

Supplemental Table 1: List of intein sequences used in the phylogenetic analysis

	1	10	20	30	40	50
Aae_RIR2_T1C	CFIEGTEVLTKRGNWEVVATEGHEFIVKVYCVSVPKGNIVVSVCIAGNC					
Aave_Aave1721_T1C	CVDDTTEALTQRGTLDMMLTPDHRVVHRVWCPIVGTSTWVAMPFITHTNT					
Ccy_Hyp_Csp1_T1C	CFTGDTLIPLVDGGEEIKCTPDHOFMLDVYCLTVPEYNNFAAGVVFVHNC					
Ceu_ClpP_T1C	CLTSDHTVLTTRGGVDLFTPNHRMYVPVYCLTG-PNNVFYKAVWTGNS					
Cme_boo_Pol_II_T1C	CFYGETKIEIFDGRVTDTNNGMHLRIRVYCLTVAEDHTLVNGVFTGQC					
Cmo_ClpP_T1C	CLTSDHTVLTTRGGVDLFTPNHRMYVPVYCLTV-PNSVFYKAVWTGNS					
Cmo_RPB2_T1C	CLTPDHDVLTTKGNINLLVTPNHRMYIPVFCLTV-PNEVFYIPVWTGNS					
Cpa_ThrRS_T1C	CLSEDMRIQTDKGGVSLLVTEHDHMYVRVWCVTVPPTGLIVAKPVVIGNC					
Cre_RPB2_T1C	CLSEEHQVLTNEGGLSLVVTTPGHDMYLRTWCVSL-P-HTYILPTILGNC					
Cst_RPB2_T1C	CLTPSHEVLTENGEIELOVTOEHYMWVSVHCVEVPEPHLFYKGVWTGNS					
Ctr_ThrRS_T1C	CLSEDMRVQTNKGSFSLFVTDNHDHMYVRVWCVTVPPTGLIIVRPTIVGNC					
Fte_RPB2_T1C	CLTPDHDVLTAEGNLSLLVTLNHRMYVKVFCLSV-PNEVFYLPVWTGNS					
Hma_Pol_II_T1C	CFHPDTRLWYEDEDRTLRTVTPDHTMLRHVYCLTVADTHRAEGTYVGQC					
Hma_PolB_T1C	CFTPDTDVLTPPEGKMDFRVTPNHRMLVGVYCVTVADNHTLMKFQFVGQS					
Hmu_PolB_T1C	CFTPDTDEVVTPPEGKIDFRVTPNHRMLVGVHCVTVEDNHTLLKFQFVGQS					
Hwa_Pol_II_1_T1C	CFHPETKIWFRTDGREITITPDHEMHIYTYCLTVSETHSLINDLSQKQC					
Hwa_PolB2_T1C	CFTPDTNVLTPNGEIDFRVTPNHRMLVGVYCVTVADNQTLLTFQWTGNS					
Lsp_GyrB_T1C	CFTGDTLVLPLADGGAEIKCTPDHOFMLDVYCLTVPEYHNFADGVVFVHNC					
Memar_MCM2_T1C	CYDDETEVLTESGQVDLAVTPNHRMYVTIYCLEVPNHTLYVIPVWCGNS					
Memar_Pol_II_T1C	CFAGDTEITVSDGGRVLTVPDPHAMLVRVYCLTVAENHTLVNGIFCGQC					
Mhu_Pol_II_T1C	CFHGDTLIEIYADGKNLLVTPDHAMLVRVYCLTVDRDHNVNNGIFTGQC					
Nma_PolB2_T1C	CFTPDTDEVLTDPDGKIDFRVTPNHRMLVGVYCVTVEDNHTLLKFQFVGQS					
Nph_PolB2_T1C	CFTPDTDEVLTDPGEIDFRVTPNHRMLVGVYCVTVEDNHTLLKFQFVGQS					
Npu_GyrB_T1C	CFIGDTLIPLADGGEEI ICTPDHOFMLDVYCLTVPEYHNFAGVVFVHNC					
Sas_RPB2_T1C	CLTPDHEVLTERGHVSLITTPNHRMWTKVHCIEV-PGHVFYKAVWTGNS					
She_RpoB_T1C	CLTVDHEVLTTKGNLDDLVTNHRMYVAVYCLSV-PNEIFYLSVWTGNS					
Sru_RIR1_T1C	CLTGDQRVVTDRGGVEQKVSARHGMPVTVYCPPTTEESVFVQALTFNC					
Ter_Ndse1_T1C	CFDSKTEILTEAGHRDFCVTPNHWMYVKVYCVNVTNHVVFVRAAWCGQC					
Tye_RNR_1T1C	CYSSDTQVLTYSGLDLLVTPNHRMLVKVYCVVNPHTLYVKACWCGNS					
StP_Twort_T1P	CISMDSMILTTEGGIELVNTYNHPLLVPPTFDVCMCKTHSFINTIVNHNT					
SoP_SO1_dpol_T1P	CLHRHTQVLTNGGVDGVLMTEDHKILTPVYDLIDEDGNRFLGFLVAHNC					
SaP_SETP12_dpol_T1P	CLHRHTQVLTGVDGVLMTEDHKILTPVYDLIDEDGNRFLGFLVAHNC					
SaP_SETP3_Helicase_T1P	CLKRGTEVIMFDGGESYTVNESHILSLGVDDYFGDGDSTRFLFTVTHNTS					
SaP_SETP5_dpol_T1P	CLHRHTQVLTGVDGVLMTEDHKILTPVYDLIDEDGNRFLGFLVAHNC					
SaP_SETP3_dpol_T1P	CLHRHTQVLTGVDGVLMTEDHKILTPVYDLIDEDGNRFLGFLVAHNC					
MP_KBG_gp53_T1CP	CSWTNARAVTRRGEINATITANHRVVGEVWCLRVPNGRFFIKIHLTGNS					
MP_U2_gp50_T1CP	CSWTNARAVTRRGEINATITANHRVVGEVWCLRVPNGRFFIKIHLTGNS					
MP_Be_gp51_T1CP	CSWTNARAVTRRGEINATITANHRVVGEVWCLRVPNGRFFIKIHLTGNS					
Mca_MupF_T1P	CFPAETPVRAAARGHRLTLTANHPVLTHVYDFETETGLIMAGGVIVHNC					
LP_PhiHSIC_Helicase_T1P	CHAYGHDIMMSDGYIRSSSTFKHRSKLGVDYDGDHLYLDFVRHHNC					
LLP_KSY1_PolA_T1CP	CFSGDTEILTPYGVYEDTNTSIHATGNDVYCVNV-PTHNIVKVISIQGNC					
CbP_D1873_RNR_T1CP	CFDGQOKTLTKSSNKEILLVTEHDHINVTYVYCFEMIEEYPFTNGVITHNC					
CP_P1201_Thy1_T1CP	CYSSDTEVLTSSEGGVDLLVTPNHKMLATVYCATVPNHTLYVPVWSSGNS					
CbP_Cst_RNR_T1CP	CFDGQOKTLTKSSNKEILLVTEHDHINVTYVYCFENIEEYPFTNGIITHNC					
BsuP_SPBc2_T1P	CVTGETLLLLTENGGYEIRATEWHKFYVDVYDTTQEDYHSLINGIVTGNC					
AP_APSE5_dpol_T1P	CLAKGTLVLTITGAYGVGMTADHQVLTEVYDLINGPRNRVFGPLIVHNC					
AP_Aaphi3_MupF_T1CP	CFPGQSELKGLPRGTVLLATPNHPILTQASCHIYQTVSGYYNSVVFVSN					
AP_APSE4_dpol_T1P	CLAKGTLVLTITGAYGVGMTADHQVLTEVYDLINGPRNRVFGPLIVHNC					
AP_APSE2_dpol_T1P	CLAKGTLVLTITGAYGVGMTADHQVLTEVYDLINGPRNRVFGPLIVHNC					
AP_APSE1_dpol_T1P	CLAKGTLVLTITGAYGVGMTADHQVLTEVYDLINGPRNRVFGPLIVHNC					
AgP_S1249_MupF_T1CP	CFPGQSELKGLPRGTVLLATPNHPILTQASCHIYQTVSGYYNSVVFVSN					
CIV_RIR1_T1V	CVAPETMILTEDGGCTLSCTPEHKFIIRVDDTYCFTEPINHNGILTGQC					
APMV_Pol_T1V	SVTGDTPIITRHQTGCIDVTEDHSLLDYVYDLETESGSFHAGEMIVKNT					
HaV01_Pol_T1V	SVTKETPLMLRTMNGYVEVTEHDHSLISYVYDVETEDGIFHAGEIIVKNT					
WIV_RIR1_T1V	CVVGETLILTEGGIFIVCTEYHRFLVNTYCFTEQLNAGVNGILTNSC					
CZIV_RIR1_T1V	CVVGETLVLTQNGGTSI ICTPYHNFSLDITYCFSEPLNAGVNGILTGC					

CV\_NY2A\_RIR1\_T1V  
IIV6\_RIR1\_T1V  
CP\_P1201\_Thy1\_T1P  
Bse\_MLS10\_TerA\_T3  
Ama\_MADE823\_T3  
Cth\_27405\_TerA\_T3  
Cth-DSM2360\_TerA\_T3  
Bvi\_IcmO\_T3  
Ckl\_PTerm\_T3  
Dra\_ATCC13939\_Snf2\_T3  
ShP\_Sfv5\_Primase\_T3  
EP\_Min27\_Primase\_T3  
Mex\_Trbc\_T3  
MP\_Catera\_gp206\_T3  
PP\_PhiEL\_Helicase\_T3  
PP\_PhiEL\_ORF11\_T3  
PP\_PhiEL\_ORF39\_T3  
PP\_PhiEL\_ORF40\_T3  
Sav\_RecG\_Helicase\_T3  
MleTN\_DnaB\_T3  
Mav35712\_DnaB\_T3  
Min1442\_DnaB\_T3  
Msm\_DnaB1\_T3  
MP\_McJw\_DnaB\_T3f  
MP\_Omega\_DnaB\_T3e  
MP\_Be\_DnaB\_T3d  
MavPT\_DnaB\_T3  
PospJS666\_DnaB\_T3  
BcePC184\_DnaB\_T3  
NspJS614\_DnaB\_T3  
MspMCS\_DnaB\_T3  
MflPYRGCK\_DnaB\_T3  
Mvan\_DnaB\_T3  
MspKMS\_DnaB\_T3  
Fal\_DnaB\_T3  
Mav104\_DnaB\_T3  
MgiPYRGCK\_DnaB\_T3  
Pnuc\_DnaB\_T3  
Mav25291\_DnaB\_T3  
RerSK121\_DnaB\_T3  
CflDSM20109\_DnaB\_T3  
KflDSM17836\_DnaB\_T3  
Ppu\_DnaB\_T1a\_1  
Ssp\_DnaB\_T1a\_1  
Rma\_DnaB\_T1a\_1  
MtuH37Rv\_DnaB\_T1a  
Gth\_DnaB\_T1a  
Msm\_DnaB2\_T1a  
Npu\_DnaB\_T1a  
MtuCDC1551\_DnaB\_T1a  
NspPCC7120\_DnaB\_T1a  
MboAF2122\_DnaB\_T1a  
CbuRSA493\_DnaB\_T1c  
Ter\_DnaB1\_T1b  
Ter\_DnaB2\_T1a  
Gvi\_DnaB\_T1b  
Cwa\_DnaB\_T1b  
Nfa\_DnaB\_T1a  
Dge\_DnaB\_T1b  
Kra\_DnaB\_T1a  
Sru\_DnaB\_T1b

CVAGDTKILTSSGGTELRCPTYPHKFYIDTYCFNEPKRHMVGVTGGQC  
CVAPETMILTEDGGCTLSCTPEHKFIIRVDDTYCFTEPINHNGILTGQC  
CYSSDTEVLTSSEGGVLLVTPNHKMLATVYCATVPNHTLYVPVWSSGNS  
QLALDTP IPTPDGGQSIVAGERHLWQVKMRCIQVSPSRLYLSMIPTHNS  
NIAHSTPVL TANRGEKIRCHANHEWTVKGHCIQVSPDGLYLKLVATHNS  
QLALDTP IPTPDGGTSIVAGERHLWKVKMRCIQVSPSRLYLSMIPTHNS  
QLALDTP IPTPDGGSSIIAGARHLWKVPMRCIQVSPSRQYLSMVP THNS  
PQPLHSLVRMADGGRSVEAARVHHWVPPARCLVDEMHCYIHDIVTHNC  
ALDLDTPIDTPTGGEFIIADTEHIWTVPVKCIADVDEKKLYLNHTATHNT  
AQPLDAKVLTP LGGASVEADA EHLWNVPAQCI AVAPDHLYVGYIVTHNT  
ACPLNEPILLADGGRYVDCAGDHLWEVECLCIMVHPRHLYVGYVATHNT  
ACPLNEPILLADGGRYVDCAGDHLWEVECLCIMVHPRHLYVGYVATHNT  
ALPLDALVHTPAGGRTA EACPEHLWEVQAACIRIHPDHLFIDYVVTHTNT  
SLACSTPILTTNGGQEVVTD A EHLWQVPVRCITVHESLSLYVGFVPTHNT  
AMPLSTRVKVPNGGRTADSHPHYHLWKTQETCCIANDEKLYVDYIVTHNT  
QQRNSAKVRI VPKGRATDVGTEHLWTL DCTCIEVNSEQLYLDFIVTHNT  
AGVLSSKIKIPGGGRYTDVSPDHLWKVETQCITIHDPHLYIDFIVTHNS  
MQPLYAKIKVPGGRYTDAGLDHLWKVEAQCISVHPDRLYIDFIVTHNT  
AQPLDSLVLAPTGGSSVECDDEHLWIVPAQCISVHPSHAYVNFTVTHNT  
ALALDTP LPTPTGGTVIVADAQH QWPTPVRCVEVNAAHLYLGMVPTHNS  
ALALDTP LPTPTGGTVIVADAAHQWLTAVRCVEVNDSHMYLAMVPTHNS  
ALALDTP LPTPTGGTVIVADAAHQWLTAVRCVEVNGSHLYLAMVPTHNS  
ALALDTP LPTPSGGTAIVADAQH QWPTPVRCVEVNPEHLYLGMVPTHNS  
ALDVETPILTGNGGAELVADHHLWAVETVCIQIHPSHVFLSLTPHNT  
ALAI DTPILSERGGQTI VADEDHLWTVPTRCITVND SRLYVGFVPTHNT  
PLALNTEVPTPSGGTEITASASHGWTTPVKCIGIT EDHLFQSRILTHNT  
ALALDTP LPTPTGGTVIVADAAHQWLTAVRCVEVNDSHMYLAMVPTHNS  
AQPLDARVRTRTGGRSTECCA EHLWQVLCQCISVHPSKLYIQDVVTHNT  
AQPLDARVRTLTGGRSAECCDEHLWC VATQCISVHPSRLYVDYVVTHTNT  
ALALDTALPTPSGGSRI VADAQH QWLT PVRCVEVNTEHLYLSMIPTHNS  
ALALDTP LPTPDGGTVIVADAQH QWLT PVRCVEVNDSHMYLSMVP THNS  
ALALDTP LPTPTGGTVLVADAEHQWLT PVRCVEVNDDHLYLSMIPTHNS  
ALALDTP LPTPTGGTVIVADAEHQWLT PVRCVEVNDDHMYLSMIPTHNS  
ALALDTP LPTPDGGTVIVADAQH QWLT PVRCVEVNDSHMYLSMVP THNS  
ALALGTALPTPTGGSVIVADGQH QWITPVRCVEVSPDHLYLTCIP THNS  
ALALDTP LPTPTGGTVIVADAAHQWLTAVRCVEVNDSHMYLAMVPTHNS  
ALALDTP LPTPTGGTVLVADAEHQWLT PVRCVEVNDDHLYLSMIPTHNS  
AQPLDAKVKTV DGGREAECCDEHLWRVQAQCVSVHPSRTYVDYVVTHTNT  
ALALDTP LPTPTGGTVIVADAAHQWLTAVRCVEVNDSHMYLAMVPTHNS  
ALALDTP LATPSGGTVLVADEQH QWLT PVRCVEVNADHMYLSMVP THNS  
ALALDTP LPTPTGGSTIVADAQH QWATPVRCVEIHQSHLYLAMIPTHNS  
ALALDTP LPTPSGGTTLVADAEHQWVVPVRCVEVNPDHLYLSMVP THNS  
CISKFSHIMWVSHVEKYLELTSNHKILT NVFDFAANP IPNF INNI IVHNS  
CISGDSLISLASTGRTIKATANHRFLTEVFDLTVPGPHNFVNDI IVHNS  
CLAGDTLITLADGGRSIRATANHRFLTEVFDLTVPGPHNFVNDI IAHNS  
CLTASTRILRADTGREVEATGSHPFMKHVFDGTVSGTHNFVNGISLHNS  
CLSYATNQPYFLKISVIVSTVEHYVLTNVYDFWIKTNNFVNALLVHNS  
CMTANTRILRADTGREVEATANHPFLTDVYDGTVPGTHNFVQGISAHNS  
CLAGDSLVTLVDSGRKIRATGNHKFLTEVFDLTVPGLHNFVNNI IVHNS  
CLTASTRILRADTGREVEATGSHPFMKHVFDGTVSGTHNFVNGISLHNS  
CLTGDSLVTLAPGGCKIRATANHKFLTEVFDLTVPGLHNF INNIIVHNS  
CLTASTRILRADTGREVEATGSHPFMKHVFDGTVSGTHNFVNGISLHNS  
CVTGDTLICLADGGRS IKATAEHRLLAEVYDLTVPKYASWLDG VVSHNS  
CAAYDTLV LQKDGGRFVETTITHPFLTQVYDLTVPE THNFVNDICLHNT  
CLTGESLICLADGGKIRATANHPFFSEVFDLTVPGLHNFVNEIIVHNS  
CCAADTP IADPVTGREVKTTLTHPYLTQVYDLTVPVTHNFVADICLHNT  
CLATNSEIILADGGRYIETTLTHPYLTQVYDLTI PKTHNFVNDICVHNT  
CLPASTRILRADTGREVEATGNHPFLTDVYDGTVPGTHNFVQGISVHNS  
CVTADTLIDVPGTGRVETTPHHPFLGQVYDLTVPGDANFIADICLHNT  
CLTAGTRVLRADTGKTVRATANHPFLTQVYDATV LGGHNFVDGIAVHNS  
CLGKGT PVMYDGGMDYRVNESHILSLVGDYFGFDGNGRFLDGTVTHNT

SspJA2_DnaB_T1a	CLAGDTQVMDADRNF SIRATGNHPFLTPVYDIEMPRHHNFVNGLLVHNS
Aeh_DnaB1_T1b	CIMAGSRLVDPRTGREIATTLTHPFLTQVYDLTVDGTHNFVEDVCVHNT
Aeh_DnaB2_T1c	CVTGDTRVLLADGGRTIRCTPEHRLRGEVFDLTVPGNACWLDGIVSHNS
ArspFB24_DnaB_T1h	GLALDTPIPAPKGGTSLVASDGNLWQALVRGIRVTKDSLFOGCQLTHNT
Pye_DnaB_T1a	CISKFNLYQTSLYGKYIQLTSNHKLLTIVFDLECKPLCNFINNFIVHNS
Mbo1173P_DnaB_T1a	CLTASTRILRADTGREVEATGSHPFMKHVFDGTVSGTHNFVNGISLHNS
Hha1_DnaB_T1b	CLAYDAEIVQADGGRIETTLTHPFLTQVYDLTIPETHNFINDVCVHNT
SuspNBC_DnaB_T1b	CLGKGTNVLMDYGGIDYRVNKSILSLVEYDKV-DDYYGFVDMTVTHNT
Mtu37HRa_DnaB_T1a	CLTASTRILRADTGREVEATGSHPFMKHVFDGTVSGTHNFVNGISLHNS
MtuF11_DnaB_T1a	CLTASTRILRADTGREVEATGSHPFMKHVFDGTVSGTHNFVNGISLHNS
MaerNIES843_DnaB_T1a	CLAGDSLVELAAPGRTIRATANHKFLTEVYDLTVDKLHNF INNIIVHNS
BceMCO3_DnaB_T3	AQPLDARVRTLTGGRSAECCDEHLWCVATQCISVHPSRLYVDYVVTHNT
Cbu_Dugway_DnaB_T1c	CVTGDTLICLADGGRSIKATAEHRLLAEVYDLTVPKYASWLDGVVSHNS
Cbu_RSA334_DnaB_T1c	CVTGDTLICLADGGRSIKATAEHRLLAEVYDLTVPKYASWLDGVVSHNS
Cbu_Goat_DnaB_T1c	CVTGDTLICLADGGRSIKATAEHRLLAEVYDLTVPKYASWLDGVVSHNS
Sare_DnaB_T1a	CLTADTRLIRADTGTKHIDATSNHPFLADV DATVLGTHNFINGIATHNS
Mtu_Haarlem_DnaB_T1a	CLTASTRILRADTGREVEATGSHPFMKHVFDGTVSGTHNFVNGISLHNS
CspPCC7425_DnaB_T1a	CLTGDSLITLADTGRAIRATGNHKFLTDFVFDLTVPEHHNFVNNIIVHNS
RmaDSM4252_DnaB_T1a	CLAGDTLITLADGGRSIRATANHRFLTEVFDLTVPGPHNFVNDIIAHNS
Pho_Pol_II_T1	CFPGDTRILVQINGSSFETTVDHPVLVVFVSLNAKKYHTVIENIVTHQC
Pab_Pol_II_T1	CFPGDTRILVQIDGRSFETTVDHPVLVVFVSLNAKKYHNVIENIVTHQC
Hsp_NRC1_Pol_II_T1	CFHPETNVWFRDEGRELTVPDHSMRRTYSLTV-AETNTLNGLFTGQC
Tko_Pol_II_T1	CFPGDTRILVQINGRSFETTVDHPVLVVFVSLNAKKYHNVIESIVTHQC
Nph_PolB1_T1	CFSGDTDVATPDGTHDFKVTENHRFLLEDHGDVITAEDNHHLQWVGQS
Sru_PolBc_T1	CFDPETEVTVDGHTDFLVTNHRFLTVDYDGMILTVADNHKFNWTGQS
Nsp_JS614_TOPRIM_T1WCT	CLAGETKVLTWDRKELFATDGRWFVEVYCAEVEDGHAF'DNILTGN
Tfus_RecA1_T1	CLTADTYVWTD RGGQVTVTRNHPLRVPTFDLMVPRTHSFLNGILSHNT
Tfus_RecA2_T1	CMHYDTLVTLADGGAQFTATPNHLIRTRKFDIEVEGNHNYFDGVMVHNS
Tfus_Tfu2914_T1WCT	CVDEETEILTTDGTSSLTTPQHRWPVRVWCVRTPNATWLATVYFTGNS

**Supplemental Table 1.** Intein Blocks A, B, F and G were concatenated to produce a single

49 amino acid sequence for each of 148 inteins. Due to the variable size of Block F, only the first and last 7 positions of Block F were used. A dash indicates the absence of an amino acid at that position. Intein names are defined in the InBase database (<http://www.neb.com/neb/inteins.html>)

(14) with the following additional suffixes: T1, Class 1 intein, T3, Class 3 intein; P, phage or prophage intein; V, viral intein; T1a, DnaB Class 1 intein present at the DnaB-a insertion site; T1b, DnaB Class 1 intein present at the DnaB-b insertion site; T1c, DnaB Class 1 intein present at the DnaB-c insertion site; T1d, DnaB Class 1 intein present at the DnaB-d insertion site; T1e, DnaB Class 1 intein present at the DnaB-e insertion site; T1f, DnaB Class 1 intein present at the DnaB-f insertion site; T1h, DnaB Class 1 intein present at the DnaB-h insertion site; T1C, Class 1 intein with CysF:4; T1CP, Class 1 intein from a phage or prophage that has CysF:4; T1WCT, Class 1 intein with the WCT motif.