



Title	Identification and characterization of a semi-dominant restorer-of-fertility 1 allele in sugar beet (<i>Beta vulgaris</i>)
Author(s)	Arakawa, Takumi; Ue, Sachiyo; Sano, Chihiro; Matsunaga, Muneyuki; Kagami, Hiroyo; Yoshida, Yu; Kuroda, Yosuke; Taguchi, Kazunori; Kitazaki, Kazuyoshi; Kubo, Tomohiko
Citation	Theoretical and Applied Genetics, 132(1), 227-240 https://doi.org/10.1007/s00122-018-3211-6
Issue Date	2019-01
Doc URL	http://hdl.handle.net/2115/76457
Rights	The final publication is available at link.springer.com . This is a post-peer-review, pre-copyedit version of an article published in Theoretical and Applied Genetics. The final authenticated version is available online at: http://dx.doi.org/10.1007/s00122-018-3211-6 .
Type	article (author version)
File Information	text Fig1-7 Supple. Fig1-11 Table1-4 .pdf



[Instructions for use](#)

1 **Title:**

2 Identification and characterization of a semi-dominant *restorer-of-fertility 1* allele in sugar beet (*Beta vulgaris*)

3

4 **Authors:**

5 Takumi Arakawa^{1,*}, Sachiyo Ue^{1,*}, Chihiro Sano¹, Muneyuki Matsunaga¹, Hiroyo Kagami¹, Yu Yoshida¹, Yosuke

6 Kuroda², Kazunori Taguchi², Kazuyoshi Kitazaki¹, Tomohiko Kubo¹

7

8 **Affiliations:**

9 ¹Research Faculty of Agriculture, Hokkaido University, N-9, W-9, Kita-ku, Sapporo 060-8589, Japan

10 ²Hokkaido Agricultural Research Center, National Agriculture and Food Research Organization, Shinsei Minami 9-4,

11 Memuro, 082-0081, Japan

12

13 **Correspondence to:**

14 Tomohiko Kubo

15 e-mail: tomohiko@abs.agr.hokudai.ac.jp

16 phone: +81-11-706-2484

17 ORCID: 0000-0001-6045-0167

18

19 *TA and SU equally contributed to this study.

20 **Abstract**

21

22 Male sterility (MS) caused by nuclear-mitochondrial interaction is called cytoplasmic male sterility (CMS) in which
23 MS-inducing mitochondria are suppressed by a nuclear gene, *restorer-of-fertility*. *Rf* and *rf* are the suppressing and
24 non-suppressing alleles, respectively. This dichotomic view, however, seems somewhat unsatisfactory to explain the
25 recently discovered molecular diversity of *Rf* loci. In the present study, we first identified sugar beet line NK-305 as a
26 new source of *Rf1*. Our crossing experiment revealed that NK-305 *Rf1* is likely a semi-dominant allele that restores
27 partial fertility when heterozygous but full fertility when homozygous, whereas *Rf1* from another sugar beet line
28 appeared to be a dominant allele. Proper degeneration of anther tapetum is a prerequisite for pollen development; thus,
29 we compared tapetal degeneration in the NK-305 *Rf1* heterozygote and the homozygote. Degeneration occurred in both
30 genotypes but to a lesser extent in the heterozygote, suggesting an association between NK-305 *Rf1* dose and
31 incompleteness of tapetal degeneration leading to partial fertility. Our protein analyses revealed a quantitative
32 correlation between NK-305 *Rf1* dose and a reduction in the accumulation of a 250 kDa mitochondrial protein complex
33 consisting of a CMS-specific mitochondrial protein encoded by MS-inducing mitochondria. The abundance of *Rf1*
34 transcripts correlated with NK-305 *Rf1* dose. The molecular organization of NK-305 *Rf1* suggested that this allele
35 evolved through intergenic recombination. We propose that the sugar beet *Rf1* locus has a series of multiple alleles that
36 differ in their ability to restore fertility and are reflective of the complexity of *Rf* evolution.

37 **Key words**

38

39 nuclear-cytoplasmic interaction, *restorer of fertility*, mitochondria, gene dose, cytoplasmic male sterility, anther
40 development.

41

42 **Author contribution statement**

43

44 TA and TK designed this study; YK, KT and TK developed plant materials; TA, SU, CS, MM, HK, YY, KK and TK
45 performed the experiments and analyzed the data; TA and TK wrote the manuscript.

46

47 **Key message**

48 The sugar beet *Rfl* locus has a number of molecular variants. We found that one of the molecular variants is a weak
49 allele of a previously identified allele.

50 **Acknowledgements**

51 This work was supported in part by JSPS KAKENHI Grant Number 18K05564 (TK) and NARO Bio-oriented
52 Technology Research Advancement Institution (BRAIN) (Research program on development of innovative technology,
53 Grant Number 30001A) (KT, KK and TK). TA is a recipient of a JSPS Research Fellowship for Young Scientists
54 (16J01146).

55

56 **Compliance with ethical standards**

57 **Conflicts of interest**

58 The authors declare that they have no conflict of interest.

59

60 This article does not contain any studies with human participants or animals performed by any of the authors.

61

62

63

64 **Introduction**

65

66 Mitochondria are not only the energy sources of eukaryotic cells but also key players in a number of other biological
67 functions (van der Blik et al. 2017). Although mitochondria have their own genomes, they are not autonomous due to
68 their reliance on the nuclear genome for such processes as DNA replication, transcription and translation (Ryan and
69 Hoogenraad 2007). Additionally, many mitochondrial gene products must interact with nuclear gene products for
70 proper function; hence, mismatches between mitochondrial and nuclear genes can cause some abnormalities in
71 development and reproduction (Lane 2011).

72 In plants, one of the best-known examples of reproductive abnormality caused by mitochondrial-nuclear mismatch
73 is cytoplasmic male sterility (CMS) (Mackenzie 2005). CMS can be explained genetically by a male-sterility inducing
74 cytoplasm (S) and a suppressor gene (*restorer-of-fertility*, *Rf*) (Budar and Pelletier 2001). The former is usually
75 characterized at the molecular level as having a unique open reading frame in its mitochondrial genome (hereafter *S-orf*)
76 and has been found in more than twenty CMS plants. In general, *S-orf* genes appear to be organized as chimeras
77 composed of fragments of known and/or unknown origin (Chen and Liu 2014), but several exceptions in which a
78 truncated mitochondrial gene or a non-coding RNA are identified as causes of CMS are also known (Ducos et al. 2001;
79 Stone et al. 2017). The primary sequences of *S-orfs* in different plants are generally less conserved (Hanson and
80 Bentolila 2004), suggesting their independent origin, i.e. they have emerged independently in plant lineages during
81 angiosperm evolution. The mechanism for how *S-orfs* cause CMS is largely unknown with a few exceptions (Luo et al.
82 2013), but presumably the *S-orfs* can exert a deleterious effect on anther tapetal tissue (the innermost cell layer of anther
83 locules) during and/or after microspore meiosis since morphological abnormalities in this tissue at these specific
84 developmental stages is evident in many CMS plants (Schnable and Wise 1998).

85 *Rfs* have been shown to encode several kinds of proteins, although the most prominent class includes those
86 encoding pentatricopeptide-repeat (PPR) proteins (Kim and Zhang 2018) (hereafter PPR-*Rfs*). PPR-*Rfs* are often
87 arrayed with homologous genes to form a gene cluster that exhibits organizational diversity in terms of the primary

88 sequence and the number of clustered genes (Kato et al. 2007; Geddy and Brown 2007; Mora et al. 2010; Melonek et al.
89 2016). This organizational diversity is thought to reflect the evolution of the gene cluster to produce novel *Rf* proteins
90 that can cope with newly emerging *S-orfs* (Fujii et al. 2011). In fact, *Rfs* for different CMSs often map to the same
91 chromosomal region (Li et al. 1998; Tang et al. 2017). On the other hand, variants of *Rf*-like gene clusters may have
92 another characteristic. Genetic analyses have suggested the presence of multiple *Rf* alleles for a single CMS in some
93 plants (Wise et al. 1996; Lee et al. 2008; Zhang et al. 2018), although evidence for diallelism of this locus is prevalent
94 in molecular genetic studies. In maize, for example, *Rf1^m* that is presumably a semi-dominant *Rf1* allele for T-type CMS
95 is known in addition to the dominant *Rf1* (Duvick 1965). The relationship between the multiple allelism of *Rf* and the
96 molecular diversity of the *Rf* locus is not fully described but is an interesting issue in terms of *Rf* evolution.

97 A semi-dominant *Rf* can restore partial fertility and full fertility when it is heterozygous and homozygous,
98 respectively. In this genetic model, the premise is that a semi-dominant *Rf* insufficiently offsets the effect of S
99 mitochondria. One may expect that the difference between fully- and partially fertile plants (and between partially
100 fertile and male sterile plants) would be smaller than the difference between fully fertile and male sterile plants. This
101 situation provides an interesting opportunity for investigating the detailed mechanism of CMS.

102 CMS is widely used for commercial hybrid seed production because CMS plants cannot self-pollinate; hence, the
103 seeds on CMS plants are absolute hybrids (Chase 2007). Fertility, on the other hand, must be restored to the hybrids to
104 ensure self-pollination if the harvest is seed or fruit. Partial fertility is practically problematic because the limited
105 amount of pollen produced is inadequate for full pollination, whereas partially fertile plants should not be used as seed
106 parents because the presence of some selfed seeds will decrease the purity of hybrid seeds. Studies on the restoration of
107 partial fertility may have practical value.

108 Sugar beet CMS was first discovered by Owen (Owen 1945) and is thus called Owen-type CMS. Two genetically
109 independent *Rfs* are known for Owen CMS (hereafter termed as *Rf1* and *Rf2*), and each has the ability to restore male
110 fertility without the other in a sporophytic manner (Pillen et al. 1993; Honma et al. 2014). Although the gene-product of
111 *Rf2* is still unknown, molecular cloning of the *Rf1* locus revealed it to encode a mitochondrial membrane protein

112 resembling OMA1 whose known functions include quality control of mitochondrial membrane proteins by degrading
113 misfolded proteins in budding yeast, and regulation of mitochondrial dynamics via processing OPA1, a central
114 component of the mitochondrial fusion machinery in mammals (Matsuhira et al. 2012 and references therein). It is
115 unlikely that the product of the *Oma1*-like gene at the *Rfl* locus has peptidase activity because its zinc-binding motif is
116 identical to a mutated form that lacks proteolytic activity (Matsuhira et al. 2012). From a restoring *Rfl* allele, a gene
117 cluster consisting of four copies of the *Oma1*-like gene was found, and one of them (termed *orf20*) exhibited the
118 strongest ability to restore male fertility to a CMS plant when expressed as a transgene (Matsuhira et al. 2012).
119 Translation products of *orf20* were shown to bind a CMS-specific protein encoded by *preSatp6*, the best candidate for
120 *S-orf* for Owen CMS (Kitazaki et al. 2015). The preSATP6 protein usually constitutes a 250 kDa complex (likely a
121 homo oligomer), and the amount of the 250-kDa complex in the anther is highly reduced in *Rfl* plants (Kitazaki et al.
122 2015). This reduction is concomitant with the expression of *orf20* and with the appearance of a novel 200 kDa complex,
123 whereas the amount of monomeric preSATP6 is apparently unchanged. From plants homozygous for a nonrestoring *rfl*,
124 a single copy of another *orf20*-like gene (*orf20L*) was found (Matsuhira et al. 2012). No binding of ectopically
125 expressed ORF20L protein with preSATP6 in transgenic suspension cells and no decrease in the amount of the 250-kDa
126 complex in *rflrfl* anthers was detected (Kitazaki et al. 2015). Therefore, *Rfl* likely encodes a type of molecular
127 chaperone that changes the higher order structure of preSATP6, whereas the gene product of the nonrestoring *rfl*
128 apparently lacks such activity. A hypothetical mechanism for fertility restoration involves protein-protein interaction
129 between preSATP6 and ORF20 to generate the 200 kDa complex, thereby reducing the amount of the 250 kDa complex
130 (i.e. the 250 kDa complex was converted into the 200 kDa complex via preSATP6-ORF20 interaction).

131 We previously reported the organizational diversity of the *Rfl* locus among sugar beet lines including copy number
132 variation of *orf20*-like genes (Moritani et al. 2013; Ohgami et al. 2016; Arakawa et al. 2018), which is reminiscent of
133 PPR-*Rf*. This observation prompted us to investigate further the possibility of functional diversity among restoring *Rfl*
134 alleles. In the present study, we report a novel restoring *Rfl* allele that confers partial fertility when the plant is
135 heterozygous. We morphologically characterized the partial fertility as unique degeneration pattern in anther tapetum.

136 In the partially fertile plant, the amount of the 250-kDa complex and *Rfl* mRNA was intermediate between fully
137 fertility restored and male sterile plants. Molecular organization of this *Rfl* suggests intergenic recombination was
138 involved in its evolution. Together with our previous data, we propose that sugar beet *Rfl* has evolved multiple alleles
139 that differ in their ability to restore male fertility to Owen CMS plants.

140

141

142 **Materials and methods**

143

144 **Plant materials**

145

146 All the sugar beet lines used in this study were developed by the Hokkaido Agricultural Research Center, National
147 Agriculture and Food Research Organization. NK-305 is an old selection of unknown origin. NK-198 is a fertility
148 restored line (Matsuhira et al. 2012). TA-33BB-CMS is a CMS line and TA-33BB-O is its maintainer inbred line
149 (Moritani et al. 2013). NK-219mm-CMS is a CMS line that can be used for transgenic experiments (Kagami et al.
150 2016). Plants were grown in the field of the Field Science Center for Northern Biosphere, Hokkaido University or in the
151 greenhouse of the Research Faculty of Agriculture, Hokkaido University (maintained at 20-25°C; natural daylight and
152 illuminated by incandescent light [150μmol/m²/sec] at night).

153

154 **Male fertility evaluation**

155

156 Male fertility was evaluated by anther morphology and color and was classified into four distinct classes: fully fertile,
157 semi-fertile (type a), semi-fertile (type b), and completely sterile as described by Arakawa et al. (2018). The four classes
158 were indexed from 0 (complete sterile) to 3 (fully fertile), respectively. The male fertility index of a plant is an average
159 of observations taken on more than four different days. In some genetic analyses, fully fertile and semi-fertile (type a)

160 types were combined into a single class (Fig. 1a-c). Anther contents were visualized by squashing the anther on a glass
161 slide in a drop of Alexander's dye (Alexander 1969).

162

163 **Cytological analysis**

164

165 Flower buds were collected and immersed into FAA solution [50% (v/v) ethanol, 3.7% (v/v) formaldehyde, 5% (v/v)
166 acetic acid]. Specimens were dehydrated in a series of isobutyl alcohol and embedded in Paraplast Plus (Sigma-Aldrich
167 Japan, Tokyo, Japan). Paraplast blocks were sectioned (10 µm thickness) using a rotary microtome HM360 (Carl Zeiss,
168 Oberkochen, Germany). Sections were stained with Toluidine blue (Chroma Gesellschaft Schmid, Stuttgart, Germany).
169 Images were collected using a light microscope BX50 (Olympus, Tokyo, Japan) equipped with a CCD camera DP21
170 (Olympus).

171

172 **DNA analysis**

173

174 Total cellular DNA from green leaves was extracted by the CTAB-method described by Doyle and Doyle (1990). DNA
175 samples were further purified by cesium chloride-ethidium bromide equilibrium centrifugation if necessary. A
176 mitochondrial DNA marker targets the TR-1 locus that shows polymorphism in the number of 32-bp repeated
177 sequences among mitotypes (Nishizawa et al. 2000). The presence/absence of a unique *Hind*III restriction site in the
178 plastid *petG-psbE* intergenic region is characteristic of male sterility-inducing cytoplasm (Ran and Michaelis 1995).
179 DNA markers for cytoplasm discrimination were detailed in Cheng et al. (2009). DNA marker s17 was detailed in
180 Taguchi et al. (2014). DNA gel blot analysis was done according to the method described in Sambrook et al. (1989).
181 DNA fragments were blotted onto Hybond-N+ (GE Healthcare, Little Chalfont, UK) according to the manufacturer's
182 instruction manual. Hybridization probe was prepared as detailed in Ohgami et al. (2016) and labeled with the DIG
183 Luminescent Detection Kit (Roche Diagnostics, Mannheim, Germany). Hybridization signals were detected on X-ray
184 film according to the manufacturer's protocol. PCR products were subjected to sequence analysis either directly or after
185 being cloned into the pBluescript (SK+) vector. Nucleotide sequences were determined using an ABI3130 Genetic

186 Analyzer (Thermo Fisher Scientific, Waltham, MA, USA). DNA sequences were analyzed using GENETYX (ver. 16)
187 (GENETYX CORPORATION, Tokyo, Japan) and Sequencher (ver. 4.5) (Hitachi Software Engineering, Tokyo, Japan).
188

189 **Protein analysis**

190

191 Mitochondria were isolated from transgenic calli according to the method of Kitazaki et al. (2015). Samples of anther
192 and crude mitochondria for blue native polyacrylamide gel electrophoresis (BN-PAGE) were prepared following the
193 method of Kitazaki et al. (2015). BN-PAGE was performed using the Native PAGE Novex BisTris Gel system (Thermo
194 Fisher Scientific) according to the manufacturer's instructions. Sample preparation and electrophoresis for SDS-PAGE
195 were performed by the method described by Yamamoto et al. (2005). Procedures for immunoblot analysis after BN- and
196 SDS-PAGE were described in Kitazaki et al. (2015). Anti-FLAG antibody (α FLAG) was purchased from Medical and
197 Biological Laboratories (Nagoya, Japan) and was diluted to 50 ng/mL for immunoblot analysis. Anti-preSATP6
198 (α preSATP6) (Yamamoto et al. 2005) was diluted to 42.5 ng/mL and 0.34 μ g/mL for blots of SDS-PAGE and
199 BN-PAGE, respectively, except for quantification of the preSATP6 protein complex in which α preSATP6 was diluted
200 to 68 ng/mL. Anti-COXI (α COXI) (Yamamoto et al. 2005) was diluted to 42.5 ng/mL. Signal intensity was quantified
201 by using Image J (<http://rsbweb.nih.gov/ij/download.html>). Conditions for the quantification were verified by
202 constructing a calibration curve.

203

204 **Real-time reverse transcription PCR analysis**

205

206 Anthers were sorted according to their developmental stages as described in Kitazaki et al. (2015). Total cellular RNA
207 from anthers was extracted with an RNeasy Plant Mini Kit (Qiagen, Valencia, CA, USA) and treated with RNase-free
208 DNase I (Takara Bio, Kusatsu, Japan). RNA samples (300 ng) were reverse-transcribed with the SuperScript III
209 First-Strand Synthesis System (Thermo Fisher Scientific) using an oligo dT primer. The resultant cDNA was mixed

210 with primers (0.2 μ M each) and PowerUp SYBR Green Master Mix (Life Technologies, Carlsbad, CA, USA). Two
211 constitutively expressed genes (*elongation factor 1 α* and *Actin*) were chosen for the internal controls (Arakawa et al.,
212 2018). Nucleotide sequences of the primers are shown in Supplementary Table 1. PCR was monitored with a Chromo 4,
213 Opticon Monitor (ver. 3.1) with a DNA-Engine PTC-200 (Bio-Rad Laboratories, Hercules, CA, USA). The PCR
214 protocol was 50 °C for 2 min, 95 °C for 2 min, and then 40 cycles of 95 °C for 15 s and 60.8 °C for 1 min. After the
215 quantification, all the reactants were heated to 95 °C (1 min) and then cooled to 50 °C. A melting-curve was then drawn
216 (53 to 87 °C, data acquisition every 0.5 °C) to verify that there was a single amplicon. The delta-Ct method was used for
217 quantification of Ct values according to the method described in Kitazaki et al. (2011).

218

219 **Generation of transgenic suspension cells**

220 Binary vectors for the transgenic experiment were constructed according to the method of Arakawa et al. (2018).
221 Briefly, for each *orf20*-like gene, the coding sequence and 5'- and 3'- untranslated regions were amplified from NK-305
222 total DNA using two pairs of primers as shown in Supplementary Table 1. The amplicons were cloned into pDONRzeo
223 (Thermo Fisher Scientific) using the BP Clonase Enzyme mix (Thermo Fisher Scientific) according to the
224 manufacturer's instructions. A FLAG tag was inserted into 3'-end of the coding region by PCR-based mutagenesis
225 described in Arakawa et al. (2018). After confirmation by nucleotide sequencing of the resultant vectors, FLAG-fused
226 genes were transferred to the binary vector pMDC Ω (Kitazaki et al. 2015) using the LR Clonase Enzyme mix (Thermo
227 Fisher Scientific). The resultant binary vectors were introduced into sugar beet suspension cells via *Agrobacterium*. The
228 procedure for transformation was as described in Kagami et al. (2015). Suspension cells were derived from callus
229 cultures of NK-219mm-CMS.

230

231 **Statistical tests**

232

233 Fisher's exact test, the χ^2 test, the Steel-Dwass test, the Tukey-Kramer test, and Welch's *t* test were conducted using the
234 website <http://aoki2.si.gunma-u.ac.jp/exact/exact.html>; accessed on Jun 27, 2018 or by using Microsoft Excel
235 (Microsoft Japan, Tokyo, Japan).

236

237 **Accession number**

238

239 Nucleotide sequence data were deposited in DDBJ/ EMBL/GenBank under the accession number LC385768.

240

241

242 **Results**

243

244 **Identification of a semi-dominant *Restorer-of-fertility 1* allele**

245

246 We observed that NK-305 flowers shed pollen grains when grown in field conditions and examined the cytoplasm type
247 of seventy-four NK-305 plants by mitochondrial and plastid DNA marker analysis. In the analysis of the mitochondrial
248 TR-1 locus, all seventy-four plants had an identical pattern to that of plants with male sterility-inducing cytoplasm
249 (Supplementary Fig. 1a). We also confirmed that their cytoplasms were of the male-sterility inducing type as evidenced
250 by the plastid marker (Supplementary Fig. 1b). The male fertile phenotype, despite having a male-sterility inducing
251 cytoplasm, indicated that NK-305 is a fertility restored line.

252 To examine whether *Rf1* or *Rf2* (or both) are involved in fertility restoration, we selected one NK-305 plant and
253 crossed it with TA-33BB-CMS, a CMS line. The resultant F₁ plants restored male fertility, but they were only partially
254 fertile at best (see Fig. 1a-c). One of the F₁ plants was self-pollinated to obtain an F₂ population of 26 plants. Pollen
255 fertility in the progeny segregated as shown in Table 1. We tested the genetic association between the phenotype and
256 two nuclear DNA markers linked to *Rf1* and *Rf2*.

257 We used DNA marker o7, a DNA fragment length polymorphic marker linked to *Rf2* (Arakawa et al. 2018) to
258 evaluate the genetic association of *Rf2* with fertility restoration. TA-33BB-CMS is homozygous for a 2.6-kbp fragment
259 (T allele), whereas the parental line NK-305 yields a 1.4 kbp fragment (K allele) (Supplementary Fig. 2a). The F₁ plant
260 had 2.6- and 1.4-kbp fragments, indicating that it was heterozygous. Plants in the F₂ population generated either 2.6, 1.4,
261 or 2.6 and 1.4 kbp fragments, indicating that they were T homozygous, K homozygous, and heterozygous, respectively
262 (Supplementary Table 2). Association between the genotype and the phenotype was statistically tested, and the null
263 hypothesis of their independence was not rejected ($p=0.150$ for fertility restoration [fully fertile+partial fertile] vs.
264 genotype, and $p=0.139$ for three classes of male fertility vs. genotype; Fisher's exact test).

265 We next genotyped the F₂ plants using the s17 marker, which is a cleaved amplified length polymorphic marker
266 linked to *Rf1* (Taguchi et al. 2014). The size and number of DNA fragments on the agarose gel of s17 were as follows:
267 TA-33BB-CMS produces 1.1 and 0.6 kbp fragments and is referred to as pattern 4 (hereafter denoted as p4), whereas
268 the parental NK-305 plant produces 1.4 and 0.5 kbp fragments, which is referred to as pattern 2 (p2) (Fig. 2a). The F₁
269 plant produced 1.4, 1.1, 0.6 and 0.5 kbp fragments, indicating that it was heterozygous (p2p4). In the F₂ population,
270 plants exhibited one of three banding patterns: 1.4 and 0.5 kbp, 1.1 and 0.6 kbp, or 1.4, 1.1, 0.6 and 0.5 kbp that
271 correspond to genotypes of p2p2, p4p4, and p2p4, respectively (Table 1). The association of fertility restoration and the
272 genotype was obvious ($p=8.69 \times 10^{-6}$; Fisher's exact test). Therefore, it seemed likely that *Rf1* was involved in the
273 observed fertility restoration.

274 We noticed that the occurrence of fully fertile plants was more frequent in the p2p2 genotype (five out of six
275 plants) compared to the p2p4 genotype (one out of 14 plants) in the F₂ population and hypothesized that NK-305 *Rf1* is
276 a weak allele but its effect is cumulative, i.e. a semi-dominant allele. We next made two additional populations, one
277 segregating NK-305 *Rf1* and the other segregating an *Rf1* from a different source and compared the patterns of their
278 fertility restoration.

279 In attempting to produce a new population segregating NK-305 *Rf1*, our concern was the small number of seeds
280 from a selfed F₁ plant of TA-33BB-CMS x NK-305 due to its low pollen fertility. Therefore, one of the F₁ plants was

281 paper-bag enclosed with TA-33BB-O, a normally fertile line with normal (N) cytoplasm but the same nuclear genotype
 282 as TA-33BB-CMS (i.e. [N]*rf1rf1rf2rf2*). As such, the obtained seeds from this F₁ plant were, in effect, an admixture of
 283 a three-way cross [i.e. (TA-33BB-CMS x NK-305) x TA-33BB-O] and the F₂ (self-pollinated TA-33BB-CMS x
 284 NK-305). The NK-305 *Rf1* was expected to segregate because both TA-33BB-O and TA-33BB-CMS are the
 285 *rf1rf1rf2rf2* genotype. In the admixture population, male fertility, and o7 and s17 genotypes segregated as shown in
 286 Supplementary Table 3. The null hypothesis for independence of fertility restoration in the o7 genotype was not rejected
 287 ($p=1$ for fertility restoration vs. genotype, and $p=0.985$ for three classes of male fertility vs. genotype; Fisher's exact
 288 test). On the other hand, that for independence of fertility restoration in the s17 genotype was rejected ($p=4.08 \times 10^{-13}$
 289 for fertility restoration vs. genotype, and $p=1.92 \times 10^{-15}$ for three classes of male fertility vs. genotype; Fisher's exact
 290 test). Pollen fertility of each plant was indexed, and its segregation is depicted in Fig. 3.

291 In our previous study, sugar beet line NK-198 was used for the molecular cloning of the *Rf1* locus (Matsuhira et al.
 292 2012). In this study, an NK-198 plant was crossed with TA-33BB-CMS, resulting in two fully fertile F₁ plants. Using
 293 fertility-restored plants as pollen parents and TA-33BB-CMS as seed parents, we repeated the back-cross to obtain BC₂.
 294 Of the four BC₂ plants, two were fully fertile, whereas the remainder were sterile. The s17 of NK-198 produces pattern
 295 1 (p1) (Fig. 2b). As expected, the two fully fertile BC₂ plants were p1p4 and the two male-sterile plants were p4p4. One
 296 of the fully fertile plants was selfed to generate a BC₂F₂ population. We examined the segregation of s17 and found that
 297 the numbers of p1p1, p1p4, and p4p4 plants were 21, 61, and 18, respectively. This result did not deviate statistically
 298 from the expected segregation (χ^2 test for 1:2:1; $p=0.081$). As eighty-three plants flowered, their male fertility was
 299 investigated. In the BC₂F₂ population, male fertility and the s17 genotypes segregated as shown in Supplementary Table
 300 4. The null hypothesis for independence of fertility restoration in the s17 genotype was rejected ($p=3.58 \times 10^{-9}$ for
 301 fertility restoration vs. genotype, and $p=5.95 \times 10^{-11}$ for three classes of male fertility vs. genotype; Fisher's exact test).
 302 Pollen fertility segregation is depicted in Fig. 3.

303 Comparing the averages of male fertility indices, groups p1p1, p1p4, and p2p2 are statistically indistinguishable
 304 from each other (class a in Fig. 3). Two groups of p4p4 (one from the admixture population and the other from the

305 BC₂F₂ population) formed a separate group (class b in Fig. 3). The group of p2p4 plants are not only statistically
306 different from those of p1p1, p1p4, and p2p2, but also from p4p4 (class c in Fig. 3).

307

308 **Tapetum degeneration in the anthers of semi-fertile plants**

309

310 In an effort to characterize the partially fertile phenotype of p2p4, we examined anthers and pollen grains
311 microscopically. Most of the pollen grains in the anthers of p2p4 plants were not stained with Alexander's dye and were
312 rather small compared to fully stainable pollen grains (compare Figs. 1e and 1g), but exine-like patterns on the pollen
313 grain surfaces were visible. Fully fertile p2p2 plants shed pollen grains that were comparable to those shed by
314 TA-33BB-O (Fig. 1d). In male-sterile p4p4 plants, the anthers were shriveled and contained only some undeveloped
315 pollen grains as was also seen in TA-33BB-CMS (Figs. 1f and 1h).

316 The internal structure of anthers was investigated histologically by collecting a developmental series of immature
317 flower buds from each of the p2p2, p2p4, p4p4 and TA-33BB-O plants. Microspore development in TA-33BB-O
318 appeared to follow the pattern previously reported in sugar beet (e. g. Majewska-Sawka et al. 1993). We defined six
319 stages as follows: the meiosis stage in which microspore mother cells enter meiosis, corresponding to Stage 6 in
320 *Arabidopsis* anther development (Sanders et al. 1999); the tetrad stage in which meiosis is completed, corresponding to
321 *Arabidopsis* Stage 7; microspore stage Sa in which individual microspores are released from tetrads, corresponding to
322 *Arabidopsis* Stage 8; microspore stage Sb-1 in which the walls of microspores are thickened and the tapetum becomes
323 thinner and less stained with Toluidine blue dye, corresponding to *Arabidopsis* Stage 9, although the onset of tapetal
324 degeneration in sugar beet appeared to be earlier; microspore stage Sb-2 in which the endothecium develops radial bars
325 of thickenings and the tapetum is invisible; and the pollen stage in which the septum degenerates and the stomium
326 breaks resulting in anther dehiscence (Fig. 4a-f). Anther development of p2p2 plants was indistinguishable from that of
327 TA-33BB-O from the meiosis stage to the pollen stage (Fig. 4g-l).

328 The morphology of p2p4 anthers was similar to that of p2p2 plants until the microspore stage Sa (Fig. 4m-o).
329 However, during the microspore Sb-1 stage, the tapetum of p2p4 plants was thicker and more intensely stained than that
330 of p2p2 plants (Fig. 4p). During the microspore Sb-2 stage, tapetal residue was still visible, although radial bars of
331 thickenings were seen in the endothecium (Fig. 4q). At the pollen stage, stomium breakage was rarely observed in p2p4
332 plants, although the septum connecting two anther locules degenerated (Fig. 4r). Pollen grains in p2p4 locules were less
333 stained compared to those in p2p2 anthers, suggesting that they were less developed.

334 Sections of p4p4 anthers were similar to those of the other genotypes until the tetrad stage (Figs. 4s and t); however,
335 in sections containing microspores, tapetal cells were attached together unlike the dissociated tapetal cells found in
336 other genotypes (Fig. 4u). Occasionally, the p4p4 tapetum was irregularly shaped and vacuolated unlike those of the
337 other genotypes (Fig. 4v). In the following stages, p4p4 anther locules were deformed (Fig. 4w). Microspores or
338 tapetum were not identifiable inside of p4p4 locules but remnants of microspore or tapetum that were misshapen by
339 deformed anther walls were seen (Fig. 4w). Radial bars of thickenings of the endothecium were observed in some
340 sections that had empty locules (Fig. 4x).

341

342 **NK-305 *Rfl* is a novel molecular variant**

343

344 We next investigated the molecular organization of NK-305 *Rfl*. Because the sugar beet *Rfl* locus contains a gene
345 cluster of *orf20*-like genes, we first examined the number of clustered genes in the NK-305 locus by DNA gel blot
346 analysis. A DNA probe corresponding to the 3'-untranslated region (UTR) that is conserved among *orf20*-like genes
347 was hybridized with total cellular DNA isolated from one of the p2p2 NK-305 plants. Two *Hind*III restriction bands of
348 7.0 and 5.4 kbp appeared on the blot (ESM Fig. 3), indicating that NK-305 *Rfl* contains two copies of *orf20*-like genes.
349 The nucleotide sequences of the previously characterized *Rfl/rfl* enabled us to PCR amplify the entire NK-305 *Rfl*
350 locus as overlapping segments whose nucleotide sequences were assembled into a contig of 20019 bp. From this

351 nucleotide sequence, two *Hind*III restriction fragments of 7167 and 5277 bp are expected to hybridize with the 3'-UTR
352 probe, a result confirmed by our DNA blot analysis.

353 From the obtained sequence, two *orf20*-like genes were found that were designated as *orf20*_{NK-305-1} and
354 *orf20*_{NK-305-2} (Fig. 5). Comparison with previously characterized *orf20*-like genes revealed two introns in each of the
355 newly identified *orf20*-like genes (Supplementary Figs. S4 and S5) as were also present in the other *orf20*-like genes.
356 The best match for *orf20*_{NK-305-1} was *orf20* in NK-198 (hereafter *orf20*_{NK-198}); the sequences differ in two nucleotide
357 residues in exon 1 (one synonymous and one nonsynonymous), one residue in exon 2 (nonsynonymous), and their
358 introns are identical (ESM Supplementary Fig. 4). Consequently, their translation products differ in two amino acid
359 residues (Supplementary Fig. 6). Sequence homology extends 1.5 kbp in the upstream region and 1.6 kbp in the
360 downstream region (Supplementary Fig. 4). In contrast, *orf20*_{NK-305-2} had the highest similarity to *orf20*_{LS} that had been
361 identified previously as an *orf20*-like gene in a specific group of *rf1rf1* lines (Ohgami et al. 2016); the two ORFs are
362 identical in exon 1, intron 1, and exon 2, but differ in intron 2 (12 nucleotide substitutions) and exon 3 (11 nucleotide
363 substitutions and a small deletion of 3 bp in *orf20*_{NK-305-2}) (Supplementary Fig. 5). Deduced protein products of the two
364 ORFs differ in 11 amino acid residues, all of which occur in the carboxyl-terminal region (Supplementary Fig. 7).
365 Sequence homology extends 5.2 kbp downstream; however, the most closely matched upstream region of *orf20*_{NK-305-2}
366 was not that of *orf20*_{LS} but *orf19*, another *orf20*-like gene accompanying *orf20*_{NK-198} (Supplementary Fig. 8).
367 Interestingly, the upstream region of *orf20*_{LS} showed sequence homology to the upstream region of *orf20*_{NK-305-1}
368 (Supplementary Fig. 9). Altogether, the molecular organization of NK-305 *Rf1* appears to be a chimera of several
369 previously characterized restoring and non-restoring alleles.

370

371 **Lower accumulation of a protein complex containing preSATP6 correlates with NK-305 *Rf1* dose**

372

373 Sugar beet CMS is associated with a 250 kDa protein complex containing preSATP6, and accumulation of the complex
374 is lower in the fertility-restored line NK-198 (Kitazaki et al. 2015). Because NK-305 was identified as another restored

line, we examined whether decreased accumulation of the 250-kDa complex also occurred. Immature anthers (meiosis and tetrad stages) of NK-198, NK-305, and TA-33BB-CMS were collected, and protein extracts were subjected to Blue Native polyacrylamide gel electrophoresis (BN-PAGE) combined with immunoblotting using antiserum against preSATP6 protein expressed in *E. coli* (α preSATP6) (Yamamoto et al. 2005) to determine the size and relative amounts of the 250-kDa complex in each genotype.

TA-33BB-CMS had an intense signal band of 250 kDa on the blot, whereas NK-198 had a smeared image with three faint bands of 250, 200 and 150 kDa (Figs. 6a and 6b). Because the signal bands detected by anti-COXI (α COXI) were very similar for TA-33BB-CMS and NK-198 on this blot (Fig. 6c), degradation of the protein sample was unlikely. NK-305 also gave a smeared signal, but the 250-kDa band was more conspicuous and the 200 and 150 kDa bands were less intense (Figs. 6a and 6b). The latter two bands emerged after longer exposure of the blot, and their signal intensity appeared to be less than that of NK-198. The total amount of preSATP6 protein was apparently different among the TA-33BB-CMS, NK-198 and NK-305 samples on the immunoblot after SDS-PAGE for an unknown reason but there was no correlation with male fertility or the reduced amount of the 250 kDa complex (Fig. 6d).

We next investigated the accumulation of the 250-kDa complex in semi-fertile anthers. Plants characterized as p2p2 (fully fertile), p2p4 (semi-fertile) and p4p4 (complete sterile) were selected from the admixture population and analyzed (Fig. 7). Because our focus was on the 250-kDa complex, antiserum concentration was minimized, hence the 200- and 150-kDa signal bands were barely visible on the blots. We quantified the signal intensity of the 250-kDa band by normalizing the signal intensity of Complex IV that was detected by α COXI. Signal intensities were highest in p4p4, followed by p2p4, and the lowest in p2p2. Compared to p4p4, the signal intensity of p2p4 was reduced by 50% and that of p2p2 was reduced by 92% (Fig. 7).

One of the two *orf20*-like genes in the locus encodes a protein capable of preSATP6 interaction and its transcript abundance correlates with NK-305 *Rf1* dose

399 A correlation between the accumulation of the 250-kDa complex and the s17 genotype (hence gene dosage of NK-305
400 *Rfl*) prompted us to investigate the quantity of mRNA for *orf20*-like genes. There are two classes of *orf20*-like genes,
401 one is preSATP6-interacting and the other is non-interacting (Kitazaki et al. 2015). We first examined whether the two
402 *orf20*-like genes in NK-305 *Rfl* interacted with preSATP6. Each of the two *orf20*-like genes was situated downstream
403 of the cauliflower mosaic virus (CaMV) 35S promoter in constructs that were introduced into suspension cells of CMS
404 sugar beet line NK-219mm-CMS via *Agrobacterium*. Because the two transgenes were tagged with FLAG, their
405 expression was checked by immunoblot analysis using anti-FLAG (α FLAG). Suspension cells expressing *orf20*_{NK-305-1}
406 had a major 43 kDa band and a minor 41 kDa band, whereas those expressing *orf20*_{NK-305-2} had only the 44 kDa band
407 (Supplementary Fig. 10). The reason for this difference is unknown but similar results were obtained when other
408 *orf20*-like genes were expressed as transgenes (Kitazaki et al., 2015; Arakawa et al., 2018). According to Kitazaki et al.
409 (2015) and Arakawa et al. (2018), no correlation was seen between the number of detected signal bands of the
410 ORF20-like protein expressed from the transgene and protein-protein interaction between preSATP6 and the
411 transgene's translation product. Mitochondrial proteins of the transgenic suspension cells were separated by BN-PAGE
412 and subjected to immunoblot analysis using α preSATP6. Proteins from *orf20*_{NK-305-1}-expressing cells had smeared
413 images with a bold 250-kDa band and faint 200- and 150-kDa bands (Supplementary Fig. 11). The 150-kDa band was
414 so faint that it was sometimes visible only with longer exposures. The smeared image containing three bands was also
415 observed from the cells expressing *orf20*_{NK-198} (Supplementary Fig. 11), whose translation product binds to preSATP6
416 protein and likely alters the higher order structure of preSATP6 (Kitazaki et al. 2015). Images obtained from
417 mitochondrial proteins of *orf20*_{NK-305-2}-expressing cells were also smeared but lacked the 200-kDa band. These images
418 were very similar to those expressing non preSATP6-interacting *orf20* homologues such as *orf20L* and the vector
419 control (Kitazaki et al. 2015; Arakawa et al. 2018; Supplementary Fig. 11). Considering that the 200 kDa protein
420 complex contains both preSATP6 and ORF20-like proteins (Kitazaki et al. 2015), *orf20*_{NK-305-1} encodes a protein that
421 can interact with preSATP6 and the protein encoded by *orf20*_{NK-305-2} does not interact with preSATP6 (Supplementary
422 Fig. 11).

423 To quantify the *orf20*_{NK-305-1} mRNA, a primer pair specific to *orf20*_{NK-305-1} was designed to discriminate *orf20*_{NK-305-1}
424 mRNA from *orf20*_{NK-305-2} mRNA. Note that the sequence homology among the four *orf20*-like gene copies in NK-198
425 *Rfl* locus is so high that specific amplification of *orf20*_{NK-198} was infeasible (see Fig. 5 and Matsuhira et al. 2012).
426 Results of our real-time reverse transcription PCR analysis are shown in Table 2. No mRNA was detected from p4p4
427 plants, confirming the specificity of the primers. In p2p2 and p2p4 plants, the *orf20*_{NK-305-1} mRNA was most abundant in
428 the tetrad stage, meiosis stage and microspore stage, in that order. Comparing the two genotypes, fold changes between
429 p2p2 and p2p4 were in the range of 1.5-2.2.

430

431

432 Discussion

433

434 Here, we show the genetic divergence of the restoring *Rfl* allele in sugar beet; although both NK-305 *Rfl* (hereafter
435 *Rfl*_{NK-305}) and NK-198 *Rfl* are restoring alleles, *Rfl*_{NK-305} is a weaker allele in its ability to restore male fertility. A single
436 *Rfl*_{NK-305} allele is insufficient to restore male fertility completely; however, the effect is cumulative, and homozygotes
437 are fully fertile. Hence, *Rfl*_{NK-305} acts as a semi-dominant allele.

438 Genetic interaction between *Rfl* and S mitochondria involves a post-translational mechanism between ORF20-like
439 protein and preSATP6 proteins (Kitazaki et al. 2015). The preSATP6-interacting *orf20*-like gene in *Rfl*_{NK-305} is
440 *orf20*_{NK-305-1}, whose mRNA accumulation level was halved in the heterozygote compared to the homozygote. In
441 accordance with the amount of *orf20*_{NK-305-1} mRNA, the amount of the 250-kDa complex was reduced by 50% and 92%
442 in the heterozygote and the homozygote, respectively. Nevertheless, the amount of preSATP6, the constituent protein of
443 the 250 kDa complex, is almost unchanged irrespective of the genotype. Therefore, the *Rfl*_{NK-305} dose possibly affects a
444 higher order structure of the preSATP6 protein. Comparing anthers of *rflrfl*, *Rfl*_{NK-305}*rfl*, and *Rfl*_{NK-305}*Rfl*_{NK-305}, the
445 amount of the 250 kDa complex was correlated with the level of male sterility. This line of evidence suggests that the
446 250 kDa complex is strongly associated with the expression of CMS.

447 The level of male sterility expression correlated with the extent of tapetal degeneration. Anther tapetum is known
448 to play a very important role in pollen development (Papini et al. 1999). In *Rfl_{NK-305}Rfl_{NK-305}* plants, the tapetum starts
449 to degenerate at the microspore Sb-1 stage and is invisible in the Sb-2 stage, as is observed in N mitochondrial plants. In
450 CMS sugar beet, the tapetum swells with large vacuoles during the microspore stage then collapses with the
451 microspores. In *Rfl_{NK-305}rfl* plants, such tapetal hypertrophy does not occur; however, the first visible difference
452 compared to fully fertile plants is that the tapetum is thicker than in *Rfl_{NK-305}Rfl_{NK-305}* at the microspore Sb-1 stage. In
453 the later Sb-2 stage, the tapetum persists in *Rfl_{NK-305}rfl* plants. This finding suggests either retardation in the onset
454 and/or irregularity in the process of tapetal degeneration. Tapetal degeneration is considered to be the crucial step for
455 pollen development as exemplified by some male sterile mutants in model plants (Rogers 2006; Wilson and Zhang
456 2009). If the tapetum provides nutrients or other substances necessary for microspore development at the expense of its
457 degeneration, there should be a linear correlation between male fertility and the extent of tapetal degeneration.

458 The extent of tapetal degeneration can be controlled by *Rfl_{NK-305}* dose in S mitochondrial plants. Tapetal
459 degeneration is an example of programmed cell death (PCD) in plants (Parish and Li 2010); hence, tapetal PCD should
460 properly occur in an *Rfl_{NK-305}Rfl_{NK-305}* plant, should be impeded in an *Rfl_{NK-305}rfl* and should be abolished in an *rflrfl*.
461 A quantitative factor derived from interaction between *Rfl_{NK-305}* and S mitochondria is assumed in this model. Our
462 results suggest that the quantity of this factor is correlated with the amount of the 250 kDa complex. The 250 kDa
463 complex is likely an oligomer of preSATP6, but its function is unknown (Yamamoto et al. 2005; Kitazaki et al. 2015).
464 Several *S-orfs* whose translation products also form oligomers are known (Rhoads et al. 1988; Duroc et al. 2005). Duroc
465 et al. (2009) proposed that the oligomers could be mild uncouplers. If the 250 kDa complex is also a mild uncoupler, in
466 accordance with its accumulation, the physiological state of mitochondria could be linearly changed, resulting in
467 changes in the production of, for example, reactive oxygen species (ROS). A relationship between ROS and the
468 progression of PCD in anthers has been reported (Hu et al. 2011). Further study is necessary to characterize tapetal PCD
469 in fertility-restored sugar beet.

470 In addition to *Rfl_{NK-305}*, we have recently identified *rfl_{Fukukoku}* from leaf beet, another allele whose effect on the
471 accumulation of the 250 kDa complex is so faint that it was genetically referred to as a non-restoring allele (Arakawa et
472 al. 2018). Another non-restoring allele selected by sugar beet breeders has apparently no effect on the accumulation of
473 the 250 kDa complex (Kitazaki et al. 2015). Thus, it appears that multiple *Rfl/rfl* alleles with different effects on the
474 accumulation of the 250 kDa complex exist in beet. Differences in the molecular effect can be explained by the amount
475 of mRNA that encodes ORF20-like proteins capable of interacting with preSATP6. Given that the amount of
476 ORF20-like protein correlates with its transcript abundance, higher levels of ORF20-like protein convert the 250 kDa
477 complex into the 200 kDa complex because the ORF20-like protein plays a principal role in this conversion process.
478 Consequently, the amount of 250 kDa complex is reduced; however, the possibility cannot be excluded that alterations
479 in the amino acid sequence of the ORF20-like protein may also affect the strength of the allele.

480 Molecular organization of the *Rfl/rfl* alleles indicates that the number of *orf20*-like genes and their primary
481 sequences are diverse (see Fig. 5). At a glance, intergenic recombination likely played a pivotal role in the
482 organizational diversity of *Rfl*. Despite the differences in the gene products, this level of diversity is reminiscent of the
483 evolution of PPR-*Rf* (Kato et al. 2007; Geddy and Brown 2007; Mora et al. 2010; Melonek et al. 2016). At present,
484 however, it remains unknown whether any of the sugar beet *Rfl/rfl* alleles are able to restore CMS lines other than
485 those of the Owen type, as has been suggested for PPR-*Rf* (Fujii et al. 2011). This question should be investigated in the
486 future. Note that beet *Rfs* for two other CMS lines have been mapped onto different chromosomes (Laporte et al. 1998;
487 Touzet et al. 2004).

488 On the other hand, a series of alleles with various strengths in their molecular effect on the 250 kDa complex
489 prompted us to think of another possibility. Perhaps sugar beet *Rfl* is in its infancy toward evolving into the most
490 suitable *Rfl* allele. In relation to this proposal, one population genetics theory predicts that plants with a restoring *Rf*
491 allele incur some negative effect on their fitness, otherwise sexual dimorphism of a population (i.e. females expressed
492 by CMS and hermaphrodites) could not be maintained (e.g. Dufay et al. 2007). Considering the molecular
493 chaperone-like function of ORF20-like proteins, their off-target effect may not be negligible when these gene products

494 are abundant. For example, excess ORF20-like protein may occasionally interact with the mitochondrial respiratory
495 complex and impair its function, leading to a slight decrease in male fitness. This phenomenon could occur more likely
496 in normal cytoplasm lacking *preSatp6*; hence, an *Rfl* allele that is too strong might not be favoured. Moreover, the
497 frequency of the Owen-type mitochondria is generally low but differs among populations (Touzet 2012; Cheng et al.
498 2011). These factors as well as other unknown influences, including the presence of numerous alleles throughout the
499 entire population of beets, could make the selection of *Rfl* alleles complicated. This notion needs some theoretical
500 support and more data about the organizational and functional variation of this gene family in order to fully describe the
501 molecular evolution of sugar beet *Rfl*.

502

503 **Supplementary Material**

504 **Supplementary Fig. 1.** Agarose (2%) gel electrophoresis of PCR products amplified with primers for cytoplasmic
505 DNA markers.

506 **Supplementary Fig. 2.** PCR products amplified with primers for DNA marker o7.

507 **Supplementary Fig. 3.** DNA gel blot analysis of an NK-305 plant probed with an *orf20*-like 3'-UTR.

508 **Supplementary Fig. 4.** Alignment of nucleotide sequences of *orf20_{NK-198}* and *orf20_{NK-305-1}* coding and flanking regions.

509 **Supplementary Fig. 5.** Alignment of nucleotide sequences of *orf20_{LS}* and *orf20_{NK-305-2}* coding and flanking regions.

510 **Supplementary Fig. 6.** Alignment of amino acid sequences of the protein products deduced from *orf20_{NK-198}* and
511 *orf20_{NK-305-1}* nucleotide sequences.

512 **Supplementary Fig. 7.** Alignment of amino acid sequences of the protein products deduced from *orf20_{LS}* and
513 *orf20_{NK-305-2}* nucleotide sequences.

514 **Supplementary Fig. 8.** Alignment of nucleotide sequences of the upstream regions of *orf19* and *orf20_{NK-305-2}*.

515 **Supplementary Fig. 9.** Alignment of nucleotide sequences of the upstream regions of *orf20_{LS}* and *orf20_{NK-305-1}*.

516 **Supplementary Fig. 10.** Immunoblot analysis of proteins from transgenic suspension cells separated by SDS-PAGE.

517 **Supplementary Fig. 11.** Immunoblot analysis of proteins from transgenic suspension cells separated by BN-PAGE.

- 518 **Supplementary Table 1.** Nucleotide sequences of primers used in this study.
- 519 **Supplementary Table 2.** Segregation of male fertility and o7 marker types in an F₂ population.
- 520 **Supplementary Table 3.** Segregation of male fertility and s17 marker types in an admixture population.
- 521 **Supplementary Table 4.** Segregation of male fertility and s17 type in a BC₂F₂ population.
- 522
- 523 **References**
- 524 Alexander MP (1969) Differential staining of aborted and nonaborted pollen. *Stain Technol* 44:117-122
- 525 Arakawa T, Uchiyama D, Ohgami T, Ohgami R, Murata T, et al. (2018) A fertility-restoring genotype of beet (*Beta*
- 526 *vulgaris* L.) is composed of a weak *restorer-of-fertility* gene and a modifier gene tightly linked to the *Rfl* locus.
- 527 PLoS ONE 13:e0198409
- 528 Budar F, Pelletier G (2001) Male sterility in plants: occurrence, determinism, significance and use. *CR Acad. Sci. Paris,*
- 529 *Life Sciences* 324:543-550
- 530 Chase CD (2007) Cytoplasmic male sterility: a window to the world of plant mitochondrial-nuclear interactions. *Trends*
- 531 *Genet* 23:81-90
- 532 Chen L, Liu Y-G (2014) Male sterility and fertility restoration in crops. *Annu Rev Plant Biol* 65:579-606
- 533 Cheng D, Kitazaki K, Xu D, Mikami T, Kubo T (2009) The distribution of normal and male-sterile cytoplasm in
- 534 Chinese sugar-beet germplasm. *Euphytica* 165:345-351
- 535 Cheng D, Yoshida Y, Kitazaki K, Negoro S, Takahashi H, et al. (2011) Mitochondrial genome diversity in *Beta*
- 536 *vulgaris* L. ssp. *vulgaris* (Leaf and Garden Beet Groups) and its implications concerning the dissemination of the
- 537 crop. *Genet Res Crop Evol* 58:553-560
- 538 Doyle JJ, Doyle JL (1990) Isolation of plant DNA from fresh tissue. *Focus* 12:13-15
- 539 Ducos E, Touzet P, Boutry M (2001) The male sterile *G* cytoplasm of wild beet displays modified mitochondrial
- 540 respiratory complexes. *Plant J* 26:171-180
- 541 Dufay M, Touzet P, Maurice S, Cuguen J (2007) Modelling the maintenance of male-fertile cytoplasm in a
- 542 gynodioecious population. *Heredity* 99:349-356

543 Duroc Y, Gaillard C, Hiard S, Defrance MC, Pelletier G, et al. (2005) Biochemical and functional characterization of
544 ORF138, a mitochondrial protein responsible for Ogura cytoplasmic male sterility in Brassiceae. *Biochimie*
545 87:1089-1100

546 Duroc Y, Hiard S, Vrielynck N, Ragu S, Budar F (2009) The Ogura sterility-inducing protein forms a large complex
547 without interfering with the oxidative phosphorylation components in rapeseed mitochondria. *Plant Mol Biol*
548 70:123-137

549 Duvick DN (1965) Cytoplasmic pollen sterility in corn. *Adv Genet* 13:1-56

550 Fujii S, Bond CS, Small ID (2011) Selection patterns on restorer-like genes reveal a conflict between nuclear and
551 mitochondrial genomes throughout angiosperm evolution. *Proc Natl Acad Sci USA* 108:1723-1728

552 Geddy R, Brown GG (2007) Genes encoding pentatricopeptide repeat (PPR) proteins are not conserved in location in
553 plant genomes and may be subject to diversifying selection. *BMC Genomics* 8:130

554 Hanson MR, Bentolila S (2004) Interactions of mitochondrial and nuclear genes that affect male gametophyte
555 development. *Plant Cell* 16:S154-S169

556 Honma Y, Taguchi K, Hiyama H, Yui-Kurino R, Mikami T, *et al.* (2014) Molecular mapping of *restorer-of-fertility*
557 2 gene identified from a sugar beet (*Beta vulgaris* L. ssp. *vulgaris*) homozygous for the
558 non-restoring *restorer-of-fertility 1* allele. *Theor Appl Genet* 127:2567-2574

559 Hu L, Liang W, Yin C, Cui X, Zong J, et al. (2011) Rice MADS3 regulates ROS homeostasis during late anther
560 development. *Plant Cell* 23:515-533

561 Kagami H, Kurata M, Matsuhira H, Taguchi K, Mikami T, *et al.* (2015) Sugar beet (*Beta vulgaris* L.). In: Wang K (ed)
562 *Agrobacterium Protocols. Methods in Molecular Biology* vol. 1223. Springer Science+ Business Media, New
563 York, pp 335-347

564 Kagami H, Taguchi K, Arakawa T, Kuroda Y, Tamagake H, et al. (2016) Efficient callus formation and plant
565 regeneration are heritable characters in sugar beet (*Beta vulgaris* L.). *Hereditas* 153:12

566 Kato H, Tezuka K, Feng YY, Kawamoto T, Takahashi H, et al. (2007) Structural diversity and evolution of the *Rf-1*
567 locus in the genus *Oryza*. *Heredity* 99:516-524

568 Kim Y-J, Zhang D (2018) Molecular control of male fertility for crop hybrid breeding. *Trends Plant Sci* 23:53-65

569 Kitazaki K, Arakawa T, Matsunaga M, Yui-Kurino R, Matsuhira H, et al. (2015) Post-translational mechanisms are
570 associated with fertility restoration of cytoplasmic male sterility in sugar beet (*Beta vulgaris*). Plant J 83:290-299

571 Kitazaki K, Kubo T, Kagami H, Matsumoto T, Fujita A, et al. (2011) A horizontally transferred tRNA^{Cys} gene in the
572 sugar beet mitochondrial genome: evidence that the gene is present in diverse angiosperms and its transcript is
573 amino acylated. Plant J 68:262-272

574 Lane N (2011) Mitonuclear match: optimizing fitness and fertility over generations drives ageing within generations.
575 Bioessays 33:860-869

576 Laporte V, Merdinoglu D, Saumitou-Laprade P, Butterlin G, Vernet P, et al. (1998) Identification and mapping of
577 RAPD and RFLP markers linked to a fertility restorer gene for a new source of cytoplasmic male sterility
578 in *Beta vulgaris* ssp. *maritima*. Theor Appl Genet 96:989–996

579 Lee J, Yoon JB, Park HG (2008) Linkage analysis between the partial restoration (*pr*) and the restorer-of-fertility (*Rf*)
580 loci in pepper cytoplasmic male sterility. Theor Appl Genet 117:383-389

581 Li X-Q, Jean M, Landry BS, Brown GG (1998) Restorer genes for different forms of *Brassica* cytoplasmic male
582 sterility map to a single nuclear locus that modifies transcripts of several mitochondrial genes. Proc Natl Acad
583 Sci USA 95:10032-10037

584 Luo D, Xu H, Liu Z, Guo J, Li H, et al. (2013) A detrimental mitochondrial-nuclear interaction causes cytoplasmic
585 male sterility in rice. Nat Genet 45:573-577

586 Mackenzie SA (2005) The influence of mitochondrial genetics on crop breeding strategies. In: Janick J (ed) Plant
587 Breeding Reviews. John Wiley & Sons, New York, pp 115-138

588 Majewska-Sawka A, Rodriguez-Garcia MI, Nakashima H, Jassen B (1993) Ultrastructural expression of cytoplasmic
589 male sterility in sugar beet (*Beta vulgaris* L.). Sex Plant Reprod 6:22-32

590 Matsuhira H, Kagami H, Kurata M, Kitazaki K, Matsunaga M, et al. (2012) Unusual and typical features of a novel
591 *restorer-of-fertility* gene of sugar beet (*Beta vulgaris* L.). Genetics 192:1347-1358

592 Melonek J, Stone JD, Small I (2016) Evolutionary plasticity of restorer-of-fertility-like proteins in rice. Sci Rep 6:35152

593 Mora JRH, Rivals E, Mireau H, Budar F (2010) Sequence analysis of two alleles reveals that intra-and intergenic
594 recombination played a role in the evolution of the radish fertility restorer (*Rfo*). BMC Plant Bio 10:35

595 Moritani M, Taguchi K, Kitazaki K, Matsuhira H, Katsuyama T, et al. (2013) Identification of the predominant
596 nonrestoring allele for Owen -type cytoplasmic male sterility in sugar beet (*Beta vulgaris* L.): development of
597 molecular markers for the maintainer genotype. Mol Breed 32:91-100

598 Nishizawa S, Kubo T, Mikami T (2000) Variable number of tandem repeat loci in the mitochondrial genomes of
599 beets. Curr Genet 37:34-38

600 Ohgami T, Uchiyama D, Ue S, Yui-Kurino R, Yoshida Y, Kamei Y, Kuroda Y, et al. (2016) Identification of molecular
601 variants of the nonrestoring *restorer-of-fertility 1* allele in sugar beet (*Beta vulgaris* L.). Theor Appl Genet
602 129:675-688

603 Owen FV (1945) Cytoplasmically inherited male-sterility in sugar beets. J Agr Res 71:423-440

604 Papini A, Mosti S, Brighigna L (1999) Programmed-cell-death events during tapetum development of angiosperms.
605 Protoplasma 207:213-221

606 Parish RW, Li SF (2010) Death of a tapetum: a programme of developmental altruism. Plant Sci 178:73-89

607 Pillen K, Steinrücken G, Herrmann RG, Jung C (1993) An extended linkage map of sugar beet (*Beta vulgaris* L.)
608 including nine putative lethal genes and the restorer gene *X*. Plant Breed 111:265-272

609 Ran Z, Michaelis G (1995) Mapping of a chloroplast RFLP marker associated with the CMS cytoplasm of sugar beet
610 (*Beta vulgaris*). Theor Appl Genet 91:836-840

611 Rhoads DM, Brunner-Neuenschwander B, Levings CS III, Siedow JN (1998) Cross-linking and disulfide bond
612 formation of introduced cysteine residues suggest a modified model for the tertiary structure of URF13 in the
613 pore-forming oligomers. Arch Biochem Biophys 354:158-164

614 Rogers HJ (2006) Programmed cell death in floral organs: how and why do flowers die? Annal Bot 97:309-315

615 Ryan MT, Hoogenraad NJ (2007) Mitochondrial-nuclear communications. Annu Rev Biochem 76:701-722

616 Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual 2nd edn. Cold Spring
617 Harbor Laboratory Press, Cold Spring Harbor

618 Sanders PM, Bui AQ, Weterings K, McIntire KN, Hsu Y-C, et al. (1999) Anther developmental defects in *Arabidopsis*
619 *thaliana* male-sterile mutants. Sex Plant Reprod 11:297-322

620 Schnable PS, Wise RP (1998) The molecular basis of cytoplasmic male sterility and fertility restoration. Trends Plant
621 Sci 3:175-180

622 Stone JD, Koloušková P, Sloan D, Štorchová H (2017) Non-coding RNA may be associated with cytoplasmic male
623 sterility in *Silene vulgaris*. J Exp Bot 68:1599-1612

624 Taguchi K, Hiyama H, Yui-Kurino R, Muramatsu A, Mikami T, et al. (2014) Hybrid breeding skewed the allelic
625 frequencies of molecular variants derived from *restorer-of-fertility 1* locus for cytoplasmic male sterility in sugar
626 beet (*Beta vulgaris* L.). Crop Sci 54:1407-1412

627 Tang H, Xie Y, Liu Y-G, Chen L (2017) Advances in understanding the molecular mechanisms of cytoplasmic male
628 sterility and restoration in rice. Plant Reprod 30:179-184

629 Touzet P (2012) Mitochondrial genome evolution and gynodioecy. In: Marechal-Drouard, L. (ed) Mitochondrial
630 Genome Evolution. Academic Press, Oxford, pp 71-98

631 Touzet P, Hueber N, Bürkholz A, Barnes S, Cuguen J (2004) Genetic analysis of male fertility restoration in wild
632 cytoplasmic male sterility *G* of beet. Theor Appl Genet 109:240-247

633 van der Blienk AM, Sedensky MM, Morgan PG (2017) Cell biology of the mitochondrion. Genetics 207:843-871

634 Wilson ZA, Zhang D-B (2009) From *Arabidopsis* to rice: pathways in pollen development. J Exp Bot 60:1479-1492

635 Wise RP, Dill CL, Schnable PS (1996) *Mutator*-induced mutations of the *rf1* nuclear fertility restorer of T-cytoplasm
636 maize alter the accumulation of T-*urf13* mitochondrial transcripts. Genetics 143:1383-1394

637 Yamamoto MP, Kubo T, Mikami T (2005) The 5'-leader sequence of sugar beet mitochondrial *atp6* encodes a novel
638 polypeptide that is characteristic of Owen cytoplasmic male sterility. Mol Genet Genom 273:342-349

639 Zhang H, Cheng X, Zhang L, Si H, Ge Y, et al. (2018) *Rf4* has minor effects on the fertility restoration of wild
640 abortive-type cytoplasmic male sterile *japonica* (*Oryza sativa*) lines. Euphytica 214:49

642 **Table 1.** Segregation of male fertility and s17 marker genotypes in F₂ populations derived from TA-33BB-CMS x
 643 NK-305

s17 (length of DNA fragment in kbp)	Male fertility			Total
	Fully fertile	Partial fertile	Sterile	
p2p2 (0.5 and 1.4)	5	1	0	6
	6			
p2p4 (0.5, 0.6, 1.1, and 1.4)	1	13	0	14
	14			
p4p4 (0.6 and 1.1)	0	0	6	6
	0			
Total	6	14	6	26
	20			

644

645

646 **Table 2.** Quantity of *orf20*_{NK-305-1} mRNA in anthers with different NK-305 *Rfl* doses

Anther stage	Reference	s17			Fold change (p2p2 vs. p2p4)
		p2p2 (n=5)	p2p4 (n=4)	p4p4 (n=3)	
Meiosis	<i>Actin</i>	0.285±0.079	0.177±0.015	ND ¹	1.61* ³
Tetrad		0.357±0.112	0.188±0.002		1.9*
Microspore		0.185±0.067	0.122±0.031	NA ²	1.52
Meiosis	<i>efla</i>	0.097±0.036	0.055±0.009	ND	1.77*
Tetrad		0.115±0.046	0.053±0.005		2.16** ⁴
Microspore		0.078±0.009	0.043±0.017	NA	1.8*

647 ¹ND, not detected; ²NA, not applicable; ³*, *p*<0.05 (Welch's *t* test); ⁴***, *p*<0.01 (Welch's *t* test).

648

649

650

651 Figure legends

652

653 **Fig 1** Sugar beet flowers with varying levels of male fertility and their anther contents. (a) A fully fertile flower. The
654 anther is well developed and has dehisced to shed abundant pollen grains. The anther color is yellow to bright yellow.
655 The semi-fertile type (type a) can be included in this class but less pollen is released. (b) A partially fertile or
656 semi-fertile (type b) flower. The anther is fairly well developed but rarely dehisces. No or very little pollen is shed. The
657 anther color is more orangish than a fully fertile anther. (c) A male-sterile flower with shriveled anthers. The anther
658 color is white or brown. (d-h) The typical anther content of plants with different genotypes and cytoplasm types. The
659 scale bar is 50 μm in length. (d) A fully-fertile plant of s17 marker type p2p2. (e) A partially-fertile plant of s17 marker
660 type p2p4. (f) A male-sterile plant of s17 marker type p4p4. (g) TA-33BB-O (normal fertile cytoplasm and *rflrflrf2rf2*).
661 (h) TA-33BB-CMS (Owen cytoplasm and *rflrflrf2rf2*).

662

663 **Fig 2** (a) Agarose gel electrophoresis of marker s17 cleaved amplified fragments from plants of TA-33BB-CMS,
664 NK-305 and representatives of their F_2 population, (b) TA-33BB-CMS, NK-198, and their BC_2F_2 . Size markers are
665 shown on the left (kbp). Marker genotypes are shown below the gel images.

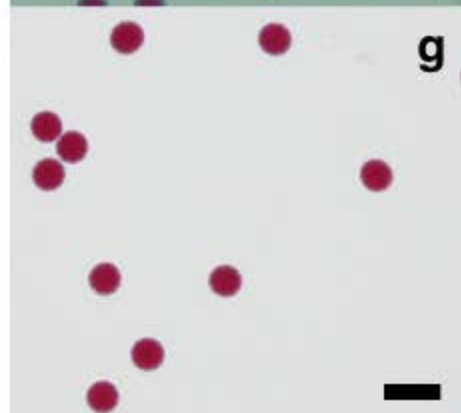
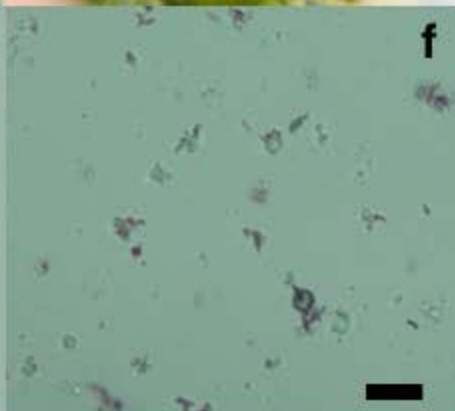
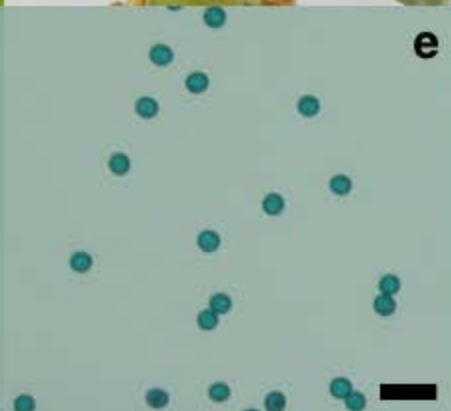
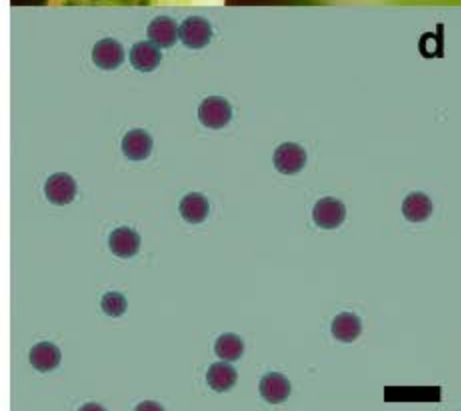
666

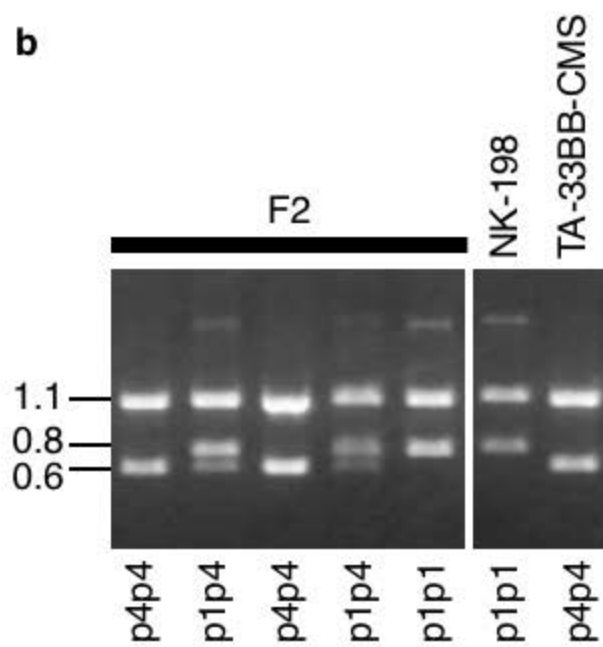
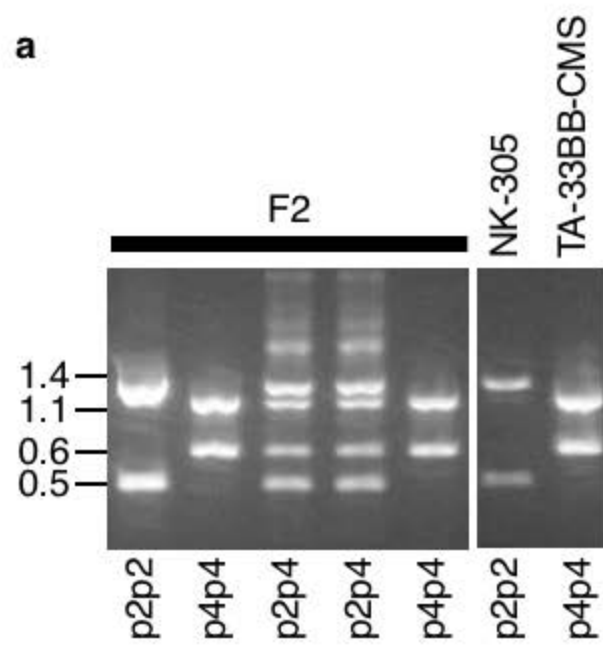
667 **Fig 3** Distribution of male fertility indices in populations segregating NK-198 *Rfl* (BC_2F_2 population) or NK-305 *Rfl*
668 (admixture population). The vertical axis indicates the male fertility index. Genotypes of the s17 marker are shown
669 under the horizontal axis. Boxes indicate the first and third quartiles, and the inside lines are median values. The upper
670 and lower external whisker lines indicate first quartiles-1.5 x interquartiles and third quartiles+1.5 x interquartiles,
671 respectively. Outliers are plotted as circles. Classes a, b and c were defined by the Steel-Dwass test ($p < 0.01$).

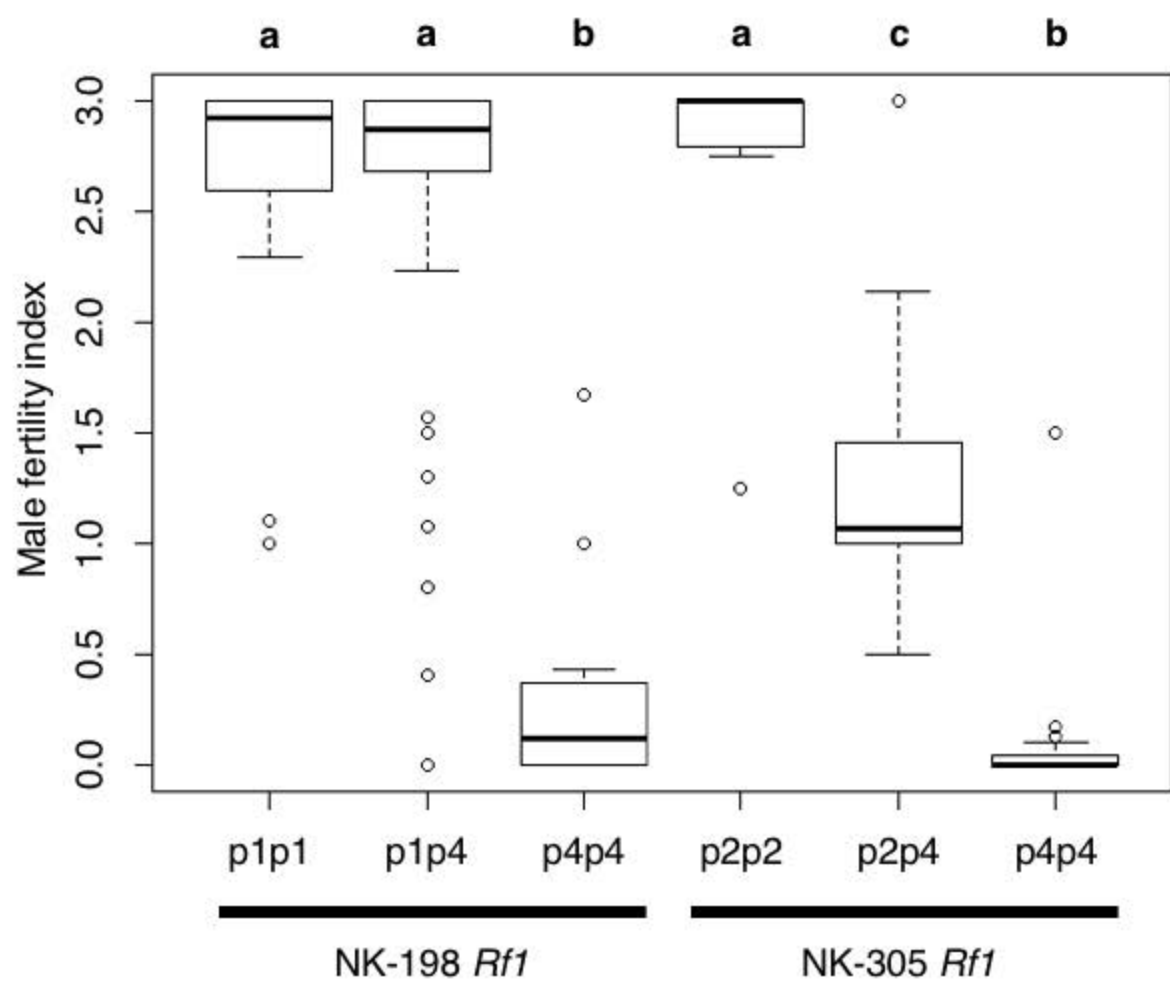
672

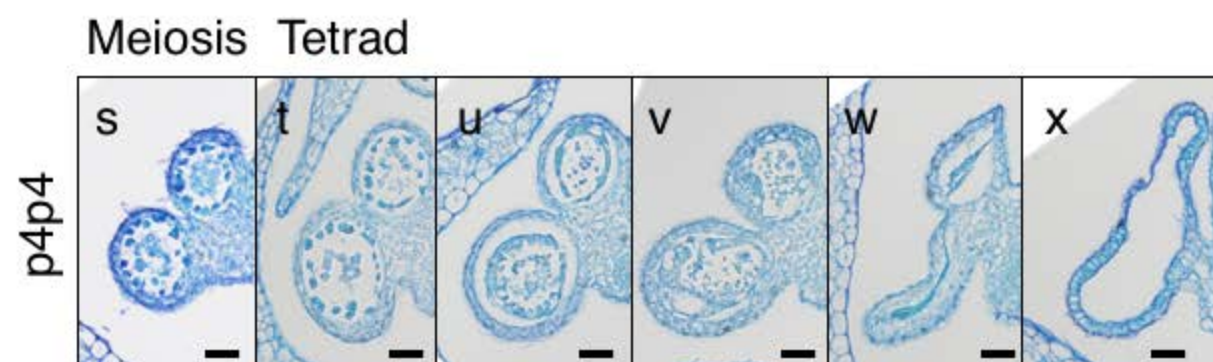
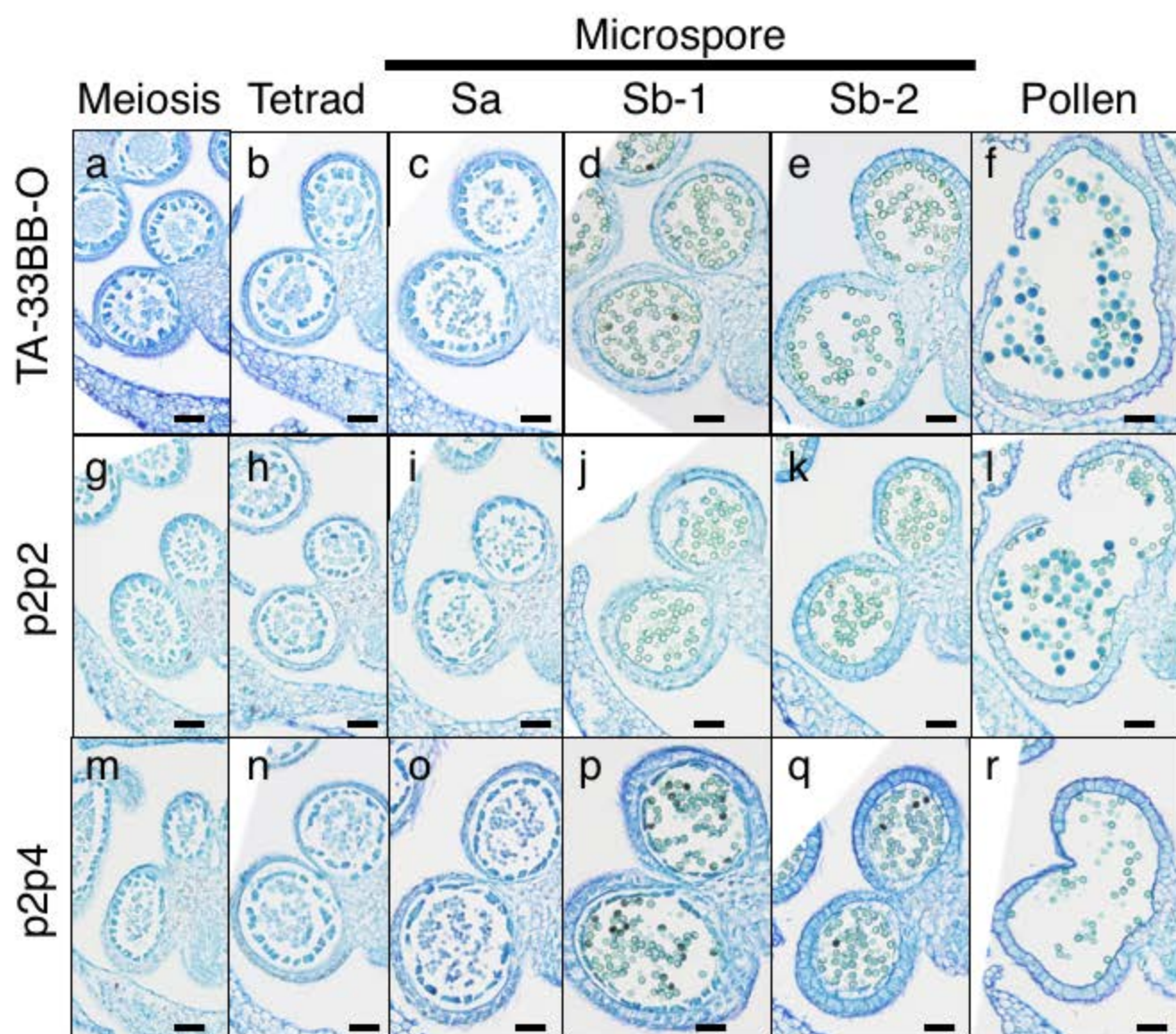
673 **Fig 4** Light microscopic observations of transverse sections of sugar beet anthers of different genotypes. TA-33BB-O
674 has normally fertile mitochondria and an *rflrflrf2rf2* genotype. Plants with s17 genotypes of p2p2, p2p4, and p4p4
675 were analyzed. Anther developmental stages defined in this study are shown at the top except for p4p4 due to its severe
676 developmental abnormalities. Scale bars are 50 μm . (a-f) TA-33BB-O, from meiosis- to pollen stages, respectively. (g-l)
677 Plants with a p2p2 genotype, from meiosis- to pollen stages, respectively. (m-r) Plants with a p2p4 genotype, from
678 meiosis- to pollen stages, respectively. (s-x) Plants with a p4p4 genotype. Panels s and t are in the meiosis and tetrad
679 stages, respectively. Panels u to x deviate from normal anther development.

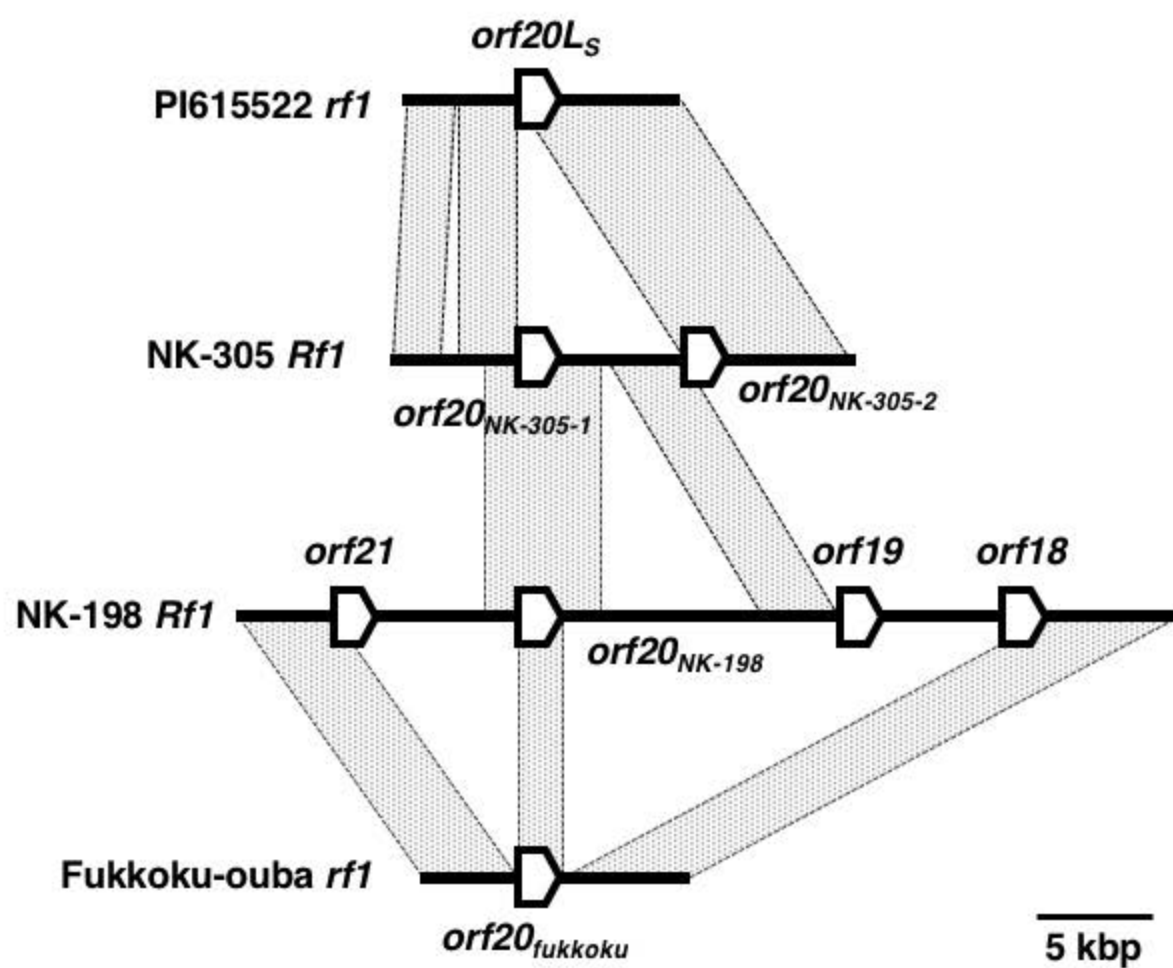
680

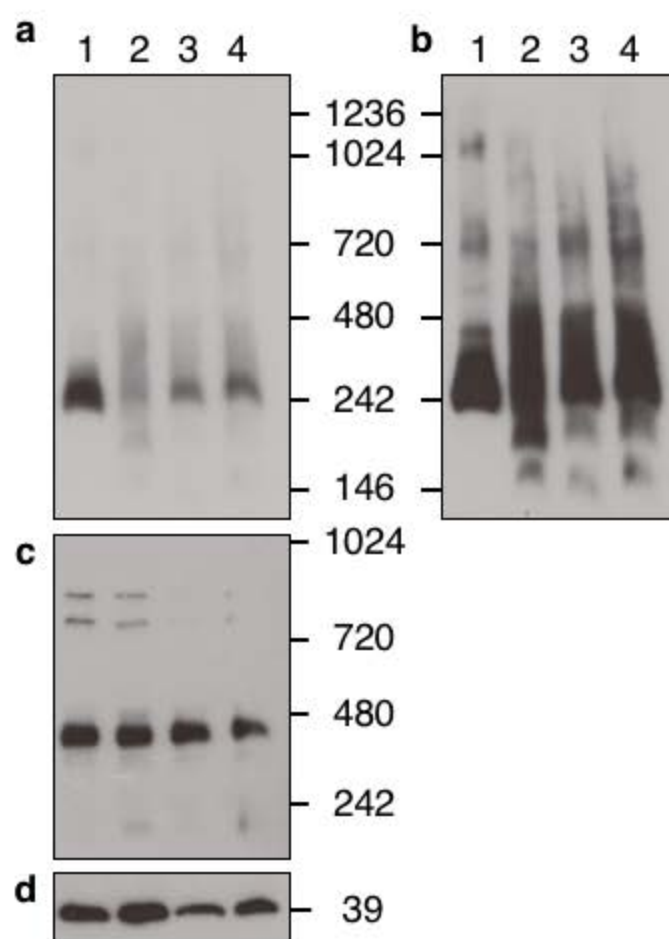


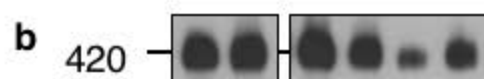
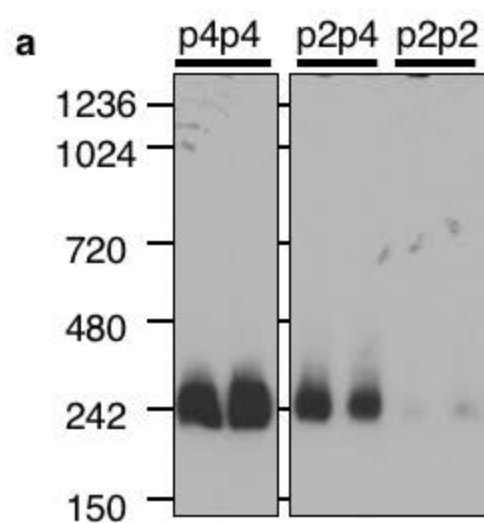






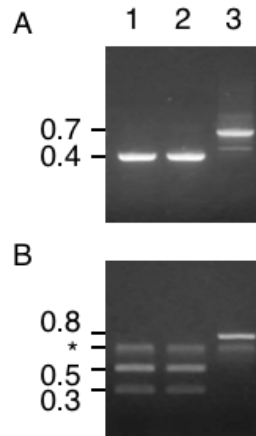




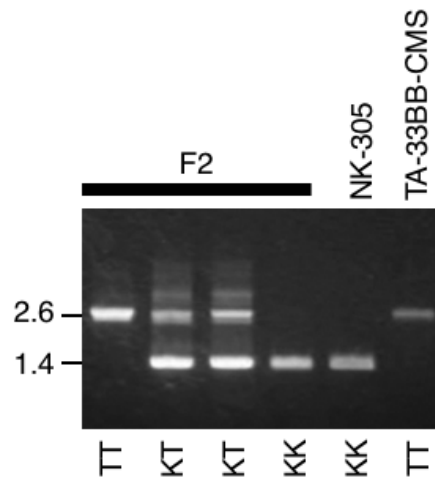


c

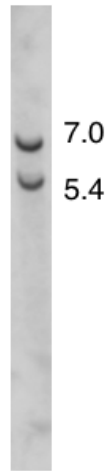
s17	n	Mean \pm SD	<i>p</i>
p4p4	3	1.568 \pm 0.164	0.0019
p2p4	3	0.857 \pm 0.068	0.0013
p2p2	2	0.118 \pm 0.014	



Supplementary Fig 1 Agarose (2%) gel electrophoresis of PCR products amplified with primers for cytoplasmic DNA markers. Lanes 1, 2, and 3 are NK-305, TA-33BB-CMS, and TA-33BB-O plants, respectively. Size markers are shown on the left (kbp). **A.** A mitochondrial marker targeting TR-1. **B.** A chloroplast marker targeting the *petG-psbE* intergenic region. Note that the PCR products of NK-305 and TA-33BB-CMS (lanes 1 and 2, respectively) are cut into two fragments of 0.3 and 0.5 kbp after *Hind*III digestion, whereas those from TA-33BB-O remain intact. An asterisk indicates the DNA bands derived from non-specific amplification.



Supplementary Fig 2 PCR products amplified with primers for DNA marker o7. Results from plants of TA-33BB-CMS, NK-305, and representatives of their F₂ population are shown. Size markers are shown on the left (kbp). Marker types are shown below the gel images.



Supplementary Fig 3 DNA gel blot analysis of an NK-305 plant probed with a 3'-UTR from *bvORF19* (an *orf20*-like gene). *HindIII*-digested total cellular DNA was electrophoresed in a 1.0% agarose gel. Size markers are shown on the right (kbp).

orf20_NK-198 -GAAACTGCTTATAAAGCTAGTGAATCAACGAAATATCAACAAGAAAATCACATTGT--- -1945
orf20_NK-305-1 ACATAATACAAATAAA--CAATAAAGTAAT-AGACATTAATACAAGTATAATATTATATA -1944
* * * * * ***** * * * * * * * * * * * * * * * * * *

orf20_NK-198 -TCATAGGAAAACTCCTAT-AACATTTGCATA----CATTGT-AGGTCTTTGT----TGC -1896
orf20_NK-305-1 ATCATTGTGGTACTTTAATTAATAATTCTAATAATAACATAATCAACTAATAGTGATATGA -1884
***** *

orf20_NK-198 ACTT-TAATCCCTCGGGCCTCGTATATTGATATTAAGTGTATTTTAATCTACGTTTTTTT -1837
orf20_NK-305-1 AATTATGAATAACAAAATAATGGACAATAATAC-AAATGTATATTAAACATTGACTATTT -1825
* *

orf20_NK-198 -----CTATTGCAACAATTACTACTTTGGATAATTTTACATCTATTGCAACAATTTCACT -1782
orf20_NK-305-1 GGACCTTATTGGACC--TTATTAGACCTGAT----TGAAACTTATTG-GACCTTATTAGA -1772
***** *

orf20_NK-198 TTTGGTATAAAAAGCAATATTTTCAAGAACAAAGCAGCATTGTGCACCCGAGA-----GA -1729
orf20_NK-305-1 CCTGAT-TGGAACCTTATTGCACCTGATTGAAACT-TATTG--CACCTGGAACCTTATTGGA -1716
* *

orf20_NK-198 ACATACCAT-TCATATAGAAGAATATGATTTTTTTTCAAC-----AACTTTTCAAGA -1678
orf20_NK-305-1 CCTTATTAGACCTTATTGGAAGTTATTGCCTTATTAGACCTTATTACAACCTTATCTGAA -1656
* *

orf20_NK-198 TAAAAAAAGCATACAATATAA-----AATTAA--AGAACATG---TAAATCTCCAAC -1631
orf20_NK-305-1 CTTATTGGACCTGAAACTTAATTTTTTTTAAAGTTGAGCAGAACGCACCCTAAATCTCCAAT -1596
* *

orf20_NK-198 AACATTTGAAAAACCTA-----AAAAAAAGCATCTATTGTAAATGAACATTTAGTCT -1577
orf20_NK-305-1 ----TATAAAATAACATAGGTGCGAATAAAAAAGAAAGTATTGTAAATGAACATTTAGTCT -1540
* *

orf20_NK-198 AAATTTAAAGAACCTAACTTTTAAATGTAAAAATTTGAAAAAGGAATCTCGCCAACAAC -1517
orf20_NK-305-1 AAATTTAAAGAACCTAACTTTTAAATGTAAA--TTAAAAAGGAATCTCGCCAACAAC -1483
***** *

orf20_NK-198 CATTTTTCTAAAAGGTAAAACAAGTTGCAAGATTTAATGAAACAGATAACAACATTTTT -1457
orf20_NK-305-1 CATTTTTCTAAAAGGTAAAACAAGTTGCAAGATTTAATGAAACAGATAACAACATTTTT -1423
***** *

orf20_NK-198 CTTAAATCATAAATTCTTAAATATTAACCTACATCGTTTAAACAGAGGTGCACCATCC -1397
orf20_NK-305-1 CTTAAATCATAAATTCTTAAATATTAACCTACATCGTTTAAACAGAGGTGCACCATCC -1363
***** *

orf20_NK-198 TTTATGCGTACTTGGATGCATGTTCTATTTGCGCTTTTCATTCCCTAAAAAACCGAGT -1337
orf20_NK-305-1 TTTATGCGTACTTGGATGCATGTTCTATTTGCGCTTTTCATTCCCTAAAAAACCGAGT -1303
***** *

orf20_NK-198 CTAAAGCTATTGTTATAATACTCTAGTAGTCCTCTCAAAAAAA--TAAGGCTTTGCTA -1278
orf20_NK-305-1 CTAAAGCTATTGTTATAATACTCTAGTAGTCGTCTCAAAAAAAATAAGGCTTTGCTA -1243
***** *

orf20_NK-198	AATTACGCCCTAAATTTTTCTGGTAACGCCCTAAATACGGTATTTTCATACCGTATTTAT	-1218
orf20_NK-305-1	AATTACGCCCTAAATTTTTCTGGTAACGCCCTAAATACGGTATTTTCATACCGTATTTAT	-1183

orf20_NK-198	ATACCTGCATTTCCTTCTCTCTTTCCCTTCACTCACTCACTTTCTCTCTCCTCTCCAC	-1158
orf20_NK-305-1	ATACCTGCATTTCCTTCTCTCTTTCCCTTCACTCACTCACTTTCTCTCTCCTCTCCA-	-1124

orf20_NK-198	CATAACTGCCACCAATTTCTCTCTCCTCTCCACCACAACCTGCCACCACACCACCACTACC	-1098
orf20_NK-305-1	-----CCACAACCTGCCACCACACCACCACTACC	-1096

orf20_NK-198	ATATTCTTCTCCAATTTACACTTCTGACTCTTATGTCTTCTAACAACATTGTGCCGGCA	-1038
orf20_NK-305-1	ATATTCTTCTCCAATTTACACTTTTGACTCCTATGTCTTCTAACAACATTGTGTCGGCA	-1036
***** ***** ***** *****		
orf20_NK-198	CCACCACCACCGACGTACCCCCGACACCCACCATCGTCGTATCCCGGCCACCACCGCCG	-978
orf20_NK-305-1	CCACCACCACCGACGTACCCCCGACACCCACCATCGTCGTATCCCGGCCACCACCGCCG	-976

orf20_NK-198	TACCCCGACACACCACCACCATCCCTTTTTCTCTACTCCGTTTTCTTTATTTTCTGAA	-918
orf20_NK-305-1	TACCCCGACACACCACCACCATCCCTTTTTCTCTACTCCGTTTTCTTTATTTTCTGA-	-917

orf20_NK-198	TTTTTTTTTAAAAAAAGGGGAGGCGTCCACCATGGACGCCCCACCACTGGTGGAGGCG	-858
orf20_NK-305-1	-TTTTTTTTTAAAAAAAGGGGAGGCGTCCACCATGGACGCCCCATCACTGGTGGAGGCG	-858
***** ***** ***** *****		
orf20_NK-198	TCCATG-CAGGCGCCTCACCATGTGGTGAAGCGCCTGCCATAGACGCCTCCCCTTTTTT	-799
orf20_NK-305-1	TCCATGGCAGGCGCCTCACCATGTGGTGAAGCGCCTGCCATAGACGCCTCCCCTTTTTT	-798
***** ***** ***** *****		
orf20_NK-198	TAAAAAAATTTCAAATTACTGTTAATGAAGAAACATAGAAGAGGAGGGTGGGTACGAC	-739
orf20_NK-305-1	TAAAAAAATTTCAAATTACTGTTAATGAAGAAACATAGAAGAGGAGGGTGGGTACGAC	-738

orf20_NK-198	GGTTGTGGTGCGGTGGCAGTGTGTGGTGGTGGCGGCAGTACTGGTGGCAGGCGACTG	-679
orf20_NK-305-1	GGTTGTGGTGCGGTGGCAGTGTGTGGTGGTGGCGGCAGTATTGGTGGCAGGCGACTG	-678
***** *****		
orf20_NK-198	GTGCGGCGGTGGTGGTGAATGGTCCCAATGGTAATGTGGGAGAAGAGGGGAGAGAAGT	-619
orf20_NK-305-1	GTGCGGCGGTGGTGGTGAACGGTCCCAATGGTAATGTGGGAGAAGAGGGGAGAGAAGT	-618
***** *****		
orf20_NK-198	GAATGTAGAAGGTGGAATCTGAAAATGAAAAATACGTATATTATACGTATTTAGGGC	-559
orf20_NK-305-1	GAATGTAGAAGGTGGAATATGAAAATGAAAAATACGTATATTATACGTATTTAGGGC	-558
***** *****		
orf20_NK-198	GTTAGTAAGATATTCTAGGGCGTCGTTAGCAAATGAGAAAAGATAATACACTCTAGTTG	-499
orf20_NK-305-1	GTTAGTAAGATATTCTAGGGCGTCGTTAGCAAATGAGAAAAATAATACACTCTAGCTG	-498
***** ***** **		

orf20_NK-198 TCGTAGACTCTCAATATGTGTCATTTAGAGACTCGTAACGCATTGACGCACTTACTCGGG -439
orf20_NK-305-1 TCGTAGACTCTCAATATGTGTCATTTAGAGACTCGTAACGCATTGACGCACTTACTCGGG -438

orf20_NK-198 GTAGAAATATTTTGTTCATTTATTAATGAGAAAATTTTCATCCCCTGATCCCAATTATC -379
orf20_NK-305-1 GTAGAAATATTTTGTTCATTTATTAATGAGAAAATTTTCATCCCCTGATCCCAATTATC -378

orf20_NK-198 AAATCAACATCTAAAAATTTAAATGAGTAGGTACGTAACGAAAAACGAATGACTCTCGA -319
orf20_NK-305-1 AAATCAACATCTAAAAATTTAAATGACTAGGTACGTAACGAAAAACGAATGACTCTCGA -318

orf20_NK-198 TAATAGTACACCCCATTAATCCATTCTTAGTTTGTTCATAGTTTGTTCGCTAGTGCATA -259
orf20_NK-305-1 TAATAGTACACCCCATTAATCCATTCTTAGTTTGTTCATAGTTTGTTCGCTAGTGCATA -258

orf20_NK-198 GCTGGCTGCAAAAGAAATCTTTTGCACAGAGAAAACTTTGCACCTTCGGAATTCAGCAG -199
orf20_NK-305-1 GCTGGCTGCAAAAGAAATCTTTTGCACAGAGAAAACTTTGCACCTTCGGAATTCAGCAG -198

orf20_NK-198 GAATATCATAACCATTTATGGAAGCAACAACCTTTGTGACCCATTTTCATCTAAAACCCCTT -139
orf20_NK-305-1 GAATATCATAGCCATTTATGGAAGCAACAACCTTTGTGACCCATTTTCATCTAAAACCCCTT -138

orf20_NK-198 AATCTCATAAATTTTACATTTTCAAGATTCAAAAATCACGTAATTTTTTTTTTGGTATGT -79
orf20_NK-305-1 AATCTCATAAATTTTACATTTTCAAGATTCAAAAATCACGTAATTTTTTTTTTGGTATGT -79

orf20_NK-198 TACTTGAACCCAGTTTCATAAAGTACGCTGAAATTCAGAAATTTGGAGCAAAGTTAGCAGC -19
orf20_NK-305-1 TACTTGAACCCAGTTTCATAAAGTACGCTGAAATTCAGAAATTTGGAGCAAAGTTAGCACC -19

orf20_NK-198 TTTTGTGTTCAAAAATC -1
orf20_NK-305-1 TTTTGTGTTCAAAAATC -1

orf20_NK-198 / Exon 1
ATGGCGTGGTACAGAAATTCAGGTTTGTCTACAATGCTTTAAACTCAACTTGGGTTCC 60
orf20_NK-305-1 ATGGCGTGGTACAGAAATTCAGGTTTGTCTACAATGCTTTAAACTCAACTTGGGTTCC 60

orf20_NK-198 AAAACATTTGGTACTATTCCAACCTCAAGAGTTCATTGCAATTCCTCATCTTTGTTTTAC 120
orf20_NK-305-1 AAAACATTTGGTACTATTCCAACCTCAAGAGTTCATTGCAATTCCTCATCTTTGTTTTAC 120

orf20_NK-198 AATCAATCTACTAATAAGTGTAGTGGGTTATTTGGGTCTGCAAAATCTGGGTATTTTAAT 180
orf20_NK-305-1 AATCAATCTACTAATAAGTGTAGTGGGTTATTTGGGTCTGCAAAATCTGGGTATTTTAAT 180

orf20_NK-198 GGGTTTAAACATCATCAAGAGATTAGCTCTTTCTCTGGTTTTCGCAAGGAGAAATTATCAT 240
orf20_NK-305-1 GGGTTTAAACATCATCAAGAGATTAGCTCTTTCTCTGGTTTTCGCAAGGAGAAATTATCAT 240
** *****

orf20_NK-198	GGTGATAAAACCGAAGTAAGTGTGAATCATGGCTGGAAAAATTCCTTGTTCCAATTGGA	300
orf20_NK-305-1	GGTGATAAAACCGAAGTAAGTGTGAATCATGGCTGGAAAAATTCCTTGTTCCAATTGGA	300

orf20_NK-198	CTAATCTTGACTTTTGGTATACTTGGTTACCCCTCATGTGCACCCAGTAGTTGTGCCATAT	360
orf20_NK-305-1	CTAATCTTGACTTTTGGTATACTTGGTTACCCCTCATGTGCACCCAGTAGTTGTGCCATAT	360

orf20_NK-198	ACAGGAAGGAAGCATTATGTGCTTATGTCAACAACCTCGTGAGAATGAAATTGGAGAAGTT	420
orf20_NK-305-1	ACAGGAAGGAAGCATTATGTGCTTATGTCAACAACCTCGTGAGAATGAAATTGGAGAAGTT	420

orf20_NK-198	GAGAAGCGGAAAATACAACCTGCTACACACCCTGATACTGATAGGGTTAGGTCAATATTC	480
orf20_NK-305-1	GAGAAGCGGAAAATACAACCTGCTACACACCCTGATACTGATAGGGTTAGGTCAATATTC	480

orf20_NK-198	CAACACATTCTTGAATCACTGGAAGAGAGATTAATCACCATGAACCTCGAACTCGAAAGA	540
orf20_NK-305-1	CAACACATTCTTGAATCACTGGAAGAGAGATTAATCACCATGAACCTCGAACTCGAAAGA	540

orf20_NK-198	GATGAACTTTCAAGGAGAAAACCATTTGGAAGGAGGAGACAGTTGATGATAAAGATAGT	600
orf20_NK-305-1	GATGAACTTTCAAGGAGAAAACCATTTGGAAGGAGGAGACAGTTGATGATAAAGATAGT	600

orf20_NK-198	AGGAAGAAGCATAGTGGGGCTAAGATAACTACTAACCATTTGGAAGGGATGAATTGGGAA	660
orf20_NK-305-1	AGGAAGAAGCATAGTGGGGCTAAGATAACTACTAACCATTTGGAAGGGATGAATTGGGAA	660

orf20_NK-198	ATTTTCGTTGTTGATAAACCGTTGGTTGAGTCCAGTTATTTATTAGGTGGGAAGATTGTT	720
orf20_NK-305-1	ATTTTCGTTGTTGATAAACCGTTGGTTGAGTCCAGTTATTTATTAGGTGGGAAGATTGTT	720

orf20_NK-198	GTTTACACCGGATTGCTCAACCATTGCAACTCTGATGCTGAATTGGCTACAATTATCGCG	780
orf20_NK-305-1	GTTTACACCGGATTGCTCAACCATTGCAACTCTGATGCTGAATTGGCTACAATTATCGCG	780

/ Intron 1		
orf20_NK-198	CATCAGGTATATAAACTATTTCATGGGACTCCAATTATGTGCTTAAGCTGATGGTTAATA	840
orf20_NK-305-1	CATCAGGTATATAAACTATTTCATGGGACTCCAATTATGTGCTTAAGCTGATGGTTAATA	840

orf20_NK-198	GAACTATACAAAAAACTGATGAATTTTAGGTTATCAGATTACATTATGAATGTCATATG	900
orf20_NK-305-1	GAACTATACAAAAAACTGATGAATTTTAGGTTATCAGATTACATTATGAATGTCATATG	900

/ Exon 2		
orf20_NK-198	TCAATTTGGTGGTATGTATTTGTTAGGTTGGGCATGCTGTGGCTCGACATGAGGCAGAGG	960
orf20_NK-305-1	TCAATTTGGTGGTATGTATTTGTTAGGTTGGGCATGCTGTGGCTCGACATGAGGCAGAGG	960

orf20_NK-198	ATTCGACAGCATTTTTCTGGTTGTTAATATCCCTCAACGTGATATTATTTAAATTCTAT	1020
orf20_NK-305-1	ATTCGACAGCATTTTTCTGGTTGTTAATATCCCTCAACGTGATATTATTTAAATTCTAT	1020

orf20_NK-198	TTACTGAGCCTGAATCTGCCAATGCAAGATCAAACTACTCTTAAGGCATCCTCTCTTGC	1080
orf20_NK-305-1	TTACTGAGCCTGAATTTGCCAATGCAAGATCAAACTACTCTTAAGGCATCCTCTCTTGC	1080

/ Intron 2		
orf20_NK-198	AAAAGTAAGTCTCTTACTCTTAAATGTTTTCTTGATGATTAAACAAACATGTGGTACTGC	1140
orf20_NK-305-1	AAAAGTAAGTCTCTTACTCTTAAATGTTTTCTTGATGATTAAACAAACATGTGGTACTGC	1140

orf20_NK-198	TACTGCATAACTGTGTTACTGCATCACATATGTTACTGCATAATTGCAAACATATATTACA	1200
orf20_NK-305-1	TACTGCATAACTGTGTTACTGCATCACATATGTTACTGCATAATTGCAAACATATATTACA	1200

orf20_NK-198	TGCCCGGACCTAGTAACCTGTTTCATTTGTCAGCGATTTTCATTTAGATATCCATTTGAGA	1260
orf20_NK-305-1	TGCCCGGACCTAGTAACCTGTTTCATTTGTCAGCGATTTTCATTTAGATATCCATTTGAGA	1260

orf20_NK-198	GCAAGTTAAATTTGTATCAAGTTGTGGAATGGAAAAGTAATAGAACTAAATAGAGAGGTG	1320
orf20_NK-305-1	GCAAGTTAAATTTGTATCAAGTTGTGGAATGGAAAAGTAATAGAACTAAATAGAGAGGTG	1320

orf20_NK-198	TGATGCTAATAAAATCTAATCCATTACTGAGTAATGGTTTTGGATCGATATATGGATTGC	1380
orf20_NK-305-1	TGATGCTAATAAAATCTAATCCATTACTGAGTAATGGTTTTGGATCGATATATGGATTGC	1380

orf20_NK-198	TATATTCACAGATTCTATCCTTTGTCGCAGATAACATTAAATTTATGTTGTTTATGCAC	1440
orf20_NK-305-1	TATATTCACAGATTCTATCCTTTGTCGCAGATAACATTAAATTTATGTTGTTTATGCAC	1440

orf20_NK-198	ATTTGACACAATAAATTTGAGTTGTGGACTATAATATATATGTGAGTTAGGTAACATATG	1500
orf20_NK-305-1	ATTTGACACAATAAATTTGAGTTGTGGACTATAATATATATGTGAGTTAGGTAACATATG	1500

/ Exon 3		
orf20_NK-198	GTGTCAATTTACAGAGTTTGAAGATTATTCAGGCTAGAGCTCCACAATTACTGCCACGA	1560
orf20_NK-305-1	GTGTCAATTTACAGAGTTTGAAGATTATTCAGGCTAGAGCTCCACAATTACTGCCACGA	1560

orf20_NK-198	ACTATCTGCTTGTCCTTGTTGGATTGTTTTCTCGGTGTTTATTCTTTATTATGGTCGG	1620
orf20_NK-305-1	ACTATCTGCTTGTCCTTGTTGGATTGTTTTCTCGGTGTTTATTCTTTATTATGGTCGG	1620

orf20_NK-198	AAGGAAATAGAAGCAGATCACATTGGAGTGCTTCTGATGGCTTCTGCTGGATACGACCCG	1680
orf20_NK-305-1	AAGGAAATAGAAGCAGATCACATTGGAGTGCTTCTGATGGCTTCTGCTGGATACGACCCG	1680

orf20_NK-198	CGAGTTGCACCTCAAGTATATGACAAGCTTGCAAAGCCACTGGGCGACTGGAAGTGTTA	1740
orf20_NK-305-1	CGAGTTGCACCTCAAGTATATGACAAGCTTGCAAAGCCACTGGGCGACTGGAAGTGTTA	1740

orf20_NK-198	GCAACTCATCATTGCAAGAATGAGAGCAAAGTTGTTAGCTCGAGCTGATGTTATGAAG	1800
orf20_NK-305-1	GCAACTCATCATTGCAAGAATGAGAGCAAAGTTGTTAGCTCGAGCTGATGTTATGAAG	1800

orf20_NK-198	GAAGCAGATAAGATATACAATGAAGTTGTAGCAGGACGTGCAATTCAAGGTCTTCAGTAA	1860
orf20_NK-305-1	GAAGCAGATAAGATATACAATGAAGTTGTAGCAGGACGTGCAATTCAAGGTCTTCAGTAA	1860

orf20_NK-198	CCATTACCAACCAGCATCTTCTTTTAGCAGCTTCGCCTGTTTATGAATTGTGGTAATCA	1920
orf20_NK-305-1	CCATTACCAACCAGCATCTTCTTTTAGCAGCTTCGCCTGTTTATGAATTATGGTAATCA	1920

orf20_NK-198	AAATTAACAGCTCATCGATCATTATATTGTCGTTATATTCATCTGTTTGACAAAGTTT	1980
orf20_NK-305-1	AAATTAACAGCTCATGGATCATTATATTGTCGTTATATTCGTCGTTTGACAAAGTTT	1980

orf20_NK-198	AGAGGTTAATTTGAGTTGAGAACTTGCTATTAGTATGTTATCCACTTGTTAATTCAATCT	2040
orf20_NK-305-1	AGAGGTTAATTTGAGTTGAGAACTTGCTATTAGT-TGTTATCCACTTGTTAATTCAATCT	2039

orf20_NK-198	AATTTGGTTCCTATGTTAATCACTGCTCTTTTTCTCCTTCAAGAGATCCTCTCTATCTTT	2100
orf20_NK-305-1	AATTTGGTTCCTATGTTAATCACTGCTCTTTTTCTCCTTCAAGAGATCCTCTCTATCTTT	2099

orf20_NK-198	TTCTTTTTTAAAGCAGTAGTAGTAGATAGCAGAGTAGTGGAGATGAGTTTCTCTTTTAGG	2160
orf20_NK-305-1	TTCTTTTTTAAAGCAGTAGTAGTAGATAGCAGAGTAGTAGAGATGAGTTTCTCTTTTAGG	2159

orf20_NK-198	TAGATATTAGTATAAACCGGGAAAGGGTCACTCAAATACAAATTAATGTTTTCAAAAC	2220
orf20_NK-305-1	TAGATATTAGTATAAACCGGGAAAGGGTCACTCAAATACAAATTAATGTTTTCAAAAC	2219

orf20_NK-198	CATCCGTTTTCTCTTTGCCTCCCTAGCTAACTATCCGCAAGGCTATTTGCTCCTCTTC	2280
orf20_NK-305-1	CATCCGTTTTCCCTTTGCCTCCCTAGCTAACTATCCGCAAGGCTATTTGCTCCTCTTC	2279

orf20_NK-198	CTTCGTGAGTTAGTCTAACGCCTTCTTGATCTTGGAGCAACTCCCTGCAAGGATCAATCA	2340
orf20_NK-305-1	CTTCGTGAGTTAGTCTAACGCCTTCTTGATCT-GGAGCAACTCCCTGCAAGGATCAATCA	2338

orf20_NK-198	AATGAGAGAGGTAATCTTTTAAAGTATTTGAATTAATAGTTGAGAGTTGAGAGTTGTTAA	2400
orf20_NK-305-1	AATGAGAGAGGTAATCTTTTAAAGTATTTGAATTAATAGTTGAGAGTTGAGAGTTGTTAA	2398

orf20_NK-198	CCCTCAATCCAGGACCCAGGTACCCAAATGATCATGGCATACCATACCCAAGCAAGCAAG	2460
orf20_NK-305-1	CCCTTAATCCAGGACCCAGGTACCCAAATGATTATGGCATACCATACCCAAGCAAGCAAG	2458
**** *****		
orf20_NK-198	ATTAGTATCCAAGAAGAAAGATGCATCAGTGTTAATTTTTAGGTGGTTTAGCGGTGGAGG	2520
orf20_NK-305-1	ATTAGTATCCAAGAAGAAAGATGCATCAGCGTTAATTTTTAGGTGGTTTAGCGGTGGAGG	2518

orf20_NK-198	AATCCAAGCTGAGGGAACAACGGGATTGGGAGCTTTACCTACCTTAGGTTGGTTTTGGGT	2580
orf20_NK-305-1	AATCCAAGCTGAGGGAACAACGGGATTGGGAGCTTTACCTACCTTAGGTTGGTTTTGGGT	2578

orf20_NK-198	AAAGAACCATTCTGTGACGGTCCAAATGGAATTTTATAAGGACATCTGATGGTATATTT	2640
orf20_NK-305-1	AAAGAATCATTCTGTGACGGTCCAAATGGAATTTTATAAGGACATCTGATGGTATATTT	2638

orf20_NK-198	TTCTTTTGA AAAAGATGTTTTGTCCACATTTTGATCTCTTGATGGGTCTGTAACATAAT	2700
orf20_NK-305-1	TTCTTTTGA AAAAGATGTTTTGTCCACATTTTGATCTCTTGATGGGTCTGTAACATAAT	2698

orf20_NK-198	AATAAACCAATCAGAAATCTGCCGTTAGAATCTTTCTTCTAAGAACACTTTACGTGAT	2760
orf20_NK-305-1	AATAAGCCAAATCAGAAATATGCTGTTAGAATCTTTCTTCTAAGAACACTTTACGTGAT	2758

orf20_NK-198	TCAAGTTTCTATATTGATTTTAGTGTTTTAGACAAACATAACCTTGAATTTTCAAAAAGT	2820
orf20_NK-305-1	TCAAGTTTCTATATTGATTTTAGTGTTTTAGACAAACATAACCTTGAGTTTTTAAAAAGT	2818

orf20_NK-198	TTTT-----ATAAAAAAAATCACCTTATATAGACACTATTATTCTAATTGAAAAAACTA	2875
orf20_NK-305-1	TTTTTATAAAAAAAATCACCTTATATAGACACTATTATTCTAATTGAAAAAACTA	2878
	**** * *****	
orf20_NK-198	TTAGTAAATTA AAAATTTAGGGCCTATTTGGTATCACTCCCAATTTAACTAACATATCT	2935
orf20_NK-305-1	TTAGTAAATTA AAAATCTCAGGGCCTATTTG-TATCACTCCCAATTTAACTAACATATCT	2937

orf20_NK-198	AGGCGTTTGATACTTCTATCCTTGAAATAATGTTAGTAATTTAGAGTCGAAATTATACCA	2995
orf20_NK-305-1	AGGCGTTTGATACTTCTATCCTTGAAATAATGTTAGTAATTTAGAGTCGAAATTATACCA	2997

orf20_NK-198	TTTCAAAGTCAGCTTCTTCCTTAATTCTATATGACATTAGACCACCATATGAAAAATTTG	3055
orf20_NK-305-1	TTTCAAAGTCAGCTTCTTCCTTAATTCTATATGACATTAGACCACCATATGATAAATTTG	3057

orf20_NK-198	GCAATAATTTGGCAATATATAGCACTATGTTCTTTGAATCTTAGCTGTAGTGGTGCAATA	3115
orf20_NK-305-1	GCAATAATTTGGCAATATATAGCACTATGTTCTTTGAATCTTTGCTGTAGTGGTGCAATA	3117

orf20_NK-198	ATAACTAATCTTGCACCTCCTTTAAGACTAAGAAGTCAAATACCTTGCTTGATAGTTTGG	3175
orf20_NK-305-1	ATAACTAATCTTGCACCTCCTTTAAGACTAAGAAGTCAAATACCTTGCTTGATAGTTTGG	3177

orf20_NK-198	TAGATAACAATCGTATCATATTAGTTGGTAATACGATGATATCTCTTTATTATAAATAGC	3235
orf20_NK-305-1	GAAATAAAAATCATATCATATCAGTTAATAATACGATGATATCTCTTCATTATAAATAGC	3237
	* **** *	
orf20_NK-198	TAGTTTTGAGGTTCAAACTTTTCATCAATTATCTCTAATAATTAATTTCCACTTATAATA	3295
orf20_NK-305-1	TAGTTTTGAGGCTCAAACTTTTCATCAATTATCTCTA-ATTAACCTTCCACTTATAATA	3294

orf20_NK-198	TA-TGGAGAGAGCTTCAGTTAAGTTAGCTTTTCTAGTTGTCCTCGTAGCTTCTGCATCAT	3354
orf20_NK-305-1	TAATGGAGAGAGCTTCAGTTAAGTTAGCTTTTCTAGTTGTCCTCGTAGCTTCTGCATCAT	3354
	** *****	

```

orf20_NK-198      GTAAGTGTCTCTGATAATCTCCAAAATTATCTTTGTATTAGCAAAATTATTTTCTACTGT 3414
orf20_NK-305-1    GTAAGTGTCTCTGATAATCTCCAAAATTATCTTTGTATTGGCAAAATTATATTCTACTGT 3414
*****
orf20_NK-198      TTGAATCTCTAATATAATATACAATCTTAGCCACTAACTCATGACG- 3460
orf20_NK-305-1    TTGAATCTCTAATACTCCCTCCGTCCCATATTTTTTATTAAT-ACGG 3460
*****          * * *          * * * * *

```

Supplementary Fig 4 Alignment of nucleotide sequences of *orf20_{NK-198}* and *orf20_{NK-305-1}* coding and flanking regions. Asterisks denote matched residues. Exon/intron boundaries are shown by forward slashes. Stop codons are underlined. Nucleotides are numbered from the first residue of the initiation codon as +1.

```

/ Exon 1
orf20L_S      ATGGCGTGGTACAGAAATTC AAGGTTTGTCTACAATGCTTTAAAACTCAACTTGC GTTCC 60
orf20_NK-305-2 ATGGCGTGGTACAGAAATTC AAGGTTTGTCTACAATGCTTTAAAACTCAACTTGC GTTCC 60
*****

orf20L_S      AAAACATTTGGTACTATTCCA AACTCCAAGAGTTCATTCGAATTCCTCATCTTTGTTT TAC 120
orf20_NK-305-2 AAAACATTTGGTACTATTCCA AACTCCAAGAGTTCATTCGAATTCCTCATCTTTGTTT TAC 120
*****

orf20L_S      AATCAATCTACTAAGTGTAGT GGGTTATTTGGGTCTGCAAAATCTGGGTATTTTAATGG G 180
orf20_NK-305-2 AATCAATCTACTAAGTGTAGT GGGTTATTTGGGTCTGCAAAATCTGGGTATTTTAATGG G 180
*****

orf20L_S      TTAAACATCATCAAGAGATT A GCTCTTTCTCTGGTTTTGCAAGGAGAAATTATCATGG T 240
orf20_NK-305-2 TTAAACATCATCAAGAGATT A GCTCTTTCTCTGGTTTTGCAAGGAGAAATTATCATGG T 240
*****

orf20L_S      GATAAAACCGAAGTAAGTG TTTGAATCATGGCTGGAAAAATTACTTCTTGGAATTGCACTA 300
orf20_NK-305-2 GATAAAACCGAAGTAAGTG TTTGAATCATGGCTGGAAAAATTACTTCTTGGAATTGCACTA 300
*****

orf20L_S      ATGTTGAGTACTGGTATATT T GCTTACCGTCATGTGCACCCAGTAGTTGTGCCATATACA 360
orf20_NK-305-2 ATGTTGAGTACTGGTATATT T GCTTACCGTCATGTGCACCCAGTAGTTGTGCCATATACA 360
*****

orf20L_S      GGAAGGAAGCATTATGTGCT TATATCAACAACCTGATGAGAATGAAAAGGGAGAAGTTGAG 420
orf20_NK-305-2 GGAAGGAAGCATTATGTGCT TATATCAACAACCTGATGAGAATGAAAAGGGAGAAGTTGAG 420
*****

orf20L_S      AAGCGGAAAATACAACCTGCT ACACACCCTGATACTGATAGGGTTAGGTCAATATTCCA A 480
orf20_NK-305-2 AAGCGGAAAATACAACCTGCT ACACACCCTGATACTGATAGGGTTAGGTCAATATTCCA A 480
*****

orf20L_S      CACATTCTTGAATCACTGGA AAGAGAGATTAATCACCATGAACTCGAACTCGAAAGAGAT 540
orf20_NK-305-2 CACATTCTTGAATCACTGGA AAGAGAGATTAATCACCATGAACTCGAACTCGAAAGAGAT 540
*****

orf20L_S      GAAACTTTCAAGGAGATAACC ATTTGGAAGGAGGAGACAGTTGATGATAAAGATAGTAGG 600
orf20_NK-305-2 GAAACTTTCAAGGAGATAACC ATTTGGAAGGAGGAGACAGTTGATGATAAAGATAGTAGG 600
*****

orf20L_S      AAGAAGCATAGTGGGGCTAAG ATAACTACTAACCATTTGGAAGGGTTGAATTGGGAAATT 660
orf20_NK-305-2 AAGAAGCATAGTGGGGCTAAG ATAACTACTAACCATTTGGAAGGGTTGAATTGGGAAATT 660
*****

orf20L_S      TTCGTTGTTGATAAACCGTTG GTTGAGTCCAGTTGTTTATTTGGTGGAAGATTGTTGTT 720
orf20_NK-305-2 TTCGTTGTTGATAAACCGTTG GTTGAGTCCAGTTGTTTATTTGGTGGAAGATTGTTGTT 720
*****

orf20L_S      TACACGGGATTGCTCAACCAT T GCAACTCTGATGCTGAATTGGCTACAATTATCGCGCAT 780

```

```

orf20_NK-305-2    TACACCGGATTGCTCAACCATTGCAACTCTGATGCTGAATTGGCTACAATTATCGCGCAT 780
*****
                  / Intron 1
orf20L_S          CAGGTATATAAACTATTCCTGGGACTCCAATTATGTGCTTAAGCTGATGGTTAATAGAA 840
orf20_NK-305-2    CAGGTATATAAACTATTCCTGGGACTCCAATTATGTGCTTAAGCTGATGGTTAATAGAA 840
*****

orf20L_S          CTATACAAAAACATGAAAGGAATTATATGAAAAAACTGATGAATTTTAGGTTATCAG 900
orf20_NK-305-2    CTATACAAAAACATGAAAGGAATTATATGAAAAAACTGATGAATTTTAGGTTATCAG 900
*****
                  / Exon 2
orf20L_S          ATTACATTATGAATGTCATATGTCAATGTGGTGGTATGATTTGTTAGGTTGGGCATGCT 960
orf20_NK-305-2    ATTACATTATGAATGTCATATGTCAATGTGGTGGTATGATTTGTTAGGTTGGGCATGCT 960
*****

orf20L_S          GTGGCTCGACATCAGGCAGAGGATCGGACAGCATTCTTCTGGTGGTCAATGTCCCTCTAC 1020
orf20_NK-305-2    GTGGCTCGACATCAGGCAGAGGATCGGACAGCATTCTTCTGGTGGTCAATGTCCCTCTAC 1020
*****

orf20L_S          GTGATAATATTTGAAGTTCTATTTACTGCGCGTAAATTTGCCAATGCAAGATCAAACTA 1080
orf20_NK-305-2    GTGATAATATTTGAAGTTCTATTTACTGCGCGTAAATTTGCCAATGCAAGATCAAACTA 1080
*****
                  / Intron 2
orf20L_S          CTCTTAAGGCATCCTCTCTTGCAAAAGTAAGTCTCTTACTCTTAAATGTTTTCTTGATG 1140
orf20_NK-305-2    CTCTTAAGGCATCCTCTCTTGCAAAAGTAAGTCTCTTACTCTTAAATGTTTTCTTGATG 1140
*****

orf20L_S          ATTAACAAACATGTGGTACTGCTACTGCATAACTGTGTTACTGCATCACATATGTTACTG 1200
orf20_NK-305-2    ATTAACAAACATGTGGTACTGCTACTGCATAACTGTGTTACTGCATCACATATGTTACTG 1200
*****

orf20L_S          CATAATTGCAAAACATATCACATTGCCCGGACCTAGTAACTTGTTTCATTTGTCAGCGAT 1260
orf20_NK-305-2    CATAATTGCAAAACATATCACATTGCCCGGACCTAGTCACTTGTTTCATTTGTCAGCGAT 1260
*****

orf20L_S          TTCATTTAGACATCCATTTGAAAGCAAGTTAAATTTGTATCAAGTTGTGGAATGGAAAAG 1320
orf20_NK-305-2    TTCATTTAGACATCCATTTGAAAGCAAGTTAAATTTGTATCAAGTTGTGGAATGGAAAAG 1320
*****

orf20L_S          TAATAGAACTAAATAGAGAGGTGTGATGCTAACAAAATCTAATCCATTACTGAGTAATGG 1380
orf20_NK-305-2    TAATAGAACTAAATAGAGAGGTGTGATGCTAACAAAATCTAATCCATTACTGAGTAATGG 1380
*****

orf20L_S          TTTTGGATCAATATATGGATTGCTATATTCCACATATTCTATACTTTGCCGCAGATAACA 1440
orf20_NK-305-2    TTTTGGATCGATATATGGATTGCTATATTCCACAGATTCTATCCTTTGTCGCAGATAACA 1440
*****

orf20L_S          TTAAATTGATGTTGTTTATTTACATTTGACACATTAATTAATTTGAGTTGTGGACTATAATAT 1500
orf20_NK-305-2    TTAAATTTATGTTGTTTATGCACATTTGACACAATAATTTGAGTTGTGGACTATAATAT 1500
*****
                  / Exon 3
orf20L_S          ATATGCGAGTTAGGTAACATAGAGTGTCAATTTACAGGGTTTGAAGATTATTCAGGCTA 1560

```

orf20_NK-305-2 ATATGTGAGTTAGGTAACATATGGTGTCAATTTACAGAGTTTGAAGATTATTCAGGCTA 1560

orf20L_S GATTTTCATCAATTACTGCCACGAACTACCTTGCACTTGGGCTTTCTTGGATTGTCTTCCT 1620
 orf20_NK-305-2 GAGCTCCACAATTACTGCCACGAACTATCT---GCTTGTCCCTTGTGGATTGTTTTCT 1617
 ** ** ***** ** **** * ** *****

orf20L_S TGGTGTTTATTCTTTATTTTGGTCGGAAGGAAATAGAAGCAGATCACATTGGAGTGCTTC 1680
 orf20_NK-305-2 CGGTGTTTATTCTTTATTATGGTCGGAAGGAAATAGAAGCAGATCACATTGGAGTGCTTC 1677

orf20L_S TGATGGCTTCTGCTGGATACGACCCGCGAGTTGCACCTCAAGTATATGACAAGCTTGCAA 1740
 orf20_NK-305-2 TGATGGCTTCTGCTGGATACGACCCGCGAGTTGCACCTCAAGTATATGACAAGCTTGCAA 1737

orf20L_S AGCCACTGGGCGACTGGAAGTGTGTTAGCAACTCATCCATTTGCAAGAATGAGAGCAAAGT 1800
 orf20_NK-305-2 AGCCACTGGGCGACTGGAAGTGTGTTAGCAACTCATCCATTTGCAAGAATGAGAGCAAAGT 1797

orf20L_S TGTTAGCTCGAGCTGATGTTATGAAGGAAGCAGATAAGATATACAATGAAGTTGTAGCAG 1860
 orf20_NK-305-2 TGTTAGCTCGAGCTGATGTTATGAAGGAAGCAGATAAGATATACAATGAAGTTGTAGCAG 1857

orf20L_S GACGTGCAATTCAAGGTCTTCAGTAA 1886
 orf20_NK-305-2 GACGTGCAATTCAAGGTCTTCAGTAA 1883

orf20L_S CCATTTACCAACCAGCATCTTCTTTTAGCAGCTTCGCCTGTTTATGAATTATGGTAATCA 1946
 orf20_NK-305-2 CCATTTACCAACCAGCATCTTCTTTTAGCAGCTTCGCCTGTTTATGAATTATGGTAATCA 1943

orf20L_S AAATTAACAGCTCATGGATCATTTTATTGTCGTTATATTTTCGTCTGTTTGACAAAGTTT 2006
 orf20_NK-305-2 AAATTAACAGCTCATGGATCATTTTATTGTCGTTATATTTTCGTCTGTTTGACAAAGTTT 2003

orf20L_S AGAGGTTAATTTGAGTTGAAAACCTTGCTATTAGTATGTTATCCTATCCCACTGACCTTGT 2066
 orf20_NK-305-2 AGAGGTTAATTTGAGTTGAAAACCTTGCTATTAGTATGTTATCCTATCCCACTGGCCTTGT 2063

orf20L_S TAATTCAATCTAATTTAGTTCCTATGTTAATTCAGTCTTTTTCTCCTTTAATAGATC 2126
 orf20_NK-305-2 TAATTCAATCTAATTTAGTTCCTATGTTAATTCAGTCTTTTTCTCCTTTAATAGATC 2123

orf20L_S CTCTCTATCCGTTTTTCTTTTTCTTTTTGAGAAAAGGAGCATCATAGATAAATATAAAT 2186
 orf20_NK-305-2 CTCTCTATCCGTTTTTCTTTTTCTTTTTGAGAAAAGGAGCATCATAGATAAATATAAAT 2183

orf20L_S CGAAGTCATTACAACTAACTAGAGCTTCCAATATAAGTCCCTGGAAGTCCGAAAGCAA 2246
 orf20_NK-305-2 CGAAGTCATTACAACTAACTAGAGCTTCCAATATAAGTCCCTGGAAGTCCGAAAGCAA 2243

orf20L_S	TAGTAGTAGATAGCATAGTAGAGATGAGTTTCTCTTTTAGGTAGATAGTAGATAGGAGGA	2306
orf20_NK-305-2	TAGTAGTAGATAGCATAGTAGAGATGAGTTTCTCTTTTAGGTAGATAGTAGATAGGAGGA	2303

orf20L_S	TCGATAGATGACCCACCATTGTGCGCGCTTTCGGCGCCACCAACACCAGGGGAAGAAGAG	2366
orf20_NK-305-2	TCGATAGATGACCCACCATTGTGCGCGCTTTCGGCGCCACCAACACCAGGGGAAGAAGAG	2363

orf20L_S	CCTCCAATCTGGCCACCACCGGATATAGGAATACCAAGCATAGGTTTGCCACCTGGATAT	2426
orf20_NK-305-2	CCTCCAATCTGGCCACCACCGGATATAGGAATACCAAGCATAGATTGTCACCTGGATAT	2423
***** *****		
orf20L_S	CAATTCACCTCCTTAACGTCCATGGATGTGATTACCCCATGCATGGATAAACGAGACTGTG	2486
orf20_NK-305-2	CAATTCACCTCCTTAACGTCCATGGATGTGATTACCCCATGCATGGATAAACGAGACTGTG	2483

orf20L_S	ATTGGCCCTGCATGTTTCGTTGTGGCAAATGCAATGGCACTTGTTGCCTCAGTAAAATGA	2546
orf20_NK-305-2	ATTGGCCCTGCATGTTTCGTTGTGGCAAATGCAATGGCACTTGTTGCCTCAGTAAAATGA	2543
***** *****		
orf20L_S	TTTATGAGTAGACACATGACATATCTTCCCTTATATTATTGTTTCATTTTGAGTTAGTTAT	2606
orf20_NK-305-2	TTTATGAGTAGACACATGAGATATCTTCCCTTATATTATTGTTTCATTTTGAGTTAGTTAT	2603
***** *****		
orf20L_S	ACTCCCTCCTCTTTTTTCTTAGTTGCTATATTCCATTTTGGATACAAAATCACATGAGA	2666
orf20_NK-305-2	ACTCCCTCCTCTTTTTTCTTAGTTGCTATATTCCATTTTGGATACAAAATCACATGAGA	2663

orf20L_S	ATTTTGACTTTCTTTAATTTTATATATGTAAGAAAAAAACATAGTATTTTATTAGATTT	2726
orf20_NK-305-2	ATTTTGACTTTCTTTAATTTTATATATGTAAGAAAAAAACATAGTATTTTATTAGATTT	2723

orf20L_S	CTCTCAAATGTGTAATTTTCATATATAGTTTTTTT-ATAATTTTCTCGTATACATAACTC	2785
orf20_NK-305-2	CTCTCAAATGTGTAATTTTCATATATAGTTTTTTTATAATTTTCTCGTATACATAACTC	2783
***** *****		
orf20L_S	AACATATTAAGGTTTGAAGTCATGACTGGGAGCAAGAAACCACGCGGGTGGGGCTCACAG	2845
orf20_NK-305-2	AACATATTAAGGTTTGAAGTCATGACTGCGAGCAAGAAACCACGCGGGTGGGGCTCACAG	2843
***** *****		
orf20L_S	CAGCGCCAGCAGGAAGGACGAGTGAAAAAAATTAGGGTGAATTAGTGTGGCTCTGATACC	2905
orf20_NK-305-2	CAGCGCCAGCAGGAAGGACGAGTGAAAAAAATTAGGGTGAATTAGTGTGGCTCTGATACC	2903

orf20L_S	ATGACAAATTGATTAGAGTTTAATGATTTGTGTTACTACTGTTAACTTAGAGTTTACAT	2965
orf20_NK-305-2	ATGACAAATTGATTAGAGTTTAATGATTTGTGTTACTATTGTTAACTTAGAGTTTATAT	2963
***** ***** **		
orf20L_S	ATATCCTAAGTTTACATCCAAGGCCTTAGGCCAATATATACAACCTCAACATATATACAA	3025

orf20_NK-305-2	ATATCCTAAGTTTACATCCAAGGCCTTAGGCCCAATATATACAACCTCAACATATATACAA 3023

orf20L_S	GGTCCAATATGTATCAACAAAATGAAAAATAAAACATAATGTAGCAACTAGCAACTAAA 3085
orf20_NK-305-2	GGTCCAATATGTATCAACAAAATGAAAAATAAAACATAATGTAGCAACTAGCAACTAAA 3083

orf20L_S	AAAAGACGGAGAGAGTATTTAAGTTAACAATATAATGATAATTAATCAGTGTAAGTATT 3145
orf20_NK-305-2	AAAAGACGGAGAGAGTATTTAATTTAACAATATAATGATAATTAATCAGTGTAAGTATT 3143
	***** *****
orf20L_S	AGTCAAGATCTAGATTGAAATTACTCTCCGACGACCTCCCGACCCATTTGCCTTTATGT 3205
orf20_NK-305-2	AGTCAAGATCTAGATTGAAATTACTCTCCGACGACCTCCCGACCCATTTGCCTTTATGT 3203

orf20L_S	AATCAATCATTATATACAACCTCCGATTCTTTTGATGGCAACTATTTGAATGTTACCTTT 3265
orf20_NK-305-2	AATCAATCATTATATACAACCTCCGATTCTTTTGATGGCAACTATTTGAATGTTACCTTT 3263

orf20L_S	ATCAACATGAAACTTTAACTATTAATATCTTTATTAATTATTAACATAATTTACTACTAA 3325
orf20_NK-305-2	ATCAACATGAAACTTTAACTATTAATATCTTTATTAATTATTAACATAATTTACTACTAA 3323

orf20L_S	TTATAAATTATTTAAATATATACCTTGAGATTAATTTTCCTAACATTTATTTACTCTA 3385
orf20_NK-305-2	TTATAAATTATTTAAATATATACCTTGAGATTAATTTTCCTAACATTTATTTACTCTA 3383

orf20L_S	ACATTTGCTTTTATATTTTATAAAGAAAATGGTGATGGTTAGAAGTATATATATATGAGT 3445
orf20_NK-305-2	ACATTTGCTTTTATATTTTATAAAGAAAATGGTGATGGTTAGAAGTATATATATATGAGT 3443

orf20L_S	AGTTC AATAGTCAAAGCCTTGCAAATTAAGAAACGGCGGAAGAGTATATTGAAAATTTA 3505
orf20_NK-305-2	AGTTC AATAGTCAAAGCCTTGCAAATTAAGAAACGGCGGAAGAGTATATTGAAAATTTA 3503

orf20L_S	TATTGTCACCTTTTTTAATAGATGAGTCGATGATGGTCTAAAATTGAGGCTTCTAATATG 3565
orf20_NK-305-2	TATTGTCACCTTTTTTAATAGATGAGTCGATGATGGTCTAAAATTGAGGCTTCTAATATG 3563

orf20L_S	CGCTAGCTTAGATTAAAGAGAACAGGAAATTGTTATGACAATTTGCGTAAGTTGACTGAG 3625
orf20_NK-305-2	CGCTAGCTTAGATTAAAGAGAACAGGAAATTGTTATGACAATTTGCGTAAGTTGACTGAG 3623

orf20L_S	ATCATCATGAAGTCGAGTATTAGCAACCTTAGTAATTTATAATTCCTCCGTTTCTTTTTA 3685
orf20_NK-305-2	ATCATCATGAAGTCGAGTATTAGCAACCTTAGTAATTTATAATTCCTCCGTTTCTTTTTA 3683

orf20L_S	ATTTACTCATTTTATTTTGGGCGAGAATCAAGGAAGAAGTACAAAAGTATGAGACAAAAA 3745
orf20_NK-305-2	ATTTACTCATTTTATTTTGGGCGAGAATCAAGGAAGAAGTACAAAAGTATGAGACAAAAA 3743

orf20L_S	CTGAAAAGATAAAGAGAAAAACTGAAAAGTGCGAGAAATAAAGTAAAAAAGTGGGTTAAA 3805

orf20_NK-305-2	CTGAAAAGATAAAGAGAAAACTGAAAAGTGCAGAGAAATAAGTAAAAAAGTGGGTAA 3803

orf20L_S	AATATTA AAAAGTGGATCACATGGGTAGAAAAGGAGGAGAAAATAATGTGACATGTTAC 3865
orf20_NK-305-2	AATATTA AAAAGTGGATCACATGGGTAGAAAAGGAGGAGAAAATAGATGTGACATGTTAC 3863

orf20L_S	CAAAAATAGAAAAAGCAGCAGTGTGCAGTGTGCAGATTA AAAAGAAACAAACCAAAAAA 3925
orf20_NK-305-2	CAAAAATAGAAAAAGCAGCAGTGTGCAGTGTGCAGATTA AAAAGAAACAAACCAAAAAA 3923

orf20L_S	AAAATGTATAAATTA AAAAGAAATGGAGGGAGTAATACTTAACACATAAATATTGTTTAT 3985
orf20_NK-305-2	AAAATGTATAAATTA AAAAGAAATGGAGGGAGTAATACTTAACACATAAATATTGTTTAC 3983

orf20L_S	TAATTGAAGCTCGAGCTTCACTCAGATACATGATTTTTAATTTTTTATGACCTATCATAG 4045
orf20_NK-305-2	TAATTGAAGCTCGAGCTTCACTCAGATACATGATTTTTAATTTTTTATGACCTATCATAG 4043

orf20L_S	TAAGTTTGCGGTGCTACTTTGATCCTTCCTTTATGATAAGACACATCCCTTTATTCTCTA 4105
orf20_NK-305-2	TAAGTTTGCGGTGCTACTTTGATCCTTCCTTTATGATAAGACACATCCCTTTATTCTCTA 4103

orf20L_S	AGAGAATCACTTAAATTTATAAGGTATAAAGTTAACGCTATAAGTTAAAGAGGTCATTTT 4165
orf20_NK-305-2	AGAGAATCACTTAAATTTATAAGGTATAAAGTTAACGCTATAAGTTAAAGAGGTCATTTT 4163

orf20L_S	TAGATAATAGGTCACCTTGATGTAATAGGTCACCTTCTACTAAAAATTTCCCGTATAG 4225
orf20_NK-305-2	TAGATAATAGGTCACCTTGATGTAATAGGTCACCTTCTACTAAAAATTTCCCGTATAG 4223

orf20L_S	CAAGTTACAATTGTGTTAACCCAAATTATGGTAACATTTTGTACTTTTATCAACTCCGAC 4285
orf20_NK-305-2	CAAGTTACAATTGTGTTAACCCAAATTATGGTAACATTTTGTACTTTTATCAACTCCGAC 4283

orf20L_S	TCATACTTGATATTGAATAACCTCACTTATCAAGTCAATGCAAGCAAATCATGTTAAAC 4345
orf20_NK-305-2	TCATACTTGATATTGAATAACCTCACTTATCAAGTCAATGCAAGCAAATCATGTTAAAC 4343

orf20L_S	TTTTGTTGCGCTTAATGTCGAGACCAGTCTATCTAAAAGAACTTTTTTTTTTTTACAAT 4405
orf20_NK-305-2	TTTTGTTGCGCTTAATGTCGAGACCAGTCTATCTAAAAGAACTTTTTTTTTTTT--ACAAT 4401

orf20L_S	TACCTAAAAGAACTTATTATTAGTTATTGGCTACAAAACCACACAAATATTCAATACAAT 4465
orf20_NK-305-2	TACCTAAAAGAACTTATTATTAGTTATTGGCTACAAAACCACACAAATATTCAATACAAT 4461

orf20L_S	CTCAAGCTACAACCTCAAAGATTAAAGATTCCCTCACCATCACATTATATACAATATA 4525
orf20_NK-305-2	CTCAAGCTACAACCTCAAAGATTAAAGATTCCCTCACCATCACATTATATACAATATA 4521

orf20L_S	TTGGTCTATCAGTGATATGATTGAATAAGTTGCACTCAATTGTATTGACCTACATAAGTA 4585

orf20_NK-305-2	TTGGTCTATCAGTGATATGATTGAATAAGTTGCACTCAATTGTATTGACCTACATAAGTA 4581

orf20L_S	TGGCCTTTATAATTACATAAAGGTGATCCAAAGTTCCAATAATATCAATGAGATTTGAAG 4645
orf20_NK-305-2	TGGCCTTTATAATTACATAAAGGTGATCCAAAGTTCCAATAATATCAATGAGATTTGAAG 4641

orf20L_S	ATTGTTTCCAGATATTCACTCAACCAAATATACCTTACGACAACCTTCATCCTGAAATCTT 4705
orf20_NK-305-2	ATTGTTTCCAGATATTCACTCAACCAAATATACCTTACGACAACCTTCATCCTGAAATCTT 4701

orf20L_S	CGGTTTGTGACAGTACTCCACTTTTAATTCAATGGTCAAATATTTATACCTTAACCCCTTA 4765
orf20_NK-305-2	CGGTTTGTGACAGTACTCCACTTTTAATTCAATGGTCAAATATTTATACCTTAACCCCTTA 4761

orf20L_S	ATTCAATGCCATAATGTTGATGATCATGTGAAGTAGTGCAATATATAGCTTTACTTCAT 4825
orf20_NK-305-2	ATTCAATGCCATAATGTTGATGATCATGTGAAGTAGTGCAATATATAGCTTTACTTCAT 4821

orf20L_S	CTAAAAATGCATTAATAAATTTAGTCAATTGATTGAATATAAACTAGCTAACTCGTGTAC 4885
orf20_NK-305-2	CTAAAAATGCATTAATAAATTTAGTCAATTGATTGAATATAAACTAGCTAACTCGTGTAC 4881

orf20L_S	TACTCAACCAAGATGGTTTCATGTCAATGATAAATAGCTCATTGTTTCTGAACCTCATAA 4945
orf20_NK-305-2	TACTCAACCAAGATGGTTTATGTCAATGATAAATAGGTCAATTGTTTCTGAACCTCATAA 4941
	***** *****
orf20L_S	ATTTTCCATATCATCTCCTTCATCACACCATGAAGGGAGCTTCTTTCAAGCTGTCGATT 5005
orf20_NK-305-2	ATTTTCCATATCATATCCTTCATCACACCATGAAGGGAGCTTCTTTCAAGCTGTCGATT 5001

orf20L_S	TGATTTTCGTCCTTGTGGCTTTGCTTCATGTAAGTGTCTTTATCTCTTATGATTCTTG 5065
orf20_NK-305-2	TGATTTTCGTCCTTGTGGCTTTGCTTCATGTAAGTGTCTTTATCTCTTATGATTCTTG 5061
	***** *****
orf20L_S	CATGTGAATTAAGATGCATATATGGTTGTCTTGA-----TATAGTTTCTTCACCTAA 5118
orf20_NK-305-2	CATGTGAATTAAGATGCATACATGGTTGTCTTGACTCTTGATATAGATTCTTCACCTAA 5121
	***** *****
orf20L_S	ACAGCTCTTCTAGTACTATTATCGATTGATGGAGTTTATACTATATTGCAGTTCTTTGTC 5178
orf20_NK-305-2	ATAGCTCTTCTAGTATTATTATTGATTGGTGGGGGCATACTATATTGCAGTTCTTTGTC 5181
	* *****
orf20L_S	CATTTACAGCAGAAGCAAGACACAAGCACGTGCATCATAGTAATTGCCACCATCGTCGTG 5238
orf20_NK-305-2	CATTTACAGCAGAAGCAAGACACAAGCACGTGCATCATAGTAATTGCCACCATCGTCGTG 5241

orf20L_S	GACGTCCTGATCCAGCACCAACACCGGCGCCTGTGCCAGCATCTAATGAAGGCATACCAC 5298
orf20_NK-305-2	GACGTCCTGATCCAGCACCAACACCGGCGCCTGTGCCAGCATCTAATGAAGGCATACCAC 5301

orf20L_S	CATATCAAATCCATGGTTGTGGTTACCCATGTAGTGACTCCAACGACTGTGATTGGCCTT	5358
orf20_NK-305-2	CATATCAAATCCATGGTTGTGGTTACCCATGTAGTGACTCGAACGACTGTGATTGGCCTT	5361

orf20L_S	GTACAGAATGCGGTGTCAACAGAACTTGTGCTTATGAAGAGCCCTTCTTTCCATCACCA-	5417
orf20_NK-305-2	GTACAGAATGCGGTGTCAACAGAACTTGTGCTTATGAAGAGCCCTTCTTTCCATCACCAT	5421

orf20L_S	-----AGTATGGAACCCCAACACAAGAATCACCCTATCACCAGCAC	5460
orf20_NK-305-2	CACCTGTTCCATACCAAGTATGGAACGCCAACACAAGAACCACCCTATCACCAGCAC	5481

orf20L_S	CGGCACCTGATAATGGCATTGGCGTAGGTGTGCCACCATATCAAATCCACGGTTGTGGCT	5520
orf20_NK-305-2	CGGCACCTGATAATGGCATTGGCGTAGGTGTGCCACCATATCAAATCCACGGTTGTGGCT	5541

orf20L_S	ACCCATGTAGTGACTCCAATGATTGTGATTGGCCTTGACAGTATGTGGTGTCAACGGAA	5580
orf20_NK-305-2	ACCCATGTAGTGACTCCAATGATTGTGATTGGCCTTGACAGTATGTGGTGTCAACGGAA	5601

orf20L_S	CTTGTGCTTTTGAAGAGCCCTTCTTTCCATCATCATCACCTGTTCCATCACCAAGTATGA	5640
orf20_NK-305-2	CTTGTGCTTTTGAAGAGCCATTCTTTCCATCATCA---CCTGTTCCATCACCAAGTATAG	5658

orf20L_S	AAACCCCAACACAAGAACCACCCTATCACCAGCACGGCACCTGATAATGGCATTGGCG	5700
orf20_NK-305-2	AAACTCCAACACAAGAACCACCCTATCACCAGCACGGCACCTGATAATGGCATTGGCG	5718
	**** *****	
orf20L_S	TAGGTGTGCCACCATATCAAATCAGGTTGTGGCTACCCATGTAGTGACTCCAACGATT	5760
orf20_NK-305-2	TAGGTGTGCCACCATATCAAATCAGGTTGTGGCTACCCATGTAGTGACTCCAACGATT	5778

orf20L_S	GTGATTGGCCTTGTACAATCTGTGGTGTGGCCAAACCTGTACTTTTGATGAGCCCTTCT	5820
orf20_NK-305-2	GTGATTGGCCTTGTACAATATGTGGTGTGACCAAACCTGTACTTTTGATGAGCCCTTCT	5838

orf20L_S	TTACATCACCTCGCTTGCTCCATTACCTCCACAGAAGCACCCATTCCACAATGGGTAC	5880
orf20_NK-305-2	TTACATCACCTCGCTTGCTCCATTACCTCCACAGAAGCACCCATTCCACAATGGGTAC	5898

orf20L_S	CTGGTAATGGCATCGCGGCACCACCATATCAGATTCATGGTTGTGGTTACCCATGTAACG	5940
orf20_NK-305-2	CTGGTAATGGCATCGCGGCACCACCATATCAGATTCATGGTTGTGGTTACCCATGTAACG	5958

orf20L_S	ACTCCAATGACTGTGATGCACCCTGTACAGTCTGTTGTGCAAACCTATACCTGCTGTTATG	6000
orf20_NK-305-2	ACTCCAATGACTGTGATGCACCCTGTACAGTCTGTTGTGCAAACCTATACCTGCTGTTATG	6018

orf20L_S	ATGTGGCTGATCCTGAGTACATGCTACCACCTATGTACCCTCTGAGCCACCGAAATTAT	6060

orf20_NK-305-2	ATGTGGCTGATCCTGAGTACATGCTACCACCTATGTCACCCTCTGAGCCAC-GAAATTAT 6077 *****
orf20L_S orf20_NK-305-2	TACCTCTTCCACCATCTCCACCTCCATCTACAGAGGATGTAGAATATGATGATATGTTTG 6120 TACCTCTTCCACCATCTCCACCTCCATCTACAGAGGATGTAGAAAATGATGATATGTTTG 6137 *****
orf20L_S orf20_NK-305-2	CACCACAACCTGCTTATGATATAGGAACGCCTGCGGAACTTCCACCACCAGGATACGAAT 6180 CACCACAACCTGCTTATGATATAGGAACGCCTGCGGAACTTCCACCACCAGGATACGAAT 6197 *****
orf20L_S orf20_NK-305-2	TCCCGCCGTATCAAATCCATGGTTGTGGCTATGGCCCCTGCATGGACTCCAACGACTGCG 6240 TCCCGCCGTATCAAATCCATGGTTGTGGCTATGGCCCCTGCATGGACTCCAACGACTGCG 6257 *****
orf20L_S orf20_NK-305-2	ATTGGCCCTGCACATCCTGCTGCTCTAATCATACATGTTGCTATGAGGAGCCTATGTTTC 6300 ATTGGCCCTGCACATCCTGCTGCTCTAATCATACATGTTGCTATGAGGAGCCTATGTTTC 6317 *****
orf20L_S orf20_NK-305-2	GATGAAAATCCTCAAACACACAAATCCAAAGAAAAAGTATAGTAACACAATGTAATAAA 6360 GATGAAAATCCTCAAACACACAAATCCAAAGAAAA-GTATAGTAACACAATGTAATAAA 6376 *****
orf20L_S orf20_NK-305-2	ACTTA-GTGTTCTGTATTACTTTAAATCACATTTGACCTATATTCTTCAATTGCTATGT 6419 ACTAAAGTGTTCCT-TATTACTTAAATCACATTTGACCTATATTCTTCAATTGCTATGT 6435 *** * ***** ***** *****
orf20L_S orf20_NK-305-2	CATTGTCTAATGATTGAAAGCAAGTACTTTTATTTCTGTGTCATACAAATGTAAGCAAGA 6479 CATTGTCTGTAATTGCAAGCAAGTACTTCTATTCTGTGTCATACAAATGTAAGCAAAA 6495 ***** * **** ***** *****
orf20L_S orf20_NK-305-2	TCAATAAAGAATATATACAACCTACGTTAATCAATTGCTACTATCAAACCTGCATTTTCTGT 6539 TTAATAAAGAATATATACAACCTACGTTAATCAATTGCTACTATCAAATTGCATTTTCTAT 6555 * ***** *
orf20L_S orf20_NK-305-2	CAAACAACAGTATAATTATTGTGAATTTGTGACGCACTCTAGTAAATAAAGTGGGTTGAA 6599 CAAACAACAGTATAATTA-----TTGTGACGTA CTCTAATAAATAAAGTGGGTTGAA 6607 ***** ***** *****
orf20L_S orf20_NK-305-2	GTAAGCATTAAATGACTGATTTCATCAGCTCTACGCCTATATCACAATTTAGAGTAGATA 6659 CTAAGCATTAAATGACTGATTTCATCAGCTCTACGCCTATATCACAATTTAGAGTAGATA 6667 *****
orf20L_S orf20_NK-305-2	CATTGCGAGATTTCTGTTGTAAGAACCATTTTGATCATGTTTCTGTAACAG-CAACATGA 6718 CATTGCGAGATTTCTGTTGTAAGAACCATTTTGATCATGTTGCTGTTACAGACAACATGA 6727 ***** **** *
orf20L_S orf20_NK-305-2	TTTCCAGTATATTTTGTGTTACAAATTTCTTCGTCGAGGAAATTGTGACAT-----TATTA 6774 TTTTCAGTAGATTTTGTGTTATAAATTTCTTCGTCGAGGAAATTGTGACATGTATTATTA 6787 *** ***** ***** ***** *****
orf20L_S	AGATACCAACT-----CTA----- 6788

orf20_NK-305-2	AGATACCAACTTAAATTCTGGTATAGGTTCCATTCTAATTAATATCATCGGCATACTCCT	6847
	*****	***
orf20L_S	-----	6788
orf20_NK-305-2	CAACTTAGATTCACTCTTAGTTGATTTACATCTTAATTCGTCAACTTTCATAGAGAACT	6907
orf20L_S	-----	6788
orf20_NK-305-2	CATCACAGGATAAGCCAGTTTAGTTCAACAATAGTATATTCCTCCGCCAAAATCCACATA	6967
orf20L_S	-----	6788
orf20_NK-305-2	ATGTTTTGAGAGTTTTAATTGCAAGTGGTATACCATCAAGCTTTTTCAGTTTTGACTTTG	7027
orf20L_S	-----	6788
orf20_NK-305-2	ATTAATCAAGATCAATATAAATCCTTTCCTCGAAAAGAAAAACGATATAAATCCTAGCT	7087
orf20L_S	-----	6788
orf20_NK-305-2	CCAATTCGTAAAAGAATTTTACCAAAGTACTTCCTCCGTTTCGTTTCAAATGCAACAAA	7147
orf20L_S	-----	6788
orf20_NK-305-2	GGGTATTATTTGTGAGATATAAAATTTCCAATTGTTGCGTTTAAAACGAGATGGAGGAAG	7207
orf20L_S	CGTTAAAAGTTGTAAGACAACAAAATGAAAACAAATGGATAAATTTCATAATATATTCAA	6848
orf20_NK-305-2	CATTAAGTTGTAAGACAACAAAATGAAAACAAATGGATAAATTTCATAATATATTCAA	7267
	* *****	
orf20L_S	CACTACCTTCTTTTGCTCAACATCACATAAACACATCCGCACCTGACAATCATTATTCAT	6908
orf20_NK-305-2	CTCTACCTTCTTTTGCTCAACATCACATAAACACATCCGCACCTGACAATCATTATTCAT	7327
	* *****	
orf20L_S	TAAACACCCGATAAAATAAATACCAGTTCTGAAATCATGGATATGAAATTCACAGTAAA	6968
orf20_NK-305-2	TAAACACCTGATAAAATAAATACCAGTTCTGAAATCATGGATATGAAACTCACAGTAAA	7387
	***** *****	
orf20L_S	AAAAAGTATTAAAAACACCAGCTGAGGCAAACACAAGATCAACAAAGAGGTAAAAATA	7028
orf20_NK-305-2	AAGA-GTATAAAAAACACCAGCTGAAGCAA-----	7418
	** * **** *****	
orf20L_S	AAAAATGGTCG	7039
orf20_NK-305-2	-----	

Supplementary Fig 5 Alignment of nucleotide sequences of *orf20L_S* and *orf20_{NK-305-2}* coding and flanking regions. Asterisks denote matched residues. Exon/intron boundaries are shown by forward slashes. Stop codons are underlined. Nucleotides are numbered from the first residue of the initiation codon as +1.

orf20_NK-198	MAWYRNSRFVYNALKNLRSKTFGTIPTPRVHSNSSLFYNQSTNKCSGLFGSAKSGYFN 60
orf20_NK-305-1	MAWYRNSRFVYNALKNLRSKTFGTIPTPRVHSNSSLFYNQSTNKCSGLFGSAKSGYFN 60

orf20_NK-198	GFKHHQEISSFSGFARRNYHGDKTEVSVESWLEKFLVPIGLILTFGILGYPHVHPVVVPY 120
orf20_NK-305-1	GFKHHQEISSFSGYARRNYHGDKTEVSVESWLEKFLVPIGLILTFGILGYPHVHPVVVPY 120
	*****:*****
orf20_NK-198	TGRKHVLMSTTRENEIGEVEKRKIQPATHPDTRVRSIFQHILESLEIREINHHELELER 180
orf20_NK-305-1	TGRKHVLMSTTRENEIGEVEKRKIQPATHPDTRVRSIFQHILESLEIREINHHELELER 180

orf20_NK-198	DETFKEKTIWKEETVDDKDSRKKHSGAKITTNHLEGMNWEIFVVDKPLVESSYLLGGKIV 240
orf20_NK-305-1	DETFKEKTIWKEETVDDKDSRKKHSGAKITTNHLEGMNWEIFVVDKPLVESSYLLGGKIV 240

orf20_NK-198	VYTGLLNHCNSDAELATIIAHQVGHAVARHEAEDSTAFFWLLISLNVILFKILFTEPESA 300
orf20_NK-305-1	VYTGLLNHCNSDAELATIIAHQVGHAVARHEAEDSTAFFWLLISLNVILFKILFTEPEFA 300
	***** *
orf20_NK-198	NARSKLLLRHPLLQKVWKI IQARAPQLLPRTICLSLVGLFSSVFILYYGRKEIEADHIGV 360
orf20_NK-305-1	NARSKLLLRHPLLQKVWKI IQARAPQLLPRTICLSLVGLFSSVFILYYGRKEIEADHIGV 360

orf20_NK-198	LLMASAGYDPRVAPQVYDKLAKPLGDWNCLATHPFARMRAKLLARADVMKEADKIYNEVV 420
orf20_NK-305-1	LLMASAGYDPRVAPQVYDKLAKPLGDWNCLATHPFARMRAKLLARADVMKEADKIYNEVV 420

orf20_NK-198	AGRAIQGLQ 429
orf20_NK-305-1	AGRAIQGLQ 429

Supplementary Fig 6 Alignment of amino acid sequences of the protein products deduced from *orf20_{NK-198}* and *orf20_{NK-305-1}* nucleotide sequences. Asterisks denote matched residues.

orf20_NK-305-2	MAWYRNSRFVYNALKNLRSKTFGTIPTPRVHSNSSLFYNQSTKCSGLFGSAKSGYFNG 60
orf20L_S	MAWYRNSRFVYNALKNLRSKTFGTIPTPRVHSNSSLFYNQSTKCSGLFGSAKSGYFNG 60

orf20L_S	MAWYRNSRFVYNALKNLRSKTFGTIPTPRVHSNSSLFYNQSTKCSGLFGSAKSGYFNG 60
orf20_NK-305-2	MAWYRNSRFVYNALKNLRSKTFGTIPTPRVHSNSSLFYNQSTKCSGLFGSAKSGYFNG 60

orf20L_S	FKHHQEISSFSGFARRNYHGDKTEVSVESWLEKLLLGIALMLSTGIFAYRHVHPVVVPYT 120
orf20_NK-305-2	FKHHQEISSFSGFARRNYHGDKTEVSVESWLEKLLLGIALMLSTGIFAYRHVHPVVVPYT 120

orf20L_S	GRKHVVLISTDENEGEVEKRKIQPATHPDTDRVRSIFQHILESLEREINHHELELERD 180
orf20_NK-305-2	GRKHVVLISTDENEGEVEKRKIQPATHPDTDRVRSIFQHILESLEREINHHELELERD 180

orf20L_S	ETFKEITIWKEETVDDKDSRKKHSGAKITTNHLEGLNWEIFVVDKPLVESSCLFGGKIVV 240
orf20_NK-305-2	ETFKEITIWKEETVDDKDSRKKHSGAKITTNHLEGLNWEIFVVDKPLVESSCLFGGKIVV 240

orf20L_S	YTGLLNHCNSDAELATIIAHQVGHAVARHQAEDRTAFFWWSMSLYVIIFEVLFTARKFAN 300
orf20_NK-305-2	YTGLLNHCNSDAELATIIAHQVGHAVARHQAEDRTAFFWWSMSLYVIIFEVLFTARKFAN 300

orf20L_S	ARSKLLLRHPLLQKVWKIIQARFHQLLPRTTLHLGFLGLSSLVFILYFGRKEIEADHIGV 360
orf20_NK-305-2	ARSKLLLRHPLLQKVWKIIQARAPQLLPRTIC-LSLVGLFSSVFILYYGRKEIEADHIGV 359
	***** * ** * *****
orf20L_S	LLMASAGYDPRVAPQVYDKLAKPLGDWNCLATHPFARMRAKLLARADVMKEADKIYNEVV 420
orf20_NK-305-2	LLMASAGYDPRVAPQVYDKLAKPLGDWNCLATHPFARMRAKLLARADVMKEADKIYNEVV 419

orf20L_S	AGRAIQGLQ 429
orf20_NK-305-2	AGRAIQGLQ 428

Supplementary Fig 7 Alignment of amino acid sequences of the protein products deduced from *orf20L_s* and *orf20_{NK-305-2}* nucleotide sequences. Asterisks denote matched residues.

orf19 --ATAAATGATAAAAAA----AAAAAACACACATCATAAGTAACCGATTAA-ACAAGT -3448
orf20_NK-305-2 AAATATAGTCATGTGGGATCTTATTAAATTCAACATAACATGTATTTTATTAATATGTAC -3441
*** * * ** * * *** **** * * *** ***** *

orf19 CACAAATAGGTTAAACTCAATAGAAGAAAACAGACATACAAACCTAACTCAACCAAGGA -3388
orf20_NK-305-2 TTTTATAATTTTAA--CAAATACAAACTAAAGATATATGTCTCAAAGTTTTGCATTGA -3383
*** ** * **** * * * *** *** * * * ** ** *

orf19 TA--ATCAAAAG-----GTAGATTTAAATGAGA--AAACGCGCAAAA----- -3349
orf20_NK-305-2 CATGCGTGAAAAGTGTAACTGTGTAGATT-AAAATGAAACGAAGGGAGTACAATCTTAGC -3324
* * ***** ***** * * * * * *

orf19 -ATTGACTTAT----TTAAAA--CAATTTTTCCAAAAAAAAAATGATATAAACTTTGTT -3297
orf20_NK-305-2 CACTAACTCATGATGTCAAGATCTCAAGTCTT----AAAAATAATGATATAA-CTTTGTT -3269
* * *** ** * * * * *** * * * ***** ***** *

orf19 A-ACATCGTCTAGGATAATATTGAAGGATGGACAACAATAAACTAAAAAGTTACAATCAC -3238
orf20_NK-305-2 ACACATAGCTAGCATAATATTGAGGGATGAACAACAATA---TAATACACTAAAA----- -3216
* ***** ***** ***** ***** ***** *** * * *

orf19 ATCATTATTTTGCAGTTGTAGTTCCATTTGCAGCAGAAGCAAGATGCAAAAATGAAGCAC -3178
orf20_NK-305-2 --AGTTAGTTTGCAGTTGTAGTTCCATTTGTAGCAGAAGCAAGATGCAAAAATGAAGCAC -3158
*** ***** *

orf19 CCCACCAATATCAGGGGTAGCAACACCAGCTCCAGCACCAGTGCCGCCGCCGCTAGCAG -3118
orf20_NK-305-2 CCCACCAATATCAGGGGTAGCAACACCAGCTCCAGCACCAGCGCCGCCGCCGCTAGCAG -3098

orf19 TAACACCTACAGGCGTACCAGCATATCAAATTCACGGTTGTGGCTACCCGTGCAGTGACT -3058
orf20_NK-305-2 TAACACCTACAGGCGTACCAACATATCAAATTCACGGTTGTGGCTACCCGTGCAGTGACT -3038

orf19 CCAATGACTGTGATTGGCCTTGACTATATGCTGTGGCCAAAACCAAACCTTGTGTTATG -2998
orf20_NK-305-2 CCAATGACTGTGATTGGCCTTGACTATATGCTGTGGCCAAAACCAAACCTTGTGTTATG -2978

orf19 ATTTGCCTTTTCGATGGTTAACATAAAACAGCACTTTTTTTTTCTAAGTAGTCATAAAGAG -2938
orf20_NK-305-2 ATTTGCCTTTTCGATGGTTAACATAAAACAGCACTTTTTTTTT-CTAAGTAGTCATAAAGAG -2919

orf19 GGTACTACAATAAATAATGTTATACAATTCCACATAAACTTAGTAATCTATCATAATAAT -2878
orf20_NK-305-2 GGTACTACAATAAATAATGTTATACAATTCCACATAAACTTAGTAAGTTGTGACAATAAT -2859
***** * * *

orf19 GTTCAATAATGAAGTACTAAATCCAGTCTCTCGTGATTCAACTAGAAAATCTCTTCTAGC -2818
orf20_NK-305-2 GTTCAATAATGAAGTACTAAATCTAGTCTCTCGTGATTCAACTAGAAAATCTCTTCTAGC -2799

orf19 TCACTAGATGGTCCTTCTACGATAGCATGGTTTCGAACTTTTCGATTCATTAGAATTGAA -2758
orf20_NK-305-2 TCACTAGATGGTCCTTCTACGATAGCATGGTTTCGAACTTTTCGATTCATTAGAATTGAA -2739

orf19	TTATGTATATTTCTACTTGTGTTTGATTCAATCCTGCCAATAATGTGATTTATACATGGT	-2698
orf20_NK-305-2	TTATGTATATTTCTACTTGTGTTTGATTCAATCCTGCCAATAATGTGATTTATACATGGT	-2679

orf19	GGTATAAGATTGTAACCGTTATTGTTGTATTGTATTCATCATATTTATAATCCCTCTATC	-2638
orf20_NK-305-2	GGTATAAGATTGTAACCGTTATTGTTGTATTGTATTCATCATATTTATAATCCCTCTATC	-2619

orf19	TCTAATTGTTTGCTTAATTTGAGGATTAATTAAGGTTGGCCTAGTGGTTGTGAGGTCCCT	-2578
orf20_NK-305-2	TCTAATTGTTTGCTTAATTTGAGGATTAATTAAGGTTGGCCTAGTGGTTGTGAGGTCCCT	-2559

orf19	CTACCACCTATTGAATAGGAGCTCGATTCTCACCTTTTGCAAATTCCTGTAAGGAAATGC	-2518
orf20_NK-305-2	CTACCACCTATTGAATAGGAGCTCGATTCTCACCTTTTGCAAATTCCTGTAAGGAAATGC	-2499

orf19	TCCTTATCTTAAAAAAGAAAAAAGAAATGCTTGCTCAATATTTTATGTTTT	-2458
orf20_NK-305-2	TCCTTATCTTAAAAAAGAAAAAAGAAATGCTTGCTCAATATTTTATGTTTT	-2440
***** *****		
orf19	TCCGATTTCAAAGACAGTATTCGAGTATTGTATTATTTTCAGTTACGTGGTAGTAAAATT	-2398
orf20_NK-305-2	TCCGATTTCAAAGACAGTATTCGAGTATTGTATTATTTTCAGTTACGTGGTAGTAAAATT	-2380

orf19	AAAAAGAAATTGTGATAGTTTGAAAATATGTGTTAAGAACTCAAAGATCTTAGATCTTA	-2338
orf20_NK-305-2	AAAAAGAAATTGTGATAGTTTGAAAATATGTGTTAAGAACTCAAAGATCTTAGATCTTA	-2320

orf19	GATCTGCCGTTGCACAACGTAACCTTAAGTACCTTTAATATCCCATTTGTAATGATTAA	-2278
orf20_NK-305-2	GATCTGCCGTTGCACAACGTAACCTTAAGTACCTTTAATATCCCATTTGTAATGATTAA	-2260

orf19	AAATTACAAAAAGTTGATATTTGGAAAGTATATTTGAGACGAATGACTATGACAGACAA	-2218
orf20_NK-305-2	AAATTACAAAAAGTTGATATTTGGAAAGTATATTTGAGACGAATGACTATGACAGACAA	-2200

orf19	ACTACATTAGATGAATAATGTCAAAATTCATATGAACCACTAAAAACGGAGGAAGTATA	-2158
orf20_NK-305-2	ACTACATTAGATGAATAATGTCAAAATTCATATGAACCACTAAAAACGGAGGAAGTATA	-2140

orf19	TATGAATTCAGATCGTCAAAATCAGGGATGTCTCAGACAATTTAGGGGTCCCGTGCCATC	-2098
orf20_NK-305-2	TATGAATTCAGATCGTCAAAATCAGGGATGTCTCAGACAATTTAGGGGTCCCGTGCCATC	-2080

orf19	CATTTTATTAATACACATCTTTTGTAAAGGTATATTAGTACATAATTTGATTCCTTAATG	-2038
orf20_NK-305-2	CATTTTATTAATACACATCTTTTGTAAAGGTATATTAGTACATAATTTGATTCCTTAATG	-2020

orf19	AACTAGAACT-TTTTATTCTAAGGTCTTGTGTTGACAAAACAAATCATGGGAAGTGA	-1979
orf20_NK-305-2	AACTAGAACTGTTTTATTCTAAGGCCTTGTGTTGACAAAACAAATCATGGGAAGTGC	-1960

orf19	AAATTCCTAGAAAGTAGAAGTATATGAATAAGTTGTTTGGTTGGCAAAATCATGGAATT	-1919
orf20_NK-305-2	AAATTCCTAGAAAGTAGAAGTATATGAATAAGTTGTTTGGTTGGCAAAATCATGGAATT	-1900

orf19	TGTAATTCTTAGGAAGTTCTACTTCCTCCAAATGGAGGAAGTTACTTACCTAGGTCCCC	-1859
orf20_NK-305-2	TGTAATTCTTAGGAAGTTCTACTTCCTCCAAATGGAGGAAGTTACTTACCTAGGTCCCC	-1840

orf19	CTAGATAAGTAGAGCTTCCCATAGGAAGTTCTACTTCCATTATCAACCAACACAATTT	-1799
orf20_NK-305-2	CTAGATAAGTAGAGCTTCCCATAGGAAGTTCTACTTCCATTATCAACCAACACAATTT	-1780

orf19	TACACTTCCTAGGAATTTCAAATACATGGAAAAATAACTTCCTAGGAAGGCTAGAACCG	-1739
orf20_NK-305-2	TACACTTCCTAGGAATTTCAAATACATGGAAAAATAACTTCCTAGGAAGGCTAGAACCG	-1720

orf19	TGCATTTTGAGTAAAAAAATTAACATAACAGGTTGAAACTTTTTATTGTATATCATATA	-1679
orf20_NK-305-2	TGCATTTTGAGTAAAAAAATTAACATAACAGGTTGAAACTTTTTATTGTATATCATATA	-1660

orf19	AAATTTTGTGTCATCAGTCAAAAGTAACAGAGGGTTCAAATTGCGGAATTCGTTCTCA	-1619
orf20_NK-305-2	AAATTTTGTGTCATCAGTCAAAAGTAACAGAGGGTTCAAATTGCGGAATTCGTTCTCA	-1600

orf19	TATACTTAGATGTACGGTTGACATGGTACACTCAATACTAATTTATGTTCAATTTGCTTAT	-1559
orf20_NK-305-2	TATACTTAGATGTACGGTTGACATGGTACACTCAATACTAATTTATGTTCAATTTGCTTAT	-1540

orf19	TTAGAATTCTCTTATTTAATATATTTTGTTCATTGAATGATTTCAAATAAATAATAGGC	-1499
orf20_NK-305-2	TTAGAATTCTCTTATTTAATATATTTTGTTCATTGAATGATTTCAAATAAATAATAGGC	-1480

orf19	AAAAAATAGCCTTTAAATGTATCATGCTCGCAACATTTAGGTATATGATAAAATTTGTA	-1439
orf20_NK-305-2	AAAAAATAGCCTTTAAATGTATCATGCTCGCAACATTTAGGTATATGATAAAATTTGTA	-1420

orf19	CAAATTTTAATGAAAAAAAATGTTGAAACAAAAATTTAACTAAGCTACAGTTGACTT	-1379
orf20_NK-305-2	CAAATTTTAATGAAAAAAAATGTTGAAACAAAAATTTAACTAAGCTACAGTTGACTT	-1361

orf19	TTAAGCTCTCTCCTTTTCTGATGCAACAAAGATTTGTTTAGCACTAGCTACTTCTTCTA	-1319
orf20_NK-305-2	TTAAGCTCTCTCCTTTTCTGATGCAACAAAGATTTGTTTCAGCACTAGCTACTTCTTCTA	-1301

orf19	TCCCATAAAATTCGCCATTTGTTTTCTCAAACCTCAAATTTTCAATTTTGATTATATT	-1259
orf20_NK-305-2	TCCCATAAAATTCGCCATTTGTTTTCTCAAACCTCAAATTTTCAATTTTGATTATATT	-1241

orf19	TTTTACCATGTAAGAAAATATCTTATCAGGAGTTTTTTGTAATCGAGGAAACCCATATA	-1199
orf20_NK-305-2	TTTTACCATGTAAGAAAATATCTTATCATGAGTTTTTTGTAATCGAGGAAACCCATATA	-1181

orf19	GGAAACTGCTTATAAAGCTAGTGAATCAACGAAATATCAACAAGAAAATCACATTGTTCA	-1139
orf20_NK-305-2	GGAAATTGCTTATAAAGCTAGTGAATCAACGAAATATCAACA-----	-1139

orf19	TAGGAAACTCCTATAACATTTGCATACATTGTAGGTCTTTGTTGCACTTTAATCCCTCG	-1079
orf20_NK-305-2	---GGAAACTCCTATAACATTTGCATACATTGTAGGTCTTTGTTGCACTTTAATCCCTCG	-1081

orf19	GGCCTCGTATATTGATATTAAGTGTATTTAATCTACGTTTTTTCTATTGCAACAATTA	-1019
orf20_NK-305-2	GGCCTCGTATATTGATATTAAGTGTATTTAATCTATATTTTTCTATTGCAACAATTA	-1021

orf19	CTACTTTGGATAATTTTACATCTATTGCAACAATTTCACTTTTGGTATAAAAAGCAATAT	-959
orf20_NK-305-2	CTACTTTGGATAATTTTACATCTATTGCAACAATTTCACTTTTGGTATAAAAAGCAATAT	-961

orf19	TTCAGAACAAGCAGCATTGTGCACCCGAGAGAACATACCATTATAGAAGAATATGA	-899
orf20_NK-305-2	TTCAGAACAAGCAGCATTGTGCACCCGAGAGAACATACCATTATAGAAGAATATGA	-901

orf19	TTTTTTTCAACAACTTTTCAAGATAAAAAAGCATACAATATAAAATTAAGAACATGT	-839
orf20_NK-305-2	TTTTTTTCAACAACTTTTCAAGATAAAAAAGCATACAATATAAAATTAAGAACATGT	-841

orf19	AAATCTCCAACAACATTTGAAAAACCTAAAAAAAGCATCTATTGTAATGAACATTT	-779
orf20_NK-305-2	AAATCTCCAACAACATTTGAAAAACCTAAAAAAAGCATCTATTGTAATGAACATTT	-781

orf19	AGTCTAAATTTAAGAACCTAACTTTTAAATGTAAAAATTTGAAAAAGGAATCTCGCCA	-719
orf20_NK-305-2	AGTCTAAATTTAAGAACCTAACTTTTAAATGTAAAAATTTGAAAAAGGAATCTCGCCA	-721

orf19	ACAACCATTTTCTAAAAGGTAAAACAAGTTGCAAGATTTAATGAAACAGATAACAAC-	-660
orf20_NK-305-2	ACAACCATTTTCTAAAAGGTAAAACAAGTTGCAAGATTTAATGAAACAGATAACAACA	-661

orf19	TTTTTCTTAAATCATAAATTCTTAAATATTAACCTACATCGTTTAAACAGAAGGTGCAC	-600
orf20_NK-305-2	TTTTTCTTAAACCATAAATTCTTAAATATTAACCTACATCGTTTAAACAGAAGGTGCAC	-601

orf19	CATCCTTTATGCGTACTTGGATGCATGTTCTATTTGCGCTTTTCCATTCCCTAAAAAAC	-540
orf20_NK-305-2	CATCCTTTATGCGTACTTGGATGCATGTTCTATTTGCGCTTTTCCATTCCCTAAAAAAC	-541

orf19	CGAGTCTAAAGCTATTGTTATAATACTCTAGAAGTCCTCTCAAAAAAAAAAAAA-CACT	-481
orf20_NK-305-2	CGAGTCTAAAGCTATTGTTATAATACTCTAGAAGTCCTCTCAAAAAAAAAAAAAACACT	-481

orf19	CTAGTTGTCGTAGACTCTCAGCATGTGTCATTTAGAGACTCGTAACGCATTGACGCATT	-421
orf20_NK-305-2	CTAGTTGTCGTAAACTCTCAGCATGTGTCATTTAAAGACTCGTAACGCATTGACGCATT	-421

orf19	ACTCGAAATATTTTGTTCATTATTAGAGAAAATTTTCATCCCCTGATCCCAATTATCAA -361
orf20_NK-305-2	ACTCGAAATATTTTGTTCATTATTAGAGAAAATTTTCATCCCCTGATCCCAATTATCAA -361

orf19	ATCAACATCTAAAAATTTAAATGACTAGGTACGTAACGAAAAACGAATGACTCTCGATA -301
orf20_NK-305-2	ATCAACATCTAAAAATTTAAATGACTAGGTACGTAACGAAAAACGAATGACTCTCGATA -301

orf19	ATAGTACACCCCATTAATCCATTCTTAGTTTGTTCGCTAGTGCATAGCTGGCTGCAAAAG -241
orf20_NK-305-2	ATAGTACACCCCATTAATCCATTCTTAGTTTGTTCGCTAGGGCATAGCTGGCTGCAAAAG -241
	***** *****
orf19	AAATCTTTTGCACAGAGAAAACCTTTTGCACCTTCGGAATTCAGTAGGAATATCATAACCA -181
orf20_NK-305-2	AAATCTTTTGCACAGAGAAAACCTTTTGCACCTTCGGAATTCAGTAGGAATATCATAACCA -181

orf19	TTTATGGAAGCAACAACCTCTTGTGACCCATTTTCATCTAAATCTTAATCTCGTAAATTTT -121
orf20_NK-305-2	TTTATGGAAGCAACAACCTCTTGGGACCCATTTTCATCTGAAATCTTAATCTCGTAAATTTT -121
	***** ***** *****
orf19	ACCTTTCAAATTCAAAAATCACATAATTTTTTTTGGTATGTTACTTGAACCCAGTTCAT -61
orf20_NK-305-2	ACCTTTCAAATTCAAAAATCACATAATTTTTTTTGGTATGTTACTTGAACCCAGTTCAT -61

orf19	AACTGACCCTGAAATTCAAGAATTTGGAGCAAAGTTAGCAGCTTTTGTGTTCAAAAATC -1
orf20_NK-305-2	AACTGACCCTGAAATTCAAGAATTTGGAGCAAAGTTAGCAGCTTTTGTGTTCAAAAATC -1

Supplementary Fig 8 Alignment of nucleotide sequences of the upstream regions of *orf19* and *orf20_{NK-305-2}*. Asterisks denote matched residues. Nucleotides are numbered from the first residue of the initiation codon as +1.

orf20L_S TGCATTTGATGATTTGAATGTGTTGGTTATTGTCTTCAAGCTTAAATGTGCCTATATTAA -4858
orf20_NK-305-1 -----

orf20L_S TACAACTTTTTTAAGGTATATTTGTACATGATTTCAATTCCTTAAGTAGTATTTGGGTTT -4798
orf20_NK-305-1 -----

orf20L_S GGGTGTGTCCCGGATAGTATTGTTTTCTGAATTTGATATGCATTTTTTTTTTTCATTT -4738
orf20_NK-305-1 -----GTTGCCCGGATAGTATTGATTCTGAATTTGATATGCATTTTTTTTTTTCATTT -5403
*** *****

orf20L_S CCATTAATATACTCCCTCCGTCCTCCAAATATAGTTTCCATTTCCATTTGGGTGTCCC -4678
orf20_NK-305-1 CCATTAATATACTCCCTCCGTCCTCCAAATATAGTTTCCATTTCCATTTGGGTGTCCC -5343

orf20L_S AAAATATAGTTCCCATATCCCATTTCCATATTTAGTTCCACGTTTTTTCGTAATTTGTCT -4618
orf20_NK-305-1 AAAATATAGTTCCCATATCCCATTTCCATATTTAGTTCCACATTTTTTCGTAATTTGTCT -5283

orf20L_S AGAAAAACCGTGTCCTCATTTATTTGCTTCTTGAATTTGGTTTTTCCTTTGTTTATT -4558
orf20_NK-305-1 AGAAAAACCGTGTCCTCATTTATTTGCTTCTTGAATTTGGTTTTTCCTTTGTTTATT -5223

orf20L_S CAACCAAAATGTTACAATTAATGCTCTTCCACCAATTATTCTCCACTCTTTCTCCTAAAA -4498
orf20_NK-305-1 CAACCAAAATGTTACAATTAATGCTCTTCTACCAATTATTCTCCACTCTTTCTCCTAAAA -5163

orf20L_S TCATCTTTTCCCATACAACATTTATTTAAATAAACAATAATCATTACTATTCTCTTAT -4438
orf20_NK-305-1 TTATCTTTTCCCATACAACATTTATTTAA-TAAACAATAATCATTACTATTCTCTTAT -5104
* *****

orf20L_S ATTCTACTTACATAAATTACCGTGAAAAAGGGAAATGGGAAGTATATTTTGGGACGGAG -4378
orf20_NK-305-1 ATTCTACTTACATAAATTACCGTGAAAAAGGGAAATGGGAAGTATATTTTGGGACGGAG -5044

orf20L_S GGAGTATTAAGTAAACTCAACATTTAAACCATACAAATATAATAATATGGAGACTTAAA -4318
orf20_NK-305-1 GGAGTATTAAGTAAACTCAACATTTAAACCATACAAATATAATAATATGGAGACTTAAA -4984

orf20L_S GCATGATTAAGTTGGTTGAGATGGTAATTGTGTCATGTATAATAACAAGAGACTACAA -4258
orf20_NK-305-1 GCATGATTAAGTTGGTTGAGATGGTAATTGTGTCATGTATAATAACAAGAGACTACAA -4924

orf20L_S GTTCAAATCTTGTGCAAGCTTATTTACTTTTGTAAATTGACATGAGATATATACACAT -4198
orf20_NK-305-1 GTTCAAATCTTGTGCAAGCTTATTTACTTTTGTAAATTGACATGAGATATATACACAT -4864

orf20L_S	TGGACAAATCTACTGAAGTAACAGAGGTGCCACGTGGTGGGTATACATTGTCACGCACAC	-4138
orf20_NK-305-1	TGGACAAATCTACTGAAGTAACAGAGGTGCCACGTGGTGGGTATACATTGTCACGCACAC	-4804

orf20L_S	CTTTAAATATATTTGTATAGATGAACTAGAACTGTTTTCTTCTAATAATAAAGAACGATA	-4078
orf20_NK-305-1	CTTTAAATATATTTGTATAGATGAACTAGAACTGTTTTCTTCTAATAATAAAGAACGATA	-4744

orf20L_S	AGGATGAAGTTTATTAACAGATGGTCTGTAGAAAATGTGATTTGCTTCAACTTTGTAAGG	-4018
orf20_NK-305-1	AGGATGAAGTTTATTAACAGATGGTCTGTAGAAAATGTGATTTGCTTCAACTTTGTAAGG	-4684

orf20L_S	TAAATCATGGCAATTAAGCTTTTAGGCGAGGAAATTAAGATCATGAATATTATTTAATT	-3958
orf20_NK-305-1	TAAATCATGGCAATTAAGCTTTTAGGCGAGGAAATTAAGATCATGAATATTATTTAATT	-4624

orf20L_S	TGTAAACTCTTTGTATATCTATCATTTTGTGCTTATAGTATGCACCATTTTCCTATGTC	-3898
orf20_NK-305-1	TGTAAACTCTTTGTATATCTATCATTTTGTGCTTATAGTATGCACCATTTTCCTATGTC	-4564

orf20L_S	TTCAAAGCCTCAAAGGAATACTCCATTTCTTTTTTTTTTTTTTTTTTTTTTTTGGTCAAGT	-3838
orf20_NK-305-1	TTCAAAGCCTCAAAGGAATACTCCATTTCTTTTTTTTTTTTTTTTTTTT-----TTGGTCAAGT	-4508

orf20L_S	GGTGATTGGTGAAGTCCCTAGAACCGTGCATTTTGAGTAAAAAATTAAGTGAACAGGTT	-3778
orf20_NK-305-1	GGTGATTGGTGAAGTCCCTAGAACCGTGCATTTTGAGTAAAAAATTAAGTGAACAGGTT	-4448

orf20L_S	GAACTTTTTATTTGTAAATCATATAAAATTTTGTGTCATCAGTCAAAA-----GTA	-3725
orf20_NK-305-1	GAACTTTTTATTTGTAAATCATATAAAATTTTGTGTCATCAGTCAAAACTCAAAAGTA	-4388

orf20L_S	ACAGAGGGTGCAAATTGCGGAATTACTTCCTCATATACTTTGATGTACCGTTGATATGGT	-3665
orf20_NK-305-1	ACAGAGGGTGCAAATTGCGGAATTACTTCCTCATATACTTTGATATACCGTTGATATGGT	-4328

orf20L_S	ACACTCAATACTAATTTATGTTTCTTTGCTTATTTGAAATTTTCTTATTTTGTTCATTTT	-3605
orf20_NK-305-1	ACACTCAATACTAATTTATGTTTCTTTGCTTATTTGAAATTTTCTTATTTTGTTCATTTT	-4268

orf20L_S	AATGATTTCAAATAAATAATAGGCAAAAAAATAGCCTTTAAATGTATCATGCTCGCAACA	-3545
orf20_NK-305-1	AATGATTTCAAATAAATAATAGGCAAAAAAATAGCCTTTAAATGTATCATGCTCGCAACA	-4208

orf20L_S	TTTAGGTATATGATAAAATTTATACAAATTTTAATGAAAAAAAATGTTGAAACAAAAAT	-3485
orf20_NK-305-1	TTTAGGTATATGATAAAATTTATACAAATTTTAATGAAAAAAAATGTTGAAACAAAAAT	-4148

orf20L_S	TTAAGCTAAGCTAAAGTTGACTTTTAAGCTCTCTCCTTTTATGATGCAACAAAGATTTTG	-3425
orf20_NK-305-1	TTAAGCTAAGCTAAAGTTGACTTTTAAGCTCTCTCCTTTTATGATGCAACAAAGATTTTG	-4088

orf20L_S	TTTGTAGCACTAGCTACTTTTTCTATCCCATATAAAATTCGCCATTTGTTTTCTCAAACCTCA	-3365
orf20_NK-305-1	TTTGTAGCACTAGCTACTTTTTCTATCCCATATAAAATTCGCCATTTGTTTTCTCAAACCTCA	-4028

orf20L_S	AATTTTCATCAATTTTGATTATGTTTTTTCACCATGTAAGAAAATATCTTATCATGTGTTT	-3305
orf20_NK-305-1	AATTTTCATCAATTTTGATTATGTTTTTTCACCATGTAAGAAAATATCTTATCATGTGTTT	-3968

orf20L_S	TTCGTAATCGAGGAAACCCATATAGGAACTGCTTATAAAGCTAGTGAATCAACGAAATA	-3245
orf20_NK-305-1	TTCGTAATCGAGGAAACCCATATAGGAACTGCTTATAAAGCTAGTGAATCAACGAAATA	-3908

orf20L_S	TCAACAGGAAAATCACATTGTTTCATAGGAAACTCCTATAACATTTGCATACATTGTAGG	-3185
orf20_NK-305-1	TCAACAGGAAAATCACATTGTTTCATAGGAAACTCCTATAACATTTGCATACATTGTAGG	-3848

orf20L_S	TCATTGTTGCACTTTATTCCTCGGCCCTCGTATATTGATATTAAGTGTATTTTAATCTA	-3125
orf20_NK-305-1	TCATTGTTGCACTTTATTCCTCGGCCCTCGTATATTGATATTAAGTGTATTTTAATCTA	-3788

orf20L_S	CGTTTTTTTCTATTGCAACAATTACTACTTTGGATAATTTTACATCTATTGCAACAATTC	-3065
orf20_NK-305-1	CGTTTTTTTCTATTGCAACAATTACTACTTTGGATAATTTTACATCTATTGCAACAATTC	-3728

orf20L_S	CATTTTTGGTATAAAAAGCAACATTTTCAGAACAAAGCATGGATTATGCACTAGGGTACCA	-3005
orf20_NK-305-1	CATTTTTGGTATAAAAAGCAACATTTTCAGAACAAAGCATGGATTATGCACTAGGGTACCA	-3668

orf20L_S	TTCATATAGAAGAATATGATTTTTTTCAACAACCTTTTCAAGATAAAAAAAGCACACAAT	-2945
orf20_NK-305-1	TTCATATAGAAGAATATGATTTTTTTCAACAACCTTTTCAAGATAAAAAAAGCACACAAT	-3608

orf20L_S	ATAAAATTAAAGAACATGTAAGAGTGC GTTTTTATTCAACTTATTGGCCCTGAACCTTATTG	-2885
orf20_NK-305-1	ATAAAATTAAAGAACATGTAAGAGTGC GTTTTTATTCAACTTATTGGCCCTGAACCTTATTG	-3548

orf20L_S	GACCTTATCTGAACCTGAATTTATTGAACCTGAACCTGAACCTTATTGGAACCTATTAAACCT	-2825
orf20_NK-305-1	GACCTTATCTGAACCTGAATTTATTGAACCTGAACCTGAACCTTATTGGAACCTATTAAACCT	-3488

orf20L_S	GATTGGACCTGATTCAACTTATTGGACCTGATTAAACCTGATT-----	-2782
orf20_NK-305-1	GATTGGACCTGATTCAACTTATTGGACCTGATTAAACCTGATTATTGAAACCTATTAGAC	-3428

orf20L_S	-----GGAACTTATTGGAAGTATTG-ACCTTATTGAAACCTATTAGAC	-2739
orf20_NK-305-1	CTTATTGGCCCTGATTGAAACTTATTAGACCTTATTGGACCTGATTGAAACTTATTAGAC	-3368
	* **** * * **** * * * **** * * **** *	
orf20L_S	CTTATTGG-----	-2731
orf20_NK-305-1	CTTATTGGACCTTATTGACAAAAACATTGACCATGAATAACATAAATATTACCACTAAC	-3308

orf20L_S	-----	-2731
orf20_NK-305-1	GTAATACTACCCCTCAAAATTTTTATGGAGTAATAATTATTATAATTCGTCCTTTAA	-3248
orf20L_S	-----	-2731
orf20_NK-305-1	AATAATGATTATTAATTATCTCTTATGATACTCCCTCCGTCCTCCATAATATAGTGCTCACT	-3188
orf20L_S	-----	-2731
orf20_NK-305-1	TCTCATTTTTCACAGAATTAAGGAAAATGGAAAGAGTTTTAGGATTTCACAAATTAGAC	-3128
orf20L_S	-----CCCTGATTGAAACTTATTAGACCTTATTGGACCTGATTGAAACTTATTAGAC	-2679
orf20_NK-305-1	CTTATTGGCCCTGATTGAAACTTATTAGACCTTATTGGACCTGATTGAAACTTATTAGAC	-3068

orf20L_S	CTTATTGGACCTTATTGACAAAAACATTGACCATGAATAACATAAATATTACCACTAA-	-2620
orf20_NK-305-1	CTTATTGGACCTTATTGACAAAAACATTGACCATGAATAACATAAATATTACCACTAAG	-3008

orf20L_S	-----CGTAAATACTACCCCTCAAAATTTTTATGGAGTAATAATTATTATAATTCGT	-2567
orf20_NK-305-1	GCACTAACGTAAATACTACCCCTCAAAATTTTTATGGAGTAATAATTATTATAATTCGT	-2948

orf20L_S	CCTTTAAAAATAATGATTATTAATTATCTCTTATGATA-----	-2529
orf20_NK-305-1	CCTTTAAAAATAATGATTATTAATTATCTCTTATGATACTCCCTCCGTCCTCCATAATATAG	-2888

orf20L_S	-----	-2529
orf20_NK-305-1	TGCTCACTTCTCATTTTTCACAGAATTAAGGAAAATGGAAAGAGTTTTAGGATTTCACA	-2828
orf20L_S	-----	-2529
orf20_NK-305-1	AAAATATATATTTAAGTAAATGTTTATTGGGTGTTTGGTCTTGGAAGTAGTTTTTTTCT	-2768
orf20L_S	-----	-2529
orf20_NK-305-1	TCTTTTTCCTTTAAAAATAAATGAAAGAGCAATAAATATACCTTGGAATGGTGTAAACA	-2708

orf20L_S	-----	-2529
orf20_NK-305-1	ATAAATAGGGGCACACATTTTACAAGAAAGTAGGGGTATTTTGAAGTTTTATGTAGAA	-2648
orf20L_S	-----	-2529
orf20_NK-305-1	ATGAAGGATAATTTAGTCCAAATAAATTGAAAAAAAAAGAAATAGGCATAACATTTTGGG	-2588
orf20L_S	-----ATTAATTTAATAA	-2516
orf20_NK-305-1	ACACCCAAAGAGGAATGCAAGCACTATATTATGGGACGGAGGGAGTAATTAATTTAATAA	-2528

orf20L_S	AAAAATTTACTATTTATATATTTGCCTATACATAACTTTCACTAATATGTTTTGATT	-2456
orf20_NK-305-1	AAAAATTTACTATTTATATATTTGCCTATACATAACTTTCACTAATATGTTTTGATT	-2468

orf20L_S	TTATAAAACACTAGTAGAAAATCAAAGTTAATTAACATTTATTGCTAACAAGTTAAAAT	-2396
orf20_NK-305-1	TTATAAAACACTAGTAGAAAATCAAAGTTAATTAACATTTATTGCTAACAAGTTAAAAT	-2408

orf20L_S	TGACACATATAAAAAATTAACATTTATTGAAGAGGGTGATGTAGAAGATGAAGAAAGATA	-2336
orf20_NK-305-1	TGACACATATAAAAAATTAACATTTATTGAAGAGGGTGATGTAGAAGATGAAGAAAGATA	-2348

orf20L_S	CCCGATGAAGAAAGATACTCTAGTGATGATAATGAAGCAATCAATTGACAACAATTATG	-2276
orf20_NK-305-1	CCCGATGAAGAAAGATACTCTAGTGATGATAATGAAGCAATCAATTGACAACAATTATG	-2288

orf20L_S	TCTTTCATTGTTATTAGTAACGAAAACATGTTATCTCTAGTTATTTAAAGACGAATTGCA	-2216
orf20_NK-305-1	TCTTTCATTGTTATTAGTAACGAAAACATGTTATCTCTAGTTATTTAAAGACGAATTGCA	-2228

orf20L_S	AATTATTGTAATTATAATTATTATTATTATTGTTAACCTTAATTATTGACCATGATTAT	-2156
orf20_NK-305-1	AATTATTGTAATTATAATTATTATTATTATTGTTAACCTTAATTATTGACCATGATTAT	-2168

orf20L_S	AATATTATTTAATAGCAATATGAATAATCAAATAATAGACAATAATACAAGTATAATACT	-2096
orf20_NK-305-1	AATATTATTTAATAGCAATATGAATAATCAAATAATAGACAATAATACAAGTATAATACT	-2108

orf20L_S	ACACATTGTGGTACTTTAATAAAAAATTCTAATAAACATAATCAGCTAATAGTAATAT	-2036
orf20_NK-305-1	ACACATTGTGGTACTTTAATAAAAAATTCTAATAAACATAATCAGCTAATAGTAATAT	-2048

orf20L_S	GAATAATAAAATAATAGACATAATACAGATAAATAACAAAATAATAGACATAATACAAAT	-1976
orf20_NK-305-1	GAATAATAAAATAATAGACATAATACAGATAAATAACAAAATAATAGACATAATACAAAT	-1988

orf20L_S	AAACAATAAAGTAATAGACATTAATACAAGTATAATATTATATAATCATTGTGGTACTTT -1916
orf20_NK-305-1	AAACAATAAAGTAATAGACATTAATACAAGTATAATATTATATAATCATTGTGGTACTTT -1928

orf20L_S	AATTAATAATTCTAATAATAACATAATCAACTAATAGTGATATGAAATTATGAATAACAAA -1856
orf20_NK-305-1	AATTAATAATTCTAATAATAACATAATCAACTAATAGTGATATGAAATTATGAATAACAAA -1868

orf20L_S	ATAATGGACAATAATACAAATGTATATTAACATTGACTATTTGGACCTTATTGGGACCT -1796
orf20_NK-305-1	ATAATGGACAATAATACAAATGTATATTAACATTGACTATTTGGACCTTATTGG-ACCT -1809
	***** ****
orf20L_S	TTATTTAGACCTGATTCAA-CTTATTGGACCTTATTAGACCTGATTGGAACCTTATTGCAC -1737
orf20_NK-305-1	TA--TTAGACCTGATTGAACTTATTGGACCTTATTAGACCTGATTGGAACCTTATTGCAC -1751
	* ***** ** *****
orf20L_S	CTGATTGAAACTTATTGCACCTGGAACCTTATTGGACCTTATTAGACCTTATTGGAAGTTA -1677
orf20_NK-305-1	CTGATTGAAACTTATTGCACCTGGAACCTTATTGGACCTTATTAGACCTTATTGGAAGTTA -1691

orf20L_S	TTGCCCTTATTAGACCTTATTACAACCTTATCTGAACCTTATTGGACCTGAAACCTTAATTTT -1617
orf20_NK-305-1	TTGCCCTTATTAGACCTTATTACAACCTTATCTGAACCTTATTGGACCTGAAACCTTAATTTT -1631

orf20L_S	TTTAAGTTGAGCAGAACGCACCCTAAATCTCCAATTATAAAATAACATAGGTCGGAATAA -1557
orf20_NK-305-1	TTTAAGTTGAGCAGAACGCACCCTAAATCTCCAATTATAAAATAACATAGGTCGGAATAA -1571

orf20L_S	AAAGAAAGTATTGTAAATGAACATTTAGTCTAAATTTAAGAACCTAACTTTTAAATGTA -1497
orf20_NK-305-1	AAAGAAAGTATTGTAAATGAACATTTAGTCTAAATTTAAGAACCTAACTTTTAAATGTA -1511

orf20L_S	AATTTAAAAAAGGAATCTCGCCAACAACCATTTTTCTAAAAGGTAAACAAGTTGCAAG -1437
orf20_NK-305-1	AATTTAAAAAAGGAATCTCGCCAACAACCATTTTTCTAAAAGGTAAACAAGTTGCAAG -1451

orf20L_S	ATTTAATGAAACAGATAACAACATTTTTCTTAAATCATAAATCTTAAATATTAACCT -1377
orf20_NK-305-1	ATTTAATGAAACAGATAACAACATTTTTCTTAAATCATAAATCTTAAATATTAACCT -1391

orf20L_S	ACATCGTTTAACAGAGGGTGCACCATCCTTTATGCGTACTTGGATGCATGTTCTATTG -1317
orf20_NK-305-1	ACATCGTTTAACAGAGGGTGCACCATCCTTTATGCGTACTTGGATGCATGTTCTATTG -1331

orf20L_S	CGCTTTTCCATTCCCTAAAAAACCGAGTCTAAAGCTATTGTTATAATACACTCTAGTAGT -1257
orf20_NK-305-1	CGCTTTTCCATTCCCTAAAAAACCGAGTCTAAAGCTATTGTTATAATACACTCTAGTAGT -1271

orf20L_S	CGTCTCAAAAAAAAAATAAGGCTTTGCTAAATTACGCCCTAAATTTTTCTGGTAACGCCCT	-1197
orf20_NK-305-1	CGTCTCAAAAAAAAAATAAGGCTTTGCTAAATTACGCCCTAAATTTTTCTGGTAACGCCCT	-1211

orf20L_S	AAATACGGTATTTTCATACCGTATTTATATACCTGCATTTCTTTCTCTCTTTCCCTTC	-1137
orf20_NK-305-1	AAATACGGTATTTTCATACCGTATTTATATACCTGCATTTCTTTCTCTCTTTCCCTTC	-1151

orf20L_S	ACTCACTCACTTTCTCTCTCTCTCCACCACAACCTGCCACCACACCACCTACCATATT	-1077
orf20_NK-305-1	ACTCACTCACTTTCTCTCTCTCTCCACCACAACCTGCCACCACACCACCTACCATATT	-1091

orf20L_S	CTTCTCCAATTTACACTTTTGACTCCTATGTCTTCTAACAACATTGTGTGGCACCACC	-1017
orf20_NK-305-1	CTTCTCCAATTTACACTTTTGACTCCTATGTCTTCTAACAACATTGTGTGGCACCACC	-1031

orf20L_S	ACCACCGACGTACCCCGACACCCACCATCGTCGTATCCCGGCCACCACCGCCGTACCC	-957
orf20_NK-305-1	ACCACCGACGTACCCCGACACCCACCATCGTCGTATCCCGGCCACCACCGCCGTACCC	-971

orf20L_S	CCGACACACCACCACCATCCCCTTTTTCTCTACTCCGTTTTCTTTATTTTCTGATTTTT	-897
orf20_NK-305-1	CCGACACACCACCACCATCCCCTTTTTCTCTACTCCGTTTTCTTTATTTTCTGATTTTT	-911

orf20L_S	TTTAAAAAAAAAGGGGAGGCGTCCACCATGGACGCCCCACCACTGGTGGAGGCGTCCATG	-837
orf20_NK-305-1	TTTAAAAAAAA-GGGGAGGCGTCCACCATGGACGCCCCATCACTGGTGGAGGCGTCCATG	-852
*** ***** *****		
orf20L_S	GCCGGCGCCTCACCATGTGGTGAAGCGCCTGCCATAGACGCCTCCCCTTTTTTTTAAAA	-777
orf20_NK-305-1	GCAGGCGCCTCACCATGTGGTGAAGCGCCTGCCATAGACGCCTCCCCTTTTTTTTAAAA	-792
** ***** *****		
orf20L_S	AAATTTCAAATTACTATTAATGAAGAAACATAGAAGAGGAGGGTGGGTTACGACGGTTGT	-717
orf20_NK-305-1	AAATTTCAAATTACTGTTAATGAAGAAACATAGAAGAGGAGGGTGGGTTACGACGGTTGT	-732
***** *****		
orf20L_S	GGTGCGGTGGCAGTGTGTGGTGGTGC GGCGCAGTA-----CTGGTGCGG	-672
orf20_NK-305-1	GGTGCGGTGGCAGTGTGTGGTGGTGC GGCGCAGTATTGGTGCGACGGCGACTGGTGCGG	-672
***** *****		
orf20L_S	CGGTGGTGGTGAACGGTCCCCAATGGTAATGTGGGAGAAGAGGGGAGAGAAGTGAATGT	-612
orf20_NK-305-1	CGGTGGTGGTGAACGGTCCCCAATGGTAATGTGGGAGAAGAGGGGAGAGAAGTGAATGT	-612

orf20L_S	AGAAGGTGGAATATGAAATGGAAAAATACGTATATTATACGTATTTAGGGCGTTAGT	-552
orf20_NK-305-1	AGAAGGTGGAATATGAAATGGAAAAATACGTATATTATACGTATTTAGGGCGTTAGT	-552

orf20L_S	AAGATATTCTAGGGCGTCGTTTAGCAAATGAGAAAAAATAATACACTCTAGTTGTCGTAG -492
orf20_NK-305-1	AAGATATTCTAGGGCGTCGTTTAGCAAATGAGAAAAAATAATACACTCTAGCTGTCGTAG -492

orf20L_S	ACTCTCAATATGTGTCATTTAGAGACTCGTAACGCATTGACGCACTTACTCGGGGTAGAA -432
orf20_NK-305-1	ACTCTCAATATGTGTCATTTAGAGACTCGTAACGCATTGACGCACTTACTCGGGGTAGAA -432

orf20L_S	ATATTTTGTTCCATTTATTAATGAGAAAATTTTCATCCCCTGATCCCAATTATCAAATCA -372
orf20_NK-305-1	ATATTTTGTTCCATTTATTAATGAGAAAATTTTCATCCCCTGATCCCAATTATCAAATCA -372

orf20L_S	ACATCTAAAAATTTAAAATGACTAGGTACGTAACGAAAAACGAATGACTCTCGATAATAG -312
orf20_NK-305-1	ACATCTAAAAATTTAAAATGACTAGGTACGTAACGAAAAACGAATGACTCTCGATAATAG -312

orf20L_S	TACACCCCATTAATCCATTCTTAGTTTGTTCATAGTTTGTTCGTAAGTGCATAGCTGGC -252
orf20_NK-305-1	TACACCCCATTAATCCATTCTTAGTTTGTTCATAGTTTGTTCGTAAGTGCATAGCTGGC -252

orf20L_S	TGCAAAGAAATCTTTTGACAGAGAAAACCTTTGCACCTTCGGAATTCAGCAGGAATAT -192
orf20_NK-305-1	TGCAAAGAAATCTTTTGACAGAGAAAACCTTTGCACCTTCGGAATTCAGCAGGAATAT -192

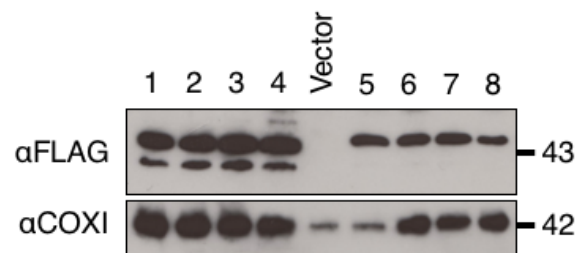
orf20L_S	CATAGCCATTTATGGAAGCAACAACCTCTTGACCCATTTTCATCTAAAACCTTAATCTC -132
orf20_NK-305-1	CATAGCCATTTATGGAAGCAACAACCTCTTGACCCATTTTCATCTAAAACCTTAATCTC -132

orf20L_S	ATAAATTTTACATTTCAGAATTCAAAAATCACGTAATTTTTTTTTTGGTATGTTACTTGA -72
orf20_NK-305-1	ATAAATTTTACATTTCAGAATTCAAAAATCACGTAATTTTTTTTTTGGTATGTTACTTGA -72

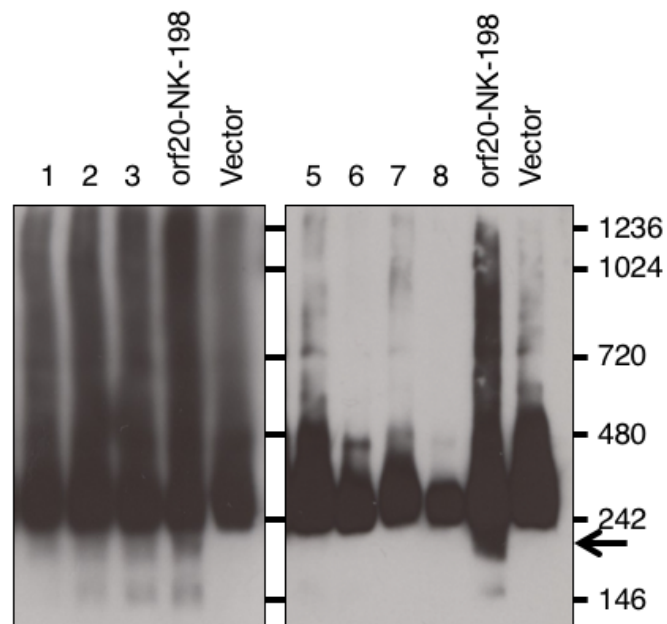
orf20L_S	ACCCAGTTCATAACTGACCCTGAAATTCAAGAATTTGGAGCAAAGTTAGCACCTTTTGTT -12
orf20_NK-305-1	ACCCAGTTCATAACTGACCCTGAAATTCAAGAATTTGGAGCAAAGTTAGCACCTTTTGTT -12

orf20L_S	GTTCAAAAATC -1
orf20_NK-305-1	GTTCAAAAATC -1

Supplementary Fig 9 Alignment of nucleotide sequences of upstream regions of *orf20L_S* and *orf20_{NK-305-1}*. Asterisks denote matched residues. Nucleotides are numbered from the first residue of the initiation codon as +1.



Supplementary Fig 10 Immunoblot analysis of proteins from transgenic suspension cells separated by SDS-PAGE. Transgenes are *orf20_{NK-305-1}::flag* (lanes 1 to 4), *orf20_{NK-305-2}::flag* (lanes 5 to 8), and a vector control. Size markers are shown on the right (kDa). Total cellular proteins were separated by SDS-PAGE and the blot was probed with αFLAG or αCOXI.



Supplementary Fig 11 Immunoblot analysis of proteins from transgenic suspension cells separated by BN-PAGE. The transgenes are *orf20_{NK-305-1}::flag* (lanes 1 to 3), *orf20_{NK-305-2}::flag* (lanes 5 to 8), *orf20_{NK-198}::flag* (*orf20*-NK-198), and a vector control. Size markers are shown on the right (kDa). Total mitochondrial proteins were separated by BN-PAGE and the blots were probed with α preSATP6. An arrow indicates the 200-kDa complex that is the hallmark of preSATP6-ORF20 interaction.

Supplementary Table 1 Nucleotide sequences of primers used in this study

Target gene	Nucleotide sequence
<i>orf20₃₀₅₋₁</i>	5'-TTCCAATTGGACTAATCTTGACTTT-3'
	5'-TTCTCCAATTTCATTCTCACG-3'
<i>ef1α</i>	5'-TGAGGCTGGTATCTCCAAGG-3'
	5'-TTGAGTACTTGGGGGTGGTG-3'
<i>Actin</i>	5'-AGACCTTCAATGTGCCTGCT-3'
	5'-ACGACCAGCAAGATCCAAAC-3'
5' UTR- <i>orf20₃₀₅₋₁</i> -3' UTR	5'-GGGGACAAGTTTGTACAAAAAAGCAGGCT AGGAATATCATAACCATT-3'
	5'-GGGGACCACTTTGTACAAGAAAGCTGGGTGGTCCTGGATTGAGGGTT-3'
5' UTR- <i>orf20₃₀₅₋₂</i> -3' UTR	5'-GGGGACAAGTTTGTACAAAAAAGCAGGCT AGGAATATCATAACCATT-3'
	5'-GGGGACCACTTTGTACAAGAAAGCTGGGTGGCAAATCTATGCTTGG-3'

Supplementary Table 2 Segregation of male fertility and o7 marker types in an F₂ population

o7	Male fertility		Sterile	Total
	Fully fertile	Partially fertile		
KK	3	2	0	5
	5			
KT	3	9	3	15
	12			
TT	0	3	3	6
	3			
Total	6	14	6	26
	20			

Supplementary Table 3 Segregation of male fertility, and o7 and s17 marker types in an admixture population

	Fully fertile			Partially fertile			Sterile			Total
	KK	KT	TT	KK	KT	TT	KK	KT	TT	
p2p2	1	4	2	0	1	0	0	0	0	8
p2p4	1	2	1	2	9	8	0	0	1	24
p4p4	0	0	0	1	0	0	3	12	7	23
Total	2	6	3	3	10	8	3	12	8	55

Supplementary Table 4 Segregation of male fertility and s17 type in a BC₂F₂ population

s17	Male fertility		Sterile	Total
	Fully fertile	Partial fertile		
p1p1	14	2	0	16
	16			
p1p4	45	4	3	52
	49			
p4p4	0	3	12	15
	3			
Total	59	9	15	83
	68			