

The Transposon-Like Correia Elements Encode Numerous Strong Promoters and Provide a Potential New Mechanism for Phase Variation in the Meningococcus

Azeem Siddique¹, Nicolas Buisine², Ronald Chalmers^{1*}

1 School of Biomedical Sciences, University of Nottingham, Queen's Medical Centre, Nottingham, United Kingdom, **2** Evolution des Régulation Endocriniennes, Museum National d'Histoire Naturelle, Paris, France

Abstract

Neisseria meningitidis is the primary causative agent of bacterial meningitis. The genome is rich in repetitive DNA and almost 2% is occupied by a diminutive transposon called the Correia element. Here we report a bioinformatic analysis defining eight subtypes of the element with four distinct types of ends. Transcriptional analysis, using PCR and a *lacZ* reporter system, revealed that two ends in particular encode strong promoters. The activity of the strongest promoter is dictated by a recurrent polymorphism (Y128) at the right end of the element. We highlight examples of elements that appear to drive transcription of adjacent genes and others that may express small non-coding RNAs. Pair-wise comparisons between three meningococcal genomes revealed that no more than two-thirds of Correia elements maintain their subtype at any particular locus. This is due to recombinational class switching between elements in a single strain. Upon switching subtype, a new allele is available to spread through the population by natural transformation. This process may represent a hitherto unrecognized mechanism for phase variation in the meningococcus. We conclude that the strain-to-strain variability of the Correia elements, and the large number of strong promoters encoded by them, allows for potentially widespread effects within the population as a whole. By defining the strength of the promoters encoded by the eight subtypes of Correia ends, we provide a resource that allows the transcriptional effects of a particular subtype at a given locus to be predicted.

Citation: 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(1): e1001277. doi:10.1371/journal.pgen.1001277

Editor: Josep Casadesús, Universidad de Sevilla, Spain

Received: August 2, 2010; **Accepted:** December 14, 2010; **Published:** January 20, 2011

Copyright: © 2011 Siddique et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was funded by grants from The Wellcome Trust to RC. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: chalmers@nottingham.ac.uk

Introduction

Neisseria meningitidis is an encapsulated Gram-negative diplococcus commensal of the human nasopharyngeal tract. Although carried asymptotically by 10–15% of the population, it occasionally crosses the epithelial cell barrier, causing bacterial meningitis and septicemia. Vaccination against serogroup A and C strains limits the impact of the disease in developed countries. However, the disease remains a significant problem in the meningitis belt of sub-Saharan Africa, where epidemics begin at the start of the dry season and may affect up to 1% of the population [1,2]. Even in the UK there are about 3000 cases each year. The disease has a rapid onset and is almost always fatal if untreated.

The *N. meningitidis* genome contains a relatively large amount of repetitive DNA. The repeats range in size from single nucleotide homopolymeric tracts, which mediate antigenic phase variation, to larger repeats of unknown or uncertain function [3–6]. One of the most abundant repeats is a miniature inverted-repeat transposable-element (MITE) first identified by Correia and colleagues in 1986 [7,8]. We refer to the repeat as the Correia element (CE), but it has also been known as NEMIS (*Neisseria* miniature insertion sequence) and CREE (Correia repeat enclosed element).

The archetypal genome sequences for the serogroup A, B and C strains of *N. meningitidis* (Z2491, MC58 and FAM18, respectively)

each contain about 250 intact CEs [3,5,6]. Insertion of the element is accompanied by duplication of a TA dinucleotide at the target site [9]. This is the hallmark of the *mariner* transposons, represented in the eubacteria by the IS630 family [10–13].

Short dispersed repeats, such as the CE, are dispersed and amplified by transposition and DNA recombination. However, their persistence in large populations of free-living bacteria, where natural selection is strong, has prompted frequent speculation that they directly benefit their hosts eg. [14]. Early chemostat experiments, in which strong nutritional selection was applied, revealed that most successful mutations were not in structural genes, but in their regulatory regions [15]. Many of these changes were due to transposons. This phenomenon is not restricted to bacteria: for example, transposon insertions upstream of the *Cyp6g1* gene in *Drosophila melanogaster* have spread to high frequency in response to the use of insecticides [16].

There are a number of ways in which transposons can change the pattern of gene expression and alter host cell physiology [15,17]. In the simplest case, an insertion may inactivate the gene encoding a transcriptional inducer or repressor. Insertions may also increase the distance between regulatory elements, interfering with activation or relieving repression. Transposons also have more direct mechanisms to control transcription. Many have constitutive promoters that drive transcription outwards from one end of the element [18–21]. Indeed, it is transposon-encoded

Author Summary

Transposons are mobile DNA elements that can jump from one location in the genome to another. They have had a profound influence on the evolutionary history of most, if not all, organisms by rearranging the order of genes and changing their expression patterns. The *mariner* family of transposons is probably the most successful group if judged by the breadth and depth of its phylogenetic distribution. One example is the Correia element, which has been amplified to a few hundred copies in *Neisseria meningitidis*. Transposons often encode promoters that drive the expression of adjacent genes. This raises the question of whether the large numbers of Correia elements in *N. meningitidis* have a significant genome-wide role in the control of gene expression. This is an interesting issue because *N. meningitidis* has evolved recently, having been first recognized in the 19th century, and is probably undergoing a period of rapid adaptation. Here we present a systematic analysis that defines eight sub-classes of Correia elements. We show that two subtypes encode strong promoters. The differential distribution of the strongest Correia promoter in the three strains provides a snapshot of evolution in action and sheds new light on the role of dispersed repeats in bacterial genomes.

promoter activity that is responsible for the successful chemostat take-over events and the spread of *Cyp6g1* alleles in *Drosophila* mentioned above.

CEs appear to influence their hosts in multiple ways. At the DNA-level, CEs are hotspots for DNA recombination and rearrangement [9,22]. At the RNA-level, CEs that are co-transcribed with adjacent genes are often targets for cleavage by RNase III [23–25]. Such processing may either stabilize or destabilize transcripts, potentially altering gene expression levels. CEs have also been proposed to act as transcriptional terminators, a consequence of their stem-loop structures and frequent presence near the 3' end of genes [26]. The Correia terminal inverted repeat (TIR) also contains a sequence resembling a –35 box for the σ^{70} class of promoters located 17 nucleotides upstream of a TATA sequence that forms at the end of the element as a result of insertion into a target site (Figure 1) [9,27]. Consequently, CEs have the potential to form outward-facing promoters at their insertion sites. In fact, CEs have been shown to contribute to the transcription of the meningococcal *lst* and *hemO* genes and the gonococcal *uvrB* gene [27–29]. Although such studies have identified transcripts emanating from individual CEs, a detailed examination of CE transcription, taking into account the variation that exists amongst copies, has not hitherto been performed.

The large number of CEs in the genome, their potential to influence gene expression patterns, and the variation in their complement between different strains, raises the question as to whether they are significant determinants of meningococcal physiology. This is important because most cases of meningococcal disease are caused by a few persistent hyper-invasive lineages and the physiological differences between these strains remain unclear. Bioinformatic analysis alone is not sufficient to settle such questions. We have therefore taken an experimental approach by measuring the strength of the promoters encoded by the CEs. We identify eight different subtypes, some of which have much higher promoter activity than others. The activity of the strongest promoter is dictated by a recurrent single polymorphism in the –35 box of the TIR. We present a genome-wide analysis of the

elements with the strongest activity, focusing on their flanking sequences and distribution in the population.

Results

The eight subtypes of Correia elements

Prior to embarking on an experimental analysis of putative CE promoters, we extended our previous bioinformatic analysis, significantly revising our classification of Correia subtypes and refining their nomenclature (Figure 1A). We searched for CEs using FASTA as described in the Materials and Methods section. In total we identified a set of 343 ‘almost-perfect’ elements, most of which are less than 2% divergent from their respective consensus sequences (Figure 1C). This set represents about half of the total number of CEs in the three genomes, the others having been excluded because of indels or other rearrangements. With some manual intervention, necessitated by the structure of the TIRs, the CEs were sorted into the eight sequence subtypes. Consensus sequences for each of the subtypes are shown in Figure 1A.

As noted previously [8,9,22,25], the CEs have a unique central region and two different types of TIRs, which we refer to as alpha (α) and beta (β) (Figure 1A, 1B). The α and β ends differ by three point mutations and a single-nucleotide indel. The precise boundary of the TIR is somewhat arbitrary and depends upon how many mismatches are tolerated. We propose to allow an inverted repeat that is 25 and 26 bp long for the α and β repeats, respectively (Figure 1B). The TIRs can be further categorized according to whether they are at the left or right end of the CE. The left and right TIRs, whether α or β , differ from each other at two positions (Figure 1B). The distinction between the two ends is important because one of the variable positions is within the predicted –35 box.

Since CEs were almost certainly amplified by a transposition mechanism, we follow a numbering convention that excludes the target site duplication from the size of the transposon. Thus, the β - β element is the longest CE with a length of 153 bp. Nucleotide positions for all of the other element subtypes are based on their alignment with this element (Figure 1A). Predicted –10 and –35 transcriptional start signals are indicated at the bottom of the alignment (Figure 1A). Note, however, that the –35 box is shifted one nucleotide further from the end of the element than proposed previously [9], for reasons that are explained below.

During our previous analysis of the CEs we identified a binding site for the IHF protein near the middle of the full-length element [9]. However, many CEs lack the IHF binding site due to a 50 bp deletion spanning this region. In our nomenclature we designate these elements using the prime symbol (').

Sequence variations within a subtype

The alignments reveal that most point mutations are scattered randomly across the CEs (not shown). However, the alignments also reveal two recurrent mutations, which are not random. The base at position 52 can be either A or G ($A \approx G$; denoted by R), while the base at position 128 is either C or T ($C > T$; denoted by Y). Henceforth, we will refer to these positions as polymorphisms. The significance of the polymorphism at position 52 is unknown. However, the polymorphism at position 128 is within the putative –35 box, and will be shown to control the strength of the CE promoter (see below). Note that although the polymorphism at position 128 may be present within the α or β end, it is unique to the right TIR of the CE.

Using the set of 121 Correia α - α elements, we determined the number of single nucleotide variants per element relative to the consensus (Figure 1C). For this analysis the polymorphisms R52

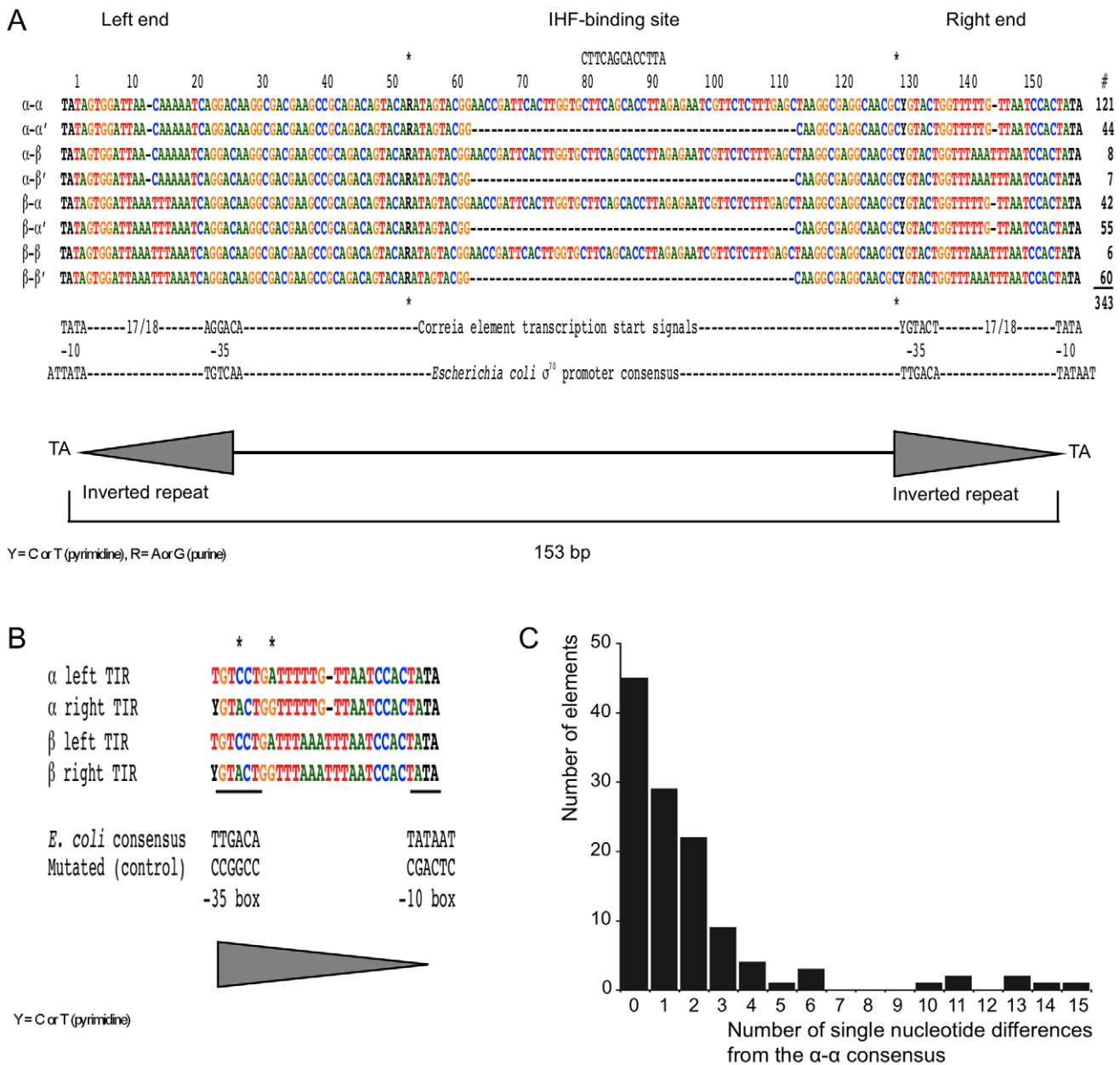


Figure 1. The eight classes of the consensus CE. (A) Nucleotide sequence alignment of the 8 CE consensus subtypes. Dashes within the alignment indicate gaps. Asterisks mark the positions of the R and Y nucleotides where the consensus sequence is polymorphic. “R” represents either A or G, with adenine more frequently present at this position. “Y” represents T, or more frequently C. Nucleotides are colored according to their identity except for the two polymorphisms and the flanking TA dinucleotide repeats, which are black. The CE -10 and -35 transcriptional start sequences and the equivalent sequences for the consensus *E. coli* σ^{70} promoter are indicated below the alignment. The total number of elements of each subclass within the *N. meningitidis* Z2491, MC58 and FAM18 genomes is indicated beside the alignment. These numbers represent approximately half of the total number of elements present in the 3 genomes. (B) Alignment of the α and β TIRs from the left and right ends of the CE. Asterisks mark the 2 nucleotides within the inverted repeat that differ between the left and right ends. The CE -10 and -35 transcriptional start sequences are underlined. For comparison purposes, the -10 and -35 transcriptional start sequences of the consensus *E. coli* σ^{70} promoter are provided below the alignment. Also given are the mutated -10 and -35 sequences constructed to replace the wild-type sequences in the Correia end-*lacZ* reporter plasmids in Table 1. (C) A graphical illustration of sequence variation, relative to the consensus sequence, within the set of 121 α - α CEs.

doi:10.1371/journal.pgen.1001277.g001

and Y128 were ignored. The number of elements in each class decreases rapidly between zero and 5 mutations. However, the decline is more gradual than the exponential decay expected if point mutations accumulate randomly. Inspection of the alignment reveals that identical point mutations occur repeatedly. For example, there are only 15 different point mutations amongst the

29 CEs with a single difference from the consensus. In our sample, point mutations that are observed more than once are always from elements of the same strain. This distribution is likely to be the result of gene conversion, in which a mutation is copied from one CE to another within a genome. There are also more elements than expected with 10 or more point mutations (Figure 1C). Many

of these mutations are tightly grouped, often at adjacent positions. These clusters of mutations were probably created during a single mutagenic episode, perhaps during natural transformation, double-strand break repair or imprecise gene conversion.

Correia repeats drive transcription

We began our study of CE transcription by assessing the promoter activity of isolated CE ends. Consensus sequences for six ends, including both Y128 variants, were cloned upstream of a promoterless *lacZ* gene in a low copy plasmid. The strength of transcription was measured using Miller's colorimetric assay for β -galactosidase activity (Table 1). The Correia α -right, β -right and the β -left sequences produced significant levels of β -galactosidase activity (75, 86 and 97 Miller Units [MU], respectively) compared to the empty vector (7 MU). In contrast, the α -left and the α -right^{Y128T} sequences were much more active, producing 540 MU and 670 MU of activity, respectively. The Y128T polymorphism had a particularly strong effect in the context of the α -right repeat where it increases activity almost 9-fold.

To confirm the position of the promoters we mutated the predicted -10 and -35 boxes of the α -right^{Y128T} and β -right^{Y128T} ends (Figure 1B and Table 1). Alteration of either sequence dramatically reduced the activity of α -right^{Y128T}. The mutations attenuated transcription from the β -right^{Y128T} repeat less severely. This suggests that the β -right repeat may provide an additional source of transcriptional activity. Inspection of the β repeat revealed a sequence, TGgTTTAAA, that is similar to an "extended -10 promoter." These promoters require no -35 box and have the consensus TGnTATAAT [30–32].

These results show that the CE TIRs possess promoter activity, but that the activity varies considerably depending on the class of CE in question. Mutational analysis demonstrates that the -10 and -35 transcriptional start signals predicted by visual inspection constitute the primary promoter of the Correia repeats. The Correia α -left and α -right^{Y128T} ends display the strongest transcriptional activity, a somewhat unexpected finding consider-

ing that the α - α element is the most common class of CE in *N. meningitidis*.

Transcription from intact CEs

To assess the role of the IHF binding site and the potential for the promoters to interfere with each other, we measured transcription from intact CEs (Table 2). The eight consensus CEs, as well as two elements incorporating both polymorphic variants (R52G and Y128T), were generated by PCR and inserted, in both orientations, upstream of a promoterless *lacZ* gene. For transcriptional analyses, chromosomal reporters are considered more reliable than multicopy plasmids. Therefore, we transferred the 21 reporter cassettes to bacteriophage lambda, which was subsequently used to make single copy phage insertions in the *Escherichia coli* chromosome.

Reporter assays performed with a strain lacking a CE insertion upstream of *lacZ* produce negligible β -galactosidase activity (0.6 MU) (Table 2). The spectrum of promoter activity for the CEs is broadly similar to that obtained from the isolated Correia repeats. Most of the α -right and β -right ends generate low levels of β -galactosidase activity (20–31 MU). The α -left end, on the other hand, generates moderately high activity whether in the context of the α - α or α - β element (116 MU or 168 MU respectively). Interestingly, the right end of the α - β element (full-length or prime) was twice as active as the right end of the β - β end (61–65 MU versus 20–31 MU). This may be an example of an interaction between promoters, whereby the activity of β -right is modulated by one adjacent promoter (α -left) but not by a different promoter (β -left).

The Y128T polymorphism has a relatively small stimulatory effect (~ 2 fold) on the promoter activity of the β -right end (Table 2). It has a much larger effect on the α -right end, raising β -galactosidase levels more than 23-fold. α -right^{Y128T} generates 469 MU of activity and is the strongest CE promoter tested. In contrast, the R52G mutation has little effect on transcriptional activity from either the α or β left end.

Table 1. β -galactosidase reporter assays for consensus Correia ends.

| Plasmid | Description ¹ | Miller units ² | Standard error ³ | Fold difference ⁴ |
|---------|---|---------------------------|-----------------------------|------------------------------|
| - | <i>lac</i> ⁻ host strain | 0.54 | 0.11 | - |
| pRC746 | vector control | 7.4 | 0.28 | 1.0 |
| pRC759 | α -left | 540 | 5.0 | 73 |
| pRC748 | α -right ⁵ | 75 | 7.2 | 10 |
| pRC750 | α -right ^{Y128T} | 670 | 22 | 90 |
| pRC761 | α -right ^{Y128T} ; -10 box altered ⁶ | 19 | 0.98 | 2.6 |
| pRC779 | α -right ^{Y128T} ; -35 box altered ⁶ | 46 | 1.9 | 6.3 |
| pRC760 | β -left | 97 | 8.8 | 13 |
| pRC749 | β -right ⁵ | 86 | 6.1 | 12 |
| pRC751 | β -right ^{Y128T} | 140 | 7.4 | 19 |
| pRC762 | β -right ^{Y128T} ; -10 box altered ⁶ | 51 | 4.3 | 6.8 |
| pRC781 | β -right ^{Y128T} ; -35 box altered ⁶ | 89 | 7.8 | 12 |

¹The consensus CE ends shown in Figure 1B (along with 4 nucleotides immediately upstream of the Correia repeat and the downstream flanking dinucleotide "AT") were cloned into a low-copy *lacZ* reporter plasmid (pRC746). Details are provided in Table S1. The *E. coli* host strain is MC4100.

² β -galactosidase activity is measured in Miller units. The following formula was used to calculate Miller units: [(spectrophotometer reading at 420 nm wavelength) \times 1000] / [(spectrophotometer reading at 600 nm wavelength) \times (cell volume (mls)) \times (time (mins))].

³Based on results from 3 independent experiments.

⁴Relative to vector control (pRC746).

⁵Unless designated otherwise, the base at position 128 is cytosine.

⁶Altered -10 and -35 box sequences are shown in Figure 1B.

doi:10.1371/journal.pgen.1001277.t001

Table 2. β -galactosidase reporter assays for chromosomally integrated *Correia* elements in log phase cultures.

| Correia element ¹ | wild-type strain ² | | <i>hipA::cat</i> strain | | Fold difference ³ | |
|--|-------------------------------|-----------|-------------------------|-----------|------------------------------|-----------|
| | ← reverse | → forward | ← reverse | → forward | ← reverse | → forward |
| α L - α R | 120±5.5 ⁴ | 20±0.8 | 140±4.2 | 20±0.6 | 0.8 | 1.0 |
| α L - α R' | 200±7.1 | 21±1.3 | 200±6.4 | 19±0.3 | 1.0 | 1.1 |
| α L ^{R52G} - α R ^{Y128T} | 97±1.0 | 470±24 | 110±5.4 | 590±7.3 | 0.9 | 0.8 |
| β L - β R | 39±5.3 | 24±0.2 | 38±0.5 | 29±0.3 | 1.1 | 0.8 |
| β L - β R' | 44±0.8 | 31±0.6 | 45±0.7 | 31±0.8 | 1.0 | 1.0 |
| β L ^{R52G} - β R ^{Y128T} | 26±0.3 | 55±4.4 | 31±0.3 | 80±1.1 | 0.8 | 0.7 |
| α L - β R | 170±7.9 | 67±1.5 | 120±5.4 | 73±2.4 | 1.4 | 0.9 |
| α L - β R' | 130±0.9 | 62±1.0 | 140±12 | 52±1.6 | 0.9 | 1.2 |
| β L - α R | 32±1.2 | 24±0.7 | 36±0.4 | 24±0.3 | 0.9 | 1.0 |
| β L - α R' | 75±0.8 | 31±1.9 | 72±0.8 | 27±0.8 | 1.0 | 1.1 |
| vector | 0.6±0.05 | | 0.6±0.01 | | 1.0 | |

¹Unless designated otherwise, the bases at positions 52 and 128 of the CE are adenine and cytosine, respectively.

²*E. coli* strain NR289.

³IHF⁺/IHF⁻.

⁴The units of measurement, Miller units, are calculated using the formula provided in Table 1. For each construct, the mean Miller unit measurement and the standard error from at least three independent experiments are provided. doi:10.1371/journal.pgen.1001277.t002

A comparison of the full-length and prime elements indicates that there is little effect of the internal rearrangement on promoter activity (Table 2). We found this surprising because we expected to see a substantial effect from the deletion of the IHF binding site. To evaluate the effect of IHF more carefully, we used P1 transduction to disrupt IHF expression in all 21 of our chromosomal CE-*lacZ* reporter strains. This had minimal effect on the transcriptional activity of the elements (Table 2).

The concentration of IHF in *E. coli* is growth-phase dependent [33,34]. Transcription from a subset of the CEs was therefore measured in stationary phase cells where the concentration of IHF is elevated (Table 3). The transcriptional profile of the selected elements was again very similar in wild-type and strains lacking IHF. We can therefore conclude that IHF binding does not significantly affect transcription from the CEs.

Identification of the *Correia* element transcriptional start point

The transcriptional start points of the CE promoters were mapped by primer extension (Figure 2A). A radiolabeled oligonucleotide primer, designed to anneal downstream of the expected start point, was hybridized to the RNA and extended by AMV reverse transcriptase. The resulting cDNA products were analyzed on a denaturing polyacrylamide gel. The 12 selected promoters produce virtually identical patterns of extension products. The most prominent band corresponds to the transcriptional start point predicted by the -10 box at the end of the CE. The shorter products presumably represent degradation products or premature termination of reverse transcription, perhaps due to secondary structure, which is strong in this region. Further analysis of the prominent band on a high resolution DNA sequencing gel revealed that it is a doublet (not shown). The two bands of the

Table 3. β -galactosidase reporter assays for chromosomally integrated *Correia* elements in stationary phase cultures.

| Correia element ¹ | wild-type strain ² | | <i>hipA::cat</i> strain | | Fold difference ³ | |
|--|-------------------------------|-----------|-------------------------|-----------|------------------------------|-----------|
| | ← reverse | → forward | ← reverse | → forward | ← reverse | → forward |
| α L - α R | 170±3.2 ⁴ | 24±1.6 | 170±5.5 | 26±1.7 | 1.0 | 0.9 |
| α L - α R' | 280±2.0 | 27±1.3 | 280±4.1 | 25±1.9 | 1.0 | 1.1 |
| α L ^{R51G} - α R ^{Y128T} | 130±2.1 | 660±26 | 140±4.2 | 670±41 | 0.9 | 1.0 |
| β L - β R | 41±2.3 | 27±1.7 | 47±3.0 | 32±2.6 | 0.9 | 0.8 |
| β L - β R' | 57±2.5 | 32±2.7 | 57±1.8 | 34±3.1 | 1.0 | 0.9 |
| β L ^{R52G} - β R ^{Y128T} | 36±2.1 | 80±9.6 | 41±3.0 | 100±11 | 0.9 | 0.8 |
| vector | 0.6±0.02 | | 0.5±0.02 | | 1.2 | |

¹Unless designated otherwise, the bases at positions 52 and 128 of the CE are adenine and cytosine, respectively.

²*E. coli* strain NR289.

³IHF⁺/IHF⁻.

⁴The units of measurement, Miller units, are calculated using the formula provided in Table 1. For each construct, the mean Miller unit measurement and the standard error from at least three independent experiments are provided. doi:10.1371/journal.pgen.1001277.t003

doublet represent products starting 10 and 13 nucleotides downstream of the CE end (Figure 2B). The initiation point 10 bp downstream of the TATA box is the more prominent of the two bands, and is identical to that identified by Black *et al.* (1995) in their transcriptional analysis of a CE upstream of the gonococcal *uvrB* gene [27].

These results demonstrate that the previously identified transcriptional start sequences at the end of the CE constitute the primary promoter of the element. Moreover, the same promoter appears to be utilized at both the α and β TIRs.

Detection of *Correia* element-derived RNA transcripts in *N. meningitidis*

To provide direct evidence that CEs drive transcription in *N. meningitidis* we used RT-PCR to analyze three α -right^{Y128T} containing loci in Z2491 (Figure 3).

One *Correia* end is located 42 bp upstream of, and in tandem orientation with, the NMA0074 ORF, which encodes GidA, a protein involved in tRNA modification. There is no obvious transcriptional terminator between the CE and the gene, so, if functional, the *Correia* α -right^{Y128T} promoter is likely to contribute substantially to the transcription of the gene. The other two α -right^{Y128T} ends are also in intergenic regions but are directed towards strong predicted transcriptional terminators. If functional, they would generate short non-coding RNA transcripts (NMA0530 and NMA0059 loci).

Three primers were used for the analysis of each locus (Figure 3). A reverse primer (primer I) that annealed 100–200 nucleotides downstream of the CE promoter was used to generate cDNA during the reverse transcription step. In the subsequent PCR step, the reverse primer was combined with one of two forward primers: one corresponding to the predicted transcriptional start point (primer II), and another immediately upstream, spanning the junction between the CE and the flanking DNA (primer III). If the promoter at the CE terminus drives transcription at these loci, we would expect a PCR product with primers I and II, but not with primers I and III.

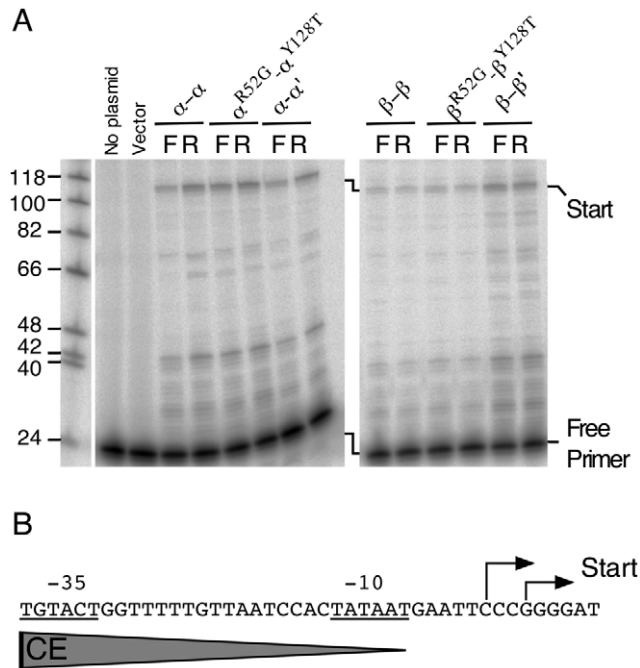


Figure 2. Transcript mapping by primer extension. (A) Primer extension analysis was performed with 20 μ g of RNA and 5 pmol of radiolabeled primer, as described in the Materials and Methods. RNA was prepared from *E. coli* MC4100 strains harboring *lacZ* reporter plasmid pRS415 [52] and derivatives containing the CEs indicated at the top of the gel. Reaction products were electrophoresed on a denaturing 8% polyacrylamide gel. The primary primer extension products and free primer are indicated beside the gel. (B) Nucleotide sequence of the Correia α -right^{Y128T} repeat and downstream sequences highlighting the transcriptional start points observed in (A). doi:10.1371/journal.pgen.1001277.g002

RT-PCR products of the correct size were obtained from the three loci with primers I and II (Figure 3). Small amounts of product were also obtained with primers I and III. This product was most abundant for the NMA0059 locus and indicates that a small amount of transcription originates upstream of the predicted start point. Control reactions were also performed with genomic DNA as the template (Figure 3). These reactions provide size standards for the respective RT-PCR products and demonstrate that the various pairs of primers perform with equal efficiency.

Genomic distribution of the α -right^{Y128T} repeat

We wished to determine the distribution of the α -right^{Y128T} repeats in the three *N. meningitidis* genomes (Z2491, MC58 and FAM18) and identify loci where they might be involved in the transcriptional regulation of nearby genes. We focused our attention on the α -right^{Y128T} repeat because it provides the strongest transcription of the Correia ends tested in this study. However, many other CE ends drive significant levels of transcription and warrant further investigation.

We performed whole-genome comparisons of the α -right^{Y128T} repeats using two different approaches as detailed in the Materials and Methods section. Each approach yielded identical results. The distribution of the α -right^{Y128T} repeat in the three *N. meningitidis* strains is shown in Figure 4. There are a total of 114 repeats, with almost 40 in each genome. Leftward and rightward facing repeats are indicated by their respective positions above and below the lines that denote each genome. Also indicated are the dinucleotides immediately flanking the TATA sequence at the end of each

repeat. These dinucleotides constitute part of the -10 box of the promoter and may have an effect on transcription depending on their divergence from the consensus (AT).

In pair-wise comparisons, less than two-thirds of the α -right^{Y128T} repeats were found to have a counterpart in the other genome (Figure 4). The synteny between pairs confirms that they are true homologs (Figure S1). For those α -right^{Y128T} repeats that are missing a counterpart, an examination of the homologous loci indicates that the counterpart is missing for one of three reasons: In a minority of cases the locus in question is absent, presumably because it has suffered a deletion. In others, the locus is present but unoccupied by a CE. In the majority of cases, however, the α -right^{Y128T} repeat has been replaced by a different type of Correia end. The most common substitution occurs when the α -right^{Y128T} is replaced by an α -right^{Y128C} repeat. However, there are cases where the α end is replaced by a β end. For example, the α -right^{Y128T} repeat at bp 1237645 in Z2491 (line 26 in Table 4) is replaced by α -right^{Y128C} in FAM18 and by β -right^{Y128T} in MC58. Since these are clearly identical CE insertions at the same dinucleotide target site, gene conversion events must account for the differences.

Genomic context of the α -right^{Y128T} repeats in Z2491

To understand the genomic context within which the repeats are found, and to identify genes that might be transcribed by the Correia promoter, we inspected the sequences downstream of the α -right^{Y128T} repeats in strain Z2491 (Table 4). Of the 39 α -right^{Y128T} ends, 7 lie within or are directed towards sequence repeat arrays (either RS-dRS3 repeats, also known as NIMEs, or ATR repeats: Table 4). For the remaining 32 repeats, the nearest significant features are ORFs, which are located up to 304 bp from the Correia end, but are often much closer. At two of these loci, the Correia end overlaps with an ORF (ORFs NMA1111 and NMA1960). Approximately two-thirds (22 of 34) of the ORFs represent hypothetical genes, genes of unknown function or probable pseudogenes (Table 4). The remaining ORFs code for proteins with diverse biological roles, including roles in metabolic processes, transcription, translation, ribosome synthesis and transport.

Many of the same ORFs are present downstream of the α -right^{Y128T} repeats in MC58 and FAM18 (Tables S3 and S4). However, these strains also have copies of α -right^{Y128T} not found in Z2491. Included amongst the ORFs downstream of these repeats are ones coding for bicyclomycin resistance (NMB0445), a TonB receptor (NMB1497), FrpA (NMB0585) and FrpC (NMB1415, NMC0527) virulence factors, a serine peptidase (NMB1998, NMC1974), and the pilus assembly protein PilG (NMC1839).

In Z2491, 18 of the 32 α -right^{Y128T} repeats driving transcription towards nearby ORFs are located in tandem to the ORF in question and will therefore produce sense transcripts (Table 4). The TransTermHP server [35] was consulted to check for the presence of rho-independent transcriptional terminators between the 18 CEs and their adjacent ORFs, but none were found. This indicates that transcription from these α -right^{Y128T} repeats is likely to contribute to the transcription of the downstream ORFs.

The 14 remaining elements are convergent, driving transcription towards the 3' end of the nearest ORF. Each of them will produce an antisense transcript unless transcription is halted by a terminator located between the CE and adjacent ORF. The TransTermHP server indicated that nine of the 14 loci had strong terminators within 300 bp of the α -right^{Y128T} repeats (TransTermHP confidence level $>80\%$; see Footnote 4 in Table 4). These RNA transcripts fulfill two key criteria used for the identification of short non-coding regulatory RNAs (sRNA) in bacteria [36–40]. However, unlike the promoters of many bona

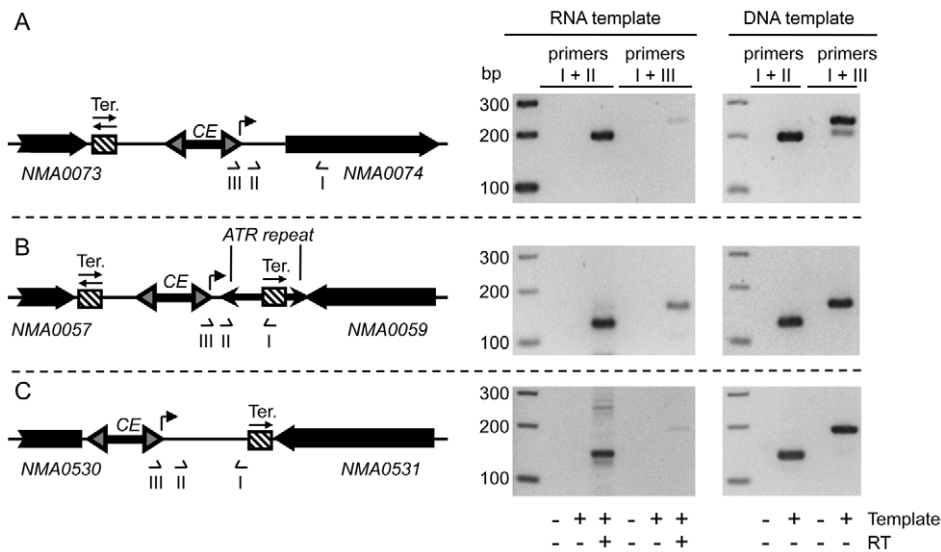


Figure 3. Transcriptional analysis of three loci containing CEs. (A) Schematic diagram of the region between the NMA0073 and NMA0074 ORFs showing the location of primers I, II and III (sequences given in Table S2) used in the accompanying PCR analysis. The solid black arrows denote genes. The hatched box represents a strong predicted transcriptional terminator, with the arrow(s) indicating its polarity: in this case, the terminator is predicted to function in both directions. The CE inverted repeats are indicated by grey arrowheads and the direction of transcription from the α -right^{Y128T} repeat is indicated by a bent arrow. Primer I was used for reverse transcription (RT). The subsequent PCR step was performed with the indicated pairs of primers and analyzed on an ethidium bromide-stained 3% Metaphor agarose TAE gel (right panels). Results from the RT-PCR are shown in the middle panel and those from the genomic DNA PCR are shown on the far right. The latter provides molecular weight standards for the RT-PCR products and a control for the efficiency of the various primer pairs. (B) The region between NMA0057 and NMA0059 is shown. Annotations are as in (A). The inverted repeats of an ATR element are depicted as black arrowheads. (C) The region between NMA0530 and NMA0531 is shown. The short NMA0530A ORF is not likely to code for a protein and has been omitted from the schematic. Annotations are as in (A). doi:10.1371/journal.pgen.1001277.g003

vide sRNAs, the CE promoters are probably not tightly regulated. Perhaps, in these examples we are witnessing the evolutionary birth of new sRNAs.

The remaining five α -right^{Y128T} repeats lack downstream terminators and would be expected to drive transcription into the 3'-end of adjacent ORFs. One of these ORFs is a pseudogene (NMA0823) and another overlaps with the CE itself (NMA1111). The remaining three ORFs encode a *metR* family transcriptional activator, a phase variable lipoprotein and a hypothetical protein (NMA0381, NMA0277 and NMA0209, respectively).

Whole-genome comparisons and Correia element annotations

During the course of this work we annotated the α -right^{Y128T} and α -left repeats, which provide the strongest promoters, and generated repeat density plots and the six pair-wise comparisons between the three meningococcal genomes. This information is provided in a format that can be viewed in the Artemis genome browser (Materials and Methods, Dataset S1 and Text S1 for simplified instructions). This will be a useful resource for future investigations. For example, a recent survey reported that meningococcal strains deleted for the CE upstream of *mtcCDE* did not have a reduced level of drug resistance [41]. This would have been anticipated by our result, which shows that the α -right^{Y128C} repeat (the relevant Correia end in this example) has low promoter activity (Table 2)

Discussion

The large number of CEs in the *N. meningitidis* genome means that it can be difficult to identify common biological themes from the analysis of individual elements. Therefore, we began our transcriptional analysis by classifying the CEs from the Z2491,

MC58 and FAM18 genomes into 8 distinct subgroups and generating a consensus sequence for each subtype. Further examination of these subtypes established that the Correia α -right^{Y128T} TIR contains by far the strongest promoter (Table 2).

Architecture of the CE promoters

The promoter activity of the α -right^{Y128T} repeat is 10 to 20-fold higher than that of α -right^{Y128C} (Table 1 and Table 2). The thymidine responsible for this dramatic difference is also present within the α -left end and it seems likely to contribute to the strong transcription from this end as well. In previous studies, the putative -35 box was positioned one nucleotide closer to the end of the CE [9,28]. However, the large effect of the Y128 polymorphism on transcription argues in favor of the new position illustrated in Figure 1. Interestingly, the Y128T mutation in the β TIR does not raise promoter activity as much as it does at the α TIR. This difference is probably due to the greater spacing between the -10 and -35 boxes of the β end relative to the α end (18 versus 17 nucleotides).

Black and colleagues predicted a stationary-phase, σ^S (*rpoS*)-dependent “gearbox” promoter in the prime version of the CE, but not in the full-length version, where a 50 bp insertion separates the -10 and -35 boxes [27]. Subsequent genome sequencing revealed that *rpoS* is absent in the meningococcus and the gonococcus [6,42]. However, a gearbox promoter could have affected our transcriptional analysis in *E. coli*, which does encode *rpoS*. This does not appear to be the case because a comparison of transcription from the right end of the full-length and prime elements shows no discernable effect of the putative gearbox promoter on the activity of any of the reporter constructs (Table 2).

An intriguing aspect of the structure of the CE is the presence of an IHF binding site in the full-length element. IHF is a histone-like protein which bends DNA by 180° upon binding [43]. In *E. coli*, it

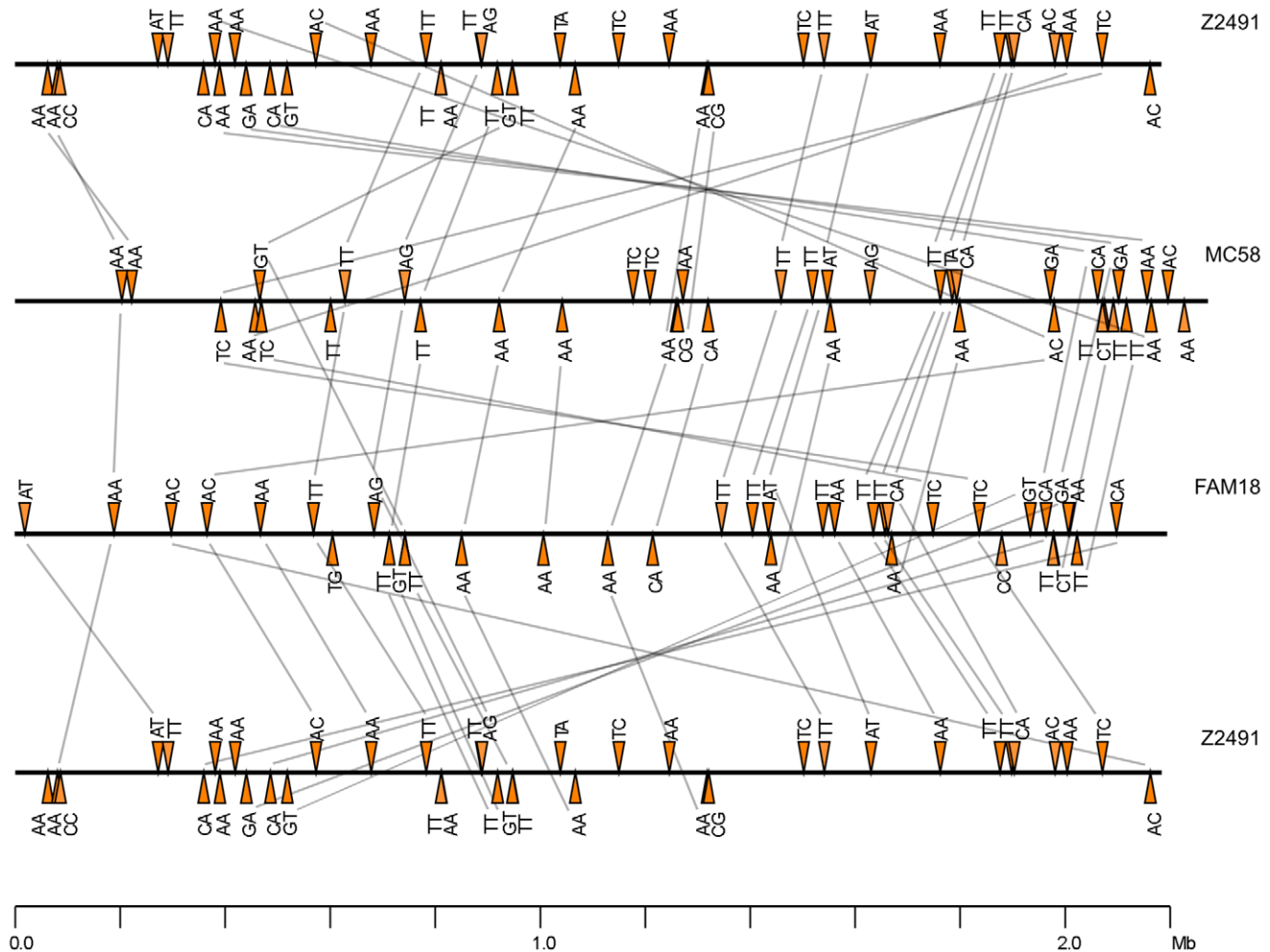


Figure 4. Distribution of the *Correia* α -right^{Y128T} repeats in the *N. meningitidis* Z2491, MC58, and FAM18 genomes. Horizontal black lines represent the *N. meningitidis* genomes. Colored triangles represent the α -right^{Y128T} promoters, of which there are almost 40 in each genome. The letters above or below each triangle refer to the dinucleotides flanking the TATA sequence at the end of each repeat. *Correia* ends represented above and below the line drive transcription to the right and left, respectively. The grey lines between genomes connect *Correia* α -right^{Y128T} repeats inserted at homologous positions in their respective genomes. The synteny of loci was established first by using a 20 bp sequence tag, and then by examining the ORF landscape surrounding each element (see Figure S1). The diagram is drawn to scale. doi:10.1371/journal.pgen.1001277.g004

has a role as an accessory protein in a variety of cellular processes including replication, recombination and transcription [44]. The *Correia* IHF-binding site has been shown to bind IHF protein from *E. coli* and *N. gonorrhoeae* in gel shift mobility assays [9,45]. Consequently, we hypothesized that IHF might modulate the activity of the *Correia* promoter. However, β -galactosidase assays performed with CE-*lacZ* reporter constructs indicated that IHF has no significant effect on CE transcription (Table 2 and Table 3). We therefore wonder whether the primary effect of IHF may be on genomic architecture and compaction of the nucleoid. In this capacity, it may alter the expression of genes at a distance by bringing distal regulatory elements together.

A potential mechanism for phase variation

A comparison of CEs from Z2491, MC58 and FAM18 reveals several conversion events in which one class of *Correia* repeat at a given locus is replaced by another. For example, the α L- α R^{Y128T} element in Z2491 on line 3 of Table 4 has been converted to α L- α R^{Y128C} in MC58. This should have the effect of reducing CE-

driven transcription of the adjacent threonine tRNA gene in MC58. Clearly, *Correia* end subtype switching has the potential to act as a mechanism for phase variation, in which the transcription of genes under the influence of a CE is modulated by the recombination-mediated switching of *Correia* promoters. Indeed, we have surveyed CE class switching in the meningococcal reference collection and find that the differences are highly correlated with the various clonal complexes (to be presented elsewhere). Class switching also has the potential to affect gene expression by altering the sensitivity of CE-containing transcripts to cleavage by RNase III. RNase III targets CE-derived stem-loop structures in transcripts, and is sensitive to point mutations that enhance or diminish the stem-loop [24].

In our analysis of the Z2491 genome we focused on the α -right^{Y128T} repeat, which provides the strongest promoter activity in our assay (Table 2). However, one should note that other classes of element, particularly the α -left repeat, also provide significant promoter activity and may be linked to important functions. In each of the three strains studied there are over 100 α -right^{Y128T}

Table 4. *Correia* α -right^{Y128T} repeat distribution in the genome of *N. meningitidis* serogroup A strain Z2491.

| | Nucleotide sequence of the <i>Correia</i> element end (29 bp)¹ | Adjacent ORF or other feature (downstream of <i>Correia</i> promoter)² | Distance of nearest ORF to <i>Correia</i> element (bp) | Direction of transcription³ | Description of putative gene product |
|----|--|--|---|---|--|
| 1 | TGTACTGGTTTTGTTAATCCACTATAaa | ATR repeat followed by NMA0059 ⁴ | 176 | convergent | modification methylase (probable pseudogene) |
| 2 | -----aa | NMA0074 (<i>gidA</i>) | 42 | tandem | tRNA uridine 5-carboxymethyl aminomethyl modification enzyme |
| 3 | -----T-----cc | <i>thr</i> tRNA gene & NMA0079 ⁵ | 69 | tandem | threonine tRNA & putative integral membrane protein ⁵ |
| 4 | -----at | NMA0277 | 155 | convergent | putative lipoprotein |
| 5 | -----G----tt | RS-dRS3 repeat array | | | |
| 6 | -----ca | NMA0381 (<i>metR</i>) | 42 | convergent | putative transcriptional activator |
| 7 | -----aa | single RS-dRS3 repeat followed by NMA0402 (<i>truA</i>) | 218 | tandem | tRNA pseudouridine synthase |
| 8 | -----aa | NMA0412 ⁴ | 79 | convergent | putative glucokinase |
| 9 | -----aa | NMA0434 | 30 | tandem | hypothetical inner membrane protein |
| 10 | -----ga | RS-dRS3 repeat array & another CE | | | |
| 11 | -----ca | NMA0495 (<i>rpmE</i>) | 107 | tandem | 50S ribosomal protein L31 |
| 12 | -----gt | NMA0530A & NMA0531 ^{4,5} | 259 | convergent | possible pseudogene & putative transposase ⁵ |
| 13 | -----ac | NMA0590 | 282 | tandem | hypothetical protein |
| 14 | -----aa | ATR repeats | | | |
| 15 | -----tt | NMA0788 ⁴ | 127 | convergent | putative secreted toxin (probable pseudogene) |
| 16 | -----G----tt | RS-dRS3 repeat array | | | |
| 17 | -----aa | NMA0823 | 28 | convergent | putative transposase (probable pseudogene) |
| 18 | -----G----tt | RS-dRS3 repeat array | | | |
| 19 | -----A-----ag | RS-dRS3 repeat array | | | |
| 20 | -----tt | NMA0940 | 118 | tandem | putative membrane protein |
| 21 | -----gt | inserted within another CE; nearest ORF is NMA0975 | 274 | tandem | putative integral membrane protein |
| 22 | -----tt | NMA0975 | 175 | tandem | putative integral membrane protein |
| 23 | -----ta | NMA1079 | 55 | tandem | hypothetical protein |
| 24 | -----aa | NMA1111 | CE end lies within ORF | convergent | hypothetical protein |
| 25 | -----tc | RS-dRS3 repeat array | | | |
| 26 | -----aa | NMA1331 | 105 | tandem | ABC transporter ATP-binding protein (probable pseudogene) |
| 27 | -----aa | NMA1411 | 95 | tandem | hypothetical protein |
| 28 | -----cg | NMA1414 | 131 | tandem | putative periplasmic protein |
| 29 | -----tc | NMA1588 (<i>truB</i>) ⁴ | 41 | convergent | tRNA pseudouridine synthase B |
| 30 | -----G-----tt | NMA1626 ⁴ | 124 | convergent | putative RTX-family exoprotein |
| 31 | -----at | NMA1696 (<i>gabD</i>) | 304 | tandem | succinate semialdehyde dehydrogenase |
| 32 | -----aa | NMA1808 (<i>metZ</i>) | 55 | tandem | O-succinylhomoserine sulfhydrolase |
| 33 | -----tt | NMA1935 & NMA1934 (<i>gcvP</i>) ^{4,5} | 24 | convergent | hypothetical protein & glycine dehydrogenase ⁵ |
| 34 | -----tt | NMA1954 | 42 | tandem | putative integral membrane protein |
| 35 | -----G--T-----ca | NMA1960 & NMA1959 (<i>rfaK</i>) ^{4,5} | CE end lies within ORF | convergent | hypothetical protein & α 1,2 N-acetylglucosamine transferase ⁵ |
| 36 | -----T-----ac | NMA2029 | 79 | convergent | hypothetical protein |
| 37 | -----aa | single RS-dRS3 repeat followed by NMA2051 | 283 | tandem | conserved hypothetical protein |

Table 4. Cont.

| | Nucleotide sequence of the Correia element end (29 bp) ¹ | Adjacent ORF or other feature (downstream of Correia promoter) ² | Distance of nearest ORF to Correia element (bp) | Direction of transcription ³ | Description of putative gene product |
|----|---|---|---|---|--|
| 38 | -----tc | NMA2108 (<i>hemN</i>) ⁴ | 138 | convergent | oxygen-independent co-protoporphyrinogen III oxidase |
| 39 | -----ac | NMA2205 | 108 | tandem | hypothetical protein |

¹Sequences are listed in order of ascending genome coordinate (coordinates not shown).

²"NMA" refers to open reading frames (ORFs) in *N. meningitidis* Z2491; the numbers indicate the physical order of the ORFs within the genome.

³Direction of transcription from the Correia α -right^{Y128T} end relative to that of the nearest ORF.

⁴Locus where a rho-independent transcriptional terminator is present no more than 300 bp from a Correia α -right^{Y128T}.

⁵The *thr* tRNA gene and NMA0530A, NMA1935 and NMA1960 ORFs are very small (75 bp, 84 bp, 147 bp, and 171 bp, respectively), so the name and description of the next ORF is given as well.

doi:10.1371/journal.pgen.1001277.t004

and α -left promoters. Could they substantially impact gene expression in the organism? We provide evidence for the transcription of *gidA*, a tRNA modification gene, from a nearby α -right^{Y128T} promoter (Figure 3). *gidA* mutants have pleiotropic effects in bacteria and include virulence defects in *Streptococcus pyogenes* and *Aeromonas hydrophila*. [46,47]. In this example, the α -right^{Y128T} end is retained in all three meningococcal strains (Table 4, line 2; Table S3, line 1; Table S4, line 2). However, it will undoubtedly be of much interest to consider whether natural variation in the distribution of CEs contributes to the development or persistence of hypervirulent lineages that are the source of most global meningococcal disease.

Potential regulatory RNAs

During our analysis of Correia α -right^{Y128T} ends in Z2491, we observed that several Correia promoters oriented towards the 3'-end of adjacent ORFs were located within short distance of a downstream transcriptional terminator. RT-PCR analysis detected RNA transcripts from two of the promoters. *N. meningitidis* does not have an extensive protein-based regulatory network for transcription and small non-coding RNAs might play a role in helping to bolster or expand this relatively skeletal network.

Certain CEs might also produce transcripts that read into the 3'-end of genes at some loci. These "antisense" transcripts could act in *cis* to modulate expression of the adjacent gene(s). Although *cis*-acting regulatory RNAs in bacteria are typically associated with extra-chromosomal and mobile elements [48], the plasticity of the meningococcal genome may favor this type of regulation.

Snapshot of evolution

CEs are not simply the degenerate remnants of transposition events that have accumulated over long periods of time. The homogeneity of CE sequences suggests that they were created relatively recently in a burst of transposition. It is not possible at present to say whether the transposition events took place in a single lineage, and were spread subsequently by genetic exchange, or whether they are the result of separate amplification events in multiple lineages. The picture is further complicated by the evidence for gene conversion between elements, as exemplified by the inter-conversion of different subtypes of CEs. These issues make it difficult to know whether CEs are under selection or evolving neutrally. However, under any model, functional elements may arise occasionally by chance. In identifying the strongest CE promoters the present work provides a way to assess the potential importance of specific CEs at loci of interest.

Materials and Methods

Bacterial strains

The following *E. coli* strains were used in this study: DH5 α [*endA1 hsdR17 glnV44 thi-1 recA1 gyrA relA1* Δ (*lacI*ZYA-*argF*)U169 *deoR* (ϕ 80 Δ lac Δ (*lacZ*M15)]; MC4100 [F⁻ *araD139* Δ (*argF-lac*)U169 *rspL150 relA1 jfbB5301 fruA25 deoC1 ptsF25*] (a kind gift from Ben Berks, University of Oxford); NR289 [MC4100 *nadA::Tn10* Δ (*gal-att-bio*)] (a kind gift from Natacha Ruiz, Princeton University); AB1157 [F⁻ *thr-1 leuB6 hisG4 thi-1 araC14* Δ (*gpt-proA*)62 *lacY1 tsx-33 glnV44 galk2 rfbC1 rpoS396 mtl-1 rpsL31 xylA5 mgl-51 argE3 kdgK5I*] (a gift from David Sherratt, University of Oxford); RC5001 [=MM294 (F⁻ *supE hsdR endA1 pro thi*)]; RC5006 [=NK9140 = MM294 *hip::CAT*]. The *N. meningitidis* strains used in this study for experimental and/or computational analyses are as follows: Z2491, serogroup A, NC_003116 [5]; MC58, serogroup B, NC_003112, [6]; FAM18, serogroup C, NC_008767, [3].

Plasmids

A list of plasmids used in this study and the details of their construction are presented in Table S1.

Bioinformatic analysis

In Figure 1 we present the consensus sequences and total numbers of almost-perfect CEs in the *N. meningitidis* serogroup A, B and C genome sequences (NC_003116, NC_003112 and NC_008767). We used the European Bioinformatics Institute (EBI) FASTA server to search the three genomes using our previous consensus sequences for CEs [9]. Visual inspection of the alignments revealed the existence of the eight discreet classes of CE represented in Figure 1. The elements were sorted manually into their respective groups and used to build eight new consensus sequences using the EBI ClustalX server. These eight 'first-round' consensus sequences were then used in a new round of FASTA searches of the three genomes. This yielded a set of 343 'almost-perfect' elements, which excludes a number of degenerate remnants and fragments that were eliminated from the analysis by the FASTA mismatch and gap penalties. During this second round of searching, it was again necessary to manually sort some of elements into their respective groups. This is because some CEs have as many differences in their central region as between their respective α and β repeats, and this leads to inconsistencies in the FASTA output. After sorting, a new set of second-round ClustalX consensus sequences was constructed from each of the groups (Figure 1A). As can be seen from the plot in Figure 1C, the great

majority of the elements differ from their respective consensus sequences by less than 2%.

The three-way whole-genome comparison of the α -right^{Y128T} repeats presented in Figure 4 was performed using two different methods, each of which gave identical results. Method 1: A BLAST search recovered a total of 114 α -right^{Y128T} repeats in the three genomes. Many of these were excluded from the set of 343 almost-perfect elements (Figure 1) because of indels or other rearrangements elsewhere in the element, but which are not expected to alter transcriptional activity from the ends. The 20 bp sequence flanking each of the 114 α -right^{Y128T} repeats was extracted and used as a sequence tag. Since a 20 bp tag is expected to be unambiguous in a 2 MB chromosome, it can be used to identify the genomic context of each element, and to evaluate the three genomes for the presence or absence of “homologous” α -right^{Y128T} sequences. Coordinates for the set of 114 α -right^{Y128T} repeats, along with the corresponding 20 bp flanking sequence tags, are provided in Table 4 and Tables S3, S4 and S5.

Method 2: CEs were extracted from the three genomes (AL157959, AE002098 and AM421808) with RepeatMasker (unpublished, www.repeatmasker.org), using the α - α consensus sequence shown in Figure 1 as a reference, under stringent parameters (-e wublast -dir. -nolow -no_is -gff -s -pa 2 -cutoff 300). The RepeatMasker output was converted to the Genbank format with a simple python script. A similar process was carried out for dSR3 and ATR repeats. Syntenic regions between *N. meningitidis* strains Z2491, MC58 and FAM18 were identified by ‘all versus all’ BLAST comparisons of these genomes. BLAST results are in tabular format (-m 8 option) and can be directly visualized with the Artemis Comparison Tool (ACT) (Dataset S1). The repeat density plots were generated by comparing each genome against itself, using BLAST (Dataset S1). Output data were parsed with a custom python script and BLAST hits (High Scoring Pairs [HSP]) with a score below 25 bits were discarded. The repeat density plot corresponds to the number of HSPs overlapping each genomic position and helps to quickly identify regions composed primarily of repetitive sequences. CEs with an α -right end were identified by successive pair-wise alignment to each of the four types of CE end, and the alignment with the best score was retained. This procedure was automated using a python script and uses the Waterman-Eggert alignment algorithm implemented in the MATCHER software provided by the EMBOSS toolkit. The α -right^{Y128T} polymorphism was scored by directly assessing position 128. Dataset S1 can be visualized in Artemis and ACT using the simplified instructions provided in Text S1.

Integration of reporter constructs in the *E. coli* chromosome

The strategy outlined below for generating chromosomal insertions is based on the procedure detailed by Hand and Silhavy (2000) [49]. *E. coli* RC5001 cells harboring pRS415 or one of twenty plasmid derivatives containing CE insertions were infected with bacteriophage λ RZ-5 and phage lysates were harvested. Each lysate contains a small fraction of recombinant phage molecules in which homologous recombination has occurred between the *lacZ*YA and *bla* gene sequences on pRS415 (or derivatives) and homologous sequences on λ RZ-5 resulting in a phage that contains the CE-*lacZ* reporter construct. The phage lysates were used to infect *E. coli* AB1157 and lysogens were selected on ampicillin-containing medium. To ensure that the lysogens contain only one prophage, P1 transduction was employed to transduce the locus (the recombinant phage and flanking chromosomal markers) to *E. coli* NR289. The recipient strain

was screened for the presence of the correct markers and for immunity to λ infection.

E. coli strains lacking IHF were constructed by P1 transduction of the 21 NR289 strains containing chromosomally-integrated CE-*lacZ* reporter constructs with phage lysates prepared from *E. coli* RC5006, a strain carrying a *cat* (chloramphenicol acetyltransferase) gene insertion in the *hip* (*himD*) gene (which encodes the β subunit of IHF).

β -galactosidase assays

The β -galactosidase detection assay was performed similarly to that first described by Jeffrey Miller (1972) [50]. *E. coli* strains were grown overnight at 37°C, diluted 1:100 in fresh LB broth and grown to mid-log phase (optical density at 600 nm of 0.5–0.7). Cells were pelleted by centrifugation, and resuspended in an equal volume of Z-buffer. Various amounts of the cell suspension were mixed with Z-buffer to a final volume of 1 ml. Cells were lysed with the addition of 50 μ l chloroform and 25 μ l 0.1% SDS. β -galactosidase activity was measured by recording the time the samples took to develop a yellow colour at 30°C after the addition of ONPG (2-Nitrophenyl β -D-galactopyranoside). Once a yellow colour was observed, reactions were stopped with 500 μ l 1 M Na₂CO₃. Cell debris was removed by centrifugation and the optical density of each sample at 420 nm was measured with a spectrophotometer.

For convenience these experiments were performed in *E. coli*. The σ^{70} promoter consensus for *Neisseria* sp. has not been defined rigorously. However, promoters from *N. meningitidis* and *E. coli* function well in each other. Sequences similar to the *E. coli* consensus are usually evident upstream of meningococcal genes and have been shown to drive comparable rates of transcription in the two organisms eg. [31,51].

Transcript mapping

The TRIzol method (Invitrogen) was employed to extract total RNA from mid-log phase *E. coli* MC4100 cells harboring pRS415 or a derivative containing one of 12 full-length or prime (Δ 50bp) CEs inserted upstream of *lacZ*YA (plasmids pRC661–pRC666 and pRC675–pRC680). 20 μ g aliquots of the RNA preparations were stored at –80°C. Prior to use, the RNA samples were treated with TURBO DNase (Ambion), extracted with phenol/chloroform/isoamyl alcohol and precipitated with ethanol to remove trace amounts of genomic DNA.

Primer extension reactions were performed with 20 μ g of cellular RNA mixed with 5 pmoles of 5′ end-labeled PAGE-purified primer (5′-GGTCATAGCTGTTTCCTGTGTG-3′) in 30 μ l of hybridization buffer (40 mM PIPES (pH 6.4), 1 mM EDTA (pH 8.0), 400 mM NaCl, 80% deionized formamide). The samples were heated to 85°C for 10 minutes, then slowly cooled to 45°C and maintained at that temperature overnight. The RNA was precipitated with ethanol and resuspended in a primer extension buffer (50 mM Tris-HCl (pH 8.3), 50 mM KCl, 10 mM MgCl₂, 10 mM DTT, 1 mM each dNTP, 0.5 mM spermidine and 2.8 mM sodium pyrophosphate) to which AMV reverse transcriptase (Promega) was added. The reactions were incubated at 42°C for 90 min, stopped with the addition of formamide-containing RNA loading buffer, boiled for 5 min, then loaded and run on a denaturing polyacrylamide gel.

Other procedures

All strains were grown on Luria-Bertani (LB) media at 37°C. The following antibiotics were used at the indicated concentrations: ampicillin, 50 μ g/ml; kanamycin, 50 μ g/ml; spectinomycin, 50 μ g/ml. Manipulations using DNA restriction and modification

enzymes were performed according to the manufacturers' recommendations. Most of these enzymes were obtained from New England Biolabs. PCR was performed either with Vent DNA polymerase or Phusion High-Fidelity DNA polymerase (both from New England BioLabs). Sequences of all cloned PCR products were confirmed by nucleotide sequencing. Reverse transcription was performed with Superscript III reverse transcriptase (Invitrogen) and 100 ng of *N. meningitidis* Z2491 RNA as template (the RNA was kindly provided by Chris Tang at Imperial College, London). The genomic locations and nucleotide sequences of the primers used for the RT-PCR reactions are provided in Figure 3 and Table S2.

Supporting Information

Dataset S1 Genome annotation files for viewing in Artemis and ACT genome browsers.

Found at: doi:10.1371/journal.pgen.1001277.s001 (25.93 MB ZIP)

Figure S1 Synteny diagrams of the ORF landscapes surrounding each of the 114 α -right^{Y128T} repeats. The ORF landscape surrounding each of the 114 α -right^{Y128T} repeats in strains Z2491, MC58, and FAM18 are shown. On each page the target element, or bait, is shown in the middle together with a repeat density track. The homologous loci in the other two strains are illustrated above and below. Pages 1–38 illustrate the 39 α -right^{Y128T} repeats in Z2491; pages 39–56 illustrate those α -right^{Y128T} repeats in MC58 that are not present in Z2491; pages 57 to 61 illustrate those α -right^{Y128T} repeats in FAM18 not present in either of the other two strains.

Found at: doi:10.1371/journal.pgen.1001277.s002 (1.94 MB PDF)

References

- Harrison LH, Trotter CL, Ramsay ME (2009) Global epidemiology of meningococcal disease. *Vaccine* 27 Suppl 2: B51–63.
- Sultan B, Labadi K, Guegan JF, Janicot S (2005) Climate drives the meningitis epidemics onset in west Africa. *PLoS Med* 2: e6. doi: 10.1371/journal.pmed.0030006.
- Correia SD, Vernikos GS, Snyder LA, Churcher C, Arrowsmith C, et al. (2007) Meningococcal genetic variation mechanisms viewed through comparative analysis of serogroup C strain FAM18. *PLoS Genet* 3: e23. doi:10.1371/journal.pgen.0030023.
- Davidsen T, Tonjum T (2006) Meningococcal genome dynamics. *Nat Rev Microbiol* 4: 11–22.
- Parkhill J, Achtman M, James KD, Bentley SD, Churcher C, et al. (2000) Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis* Z2491. *Nature* 404: 502–506.
- Tettelin H, Saunders NJ, Heidelberg J, Jeffries AC, Nelson KE, et al. (2000) Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58. *Science* 287: 1809–1815.
- Correia FF, 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 FF, 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.
- Buisine N, Tang CM, Chalmers R (2002) Transposon-like Correia elements: structure, distribution and genetic exchange between pathogenic *Neisseria* sp. *FEBS Lett* 522: 52–58.
- Clayey Bouuaert C, Chalmers RM (2010) Gene therapy vectors: the prospects and potentials of the cut-and-paste transposons. *Genetica* 138: 473–484.
- Clayey Bouuaert C, Chalmers R (2010) Transposition of the human *Hsmar1* transposon: rate-limiting steps and the importance of the flanking TA dinucleotide in second strand cleavage. *Nucleic Acids Res* 38: 190–202.
- Munoz-Lopez M, Siddique A, Bischerour J, Lorite P, Chalmers R, Palomeque T (2008) Transposition of *Mboumar-9*: identification of a new naturally active *mariner*-family transposon. *J Mol Biol* 382: 567–572.
- Liu D, Bischerour J, Siddique A, Buisine N, Bigot Y, Chalmers R (2007) The human SETMAR protein preserves most of the activities of the ancestral *Hsmar1* transposase. *Mol Cell Biol* 27: 1125–1132.
- Aziz RK, Breitbart M, Edwards RA (2010) Transposases are the most abundant, most ubiquitous genes in nature. *Nucleic Acids Res* 38: 4207–4217.
- Chalmers R, Blot M (1999) Insertion Sequences and Transposons. In: Charlebois RL, ed. *Organization of the Prokaryotic Genome*. Washington, D.C.: American Society for Microbiology. pp 151–169.
- Schmidt JM, Good RT, Appleton B, Sherrard J, Raymant GC, et al. (2010) Copy number variation and transposable elements feature in recent, ongoing adaptation at the *Cyp6g1* locus. *PLoS Genet* 6: e1000998. doi:10.1371/journal.pgen.1000998.
- Mahillon J, Chandler M (1998) Insertion sequences. *Microbiol Mol Biol Rev* 62: 725–774.
- Simons RW, Hoopes BC, McClure WR, Kleckner N (1983) Three promoters near the termini of *IS10* - pIN, pOUT, and pIII. *Cell* 34: 673–682.
- Glandsdorff N, Charlier D, Zafarullah M (1981) Activation of gene expression by IS2 and IS3. *Cold Spring Harb Symp Quant Biol* 45 Pt 1: 153–156.
- Hinton DM, Musso RE (1982) Transcription initiation sites within an IS2 insertion in a Gal-constitutive mutant of *Escherichia coli*. *Nucleic Acids Res* 10: 5015–5031.
- Prentki P, Teter B, Chandler M, Galas DJ (1986) Functional promoters created by the insertion of transposable element *IS1*. *J Mol Biol* 191: 383–393.
- Liu SV, Saunders NJ, Jeffries A, Rest RF (2002) Genome analysis and strain comparison of Correia repeats and Correia repeat-enclosed elements in pathogenic *Neisseria*. *J Bacteriol* 184: 6163–6173.
- De Gregorio E, Abrescia C, Carlomagno MS, Di Nocera PP (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 MS, Di Nocera PP (2003) Ribonuclease III-mediated processing of specific *Neisseria meningitidis* mRNAs. *Biochem J* 374: 799–805.
- Mazzone M, De Gregorio E, Lavitola A, Pagliarulo C, Alifano P, Di Nocera PP (2001) Whole-genome organization and functional properties of miniature DNA insertion sequences conserved in pathogenic *Neisseriae*. *Gene* 278: 211–222.
- Francis F, Ramirez-Arcos S, Salimnia H, Victor C, Dillon JR (2000) Organization and transcription of the division cell wall (*dac*) cluster in *Neisseria gonorrhoeae*. *Gene* 251: 141–151.
- Black CG, Fyfe JA, Davies JK (1995) A promoter associated with the neisserial repeat can be used to transcribe the *uvrB* gene from *Neisseria gonorrhoeae*. *J Bacteriol* 177: 1952–1958.
- Packiam M, Shell DM, Liu SV, Liu YB, McGee DJ, et al. (2006) Differential expression and transcriptional analysis of the alpha-2,3-sialyltransferase gene in pathogenic *Neisseria* spp. *Infect Immun* 74: 2637–2650.

29. Zhao S, Montanez GE, Kumar P, Sannigrahi S, Tzeng YL (2010) Regulatory role of the MisR/S two-component system in hemoglobin utilization in *Neisseria meningitidis*. *Infect Immun* 78: 1109–1122.
30. Kumar A, Malloch RA, Fujita N, Smillie DA, Ishihama A, Hayward RS (1993) The minus 35-recognition region of *Escherichia coli* sigma 70 is inessential for initiation of transcription at an “extended minus 10” promoter. *J Mol Biol* 232: 406–418.
31. Swartley JS, Ahn JH, Liu LJ, Kahler CM, Stephens DS (1996) Expression of sialic acid and polysialic acid in serogroup B *Neisseria meningitidis*: divergent transcription of biosynthesis and transport operons through a common promoter region. *J Bacteriol* 178: 4052–4059.
32. Ponnambalam S, Webster C, Bingham A, Busby S (1986) Transcription initiation at the *Escherichia coli* galactose operon promoters in the absence of the normal –35 region sequences. *J Biol Chem* 261: 16043–16048.
33. Azam TA, Iwata A, Nishimura A, Ueda S, Ishihama A (1999) Growth phase-dependent variation in protein composition of the *Escherichia coli* nucleoid. *J Bacteriol* 181: 6361–6370.
34. Ditto MD, Roberts D, Weisberg RA (1994) Growth phase variation of integration host factor level in *Escherichia coli*. *J Bacteriol* 176: 3738–3748.
35. Kingsford CL, Ayanbule K, Salzberg SL (2007) Rapid, accurate computational discovery of rho-independent transcription terminators illuminates their relationship to DNA uptake. *Genome Biol* 8: R22.
36. Gottesman S (2002) Stealth regulation: biological circuits with small RNA switches. *Genes Dev* 16: 2829–2842.
37. Gottesman S (2004) The small RNA regulators of *Escherichia coli*: roles and mechanisms*. *Annu Rev Microbiol* 58: 303–328.
38. Argaman L, Hershberg R, Vogel J, Bejerano G, Wagner EGH, Margalit H, Altuvia S (2001) Novel small RNA-encoding genes in the intergenic regions of *Escherichia coli*. *Curr Biol* 11: 941–950.
39. Chen S, Lesnik EA, Hall TA, Sampath R, Griffey RH, Ecker DJ, Blyn LB (2002) A bioinformatics based approach to discover small RNA genes in the *Escherichia coli* genome. *BioSystems* 65: 157–177.
40. Vogel J, Sharma CM (2005) How to find small non-coding RNAs in bacteria. *Biol Chem* 386: 1219–1238.
41. Enriquez R, Abad R, Chanto G, Corso A, Cruces R, et al. (2010) Deletion of the Correia element in the *mtr* gene complex of *Neisseria meningitidis*. *J Med Microbiol* 59: 1055–1060.
42. Snyder LA, Shafer WM, Saunders NJ (2003) Divergence and transcriptional analysis of the division cell wall (*dew*) gene cluster in *Neisseria* spp. *Mol Microbiol* 47: 431–442.
43. Swinger KK, Rice PA (2004) IHF and HU: flexible architects of bent DNA. *Curr Opin Struct Biol* 14: 28–35.
44. Dorman CJ (2009) Nucleoid-associated proteins and bacterial physiology. *Adv Appl Microbiol* 67: 47–64.
45. Rouquette-Loughlin CE, Balthazar JT, Hill SA, Shafer WM (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.
46. Sha J, Kozlova EV, Fadl AA, Olano JP, Houston CW, Peterson JW, Chopra AK (2004) Molecular characterization of a glucose-inhibited division gene, *gidA*, that regulates cytotoxic enterotoxin of *Aeromonas hydrophila*. *Infect Immun* 72: 1084–1095.
47. Cho KH, Caparon MG (2008) tRNA modification by GidA/MnmE is necessary for *Streptococcus pyogenes* virulence: a new strategy to make live attenuated strains. *Infect Immun* 76: 3176–3186.
48. Carpousis AJ (2003) Degradation of targeted mRNAs in *Escherichia coli*: regulation by a small antisense RNA. *Genes Dev* 17: 2351–2355.
49. Hand NJ, Silhavy TJ (2000) A practical guide to the construction and use of *lac* fusions in *Escherichia coli*. *Methods Enzymol* 326: 11–35.
50. Miller JH (1972) *Experiments in Molecular genetics*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
51. Sawaya R, Arhin FF, Moreau F, Coulton JW, Mills EL (1999) Mutational analysis of the promoter region of the *porA* gene of *Neisseria meningitidis*. *Gene* 233: 49–57.
52. Simons RW, Houman F, Kleckner N (1987) Improved single and multicopy *lac*-based cloning vectors for protein and operon fusions. *Gene* 53: 85–96.