

Novel Betacoronavirus in Dromedaries of the Middle East, 2013

Technical Appendix

Technical Appendix Table 1. Coding potential and putative TRSs of the genomes of DcCoV UAE-HKU23

CoV, strain no., ORF	Location, nt	Length		Frame	TRS location, nt	Putative TRS TRS sequence(s), distance in bases to AUG*
		nt	aa			
DcCoV UAE-HKU23 (265F)						
1ab	215–21498	21284	7094	+2,+1	68	UCUAAAC (140)AUG
NS2	21508–22344	837	278	+1	21494	UCUAAACUUUAAGAAUG
HE	22356–23630	1275	424	+3	22340	ACUAAACUCAGUGAAAAUG
S	23645–27745	4101	1366	+2	23638	UCUAAACAUG
NS5	28114–28443	330	109	+1	28056	GGUGAAC (51)AUG
E	28430–28684	255	84	+2	28300	UCCAAAC (123)AUG
M	28699–29391	693	230	+1	28689	UCCAAACAUUAUG
N	29401–30747	1347	448	+1	29387	UCUAAACUUUAAGGAUG
N2	29462–30085	624	207	+2		
DcCoV UAE-HKU23 (362F)						
1ab	215–21498	21284	7094	+2,+1	68	UCUAAAC (140)AUG
NS2	21508–22344	837	278	+1	21494	UCUAAACUUUAAGAAUG
HE	22356–23630	1275	424	+3	22340	ACUAAACUCAGUGAAAAUG
S	23645–27745	4101	1366	+2	23638	UCUAAACAUG
NS5	28114–28443	330	109	+1	28056	GGUGAAC (51)AUG
E	28430–28684	255	84	+2	28300	UCCAAAC (123)AUG
M	28699–29391	693	230	+1	28689	UCCAAACAUUAUG
N	29401–30747	1347	448	+1	29387	UCUAAACUUUAAGGAUG
N2	29462–30085	624	207	+2		
DcCoV UAE-HKU23 (368F)						
1ab	215–21498	21284	7094	+2,+1	68	UCUAAAC (140)AUG
NS2	21508–22344	837	278	+1	21494	UCUAAACUUUAAGAAUG
HE	22356–23630	1275	424	+3	22340	ACUAAACUCAGUGAAAAUG
S	23645–27745	4101	1366	+2	23638	UCUAAACAUG
NS5	28114–28443	330	109	+1	28056	GGUGAAC (51)AUG
E	28430–28684	255	84	+2	28300	UCCAAAC (123)AUG
M	28699–29391	693	230	+1	28689	UCCAAACAUUAUG
N	29401–30747	1347	448	+1	29387	UCUAAACUUUAAGGAUG
N2	29462–30085	624	207	+2		

*Boldface indicates putative TRS sequences. TRS, transcription regulatory sequences; DcCoV, dromedary camel coronavirus; CoV, coronavirus.

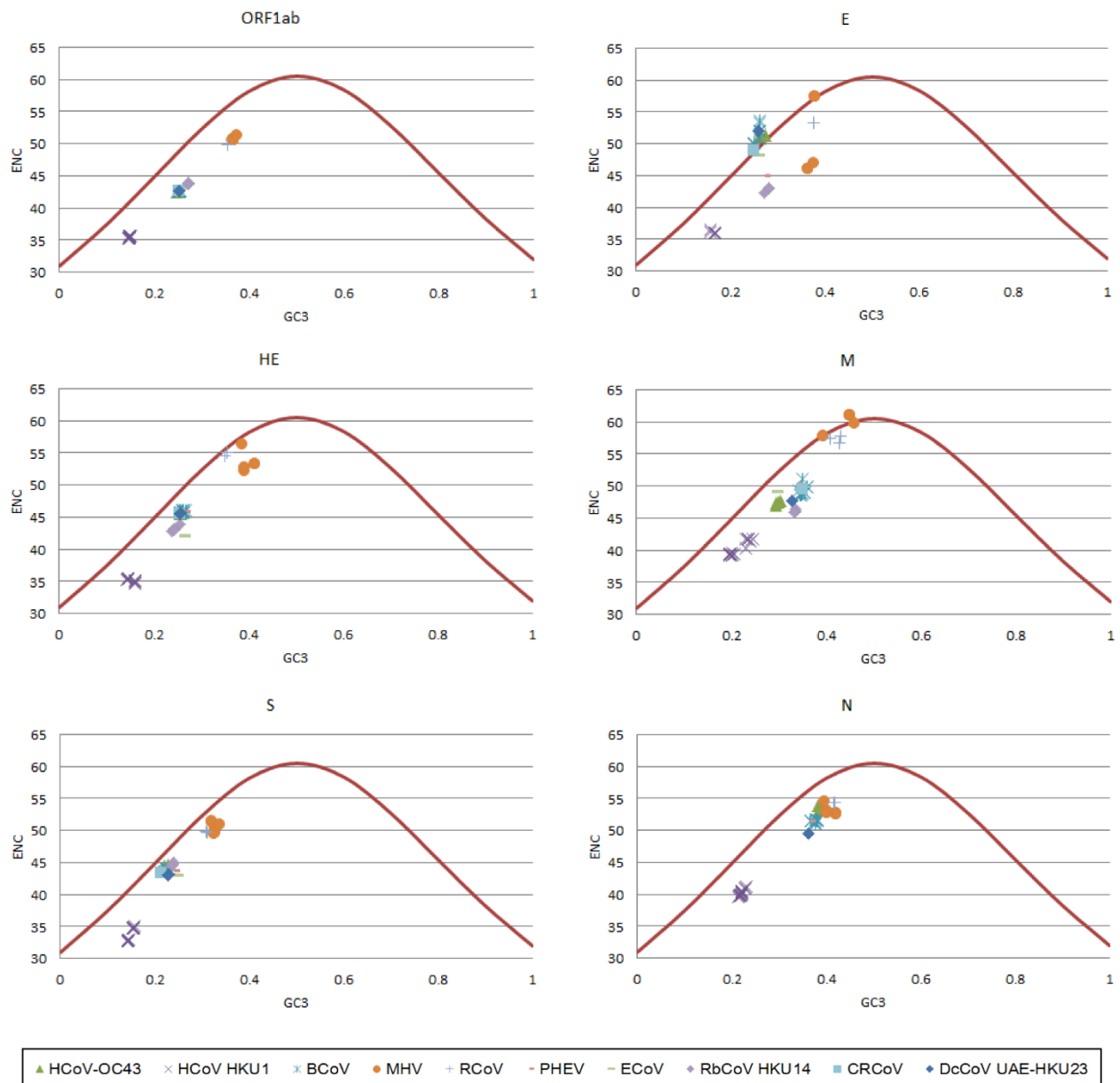
Technical Appendix Table 2. Characteristics of putative nonstructural proteins of ORF1ab in DcCoV UAE-HKU23*

nsp	Putative function/domain	First amino acid residue ^{position}	Last amino acid residue ^{position}	Length, aa	Length of corresponding protein in HCoV-OC43, BCoV, ECoV, and PHEV
nsp1	Leader protein	M ¹	G ²⁴⁶	246	246 (244 in ECoV)
nsp2	MHV p65-like protein	V ²⁴⁷	A ⁸⁵¹	605	605 (601 in ECoV)
nsp3	PL1pro, PL2pro, AC, ADRP, HD	G ⁸⁵²	G ²⁷⁵⁰	1899	1899 (1951 in ECoV)
nsp4	HD	A ²⁷⁵¹	Q ³²⁴⁶	496	496
nsp5	3CLpro	S ³²⁴⁷	Q ³⁵⁴⁹	303	303
nsp6	HD	S ³⁵⁵⁰	Q ³⁸³⁶	287	287
nsp7	Unknown	S ³⁸³⁷	Q ³⁹²⁵	89	89
nsp8	Unknown	A ³⁹²⁶	Q ⁴¹²²	197	197
nsp9	Unknown	N ⁴¹²³	Q ⁴²³²	110	110
nsp10	Unknown	A ⁴²³³	Q ⁴³⁶⁹	137	137
nsp11	Unknown (short peptide at the end of ORF1a)	S ⁴³⁷⁰	V ⁴³⁸³	14	14
nsp12	RdRp	S ⁴³⁷⁰	Q ⁵²⁹⁷	928	928
nsp13	Hel	S ⁵²⁹⁸	Q ⁵⁹⁰⁰	603	603 (599 in ECoV)
nsp14	ExoN	C ⁵⁹⁰¹	Q ⁶⁴²¹	521	521
nsp15	XendoU	S ⁶⁴²²	Q ⁶⁷⁹⁵	374	375 (366 in ECoV and 374 in BCoV)
nsp16	2'-O-MT	A ⁶⁷⁹⁶	I ⁷⁰⁹⁴	299	299

*ORF, open-reading frame; DcCoV, dromedary camel coronavirus; PL1pro, PL2pro, papain-like protease 1 and papain-like protease 2; AC, acidic domain; ADRP, adenosine diphosphate-ribose 1'-phosphatase; HD, hydrophobic domain; 3CLpro, 3C-like protease; RdRp, RNA-dependent RNA polymerase; Hel, helicase; ExoN, 3'-to-5' exonuclease; XendoU, poly(U)-specific endoribonuclease; 2'-O-MT, 2'-O-ribose methyltransferase.



Technical Appendix Figure 1. Estimation of tMRCA (time to most common ancestor) of dromedary camel coronavirus (DcCoV) UAE-HKU23 strains, DcCoV UAE-HKU23 strains/bat CoV, and DcCoV UAE-HKU23 strains/bat CoV/HCoV-OC43 based on the RNA-dependent RNA polymerase gene. The mean estimated dates are labeled and are represented by gray squares. The taxa are labeled with their sampling years.



Technical Appendix Figure 2. A plot of effective number of codons (ENC) against use of G or C at third position of codons (GC3) of open-reading frame (ORF)1ab, HE, S, E, M, and N genes in lineage A of *Betacoronavirus*. The continuous curve plots the relationship between ENC and GC3 in the absence of selection.