9. Datta, P. and Bhattacharyya, D., Analysis of fluorescence excitation–emission matrix of multicomponent drugs: A case study with human placental extract used as wound healer. *J. Pharm. Biomed. Anal.*, 2004, **36**, 211–218.
10. Corre, J., Lucchini, J. J., Mercier, G. M. and Cremieux, A., Antibacterial activity of phenethyl alcohol and resulting membrane alterations. *Res. Microbiol.*, 1990, **141**, 483–497.
11. Dominguez, M. C., de La Rosa, M. and Borobio, M. V., Application of a spectrophotometric method for the determination of postantibiotic effect and comparison with viable counts in agar. *J. Antimicrob. Chemother.*, 2001, **47**, 391–398.
12. Beauvoit, B., Liu, H., Kang, K., Kaplan, P. D., Miwa, M. and Chance, B., Characterization of absorption and scattering properties for various yeast strains by time-resolved spectroscopy. *Cell Biophys.*, 1993, **23**, 91–109.
13. Muratore, O. *et al.*, Evaluation of the trophic effect of human placental polydeoxyribonucleotide on human knee skin fibroblasts in primary culture. *Cell. Mol. Lifesci.*, 1997, **53**, 279–285.
14. Failla, C. M., Odorisio, T., Cianfarani, F., Schietroma, C., Puddu, P. and Zambruno, G., Placenta growth factor is induced in human keratinocytes during wound healing. *J. Invest. Dermatol.*, 2000, **115**, 388–395.
15. Scane, T. M., Guest, J. F. and Hawkins, D. F., Extraction of a low molecular weight antibacterial peptide from human placenta. *Br. J. Obstet. Gynaecol.*, 1988, **95**, 633–634.
16. Sharma, U. K. *et al.*, A novel factor produced by placental cells with activity against HIV-1. *J. Immunol.*, 1998, **161**, 6401–6406.
17. Rodgers, G. L., Mortensen, J. E., Fisher, M. C. and Long, S. S., *In vitro* susceptibility testing of topical antimicrobial agents used in pediatric burn patients: Comparison of two methods. *J. Burn Care Rehabil.*, 1997, **18**, 406–410.
18. Bernardi, G., Chromatography of nucleic acids on hydroxyapatite columns. *Methods Enzymol.*, 1971, **21**, 95–139.
19. Sur, T. K., Biswas, T. K., Ali, L. and Mukherjee, B., Anti-inflammatory and anti-platelet aggregation activity of human placental extract. *Acta Pharmacol. Sin.*, 2003, **24**, 187–192.
20. Magi, G., Nardoianni, V., Antonelli, P., Cardini, P. and Rovere, V., Valutazione clinica dell'effetto del polidesossiribonucleotide nelle lesioni trofiche degli arti inferiori di soggetti vasculopatici cronici. *Minerva Anestesiol.*, 1991, **16**, 2–5.
21. Bianchini, P., Tellini, N., Morani, A. M. and Follani, M. G., Pharmacological data on polydeoxyribonucleotide of human placenta. *Int. J. Tiss. React.* 1981, **3**, 151–154.
22. Bigliardi, P., Treatment of acute radiodermatitis of first and second degree with semi-greasy placenta ointment. *Int. J. Tiss. React.*, 1982, **4**, 153–154.
23. Pati, S., Saumandal, B. K., Bhattacharyya, A. K. and Saumandal, P., Clinical evaluation of effect of dressing with placental extract (Inj Placentrex) in the treatment of infected wounds. *J. Obstet. Gynaecol. India*, 2001, **51**, 124–126.
24. Shukla, V. K., Rasheed, M. A., Kumar, M., Gupta, S. K. and Pandey, S. S., A trial to determine the role of placental extract in the treatment of chronic non-healing wounds. *J. Wound Care*, 2004, **13**, 177–179.
25. Wright, G. E. and Brown, N. C., Deoxyribonucleotide analogs as inhibitors and substrates of DNA polymerases. *Pharmacol. Ther.*, 1990, **47**, 447–497.


ACKNOWLEDGEMENTS. Financial assistance from M/s Albert David Ltd, Calcutta, India is acknowledged. P.D. was supported by a research fellowship from the same source. Mr Chinmoy Sarkar of IICB helped with the HPLC analysis.

Received 4 August 2003; revised accepted 9 November 2004

# An altered G + C% region within potential filamentous hemagglutinin open reading frames of *Ralstonia solanacearum*

**Debojyoti Das, Sujit Kumar Verma and Suvendra Kumar Ray***

Department of Molecular Biology and Biotechnology, Tezpur University, Tezpur 784 028, India

**Analysis of hemagglutinin open reading frames (ORFs) of *Ralstonia solanacearum* using FramePlot exhibited an altered G + C% (altered GC regions) at the 3′ regions of most large ORFs. The average G + C% values of these hemagglutinin ORFs and the altered GC regions are found to be 66.51 and 51.37 respectively. The amino acid usage of this altered GC region is similar to the whole protein, while the codon usage pattern is different. We reason that this altered GC region is an evolutionary adaptation rather than acquisition by horizontal gene transfer. The codon usage in the altered GC region could have a regulatory effect in the rate of translation of these large surface proteins.**

FILAMENTOUS hemagglutinin genes were first discovered in the human pathogenic bacterium *Bordetella pertussis* and play an important role in the attachment of this bacterium to host cells[1]. The genome sequence of *Xylella fastidiosa* (the first phytopathogenic bacterium whose genome was completely sequenced) revealed the occurrence of three filamentous hemagglutinin genes in this plant pathogen[2]. Genome sequences of plant and animal pathogenic bacteria have revealed the wide occurrence of these genes among pathogenic bacteria[3]. *Ralstonia solanacearum* is a Gram-negative bacterium that causes a lethal wilt disease in more than 200 plants. One of the striking features of the *R. solanacearum* genome sequence is the presence of twenty-seven hemagglutinin genes, of which thirteen are called probable hemagglutinins as they exhibit homology with the filamentous hemagglutinin (FhaB) of *B. pertussis*, and the HMW1A/HMW2A adhesins of *Haemophilus influenzae*[4]. The remaining fourteen are called hemagglutinin-like proteins because these open reading frames (ORFs) contain variable internal repeats that are structurally related to filamentous hemagglutinins[4]. Here we discuss the probable hemagglutinins as potential filamentous hemagglutinins. Filamentous hemagglutinins are surface-localized adhesins encoded by large ORFs. After their synthesis in the cytosol, the huge polypeptides are translocated to the surface across the cytoplasmic membrane using the sec-protein apparatus[5]. This is followed by translocation across the outer membrane using the two-partner secretion system[6]. It has been speculated that during transport, the protein remains in an

*For correspondence. (e-mail: suven@tezu.ernet.in)

extended conformation[5,6]. While analysing these gene sequences of *R. solanacearum,* we observed a variation of GC% at the 3′ end of several ORFs. GC% variation observed in the genome has been attributed mainly to the recent acquisition of genes by horizontal gene transfer. Our observation suggests that this altered GC region is not due to horizontal gene transfer; rather it is due to an evolutionary adaptation of the organism for some important functions.

FramePlot[7] analysis of twenty-seven potential hemagglutinin genes[4] of *R. solanacearum* revealed nine ORFs with altered GC% variation (within the ORFs), referred to as altered GC regions. These nine ORFs include most of the large ORFs and belong to the potential filamentous hemagglutinin class (Table 1). In case of eight ORFs, the altered GC regions are present at the 3′ end (Figure 1), while RSp1620 has the altered GC region present at the 5′ end and hence is excluded in our future study. The GC% of *R. solanacearum* genome[4] is 66.97. The average GC% of the above eight hemagglutinin ORFs is 66.51, which is close to the genome GC% value. However, the average GC% value of the ~500 bp (min 465 bp and max 588 bp) at the 3′ end is found to be 51.37 (Table 1). The average GC% value for the unaltered GC region (Table 1) in the ORF is not significantly different from that of the whole ORF because the altered GC region is only limited to ~5% of the whole ORF, except RSp1071. The average GC% at the third nucleotide position of codons in the eight hemagglutinin ORFs was found to be 84.18. This value is concordant with the observation of ORFs found in GC-rich organisms[7]. The average GC% at the third nucleotide position of codons at the altered GC region in eight ORFs was found to be 48.63, which is significantly lower than 84.18 (Table 1). The $Z$ values for both the altered GC region and the unaltered GC region were compared (Table 1) and found to be significant with our observation. The maximum $Z$ value for the unaltered GC region was found to be 2.05. The $P$ value derived from the standard $Z$ table is 0.02, which is greater than 0.01. The $Z$ value for the altered GC region varies within –11.26 to –6.71. The $P$ values for these are much less than 0.0001. This suggests that the difference in GC% observed in the altered GC region in comparison to the whole ORF is statistically significant. We compare the amino acid as well as codon usage pattern of the altered GC regions with the whole protein sequence that was obtained from the translated sequence using the software 'DNA for Windows' (Crop Protection Group, Durham, UK). From Figure 2 *a*, it is evident that amino acid contents encoded in the altered GC region are similar to those in the whole protein, though there is a difference in lysine and threonine. This difference is not significant enough to give rise to the altered GC region observed here. To study the codon usage pattern, codons for alanine (GCT, GCC, GCA, GCG), glycine (GGT, GGC, GGA, GGG), lysine (AAA, AAG), asparagine (AAT, AAC) and isoleucine (ATT, ATC, ATA) were considered, as these amino acids are encoded either by GC-rich or AT-rich codons. From Figure 2 *b*, it is evident that the codon usage pattern is significantly different between the altered GC region and the whole ORF. In the case of the whole ORF, codons ending with G/C are used more frequently in comparison to the synonymous codons ending with A/T. In the altered GC region, there is a tendency of the more frequent use of codons ending with A/T over the synonymous codons ending with G/C in relation with the whole ORF. Codon usage studies for other amino acids also give similar results (data not shown). This confirms that the altered GC region present within the filamentous hemagglutinin ORFs is due to differential codon usage rather than differential amino acid contents encoded in this region.

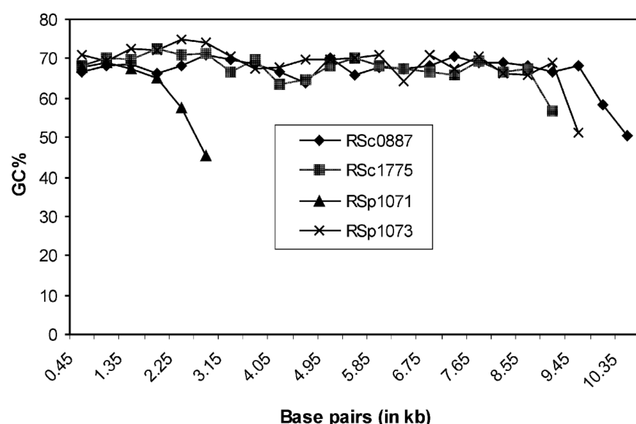


**Figure 1.** GC% through the hemagglutinin ORFs. GC% through the ORF was plotted for RSc0887, RSc1775, RSp1071 and RSp1073 ORFs. In all the ORFs, GC% clearly falls at the 3′ end. For plotting GC% through the ORF, a stretch of 450 nucleotides was taken at a time and its GC% was calculated using the software 'DNA for Windows'. This was repeated till the end of the ORF. In case the total number of nucleotides is not 450 at the end of the ORF, GC% was calculated by adding the sequence to the previous 450 nucleotides and the combined GC% was plotted as one point.

Previously known altered GC regions within the genome of an organism have been mainly reasoned to be acquired through horizontal gene transfer[8]. Therefore, we tried to find out whether the altered GC regions had arisen due to horizontal gene transfer or due to evolutionary forces to achieve novel functionalities. The altered GC regions among different ORFs were compared using BLAST analysis[9]. No similarity was observed either at the nucleotide or amino acid sequence level. In fact, RSc0887 and RSp0540 that exhibit 95% homology between each other at the nucleotide level, surprisingly share no homology at their 3′ ends where the altered GC regions are present. Immediate downstream ORFs to this altered GC region exhibit GC% similar to the *R. solanacearum* genome, suggesting the altered GC region is only restricted to hemagglutinin ORFs. In addition, BLAST analysis also did not reveal any sequence conservation site among hemagglutinin genes where this altered GC region begins. These eliminate a possibility that the origin of this region is due to horizontal gene transfer. Furthermore, we analysed filamentous hemagglutinin

**Table 1.** Filamentous hemagglutinin genes of *Ralstonia solanacearum*

| ORF | Length (bp) | GC% variation[1] | GC% of the whole ORF | GC% of the unaltered region | GC% of the altered region | *Z1* score | *Z2* score | Signal peptide[3] |
|---|---|---|---|---|---|---|---|---|
| RSc0887 | 10506 | + (474)[2] | 66.60 (85.1)[4] | 67.31 (86.8)[4] | 51.47 (48.1)[4] | 0.47 | –9.00 | – |
| RSc1775 | 9498 | + (486) | 67.00 (85.3) | 67.64 (87.2) | 55.30 (50.0) | 0.67 | –6.71 | – |
| RSc3162 | 3039 | – | 66.57 (89.4) | – | – | – | – | + |
| RSc3183 | 2541 | – | 68.67 (89.6) | – | – | – | – | + |
| RSc3188 | 8256 | + (516) | 66.67 (80.7) | 67.49 (83.1) | 54.40 (45.3) | 0.58 | –7.25 | – |
| RSp0540 | 10659 | + (528) | 66.50 (85.0) | 67.49 (87.1) | 48.20 (44.9) | 0.58 | –10.96 | – |
| RSp1071 | 2547 | + (465) | 63.00 (80.4) | 66.43 (88.2) | 47.70 (45.8) | 0.04 | –11.26 | – |
| RSp1073 | 9969 | + (585) | 68.98 (90.7) | 69.94 (92.8) | 53.60 (56.4) | 2.05 | –7.73 | – |
| RSp1444 | 4116 | – | 69.36 (88.0) | – | – | – | – | + |
| RSp1536 | 2541 | – | 68.83 (88.5) | – | – | – | – | + |
| RSp1539 | 8214 | + (588) | 67.40 (84.8) | 68.59 (87.1) | 52.04 (54.1) | 1.24 | –8.66 | – |
| RSp1545 | 8076 | + (492) | 66.00 (81.5) | 67.15 (83.9) | 48.30 (44.5) | 0.38 | –10.9 | – |
| RSp1620 | 3930 | + (300)[#] | 63.30 (86.9) | 64.1 (89.6) | 53.70 (54.00) | – | – | + |
| Mean[5] | | | 66.51 (84.18) | 67.75 (87.02) | 51.37 (48.6) | 0.75 | –9.05 | |
| Standard deviation[5] | | | 1.67 (3.36) | 1.06 (3.12) | 2.99 (4.51) | 0.62 | 1.79 | |

*R. solanacearum* sequences can be downloaded from Genome Information Broker site (www.gib.genes.nig.ac.jp or www.gib.genes.nig.ac.jp/single/index.php?spid=Rsol_GMI1000).
+, Indicates presence; –, Indicates absence.
[1]GC% variation within ORFs found by FramePlot analysis.
[2]Number in bracket indicates the length of altered GC regions at the 3′ end of the ORF (except RSp1620).
[3]Signal peptide analysis was done by SignalP 3.0 server.
[4]Number in bracket indicates the GC% of the third nucleotide of the codons in the region.
[5]Calculations were made only with the values of hemagglutinin ORFs having altered GC regions at 3′ end.
[#]Variation at the 5′ end.
*Z1* score is the *Z* value of the unaltered GC region within the ORF. This was calculated by deducting the mean GC% of the whole ORFs with the GC% of the unaltered region followed by division with the standard deviation value of the whole ORF.
*Z2* score is the *Z* value of the altered GC region within the ORF. (This was calculated as discussed for *Z1*, but the GC% of the altered region was taken instead of GC% of the unaltered region.)

genes of other plant and animal pathogenic bacteria (Table 2). Consistent with our observation, filamentous hemagglutinin genes from the bacteria *Erwinia chrysanthemi*, *Neisseria menningitis*, *Pseudomonas aeruginosa*, *Xanthomonas axonopodis* pv. *citri*, *Xyllela fastidiosa*, and *Yersinia pestis* exhibit altered GC region at the 3′ end of the ORF. In addition, *E. chrysanthemi* has another altered GC region towards the mid-region of the ORF. In *Escherichia coli O157 : H7*, the altered GC region is present at the 5′ end of the ORF. In all cases it has been observed that the altered GC region has a lower GC% in comparison to the whole ORF. This further supports that the altered GC regions have not arisen due to horizontal gene transfer. The filamentous hemagglutinin gene of *B. pertussis* has uniform GC% throughout the ORF like some ORFs in *R. solanacearum*. So filamentous hemagglutinin genes can be subdivided into two groups; one having uniform GC region throughout the ORF; and other having altered GC region within the ORF. To understand the functional significance of this altered GC region, we analysed the amino terminal region

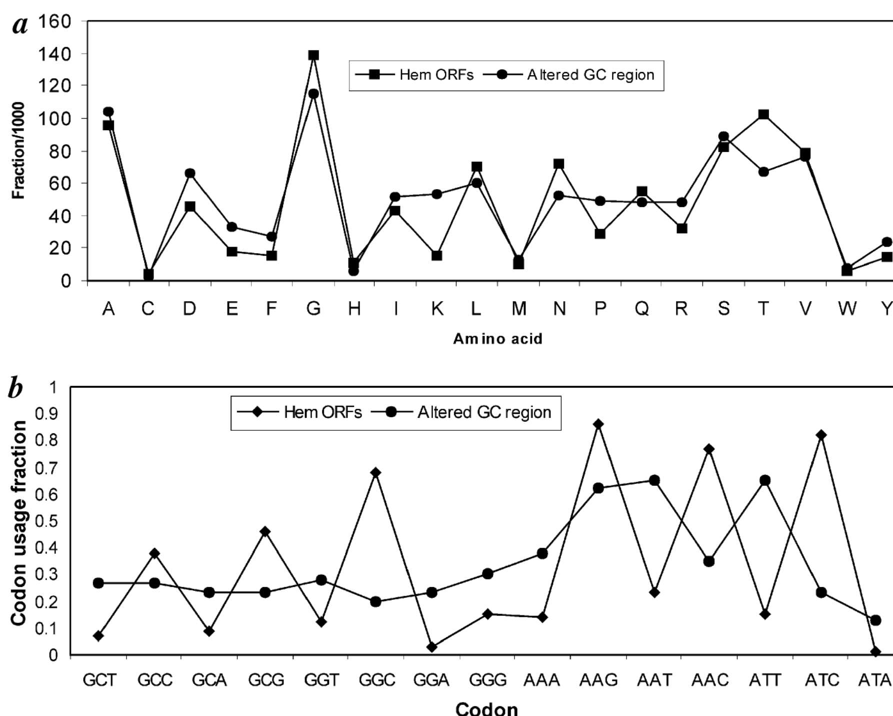


**Figure 2 *a* and *b*.** Amino acid and codon usage pattern. Comparison of amino acid usage (***a***) and codon usage pattern (***b***) between hemagglutinin (Hem) ORFs and altered GC region. ***a***, One-letter code of all 20 amino acids has been given in the *X*-axis. In the *Y*-axis, fraction of particular amino acid per thousand has been given for hemagglutinin genes as well as for the altered GC region. ***b***, Codons for alanine, glycine, lysine, asparagine and isoleucine have been given in the *X*-axis. In the *Y*-axis codon usage fraction for a particular amino acid in relation to other synonymous codons has been plotted.

of filamentous hemagglutinin genes of *R. solanacearum* using SignalP 3.0 software[10]. Interestingly, this program detected a potential signal peptide in filamentous hemagglutinin genes that have uniform GC% throughout the ORF, whereas it could not detect a classical signal peptide in hemagglutinin genes that has an altered GC region within the ORF (Table 1). Though similar studies with the filamentous hemagglutinin genes of other bacteria (Table 2) indicated that most of the genes exhibiting an altered GC region lack a potential signal peptide, it is not true in all cases. In case of *E. chrysanthemi*, *P. aeruginosa* and *X. axonopodis* hemagglutinin genes (Table 2), a potential signal peptide was detected at the amino terminal end of the derived protein sequences though an altered GC region is present at the 3′ region of these ORFs. Since in our observation most of the hemagglutinin genes having altered GC region in their ORF lack a potential signal peptide, we think the altered GC region might play a regulatory role that is related with the transport of the protein.

Our results suggest the presence of an altered GC% and codon usage region at the 3′ end of filamentous hemagglutinin ORFs of *R. solanacearum*. Previously, GC% variation and codon usage regions were reported within different regions of the genome in an organism. This variation is attributed to either the acquisition of genes from other organisms by horizontal gene transfer[8] or location of genes near the replication termination region in the genome[11]. However, presence of the altered GC region within an ORF is surprising and interesting. The possibility of acquisition of these altered GC regions through horizontal gene transfer is eliminated due to: (i). no similarity among sequences present in an organism; (ii) in all the cases the GC% is lower than the whole ORF; (iii) wide occurrence in several ORFs, and (iv) its restriction to a particular region within the ORF. Each organism has a characteristic GC% in its genome, whose functional significance is yet to be revealed. The differential GC% of the genome of different organisms has been attributed to the nucleotide pool present in the cytosol, DNA replication system and DNA repair system[12]. The altered GC region discussed here seems to be not incongruous with the GC% of the genome, possibly due to some functional significance. One of the reasonable explanations we suggest is that the differential nucleotide content might be playing a regulatory function in the hemagglutinin gene expression. Hemagglutinin genes are huge in size. Tertiary/secondary structure of these proteins suggests the presence of β-rich regions and extended conformation[13]. Since these are surface localized proteins, after their synthesis in the cytosol, these proteins need to be transported to the periplasm from where they move to the surface. During its transport, the polypeptide has to remain in an extended conformation. Unlike eukaryotes, in prokaryotes transcription

**Table 2.** Filamentous hemagglutinin genes in other bacteria

| ORF (organism) | Length (bp) | GC% variation[1] | GC% of the genome | GC% of the whole ORF | GC% of the unaltered region | GC% of the altered region | Signal peptide[3] |
|---|---|---|---|---|---|---|---|
| BP1879 (*Bordetella pertusis*) | 10773 | – | 67.6 | 67.5 | – | – | – |
| HecA (*Erwinia chrysanthemi*) | 11547 | + (i) (630)[2] (ii) (700)[@] | 55.1–57 | 68.8 (82.9)[4] | 70.07 (85.24)[4] | (i) 47.14 (42.4)[4] (ii) 47.7[@] (47.57) | + |
| Z1542 (*Escherichia coli*) | 3813 | + (480)[#] | 50.79 | 49.22 (57.5) | 50.26 (60.04) | 42.5 (40.63) | + |
| NMB0493 (*Neisseria meningitidis*) | 8112 | + (471) | 51.4 | 49.12 (47.9) | 49.68 (48.96) | 40.1 (31.21) | – |
| NMB1214 (*N. meningitidis*) | 6822 | + (417) | 51.4 | 50.6 (54.4) | 50.88 (55.22) | 46.28 (42.4) | – |
| NMB1768 (*N. meningitidis*) | 7545 | + (645) | 51.4 | 45.68 (40.5) | 46.9 (42.6) | 32.7 (18.1) | – |
| PA0041 (*Pseudomonas aeruginosa*) | 10608 | + (450) | 66.6 | 68.57 (87.8) | 69.39 (89.81) | 50.0 (43.3) | + |
| XAC1815 (*Xanthomonas axonopodis*) | 14262 | + (720) | 64.7 | 63.15 (72.6) | 64.01 (74.2) | 47.0 (42.9) | + |
| XAC1816 (*X. axonopodis*) | 4293 | + (948) | 64.7 | 61.47 (75.5) | 63.47 (81.61) | 54.43 (53.8) | – |
| XF0889 (*Xyllela fastidiosa*) | 9849 | + (330) | 52.7 | 66.12 (84.0) | 66.6 (85.16) | 52.42 (51.8) | – |
| XF2196 (*X. fastidiosa*) | 10329 | + (930) | 52.7 | 65.29 (82.2) | 67.03 (85.57) | 47.74 (48.1) | – |
| XF2775 (*X. fastidiosa*) | 10368 | + (990) | 52.7 | 65.3 (82.2) | 67.03 (85.54) | 49.29 (50.9) | – |
| YPO0599 (*Yersinia pestis*) | 9888 | + (688) | 46 | 57.23 (63.08) | 58.05 (65.05) | 46.28 (36.7) | – |

DNA sequences can be downloaded from Genome Information Broker site (www.gib.genes.nig.ac.jp or www.gib.genes.nig.ac.jp/single/index.php?spid=Rsol_GMI1000).
+, Indicates presence; –, Indicates absence.
[1]GC% variation within ORFs found by FramePlot analysis.
[2]Number in bracket indicates the length of altered GC regions at the 3′ end of the ORF (except Z1542).
[3]Signal peptide analysis was done by SignalP 3.0 server.
[4]Number in bracket indicates the GC% of third nucleotide of the codons in the region.
[#]Variation at the 5′ end.
[@]Altered GC region present at the mid-portion of the ORF.

and translation are coupled. So translocation of the proteins to the extracytosolic region takes place after their synthesis[14]. This is a problem for the large surface proteins because the polypeptide coming out from the ribosome tends to fold in the cytosol itself. Folding and unfolding of a huge protein in the cytosol would have temporal constraints. However, slow release of the proteins from the translating ribosome to the cytosol might delay the folding process which would invite cytosolic chaperones to bind to the nascent protein and block the folding process prior to their translocation. Altered codon usage at the C-terminal region would affect the rate of translation[12], which would ultimately slow the synthesis and release of these proteins to the cytosol. It is pertinent to say that gene regulation mediated by slowing the rate of translation has been exemplified long before in attenuation mode of regulation in *trp* operon of *E. coli*[15]. During evolution the cell has developed different strategies to regulate translation rate of genes depending upon the requirements. The absence of these altered GC regions in other filamentous hemagglutinin genes (Table 1), suggests that there might be different mechanisms involved in keeping these proteins in an unfolded state. The detection of a potential signal peptide sequence in these protein sequences supports our hypothesis. Consistent with this idea, the presence of an altered GC region as well as signal peptide in some of the ORFs suggests that these proteins have both strategies available to remain in an unfolded state in the cytosol during the translation. However, the filamentous hemagglutinin gene of *B. pertussis* possesses neither an altered GC region nor a potential signal peptide. We expect this protein might be using a different strategy to remain in an unfolded state during

**Table 3.** ORFs of *Ralstonia solanacearum* that do not exhibit altered GC region

| ORF | Length (bp) | GC% variation* | GC% | Probable function | Probable localization |
|---|---|---|---|---|---|
| RSc0049[1] | 1257 | – | 72.16 | Hemagglutinin related protein | Cell surface |
| RSc0115 | 5019 | – | 71.57 | Hemagglutinin related protein | Cell surface |
| RSc0127 | 1443 | – | 69.16 | Hemagglutinin related protein | Cell surface |
| RSc1495 | 1875 | – | 71.57 | Hemagglutinin related protein | Cell surface |
| RSc2796 | 750 | – | 72.67 | Hemagglutinin related protein | Cell surface |
| RSc2797 | 1086 | – | 72.65 | Hemagglutinin related protein | Cell surface |
| RSp0116 | 1227 | – | 72.66 | Hemagglutinin related protein | Cell surface |
| RSp0183 | 1233 | – | 71.05 | Hemagglutinin related protein | Cell surface |
| RSp0808 | 1743 | – | 68.16 | Hemagglutinin related protein | Cell surface |
| RSp0820 | 1341 | – | 68.23 | Hemagglutinin related protein | Cell surface |
| RSp1093 | 1062 | – | 73.16 | Hemagglutinin related protein | Cell surface |
| RSp1094 | 1380 | – | 73.55 | Hemagglutinin related protein | Cell surface |
| RSp1180 | 12321 | – | 69.87 | Hemagglutinin related protein | Cell surface |
| RSp1605 | 7974 | – | 68.6 | Hemagglutinin related protein | Cell surface |
| RS05500 | 489 | – | 61.6 | *Agrobacterium tumefaciens* AttZ | Cell surface |
| RS05353 | 414 | – | 67.9 | *Agrobacterium tumefaciens* AttZ | Cell surface |
| RS01559 | 1905 | – | 70.7 | Outer membrane channel protein | Outer membrane |
| RS02594 | 1365 | – | 69.5 | Outer membrane channel protein | Outer membrane |
| RSp0641 | 20670 | – | 69.16 | Probable peptide synthetase protein | Cytosolic |
| RSp0642 | 17862 | – | 67.79 | Probable peptide synthetase protein | Cytosolic |
| RSc1806 | 12807 | – | 73.73 | Probable polyketide synthase protein | Cytosolic |
| RSc1810 | 7143 | – | 73.57 | Probable polyketide synthase protein | Cytosolic |

[1]*R. solanacearum* sequences can be downloaded from Genome Information Broker site (www.gib.genes.nig.ac.jp or www.gib.genes.nig.ac.jp/single/index.php?spid=Rsol_GMI1000).
*GC% variation within ORFs found by FramePlot analysis.
–, Indicates no variation.

translation. In corroboration with our idea, there is no altered GC region in the large ORFs that encode cytosolic proteins which are much larger than hemagglutinin proteins (RSp0641 and RSp0642; Table 3). Our observation suggests that this altered GC region is limited to filamentous hemagglutinin ORFs and not present in other hemagglutinin-related proteins of *R. solanacearum* (Table 3), though there are large ORFs belonging to this class of genes whose encoded proteins are predicted to be surface-localized. This difference might be due to differential translocation process occurring in the cell[16]. The presence of this altered GC% region at the 3′ end rather than in any site in the ORF, may be to avoid transcription termination by the cryptic ρ-binding site, as it is known that the translating ribosome sterically hinders internal ρ-binding sites that are present in large ORFs due to co-transcriptional translation. Experimental analysis of this region will reveal its exact function in filamentous hemagglutinin biosynthesis.


1. Smith, A. M., Guzman, C. A. and Walker, M. J., The virulence factors of *Bordetella pertussis*: A matter of control. *FEMS Microbiol. Rev.*, 2001, **25**, 309–333.
2. Simpson, A. J. *et al.*, The genome sequence of the plant pathogen *Xylella fastidiosa*. *Nature*, 2000, **406**, 151–157.
3. Rojas, C. M., Ham, J. H., Deng, W-L., Doyle, J. J. and Collmer, A., HecA, a member of a class of adhesin produced by diverse pathogenic bacteria, contributes to the attachment, aggregation, epidermal cell killing, and virulence phenotypes of *Erwinia chrysanthemi* EC16 on *Nicotiana cievelandii* seedlings. *Proc. Natl. Acad. Sci. USA*, 2002, **99**, 13142–13147.
4. Salanoubat, M. *et al.*, Genome sequence of the plant pathogen *Ralstonia solanacearum*. *Nature*, 2002, **415**, 497–502.
5. Manting, E. K. and Driessen, J. M., *Escherichia coli* translocase: The unravelling of a molecular machine. *Mol. Microbiol.*, 2000, **37**, 226–238.
6. Jacob-Dubuisson, F., Locht, C. and Antonie, R., Two partner secretion in Gram-negative bacteria: A thrifty, specific pathway for lage virulence proteins. *Mol. Microbiol.*, 2001, **40**, 306–313.
7. Ishikawa, J. and Hotta, K., FramePlot: A new implementation of the frame analysis for predicting protein-coding regions in bacteria DNA with a high G + C content. *FEMS Microbiol. Lett.*, 1999, **174**, 251–253; http://www.nih.go.jp/~jun/cgi-bin/frameplot.pl
8. Solnick, J. V. and Young, G. M., Bacterial pathogenicity islands and infectious disease. In *Horizontal Gene Transfer* (eds Syvanen, M. and Kado, C. I.), Academic Press, New York, 2002, pp. 111–121.
9. Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D. L., Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.*, 1997, **25**, 3389–3402; http://www.ncbi.nlm.nih.gov/BLAST/Blast.cgi
10. Bendtsen, J, D., Nielsen, H., von Heijne, G. and Brunak, S., Improved predition of signal peptides: SingalP 3.0. *J. Mol. Biol.*, 2004, **340**, 783–795; http://www.cbs.dtu.dk/services/SignalP
11. Daubin, V. and Perriere, G., G + C3 structuring along the genome: A common feature in prokaryotes. *Mol. Biol. Evol.*, 2003, **20**, 471–483.

12. Osawa, S., Jukes, T. H., Watanabe, K. and Muto, A., Recent evidence for evolution of the genetic code. *Microbiol. Rev.*, 1992, **56**, 229–264.
13. Makhov, A. M. *et al.*, Filamentous hemagglutinin of *Bordetella pertussis*. A bacterial adhesin formed as a 50-nm monomeric rigid rod based on a 19-residue repeat motif rich in beta strands and turns. *J. Mol. Biol.*, 1994, **241**, 110–124.
14. von Heijne, G., Life and death of a signal peptide. *Nature*, 1998, **396**, 111–113.
15. Landick, R., Turnbough, Jr. C. and Yanofsky, C., Transcription attenuation. In *Escherichia coli and Salmonella: Cellular and Molecular Biology* (eds Neidhardt, F. C. *et al.*), ASM Press, Washington DC, 1996, 2nd edn, pp. 1263–1286.
16. Thanassi, D. G. and Hultgren, S. J., Multiple pathways allow protein secretion across the bacteria outer membrane. *Curr. Opin. Cell Biol.*, 2000, **12**, 420–430.

ACKNOWLEDGEMENTS. We thank Prof. Bolin K. Konwar, Head, Department of Molecular Biology and Biotechnology, Tezpur University for support and Dr Anand Ramteke for providing the necessary software. We also thank Dr Alak K. Buragohain, Mr Prabhu B. Patil, Dr A. Prakash and Dr R. Rajeshwari for their comments on the manuscript. We are grateful to the referee for his critical comments that helped us to better revise the manuscript.

Received 28 July 2004; revised accepted 19 October 2004

# Identification of differentially expressed ripening-related cDNA clones from tomato (*Lycopersicon esculentum*) using tomato EST array

**K. N. Anjanasree[1], Alka Srivastava[2], Avtar Handa[2] and K. C. Bansal[1,*]**

[1]National Research Centre on Plant Biotechnology, Indian Agricultural Research Institute, New Delhi 110 012, India
[2]Department of Horticulture, Purdue University, West Lafayette, IN 47907-2010, USA

**Qualitative changes in ripening tomato fruits are invariably associated with differential gene expression mediated by the phytohormone ethylene and certain developmental and environmental cues. In the present study, we compared, using the tomato expressed sequence tag (EST) array, the cDNA expression patterns of 'breaker'-stage fruits of two cultivars, Pusa Ruby and Pusa Uphar (DT-2), which significantly differed in their rate of ripening and shelf-life. Screening of bacterial EST microarray filters revealed the presence of 1536 differentially expressed cDNA clones, among which 200 were found up-regulated and 42 down-regulated. The results revealed significant differences in the ripening-related gene expression between the two cultivars which differed in the rate of ripening, and demonstrated the utility of EST-screening approaches as an initial step towards identifying key genes that affect fruit storability, shelf-life and post-harvest characteristics.**

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

THE development of powerful, high throughput genomic technologies like differential display[1], serial analysis of gene expression[2] and microarrays[3] facilitates efficient and accurate monitoring of genetic variations and gene expression patterns at a global level, thus providing new vistas for genome analysis. DNA microarray technology is the key element in functional genomic studies capable of miniaturization and automation, and facilitates a holistic integration of biological knowledge. Arrays based on expressed sequence tags (ESTs) generated by single-pass sequencing of 300–900 bases from the 5′-end of the cDNA clones provide obvious advantage in expression studies compared to anonymous clones, as immediate functional implications can be drawn based on sequence homologies.

Tomato fruit development is a complex process, which is genetically programmed and characterized by distinct stages. Early fruit development occurs in three phases, i.e. fruit set, a stage of rapid cell division and a period of cell enlargement. The later phase is the dynamic period of fruit ripening, where a series of coordinated changes in metabolism occur. In climacteric fruits, system 2 ethylene synthesis occurs during the onset of ripening which is autocatalytic in nature, concomitantly with the initiation of lycopene synthesis. In tomato fruits, this transition from system 1 to system 2 occurs at the breaker stage, which is characterized by the appearance of first visible red colour at the blossom end of the fruit. At this stage, the fruit would have attained a certain degree of maturity, initiated softening and associated changes, including colour and taste development. Accompanying these two distinct phases of development are the dramatic differences in the cellular processes, biosynthetic pathways and above all differential gene expression. Considerable evidences suggest that ripening of tomato requires differential expression of genes encoding ripening-related enzymes catalysing different aspects of the ripening programme. The existence and activity of a group of developmentally regulated genes may be considered as a general feature of the ripening process in all fruits[4]. Subsets of genes expressed during early and late stages of development reflect the nature of physiological and biochemical events unique to each stage.

Although defined patterns of differential expression exist at each stage of fruit development, the number of genes exhibiting altered expression shows a dramatic increase at the commencement of ripening[5]. Genes or multigene families which are differentially expressed during ripening include housekeeping genes like those coding for respiratory enzymes and other genes involved in different facets of ripening and senescence, like chlorophyll and cell-wall degradation, genes mediating ethylene biosynthesis and perception, lycopene biosynthesis and production of taste and aroma volatiles. Apart from hitherto known genes, efforts are also