

Human astrovirus, complete genome

NCBI Reference Sequence: NC_001943.1

[FASTA Graphics](#)

[Go to:](#)

LOCUS NC_001943 6813 bp ss-RNA linear VRL 13-AUG-2018

DEFINITION Human astrovirus, complete genome.

ACCESSION NC_001943

VERSION NC_001943.1

DBLINK BioProject: [PRJNA485481](#)

KEYWORDS RefSeq; capsid protein; nonstructural protein.

SOURCE Human astrovirus

ORGANISM [Human astrovirus](#)
Viruses; Riboviria; Orthornavirae; Pisuviricota; Stelpaviricetes;
Stellavirales; Astroviridae.

REFERENCE 1 (bases 1 to 6813)

AUTHORS Geigenmuller,U., Chew,T., Ginzton,N. and Matsui,S.M.

TITLE Processing of nonstructural protein 1a of human astrovirus

JOURNAL J. Virol. 76 (4), 2003-2008 (2002)

PUBMED [11799197](#)

REFERENCE 2 (bases 1 to 6813)

AUTHORS Bass,D.M. and Qiu,S.

TITLE Proteolytic processing of the astrovirus capsid

JOURNAL J. Virol. 74 (4), 1810-1814 (2000)

PUBMED [10644354](#)

REFERENCE 3 (bases 1 to 6813)

AUTHORS Willcocks,M.M., Brown,T.D., Madeley,C.R. and Carter,M.J.

TITLE The complete sequence of a human astrovirus

JOURNAL J. Gen. Virol. 75 (Pt 7), 1785-1788 (1994)

PUBMED [8021608](#)

REMARK (sites)

REFERENCE 4 (bases 4000 to 6813)

AUTHORS Willcocks,M.M. and Carter,M.J.

TITLE Identification and sequence determination of the capsid protein
gene of human astrovirus serotype 1

JOURNAL FEMS Microbiol. Lett. 114 (1), 1-7 (1993)

PUBMED [8293952](#)

REFERENCE 5 (bases 5779 to 6813)

AUTHORS Willcocks,M.M. and Carter,M.J.

TITLE The 3' terminal sequence of a human astrovirus

JOURNAL Arch. Virol. 124 (3-4), 279-289 (1992)

PUBMED [1605739](#)

REFERENCE 6 (bases 1 to 6813)

AUTHORS Willcocks,M.M. and Carter,M.J.

TITLE Sequence analysis of a human Astrovirus

JOURNAL Unpublished

REFERENCE 7 (bases 1 to 6813)

CONSRM NCBI Genome Project

TITLE Direct Submission

JOURNAL Submitted (01-AUG-2000) National Center for Biotechnology Information, NIH, Bethesda, MD 20894, USA

REFERENCE 8 (bases 1 to 6813)

AUTHORS Willcocks,M.

TITLE Direct Submission

JOURNAL Submitted (24-AUG-1993) Willcocks M., University of Newcastle upon Tyne, Framlington Place, Newcastle upon Tyne, Tyne and Wear, UK

COMMENT VALIDATED [REFSEQ](#): This record has undergone validation or preliminary review. The reference sequence was derived from [Z25771](#).

The mature peptides were added by the NCBI staff as indicated (predicted) in the publications.

COMPLETENESS: full length.

FEATURES Location/Qualifiers

source	1..6813
	/organism="Human astrovirus"
	/mol_type="genomic RNA"
	/db_xref="taxon: 1868658 "
	/type="1"
gene	86..4380
	/locus_tag="HAstVgpl"
	/db_xref="GeneID: 29030992 "
CDS	join(86..2845,2845..4380)
	/locus_tag="HAstVgpl"
	/function="2b like nuclear targetting sequence, serine protease"
	/standard_name="Non structural gene"
	/note="The positions of the mature protein products of orfla are provided as speculated by the authors based on the deletion analyses; Non-structural gene composed of two sections 1a and 1b separated by ribosomal frame shift motif at appx position 2840 leading to expression of RNA-dependent RNA polymerase."
	/citation=[3]
	/codon_start=1
	/product="Non-structural protein"
	/protein_id=" NP_059443.2 "
	/db_xref="UniProtKB/TrEMBL: Q67726 "
	/db_xref="GeneID: 29030992 "

```

/translation="MAYGEPYSSKPKDFNFGSTMARRQMTPTMVTKLPKFVRNSPQ
AYDWIVRGLIFPTIGKTYFQRVVVITGGLEDGTYSFAFDGKEWVGIIPIEHLNLMSS
LKLIHKANALQERLRLSQEEKATLALDVQFLQHENVRLKEMIPKPEPRKIQMKWIIMG
AVLTFLSLIPGGYAHSQTNNTIFTDMAACKYSTETLTENLDLRIKLALANITISDKL
DAVRQILNFAFVPRAHWLRVTFYIIHYEMWNI FMFVLAIGTVMRSARPGTDLVTLAT
SHLSGFRMAVLPTIPFHTTMTLWMNTLMVCYYFDNLLAITLAILAPILGIIFLCFME
DSNYVSQIRGLIATAVLIAGGHACLTLTGTTTSLFVVILTCTRFIRMATVFIGTRFEIR
DANGKV VATVPTRIKNVAFFQKLKQSGVRVGVNEFVVIKPGALCVIDTPEGKGTGF
FSGNDIVTAAHVVGNNFTFVNVCYEGLMYEAKVRYMPEKDIAFLTCPGDLHPTARLKLS
KNPDYSCVTVMAYVNEDLVVSTAAAMVHGNTLSYAVRTQDGMSGAPVCDKYGRVLAVH
QTNTGYTGGAVIDPADFHPVKAPSQVELLKEEIERLKAQLNSATENATTVVTQQPSA
ALEQKSVSDSDVVDLVRTAMEREMKVL RDEINGILAPFLQKKKGKTKHGRGRVRRNLR
KGVKLLTEEEYRELLEKGLDRETFLDLIDRIIGERSGYPDYDDEDYDDEDDGWMVG
DDVEFDYTEVINFDQAKPI PAPRTTKQKICPEPEVESQPLDLSQKKEKQSEYEQQVVK
STKPQQLEHEQQVVKPIKPQKSEPQPYSTYQYKAPIWESYDFDWEDDAKFILPAPHR
LTKADEIVLGSKIVKLRTIIETAIKTQNY SALPEAVFELDKAAYEAGLEGFLQRVSKS
NKAPKNLQRAPEDQGAQNYHSLDAWKLLLEPPRRRCVPANFPLLGHLPINRPIFDDK
KPRDDLGLLPEPTWHAFFEYGP TTWGPQAFIKSFDKFFYAEPIDFFSEYPQLCAFAD
WATYREFRYLEDTRVIHITATEKNTDSTPAYPKMNYFDTEENYLEAHGWAPYIREFTR
VYKGDKPEVLWYLF LKKEIIKEEKIRNSDIRQIVCADPIYTRIGACLEAHQNALMKQH
TDTSVGQCGWSPMEGGFKTMQRLVNKG NKHFIEFDWTRYDGTIPPALFKHIKEIRWN
FINKDQREKYRHVHEWYVNNLLNRHVLLPSGEVTLQTRGNPSGQFSTMDNNMVNFWL
QAFEFAYFNGPDRDLWKTYD TVVYGDDRLSTTPSPDDYEERVITMYRDI FGMWVKPG
KVICRDSIVGLSFCGFTVNENLEPVPTSPEKLMASLLKPYKILPDLES LHGKLLCYQL
LAAFMAEDHPFKVYVEHCLSR TAKQLRDSGLPARLTEEQ LHRIWRGGPKKCDG"

```

```

mat_peptide      86..>595
/locus_tag="HAstVgp1"
/product="N-terminal orflab cleavage product"
/inference="non-experimental evidence, no additional
details recorded"
/note="This product does not appear to depend on the
activity of the viral 3C-like proteinase (NSMP-3); NSMP-1,
non-structural mature protein 1"
/protein_id="NP_751902.1"
mat_peptide      596..1312
/locus_tag="HAstVgp1"
/product="NSMP-2, non-structural mature protein 2"
/inference="non-experimental evidence, no additional
details recorded"
/protein_id="NP_751906.1"
mat_peptide      <1313..>2047
/locus_tag="HAstVgp1"
/product="3C-like serine proteinase"
/inference="non-experimental evidence, no additional

```

```

details recorded"
/note="NSMP-3, non-structural mature protein 3"
/protein_id="NP_751903.1"
mat_peptide    join(<2048..2845,2845..>2850)
/locus_tag="HAstVgp1"
/product="C-terminal part of the orfla product"
/inference="non-experimental evidence, no additional
details recorded"
/note="contains nuclear localization signal; NSMP-4,
non-structural mature protein 4"
/protein_id="NP_751904.1"
mat_peptide    <2851..4377
/locus_tag="HAstVgp1"
/product="orflb product"
/inference="non-experimental evidence, no additional
details recorded"
/protein_id="NP_751905.1"
gene           4373..6733
/locus_tag="HAstVgp2"
/db_xref="GeneID:29030991"
CDS            4373..6733
/locus_tag="HAstVgp2"
/function="structural"
/citation=[4]
/codon_start=1
/product="capsid precursor protein"
/protein_id="NP_059444.1"
/db_xref="UniProtKB/TrEMBL:O12792"
/db_xref="GeneID:29030991"
/translation="MASKSNKQVTVEVSNNGRSRSKSRARSQSRGRDKSVKITVNSRN
RARRQPGRDKRQSSQVRNIVNKQLRKQGVTPGPKPAICQRATATLGTVGSNTSGTTEI
EACILLNPVLVKDATGSTQFGPVQALGAQYSMWKLKYLNVKLTSMVGASAVNGTVSGV
SLNPPTTPTSTSWSGLGARKHLDVTVGKNATFKLKPSDLGGPRDGGWLTNTNDNASDT
LGPSIEIHTLGRMTSSYKNEQFTGGFLVELASEWCFTGYAANPNLVNLVKSTDNQVS
VTFEKSAGSPLIMNVPEGSHFARTVLRSTTPTTLARAGERTTSDTVWQVLNTAVSAA
ELVTPPPFNWLKGGWWFVKLIAGRTRTGSRSFYVPSYQDALSNKPALCTGSTPGGM
RTRNPVTTTLQFTQMNQPSLGHEAPAAFGRSIPAPGEEFKVVLTFGAPMSPNANNKQ
TWNKPLDAPSGHYNVKIAKDVDHYLTMQGFTSIASVDWYTIIDFQPSAPAPIQGLQV
LVNSSKKADVYAIKQFVTAQTNNKHQVTSFLVKVTTGQVNNYLSYFYRASATGDAT
TNLLVRGDTYTAGISFTQGGWYLLTNTSIVDGAMPPGWVWNNVELKTNTAYHMDKGLV
HLIMPLPESTQMCYEMLTSIPRSRASGHGYESDNTEYLDAPDSADQFKEDIETDTDIE
STEDEDEADRFDIIDTSDEEDENETDRVTLLSTLVNQGMTMTRATRIARRAFPTLSDR
IKRGVYMDLLVSGASPGNAWSHACEEARKAAGEINPCTSGSRGHAE"
proprotein     4583..6730

```

```

        /locus_tag="HAstVgp2"
        /product="79 kDa capsid proteins precursor"
        /note="Processing of this capsid protein into smaller
        capsid proteins by trypsin enhances the infectivity of
        virions"
mat_peptide      4583..>5554
        /locus_tag="HAstVgp2"
        /product="VP34 precursor"
        /protein_id="NP_751899.1"
mat_peptide      5456..6730
        /locus_tag="HAstVgp2"
        /product="VP26 precursor"
        /protein_id="NP_751901.1"
mat_peptide      5555..6730
        /locus_tag="HAstVgp2"
        /product="VP29 precursor"
        /protein_id="NP_751900.1"

```

ORIGIN

```

1 ccaagagggg ggtggtgatt ggcctttggc ttatcagtgt atagataaca tttctctgac
61 cgtttaccac acaattaaca acaatatggc atacggtgag ccatactata gctctaaacc
121 tgacaaagat ttcaattttg gaagcacaat ggcacgtagg cagatgacac ctactatggt
181 aacaaagctt cccaaatttg ttaggaattc tccacaagct tatgattgga tcgtaagagg
241 tctgatcttt ccacaccattg gtaaaactta tttccaacga gttgttgtga ttactggtgg
301 gcttgaggat ggaacatatg gctcattcgc atttgacggt aaagagtggg tagggatcta
361 cccaatagag cacttaaate tcatgtcatc tttgaaactg atacacaaag ccaacgctct
421 tcaggagaga ctgcgtctct cccaagaaga gaaagccacc ctgcgtctgg atgtgcaatt
481 ctttcaacat gaaaacgtgc gattgaagga gatgatccca aagccagagc cacggaat
541 acagatgaag tggataataa tgggagcagt gcttacattc ttatctctga tacctggggg
601 ttatgcgcac agccagacca acaacaccat atttactgac atgatagctg cctgcaagta
661 ctcaactgag acactaacag aaaatcttga ccttagaatc aagcttgac tagcaaacat
721 aaccattagt gataagctag atgctgtgag gcaaattctt aactttgcct ttgtgcccag
781 agcccattgg ttgagaactg tggtctatta tatccattac tatgaaatgt ggaatatatt
841 tatgtttgtt cttgctattg gcaactgcat gaggagcgcc cgccctggta cagacttggg
901 tacacttgca acatcccact tgtctggttt taggatggct gtcctacca caattccatt
961 tcacaccact atgactttgt gggttatgaa cacacttatg gtttgttatt attttgacaa
1021 cttgctagca ataacattgg caatcttagc accaattctt ggcattatct tcttgtgctt
1081 catggaagac tccaactatg tgagccagat acgtggcctt attgctacag cagtattaat
1141 tgctggtggg catgcatgtt tgacactcac aggcacgacc acgtcattgt ttgttgtcat
1201 actaacctgt aggttcatac gtatggcaac tgttttcatt ggcaccaggt tcgagatccg
1261 tgacgctaata gaaagggttg tggccactgt accaactagg attaaaaatg ttgcatttga
1321 cttttttcag aagctgaagc agtcaggggt gcgagttgga gtcaacgaat tcgttgtcat
1381 aaaaccaggt gcattatgtg ttatagacac ccctgaagga aaaggacag gtttcttttc
1441 tggcaatgac atagtaacag cagcacatgt tggttgcaat aatacttttg tgaatgtgtg
1501 ctatgagggg ttgatgtatg aagcgaaggt gcggtacatg cccgaaaagg atatagcatt

```

1561 ctttaacttgt cctggtgacc tgcattccaac agcaagatta aaattatcaa agaaccaga
1621 ttatagttgt gtcacagtta tggcttatgt gaatgaggat cttgtggttt caaccgcagc
1681 agctatggtg catggcaaca ctctttcata tgcagttcgc acccaggacg ggatgtcggg
1741 tgcaccagtc tgtgacaaat atggtcgggt gttagcagtc catcaaaca acactgggta
1801 cactggaggt gctgtcataa tagaccagc agacttccat ccagtgaag ccccatctca
1861 ggtggaattg ctcaaagagg aaatagagcg gctaaaagct caactgaact ctgccactga
1921 gaacgcaacg actgtagtta cacaacaacc tagtgctgca ctagaacaga aaagtgtcag
1981 cgatagtgat gtagttgacc ttgtcagaac tgcaatggaa cgtgagatga aggtgctgcg
2041 tgatgaaatc aatggaatac ttgcaccatt cctacaaaaa aagaaaggta agaccaagca
2101 tggtaggggt agagtcagcg gtaaccttag aaaagggtgtg aaacttctta ccgaggaaga
2161 gtatcgagaa ctcttagaga aaggtcttga tctgagaca ttccttga **tc tcatagaccg**
2221 cattattggt gagaggtctg gctaccctga ctatgatgat gaagattact atgatgaaga
2281 tgatgatggc tggggaatgg ttggtgatga tgtagaattt gattataactg aagtaattaa
2341 ctttgacca gcaaaacca ttctgcccc gagaacaacc aagcaaaaaa tttgccccga
2401 accagaagtc gaatcacaac cacttgattt gtcccaaaag aaagaaaaac aatcagaata
2461 tgaacaacaa gtggtgaagt ctaccaagcc tcaacaatta gaacatgaac aacaagtggg
2521 gaagcctatc aagcctcaga agagtgcagc tcaaccatac tcacaaactt acggcaaggc
2581 accaatctgg gaatcttacg attttgactg ggatgaggat gatgccaagt ttattctgcc
2641 agcgccacac cggttgacta aggcagatga aatagtcctt ggatccaaaa tctgtaagct
2701 tagaacgatt attgaaacag ccataaagac tcagaattat agtgcattac ctgaagcagt
2761 atttgagctc gacaaagcag cttatgaagc aggttttgaa ggttttctcc aaagggttaa
2821 atcgaaaaac aaggcccaa aaaactaca agggccccag aagaccaagg ggcccaaac
2881 taccactcat tagatgcag gaaattgtt ctagagcctc cgcgggagcg aaggtgctg
2941 cctgcgaatt ttccactatt aggccattta ccaattaata gacctatctt tgatgataag
3001 aaaccaggg acgatctcct tggcctactt ccagaacca cctggcatgc tttcgaggaa
3061 tacggacca ccacatgggg ccacaaagct tttatcaaat cttttgataa atttttttat
3121 gcagaacca ttgacttttt ctcaaatat ccacagtgtg gtgctttcgc tgattgggca
3181 acttatcgcg agtttcggta tctagaggat actagagtga tacacataac tgcaactgag
3241 aaaaacactg attcaacacc tgcttatcct aaaatgaatt attttgacac tgaagaaaat
3301 tacctggaag cacatgggtg ggcaccatat attagagaat tcaactagggt ctacaaagga
3361 gacaaacctg aagtactttg gtacctatct ctttaagaaag agatcattaa ggaggaaaaa
3421 attaggaatt ctgatatccg gcagatagtg tgtgccgacc ccatttacac caggataggg
3481 gcgtgcttag aagcgcatca gaatgcctta atgaaacagc ataccgacac ctcaattggt
3541 caatgtgggt ggtcaccaat ggaaggcggc tttaaaaaaa ccatgcagcg cctggtaaat
3601 aaagggaata aacacttcat tgagttcgac tggaccgct atgatggaac tataccacca
3661 gcaactctta aacacatcaa agaaattagg tggaaatttca tcaataaaga ccaacgtgaa
3721 aagtacagac atgtacatga atggtatgtt aacaacctcc ttaatcgcca tgtacttcta
3781 ccatctgggt aagtcacctt gcagacgca ggtaatccat ctggtcagtt ttcaacaaca
3841 atggataaca acatgggttaa cttttgggta caggcttttg agtttgctta ttttaatgga
3901 ccagacagag acctttggaa gacctatgac actgtagttt atggagatga caggctttct
3961 acaacacctt cggtgcccg tgattatgag gagagagtga tcaactatga tagagacatc
4021 tttggcatgt gggtaagcc tgggaagggt atctgtagag acagcatagt tggattgtcc
4081 ttttgtggct tcaactgtta tgaaaacct gagcctgtgc caacttctcc tgaaaagtta
4141 atggcatcac tgttaaaacc ttacaaaata ttacctgatc ttgaatcact ccatgggaaa

4201 ctccctatgct atcagttgct tgctgcgttc atggcagagg accacccctt taagggtgat
4261 gtggagcact gcctctcgcg gactgcaaag cagcttcgtg actctggcct tccagccaga
4321 ctacacagaag agcaactcca tcgcatttgg aggggaggac caaagaagtg tgatggctag
4381 caagtccaac aagcaagtaa ctgttgaggt cagtaataat ggccgcagca ggagtaaattc
4441 aagggcccggt tcacaatcta ggggccgaga caaatcagtc aagatcacag tcaattcaag
4501 aaacagagcc aggagacagc ccggacgcga caaacgtcaa tcttctcaac gtgtccgtaa
4561 cattgtcaat aagcaactca ggaaacaggg tgtcacagga ccaaacctg caatatgtca
4621 gagagcaaca gcaacccttg ggacagtcgg gtcaaacc accagggacca ctgagattga
4681 ggcggtgtatt ctccctcaacc ctgtcctcgt taaggacgct actggaagca ctcagtttgg
4741 ccctgtgcag gcgctaggtg cacagtattc catgtggaag ttgaagtatt tgaatgtcaa
4801 attgacctct atggtttgtg catctgctgt taatggtact gtctcagggg totcacttaa
4861 cccactacc acgcctactt ctactagttg gtcaggatta ggtgcacgta aacatcttga
4921 tgttacggtt ggtaagaatg caacatttaa attgaaacct tctgaccttg gtggacctag
4981 ggatggctgg tggctcaca acaccaacga caatgcatct gacaccctgg gaccctccat
5041 tgaaatacac aactcggac gaacaatgtc ctcatacaag aatgagcagt tcacaggagg
5101 gctatTTTTTgtt gttgaacttg cttcagagtg gtgcttcaca ggttatgctg ccaacccaaa
5161 ccttgTTaTaaC ttggtcaagt caacagataa tcaagtatct gtgacatttg aaggaagtgc
5221 aggatcacca ttaataatga atgtgccaga ggggagccat tttgcacgaa cagttcttgc
5281 acgctcaaca acaccacca ctctagcgcg tgcaggagag agaaccacct cagacacagt
5341 atggcaggtg ctcaatacag ctgtatctgc tgctgagctt gtcacgcctc ctccattcaa
5401 ttggttagtt aaaggaggtt ggtggtttgt gaaactcatt gccgggagaa ccagaactgg
5461 ctctcgtagc ttttatgtgt atcccagcta ccaagatgct ttgtccaaca agccagcact
5521 ctgcaccggc agcactccag gaggcagtag gacgcggaat cctgtgacaa ccaccctaca
5581 attcacacag atgaaccaac ctagtctagg gcacggtgaa gcaccagctg cgttttgtag
5641 atccattcca gcacctggtg aggagtttaa agtcgtcctc acgtttggag ccccaatgag
5701 ccctaattgca aataacaaac agacttggtt taataaacct cttgatgcgc cttcaggcca
5761 ttacaatgtg aagattgcaa aggatgttga ccaactatcta accatgcagg gtttcacttc
5821 tatagcttct gttgactggt acaccataga ttttcaacca tctgaggccc ctgccccaat
5881 acaaggcctg caggtacttg tgaacagctc gaaaaaagct gatgtgtatg ccatcaaaca
5941 atttgttaca gcgcagacca acaacaagca ccagggttaca agcctgttcc tagtaaaagt
6001 gacaactggt tttcaggtga acaactacct gagttacttt tacagggcgt ctgccactgg
6061 ggatgccaca actaacctgt tggtagagag agacacatac acagcagga taagttttac
6121 ccagggtgga tggtagttgt tgacaaatac atctattgtt gatggggcta tgccacctgg
6181 ttgggtctgg aataacgtgg aacttaaaac taacacagca tatcacatgg acaaaggctt
6241 ggtccattta ataatgcctt tgccctgagtc cagcgaatg tgttatgaga tgctgacatc
6301 tataccacgc tccagggcat ctggatcatg ttatgagagt gacaacaccg aatacttggg
6361 tgctccagac tctgctgacc agtttaaaga agacatagag acagatacag acattgagag
6421 tacagaggac gaagacgaag cggacaggtt tgacatcata gacacttctg atgaagaaga
6481 tgaaaatgag acagaccgtg taaccctcct ctcaactctc gtaaaccaag gaatgacaat
6541 gacgcgtgcc acaaggatag caaggcgcgc attccccacg ctttccgata ggatcaagcg
6601 tggagtatac atggacctgc ttgtctcggg ggcaagccca ggcaatgcat ggtcccatgc
6661 gtgtgaagaa gcacgcaaag cagcagggga aatcaatccc tgcacatctg gaagccgcgg
6721 ccacgccgag taggaacgag ggtacagctt ctttcttttc tgtctctgtt tagattattt
6781 taatcaccat ttaaaattga tttaatcaga agc

