BMC
Genomics

High-throughput sequencing reveals small RNAs involved in ASGV infection

Marike Visser[1,2], Hans J Maree[2,3], D Jasper G Rees[1] and Johan T Burger[2*]

Abstract

Background: Plant small RNAs (sRNAs) associated with virulent virus infections have been reported by previous studies, while the involvement of sRNAs in latent virus infection remains largely uncharacterised. Apple trees show a high degree of resistance and tolerance to viral infections. We analysed two sRNA deep sequencing datasets, prepared from different RNA size fractions, to identify sRNAs involved in *Apple stem grooving virus* (ASGV) infection.

Results: sRNA analysis revealed virus-derived siRNAs (vsiRNAs) originating from two ASGV genetic variants. A vsiRNA profile for one of the ASGV variants was also generated showing an increase in siRNA production towards the 3′ end of the virus genome. Virus-derived sRNAs longer than those previously analysed were also observed in the sequencing data. Additionally, tRNA-derived sRNAs were identified and characterised. These sRNAs covered a broad size-range and originated from both ends of the mature tRNAs as well as from their central regions. Several tRNA-derived sRNAs showed differential regulation due to ASGV infection. No changes in microRNA, natural-antisense transcript siRNA, phased-siRNA and repeat-associated siRNA levels were observed.

Conclusions: This study is the first report on the apple sRNA-response to virus infection. The results revealed the vsiRNAs profile of an ASGV variant, as well as the alteration of the tRNA-derived sRNA profile in response to latent virus infection. It also highlights the importance of library preparation in the interpretation of high-throughput sequencing data.

Keywords: *Apple stem grooving virus*, Next-generation sequencing, Plant-virus interaction, tRNA-derived fragment, tRNA-half, Virus-derived small interfering RNA

Background

The domesticated apple, *Malus* x *domestica* (*M.* x *domestica*), has a wide range of infectious agents, which include fungi, bacteria, phytoplasma, viruses and viroids. One such virus, *Apple stem grooving virus* (ASGV), is the type member of the genus *Capillovirus* (family *Flexiviridae*) [1]. It is a positive-sense RNA virus with a genome of approximately 6.5 kb, which is organised into two overlapping open reading frames (ORFs) [2]. ASGV infection is mostly symptomless (latent) in apple cultivars, depending on the virus strain, however some cultivars are susceptible and may develop severe symptoms such as xylem pitting and grooving, phloem necrosis and the complete decay of the tree [3].

During infection the replication of RNA viruses generate long dsRNA intermediate molecules that triggers the synthesis of small interfering RNAs (siRNAs) [4]. Furthermore, the folded duplex regions of single stranded viral RNAs can also result in siRNA synthesis [5]. These virus-derived siRNAs (vsiRNAs) subsequently regulate viral RNA expression through a process known as RNA silencing. In addition to vsiRNA production, plants' endogenous small RNA (sRNA) pathways are also affected by viral infection [6-8].

With the introduction of next-generation sequencing the knowledge of sRNA species has been extended beyond the well-characterised miRNA, trans-acting siRNA (tasiRNA) and natural-antisense transcript (NAT) siRNA (nat-siRNA) groups. Although sRNAs were shown to originate from tRNA before, Lee et al. [9] was the first to illustrate that these molecules were not produced by non-systematic tRNA degradation [9]. Small RNAs associated with tRNAs have been divided into two categories based on the tRNA region they originate from. The first group, called tRNA halves (tsRNA/tiRNA), are derivatives of

* Correspondence: jtb@sun.ac.za
[2]Genetics Department, Stellenbosch University, Stellenbosch, South Africa
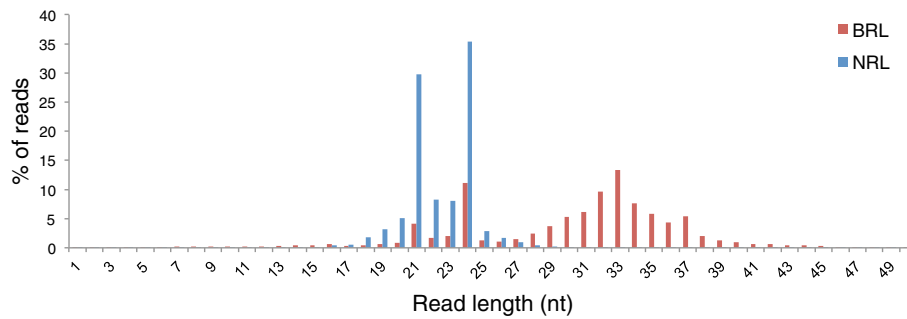Full list of author information is available at the end of the article

Figure 1 Size-distribution of the two sRNA sequencing library types. Histogram illustrating the number of reads, 1 to 50 nt in length, as a percentage of the reads in this size-range for the BRL and NRL data respectively.

mature tRNAs cleaved in the anticodon loop, resulting in functional sRNAs of around 28 to 36 nucleotides in size. Enzymes involved in their biogenesis have been identified for humans [10], yeast [11] and bacteria [12], but are still unknown in plants.

Transfer RNA cleaved in the D or T loop give rise to a second group of sRNAs, called tRNA-derived RNA fragments (tRFs). This group can be further divided into sRNAs stemming from (a) the 5′ end of mature tRNAs, (b) the 3′ end of mature tRNAs and (c) the 3′ end of immature tRNAs, called 5′-tRFs, 3′ CCA tRFs and 3′ U tRFs respectively [13]. Several synonyms have been used for the different sub-groups [9,14].

In this study a next-generation sequencing approach was followed to identify sRNAs that are associated with a latent virus infection in apple plants. In addition to illustrating the vsiRNA profiles associated with an ASGV genetic variant the results from this study demonstrate the involvement of tRNA-derived sRNAs in plant-virus interaction. The lack of differential regulation of miR-NAs, phasiRNAs, nat-siRNAs and rasiRNAs in leaf material is also shown.

Results and discussion

sRNA sequencing libraries

Two library preparation approaches were followed. The first approach made use of total RNA to produce a broad range library (BRL) for each sample, with individual sequenci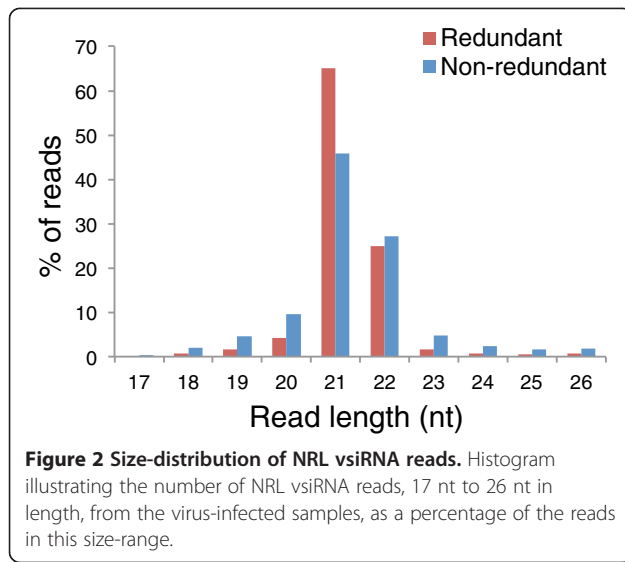ng datasets comprising of between 7,543,861 and 11,648,479 reads. Reads of 27 nt and longer contributed to 73% of all BRL reads. Since sRNAs involved in gene silencing are mostly considered to fall within the 17 to 26 nt size-range, a second narrow range library (NRL) was prepared for each sample using size-selected sRNAs to increase the sequence depth of these sRNAs. These libraries generated 7,235,867 to 14,896,610 high quality reads per sample. The size-range 17 to 26 nt in length represented 97% of all the reads and were used for downstream analysis. Figure 1 illustrates the size distribution of the sRNA reads (1 to 50 nt in size) for the pooled BRL and pooled NRL datasets. The histogram not only highlights the increase in the percentage of reads 17 to 26 nt in length in the NRL datasets when compared to the BRL datasets, but also shows a change in ratios between the different size groups, in particular when comparing the ratio between the 21 and 24 nt groups. Since the same total RNA extract was used to prepare both libraries, this observation demonstrates the effect of library preparation on the final sequencing data and highlights the difficulty of comparing data generated by different protocols.

vsiRNAs resulting from ASGV infection

The NRL data was first used to analyse the production of vsiRNAs, since these datasets were enriched for sRNAs in the size range known to be associated with vsiRNAs. Reads, which did not align to the apple nuclear, chloroplast or mitochondrial genomes, were mapped (allowing a single mismatch) onto the complete genomes of six ASGV

Table 1 Results for the virus-infected NRL sRNA read-mapping against ASGV genomes

Isolate	GenBank Accession number	Country	Host	Genome size (nt)	Total number of reads mapped	Non-redundant number of reads mapped	Genome coverage (%)
ASGV-AC	JX080201.1	Germany	*M. domestica*	6496	27069	5897	98.04
ASGVp12	HE978837.1	India	*M. domestica*	6478	25256	5297	96.34
ASGV P-209	NC_001749.2	Japan	*M. domestica*	6495	14341	3810	88.96
ASGV	D14995.2	Japan	*M. domestica*	6495	14341	3810	88.96
ASGV-HH	JN701424.1	China	*Pyrus pyrifolia*	6496	8591	1872	54.63
ASGV-CHN	JQ308181.1	China	*M. domestica*	6495	6555	1659	52.73

Figure 2 Size-distribution of NRL vsiRNA reads. Histogram illustrating the number of NRL vsiRNA reads, 17 nt to 26 nt in length, from the virus-infected samples, as a percentage of the reads in this size-range.

isolates. sRNA read-mapping results for the pooled NRL sequencing data from the virus-infected samples are shown in Table 1. In total, 0.59% of all non apple-derived reads (17 to 26 nt in length) from the infected samples mapped onto at least one of the ASGV genomes. The large number of reads mapped onto the genome of the German isolate ASGV-AC, with 98% coverage. The majority of the virus-derived reads from the NRL data were 21 nt long followed by reads 22 nt in length (Figure 2), which is often seen for positive-sense RNA viruses [15,16].

The occurrence of mixed ASGV infection was analysed using the genomes of three isolates (ASGV-AC, ASGV P-209 and ASGV-HH). These isolates each had an equal or higher sRNA read count (Table 1) than their closest relative (Figure 3). To determine the sRNA reads associated only with a specific variant genome, reads with a uniquely mapped position and genome were reported (Table 2). The large majority of variant-specific reads were associated with ASGV-AC, followed by the Japanese isolate (P-209). These variant-specific reads were distributed along the length of each of the three genomes (Figure 4), indicating that more than one ASGV

variant was present with distinct genome sequences, rather than a single recombinant virus. Given their reasonably large number of total, as well as variant-specific reads, we suggest that at least two ASGV genetic variants, closely related to ASGV-AC and P-209 respectively, were present in the samples. Closer assessment of the read-mapping profiles for the individual samples suggested that two samples contained a mixed infection of the two variants, while the third was singly infected with a genetic variant of ASGV-AC.

Since only one sample was confirmed to be infected with a single ASGV variant, only reads from this sample could be used to generate a complete vsiRNA profile of this variant. Figure 5 shows the mapping distribution of the vsiRNA reads along the virus genome. In general the 3′ end of the genomes showed regions of higher genome coverage by vsiRNAs. The increase in vsiRNAs production toward the 3′ end of the genome has previously been ascribed to the presence of viral subgenomic RNAs (sgRNAs) [17,18]. Both the ASGV movement and coat proteins are expressed from 3′ sgRNAs [19-21] and may explain the increase in vsiRNA originating from the 3′ end. The non-redundant reads were also plotted onto the ASGV genomes (Figure 5). The majority of the genome sequences were associated with the production of more than one unique vsiRNA, illustrating that multiple Dicer-like (DCL) cleavage sites are in close proximity to each other on a virus genome.

Altogether, from the three BRL datasets of the infected samples, 0.24% of the non apple-derived reads (>16 nt in length) mapped onto the six ASGV genomes. Although less than the NRL viral reads, these reads still covered 97% and 82% of the ASGV-AC and P-209 genomes, respectively (Table 3). Similar to the NRL datasets, the 21 nt long reads also dominated the viral reads in the BRL datasets (Figure 6). The second most abundant size group was the 22-nt group, closely followed by reads 33 nt in length. To our knowledge this is the first report of plant virus-derived sRNA reads larger than 30 nt in length. These larger sRNAs contributed significantly to the number of virus-associated reads and may point
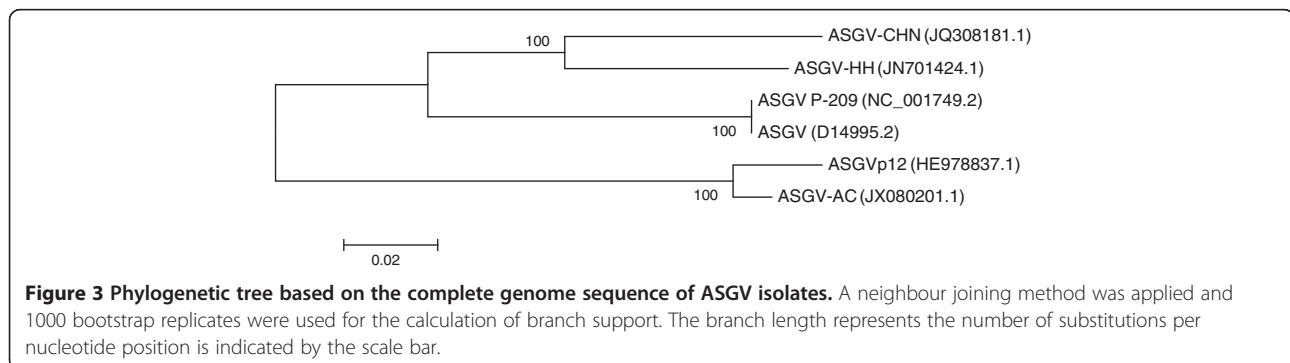

Figure 3 Phylogenetic tree based on the complete genome sequence of ASGV isolates. A neighbour joining method was applied and 1000 bootstrap replicates were used for the calculation of branch support. The branch length represents the number of substitutions per nucleotide position is indicated by the scale bar.

Table 2 Results for the vsiRNA variant-specific read-mapping

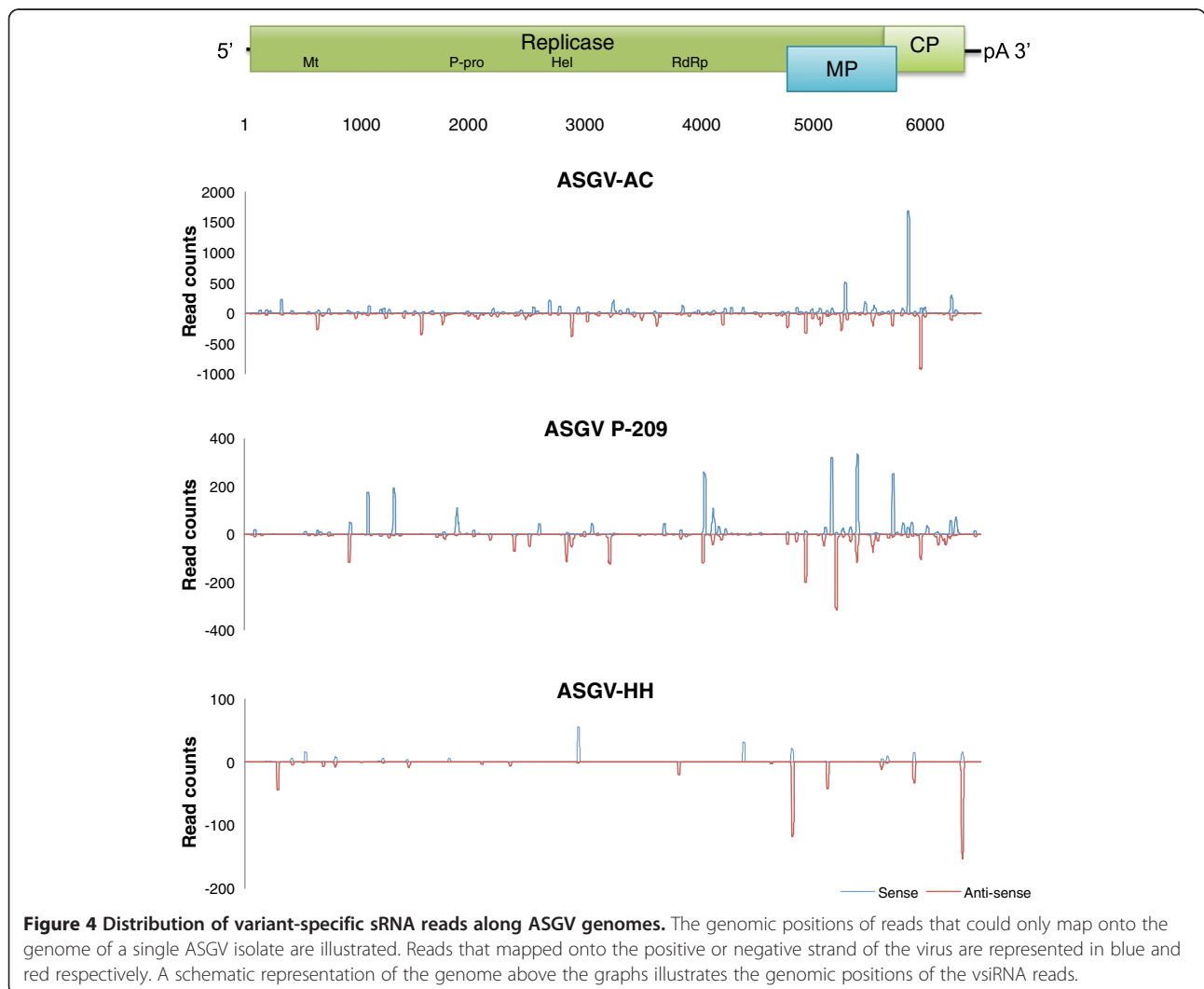| Isolate	GenBank Accession number	ASGV-infected sample 1	ASGV-infected sample 2	ASGV-infected sample 3	All ASGV-infected samples
ASGV-AC	JX080201.1	4074 (1326)	4905 (1544)	6507 (1832)	15486 (3111)
ASGV P-209	NC_001749.2	2469 (768)	2448 (703)	83 (20)	5000 (1120)
ASGV-HH	JN701424.1	269 (95)	267 (118)	141 (34)	677 (191)

The number of reads which mapped only onto a specific ASGV genome for each virus-infected sample as well as for the pooled data are shown.
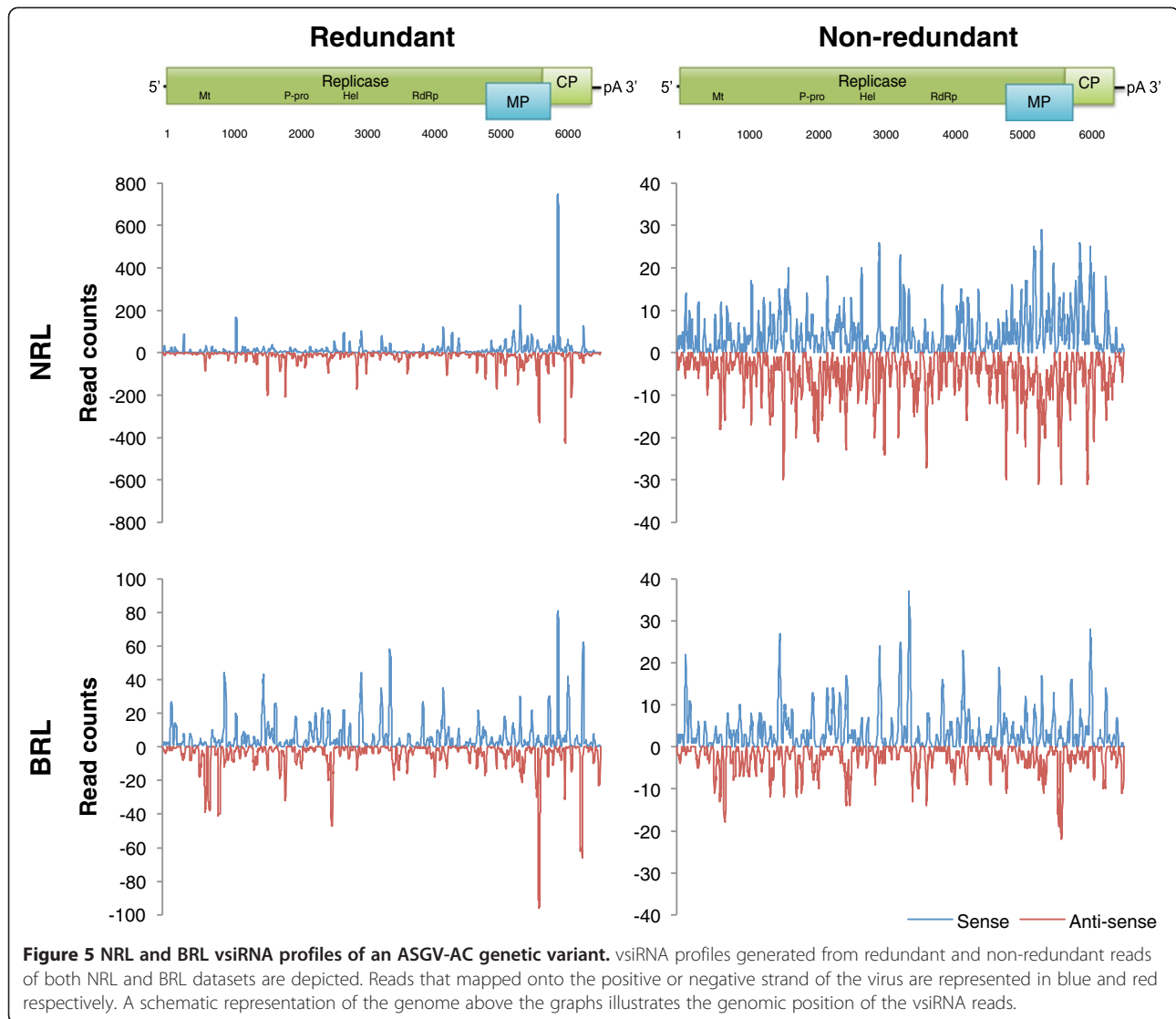The non-redundant read counts are given within brackets.

towards their biological importance. Alternatively, these reads possibly represent remnants of siRNA-directed ASGV genome degradation. The distribution of BRL reads along the ASGV genome was also examined (Figure 5). The presence of a substantial number of larger sRNAs in the BRL data resulted in a change in the vsiRNA profiles. The dominant areas of higher coverage by the conventional vsiRNAs (as can been seen from the NRL data) are surpassed (in the BRL data) by the additional areas of higher coverage, which are generated by

the longer vsiRNAs. Furthermore, the 3′ vsiRNA bias was also less evident in the BRL data, compared to the NRL data. This observation once again demonstrates the effect of library preparation on sequencing results and the interpretation thereof.

tRNA-derived sRNAs show differential regulation due to ASGV infection

Previous studies have established that sRNAs are generated from tRNAs in a non-random manner and that

Figure 4 Distribution of variant-specific sRNA reads along ASGV genomes. The genomic positions of reads that could only map onto the genome of a single ASGV isolate are illustrated. Reads that mapped onto the positive or negative strand of the virus are represented in blue and red respectively. A schematic representation of the genome above the graphs illustrates the genomic positions of the vsiRNA reads.

Figure 5 NRL and BRL vsiRNA profiles of an ASGV-AC genetic variant. vsiRNA profiles generated from redundant and non-redundant reads of both NRL and BRL datasets are depicted. Reads that mapped onto the positive or negative strand of the virus are represented in blue and red respectively. A schematic representation of the genome above the graphs illustrates the genomic position of the vsiRNA reads.

they play a regulatory role similar to other sRNA species [9,22,23]. In the BRL data the tRNA-derived sRNAs represented 23% and 19% of the reads larger than 16 nt for the infected and healthy samples respectively. These sRNAs varied in size from 17 to 59 nt in length, representing both tRFs and tRNA-halves. The broad size-range of the tRNA-derived sRNA reads in this study demonstrates that it is not always possible to clearly distinguish between these two classes only based on sequence length and origin. The larger species also stretched beyond the recognized tRNA-half size-range, spanning the anticodon loop, similar to previous reports

Table 3 Results for the virus-infected BRL sRNA read-mapping against ASGV genomes

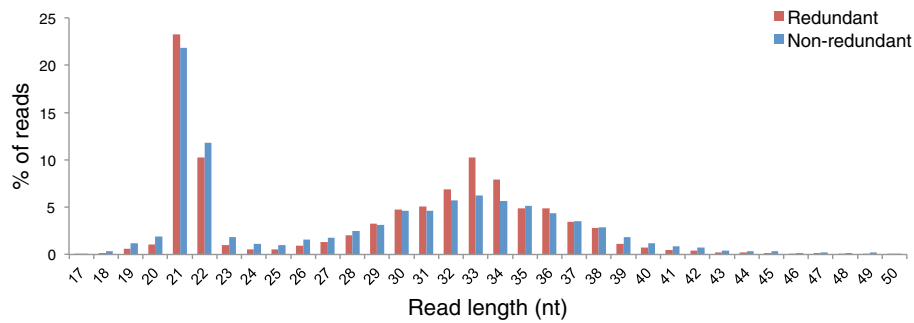| Isolate	GenBank Accession number	Genome size (nt)	Total number of reads mapped	Non-redundant number of reads mapped	Genome coverage (%)
ASGV-AC	JX080201.1	6496	7795	3751	97.44
ASGVp12	HE978837.1	6478	6628	3201	93.96
ASGV P-209	NC_001749.2	6495	4456	2059	82.03
ASGV	D14995.2	6495	4456	2059	82.03
ASGV-HH	JN701424.1	6496	1945	886	42.13
ASGV-CHN	JQ308181.1	6495	1341	721	39.98

Figure 6 Size-distribution of BRL vsiRNA reads. Histogram illustrating the number of BRL vsiRNA reads, 17 nt and longer, from the virus-infected samples as a percentage of the reads in this size-range.

[24,25]. The majority of tRNA-derived sRNAs in apple were 33 nt in length followed by reads of 32 and 37 nt (Figure 7). The dominant single tRNA-derived sRNA was a 5′ tRNA-half (33 nt long) originating from tRNA-Asp^{GTC}, and was represented by a total of 1,814,310 reads in the BRL datasets. In contrast, for the NRL dataset only 1.6% of all reads (17 tot 26 nt), originated from tRNAs.

sRNAs, originating from both 5′ and 3′ ends of mature tRNAs, were identified in datasets from both library types (Additional file 1: Table S1 and S2). Additionally, and in agreement with previous studies [23,26], sRNAs were also identified originating from the central part of tRNAs. These internal species were especially prominent in the cluster of sRNAs (in the BRL data) spawning from tRNA-Gln^{CTG}.

When the potential involvement of tRNA-derived sRNAs in ASGV infection was investigated, several tRNA-derived sRNAs showed significant variation in expression levels between infected and healthy samples (Additional file 2: Table S3 to S6). Not only did individual sRNAs show differential expression, but the total number of sRNAs spawned by some of these tRNAs was found to significantly vary between the two groups. One

tRNA, tRNA-Tyr^{GTA}, in particular displayed an interesting altered sRNA arrangement in the ASGV-infected samples (Figure 8). The BRL data revealed an increase in sRNAs derived from its 3′ end, extending into the variable region, while both BRL and NRL datasets showed a decrease in sRNAs that were generated from the central part of the tRNA. The 5′ ends of these internal sRNAs consistently coincided with the 5′ ends of the anticodon stems and extended into the variable regions. Due to the uniformity of their 5′ ends, these internal sRNAs cannot originate as a result of random tRNA degradation. From the inverse regulation of the two fragment types, it can be argued that a single tRNA molecule did not give rise to both species, but rather that they were generated through separate pathways. However, the possibility exists that these two processes are linked and that the production of the one species affects that of the other.

The biogenesis of tRNA-derived sRNAs, as well as the way in which they affect other molecular pathways remains to be elucidated. Earlier reports speculated that tRFs bind, to ribosomes resulting in a down-regulation of gene expression [27]. Through their association with argonaute proteins a possible role in post-transcriptional
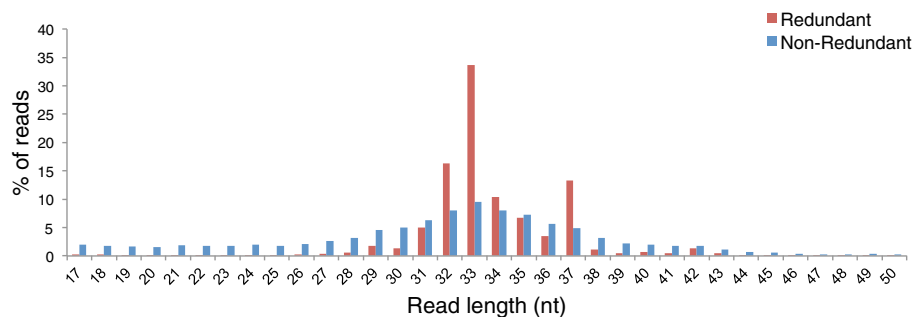

Figure 7 Size-distribution of BRL tRNA-derived sRNA reads. Histogram illustrating the number of BRL tRNA-derived sRNA reads, 17–50 nt in length, as a percentage of the reads in this size-range.
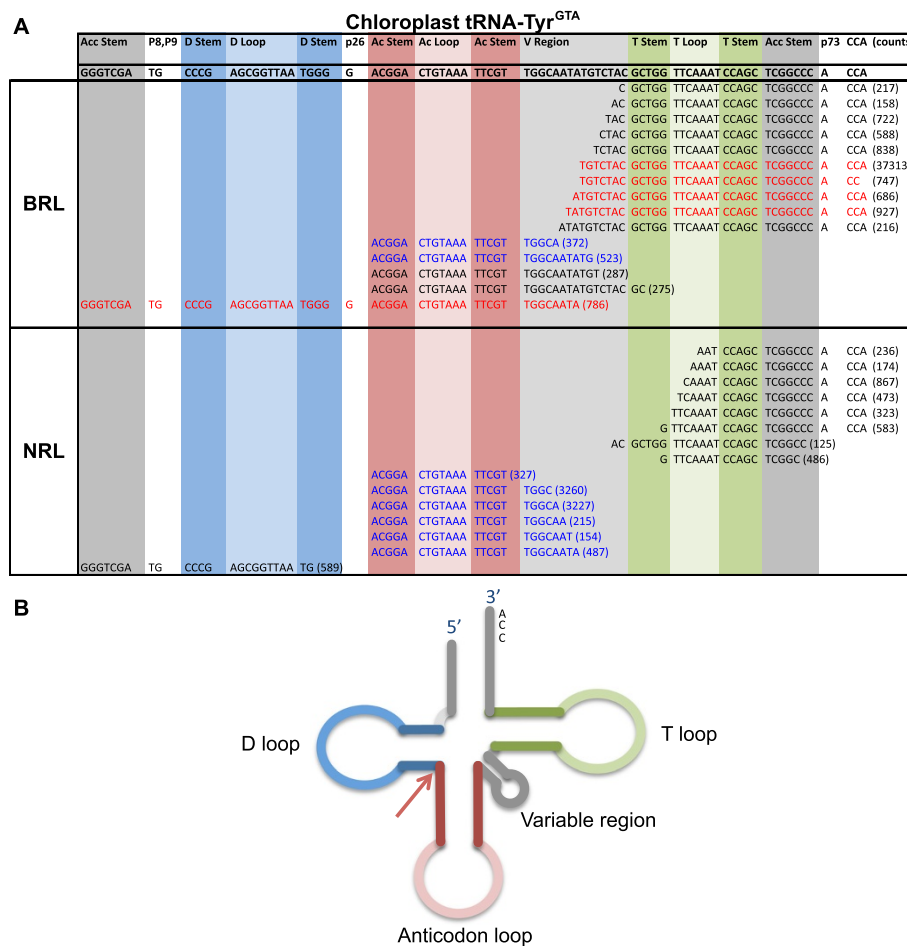
Figure 8 Variation in tRNA-derived sRNA profiles. A) Diagram showing the sRNA reads with the highest read count for each of the two types of data, which are associated with the chloroplast tRNA-Tyr^{GTA}. The sRNAs, which were up- or down-regulation due to ASGV infection, are indicated in red and blue respectively. The total read count of each sRNA is indicated within brackets. **B)** The red arrow illustrates the 5′ start position of a cluster of central sRNAs, originating from tRNA-Tyr^{GTA}, which are down-regulated during ASGV infection.

gene silencing was also suggested [23]. The biological function of the differentially regulated tRNA-derived sRNAs in the current study remains to be determined.

The involvement of other endogenous sRNAs in ASGV infection

Besides the vsiRNAs and tRNA-derived sRNAs involved in ASGV infection, differential expression analysis showed no variation in phasiRNA and miRNA levels as a result of ASGV infection; neither did the nat-siRNAs or rasiRNAs show any change in expression levels (Additional file 2: Table S7 to S17). In addition to their regulatory role during plant development, miRNAs are often linked to stress response. The latent nature of ASGV may therefore explain what seems to be a lack of miRNA involvement during ASGV infection.

Conclusions

In this study next-generation sequencing of sRNAs was used to investigate plant responses to latent virus infection. Two different sRNA libraries were generated per sample. Both datasets illustrated the synthesis of virus-derived sRNAs in response to ASGV infection. Along with earlier reported tRNA-derived sRNAs of more than 30 nt in length, BRL data from this study additionally suggested virus-derived RNAs larger than the well-characterised vsiRNAs of around 21 nt. The vsiRNA profiles varied depending on the method of library preparation used, illustrating the importance of consistency when comparing different samples. Additionally, the results showed that ASGV-infection resulted in a change in the expression of tRNA-derived sRNAs, although the biological function of these sRNAs remains to be elucidated. This study is the first to report on sRNAs involved in ASGV-infection in the domesticated apple.

Methods

Sequencing library construction and data preparation

Sample material was collected from three healthy and three asymptomatic ASGV-infected (as confirmed by RT-PCR), greenhouse-grown, *M.* x *domestica* cv. 'Golden Delicious' (NIVV) seedlings, grafted onto MM.109 rootstocks. The viral status was confirmed by two multiplex RT-PCR reactions described in Menzel et al. [28]. The primers for *Apple mosaic virus* detection were replaced with those described in Hassan et al. [29]. See Additional file 3: Table S18 for primer information. Total RNA was extracted from mature leaf material using the Plant RNA Reagent Kit (Invitrogen) and used for library (BRL) preparation by means of the TruSeq Small RNA library preparation kit from Illumina. For each sample a second library (NRL) was prepared using the small RNA fraction (17–29 nt) purified from total RNA using a 15% TBE-urea polyacrylamide gel. The final BRL and NRL libraries were size-selected by means of a 3% Pippin Prep cassette (Sage) and a 6% polyacrylamide gel (Invitrogen), respectively, and sequenced on an Illumina HiScan SQ instrument. The software cutadapt (V 1.0) [30] was used to remove adapter sequences and the reads were filtered for quality (phred score ≥ 20) using FASTX-toolkit (V 0.0.13) [31]. For the NRL, reads less than 17 or longer than 26 nt in length were discarded, while all filtered reads 17 nt and longer were used for the analysis of the BRL data.

vsiRNA analysis

Reads from the three NRL virus-infected datasets were combined for vsiRNA analyses. Reads that could map with less than two mismatches onto the apple nuclear, chloroplast or mitochondrial genomes, obtained from the Genome Database for Rosaceae [32,33], were removed. Bowtie (V 0.12.7) [34] was used to perform all read-mapping analyses. The filtered reads were then mapped onto six ASGV genomes, allowing only a single mismatch. Similar analyses were performed for the pooled BRL virus-infected samples. Variant-specific reads were identified as those reads that uniquely mapped (using Bowtie) onto one of the six ASGV genomes, when only allowing perfect matches between the sRNA read and the genome.

tRF and tRNA-half identification

Mature tRNA sequences of five angiosperms (*Arabidopsis thaliana*, *Brachypodium distachyon*, *Medicago truncatula*, *Oryza sativa* and *Populus trichocarpa*) were retrieved from the PlantRNA database [35]. To identify apple tRFs present, the six NRL datasets were combined and mapped, with Bowtie, onto the retrieved mature tRNA sequences, allowing two mismatches. tRNA-halves were correspondingly identified using the pooled BRL datasets.

Differential expression analysis of apple sRNA species

The standalone differential expression tool of miRanalyzer [36,37], which implements the R package, DESeq2 [38], was used to determine variation in sRNA expression levels between the healthy and the ASGV-infected samples. Five distinct sRNA species were investigated using the NRL data, namely miRNAs, phasiRNAs, nat-siRNAs, rasiRNAs and tRFs. The BRL data was used for tRNA-halves differential expression analysis. miRNA analysis was based on miRBase (version 20) [39-42] apple entries, as well as recently predicted novel miRNAs [43]. The phasiRNAs, nat-siRNAs and rasiRNAs analysed were also previously identified [43], while the tRFs and tRNA-haves were identified during the current study. The phasiRNAs included a group of apple tasiRNAs available on the tasiRNAdb [44-46].

Availability of supporting data

The datasets supporting the results of this article are available in the BioProject repository of the National Centre for Biotechnology Information, BioProject: PRJNA235941 in http://www.ncbi.nlm.nih.gov/bioproject/.

Additional files

Additional file 1: Table S1. tRNA-derived sRNAs identified in the BRL data. The 15 sRNA reads with the highest read counts associated with a tRNA are shown and those differentially regulated (|log2fold change| > =1 and padj < =0.05) as a result of ASGV infection are indicated. **Table S2.** tRNA-derived sRNAs identified in the NRL data. The 15 sRNA reads with the highest read counts associated with a tRNA are shown and those differentially regulated (|log2fold change| > =1 and padj < =0.05) as a result of ASGV infection are indicated.

Additional file 2: Table S3. Results for the differential expression analysis of clusters of sRNAs originating from tRNAs, based on BRL data. **Table S4.** Results for the differential expression analysis of the individual tRNA-derived sRNAs, based on BRL data. **Table S5.** Results for the differential expression analysis of clusters of sRNAs originating from tRNAs, based on NRL data. **Table S6.** Results for the differential expression analysis of the individual tRNA-derived sRNAs, based on NRL data. **Table S7.** Results for the differential expression analysis of apple miRNAs present in miRBase. **Table S8.** Results for the differential expression analysis of recently predicted apple miRNAs. **Table S9.** Results for the differential expression analysis of the cluster of nat-siRNAs originating from both strands of the overlapping region of NAT pairs. **Table S10.** Results for the differential expression analysis of the cluster of nat-siRNAs, from the first transcript, originating from the overlapping region of NAT pairs. **Table S11.** Results for the differential expression analysis of the cluster of nat-siRNAs, from the second transcript, originating from the overlapping region of NAT pairs. **Table S12.** Results for the differential expression analysis of all the sRNAs originating from a phased cluster. **Table S13.** Results for the differential expression analysis of phasiRNAs. **Table S14.** Results for the differential expression analysis of tasiRNAs. **Table S15.** Results for the differential expression analysis of the cluster of rasiRNAs originating from both strands of a repetitive sequence. **Table S16.** Results for the differential expression analysis of the cluster of rasiRNAs originating from the forward strand of a repetitive sequence. **Table S17.** Results for the differential expression analysis of the cluster of rasiRNAs originating from the reverse strand of a repetitive sequence.

Additional file 3: Table S18. Diagnostic RT-PCR primers. Multiplex-primers used to determine the viral status of apple plants.

Abbreviations

ASGV: *Apple stem grooving virus*; BRL: Broad range library; DCL: Dicer-like; *M. x domestica*: *Malus x domestica*; miRNA: microRNA; NRL: Narrow rang library; nat-siRNA: Natural-antisense transcript siRNA; NGS: Next-generation sequencing; phasiRNA: phased-siRNA; rasiRNAs: Repeat-associated siRNA; siRNA: Small interfering RNA; sRNA: small RNA; sgRNA: subgenomic RNA; tasiRNA: trans-acting siRNA; tRF: tRNA-derived fragment; vsiRNA: virus-derived siRNA.

Competing interest

The authors declare that they have no competing interests.

Authors' contributions

MV participated in the design of the study, collected and screened the samples for virus infection, carried out the sequencing library preparation, performed data analysis and drafted the manuscript. HJM participated in the design of the study and in sample collection, and contributed to drafting the manuscript. DJGR participated in the design of the study and performed sequencing runs. JTB participated in the design of the study and contributed to drafting the manuscript. All authors read and approved the final manuscript.

Author details

[1]Biotechnology Platform, Agricultural Research Council, Pretoria, South Africa. [2]Genetics Department, Stellenbosch University, Stellenbosch, South Africa. [3]Infruitec-Nietvoorbij, Agricultural Research Council, Stellenbosch, South Africa.

References

1. Adams MJ, Antoniw JF, Bar-Joseph M, Brunt AA, Candresse T, Foster GD, Martelli GP, Milne RG, Zavriev SK, Fauquet CM: **The new plant virus family Flexiviridae and assessment of molecular criteria for species demarcation.** *Arch Virol* 2004, **149**:1045–1060.
2. Yoshikawa N, Sasaki E, Kato M, Takahashi T: **The nucleotide sequence of Apple stem grooving Capillovirus genome.** *Virology* 1992, **191**:98–105.
3. Nickel O, Fajardo TV, Jelkmann W, Kuhn GB: **Sequence analysis of the capsid protein gene of an isolate of Apple stem grooving virus, and its survey in Southern Brazil.** *Fitopatol Bras* 2001, **26**:655–659.
4. Hamilton AJ, Baulcombe DC: **A species of small antisense RNA in posttranscriptional gene silencing in plants.** *Science* 1999, **286**:950–952.
5. Molnar A, Csorba T, Lakatos L, Varallyay E, Lacomme C, Burgyan J: **Plant virus-derived small interfering RNAs originate predominantly from highly structured single-stranded viral RNAs.** *J Virol* 2005, **79**:7812–7818.
6. Romanel E, Silva TF, Corrêa RL, Farinelli L, Hawkins JS, Schrago CEG, Vaslin MFS: **Global alteration of microRNAs and transposon-derived small RNAs in cotton (Gossypium hirsutum) during Cotton leafroll dwarf polerovirus (CLRDV) infection.** *Plant Mol Biol* 2012, **80**:443–460.
7. Shivaprasad PV, Chen H-M, Patel K, Bond DM, Santos BACM, Baulcombe DC: **A microRNA superfamily regulates nucleotide binding site-leucine-rich repeats and other mRNAs.** *Plant Cell* 2012, **24**:859–874.
8. Xiao B, Yang X, Ye C-Y, Liu Y, Yan C, Wang Y, Lu X, Li Y, Fan L: **A diverse set of miRNAs responsive to begomovirus-associated betasatellite in Nicotiana benthamiana.** *BMC Plant Biol* 2014, **14**:60.
9. Lee YS, Shibata Y, Malhotra A, Dutta A: **A novel class of small RNAs: tRNA-derived RNA fragments (tRFs).** *Genes Dev* 2009, **23**:2639–2649.
10. Fu H, Feng J, Liu Q, Sun F, Tie Y, Zhu J, Xing R, Sun Z, Zheng X: **Stress induces tRNA cleavage by angiogenin in mammalian cells.** *FEBS Lett* 2009, **583**:437–442.
11. Thompson DM, Parker R: **Stressing out over tRNA cleavage.** *Cell* 2009, **138**:215–219.
12. Ogawa T, Tomita K, Ueda T, Watanabe K, Uozumi T, Masaki H: **A cytotoxic ribonuclease targeting specific transfer RNA anticodons.** *Science* 1999, **283**:2097–2100.
13. Sobala A, Hutvagner G: **Transfer RNA-derived fragments: origins, processing, and functions.** *Wiley Interdiscip Rev RNA* 2011, **2**:853–862.
14. Haussecker D, Huang Y, Lau A, Parameswaran P, Fire AZ, Kay MA: **Human tRNA-derived small RNAs in the global regulation of RNA silencing.** *RNA* 2010, **16**:673–695.
15. Donaire L, Wang Y, Gonzalez-Ibeas D, Mayer KF, Aranda MA, Llave C: **Deep-sequencing of plant viral small RNAs reveals effective and widespread targeting of viral genomes.** *Virology* 2009, **392**:203–214.
16. Jakubiec A, Yang SW, Chua N-H: **Arabidopsis DRB4 protein in antiviral defense against Turnip yellow mosaic virus infection: DRB4 function in antiviral defense.** *Plant J* 2012, **69**:14–25.
17. Ruiz-Ruiz S, Navarro B, Gisel A, Peña L, Navarro L, Moreno P, Serio FD, Flores R: **Citrus tristeza virus infection induces the accumulation of viral small RNAs (21–24-nt) mapping preferentially at the 3'-terminal region of the genomic RNA and affects the host small RNA profile.** *Plant Mol Biol* 2011, **75**:607–619.
18. Silva TF, Romanel EA, Andrade RR, Farinelli L, Østerås M, Deluen C, Corrêa RL, Schrago CE, Vaslin MF: **Profile of small interfering RNAs from cotton plants infected with the polerovirus Cotton leafroll dwarf virus.** *BMC Mol Biol* 2011, **12**:40.
19. Hirata H: **A single silent substitution in the genome of Apple stem grooving virus causes symptom attenuation.** *J Gen Virol* 2003, **84**:2579–2583.
20. Tatineni S, Afunian MR, Gowda S, Hilf ME, Bar-Joseph M, Dawson WO: **Characterization of the 5'- and 3'-terminal subgenomic RNAs produced by a capillovirus: Evidence for a CP subgenomic RNA.** *Virology* 2009, **385**:521–528.
21. Komatsu K, Hirata H, Fukagawa T, Yamaji Y, Okano Y, Ishikawa K, Adachi T, Maejima K, Hashimoto M, Namba S: **Infection of capilloviruses requires subgenomic RNAs whose transcription is controlled by promoter-like sequences conserved among flexiviruses.** *Virus Res* 2012, **167**:8–15.
22. Li Z, Ender C, Meister G, Moore PS, Chang Y, John B: **Extensive terminal and asymmetric processing of small RNAs from rRNAs, snoRNAs, snRNAs, and tRNAs.** *Nucleic Acids Res* 2012, **40**:6787–6799.
23. Loss-Morais G, Waterhouse PM, Margis R: **Description of plant tRNA-derived RNA fragments (tRFs) associated with argonaute and identification of their putative targets.** *Biol Direct* 2013, **8**:6.
24. Zhang S, Sun L, Kragler F: **The phloem-delivered RNA pool contains small noncoding RNAs and interferes with translation.** *Plant Physiol* 2009, **150**:378–387.
25. Nowacka M, Strozycki PM, Jackowiak P, Hojka-Osinska A, Szymanski M, Figlerowicz M: **Identification of stable, high copy number, medium-sized RNA degradation intermediates that accumulate in plants under non-stress conditions.** *Plant Mol Biol* 2013, **83**:191–204.
26. Wang Q, Lee I, Ren J, Ajay SS, Lee YS, Bao X: **Identification and functional characterization of trna-derived RNA fragments (tRFs) in respiratory syncytial virus infection.** *Mol Ther* 2013, **21**:368–379.
27. Gebetsberger J, Zywicki M, Künzi A, Polacek N: **tRNA-Derived fragments target the ribosome and function as regulatory non-coding RNA in Haloferax volcanii.** *Archaea* 2012, **2012**:1–11. Article ID 260909.
28. Menzel W, Jelkmann W, Maiss E: **Detection of four apple viruses by multiplex RT-PCR assays with coamplification of plant mRNA as internal control.** *J Virol Methods* 2002, **99**:81–92.
29. Hassan M, Myrta A, Polak J: **Simultaneous detection and identification of four pome fruit viruses by one-tube pentaplex RT-PCR.** *J Virol Methods* 2006, **133**:124–129.
30. Martin M: **Cutadapt removes adapter sequences from high-throughput sequencing reads.** *EMBnet J* 2011, **17**:10–12.
31. *FASTX-Toolkit Software Package.* http://hannonlab.cshl.edu/fastx_toolkit/.
32. Jung S, Jesudurai C, Staton M, Du Z, Ficklin S, Cho I, Abbott A, Tomkins J, Main D: **GDR (Genome Database for Rosaceae): integrated web resources for Rosaceae genomics and genetics research.** *BMC Bioinformatics* 2004, **5**:130.
33. Jung S, Staton M, Lee T, Blenda A, Svancara R, Abbott A, Main D: **GDR (Genome Database for Rosaceae): integrated web-database for Rosaceae genomics and genetics data.** *Nucleic Acids Res* 2008, **36**:D1034–1040.
34. Langmead B, Trapnell C, Pop M, Salzberg SL: **Ultrafast and memory-efficient alignment of short DNA sequences to the human genome.** *Genome Biol* 2009, **10**:R25.

35. Cognat V, Pawlak G, Duchene A-M, Daujat M, Gigant A, Salinas T, Michaud M, Gutmann B, Giege P, Gobert A, Marechal-Drouard L: **PlantRNA, a database for tRNAs of photosynthetic eukaryotes.** *Nucleic Acids Res* 2012, **41**:D273–D279.
36. Hackenberg M, Sturm M, Langenberger D, Falcón-Pérez JM, Aransay AM: **miRanalyzer: a microRNA detection and analysis tool for next-generation sequencing experiments.** *Nucleic Acids Res* 2009, **37**:W68–W76.
37. Hackenberg M, Rodríguez-Ezpeleta N, Aransay AM: **miRanalyzer: an update on the detection and analysis of microRNAs in high-throughput sequencing experiments.** *Nucleic Acids Res* 2011, **39**:W132–138.
38. Anders S, Huber W: **Differential expression analysis for sequence count data.** *Genome Biol* 2010, **11**:R106.
39. Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ: **miRBase: microRNA sequences, targets and gene nomenclature.** *Nucleic Acids Res* 2006, **34**:D140–D144.
40. Griffiths-Jones S, Saini HK, van Dongen S, Enright AJ: **miRBase: tools for microRNA genomics.** *Nucleic Acids Res* 2008, **36**:D154–D158.
41. Kozomara A, Griffiths-Jones S: **miRBase: integrating microRNA annotation and deep-sequencing data.** *Nucleic Acids Res* 2011, **39**:D152–D157.
42. Kozomara A, Griffiths-Jones S: **miRBase: annotating high confidence micro-RNAs using deep sequencing data.** *Nucleic Acids Res* 2014, **42**:D68–D73.
43. Visser M, van der Walt AP, Maree HJ, Rees DJG, Burger JT: **Extending the sRNAome of apple by next-generation sequencing.** *PLoS One* 2014, **9**:e95782.
44. Xia R, Zhu H, An Y, Beers EP, Liu Z: **Apple miRNAs and tasiRNAs with novel regulatory networks.** *Genome Biol* 2012, **13**:R47.
45. Xia R, Meyers BC, Liu Z, Beers EP, Ye S, Liu Z: **MicroRNA superfamilies descended from miR390 and their roles in secondary small interfering RNA biogenesis in eudicots.** *Plant Cell* 2013, **25**:1555–1572.
46. Zhang C, Li G, Zhu S, Zhang S, Fang J: **tasiRNAdb: a database of ta-siRNA regulatory pathways.** *Bioinforma Oxf Engl* 2014, **30**:1045–1046.