

21. Navarro, G. and Raffinot, M., *Lecture Notes in Computer Science*, Springer-Verlag, Berlin, 1998, vol. 1448, pp. 14–31.
22. Karp, R. M. and Rabin, M. O., *IBM J. Res. Dev.*, 1987, **31**, 249–260.
23. Baeza-Yates, R. A. and Gonnet, G. H., *Commun. ACM*, 1992, **35**, 74–82.
24. Hancart, C., *Actes des 2<sup>e</sup> Journées Franco-Belges* (ed. Krob, D.), 1991, PUR 176, Rouen, France, 1992, pp. 99–110.
25. Raita, T., *Software – Practice Experience*, 1992, **22**, 879–884.
26. Galil, Z. and Seiferas, J., *J. Comput. Syst. Sci.*, 1983, **26**, 280–294.
27. Crochemore, M. and Perrin, D., *J. ACM*, 1991, **38**, 651–675.
28. Sheik, S. S., Aggarwal, S. K., Poddar, A., Balakrishnan, N. and Sekar, K., *J. Chem. Inf. Comp. Sci.*, 2004, **44**, 1251–1256.
29. Bairoch, A. and Apweiler, R., *Nucleic Acids Res.*, 1998, **26**, 38–42.
30. Barker, W. C. *et al.*, *Nucleic Acids Res.*, 1998, **28**, 27–32.
31. Bernstein, F. C. *et al.*, *J. Mol. Biol.*, 1977, **112**, 535–542.
32. Charras, C. and Lecroq, T., *Handbook of Exact String Matching Algorithms* (available in the website: <http://www-igm.univ-mlv.fr/~lecroq/string/>).

ACKNOWLEDGEMENTS. We thank the Bioinformatics Centre (DIC), the Interactive Graphics Based Molecular Modelling facility (IGBMM) and the Supercomputer Education and Research Centre, IISc, Bangalore for support. This work is completely supported by the Institute-wide Computational Genomics Project supported by the Department of Biotechnology (DBT), New Delhi. DIC and IGBMM are supported by DBT.

Received 20 July 2004; revised accepted 1 February 2005

## Compositional symmetry of DNA duplex in bacterial genomes

Sujit Kumar Verma<sup>1,\*</sup>, Debojyoti Das<sup>1</sup>,  
Siddharatha Sankar Satapathy<sup>2</sup>,  
Alak Kumar Buragohain<sup>1</sup> and  
Suvendra Kumar Ray<sup>1</sup>

<sup>1</sup>Department of Molecular Biology and Biotechnology, and

<sup>2</sup>Department of Computer Science and Information Technology, Tezpur University, Tezpur 784 028, India

**Computational analysis of seven bacterial genomes revealed that both the DNA strands in a genome exhibit compositional symmetry in terms of the abundance of a non-palindromic oligonucleotide (di, tri, tetra, penta, hexa, hepta and octa). This symmetry in DNA duplex suggests that both strands in the duplex possess similar compositional characteristics, though the nucleotide sequences in the DNA strands are different (complementary). This compositional symmetry between DNA strands in genomes may be due to the abundance of**

**coding sequences in both the strands and to ensure the synchronous completion of replication of the two strands.**

**Keywords:** Genome, DNA duplex, non-palindromic sequence, oligonucleotide composition, coding sequences, DNA replication.

A chromosome is made up of a DNA duplex in which the two DNA strands are antiparallel and complementary to each other: adenine (A) of one strand pairs with thymine (T) of the other strand, and guanine (G) of one strand pairs with cytosine (C) of the other strand<sup>1</sup>. Thus sequences of both the strands are different, except at the palindromic regions (symmetrical DNA sequence). In the case of a palindromic oligonucleotide, its abundance in both the DNA strands in a genome is identical. However, in the case of a non-palindromic oligonucleotide (non-symmetrical DNA sequence), its abundance in both the DNA strands in a genome might be different. In this study we have analysed seven bacterial genomes (*Bacillus subtilis*<sup>2</sup>, *Escherichia coli*<sup>3</sup>, *Haemophilus influenzae*<sup>4</sup>, *Pseudomonas aeruginosa*<sup>5</sup>, *Pseudomonas syringae*<sup>6</sup>, *Ralstonia solanacearum*<sup>7</sup> (mega plasmid and chromosome) and *Xanthomonas campestris* pv. *campestris*<sup>8</sup> (*Xcc*) for abundance of non-palindromic oligonucleotides in both the DNA strands. We present evidence that abundance of a non-palindromic oligonucleotide in both the DNA strands in a genome is similar. This compositional symmetry of the DNA duplex in genomes is interesting and suggests that there is a tendency in the genome to maintain similarity between both the DNA strands, though functionally the two strands have different attributes.

Complete genome sequences of *B. subtilis*, *E. coli*, *H. influenzae*, *P. aeruginosa*, *P. syringae*, *R. solanacearum* and *Xcc* were downloaded from the ‘genome information broker site’ ([www.gib.genes.nig.ac.jp](http://www.gib.genes.nig.ac.jp)). These sequences were analysed for the occurrence of non-palindromic oligonucleotide sequences (di, tri, tetra, penta, hexa, hepta and octa nucleotides; Table 1) in both DNA strands using a computer program ‘seqsearch’ (developed by the authors). Since DNA strands are complementary to each other, by studying nucleotide composition of one of the strands, composition of the other strand can be determined, e.g. the number of As present in one of the strands is equal to the number of Ts present in the other strand, and the number of Gs present in one strand is equal to the number of Cs present in the other strand. Similar logic can also be applied for studying oligonucleotide composition as well, e.g. total number of TG dinucleotides present in one of the strands of a genome is same as the total number of CA dinucleotides present in the other strand of the genome. Thus if we count the total number of TG and CA in one of the strands of the genome, then we would be able to compare between the number of TGs present in both the DNA strands. In this study, we have analysed one of the DNA strands in the genomes for comparing the abundance of nucleotides/oligonucleotides

\*For correspondence. (e-mail: [suven@tezu.ernet.in](mailto:suven@tezu.ernet.in))

**Table 1.** Composition of oligonucleotides in one of the strands of bacterial genomes

Sl. no.	Nucleotide repeats	<i>B. subtilis</i>	<i>Eco</i>	<i>Hin</i>	<i>Pae</i>	<i>Psy</i>	<i>Rs(c)</i>	<i>Rs(p)</i>	Xcc
Dinucleotides									
1	AT*	343199	309819	166837	204598	301703	139299	80664	202312
2	TA*	218056	211961	131955	94404	147196	46670	27618	75791
3	GC*	253807	383931	95529	814925	676674	519714	291198	701738
4	CG*	206553	346670	72523	763123	556309	502005	278148	608389
5	<b>TG</b>	<b>280816</b>	<b>322239</b>	<b>119996</b>	<b>373698</b>	<b>473039</b>	<b>241795</b>	<b>136350</b>	<b>372038</b>
6	<b>CA</b>	<b>280808</b>	<b>325149</b>	<b>121618</b>	<b>389849</b>	<b>470563</b>	<b>238524</b>	<b>136784</b>	<b>376048</b>
7	GT	195120	255608	91314	294581	347686	188156	103417	263761
8	AC	194108	256663	92410	306969	351045	181824	103621	264451
9	<b>AG</b>	<b>235331</b>	<b>237877</b>	<b>88457</b>	<b>353933</b>	<b>353389</b>	<b>181197</b>	<b>101512</b>	<b>254017</b>
10	<b>CT</b>	<b>237815</b>	<b>236061</b>	<b>88551</b>	<b>353792</b>	<b>358132</b>	<b>179006</b>	<b>100745</b>	<b>249846</b>
11	GA	273773	267247	94125	381247	388378	217126	121396	268939
12	TC	277261	267288	94745	384807	387285	217997	120858	268093
13	<b>AA</b>	<b>415077</b>	<b>337870</b>	<b>219880</b>	<b>190634</b>	<b>325136</b>	<b>106295</b>	<b>61665</b>	<b>168484</b>
14	<b>TT</b>	<b>416667</b>	<b>339482</b>	<b>217512</b>	<b>185979</b>	<b>322482</b>	<b>109965</b>	<b>61692</b>	<b>167994</b>
15	CC	193951	271673	68014	595923	454529	318903	184318	416580
16	GG	192288	270137	66448	515878	453580	327936	184522	417685
Trinucleotides									
1	<b>TTT</b>	<b>161011</b>	<b>109831</b>	<b>82070</b>	<b>27694</b>	<b>84064</b>	<b>18492</b>	<b>10823</b>	<b>28810</b>
2	<b>AAA</b>	<b>159595</b>	<b>108924</b>	<b>82918</b>	<b>28948</b>	<b>85103</b>	<b>16997</b>	<b>10467</b>	<b>29035</b>
3	GGG	33529	47495	10828	101892	83998	59304	34623	68919
4	CCC	34321	47775	11437	108533	84262	56266	34767	68652
5	<b>TTA</b>	<b>70144</b>	<b>68828</b>	<b>51205</b>	<b>8017</b>	<b>30873</b>	<b>4316</b>	<b>2824</b>	<b>9666</b>
6	<b>TAA</b>	<b>70968</b>	<b>68838</b>	<b>51470</b>	<b>8039</b>	<b>30314</b>	<b>4308</b>	<b>2840</b>	<b>9897</b>
7	TCT	69991	55472	25333	49213	64440	28643	15559	37319
8	AGA	68722	56621	25721	49653	64409	28578	15799	37906
9	<b>TCG</b>	<b>50391</b>	<b>71739</b>	<b>18778</b>	<b>171769</b>	<b>138310</b>	<b>100517</b>	<b>54029</b>	<b>112939</b>
10	<b>CGA</b>	<b>49683</b>	<b>70938</b>	<b>18835</b>	<b>171446</b>	<b>139095</b>	<b>99706</b>	<b>54286</b>	<b>113135</b>
11	TGC	68203	95232	32368	134551	156763	92681	52974	151168
12	GCA	68102	96028	32560	137824	155838	92945	53513	152583
13	<b>TGG</b>	<b>49849</b>	<b>85141</b>	<b>27034</b>	<b>129229</b>	<b>127886</b>	<b>74483</b>	<b>41607</b>	<b>115332</b>
14	<b>CCA</b>	<b>50018</b>	<b>86436</b>	<b>27479</b>	<b>137888</b>	<b>125645</b>	<b>72860</b>	<b>41488</b>	<b>116321</b>
15	GGC	56701	92144	18503	262448	189124	159263	88319	196357
16	GCC	57342	92973	18820	268845	188502	155836	88081	195586
17	<b>GAC</b>	<b>44947</b>	<b>54737</b>	<b>11123</b>	<b>108034</b>	<b>104621</b>	<b>63934</b>	<b>35767</b>	<b>72479</b>
18	<b>GTC</b>	<b>45603</b>	<b>54221</b>	<b>11248</b>	<b>106680</b>	<b>103658</b>	<b>65525</b>	<b>35894</b>	<b>72738</b>
19	CAC	42103	66751	23538	89398	103597	64300	36504	105793
20	GTG	41827	66117	23008	82328	103091	67107	36591	104532
Tetranucleotides									
1	<b>TTTT</b>	<b>62058</b>	<b>35609</b>	<b>30236</b>	<b>6660</b>	<b>24522</b>	<b>4784</b>	<b>2703</b>	<b>6974</b>
2	<b>AAAA</b>	<b>61191</b>	<b>35134</b>	<b>30845</b>	<b>6858</b>	<b>25003</b>	<b>4287</b>	<b>2471</b>	<b>6929</b>
3	CCCC	6713	8855	2235	18417	13927	9821	6872	10597
4	GGGG	6408	8719	2077	16526	13802	10905	6736	10752
5	<b>TAAA</b>	<b>29836</b>	<b>22270</b>	<b>21123</b>	<b>1784</b>	<b>9016</b>	<b>1012</b>	<b>747</b>	<b>2130</b>
6	<b>TTTA</b>	<b>29633</b>	<b>22383</b>	<b>21310</b>	<b>1782</b>	<b>8969</b>	<b>1091</b>	<b>730</b>	<b>2144</b>
7	GGGC	8780	15303	3035	45604	36419	31450	17836	33315
8	GCCC	9026	15743	3200	48714	35960	29794	17757	33640
9	<b>TGGC</b>	<b>14883</b>	<b>31148</b>	<b>7775</b>	<b>53097</b>	<b>52055</b>	<b>31533</b>	<b>17432</b>	<b>48656</b>
10	<b>GCCA</b>	<b>15093</b>	<b>31796</b>	<b>7823</b>	<b>57035</b>	<b>51599</b>	<b>31058</b>	<b>17496</b>	<b>48982</b>
11	ATCG	18344	24354	6785	35443	37775	25587	14661	34854
12	CGAT	18146	24248	6828	35285	37534	25993	14508	35017
13	<b>GTAG</b>	<b>5607</b>	<b>9451</b>	<b>3523</b>	<b>17654</b>	<b>12113</b>	<b>7446</b>	<b>3731</b>	<b>10173</b>
14	<b>CTAC</b>	<b>5484</b>	<b>9122</b>	<b>3495</b>	<b>17261</b>	<b>12936</b>	<b>7010</b>	<b>4050</b>	<b>9933</b>
15	GCAC	10595	18150	6777	31632	35330	25576	14098	40769
16	GTGC	10320	17947	6700	29623	35287	26054	14190	40496
17	<b>GCCG</b>	<b>19017</b>	<b>28577</b>	<b>3235</b>	<b>114594</b>	<b>63009</b>	<b>73486</b>	<b>40787</b>	<b>83459</b>
18	<b>CGGC</b>	<b>19215</b>	<b>28919</b>	<b>3112</b>	<b>115297</b>	<b>63402</b>	<b>74054</b>	<b>40677</b>	<b>84077</b>
19	GGGT	6681	12232	2805	23662	22414	9719	5572	15719
20	ACCC	6819	12200	2971	24890	23102	9246	5572	15461

(contd...)

# RESEARCH COMMUNICATIONS

**Table 1.** (contd...)

Sl. no.	Nucleotide repeats	<i>B. subtilis</i>	<i>Eco</i>	<i>Hin</i>	<i>Pae</i>	<i>Psy</i>	<i>Rs(c)</i>	<i>Rs(p)</i>	Xcc
<b>Pentanucleotides</b>									
1	<b>TTTTT</b>	<b>22884</b>	<b>11653</b>	<b>10490</b>	<b>1904</b>	<b>7516</b>	<b>1513</b>	<b>846</b>	<b>2043</b>
2	<b>AAAAA</b>	<b>22389</b>	<b>11474</b>	<b>10727</b>	<b>2050</b>	<b>7737</b>	<b>1330</b>	<b>773</b>	<b>1933</b>
3	CCCCC	1226	1518	364	2883	2376	1757	1331	1524
4	GGGGG	1139	1456	396	2506	2401	2076	1329	1647
5	<b>TTAAT</b>	<b>6455</b>	<b>6369</b>	<b>5909</b>	<b>264</b>	<b>1203</b>	<b>159</b>	<b>137</b>	<b>296</b>
6	<b>ATTAA</b>	<b>6403</b>	<b>6360</b>	<b>5842</b>	<b>291</b>	<b>1226</b>	<b>156</b>	<b>145</b>	<b>298</b>
7	TGTTT	10507	6341	3830	1135	5155	1002	617	1789
8	AAACA	10397	6402	3821	1183	5030	870	585	1809
9	<b>TTCTT</b>	<b>11538</b>	<b>5672</b>	<b>4670</b>	<b>4219</b>	<b>5305</b>	<b>2471</b>	<b>1170</b>	<b>3191</b>
10	<b>AAGAA</b>	<b>11285</b>	<b>5641</b>	<b>4673</b>	<b>4197</b>	<b>5355</b>	<b>2368</b>	<b>1169</b>	<b>3254</b>
11	TTCAA	9928	5309	4776	2954	7714	1966	1128	3039
12	TTGAA	9937	5402	4691	3096	7723	1937	1112	3069
13	<b>GGACA</b>	<b>3471</b>	<b>2181</b>	<b>633</b>	<b>5800</b>	<b>5154</b>	<b>3067</b>	<b>1860</b>	<b>4128</b>
14	<b>TGTCC</b>	<b>3554</b>	<b>2229</b>	<b>590</b>	<b>5679</b>	<b>5165</b>	<b>3118</b>	<b>1816</b>	<b>4087</b>
15	GCATC	4667	7072	1659	11285	10870	9523	5303	13396
16	GATGC	4590	7222	1650	11046	10907	9597	5388	13485
17	<b>GCCCG</b>	<b>2844</b>	<b>5504</b>	<b>501</b>	<b>16185</b>	<b>12615</b>	<b>13042</b>	<b>7607</b>	<b>12604</b>
18	<b>CGGGC</b>	<b>2725</b>	<b>5358</b>	<b>458</b>	<b>15475</b>	<b>12603</b>	<b>13517</b>	<b>7745</b>	<b>12432</b>
19	GCGAT	4443	8873	2323	12052	10972	8250	4605	12856
20	ATCGC	4519	8852	2437	12078	10942	8059	4678	12551
<b>Hexanucleotides</b>									
1	<b>TTTTTT</b>	<b>7507</b>	<b>3213</b>	<b>3133</b>	<b>550</b>	<b>2070</b>	<b>488</b>	<b>274</b>	<b>584</b>
2	<b>AAAAAA</b>	<b>7234</b>	<b>3189</b>	<b>3213</b>	<b>616</b>	<b>2145</b>	<b>409</b>	<b>256</b>	<b>476</b>
3	CCCCCC	163	240	63	402	382	268	205	245
4	GGGGGG	135	218	78	304	458	333	259	283
5	<b>TTAAAA</b>	<b>3580</b>	<b>1727</b>	<b>2378</b>	<b>102</b>	<b>488</b>	<b>42</b>	<b>44</b>	<b>106</b>
6	<b>TTTTAA</b>	<b>3507</b>	<b>1779</b>	<b>2383</b>	<b>110</b>	<b>508</b>	<b>75</b>	<b>46</b>	<b>100</b>
7	CCGGGG	394	843	11	3082	1635	1445	914	1331
8	CCCCGG	396	804	22	3218	1628	1321	971	1365
9	<b>TGCATC</b>	<b>1087</b>	<b>1332</b>	<b>560</b>	<b>1245</b>	<b>1947</b>	<b>1051</b>	<b>622</b>	<b>2150</b>
10	<b>GATGCA</b>	<b>1093</b>	<b>1424</b>	<b>493</b>	<b>1272</b>	<b>1867</b>	<b>1133</b>	<b>617</b>	<b>2192</b>
11	GTGTCA	883	653	190	185	909	325	196	463
12	TGACAC	841	681	163	222	972	313	206	485
13	<b>CAAACG</b>	<b>1573</b>	<b>1574</b>	<b>508</b>	<b>358</b>	<b>1732</b>	<b>454</b>	<b>283</b>	<b>935</b>
14	<b>CGTTTG</b>	<b>1583</b>	<b>1539</b>	<b>514</b>	<b>354</b>	<b>1765</b>	<b>441</b>	<b>329</b>	<b>942</b>
15	CACACA	530	521	241	167	794	187	119	539
16	TGTGTG	473	532	237	128	746	210	113	455
17	<b>ACGACG</b>	<b>416</b>	<b>1203</b>	<b>282</b>	<b>2497</b>	<b>2011</b>	<b>1860</b>	<b>1153</b>	<b>1789</b>
18	<b>CGTCGT</b>	<b>446</b>	<b>1166</b>	<b>297</b>	<b>2414</b>	<b>1972</b>	<b>1798</b>	<b>1146</b>	<b>1788</b>
19	CGGTTA	768	1327	465	347	903	179	74	380
20	TAACCG	678	1376	436	396	880	145	96	361
<b>Heptanucleotides</b>									
1	<b>TTTTTTT</b>	<b>1872</b>	<b>702</b>	<b>685</b>	<b>138</b>	<b>507</b>	<b>143</b>	<b>89</b>	<b>183</b>
2	<b>AAAAAAA</b>	<b>1855</b>	<b>711</b>	<b>675</b>	<b>171</b>	<b>509</b>	<b>110</b>	<b>84</b>	<b>120</b>
3	CCCCCCC	14	48	14	61	75	48	35	68
4	GGGGGGG	9	35	17	32	111	72	61	73
5	<b>TTTAACC</b>	<b>246</b>	<b>435</b>	<b>308</b>	<b>21</b>	<b>99</b>	<b>9</b>	<b>8</b>	<b>16</b>
6	<b>GGTAAA</b>	<b>283</b>	<b>460</b>	<b>312</b>	<b>18</b>	<b>90</b>	<b>13</b>	<b>10</b>	<b>26</b>
7	CCGGTTA	188	318	10	74	243	44	27	112
8	TAACCGG	162	317	6	88	255	44	28	102
9	<b>TGCAAGC</b>	<b>380</b>	<b>190</b>	<b>161</b>	<b>285</b>	<b>618</b>	<b>202</b>	<b>128</b>	<b>367</b>
10	<b>GCTTGCA</b>	<b>327</b>	<b>203</b>	<b>185</b>	<b>284</b>	<b>562</b>	<b>217</b>	<b>148</b>	<b>419</b>
11	GGGTTTA	170	214	110	27	160	17	3	31
12	TAAACCC	183	193	108	18	146	17	10	41
13	<b>AGGCCCG</b>	<b>122</b>	<b>155</b>	<b>16</b>	<b>778</b>	<b>667</b>	<b>504</b>	<b>302</b>	<b>402</b>
14	<b>CGGGCCT</b>	<b>131</b>	<b>147</b>	<b>9</b>	<b>788</b>	<b>672</b>	<b>524</b>	<b>287</b>	<b>389</b>
15	ATCGACG	126	465	92	1291	1051	934	501	912
16	CGTCGAT	140	512	26	1260	1073	959	494	899

(contd...)

**Table 1.** (contd...)

Sl. no.	Nucleotide repeats	<i>B. subtilis</i>	<i>Eco</i>	<i>Hin</i>	<i>Pae</i>	<i>Psy</i>	<i>Rs(c)</i>	<i>Rs(p)</i>	<i>Xcc</i>
17	<b>AGCTAAC</b>	<b>110</b>	<b>146</b>	<b>77</b>	<b>20</b>	<b>62</b>	<b>13</b>	<b>7</b>	<b>25</b>
18	<b>GTTAGCT</b>	<b>66</b>	<b>127</b>	<b>88</b>	<b>14</b>	<b>61</b>	<b>15</b>	<b>5</b>	<b>34</b>
19	GCTAGGT	36	12	40	38	51	18	13	38
20	ACCTAGC	33	15	22	27	43	13	11	35
Octanucleotides									
1	<b>TTTTTTT</b>	<b>243</b>	<b>119</b>	<b>92</b>	<b>31</b>	<b>86</b>	<b>46</b>	<b>30</b>	<b>53</b>
2	<b>AAAAAAA</b>	<b>228</b>	<b>123</b>	<b>85</b>	<b>39</b>	<b>73</b>	<b>29</b>	<b>24</b>	<b>31</b>
3	CCCCCCC	3	9	4	10	23	12	5	31
4	GGGGGGG	1	6	2	5	55	20	14	26
5	<b>GGTCCAA</b>	<b>53</b>	<b>16</b>	<b>14</b>	<b>27</b>	<b>34</b>	<b>13</b>	<b>7</b>	<b>25</b>
6	<b>TTGGAACC</b>	<b>43</b>	<b>15</b>	<b>7</b>	<b>21</b>	<b>44</b>	<b>19</b>	<b>7</b>	<b>34</b>
7	GTAAGCAA	73	84	48	7	45	19	10	13
8	TTGCTTAC	80	59	61	10	35	16	10	16
9	<b>GCTAGGAA</b>	<b>9</b>	<b>0</b>	<b>9</b>	<b>14</b>	<b>8</b>	<b>6</b>	<b>0</b>	<b>9</b>
10	<b>TTCCTAGC</b>	<b>22</b>	<b>3</b>	<b>7</b>	<b>4</b>	<b>9</b>	<b>3</b>	<b>4</b>	<b>9</b>
11	AACCGGTC	44	57	5	65	148	50	35	74
12	GACCGGTT	37	71	4	95	156	49	32	79
13	<b>GGGAAATT</b>	<b>88</b>	<b>108</b>	<b>57</b>	<b>32</b>	<b>64</b>	<b>20</b>	<b>11</b>	<b>25</b>
14	<b>AATTTCCC</b>	<b>91</b>	<b>95</b>	<b>69</b>	<b>29</b>	<b>54</b>	<b>22</b>	<b>13</b>	<b>23</b>
15	AATTTGGG	48	33	55	1	16	4	1	15
16	CCCAAATT	42	26	64	4	15	4	1	10
17	<b>AAGCCGGC</b>	<b>144</b>	<b>16</b>	<b>7</b>	<b>504</b>	<b>248</b>	<b>363</b>	<b>152</b>	<b>437</b>
18	<b>GCCGGCTT</b>	<b>123</b>	<b>13</b>	<b>3</b>	<b>530</b>	<b>236</b>	<b>348</b>	<b>168</b>	<b>460</b>
19	TTGCTTAG	38	24	23	9	15	4	3	5
20	CTAAGCAA	20	22	19	2	14	3	3	8

*Eco*, *E. coli*; *Hin*, *H. influenzae*; *Pae*, *P. aeruginosa*; *Psy*, *P. syringae*; *Rs(c)*; *R. solanacearum* (chromosome); *Rs(p)*, *R. solanacearum* (megaplasmid); *Xcc*, *X. campestris* pv. *campestris*.

\*Palindromic sequences.

Complementary oligonucleotides are either shown in bold or regular letters.

in both the DNA strands. In Table 2, genome length of the seven bacteria and nucleotide composition in one of the DNA strands of all the genomes are presented. Compositional values of complementary nucleotides (A and T, G and C) are similar, whereas the same differs significantly between non-complementary nucleotides. This suggests that there is compositional symmetry between both DNA strands in genomes at the mononucleotide level.

Table 1 represents the compositional value of non-palindromic oligonucleotides in seven bacterial genomes. Observation of the composition of all sixteen possible dinucleotides in the seven bacterial genomes revealed that dinucleotides made up of G and/or C (GG, GC, CG and CC) are more abundant than dinucleotides made up of A and/or T (AA, AT, TA and TT) in *P. aeruginosa*, *P. syringae*, *R. solanacearum*, *Xcc*. This is most likely due to the GC-rich nature of the genomes of these species (Table 2). Similarly, in the case of *B. subtilis*, *E. coli* and *H. influenzae*, dinucleotides made up of A and/or T are more abundant than those made up of G and/or C, as the genomes are either relatively less GC-rich or AT-rich. However, the abundance of dinucleotides like AG, TG, TC and AC (made up of 50% A/T and 50% G/C) in a genome exhibits a significant difference among them. This indicates that the oligonucleotide occurrence frequency in the genome

is not always concordant with GC%. So the abundance of an oligonucleotide in a genome is not only dependent upon the genome GC%, but the impact of evolutionary selection pressure on the sequence may also have an influence on its abundance value. To find out whether the abundance of an oligonucleotide in a genome is concordant with the genome GC%, we calculated the relative abundance of an oligonucleotide (Figure 1). If the relative abundance value is 1, then oligonucleotide abundance is concordant with the GC% of the genome. Any deviation from the value 1, in either direction (positive or negative), is a reflection of the extent of evolutionary selection pressure on the abundance of the oligonucleotide in the genome. Out of the total 16 possible dinucleotides, four are palindromes (sl. no. 1–4 of dinucleotides; Table 1) and the rest 12 (sl. no. 5–16 of dinucleotides; Table 1) are non-palindromes. Table 1 shows that the number of TG dinucleotides present in one strand of the genome is similar to the total number of CA dinucleotides present in the same strand. As discussed above, the number of CAs present in one strand is same as the number of TGs present in the complementary strand. Therefore, TG dinucleotide composition is similar in both DNA strands of the genomes. Similar observations were made with respect to other non-palindromic dinucleotides (Table 1). The relative abundance of all sixteen dinucleotides in one of the strands

**Table 2.** Nucleotide composition in one of the DNA strands of genomes

Organism	Genome size <sup>a</sup> (nt)	A	T	C	G	Total	GC%	Others <sup>b</sup>
<i>B. subtilis</i>	4214630	1187714 (0.282) <sup>d</sup>	1192801 (0.283)	919127 (0.218)	914988 (0.217)	4214630	43.51	0
<i>E. coli</i>	4639675	1142228 (0.246)	1140970 (0.245)	1179554 (0.254)	1176923 (0.253)	4639675	50.78	0
<i>H. influenzae</i>	1830138	567623 (0.310)	564241 (0.308)	350723 (0.191)	347436 (0.189)	1830023	38.15	115 <sup>c</sup>
<i>R. solanacearum</i> (plasmid)	2094509	347463 (0.165)	346518 (0.165)	699995 (0.334)	700533 (0.334)	2094509	66.86	0
<i>R. solanacearum</i> (chromosome)	3716413	608615 (0.163)	616427 (0.165)	1238438 (0.333)	1252933 (0.337)	3716413	67.03	0
<i>P. aeruginosa</i>	6264403	1056134 (0.169)	1038950 (0.166)	2102687 (0.336)	2066632 (0.330)	6264403	66.55	0
<i>P. syringae</i>	6397126	1331273 (0.208)	1330003 (0.208)	1869533 (0.292)	1866317 (0.292)	6397126	58.39	0
<i>X. campestris</i> pv. <i>campestris</i>	5076188	889264 (0.175)	883916 (0.174)	1650865 (0.325)	1652127 (0.325)	5076172	65.06	16 (N)

<sup>a</sup>Total number of nucleotides present in one of the DNA strands in the genome.

<sup>b</sup>Nucleotide sequence presented in the genome sequence apart from A/T/C/G.

<sup>c</sup>For example, K/M/N/R/S/W/Y.

<sup>d</sup>Frequency of nucleotide in one of the DNA strands (nucleotide abundance divided by the genome size).

in the genomes is presented in Figure 1a. The relative abundance values were found to deviate from 1, suggesting that the values are not concordant with the genome GC%. However, for two complementary oligonucleotides, the deviation value from 1 is similar. In order to determine the abundance of non-palindromic oligonucleotides comprising tri, tetra, penta, hexa, hepta and octa nucleotides, we studied the abundance of twenty randomly chosen oligonucleotides in each of the cases. It is evident from Table 1 that the abundance of the complementary oligonucleotides in one of the DNA strands in a genome is similar. Thus both the DNA strands in a genome exhibit compositional symmetry of an oligonucleotide. The relative abundance of tri, tetra and penta oligonucleotides in the genomes is given in Figure 1b–d. It is evident that there is a selection pressure on the occurrence of an oligonucleotide in the genome, which is similar in both the strands of a DNA duplex. A similar study was carried out in one of the Archaea genomes (*Methanococcus maripaludis*), and one of the chromosomes of several eukaryotic genomes (*Arabidopsis thaliana*, *Drosophila melanogaster*, *Oryza sativa*, *Homo sapiens*, *Plasmodium falciparum* and *Saccharomyces cerevisiae* (Table 3). The compositional symmetry was also observed in the above genomes, suggesting that it is a common feature of all genomes.

This compositional symmetry of oligonucleotides in the DNA duplex has given rise to similar number of complementary oligonucleotides in a single DNA strand in the genome, similar number of a particular oligonucleotide in complementary DNA strands in the genome and similar number of direct and inverted repeats of a oligonucleotide base pair in the genome (Figure 2). Why this compositional

symmetry in genomes? It is evident from the genome sequences that coding sequences are present in similar proportions in both the DNA strands in a genome. This compositional symmetry of the coding sequences in the genome might be responsible for the compositional symmetry of oligonucleotides discussed here. This is because, if the proportion of coding sequences in both the DNA strands is similar, the abundance of individual amino acids would also be similar in the proteomes of both the strands. Also, the abundance of a di or tri peptide in the proteomes of both strands would be similar. So, if the total number of a dipeptide encoded by one of the strands in a genome is similar to the total number of the dipeptide encoded by the other strand, the abundance of the hexanucleotide encoding the dipeptide will be similar in both the DNA strands, as the codon usage pattern in an organism remains same for both the DNA strands in a genome. This will lead to a compositional symmetry of the hexanucleotide in the DNA duplex, even if it is a non-palindrome.

In eukaryotes, the proportion of coding sequences in the genome is relatively less in comparison to the non-coding sequences. This is because of the presence of intervening sequences (introns)<sup>9</sup> within the genes and the genome is replete with repetitive sequences<sup>10</sup>. In our analysis we observed a compositional symmetry in these eukaryotic genomes, in spite of the presence of a significant proportion of non-coding sequences. How is it possible with our above presumption that the presence of coding sequences plays an important role in evolving compositional symmetry in genomes? It is pertinent to point out that repetitive DNA sequences have appeared in the genomes by the activity of transposons<sup>11</sup>. The transposons get inserted randomly in the genome and

there is an equal chance that these get inserted in either of the orientations. So the repetitive elements (non-coding sequence) generated in the genome by random integration of transposons would not destroy the compositional symmetry created by coding sequences. Regarding the presence of introns, there must be a strong sequence conservation among many intervening sequences in genes in a genome, because there are few ribonucleoproteins present in the cell that are responsible for removing the intronic regions from different genes. Wide variability in the intronic sequences would

invariably create difficulty in splice site recognition by the rather small repertoire of ribonucleoproteins involved in subsequent spliceosome complex formation<sup>12</sup>. However, the occurrence of fewer variants in the intervening sequences would not destroy the compositional symmetry evolved in the genome by the presence of coding sequences. Again, it is worthwhile to point out about one of the speculations that introns are degenerate transposable elements which were inserted into the gene in the very distant past, but which have lost the capacity for transposition<sup>13</sup>.

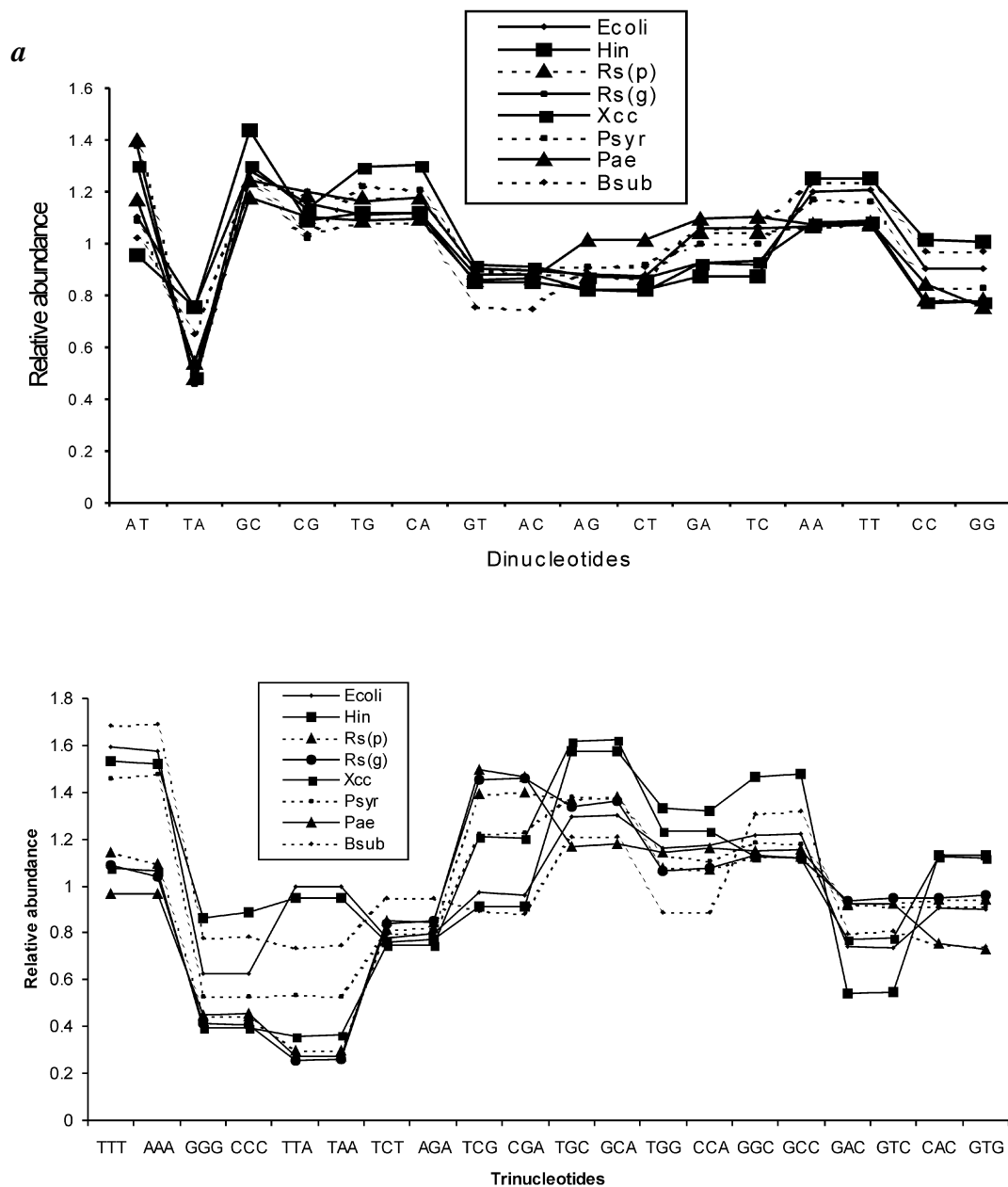
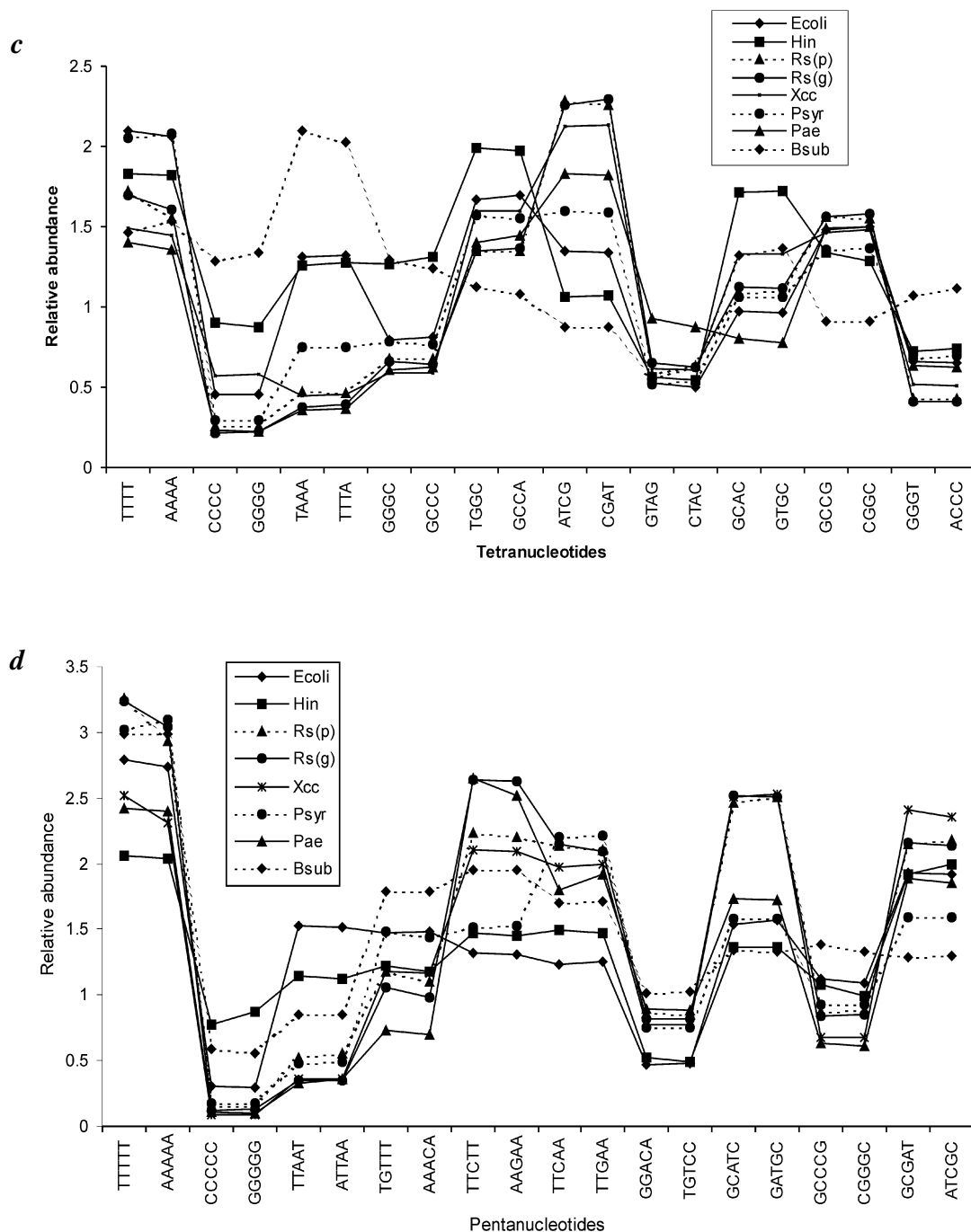


Figure 1. a-b (Contd...)



**Figure 1.** *a-d*, Relative abundance of oligonucleotides (di, tri, tetra and penta) in one of the DNA strands in bacterial genomes. Relative abundance of an oligonucleotide is calculated by dividing the observed abundance of an oligonucleotide in the DNA strand (found by 'seqsearch' software) with the expected abundance of the oligonucleotide (calculated by multiplying the frequency of the individual nucleotides with the length of the DNA strand; frequency of a nucleotide is calculated by dividing the abundance of the nucleotide in a DNA strand with the length of the DNA strand). If the relative abundance value of an oligonucleotide is more than 1, the oligonucleotide is more preferred in the genome and vice versa. It is clear that two complementary oligonucleotides have similar relative abundance in one of the DNA strands in a genome, suggesting compositional symmetry of the DNA duplex.

This compositional symmetry in the DNA duplex will also lead to the synchronous completion of DNA replication in both the strands. According to the recent model of replication,

each of the DNA strands within a replicon is being synthesized independently by a separate replisome complex<sup>14</sup>. During replication, nucleotides are diffused to the

**Table 3.** Oligonucleotide abundance in one of the DNA strands of the genomes in one Archaea and in six eukaryote species (one chromosome in each of the eukaryotic genomes)

Sl. no.	Nucleotide repeats	<i>M. maripaludis</i>	<i>A. thalina</i> (chr. 1)	<i>A. thalina</i> (chr. 2)	<i>D. melanogaster</i>	<i>H. sapiens</i> (contig 4; chr. 1)	<i>O. sativa</i> (chr. 1)	<i>P. falciparum</i> (chr. 1)	<i>S. cerevisiae</i> (chr. 1)
<b>Dinucleotides</b>									
1	AT*	168913	981134	666275	123053	5648	812643	107431	19769
2	TA*	133503	802885	559838	112619	4222	670760	102009	16181
3	GC*	50592	336794	214166	46764	7123	533726	6008	8910
4	CG*	40498	257186	171048	39396	1673	412757	5137	7089
5	<b>TG</b>	<b>98416</b>	<b>693038</b>	<b>444588</b>	<b>75027</b>	<b>9525</b>	<b>661079</b>	<b>27825</b>	<b>15617</b>
6	<b>CA</b>	<b>98661</b>	<b>695525</b>	<b>447356</b>	<b>74442</b>	<b>8900</b>	<b>665514</b>	<b>28109</b>	<b>15224</b>
7	GT	75391	565329	370673	65827	6083	529951	24390	12938
8	AC	76182	565375	371722	65842	5466	533790	25251	12493
9	<b>AG</b>	<b>81069</b>	<b>649250</b>	<b>421496</b>	<b>62283</b>	<b>9578</b>	<b>594086</b>	<b>23338</b>	<b>13621</b>
10	<b>CT</b>	<b>79963</b>	<b>649834</b>	<b>426005</b>	<b>63263</b>	<b>8855</b>	<b>593494</b>	<b>21803</b>	<b>13112</b>
11	GA	94000	697351	452293	64116	7570	604245	25902	14478
12	TC	92348	700374	458529	64496	6839	604249	23789	14021
13	<b>AA</b>	<b>232928</b>	<b>1225804</b>	<b>823771</b>	<b>147956</b>	<b>7212</b>	<b>910948</b>	<b>105131</b>	<b>23947</b>
14	<b>TT</b>	<b>227936</b>	<b>1228270</b>	<b>826130</b>	<b>149067</b>	<b>7805</b>	<b>902496</b>	<b>96348</b>	<b>24151</b>
15	CC	55085	368604	258091	41480	8728	509677	10234	9219
16	GG	55652	367678	251115	41237	10059	510797	10586	9437
<b>Trinucleotides</b>									
1	<b>TTT</b>	<b>96005</b>	<b>462900</b>	<b>313809</b>	<b>60501</b>	<b>2906</b>	<b>330538</b>	<b>46734</b>	<b>8845</b>
2	<b>AAA</b>	<b>99564</b>	<b>462644</b>	<b>312603</b>	<b>60240</b>	<b>2595</b>	<b>335168</b>	<b>52053</b>	<b>8576</b>
3	GGG	10608	59864	44574	7760	3140	111805	1572	1593
4	CCC	10848	59927	46600	7739	2591	110396	1381	1622
5	<b>TTA</b>	<b>52338</b>	<b>263905</b>	<b>186741</b>	<b>37773</b>	<b>1202</b>	<b>192929</b>	<b>30566</b>	<b>4694</b>
6	<b>TAA</b>	<b>52337</b>	<b>263802</b>	<b>187025</b>	<b>37608</b>	<b>1117</b>	<b>194557</b>	<b>31939</b>	<b>4787</b>
7	TCT	27101	254403	160655	19425	2250	177887	8604	4424
8	AGA	28405	254220	158664	19461	2446	178096	9667	4537
9	<b>TCG</b>	<b>22321</b>	<b>87448</b>	<b>61690</b>	<b>12460</b>	<b>336</b>	<b>102739</b>	<b>1803</b>	<b>2200</b>
10	<b>CGA</b>	<b>13690</b>	<b>86849</b>	<b>61024</b>	<b>12553</b>	<b>340</b>	<b>103283</b>	<b>1987</b>	<b>2158</b>
11	TGC	22321	114532	71456	15715	2002	152664	2602	2945
12	GCA	22355	114645	71138	15400	2088	153055	2670	2993
13	<b>TGG</b>	<b>21149</b>	<b>143233</b>	<b>93015</b>	<b>15086</b>	<b>3145</b>	<b>154568</b>	<b>3739</b>	<b>3691</b>
14	<b>CCA</b>	<b>20645</b>	<b>143646</b>	<b>95252</b>	<b>15094</b>	<b>2774</b>	<b>155448</b>	<b>3725</b>	<b>3696</b>
15	GGC	8389	62068	41136	9599	2521	135050	782	1905
16	GCC	8243	61826	41763	9367	2373	134094	733	1960
17	<b>GAC</b>	<b>10454</b>	<b>100326</b>	<b>67032</b>	<b>11369</b>	<b>1345</b>	<b>108024</b>	<b>2939</b>	<b>2384</b>
18	<b>GTC</b>	<b>9732</b>	<b>100840</b>	<b>67527</b>	<b>11281</b>	<b>1336</b>	<b>108070</b>	<b>2668</b>	<b>2455</b>
19	CAC	12732	118156	76070	12974	2059	139995	3713	2722
20	GTG	12613	116882	74974	13369	2410	138061	3778	2798
<b>Tetranucleotides</b>									
1	<b>TTTT</b>	<b>41631</b>	<b>177161</b>	<b>120974</b>	<b>24678</b>	<b>1257</b>	<b>129896</b>	<b>26119</b>	<b>3465</b>
2	<b>AAAA</b>	<b>43313</b>	<b>177329</b>	<b>120137</b>	<b>24382</b>	<b>1135</b>	<b>132462</b>	<b>29241</b>	<b>3255</b>
3	CCCC	2347	9659	8304	1403	678	29013	246	299
4	GGGG	2393	9817	7982	1379	853	29626	332	279
5	<b>TAAA</b>	<b>24595</b>	<b>96496</b>	<b>71495</b>	<b>15869</b>	<b>389</b>	<b>70302</b>	<b>12989</b>	<b>1696</b>
6	<b>TTTA</b>	<b>23816</b>	<b>96577</b>	<b>71458</b>	<b>15508</b>	<b>424</b>	<b>69523</b>	<b>12139</b>	<b>1657</b>
7	GGGC	2017	9146	7038	1872	750	26334	132	316
8	GCCC	2112	9072	7193	1762	656	25777	110	352
9	<b>TGGC</b>	<b>2883</b>	<b>23793</b>	<b>14801</b>	<b>3576</b>	<b>744</b>	<b>36598</b>	<b>301</b>	<b>805</b>
10	<b>GCCA</b>	<b>2685</b>	<b>23760</b>	<b>15081</b>	<b>3488</b>	<b>715</b>	<b>36318</b>	<b>285</b>	<b>800</b>
11	ATCG	5290	24876	17052	3506	60	25163	603	624
12	CGAT	5287	24542	16840	3615	61	25212	682	644
13	<b>GTAG</b>	<b>3502</b>	<b>25835</b>	<b>17765</b>	<b>2735</b>	<b>289</b>	<b>29649</b>	<b>1024</b>	<b>764</b>
14	<b>CTAC</b>	<b>3463</b>	<b>25891</b>	<b>17853</b>	<b>2851</b>	<b>271</b>	<b>29283</b>	<b>1085</b>	<b>672</b>
15	GCAC	2915	15668	10008	2687	461	31245	525	531
16	GTGC	2857	15559	9820	2756	500	31141	434	517
17	<b>GCCG</b>	<b>1142</b>	<b>10614</b>	<b>6854</b>	<b>1831</b>	<b>129</b>	<b>43009</b>	<b>88</b>	<b>351</b>
18	<b>CGGC</b>	<b>1142</b>	<b>10671</b>	<b>6885</b>	<b>1870</b>	<b>134</b>	<b>43202</b>	<b>83</b>	<b>299</b>
19	GGGT	1219	18293	13863	2214	576	23204	550	485
20	ACCC	2959	18316	14568	2263	508	23087	543	490

(contd...)

# RESEARCH COMMUNICATIONS

**Table 3.** (contd...)

Sl. no.	Nucleotide repeats	<i>M. maripaludis</i>	<i>A. thalina</i> (chr. 1)	<i>A. thalina</i> (chr. 2)	<i>D. melanogaster</i>	<i>H. sapiens</i> (contig 4; chr. 1)	<i>O. sativa</i> (chr. 1)	<i>P. falciparum</i> (chr. 1)	<i>S. cerevisiae</i> (chr. 1)
<b>Pentanucleotides</b>									
1	TTTTT	16491	70542	47383	10000	642	54569	16220	1453
2	AAAAA	16944	70744	46910	9649	635	55846	17965	1305
3	CCCCC	397	1536	1703	261	162	8507	56	43
4	GGGGG	393	1607	1700	272	191	8615	96	59
5	TTAAT	6322	25226	17994	3979	79	20699	3340	385
6	ATTAA	6703	25336	17945	4108	54	20689	3450	375
7	TGTTT	4077	32247	20963	3386	208	20606	1673	444
8	AAACA	4123	32288	21243	3331	167	21133	1913	419
9	TTCTT	5685	40672	24432	2659	159	21995	2348	723
10	AAGAA	6089	40465	24369	2841	173	22270	2770	772
11	TTCAA	6653	25698	17108	2711	122	16875	1044	579
12	TTGAA	6765	25881	17219	2608	122	16571	1107	627
13	GGACA	811	6077	3957	788	144	7426	155	159
14	TGTCC	762	6166	4023	885	141	7428	177	167
15	GCATC	1270	7508	4680	760	83	9412	143	190
16	GATGC	1398	7478	4731	743	85	9408	129	186
17	GCCCG	319	1067	1012	305	54	5645	11	47
18	CGGGC	290	1068	1030	288	57	5769	7	53
19	GCGAT	732	3905	2523	620	31	5012	54	110
20	ATCGC	725	3933	2576	614	27	5047	65	109
<b>Hexanucleotides</b>									
1	TTTTTT	5892	33094	20927	4113	395	26704	11337	705
2	AAAAAA	5975	33221	20789	3960	409	27345	12184	595
3	CCCCCC	54	248	436	54	32	2859	13	11
4	GGGGGG	62	270	437	53	26	2909	38	18
5	TTAAAA	5334	11816	10437	2583	38	9435	1789	161
6	TTTTAA	5127	11816	10461	2549	66	9300	1739	163
7	CCGGGG	55	268	244	31	13	1448	2	4
8	CCCCGG	44	261	237	47	18	1460	2	4
9	TGCATC	695	2833	1691	244	28	3346	46	53
10	GATGCA	747	2800	1771	211	23	3305	46	73
11	GTGTCA	112	1877	1239	172	37	2090	49	51
12	TGACAC	135	1865	1167	202	29	2030	50	44
13	CAAACG	351	1998	1219	287	6	1745	56	52
14	CGTTTG	334	2097	1149	301	3	1727	33	45
15	CACACA	187	3613	1920	383	102	4330	188	68
16	TGTGTG	173	3587	1930	389	89	3900	203	38
17	ACGACG	111	1286	678	109	2	3053	17	38
18	CGTCGT	81	1418	684	107	4	2913	15	32
19	CGGTTA	230	1367	1051	146	1	1004	24	25
20	TAACCG	214	1383	1125	151	1	960	25	18
<b>Heptanucleotides</b>									
1	TTTTTTT	1620	18793	11166	1863	285	14473	8706	403
2	AAAAAAA	1620	18841	11199	1826	311	14689	9141	325
3	CCCCCCC	9	68	218	18	4	1508	5	2
4	GGGGGGG	6	87	213	27	1	1504	18	8
5	TTTAACC	193	963	760	96	2	639	29	18
6	GGTAAA	193	1003	732	103	10	686	40	13
7	CCGGTTA	43	393	235	17	0	215	0	4
8	TAACCGG	31	391	246	17	1	204	3	3
9	TGCAAGC	141	556	266	73	4	755	6	8
10	GCTTGCA	145	513	305	80	11	795	4	10
11	GGGTTTA	108	901	693	48	4	480	104	8
12	TAAACCC	133	827	626	73	2	466	56	10
13	AGGCCCG	26	71	70	8	8	282	1	1
14	CGGGCCT	20	59	63	12	8	290	0	3
15	ATCGACG	44	309	170	37	0	422	2	9
16	CGTCGAT	36	349	171	33	0	451	6	6

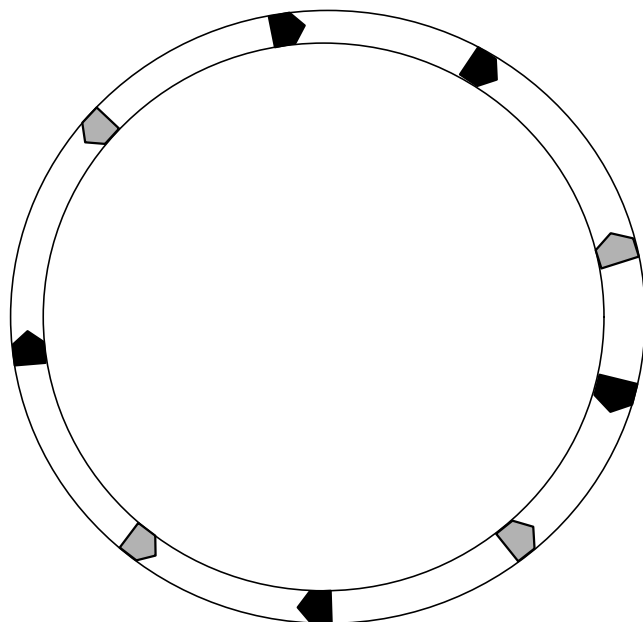
(contd...)

Table 3. (contd...)

Sl. no.	Nucleotide repeats	<i>M. maripaludis</i>	<i>A. thalina</i> (chr. 1)	<i>A. thalina</i> (chr. 2)	<i>D. melanogaster</i>	<i>H. sapiens</i> (contig 4; chr. 1)	<i>O. sativa</i> (chr. 1)	<i>P. falciparum</i> (chr. 1)	<i>S. cerevisiae</i> (chr. 1)
17	AGCTAAC	23	571	403	32	0	468	12	11
18	GTTAGCT	30	569	380	54	3	488	17	9
19	GCTAGGT	8	224	257	28	5	361	1	4
20	ACCTAGC	7	231	240	27	4	368	1	7
Octanucleotides									
1	TTTTTTTT	237	11479	6490	947	224	7610	7111	264
2	AAAAAAAA	226	11529	6593	957	256	7676	7298	207
3	CCCCCCCC	1	44	149	11	2	1078	2	0
4	GGGGGGGG	1	60	149	20	0	1056	9	4
5	GGTTCCAA	52	193	136	15	0	116	6	4
6	TTGGAACC	71	199	119	10	2	129	3	2
7	GTAAGCAA	22	265	121	19	0	176	3	7
8	TTGCTTAC	16	230	132	19	0	164	5	9
9	GCTAGGAA	2	90	95	13	2	132	2	2
10	TTCCTAGC	4	80	90	14	1	131	1	2
11	AACCGGTC	8	103	74	6	0	90	1	1
12	GACCGGTT	8	87	65	11	0	59	0	0
13	GGGAAATT	71	230	141	23	3	205	6	9
14	AATTTCCC	90	214	137	36	3	185	9	8
15	AATTTGGG	53	302	178	28	3	234	9	10
16	CCCAAATT	59	283	202	34	2	221	7	3
17	AAGCCGGC	0	43	20	8	0	132	0	0
18	GCCGGCTT	2	35	20	11	0	140	0	3
19	TTGCTTAG	2	235	153	23	0	162	1	5
20	CTAAGCAA	3	197	139	19	2	165	1	3

\*Palindromic sequences.

Complementary oligonucleotides have either been shown in bold or in regular letters.



**Figure 2.** Schematic representation of oligonucleotide compositional symmetry in bacterial genomes. Bacterial genome is represented as two circles, each circle representing one strand of the DNA duplex. Black arrows represent a non-palindromic oligonucleotide base pair present in one direction. Shaded arrows represents the same non-palindromic oligonucleotide base pair present in reverse orientation in the genome. The number of black arrows (direct repeat) is similar to the number of shaded arrows (inverted repeat), suggesting compositional symmetry of the genome.

site of synthesis from the cytosol. Only the complementary nucleotide is incorporated in the newly synthesized strand, whereas non-complementary nucleotides are not accepted. The possibility of a particular nucleotide coming to the site of synthesis is dependent on its relative abundance in the cytosol. If the DNA duplex is asymmetric, e.g. if one of the strands is purine rich, then during the replication more number of pyrimidines would have to come to the site of synthesis for the formation of the daughter strand. Similarly, in case of the complementary pyrimidine-rich strand, more number of purines would have to come to the site of replication. If in the nucleotide pool, purines are more than pyrimidines, then the daughter strand being synthesized against the pyrimidine-rich strand, would be synthesized earlier in comparison to the daughter strand being synthesized against the complementary purine-rich strand. This problem would not arise in the case where both DNA strands would have similar distribution of purines and pyrimidines.

Studies on several bacterial genomes have revealed that there is a compositional asymmetry of nucleotides between leading and lagging strands in the genome<sup>15</sup>. The leading strand is rich in keto bases (G/T), whereas the lagging strand is rich in amino bases (A/C). This does not contradict our observation because in bacteria, the usual mode of replication is bi-directional. In this mode of replication, from the origin to the termination of replication, each strand gets synthesized

continuously (leading strand) in one direction and discontinuously (lagging strand) in the other direction. Usually, the replication termination point from the origin of replication is present in a symmetrical manner to complete the replication simultaneously from the opposite fork. The localized asymmetry referred to above, therefore, would not affect the overall symmetry between the DNA strands. DNA replication rate is an important phenomenon in the cell. During replication, though the leading and the lagging strands are being synthesized by the same replisome complex, the directional advantage of the leading strand synthesis might lead to a net delay in the synthesis of the lagging strand. To overcome this delay and complete both strands simultaneously, the cell might have adopted a strategy of maintaining the leading strand rich in keto bases and vice versa. Further work in this aspect will reveal the mechanism and reason for the leading and lagging strand compositional asymmetry within a replisome. It is pertinent to note that the composition of oligonucleotides in the genome has been analysed earlier by several workers<sup>16,17</sup>, but their objective was to analyse differential oligonucleotide compositional pattern in the genome and also how this can be used to find out alien DNA sequences in the genome. More studies on this aspect will reveal the exact reason for the symmetrical nature of the genome.

1. Watson, J. D. and Crick, F. H. C., Molecular structure of nucleic acids: A structure for deoxyribonucleic acid. *Nature*, 1953, **171**, 737–738.
2. Kunst, F. *et al.*, The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*. *Nature*, 1997, **390**, 249–256.
3. Blattner, F. R. *et al.*, The complete genome sequence of *Escherichia coli* K-12. *Science*, 1997, **277**, 1453–1474.
4. Fleischmann, R. D. *et al.*, Whole genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science*, 1995, **406**, 496–512.
5. Stover, C. K. *et al.*, Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic pathogen. *Nature*, 2000, **390**, 959–964.
6. Buell, C. R. *et al.*, The complete genome sequence of the *Arabidopsis* and tomato pathogen *Pseudomonas syringae* pv. tomato DC3000. *Proc. Natl. Acad. Sci. USA*, 2003, **100**, 10181–10186.
7. Salanoubat, M. *et al.*, Genome sequence of the plant pathogen *Ralstonia solanacearum*. *Nature*, 2002, **415**, 497–502.
8. da Silva A. C. *et al.*, Comparison of the genomes of two *Xanthomonas* pathogens with different host specificities. *Nature*, 2002, **417**, 459–463.
9. Lewin, B., Exons and introns. In *Genes VI*, Oxford University Press, New York, 1997, pp. 644–662.
10. Lewin, B., Genomes. In *Genes VI*, Oxford University Press, New York, 1997, pp. 663–686.
11. Bowen, N. J. and Jordan, I. K., Transposable elements and the evolution of eukaryotic complexity. *Curr. Issues Mol. Biol.*, 2002, **4**, 65–76.
12. Smith, J. M., The evolution of the eukaryotic genome. In *Evolutionary Genetics*, Oxford University Press, New York, 1999, 2nd edn, pp. 201–223.
13. Watson, J. D., Baker, T. A., Bell, S. P., Gann, A., Levine, M. and Losick, R., RNA splicing. In *Molecular Biology of the Gene*, Pearson-Education, Singapore, 2004, 5th edn, pp. 379–410.

14. Watson, J. D., Baker, T. A., Bell, S. P., Gann, A., Levine, M. and Losick, R., The replication of DNA. In *Molecular Biology of the Gene*, Pearson-Education, Singapore, 2004, 5th edn, pp. 181–234.
15. Rocha, E. P. C., Danchin, A. and Viari, A., Universal replication biases in bacteria. *Mol. Microbiol.*, 1999, **32**, 11–16.
16. Karlin, S., Mrazek, J. and Campbell, A. M., Compositional biases of bacterial genomes and evolutionary implications. *J. Bacteriol.*, 1997, **179**, 3899–3913.
17. Coenye, T. and Vandamme, P., Simple sequence repeats and compositional bias in the bipartite *Ralstonia solanacearum* GM11000 genome. *BMC Genomics*, 2003, **4**, 10.

ACKNOWLEDGEMENTS. We thank Prof. Bolin K. Konwar, Head, Molecular Biology and Biotechnology, Department, Tezpur University, for support and Mr Manabendra Mandal for useful discussions. We also thank Drs A. Prakash and Malay K. Basu for comments on the manuscript.

Received 27 December 2004; revised accepted 19 April 2005

## Evaluation of heat acclimation and salicylic acid treatments as potent inducers of thermotolerance in *Cicer arietinum* L.

Usha Chakraborty\* and Cyaria Tongden

Plant Biochemistry Laboratory, Department of Botany, University of North Bengal, Siliguri 734 430, India

**Efficacy of heat acclimation and salicylic acid (SA) treatments in induction of thermotolerance was tested in six different genotypes of *Cicer arietinum* L. Remarkable reduction in relative injury of membranes was observed in plants pre-treated with SA in comparison to heat-acclimatized and untreated control seedlings subjected to lethal temperature treatment. Both treatments resulted in increase in protein and proline content over control seedlings, which was more significant in SA pre-treatments, with the maximum increase being recorded in ICC 4918 and 1852. Both treatments led to the induction of peroxidase (POX), ascorbate peroxidase (APOX) and catalase (CAT) activities. Activities of POX and APOX increased remarkably, while CAT showed a reduction in activity.**

**Keywords:** Thermotolerance, salicylic acid, heat acclimation, antioxidative enzymes, *Cicer arietinum*.

HIGH surface temperatures are common to soils during periods of drought. Seedlings frequently experience high temperature during emergence and establishment in many regions of the world, which leads to reduction in yield<sup>1</sup>. When

\*For correspondence. (e-mail: chakrabortyusha@hotmail.com)