DOI: 10.1515/ap-2017-0090

Acta Parasitologica, 2017, 62(4), 748–761; ISSN 1230-2821

DE GRUYTER

Comparative analysis of codon usage pattern and its influencing factors in *Schistosoma japonicum* and *Ascaris suum*

Gulshana A. Mazumder[1], Arif Uddin[2] and Supriyo Chakraborty[1]*

[1]Department of Biotechnology Assam University, Silchar-788011, Assam, India;
[2]Moinul Hoque Choudhury Memorial Science College Algapur, Hailakandi-788150, Assam, India

Abstract

Schistosoma japonicum and *Ascaris suum* are considered as the major parasites of human which cause various life threatening diseases such as schistomiasis and ascariasis. The codon usage bias (CUB) is known as the phenomenon of more usage of a specific codon than the other synonymous codons for an amino acid. The factors that influence the codon usage bias are mutation pressure, natural selection, gene expression, gene length, GC content, RNA stability, recombination rates, codon position etc. Here we had used various bioinformatic tools and statistical analyses to understand the compositional features, expression level and codon usage bias in the genes of these two species.After estimating the effective number of codon (ENC) in both the species, codon usage bias was found to be low and gene expression was high. The nucleobase A and T were used most often than C and G. From neutrality plot and correspondence analysis it was found that both natural selection and mutation pressure played an important role in shaping the codon usage pattern of both species. Moreover, natural selection played a major role while mutation pressure played a minor role in shaping the codon usage bias in *S. japonicum and A.suum.* This is the first report on the codon usage biology in *S. japonicum and A.suum*, and the factors influencing their codon usage bias. These results are expected to be useful for genetic engineering and evolutionary studies.

Keywords

Schistosoma japonicum, *Ascaris suum*, codon usage bias, compositional properties, expression level, neutrality plot

Introduction

The genetic code is degenerate *i.e.*, 64 codons code for 20 standard amino acids, meaning that most of the amino acids are encoded by more than one codon. The codons generally differ at the third position. The non-uniform usage of synonymous codons *i.e.*, some codons are more frequently used than others in protein coding genes is known as codon usage bias. The frequency of the usage of the synonymous codons is usually unequal within and among different organisms. It has been found that various factors such as mutational bias, selection, intron splicing, gene conversion, protein secondary structures, and DNA replication are strongly related to synonymous codon usage bias (Drummond and Wilke 2008, Kahali *et al.* 2007, Warnecke and Hurst 2007). The balance between selection and mutational bias in prokaryotes or unicellular eukaryotes also determines the codon usage (Gouy and Gautier 1982, Sharp *et al.* 2005). In multicellular eukaryotes, such as *Caenorhabditis elegans* and *Drosophila melanogaster*, it is mostly determined by the selection for translational efficiency (Stenico *et al.* 1994). The equilibria among various forces such as mutation pressure, translational selection and genetic drift are considered as important factors responsible for explaining the codon usage patterns(Shah and Gilchrist 2011). Various other factors known to influence codon usage bias within and among species include geneexpression level, GC content, RNA stability, recombination rates, gene length, codon position, and others (which may include environmental stress and population size) (Behura and Severson 2013).

In this study our objective is to analyze the codon usage bias in the nuclear genes of *S. japonicum* and *A. suum. S. japonicum* is found in China and the Philippines. It is known to cause schistomiasis which poses a great threat to human health. Schistomiasis is an infection which is caused by three known species of Schistosoma (*S. mansoni*, *S. japonicum and S. haematobium*). Among these *S. japonicum* is considered to be the most infectious(Jia Tie-Wu *et al.* 2007). Infection caused by this species causes katayama fever (Ishii

*Corresponding author: supriyoch_2008@rediffmail.com

et al. 2003). It can be treated with the usage of praziquantel, a quinolone derivative(Shuhua *et al.* 2002). Schistosomiasis infection results into fibrosis which can be evaluated by ultrasound, a non-invasive method (Carlton *et al.* 2010). The life cycle of *Schistosoma japonicum* starts when the eggs of the parasite are released with the feces and when they come in contact with water they hatch into miracidia (free-swimming larva). This larva afterwards has to infect a snail of the genus *Oncomelania* (such as *Oncomelaniahupensis)* within one or two days. Here inside the snail the larva undergoes asexual reproduction and passes through a series of stages known as sporocysts. After the asexual reproduction stage the cercaria are generatedin large quantities.These are free-swimming larvaewhich leave (shed into the environment) the snail and then infect a vertebrate host. When the cercarium penetrates the skin of itstarget host it loses its tail and becomes a schistosomule. The worms then migrate through the blood circulation of the host and end at the mesenteric veins where they mate and lay eggs. Each pair laysabout 1500–3500 eggs per day. Through the tissues the eggs infiltrate and pass to the feces. Schistosome eggs are used in various ways. They can be used to perform the circumoval test for bilharziasis and also in the preparation of antigens for the intradermal test (Oliver-González *et al.* 1955).

A. suum (lives in small intestine of pigs) is also a type of helminth, and one of the species known as *A. suum lumbricoides* (lives in small intestine of human), affects humans and causes the disease ascariasis. This may be followed by symptoms of abdominal swelling, pain and diarrhea (Dold and Holland 2011). According to Liu GH *et al.* (2012), *A. lumbricoides* and *A. suum* are closely related. Their mt genome data provide evidence that *A. lumbricoides* and *A. suum* may represent the same species and they also have a nucleotide composition high in A and T (71.7% for *A. lumbricoides* and 71.8% for *A. suum*) (Liu Guo-Hua *et al.* 2012). ts life cycle starts in stools when the first appearance of eggs occursin 60–70 days. In the larval ascariasis, symptoms usually occur 4–16 days after infection. The final symptoms appear as gastrointestinal discomfort, colic and vomiting, fever, and also the presence of live worms in stools. Obstruction of intestine may be caused by a bolus of worms and the migrating larvae may cause pneumonitis and eosinophilia. Adult worms have a life-span of 1–2 years which indicates that the target individual may be infected all its life as worms die and new worms are acquired (Anderson *et al.* 1992). Eggs can survive for 15 years and a single worm is known to produce 200 thousand eggs per day and they maintain their position by swimming against the intestinal flow (Crompton *et al.* 1985). For the prevention of ascariasis, broad-spectrum benzimidazoles such as mebendazole and albendazole are the drug of choice which are recommended by WHO. According to Halina W (2015), genomic and proteomic technologies provided various opportunities to discover and improve DNA vaccines which are found to be stable at room temperatures. But poor immunogenicity is found to be the main limitation, thus it became necessary to couple the antigens with adjuvant molecules (Wedrychowicz 2015).

Codon usage bias is a well-established tool for exploring the molecular biology, genetics and evolutionary relationship of an organism. We have selected these two species namely *S. japonicum* and *A. suum* because these two parasites cause serious threat to human being and that the work on codon usage bias of the genes of these two parasites is not yet done. In our current study, wehave investigated nucleotide composition, expression level and codon usage bias of nuclear genes of these two parasites. Also, this study would give us insightsinto the various factors which may influence the codon usage pattern in the species under study.

Materials and Methods

Retrieval of Sequences

The complete coding sequences (cds) of nuclear genes for *S. japonicum* and *A. suum* were retrieved from the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/Genbank/). The species with their accession numbers are presented in **S1 and S2.** We have selected only those cds which are exact multiple of three with correct start and stop codon, and without any unknown base (N).

Relative synonymous codon usage (RSCU)

RSCU values close to 1 indicate that the synonymous codons for an amino acid are used equally, and RSCU values >1.0 indicate a strong bias for the corresponding codons. RSCU was estimated as:

$$RSCUij = \frac{X_{ij}}{\frac{1}{ni}\sum_{j=1}^{ni} X_{ij}}$$

Where Xij is the observed number of the 'i'th codon for the 'j'th amino acid, 'ni' is the total number of synonymous codons that encode the 'j'th amino acid.

Effective number of codons (ENC)

For quantifying the codon usage bias at an amino acid 'a' in standard genetic code, the homozygosity of codon usage (Fa) was calculated according to Wright (1990) as:

$$F_a = \left(na\sum_{i=1}^{k} p_i^2 - 1 \right) / (n_a - 1)$$

Here, *na* is the observed number of codons for the particular amino acid in coding sequence, pi is the frequency of 'i' th codon and k is the number of synonymous codons for the amino acid (i.e., degeneracy level).

The average of the Fa for each redundancy class (2-fold, 3-fold, 4-fold and 6-fold) was then computed as:

$$\overline{F}_r = \frac{1}{nRC}\sum_{a \in RC} F_a$$

Where nRC is the number of amino acids in the RC redundancy class. Finally, observed ENC was computed as follows:

$$Nc = 2 + (9/\overline{F2}) + (1/\overline{F3}) + (5/\overline{F4}) + (3/\overline{F6})$$

Gravy

Gravy score is the average hydrophobicity and hydrophilicity of proteins. Hydrophobicity score i.e., arbitrary unit below 0 are found to be globular (hydrophilic protein), while scores above 0 are membranous (hydrophobic).

Neutrality Plot

In the neutrality plot, GC12 was used as the ordinate and GC3 as abscissa. An independent gene was represented by each dot. If all the points of the plot show narrow GC3 distribution, it suggests low mutation bias or high conservation of GC content. On the other hand, if the curve of the neutrality plot tends to be parallel to the horizontal axis or sloped, it suggests low-correlation between GC12 and GC3. The neutrality plot is known to measure the degree of neutrality when selection pressure plays a prime role in evolution(Sueoka 1988, 1999).

Correspondence analysis (COA)

Correspondence analysis is a statistical technique used for studying the major trends in codon usage variation in coding sequences and it distributes the codons in axes with these trends (Greenacre 1984, Shields and Sharp 1987).

Box plot

Box plot was created by John W. Tukey. It is a way to display the distribution of data based on the five number summaries: minimum, first quartile, median, third quartile and maximum. Segment inside the rectangle shows the median and whiskers above and below the box show the locations of the minimum and maximum.

Statistical analysis

Correlation analysis was performed to estimate the relationship between the overall nucleotide compositions and the nucleotide compositions at 3rd codon position. Further, correlation coefficient of ENC was estimated with GC, GC3 and between GC12 and GC3. All the statistical analyses were done using the SPSS software, XL STAT and PAST.

Results

Codon usage bias of *A. suum* and *S. japonicum*

The effective number of codons (ENC) was used to determine the degree of codon usage bias in nuclear genes of *A. suum* and *S. japonicum*. In *A. suum* and *S. japonicum*, the mean ENC values were 59.13 and 54.03 respectively. The ENC values in these species were higher which suggest that codon usage bias was low in both the species.The expression level of nuclear gene in *S. japonicum* was found to be higher than *A. suum*. Jia *et al.* also found high expression level in genes of *B. mori* (Jia Wenli and Higgs 2008).

Nucleotide composition analysis in nuclear genes of *A. suum* and *S. japonicum*

It was earlier reported that the codon usage pattern of a gene might be affected by the nucleotide composition (Knight *et al.* 2001). Thus, we analyzed the compositional properties of nuclear genes in *A. suum* and *S. japonicum*. The composition of overall nucleotide and composition of nucleotide at the 3rd position of codons confirm that compositional constraints influenced the codon usage pattern in these species. From Fig 1, it was found that the distribution of A, T, G, and C%among the codons was unequal in the species *S. japonicum* and *A. suum*. The overall GC content was lower than AT contents in both the species *i.e.*the genes were AT rich. The overall GC% was the highest in *A. suum* followed by *S. japonicum*. In *S. japonicum* the greatest difference of GC content was found between the 1st position and the 3rd codon

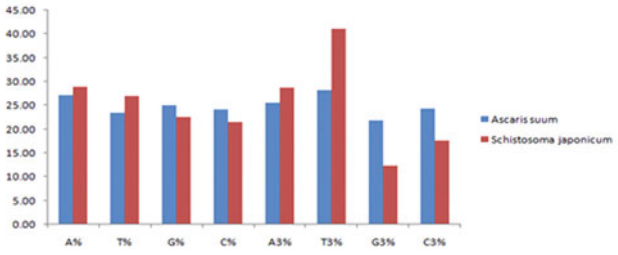


Fig. 1. Comparison of nucleotide composition and its 3rd codon position in *A. suum* and *S. japonicum*

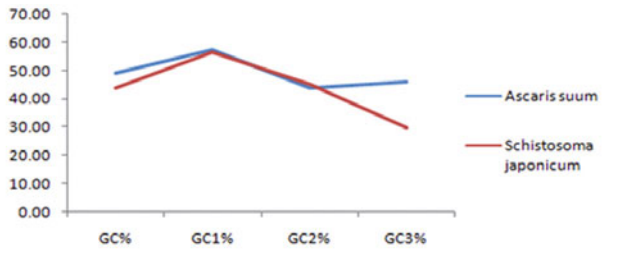


Fig. 2. Comparison of overall GC content and at its 1st 2nd and 3rd position of *A. suum* and *S. japonicum*

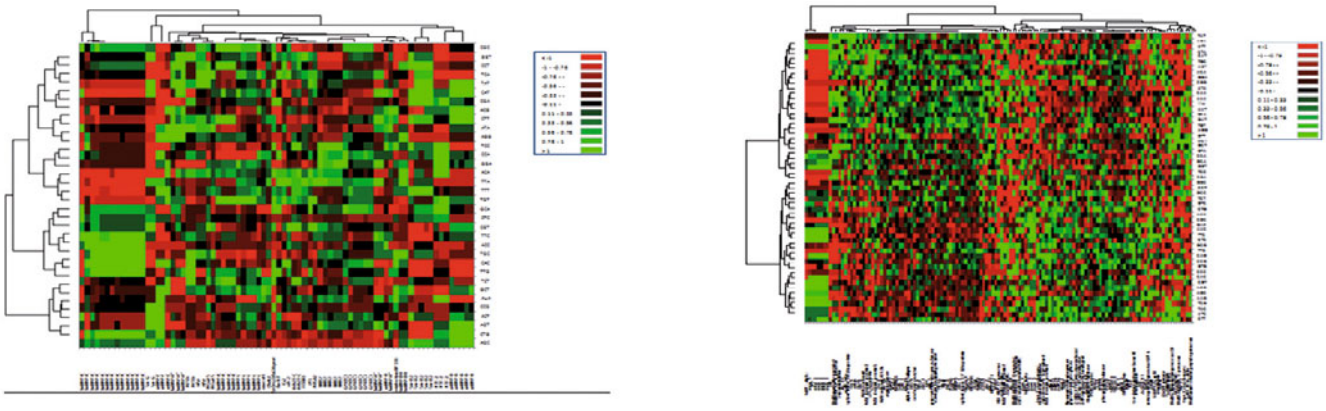


Fig. 3. Hierarchial clustering of heat map for nuclear genes in (a) *S. japonicum* and (b) *A. suum*. Each rectangular box on the map represents the RSCU value of a codon (shown in rows) corresponding to different species (shown in columns). The color and the degree of intensity represent the RSCU value

position while in *A. suum* the greatest difference of GC content was found between the 1st position and the 2ndcodon position as shown in Fig 2.

Pattern of codon usage

Patterns of synonymous codon usage in the coding sequences of *S. japonicum* and *A. suum* were assessed by RSCU analysis. In *S. japonicum*, 29 codons were used most frequently and the nucleobase T at the 3rd position occurred more frequently whereas, in *A. suum* 28 codons were used most frequently and the nucleobases A and T at the 3rd position occurred most frequently. Most abundantly used codons were A/T ended (Table I). Analysis of over-representation of codons showed that 8 codons had RSCU values> 1.6 in *S. japonicum*, and only one codon in *A. suum* (Fig 3). From the nucleotide composition and RSCU analyses it was found that selection of the preferred codons was mostly influenced by compositional constraints (*i.e.*, A and T), which accounts for the presence of mutation pressure. Comparison of RSCU values in *S. japonicum* and *A. suum* which were greater than 1 i.e. the most frequently used codons were depicted in the
Fig. 4.

Interrelationships among different compositional features

The factors namely mutation pressure and natural selection are known to affect the codon usage pattern in a species. The whole genome can be affected by the mutational pressure, and it accounts for the majority of codon usage.We performed correlation analysis between overall nucleotide composition and nucleotide composition at 3rd codon position to determine whether the evolutionary process was driven alone by the mutational pressure or by both mutation pressure and natural selection. In both *S. japonicum* and *A. suum* highly significant positive correlation was found almost among overall nucleotide composition and its composition at 3rd codon position as shown in Table II. These results suggest the compositional

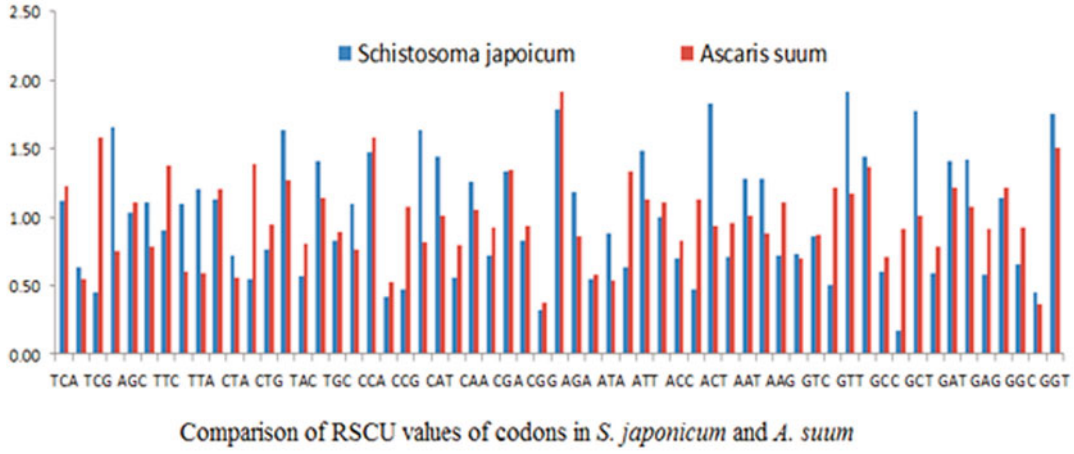


Fig. 4. Comparison of RSCU values of codons in *S. japonicum* and *A. suum*, where RSCU value >1.6 indicates over-represented codons and RSCU <0.6 indicates under-represented codons

Table I. RSCU values of codons in *A. suum* and *S. japonicum*

RSCU	*A. suum*	*S. japonicum*
TCA	1.23	1.12
TCC	0.54	0.64
TCG	1.59	0.46
TCT	0.76	1.66
AGC	1.11	1.03
AGT	0.78	1.10
TTC	1.38	0.90
TTT	0.61	1.10
TTA	0.59	1.21
TTG	1.21	1.13
CTA	0.56	0.73
CTC	1.38	0.54
CTG	0.95	0.76
CTT	1.27	1.63
TAC	0.80	0.57
TAT	1.14	1.41
TGC	0.90	0.83
TGT	0.77	1.09
CCA	1.59	1.48
CCC	0.53	0.41
CCG	1.07	0.47
CCT	0.82	1.63
CAT	1.01	1.44
CAC	0.79	0.56
CAA	1.05	1.26
CAG	0.92	0.72
CGA	1.35	1.33
CGC	0.94	0.83
CGG	0.37	0.33
CGT	1.92	1.79
AGA	0.86	1.18
AGG	0.58	0.54
ATA	0.54	0.88
ATC	1.33	0.64
ATT	1.13	1.49
ACA	1.11	1.00
ACC	0.82	0.70
ACG	1.13	0.48
ACT	0.94	1.83
AAC	0.96	0.71
AAT	1.02	1.29
AAA	0.88	1.28
AAG	1.11	0.72
GTA	0.70	0.73
GTC	0.87	0.86
GTG	1.22	0.50
GTT	1.18	1.91
GCA	1.37	1.44
GCC	0.70	0.60
GCG	0.91	0.18
GCT	1.02	1.78
GAC	0.78	0.60
GAT	1.22	1.41
GAA	1.07	1.42
GAG	0.91	0.58
GGA	1.21	1.14
GGC	0.92	0.65
GGG	0.36	0.46
GGT	1.51	1.75

constraint arising from mutation pressure and natural selection determine the pattern of codon usage in these two species. Furthermore, significant correlation between ENC and various GC contents as shown in Table III, suggests that the nucleotide composition under natural selection and mutational pressure affects the synonymous codon usage in nuclear genes of *S. japonicum* and *A. suum.* If mutation pressure governs the pattern of synonymous codon usage then the frequency of nucleotide A and T should be equal to that of G and C at synonymous 3rd codon position. But the frequencies of these nucleotides were not same in *S. japonicum* and *A. suum* which indicate that other factor such as natural selection might have played a role in codon usage pattern.

Correspondence analysis in the nuclear gene of *S. japonicum* and *A. suum*

Correspondence analysis (COA) is a statistical and multivariate analytical techniquewhich can explore the relationship between variables in the contingency table. It can be used for investigating the variation in RSCU values among genes(Greenacre 1984).It is widely used for studying the major trends insequence variation and for distributing the coding sequences along these trends. In *S. japonicum*, (Fig. 5) it was revealed that the 1st axis contributed 15.22% and the 2nd axis contributed 14.09% of the total variation, while in *A .suum* the 1st axis contributed 14.69% and the 2nd axis contributed 7.65% of the total variation. This suggests the 1st axis as the major contributor in synonymouscodon usage pattern. Most of the points (codons positions) are found to be closer to the axeswhich indicate that compositional constraintunder mutation might correlate to the codon usage pattern in the nuclear gene. Butsome other codons are found in scattered distribution which indicates that other factors like natural selection might also affect the codon usage pattern.

Neutrality Plot

In different organisms the major factors that contribute to codon usage bias are considered to be mutation bias and translation selection. To analyze the influences of mutation bias and translation selection on codon usage, we had drawn neutrality plots (GC12 vs GC3). Mutational bias is considered to be the main force in shaping codon usage when the correlation between GC12 and GC3 is statistically significant and the slope of the regression line is close to 1. On the otherhand, if selection is the major factor, then the slope of the regression

Table II. Correlation between overall nucleotide composition (%) and its 3rd codon position in *A. suum* and *S. japonicum*

A. suum					
Nucleotide (%)	A3%	T3%	G3%	C3%	GC3%
A%	0.486**	0.003	–0.261**	–0.277**	–0.339**
T%	0.049	0.732**	–0.443**	–0.326**	–0.486**
G%	–0.350**	–0.327**	0.496**	0.197*	0.444**
C%	–0.240**	–0.459**	0.298**	0.424**	0.449**
GC%	–0.352**	–0.493**	0.469**	0.401**	0.548**
S. japonicum					
Nucleotide (%)	A3%	T3%	G3%	C3%	GC3%
A%	0.625**	0.081	–0.300**	–0.503**	–0.535**
T%	0.192	0.575**	–0.335**	–0.509**	–0.556**
G%	–0.589**	–0.380**	0.430**	0.660**	0.719**
C%	0.019	–0.078	0.005	0.053	0.043
GC%	–0.467**	–0.346**	0.352**	0.559**	0.603**

**p<0.01

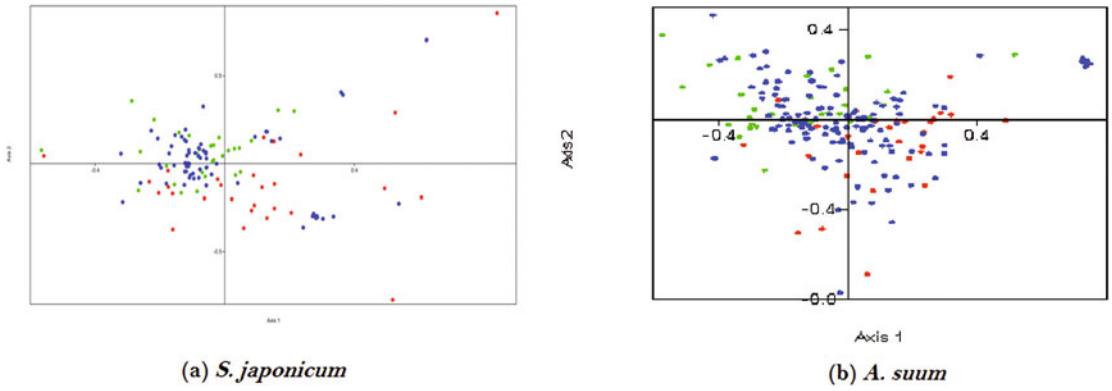


Fig. 5. Correspondence analysis of codon usage patterns of nuclear gene in (a) *S. japonicum* and (b) *A. suum*

Table III. Correlation coefficients between ENC and different GC contents

Correlation between	***S. japonicum***	***A. suum***
ENC and GC	0.597**	0.445**
ENC and GC1	0.441**	0.290**
ENC and GC2	0.377**	0.043
ENC and GC3	0.995**	0.641**

**p<0.01

line is close to 0. The results revealed significant correlations between GC12 and GC3 (Fig 6). The slope of the regression line in *S. japonicum* was 0.841, indicating relative neutrality of 84.1% which indicates the role ofmutation pressure. In case of *A. suum*, the slope of the regression line was 0.094 indicating relative neutrality of 9.4%,this hints the role of natural selection.

To test the linearity of the regression model, we performed the residual analysis (Fig 7). For both the datasets the residual

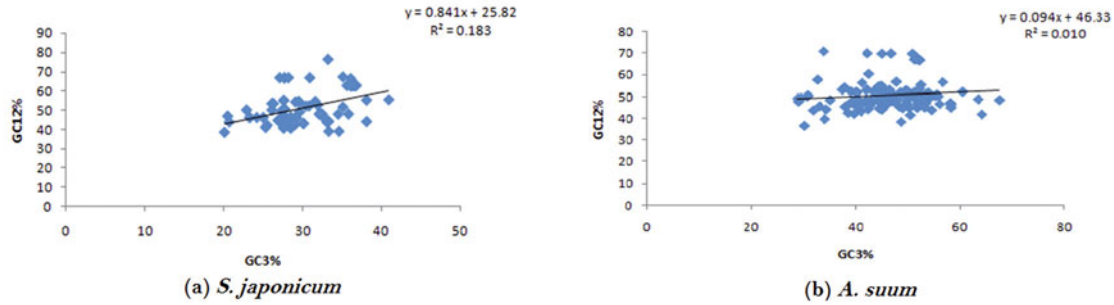
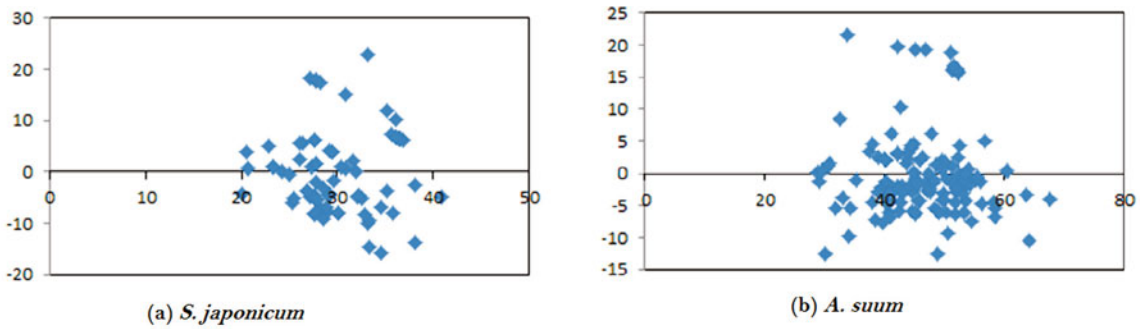


Fig. 6. Neutrality plot in (a) *S. japonicum* and (b) *A. suum*



(a) *S. japonicum* (b) *A. suum*

Fig. 7. Residual plot for (a) *S. japonicum* and (b) *A. suum*

Table IV. Correlation between ENC and various skews in *S. japonicum* and *A. suum*

	S. japonicum	*A. suum*
GC (r)	0.632**	–0.064
AT (r)	–0.006	–0.015
Purine (r)	–0.682**	–0.431**
Pyrimidine (r)	–0.344**	–0.352**
Amino (r)	–0.692**	–0.377**
Keto (r)	–0.318**	–0.405**

plots were in scattered form which indicates that the linear regression model explains the relationship between GC12 and GC3 for both the species.

Correlation between codon usage and skewness

It was found that the GC skews, AT skews, purine skews, pyrimidine skews, amino skews and keto skews in both the species *S. japonicum* and *A. suum* were all negative except that of GC skew in *S. japonicum* as shown in Table 4. These suggest asymmetrical nucleotide composition between the two strands of DNA. In *S. japonicum* significant negative correlation of ENC was found withpurine skew (–0.682**), pyrimidine skew (–0.344**), amino skew (–0.692**) and keto skew (–0.318**) respectively, therefore in *S. japonicum* purine skew, pyrimidine skew, amino skew and keto skew significantly affect the codon usage pattern.

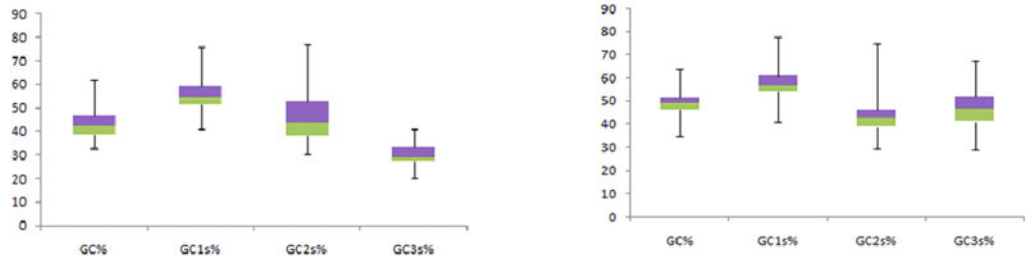
In *A. suum*, significant correlation of ENC was found withpurine skew (–0.431**), pyrimidine skew (–0.352**), amino skew (–0.377**) and keto skew (–0.405**) **(Table 4).**This suggests that in *A. suum*, the purine skew, pyrimidine skew, amino skew and keto skew significantly affect the codon usage.

Relationship between codon bias and hydropathicity index (Gravy) and aromaticity score (Aromo)

From earlier studies it was found that the hydropathicity and aromaticity of encoded protein used to play an important role in shaping codon usage of some species (Chen *et al.* 2013). Here we conducted a correlation analysis between codon usage bias and hydropathicity index and aromaticity score values in both *S. japonicum* and *A. suum*. The correlation coefficients for the Gravy and Aromo scores in *S. japonicum* were (r = –0.133; r = 0.051 respectively) whereas, in *A. suum* (r = –0.088; r = –0.090 respectively) which strongly indicate that in both *S. japonicum* and *A. suum* the hydropathicity and aromaticity of the encoded proteins were associated with the codon usage bias.

Box plot analysis

It is an exploratory graphic which is used to identify outliers and to compare the distributions. It is used for showing overall patterns of a group. It helps visualize the range for a large group of genes.



(a) *S. japonicum* (b) *A. suum*

Fig. 8. Box plot analysis between GC, GC1, GC2 and GC3 in (a) *S. japonicum* and (b) *A. suum*

From Fig. 8, we found that in both *A. suum* and *S. japonicum*, the overall GC content and GC content at the 1st, 2nd and 3rd position were at high level of agreement with each other while the presence of one larger box plot indicates that there might be some difference between the groups.

Discussion

Codon usage bias is a molecular tool for the analysis of genetic information and paves the way for understanding the evolutionary relationship of an organism. Most of the studies of codon usage bias have been reported on higher organisms, microorganisms, viruses like hepatitis virus, zika virus etc, parasites namely *W.bancrofti* and *S. haematobium* butno work was reported earlier on the codon usage bias of *S. japonicum* and *A. suum.* This study would contribute towards elucidating the genetic and molecular evolution of these two parasites. This study would also highlight the codon usage variation and the factors influencing the codon usage bias in *S. japonicum* and *A. suum.*

ENC is widely used to measure the degree of codon usage bias of a gene. Its value differs from species to species and its elevated value indicates low codon bias and vice-versa. In *A. suum* and *S. japonicum*, the mean ENC values were found to be 59.13 and 54.03 respectively, which are considered to be high and thus it indicates low codon usage bias. The low codon usage bias has also been observed in some other parasites, such as *Plasmodium vivax* (ENC: 55.54), *Plasmodium knowlesi* (ENC: 55.28) (Yadav and Swati 2012). Liu Xudong *et al.* found that the codon usage bias was low in H9N2 virus (Liu Xudong *et al.* 2010). Zhang *et al.* reported that the ENC values of TTSuV2 varied from 55.20 to 58.18 with a mean value of 56.21 which revealed that codon usage bias was also not remarkable in TTSuV2 genomes (Zhang *et al.* 2013). Jenkins *et al.* added that for efficient replication in vertebrates the low codon usage bias might be very helpful (Jenkins and Holmes 2003).Also in *Taeniasaginata*, the low ENC value was observed which indicates strong codon usage bias (Yang *et al.* 2014). But high ENC value was observed in *Wuchereria bancrofti* and *Schistosoma haematobium* (Mazumder *et al.* 2016a).

Several previous studies have shown that gene expression and ENC are inversely correlated with each other *i.e.* lower ENC value indicates higher codon usage preference and higher gene expression and vice-versa. Significant correlation between ENC and various GC contents in *S. japonicum* and *A. suum* suggests that the nucleotide composition under natural selection and mutation pressure affect the synonymous codon usage in these two species.

In our study we found that the overall GC% was higher in *S. japonicum* followed by *A. suum* and the average GC content was lower than AT content in both species. The average GC content in *S. haematobium* was 36.37%, and 43.41% in *W. Bancrofti* which revealed that the genes are AT rich (Mazumder *et al.* 2016b). In *T. saginata* the GC content of the genes varied from 31 to 80.2% (Yang *et al.* 2014). From nucleotide composition analysis it was shown that A and T nucleotides constitute the majority of overall nucleotide composition in both *S. japonicum* and *A. suum*. It was also supported from the study of codon usage in *Ascaris* that their genome had more A and T than that of G and C (Fadiel *et al.* 2002). According to box plot analysis,the overall GC content and GC content at the 1st, 2nd and 3rd position were found to be at high level of agreement with each other but it also indicates there might be some difference between the groups.

The RSCU analysis revealed that the codon usage bias was toward A- and T- ended codons. The pattern of codon usage varies among different genes and also varies between two species supporting our earlier studies.In mitochondrial genome of some species of platyhelminthes, it had also been found that T-ending codons were predominant, which also suggests AT richness at the third codon position (Chen *et al.* 2013).

We found significant correlation between overall nucleotide composition and its composition at 3rd codon position in both the species. Further, we also found significant correlation between ENC and various GC contents in *S. japonicum* and *A. suum* which suggests that the nucleotide composition was under natural selection and that mutational pressure affects the synonymous codon usage supporting the result of Zhang *et.al* (Zhang *et al.* 2013). Chen *et.al* 2013 found the compositional constraint as one of the major factors which shapes the codon usage in the mitochondrial genomes of some species of platyhelminthes (Chen *et al.* 2013). Codon usage bias examined in *Schistosoma mansoni* revealed that codon bias had been dependent on the overall base composition of the genes (Ellis and Morrison 1995). In different organisms, translational selection and compositional constraints are suggested to be the major factors for codon usage variation between genes (Karlin and Mrázek 1996).

Among various influencing factors, mutational pressure and natural selection are considered to be the two major factors that shape codon usage patterns. To determine the share of each factor on the codon usage patterns of both the species, the neutrality plot analysis was performed which showed that the influence of natural selection dominates over mutation pressure in both species. Similar result was also found in *S.haematobium*, *W.bancrofti*, *T. solium*, *T. saginata*(Yang *et al.* 2014) and *B.mori* (Jia Xian *et al.* 2015).

The genetic and chemical features of the double-stranded DNA molecule may be usually implied by its specific structure. According to Chargaff's rule, A = T and G = C in a DNA duplex. Now if we assume that there are no mutational pressure or selection pressure to influence the composition of the two DNA strands, the Chargaff's rules should be in force for each of the two strands and not only for double-stranded DNA. These rules are known as parity rules type 2 (PR2) (Lobry 1995). Deviation from this PR2 leads to asymmetric

substitution patterns and DNA asymmetry. The asymmetry is observed even at the level of codons and amino acids (Lafay *et al.* 1999, Mackiewicz *et al.* 1999, McInerney 1998, Perrière *et al.* 1996, Rocha *et al.* 1999, Romero *et al.* 2000). Deviations from PR2 are usually analyzed in terms of an excess of the number of guanines relative to cytosines or adenines relative to thymines and the bias is measured by GC and AT skews, (G-C)/(G+C) and (A-T)/(A+T), respectively. Thusin the present study we found that in both the species the purine skew, pyrimidine skew, amino skew and keto skew significantly affect the codons usage. Uddin and Chakraborty (2014) also found in their study that AT skew was positive but GC skew was negative which suggests that in most species C was abundant over G while in some others A was abundant over T and also there was asymmetry in nucleotide composition between the two DNA strands (Uddin and Chakraborty 2014) . In the present study we found that in *A. suum* the GC skew and the AT skew were negative but in *S. japonicum* GC skew and AT skew were positive. Wan *et al* also showed that at the third codon position,the asymmetric distribution of G over C and A over T may increase codon usage bias (Wan *et al.* 2004)**.**

Conclusion

The present study of codon bias in *Schistosoma* and *Ascaris* species revealed that most of the frequent codons end with A or T at the 3rd codon position and that the expression level of the genes was high.It was also found that factors such as natural selection and mutation pressure played an important role in shaping the codon usage pattern in the nuclear genes of these species.This study enhanced our understanding of the mechanisms underlying the codon usage and the various factors influencing it.

Conflict of interest

There is no conflict of interest in this research work.

Acknowledgements. We are thankful to Assam University, Silchar, Assam for providing the necessary facilities for this research work.

Reference

Anderson R.M., May R.M., Anderson B. 1992. Infectious diseases of humans: dynamics and control: *Wiley Online Library*, 16, 208–2012. DOI: 10.1111/j.1753-6405.1992.tb00056.x

Behura S.K., Severson D.W. 2013. Codon usage bias: causative factors, quantification methods and genome wide patterns: with emphasis on insect genomes. *Biological Reviews*, 88, 49–61.DOI: 10.1111/j.1469-185X.2012.00242.x

Carlton E.J., Hsiang M., Zhang Y., Johnson S., Hubbard A., Spear R.C. 2010. The impact of Schistosoma japonicum infection and treatment on ultrasound-detectable morbidity: a five-year cohort study in Southwest China. *PLoS Negl Trop Dis*, 4, e685. DOI: https://doi.org/10.1371/journal.pntd.0000685

Chen L., Yang D., Liu T., Nong X., Huang X., Xie Y., *et al*. 2013. Synonymous codon usage patterns in different parasitic platyhelminth mitochondrial genomes. *Genetics and Molecular Research*, 12, 587–596. DOI: http://dx.doi.org/10.4238/2013.February.27.8

Dold C., Holland C.V. 2011. Ascaris and ascariasis. *Microbes and Infection*, 13, 632–637. DOI: https://doi.org/10.1016/j.micinf.2010.09.012

Drummond D.A., Wilke C.O. 2008. Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. *Cell*, 134, 341–352. DOI: https://doi.org/10.1016/j.cell.2008.05.042

Ellis J, Morrison D. 1995. Schistosoma mansoni: patterns of codon usage and bias. *Parasitology*, 110, 53–60. DOI: https://doi.org/10.1017/S003118200008104X

Fadiel A., Lithwick S., Gamra M. 2002. Codon usage analysis of Ascaris species influence of base and intercodon frequencies on the synonymous codon usage. *Journal of the Egyptian Society of Parasitology*, 32, 625–638. DOI: 0000-0002-5759-6891

Gouy M., Gautier C. 1982. Codon usage in bacteria: correlation with gene expressivity. *Nucleic Acids Research*, 10, 7055–7074. DOI: https://doi.org/10.1093/nar/10.22.7055

Greenacre M.J. 1984. Theory and applications of correspondence analysis

Hall A., Hewitt G., Tuffrey V., De Silva N. 2008. A review and meta analysis of the impact of intestinal worms on child growth and nutrition. *Maternal and Child Nutrition* 4, 118–236. DOI: 10.1111/j.1740-8709.2007.00127.x

Ishii A., Tsuji M., Tada I. 2003. History of Katayama disease: schistosomiasis japonica in Katayama district, Hiroshima, Japan. *Parasitology International*, 52, 313–319. DOI: https://doi.org/10.1016/S1383-5769(03)00046-1

Jenkins G.M., Holmes E.C. 2003. The extent of codon usage bias in human RNA viruses and its evolutionary origin. *Virus Research*, 92, 1–7. DOI: https://doi.org/10.1016/S0168-1702(02)00309-X

Jia T-W., Zhou X-N., Wang X-H., Utzinger J., Steinmann P., Wu X-H. 2007. Assessment of the age-specific disability weight of chronic schistosomiasis japonica. *Bulletin of the World Health Organization*, 85, 458–465. DOI: http://dx.doi.org/10.1590/S0042-96862007000600012

Jia W., Higgs P.G. 2008. Codon usage in mitochondrial genomes: distinguishing context-dependent mutation from translational selection. *Molecular Biology and Evolution*, 25, 339–351. DOI: https://doi.org/10.1093/molbev/msm259

Jia X., Liu S., Zheng H., Li B., Qi Q., Wei L., *et al*. 2015. Non-uniqueness of factors constraint on the codon usage in Bombyx mori. *BMC genomics*, 16, 356. DOI: 10.1186/s12864-015-1596-z

Kahali B., Basak S., Ghosh T.C. 2007. Reinvestigating the codon and amino acid usage of S. cerevisiae genome: a new insight from protein secondary structure analysis. *Biochemical and biophysical research communications*, 354, 693–699. DOI: https://doi.org/10.1016/j.bbrc.2007.01.038

Karlin S., Mrázek J. 1996. What drives codon choices in human genes? *Journal of Molecular Biology*, 262, 459–472. DOI: https://doi.org/10.1006/jmbi.1996.0528

Knight R.D., Freeland S.J., Landweber L.F. 2001. A simple model based on mutation and selection explains trends in codon and amino-acid usage and GC composition within and across genomes. *Genome Biology*, 2, 1. DOI: 10.1186/gb-2001-2-4-research0010

Lafay B, Sharp PM, Lloyd AT, McLean MJ, Devine KM, Wolfe KH. 1999. Proteome composition and codon usage in spirochaetes: species-specific and DNA strand-specific mutational biases. *Nucleic acids research*, 27, 1642–1649. DOI: https://doi.org/10.1093/nar/27.7.1642

Liu G-H., Wu C-Y., Song H-Q., Wei S-J., Xu M-J., Lin R-Q., *et al*. 2012. Comparative analyses of the complete mitochondrial genomes of Ascaris lumbricoides and Ascaris suum from humans and pigs. *Gene*, 492, 110–116. DOI: https://doi.org/10.1016/j.gene.2011.10.043

Liu X., Wu C., Chen AY-H. 2010. Codon usage bias and recombination events for neuraminidase and hemagglutinin genes in Chinese isolates of influenza A virus subtype H9N2. *Archives of Virology*, 155, 685–693. DOI: https://doi.org/10.1007/s00705-010-0631-2

Lobry J. 1995. Properties of a general model of DNA evolution under no-strand-bias conditions. *Journal of Molecular Evolution*, 40, 326–330. DOI: https://doi.org/10.1007/BF00163237

Mackiewicz P., Gierlik A., Kowalczuk M., Dudek M.R., Cebrat S. 1999. How does replication-associated mutational pressure influence amino acid composition of proteins? *Genome Research*, 9, 409–416. DOI: 10.1101/gr.9.5.409

Mazumder G., Uddin A., Chakraborty S. 2016a. Expression levels and codon usage patterns in nuclear genes of the filarial nematode Wucheraria bancrofti and the blood fluke Schistosoma haematobium. *Journal of Helminthology*, 91, 72–79. DOI: https://doi.org/10.1017/S0022149X16000092

McInerney J.O. 1998. Replicational and transcriptional selection on codon usage in Borrelia burgdorferi. *Proceedings of the National Academy of Sciences*, 95, 10698–10703. PMCID: PMC27958

Oliver-González J., Bauman P.M., Benenson A. 1955. Immunological aspects of infections with Schisiosoma mansoni. *American Journal of Tropical Medicine and Hygiene*, 4, 443–452. DOI: https://doi.org/10.4269/ajtmh.1955.4.443

Perrière G., Lobry J.R, Thioulouse J. 1996. Correspondence discriminant analysis: a multivariate method for comparing classes of protein and nucleic acid sequences. *Computer Applications in the Biosciences:CABIOS*, 12, 519–524. DOI: https://doi.org/10.1093/bioinformatics/12.6.519

Rocha E.P., Danchin A., Viari A. 1999. Universal replication biases in bacteria. *Molecular Microbiology*, 32, 11–16. DOI: 10.1046/j.1365-2958.1999.01334.x

Romero H, Zavala A, Musto H. 2000. Codon usage in Chlamydia trachomatis is the result of strand-specific mutational biases and a complex pattern of selective forces. *Nucleic acids research* 28: 2084–2090. DOI: https://doi.org/10.1093/nar/28.10.2084

Shah P., Gilchrist M.A. 2011. Explaining complex codon usage patterns with selection for translational efficiency, mutation bias, and genetic drift. *Proceedings of the National Academy of Sciences*, 108, 10231–10236. DOI: 10.1073/pnas.1016719108

Sharp P.M., Bailes E., Grocock R.J., Peden J.F., Sockett R.E. 2005. Variation in the strength of selected codon usage bias among bacteria. *Nucleic Acids Research*, 33, 1141–1153. DOI: https://doi.org/10.1093/nar/gki242

Shields D.C., Sharp P.M. 1987. Synonymous codon usage in Bacillus subtilis reflects both translational selection and mutational biases. *Nucleic Acids Research*, 15, 8023–8040. DOI: https://doi.org/10.1093/nar/15.19.8023

Shuhua X., Tanner M., N'Goran E.K., Utzinger J., Chollet J., Bergquist R. *et al*. 2002. Recent investigations of artemether, a novel agent for the prevention of schistosomiasis japonica, mansoni and haematobia. *Acta Tropica*, 82, 175–181. DOI: https://doi.org/10.1016/S0001-706X(02)00009-8

Stenico M., Lloyd A.T., Sharp P.M. 1994. Codon usage in Caenorhabditis elegans: delineation of translational selection and mutational biases. *Nucleic Acids Research*, 22, 2437–2446. DOI: https://doi.org/10.1093/nar/22.13.2437

Sueoka N. 1988. Directional mutation pressure and neutral molecular evolution. *Proceedings of the National Academy of Sciences*, 85, 2653–2657. PMID:3357886

Sueoka N. 1999. Two aspects of DNA base composition: G+ C content and translation-coupled deviation from intra-strand rule of A = T and G = C. *Journal of Molecular Evolution*, 49, 49–62. DOI: https://doi.org/10.1007/PL00006534

Uddin A., Chakraborty S. 2014. Mutation Pressure Dictates Codon Usage Pattern in Mitochondrial Atpase8 in Some Mammalian Species. *International Journal Scientific Research* 3, 2206–2212

Wan X-F., Xu D., Kleinhofs A., Zhou J. 2004. Quantitative relationship between synonymous codon usage bias and GC composition across unicellular genomes. *BMC Evolutionary Biology*, 4, 19. DOI: 10.1186/1471-2148-4-19

Warnecke T., Hurst L.D. 2007. Evidence for a trade-off between translational efficiency and splicing regulation in determining synonymous codon usage in Drosophila melanogaster. *Molecular Biology and Evolution*, 24, 2755–2762. DOI: https://doi.org/10.1093/molbev/msm210

Wedrychowicz H. 2015. Antiparasitic DNA vaccines in 21st century. *Acta Parasitologica*, 60, 179–189. DOI: https://doi.org/10.1515/ap-2015-0026

Yadav M.K., Swati D. 2012. Comparative genome analysis of six malarial parasites using codon usage bias based tools. *Bioinformation*, 8, 1230–1239. DOI: 10.6026/97320630081230

Yang X., Luo X., Cai X. 2014. Analysis of codon usage pattern in Taenia saginata based on a transcriptome dataset. *Parasites & Vectors*, 7, 1. DOI: 10.1186/s13071-014-0527-1

Zhang Z., Dai W., Dai D. 2013. Synonymous codon usage in TTSuV2: analysis and comparison with TTSuV1. *PloS One*, 8, e81469. DOI: https://doi.org/10.1371/journal.pone.0081469

Received: February 4**,** 2017
Revised: July 20, 2017
Accepted for publication: July 21, 2017

Supplementary Tables

S1. Gene name, gene length and accession no. of *A. suum*

Gene Name	Accession No.
AsSLR8.97	AAB07017.1_1
ALG-5	AEF32753.1_1
iron-sulfur subunit of succinate dehydrogenase	BAA23716.1_1
P34	AAP94888.1_1
AF2 peptide	AAQ90307.1_1
MFP2	AAP94886.1_1
cytochrome b small subunit of mitochondrial fumaratereductase	BAA11233.1_1
phosphofructokinase	AAR16088.1_1
mdp-1	AAT77425.1_1
dhs-16	AEF32525.1_1
mdp-1	AAT77422.1_1
afp-10	AEX09240.1_1
hemoglobin	AAA29374.1_1
tps1	AEX60788.1_1
WAGO-2	AEF32758.1_1
neuropeptide-like protein precursor 21	AKB91125.1_1
FMRFamide-related peptide FLP-14 precursor	AAQ23190.1_1
spliced leader 30 kDa protein	AAM12418.1_1
DROSHA	AEF32762.1_1
hsp90	ACO55134.1_1
GAR-1	ACM78885.1_1
flavoprotein subunit of complex II	BAA21636.1_1
AsSLR8.60	AAB07018.1_1
pyruvate dehydrogenase kinase	AB52573.1_1
Pxy14	AAB40605.1_1
WAGO-1	AEF32757.1_1
carboxypeptidase inhibitor precursor	ABW38008.1_1
G15-6A protein	AAQ87881.1_1
thymidylate synthase	AAC97507.1_1
2-methyl branched-chain enoyl CoA reductase isoform I	AAC48316.1_1
MSP-domain protein 5	AAN08883.1_1
PASHA	AEF32763.1_1
tuf-1	BAF31894.1_1
jh1	AAN08879.2_1
eft-1	DAA05869.1_1
tps2	AEX60787.1_1
asabf-zeta	BAC57992.1_1
lipoamide dehydrogenase	AAD30450.1_1
RdRP3	AEF32766.1_1
Eif4g	ACX37244.1_1
type-2 cytochrome c	BAA11132.1_1
MSP-domain protein 4	AAN08882.1_1
thioredoxin peroxidase	BAA90476.1_1
C-type lectin	ADM49197.1_1
SDHA2	BAB84191.1_1
AsEP36	BAB68543.1_1
afp-2	AEX09239.1_1

Ascec-4	BAD89088.1_1
neuropeptide precursor AFP-6	AAU10528.1_1
mdp-1	AAT77426.1_1
AsMIF	BAD24819.1_1
SDHD2	BAB84192.1_1
fatty acid binding protein homolog As-p18	AAA98565.1_1
enol-1	ADQ00605.1_1
srp-1	AEH42098.1_1
CSR-1	AEF32756.1_1
MSP-domain protein 2	AAN08880.1_1
pyruvate dehydrogenase beta subunit	AAA29379.1_1
As22	BAD42839.1_1
pyruvate dehydrogenase	AAA29376.1_1
FMRFamide-related peptide FLP-8 precursor	AAQ23188.1_1
pepsin inhibitor	CAA12072.1_1
ASABF	BAA11943.1_1
dihydrolipoyl dehydrogenase-binding protein	AAD30034.1_1
ALG-4	AEF32752.1_1
Ascec-2	BAD89086.1_1
L2R59	BAB67769.1_1
asabf-epsilon	BAC41495.1_1
ASABF-gamma	BAC00498.1_1
ASABF-delta	BAC00499.1_1
cytochrome b-large subunit	BAA11232.2_1
NRDE-3	AEF32760.1_1
myoglobin	AAA64695.1_1
NLP22	AIB09454.1_1
alpha-2 (IV) collagen	AAA18014.1_1
nicotinic acetylcholine receptor alpha subunit	ABS95448.1_1
arnt	BAJ17132.1_1
type-1 cytochrome c	BAA11131.1_1
immunosuppressive ovarian message protein	AAB53808.1_1
FMRFamide-like peptide 8	AAQ90306.1_1
try-5	AEH42099.1_1
FMRFamide-related peptide FLP-6 precursor	AAQ23187.1_1
acr-16	AKR16139.1_1
afp-11	AEX09241.1_1
scavenger decapping enzyme	ADB92583.1_1
MFP2c	AAW29197.1_1
MFP1-beta	AAP94885.1_1
MSP-domain protein 3	AAN08881.1_1
oaras	AAS55869.1_1
GAR-1	ACM78886.1_1
MFP3	AAP94887.1_1
Ascec-1	BAD89085.1_1
HIF1A	BAJ17131.1_1
2-methyl branched-chain enoyl CoA reductase	AAA16096.1_2
neuropeptide-like protein NLP-12 precursor	AAQ23186.1_1
AAC	AAD30505.1_1
serotonin receptor	AAC78396.1_1
FMRFamide-related peptide FLP-12 precursor	AAQ23189.1_1

asabf-6Cys-alpha	BAC41496.1_1
RdRP1	AEF32764.1_1
mdp-1	AAT77427.1_1
As37	BAC06575.1_1
RdRP2	AEF32765.1_1
apb-2	AAC38844.1_1
L2R37	BAC66614.1_1
spliced leader 175 kDa protein	AAM12419.1_1
translation initiation factor 4E	AAT09130.1_1
PEPCK	AAA29378.1_1
col-6	AAD17458.1_1
nicotinic acetylcholine receptor subunit ACR-26	ACZ37230.1_1
tuf-2	BAF30979.1_1
nicotinic acetylcholine receptor non-alpha subunit	ABV59997.1_1
WAGO-3	AEF32759.1_1
FMRFamide-related peptide precursor	AAW78865.1_1
slo-1	ACC68842.1_1
AFP-13	ADI99988.1_1
Ascec-3	BAD89087.1_1
ASABF-beta	BAC00497.1_1
MFP1-alpha	AAP94884.1_1
ALG-1	AEF32751.1_1
FMRFa-like peptide precursor 14	AKB91124.1_1
arginine kinase	ACO56119.1_1
MFP2b	AAP94889.1_1
ALG-6	AEF32754.1_1
trx-1	AAS78778.1_1
ATP synthase F6 family protein	ADJ68251.1_1
mdp-1	AAT77423.1_1
mdp-1	AAT77424.1_1
DCR-1	AEF32761.1_1

S2. Gene name, gene length and accession no. of *S. japonicum*

Gene Name	Accession No.
snRNP-B	CAX76474.1_1
hnRNP-L	CAX73739.1_1
hnRNP-K	CAX77156.1_1
reinfection related protein 338	AAM03145.1_1
FOXK1	CAX72988.1_1
nfyc	CAX74042.1_1
snRNP-F	CAX75481.1_1
Es2	CAX72937.1_1
Sm-D2	CAX76811.1_1
hnRNP-H’	CAX75604.1_1
hnRNP-L	CAX73599.1_1
hnRNP-K	CAX77158.1_1
hnRNP-H’	CAX75601.1_1
Sm-D2	CAX76810.1_1
GPATC3	CAX75850.1_1
hnRNP-K	CAX77151.1_1
hnRNP-K	CAX71393.1_1
hnRNP-H’	CAX75602.1_1
hnRNP-D0	CAX70184.1_1
hnRNP-H’	CAX75603.1_1
BRD2	CAX75271.1_1
hnRNP-K	CAX77149.1_1
hnRNP-K	CAX71394.1_1
snRNP-F	CAX75478.1_1
Sm-D2	CAX71287.1_1
hnRNP-K	CAX77150.1_1
onecut1	CAX73635.1_1
hnRNP-H’	CAX75600.1_1
GPATC3	CAX75847.1_1
hnRNP-K	CAX75227.1_1
Chek2	CAX69705.1_1
hnRNP-D0	CAX74648.1_1
BRD2	CAX75273.1_1
hnRNP-K	CAX77152.1_1
hnRNP-H’	CAX75599.1_1
Sm-D2	CAX76818.1_1
snRNP-B	CAX76473.1_1
nfyc	CAX74043.1_1
GPATC3	CAX75845.1_1
NTF-2	CAX75954.1_1
SJCHGC05630 protein	AAP06387.1_1
NCoA-5	CAX73999.1_1
Rack7	CAX82429.1_1
NFYA	CAX69967.1_1
snRNP-E	CAX70440.1_1
snRNP-B	CAX76476.1_1
Hrb27-C	CAX74025.1_1
Sm-G	CAX75268.1_1
hnRNP-K	CAX70457.1_1
hnRNP-K	CAX75228.1_1
Pf1	CAX73573.1_1
hnRNP-K	CAX77160.1_1
BRD2	CAX75274.1_1
NFYA	CAX74274.1_1
Sm-G	CAX75266.1_1
Sm-D2	CAX71284.1_1
eft-2	CAX73469.1_1
hnRNP-K	CAX77153.1_1
NTF-2	CAX75952.1_1
GPATC3	CAX75849.1_1
NTF-2	CAX75953.1_1
hnRNP-L	CAX73738.1_1
snRNP-B	CAX76480.1_1
GPATC3	CAX75846.1_1
GPATC3	CAX75848.1_1
BRD2	CAX75275.1_1
snRNP-B	CAX76477.1_1
hnRNP-L	CAX73740.1_1
Hrb27-C	CAX74024.1_1
NR5A2	CAX73127.1_1
hnRNP-K	CAX77148.1_1
NCoA-5	CAX73998.1_1
snRNP-B	CAX76475.1_1
hnRNP-K	CAX75226.1_1
BRD2	CAX75272.1_1
hnRNP-L	CAX69737.1_1
hnRNP-K	CAX77155.1_1
hnRNP-K	CAX77159.1_1