

Resources for Instructors

1. Bioinformatics Tools to Study Bacteriophages

Resource Name	URL	Description
PHASTER	https://phaster.ca/	Tool to search for bacteriophages in bacterial genomes
VIRFAM	http://biodev.cea.fr/virfam/	Detection and classification of bacteriophages using head and tail proteins
MVP Database	https://mvp.medgenius.info/home	A microbe-phage interactions database. Allows identification of phages sequences and potential bacterial hosts
PhiSiGns	http://phisigns.sourceforge.net/	A tool to identify signature genes in phages and design primers
SEA PAHGES	https://seaphagesbioinformatics.helpdocsonline.com/home	Bioinformatics guide from the SEA PHAGES Programs
CRISPRFINDER	https://crispr.i2bc.paris-saclay.fr/Server/	A program to find CRISPRs in DNA sequences
iTOL	https://itol.embl.de/	Tool to visualize, store and annotate phylogenetic trees
MAFFT	https://mafft.cbrc.jp/alignment/server/	Multiple sequence alignment program for large data sets. Creates alignments and phylogenetic trees
CLUSTAL OMEGA	https://www.ebi.ac.uk/Tools/msa/clustalo/	Multiple sequence alignment program. Creates alignments and phylogenetic trees

2. **Accession numbers for PHASTER Exercises.** The accession numbers in the table below can be used to run PHASTER and compare the number of phages in diverse bacterial species.

Strain	Accession Number
<i>Sinorhizobium meliloti</i> AK83	NC_015590.1
<i>Sinorhizobium medicae</i> WSM419	NC_009636.1
<i>Rhizobium leguminosarum</i> bv. <i>trifolii</i> WSM1689	NZ_CP007045.1
<i>Mesorhizobium</i> sp. B7 complete genome	NZ_CP018171.1
<i>Bradyrhizobium oligotrophicum</i> S58	NC_020453.1
<i>Escherichia coli</i> str. K-12 substr. MG1655	NC_000913.3
<i>Escherichia coli</i> O83:H1 str. NRG 857C	NC_017634.1
<i>Escherichia coli</i> UMN026 chromosome	NC_011751.1
<i>Escherichia coli</i> IAI39 chromosome	NC_011750.1
<i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Chester str. ATCC	NZ_CP019178.1
<i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Minnesota str. ATCC .4592393	NZ_CP019184.1
<i>Shigella flexneri</i> 2a str. 301	NC_004337.1
<i>Bacillus cereus</i> ATCC 14579	NC_004722.1
<i>Bacillus subtilis</i> BSn5	NC_004722.1
<i>Bacillus anthracis</i> str. Ames	NC_003997.3
<i>Bacillus anthracis</i> str. Turkey32	NC_003997.3
<i>Bacillus thuringiensis</i> serovar konkukian str. 97-27	NC_003997.3
<i>Bacillus subtilis</i> PY79	NC_003997.3
<i>Bacillus horikoshii</i> strain 20a	NC_003997.3

3. Integrase Protein Sequences for Alignments (PhageIntegrases_Set1)

>1_EDL933
MLNGYIDEKAASAKLIRSTLSDAFREIAEGHITTNPVAATRAAKSEVRRSLTADEYL
KIYQAAESSPCWLRLAMELAVVTGQRVGDLCEMKWSDIVDGLYVEQSXTGVKIAIPTTL
HVDALGISMKETLDKCKEILGETTIIASTRREPLSSGTVSRYFMRARKASGLSFEGDPPT
FHELRSL SARLYEKQISDKFAQHLLGHKSDTMASQYRDDRGEWDRKIEIK
>2_EDL933
MRKTSVAKWQNYELEKAKLHNIMTVAKLWHMFMDSPAFTELAPRTQKDYRQHQKALLMV
FGKVLA DVNKTEQVRIFMDKRGLESKTQANHELASLSRVYWGTYERGVKNNPCKGVRKF
SLKARTVYITDEQYAAIYAEAIPOQLRIAMEISYLCARLGVDLELKWDQIMDKGIYIEQN
KTGTKQIKEWSPRLRTAIQLARNVSSCTCEYVINTTKGGKVIAKTLNNWWNQAKRAAEQK
VGVPFCNFHDIKAKGISDYEGSSRDQIFSGHKTENQVLIYDRKTKITPTLDLPLVVSK
>3_EDL933
MANSAYPAGVENHGGKLRITFKYRGKRVRENLRVPDTPKRNKIA GELRASVCFAIRTGTF
DYADRFDPDSPNLKLFGLVKKDITVGELAQWKLT KAMEIGSNALNRYQSVMKNMLPRLGP
GR LASSITKEDLLFIRKDLLTGEKGSRKTSRKGR TVPTVNYYM TTAGMFSFAAENGY
LEKNPFPNSITPLRKSKPVPDPLTRDEF SRLIDACHHQQT KNLWTVAVFTGMRHGEIAALA
WEDIDLKAGTITVRRNFTKIGDFTLPKTDAGTNRVIHLLAPAIEALKNQAMLTRLSRHQ
ITVQLREYGR TILHECTFVFCPQIVRKNHKAGINYAVSSIGATWDSA IKRAGIRS RKA
SRHTYACWALSSGANPTFIASQMGHSSASMVNVYGAWMPECSVTVQVAM LNNVLA
RAPD VPQSDQ EDEIKLYFSK
>4_EDL933
MSPRPRKNSTDVAGLYEKFD RRTGRVYYQYKNPVTGKFHGLGTDKGKA EKIA STAN QRIA
AAEA EYFMRKIDE SP SATKRGIRLKAWVDRYLKIQDTRLKNGDIAATTHKEKTRMAAYL
VSRLGNHPLKELEV RDFALI LDEWLDKDMVSTARVN RGLWV D IYKEA QHAGEVPPGWNPP
EATRKPIPKVTRARLT MEDWQK IYNA TPEKHFIRNAMLLAIVTGQRRDICHMRFSDVWN
EHLHITOQKTGMRLALPLTLRCDAIGITLKEVIDGCRDRILSPYLIHSRHOKQPKPM SKD
NLSDYFAKARDLAGVIPPAGKTPPTFHEQRSLSERLYRAQGIDTKTLLGHKVQATTDRYN
DTRGQE WVKLVI
>5_EDL933
MAASPRSHKISIPNLYCKL DKRTGKVWQYKHPLSGRFHSLGT DENEAKQVATEANTIIA
EQRT RQI LS VNERLERMKG RRS DITVTEWLDK YNSIQEDRLQHNE LRPN SYRQKGKPI RL
FREHCGMQHLKD ITALDIA EII DAVKAEGHNRMAQVVRMVLIDVFKEA QHAGHVPPGFNP
AQATKQPRNRVNRQRLS LPEWQAI FDSV SRRQPYLKCGM LAL VTGQRL SDICNLKFSDI
WDDMLHITQEK TGSK LAIPLNLKCD ALN ITLREV ISQC RD AVVSKYLVHYRHTSQANRG
DQVSANTLT A FKAREKCGIKWEQGTAPT FHEQRSLSERLYREQGLDTQKLLGHKSRKM
TDRYND DRGKD WIIVDIKTA
>6_EDL933
MARPRKYKTDVPG LSPYFDK RNNK VY WRYRHPITGKNHGLGSIDQKLAETIAAEANSRLA
RQQMEQMLSLQEKIIISDTGG SSTVTIFLNNYRKIQ QERYENGEIKLNTLKQKAAPL RVD
ERFGTRPLDAITVKDV VSVLE EYKARGHNR MGQIFRKVLIDV FREA QQTGDVPPGFNP
SAKKPQVRISQR QLTFDEWMMIYNAAEKDGYFLQ RGML LAL MTGQRL SDICKMQFSDIRD
GYLHVEQ QKTGTRIAIPLALRCDKLNLT LDDVVSSCRDCVLS P WLLHHHHAKGTA KRG
VKPATLTVA FKKARD SVDYNWRANGTPPSFHEQRSLSERL FREQVDT KILLGH SNQKMT
DIYNDARGKEWKKLVI
>7_EDL933
MSVRKIPSGKWLCECYP GASGKR IRKQFATKSEALS YERRLMNSRVGDEFQDGSGPRLS
ELIARWFEMYGKTLSSGAER KV KLEAIC SRLGDPF ASQFDK NMFATYRERRLSGEWNP
KKKLSEATVNRE QSYL HAVFA ELKRLGEW SGENPLT GIRK FREEE KE LAFLYV D EIERLL
IACDES RNKDLGV VV RIGLATG ARWSEA EGLKOSQV LPGRITF VKT KGK NRTV P
ISPOL QAMLPK KRGAL FSP CYEA F DAAIK RAKI ELPDG QLTHV L RHTFASH FM MRGGN ILV LQ
LGHSDIKMTMRYAHFAPGHLEAA VELNPFDNRG
>8_EDL933
MGRRRK NPEHEK LPPKV VP NKYSV WKPTSRES VTLTAIEDGLA ALWKKYEETVNHRDRAM
TFGRLWEKFLAS AYYSELSP RTQKDYLQHQ KLLAVFGKVLAD SVKPEH IRRYMDKRGEQ
SKTQANHEKSSMSRVY SWGY ERGYV KANPCAGVSKFKAKN RERYV TDKEYQAVLSVAPLP
VFIAMEIAYLCAARVSDVLSLKWEQIGNDGIFIQ QGKTGKKQIKA WS PRLQAAIEKAKQ

PTSAYVISNQYGNRYMYKGFNEMWVEARNHAGKISGILTDFTFHDLAKGISDYEGSSRD
KQLFSGHKTEGQVLITYDRKVVKVSPTLDVPLPENIPRKYSK

>9_Sakai8

MLNTYVAEGKAVSARVIRSTLVDVFRAIAEGHVATNPVTTTRAAKSEVRRSLTANEYV
AIYHAAEHLPIWLRLSMDLAVVTGQRVGDLCRMKWDINDGHLHIGQSKTGAKIAIPLAL
TIDALDISLVDTLQKCREASSETIIASTYHEPLSPATVSRYLTKARNASGISFDGDPPT
FHELRSLSARLYRNQIGYKFAQRLLGHKSMAAHYRDSRGREWDKIEIG

>10_Unknown

MSNASYPTGVENHGGSLRIWFHYNGKRVRENLCVPDTAKNRKIAGELRTSVCFAIRMGSF
DYAAQFPNSPNLKHFGLGKREITVKALSEKWLDLKKIEICANALNRYQSVIKNMPLMGE
KKLVSSITKEDLLFVRRDLLTGYQKLSNGKTSSIKGRSVVTNVYMMTTIAGMFQFATDNG
YTSGNPNGLAPLKKSKVKPDPLTRDEFIRFIEACRHQQTKNLWILAVYTGIRHGEVSL
AWEDIDLKARTITIRRNYTKLGEFTPPTDAGTGRTIHLVQPAIDALKSQAEMTMLGKQH
SVEVKQREYGRATAVHKCTFVFSPOVTKQQQLSGPHYKVDSIRESWTSILKRALRHKSY
QSRHTYACWSLAAGANPSFIASQMGTNAQMVFNVYGAWMKDNNHEQIELLNKRLSESVP
CMPHKKAG

>11_O104H4

MDKVKYPTGVENHGGTLRIWFNFKGKRVRENLCVPDTAKNRKIAGELRTSVCFAIRTSF
DYAAQFPDSPNLQAFGVSKKEITVKELEEKWLDLKRMEISANAFNRYESVARTMVPKIGG
SRLVSSVTKEELLYIRKDLLTGYQNSMKNKAAKGRSVVTNVYMMTTIAGMFQFAADHGYL
EANPFQGIKPLKRARAEPDPLTRDEFIRLIDACRHOQTKNLWSLAVYTGMRHGEVSLAW
EDIDLKAGTTITIRRNYTKLGEFTLPKTEASTNRVHVLIQPAISVLKNQAEMLRGKQYNI
KVQLREYGRSVNHECTFVFPQVVRKSEQVGFVYKVDSVGDSWETAIKRAGIRHRKAYQS
RHTYACWSLSAGANPSFIASQMGTNAQMVFNVYGAWMTDNSAEQIAMLNQKLTDYVPMM
PHSHQSDFTRGLLKSVS

>12_O104H4

MRPEGRKGRKIRKKFKTKSDAVLYERWVLAQQHNNEWKNSIDRRPLSVLIDLWWKYHGQ
LMKSGHNTRLKLLRLSEAMDPCVHKLNTMLTELRSRIEQIQPSTINREIGALSAMF
TALISSGHFLNDNPVQGLKGMKVNEREMGYLSKSECVQLLDALAENPDERLAVEILLSTG
ARWGEVAVLEQRRLVLCRITSKTKNSKNRTVPISESLEFEKIKKRGKGKLFPTLDYPLVR
DVIKTVASDVDPDGQAVHALRHTFASHFMMNGGNILTLQKILGHAKIQTMTIYAH LAPDYL
QDAVRFNPIAG

>13_O104H4

MSLFRRGEIWYASYSLPGGKRIKESLGTDKRQAQELHDKRKAELWRVDRLGMPDVTFE
EA CLR WLEEKADKKSIDS DK SRIA FWIEHF EGIRIKD ISEAMI YS V IS KAY NR KT KER WK
LQVEAALRKGKEPPAYIPKS VST QTKATH LAMIA IL RAA ERD W KWL E KAP VIKI PAV KN
KRV RWE KE EARR LID AC SDPL KSVV K FAL AT GLR RS NI IN LEW QQ ID M Q RR V A W VP ED
SKSNRAIGVALNDTACKVLRDQIGKHHRWVFTTAARRPDGTMTPSIRKMR LDY NT SWL
TACRRAGIENFRFHDLRHTWASWLIQSGVPLS VLQEMGGWE SI EMV RRYA HLAP N L TEH
ARKIDDILGDDVPNLSHPEVFEDAKKA

>14_O104H4

MARPRKYKTDVPGSPYFDKRNK VY WRYRHPITGKNHGLGSIDQKLAETIAAEANSRLA
RQOMEQMLSLOEKIIISDTGGSTVTIFLNNYRKIQOERYENGEIKLNTLKQKAAPLRVD
ERFGTRRLDAITVKDVSVL EYKARGHNRMGQIFRKVLIDVFREAQQTGDVPPGFNPAE
SAKKPQVRISRQLTFDEWMMIYNAAEKDGYFLQRGMLLALTGQRLSDICKMQFSDIRD
GYLHVEQQKTGTRIAIPLALRCDKLNLTLDDVVSCRDCVLSPWLLHHHHAKGTAKRGGM
VKPATLTVAFKKARDSDV DYNWRANGTPSFHEQRLSERLFREQVDTNILLGHSNQKMT
DIYNDARGKEWKKLVI

>15_O104H4

MANSAYPAGVENHGGKL RITFKYRGKRVRENLCVPDTPKNRKIAGELRASVCFAIRGTF
DYAERFPDSPNLKLFGLVKKDITVGE LAQKWLTLKAMEISSNALNRYQSVIKNMPLRLGP
GRLASSITKEDLLFIRKDLLTGEKGSRKTSRKGR TVPTVNVYMMTTAGMFSFAAENGY
LEKNPFSNISITPLRKSKPVDPPLTRDEF SRLIDACHHQQTKNLWTVAFTGMRHGEIAALA
WEDIDLKAGTTITVRRNFTKIGDFTLPKT DAGTNRVIHLLAPAI EALKNQAMLTRL SRQHQ
ITVQLREYGR TILHECTFVFCPQIVRKNHKAGINYAVSSIGATWDSA IKRAGIRSRKAYQ
SRHTYACWALSAGANPTFIASQMGTNAQMVFNVYGAWMPECSVTVQVAMLNNVLA RD
VPQSDQEDEIKLYFSK

>16_O104H4
MGRRRSHERRDLPNLYIRNNNGYYCYRDPRTGKEFGLGRDRRIAITEAIQANIELFSGHK
HKPLTARINSDNSVTLHSWLDREYEKILASRGIKQTLINYMSKIKAIRRGLPDAPLEDIT
TKEIAAMLNGYIDEGKAASAKLIRSTLSDAFREAIAEGHITNPVAATRAAKSEVRRSRL
TADEYLKIYQAAESSPCWLRLAMELAVVTGQRVGDLCEMKWSDIVDGYLVEQSKTGVKI
AIPTALHVDALGISMKETLDKCKEILGETIIASTRREPLSSGTVSRYFMRARKASGLSF
EGDPPTFHELRSLSARLYEKQISDKFAQHLLGHKSDTMASQYRDDRGREWDKIEIK

