Siddique, Azeem and Buisine, Nicolas and Chalmers, Ronald (2011) The transposon-like correia elements encode numerous strong promoters and provide a potential new mechanism for phase variation in the meningococcus. PLoS Genetics, 7 (1). 13/1-13/13. ISSN 1553-7390

Access from the University of Nottingham repository:
http://eprints.nottingham.ac.uk/2471/1/Chalmerscorreia.pdf

Copyright and reuse:

The Nottingham ePrints service makes this work by researchers of the University of Nottingham available open access under the following conditions.

This article is made available under the Creative Commons Attribution licence and may be reused according to the conditions of the licence. For more details see: http://creativecommons.org/licenses/by/2.5/

A note on versions:

The version presented here may differ from the published version or from the version of record. If you wish to cite this item you are advised to consult the publisher's version. Please see the repository url above for details on accessing the published version and note that access may require a subscription.

For more information, please contact eprints@nottingham.ac.uk
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 Chalmers1*

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

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

Introduction

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.

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


Editor: Josep Casadesús, Universidad de Sevilla, Spain

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

Neisseria meningitidis is an encapsulated Gram-negative diplococcus commensal of the human nasopharyngeal tract. Although carried asymptomatically 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 subtypes 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 25 and 26 bp long TATA box for transcription. 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 subtypes 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.

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

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 adenosine 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 σ70 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 σ70 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
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-rightY128T 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-rightY128T and β-rightY128T ends (Figure 1B and Table 1). Alteration of either sequence dramatically reduced the activity of α-rightY128T. The mutations attenuated transcription from the β-rightY128T 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 TGGTATAAT [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-rightY128T ends display the strongest transcriptional activity, a somewhat unexpected finding considering that the α-α element is the most common class of CE in N. meningitidis.

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

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Description</th>
<th>Miller units²</th>
<th>Standard error³</th>
<th>Fold difference⁴</th>
</tr>
</thead>
<tbody>
<tr>
<td>pRC746</td>
<td>lacZ host strain</td>
<td>0.54</td>
<td>0.11</td>
<td>-</td>
</tr>
<tr>
<td>pRC759</td>
<td>α-left</td>
<td>7.4</td>
<td>0.28</td>
<td>1.0</td>
</tr>
<tr>
<td>pRC748</td>
<td>α-right⁵</td>
<td>540</td>
<td>5.0</td>
<td>73</td>
</tr>
<tr>
<td>pRC750</td>
<td>α-rightY128T</td>
<td>75</td>
<td>7.2</td>
<td>10</td>
</tr>
<tr>
<td>pRC761</td>
<td>α-rightY128T; −10 box altered⁶</td>
<td>670</td>
<td>22</td>
<td>90</td>
</tr>
<tr>
<td>pRC779</td>
<td>α-rightY128T; −35 box altered⁶</td>
<td>46</td>
<td>1.9</td>
<td>6.3</td>
</tr>
<tr>
<td>pRC760</td>
<td>β-left</td>
<td>97</td>
<td>8.8</td>
<td>13</td>
</tr>
<tr>
<td>pRC749</td>
<td>β-right⁵</td>
<td>86</td>
<td>6.1</td>
<td>12</td>
</tr>
<tr>
<td>pRC751</td>
<td>β-rightY128T</td>
<td>140</td>
<td>7.4</td>
<td>19</td>
</tr>
<tr>
<td>pRC762</td>
<td>β-rightY128T; −10 box altered⁶</td>
<td>51</td>
<td>4.3</td>
<td>6.8</td>
</tr>
<tr>
<td>pRC781</td>
<td>β-rightY128T; −35 box altered⁶</td>
<td>89</td>
<td>7.8</td>
<td>12</td>
</tr>
</tbody>
</table>

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 1. The E. coli host strain is MC4100.

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

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/λ 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-rightY128T 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.
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 PI transduction to disrupt IHF expression in all 21 of our chromosomal CE-loci 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 2). 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 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

<table>
<thead>
<tr>
<th>Correia element</th>
<th>wild-type strain</th>
<th>hipA::cat strain</th>
<th>Fold difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>L</td>
<td>reverse</td>
<td>forward</td>
<td>reverse</td>
</tr>
<tr>
<td>LR51G</td>
<td>LR51G</td>
<td>LR51G</td>
<td>LR51G</td>
</tr>
<tr>
<td>R</td>
<td>30</td>
<td>0.3</td>
<td>1.0</td>
</tr>
<tr>
<td>LR52G</td>
<td>LR52G</td>
<td>LR52G</td>
<td>LR52G</td>
</tr>
<tr>
<td>a</td>
<td>0.5</td>
<td>0.6</td>
<td>1.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Correia element</th>
<th>wild-type strain</th>
<th>hipA::cat strain</th>
<th>Fold difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>L</td>
<td>reverse</td>
<td>forward</td>
<td>reverse</td>
</tr>
<tr>
<td>LR51G</td>
<td>LR51G</td>
<td>LR51G</td>
<td>LR51G</td>
</tr>
<tr>
<td>R</td>
<td>30</td>
<td>0.3</td>
<td>1.0</td>
</tr>
<tr>
<td>LR52G</td>
<td>LR52G</td>
<td>LR52G</td>
<td>LR52G</td>
</tr>
<tr>
<td>a</td>
<td>0.5</td>
<td>0.6</td>
<td>1.0</td>
</tr>
</tbody>
</table>

1Unless designated otherwise, the bases at positions 52 and 128 of the CE are adenine and cytosine, respectively.
2E. coli strain NR289.
3Promoter.
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

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-rightV120T 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-rightV120T promoter is likely to contribute substantially to the transcription of the gene. The other two α-rightV120T 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.
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 downstream ORFs. The TransTermHP server [35] was consulted to check for the transcriptional start points observed in (A). Each of them will contribute to the transcription of the downstream ORFs.

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)A28C 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)A128T 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)A28C 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 PiiG (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 CEIs 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). The inverted repeats of an ATR element are depicted as black arrowheads.

Dispersed Repeats in the Meningococcal Genome

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 \(\text{metR}^{\text{CDE}}\) 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 \(N. \text{meningitidis}\) 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, \(\sigma^5\) (\(\text{rpoD}\))-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 \(\text{rpoD}\) is absent in the meningococcus and the gonococcus [6,42]. However, a gearbox promoter could have affected our transcriptional analysis in \(E. \text{coli}\), which does encode \(\text{rpoD}\). 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 discernible 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 \(E. \text{coli}\), it...
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 2L-2R*Y128T* element in Z2491 on line 3 of Table 4 has been converted to 2L-2R*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 2L-2R*Y128T* repeat, which provides the strongest promoter activity in our assay (Table 2). However, one should note that other classes of element, particularly the 2L-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 2L-2R*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.

**Figure 4. Distribution of the Correia 2L-2R*Y128T* repeats in the N. meningitidis Z2491, MC58, and FAM18 genomes.** Horizontal black lines represent the *N. meningitidis* genomes. Colored triangles represent the 2L-2R*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 are represented above and below the line drive transcription to the right and left, respectively. The grey lines between genomes connect Correia 2L-2R*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
Table 4. Correia α-right128T repeat distribution in the genome of *N. meningitidis* serogroup A strain Z2491.

<table>
<thead>
<tr>
<th>Nucleotide sequence of the Correia element end (29 bp)1</th>
<th>Adjacent ORF or other feature (downstream of Correia promoter)2</th>
<th>Distance of nearest ORF to Correia element (bp)</th>
<th>Direction of transcription3</th>
<th>Description of putative gene product</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 TGTACTGGTTTTGTTAATCATAa</td>
<td>ATR repeat followed by NMA00594</td>
<td>176</td>
<td>convergent</td>
<td>modification methylase (probable pseudogene)</td>
</tr>
<tr>
<td>2 ---------------------------aa</td>
<td>NMA0074 (gidA)</td>
<td>42</td>
<td>tandem</td>
<td>tRNA uridine 5-carboxymethyl aminomethyl modification enzyme</td>
</tr>
<tr>
<td>3 ------------------------Tcccc</td>
<td>thr tRNA gene &amp; NMA00795</td>
<td>69</td>
<td>tandem</td>
<td>threonine tRNA &amp; putative integral membrane protein5</td>
</tr>
<tr>
<td>4 -------------------------------at</td>
<td>NMA0277</td>
<td>155</td>
<td>convergent</td>
<td>putative lipoprotein</td>
</tr>
<tr>
<td>5 -----------------------------Gtt</td>
<td>RS-drS3 repeat array</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 ---------------------------ca</td>
<td>NMA0381 (metR)</td>
<td>42</td>
<td>convergent</td>
<td>putative transcriptional activator</td>
</tr>
<tr>
<td>7 ---------------------------aa</td>
<td>single RS-drS3 repeat followed by NMA0402 (truA)</td>
<td>218</td>
<td>tandem</td>
<td>tRNA pseudouridine synthase</td>
</tr>
<tr>
<td>8 ---------------------------aa</td>
<td>NMA0412a</td>
<td>79</td>
<td>convergent</td>
<td>putative glucokinase</td>
</tr>
<tr>
<td>9 ---------------------------aa</td>
<td>NMA0434</td>
<td>30</td>
<td>tandem</td>
<td>hypothetical inner membrane protein</td>
</tr>
<tr>
<td>10 --------------------------ga</td>
<td>RS-drS3 repeat array &amp; another CE</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11 ---------------------------ca</td>
<td>NMA04955 (rpmE)</td>
<td>107</td>
<td>tandem</td>
<td>50S ribosomal protein L31</td>
</tr>
<tr>
<td>12 --------------------------gt</td>
<td>NMA0530A &amp; NMA0531a</td>
<td>259</td>
<td>convergent</td>
<td>possible pseudogene &amp; putative transposase5</td>
</tr>
<tr>
<td>13 --------------------------ac</td>
<td>NMA0590</td>
<td>282</td>
<td>tandem</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>14 --------------------------at</td>
<td>ATR repeats</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15 --------------------------ct</td>
<td>NMA0788a</td>
<td>127</td>
<td>convergent</td>
<td>putative secreted toxin (probable pseudogene)</td>
</tr>
<tr>
<td>16 --------------------------Gtt</td>
<td>RS-drS3 repeat array</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17 --------------------------aa</td>
<td>NMA0823</td>
<td>28</td>
<td>convergent</td>
<td>putative transposase (probable pseudogene)</td>
</tr>
<tr>
<td>18 --------------------------Gtt</td>
<td>RS-drS3 repeat array</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19 --------------------------Aag</td>
<td>RS-drS3 repeat array</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20 --------------------------Gtt</td>
<td>inserted within another CE; nearest ORF is NMA0975</td>
<td>274</td>
<td>tandem</td>
<td>putative integral membrane protein</td>
</tr>
<tr>
<td>21 --------------------------gta</td>
<td>NMA0975</td>
<td>175</td>
<td>tandem</td>
<td>putative integral membrane protein</td>
</tr>
<tr>
<td>22 --------------------------tt</td>
<td>NMA1079</td>
<td>55</td>
<td>tandem</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>23 --------------------------ta</td>
<td>NMA1111</td>
<td></td>
<td>convergent</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>24 --------------------------aa</td>
<td>NMA1331</td>
<td>105</td>
<td>tandem</td>
<td>ABC transporter ATP-binding protein (probable pseudogene)</td>
</tr>
<tr>
<td>25 --------------------------tc</td>
<td>RS-drS3 repeat array</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26 --------------------------aa</td>
<td>NMA1808 (metZ)</td>
<td>55</td>
<td>tandem</td>
<td>O-succinylhomoserine sulfhydrolase</td>
</tr>
<tr>
<td>27 --------------------------tt</td>
<td>NMA1935 &amp; NMA1934 (gcvP)</td>
<td>24</td>
<td>convergent</td>
<td>hypothetical protein &amp; glycine dehydrogenase5</td>
</tr>
<tr>
<td>28 --------------------------tt</td>
<td>NMA1954</td>
<td>42</td>
<td>tandem</td>
<td>putative integral membrane protein</td>
</tr>
<tr>
<td>29 --------------------------tc</td>
<td>NMA1960 &amp; NMA1959 (rfaT)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30 --------------------------tt</td>
<td>NMA2029</td>
<td>79</td>
<td>convergent</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>31 --------------------------at</td>
<td>NMA2051</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>32 --------------------------aa</td>
<td>single RS-drS3 repeat followed by NMA2051</td>
<td>283</td>
<td>tandem</td>
<td>conserved hypothetical protein</td>
</tr>
</tbody>
</table>
Table 4. Cont.

<table>
<thead>
<tr>
<th>Nucleotide sequence of the Correia element end (29 bp)</th>
<th>Adjacent ORF or other feature (downstream of Correia promoter)</th>
<th>Distance of nearest ORF to Correia element (bp)</th>
<th>Direction of transcription</th>
<th>Description of putative gene product</th>
</tr>
</thead>
<tbody>
<tr>
<td>38 --------tc</td>
<td>NMA2108 (hemN)</td>
<td>138</td>
<td>convergent</td>
<td>oxygen-independent co-proporphyrinogen III oxidase</td>
</tr>
<tr>
<td>39 --------ac</td>
<td>NMA2205</td>
<td>108</td>
<td>tandem</td>
<td>hypothetical protein</td>
</tr>
</tbody>
</table>

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-right128R 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-right128R.
5The th4 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 gldA, a tRNA modification gene, from a nearby α-right128R promoter (Figure 3). gldA mutants have pleiotropic effects in bacteria and include virulence defects in Streptococcus pneumoniae and Aeromonas hydrophila. [46,47]. In this example, the α-right128R 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-right128R 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: DH5α [endA1 hsdR17 glnV44 thi-1 recA1 gyrA relA1 Δ(lacI258-argF)U169 (F’ proAB LacZM15)]; MC4100 [F’ amD1939 Δ(argF-lac)], U169 rphL150 relA1 fbb5301 pmaA25 dcmC1 phoP23 [a kind gift from Ben Berka, University of Oxford]; NR298 [MC4100 mfdE:Tsα10 Δ(gal-att-bio)] [a kind gift from Natacha Ruiz, Princeton University]; AB1157 [F’ thi-1 lacYI6 hisG4 thi-1 araC14 Δ(lac-proA)62 lacZΔM15 tetR33 glnV44 gaiIK2 rfbC1 rps3966 met-l1 rpsL31 yajA3 mgl-31 argE3 kdgK31] [a gift from David Sherratt, University of Oxford]; RC5001 [= MM294 (F’ supE hsdR endA1 pro thi)]; RC5006 [= NK9140 = MM294 hsdR::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_006767, [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 \( \sigma^{\text{right}} \) 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 \( \sigma^{\text{right}} \) 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 \( \sigma^{\text{right}} \) 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” \( \sigma^{\text{right}} \) sequences. Coordinates for the set of 114 \( \sigma^{\text{right}} \) 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 (AI157959, AE002090 and AM421800) with RepeatMasker (unpublished, www.repeatmasker.org), using the \( \sigma^{\alpha} \) consensus sequence shown in Figure 1 as a reference, under stringent parameters (-e -wblastl -dir -nomol -no_is_gff -s -p2 -cutoff 300). The RepeatMasker output was converted to the Genbank format with a simple python script. A similar process was carried out for dSRS and ATR repeats. Syntenic regions between \( N. meningitidis \) strains Z2491, MC58 and FA1090 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 hits 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 \( \sigma^{\text{right}} \) 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 algorithm implemented in the MATCHEHS software provided by the EMBOSS toolkit. The \( \sigma^{\text{right}} \) repeat density plot was generated by directly accessing 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 20 plasmid derivatives containing CE insertions were infected with bacteriophage \( \lambda \) and phage lysates were harvested. Each lysate contains a small fraction of recombinant phage molecules in which homologous recombination has occurred between the lac\( \Sigma \) and \( bia \) gene sequences on pRS415 or one of the plasmids pRC651–pRC660. 20 \( \mu \)l aliquots of the RNA preparations were stored at \(-80^\circ\)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.

**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 (\( \Delta bop \)) CE inserts upstream of lac\( \Sigma \) (plasmids pRC661–pRC666 and pRC675–pRC680). 20 \( \mu \)l aliquots of the RNA preparations were stored at \(-80^\circ\)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.

**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 \( \mu \)g/ml; kanamycin, 50 \( \mu \)g/ml; spectinomycin 50 \( \mu \)g/ml. Manipulations using DNA restriction and modification

---

**\( \beta \)-galactosidase assays**

The \( \beta \)-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 \( \mu \)l chloroform and 25 \( \mu \)l 0.1% SDS. \( \beta \)-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 \( \beta \)-D-galactopyranoside). Once a yellow colour was observed, reactions were stopped with 500 \( \mu \)l 1 M Na\(_2\)CO\(_3\). 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 \( \sigma^{\text{70}} \) promoter consensus for \( N. meningitidis \) 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 (\( \Delta bop \)) CE inserts upstream of lac\( \Sigma \) (plasmids pRC661–pRC666 and pRC675–pRC680). 20 \( \mu \)l aliquots of the RNA preparations were stored at \(-80^\circ\)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.

**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 \( \mu \)g/ml; kanamycin, 50 \( \mu \)g/ml; spectinomycin, 50 \( \mu \)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 a-rightY128T repeats. The ORF landscape surrounding each of the 114 a-rightY128T 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 a-rightY128T repeats in Z2491; pages 39–56 illustrate those a-rightY128T repeats in strains Z2491; pages 57 to 61 illustrate those a-rightY128T 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

10. Claeyys Bouuaert C, Chalmers RM (2010) Gene therapy vectors: the prospects and below. Pages 1–38 illustrate the 39 a-rightY128T 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 a-rightY128T repeats in Z2491; pages 39–56 illustrate those a-rightY128T repeats in strains Z2491; pages 57 to 61 illustrate those a-rightY128T 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)

Table S1 Plasmids used in this study.

Table S2 Primers for RT-PCR.

Table S3 Correia a128T right repeat distribution in the genome of N. meningitidis serogroup B strain MC58.

Table S4 Correia a128T right repeat distribution in the genome of N. meningitidis serogroup C strain FAM18.

Table S5 The coordinates and sequences of the 20 bp sequence tag flanking each of the 114 a-rightY128T repeats in strains Z2491, MC58, and FAM18.

Text S1 Instructions for loading Dataset S1 into Artemis and ACT. Found at: doi:10.1371/journal.pgen.1001277.s008 (0.03 MB PDF)

Acknowledgments

We would like to thank Dr Shuqin Hu for valuable assistance with some of the plasmid constructs and Natacha Ruiz of Princeton University for advice on engineering the transcriptional reporter strains.

Author Contributions

Conceived and designed the experiments: AS NB RC. Performed the experiments: AS NB. Analyzed the data: AS NB RC. Wrote the paper: AS RC.

PLoS Genetics | www.plosgenetics.org 12 January 2011 | Volume 7 | Issue 1 | e1001277

Dispersed Repeats in the Meningococcal Genome