Promoter reuse in prokaryotes

Harm Nijveen,1 Mariana Matus-Garcia1,2,† and Mark Willem John van Passel2,*

1Laboratory of Bioinformatics; Wageningen University; Wageningen, The Netherlands; 2Systems and Synthetic Biology; Wageningen University; Wageningen, The Netherlands
†Current affiliation: Massachusetts Institute of Technology; Cambridge, MA USA

A necdotal evidence shows promoters being reused separate from their downstream gene, thus providing a mechanism for the efficient and rapid rewiring of a gene’s transcriptional regulation. We have identified over 4,000 groups of highly similar promoters using a conservative sequence similarity search in all fully sequenced prokaryotic genomes. About 6% of those groups are shared between bacteria from different taxonomic depth, including different genera, families, orders, classes and even phyla. Database searches against known mobile elements and RNA motifs have indicated that regulatory motifs such as riboswitches could be moved around on putative mobile promoters.

Reuse of protein coding DNA sequences through gene duplication and horizontal gene transfer is a well-known and profound innovative force in nature;1 in sharp contrast to this, the mobility of a gene’s transcription regulatory function encapsulated in its promoter region is much less known. There are a few well studied classes of mobile genetic elements that harbour functional promoters, like Correia elements;2 ERICs3 and REPIN.4 But also examples of duplicated promoters not associated with known mobile elements5,6 suggest that promoter reuse could represent a rampant and rapid mechanism of gene rewiring. In a recent publication Blount et al.7 identified a promoter capture event as a crucial step in the evolution of aerobic citrate utilization by a population of Escherichia coli in a long-term evolution experiment, and speculate that promoter capture may be an important and little appreciated adaptive force in genome evolution. Similarly, Bongers et al.8 described the activation of a silent lactate dehydrogenase gene by promoter recruitment in Lactococcus lactis. In both studies insertion sequences (IS) were involved in promoter mobility, though Blount et al. also found cases that were not associated with IS elements. In order to estimate the relevance of promoter recruitment in genome evolution we made a conservative inventory of such events in prokaryotes, which was recently published in Nucleic Acids Research.9

Tip of the Iceberg

To assess the extent of promoter reuse in bacteria we looked for groups of bacterial genes per genome that share highly similar sequences upstream of their transcriptional start site, but do not have obvious flanking paralogous coding sequences. More specifically, we extracted in silico the DNA region between positions -150 and -50 relative to the translational start site,10 so most of our upstream sequence fragments will not contain the important -10 (Pribnow) box, but should include the -35 element. Using BLAST11 we then searched for sequence pairs that matched with 80% or more nucleotide identity over at least 50 nucleotides, to select for short conserved

Keywords: promoters, rewiring, mobile elements, prokaryotes, horizontal gene transfer

Submitted: 11/01/12
Revised: 12/09/12
Accepted: 12/10/12
http://dx.doi.org/10.4161/mge.23195

*Correspondence to: Mark Willem John van Passel; Email: mvanpassel@gmail.com

Commentary to: Matus-Garcia M, Nijveen H, van Passel MWJ. Promoter propagation in prokaryotes. Nucleic Acids Res 2012;40:10032-40; PMID:22933716; http://dx.doi.org/10.1093/nar/gks787

www.landesbioscience.com
Mobile Genetic Elements 279–281; November/December 2012; © 2012 Landes Bioscience
DNA elements. Sequences with more than one hit in the database were clustered into families. Sequence pairs that in addition showed a high nucleotide identity in their adjacent coding sequences were assumed to be paralogs and excluded because for this study we were interested in the independent mobility of promoters, not duplicated regions (for details see the Materials and Methods in our Nucleic Acids Research paper).

We analyzed all available complete prokaryotic genomes (1,362; July 2011) and even with our strict selection criteria found over 4,000 families of highly similar sequences upstream of apparently unrelated coding sequences. The majority of these families actually consist of pairs that on average share 92% nucleotide identity, meaning that at least 46 out of 50 base pairs were conserved, but we also found pairs that were completely identical over 100 base pairs. Whether this level of high identity is the result of a strong selective pressure, or indicative of recent duplication events remains to be investigated. We termed these homologous non-coding sequences Putative Mobile Promoters, PMPs. In fact, some of these sequences likely are not promoters but have a different function that causes their conservation. Looking for known elements in our PMP set we actually found 42 tRNAs, 83 resembled other RNA families and Bacteria, no transfer events were observed. Some non-coding sequence elements like tRNAs are very well conserved over large evolutionary distances, but if highly similar sequences are found only in small number of distantly related species horizontal gene transfer is a more likely scenario. The large majority of the PMPs are located on a chromosome, but for one group of PMPs all members are in fact on plasmids. These plasmids are associated with multiple-drug resistance in pathogenic Salmonella and are frequently transferred between bacterial species.

More surprising even than the large number of promoter pairs sharing high nucleotide identity within one bacterial genome is that about 6% are shared between distantly related species. Clustering these based on sequence similarity resulted in 62 distinct groups, of which four are present in species that are related only by belonging to the same phylum (Table 1). As expected, inter-taxon transfers seem to decrease with phylogenetic distance and at the domain level, i.e., between Archaea and Bacteria, no transfer events were observed. Some non-coding sequence elements like tRNAs are very well conserved over large evolutionary distances, but if highly similar sequences are found only in small number of distantly related species horizontal gene transfer is a more likely scenario. The large majority of the PMPs are located on a chromosome, but for one group of PMPs all members are in fact on plasmids. These plasmids are associated with multiple-drug resistance in pathogenic Salmonella and are frequently transferred between bacterial species.

Although the genetic code for translating DNA to protein is extremely well conserved between species as distant as *Escherichia coli* and *Homo sapiens*, transcriptional cis-regulatory elements are much more variable and their activity can even differ between strains of the same species. It can therefore be expected that the 62 homologous PMPs are not primarily transcription factor binding sites, but rather have other (regulatory) functions causing their high conservation. Indeed, two of the PMPs that are shared between families of bacteria are known S-adenosylmethionine (SAM) binding riboswitches. The other 60 PMPs however did not resemble any of the RNA families included in the RFam database, so their function at present remains uncovered.

We conclude that we have uncovered a large number of putative mobile promoter families, present in numerous bacterial genomes. These may be involved in rapid adaptive processes via transcriptional rewiring, or include post-transcriptional regulatory functions. The ways these PMPs move within and between genomes is still unknown, but due to the large number of families, this may include diverse mobilization mechanisms.

Finally, although transcription regulation in euukaryotes is more complex than in bacteria, it seems obvious that also in euukaryotes promoter reuse offers a mechanism for rapid adaptation of gene expression. It would therefore be very interesting to extent our analysis to this domain, especially now more genomes and transcriptomes are becoming available that greatly facilitate the mapping of the core promoters.

**Disclosure of Potential Conflicts of Interest**

No potential conflicts of interest were disclosed.

**References**

1. van Passel MWJ, Marr RJ, Ochman H. The emergence and fate of horizontally acquired genes in *Escherichia coli*. PLoS Comput Biol 2008; 4:e1000059; PMID:18404206; http://dx.doi.org/10.1371/journal.pcbi.1000059.
2. Siddique A, Buisson N, Chalmers R. The transposon-like Correia elements encode numerous strong promoters and provide a potential new mechanism for phase variation in the meningococcus. PLoS Genet 2011; 7:e1001277; PMID:21283790; http://dx.doi.org/10.1371/journal.pgen.1001277.
3. De Gregorio E, Silvestro G, Perillo M, Carlomagno MS, Di Nocera PP. Enterobacterial repetitive intergenic consensus sequence repeats in yersiniae: genom ic organization and functional properties. J Bacteriol 2005; 187:7945-54; PMID:16291667; http://dx.doi.org/10.1128/JB.187.23.7945-7954.2005.
4. Bertels F, Rainey PB. Within-genome evolution of REPINs: a new family of miniature mobile DNA in bacteria. PLoS Genet 2011; 7:e1001232; PMID:21068159; http://dx.doi.org/10.1371/journal.pgen.1001232.
5. Usakin LA, Kogan GL, Kalmykova AI, Gvozdev VA. An alien promoter capture as a primary step of the evolution of testes-expressed repeats in the Drosophila melanogaster genome. Mol Biol Evol 2005; 22:1955-66; PMID:15829659; http://dx.doi.org/10.1093/molbev/msi47.
6. Vandeponde K, Andries V, van Roy F. The NBPF1 promoter has been recruited from the unrelated EVIS gene before sitemap radiation. Mol Biol Evol 2009; 26:1321-32; PMID:19282512; http://dx.doi.org/10.1093/molbev/msp047.
7. Blount ZD, Barrick JE, Davidson CJ, Lenski RE. Genomic analysis of a key innovation in an experimental Escherichia coli population. Nature 2012; 489:513-8; PMID:22992527; http://dx.doi.org/10.1038/nature11514.

8. Bongers RS, Hoeftnagel MHN, Starrenburg MJG, Steermink MA, Arends JG, Hugenholtz J, et al. IS981-mediated adaptive evolution recovers lactate production by ldhB transcription activation in a lactate dehydrogenase-deficient strain of Lactococcus lactis. J Bacteriol 2003; 185:4499-507; PMID:12867459; http://dx.doi.org/10.1128/JB.185.15.4499-4507.2003.

9. Matus-Garcia M, Nijveen H, van Passel MWJ. Promoter propagation in prokaryotes. Nucleic Acids Res 2012; 40:10032-40; PMID:22933716; http://dx.doi.org/10.1093/nar/gkr787.

10. Mendoza-Vargas A, Olvera L, Olvera M, Grande R, Vega-Alvarado L, Taboada B, et al. Genome-wide identification of transcription start sites, promoters and transcription factor binding sites in E. coli. PLoS One 2009; 4:e7526; PMID:19383895; http://dx.doi.org/10.1371/journal.pone.0007526.

11. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 1997; 25:3389-402; PMID:9254694; http://dx.doi.org/10.1093/nar/25.17.3389.

12. Siguer P, Perochon J, Lestrade L, Mahillon J, Chandler M. ISfinder: the reference centre for bacterial insertion sequences. Nucleic Acids Res 2006; 34(Database issue):D32-6; PMID:16381877; http://dx.doi.org/10.1093/nar/gkj014.

13. Nielsen P, Krogh A. Large-scale prokaryotic gene prediction and comparison to genome annotation. Bioinformatics 2005; 21:4322-9; PMID:16249266; http://dx.doi.org/10.1093/bioinformatics/bti701.

14. Saks ME, Conery JS. Anticodon-dependent conservation of bacterial tRNA gene sequences. RNA 2007; 13:651-60; PMID:17379816; http://dx.doi.org/10.1261/rna.345907.

15. Fricke WF, Welch TJ, McDermott PF, Mammel MK, LeClerc JE, White DG, et al. Comparative genomics of the IncA/C multidrug resistance plasmid family. J Bacteriol 2009; 191:4750-7; PMID:19482926; http://dx.doi.org/10.1128/JB.00189-09.

16. Deniger SW, Fay JC. Frequent gain and loss of functional transcription factor binding sites. PLoS Comput Biol 2007; 3:e99; PMID:17530920; http://dx.doi.org/10.1371/journal.pcbi.0030099.

17. Hendriksen WT, Silva N, Bootsm JH, et al. Regulation of Gene Expression in Streptococcus pneumoniae by Response Regulator 09 Is Strain Dependent. mmb.asm.org.

18. van Hijum SAFT, Medema MH, Kuipers OP. Mechanisms and evolution of control logic in prokaryotic transcriptional regulation. [– Table of Contents.]. Microbiol Mol Biol Rev 2009; 73:481-509; PMID:19721087; http://dx.doi.org/10.1128/MMBR.00037-08.

19. Weinberg Z, Barrick JE, Yao Z, Roth A, Kim JN, Gore J, et al. Identification of 22 candidate structured RNAs in bacteria using the CMfinder comparative genomics pipeline. Nucleic Acids Res 2007; 35:4809-19; PMID:17621584; http://dx.doi.org/10.1093/nar/gkm487.