

Evolution of eukaryotic single-stranded DNA viruses of the *Bidnaviridae* family from genes of four other groups of widely different viruses

Mart Krupovic^{a*} and Eugene V. Koonin^b

^a Institut Pasteur, Unité Biologie Moléculaire du Gène chez les Extrêmophiles, Department of Microbiology, Paris 75015, France

^b National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA

* - corresponding author

E-mail: krupovic@pasteur.fr

478429021|BmBDV 14 VKITEDTTPATPGYYSKSD----ELASFLKELVDN-----TPIGQLSFSHFVAGSFLKDDER-KIHTTDPVFNHISDGIIPIDVDIDNFI--EYLDPEKSSFSKYS 110

Polinton-GlyFla 18 IKWQDIESAFEGRIKGCVINLRHTDLKFLNDAENLVIEKQVAMQHGSLKNTLLARFECKRN--DAVVIETKSFQ--KSTTLASTGLHEWQESVNVQTLRKVDLQ 127

Polinton-3_NV 162 FTPKBEQQTALGKYLKHCIAQORGYDKPTFIARIKRQVLELIS--RQKKIKVKVFIPTCGFIKEDPATAQIEEELGYFHTEKEIIVTESTDLSLDF--DMTMYLLGLVLFQ 270

Polinton-1_HM 125 IEFKLSQSSLKNVTKQFSAEAGVKGYDALSFMKSAENNVIKILN---SNKGS-KIYIVLSCEMERTDLKGETITTTASFS--KAEVVLSTDLNDFY--ERAEQKILESLSAFQ 231

Polinton-1_CB 76 SEYKEVGDVLRDTHHGNSSN---NIQSFNLNKESDVFDKIY---KDSVVKIYFSLSCEMKKE---GTEEARLDHITTTSKALHKGENRK--SFY--EIVKLDLNLRLQOHO 173

Polinton-1_CR 69 TDYKEVGNLRLDRSHHGNSK----NIEFLDLKKEAVLNDIY---EDNVKLYVFLCEMLKE---ETEEMQIAHITLTKENKPLYIGDDKV--EWF--KSLKLDLNLRLHQA 165

478429021|BmBDV 111 --GGFSFQVILSYWVNIIVPSSPNNAIVPTSRNTQDYRDGTPAPELGLSSS--IRRNFIGVIAQHFTFPGYTKSGKVKSTVPLEVIEEFTKQGFERYAEVLAVNVTSTPAIHA 220

Polinton-GlyFla 128 ENQSGWSLSETLNLTVNISRYTTPLOVGISTFVAVPDDIRKKEGVV---NIIKK-DHYCFWLVCVVAALY--P-VKAHRNLPSSMPYPS--SVLIDHTGI----KFPTELKDIPLFEK 229

Polinton-3_NV 271 KQSGGWQDFDOVEYFDINIDPPEPLSG--SSYIPLPKLASKAKAI--VVKENDHECFKWAVTSAVY--PREKDPQRFKSKMIENS--EKFDSWGI----EFPVSLKQIDKFEK 372

Polinton-1_HM 232 QLGSSWIFVSVKMKDINIIEYKPIKG--KSYIPLPKELAAKAI--NMKNE--DNECFKWVAVFFN--PKEKNSERVDKDLKEQS--EKLNWEKI----KFPVSLQIITQFEK 332

Polinton-1_CB 174 GRGSGFRNLRIIGLRMYQTRIRLSG--SKYIELPDWIKNKKAVI--IQNK--DDKCFMWCILAHLF--PVEQNPRV--SKVKDHV--SKVNFDFG--DFFPQVKVDVDFEK 273

Polinton-1_CR 166 ARGSGFRNLRLGLRVQTKIRPLSG--SKYMELPDWIKNKKAVI--IQNK--DNMCFMWCILAHLF--PVEKNAERT--SKKKEHV--NNIDFNGE--EFPVQVKNLDAFEK 265

478429021|BmBDV 221 ELNEFNIIISW-----ITGPIYAKILGEVDE--VPIHML-----KKGDN-----LHLIRSRRAFLNE--NKTQNRVFECLCKKFIHPNACNITGETIICID 303

Polinton-GlyFla 230 MND--LRINNVYIEPKKNEEDPSIIVLYLSSNLECOQDT--IHLMLIETSFTLNNDGKIKNKHPIHFAWIRNLSRLVKS--SITKQKHLFFCDRCLCHFKLEASFQRHSDLQ 340

Polinton-3_NV 373 QNQ--YTVNVFGEYETK----IYPLRISEK--DPDNAINLLLI-----SDDET--HYCWIKNMMLVSTQIDEFHFRPLCCRCNLNFRCKQSLKHSKCCGN 460

Polinton-1_HM 333 NNQDIDSNVNVGYES-----YPLRISSEKNN--RQH--IDLILL-----SNDETN--HYCLIKLSRLSSQSKSNEHEMVFYCRNLGFCFEESLNHKLQDT 422

Polinton-1_CB 274 RND--LAINIIVLHVNDVEDEK---INSYKYSKVNVDVDRVLLLV-----TDSGEG---HYCLIKNIDRLMNT---QNGRNRKFCIRCNHNFSEKYEENHLSYDM 366

Polinton-1_CR 266 KNN--LAINIIEHLEGQOK---FQPKYKSKNTKVELNRVINLLV-----TDSSGEG---HYCLIKNIDRLMNT---KDSYSTKFCIRCHNHFYSAEKYDKHLSDMS 357

478429021|BmBDV 304 D---IKDNEKDKGLISPLSKPKIYIYPTKHLVAYADFAITDNN---IHKPASYSLIETIAGDPPAENFI---FTKSVNLNLEDEVFN---EDGEIIFKDNIIEMFLTDL 401

Polinton-GlyFla 341 HNKVRMTLPEDEKHK---ILSRDRHYKDPVFPVYVADLESVLEEE-----TENIQKHTPHSIAYYLHCSYDNLKSKFNLN---SEDCIKWFKVELELISHKV 432

Polinton-3_NV 461 HEAVGIEMPKIKDKDGNLPHIKFKFNYNRKMVRVFPVYVADDFESFENITDTCSPDESCKSFTKQYQKHKPSGFCYLKIC--FDGDISPSELVRYTAESPDIDIPQLFVESLESDIKIKI 572

Polinton-1_HM 423 HDSVRIELPKPNT----MIEFKNYNKMVRVFPVYVADDFESFIKPIYNTCSNPNESYTKQYQKHTPSSFCYYIKC--FDEKTYQSKLVTPTASNEDEVDQAKFNMLLEDVKKI 529

Polinton-1_CB 367 NSPLQMIKPSKD-----YIEFDGIQKTKQRYVYADDFESVIYKIDNVSRSSESNNENFGKHVSAFQVVDVSTNTIY---DMKSYI----GYDTILKFNNEYLVEVCEKL 467

Polinton-1_CR 358 NAPIQMIKPSKD-----YIEFSGVQKTKHRYVYADDFESVIYKISNSTSNPNKSWSENIKGHVSAFQVVDVSTNTIY---EMKSYV---GYDSVFKFNNEYLDVCRRL 458

478429021|BmBDV 402 YSIIKKFQNTDYKGEANLDEAGERFKCLACNKNKGG---KYYARAYGLG--IFGYCRSCFLAH--NNTFIVYFHNFKGYDHHILEQLLNKDSKHNTRCGKSINKMVDI 506

Polinton-GlyFla 433 --SGYLNPLPLPKLTDDEQKSFNAVNCIENKPLNSANDKKVRDHCHFTGK--FRGAHNSCNLNFQOQSHIIPVVFHNLSGYDCHFELESVAIVFNGSV---EILPIN 535

Polinton-3_NV 573 YDKF--RFP--KVKMTPKDKIAVNDATHCHICEGELGED--KVLDDHCHLTGK--YRGAHANACNLDHRIKFPFVIFHNLGSDSHLFTKMLGTSEGKI-----NCIPNN 670

Polinton-1_HM 530 YNDYLFKP--KKMFTMKDKNDFNAKICTACEKLNDE--RVRDHCHITGK--YRGAHANACNLDHRIKFPFVIFHNLGSDSHLFTKMLGTSEGKI-----NCIPNN 670

Polinton-1_CB 468 LSIADM--RMQKLTFEVESFHSCECPACNCGFGD--KVRDHCHITGK--YRGLPCNACNLLKRRNNTFPVFFHNLSGYDSHLFTKMLGTSEGKI-----NCIPNN 627

Polinton-1_CR 459 LNMSDV---EMNKLTEKEWKFENDECFPCACGKNEFEDNGLPTEKVRD--DHWGTGEYRGLPCNENLLKRRNNTFPVFFHNLSGYDSHLFTKMLGTSEGKI-----NCIPNN 566

478429021|BmBDV 507 THKDLISD-----FIRITFKDTPNPLPELASLANKLITLKYTP-----DKFKEAFNSGKGEFFYEWFDNFMLEEIEVQDPADWDRLTNKKGIE 594

Polinton-GlyFla 536 KEKYIEFTKTVD-----DNIHLRFIDSRFRMASIDKLSYLLTDDDMITTRQFYSDEPQLKLVTSKGIFFPEYIDCLEKLEDEPOLDK--QAFYSRLNDKHHVDE 634

Polinton-3_NV 671 EEKYISFTKQIVVDSFTNK--EGNKIDVYKRDIFRIDSFRFMSASLDSLVGNMSRECFKNLAE--YEGEELQLLLRKGVPYDFWDFGFSKLDATQOLQR--EAFHSLKNDTDEE 780

Polinton-1_HM 628 EEKYISFSKELKVNEMFNR--EGKKVEVKLYLRFDSFKMAASLDSLTKNLSKQCKNISR--YSGNELNLLLRKGVVYFVWVSDIKLNEPOLPPK--ESFYSRLNDEGISDE 737

Polinton-1_CB 571 IEKFIISFHYHFDK-----RKKYIEIRFLDSLGFMPSSLDLYLDELEDECCITKQYNNHEHLFHLMRKGVVYDFMDSFKKYSNTELEPI--EDFYNSLSGETISNE 672

Polinton-1_CR 567 IEKFIISFHYHFDK-----RKKYIEIRFLDSLGFMACLDHLSLTKLQCAITKQYNNHEHLFHLMRKGVVYDFMDSFKKYSNTELEPI--ESFYNTLTDENISNE 676

478429021|BmBDV 595 IIKKANQIWDNMQIFHDYVLLNELDVLWLLLEVFHAFDFTVNEKDIDPVYFDGAPGLTFYLRMYENSLLDMHVLPDKNVYLDVSRNIRGGVTQVVTKYAN----- 697

Polinton-GlyFla 635 DYSFACTVWKNFVQTLGEYSDDLKTDVLLADIFENFDFASCFTYELDALHYHTAPGLA--FASMLKHSVQLELLELPEKILFFKGLRGGVQCCANRYAKANNRYMGEDF 746

Polinton-3_NV 781 DHLHARRVWEVFGMGTRDYNHLYLESVLLADVFENFDFVCLKKNYGLDPAWYHTAPGLA--WDAALKTQVRLLELTDYDMLMIEKGIIRGGVSMISNRHGANNPY--MKEY 891

Polinton-1_HM 738 DYLAQNVKWEKFCETFRDYHNLNESDVLLLADVFENFDFDXCNXYLDPAWYHTAPGLA--WDAALKTQVRLLELTDYDMLMIEKGIIRGGVSMISNRHGANNPY--MKEY 891

Polinton-1_CB 673 SVEYAAQVWDMSCOTLEDYTRTMINDVLLADVFETFNVSLEKYEKLDPCWYHTAPGLA--WDAALKTQVRLLELTDYDMLMIEKGIIRGGVSMISNRHGANNPY--MKEY 783

Polinton-1_CR 677 SPEYACKVWEKFCETFRDYHNLNESDVLLLADVFESFQKVSLEKYLHDCPCWYHTAPGLA--WDAALKTQVRLLELTDYDMLMIEKGIIRGGVSMISNRHGANNPY--MKEY 787

478429021|BmBDV 698 --IEDVDETIYVLDVNTMYSYCMKQKLANLYLGLDALPD-----NYDSDDNFCYFKIGDYSVEYLHDLPAHLSMPLMHPQY-----NNKLCITFLDKKMLIHSKVEKYLL 798

Polinton-GlyFla 747 NPEDESIVYLDVNLGAMSMPLPQGSFEWEIETVE--NISSFNSDESTGVYLVKLEVPEDLHE--LQVMPLECPHEHTPPGSKTCLATLTHFKNYVHYKLNLEQL 856

Polinton-3_NV 842 DNPLTKYIYLDANLLYGWAMSPLPTHGFWRMGDQELS-----GWRDRPCIDLEDEYPHHLHD--LHNDYPLAPEISIGL--NVDKLVNPLNKHNYVHHETRLRVL 993

Polinton-1_HM 849 DESKESTYIYLDANLLYGWAMSPLPTHGFWRMGDQELS-----GWRDRPCIDLEDEYPHHLHD--LHNDYPLAPEISIGL--NVDKLVNPLNKHNYVHHETRLRVL 993

Polinton-1_CB 784 NPEDESIVYLDANLLYGWAMSQKLYDEFEFFENAADESMEMENLNSQKGGKILVLEVDYPSLHD--KNDLDFCENKRVG--TNNKLDVDFSPKRYVHYKLNLEQL 892

Polinton-1_CR 788 NLEESKYLLYLDANLLYGWAMSQKLYDEFEFFENFEL--EMIDDLTANGKGCILEVLDYPSLHD--KNDLDFCENKRVG--TNSKLSDFSPKRYVHYKLNLEQL 894

478429021|BmBDV 799 SRGLVCDRIHYVYKFKQEIYIKDYVETNIQKNS--TDPGTKDYKLNKALFGKTCENVFKY--KIFSVTN---VNSGDRENKCMKSAKSHITLGCILYIEECVTRVLLDKPIQ 907

Polinton-GlyFla 857 QLGKMLVNRVRLKFAQSPWLKSYIDKNECKNAANEFKFNFKLMNNAVFGKTMENVRKHKDIKLVNKLGRYGAKSLISKSNFNSFMIFGELNLVIELNKLNVYFNKPIY 969

Polinton-3_NV 994 SFGLKVTIHRGVTFEESAWLEPYIDLNTDLAAKATNDFEEDFFKLMNNSVFGKTMENVRKHKDIKLVNKLGRYGAKSLISKSNFNSFMIFGELNLVIELNKLNVYFNKPIY 1103

Polinton-1_HM 951 RLGIKLTIRHRGKFEESAWLSKIKLNTDLTKATNDFEEDFFKLMNNSVFGKTMENVRKHKDIKLVNKLGRYGAKSLISKSNFNSFMIFGELNLVIELNKLNVYFNKPIY 1060

Polinton-1_CB 893 QOGLVKKIRRVVTFKESNWLASYIELNTLKKAKNEFEKDFKLMNNSVFGKTMENVRKHKDIKLVNKLGRYGAKSLISKSNFNSFMIFGELNLVIELNKLNVYFNKPIY 1002

Polinton-1_CR 895 DRLMLKLTIRRVVTFKESNWLSSYIELNTLKKAKNEFEKDFKLMNNSVFGKTMENVRKHKDIKLVNKLGRYGAKSLISKSNFNSFMIFGELNLVIELNKLNVYFNKPIY 1004

478429021|BmBDV 908 IGFTILELAKLMIYEFIEYLEDVIP--EGSTATMLYTDTSVIFKFKNGFVHPYKYL--TTSLSLAKLID---PINKDGSFGSATKTPGLWSDDTKYKTIIEFVGLRAKQVAY 1014

Polinton-GlyFla 970 VGFSLDLSLIFVYDFHNYVYK--RNFSNDTSKLLYTDTSVLYHFKVPDFVQIIRKDI--HKFDTSDFTI---NNDYGMQPKNKKVGLMKDENNKIIVSEFVGLRAKLVAF 1075

Polinton-3_NV 1104 LGMSILDLSKTLMYDFHNYVYK--PKYGEDAKLFTDTSVLYHFKVPDFVQIIRKDI--HKFDTSDFTI---NNDYGMQPKNKKVGLMKDENNKIIVSEFVGLRAKLVAF 1210

Polinton-1_HM 1061 LGMCILDLSKTLMYDFHNYVYK--NKYGDKAKLLYTDTSVLYHFKVPDFVQIIRKDI--HKFDTSDFTI---NNDYGMQPKNKKVGLMKDENNKIIVSEFVGLRAKLVAF 1170

Polinton-1_CB 1003 VGMSILDLSKTLMYDFHNYVYK--PKYSSNRLCYQDTSVLYHFKVPDFVQIIRKDI--HKFDTSDFTI---NNDYGMQPKNKKVGLMKDENNKIIVSEFVGLRAKLVAF 1106

Polinton-1_CR 1005 VGMSILDLSKTLMYDFHNYVYK--PKYGNLKLKYQDTSVLYHFKVPDFVQIIRKDI--HKFDTSDFTI---NNDYGMQPKNKKVGLMKDENNKIIVSEFVGLRAKLVAF 1108

478429021|BmBDV 1015 STANDR--DILKHKGIKPNALKDDNNPNNVDFRNVLFEMKDLTVNIAQIRATKNNVLTSTVSKKLLALSTKD----NKRITYTDKVTLPFGYKGEYSNYMDDVVE---- 1115

Polinton-GlyFla 1076 KLHEKDEVKKRAKRVKGS---LRKITFDKPKCLLDHVNISKEQFLIRSEKHQVTKIKQNKLALSWDD----DKRQLLDDSTDLPWGKVDMTPPAKRRK----- 1172

Polinton-3_NV 1211 KMDEGK--EEKKCKGVKRAV---VKKSIQFDYDKCLFGKPKQMLRMLNVRSHKHHDVYETETVKNVALSHED---DKRVICDDGIHFAHGFRFTLLGGGSVSSGTT---- 1308

Polinton-1_HM 1171 KIDEE--DKKCKGVKRVN---VKNYITHEDYKCLMKNKQMRKMNVRSHKHHDVYETETVKNVALSHED---DKRVICDDGIHFAHGFRFTLLGGGSVSSGTT---- 1256

Polinton-1_CB 1107 KIDDSG--EQKKNKGIKKS---VKKEMTFDNYKNCLENRTIERKQQLTSSSKHEISSVRQKVVLLNIIIGDKKETKRYFVDNI--NSLAFGHYRIK----- 1198

Polinton-1_CR 1109 KIQDGL--ETKKNKGVKKN---VKKEMTFDNYKNCLENRTIERKQQLTSSSKHEISSVRQKVVLLNIIIGDKKETKRYFVDNI--NSLAFGHYRIK----- 1200

Figure S1. Multiple sequence alignment of the type B DNA polymerases from BmBDV and selected polintoviruses.