



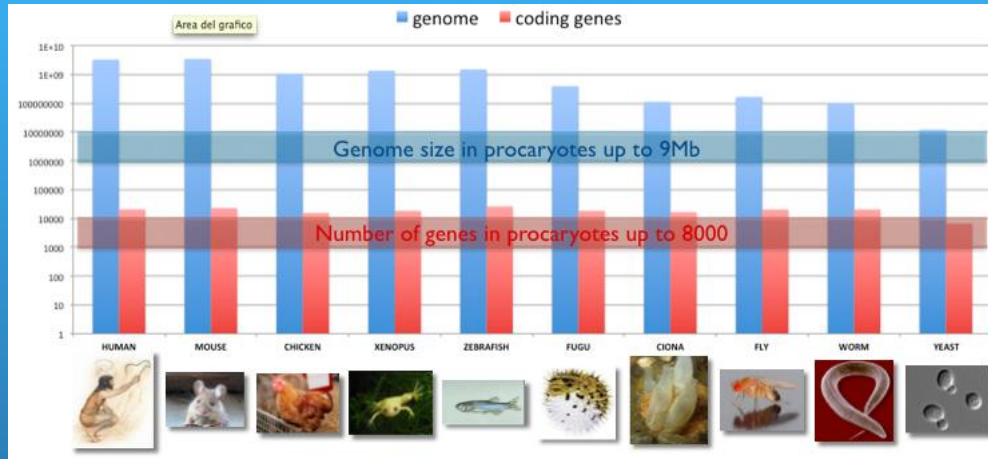
Platform for functional ncRNA analysis of NGS data

Andreas Gisel

CNR - ITB

Bari

The RNA World



non-coding RNA (ncRNA) genes, which produce transcripts that function as structural, catalytic or regulatory RNAs

- ✓ MicroRNA (miRNA) : post-transcriptional regulatory genes
- ✓ PIWI-interacting RNA (piRNA): germline transposon silencing
- ✓ Small interfering RNA (siRNA) : active molecules in RNA interference
- ✓ Small nuclear RNA (snRNA) : includes spliceosomal RNAs.
- ✓ Small nucleolar RNA (snoRNA): most known snoRNA are involved in rRNA modification
- ✓ Long non-coding RNA (lncRNA) : Little is known about them, involved in mRNA regulation

The RNA World

1986

Proc. Natl. Acad. Sci. USA
Vol. 83, pp. 5372-5376, August 1986
Biochemistry

Inhibition of gene expression in plant cells by expression of antisense RNA

(chimeric genes/electroporation/plant transformation/transient chloramphenicol) *The Plant Cell*, Vol. 2, 279-289, April 1990 © 1990 American Society of Plant Physiologists

JOSEPH R. ECKER AND RONALD W. DAVIS

Department of Biochemistry, Stanford University School of Medicine, Stanford, CA 94305

Introduction of a Chimeric Chalcone Synthase Gene into *Petunia* Results in Reversible Co-Suppression of Homologous Genes *in trans*

Carolyn Napoli,¹ Christine Lemieux, and Richard Jorgensen²

DNA Plant Technology Corporation, 6701 San Pablo Avenue, Oakland, California 94608

1990



Molecular Microbiology (1992) 6(22), 3343-3353

1992

Quelling: transient inactivation of gene expression in *Neurospora crassa* by transformation with homologous sequences

Nicoletta Romano and Giuseppe Macino*

Dipartimento di Biopatologia Umana, Sezione di Biologia Cellulare, Policlinico Umberto 1, Università di Roma 'La Sapienza', 00161 Rome, Italy.

whether these are tandemly arranged or located on different chromosomes (Faugeron *et al.*, 1990; Selker, 1990). Pre-meiotic inactivation appears to involve at least two different steps: an initial interaction between homologous sequences followed by sequence modifications, either cytosine methylation as in *A. immersus*, or both methyl-

The RNA World

1993

The *C. elegans* Heterochronic Gene *lin-4* Encodes Small RNAs with Antisense Complementarity to *lin-14*

Rosalind C. Lee,^{*†} Rhonda L. Feinbaum,^{*†} and Victor Ambros[†]

Harvard University
Department of Cellular and Developmental Biology
Cambridge, Massachusetts 02138

Summary

lin-4 is essential for the normal temporal control of diverse postembryonic developmental events in *C. elegans*. *lin-4* acts by negatively regulating the level of LIN-14 protein, creating a temporal decrease in LIN-14 protein starting in the first larval stage (L1). We have cloned the *C. elegans lin-4* locus by chromosomal walking and transformation rescue. We used the *C. elegans* clone to isolate the gene from three other *Caenorhabditis* species; all four *Caenorhabditis* clones functionally rescue the *lin-4* null allele of *C. elegans*. Comparison of the *lin-4* genomic sequence from these four species and site-directed mutagenesis of potential open reading frames indicated that *lin-4* does not encode a protein. Two small *lin-4* transcripts of approximately 22 and 61 nt were identified in *C. elegans* and found to contain sequences complementary to a repeated sequence element in the 3' untranslated region (UTR) of *lin-14* mRNA, suggesting that *lin-4* regulates *lin-14* translation via an antisense RNA-RNA interaction.

Ambros and Horvitz, 1987). A loss-of-function (*lf*) mutation, *lin-4(lf)*, causes the worm to adopt early fates at inappropriately early cell lineage patterns normally initiated at later stages, and the worm molts (Chalfie et al., 1981). The *lin-4* heterochronic developmental phenotype is the presence of adult structures (such as the vulva) and the prevention of L1-specific programs. *lin-4* null (*0*) mutations cause a phenotype similar to that of *lin-4(lf)* and are completely rescued by a wild-type *lin-4*. This is consistent with *lin-4* acting as a repressor of *lin-14* (Ambros and Horvitz, 1987). *lin-4* mutants skip the expression of L1-specific programs and precociously execute programs normally associated with L3, L4, and adult stages. *lin-4* mutants exhibit a phenotype that is virtually identical to that of *lin-4(lf)* (Ambros and Horvitz, 1987). These observations indicate that *lin-4* develops a high level of *lin-14* mRNA in the late L1 stage, which specifies L1-specific programs. Thus, the normal development of *C. elegans* requires the downregulation of L1 programs to later stages. The temporal decrease in *lin-14* mRNA on the *lin-4*-dependent decrease in LIN-14 protein is essential for the normal development of *C. elegans*.

The temporal decrease in *lin-14* mRNA is caused by a decrease in the level of LIN-14 protein. LIN-14 is normally abundant in the nucle-

The RNA World



Cell, Vol. 81, 611–620, May 19, 1995, Copyright © 1995 by Cell Press

***par-1*, a Gene Required for Establishing Polarity in *C. elegans* Embryos, Encodes a Putative Ser/Thr Kinase That Is Asymmetrically Distributed**

Su Guo and Kenneth J. Kemphues
Section of Genetics and Development
Cornell University
Ithaca, New York 14853

the ZC22 cDNA may be derived from *par-1*.

Phenocopying par-1 in Wild Type via Antisense RNA Injection

Germline transformation rescue with a genomic clone containing only the gene of interest has been routinely used

1995

The RNA World

scientific correspondence

Nature 385, 781 - 782 (27 February 1997); doi:10.1038/385781a0

Plants combat infection by gene silencing

1997

SIMON N. COVEY, NADIA S. AL-KAFF, AMAGOIA LÂNGARA & DAVID S. TURNER

Department of Virus Research, John Innes Centre, Norwich Research Park, Colney, Norwich NR4 7UH, UK e-mail: covey@bbsrc.ac.uk

A Similarity Between Viral Defense and Gene Silencing in Plants

Frank Ratcliff, Bryan D. Harrison, David C. Baulcombe*

Gene silencing in plants, in which an endogenous gene is suppressed by introduction of a related transgene, has been used for crop improvement. Observations that viruses are potentially both initiators and targets of gene silencing suggested that this phenomenon may be related to natural defense against viruses. Supporting this idea, it was found that nepovirus infection of nontransgenic plants induces a resistance mechanism that is similar to transgene-induced gene silencing.

Virus-induced gene silencing (VIGS)

Post transcriptional gene silencing (PTGS)

The RNA World

Cell, Vol. 90, 479–490, August 8, 1997, Copyright ©1997 by Cell Press

Cosuppression in *Drosophila*: Gene Silencing of *Alcohol dehydrogenase* by *white-Adh* Transgenes Is *Polycomb* Dependent

1997

Manika Pal-Bhadra,* Utpal Bhadra,*
and James A. Birchler†
Division of Biological Sciences
117 Tucker Hall
University of Missouri
Columbia, Missouri 65211

Summary

When two to six copies of a *white* promoter-*Alcohol dehydrogenase* (*Adh*) reporter fusion gene are introduced into the genome, the expression is progressively reduced both in larvae and adults rather than the expected gene dosage effect. In addition, multiple transgenes reduce endogenous *Adh* transcripts, a result that is strongly analogous to “cosuppression” phenomena described in many plant species but which has not been previously observed in animals. Silencing of the *Adh* gene is not influenced by *zeste*-dependent transvection but strongly affected by the *Polycomb* and *Polycomblike* mutations. *Polycomb* and *polyhomeotic* proteins are bound to the chromatin at the sites of the repressed *w-Adh* transgenes.



The RNA World

Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*

Andrew Fire*, SiQun Xu*, Mary K. Montgomery*, Steven A. Kostas*†, Samuel E. Driver‡ & Craig C. Mello‡

* Carnegie Institution of Washington, Department of Embryology, 115 West University Parkway, Baltimore, Maryland 21210, USA

† Biology Graduate Program, Johns Hopkins University, 3400 North Charles Street, Baltimore, Maryland 21218, USA

‡ Program in Molecular Medicine, Department of Cell Biology, University of Massachusetts Cancer Center, Two Biotech Suite 213, 373 Plantation Street, Worcester, Massachusetts 01605, USA

Experimental introduction of RNA into cells can be used in certain biological systems to interfere with the function of an endogenous gene^{1,2}. Such effects have been proposed to result from a simple antisense mechanism that depends on hybridization between the injected RNA and endogenous messenger RNA transcripts. RNA interference has been used in the nematode *Caenorhabditis elegans* to manipulate gene expression^{3,4}. Here we investigate the requirements for structure and delivery of the interfering RNA. To our surprise, we found that double-stranded RNA was substantially more effective at producing interference than was either strand individually. After injection into adult animals, purified single strands had at most a modest effect, whereas double-stranded mixtures caused potent and specific interference. The effects of this interference were evident in both the injected animals and their progeny. Only a few molecules of injected double-stranded RNA were required per affected cell, arguing against stoichiometric interference with endogenous mRNA and suggesting that there could be a catalytic or amplification component in the interference process.

1997



The Nobel Prize in Physiology or Medicine 2006
Andrew Z. Fire, Craig C. Mello

The Nobel Prize in Physiology or Medicine 2006

Summary [Illustrated Information](#)

Prize Announcement

Press Release

Advanced Information

Popular Information

Nobel Prize Award Ceremony

Andrew Z. Fire

Craig C. Mello



Nobelforsamlingen

The Nobel Assembly at Karolinska Institutet



Karolinska
Institutet

English
Swedish

Press Release

2 October 2006

The Nobel Assembly at Karolinska Institutet has today decided to award The Nobel Prize in Physiology or Medicine for 2006 jointly to

Andrew Z. Fire and Craig C. Mello

for their discovery of "RNA interference – gene silencing by double-stranded RNA"



The RNA World


1999

A Species of Small Antisense RNA in Posttranscriptional Gene Silencing in Plants

Andrew J. Hamilton and David C. Baulcombe*

Posttranscriptional gene silencing (PTGS) is a nucleotide sequence-specific defense mechanism that can target both cellular and viral mRNAs. Here, three types of transgene-induced PTGS and one example of virus-induced PTGS were analyzed in plants. In each case, antisense RNA complementary to the targeted mRNA was detected. These RNA molecules were of a uniform length, estimated at 25 nucleotides, and their accumulation required either transgene sense transcription or RNA virus replication. Thus, the 25-nucleotide antisense RNA is likely synthesized from an RNA template and may represent the specificity determinant of PTGS.

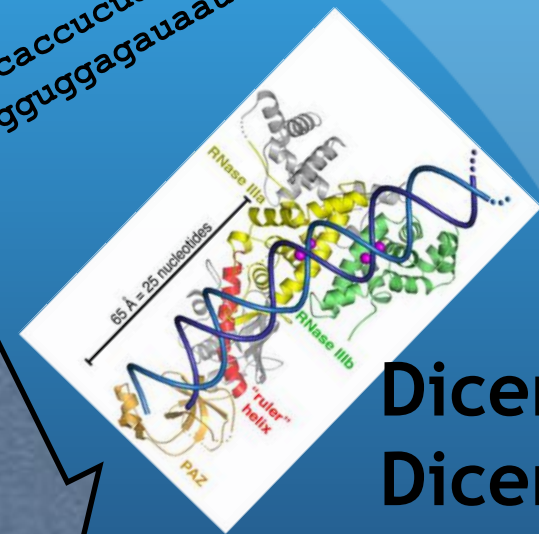
The RNA World


 transgene
 antisenes RNA


 replicating virus


 secondary structures

5' -auggcugaucaccucucuaaauagauggggccccacauauac-3'
 3' -uaacgacuaguggagagaauuaucuaaccgggguguaauaugaccacuccgguggagauaaucucucuaaccagucguuacac-5'



Dicer
Dicer-like

5' -auagauggggccccacauauac-3'

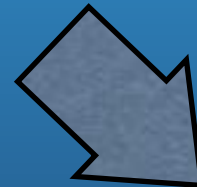
The RNA World

5' -auagaugggccccacauauac-3'

small RNA

RISC

RNA induced silencing complex



block



3' -uaacgacuaguggagagauuauaucua...uaugaccacuccgguggagauaau...

cleave



acgacuaguggagagauuauaucuaccgg...gguguauaugaccacuccgguggagauaauaucucucuaccagucguuacac-5'



The RNA World

Why non-coding RNA?

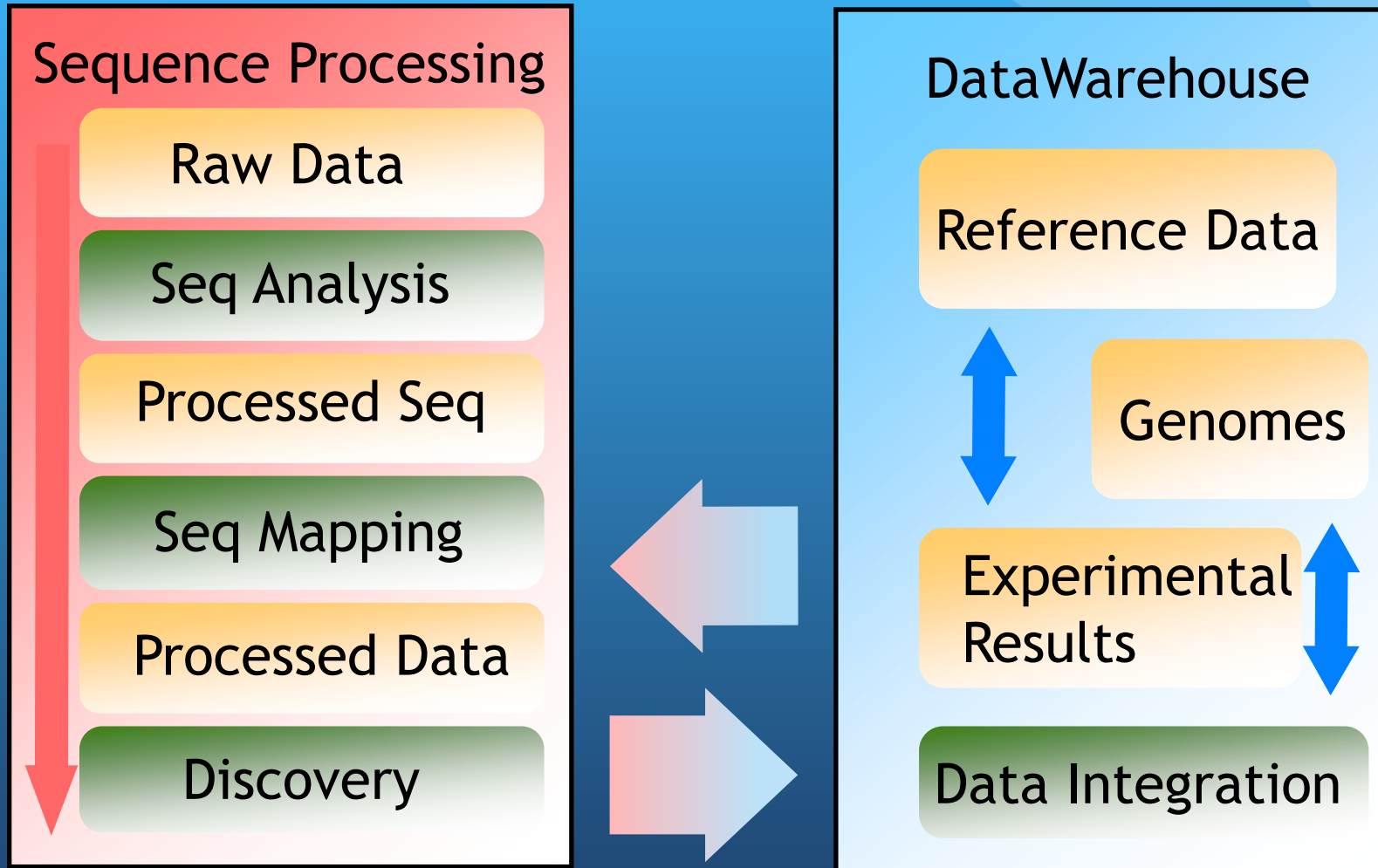
- How many ncRNA genes are there?
- How important are they?
- What functions does a cell 'delegate' to RNA instead of protein and why?

To address these questions, new systematic gene-discovery approaches need to be developed that are specifically aiming the discovery ncRNAs

Next Generation Sequencing (NGS) technologies with the vast amount and sensitivity enable such a research.

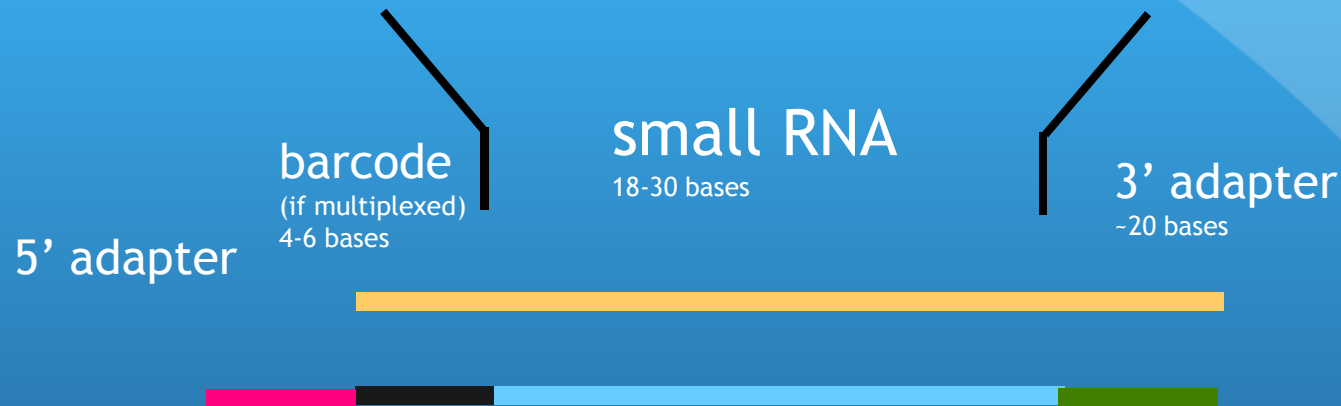
The new spectrum of NGS applications together with the massive amount of data requires a focused investment and development of bioinformatics tools managing and analysing such complex and large datasets to infer biological meaning.

The ITB ncRNA Platform



Sequence Processing

Small RNA Sequence Analysis

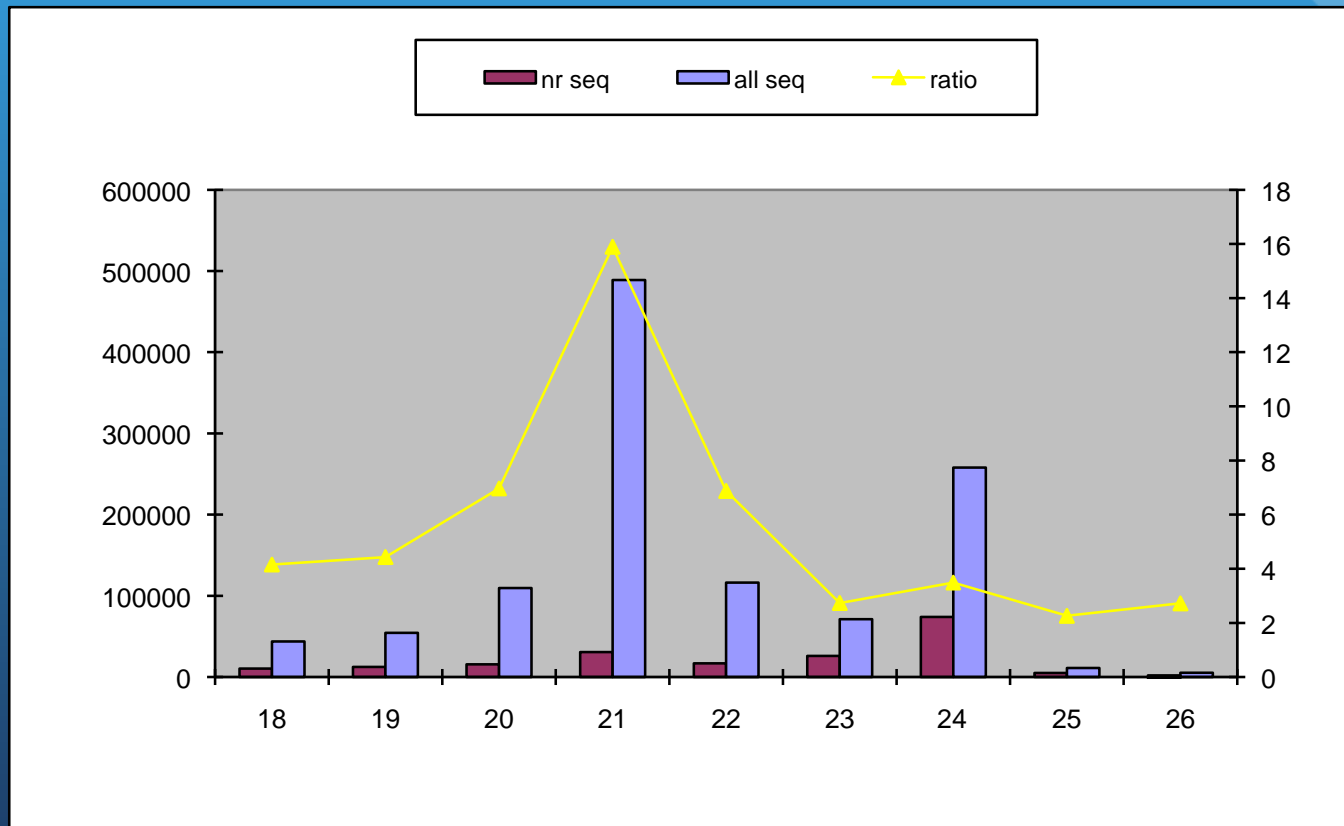


Reads are typically around 50 bases and therefore read into the 3'-adaptor

- ✓ If multiplexed detect barcode sequence and separate experiments
- ✓ Trimming barcode and 3'-adaptor fragment (adaptor sequence not always present since "small RNA" too long ==> discharge!!!)
- ✓ Since sequencing errors tends to be more frequent in 3' part, 3'-adaptor trimming need to accept mismatches

Sequence Processing

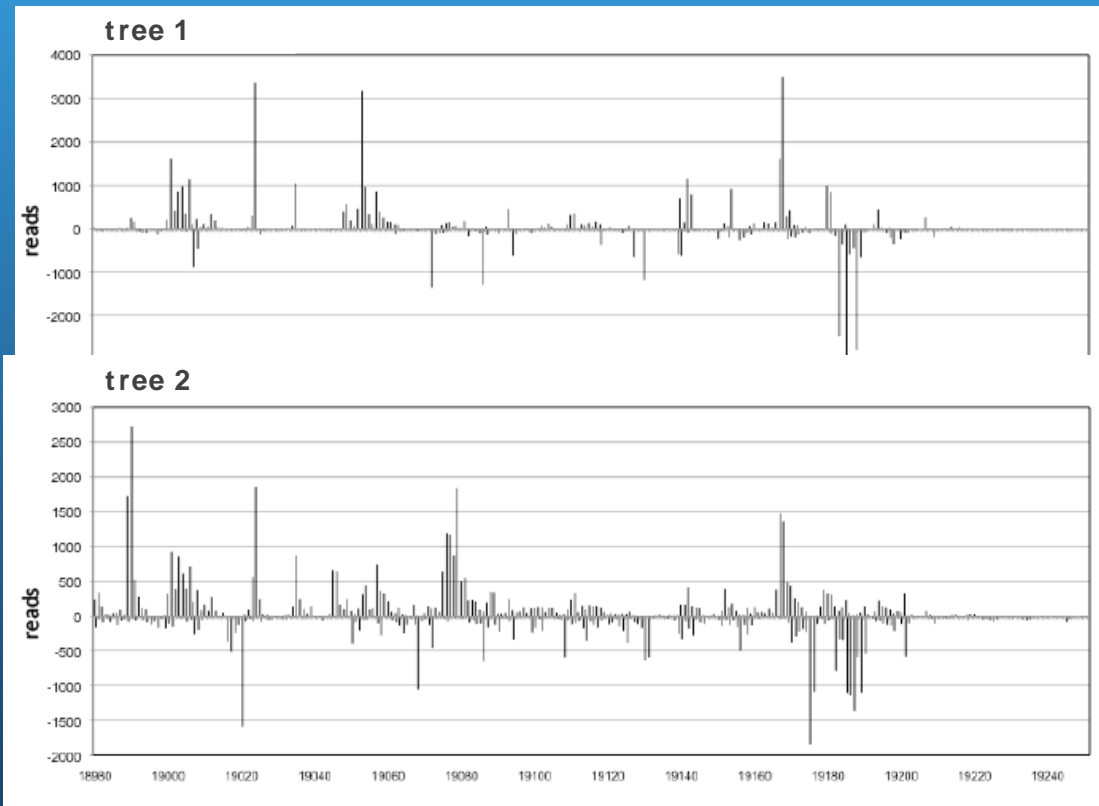
Sequence Analysis



Sequence Processing

Sequence Mapping

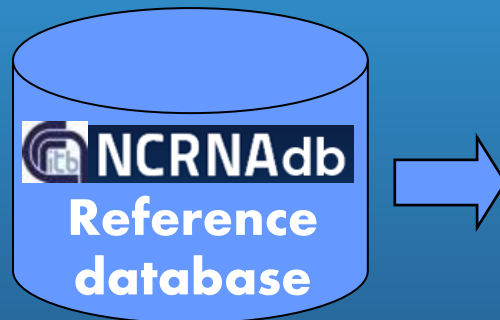
BWA mapping on Genome(s)



Sequence Processing

Sequence Mapping

BWA mapping on ncRNA Reference

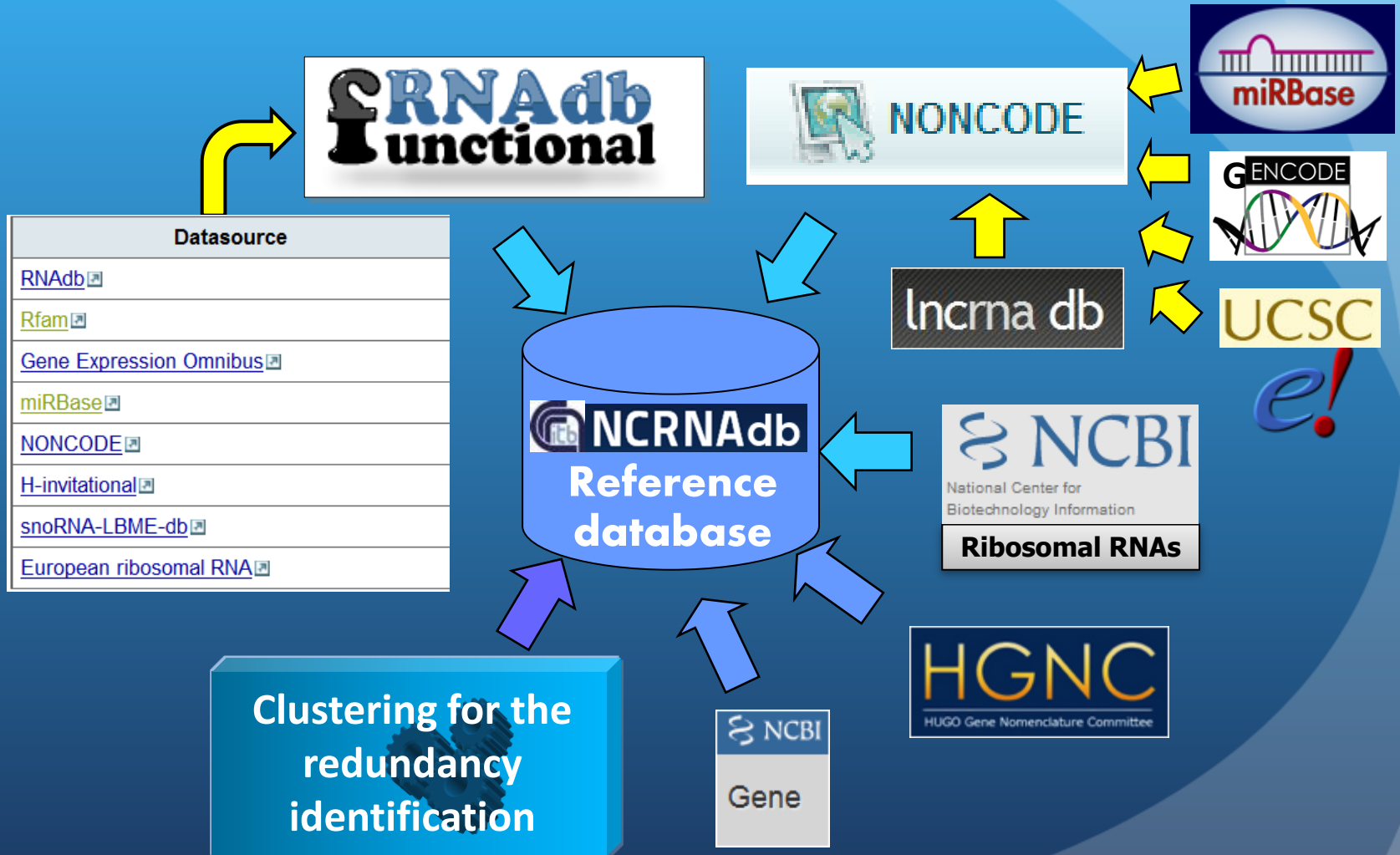


```

1524329 *
551158 n3435 |miRNA |Homo
452969 n408449 |lncRNA |Homo
452890 n402469 |lncRNA |Homo
452670 n121773 |piRNA |Homo
452589 n203154 |pre_miRNA |Monodelphis
452347 n200513 |pre_miRNA |Pongo
452164 n200465 |pre_miRNA |Macaca
451608 n201030 |pre_miRNA |Homo
451384 n387643 |lncRNA |Homo
411580 n197338 |pre_miRNA |Monodelphis
411074 n201230 |pre_miRNA |Pan
411030 n200698 |pre_miRNA |Bos
410589 n200478 |pre_miRNA |Gorilla
410543 n408444 |lncRNA |Homo
410430 n414739 |lncRNA |Mus
410385 n199475 |pre_miRNA |Macaca
410251 n423547 |lncRNA |Mus
410237 n204227 |pre_miRNA |Rattus
410118 n200501 |pre_miRNA |Mus
409989 n387171 |lncRNA |Homo
409446 n201230 |pre_miRNA |Homo
387835 n200574 |miRNA |Macaca
386659 n204819 |miRNA |Monodelphis
262091 n205000 |pre_miRNA |Pan
261353 n205000 |pre_miRNA |Macaca
252674 n331895 |miRNA |Macaca
252282 n331894 |miRNA |Pan
252115 n200574 |miRNA |Homo
251582 n331896 |miRNA |Gorilla
250300 n200574 |miRNA |Gorilla
  
```


DataWarehouse

“ncRNADB” Reference Database



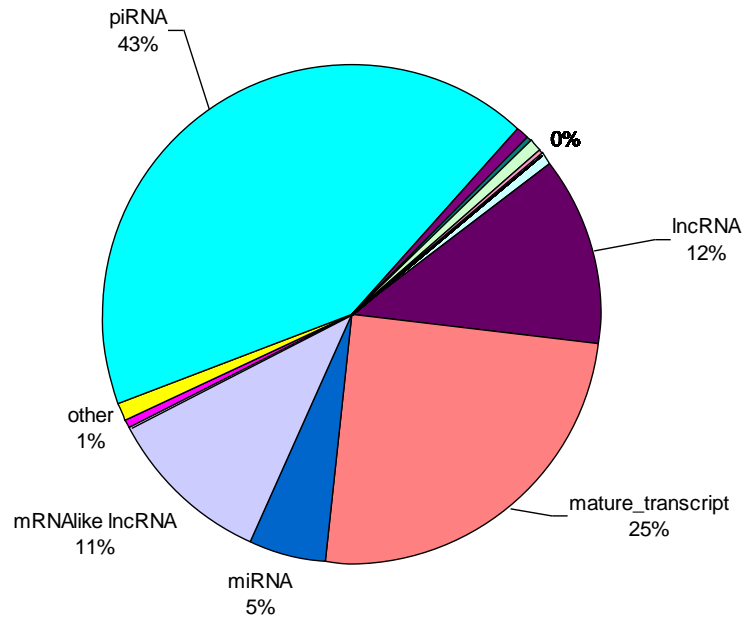
DataWarehouse

Some “ncRNADB” Statistics

Total sequences: 411552



Classes



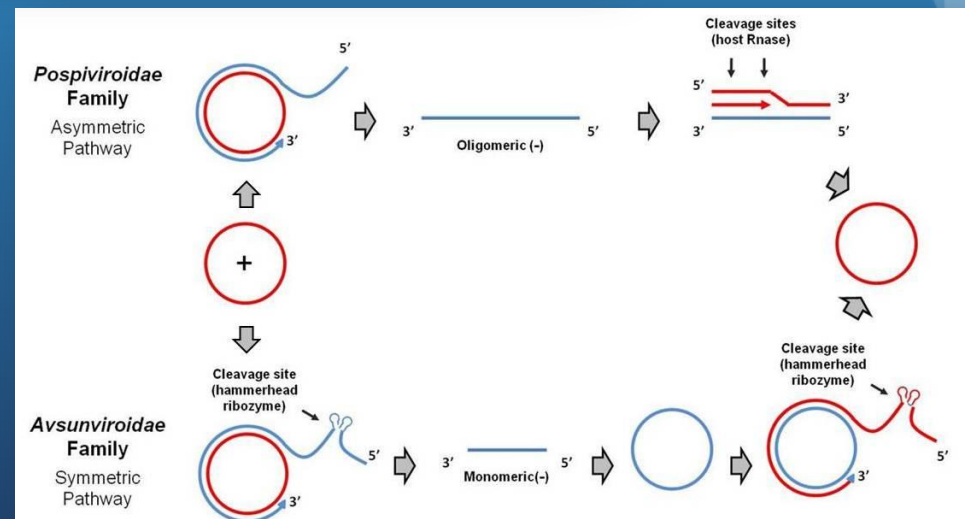
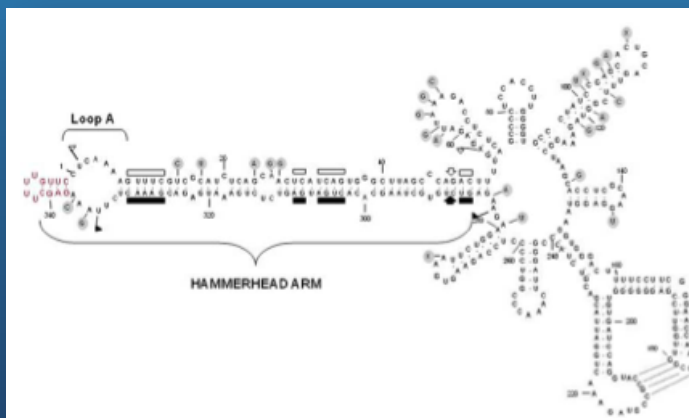
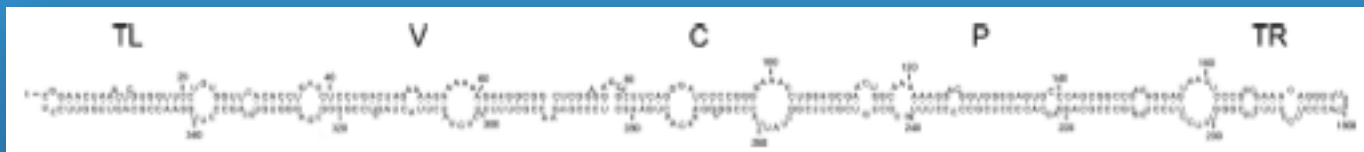
- 4.5S
- antis
- gRN
- lincR
- lncR
- mat
- miR
- mRN
- ncR
- Non
- othe
- piRN
- pre.
- Rep
- RNA
- scR
- self
- snm
- sno
- snR
- SRP
- SRP
- telor
- tmR
- VA
- Y R

organism	total
Mus musculus	119777
Drosophila melanogaster	102581
Homo sapiens	91286
Rattus norvegicus	66760
Caenorhabditis elegans	4900
Oryza sativa	869
Arabidopsis thaliana	781
Danio rerio	454
Populus trichocarpa	292
Xenopus tropicalis	275
Pan paniscus	274
Caenorhabditis briggsae	271
Pongo pygmaeus	263
Gorilla gorilla	263
Gallus gallus	256
Pan troglodytes	254
Macaca nemestrina	231
Takifugu rubripes	221
Macaca mulatta	221
Bos taurus	211
Tetraodon nigroviridis	209
Saccharomyces cerevisiae	187
Zea mays	165
Monodelphis domestica	159
Lagothrix lagotricha	145

Results

Viroid plant interaction

Viroids ("virus-like") are small circular RNA plant pathogens. They do not code for any protein but can cause heavy symptoms in some ornamental and crop plants.



Results

Viroid plant interaction

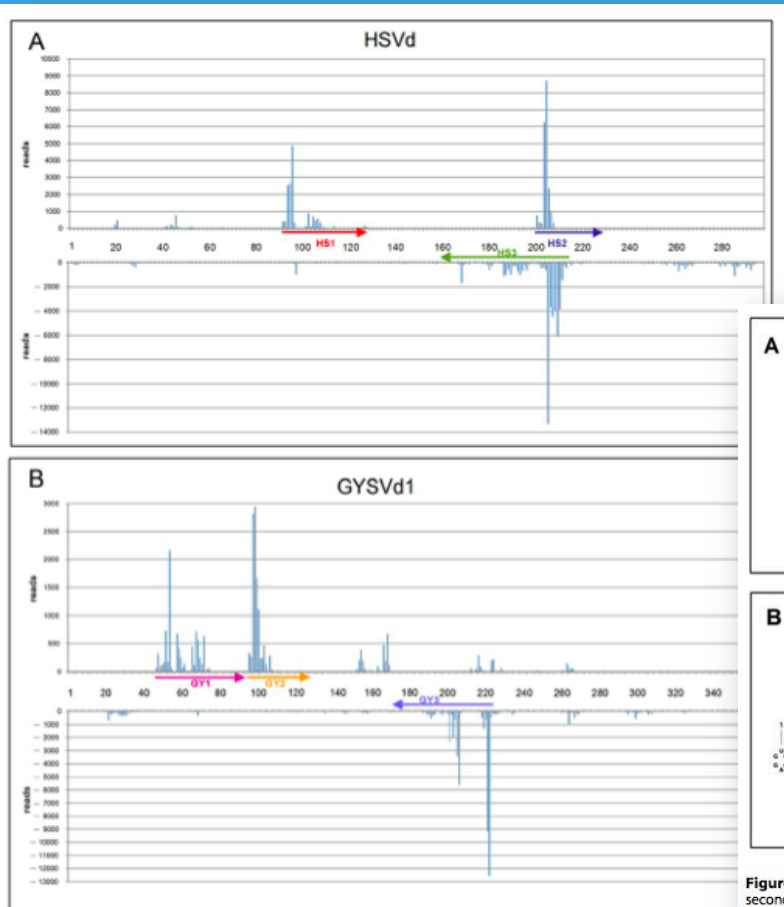


Figure 4. Most HSVd- and GYSVd1-sRNAs derive from restricted regions of the genomic (+) and (-) RNAs. Location of the 5' and frequency of the HSVd-sRNAs (A) and GYSVd1-sRNAs (B) from tendrils in their corresponding (+) and (-) genomic RNAs. Positives correspond to vd-sRNAs of (+) polarity and negative values correspond to the vd-sRNAs of (-) polarity. Note that the scale is different in the and that the same numbers are used in the (+) polarity (5'→3' orientation is from left to right) and in the (-) polarity (5'→3' orientation is from right to left). For the location of the 5' termini of vd-sRNAs we have considered the HSVd and GYSVd1 sequence variant with the accession n X06873 [48] and GQ995473, respectively. The viroid sequences covered by vd-sRNAs following to hotspots (HS1, HS2 and HS3 for HSVd and GY1 and GY3 for GYSVd1) are denoted by arrows whose sense indicates 5'→3' orientation. doi:10.1371/journal.pone.0007686.g004

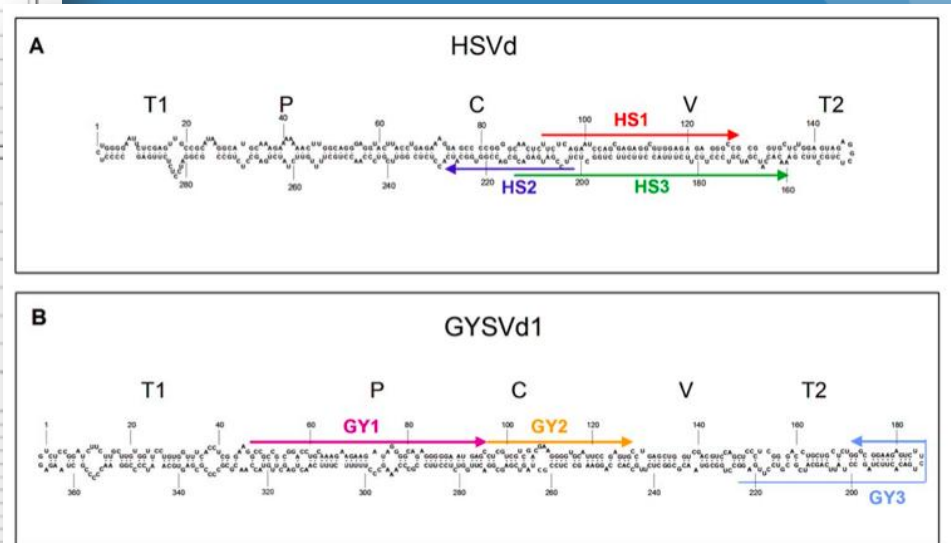


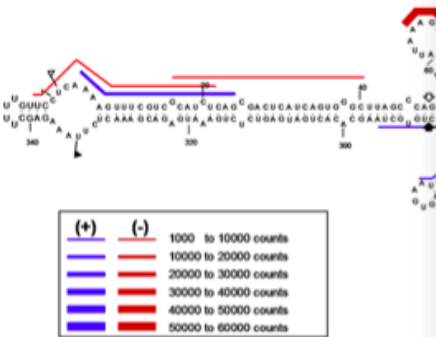
Figure 5. HSVd-sRNAs and GYSVd1-sRNAs do not cover the same viroid structural domains. Sequence and computer-predicted secondary structure for the (+) strand of the HSVd (sequence variant X06873) (A) and the GYSVd1 (sequence variant GQ995473) (B), corresponding to the master and the consensus variants in the grapevine sequenced viroid populations, respectively. The viroid sequences covered by vd-sRNAs corresponding to hotspots (HS1, HS2 and HS3 for HSVd, and GY1, GY2 and GY3 for GYSVd1) are denoted by arrows whose sense indicates 5'→3' orientation. The position of five structural domains proposed for PSTVd and closely-related viroids [80] are indicated (P: pathogenic; V: variable; C: central; T1: terminal left; T2: terminal right), although no data on the functional properties of these regions in HSVd and GYSVd1 are available. The secondary structures were obtained by the program Mfold [81]. doi:10.1371/journal.pone.0007686.g005

Results

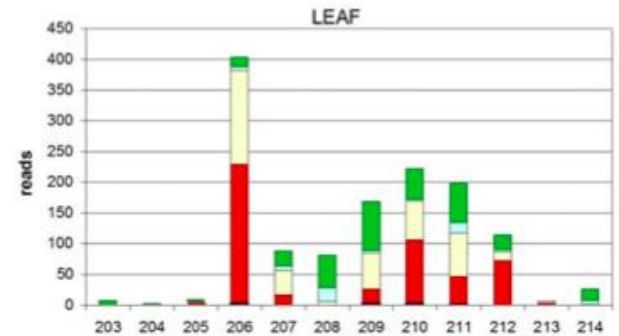
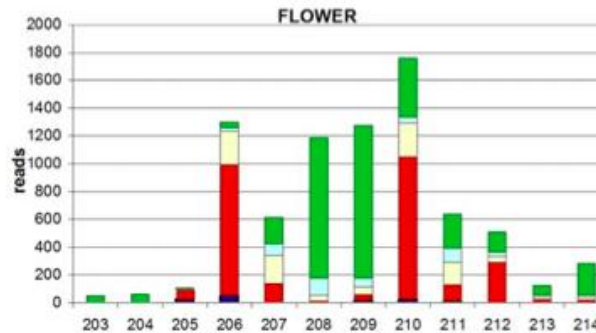
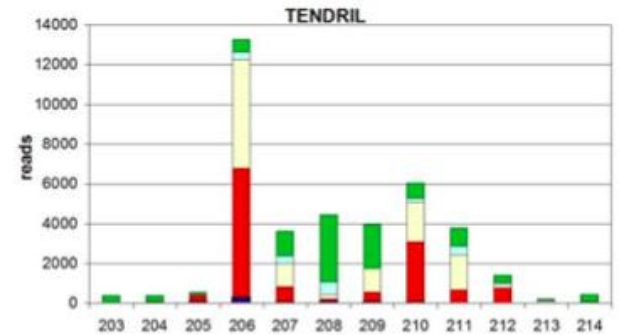
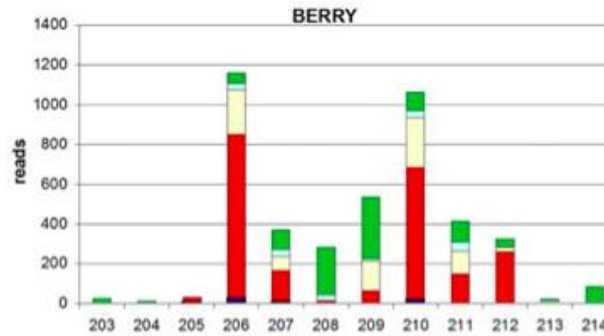
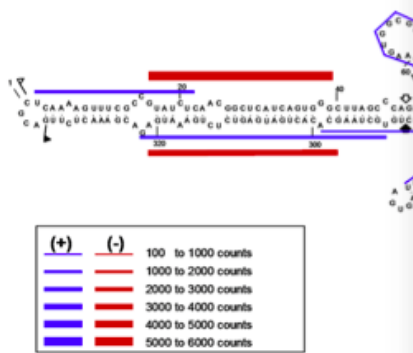
Viroid plant interaction



PC-C40



GDS6



■ 20 ■ 21 ■ 22 ■ 23 ■ 24

Figure 6. Specific size-classes vd-sRNAs may largely prevail at certain genomic positions. Histograms comparing location of the 5' termini, frequency and size distribution of (-) vd-sRNAs corresponding to the HSVd hotspot 3 (HS3) and recovered from the different grapevine tissues. Numbers are referred to HSVd sequence variant with the accession number X06873. 5'→3' orientation is from right to left.
doi:10.1371/journal.pone.0007686.g006

Figure 6. Location and frequency of the 5' termini of the PLMVd-sRNA. Note that the scale of counts is different in both samples and that only frequent included. The PLMVd-sRNAs of both polarities are referred to the secondary thermodynamic, *in vitro*, and *in vivo* data.
doi:10.1371/journal.pone.0007539.g006

Results

Viroid smallRNA in RDR6 silenced N.benth

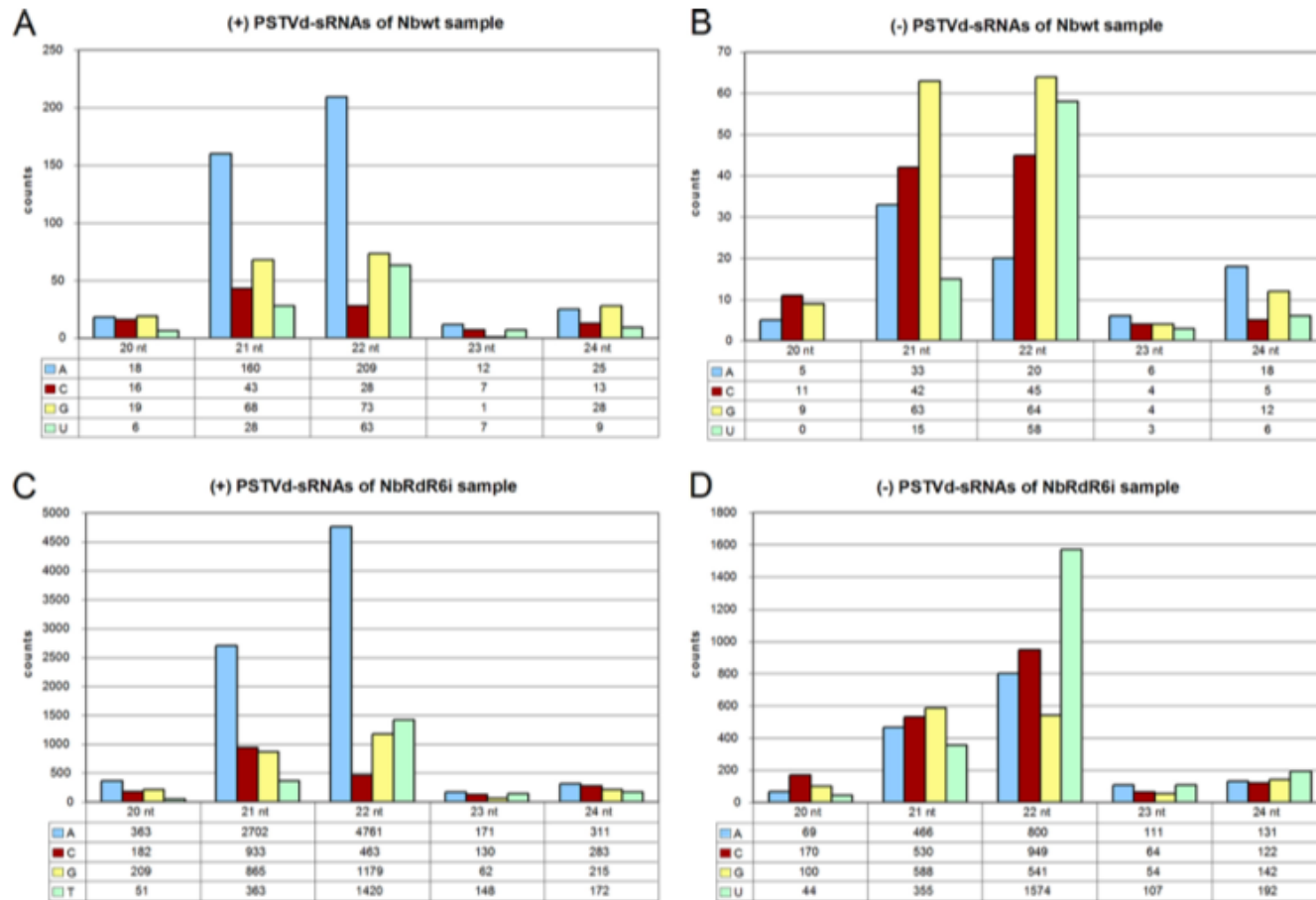


FIG. 6. (unique) p

FIG. 8. Biased distribution of the 5' nucleotide in plus- and minus-strand PSTVd sRNAs. The histograms compare in NbwT (A and B) and NbRDR6i (C and D) the total counts corresponding to PSTVd sRNAs (20 to 24 nt) of plus polarity (A and C) and minus polarity (B and D) with different 5' termini.

nonredundant
RDR6i samples.

Results



CTV infected orange plants

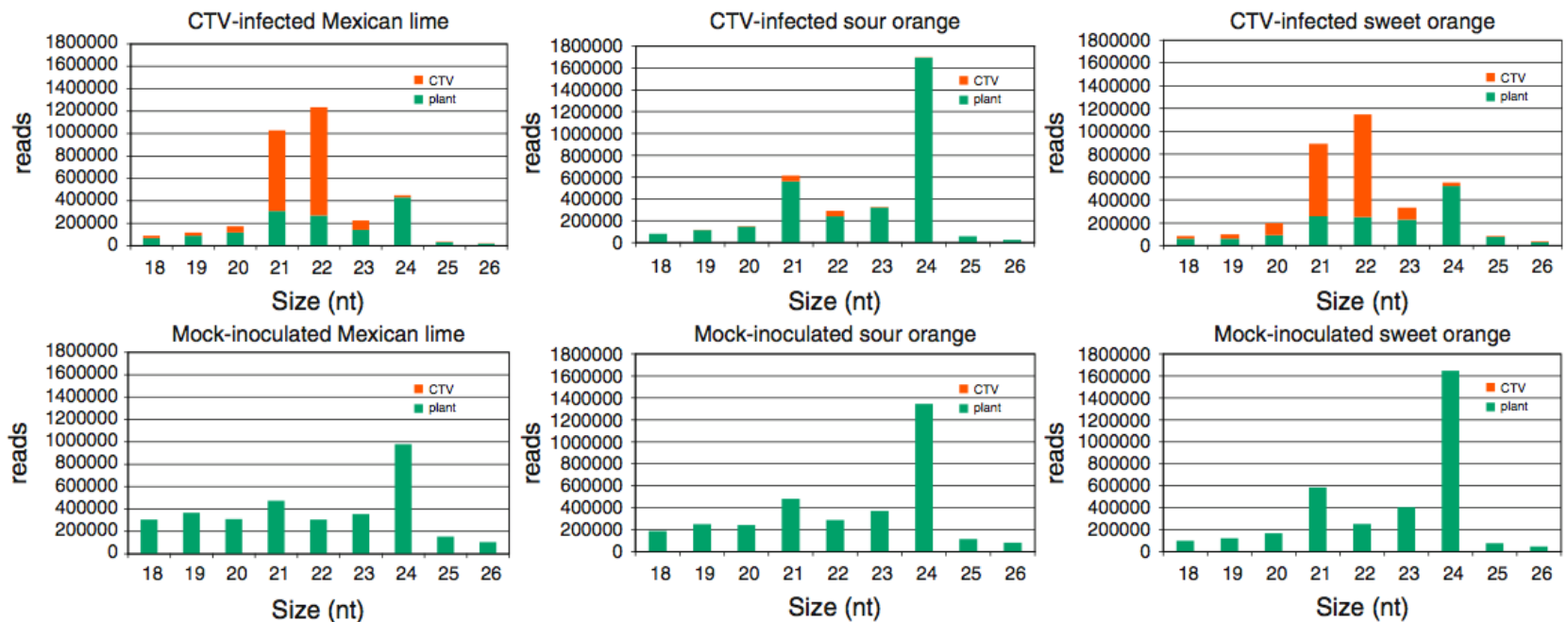
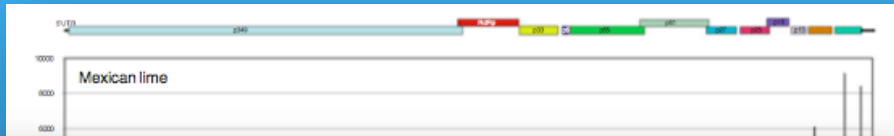


Fig. 1 Size distribution of CTV and plant sRNAs (*orange* and *green*, respectively) in CTV-infected and mock-inoculated Mexican lime, sour orange and sweet orange. The histograms compare the distribution of

18–26-nt total sRNA reads. A minor contamination of CTV-sRNAs observed in the mock-inoculated Mexican lime (representing 0.1% of the total sRNAs of this size) has not been represented

Results



CTV infected orange plants

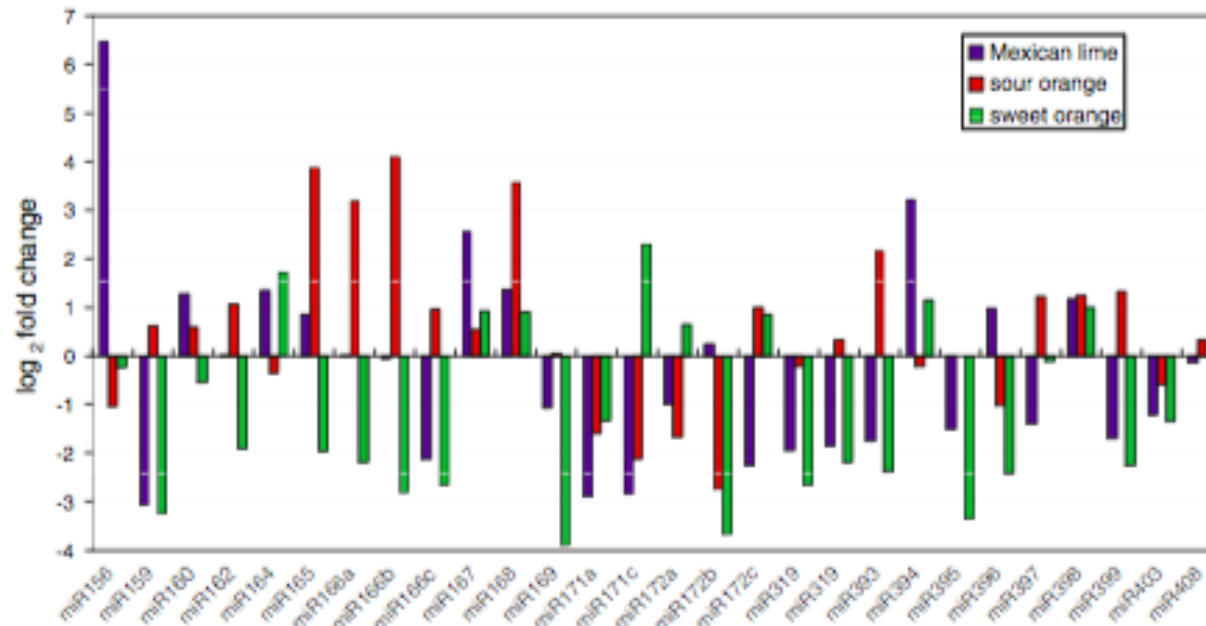


Fig. 8 Effect of CTV infection on the accumulation of specific miRNA from Mexican lime, sour orange and sweet orange. Only miRNAs with more than 10 reads in at least one of the three citrus hosts are considered, and the expression levels of miRNAs are plotted as \log_2 of fold change (reads of CTV-infected versus mock-inoculated samples). Some miRNA sequences predicted by comparative analyses (Song et al. 2007) exhibited minor variations with respect to those retrieved from our deep sequencing libraries. In particular: (i) the miR156 and miR393 most frequently sequenced in Mexican lime, sweet and sour orange have an additional nucleotide at the 5'-end

when compared with the sequence deposited in the miRbase and predicted, respectively, (ii) the most abundant miR166 and miR172 have their 5' termini shifted one or two nucleotides downstream with respect to those predicted and deposited in the miRbase, respectively, and (iii) the prevalent miR166 in the three citrus species analyzed is 21-nt long and does not contain the additional 3'-nucleotide proposed in some members of this family (csi-miR166a, csi-miR166b, ctv-miR166, crt-miR166a, crt-miR166b) deposited in the miRbase (see also Supplementary Table S3)

Results



Mouse miRNA fold-change (graphic interface)

CTRL-T1-T2 (pvalue <= 0,05) [Export](#) [Print all pages](#) [Print current page](#)

1 2 ... 5 [Define page size](#)

Refresh

CTRL-T1-T2 (pvalue <= 0,05) [Export](#) [Print all pages](#) [Print current page](#)

CTRL => T1 [Export](#) [Print all pages](#) [Print current page](#)

1 2 ... 10 [Define page size](#)

Refresh Quick search

+	Actions	Sequence	miRNA 1 ID	Experiment	Differential Expression	Fold Change	pValue	RPM Exp1	RPM Exp2	miRNA 2 ID
+		GTGGGAAGGAAGCTACAAGACAGCT	-	S1->S2	over	4.870	1.63251e-42	3.73	109.06	mmu-miR-6368
+		AGGAGAAAGGCCTCTCTCTC	mmu-miR-124-5p	S1->S2	over	2.612	0.000588264	1.68	10.28	-
+		AGGAGAAAGGCCTCTCTCTC	mmu-miR-124-5p	S1->S2	over	2.612	0.000588264	1.68	10.28	-
+		AGGAGAAAGGCCTCTCTCTC	mmu-miR-124-5p	S1->S2	over	2.612	0.000588264	1.68	10.28	-
+		AGAGGTAGTAGGTTGCATAGTTTT	mmu-let-7d-5p	S1->S2	over	1.869	0.0661547	0.00	3.65	hsa-let-7a-5p
+		AGAGGTTTTCTGGGTCTCTG	mmu-miR-329-5p	S1->S2	over	1.869	0.0661547	0.00	3.65	rno-miR-329-5p
+		AAACATTGCGGGTGCCTTC	mmu-miR-543-5p	S1->S2	over	1.745	0.0384001	1.68	5.64	mmu-miR-543-3p
+		AACTGGCCACAAAGTCCCGCTT	mmu-miR-193b-3p	S1->S2	over	1.745	0.0384001	1.68	5.64	mmu-miR-193a-3p
+		TGAAATGTTTAGGACCACTAGA	mmu-miR-203-5p	S1->S2	over	1.615	0.00000107916	13.96	42.76	-

DataWarehouse

Genomic References

Sequences and Features of:

- ✓ Homo sapiens
- ✓ Mus Musculus
- ✓ Arabidopsis thaliana

ncRNA Projects involved

- ✓ Plant - Viroid interactions in peach tree, grapevine and tobacco
- ✓ Plant - Virus interaction in orange tree
- ✓ Multidrug resistance in dog cell lines
- ✓ Immune response in mouse
- ✓ miRNA driven methylation profile in Arabidopsis

The Team



**Consiglio Arianna
De Caro Giorgio
D'Elia Domenica
Gisel Andreas
Grillo Giorgio
Licciulli Flavio
Liuni Sabino
Tulipano Angelica

Losito Nicola**

Collaboration



University of Bari

- **Computer Science**
 - Malerba Donato
 - Ceci Michelangelo
 - Corrado Loglisci
- **Dept. of Pharmacy**
 - Nicola Colabufo
 - Antonio Carrieri

CNR - IVV, Bari

- Francesco Di Serio
- Beatriz Navarro

CSIC - UPV, Valencia, Spain

- Ricardo Flores