

This is the peer reviewed version of the following article: Elbeyioglu, Firat, Roberts, Sabrina B., Spencer-Smith, Russell, Pulijala, Madhuri, Zelewska, Marta, Nebel, Jean-Christophe and Snyder, Lori AS (2016) Inversion of correia repeat enclosed elements in '*Neisseria gonorrhoeae*', which will be published in final form at <http://dx.doi.org/10.1099/mic.0.000394>. This article may be used for non-commercial purposes in accordance with the publisher restrictions.

**Title:**

Inversion of Correia Repeat Enclosed Elements in *Neisseria gonorrhoeae*.

**Authors:**

Firat Elbeyioglu\*, Sabrina B. Roberts\*, Russell Spencer-Smith<sup>^</sup>, Madhuri Pulijala, Marta A. Zelewska, Jean-Christophe Nebel<sup>‡</sup>, and Lori A. S. Snyder<sup>†</sup>

**Affiliations:**

School of Life Sciences, Pharmacy, and Chemistry, Kingston University, Kingston upon Thames, UK. KT1 2EE

<sup>‡</sup>School of Computer Science and Mathematics, Kingston University, Kingston upon Thames, UK. KT1 2EE.

\* These authors contributed equally.

<sup>^</sup> Current address: University of Illinois at Chicago, Chicago, Illinois, USA

<sup>†</sup> Corresponding Author: Kingston University, School of Life Sciences, Penrhyn Road, Kingston upon Thames, UK. KT1 2EE; +44 208 417 2003; l.snyder@kingston.ac.uk

**Short title:**

Invertible elements in *Neisseria gonorrhoeae*.

**Keywords:**

Gonococcus; phase variation; inversion; Correia Repeat Enclosed Elements; CREE

**Subject category:**

Genomics and systems biology.

**Word count:**

2,652

**Depositories:**

The GenBank accession numbers for the re-sequencing data of *Neisseria gonorrhoeae* strain NCCP11945 are SRR3547950 and (submitted SUB2025235).

**Abbreviations:**

CREE: Correia Repeat Enclosed Element

**Abstract:**

*Neisseria gonorrhoeae* is capable of causing gonorrhoea and more complex diseases in the human host. Within the gonococcal genome are over 100 copies of the IS-like Correia Repeat Enclosed Element, which has been predicted to be mobile within the neisserial genomes. Although there is evidence of ancestral movement of these elements, no previous study has provided evidence for current mobilisation. The Correia Repeat Enclosed Element has the ability to alter gene expression and regulation in many ways: by insertional mutagenesis; by introducing promoter elements; by generating mRNA processing sites, and by association with ncRNAs. Previous studies have compared the genomic locations of Correia Repeat Enclosed Elements in the *Neisseria* spp., demonstrating that otherwise identical regions have either the element or the target TA insertion site. In this study, we report for the first time movement of Correia Repeat Enclosed Elements, through inversion of the element at its chromosomal location. Analysis of Ion Torrent generated genome sequence data from *Neisseria gonorrhoeae* strain NCCP11945 passaged for 8 weeks in the laboratory under standard conditions and stress conditions revealed a total of 37 inversions: 24 were exclusively seen in the stressed sample; 7 in the control sample; and the remaining 3 were seen in both samples. These inversions have the capability to alter gene expression in *N. gonorrhoeae* through the previously determined activities of the sequence features of these elements, potentially resulting in reversible phase variable gene expression.

## Introduction:

*Neisseria gonorrhoeae* contains within its genome just over 100 copies of an inverted repeat (Snyder *et al.*, 2009), called the Correia Repeat Enclosed Element (CREE) (Liu *et al.*, 2002). The CREE have inverted repeats (Correia *et al.*, 1986 & 1988) and a characteristic core region that defines their length (Fig. 1) (Snyder *et al.*, 2009).

Longer CREE contain an IHF binding site (Fig. 1), which can influence nearby gene expression (Rouquette-Loughlin *et al.*, 2004). Transcripts containing CREE can be bound by RNase III (DeGregorio *et al.*, 2002), impacting upon the expression of genes encoded on the mRNA. CREE are involved in gene deletion events (van der Ende *et al.*, 1999; Claus *et al.*, 2007) and in gene inactivation (Mahillon & Chandler, 1998; Klee *et al.*, 2000; Parkhill *et al.*, 2000; Snyder *et al.*, 2004). CREE are also located in proximity to ncRNAs (Roberts *et al.*, 2016).

There are two promoters associated with the CREE. The Black promoter drives transcription from the end of the CREE, combining chromosomal sequence and CREE sequence to generate the -10 promoter element (Fig. 1) (Black *et al.*, 1995). These have been investigated previously to define promoter strength (Siddique *et al.*, 2011). The Snyder promoter drives transcription through the length of the CREE (Fig. 1) (Snyder *et al.*, 2003) generating targets for RNase III. Both types of promoter can be present at either, both, or neither end of the CREE. This, coupled with RNase III processing (DeGregorio *et al.*, 2003), indicates that CREE have an impact upon gene expression. Indeed, expression of the *mtrCDE* efflux pump genes in *Neisseria meningitidis* is driven by a CREE promoter and subject to RNase III processing (Rouquette-Loughlin *et al.*, 2004).

Mobile genetic elements are segments of DNA that are able to move to different locations within the genome. CREE share similar features with classic mobile elements such as transposons and IS elements, including the presence of inverted repeat ends, differential distribution between strains, and a target sequence for insertion (Correia *et al.*, 1986; Correia *et al.*, 1988; Liu *et al.*, 2002). Comparative analyses of homologous regions of chromosomes have revealed locations with and without CREE, with those without having a target TA sequence (Liu *et al.*, 2002; Snyder *et al.*, 2009). Unlike recognized mobile elements (Mahillon & Chandler, 1998;

Hurst & Werren, 2001), CREE do not code for the enzyme needed for transposition (Buisine *et al.*, 2002; Liu *et al.*, 2002).

In this study, we used next-generation sequencing to seek evidence of IS-like movement of CREE, which is hypothesized to have been involved in altering the regulatory networks, such as in the meningococcal *mtrCDE* operon. The data generated from these studies identify a never before described type of phase variable expression switching in *Neisseria* spp. due to inversion of the CREE sequence.

### **Methods:**

Ion Torrent genome sequencing. *N. gonorrhoeae* strain NCCP11945 was obtained from the original genome sequencing laboratory (Chung *et al.*, 2008) having been minimally passaged between sequencing and arriving in our laboratory. Following one passage, stock cultures were frozen at -80°C for use in these experiments. The MIC of this strain for nalidixic acid was determined to be 256 µg/ ml. A common starting inoculum of *N. gonorrhoeae* strain NCCP11945 was grown in standard conditions on GC agar (Oxoid) and in stress conditions on GC agar with 128 µg/ ml nalidixic acid at 37°C in a candle tin for a length of 8 weeks with passage of multiple colonies to fresh agar plates every 2 days. No changes in growth rate or viability were noted over this period. DNA extraction, from a sweep of colonies directly from the last passage plate for each of the two conditions, was done using the Gentra Puregene Yeast/Bact. Kit (Qiagen). Using the Ion Personal Genome Machine, Ion Express Fragment Library kit, Ion Express Template kit, and Ion Sequencing kit (Life Technologies), 1 µg of the extracted DNA was genome sequenced.

Ion Torrent data analysis. CLC Genomics Workbench version 4 and 5 (CLC bio) was used to map Ion Torrent (Life Technologies) generated sequence reads to the original reference genome sequence of *N. gonorrhoeae* strain NCCP11945 (CP001050) (Chung *et al.*, 2008) to which was added annotations of the CREE locations in that strain (Snyder *et al.*, 2009). Within the CLC Genomics Workbench, the mapping tool was used to map individual sequence reads against the reference genome. Each CREE location was then manually analysed for possible rearrangements using CLC Genomics Workbench. Sequence data for the inversions, extracted from reads aligned against the strain NCCP11945 reference genome, are

included in Supplementary Material 1. The standard growth condition genome sequence data had an average read length of 123 bp, 50.5 M aligned bases, and 22.6X coverage of the *N. gonorrhoeae* strain NCCP11945 genome sequence (accession SRR3547950). The stress growth condition genome sequence data had an average read length of 119 bp, 22.5 M aligned bases, and 10.1X coverage (accession submitted SUB2025235).

## **Results:**

The locations of all 131 previously identified CREE in *N. gonorrhoeae* strain NCCP11945 (Supplementary Table 1; Snyder *et al.*, 2009) were analyzed at the read level in the Ion Torrent next-generation sequence data. The sequenced DNA came from laboratory cultures that had been passaged for 8 weeks in standard growth conditions and for 8 weeks in stress conditions, with the culture grown in a sub-lethal concentration of antibiotic. The starting culture was minimally passaged from the strain used to generate the original genome sequence data (Chung *et al.*, 2008). This analysis identified 37 instances where read data indicated an inversion of the CREE (Fig. 2). Sequence data flanking each of these CREE is consistent with the published genome sequence for strain NCCP11945. However, the directionality in the core of CREE between the inverted Correia Repeats is inverted in a portion of the next-generation sequencing reads. Of the 37 inversions noted, 24 were exclusively seen in the sample from growth in stress conditions, 7 were seen exclusively in the standard growth condition sample, and the remaining 3 inversions were seen in both samples (Table 1). The majority (26 of 37) of the CREE inversions involved those repeats that are 104-110 bp in length, with only 4 longer IHF-binding site containing CREE showing inversions (Fig. 1).

Most CREE have a TATA sequence at both ends, however 36 of the 131 CREE in *N. gonorrhoeae* strain NCCP11945 lack the full TATA at either end of the CREE and 14 lack a terminal TA at either end (Snyder *et al.*, 2009). All but 3 of the CREE that demonstrated an inversion in the next-generation sequencing data have TATA at both ends; CREE 9244-9340 has a TATA...TAAA structure, whilst CREE 141888-142043 and CREE 570722-570875 have TATG...TATA structures. In addition, the symmetry of the CREE inverted repeats is not always perfect; as noted previously there are two types of repeat structures, with the terminal 11 bp

containing the TATA sequence being followed by either AACAAAAA or AAATTTAAA (Snyder *et al.*, 2009). Sequence variations from these consensus are also observed (Snyder *et al.*, 2009). In CREE 9244-9340, in addition to the TATA...TAAA structure, the TATA end has an adjacent sequence that is a hybrid of the two consensus sequences: AACTTTAAA, while the TAAA end has the conserved AACAAAAA structure (Supplementary Table 1). CREE 141888-142043 also has a hybrid AACTTTAAA structure, however this is at the TATG end of the CREE, whereas CREE 570722-570875, which also has a TATG end is symmetrical (Supplementary Table 1). In CREE 409189-409295 polymorphisms exist in this repeat such that it has AAATCTAAA...TTTGAATTT (Supplementary Table 1). Some of the inverted CREE have each of the two different types of repeats, one at each end, including CREE 665821-665926, CREE 933079-933183, CREE 1560064-1560168, and CREE 2106628-2106696 (Supplementary Table 1). The other CREE reported here to invert have symmetrical inverted repeats with approximately equal numbers having AACAAAAA...TTTTTGTT and AAATTTAAA...TTTAAATTT (Supplementary Table 1).

### **Discussion:**

Correia Repeat Enclosed Elements are unique to the *Neisseria* spp.; comparative evidence supports mobilization of these elements and their involvement in chromosomal rearrangement (Liu *et al.*, 2002; Snyder *et al.*, 2009; Spencer-Smith *et al.*, 2012). However, active mobilisation has never been demonstrated, therefore *N. gonorrhoeae* strain NCCP11945 was repeatedly passaged in the laboratory in an attempt to capture CREE mobilisation events. Unexpectedly, no evidence of deletion or insertion of CREE was found despite evolutionary evidence of the IS-like movement of CREE (Mazzone *et al.*, 2001; Liu *et al.*, 2002; Snyder *et al.*, 2009).

However, CREE inversions were identified within Ion Torrent generated genome sequence data for *N. gonorrhoeae* strain NCCP11945 (Table 1). It is interesting to note that the vast majority of inversions were identified in sequence data from the stress growth condition sample. Comparatively, the number of individual sequence reads that mapped to each region in the next-generation sequencing data (sequencing data depth) was actually greater in the standard conditions. There was therefore more sequence data representing the CREE regions

in the standard conditions, although more inversions were observed in the stress conditions. We are unable to conclude whether this observed difference in inversions between the two samples is due to growth in stress conditions or chance.

IHF is thought to facilitate end synapses (Buisine *et al.*, 2002), therefore the presence of the IHF binding site in the longest form of the CREE (Snyder *et al.*, 2009) could be part of the inversion mechanism. However, only 4 of the 37 inverted CREE were the larger IHF-binding site containing-CREE (Table 1). Most of the inversions were found to be in CREE  $\leq 110$ bp in length, with the majority (26 of 34 unique CREE) being 104-110 bp in length (Table 1). This suggests that IHF is not required for an inversion to occur.

The initial 11 bp at the ends of the CREE are conserved, followed by sequences of either 8 or 9 bp. Of the CREE that are reported here to have inverted, some are symmetrical, with the same 8 or 9 bp sequence at both ends, some are asymmetrical with one of each, some have polymorphisms in these sequences, and a few have one with a hybrid of the two types (Supplementary Table 1). This suggests that inverted repeat symmetry of these more internal sequences, past the terminal 11 bp, does not influence inversion.

In *N. gonorrhoeae* strain NCCP11945, 36 of the 131 CREE in the genome lack the terminal TATA at either end, and 14 of the 131 CREE lack TA at either end (Snyder *et al.*, 2009). Of the CREE that have inverted in the sequence data, most have the TATA at both ends of the sequence; those that did not, have a TATG.....TATA or TATA...TAAA end structure (2123050 – 2123202). Therefore, specificity for an inversion could lie within either of the TA nucleotides near the end of the CREE and may account for the identified TA target sequence (Liu *et al.*, 2002; Snyder *et al.*, 2009).

It is interesting to find inversions present in the standard growth condition sample, indicating that antibiotic stress is not required for an inversion event to take place within the gonococcus. Future work would be needed to understand the role of CREE inversions under normal conditions to be able to appropriately evaluate the effect of increased inversions seen in the stress sample. While these observations may be chance alone, the increased number of inversions seen in the stress sample could be a way the gonococcus adapts to the stressful conditions induced by the



quinolone, nalidixic acid, which is known to activate bacteriophage in *E. coli* (DeMarini & Lawrence, 1992). Indeed, it has been suggested that “transposition bursts” in response to stress, in which IS elements have elevated transposition activity, can contribute to the generation of advantageous genetic changes (Wu *et al.*, 2015).

At the time the sequence data was generated, the maximum read length capable on this technology was 100 bp, therefore the average read length for the two sequencing runs that were investigated here are 119 bp and 123 bp. It is possible that within the genome a more complex rearrangement than the simple inversions described here may have occurred, however no evidence of this has been uncovered from our analysis of the data.

. Orientation of the CREE-associated promoters (Black *et al.*, 1995; Snyder *et al.*, 2003; Siddique *et al.*, 2011) (Fig. 1) would change with inversion (Supplementary Figures 1-7). CREE-associated promoters drive gene expression at different levels, depending on their sequences (Siddique *et al.*, 2011). Inversion of CREE (Table 1), may be a heretofore unexplored mechanism of phase variable expression of genes in a species otherwise rich in phase variation (Snyder *et al.*, 2001; Zelewska *et al.*, 2016). In the related genus *Moraxella*, inversion events occur mediated by Piv, which inverts a region of the chromosome containing pilin-encoding genes, reversibly switching between expressed genes (Heinrich & Glasgow, 1997). Although the *irg* genes of the *Neisseria* spp. are homologous in sequence to *piv*, they have no role in neisserial pilin variation (Rozsa *et al.*, 1997; Skaar *et al.*, 2005). Whether over-expression of *irg* would increase the inversion of CREE has yet to be determined.

From the CREE shown here that inverted during our experiment, seven show a notable difference in the orientations of promoters associated with the CREE. The CREE between *leuS* and *drg* has a Black promoter 5' of the *drg* coding region (NGK\_0010; Supplementary Figure 1); inversion leaves *drg* without a CREE promoter. This would suggest phase variable expression of *drg* is mediated by inversion of the CREE. A similar situation occurs in hypothetical NGK\_0961 (Supplementary Figure 3) and hypothetical NGK\_2374 (Supplementary Figure 7), where a CREE promoter is only present in one orientation.

From previous comparisons between strains (Siddique *et al.*, 2011), it was believed that different classes of CREE were present in identical locations between the strains and that these classes would influence promoter strength and gene expression. For example, the CREE upstream of *thr* tRNA and NMA0079 is different between strains Z2491 and MC58, believed to be a result of recombination-mediated Correia end subtype switching, postulated to be a type of phase variation (Siddique *et al.*, 2011). Such a recombination-mediated event involving a distal sequence would not appear to adhere to the phase variation model that changes are readily reversible and stochastic in nature. Further, the CREE in question is accompanied by an adjacent CC sequence, therefore it would not generate the typical -10 of a promoter element as defined by Black *et al.* (1995). At the opposite end of this CREE, however, the terminal CREE TATA sequence is adjacent to TAAT, which could form a TATAAT -10 consensus. Inversion of this CREE would then alter the -35 sequence for the Black promoter toward NMA0078, potentially changing the strength of this promoter; inversion would not alter the TATACC at the NMA0079 end of the CREE.

The presence of a CREE 5' of a gene may not only remove it from its ancestral regulatory network, it may also introduce an invertible switch in expression from CREE-associated promoters. Certainly, we know that the ancestral regulatory network controlling *mtrCDE* has been altered in meningococci due to the presence of a CREE in the promoter region, which drives transcription (Rouquette-Loughlin *et al.*, 2004). The potential for such CREE and their associated promoters to invert introduces a potential phase variable gene expression system not previously described in the *Neisseria* spp.

## **Conclusions:**

This study is the first showing evidence for any form of CREE mobilisation within a single strain and the first to demonstrate rearrangements in the form of localised inversions within a single strain of *Neisseria*. Although CREE were not found to have moved to a new location, the ability of CREE to invert is an intriguing event in itself and impacts upon gene expression from CREE-associated promoters. This is

proposed here to represent a heretofore unexplored reversible phase variation gene expression switching mechanism in the *Neisseria* spp.

**Acknowledgements:**

S.B.R. was the recipient of a Kingston University PhD studentship. R.S.-S. was the recipient of a Kingston University MSc studentship.

## References:

- Black, C.G., Fyfe, J.A. & Davies, J.K. (1995).** A promoter associated with the neisserial repeat can be used to transcribe the *uvrB* gene from *Neisseria gonorrhoeae*. *J Bacteriol* **177**, 1952-1958.
- Buisine, N., Tang, C.M. & Chalmers, R. (2002).** Transposon-like Correia elements: structure, distribution and genetic exchange between pathogenic *Neisseria* sp. *FEBS Lett* **522**, 52-58.
- Chung, G.T., Yoo, J.S., Oh, H.B., Lee, Y.S., Cha, S.H., Kim, S.J., Yoo, C.K. (2008).** Complete genome sequence of *Neisseria gonorrhoeae* NCCP11945. *J Bacteriol* **190**, 6035-6036.
- Claus, H., Elias, J., Meinhardt, C., Frosch, M. & Vogel, U. (2007).** Deletion of the meningococcal *fetA* gene used for antigen sequence typing of invasive and commensal isolates from Germany: frequencies and mechanisms. *J Clin Microbiol* **45**, 2960-2964.
- Correia, F.F., Inouye, S. & Inouye, M. (1986).** A 26-base-pair repetitive sequence specific for *Neisseria gonorrhoeae* and *Neisseria meningitidis* genomic DNA. *J Bacteriol* **167**, 1009-1015.
- Correia, F.F., Inouye, S. & Inouye, M. (1988).** A family of small repeated elements with some transposon-like properties in the genome of *Neisseria gonorrhoeae*. *J Biol Chem* **263**, 12194-12198.
- De Gregorio, E., Abrescia, C., Carlomagno, M.S. & Di Nocera, P.P. (2002).** The abundant class of nemis repeats provides RNA substrates for ribonuclease III in *Neisseriae*. *Biochim Biophys Acta* **1576**, 39-44.
- De Gregorio, E., Abrescia, C., Carlomagno, M.S. & Di Nocera, P.P. (2003).** Ribonuclease III-mediated processing of specific *Neisseria meningitidis* mRNAs. *Biochem J* **374**, 799-805.
- DeMarini, D.M. & Lawrence, B.K. (1992).** Prophage induction by DNA topoisomerase II poisons and reactive-oxygen species: role of DNA breaks. *Mutat Res* **267**, 1-17.

**Heinrich, D.W. & Glasgow, A.C. (1997).** Transcriptional regulation of type 4 pilin genes and the site-specific recombinase gene, *piv*, in *Moraxella lacunata* and *Moraxella bovis*. *J Bacteriol* **179**, 7298-7305.

**Hurst, G.D. & Werren, J.H. (2001).** The role of selfish genetic elements in eukaryotic evolution. *Nat Rev Genet* **2**, 597-606.

**Klee, S.R., Nassif, X., Kusecek, B., Merker, P., Beretti, J.L., Achtman, M. & Tinsley, C.R. (2000).** Molecular and biological analysis of eight genetic islands that distinguish *Neisseria meningitidis* from the closely related pathogen *Neisseria gonorrhoeae*. *Infect Immun* **68**, 2082-2095.

**Liu, S.V., Saunders, N.J., Jeffries, A. & Rest, R.F. (2002).** Genome analysis and strain comparison of *correia* repeats and *correia* repeat-enclosed elements in pathogenic *Neisseria*. *J Bacteriol* **184**, 6163-6173.

**Mahillon, J. & Chandler, M. (1998).** Insertion sequences. *Microbiol Mol Biol Rev* **62**, 725-774.

**Mazzone, M., De Gregorio, E., Lavitola, A., Pagliarulo, C., Alifano, P. & Di Nocera, P.P. (2001).** Whole-genome organization and functional properties of miniature DNA insertion sequences conserved in pathogenic *Neisseriae*. *Gene* **278**, 211-222.

**Parkhill, J., Achtman, M., James, K.D., Bentley, S.D., Churcher, C., Klee, S.R., Morelli, G., Basham, D., Brown, D., Chillingworth, T. & other authors (2000).** Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis* Z2491. *Nature* **404**, 502-506.

**Roberts, S.B., Spencer-Smith, R., Shah, M., Nebel, J-C., Cook, R.T. & Snyder, L.A.S. (2016).** *Correia* Repeat Enclosed Elements and non-coding RNAs in the *Neisseria* species. *Microorganisms* **4(3)**, 31.

**Rouquette-Loughlin, C.E., Balthazar, J.T., Hill, S.A. & Shafer, W.M. (2004).** Modulation of the *mtrCDE*-encoded efflux pump gene complex of *Neisseria meningitidis* due to a *Correia* element insertion sequence. *Mol Microbiol* **54**, 731-741.

- Rozsa, F.W., Meyer, T.F. & Fussenegger, M. (1997).** Inversion of *Moraxella lacunata* type 4 pilin gene sequences by a *Neisseria gonorrhoeae* site-specific recombinase. *J Bacteriol* **179**, 2382-2388.
- Siddique, A., Buisine, N. & Chalmers, R. (2011).** The transposon-like Correia elements encode numerous strong promoters and provide a potential new mechanism for phase variation in the meningococcus. *PLoS Genet* **7**, e1001277.
- Skaar, E.P., Lecuyer, B., Lenich, A.G., Lazio, M.P., Perkins-Balding, D., Seifert, H.S. & Karls, A.C. (2005).** Analysis of the Piv recombinase-related gene family of *Neisseria gonorrhoeae*. *J Bacteriol* **187**, 1276-1286.
- Snyder, L.A., Butcher, S.A. & Saunders, N.J. (2001).** Comparative whole-genome analyses reveal over 100 putative phase-variable genes in the pathogenic *Neisseria* spp. *Microbiology* **147**, 2321-2332.
- Snyder, L.A., Shafer, W.M. & Saunders, N.J. (2003).** Divergence and transcriptional analysis of the division cell wall (*dcw*) gene cluster in *Neisseria* spp. *Mol Microbiol* **47**, 431-442.
- Snyder, L.A., Davies, J.K. & Saunders, N.J. (2004).** Microarray genotyping of key experimental strains of *Neisseria gonorrhoeae* reveals gene complement diversity and five new neisserial genes associated with Minimal Mobile Elements. *BMC Genomics* **5**, 23.
- Snyder, L.A., Cole, J.A. & Pallen, M.J. (2009).** Comparative analysis of two *Neisseria gonorrhoeae* genome sequences reveals evidence of mobilization of Correia Repeat Enclosed Elements and their role in regulation. *BMC Genomics* **10**, 70.
- Spencer-Smith, R., Varkey, E.M., Fielder, M.D. & Snyder, L.A. (2012).** Sequence features contributing to chromosomal rearrangements in *Neisseria gonorrhoeae*. *PLoS One* **7**, e46023.
- Van der Ende, A., Hopman, C.T. & Dankert, J. (1999).** Deletion of *porA* by recombination between clusters of repetitive extragenic palindromic sequences in *Neisseria meningitidis*. *Infect Immun* **67**, 2928-2934.

**Wu, Y., Aandahl, R.Z. & Tanaka M.M. (2015).** Dynamics of bacterial insertion sequences: can transposition bursts help the elements persist? *BMC Evol Biol* **15**, 288.

**Zelevska, M.A., Pulijala, M., Spencer-Smith, R., Mahmood, H.A., Norman, B., Churchward, C.P., Calder, A. & Snyder, L.A.S. (2016).** Phase variable DNA repeats in *Neisseria gonorrhoeae* influence transcription, translation, and protein sequence variation. *Microbial Genomics* **2**.

### Figure Legends:

Figure 1. CREE structure and associated promoters. CREE are made up of inverted repeats (blue) and a directional core region made up of segments that are present in all CREE (grey), present in all but the shortest CREE (brown), and present only in the longest CREE containing an IHF binding site (red). There are two types of functional promoters that can be associated with the CREE inverted repeats, Black promoters (black) and Snyder promoters (purple). These can be present on either strand and each CREE can potentially have all four possible promoters, none, or any combination, depending on the specific CREE sequence and adjoining sequence. The final 2 bp of the Black promoter -10 are made up of bases outside of the CREE, as is the entirety of the Snyder promoter -35 region.

Figure 2. Evidence of CREE inversion in sequence data. The assembled sequence data (PGM) is aligned against the reference genome sequence for *N. gonorrhoeae* strain NCCP11945 (NCCP11945). For this 71 bp CREE, from position 378,769 to 378,839, there are inverted repeats (blue arrows) and a core region that has directionality (grey arrow). Flanking sequence outside the CREE is in lower case.

Some assembled sequence data aligns with the genome sequence (PGM-F), but some show an inverted orientation to the CREE (PGM-R). Note that the within the CREE mismatches in the PGM-R sequence match the NCCP11945 reference sequence when its reverse complement, PGM-RC, is aligned, but then the flanking sequence does not align (underlined). The flanking sequence in PGM-R aligns, but the core has reversed directionality, therefore the CREE has inverted.

**Tables:**

Table 1. Inversion of CREE in *N. gonorrhoeae* strain NCCP11945.

Sample*	CREE location†	CREE size	Associated coding sequence(s)‡
Stress	9244 - 9340	97 bp	NGK_0009 <i>leuS</i> ; NGK_0010 <i>drg</i>
Standard	13475 - 13580	106 bp	NGK_0020; NGK_0221
Standard	137920 - 138026	107 bp	NGK_0171; NGK_0172
Stress	141888 - 142043	156 bp	NGK_0177; NGK_0178
Stress	210653 - 210757	105 bp	NGK_0263 <i>marR</i> family; NGK_0264 <i>mmsB</i>
Both**	315775 - 315881	107 bp	NGK_0377; NGK_0378
Stress	332258 - 332361	104 bp	NGK_0402; NGK_0403
Stress	378769 - 378839	71 bp	NGK_0456
Stress	409189 - 409295	107 bp	NGK_0499; NGL_0500 <i>lep</i>
Stress	518804 - 518908	105 bp	NGK_0622 <i>dnaB</i>
Stress	570722 - 570875	154 bp	NGK_0693 <i>opa</i> ; NGK_0694; NGK_0695
Stress	636542 - 636648	107 bp	NGK_0784; NGK_0785
Standard	665821 - 665926	106 bp	NGK_0816; NGK_0817
Both	684759 - 684863	105 bp	NGK_0842 <i>irg7</i> ; NGK_0843
Stress	748921 - 749074	154 bp	NGK_0920; NGK_0921
Stress	773275 - 773380	106 bp	NGK_0951; NGK_0952
Stress**	782431 - 782536	106 bp	NGK_0960; NGK_0961
Standard	933079 - 933183	105 bp	NGK_1137; NGK_1138 <i>recA</i>
Stress	949754 - 949858	105 bp	NGK_1162; NGK_1163; NGK_1164
Standard	958334 - 958439	106 bp	NGK_1175; NGK_1176 <i>edd</i>
Standard	1073168 - 1073274	107 bp	NGK_1304; NGK_1305 <i>plsC</i>
Both	1253475 - 1253579	105 bp	NGK_1511; NGK_1512
Stress	1324830 - 1324935	106 bp	NGK_1591; NGK_1592; NGK_1593



Stress	1465780 - 1465934	155 bp	NGK_1767 <i>potD</i>
Stress	1560064 - 1560168	105 bp	NGK_1874; NGK_1875; NGK_1876 <i>ompA</i>
Stress	1577266 - 1577372	107 bp	NGK_1899; NGK_1900
Stress	1753230 - 1753334	105 bp	NGK_2117; NGK_2118 <i>mviN</i>
Stress**	1882841 - 1882913	73 bp	NGK_2269; NGK_2270 <i>mafA</i>
Standard	1974277 - 1974383	107 bp	NGK_2369; NGK_2370 <i>gltS</i>
Stress	1977219 - 1977328	110 bp	NGK_2373; NGK_2374
Stress	2106628 - 2106696	69 bp	NGK_2511
Stress	2161831 - 2161937	107 bp	NGK_2573; NGK_2574 <i>app</i>
Stress	2172837 - 2172941	105 bp	NGK_2581 <i>hpuA</i> ; NGK_2582
Stress**	2189554 - 2189660	107 bp	NGK_2603 <i>arnT</i> ; NGK_2604 <i>regF</i>

\* The inversion was detected in a sample cultured in either standard or stress growth conditions, or both.

\*\* Sequence read data spans the CREE to include flanking sequence data from both ends.

† From GenBank Accession Number CP001050

‡ CDSs from CP001050 annotated with NGK numbers.

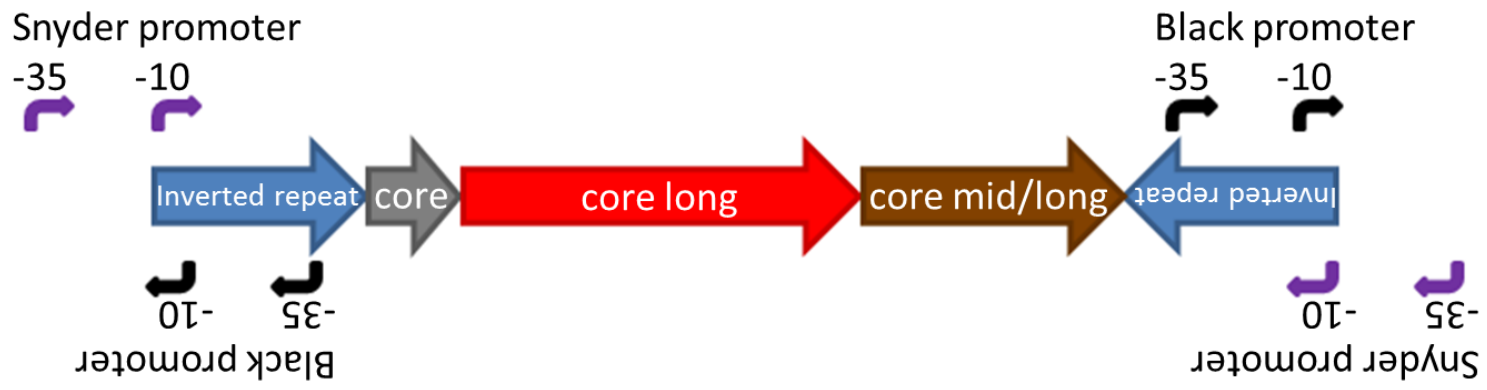


Figure 1

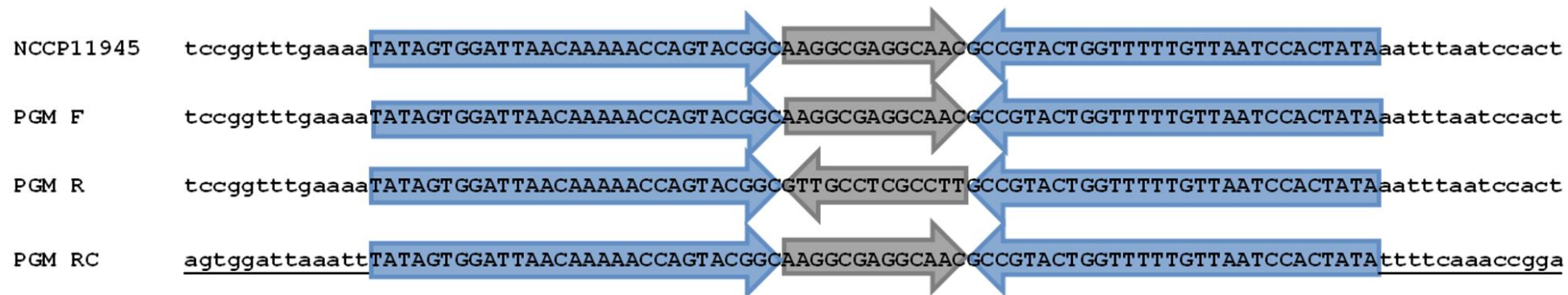


Figure 2

# Supplementary Material 1

>CREE 9244-9340 NCCP11945 Stress

taaggcagattggtgtctgcaagacagacctcaaatatagcggattaacaaaaatcaggacaaggagacgaagccgagacagcaggcaaggcgggcaacccgtaccggtttaaagtaacccg

>CREE 141888-142043 NCCP11945 Stress

taaatgccgtctgaaaaacggggaatccgtgcaggagaaaaaacccgttaaggtatggtggattaactttaaaccggtacggcgttgctcgctcagatcaaagagaacgattttctaaggtgctgaagcaccaagtgat

>CREE 210653-210757 NCCP11945 Stress

ttatgactatagtgaaatacaaaaaccggtacggcgttgctcgcttgcgtactattgtactgtctgaggcttgcctgccttgctgattttg

>CREE 315775-315881 NCCP11945 Stress

tgcagcccgatcgggattttatagtggttaaaatcaaatcaggacaaggcggcgaagccgagacagatacaaatagtagcgaaggcaggcaacccgtaccggtttaaatttaaccactataccgccgagatgatgccgtaactctccgtaaac

tcgggcagatgatgtcgccgattttccctgacaatgcag

>CREE 332258-332361 NCCP11945 Stress

agttttcaaaattcatcagaaattatgatttaattggcataaattttgagattatagtggttaacaaaaaccagtagcggcgttgctcgcttgcctgactattgtactgtctgaggctcggccttgctgattttt

>CREE 378769-378839 NCCP11945 Stress

caaaaaagacgaataaatatgaatggaatgccgtctgaaggtcagacggcattttccggttggaaatagtggttaacaaaaaccagtagcggcgttgctcgcttgcctgactg

>CREE 409189-409295 NCCP11945 Stress

tcgataatgccggcaaggcgtgcatatagtggttaaaatcaaaagtcagcctggcctgccttgcctgactattgtactgtcggcgttgcctgccttggcc

>CREE 518804-518908 NCCP11945 Stress

ctaataattttttactttatagtggttaacaaaa-ccagtagcggcgttgctcgcttgcctgactatctgtaccgtctgaggccttgcctgctgattttt

>CREE 570722-570875 NCCP11845 Stress

taatcagcgtatagcggattaacaaaaccggtacggcgttggccccccccgctcaaagggaacgcttctaggtgctgaagcaccgggaggatcggaaccgtactattgtactgctgaggcccc

>CREE 636542-636648 NCCP11945 Stress

ctgacggcgtaaacacttgcgcaagaagccttggcagcctgcccgtggcgttatagtggttaaaatcaaaaccaggaacggcggcgaaccgcagacagatacaaatagtagcgaaggcaggcaacg

>CREE 684759-684863 NCCP11945 Stress

atttatgagtcgtcattcccgaaaagcgggaatcagtttttatagtggttaacaaaaaccaggaacggcgaagcgtcagacagatacaaatagtagcgaaggcaggcaacgc

>CREE 748921-749074 NCCP11945 stress

gaaaaatagcggattaacaaaaaccggtacggcgttccccccccggccaaagggaacgattccctaaggtaatgaagcaccggggaaccggtccgtaccattgtactgctgaggccgcttgcctgattttgta

>CREE 773275-773380 NCCP11945 Stress

taatctataatggcttgcgttgaataactgctgtgccaattatagtggttaaaatcaaaaccggtacggcgttgctcgcttgcctgactatctgtactgtctgaggcctcggccttgc

>CREE 782431-782536 NCCP11945 Stress

gttttgccttattatagtggttaaaatcaaaaccggtacggcgttgctcgcttgcctgactattgtactgctgaggccttgcctgatttaaatttaaccactataatccgccaatgccgtctgaagccttttatgcttca

>CREE 949754-949858 NCCP11945 Stress

atcaaaaatcaggacaaggcgacgaagccgagacagatacaaatagtagcgaaggcaggtaaacccgtaccggtttttgtaaccactataaaatctccagcgggtttgggtgttggcgagccgcgcccgatagg

>CREE 1253475-1253579 NCCP11945 Stress

gacacataaagcaccgcctatgtgctcctgatttgaagggttacaccctcccaaatatagtggttaacaaaaaccggtacggcgaagccgagacagatacaaatagtagcgaaggcaggcaac

>CREE 1324830-1324935 NCCP11945 Stress

cgttcagcgttggcatattcggatgtatcggtatagtgaaatcaaaatcaaaaccggtacggcgaagccgagacagatacaaatagtagcgaaggcaggcaacgcctgac

>CREE 1465780-1465934 NCCP11945 Stress

gtcttccctgattatagcggattaacaaaaaccggtacggcgttggcccccttagctcaagagaacgattctctaaggtgctgaagcaccgaagtgatcggttccgtactattgtactgtctga

>CREE 1560064-1560168 NCCP11945 Stress

aaatttaaccggtacggcgttgctcgcttgcctgactattgtactgtctgaggccttgcctgattttgtaaccactataatcggttgaaaactctgattttaaggcggtaggatgtgggt

>CREE 1577266-1577372 NCCP11945 Stress

aaatttaaccagtagcgttgcctcgcttgcctgactatctgtactgtctgaggccttgcctgatttaaatttaaccactatacagccaacaa

>CREE 1753230-1753334 NCCP11945 Stress  
ttgaaacctgcccttagggcgtagaatcagactttatagtgattaacaaaaaccggtacggcggtgcctgccttgccgtactattgtactgtctgcggttgggcgcttgcctgattttgttaa  
>CREE 1882841-1882913 NCCP11945 Stress  
cacactggtccactgctgattgatgattgacatagggctggcagatatagtcgcatatggcttcagacagccatttattatgagattatagtgattaaattaaaccggtacggcaaggcgaggcaacgccgtaccggtttaaatttaaccactatattacaattttgatacaattgttttcatcaaaggagaaatctatgcgggc  
>CREE 1977219-1977328 NCCP11945 Stress  
tggaaaattcttctcaggctttccctgccgaatggacaatagcggattaacaaaaaccggtacggcggtgcctgccttgccgtactattgtactgtctgcggttggccttgcctgatttttg  
>CREE 2106682-2106696 NCCP11945 Stress  
tcgggaaagtttctgctgggttgaaccggtgggagtggtcggggaatgacgggatatttgcgtttatagtgattaaattaaaccggtacggcggtgcctgccttgctgattttgttaatcc  
>CREE 2161831-2161937 NCCP11945 Stress  
gatttttagccatgttcggcgagtggggaaatctatagtgattaaattaaaccggtacggcggtgcctgccttgccgtactattttactgtctgcggttggccttgcctgattta  
>CREE 2172837-2172941 NCCP11945 Stress  
aaggcgaggcaacgccgtaccggttttggtaaccactatgatgacggcggtccggcgcggtggggcttgcggcg  
>CREE 2189554-2189660 NCCP11945 Stress  
ctctgtgtgatcctgcaaaaaccgctgatttatagtgattaaattaaaccggtacggcgaggcaacgccgtaccggtttaaatttaaccactatataaaggtttaacgccgcaattataacgcatttccga  
>CREE 13475-13580 NCCP11945 Standard  
caaatagtagcggcaaggcgaggcaacgccgtactattgtactgtctgcggttcgcccttgccgattaaattaattcactatatttcatatacttattatctaactttccgtaca  
>CREE 137920-138026 NCCP11945 Standard  
tccggtacggcggtgcctgccttgccgtactactgtactgtctgcggttcgcccttgcctgattaaatttaaccactataaacatcgtcctgcctgacctgcactcaaattgcggata  
>CREE 315775-315881 NCCP11945 Standard  
tcggacatcatctgcccagtttacggaaggttacggcgcatcatctgcggggtatagtgattaaattaaaccggtacggcggtgcctgccttgccgtactattgtactgtct  
>CREE 665821-665926 NCCP11945 Standard  
aacaaaaatcaggacaaggcgacgaagccgcagacagtacaaatagtagcggcaaggcgaggcaacgcgtgaccggtttaaatttaaccactatccggctttccatgcttaaacagcactacctttca  
>CREE 684759-684863 NCCP11945 Standard  
ggacaaggcgccgaagctgcagacagtacaaatagtagcggcaaggcgaggcaacgccgtacctggtttgtaatacactataaattcagccatccgataaattcctgtgcttagca  
>CREE 933079-933183 NCCP11945 Standard  
ccgtttttttgagccctccggacggctgatttgattttggtatagtgattaaattaaaccagtagcggcggtgcctgccttgccgtactattgtactgtctgcggtctccgcttgcctgatttt  
>CREE 958334-958439 NCCP11945 Standard  
ggacaaggcgggcgagccgcagacagcacaatagtagcggcaaggcgaggcaacgccggaccggtttaaatttaaccactatattttaaaggtagccgtctgaaagtgggaagcgggtgctata  
>CREE 1073168-1073274 NCCP11945 Standard  
tcaaaagaaatgccgaaggatgtttgttatagtgattaaatttagaccggtacggcggtgcctgccttgccgtactattttactgtctgcggttgcggccttgcctgattaaattta  
>CREE 1253475-1253579 NCCP11945 Standard  
tcaaatagtagcggcaaggcgaggcaacgccgtaccggttttggtaaccactataaagtctgattctaccgccctaaagggcaggttcaaccgaaaaggaacagatgaaagcata  
>CREE 1974277-1974383 NCCP11945 Standard  
gacaaggcgacgaagccacagacagtacagatagtagcggcaaggcgagtcaacgatgactggtttaaatttaaccactatacaactatttttacaacaaaatgccgtccgaaactccgttcagac

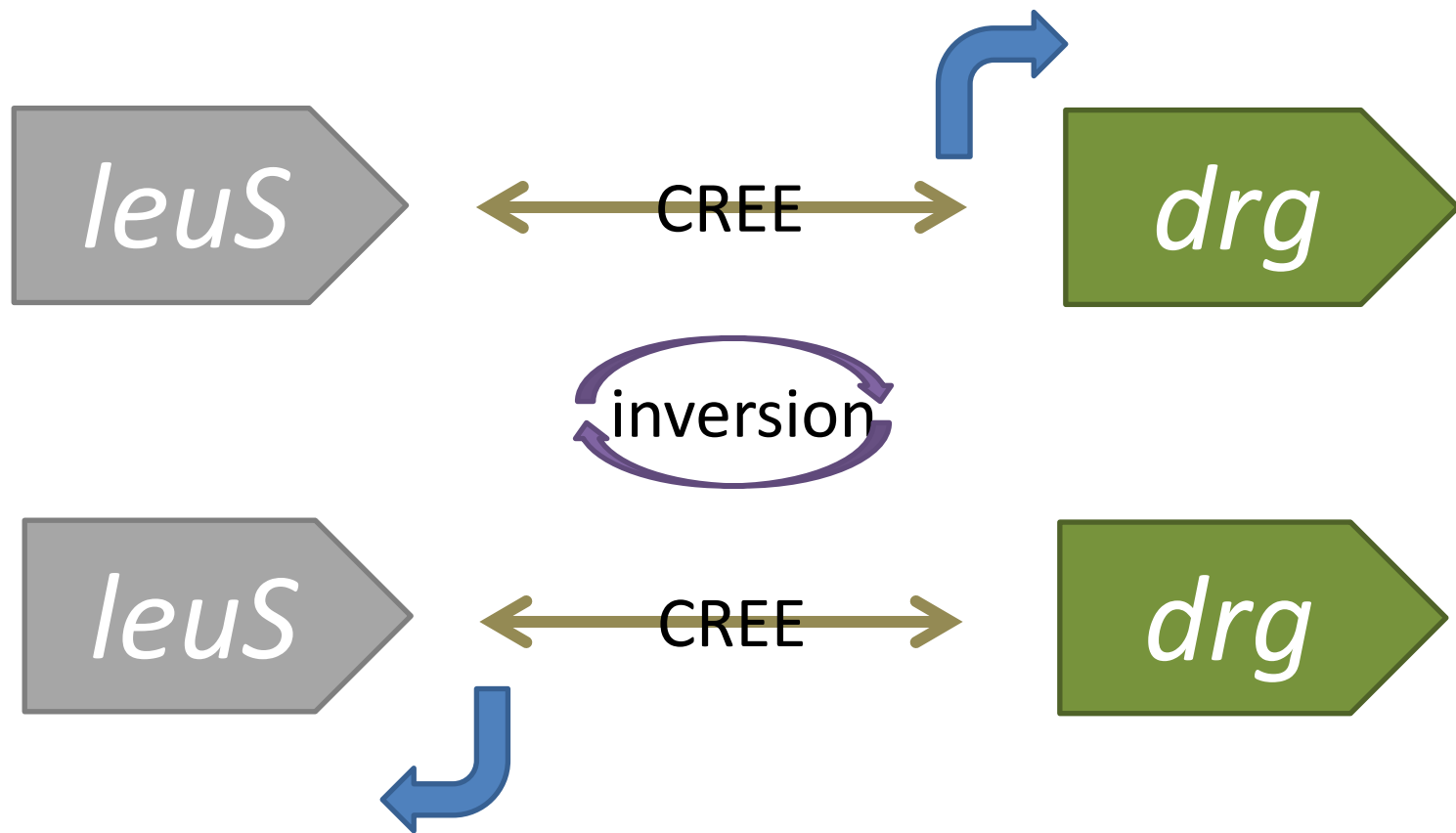
Supplementary Table 1. CREE in <i>N. gonorrhoeae</i> strain NCCP11945 as reported in Roberts <i>et al.</i> , <i>Microorganisms</i> 2016, 4(3), 31.			
beginning	end	size	sequence*
9244	9340	97	tatagcggattaactttaaacccggtacggcgttgccccgccttgccctgctgtctcgcgcttctgtcctgtcctgat ttttgttaactactataa
13475	13580	106	tatagtgaatataaatttaaacccggtacggcgttgccctgccttgccgtactatttgtactgtctgcggttgcgcccctgtcccgatttaaat ttaattccactata
19526	19680	155	tatagcggattaactttaaacccggtacggcgttgccccccccggctcaaagggaaacggttccctaaggcgcccaagcaccgggcgcaaccgggtccggtaccatttgtactgctgcgcccgccgcttgcctgat ttttgttaactcctctatg
36007	36112	106	tatagtggattaaatttaaacccagtcacagcgttgccctgccttgccgtactatttgtactgtctgcggtctgcctgcctgatttttgttaaccactata
38820	38973	154	tatagcggattaaacaaaaaccggtacggcgttgccccccccggctcaaagggaaacggttccctaagggtgatggagcgcggcggaatcggttccgtaccattcgtactgctgcgcccgccgcttgcctgat ttttgttaactccgctatg
77264	77417	154	tatggcggattaacaaaaaccggtacggcgttgccccccccggctcaaagggaaacggttccctaagggtgctgaagcaccgggcggaatcggttccgtactatttgtactgctgcgcccgccgcttgcctgat ttttgttaactccgctata
83102	83256	155	tatagcggattaaacaaaaaccggtacggcgttgccccccccgctcaaagggaaacggttccctaaggcaccacaagcacaagtgatcggttccgtactatttgtactgtctgcggttcgctgccttgcctgat tttcgttaactccgctata
99783	99937	155	tgttgcggattaacaaaaaccggtacggcgttgccccccccggctcaaagggaaacggttccctaaggcgcccaagcaccgggcaaccgggttccgtaccatcgtactgctgcggttcgccccttgcctgat ttttgttaactccgctata
102100	102254	155	tatagcggattaaacaaaaaccggtacggcgttgccccccccgctcaaagggaaacggttccctaagggttcaagcaccgggcaaccgggttccgtaccatcgtactgctgcggttcgccccttgcctgat ttttgttaactccgctata
130155	130309	155	tatagtggattaaacaaaaaccggtacggcgttgccccccccgctcaaagggaaacggttccctaagggtgatggagcaccagtgaaatcggttccgtactatttgtactgctgcggttcgccccttgcctgat ttttgttaactccactata
137920	138026	107	tatagtggattaaatttaaacccggtacggcgttgccctgccttgccgtactatctgtactgtctgcggttcgccccttgcctgat ttaaat ttaactccactata
141888	142043	156	tatggtggattaactttaaacccggtacggcgttgccctgcctcagatcaaagagaacgatttctaaaggtgctgaagcaccagtgatcggttccgtactatctgtactgtctgcggttcgccccttgcctgat ttttgttaactccactata
156099	156252	154	tatagcggattaaacaaaaaccggtacggcgttgccccccccggctcaaagggaaacggttccctaagggtgatggagcgcggcggaatcggttccgtaccattcgtactgctgcggttcgccccttgcctgat ttttgttaactccgctatg
163256	163410	155	tatagtggattaacaaaaaccggtacggcgttgcctgccttgccgtactatttgtactgtctgcggttcgctgccttgcctgat ttttgttaactccactata
164786	164891	106	tatggtggattaaatttaaacccggtacagcgttgccctgccttgccgtactatctgtactgtctgcggttcgctgccttgcctgat ttttgttaactccgctta
198812	198916	105	tagagtgattaaatttaaacccggtacggcgttgccctgccttgccgtactatttgtactgtctgcggttcgccccttgcctgat ttttgttaactactata
210653	210757	105	tatagtggattaaacaaaaaccggtacggcgttgcctgccttgcctactatttgtactgtctgcggttcgctgccttgcctgat ttttgttaactccactata
217937	218043	107	TATAGTGGATTAAATTAAACCCGGTACGGCGTTGCCCTGCCTTGCCGTACTACTGTCTGCGGCTTCGTCGTTTTGTCCTGATTTAAATTTAATCCACTATA
22619	226324	106	tatagtggattaaatttaaacccggtacggcgttgccctgccttgccgtactatttgtactgtctgcggttcgccccttgcctgat ttttgttaactccactata
234038	234145	108	tatagtggattaaacaaaaaccggtacggcgttgcctgccttgcctactatttgtactgtctgcggttcgctgccttgcctgat ttaaat ttaactccactata
277340	277445	106	tatagtggattaaacaaaaaccggtacggcgttgccccccccgctcaaagggaaacggttccctaagggtgatggagcgcggcggaatcggttccgtaccattcgtactgctgcggttcgccccttgcctgat ttttgttaactccactata
296830	296935	106	tatagtggattaaatttaaacccggtacggcgttgccctgccttgccgtactatttgtactgtctgcggttcgctgccttgcctgat ttttgttaactccactatt
315775	315881	107	tatagtggattaaatttaaacccggtacggcgttgccctgccttgccgtactatttgtactgtctgcggttcgccccttgcctgat ttaaat ttaactccactata
321931	322084	154	tatagcggattaaacaaaaaccggtacggcgttgccccccccggctcaaagggaaacggttccctaagggtgatggagcgcggcggaatcggttccgtaccattcgtactgctgcgcccgccgcttgcctgat ttttgttaactccgctatg
332258	332361	104	tatagtggattaacaaaaaccggtacggcgttgcctgccttgccgtactatttgtactgtctgcggttcgccccttgcctgat ttttgttaactccactata
378769	378839	71	tatagtggattaaacaaaaaccggtacggcgttgcctgccttgccgtactatttgtactgtctgcggttcgccccttgcctgat ttttgttaactccactata
409189	409295	107	tatagtggattaaatttaaacccggtacagcgttgccctgccttgccgtactatttgtactgtctgcggttcgctgccttggcctgat tttgaat ttaactccactata
465354	465507	154	tatagcggattaaacaaaaaccggtacgtcggttgcctgccttgccgtactatttgtactgtctgcggttcgctgccttgcctgat ttttgttaactccgctata
472397	472550	154	tatagtgaattaaacaaaaaccggtactcggttggtcgccttgccgtactatttgtactgtctgcggttcgccccttgcctgat ttttgttaactccactata
495688	495844	157	tatagtggattaaatttaaacccggtacggcattacctgccttgccgtactatttgtactgtctgcggttcgctgccttgcctgat ttaaat ttaactccactata
497451	497605	155	tatagtgaattaaacaaaaaccggtacggcgttgccccccccggctcaaagggaaacggttccctaagggtgctgaagcaccagtgatcggttccgtactatctgtactgtctgcggttcgctgccttgcctgat ttttgttaactccactata
513116	513222	107	tatagtggattaaatttaaacccggtacggcgttgcctgccttgccgtactatttgtactgtctgcggttcgctgccttgcctgat ttaaat ttaactccactata
518804	518908	105	tatagtggattaaacaaaaaccggtacggcgttgcctgccttgccgtactatctgtactgtctgcggttcgccccttgcctgat ttttgttaactccactata
562894	563048	155	tatagtggattaaacaaaaaccggtacggcgttgccccccccgctcaaagggaaacggttccctaagggtgctgaagcaccagcgaaccgggtccgtactatttgtactgtctgcggttcgctaccattgcctgat ttttgttaactccactata
570722	570875	154	tatggcggattaaacaaaaaccggtacggcgttgccccccccggctcaaagggaaacggttccctaagggtgctgaagcaccggcggaatcggttccgtactatttgtactgctgcgcccgccgcttgcctgat ttttgttaactccgctata
611024	611177	154	tatggcggattaaacaaaaaccggtacggcgttgcctgccttgccgtactatttgtactgtctgcggttcgccccttgcctgat ttttgttaactccgctata
613437	613509	73	tatagtggattaaatttaaacccggtacggcgttgcctgccttgccgtactatttgaat ttaactccactata
625567	625673	107	tatagtggattaaatttaaacccggtacggcgttgccctgccttgccgtactatttgtactgtctggttgcctgccttgcctgat ttaaat ttaactccactata
627722	627827	106	tatagtggattaaatttaaacccggtacggcgttgccctgccttgccgtactatttgtactgtctgcggttcgccccttgcctgat ttttgttaactccactata
634929	635082	154	tatagcggattaaacaaaaaccggtacggcgttgccccccccggctcgaagaggacgattccctaagggtgatggagcgcggcggaatcggttccgtaccattcgtactgctgcgcccgccgcttgcctgat ttttgttaactccgctatg
636542	636648	107	tatagtggattaaatttaaacccggtacggcgttgccctgccttgccgtactatttgtactgtctgcggttcgccccttgcctgat ttaaat ttaactccactata
665821	665926	106	tatagtggattaaatttaaacccggtacagcgttgccctgccttgccgtactatttgtactgtctgcggttcgctgccttgcctgat ttttgttaactccactata
684759	684863	105	tatagtggattaaacaaaaaccggtacggcgttgcctgccttgccgtactatttgtactgtctgcggttcgctgccttgcctgat ttttgttaactccactata
699814	699920	107	tatagtggattaaatttaaacccggtacagcgttgccctgccttgccgtactatctgtactgtctgcggttcgctgccttgcctgat ttaaat ttaactccactata
723573	723726	154	tatagcggattaaacaaaaaccggtacggcgttgccccccccggctcaaagggaaacggttccctaagggtgatgaagcaccgggcggaatcggttccgtaccatttgtaccgtctgcgcccgccgcttgcctgat ttttgttaactccgctatg
724164	724317	154	tatagtgaattaaacaaaaaccggtacggcgttgccccccccggctcaaagggaaacggttccctaagggtgatggagcgcggcggaatcggttccgtaccatttgtactgctgcgcccgccgcttgcctgat ttttgttaactccactata
748921	749074	154	tatagcggattaaacaaaaaccggtacggcgttgcctgccttgccgtactatttgtactgctgcggttcgctgccttgcctgat ttttgttaactccactata
773275	773380	106	tatagtggattaaatttaaacccggtacggcgttgccctgccttgccgtactatctgtactgtctgcggttcgctgatttaaat ttaactccactata
777827	777932	106	tatagtggattaaatttaaacccggtacggcgttgccctgccttgccgtactatctgtactgtctgcggttcgccccttgcctgat ttttgttaactccactata
782431	782536	106	tatagtggattaaatttaaacccggtacggcgttgccctgccttgccgtactatttgtactgctgcggttcgccccttgcctgat ttaaat ttaactccactata
790334	790439	106	tatagtggattaaatttaaacccggtacggcgttgccctgccttgccgtactatctgtactgtctgcggttcgctgccttgcctgat ttttgttaactccactata
797704	797808	105	tatagtgaattaaatttaaacccggtacggcgttgccctgccttgccgtactatttgtactgtctgcggttcgccccttgcctgat ttttgttaactccactata

816422	816575	154	tatagcggattaacaaaaaccggtacggcggtgccccgccccggctcaaagggaaacggttccctaaaggcgccaagcaccgggccaacgggtccataccattcgtactgctgcggccccgcgcctgtcctgatttttgtaataccgctata
924580	924733	154	tatagcggattaacaaaaaccggtacggcggtgccccgccccggctcaaagggaaacgattccctaaaggctgatgaagcaccgggccaacgggtccgtaaccattcgtactgcctgcggtccgcgcctgtcctgatttttgtaataccgctatg
933079	933183	105	tatagtggtataaatttaaaccagttacggcggtgctctgccttgcgcgtactatttgtactgtctgcggctctcgcctgtcctgatttttgtaataccgctata
949754	949858	105	tatagtggtataaacaaccggtacggcggttacctgccttgcgcgtactatttgtactgtctgcggctctcgtcgcctgtcctgatttttgtaataccactata
958334	958439	106	tatagtggtataaatttaaaccggtccggcggttgctctgccttgcgcgtactatttgtgctgtctgcggctcgcgccttgcctgatttaaatttaattccactata
961596	961701	106	tatagtggtataaatttaaaccagttacggcggttacctgccttgcgcgtactatttctactgtcggctctcgttgccttgcctgatttaaatttaattccactata
1033492	1033597	106	tatagtggtataaatttaaaccggtacggcggttgctctgccttgcgcgtactatctgtactgtctgcggctctcgtcgccttgcctgatttttgtaataccactata
1073168	1073274	107	tatagtggtataaatttagaccggtacggcggttgctctgccttgcgcgtactatttttactgtctgcggctctgcgccttgcctgatttaaatttaattccactata
1115108	1115213	106	tatagtagattaaatttaaaccggtccggcggttacctgccttgcgcgtactatttgtactgtctgcggctctcgtcgccttgcctgatttttgtaataccactata
1161730	1161883	154	tatggtgaattaacaaaaaccggtacggcggtgccccgccttagctcaaagggaaacgattccctaaaggctgatgaagcaccaggccaacgggtccgtaaccatttgtactgcctgcggccccgcgccttgcctgatttttgtaataccactata
1171932	1172036	105	tagagtggtataaatttaaaccggtacggcggttgctctgccttgcgcgtactatttgtactgtctgcggctcgcgccttgcctgatttttgtaataccgctata
1235244	1235314	71	tatagtggtatacaaaaaaccggtacggcggttgctctgccttgcgcgtactgtttttgtaataccgctata
1236171	1236324	154	tatggcggatgaacaaaaaccggtacgtcgttgccccgccccggctcaaagggaaacggttccccaaaggatgatgaagcaccgggccaacgggtccgtaaccatttgtactgcctgcggccccgcgccttgcctgatttttgtaataccgctata
1253475	1253579	105	tatagtggtatacaaaaaaccggtacggcggttgctctgccttgcgcgtactatttgtactgtctgcggctctcgcgccttgcctgatttttgtaataccactata
1262895	1263048	154	tatggcggattaacaaaaaccggtacggcggtgccccgccccggctcaaagggaaacggttccctaaaggatgatgaagcaccgggccaacgggtccgtaaccattcgtactgcctgcggccccgcgccttgcctgatttttgtaataccgctata
1290634	1290787	154	tatagcggattaacaaaaaccggtacggcggtgccccgccccggctcaaagggaaacgattccctaaaggatgatgaagcaccgggccaacgggtccgtaaccattcgtactgcctgcggccccgcgccttgcctgatttttgtaataccgctatg
1313578	1313684	107	tatagtggtataaatttaaaccggtacggcggttgctctgccttgcgcgtactatttgtactgtctgcggctctcgcgccttgcctgatttaaatttaattccactata
1321069	1321174	106	tatagtggtatacaaaaaaccggtacggcggttgctctgccttgcgcgtactatctgtactgtctgcggctctcgtcgccttgcctgatttaaatttaattccactata
1324830	1324935	106	tatagtgactataaatttaaaccagttacggcggttgctctgccttgcgcgtactatttgtactgtctgcggctcgcgccttgcctgatttaaatttaattccactata
1326495	1326648	154	tatagtggtatacaaaaaaccggtacggcggtgccccgccttagctcaaagagaacgattccctaaaggcgccaagcaccgggccaacgggtccgtaaccatttgtactgtctgcggccccgcgccttgcctgatttttgtaataccactata
1347106	1347260	155	tatagcggattaacaaaaaccggtacggcggttgctctgccttgcgcgtactaaagagaacgattcctaaaggatgatgaagcaccaggccaacgggtccgtaaccatttgtactgtctgcggctctcgttgccttgcctgatttttgtaataccactata
1426873	1427026	154	tatagcggattaacaaaaaccggtacggcggtgccccgccccggctcaaagggaaacgattccctaaaggatgatgaagcaccgggccaacgggtccgtaaccattcgtactgcctgcggccccgcgccttgcctgatttttgtaataccgctatg
1435613	1435768	156	tatagtggtataactttaaaccagttacggcggttgctctgccttagctcaaagagaacgattccctaaaggatgatgaagcaccaggccaacgggtccgtaaccatttgtactgtctgcggctctcgccttgcctgatttttgtaataccgctata
1465780	1465934	155	tatagcggattaacaaaaaccggtacggcggtgccccgccttagctcaaagagaacgattcctaaaggatgatgaagcaccaggccaacgggtccgtaaccatttgtactgtctgcggctctcgccttgcctgatttttgtaataccactata
1490568	1490721	154	tatggcggattaacaaaaaccggtacggcggtgccccgccccggctcaaagggaaacggttccccaaaggatgatgaagcaccgggccaacgggtccgtaaccatttgtactgcctgcggccccgcgccttgcctgatttttgtaataccgctata
1499551	1499703	153	tatagtggtataaatttaaaccggtacgggttgccccgccccggctcaaagggaaacgattccctaaaggcgccaagcaccgggccaacgggtccgtaaccatttgtactgcctgcggccccgcgccttgcctgatttttgtaataccgctata
1519344	1519497	154	tatagcggattaacaaaaaccggtacgtcgttgccccgccccggctcaaagggaaacggttccctaaaggcgccaagcaccgggccaacgggtccgtaaccattcgtactgcctgcggccccgcgccttgcctgatttttgtaataccgctata
1540112	1540216	105	tatagtggtataaatttaaaccagttacggcggttgccccgccttgcctgactatttgtactgtctgcggctcgcgccttgcctgatttttgtaataccactata
1541956	1542110	155	tatagtggtatacaaaaaaccggtacggcggttgccccgccttagctcaaagagaacgattcctaaaggatgatgaagcaccaggccaacgggtccgtaaccatttgtactgtctgcggctctcgccttgcctgatttttgtaataccactata
1544817	1544971	155	tatagtggtatacaaaaaaccggtacggcggttgccccgccccggctcaaagagaacgattcctaaaggatgatgaagcaccaggccaacgggtccgtaaccatttatactgtctgcggctctcgccttgcctgatttttgtaataccactata
1553685	1553853	169	tatagtggtatacaaaaaaccggtacggcggttgctctgccttgcgcgtactatctgtactgtctgcggctctcgtcgccttgcctgatttttgtaataccagccttgcgtactatctgtactgtctgcggctctcgccttgcctgatttttgtaataccgctata
1560064	1560168	105	tatagtggtataaatttaaaccggtacggcggttgctctgccttgcgcgtactattgtactgtctgcggctctcgtcgccttgcctgatttttgtaataccactata
1577266	1577372	107	tatagtggtataaatttaaaccagttacggcggttgctctgccttgcgcgtactatttgtactgtctgcggctctcgtcgccttgcctgatttaaatttaattccactata
1584653	1584759	107	tatagtggtataaatttaaaccggtacggcggttgccccgccttgcgcgtactatctgtactgtctgcggctctcgccttgcctgatttttgtaataccactata
1620383	1620536	154	tatggcggattaacaaaaaccggtacggcggtgccccgccccggctcaaagggaaacggttccctaaaggatgatgaagcaccgggccaacgggtccgtaaccattcgtactgcctgcggccccgcgccttgcctgatttttgtaataccgctata
1698964	1699117	154	tatagtgattaacaaaaaccggtacggcggttgccccgccccggctcaaagaggacgattccctaaaggatgatgaagcaccaggccaacgggtccgtaaccattcgtactgcctgcggccccgcgccttgcctgatttttgtaataccactata
1724282	1724435	154	tatggcggattaacaaaaaccggtacggcggtgccccgccccggctcaaagggaaacgattccctaaaggatgatgaagcaccgggccaacgggtccgtaaccattcgtactgcctgcggccccgcgccttgcctgatttttgtaataccgctata
1753230	1753334	105	tatagtggtatacaaaaaaccggtacggcggttgctctgccttgcgcgtactatttgtactgtctgcggctctcgcgccttgcctgatttttgtaataccactata
1760746	1760852	107	tatagtggtataaatttaaaccggtacggcggttgctctgccttgcgcgtactatctgtactgtctgcggctctcgcgccttgcctgatttaaatttaattccactatg
1787782	1787935	154	tatagcggattaacaaaaaccggtacggcggtgccccgccttagctcaaagggaaacgattccctaaaggcgccaagcaccgggccaacgggtccgtaaccatttgtactgcctgcggccccgcgccttgcctgatttttgtaataccgctata
1791195	1791348	154	tatagcggattaacaaaaaccggtacggcggttgccccgccccggctcaaagggaaacggttccctaaaggcgccaagcaccgggccaacgggtccgtaaccatttgtactgcctgcggccccgcgccttgcctgatttttgtaataccgctata
1792751	1792855	105	tatagtgattaaatttaaaccagttacggcggttgctctgcctttcgtactatccgtactgcctgcggccccgcgccttgcctgattttcgttaataccgctata
1796016	1796169	154	tatggcggattaacaaaaaccggtacggcggtgccccgccccggctcaaagggaaacggttccccaaaggatgatgaagcaccgggccaacgggtccgtaaccatttgtactgcctgcggccccgcgccttgcctgatttttgtaataccgctata
1806866	1807019	154	tatagcggattaacaaaaaccggtacggcggttgccccgccccggctcaaagggaaacggttccctaaaggcgccaagcaccgggccaacgggtccgtaaccatttgtactgcctgcggccccgcgccttgcctgatttttgtaataccgctata
1826329	1826482	154	tatagtggtatacaaaaaaccggtacggcggttgccccgccttgccttgccttgcctaaagggaaacgattccctaaaggatgatgaagcaccgggccaacgggtccgtaaccatttgtactgcctgcggccccgcgccttgcctgatttttgtaataccgctata
1855254	1855408	155	tatagcggattaacaaaaaccggtacggcggtgccccgccccggctcaaagggaaacgattccctaaaggatgatgaagcaccgggccaacgggtccgtaaccatttgtactgtctgcggctctcgcgccttgcctgatttttgtaataccactata
1882841	1882913	73	tatagtggtataaatttaaaccggtacggcggttgctctgccttgcgcgtactgtttaaatttaattccactata
1905597	1905751	155	tatggtggtataactttaaaccggtacggcggtgccccgccccggctcaaagggaaacggttccctaaaggcgccaagcaccgggccaacgggtccgtaaccatttgtactgcctgcggccccgcgccttgcctgatttttgtaataccgctata
1927438	1927591	154	tatagcggattaacaaaaaccggtacggcggttgccccgccccggctcaaagggaaacgattccctaaaggatgatgaagcaccgggccaacgggtccgtaaccatttgtactgcctgcggccccgcgccttgcctgatttttgtaataccgctata
1941597	1941750	154	tatagcggattaacaaaaaccggtacggcggttgccccgccccggctcaaagggaaacggttccctaaaggcgccaagcaccgggccaacgggtccgtaaccatttgtactgcctgcggccccgcgccttgcctgatttttgtaataccgctatg
1959739	1959892	154	tatggcggattaacaaaaaccggtacggcggtgccccgccccggctcaaagggaaacggttccctaaaggatgatgaagcaccgggccaacgggtccgtaaccatttgtactgcctgcggccccgcgccttgcctgatttttgtaataccgctata
1974277	1974383	107	tatagtggtataaatttaaaccagttacgttgccttgcgcgtactatctgtactgtctgtggctctcgtcgccttgcctgatttaaatttaattccactata
1977219	1977328	110	tatagtggtatacaaaaaaccggtacggcggttgctctgccttgccttgcgcgtactatttgtactgtctgcggcttgcctgccttgcctgatttttgtaataccgctata
1988461	1988567	107	tatagtggtataaatttaaaccagttacggcggttgctctgccttgcgcgtactatttgtactgtctgcggctctcgtcgccttgcctgatttaaatttaattccactata

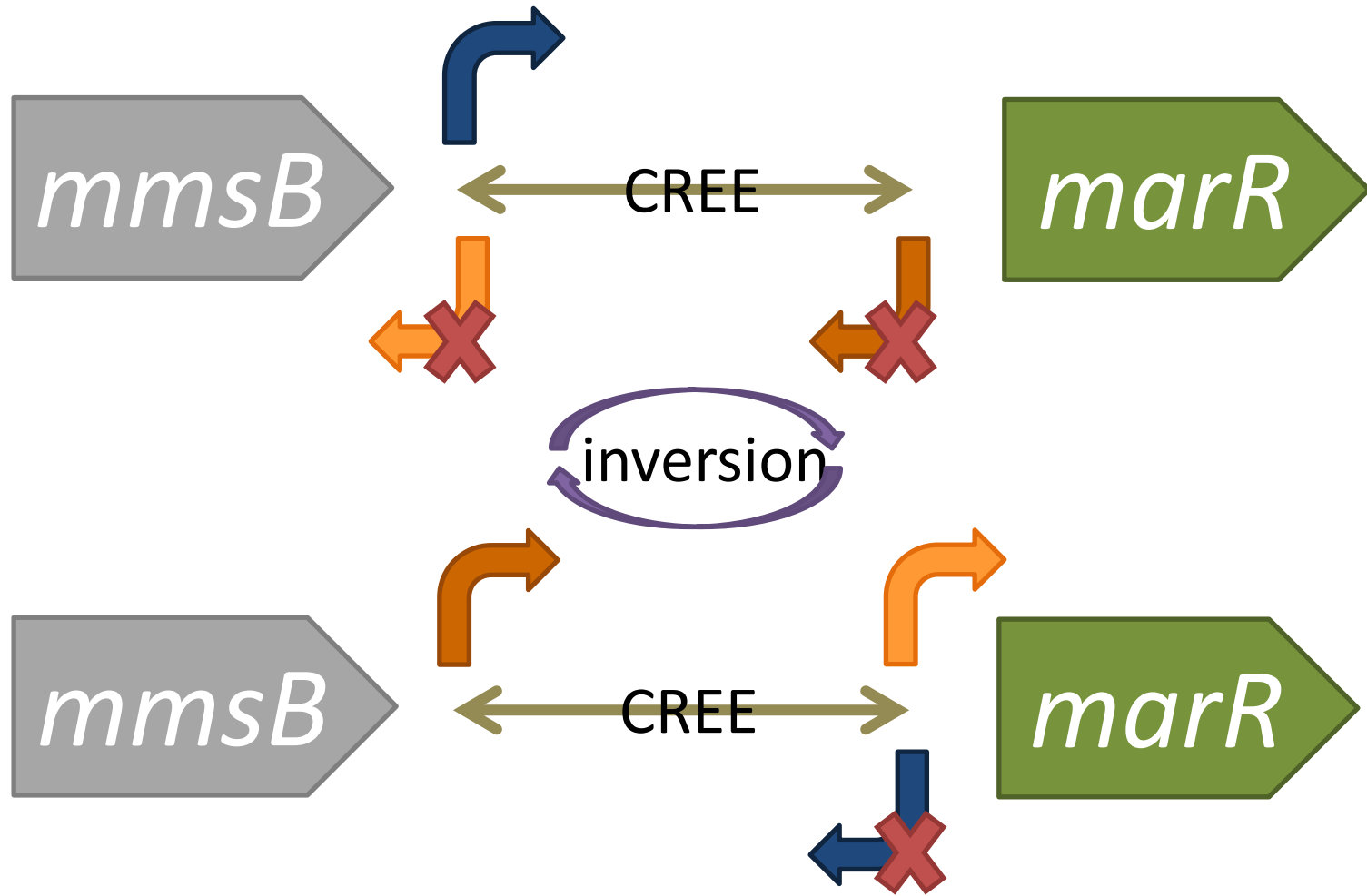
1989216	1989358	143	tatagtggtataacaaaaaccggtacggcggtgccccgccttagctcaaaagagaacgattccctaaggtgctgaagcaccgaagtgaatcggtccggtactatttgactgtctgcgccccgccttgctgattactata
1990334	1990488	155	tatagtggtataacaaaaaccggtacggcggtgccccgccttagctcaaaagagaacgattccctaaggtgctgaagcaccgaagtgaatcggtccggtactatttgactgtctgcggtccggtcgcttgctgatttttggtaataccactata
1994050	1994203	154	tatagcgaattaacaaaaaccggtacggcggtgccccgccttagctcaaaagagaacgattccctaaggtgctgaagcaccgaagtgaatcggtccggtactatttgactgtctgcggtccggtcgcttgctgatttttggtaatactactata
2031968	2032074	107	tatagtggtataaatttaaacccggtacggcggtgccccgccttagctcaaaagagaacgattccctaaggtgctgaagcaccgaagtgaatcggtccggtactatttgactgtctgcggtccggtcgcttgctgatttttggtaatactactata
2050694	2050847	154	tatagcggattaacaaaaaccggtacggcggtgccccgccttagctcaaaagagaacgattccctaaggtgctgaagcaccgaagtgaatcggtccggtactatttgactgtctgcggtccggtcgcttgctgatttttggtaataccactata
2052289	2052442	154	tatagcggattaacaaaaaccggtacggcggtgccccgccttagctcaaaagagaacgattccctaaggtgctgaagcaccgaagtgaatcggtccggtactatttgactgtctgcggtccggtcgcttgctgatttttggtaataccgcata
2059092	2059198	107	tatagtggtataaatttaaacccggtacggcggtgccccgccttagctcaaaagagaacgattccctaaggtgctgaagcaccgaagtgaatcggtccggtactatttgactgtctgcggtccggtcgcttgctgatttttggtaataccactata
2075962	2076115	154	tatagcggattaacaaaaaccggtacggcggtgccccgccttagctcaaaagagaacgattccctaaggtgctgaagcaccgaagtgaatcggtccggtactatttgactgtctgcggtccggtcgcttgctgatttttggtaataccgctata
2086435	2086539	105	tatagtggtataacaaaaaccggtacggcggtgccccgccttagctcaaaagagaacgattccctaaggtgctgaagcaccgaagtgaatcggtccggtactatttgactgtctgcggtccggtcgcttgctgatttttggtaataccactata
2091043	2091148	106	tatagtggtataaatttaaacccggtacggcggtgccccgccttagctcaaaagagaacgattccctaaggtgctgaagcaccgaagtgaatcggtccggtactatttgactgtctgcggtccggtcgcttgctgatttttggtaataccactata
2096299	2096453	155	tatagtggtataacaaaaaccggtacggcggtgccccgccttagctcaaaagagaacgattccctaaggtgctgaagcaccgaagtgaatcggtccggtactatttgactgtctgcggtccggtcgcttgctgatttttggtaataccactata
2106628	2106696	69	tatagtggtataaatttaaacccggtacggcggtgccccgccttagctcaaaagagaacgattccctaaggtgctgaagcaccgaagtgaatcggtccggtactatttgactgtctgcggtccggtcgcttgctgatttttggtaataccactata
2112852	2113006	155	tatagtggtataaatttaaacccggtacggcggtgccccgccttagctcaaaagagaacgattccctaaggtgctgaagcaccgaagtgaatcggtccggtactatttgactgtctgcggtccggtcgcttgctgatttttggtaataccactata
2123050	2123202	153	tatagcggattaacaaaaaccggtacggcggtgccccgccttagctcaaaagagaacgattccctaaggtgctgaagcaccgaagtgaatcggtccggtactatttgactgtctgcggtccggtcgcttgctgatttttggtaataccgcata
2134283	2134436	154	tatagtggtataacaaaaaccggtacggcggtgccccgccttagctcaaaagagaacgattccctaaggtgctgaagcaccgaagtgaatcggtccggtactatttgactgtctgcggtccggtcgcttgctgatttttggtaataccgctata
2146958	2147113	156	tatagtggtataacaaaaaccggtacggcggtgccccgccttagctcaaaagagaacgattccctaaggtgctgaagcaccgaagtgaatcggtccggtactatttgactgtctgcggtccggtcgcttgctgatttaatttaataccactata
2149525	2149680	156	tatagtggtactaaatttaaacccggtacggcggtgccccgccttagctcaaaagagaacgattccctaaggtgctgaagcaccgaagtgaatcggtccggtactatttgactgtctgcggtccggtcgcttgctgatttttggtaataccactata
2151965	2152118	154	tatagcggattaacaaaaaccggtacggcggtgccccgccttagctcaaaagagaacgattccctaaggtgctgaagcaccgaagtgaatcggtccggtactatttgactgtctgcggtccggtcgcttgctgatttttggtaataccgctata
2161831	2161937	107	TATAGtggtataaatttaaacccggtacggcggtgccccgccttagctcaaaagagaacgattccctaaggtgctgaagcaccgaagtgaatcggtccggtactatttgactgtctgcggtccggtcgcttgctgatttaatttaataccactata
2172837	2172941	105	tatagcggattaacaaaaaccggtacggcggtgccccgccttagctcaaaagagaacgattccctaaggtgctgaagcaccgaagtgaatcggtccggtactatttgactgtctgcggtccggtcgcttgctgatttttggtaataccactata
2185647	2185800	154	tataatggattaacaaaaaccggtacggcggtgccccgccttagctcaaaagagaacgattccctaaggtgctgaagcaccgaagtgaatcggtccggtactatttgactgtctgcggtccggtcgcttgctgatttttggtaataccactata
2186353	2186458	106	tatagtggtataaatttaaacccggtacggcggtgccccgccttagctcaaaagagaacgattccctaaggtgctgaagcaccgaagtgaatcggtccggtactatttgactgtctgcggtccggtcgcttgctgatttttggtaataccactata
2186864	2186967	104	tatagtggtataacaaaaaccggtacggcggtgccccgccttagctcaaaagagaacgattccctaaggtgctgaagcaccgaagtgaatcggtccggtactatttgactgtctgcggtccggtcgcttgctgatttttggtaataccactata
2189554	2189660	107	tatagtggtataaatttaaacccggtacggcggtgccccgccttagctcaaaagagaacgattccctaaggtgctgaagcaccgaagtgaatcggtccggtactatttgactgtctgcggtccggtcgcttgctgatttaatttaataccactata
2224365	2224519	155	tatagtggtataacaaaaaccggtacggcggtgccccgccttagctcaaaagagaacgattccctaaggtgctgaagcaccgaagtgaatcggtccggtactatttgactgtctgcggtccggtcgcttgctgatttttggtaataccactata

\* Regions where the CRE or Correia Repeats overlap an annotated coding region are shown in upper case letters.

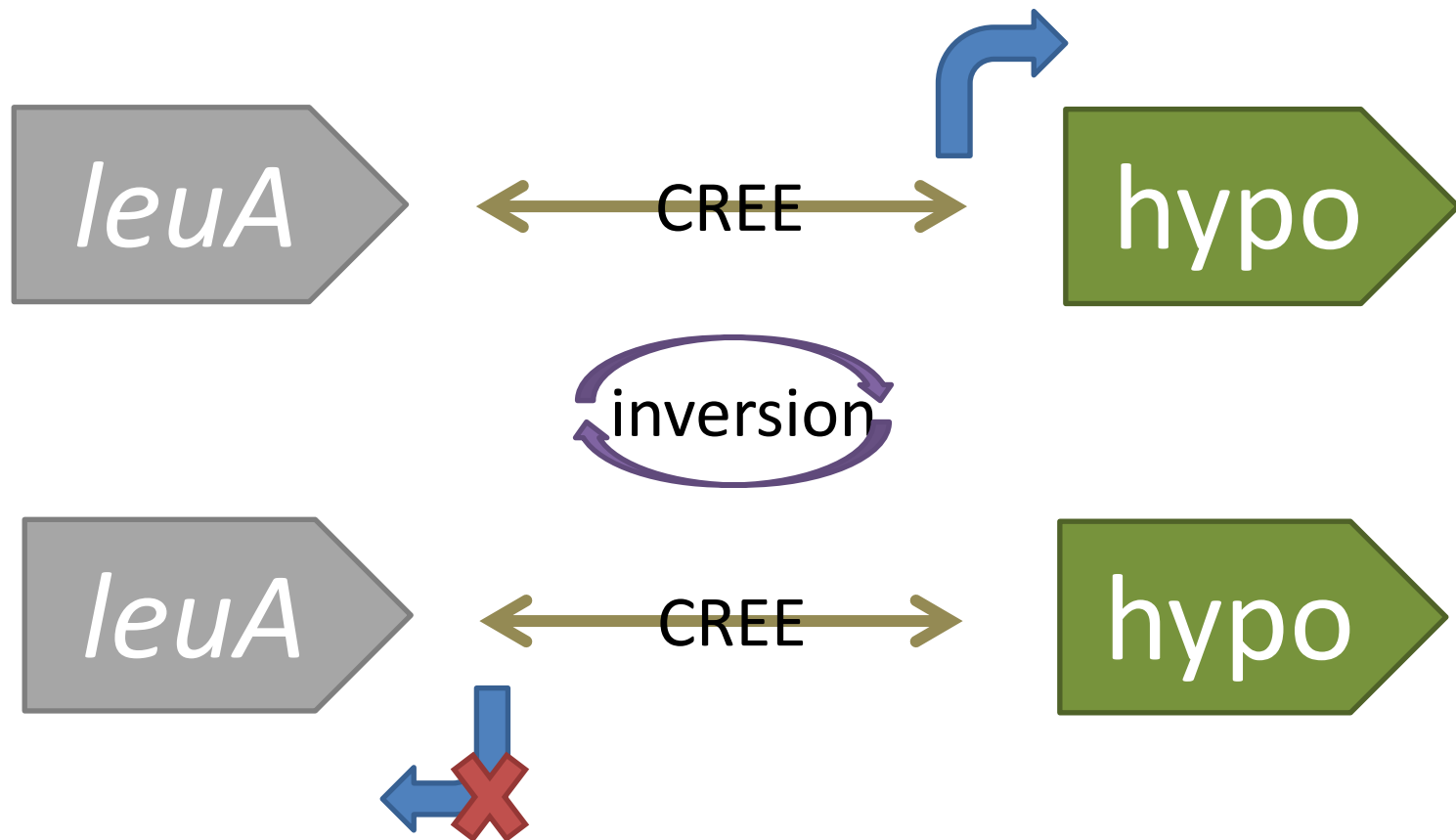




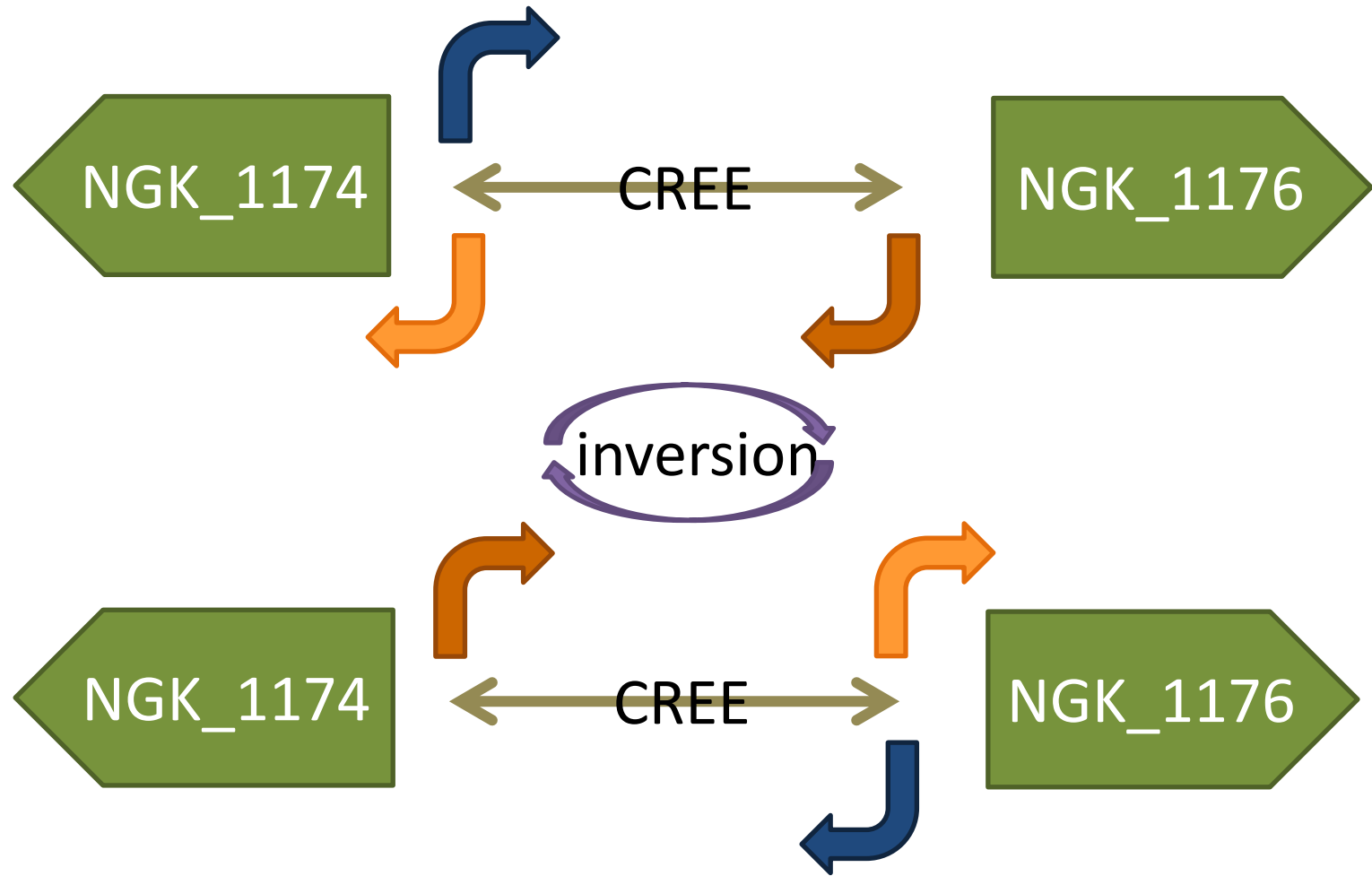
Supplementary Figure 1. CREE promoter 5' of *drg* (NGK\_0010). A functional Black promoter was described previously (Cantalupo *et al.*, FEBS Lett. 2001), shown by the blue arrow. No other CREE associated promoters are evident from the sequence, therefore on inversion, as observed here (Table 1), the original Black promoter would no longer transcribe *drg* (blue arrow) and no new Black promoter would be formed by the inversion.



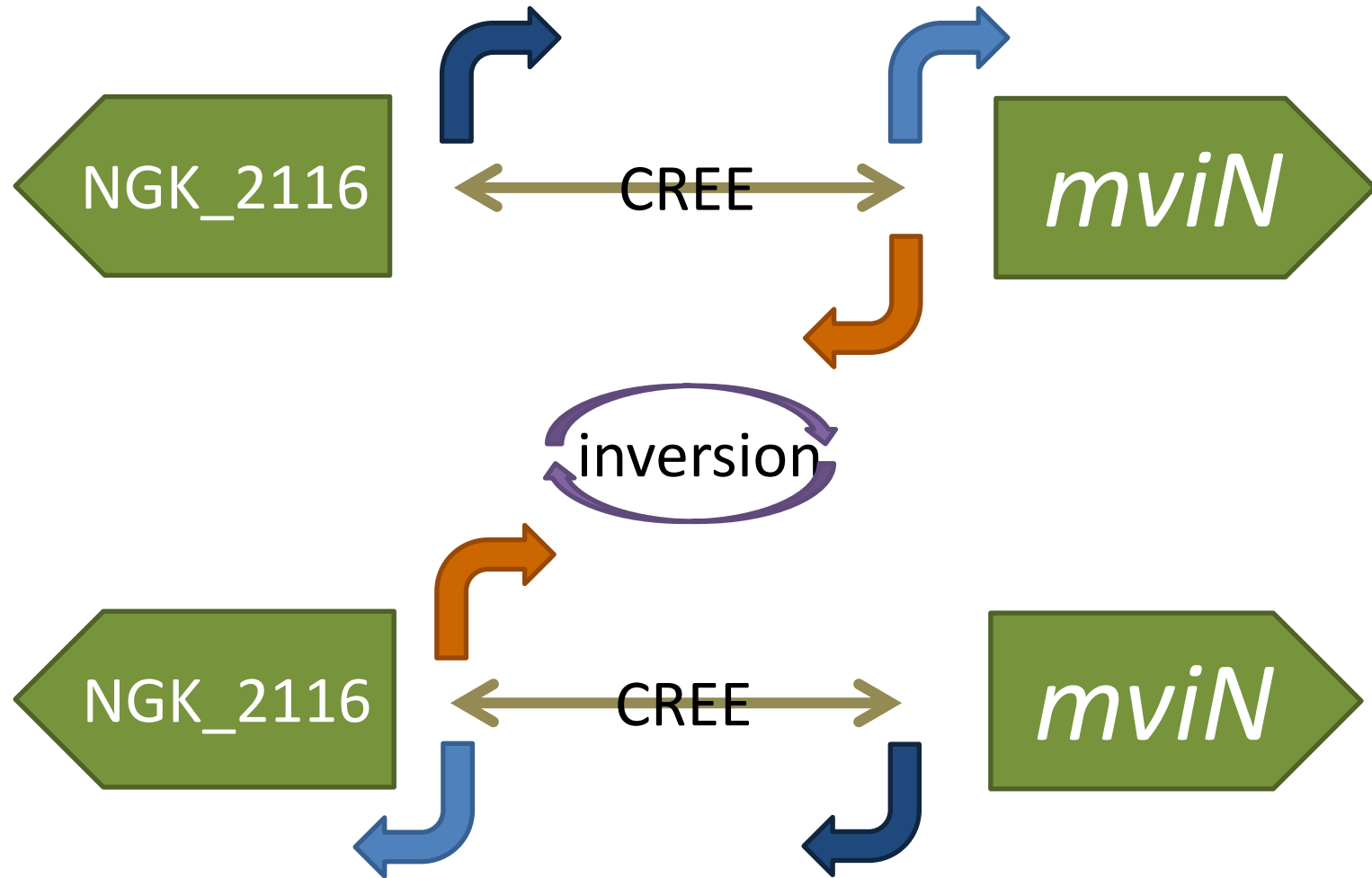
Supplementary Figure 2. CREE promoters 5' of *marR* (NGK\_0263). A potential Snyder promoter is found in the sequence (blue arrow). Upon inversion, as observed here (Table 1), potential Snyder and Black promoters would be generated from the CREE and native chromosomal sequence (orange arrows). Red crosses indicate a lack of either complete -10 or -35 sequences in that orientation.



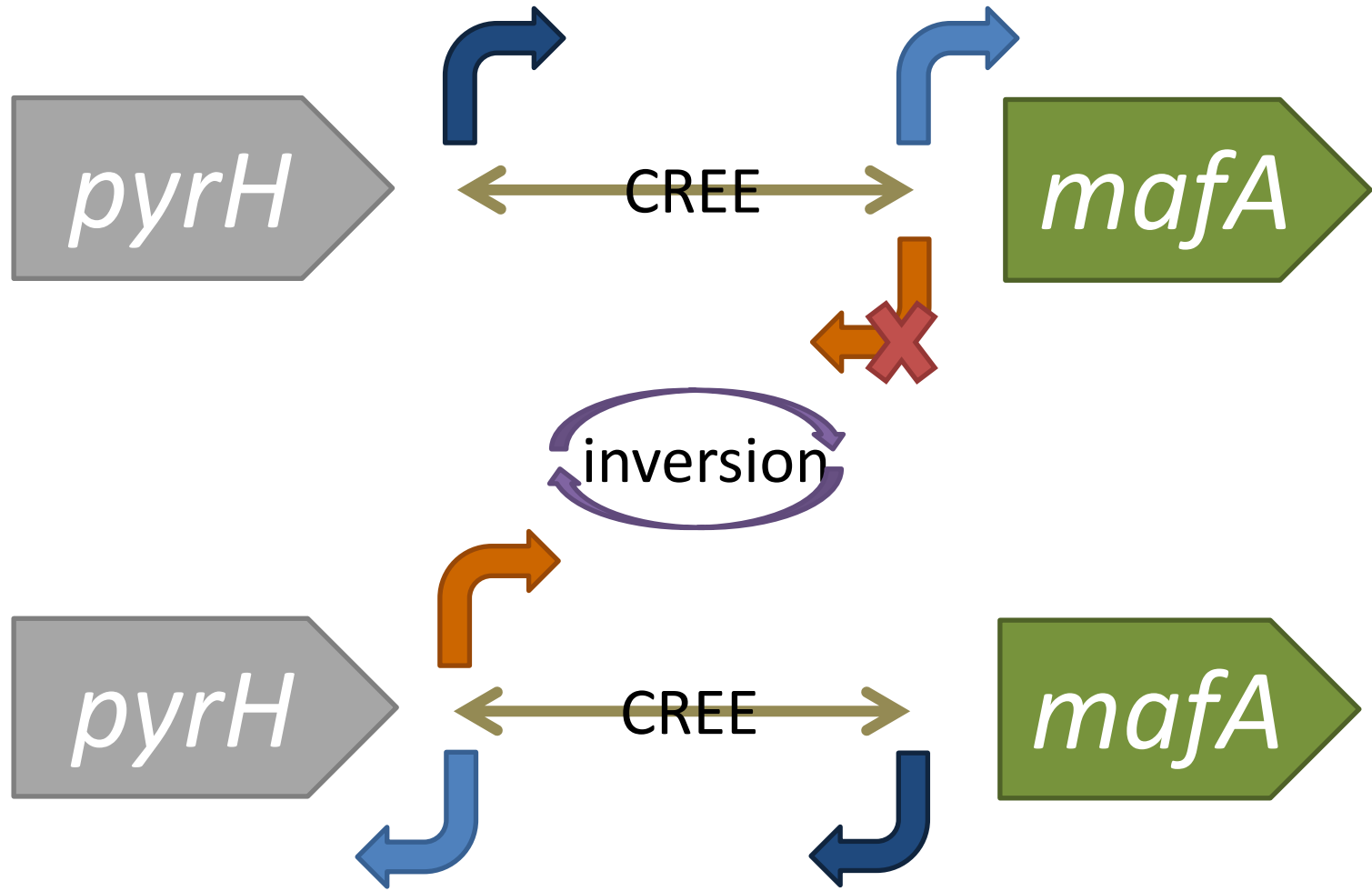
Supplementary Figure 3. CREE promoter 5' of hypothetical gene NGK\_0961. A single potential Black promoter was identified from the sequence (blue arrow). Therefore on inversion, as observed here (Table 1), the original Black promoter would no longer transcribe NGK\_0961 (blue arrow) and would no longer form a -10 with the native sequence (red cross). No new Black promoter would be formed by the inversion and no Snyder promoters have been found in the sequence.



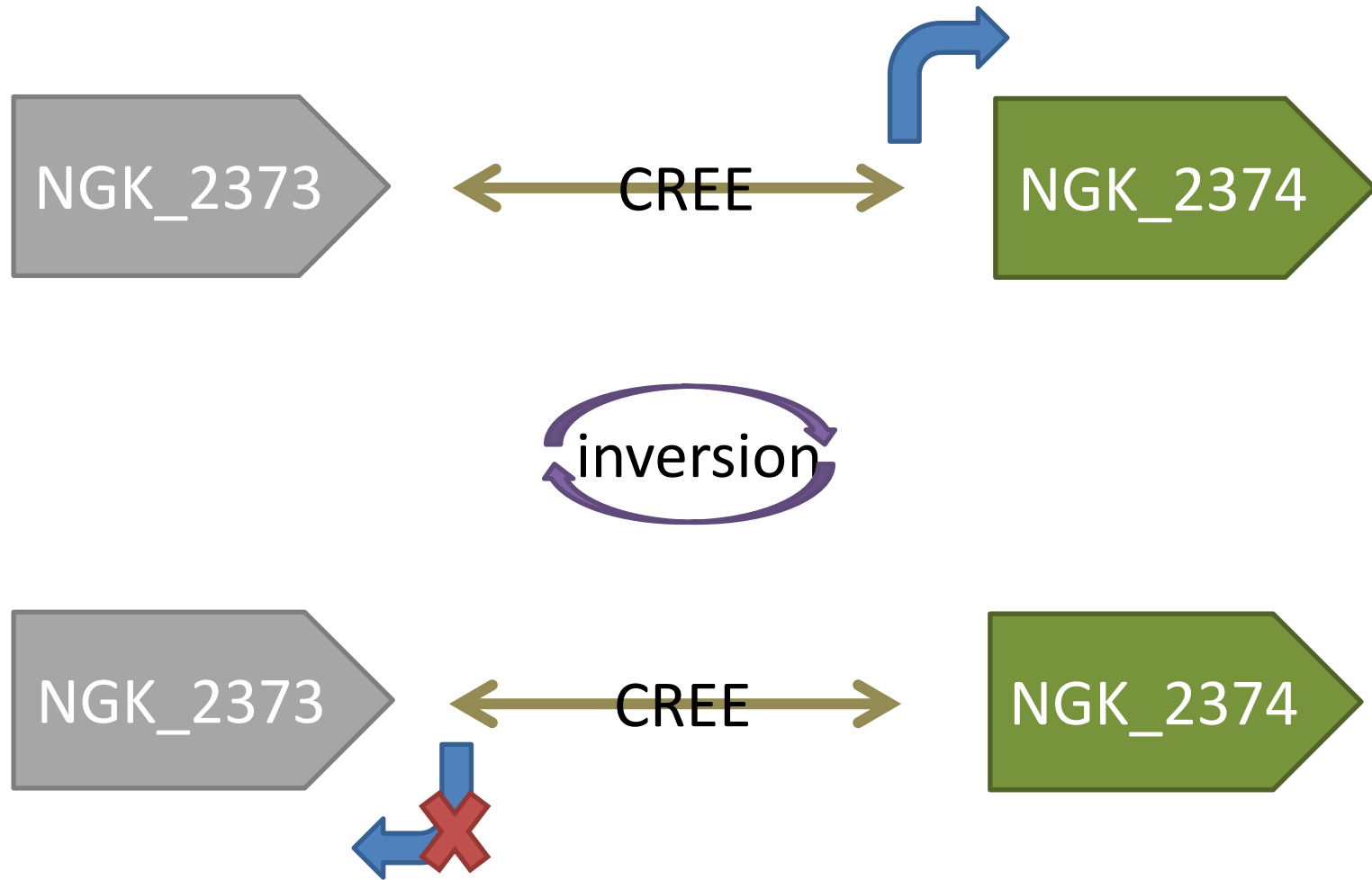
Supplementary Figure 4. CREE promoters between NGK\_1174 (glucose-6-phosphate 1-dehydrogenase) and NGK\_1176 (phosphogluconate dehydratase). Potential Snyder promoters are found in the sequence (blue and dark orange arrows), as well as one potential Black promoter (light orange arrow). Upon inversion, as observed here (Table 1), the sequences retain their potential -10 and -35 regions. The promoters can therefore be functional in either orientation, with both a Snyder and Black promoter potentially driving transcription (orange arrows) or just a Snyder promoter (blue arrow). This may result in differences in RNaseIII processing of the transcript, as suggested previously (DeGregorio *et al.*, Biochem J, 2003).



Supplementary Figure 5. CREE promoters between NGK\_2116 (hypothetical gene) and *mviN*, an essential gene in *E. coli* (Kato & Hashimoto, Mol Sys Biol, 2007) involved in murien synthesis (Inoue *et al.*, J Bacteriol, 2008). Potential Snyder promoters (blue and dark orange arrows) and a potential Black promoter (light blue arrow) are identified. Inversion (Table 1), retains the potential -10 and -35 regions. The promoters can therefore be functional in either orientation, with both a Snyder and Black promoter potentially driving transcription (blue arrows) or just a Snyder promoter (orange arrow). This may result in differences in RNaseIII processing of the transcript, as suggested previously (DeGregorio *et al.*, Biochem J, 2003).



Supplementary Figure 6. CREE promoters 5' of *mafA* (NGK\_2270), believed to be an adhesin (Paruchuri *et al.*, PNAS, 1990). Potential Snyder and Black promoters (blue arrows), may drive transcription of *mafA*; after inversion there is only a Snyder promoter (orange arrow, lacking -35 in the opposite orientation), with implications for RNaseIII processing.



Supplementary Figure 7. CREE promoters 5' of NGK\_2374 (hypothetical gene). A potential Black promoter (blue arrow), is present in only one orientation of the CREE. No other CREE associated promoters are evident from the sequence. Upon inversion (Table 1), the associated -10 sequence is not retained and no new potential promoters are generated. Transcription of the hypothetical gene would therefore be dependent on the orientation of this CREE.