Virology 486 (2015) 2–6

Contents lists available at ScienceDirect

Virology

journal homepage: www.elsevier.com/locate/yviro

Brief Communication

Mutational bias of *Turnip Yellow Mosaic Virus* in the context of host anti-viral gene silencing

Jinmin Ma [a], Denise Pallett [b], Hui Jiang [a], Yong Hou [a], Hui Wang [a,b,c,*]

[a] BGI-shenzhen, Beishan Road, Yantian, Shenzhen 518083, China
[b] NERC/Centre for Ecology & Hydrology, Benson Lane, Wallingford, Oxfordshire OX10 8BB, UK
[c] Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK

ARTICLE INFO

Article history:
Received 4 April 2015
Returned to author for revisions
1 May 2015
Accepted 21 August 2015

Keywords:
Anti-viral post-transcriptional gene silencing
Turnip Yellow Mosaic Virus
Brassica juncea
Virus population
Mutational bias

ABSTRACT

Plant Dicer-like (DCL) enzymes exhibit a GC-preference during anti-viral post-transcriptional gene silencing (PTGS), delivering an evolutionary selection pressure resulting in plant viruses with GC-poor genomes. However, some viruses, *e.g. Turnip Yellow Mosaic Virus* (TYMV, genus *Tymovirus*) have GC-rich genomes, raising the question as to whether or not DCL derived selection pressure affects these viruses. In this study we analyzed the virus-derived small interfering RNAs from TYMV-infected leaves of *Brassica juncea* showed that the TYMV population accumulated a mutational bias with AU replacing GC (GC–AU), demonstrating PTGS pressure. Interestingly, at the highly polymorphic sites the GC–AU bias was no longer observed. This suggests the presence of an unknown mechanism preventing mutational drift of the viral population and maintaining viral genome stability, despite the host PTGS pressure.

Introduction

Plant virus infections trigger host anti-viral responses which are mediated by post-transcriptional gene silencing (PTGS, also known as RNA interference in animals) (Ding and Voinnet, 2007; Pumplin and Voinnet, 2013). Plant PTGS is mediated by the Dicer-like (DCL) RNAse III enzymes which cleave double-stranded (ds) RNAs. There are four families of plant DCL enzymes involved in anti-viral PTGS that produces virus-derived small interfering (vsi) RNAs in different lengths (DCL-4: 21 nt, DCL-2: 22 nt, DCL-3: 24 nt) (Deleris et al., 2006). Unlike animal Dicers, plant DCLs operate with a GC bias, *i.e.*, producing vsiRNAs with a bias for GC-rich regions of the viral genomes (Donaire et al., 2009; Ho et al., 2006, 2007, 2010; Miozzi et al., 2013; Yan et al., 2010; Zhang et al., 2014). It has been hypothesized that a GC-bias in vsiRNA production results in a selection pressure on plant virus genome evolution. Evidence obtained by analyzing the vsiRNA populations supports the hypothesis that this pressure results in the evolution of GC-poor (AT/U-rich) genomes and reduces host anti-viral PTGS responses during infection (Ho et al., 2010). However, despite the fact that majority of the plant viruses have GC-poor genomes, there are some plant viruses that display GC-rich genomes. This raises the question as to how the GC-rich viruses deal with the plant anti-viral PTGS attack and how they maintain genomic integrity. In this study we analyzed the mutation profiles of a GC-rich, positive single-stranded (+ss)RNA virus, *Turnip Yellow Mosaic Virus* (TYMV, genome size 6318–6320 nt long, genus *Tymovirus*, family *Tymoviridae*) using the vsiRNA population defined by the small RNA sequencing protocol based on the Illumina Hi-Seq platform. Next generation sequencing (NGS) technology has previously been shown to be a powerful tool for studying viral ecology (Stobbe and Roossinck, 2014; Virgin, 2014) and viral populations (Beerenwinkel and Zagordi, 2011; Skums et al., 2014, 2013; Watson et al., 2013; Willerth et al., 2010; Wright et al., 2011). Small (s)RNA sequencing is particularly useful for characterizing host anti-viral immunity mediated by PTGS (Donaire et al., 2009; Pallett et al., 2010; Wu et al., 2010). Infection with TYMV triggers vsiRNA production by DCLs (Jakubiec et al., 2012) and the virus encodes a PTGS suppressor protein, P69 (Chen et al., 2004). In this study, more GC to AU (GC–AU) mutations were observed at the GC-dominant positions than the AU–GC mutations at the AU-dominant positions, demonstrating that PTGS derived selection pressure operated within the TYMV population. However, this GC–AU bias was not detected in the highly polymorphic sites in the

* Corresponding author at: NERC/Centre for Ecology & Hydrology, Benson Lane, Wallingford, Oxfordshire OX10 8BB, UK.
E-mail addresses: huw@ceh.ac.uk, huiwang789@gmail.com (H. Wang).

http://dx.doi.org/10.1016/j.virol.2015.08.024
0042-6822/

TYMV genome, suggesting that the viral genome composition may be maintained by an unknown mechanism.

Results and discussion

In all three samples, approximately 10 million sRNA reads were obtained (SRR2017660, TYMV-Pre, 1,911,164 unique reads; SRR1867781, TYMV-NGS, 4,691,814 unique reads; SRR1867782, Control, 3,022,140 unique reads). All libraries were dominated by 24 nt long sRNA species (Supplementary Fig. S2), indicating good sRNA quality.

In both TYMV samples, TYMV vsiRNAs covered the virus genome fully. The GC contents of the TYMV vsiRNAs were 56.32% and 56.09% for TYMV-Pre and TYMV-NGS respectively, comparable with that of the TYMV_Ox_NGS consensus sequence (56.32%). There was no GC enrichment observed for the TYMV vsiRNAs, in contrast to reports of other plant viruses with GC-poor genomes (Donaire et al., 2009; Ho et al., 2006, 2007, 2010; Miozzi et al., 2013; Yan et al., 2010; Zhang et al., 2014). It is likely the high GC content of the TYMV genome masked the GC-preference of the plant DCLs, as previously observed in the micro(mi)-RNA productions from GC-rich precursor miRNAs (pre-miRNA) in *Gramineae* species (Ho et al., 2006, 2007). However, there were significant GC increments in vsiRNA species that were detected repeatedly (Fig. 1A, TYMV-Pre, df=3, $F=8.25$, $P<0.001$; TYMV-NGS, df=3, $F=10.33$, $P<0.001$, ANOVA, MiniTab), demonstrating evidence of plant DCL mediated GC-bias in vsiRNA production (Donaire et al., 2009; Ho et al., 2006, 2007, 2010; Miozzi et al., 2013; Yan et al., 2010; Zhang et al., 2014). The minor nucleotide pair rates at the TYMV genome positions were also significantly different (Fig. 1B). The GC–AU were higher than the AU–GC in both TYMV-Pre (df=1, $F=146.16$, $P<0.001$, ANOVA, MiniTab) and TYMV-NGS (df=1, $F=39.17$, $P<0.001$) (Fig. 1B), indicating that there were higher rates of AU mutations detected at GC dominant sites than the rates of GC mutations detected at AU dominant sites. This suggests that the TYMV populations in both samples were under a selection pressure mediated by the GC-preference of plant PTGS. However, due to the possibility that experimental factors may affect the NGS performance (Sims et al., 2014), we further investigated the viral mutation profiles (Supplementary Tables S1 and S2) of the TYMV samples.

If the accumulation of GC–AU mutations (Fig. 1B) was due to genuine biological processes it would positively correlate to Pi, *i.e.*, the higher Pi the greater GC–AU > AU–GC. The GLM analysis showed such evident in both TYMV samples (Supplementary Fig. S3). However, we also detected experimental factors that influenced the GC–AU and AU–GC. Due to the sequence quality of the two TYMV datasets (described in following paragraphs), we used the TYMV-NGS library to further clarify the mutational bias of TYMV.

From the TYMV-NGS library, 1,091,704 TYMV derived vsiRNAs (74,716 unique) were mapped against the TYMV_Ox_NGS consensus sequence (Fig. 2). The TYMV vsiRNAs were dominated by the 21 nt species (Fig. 2A) compared to the 24 nt domination of the total sRNA (Supplementary Fig. S2B) indicating that the host DCL-4 pathway was dominant in TYMV vsiRNA production (Deleris et al., 2006). The TYMV vsiRNAs were generated from both plus (32.3%) and minus (67.7%) strands of the virus genome (Fig. 2B), suggesting that the double-stranded viral intermediates severed as the DCL targets. In addition to the result shown in Fig. 1B, the Kruskal–Wallis test also showed that the AU mutation rate at the GC-dominant sites (GC–AU, $n=3554$, Mid=0.23%) was significantly higher ($H=704.54$, df=1, $P=0.000$) than the GC mutation rate at the AU-dominant sites (AU–GC, $n=2765$, Mid=0.12%), resulting in the rejection of the null hypothesis that GC–AU is

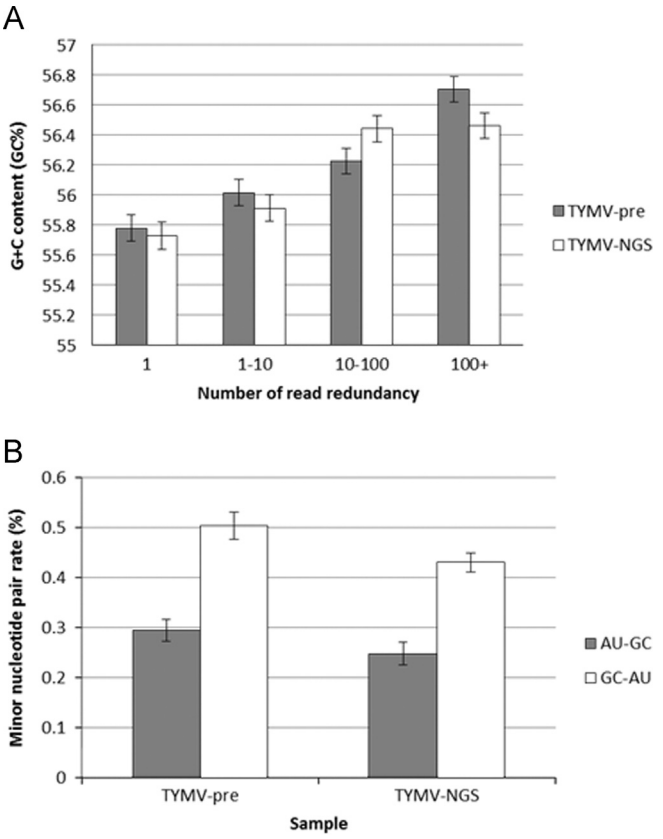

Fig. 1. Plant GC biases in TYMV vsiRNA production and GC–AU bias in TYMV mutagenesis. Data was based on both of the TYMV infected samples (TYMV-Pre and TYMV-NGS). Panel A shows the increments of GC contents in repeatedly detected TYMV reads (redundant reads that have identical sequences, $P<0.001$ for both TYMV-Pre and TYMV-NGS series), indicating that the TYMV vsiRNA hotspots (Fig. 2B) had higher GC contents than the non-hotspot regions. The *X*-axis numbers indicate read redundancy categories of $n=1$, $1<n\leq 10$, $10<n\leq 100$ and $n>100$. Panel B shows that the rates of AU mutations detected at the GC dominant sites (GC–AU) were higher ($P<0.001$ for both the samples) than those of GC mutations detected at the AU dominant sites (AU–GC), suggesting that the TYMV populations accumulated mutations with a GC–AU bias.

equal to AU–GC. Although the plants were sampled at a single time point, the observed single nucleotide polymorphism profile (Supplementary Table S2) represented a snapshot of the TYMV population. The significantly elevated GC–AU rate (compared to AU–GC) suggested that the TYMV population was biased towards accumulating GC–AU mutations.

If there were significant experimental artifacts generated due to procedure, *e.g.*, sequencing errors (Sims et al., 2014), the GC–AU mutational bias would be observed in relation to the sequencing depth of each site, *i.e.*, the more time a site was subjected to analysis, the more (or less) experimental artifacts would be displayed. Therefore, we divided the TYMV sites into 10 groups according to the sequence coverage (Depth) in descending order, *i.e.*, Depth-Group-A represented the top 10% of the mostly sequenced sites and Depth-Group-B represented the next 10% of sites *etc.* (Supplementary Table S2). Analysis with GLM detected statistically significant interactions (df=9, $F=2.34$, $P=0.012$) between the site domination (Code-AU–GC) and the sequencing depth (Code-depth) (Supplementary Fig. S4) showing that the sequencing depth (ranging from 10 to 41,254, $n=6319$ sites) was a factor contributing to the difference between GC–AU and AU–GC (Fig. 3A). When sequencing depth > 2886 (Supplementary Table S2) in the Depth-Groups A–D (Fig. 3B), the difference between GC–AU and AU–GC became stable indicating that sequencing depth was no longer a significant factor (df=3, $F=0.02$, $P=0.995$, GLM,