A Study on Mutational Dynamics of Simple Sequence Repeats in Relation to Mismatch Repair System in Prokaryotic Genomes

Pankaj Kumar · H. A. Nagarajaram

Abstract Mutational bias toward expansion or contraction of simple sequence repeats (SSRs) is referred to as directionality of SSR evolution. In this communication, we report the mutational bias exhibited by mononucleotide SSRs occurring in the non-coding regions of several prokaryotic genomes. Our investigations revealed that the strains or species lacking mismatch repair (MMR) system generally show higher number of polymorphic SSRs than those species/strains having MMR system. An exception to this observation was seen in the mycobacterial genomes that are MMR deficient where only a few SSR tracts were seen with mutations. This low incidence of SSR mutations even in the MMR-deficient background could be attributed to the high fidelity of the DNA polymerases as a consequence of high generation time of the mycobacteria. MMR system-deficient species generally did not show any bias toward mononucleotide SSR expansions or contractions indicating a neutral evolution of SSRs in these species. The MMR-proficient species in which the observed mutations correspond to secondary mutations showed bias toward contraction of polymononucleotide tracts, perhaps, indicating low efficiency of MMR system to repair SSR-induced slippage errors on template strands. This bias toward deletion in the mononucleotide SSR tracts might be a probable reason behind scarcity for long poly A|T and G|C tracts in prokaryotic systems which are mostly MMR proficient. In conclusion, our study clearly demonstrates mutational dynamics of SSRs in relation to the presence/absence of MMR system in the prokaryotic system.

Keywords Simple sequence repeats · Microsatellites · Polymorphism · Directionality · Prokaryotes · Mutation

Introduction

The DNA replication process can involve introduction of wrong bases into the newly synthesized strand at a rate of one wrong base per 10^3–10^6 bases (Kunkel 2004; Kunkel and Bebenek 2000; Kunkel and Loeb 1980; Umar and Kunkel 1996). Wrongly incorporated base (mismatch or unpaired base) is referred to as primary mutation because this depends on the enzymatic property of DNA polymerase (Harr et al. 2002). To correct wrongly incorporated bases, organisms have evolved a system called mismatch repair system (MMRS). MMRS reduces the number of mismatch errors by 100–1000 times thereby maintaining the genomic stability (Cox 1995). The MMRS in *Escherichia coli* is constituted by three enzymes MutS, MutL, and MutH. Homologs of MutS and MutL are present in almost every prokaryote and eukaryote (Eisen 1998; Lin et al. 2007) whereas the omnipresence for MutH has been doubted (Claverys and Lacks 1986). Structure and function of MutS have been studied in *Escherichia coli* and *Bacillus subtilis* (Sixma 2001; Smith et al. 2001; Yang 2000). It is now known that the functional form of MutS is a homodimer (albeit structurally heterodimer) that recognizes the

P. Kumar · H. A. Nagarajaram (✉)
Laboratory of Computational Biology, Centre for DNA Fingerprinting and Diagnostics (CDFD), Tuljaguda Complex, 4-1-714, Mozamjahi Rd, Nampally, Hyderabad 500 001, India
e-mail: han@cdfd.org.in

Present Address:
P. Kumar
Biochemistry and Molecular Genetics, University of Virginia School of Medicine, Charlottesville, VA 22908, USA

mismatched nucleotide and INDELs (insertions and deletions), and forms a complex with the DNA (Hsieh 2001; Iyer et al. 2006; Lamers et al. 2000; Li 2008; Obmolova et al. 2000). The MutS–DNA complex further interacts with the MutL homodimer protein in an energy-dependent manner (Acharya et al. 2003; Selmane et al. 2003). The interaction of these two complexes activates MutH which is an endonuclease, that cleaves the newly synthesized strand by recognizing the methylation pattern of "A" in the motif GATC on template strand and finally removes the mismatched bases by the excision of incorrect nucleotides on nascent strand (primer strand) (Lee et al. 2005). MMR system not only repairs mismatches but also some short INDELs in simple sequence repeats (SSRs) (Jaworski et al. 1995; Modrich and Lahue 1996; Parker and Marinus 1992; Schofield and Hsieh 2003). Mutation in the MMR system has, generally, been shown to destabilize SSRs in eukaryotes (Acharya et al. 1996; Strand et al. 1993) as well as in prokaryotes (Jaworski et al. 1995; Vogler et al. 2006). For example, it has been reported that in *Neisseria meningitides,* MutS mutation destabilizes the homopolymeric tracts (Martin et al. 2004). Similarly, the MutS mutation has been shown to destabilize dinucleotide (5′ AT) repeats in *Haemophilus influenzae* genome (Bayliss et al. 2002). However, in the same genome it has been shown that only inactivation of polI destabilized tetranucleotide (5′ AGTC) repeat tracts, but not MutS (Bayliss et al. 2002).

SSRs, also known as microsatellites, are the nucleotide sequences of repeats of 1–6 bp (Schlotterer 2000). These sequences are highly polymorphic characterized by high rates of insertion and deletion (INDEL) mutations of their repeat units (Garcia-Diaz and Kunkel 2006; Kunkel 2004; Sreenu et al. 2006; van Belkum et al. 1998). INDELs of repeat units in a SSR arise as a consequence of slipped strand mispairing during the process of DNA replication (Levinson and Gutman 1987; Schlotterer and Tautz 1992; Streisinger et al. 1966). Slippage on the template strand leads to contraction (deletion of repeat units) of SSRs whereas slippage on the nascent strand manifests into expansion (insertion of repeat units) of SSRs (Garcia-Diaz and Kunkel 2006; Harr et al. 2002; Levinson and Gutman 1987; Mirkin 2005; Streisinger et al. 1966). The bias in SSRs either toward expansion or contraction is referred to as the directionality of SSR evolution. Ever since it became known that several hereditary diseases are associated with expansion of triplet repeats (Ashley and Warren 1995) and colon cancer is associated with instability of certain mono- and di-nucleotide repeats (Aaltonen et al. 1993; Ionov et al. 1993; Marra and Boland 1995; Thibodeau et al. 1993), there has been a lot of interest on the discovery of the mechanisms behind directionality of SSR mutations. Despite some investigations into the directional evolution of the SSRs (Amos and Rubinstzein 1996; Amos et al.

1996; Ellegren 2002; Ellegren et al. 1995; Harr et al. 2002; Mirkin 2005; Primmer et al. 1996; Rubinsztein et al. 1995a, b; Webster et al. 2002), there is still an insufficient understanding of the factors influencing directionality of SSR mutations. There have been conflicting reports with regard to the directionality of SSRs in prokaryotes as well as in eukaryotes (Henderson and Petes 1992; Metzgar et al. 2002). In most of the cases dinucleotide repeats were the subjects of studies which reported that mutations are biased toward expansion in human as well as in swallow (Ellegren et al. 1997). Twerdi et al. (1999) reported expansion bias in SSRs in the MMRS-deficient as well as the MMRS-proficient mammalian cell lines. Similar results (expansion bias) were reported by Yamada et al. (2002) in the human MMRS-deficient and the proficient cell lines. Webster et al. (2002) reported bias of certain dinucleotide SSRs for expansion by comparing the human and chimpanzee genomes. Xu et al. (2000) reported that long alleles of a tetranucleotide SSR tracts show bias toward contraction. Huang et al. (2002) reported that the long alleles of a dinucleotide repeat show bias toward contraction whereas short alleles show bias toward expansion. Even mononucleotides repeats were also shown to be biased toward expansion in the MMRS-proficient as well as the MMRS-deficient mammalian cell lines. It has also been argued that the MMRS does not influence the direction of mutation (Boyer et al. 2002). However, in-depth study by Harr et al. (2002) in drosophila showed the role of the MMRS in the directionality of SSR mutation. In the wild-type cell lines SSR mutations were significantly biased toward contraction whereas in the *spellchecker* mutation accumulation cell lines deficient for MMRS SSR mutations were slightly biased toward expansion (Ellegren 2002; Harr et al. 2002).

While most of the studies focused on eukaryotic systems only very few have focused on prokaryotes. Studies on *Haemophilus influenza* and *Escherichia coli* genomes (De Bolle et al. 2000; Morel et al. 1998) suggested that bacterial SSRs do show directionality during evolution. The lengths of the SSR tracts studied were $(AGTC)_{17-38}$ and $(AC)_{51}$ in *Haemophilus influenzae* and *Escherichia coli*, respectively. It is to be noted that *Escherichia coli* has the functional MMR system (Levy and Cebula 2001) and its homolog is present in *Haemophilus influenza* as well. In *Mycoplasma gallisepticum* where the MMRS appears to be absent (Carvalho et al. 2005; Himmelreich et al. 1996), it was observed that a trinucleotide SSR tract $(GAA)_{12}$ is biased toward contraction (Metzgar et al. 2002). Furthermore, a study in *Escherichia coli* on the tetranucleotide tract $(AAGG)_9$ revealed that the tract is expansion biased (Eckert and Yan 2000). It should be noted that the lengths of SSRs considered in the aforementioned studies are much longer than the typical lengths of SSRs observed in the prokaryotic genomes (Field and Wills 1998; Mrazek et al.

2007; Sreenu et al. 2007) and hence observations made in those studies may not adequately represent the mutational bias of SSRs in the prokaryotic genomes. The reported contraction bias of these long tracts can be due to the fact that long SSRs experience, on an average, downward mutational bias. Furthermore, the studies were conducted on individual SSR tracts and were carried out under defined laboratory conditions. None of these studies reported mutational dynamics of mononucleotide repeats which are the most abundant repeat types in prokaryotic genomes (Coenye and Vandamme 2005). Lack of a global study on behavior of SSR mutations as well as the availability of whole genome sequences of variety of strains belonging to a number of species gave us tremendous advantage to analyze mutational pattern of SSRs in relation to the presence and absence of the MMRS. Studies were carried out on the SSRs found in the non-coding regions as these are thought to be neutral as far as the selection is concerned. Our studies revealed that SSRs in the genomes where the MMRS was present show mutational bias toward contraction. In the genomes where the MMRS is absent SSRs do not show bias toward either expansion or contraction.

Materials and Methods

Systematic Search for MMRS in Bacteria and Archaea Genomes

The whole genome sequences of prokaryotes were downloaded from NCBI's ftp site (ftp://ftp.ncbi.nih.gov/genomes/Bacteria/). We considered only those species of prokaryotes for which the complete genome sequences are available for at least three strains. The annotations available along with the genome sequences were searched for the presence of keywords "MutS" and "MutL". We did not consider MutH because its universal presence as a part of the MMRS in the prokaryotic genomes has been doubted (Claverys and Lacks 1986). It should be noted that in some prokaryotic genomes two types of MutS genes are present viz., MutS1 and MutS2. MutS1 is equivalent to *Escherichia coli* MutS and MutS2 counterpart is absent in *Escherichia coli*. The MutS1 has four functional conserved domains viz., MutSI, MutSII, MutSd, and MutSac. MutS2 lacks the MutSI, MutSII domains but has an additional SMR domain at its C-terminal end (Lin et al. 2007). In the *Helicobacter pylori* it has been shown that MutS2 does not play any role in MMR (Kang et al. 2005). Hence, we looked for the presence of MutS1 among the MutS homologs as it is the one that is involved in MMR. We therefore refer MutS1 as MutS unless otherwise it is mentioned specifically.

The protein sequences corresponding to the ORFs annotated as MutS and MutL were searched for the presence of functionally conserved domains characteristic of MutS and MutL using Pfam (Finn et al. 2008) and SMART databases (Schultz et al. 1998). Only those sequences containing all the characteristic functional conserved domains were considered as MutS and MutL homologs. Genomes where MutS and MutL homologs were absent and those in which the homologs were found but missing one of the characteristic functional domains was subjected to tblastn searches using full length MutS and MutL ORFs as queries. This enabled us to discover in some genomes DNA sequences very similar to full length MutS/MutL ORFs but containing frameshift mutations causing premature terminations. Those genomes which failed to yield sequences similar to MutS/MutL were further subjected to profile-based searches using PROFILE-SS to reconfirm the complete absence of the two genes. The profiles for MutS and MutL were constructed from their respective multiple sequence alignments of all the homologs gathered by us. The results pertaining to the presence/absence of MutS/MutL homologs in some genomes were also reconfirmed by comparing with the available published literature (Eisen 1998; Lin et al. 2007).

Extraction of SSRs and Comparison of Equivalent SSRs

SSRF (Sreenu et al. 2003) was used for the extraction of SSRs from the complete genome sequences. SSRF reports the motif type, number of iterations, start and end of SSR in the genome, and the region where SSRs are observed (coding or non-coding). The equivalent SSRs in the genomes belonging to a species were identified by comparing the flanking regions. SSRs were considered as equivalent provided the flanking regions of at least 50 bp were identical to each other in all the compared genomes. The equivalent SSRs showing length variation were considered as polymorphic SSRs (PSSRs). The entire process of genome sequence comparisons for identification of PSSRs has been coded in the form of a computer program called polymorphic simple sequence repeat finder (PSSRFinder) (Kumar et al. 2011). PSSRs harbored within the non-coding regions of all the compared genomes were categorized as non-coding PSSRs. Of these PSSRs, we considered only those showing two polymorphic states (i.e., two types of alleles) and among these equally populated allelic cases were discarded. In the cases (species) where multiple sub-strains were available only one sub-strain was considered for analysis. Information pertaining to the polymorphic SSRs identified from whole genomes belonging to 85 species have been made available in the form of a relational database called PSSRdb (pssrdb.cdfd.org.in) (Kumar et al. 2011).

Species-Wise Identification of Contractions/Expansions in PSSRs

To decide whether a given PSSR in a species is a case of expansion or contraction we examined the length distribution of PSSRs across strains. If the major allele is longer than the minor allele then the PSSR was considered as a case of contraction whereas if the major allele is shorter than the minor allele then the PSSR was considered as a case of expansion. However, in the case of Yersinia species the status of PSSRs as expansion or contraction was decided with reference to its ancestor *Yersinia pseudotuberculosis*. Furthermore, the PSSRs with reference alleles (the major or ancestor allele) longer than two repeat units were only considered for further analysis.

Statistical Significance of Contraction or Expansion Bias

The significance of contraction or expansion events was tested by performing χ^2 test with one degree of freedom (any χ^2 value >3.84 is considered as significant). χ^2 value is calculated as given below.

$$\chi^2 = \sum_{i=1}^{k} (O_i - E_i)^2 / E_i$$

where O_i is the observed number of events, E_i is the expected number of events (total number of contraction and contraction divided by 2), $i = 1$, $k = 2$ [only two events are possible (expansion or contraction)].

Results and Discussion

MMR System in Bacterial and Archaeal genomes

Our search for the *MutS and MutL* genes in the bacterial and archaeal genomes yielded three groups of species: (a) the species where one of the strains lacks the MMRS (Table 1), (b) the species where all the strains lack MMRS (Table 1), and (c) the species where all the strains have MMRS. Additional information such as name of the strain and protein IDs of MutS and MutL of each strain are given as Supplementary Table I.

The Species Where One of the Strains Lacks MMRS

Of the species screened we found *Haemophilus influenza* and *Acinetobacter baumannii* where one of the strains (pittGG strain in *Haemophilus influenzae*, and ATCC 17978 strain in *Acinetobacter baumannii*) lacked functional MutS due to a frameshift mutation. In *Haemophilus influenzae* pittGG strain MutS is prematurely terminated due to the deletion of one repeat unit in a 7-bp mononucleotide, a poly "A" SSR tract positioned at 2,155 bp as compared to its homologs in the other strains of *Haemophilus influenzae*. Similarly in *Acinetobacter baumannii* ATCC 17978 strain MutS is frame shifted at position 1,989 due to the insertion of one nucleotide in the mononucleotide tract of a poly "A" tract, which is 6 bp in length in all other strains of this species. Due to this frame shift the ORF has split into two parts, A1S_1251 (573) and A1S_1252 (206). In the ORF A1S_1251, the ATPase and MutSIII domains are intact while in the second ORF the MutSII

Table 1 Species where *MutS* and *MutL* genes are present ("+") and absent ("−")

A				B			
Species name	Name of strains	MutS	MutL	Species name	#Strains	MutS	MutL
Acinetobacter baumannii	AB0057, AB307, ACICU, and SDF	+	+	*Campylobacter jejuni*	5	−	−
	ATCC	−	−	*Helicobacter pylori*	6	−	−
Haemophilus influenzae	RD_KW20, 028NP, F3031, F3047 and PittEE	+	+	*Mycobacterium tuberculosis*	3	−	−
	PittGG	−	+	*Mycobacterium bovis*	3	−	−
Staphylococcus aureus	MRSA, MSSA476, MW2, JH1, Mu3, Mu50, N315, COL, Newmann, NCTCT, USA300	+	+	*Mycoplasma hyopneumoniae*	3	−	−
				Corynebacterium glutamicum	3	−	−
	RF122	+	−	*Methanococcus maripaludis*	4	−	−
				Sulpholobus islandicus	6	−	−

In *Acinetobacter baumannii*, *Haemophilus influenzae*, and *Staphylococcus aureus* one of the strains lacks functional MMRS because of a frameshift mutation and therefore the corresponding strains are MMRS deficient

domain which is believed to be a connector domain between mismatch recognition domain and ATPase domain (Lamers et al. 2000; Obmolova et al. 2000) has been deleted. Although it is annotated as a coding region in the genome annotation file (Smith et al. 2007), splitting of domain suggests that the MutS protein may not be functional in this strain. In addition to this we also found in *Acinetobacter baumannii* and *Staphylococcus aureus* where one of the strains (ATCC 17978 strain in *Acinetobacter baumannii* and RF122 strains of *Staphylococcus aureus*) lacked functional MutL due to a frameshift mutation. In *Acinetobacter baumannii* ATCC 17978 *MutL* gene is frame shifted due to the deletion of one "A" in a poly "A" tract of 7 bp. In *Acinetobacter baumannii* wild-type gene of *MutL* codes for a 650 amino acid long protein, however, due to frame-shift mutation in strain ATCC 17978, it has resulted in two ORFs of which the second ORF translates into a 369 amino acid sequence corresponding to a truncated TopoII domain which is a MMR domain. On the other hand, in *Staphylococcus aureus* RF122 strain *MutL* gene is frame shifted due to the deletion of 1 bp at position 1,450 in a non-SSR tract and hence the homologous region in the genome is annotated as non-coding region.

Species Where all the Strains Lack MMRS

In addition to the above-mentioned species we also found eight species (two belonging to archaea and six belonging to bacteria) where both *MutS* and *MutL* genes (the complete DNA sequence) are absent in all their known strains (Table 1) thus offering as suitable systems to study SSR mutations in the light of MMRS deficiency. Among these, MMRS deficiency in five species viz., *Mycobacterium tuberculosis*, *Mycobacterium bovis*, *Mycoplasma hyopneumoniae*, *Helicobacter pylori*, and *Campylobactor jejuni* has already been reported (Carvalho et al. 2005; Dos Vultos et al. 2009; Eisen 1998; Lin et al. 2007; Mizrahi and Andersen 1998).

We looked into the phylogenetic relationships among the MMRS-deficient species (Fig. 1). It can be seen that MMRS-deficient species make four separate clusters but distinct from the clusters made by MMRS-proficient species. The fact that MMRS-deficient species form four clusters distinct from MMRS-proficient species indicates that the loss of MMRS is not an individual character but inherited from a common ancestor which might have selected for loss of MMRS. The phylogenic relationship obtained by our study is similar to the phylogenic relationship of prokaryotes generated on the basis of whole genome sequences by Henz et al. (2005). MMR-deficient species considered in our study are clustered in four different clusters in Henz et al. studies too. A total of 14

MMRS-deficient species are present in four different clusters in our study. However, of these 14 species, complete genome sequences of at least three strains in each species were only available for 7 species (please see the "Materials and Methods" for species selection criteria).

Polymorphic SSRs

Figure 2 shows the number of PSSRs found in each species as well as their densities in coding and non-coding regions (please see Supplementary Table I for the number and names of the strains in each species). In all the species the tract density in non-coding regions is strikingly higher than (on average ∼14 times) that in coding regions. Higher incidence of SSR polymorphism in non-coding regions as compared to coding regions indicates relatively unrestrained polymorphism in non-coding regions. Restraint on SSR polymorphism in coding regions can be attributed to selection pressures. On the other hand, non-coding regions do not have such selection pressures and hence the possibility of unrestrained SSR polymorphism. We therefore considered only the PSSRs found in non-coding regions. We found more than 11,000 non-coding PSSRs from 41 species (see "Materials and Methods" for species selection criteria) of which a large majority is constituted by mononucleotide tracts. The numbers of PSSRs of each repeat type in different species are given in Supplementary Table II. About 99% of the PSSRs have undergone INDEL mutations of one repeat unit. Since 95% of the PSSRs are the mononucleotide tracts we considered only these tracts for further expansion/contraction studies.

MMR Deficiency Leads to Destabilization of SSR Tracts

The number of PSSRs found in *Haemophilus influenza*, *Acinetobacter baumannii*, and *Staphylococcus aureus* are shown in Fig. 3. Most of the MMRS-deficient strains show significantly higher ($Z > 4$) number of PSSRs as compared to the MMRS-proficient strains indicating MMRS deficiency has led to destabilization of SSR tracts. This is in agreement with earlier reports (Levy and Cebula 2001; Watson et al. 2004). It is pertinent to note that MutS/MutL genes themselves harbor SSR tracts and as already noted the MMRS deficiency has arisen because of the frameshift mutations caused by SSR tracts and therefore in these species MMRS deficiency/proficiency is a matter under the control of SSR mutations. Unless fixed it is quite probable that mutating SSRs yield functional MMRS for some generations and non-functional MMRS for some other generations. An evidence to this argument can be seen from Fig. 3 where some of the strains ACICU of *Acinetobacter baumannii* and pittEE strain of *Haemophilus*

Fig. 1 The phylogenetic relationship among all the species considered in this study is shown. The phylogenetic tree was constructed based on the 16S rRNA sequences. The tree also includes some species which were not included in this analysis because they were not meeting the required criteria (minimum three strains complete genome sequence) as mentioned in "Materials and Methods". These extra species were included in the phylogenetic analysis because they are found to be MMRS deficient and cluster together with other MMRS-deficient species. MMR-deficient species are clustered into four different clusters (marked by *X*). Multiple sequence alignment was generated by CLUSTALX. Phylogeny was constructed by using the PHYLIP package. The un-rooted tree was drawn using draw-tree. Boot strap value used was 1,000
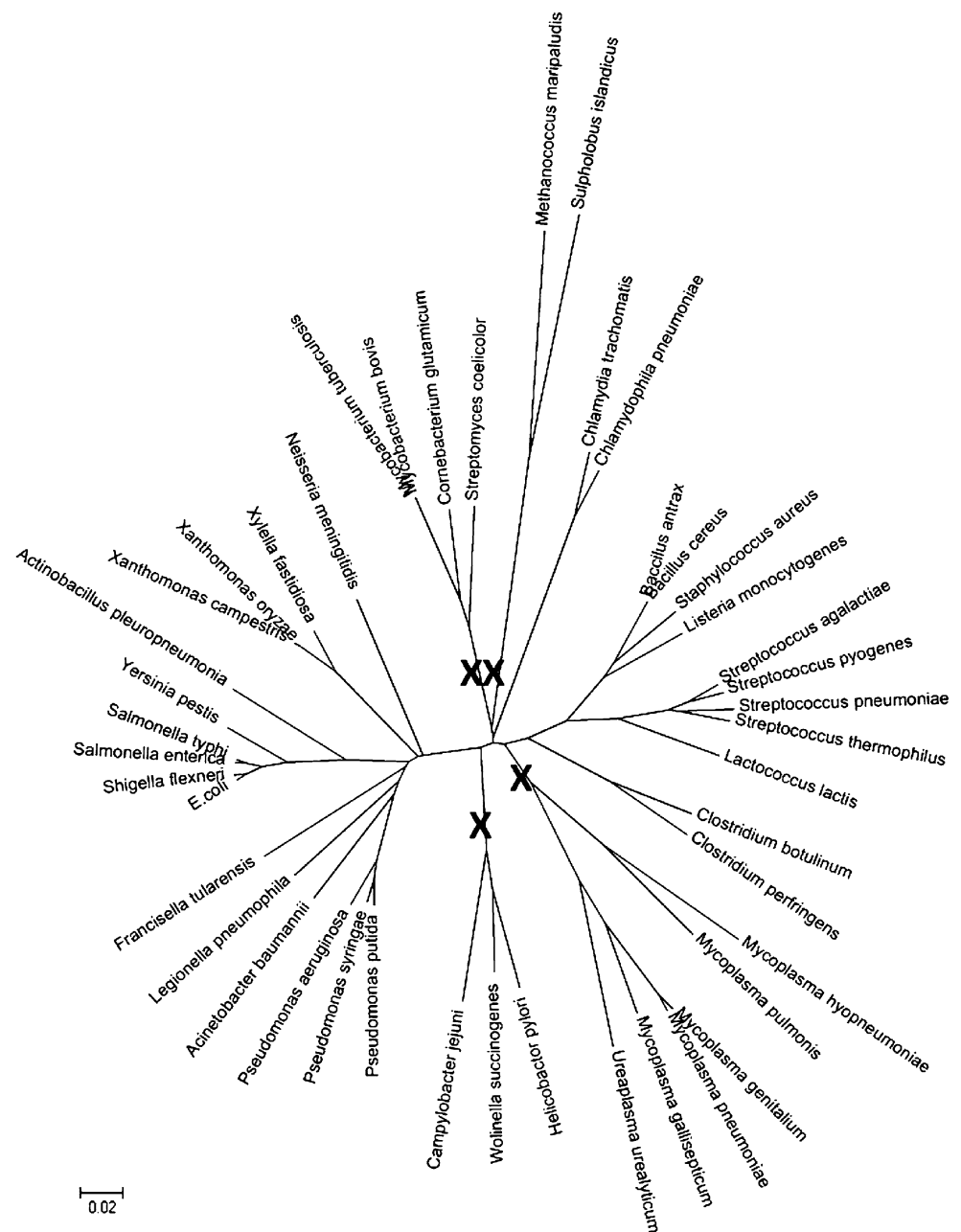
influenzae also show high level of SSR instability (however, with low significance with *Z* score < 4.0) despite being MMRS proficient.

We also found eight species where the genes of MMRS were completely absent in their respective genomes. MMRS deficiencies in the following five species viz., *Mycobacterium tuberculosis*, *Mycobacterium bovis*, Mycoplasma, Helicobactor and Campylobactor have already been reported (Carvalho et al. 2005; Dos Vultos et al. 2009; Eisen 1998; Lin et al. 2007; Mizrahi and Andersen 1998).

The densities of non-coding PSSRs (the number of PSSRs per Mb per strain) found in the MMRS-deficient

and MMRS-proficient species are shown in Fig. 4. Though we have not defined a threshold for PSSR density to discriminate species as "high" or "low" in PSSRs, it is pertinent to note that nearly half of the MMR-deficient species have high PSSR densities in the non-coding regions. However, as the PSSR density relates to SSR stability and hence we can say higher the PSSR density, lower is the SSR stability. From the figure it can be seen that *Mycoplasma hypopneumoniae*, *Methanococcus maripalusis*, *Helicobacter pylori*, and *Campylobacter jejuni* are positioned at the high end of the PSSR tract density chart. The other deficient species viz., Mycobacterium (includes both *Mycobacterium tuberculosis* and *Mycobacterium bovis*)
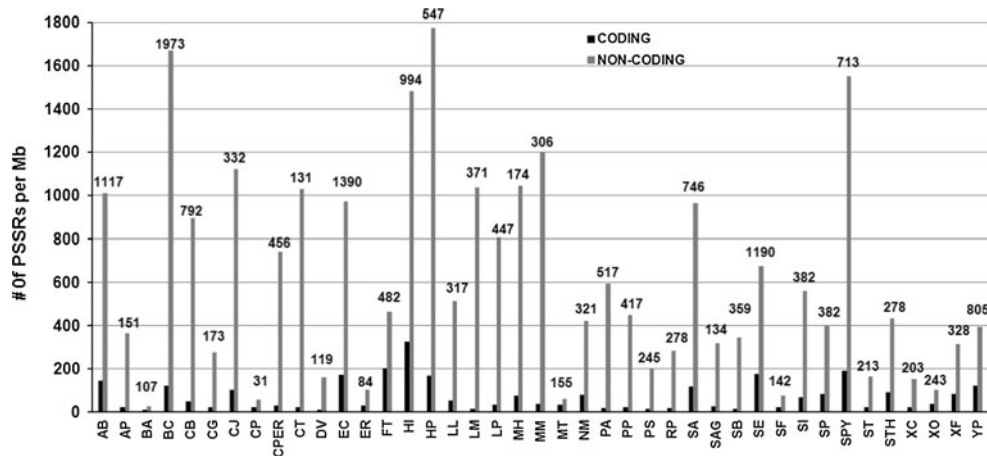
Fig. 2 The tract densities of PSSRs in coding and non-coding regions in various prokaryotic species are shown. The total number of PSSRs found in each species is shown on *top of the bar*. The tract density was calculated as number of PSSRs per Mb in given regions (coding or non-coding) in species. Total number of PSSRs is the number of non-redundant polymorphic loci observed in a species. Average of base pair in coding and non-coding regions in all strains in each species was considered as total number of bases in coding and non-coding regions, respectively. *AB Acinetobacter baumannii, AP Actinobacillus pleuropneumoniae, BA Bacillus anthracis, BC Bacillus cereus, CB Clostridium botulinum, CG Corynebacterium glutamicum, CJ Campylobacter jejuni, CP Chlamydophila pneumoniae, CPER Clostridium perfringens, CT Chlamydia trachomatis, DV Desulfovibrio vulgaris, EC*

Escherichia coli, ER Ehrlichia ruminantium, FT Francisella tularensis, HI Haemophilus influenzae, HP Helicobacter pylori, LL Lactococcus lactis, LM Listeria monocytogenes, LP Legionella pneumophila, MH Mycoplasma hyopneumoniae, MM Methanococcus maripaludis, MT Mycobacterium tuberculosis and Mycobacterium bovis, NM Neisseria meningitidis, PA Pseudomonas aeruginosa, PP Pseudomonas putida, PS Pseudomonas syringae, RP Rhodopseudomonas palustris, SA Staphylococcus aureus, SAG Streptococcus agalactiae, SB Shewanella baltica, SE Salmonella enterica, SF Shigella flexneri, SI Sulpholobus islandicus, SP Streptococcus pneumoniae, SPY Streptococcus pyogenes, ST Salmonella typhi, STH Streptococcus thermophilus, XC Xanthomonas campestris, XO Xanthomonas oryzae, XF Xylella fastidiosa, YP Yersinia pestis
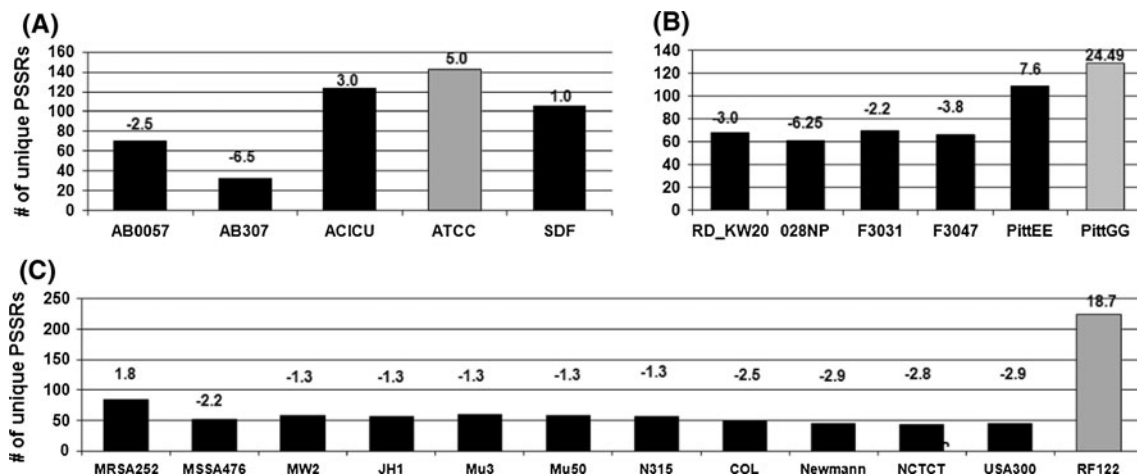

Fig. 3 The number of PSSRs (Obs) found in each strain of **a** *Acinetobacter baumannii*, **b** *Haemophilus influenzae*, and **c** *Staphylococcus aureus* is shown. The MMRS-deficient and MMRS-proficient strains are shown in *gray* and *black bars*, respectively.

Z scores $\left(Z - \text{Score} = \frac{(\text{Obs}-\text{Exp})^2}{\text{Exp}}\right)$ are indicated above the *bars*. Average number of PSSRs in each species was considered as expected number (Exp) of PSSRs in each species

and *Sulpholobus islandicus* and *Corynebacterium glutamicum* harbor relatively low density of PSSRs and seem as "outliers" among MMRS-deficient species. Mycobacteria are known for low mutation rates which have been linked to the slow rate of DNA synthesis in these bacteria (Dos Vultos et al. 2009; Hiriyanna and Ramakrishnan 1986). Slow DNA synthesis may promote increased fidelity of the DNA polymerases even in the MMRS-deficient

background (Radman 1998). The other outlier *Sulpholobus islandicus* is also a slow growing bacterium with the doubling time of 7–8 h. From the above observations it can be concluded that MMRS deficiency generally leads to the destabilization of SSRs.

However, we would like to point out an observation made in *Escherichia coli*, which is MMRS-proficient species, the F-plasmid can destabilize SSR tracts (Schlotterer
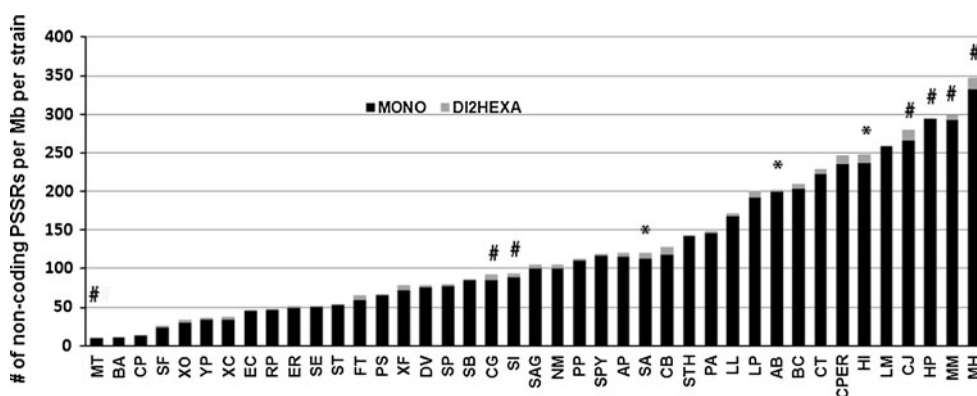
Fig. 4 Number of non-coding PSSRs per Mb per strain found in the non-coding regions of different prokaryotic species is shown. Due to heterogeneity in the number of strains in different species PSSR density was normalized to the total number of PSSR per mega base pairs per strain. The densities of PSSRs (the number of PSSRs per Mb per strain) found in the MMRS-deficient and MMRS-proficient species are shown. In each bar, the *darkly shaded portion* represents mononucleotide SSRs and the *lightly shaded portion* represents the other SSRs (di to hexa). The MMR-deficient species are marked with *number symbol* and species where one of the strains is deficient in MMR are marked with *asterisk*. For the x axis legend abbreviation please see Fig. 2 legend

et al. 2006). Given this, we can assume that MMR-proficient species too can harbor high number of PSSRs as a consequence of yet to be discovered mechanisms some of which may be species specific.

MMR is Strand Biased

During replication, there is an equal probability for the nascent strand and the template strand to harbor slippage errors and therefore one can expect equal number of contraction and expansion events in PSSRs. Results shown in Table 2 reveal that PSSRs are contraction biased in MMRS-proficient strains (*Haemophilus influenzae*, *Acinetobacter baumannii*, and *Staphylococcus aureus*). On the other hand, MMRS-deficient strains do not show significant bias toward either expansion or contraction. These results suggest that MMRS is less efficient in repairing slippage mutation on the template strand as compared to the nascent strand. It can also be argued that the contraction bias observed in MMRS-proficient species might also be a

consequence of high frequency of primary mutations (slip out) in the template strand as compared to the nascent strand. To test which one of the aforementioned arguments is correct we examined contraction/expansion bias of PSSRs in the species lacking MMRS.

Primary SSR Mutations are Unbiased Toward Expansion/Contractions in MMR-Deficient Species

The SSR mutations observed in the absence of the MMRS are referred to as primary SSR mutations. Table 1 gives the list of species lacking the MMRS. The numbers of PSSRs in all the six species (*Campylobacter jejuni*, *Helicobacter pylori*, *Mycoplasma hypopneumoniae*, *Corynebacterium glutamicum*, *Methanococcus maripalusis*, and *Sulpholobus islandicus*) along with the number of times they show contractions and expansions are given in Table 3. It can be seen from the table that most of the MMRS-deficient species do not show significant differences between the numbers of expanded and contracted PSSRs. However, it

Table 2 PSSRs found in MMR-proficient and MMR-deficient strains of *Acinetobacter baumannii*, *Haemophilus influenza*, and *Staphylococcus aureus*

Species name	Name of strain(s)	#PSSR	Cont	Exp	Cont/Exp	χ^2 1:1	P value
Acinetobacter baumannii	**AB0057, AB307_0294, ACICU and SDF**	**312**	**180**	**132**	**1.36**	**7.38**	**0.006**
	ATCC_17978	139	76	63	1.20	1.22	NS
Haemophilus influenzae	**RD_KW20, 028NP, F3031, F3047 and PittEE**	**319**	**180**	**139**	**1.29**	**5.27**	**0.002**
	PittGG	116	59	57	1.0	0.02	NS
Staphylococcus aureus	**MRSA252, MSSA476, MW2, JH1, Mu3, Mu50, N315, COL, Newmann, NCTCT, USA300**	**312**	**183**	**129**	**1.41**	**9.35**	**0.002**
	RF122	180	89	91	1.0	0.02	NS

NS not significant

Number of instances of PSSR expansions and contractions are also given. The statistically significant ones are shown in bold

Table 3 Number of PSSRs found in MMR-deficient species

Name of species	#Strain	GC%	Size (106)	#PSSR (poly G|C)	Cont	Exp	Cont/Exp	χ^2 1:1	*P* value
Campylobactor jejuni	5	31	1.7	153 (1)	67	86	0.78	2.36	NS
Helicobacter pylori	6	39	1.6	189 (28)	96	93	1.01	0.05	NS
Methanococcus maripaludi[a]	4	33	1.7	174 (8)	84	90	0.93	0.21	NS
Mycoplasma hypopneumoniae	3	29	0.9	126 (2)	43	83	0.51	12.70	0.0003
Corynebacterium glutamicum	3	54	3.3	107 (36)	54	53	1.0	0.02	NS
Sulpholobus islandicus[a]	6	35	2.7	141 (21)	61	80	0.76	2.56	NS

Cont contraction, *Exp* expansion

Table also show the number of expansion and contraction events in each species. Number of poly G|C tract is given in parentheses

[a] Archaea

can be seen in the case of *Mycoplasma hypopneumoniae* SSRs are significantly biased toward expansion. The expansion bias of SSRs in this MMRS-deficient species could be a consequence of selection for long SSR tracts. We would like to state that as compared to the other

bacterial species *Mycoplasma hypopneumoniae* SSR tracts are longer. The most of the long polymorphic tracts in *Mycoplasma hypopneumoniae* show more than two polymorphic states and hence we could not unequivocally infer contraction/expansion events of SSRs (please see the

Table 4 Number of PSSRs found in MMR-proficient species

Name of species	#Strain	@GC%	@ Genome size in bps (10^6)	#PSSR (poly G|C)	Cont	Exp	Cont/Exp	χ^2 1:1	*P* value
Actinobacillus pleuropneumoniae	3	41	2.2	94 (9)	53	41	1.30	1.53	0.21
Bacillus cereus	8	35	5.1	962 (127)	526	436	1.20	8.42	0.0037
Clostridium botulinum	7	28	3.8	349 (35)	201	148	1.36	8.05	0.0046
Clostridium perfringens	3	28	3.0	275 (11)	156	119	1.31	4.98	0.025
Chlamydia trachomatis	6	41	1.04	87 (15)	50	37	1.35	1.94	0.16
Escherichia coli	21	51	5.0	542 (129)	324	218	1.48	20.73	0.0001
Francisella tularensis	7	32	1.9	141 (6)	89	52	1.71	9.71	0.0018
Lactococcus lactis	3	36	2.4	95 (4)	53	42	1.26	1.27	0.26
Listeria monocytogenes	4	38	2.9	216 (7)	122	94	1.29	3.63	0.05
Legionella pneumophila	4	38	3.5	194 (24)	119	75	1.58	9.98	0.0016
Neisseria meningitides	4	52	2.2	99 (37)	55	44	1.25	1.22	0.27
Pseudomonas aeruginosa	4	66	6.5	394 (241)	250	144	1.73	28.52	0.0001
Pseudomonas putida	4	62	6.0	131 (54)	84	47	1.78	10.45	0.0012
Pseudomonas syringae	4	66	6.5	98 (28)	69	29	2.37	16.33	0.0001
Streptococcus agalactiae	3	36	2.1	74 (6)	41	33	1.24	0.86	0.35
Shewanella baltica	4	46	5.2	231 (45)	151	80	1.9	21.82	0.0001
Salmonella enteric	13	52	4.7	463 (108)	269	194	1.39	12.15	0.0005
Shigella flexneri	3	51	4.6	51 (22)	25	26	1.00	0.02	0.88
Streptococcus pneumonia	6	40	2.1	106 (8)	62	44	1.40	3.06	0.08
Streptococcus pyogenes	13	39	1.9	397 (40)	214	182	1.18	2.59	0.10
Salmonella typhi	3	52	4.8	113 (24)	48	65	0.73	2.56	0.10
Streptococcus thermophilus	3	39	1.8	109 (12)	60	49	1.22	1.11	0.29
Xanthomonas campestris	4	65	5.1	74 (50)	55	19	2.89	17.51	0.0001
Xanthomonas oryzae	3	64	5.0	57 (65)	25	32	0.78	0.86	0.35
Xylella fastidiosa	4	52	2.6	92 (37)	49	43	1.13	0.39	0.53
Yersinia pestis	20	48	4.6	282 (68)	164	118	1.40	7.50	0.006

Cont contraction, *Exp* expansion, @ average

In most of the species mutation in SSRs is biased toward contraction. Number of poly G|C tract is given in parentheses

"Materials and Methods") and these long tracts were not analyzed in our studies.

PSSRs are Contraction Biased in MMRS-Proficient Species

Mutations in SSRs which escape MMRS are referred to as secondary mutations. We examined the directionality (contraction or expansion) of these secondary mutations (PSSRs) in MMRS-proficient species. The numbers of PSSRs found in the MMR-proficient species are given in Table 4. It can be seen that in most of the species PSSRs are significantly contraction biased.

PSSR Contraction Bias is Independent of Repeat Count as well as Repeat Types

In several studies, it has been reported that the SSRs with high repeat counts experience a downward mutation bias whereas those with low repeat counts are prone for expansions (Huang et al. 2002; Xu et al. 2000). In the studies reported elsewhere the length of SSR tracts were very large compared to the observed length of polymorphic SSRs in this study. Most of the PSSRs observed in our study are of 3- to 8-bp long and hence on this basis we have considered 3–5 bp in one group of PSSRs and more than 5 bp as the other group of PSSRs. We have also examined the data over one repeat unit shift (Fig. 5; Supplementary Table III). In most of the species, the PSSRs were observed with bias toward contraction suggesting that generally the SSR mutations are biased toward contraction irrespective

of their repeat counts suggesting that the mutational bias of SSRs is related to MMRS rather than their repeat counts.

To check the difference in mutation pattern in poly A|T and poly G|C tract we have analyzed the range of genomes with low GC content to high GC content. It should be noted that the A/T-rich genomes harbor higher number of polymorphic poly A|T tracts as compared to the number of polymorphic poly G|C tracts. Similarly in GC rich genomes it is expected to see higher number poly G|C PSSR tracts as compared to the number of poly A|T PSSR tracts. To examine the mutational bias, if any, between poly G|C and poly A|T tracts we have considered seven species (Table 5) of which three are AT-rich genomes, one is a neutral genome (AT and GC content 50% each) and the remaining three are GC-rich genomes. It can be seen from Table 5 that there is no difference between poly A|T and poly G|C PSSR tracts with regard to their mutational bias.

Directionality of SSR Mutation is Independent of Physical Properties of Genomes

Table 6 gives a comparison between the directionality of SSRs mutation between the MMRS-proficient and MMRS-deficient species having similar GC content of the genome. It can be seen that the genomes having similar genome compositions have different directionality of SSR mutation and hence it can be argued that directionality in SSR mutations between the MMRS-proficient and MMRS-deficient species is independent of the composition of genomes but dependent on the presence or absence of the MMRS.
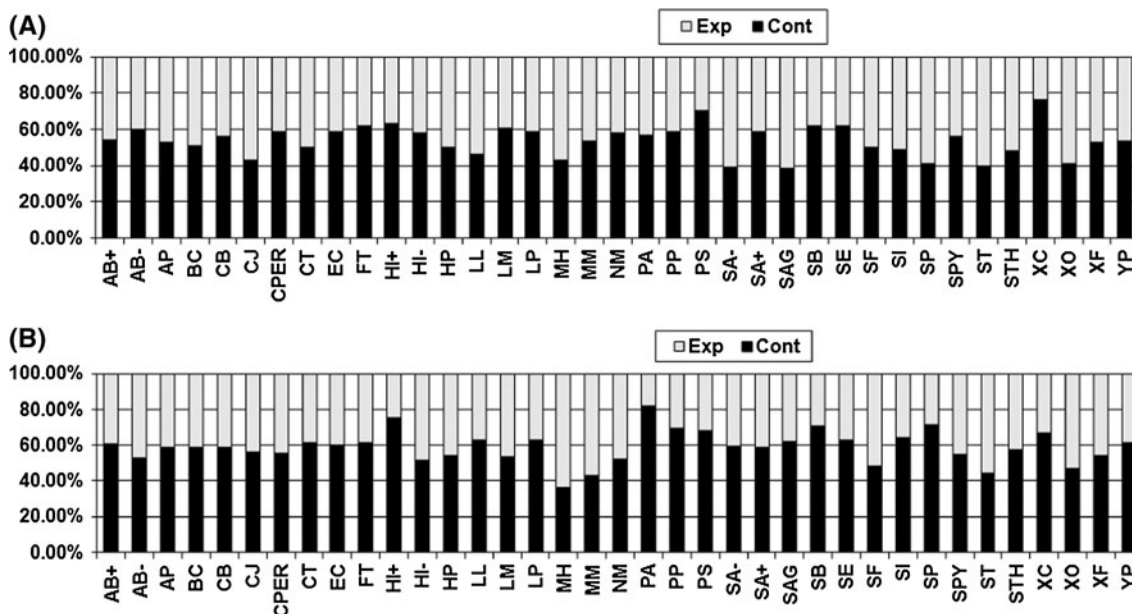

Fig. 5 Mutation bias is independent of repeat count. Illustration of contraction bias of SSR mutations is not related to repeat count. Count ≤5 bp (**a**) and count ≥6 bp (**b**). For the *x* axis legend abbreviation please see Fig. 2 legend

Table 5 Contraction bias observed irrespective of repeat types (poly A|T or G|C)

Species name	GC%	A|T		G|C	
		Contraction	Expansion	Contraction	Expansion
Francisella tularensis	32	85	50	4	2
Listeria monocytogenes	38	121	88	1	6
Legionella pneumophila	38	102	68	17	7
Escherichia coli	51	242	171	82	47
Pseudomonas aeruginosa	66	89	64	161	80
Pseudomonas putida	62	49	28	35	19
Xanthomonas campestris	65	17	7	38	12

Table 6 Comparison between MMR-proficient and MMR-deficient species with similar GC percentage

Species name	GC%	Cont/Exp	χ^2 1:1
HP	39	1.01	0.05
CJ	36	0.78	2.36
MM	33	0.93	0.21
MH	29	0.53	7.20
SI	35	0.76	2.56
LM	38	1.31	3.84
SAG	36	1.6	16.06
SA	33	1.41	9.35
CPER	28	1.31	4.98
BC	35	1.20	8.42

Data related to MMR-proficient species are indicated in bold

Conclusions

In this study, we have investigated the mutational bias of SSRs toward expansion or contraction in relation to the presence and absence of MMRS in prokaryotes. Our investigations revealed that the MMR deficiency as a consequence of frameshift mutation in MMR gene leads to an increase in the number of SSR tracts undergoing slippage-related INDEL mutations. The strain-specific INDEL mutations in MMRS in *Acinetobacter baumannii*, *Haemophilus influenza*, and *Staphylococcus aureus* indicate that these mutations are not inherited during speciation. The strain-specific mutations have arisen during different points of time during evolution. If the mutation event is relatively new then all the characteristics of a mutation may not be very evident. The number of SSR mutations is not very high in case of *Acinetobacter baumannii* ATCC_17978 as compared to the other strains of *Acinetobacter baumannii* and this could be due to the recently happened mutation in MMRS or sequencing errors. The indication of this is also evident in the mutation bias of the SSRs in this strain [Table 2 (contraction/expansion ratio is 1.2)]. In the case of the other two mutated strains Viz., *Haemophilus influenza* PittGG and *Staphylococcus aureus*

RF122 SSR mutations are not biased toward either of contraction or expansion (contraction/expansion ratio is 1).

Most of the MMR-deficient species show high number of mutated SSR tracts. However, despite the MMR deficiency mycobacterium showed high stability in non-coding SSR tracts which could be due to increased fidelity of the DNA polymerases even in the MMR-deficient background due to high generation time. The variations in SSR stabilities between the MMR-deficient species are also due to the fact the differences in the living environment between the species. The species where MMRS is present SSR mutations show significant bias toward deletions. However, we could not find an unequivocal bias of SSRs either toward expansion or contraction in species deficient in MMRS. The directionality of evolution of SSRs seemed to be independent of their repeat counts [long (6–8 bp) vs. short (3–5 bp) SSR] as well as the repeat types (A|T vs. G|C). Directionality of SSR mutation is also independent of the physical properties (GC content) of genome. Contraction bias of SSRs explains the enrichment of short tracts in prokaryotic genomes which are mostly MMRS proficient.

References

Aaltonen LA, Peltomaki P, Leach FS, Sistonen P, Pylkkanen L, Mecklin JP, Jarvinen H, Powell SM, Jen J, Hamilton SR et al (1993) Clues to the pathogenesis of familial colorectal cancer. Science 260:812–816

Acharya S, Wilson T, Gradia S, Kane MF, Guerrette S, Marsischky GT, Kolodner R, Fishel R (1996) hMSH2 forms specific mispair-binding complexes with hMSH3 and hMSH6. Proc Natl Acad Sci USA 93:13629–13634

Acharya S, Foster PL, Brooks P, Fishel R (2003) The coordinated functions of the *E. coli* MutS and MutL proteins in mismatch repair. Mol Cell 12:233–246

Amos W, Rubinsztein DC (1996) Microsatellites are subject to directional evolution. Nat Genet 12:13–14

Amos W, Sawcer SJ, Feakes RW, Rubinsztein DC (1996) Microsatellites show mutational bias and heterozygote instability. Nat Genet 13:390–391

Ashley CT Jr, Warren ST (1995) Trinucleotide repeat expansion and human disease. Annu Rev Genet 29:703–728

Bayliss CD, van de Ven T, Moxon ER (2002) Mutations in polI but not mutSLH destabilize *Haemophilus influenzae* tetranucleotide repeats. EMBO J 21:1465–1476

Boyer JC, Yamada NA, Roques CN, Hatch SB, Riess K, Farber RA (2002) Sequence dependent instability of mononucleotide microsatellites in cultured mismatch repair proficient and deficient mammalian cells. Hum Mol Genet 11:707–713

Carvalho FM, Fonseca MM, Batistuzzo De Medeiros S, Scortecci KC, Blaha CA, Agnez-Lima LF (2005) DNA repair in reduced genome: the mycoplasma model. Gene 360:111–119

Claverys JP, Lacks SA (1986) Heteroduplex deoxyribonucleic acid base mismatch repair in bacteria. Microbiol Rev 50:133–165

Coenye T, Vandamme P (2005) Characterization of mononucleotide repeats in sequenced prokaryotic genomes. DNA Res 12:221–233

Cox EC (1995) Recombination, mutation and the origin of species. BioEssays 17:747–749

De Bolle X, Bayliss CD, Field D, van de Ven T, Saunders NJ, Hood DW, Moxon ER (2000) The length of a tetranucleotide repeat tract in *Haemophilus influenzae* determines the phase variation rate of a gene with homology to type III DNA methyltransferases. Mol Microbiol 35:211–222

Dos Vultos T, Mestre O, Tonjum T, Gicquel B (2009) DNA repair in *Mycobacterium tuberculosis* revisited. FEMS Microbiol Rev 33:471–487

Eckert KA, Yan G (2000) Mutational analyses of dinucleotide and tetranucleotide microsatellites in *Escherichia coli*: influence of sequence on expansion mutagenesis. Nucleic Acids Res 28:2831–2838

Eisen JA (1998) A phylogenomic study of the MutS family of proteins. Nucleic Acids Res 26:4291–4300

Ellegren H (2002) Mismatch repair and mutational bias in microsatellite DNA. Trends Genet 18:552

Ellegren H, Primmer CR, Sheldon BC (1995) Microsatellite 'evolution': directionality or bias? Nat Genet 11:360–362

Ellegren H, Lindgren G, Primmer CR, Moller AP (1997) Fitness loss and germline mutations in barn swallows breeding in Chernobyl. Nature 389:593–596

Field D, Wills C (1998) Abundant microsatellite polymorphism in *Saccharomyces cerevisiae*, and the different distributions of microsatellites in eight prokaryotes and *S. cerevisiae*, result from strong mutation pressures and a variety of selective forces. Proc Natl Acad Sci USA 95:1647–1652

Finn RD, Tate J, Mistry J, Coggill PC, Sammut SJ, Hotz HR, Ceric G, Forslund K, Eddy SR, Sonnhammer EL, Bateman A (2008) The Pfam protein families database. Nucleic Acids Res 36:D281–D288

Garcia-Diaz M, Kunkel TA (2006) Mechanism of a genetic glissando: structural biology of indel mutations. Trends Biochem Sci 31:206–214

Harr B, Todorova J, Schlotterer C (2002) Mismatch repair-driven mutational bias in *D. melanogaster*. Mol Cell 10:199–205

Henderson ST, Petes TD (1992) Instability of simple sequence in *Saccharomyces cerevisiae*. Mol Cell Biol 12:2749–2757

Henz SR, Huson DH, Auch AF, Nieselt-Struwe K, Schuster SC (2005) Whole-genome prokaryotic phylogeny. Bioinformatics 21:2329–2335

Himmelreich R, Hilbert H, Plagens H, Pirkl E, Li BC, Herrmann R (1996) Complete sequence analysis of the genome of the bacterium *Mycoplasma pneumoniae*. Nucleic Acids Res 24:4420–4449

Hiriyanna KT, Ramakrishnan T (1986) Deoxyribonucleic acid replication time in *Mycobacterium tuberculosis* H37 Rv. Arch Microbiol 144:105–109

Hsieh P (2001) Molecular mechanisms of DNA mismatch repair. Mutat Res 486:71–87

Huang QY, Xu FH, Shen H, Deng HY, Liu YJ, Liu YZ, Li JL, Recker RR, Deng HW (2002) Mutation patterns at dinucleotide microsatellite loci in humans. Am J Hum Genet 70:625–634

Ionov Y, Peinado MA, Malkhosyan S, Shibata D, Perucho M (1993) Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for colonic carcinogenesis. Nature 363:558–561

Iyer RR, Pluciennik A, Burdett V, Modrich PL (2006) DNA mismatch repair: functions and mechanisms. Chem Rev 106:302–323

Jaworski A, Rosche WA, Gellibolian R, Kang S, Shimizu M, Bowater RP, Sinden RR, Wells RD (1995) Mismatch repair in *Escherichia coli* enhances instability of (CTG)*n* triplet repeats from human hereditary diseases. Proc Natl Acad Sci USA 92:11019–11023

Kang J, Huang S, Blaser MJ (2005) Structural and functional divergence of MutS2 from bacterial MutS1 and eukaryotic MSH4–MSH5 homologs. J Bacteriol 187:3528–3537

Kumar P, Chaitanya PS, Nagarajaram HA (2011) PSSRdb: a relational database of polymorphic simple sequence repeats extracted from prokaryotic genomes. Nucleic Acids Res 39:D601–D605

Kunkel TA (2004) DNA replication fidelity. J Biol Chem 279:16895–16898

Kunkel TA, Loeb LA (1980) On the fidelity of DNA replication. The accuracy of *Escherichia coli* DNA polymerase I in copying natural DNA in vitro. J Biol Chem 255:9961–9966

Kunkel TA, Bebenek K (2000) DNA replication fidelity. Annu Rev Biochem 69:497–529

Lamers MH, Perrakis A, Enzlin JH, Winterwerp HH, de Wind N, Sixma TK (2000) The crystal structure of DNA mismatch repair protein MutS binding to a G × T mismatch. Nature 407:711–717

Lee JY, Chang J, Joseph N, Ghirlando R, Rao DN, Yang W (2005) MutH complexed with hemi- and unmethylated DNAs: coupling base recognition and DNA cleavage. Mol Cell 20:155–166

Levinson G, Gutman GA (1987) Slipped-strand mispairing: a major mechanism for DNA sequence evolution. Mol Biol Evol 4:203–221

Levy DD, Cebula TA (2001) Fidelity of replication of repetitive DNA in mutS and repair proficient *Escherichia coli*. Mutat Res 474:1–14

Li GM (2008) Mechanisms and functions of DNA mismatch repair. Cell Res 18:85–98

Lin Z, Nei M, Ma H (2007) The origins and early evolution of DNA mismatch repair genes—multiple horizontal gene transfers and co-evolution. Nucleic Acids Res 35:7591–7603

Marra G, Boland CR (1995) Hereditary nonpolyposis colorectal cancer: the syndrome, the genes, and historical perspectives. J Natl Cancer Inst 87:1114–1125

Martin P, Sun L, Hood DW, Moxon ER (2004) Involvement of genes of genome maintenance in the regulation of phase variation frequencies in *Neisseria meningitidis*. Microbiology 150:3001–3012

Metzgar D, Liu L, Hansen C, Dybvig K, Wills C (2002) Domain-level differences in microsatellite distribution and content result from different relative rates of insertion and deletion mutations. Genome Res 12:408–413

Mirkin SM (2005) Toward a unified theory for repeat expansions. Nat Struct Mol Biol 12:635–637

Mizrahi V, Andersen SJ (1998) DNA repair in *Mycobacterium tuberculosis*. What have we learnt from the genome sequence? Mol Microbiol 29:1331–1339

Modrich P, Lahue R (1996) Mismatch repair in replication fidelity, genetic recombination, and cancer biology. Annu Rev Biochem 65:101–133

Morel P, Reverdy C, Michel B, Ehrlich SD, Cassuto E (1998) The role of SOS and flap processing in microsatellite instability in *Escherichia coli*. Proc Natl Acad Sci USA 95:10003–10008

Mrazek J, Guo X, Shah A (2007) Simple sequence repeats in prokaryotic genomes. Proc Natl Acad Sci USA 104:8472–8477

Obmolova G, Ban C, Hsieh P, Yang W (2000) Crystal structures of mismatch repair protein MutS and its complex with a substrate DNA. Nature 407:703–710

Parker BO, Marinus MG (1992) Repair of DNA heteroduplexes containing small heterologous sequences in *Escherichia coli*. Proc Natl Acad Sci USA 89:1730–1734

Primmer CR, Saino N, Moller AP, Ellegren H (1996) Directional evolution in germline microsatellite mutations. Nat Genet 13:391–393

Radman M (1998) DNA replication: one strand may be more equal. Proc Natl Acad Sci USA 95:9718–9719

Rubinsztein DC, Amos W, Leggo J, Goodburn S, Jain S, Li SH, Margolis RL, Ross CA, Ferguson-Smith MA (1995a) Microsatellite evolution—evidence for directionality and variation in rate between species. Nat Genet 10:337–343

Rubinsztein DC, Leggo J, Amos W (1995b) Microsatellites evolve more rapidly in humans than in chimpanzees. Genomics 30:610–612

Schlotterer C (2000) Evolutionary dynamics of microsatellite DNA. Chromosoma 109:365–371

Schlotterer C, Tautz D (1992) Slippage synthesis of simple sequence DNA. Nucleic Acids Res 20:211–215

Schlotterer C, Imhof M, Wang H, Nolte V, Harr B (2006) Low abundance of *Escherichia coli* microsatellites is associated with an extremely low mutation rate. J Evol Biol 19:1671–1676

Schofield MJ, Hsieh P (2003) DNA mismatch repair: molecular mechanisms and biological function. Annu Rev Microbiol 57:579–608

Schultz J, Milpetz F, Bork P, Ponting CP (1998) SMART, a simple modular architecture research tool: identification of signaling domains. Proc Natl Acad Sci USA 95:5857–5864

Selmane T, Schofield MJ, Nayak S, Du C, Hsieh P (2003) Formation of a DNA mismatch repair complex mediated by ATP. J Mol Biol 334:949–965

Sixma TK (2001) DNA mismatch repair: MutS structures bound to mismatches. Curr Opin Struct Biol 11:47–52

Smith BT, Grossman AD, Walker GC (2001) Visualization of mismatch repair in bacterial cells. Mol Cell 8:1197–1206

Smith MG, Gianoulis TA, Pukatzki S, Mekalanos JJ, Ornston LN, Gerstein M, Snyder M (2007) New insights into *Acinetobacter baumannii* pathogenesis revealed by high-density pyrosequencing and transposon mutagenesis. Genes Dev 21:601–614

Sreenu VB, Ranjitkumar G, Swaminathan S, Priya S, Bose B, Pavan MN, Thanu G, Nagaraju J, Nagarajaram HA (2003) MICAS: a fully automated web server for microsatellite extraction and analysis from prokaryote and viral genomic sequences. Appl Bioinform 2:165–168

Sreenu VB, Kumar P, Nagaraju J, Nagarajaram HA (2006) Microsatellite polymorphism across the *M. tuberculosis* and *M. bovis* genomes: implications on genome evolution and plasticity. BMC Genomics 7:78

Sreenu VB, Kumar P, Nagaraju J, Nagarajam HA (2007) Simple sequence repeats in mycobacterial genomes. J Biosci 32:3–15

Strand M, Prolla TA, Liskay RM, Petes TD (1993) Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. Nature 365:274–276

Streisinger G, Okada Y, Emrich J, Newton J, Tsugita A, Terzaghi E, Inouye M (1966) Frameshift mutations and the genetic code. This paper is dedicated to Professor Theodosius Dobzhansky on the occasion of his 66th birthday. Cold Spring Harb Symp Quant Biol 31:77–84

Thibodeau SN, Bren G, Schaid D (1993) Microsatellite instability in cancer of the proximal colon. Science 260:816–819

Twerdi CD, Boyer JC, Farber RA (1999) Relative rates of insertion and deletion mutations in a microsatellite sequence in cultured cells. Proc Natl Acad Sci USA 96:2875–2879

Umar A, Kunkel TA (1996) DNA-replication fidelity, mismatch repair and genome instability in cancer cells. Eur J Biochem 238:297–307

van Belkum A, Scherer S, van Alphen L, Verbrugh H (1998) Short-sequence DNA repeats in prokaryotic genomes. Microbiol Mol Biol Rev 62:275–293

Vogler AJ, Keys C, Nemoto Y, Colman RE, Jay Z, Keim P (2006) Effect of repeat copy number on variable-number tandem repeat mutations in *Escherichia coli* O157:H7. J Bacteriol 188:4253–4263

Watson ME Jr, Burns JL, Smith AL (2004) Hypermutable *Haemophilus influenzae* with mutations in mutS are found in cystic fibrosis sputum. Microbiology 150:2947–2958

Webster MT, Smith NG, Ellegren H (2002) Microsatellite evolution inferred from human–chimpanzee genomic sequence alignments. Proc Natl Acad Sci USA 99:8748–8753

Xu X, Peng M, Fang Z (2000) The direction of microsatellite mutations is dependent upon allele length. Nat Genet 24:396–399

Yamada NA, Smith GA, Castro A, Roques CN, Boyer JC, Farber RA (2002) Relative rates of insertion and deletion mutations in dinucleotide repeats of various lengths in mismatch repair proficient mouse and mismatch repair deficient human cells. Mutat Res 499:213–225

Yang W (2000) Structure and function of mismatch repair proteins. Mutat Res 460:245–256