Computational analysis of high-throughput small RNA sequencing reveals microRNAs in a single-celled organism

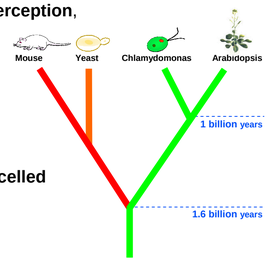
Frank Schwach[1,2], David J. Studholme[2], Attila Molnar[2], Eva Thuenemann[2], Michael Burrell[2], Alvaro Perez-Martinez[2], Vincent Moulton[1] and David C. Baulcombe[2]

[1]University of East Anglia and [2]The Sainsbury Laboratory/John Innes Centre, Norwich, UK.

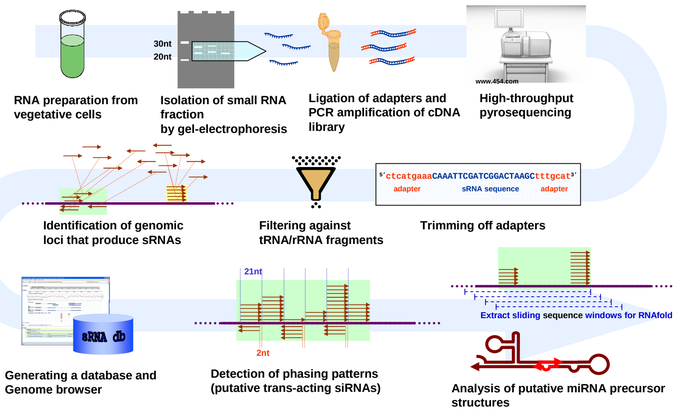
Micro (mi)RNAs are short (mostly 21-23nt) RNAs with the ability to regulate target genes post-transcriptionally. Many known miRNAs are involved in tissue development and maintenance and, until now, miRNAs appeared to be absent altogether from unicellular organisms. This has often led to the speculation that miRNAs have co-evolved with multicellularity in plants and animals. In contrast, we found that miRNA precursors are present in the single-celled green alga *Chlamydomonas reinhardtii* and that they give rise to mature miRNAs that regulate target genes by post-transcriptional RNA cleavage.
This shows that miRNA evolution began earlier than previously thought and was not a consequence of the onset of multicellularity.
Our database of *Chlamydomonas* small RNAs is publicly available at *www.cresirna.tsl.ac.uk*.

Chlamydomonas reinhardtii: a single-celled green alga

- **Model system** for studying biological processes including **photosynthesis**, **light perception**, **flagella** function and assembly, **circadian rhythms** etc.
- Well established protocols for cultivation and transformation
- Can grow in the dark and **shares** many **characteristics with animals**
- ~95% of **genome sequenced**
- **Transcriptional** and **post-transcriptional gene silencing** have been described in *Chlamydomonas*
- Our aim was to **characterise** the **endogenous small RNA population** of a **single-celled** organism and to find out whether **miRNA genes** are present

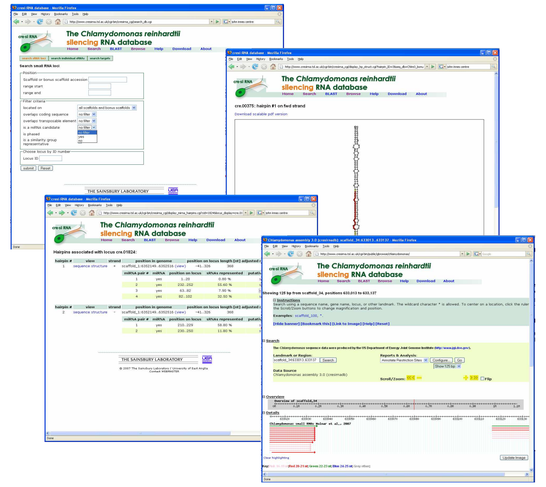


Preparation, sequencing and computational analysis of small RNAs



- **Genomic loci** were defined as local accumulations of small RNA matches to the genome with **at least 4 matches** and a **maximum "gap"** length of **300bp**.
- Analysis of **phasing patterns**: the percentage of small RNAs fitting into a perfect 21nt phasing pattern was assessed by **randomisation** experiments. **Four loci were identified** that could be similar to **trans-acting (ta)-siRNA** loci in higher plants.
- To find **potential miRNA-precursors**, a number of windows containing the locus sequence and varying lengths of flanking sequence were folded using **RNAfold[1]** and the significance assessed by the **randfold[2]** program.
- **Sequences** were **mapped** back onto the **secondary structures** and assembled into **overlap groups**. These had to pass **filters** including minimum percentage of **base-pairing**, maximum length of **unpaired regions**, **internal hairpins** and maximum **length** of overlap group.
- A miRNA candidate locus had to have at least **80% of small RNAs pass** these **filters**

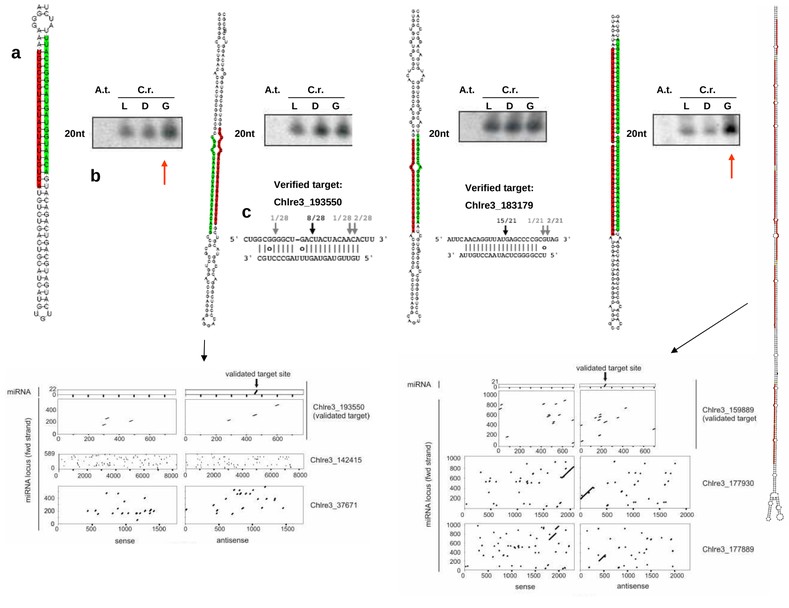
cre-siRNAdb: a database for *Chlamydomonas* small RNAs



- The database contains **46,672 small RNA** sequence reads
- **Searches** can be performed on **individual small RNAs**, small RNA **loci** and **targets**
- Lists of small RNA loci are linked to a **genome browser** showing the small RNA matches to the *Chlamydomonas* draft genome (version 3), obtained from the Joint Genome Institute (US Department of Energy)
- Detailed information about **miRNA candidate loci** is available, including **sequences** of predicted miRNAs/precursors and the precursor **secondary structures.**
- **BLAST** searches can be performed against the small RNA sequences or genomic small RNA loci
- Available online at **www.cresirna.tsl.ac.uk**

Many miRNA precursors in *Chlamydomonas* resemble proposed early stages of miRNA evolution in higher plants

- We found **68** candidate **miRNA precursors**
- Four miRNA candidates out of 8 tested by northern blot showed **differential expression** in gametes and vegetative cells
- **Potential targets** for four out of 18 tested candidate miRNAs could be **verified by 5' RACE** analysis, showing **cleavage** in the **centre** of the predicted miRNA **target site**. This is similar to what is found in most miRNA targets of **higher plants**
- Two verified targets are associated with cell motility, one is a sugar epimerase
- **No conserved homologues** of the *Chlamydomonas* miRNAs were found in other plant or animal species
- **miRNA precursors** in plants are thought to **originate** from multiple **gene duplication** events, leaving (partial) copies of a gene in an **inverted repeat configuration[3]**. Young miRNA precursors form **long near-perfect hairpins** with the potential to produce **multiple mature miRNAs** that target the gene of origin. At later stages, the hairpin accumulates mismatches and produces only a single pair of miRNA/miRNA*, not necessarily targeting the original gene of origin.
- **47** candidate precursors had **long** (>150nt) **near-perfect hairpins**, often producing **multiple mature miRNA/miRNA* pairs** (some "in phase")
- We searched for **extended regions of similarity** between **miRNA loci** and **ESTs** to find **potential genes of origin** for long hairpins.



a) Predicted hairpin structures

miRNA candidates indicated in red, miRNA*s in green (based on abundance)

b) Northern blots for miRNA expression profiles:

A.t: Arabidopsis thaliana
C.r.: Chlamydomonas reinhardtii
L: grown in light
D: grown in dark
G: gametes
♦: miRNA upregulated in gametes
20nt size marker indicated

c) Race target cleavage analysis: Upper sequence: predicted target Lower sequence: miRNA

Arrows indicate number of times a cleavage site was found in a given number of 5' RACE experiments

d) Sequence similarity between two miRNA loci and their two highest-scoring FASTA matches among *Chlamydomonas* ESTs as well as the validated targets. Sequence similarity between the targets and one of the mature miRNAs each is also shown. The plots were generated using DOTMATCHER[4].

Diagonal lines indicate stretches of sequence similarity above a threshold score.

Long diagonal lines represent extended regions of similarity between the miRNA locus and an EST. The short-hairpin locus (left) has no such extended similarity even to its best overall matches in the EST collection. This could indicate that this locus is at a later stage of miRNA evolution than the long-hairpin locus on the right, which still shows extended regions of similarity to two potential genes of origin but not the verified target of one of its mature miRNA sequences.

This work was supported by BBSRC grant BB/E004091/1 and the Gatsby Charitable Foundation

References

This work is published in: Molnar A, Schwach F, Studholme DJ, Thuenemann EC, Baulcombe DC (2007), Nature 447:1126-9

[1] Hofacker IL,Fontana W, Stadler PF, Bonhoeffer S, Tacker M, Schuster P (1994) Monatshefte f. Chemie 125: 167-188
[2] Bonnet E, Wuyts J, Rouze P, Van de Peer Y (2004) Bioinformatics 20:2911-7
[3] Chapman EJ, Carrington JC (2007) Nat Rev Genet 8:884-96
[4] Rice P, Longden I, Bleasby A (2000) Trends in Genetics 16:276--277