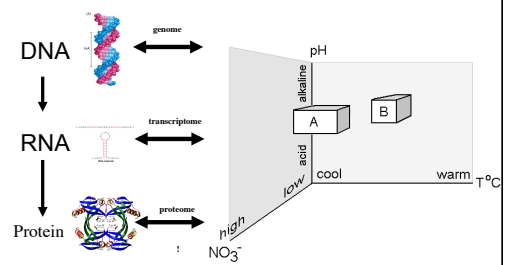
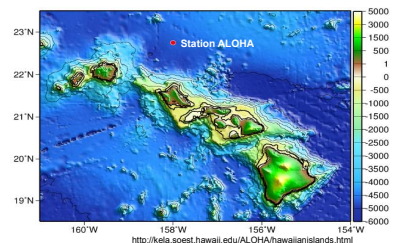


HIERARCHY OF MACROMOLECULAR INFORMATION ALSO CONTAINS INFORMATION ABOUT ENVIRONMENT & ADAPTATION

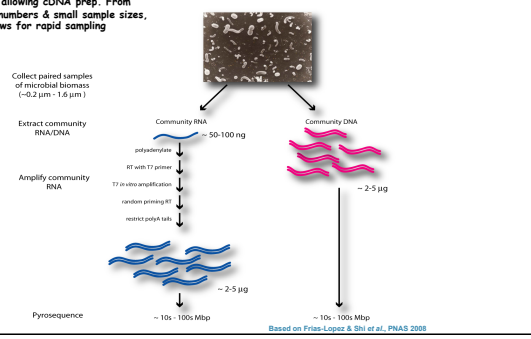


Station ALOHA - North Pacific Subtropical Gyre



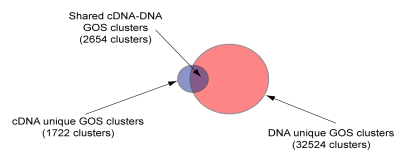
METATRANSCRIPTOMIC SURVEYS

Our initial goal was to develop a protocol allowing cDNA prep. From low cell numbers & small sample sizes, that allow for rapid sampling



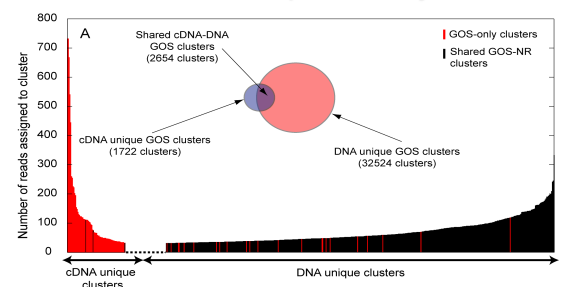
Station ALOHA - HOT179 - 75 m Metatranscriptomic analysis

Reads with homologues in existing peptide databases



Frias-Lopez and Shi et al, PNAS March 2008

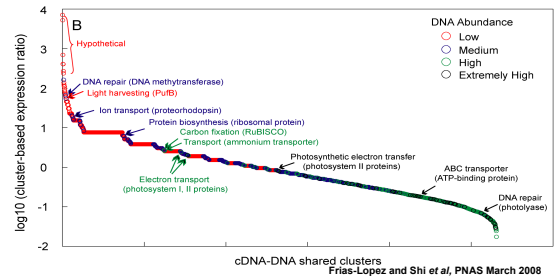
Station ALOHA - HOT179 - 75 m Metatranscriptomic analysis



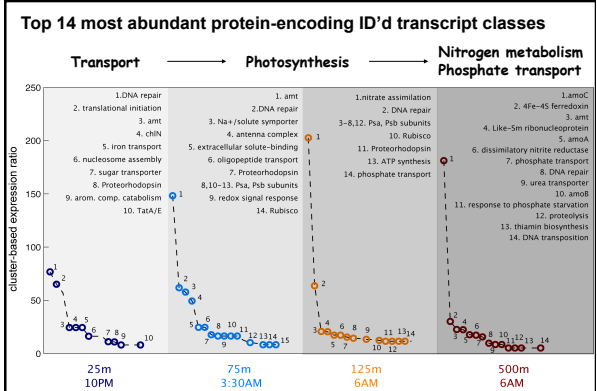
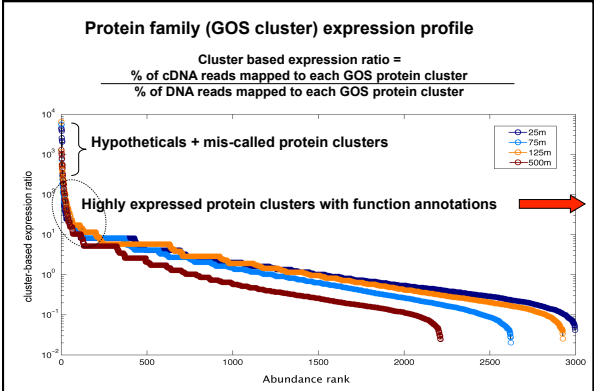
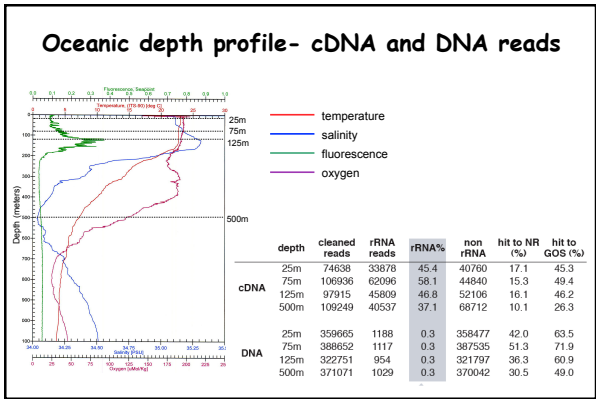
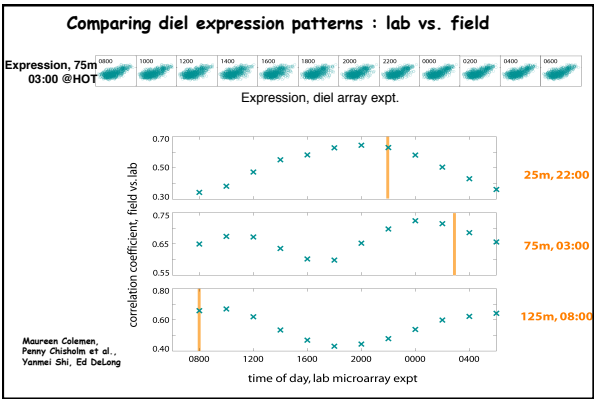
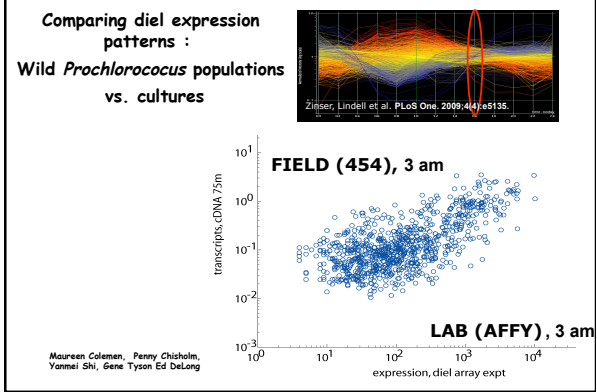
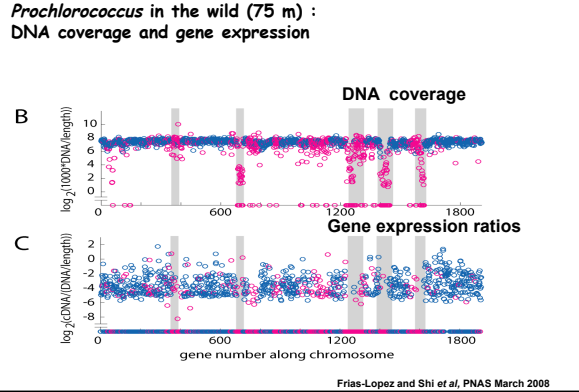
Frias-Lopez and Shi et al, PNAS March 2008

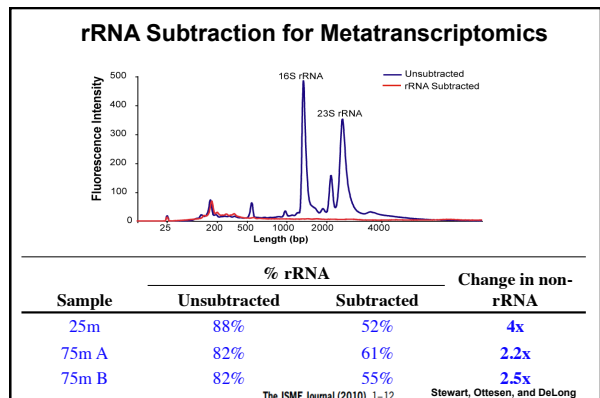
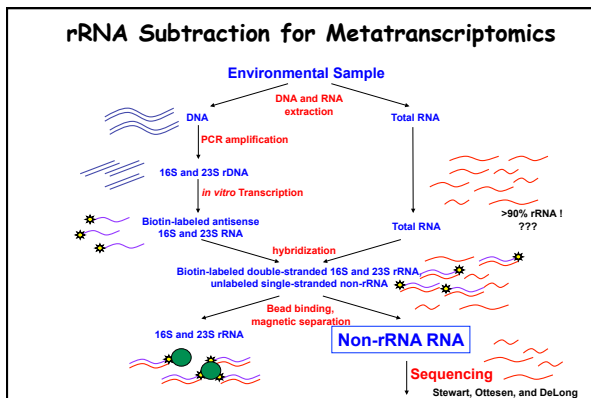
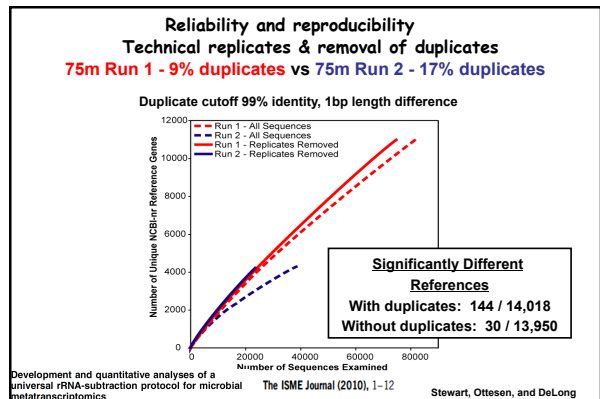
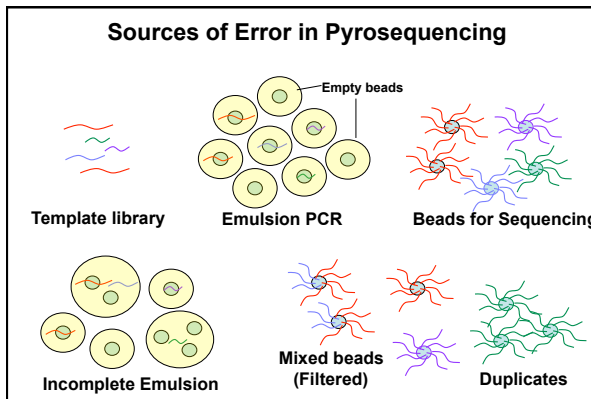
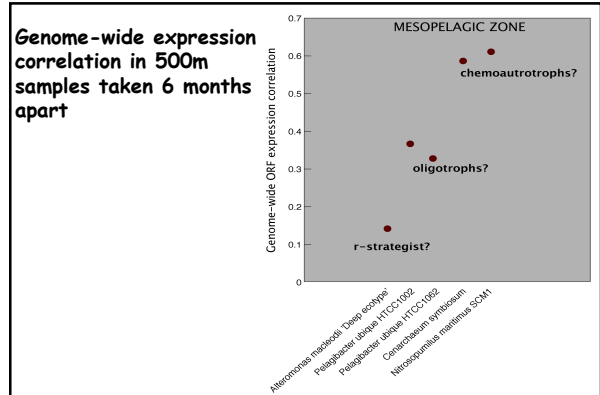
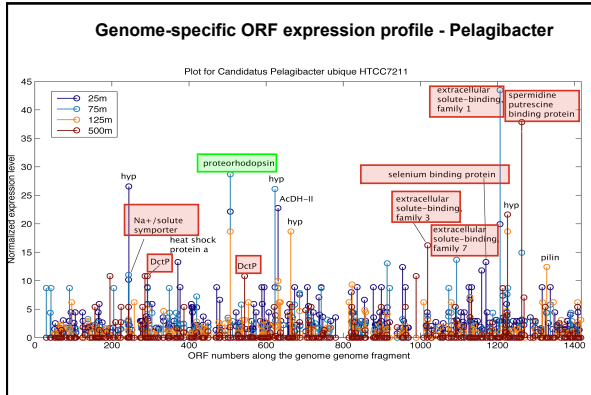
Community Gene Expression Station ALOHA H179 - 75 m

Cluster based expression ratio = number of cDNA reads mapped to each GOS protein cluster / number of DNA reads mapped to each GOS protein cluster

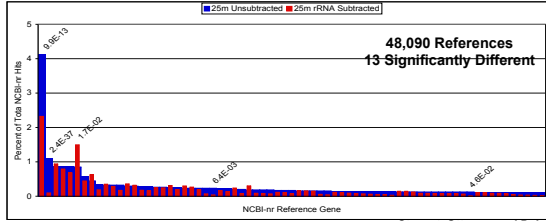


Frias-Lopez and Shi et al, PNAS March 2008





Variation in Transcript Abundance 25m Unsubtracted vs rRNA Subtracted



The ISME Journal (2010), 1-12

Reads mapped to top hit in NCBI-nr database, e-value cutoff 1×10^{-05}
References representing >0.1% of 25m Unsubtracted shown in descending abundance
P-values <0.05 marked (Audric and Claverie significance test, FDR multiple comparison test)

rRNAs in metatranscriptomic datasets after subtractive hybridization

| sample | primers ¹ | total reads | rRNA reads | % rRNA |
|------------------------------|----------------------|-------------|------------|-------------|
| Pure culture | | | | |
| <i>Dokdonia</i> sp., rep 1 | B | 630 260 | 65 339 | 10.4 |
| <i>Dokdonia</i> sp., rep 2 | B | 195 278 | 4 859 | 2.5 |
| <i>Dokdonia</i> sp., rep 3 | B | 91 437 | 10 784 | 11.8 |
| Bacterioplankton | | | | |
| Bermuda, tropical | | | | |
| 20m | B | 511 525 | 146 530 | 28.6 |
| 50m | B | 365 838 | 87 240 | 23.8 |
| 100m | B | 519 951 | 143 907 | 27.7 |
| OMZ, experimental incubation | | | | |
| OMZ t0 | BAE | 27 300 | 9 805 | 35.9 |
| OMZ t1 | BAE | 105 274 | 58 240 | 55.3 |
| OMZ t2 | BAE | 64 463 | 29 590 | 45.9 |
| Monterey Bay ² | | | | |
| 10m sample WCR3 | BAE | 248 016 | 82 932 | 33.4 |
| 10m sample WCR5 | BAE | 238 635 | 90 767 | 38.0 |
| 10m sample WCR6 | BAE | 235 339 | 82 501 | 35.1 |
| 10m sample BAC16 | BAE | 102 024 | 40 833 | 40.0 |

¹ probe: B = Bacterial 16S and 23S (primers in Table 1); A = Archaeal 16S and 23S (Table S2);

E = Eukaryotic 18S and 28S (Table S2)

² Archaeal 16S probe excluded - PCR yielded multiple bands Stewart, Ottesen, and DeLong, ISME J.

Comparison of subtracted vs. unsubtracted and replicates...

Table 3. Dataset (DS) comparisons - non-rRNA sequences mapped to non-redundant (nr) NCBI reference sequences

| DS compared ¹ | total refs ² | | refs unique to DS ³ | | % refs shared ⁴ | % reads in sig. diff. abundance ⁵ | % reads in sig. diff. refs ⁶ | |
|--------------------------|-------------------------|--------|--------------------------------|--------|----------------------------|--|---|------|
| | DS1 | DS2 | DS1 | DS2 | | | DS1 | DS2 |
| 25m S1 25m S2 | 21 011 | 33 097 | 13 261 | 25 347 | 16.7 | 0 | 0 | 0 |
| 25m U 25m S | 4 110 | 46 358 | 1 732 | 43 980 | 4.9 | 13 | 7.0 | 4.0 |
| 75m 75m | | | | | 9.3 | | | |
| AS1* AS2* | 4 278 | 11 040 | 2 978 | 9 740 | 144 | 27.0 | 4.1 | |
| 75m AS1 75m AS2 | 4 231 | 11 011 | 2 939 | 9 719 | 9.3 | 30 | 11.0 | 7.5 |
| 75m AU1 75m AU2 | 1 275 | 4 193 | 975 | 3 893 | 5.8 | 6 | 2.7 | 0 |
| 75m BU1 75m BU2 | 1 086 | 6 794 | 747 | 6 455 | 4.5 | 2 | 0.8 | 0 |
| 75m AS1 75m BS1 | 14 018 | 14 860 | 10 434 | 11 276 | 14.2 | 75 | 15.0 | 8.5 |
| 75m AS 75m BS | 13 950 | 14 790 | 10 384 | 11 224 | 14.2 | 7 | 7.2 | 5.8 |
| 75m AU* 75m BU* | 5 213 | 7 586 | 3 955 | 6 328 | 10.9 | 14 | 3.0 | 3.5 |
| 75m AU 75m BU | 5 168 | 7 541 | 3 918 | 6 291 | 10.9 | 0 | 0 | 0 |
| 75m U 75m S | 11 459 | 25 174 | 7 166 | 20 881 | 13.3 | 3 | 1.5 | 1.9 |
| 25m all 75m all | 48 090 | 32 340 | 36 341 | 20 591 | 17.1 | 306 | 18.0 | 22.0 |

¹ as listed in Table 2, where * represents dataset comparisons without removal of replicate sequences

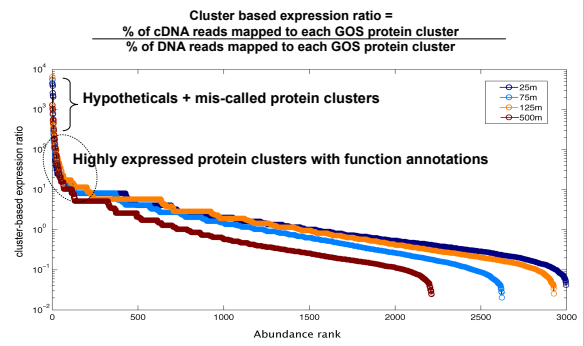
² total number of unique reference genes identified via BLASTX of non-rRNA reads against the NCBI

³ non-redundant (nr) database (e-value $\leq 1 \times 10^{-5}$)

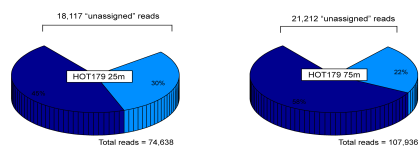
⁴ reference genes present in only one dataset

The ISME Journal (2010), 1-12

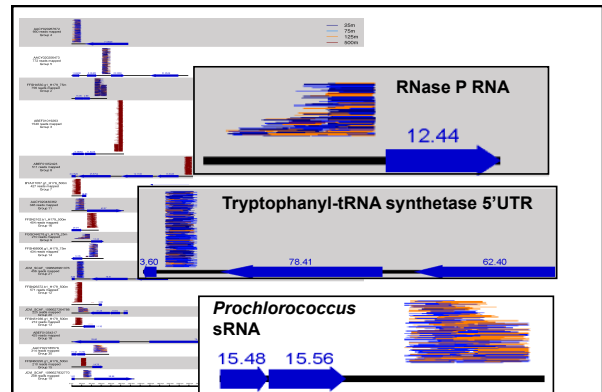
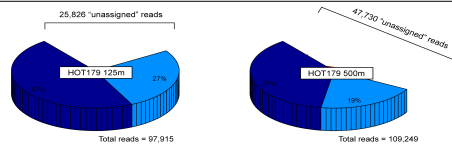
Protein family (GOS cluster) expression profile



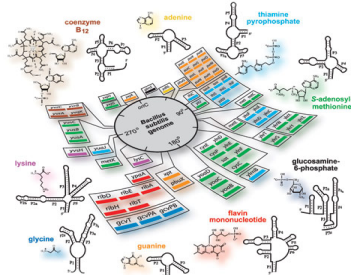
UNIDENTIFIED RNAs IN METATRANSCRIPTOMIC DATASETS



What do the mystery RNAs represent ???



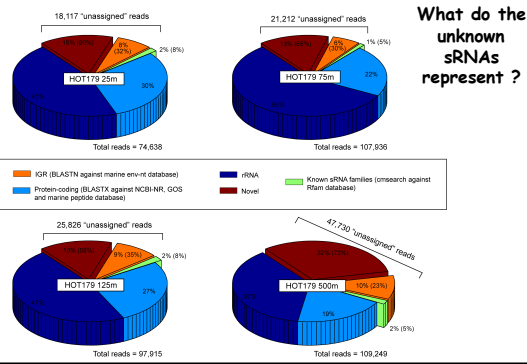
Bacterial Riboswitches



Known sRNA families in HOT community transcript data sets

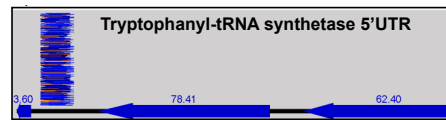
| Rfam ID | Function | Total count | 25m | 75m | 125m | 500m |
|-----------------------|---|-------------|-----------|-----------|-----------|-----------|
| RF00162.SAM | Riboswitch; methionine/cysteine biosynthesis | 410 | 010 | 010 | 010 | 410 |
| RF00029.Intron_gpII | Self-splicing ribozyme | 210 | 010 | 010 | 010 | 210 |
| RF00016.SNORD14 | Cleavage of eukaryotic precursor rRNA | 210 | 010 | 110 | 110 | 010 |
| RF00169.SRP_bact | Translation and targeting of proteins to cell membranes | 474 30 | 101 9 | 94 11 | 148 4 | 131 6 |
| RF00044.U2 | Pre-mRNA splicing in eukaryotes | 1411 | 111 | 510 | 110 | 710 |
| RF00010.RNaseP_bact_a | Generation of mature tRNA | 933 63 | 238 27 | 267 21 | 194 12 | 134 3 |
| RF00023.tmrRNA | Rescue of stalled ribosomes; cell cycle regulation | 1961 200 | 242 38 | 413 49 | 539 50 | 766 63 |
| RF00013.6S | Gene regulation during stationary phase | 71118 | 1217 | 23110 | 3311 | 3114 |
| RF00054.Glycine | Riboswitch; glycine metabolism | 29117 | 1713 | 613 | 517 | 114 |
| RF00005.sRNA | Protein synthesis | 1036 874 | 175 214 | 138 259 | 460 205 | 232 196 |
| RF00059.TPP | Riboswitch; gene regulation | 7115 | 112 | 014 | 410 | 219 |
| RF00174.Cobalamin | Riboswitch; gene regulation | 216 | 112 | 111 | 011 | 012 |
| RF00017.SRP_euk_arch | Translation and targeting of proteins to cell membranes | 4143 | 1119 | 115 | 113 | 1116 |
| RF00519.suH | Putative sRNA with unknown function | 012 | 010 | 010 | 010 | 012 |
| RF00066.U7 | Pre-mRNA splicing in eukaryotes | 015 | 011 | 010 | 013 | 011 |
| RF00582.SCARN14 | small nuclear RNA in eukaryotes | 018 | 013 | 010 | 012 | 013 |
| RF00521.SAM_alpha | Riboswitch; methionine biosynthesis in | 0110 | 012 | 010 | 012 | 016 |

Inventory of oceanic cDNA sequences



Metatranscriptomics reveals unique microbial small RNAs in the ocean's water column

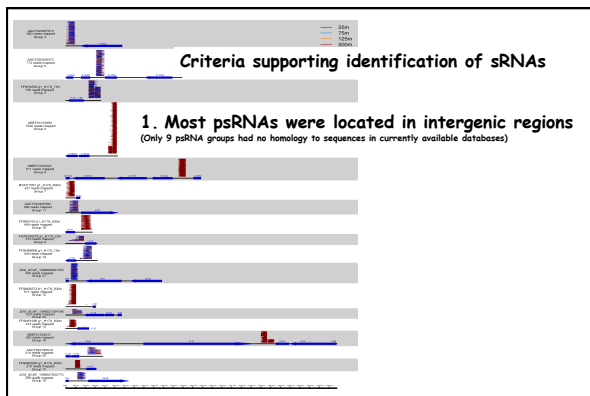
Yanmei Shi¹, Gene W. Tyson¹ & Edward F. DeLong^{1,2} | Vol 459 | 14 May 2009 | doi:10.1038/nature08055



Self clustering approach identified 66 novel sRNA groups: $\geq 85\%$ id, $\geq 90\%$ * length, ≥ 45 bp, > 100 sequences

Criteria supporting identification of sRNAs

1. Most psRNAs were located in intergenic regions (Only 9 psRNA groups had no homology to sequences in currently available databases)



Criteria supporting identification of sRNAs

2. Typical conserved 2^o structure for known sRNAs evident in psRNA clusters

