Human endemic coronavirus emergence in the context of past and recent zoonotic outbreaks

Diego Forni, Rachele Cagliani, Federica Arrigoni, Martino Benvenuti, Chiara Pontremoli, Uberto Pozzoli , Mario Clerici, Luca De Gioia, Manuela Sironi

Supplementary Material

Supplementary Table 1. List of endemic coronavirus strains.

HCoV-OC43

Accession ID	Species	Collection Date
NC_006213	Human coronavirus OC43	1967
KY014281	Human coronavirus OC43	2002
JN129834	Human coronavirus OC43	2004
JN129835	Human coronavirus OC43	2004
AY903459	Human coronavirus OC43	2004
AY903460	Human coronavirus OC43	2004
KY014282	Human coronavirus OC43	2007
KY983583	Human coronavirus OC43	2015
KY983585	Human coronavirus OC43	2015
KY983588	Human coronavirus OC43	2015
KY967356	Human coronavirus OC43	2015
KY967358	Human coronavirus OC43	2015
KY967359	Human coronavirus OC43	2015
KY967360	Human coronavirus OC43	2015
KY967361	Human coronavirus OC43	2015
KY554972	Human coronavirus OC43	2016
KY554973	Human coronavirus OC43	2016
KY554974	Human coronavirus OC43	2016
KY554975	Human coronavirus OC43	2016
KY674917	Human coronavirus OC43	2016
KY674918	Human coronavirus OC43	2016
KY674920	Human coronavirus OC43	2016
KY684759	Human coronavirus OC43	2016
KY369905	Human coronavirus OC43	2016
KY369906	Human coronavirus OC43	2016
KY369907	Human coronavirus OC43	2016
MN306036	Human coronavirus OC43	2019
MN306041	Human coronavirus OC43	2019
MN306042	Human coronavirus OC43	2019
MN306043	Human coronavirus OC43	2019
MN306053	Human coronavirus OC43	2019
MN310476	Human coronavirus OC43	2019
MN310478	Human coronavirus OC43	2019
KF530093	Human coronavirus OC43	1983
KF530060	Human coronavirus OC43	1985
KF530085	Human coronavirus OC43	1987
KF530086	Human coronavirus OC43	1987
KF530087	Human coronavirus OC43	1987
KF530077	Human coronavirus OC43	1987
KF530083	Human coronavirus OC43	1987
KF530073	Human coronavirus OC43	1989
KF530066	Human coronavirus OC43	1990

KF530065	Human coronavirus OC43	1990
KF530061	Human coronavirus OC43	1990
KF530088	Human coronavirus OC43	1990
KF530076	Human coronavirus OC43	1991
KF530096	Human coronavirus OC43	1991
KF530091	Human coronavirus OC43	1991
KF530089	Human coronavirus OC43	1991
KF530095	Human coronavirus OC43	1991
KF530067	Human coronavirus OC43	1991
KF530082	Human coronavirus OC43	1991
KF530094	Human coronavirus OC43	1991
KF530079	Human coronavirus OC43	1991
KF530071	Human coronavirus OC43	1992
KF530097	Human coronavirus OC43	1992
KF530074	Human coronavirus OC43	1992
KF530090	Human coronavirus OC43	1993
KF530059	Human coronavirus OC43	1995
KF530084	Human coronavirus OC43	1995
KF530062	Human coronavirus OC43	1995
KF530075	Human coronavirus OC43	1995
KF530098	Human coronavirus OC43	1996
KF530064	Human coronavirus OC43	1996
KF530078	Human coronavirus OC43	1996
KF530063	Human coronavirus OC43	1996
KF530099	Human coronavirus OC43	1997
KF530072	Human coronavirus OC43	1997
KF530080	Human coronavirus OC43	1997
KF530069	Human coronavirus OC43	1998
KF530081	Human coronavirus OC43	1999
KF530070	Human coronavirus OC43	1999
KF530068	Human coronavirus OC43	2000
KF530092	Human coronavirus OC43	2000
KF923905	Human coronavirus OC43	2005
KF923889	Human coronavirus OC43	2006
KF923899	Human coronavirus OC43	2006
KF923900	Human coronavirus OC43	2006
KF923890	Human coronavirus OC43	2007
KF923892	Human coronavirus OC43	2007
KF923894	Human coronavirus OC43	2007
KF923907	Human coronavirus OC43	2007
KF923921	Human coronavirus OC43	2007
KF923891	Human coronavirus OC43	2007
KF923901	Human coronavirus OC43	2007
KF923908	Human coronavirus OC43	2007
KF923909	Human coronavirus OC43	2007
KF923910	Human coronavirus OC43	2007
KF923911	Human coronavirus OC43	2007
KF923912	Human coronavirus OC43	2007
KF923913	Human coronavirus OC43	2007

KF923914	Human coronavirus OC43	2007
KF923915	Human coronavirus OC43	2007
KF923916	Human coronavirus OC43	2007
KF923917	Human coronavirus OC43	2007
KF923919	Human coronavirus OC43	2007
KF923920	Human coronavirus OC43	2007
KF923923	Human coronavirus OC43	2008
KF923922	Human coronavirus OC43	2009
KP198611	Human coronavirus OC43	2010
KF923886	Human coronavirus OC43	2010
KF923887	Human coronavirus OC43	2010
KF923918	Human coronavirus OC43	2010
KP198610	Human coronavirus OC43	2010
KF923893	Human coronavirus OC43	2010
KF923895	Human coronavirus OC43	2010
KF923924	Human coronavirus OC43	2010
KF923888	Human coronavirus OC43	2010
KF923925	Human coronavirus OC43	2010
KX344031	Human coronavirus OC43	2011
KU131570	Human coronavirus OC43	2011
KJ958218	Human coronavirus OC43	2011
KJ958219	Human coronavirus OC43	2011
KF923896	Human coronavirus OC43	2012
KX538964	Human coronavirus OC43	2012
KF923898	Human coronavirus OC43	2012
KF923906	Human coronavirus OC43	2012
KX538965	Human coronavirus OC43	2012
KX538966	Human coronavirus OC43	2012
KF923902	Human coronavirus OC43	2012
KF923903	Human coronavirus OC43	2012
KF923904	Human coronavirus OC43	2012
KX538967	Human coronavirus OC43	2012
KX538968	Human coronavirus OC43	2012
KX538969	Human coronavirus OC43	2012
KF923897	Human coronavirus OC43	2012
KX538970	Human coronavirus OC43	2012
KX538971	Human coronavirus OC43	2012
KX538972	Human coronavirus OC43	2012
KX538973	Human coronavirus OC43	2012
KX538974	Human coronavirus OC43	2012
KX538975	Human coronavirus OC43	2012
KX538976	Human coronavirus OC43	2012
KX538977	Human coronavirus OC43	2012
KX538978	Human coronavirus OC43	2013
KX538979	Human coronavirus OC43	2013
MG197718	Human coronavirus OC43	2015
MG197709	Human coronavirus OC43	2015
MG197710	Human coronavirus OC43	2015
MG197715	Human coronavirus OC43	2015

MG197719	Human coronavirus OC43	2015
MG197720	Human coronavirus OC43	2015
MG197716	Human coronavirus OC43	2015
MG197711	Human coronavirus OC43	2015
MG197712	Human coronavirus OC43	2015
MG197721	Human coronavirus OC43	2015
MG197717	Human coronavirus OC43	2015
MG197714	Human coronavirus OC43	2015
MG197713	Human coronavirus OC43	2015
MG197722	Human coronavirus OC43	2015
MG977446	Human coronavirus OC43	2016
MF374983	Human coronavirus OC43	2016
MF314143	Human coronavirus OC43	2016
MG197723	Human coronavirus OC43	2016
MG977451	Human coronavirus OC43	2016
MG977452	Human coronavirus OC43	2016
MH121121	Human coronavirus OC43	2016
MG977447	Human coronavirus OC43	2016
MG977448	Human coronavirus OC43	2016
MG977449	Human coronavirus OC43	2016
MG977450	Human coronavirus OC43	2016
MG977445	Human coronavirus OC43	2016
MF374984	Human coronavirus OC43	2017
MG977444	Human coronavirus OC43	2017
MF374985	Human coronavirus OC43	2017
MN026165	Human coronavirus OC43	2017
MN026164	Human coronavirus OC43	2018
NC_003045	Bovine coronavirus (Betacoronavirus 2)	

HCoV-229E

Accession ID	Species	Collection Date
NC_002645	Human coronavirus 229E	1962
JX503061	Human coronavirus 229E	2009
JX503060	Human coronavirus 229E	2010
KY983587	Human coronavirus 229E	2015
KY967357	Human coronavirus 229E	2015
KU291448	Human coronavirus 229E	2015
KY621348	Human coronavirus 229E	2016
KY674914	Human coronavirus 229E	2016
KY674919	Human coronavirus 229E	2016
KY684760	Human coronavirus 229E	2016
KY369908	Human coronavirus 229E	2016
KY369909	Human coronavirus 229E	2016
KY369910	Human coronavirus 229E	2016
KY369911	Human coronavirus 229E	2016
KY369912	Human coronavirus 229E	2016

KY369913	Human coronavirus 229E	2016
KY369914	Human coronavirus 229E	2016
MT438696	Human coronavirus 229E	2016
MF542265	Human coronavirus 229E	2016
MN369046	Human coronavirus 229E	2018
MN306046	Human coronavirus 229E	2019
KF514429	Human coronavirus 229E	1989
KF514432	Human coronavirus 229E	1993
KF514433	Human coronavirus 229E	1993
KF514430	Human coronavirus 229E	1993
KF514431	Human coronavirus 229E	1993
MT438700	Human coronavirus 229E	2017
MT438697	Human coronavirus 229E	2017
MT438698	Human coronavirus 229E	2017
MT438699	Human coronavirus 229E	2017
MW039392	Human coronavirus 229E	2020
AY386390*	Human coronavirus 229E	1967
AB691764*	Human coronavirus 229E	1979
DQ243964*	Human coronavirus 229E	1979
DQ243965*	Human coronavirus 229E	1982
DQ243966*	Human coronavirus 229E	1982
DQ243967*	Human coronavirus 229E	1982
DQ243968*	Human coronavirus 229E	1982
DQ243969*	Human coronavirus 229E	1982
DQ243970*	Human coronavirus 229E	1982
DQ243971*	Human coronavirus 229E	1984
DQ243972*	Human coronavirus 229E	1984
DQ243973*	Human coronavirus 229E	1990
DQ243974*	Human coronavirus 229E	1992
DQ243975*	Human coronavirus 229E	1992
DQ243976*	Human coronavirus 229E	1992
AY386388*	Human coronavirus 229E	1999
AY386389*	Human coronavirus 229E	1999
AY386386*	Human coronavirus 229E	1999
AY386383*	Human coronavirus 229E	1999
DQ243977*	Human coronavirus 229E	2001
DQ243978*	Human coronavirus 229E	2001
DQ243979*	Human coronavirus 229E	2002
DQ243980*	Human coronavirus 229E	2003
DQ243983*	Human coronavirus 229E	2003
DQ243984*	Human coronavirus 229E	2003
DQ243985*	Human coronavirus 229E	2003
DQ243986*	Human coronavirus 229E	2003
AB691765*	Human coronavirus 229E	2004
KM055557*	Human coronavirus 229E	2005
KM055559*	Human coronavirus 229E	2005
KM055551*	Human coronavirus 229E	2007
KM055552*	Human coronavirus 229E	2007
GU068546*	Human coronavirus 229E	2007

AB691767*	Human coronavirus 229E	2008
KM055554*	Human coronavirus 229E	2008
KM055555*	Human coronavirus 229E	2008
KM055556*	Human coronavirus 229E	2008
GU068547*	Human coronavirus 229E	2008
GU068548*	Human coronavirus 229E	2008
GU068549*	Human coronavirus 229E	2008
KM055532*	Human coronavirus 229E	2009
KM055533*	Human coronavirus 229E	2009
KM055534*	Human coronavirus 229E	2009
KM055535*	Human coronavirus 229E	2009
KM055536*	Human coronavirus 229E	2009
KM055537*	Human coronavirus 229E	2009
KM055545*	Human coronavirus 229E	2009
KM055546*	Human coronavirus 229E	2009
KM055547*	Human coronavirus 229E	2009
KM055553*	Human coronavirus 229E	2009
KM055558*	Human coronavirus 229E	2009
KM055560*	Human coronavirus 229E	2009
KM055538*	Human coronavirus 229E	2010
KM055539*	Human coronavirus 229E	2010
KM055540*	Human coronavirus 229E	2010
KM055541*	Human coronavirus 229E	2010
KM055542*	Human coronavirus 229E	2010
KM055543*	Human coronavirus 229E	2010
KM055531*	Human coronavirus 229E	2011
KM055544*	Human coronavirus 229E	2011
KM055548*	Human coronavirus 229E	2011
KM055549*	Human coronavirus 229E	2011
KM055550*	Human coronavirus 229E	2011
JX513245*	Human coronavirus 229E	2011
KJ866102*	Human coronavirus 229E	2013
NC_028752	Camel alphacoronavirus	2015

* strain used only in the spike RBD analyses

HCoV-HKU1

Accession ID	Release_Date	Species
NC_006577	Human coronavirus HKU1	2004
DQ415903	Human coronavirus HKU1	2003
DQ339101	Human coronavirus HKU1	2004
DQ415896	Human coronavirus HKU1	2004
DQ415897	Human coronavirus HKU1	2004
DQ415898	Human coronavirus HKU1	2004
DQ415904	Human coronavirus HKU1	2004
DQ415905	Human coronavirus HKU1	2004
DQ415906	Human coronavirus HKU1	2004
DQ415907	Human coronavirus HKU1	2004

DQ415908	Human coronavirus HKU1	2004
DQ415909	Human coronavirus HKU1	2004
DQ415910	Human coronavirus HKU1	2004
DQ415911	Human coronavirus HKU1	2004
DQ415912	Human coronavirus HKU1	2004
DQ415913	Human coronavirus HKU1	2004
DQ415914	Human coronavirus HKU1	2004
DQ415899	Human coronavirus HKU1	2005
DQ415900	Human coronavirus HKU1	2005
DQ415902	Human coronavirus HKU1	2005
KY983584	Human coronavirus HKU1	2015
KY674943	Human coronavirus HKU1	2016
MK167038	Human coronavirus HKU1	2017
KF686338	Human coronavirus HKU1	2005
HM034837	Human coronavirus HKU1	2005
KF850450	Human coronavirus HKU1	2005
KT779555	Human coronavirus HKU1	2009
KT779556	Human coronavirus HKU1	2009
KF686340	Human coronavirus HKU1	2009
KF686342	Human coronavirus HKU1	2009
KF686344	Human coronavirus HKU1	2009
KF430202	Human coronavirus HKU1	2010
KF686343	Human coronavirus HKU1	2010
KF686341	Human coronavirus HKU1	2010
KF430201	Human coronavirus HKU1	2010
KF686345	Human coronavirus HKU1	2010
MH940245	Human coronavirus HKU1	2017
NC_012936	Murine coronavirus (Rat coronavirus Parker)	1970

HCoV-NL63

Accession ID	Species	Collection Date
NC_005831	Human coronavirus NL63	2003
KF530110	Human coronavirus NL63	1983
KF530106	Human coronavirus NL63	1987
KF530114	Human coronavirus NL63	1989
KF530108	Human coronavirus NL63	1989
KF530111	Human coronavirus NL63	1990
KF530109	Human coronavirus NL63	1990
KF530104	Human coronavirus NL63	1990
KF530113	Human coronavirus NL63	1990
KF530107	Human coronavirus NL63	1991
KF530105	Human coronavirus NL63	2001
KF530112	Human coronavirus NL63	2001
JX504050	Human coronavirus NL63	2004
JQ765568	Human coronavirus NL63	2005
JQ765569	Human coronavirus NL63	2005

JQ765570	Human coronavirus NL63	2005
JQ765571	Human coronavirus NL63	2005
JQ900258	Human coronavirus NL63	2005
JQ765572	Human coronavirus NL63	2005
JQ900259	Human coronavirus NL63	2005
JQ765573	Human coronavirus NL63	2005
JQ900260	Human coronavirus NL63	2005
JQ765574	Human coronavirus NL63	2005
JQ765575	Human coronavirus NL63	2005
JQ765566	Human coronavirus NL63	2008
JX104161	Human coronavirus NL63	2008
JX524171	Human coronavirus NL63	2009
JQ765565	Human coronavirus NL63	2009
JQ900257	Human coronavirus NL63	2009
JQ900255	Human coronavirus NL63	2009
JQ765564	Human coronavirus NL63	2009
JQ900256	Human coronavirus NL63	2009
JQ765567	Human coronavirus NL63	2009
JQ765563	Human coronavirus NL63	2009
MG428702	Human coronavirus NL63	2010
MG428699	Human coronavirus NL63	2010
MG428704	Human coronavirus NL63	2010
MG428707	Human coronavirus NL63	2010
MG428705	Human coronavirus NL63	2010
MG428706	Human coronavirus NL63	2010
MG428700	Human coronavirus NL63	2010
MG428703	Human coronavirus NL63	2010
MG428701	Human coronavirus NL63	2010
MG772808	Human coronavirus NL63	2014
KY983586	Human coronavirus NL63	2015
KY829118	Human coronavirus NL63	2015
KT381875	Human coronavirus NL63	2015
KT266906	Human coronavirus NL63	2015
KX179500	Human coronavirus NL63	2015
KU521535	Human coronavirus NL63	2015
KY554967	Human coronavirus NL63	2016
KY554968	Human coronavirus NL63	2016
KY554969	Human coronavirus NL63	2016
KY554970	Human coronavirus NL63	2016
KY554971	Human coronavirus NL63	2016
KY674915	Human coronavirus NL63	2016
KY674916	Human coronavirus NL63	2016
MN306018	Human coronavirus NL63	2018
MN026166	Human coronavirus NL63	2018
MK334044	Human coronavirus NL63	2018
MK334047	Human coronavirus NL63	2018
MK334045	Human coronavirus NL63	2018
MK334043	Human coronavirus NL63	2018
MK334046	Human coronavirus NL63	2018

MN306040	Human coronavirus NL63	2019
DQ445911	Human coronavirus NL63	2002
DQ445912	Human coronavirus NL63	2003
AY518894	Human coronavirus NL63	1988
NC_032107	NL63-related bat coronavirus	2010

Supplementary Table 2. HCoV-229E and HCoV-OC43 positively selected sites.

HCoV-229E					
Protein	**aa position**	**Ancestral allele**	**Major allele**	**Major allele frequency**	**Posterior probability**
ORF1a	1253	Gly	Ile	1	0.94
	1255	Pro	Asn	1	1.00
	1256	Pro	Thr	1	1.00
	1257	Ala	Val	1	1.00
	1258	Glu	Asp	1	1.00
	1259	Leu	Thr	1	1.00
	1260	Ala	Ile	0.89	1.00
	1262	Ile	Lys	1	1.00
	1263	Asp	Glu	1	0.99
	1265	Thr	Phe	1	0.99
	1269	Val	Glu	1	0.98
	1273	Ser	Ala	1	0.97
	1276	Thr	Val	1	0.97
	1277	Cys	His	0.89	0.97
	1278	Asn	Gly	0.89	0.97
	1279	Ser	Val	0.69	0.93
Spike	21	Val	Gly	1	0.80
	22	Leu	Thr	0.74	0.79
	307	Asp	Lys	0.54	0.85
	309	Lys	Glu	0.81	0.90
	310	Pro	His	0.7	0.93
	312	Ser	Arg	0.77	0.92
	314	Ala	Pro	0.81	0.88
	316	Thr	Lys	0.74	0.78
	349	Lys	Gln	0.94	0.81
	350	Tyr	Phe	0.94	0.89
	353	Ala	Asp	0.68	0.98
	355	Ser	Val	0.81	0.99
	356	Asn	Lys	0.94	0.99
	357	Val	Leu	0.71	0.99
	358	Gly	Ala	0.65	0.98
	404	Phe	Leu	0.97	0.91
	405	Ala	Val	0.74	0.95
	406	Tyr	Asn	0.97	0.97
	407	Ile	His	0.61	0.98
	408	Asn	Lys	0.77	0.96
	410	Tyr	His	0.97	0.92
	411	Thr	Asn	0.9	0.90
	519	Ser	Phe	1	0.83
	520	Ser	Thr	1	0.79
	970	Asn	Gly	1	0.90
	971	Thr	Ile	0.65	0.87
	1056	Gly	Val	1	0.77

	1057	Ile	Val	1	0.77
	1058	Asp	Glu	1	0.76
M	2	Thr	Ser	1	0.97
	3	Glu	Asn	1	0.98
	4	Ala	Asp	1	0.98
	6	Ile	Cys	1	0.98
	8	Glu	Gly	1	0.97
	9	Glu	Asp	1	0.96
	10	Leu	Ile	1	0.95

HCoV-OC43

Protein	aa position	Ancestral allele	Major allele	Major allele frequency	Posterior probability
ORF1a	864	Asn	Lys	0.99	0.76
	867	Ala	Ile	0.99	0.81
	869	Thr	Met	0.95	0.80
E	8	Phe	Leu	1	0.96
	29	Ile	Thr	1	0.95
	74	Val	Ile	0.75	0.81
HE	114	Thr	Asn	0.98	0.98
	115	Thr	Arg	0.98	0.99
	177	Arg	Pro	0.98	0.83
	178	Glu	Gln	0.98	0.85
	180	Asn	Lys	0.62	0.86
	181	Phe	Ser	0.98	0.86
	184	Tyr	Ile	0.74	0.91
	185	Tyr	Ile	0.79	0.90
	186	Tyr	Thr	0.79	0.89
Spike	22	Thr	Pro	0.62	0.81
	23	Thr	Leu	0.86	0.99
	24	Val	Asp	0.99	1.00
	26	Ile	Phe	0.74	0.99
	28	Asp	Asn	0.83	0.97
	29	Val	Arg	0.62	0.97
	33	Val	Pro	0.68	0.83
	84	Leu	Asp	0.86	0.93
	143	His	Arg	1	0.89
	145	Thr	Ile	1	0.95
	146	Ile	Asn	1	0.96
	147	Leu	Ser	0.83	0.97
	148	Gly	Thr	0.81	0.95
	183	Gln	His	1	0.86
	184	Arg	Phe	0.86	0.98
	185	Val	Lys	0.86	0.99

	259	Ala	Arg	0.86	0.99
	260	Leu	Phe	0.99	0.97
	533	Ser	Phe	0.71	0.78
	536	Pro	Thr	0.7	0.88
	617	Ile	Glu	0.58	0.84
	630	Thr	Ser	0.99	0.88
	708	Ser	Thr	0.95	0.76
	732	Ser	Ala	0.99	0.91
	733	Ser	Ile	1	0.95
	734	Val	Ser	1	0.97
	912	Ser	Asp	1	0.96
	1305	Gly	Cys	0.74	0.81
	1306	Phe	Leu	1	0.80
ORF5	5	Lys	Arg	1	0.94
	8	Ile	Lys	1	0.99
	10	Tyr	Phe	0.74	0.83
	14	Thr	Ile	0.99	0.92
	32	Lys	Gln	1	0.90
	52	Ser	Arg	0.95	0.92
	57	Leu	Phe	1	0.90
	74	Thr	Asn	1	0.90
	85	Glu	Asp	1	0.89
M	114	Trp	Phe	0.99	0.97
	183	Leu	Cys	0.99	0.87

Supplementary Table 3. List of predicted linear B cell epitopes for the spike proteins of HCoV-229E and HCoV-OC43.

HCoV-229E		
Isolate genBank ID (Class)	Sequence	Length
NC_002645 (Class I)	KPQSGGGKC	9
	CPFSFGKVNNFVKF	14
	LKDIPGG	7
	WAYSKY	6
DQ243968 (Class I/II)	KLQSGVGR	8
	GNCPFSFGKVNNFVKFG	17
	LKDIPGG	7
DQ243976 (class IV)	RSGVGRCYNC	10
	NCPFSFGKVNNFVK	14
	KDIPGG	6
	LANLNSH	7
KM055554 (class V)	RGPGRCYNCRP	11
	FNETKG	6
	NCPFSFGKVNNFVKF	15
	LKDIPGG	7
	LVNHKSH	7
KY369908 (Class VI)	FEHQRGPGK	9
	NVKLAR	6
	NCPFSFGKVNNFVKF	15
	LKDIPGG	7
	LVNHKS	6

HCoV-OC43		
Isolate genBank ID	Sequence	Length
NC_006213	TSDNINDKDTGPPPIST	17
	PTSGSTYRNMALK	13
	FKPPFL	6
	KVIKDRVMY	9
	TINSTQDGDNKL	12

	EYPQTICHPNLGNHRKELWHLDTGV	25
	MSDFMSEI	8
	MSDFMSEI	8
KF530093	TSDTSYINDVDTGVPPIST	19
	PTSGSTYRNMAL	12
	KVIKDDVL	8
	TINLDNKL	8
	EYPHTICHPNLGNHRTELWHLDTGV	25
	KRKDGFT	7
	MSDFMSEI	8
KF530071	LKCPLDTSYKGTFNNKDTGPPFIST	25
	PTSGSTYRNMALKGTDK	17
	TLWFKPPFL	9
	NTKVFKDGVMY	11
	TINSQDGDNKL	11
	EYPHTICHPKLGNHFKELWHLDTGV	25
	RRDIGFT	7
	MSDFMSEI	8
KF923918	NCPLDPRLKGSFNNRDTGPPSIST	24
	PTSGSTYRNMALKGTDL	17
	WFKPPFL	7
	NTKVFKDGVMY	11
	TINSTQDGVNKL	12
	EYPHTICHPKLGNHFKELWHSDTGV	25
	RLDIGFT	7
KY983583	CTLDPRLKGSFNNRDTGPPSISI	23
	PTSGSTYRNMALK	13
	FKPPFL	6
	TKVFKDGVMY	10
	TINSTQDGVNKL	12
	YPHTICHPNLGNHFKELWHYDTGV	24
	CMSDFM	6

Supplementary Table 4: List of BCoV strains.

Accession ID	Species	Collection Date
NC_003045.1	Bovine coronavirus	1997
DQ811784.2	Bovine coronavirus	1983
FJ425184.1	Waterbuck coronavirus	1994
FJ425185.1	Waterbuck coronavirus	1994
FJ425186.1	Waterbuck coronavirus	1994
FJ425187.1	White-tailed deer coronavirus	1994
FJ425188.1	Sambar deer coronavirus	1994
FJ425189.1	Sambar deer coronavirus	1994
FJ425190.1	Sambar deer coronavirus	1994
FJ938066.1	Bovine respiratory coronavirus	1996
FJ938063.1	Bovine coronavirus	1996
FJ938064.1	Bovine coronavirus	2000
FJ938065.1	Bovine respiratory coronavirus	2000
LC494164.1	Bovine coronavirus	2006
LC494162.1	Bovine coronavirus	2006
LC494163.1	Bovine coronavirus	2006
LC494160.1	Bovine coronavirus	2007
LC494161.1	Bovine coronavirus	2007
LC494172.1	Bovine coronavirus	2007
LC494173.1	Bovine coronavirus	2008
LC494174.1	Bovine coronavirus	2008
LC494177.1	Bovine coronavirus	2008
LC494165.1	Bovine coronavirus	2008
LC494166.1	Bovine coronavirus	2008
LC494167.1	Bovine coronavirus	2008
LC494168.1	Bovine coronavirus	2009
LC494169.1	Bovine coronavirus	2009
LC494170.1	Bovine coronavirus	2010
LC494171.1	Bovine coronavirus	2010
LC494129.1	Bovine coronavirus	2010

LC494130.1	Bovine coronavirus	2010
LC494131.1	Bovine coronavirus	2011
LC494132.1	Bovine coronavirus	2011
LC494133.1	Bovine coronavirus	2011
LC494135.1	Bovine coronavirus	2012
LC494136.1	Bovine coronavirus	2012
LC494127.1	Bovine coronavirus	2012
LC494128.1	Bovine coronavirus	2012
LC494134.1	Bovine coronavirus	2012
LC494137.1	Bovine coronavirus	2013
MG757138.1	Bovine coronavirus	2014
MG757139.1	Bovine coronavirus	2014
MG757140.1	Bovine coronavirus	2014
MG757141.1	Bovine coronavirus	2014
MG757142.1	Bovine coronavirus	2014
KX982264.1	Bovine coronavirus	2014
LC494154.1	Bovine coronavirus	2014
LC494155.1	Bovine coronavirus	2014
LC494156.1	Bovine coronavirus	2014
LC494138.1	Bovine coronavirus	2015
KU886219.1	Bovine coronavirus	2015
LC494139.1	Bovine coronavirus	2015
LC494157.1	Bovine coronavirus	2015
LC494140.1	Bovine coronavirus	2015
LC494175.1	Bovine coronavirus	2016
LC494176.1	Bovine coronavirus	2016
LC494126.1	Bovine coronavirus	2016
LC494141.1	Bovine coronavirus	2016
LC494158.1	Bovine coronavirus	2016
LC494146.1	Bovine coronavirus	2016
MH043955.1	Bovine coronavirus	2016
LC494147.1	Bovine coronavirus	2016
LC494142.1	Bovine coronavirus	2016
LC494143.1	Bovine coronavirus	2016

LC494144.1	Bovine coronavirus	2016
LC494145.1	Bovine coronavirus	2016
LC494148.1	Bovine coronavirus	2016
LC494149.1	Bovine coronavirus	2016
LC494178.1	Bovine coronavirus	2016
LC494179.1	Bovine coronavirus	2016
LC494180.1	Bovine coronavirus	2016
LC494181.1	Bovine coronavirus	2016
LC494182.1	Bovine coronavirus	2016
LC494150.1	Bovine coronavirus	2017
LC494151.1	Bovine coronavirus	2017
LC494152.1	Bovine coronavirus	2017
LC494153.1	Bovine coronavirus	2017
MG518518.1	Water deer coronavirus	2017
MH043952.1	Bovine coronavirus	2017
MH043953.1	Bovine coronavirus	2017
MH043954.1	Bovine coronavirus	2017
LC494159.1	Bovine coronavirus	2017
LC494183.1	Bovine coronavirus	2017
LC494184.1	Bovine coronavirus	2017
LC494185.1	Bovine coronavirus	2017
LC494186.1	Bovine coronavirus	2017
LC494187.1	Bovine coronavirus	2017
LC494188.1	Bovine coronavirus	2017
LC494189.1	Bovine coronavirus	2017
LC494190.1	Bovine coronavirus	2017
LC494191.1	Bovine coronavirus	2017
LC494192.1	Bovine coronavirus	2017

Supplementary Table 5. Model comparison using BEAST path sampling.

HCoV-OC43								
			Strict Clock			Uncorrelated Relaxed Clock		
Model			Constant Size	Exponential Growth	Bayesian Skyline	Constant Size	Exponential Growth	Bayesian Skyline
		Likelihood	-28182.79	-28222.07	-28208.74	-28203.88	-28424.75	-28253.29
Strict Clock	Constant Size	-28182.79	0	39.28	25.95	21.09	241.96	70.5
	Exponential Growth	-28222.07	-39.28	0	-13.33	-18.19	202.68	31.22
	Bayesian Skyline	-28208.74	-25.95	13.33	0	-4.86	216.01	44.55
Uncorrelated Relaxed Clock	Constant Size	-28203.88	-21.09	18.19	4.86	0	220.87	49.41
	Exponential Growth	-28424.75	-241.96	-202.68	-216.01	-220.87	0	-171.46
	Bayesian Skyline	-28253.29	-70.5	-31.22	-44.55	-49.41	171.46	0
HCoV-229E								
			Strict Clock			Uncorrelated Relaxed Clock		
Model			Constant Size	Exponential Growth	Bayesian Skyline	Constant Size	Exponential Growth	Bayesian Skyline
		Likelihood	-33463.83	-33470.38	-33468.54	-33466.82	-33489.96	-33484.14
Strict Clock	Constant Size	-33463.83	0	6.55	4.71	2.99	26.13	20.31
	Exponential Growth	-33470.38	-6.55	0	-1.84	-3.56	19.58	13.76
	Bayesian Skyline	-33468.54	-4.71	1.84	0	-1.72	21.42	15.6
Uncorrelated Relaxed Clock	Constant Size	-33466.82	-2.99	3.56	1.72	0	23.14	17.32
	Exponential Growth	-33489.96	-26.13	-19.58	-21.42	-23.14	0	-5.82
	Bayesian Skyline	-33484.14	-20.31	-13.76	-15.6	-17.32	5.82	0
HCoV-NL63								
			Strict Clock			Uncorrelated Relaxed Clock		
Model			Constant Size	Exponential	Bayesian	Constant Size	Exponential	Bayesian Skyline

				Growth	Skyline		Growth	
		Likelihood	-7082.57	-7062.08	-7071.20	-7061.60	-7048.26	-7052.97
Strict Clock	**Constant Size**	-7082.57	0	-20.49	-11.37	-20.97	-34.31	-29.6
	Exponential Growth	-7062.08	20.49	0	9.12	-0.48	-13.82	-9.11
	Bayesian Skyline	-7071.20	11.37	-9.12	0	-9.6	-22.94	-18.23
Uncorrelated Relaxed Clock	**Constant Size**	-7061.60	20.97	0.48	9.6	0	-13.34	-8.63
	Exponential Growth	-7048.26	34.31	13.82	22.94	13.34	0	4.71
	Bayesian Skyline	-7052.97	29.6	9.11	18.23	8.63	-4.71	0

Note: A Bayes factor >3 indicates positive support

Supplementary Table 6. List of analyzed viral disease outbreaks.

Zoonotic with no arthropod vector
Argentine hemorrhagic fever
Bolivian hemorrhagic fever
COVID19
Borna virus encephalitis
Ebola
Hantavirus pulmonary syndrome
Hantaviruses - Old World
Hendra virus disease
Hepatitis E
Herpes B infection
Influenza
Lassa fever
Lujo virus infection
Lymphocytic choriomeningitis
Marburg virus disease
MERS
Murray Valley and Kunjin
Monkeypox
Orf
Pseudocowpox
Rabies
SARS
Venezuelan hemorrhagic fever
Arthropod-borne
Barmah Forest disease
Bunyaviridae infections - misc.
California encephalitis group
Chandipura and Vesicular stomatitis viruses
Chikungunya
Crimean-Congo hemorrhagic fever
Dengue
Eastern equine encephalitis

Japanese encephalitis

Kyasanur Forest disease

Louping ill

Mayaro

Onyong nyong

Ockelbo disease

Oropouche

Phleboviruses - Old World

Pogosta disease

Powassan

Rift Valley fever

Rocio

Ross River disease

Severe fever with thrombocytopenia syndrome

Sindbis

St. Louis encephalitis

Tick-borne encephalitis

Venezuelan equine encephalitis

Wesselsbron

West Nile fever

Western equine encephalitis

Zika

Non-zoonotic
Adenovirus infection
Conjunctivitis - viral
Cytomegalovirus infection
Enterovirus infection
Gastroenteritis - viral
Hepatitis D
Herpes simplex infection
Human herpesvirus 6 infection
Infectious mononucleosis or EBV infection
Laryngotracheobronchitis

Meningitis - aseptic (viral)

Molluscum contagiosum

Parainfluenza virus infection

Parvovirus B19 infection

Pleurodynia

Respiratory syncytial virus infection

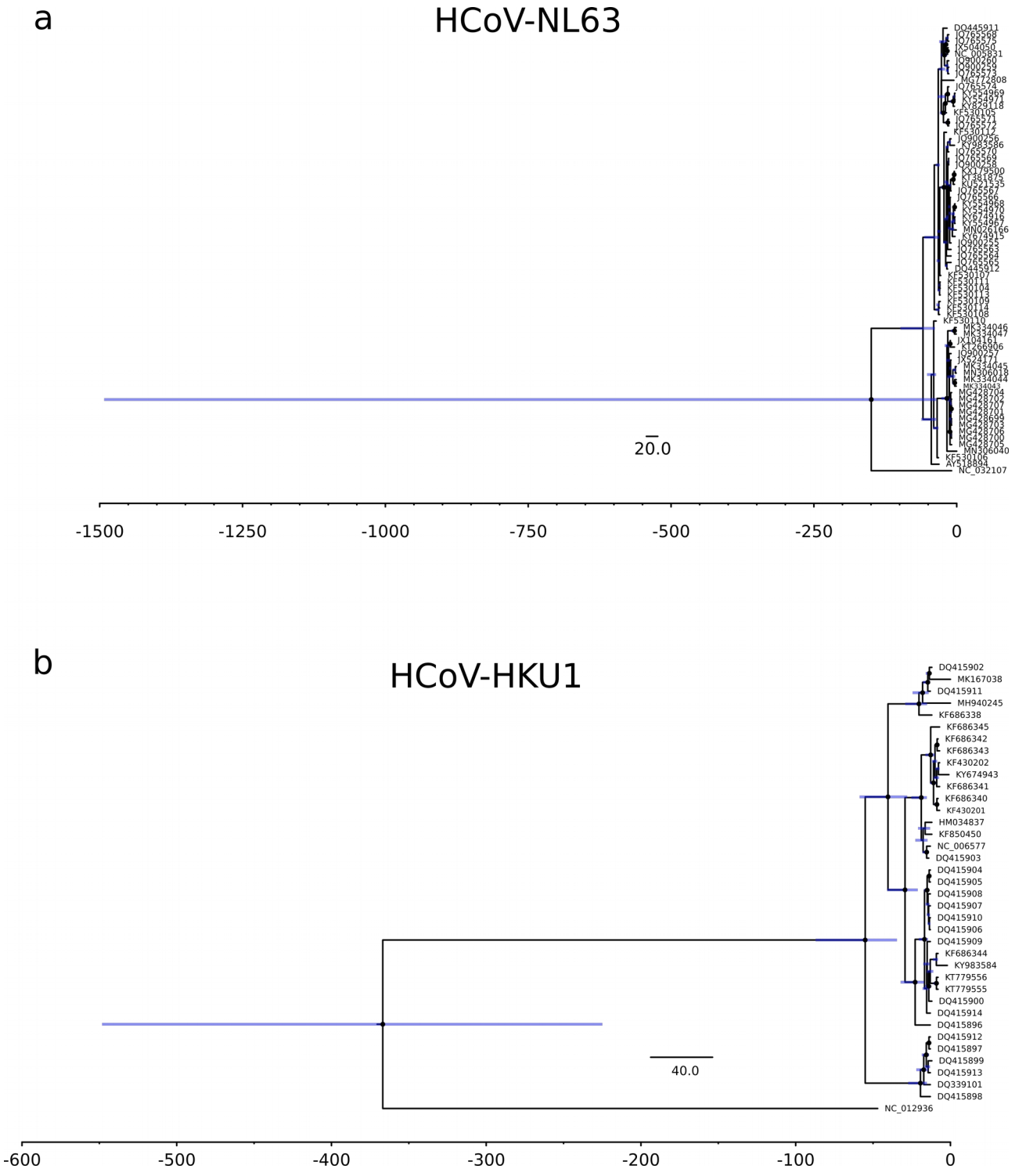
Supplementary Table 7. Viral disease outbreaks reported in GIDEON.

	Zoonotic with no arthropod vector		Arthropod-borne		Non-zoonotic	
year	All	At least 50 cases	All	At least 50 cases	All	At least 50 cases
1945	0	0	2	1	0	0
1946	0	0	0	0	2	1
1947	1	0	1	0	1	0
1948	0	0	2	0	1	0
1949	0	0	2	1	1	0
1950	0	0	1	0	1	1
1951	1	0	2	0	1	0
1952	1	0	2	1	2	0
1953	0	0	2	0	3	0
1954	1	0	3	1	5	0
1955	2	1	1	1	4	1
1956	0	0	4	2	4	3
1957	1	1	4	2	5	1
1958	1	0	1	1	6	2
1959	2	0	3	1	4	0
1960	0	0	3	0	4	1
1961	0	0	3	1	4	3
1962	1	0	7	1	5	1
1963	1	0	3	2	5	0
1964	2	1	3	2	7	0
1965	2	0	3	0	5	2
1966	0	0	2	1	7	1
1967	5	3	3	2	5	1
1968	2	1	3	1	5	4
1969	1	1	4	1	5	4
1970	1	1	4	0	7	3
1971	1	0	4	2	5	4
1972	2	1	3	1	5	3
1973	1	1	4	3	5	5
1974	3	1	6	3	7	3

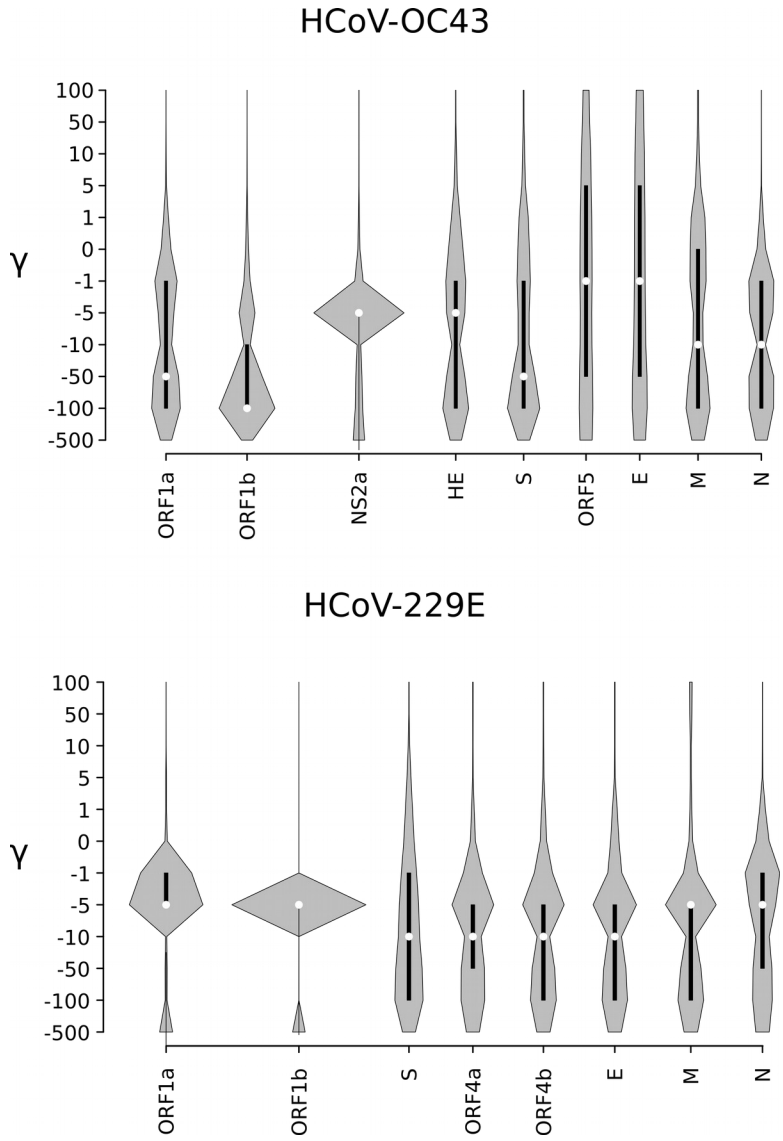
1975	4	1	6	3	7	6
1976	5	2	7	2	9	2
1977	2	0	7	5	6	3
1978	2	1	5	5	6	5
1979	3	1	3	2	6	4
1980	3	1	4	1	8	3
1981	6	1	7	1	7	5
1982	3	1	6	2	9	3
1983	2	1	7	4	8	6
1984	3	2	7	2	9	4
1985	3	2	4	2	7	2
1986	2	2	4	1	7	7
1987	3	1	6	4	9	6
1988	4	1	4	2	7	5
1989	5	2	7	2	8	6
1990	3	1	8	4	8	6
1991	3	2	6	2	11	4
1992	4	1	6	1	7	1
1993	2	1	6	2	8	5
1994	5	2	7	3	8	5
1995	7	4	9	5	8	5
1996	6	3	11	2	8	5
1997	7	4	6	4	6	4
1998	3	2	8	2	9	6
1999	6	3	8	3	8	5
2000	5	3	8	4	8	5
2001	4	3	8	5	7	5
2002	9	5	11	8	7	7
2003	8	6	11	7	7	5
2004	8	4	11	5	10	6
2005	8	3	11	7	8	5
2006	6	2	8	6	8	5
2007	10	4	12	7	9	6
2008	10	4	9	5	8	5
2009	9	3	14	7	8	4

2010	7	5	14	8	7	6
2011	11	4	13	6	8	6
2012	10	6	12	6	8	7
2013	10	2	13	6	7	6
2014	10	5	14	7	8	5
2015	7	4	13	8	7	7
2016	8	4	15	10	8	6
2017	11	6	12	6	8	6
2018	6	5	10	9	6	5
2019	8	4	12	6	5	3
2020	6	3	8	4	2	2

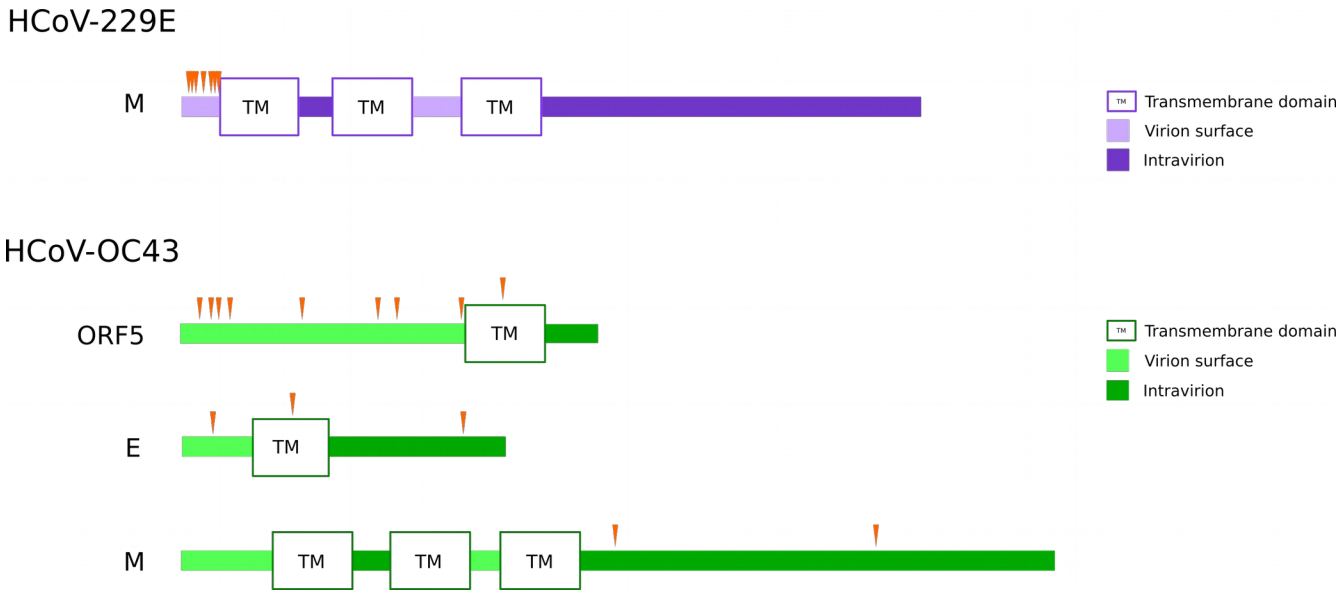
Supplementary Figure 1. HCoV-NL63 and HCoV-HKU1 timescaled phylogenetic trees. Maximum likelihood trees estimated for the non-recombinant region of HCoV-NL63 (a) and HCoV-HKU1 (b). Branch lengths represent the evolutionary time measured by the grids corresponding to the timescale shown at the tree base (in years). For internal nodes 95% credible interval bars are shown and black dots indicate a posterior probability >0.80 for that node. The gray dots indicate the most related animal coronaviruses.



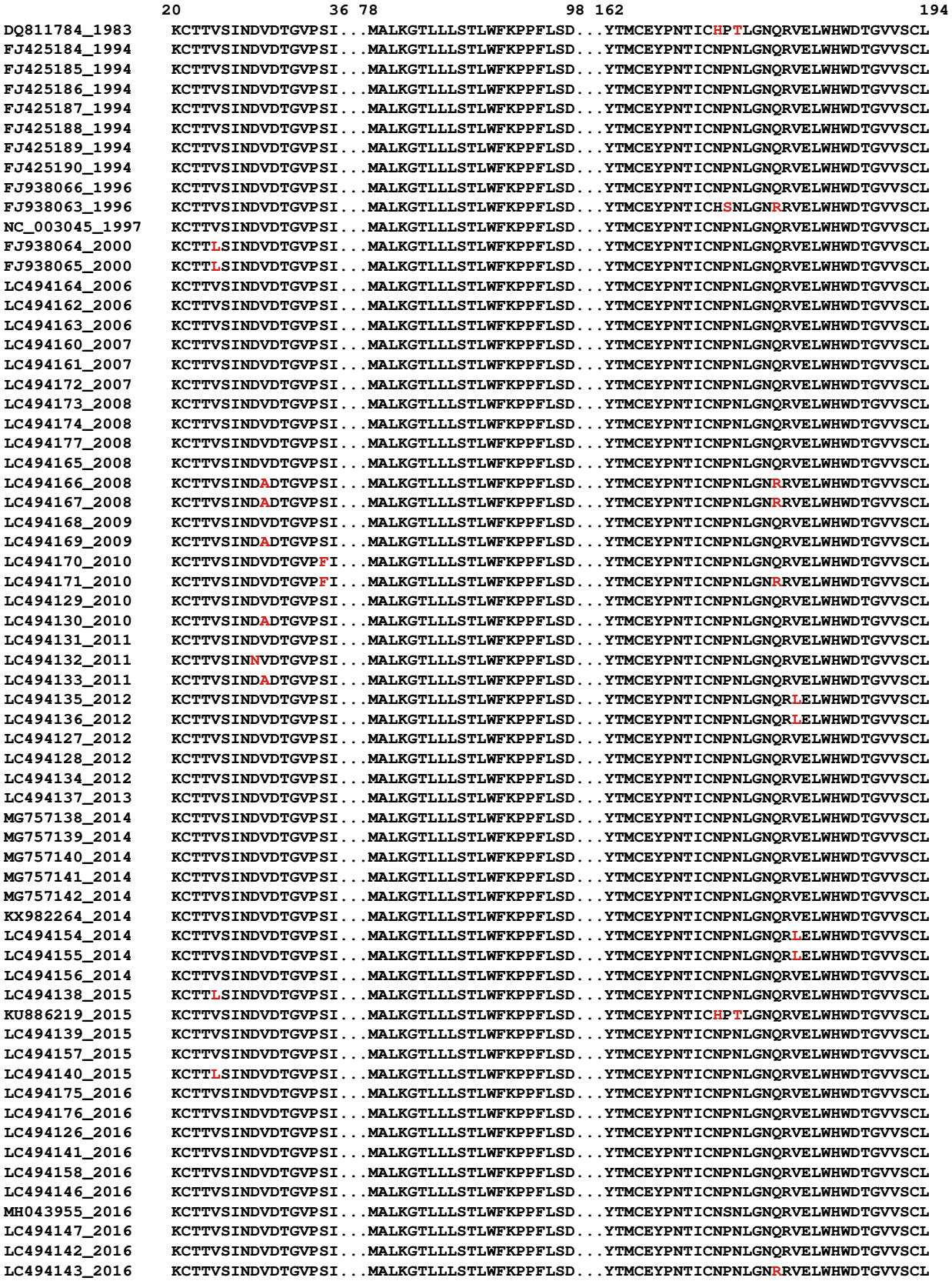
Supplementary Figure 2. Population genetics-phylogenetics analysis of HCoV-OC43 and HCoV-229E genes. Violin plots (median, white dot; interquartile range, black bar) of selection coefficients calculated for all HCoV-OC43 and HCoV-229E ORFs. Selection coefficients (γ) are classified as strongly beneficial (100, 50), moderately beneficial (10, 5), weakly beneficial (1), neutral (0), weakly deleterious (−1), moderately deleterious (−5, −10), strongly deleterious (−50, −100), and inviable (−500).



Supplementary Figure S3. Schematic representation of HCoV-OC43 and HCoV-229E proteins showing evidence of positive selection. Positions of positively selected sites detected by gammaMap are indicated with orange triangles.

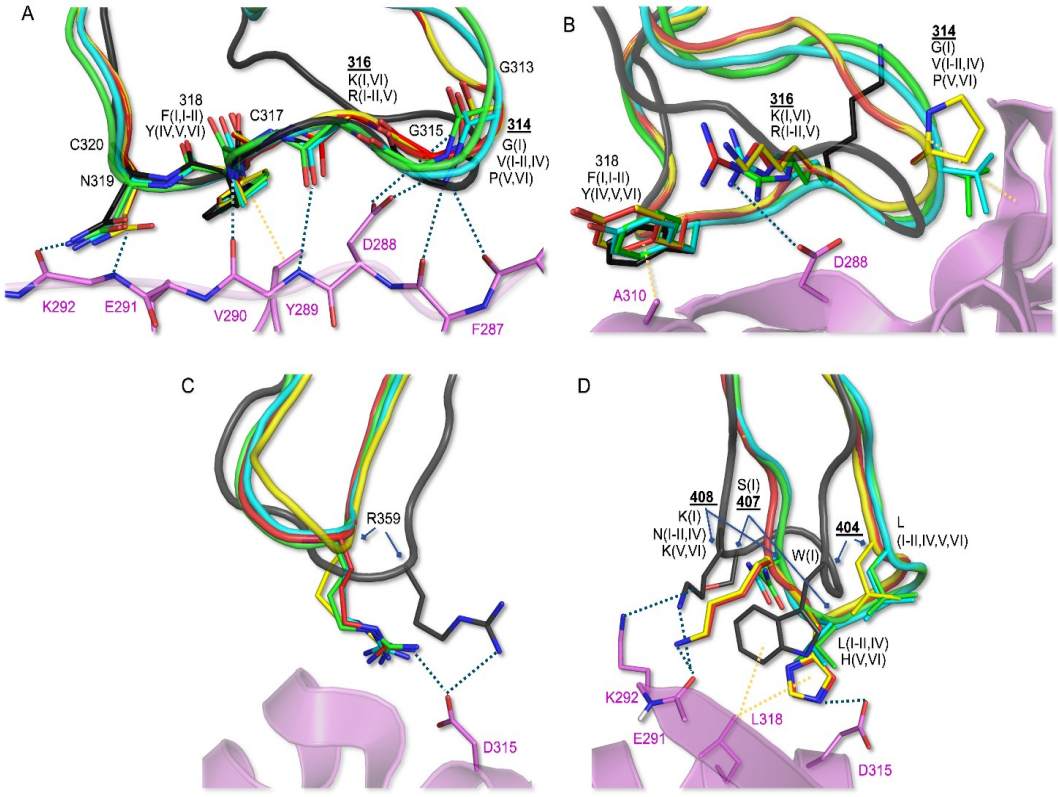


Supplementary Figure 4. Protein alignment of the three RBD loops in the spike protein of BCoV. Positions refer to NC_003045.

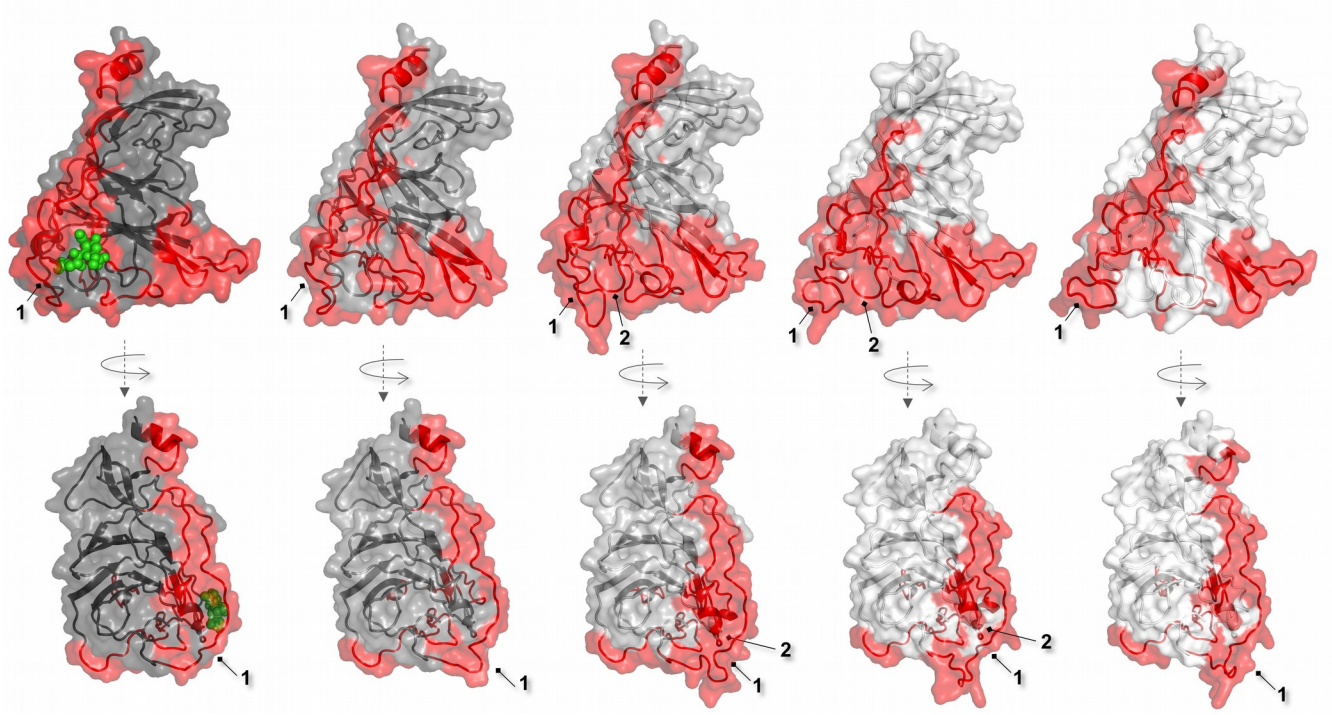


```
LC494144_2016    KCTTVSINDVDTGVPSI...MALKGTLLLSTLWFKPPFLSD...YTMCEYPNTICNPNLGNRRVELWHWDTGVVSCL
LC494145_2016    KCTTVSINDVDTGVPSI...MALKGTLLLSTLWFKPPFLSD...YTMCEYPNTICNPNLGNQRVELWHWDTGVVSCL
LC494148_2016    KCTTVSINDVDTGVPSI...MALKGTLLLSTLWFKPPFLSD...YTMCEYPNTICNPNLGNQRVELWHWDTGVVSCL
LC494149_2016    KCTTVSINDVDTGVPSI...MALKGTLLLSTLWFKPPFLSD...YTMCEYPNTICNPNLGNQRVELWHWDTGVVSCL
LC494178_2016    KCTTVSINDVDTGVPSI...MALKGTLLLSTLWFKPPFLSD...YTMCEYPNTICNPNLGNQRVELWHWDTGVVSCL
LC494179_2016    KCTTVSINDVDTGVPSI...MALKGTLLLSTLWFKPPFLSD...YTMCEYPNTICNPNLGNQRVELWHWDTGVVSCL
LC494180_2016    KCTTVSINDVDTGVPSI...MALKGTLLLSTLWFKPPFLSD...YTMCEYPNTICNPNLGNQRVELWHWDTGVVSCL
LC494181_2016    KCTTVSINDVDTGVPSI...MALKGTLLLSTLWFKPPFLSD...YTMCEYPNTICNPNLGNQRVELWHWDTGVVSCL
LC494182_2016    KCTTVSINDVDTGVPSI...MALKGTLLLSTLWFKPPFLSD...YTMCEYPNTICNPNLGNQRVELWHWDTGVVSCL
LC494150_2017    KCTTVSINDVDTGVPSI...MALKGTLLLSTLWFKPPFLSD...YTMCEYPNTICNPNLGNQRVELWHWDTGVVSCL
LC494151_2017    KCTTVSINDVDTGVPSI...MALKGTLLLSTLWFKPPFLSD...YTMCEYPNTICNPNLGNQRVELWHWDTGVVSCL
LC494152_2017    KCTTVSINDVDTGVPSI...MALKGTLLLSTLWFKPPFLSD...YTMCEYPNTICNPNLGNQRVELWHWDTGVVSCL
LC494153_2017    KCTTVSINDVDTGVPSI...MALKGTLLLSTLWFKPPFLSD...YTMCEYPNTICNPNLGNQRVELWHWDTGVVSCL
MG518518_2017    KCTTVSINDVDTGVPTI...MALKGTLLLSTLWFKPPFLSD...YTMCEYPNTICNPNLGNQRVELWHWDTGVVSCL
MH043952_2017    KCTTVSINDVDTGVPSI...MALKGTLLLSTLWFKPPFLSD...YTMCEYPNTICNSNLGNQRVELWHWDTGVVSCL
MH043953_2017    KCTTVSINDVDTGAPSI...MALKGTLLLSTLWFKPPFLSD...YTMCEYPHTICHPNLGNKRVELWHWDTGVVSCL
MH043954_2017    KCTTVSINDVDTGVPSI...MALKGTLLLSTLWFKPPFLSD...YTMCEYPNTICNSNLGNQRVELWHWDTGVVSCL
LC494159_2017    KCTTVSINDVDTGVPSI...MALKGTLLLSTLWFKPPFLSD...YTMCEYPNTICNPNLGNQRVELWHWDTGVVSCL
LC494183_2017    KCTTVSINDVDTGVPSI...MALKGTLLLSTLWFKPPFLSD...YTMCEYPNTICNPNLGNQRVELWHWDTGVVSCL
LC494184_2017    KCTTVSINDVDTGVPSI...MALKGTLLLSTLWFKPPFLSD...YTMCEYPNTICNPNLGNQRVELWHWDTGVVSCL
LC494185_2017    KCTTVSINDVDTGVPSI...MALKGTLLLSTLWFKPPFLSD...YTMCEYPNTICNPNLGNQRVELWHWDTGVVSCL
LC494186_2017    KCTTVSINDVDTGVPSI...MALKGTLLLSTLWFKPPFLSD...YTMCEYPNTICNPNLGNQRVELWHWDTGVVSCL
LC494187_2017    KCTTVSINDVDTGVPSI...MALKGTLLLSTLWFKPPFLSD...YTMCEYPNTICNPNLGNQRVELWHWDTGVVSCL
LC494188_2017    KCTTVSINDVDTGVPSI...MALKGTLLLSTLWFKPPFLSD...YTMCEYPNTICNPNLGNQRVELWHWDTGVVSCL
LC494189_2017    KCTTVSINDVDTGVPSI...MALKGTLLLSTLWFKPPFLSD...YTMCEYPNTICNPNLGNQRVELWHWDTGVVSCL
LC494190_2017    KCTTVSINDVDTGVPSI...MALKGTLLLSTLWFKPPFLSD...YTMCEYPNTICNPNLGNKRVELWHWDTGVVSCL
LC494191_2017    KCTTVSINDVDTGVPSI...MALKGTLLLSTLWFKPPFLSD...YTMCEYPNTICNPNLGNQRVELWHWDTGVVSCL
LC494192_2017    KCTTVSINDVDTGVPSI...MALKGTLLLSTLWFKPPFLSD...YTMCEYPNTICNPNLGNRRVELWHWDTGVVSCL
                 ****:***:.***.* *   ********************   *******:***:..***:*:**************
```

Supplementary Figure 5. Atomic details of the interaction between the RDB domain of HCoV-229E spike protein (class I in black, I-II in light green, IV in light blue, V in salmon and VI in yellow) and hANPEP (in violet) at the binding interface. Salt bridges and H-bonds are represented by dashed blue lines, whereas hydrophobic contacts are indicated as dashed yellow lines. N atoms are colored in blue, O atoms in red. Positively selected sites are underlined. RBDs-hANPEP conserved interaction pattern, mainly involving the 314-320 segment of RBD loop 1 and the 287-292 portion of hANPEP domain II, is shown in panel A. Side chains not involved in the reported interactions have been omitted for clarity, as well as the C317-C320 disulfide bond. Other relevant structural differences among the various RBDs and details on their interactions at the binding interface, involving loop1, loop2 and loop3, are represented in panels B, C and D, respectively. In particular, in loop1 the motif GGG (313-315) of class I is converted in GVG in classes I-II and IV and GPG in classes V and VI. The increased steric hinderance should favor dispersive interactions with the receptor. Additionally, the presence of proline in place of glycine/valine imposes a slight distortion of loop1 that should strengthen the interactions with hANPEP. In general, the different conformation of the entire loop 1 as found in classes I-II, IV, V and VI with respect to class I, strongly alters the orientation of the 316 residue (lysine or arginine) side chain, allowing the formation of a salt bridge with hANPEP D288 that is not observed in class I RBD. For what concerns loop2, the change of its conformation passing from class I RBD to the other variants stabilizes the salt bridge formed between D315 of hANPEP and site 359 (which is arginine for all classes). As for loop3, in class I RBD S407 forms a long-range polar interaction with K292 of hANPEP while K408 interacts with the carbonyl backbone of E291. Both interactions are lost in classes I-II and IV RBDs. Here, however, there are two leucine residues at sites 404 and 408, which could play a role in forming and/or optimizing hydrophobic interactions (with L318, for instance). Strong polar interactions are restored in classes V and VI, where we find a lysine at 408 position, intercepting E291 backbone in hANPEP, and a histidine at site 407, which could form an additional polar interaction with the spatially close D315 of hANPEP. All these observations indicate that specific mutations (especially in loop 3) and changes in the three loops conformation should progressively favor the formation of the hRBD and hANPEP complex going from class I to VI.



Supplementary Figure 6. Epitopes mapped on the spike protein of different HCoV-OC43 strains, representing different time-frames (from left to right: reference strain, 1967-1990, 1991-2000, 2002-2010, 2011-2019). The proteins are in different shades of gray, whereas the epitopes are in red. The 9-O-acetylated sialic is shown with green spheres.



Supplementary Figure 7. Regression plot of the root-to-tip distance against sampling dates for the RBD domain of the HCoV-229E spike protein. Each point corresponds to a viral sequence and the line is the linear regression calculate using a method that minimizes the residual mean squares. The r coefficient and the *p* value calculated on 1000 permutations are shown. A phylogenetic tree with the generated clusters based on tip distributions is also reported.