTTACTTGTACAGCTCGTCCATG
pLG5088	CAGGGATCCCGCAGCTGCCGCAGCTGCAATGGTGAGCA AGGGCGAG CCAGAATTCTTACTTGTACAGCTCGTCCATG
pLG5101	CAGGGATCCCGCGTGCACTAACTTCATGATGACCCCTACGT GGTGACAATGGTGAGCAAGGGCGAG CCAGAATTCTTACTTGTACAGCTCGTCCATG
pLG5102	CAGGGATCCTCATGATGACCCCTACGTGGTGACACGGTA CTACCGGATGGTGAGCAAGGGCGAG CCAGAATTCTTACTTGTACAGCTCGTCCATG
pLG5103	CAGGGATCCTGTGATTCGGCCTGGCCCGGATTGCCGATCC TGAGCATGACCACACCGGCTCCTGACGGAGTATGTGGCTA CGCGCTGGATGGTGAGCAAGGGCGAG CCAGAATTCTTACTTGTACAGCTCGTCCATG CCAGAATTCTTACTTGTACAGCTCGTCCATG
pSK4702	ATCGGACCCGGATGCCGCGTCTTAAAC AAATGCATTATTACGAAGCAGTGGATGTG
pLG5186	CACTAACTTCATGATGACCGAATACGTGGTGACACGGTAC GTACCGTGTCAACCACGTATTGGTCATCATGAAGTTAGTG
pSK4611	AGCGTCGACATGCCCTACATCCCTTAAAC CAGCATGCTAGCTAAAGTAATGGATGCAG
pKB4505	AGAATTCCCTTAATGCCCTACATCCC ACCGTCGACCAACGTAGCTAAAGTAACGG
pKB4506	AGAATTCAATGGTATTTCAGCCAAACC ACCGTCGACCAACGTAGCTAAAGTAACGG
pKB4699	CAGGGATCCATGCTTGACTTGGCCTGGCCCGGACAGCGTG CACTAACTTCATGATGACCCCTACGTGGTGACACGGTACT ACCGGATGGTGAGCAAGGGCGAG CCAGAATTCTTACTTGTACAGCTCGTCCATG
pLG6728	ATGTCAGATAAGATATCGAGTCTCTGTAAAAG GCCCTGAAAATACAGGTTTCGTGATGATGATGATGGC TGC

pLG6725	TTCATGATGACACCATAATGTGGTCACACGTTATTACAGAGC TCCAGAAATTATGTTGAATTCC
	ATTGTACATGCGGTACGGGCCAGGCCAAAGTCCAAGATCT TGAGATCACAAAGTGGTGTTCAG
pLG6726	CATACAGACGACGAAATGACAGGATATGTAGCCACCGCGTT GGTACAGAG
	ACGGGCAAGGCCAAAGTCCAAGATCTTGAGATCACAAAGTG GTGTTCAAG
pLG6727	GATCCAGACCATAATGATCATACAGGGTTCTTGACAGAGTACGT GGCCACTAGGTGGTACAGG
	TGCAACCCGAGCCAGTCCAAAATCAAATCTTCAGCTCAC AGTCTTCATTACAG
pLG6680	GTATGCCCGCCGAAGCAGAAGC CGGGGGCATACAAACAGTGTATGGG
pLG6681	CGCCGCAGCAGAAGCCCCACCACC CTGCTGCGGGGGTCATACAAAC
pLG6682	AGCAGCAGCCCCACCACCTCAAATTATG GGGCTGCTGCTCGGGGGTCATAC
pLG6683	CCCCCCGCAGCAGCAGCCCCACCACCTCAAATTATG GCTGCTGCGGGGGCATACAAACAGTGTATGGGT G

References

- Baruch, K., Gur-Arie, L., Nadler, C., Koby, S., Yerushalmi, G., Ben-Neriah, Y., Yoge, O., Shaulian, E., Guttman, C., Zarivach, R., *et al.* (2011). Metalloprotease type III effectors that specifically cleave JNK and NF-kappaB. *The EMBO journal* 30, 221-231.
- Schlosser-Silverman, E., Elgrably-Weiss, M., Rosenshine, I., Kohen, R., and Altuvia, S. (2000). Characterization of Escherichia coli DNA lesions generated within J774 macrophages. *Journal of bacteriology* 182, 5225-5230.
- Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J., *et al.* (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular systems biology* 7, 539.