Identification of AGO3-Associated miRNAs and Computational Prediction of Their Targets in the Green Alga *Chlamydomonas reinhardtii*

Adam Voshall,* Eun-Jeong Kim,* Xinrong Ma,* Etsuko N. Moriyama,*[†] and Heriberto Cerutti*^{†,1}

*School of Biological Sciences and [†]Center for Plant Science Innovation, University of Nebraska, Lincoln, Nebraska 68588

ABSTRACT The unicellular green alga *Chlamydomonas reinhardtii* harbors many types of small RNAs (sRNAs) but little is known about their role(s) in the regulation of endogenous genes and cellular processes. To define functional microRNAs (miRNAs) in *Chlamydomonas*, we characterized sRNAs associated with an argonaute protein, AGO3, by affinity purification and deep sequencing. Using a stringent set of criteria for canonical miRNA annotation, we identified 39 precursor miRNAs, which produce 45 unique, AGO3-associated miRNA sequences including 13 previously reported miRNAs and 32 novel ones. Potential miRNA targets were identified based on the complementarity of miRNAs with candidate binding sites on transcripts and classified, depending on the extent of complementarity, as being likely to be regulated through cleavage or translational repression. The search for cleavage targets identified 74 transcripts. However, only 6 of them showed an increase in messenger RNA (mRNA) levels in a mutant strain almost devoid of sRNAs. The search for translational repression targets, which used complementarity criteria more stringent than those empirically required for a reduction in target protein levels, identified 488 transcripts. However, unlike observations in metazoans, most predicted translation repression targets did not show appreciable changes in transcript abundance in the absence of sRNAs. Additionally, of three candidate targets examined at the protein level, only one showed a moderate variation in polypeptide amount in the mutant strain. Our results emphasize the difficulty in identifying genuine miRNA targets in *Chlamydomonas* and suggest that miRNAs, under standard laboratory conditions, might have mainly a modulatory role in endogenous gene regulation in this alga.

KEYWORDS miRNA; argonaute; miRNA target; transcript cleavage; translation repression

RNA interference (RNAi) provides eukaryotic cells with a precise means of gene expression regulation through a variety of mechanisms (Kim 2005; Bartel 2009; Voinnet 2009). MicroRNAs, key components of the RNAi machinery, consist of endogenously encoded 20- to 24-nucleotide (nt) RNA molecules (Kim 2005). In plants and some green algae, these molecules are predominantly derived from RNA hairpins up to ~300 bp in length that are processed by RNase III enzymes, the Dicer-like (DCL) family of proteins (Casas-Mollano *et al.* 2008; Voinnet 2009; Tarver *et al.* 2012). In *Arabidopsis thaliana*, most microRNAs (miRNAs) are generated by DCL1 in two processing steps, from the initial transcript [the primary

miRNA (pri-miRNA)] to a smaller hairpin miRNA precursor (pre-miRNA) and then to a short duplex consisting of the miRNA and its complementary sequence (the miRNA* or passenger strand) (Voinnet 2009; Zhu *et al.* 2013). The short RNA duplex is loaded into the RNA-induced silencing complex (RISC), which contains an argonaute (AGO) protein as a central component and becomes active upon removal of the miRNA* strand. Active RISC can repress expression of target messenger RNAs (mRNAs) by preventing translation and/or by triggering transcript degradation (Bartel 2009; Voinnet 2009).

The extent of complementarity between miRNA and target mRNA has been proposed to play a major role in determining whether silencing occurs through transcript cleavage or translational repression/mRNA destabilization (Bartel 2009). The first nucleotide at the 5' end of a miRNA is secured within the middle (Mid) domain of argonaute, making this nucleotide inaccessible for binding with target transcripts, although it is important for determining miRNA loading into distinct AGOs (Mi *et al.* 2008; Wang *et al.* 2008a; Frank *et al.* 2010, 2012; Zha *et al.* 2012; Endo *et al.* 2013). Nucleotides 2–8 from the

Copyright © 2015 by the Genetics Society of America
doi: 10.1534/genetics.115.174797

Manuscript received January 21, 2015; accepted for publication March 11, 2015;
published Early Online March 13, 2015.

Supporting information is available online at <http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.115.174797/-/DC1>.

¹Corresponding author: School of Biological Sciences and Center for Plant Science Innovation, University of Nebraska, E211 Beadle Center, Lincoln, NE 68588-0666.
E-mail: hcerutti1@unl.edu

5' end of a miRNA (the seed region, see [Supporting Information, Figure S1](#)) are often almost entirely complementary to a target mRNA, while the remaining nucleotides may contribute variably to the pairing interaction (Wang *et al.* 2008b; Bartel 2009; Sheng *et al.* 2014). In land plants, the complementarity of miRNAs and targets is commonly nearly perfect, resulting in cleavage of the target sequence between nucleotides 10 and 11 of the miRNA (Rogers and Chen 2013; Shen *et al.* 2013; Zhou and Luo 2013; Liu *et al.* 2014). In animals, by contrast, miRNAs frequently display extensive mismatching with their targets and regulate target expression through translational repression and/or transcript destabilization independent of AGO-mediated cleavage (Carthew and Sontheimer 2009; Djuranovic *et al.* 2012; Fujiwara and Yada 2013; Zheng *et al.* 2013). There is a growing body of evidence indicating that plants can perform this type of regulation as well; however, it still appears to be much less common than regulation through direct cleavage (Lanet *et al.* 2009; Beauclair *et al.* 2010; Yang *et al.* 2012; Axtell 2013; Li *et al.* 2013).

In the unicellular alga *Chlamydomonas reinhardtii*, miRNAs have been described relatively recently, based on deep sequencing of total cellular small RNA (sRNA) populations (Molnar *et al.* 2007; Zhao *et al.* 2007; Ibrahim *et al.* 2010; Shu and Hu 2012; Lv *et al.* 2013). However, the validity of many predicted *Chlamydomonas* miRNAs has been questioned of late because of their large hairpin precursor structures and the apparent imprecise processing of the sRNAs (Nozawa *et al.* 2012; Tarver *et al.* 2012). *C. reinhardtii* possesses core miRNA/sRNA processing and effector machinery similar to that in higher plants (Casas-Mollano *et al.* 2008). This machinery includes three DCL proteins along with three AGO proteins. DCL1 and AGO1 appear to be primarily involved in the silencing of transposable elements whereas AGO3 seems to be most extensively involved in miRNA functions (Casas-Mollano *et al.* 2008). MicroRNA quality control is at least partially directed by the MUT68 protein, which functions by placing untemplated nucleotides (generally uridyls) at the 3' end of the miRNA molecules, flagging them for degradation (Ibrahim *et al.* 2010).

Little is known about miRNA/target interactions in green algae. Early studies suggested that miRNAs in *C. reinhardtii* might function primarily by triggering transcript cleavage, in a manner similar to those in land plants (Molnar *et al.* 2007; Zhao *et al.* 2007). However, very few of the 86 mature *Chlamydomonas* sRNAs deposited in miRBase (Kozomara and Griffiths-Jones 2014) have identifiable targets with near perfect complementarity. This lack of easily recognizable targets leaves open the possibility that the miRNAs may instead largely operate by a mechanism(s) more akin to that in animals. This hypothesis is strengthened by recent evidence, using artificial constructs, demonstrating that miRNA/sRNA-mediated translation repression can occur in *C. reinhardtii* (Ma *et al.* 2013; Yamasaki *et al.* 2013). Thus, it remains unclear whether the majority of the described *Chlamydomonas* sRNAs are truly miRNAs, what their potential targets are, and by which mechanism(s) they may regulate endogenous gene expression.

In animals, the non-RISC-associated passenger strand is very labile and it has been suggested that the high stability of miRNAs reflects their RISC association, implying that sequencing of total cellular sRNA populations accurately reflects the RISC-associated sRNA populations (Winter and Diederichs 2011). This assumption has been recently challenged by studies indicating that only a fraction of mature miRNAs are in fact AGO associated (Janas *et al.* 2012; Schug *et al.* 2013; Stalder *et al.* 2013; Flores *et al.* 2014). It remains uncertain whether this also applies to land plants and green algae. Nonetheless, to increase the likelihood of defining a set of functional, RISC-bound miRNAs in *Chlamydomonas*, we identified AGO3-associated sRNAs by deep sequencing. From these sequences, candidate miRNAs, meeting canonical criteria, were selected and used to predict interacting targets computationally. We then examined the mechanism(s) of regulation for a few putative endogenous targets. Our results suggest the existence of a limited set of genuine AGO3-associated miRNAs in *C. reinhardtii*, some of which may control target transcripts by cleavage or translation repression. However, under standard laboratory conditions, miRNAs/sRNAs do not appear to play a major role in the regulation of essential cellular processes in this green alga.

Materials and Methods

Transgenic strains, mutants, and culture conditions

Chlamydomonas cells were routinely grown in tris-acetate-phosphate (TAP) medium (Van Dijk *et al.* 2005; Ma *et al.* 2013). Strain Maa7-IR44s, containing an inverted repeat transgene targeting the 3'-UTR of the *MAA7* gene (encoding Tryptophan Synthase β -subunit), has been previously described (Ma *et al.* 2013). Strain Mut-20, deleted for *Tudor Staphylococcal Nuclease 1 (TSN1)* and almost devoid of miRNAs/sRNAs, was obtained in an insertional mutagenesis screen designed to isolate mutants defective in RNAi-mediated translation repression (Ibrahim *et al.* 2010; Ma *et al.* 2013). For the isolation of AGO3-associated sRNAs, we fused the FLAG tag (Einhauser and Jungbauer 2001) to the N-terminal end of the *Chlamydomonas* AGO3 (g16859) coding sequence. This construct was placed under the control of *PsaD* regulatory sequences and transformed into Maa7-IR44s. A similar construct was made fusing the FLAG tag to the coding sequence of the *Ble* gene [confering bleomycin resistance (Stevens *et al.* 1996)] and transformed into Maa7-IR44s. *PsaD* regulatory sequences were used because they have been reported to allow reliable expression, with reduced incidence of gene silencing, of complementary DNA (cDNA) transgenes in *Chlamydomonas* (Fischer and Rochaix 2001).

Isolation of AGO3-associated sRNAs, library preparation, and sequencing

FLAG-tagged AGO3 was affinity purified from a cell lysate as previously described for a TAP-tagged protein (Van Dijk *et al.* 2005). RNAs associated with AGO3 were purified with TRI reagent (Molecular Research Center) and contaminant DNA

was removed by DNase I treatment (Ambion) (Ibrahim *et al.* 2010; Ma *et al.* 2013). Construction of cDNA libraries and Illumina sequencing were then carried out as previously reported (Ibrahim *et al.* 2010). All Illumina reads from sRNA sequencing were deposited into the NCBI Sequence Read Archive (SRA) under accession no. SRR1747077.

sRNA mapping and profiling

Sequenced reads were first mapped to version 9.1 of the Phytozome *C. reinhardtii* genome (Merchant *et al.* 2007), by using version 3.02 of Novoalign (www.novocraft.com) with the miRNA flag and with a score threshold of 15. Mapped reads were filtered to remove those showing alignments with gaps or mismatches as well as those that mapped to more than five locations in the genome. Both the total population of sequenced reads and the population of reads mapped to the genome were profiled based on length. Genome-mapped reads were also classified as matching the chloroplast or the mitochondrial genomes, ribosomal RNAs (rRNAs), small nuclear RNAs (snRNAs), small nucleolar RNAs (snoRNAs), transfer RNAs (tRNAs), other noncoding RNAs (ncRNAs), or transposable elements. All sequences except for transposable elements were taken from GenBank (accession nos. BK000554.2 and NC_001638.1 for the chloroplast and mitochondrial genomes, respectively). Transposon sequences were taken from Repbase (Kapitonov and Jurka 2008). The sequence logo to show nucleotide composition bias by position was made using version 3.3 of Weblogo, analyzing all unique reads 21 nt in length (Crooks *et al.* 2004).

Genomic clustering of sRNAs and miRNA identification

After removing those reads that mapped to known ncRNA categories, transposons, the chloroplast, or the mitochondrial genomes, remaining sRNA reads were processed following the steps illustrated in Figure 1A. The reads were first clustered by genomic location such that there was no more than 200 nt between adjacent reads, regardless of strand, in a given cluster. The reads on both strands in the same genomic location were placed in the same cluster. These clusters were analyzed further to differentiate pre-miRNA structures from other sRNA structures. The genomic sequence for each strand of the cluster was folded using version 2.1.5 of RNAfold from the Vienna RNA package (Lorenz *et al.* 2011). Clusters with a genomic sequence <50 nt or >1000 nt in length were excluded from folding, as they were deemed unlikely to code for canonical miRNAs (see Tarver *et al.* 2012). Clusters containing sequence gaps (*i.e.*, unsequenced genomic regions) were also excluded since the secondary structure folding of these regions cannot be unambiguously predicted. The obtained secondary structures were then parsed to determine whether they folded into a hairpin. However, clusters whose highest expressed reads had more than four mismatches with the complementary arm of the hairpin were removed from further analyses, following the criteria set up by Tarver *et al.* (2012) for the annotation of plant miRNAs. Clusters remaining after this filtering were manually curated based on the accuracy of processing of

the 5' end of the predominant reads and the frequency of predominant reads (Tarver *et al.* 2012; Coruh *et al.* 2014). The putative miRNA* sequence for the highest expressed read(s) of each cluster was determined using the hairpin fold and the genomic sequence, with 1- or 2-nt offset to allow for the 3' overhang in the miRNA/miRNA* duplex.

MicroRNA target prediction

Figure 1B illustrates the process used for the prediction of miRNA targets. Potential miRNA binding sites were determined by searching version 9.1 of the Phytozome *C. reinhardtii* transcriptome, using version 2.1 of RNAhybrid (Kruger and Rehmsmeier 2006). This algorithm returns the most energetically favorable pairing for each miRNA/target pair. The obtained hybrid pairs were filtered based on the number and location of gaps, wobbles, and mismatches to limit the predictions to those that could regulate targets in a foreseeable manner. For cleavage targets, this search required perfect matching for nucleotides 2–8 (the seed region, see Figure S1) and nucleotides 9–12 (the catalytic or center region, see Figure S1) and no more than three GU wobbles and three mismatches or a gap of >1 nt in the remaining sequence. For translational repression targets, the constraints for the catalytic region were relaxed to allow up to three mismatches or wobbles, with the constraints for the seed region and the 3'-end pairing remaining the same. Additionally, translational repression targets needed at least one mismatch or wobble in the catalytic region to keep the two sets of predicted targets nonoverlapping. MicroRNA-encoding transcripts, which may also be targeted for degradation by their encoded miRNAs, were removed from the analysis.

Putative functions of predicted targets were evaluated by using the annotations of *Chlamydomonas* genes (if available) as well as conserved protein domains (Finn *et al.* 2014). Functional annotations were taken from the Algal Functional Annotation site (Lopez *et al.* 2011) and are mostly based on those on Phytozome (Goodstein *et al.* 2012). For each protein, even if a domain is repeated, each domain is reported only once.

RNA analyses and quantitative PCR assays

Total cell RNA was purified with TRI reagent (Molecular Research Center), following the manufacturer's instructions. For sRNA analyses, total RNA samples were resolved in 15% polyacrylamide/7-M urea gels and electroblotted to Hybond-XL membranes (GE Healthcare). Blots were hybridized with ³²P-labeled DNA probes, using the High Efficiency Hybridization System as previously described (Ibrahim *et al.* 2010). Specific miRNAs were detected by hybridization with DNA oligonucleotides labeled at their 5' termini with [γ -³²P]ATP and T4 Polynucleotide Kinase (New England Biolabs, Beverly, MA) (Ibrahim *et al.* 2010; Ma *et al.* 2013). For examination of sRNAs by SYBR-Gold (Life Technologies) staining, TRI reagent isolated total RNA was fractionated through Microcon YM-100 centrifugal devices (Millipore, Bedford, MA) to remove high molecular weight transcripts. sRNAs were concentrated from the filtrate by ethanol precipitation and resolved in denaturing polyacrylamide gels. Putative cleavage sites of the *Cre17*.

Figure 1 (A and B) Criteria and flowcharts for predicting canonical miRNAs (A) and their targets (B).

g697550 transcript were examined using a 5'-RACE approach with the GeneRacer kit (Life Technologies), as previously described (Molnar *et al.* 2007; Yamasaki *et al.* 2013). First and nested PCR amplifications were performed using the following primers: *g697550*(Race)-R1 (5'-CCTTGCACTTGAGGCACTGCA CAA-3') and *g697550*(Race)-R2 (5'-CGTGTGGGGGCGGGAT GAT-3'). For quantitative PCR analyses, DNase I-treated RNA samples were used as template for first-strand cDNA synthesis, using an oligo(dT)₁₈ primer and SuperScript III reverse transcriptase (Life Technologies). Primer pairs for the quantitative PCR amplifications were as follows: for *Cre17.g697550*, *g697550*-F (5'-GAGAGGATCGCGGACAACC-3') and *g697550*-R (5'-AGGACCGGTAGATGCTCTTGG-3'); and for *Cre16.g683650*, *g683650*-F (5'-CAGTTTGAGCCCGACCTACG-3') and *g683650*-R (5'-CCACGCCGCACTCCAGC-3'). DNA fragments were

amplified and quantified with the RT² SYBR Green/Fluorescein qPCR mastermix (QIAGEN, Valencia, CA), using the iCycler Real Time PCR Detection System (Bio-Rad, Hercules, CA).

Differential gene expression analysis

RNA samples from Maa7-IR44s and Mut-20, grown photoheterotrophically in TAP medium, were purified with TRI reagent and treated with DNase I. Standard RNA-Seq of the samples was then performed using an Illumina Genome Analyzer II at the University of Nebraska (Lincoln) Genomics facility (Brueggeman *et al.* 2012). All reads obtained by RNA-Seq were deposited into the NCBI SRA under accession nos. SRR1747017 and SRR17470742. Sequencing reads from both strains were mapped to the *C. reinhardtii* transcript models in version 9.1 of Phytozome (Merchant *et al.* 2007). Reads were

first trimmed to 36 nt in length, to remove low-quality 3'-end sequences before mapping. The alignment was performed using Bowtie2 (Langmead and Salzberg 2012) with a seed length of 25 nt, allowing two mismatches, and a maximum hit of 1. Only reads that uniquely matched with a single transcript were counted. Raw gene counts were determined by counting the number of reads aligned to each transcript. Transcript abundance was examined as reads per kilobase per million (RPKM), which normalizes the read counts based on both transcript length and total number of reads, using the formula

$$\text{RPKM} = \frac{10^9 C}{NL},$$

where C is the number of reads mapped to the transcript, N is the total number of mapped reads in the library, and L is the length of the transcript in nucleotides (Mortazavi *et al.* 2008). To determine changes in gene expression, transcript abundance for each gene was compared pairwise between Mut-20 and its parental strain Maa7-IR44s. Differences in gene expression were examined as $\log_2$ (fold change), where fold change refers to the ratio of RPKM values between the Mut-20 and the Maa7-IR44s (control) strains. Statistical analysis to determine the significance of gene expression changes was performed using version 1.16 of DESeq (Anders and Huber 2010).

Immunoblot analysis

The *Chlamydomonas* Cre16.g683650 protein was immunodetected, following a standard procedure (Ma *et al.* 2013), by overnight incubation at 4° with a 1:10,000 dilution of a rabbit antibody raised against a C-terminal peptide (GIKPSAHKRGGVRM) conjugated to KLH (GenScript). A modification-insensitive polyclonal antibody (Abcam ab1791) was used to detect histone H3.

Results

Identification of AGO3-associated sRNAs

A FLAG-tagged AGO3, under the control of the *PsaD* promoter, was introduced into the Maa7-IR44s strain (Ma *et al.* 2013) and its functional competence evaluated by examining its subcellular localization (see Figure S2) and its association with the translationally repressed *MAA7* transcript (see Figure S3). sRNAs associated with AGO3 were then isolated by co-immunoprecipitation with the FLAG-tagged protein. Deep sequencing of the corresponding sRNA library generated 20,589,616 reads (675,917 unique), of which 12,463,249 (117,241 unique) mapped perfectly to the nuclear genome. Unique reads refer to nonredundant sRNA sequences. Both the full library and the genome-matching reads revealed that 93% of the AGO3-associated sRNAs are between 20 and 22 nt in length (Figure 2A). Additionally, 98% of the sequenced sRNAs have a U as the 5' nucleotide (Figure 2E), suggesting a strong preference by AGO3 for both length and starting nucleotide. Only 3% of all mapped reads matched known non-coding RNAs or transposable elements (Figure 2C).

MicroRNA prediction

MicroRNAs were predicted based on the criteria outlined by Tarver *et al.* (2012) (Figure 1A). To differentiate miRNAs from other sRNAs, all mapped reads were clustered by genomic location such that within each cluster adjacent reads were no more than 200 nt apart, regardless of strand. Genomic sequences for each strand of each cluster were folded using RNAfold to determine the secondary structure of putative pre-miRNAs (Lorenz *et al.* 2011). To be classified as a miRNA precursor, a cluster was required to fold into a hairpin and have no more than two predominant 5'-processing sites. In addition, the main reads (*i.e.*, the greatest abundance reads in each cluster) were required to have no more than four mismatches in the complementary arm of the hairpin (Tarver *et al.* 2012). Using these criteria, we identified 39 suitable pre-miRNAs, which produce 45 unique miRNA sequences (Table 1). Of these 45 miRNAs, 29 had their respective miRNA* present in the library, while the remaining 16 did not (see Discussion). We found 13 previously reported *C. reinhardtii* miRNAs and identified 32 novel ones. Although three clusters matched previously described precursor miRNAs (pre-miR919, pre-miR1148, and pre-miR1164), the identified predominant reads differed from the annotated miRNA sequences (Table 1). Additionally, cluster 17,755 corresponded to the antisense strand of pre-miR1154 (Table 1). The 45 miRNAs accounted for 3,467,899 reads (nearly 28%) of the 12,463,249 genome-mapped reads (Figure 2C and Table 1) and correspond to some of the most abundant sequences in the library, with a high degree of redundancy. As an example of the obtained results, the putative pre-miRNA folding and matching reads for cluster 19,166 are shown in Figure 3A. The folding data and matching reads for all the miRNA-candidate clusters can be found in Figure S4.

The expression of a subset of miRNAs was verified by Northern blot analyses (Figure 3B). Intriguingly, a very abundant sequence in the sRNA library (based on read frequency) did not have an identifiable precursor in the *Chlamydomonas* genome. Yet the corresponding sRNA (designated miR B) was easily detected in RNA blots (Figure 3B) and appears to be a functional miRNA-like molecule (see below). Since the current version of the *C. reinhardtii* genome contains sequence gaps, we hypothesize that the gene encoding miR B might be located in an unsequenced region of the genome. In support of this interpretation, for another sRNA (designated miR C), which also is detectable in Northern blots (Figure 3B) and appears to be functional (see below), the putative pre-miRNA includes a sequence gap that precludes its unambiguous folding into the expected stem-loop structure (see Figure S5).

Many previously identified *Chlamydomonas* sRNAs reported as miRNAs (Molnar *et al.* 2007; Zhao *et al.* 2007; Shu and Hu 2012; Kozomara and Griffiths-Jones 2014), in addition to those listed in Table 1, were also found associated with AGO3 (Table S4). However, these sequences did not meet the criteria to be classified as miRNAs (Tarver *et al.* 2012) since

Figure 2 Classification of AGO3-associated *Chlamydomonas* small RNAs. (A) Length distribution of AGO3-associated sRNAs, displaying the frequencies in counts per million reads (CPM) for all reads in the library (blue) and for the reads that map to the genome (red). (B) Length distribution of AGO3-associated unique (nonredundant) sRNA sequences, displaying frequencies in CPM for all unique sequences in the library (blue) and for the unique sequences that map to the genome (red). (C) Abundance (as percentage of all genome-mapped reads) of AGO3-associated sRNAs matching different categories of sequences. The black circle indicates sRNAs matching to the chloroplast genome and all other categories are shown clockwise in the same order as in the key. The remaining sRNA category includes non-miRNA sequences processed from predicted miRNA precursors, those derived from imprecise processing of long hairpins, and other small RNAs with undefined biogenesis. (D) Number of unique AGO3-associated sRNAs (nonredundant sequences) matching different categories of sequences. (E) Relative nucleotide composition bias by position for all unique sRNAs 21 nt in length. The graph was made with Weblogo (Crooks *et al.* 2004).

some appear to derive from long RNA hairpins that produce multiple sRNAs with imprecise 5' ends (typical of small interfering RNAs) and others seem to be encoded next to sequence gaps preventing the unequivocal prediction of secondary structures (Figure 2, C and D, remaining sRNA category). This annotation problem is also becoming apparent in other eukaryotes, as the depth of sequenced sRNAs increases and miRNA prediction criteria are improved (Tarver *et al.* 2012; Coruh *et al.* 2014; Kozomara and Griffiths-Jones 2014). Additionally, a few potentially genuine miRNAs (e.g., miR1151b) were present at very low levels in our library (Table S4) and may not associate meaningfully with AGO3. These sequences

could correspond to real miRNAs bound to other *Chlamydomonas* argonautes.

MicroRNA target prediction

Target prediction for the identified 45 miRNA sequences relied on complementarity between the miRNAs and putative binding sites on transcripts (Figure 1B). The binding sites were uncovered using RNAhybrid to search against the *Chlamydomonas* transcriptome in Phytozome version 9.1 (Kruger and Rehmsmeier 2006; Goodstein *et al.* 2012). Potential targets were divided into two categories, depending on the extent of complementarity to a miRNA, those likely to be regulated

Table 1 AGO3-associated candidate miRNAs in *Chlamydomonas reinhardtii*

Cluster ^a	Alias ^b	miRNA	O ^c	Chr ^d	Start ^e	Location ^f	Gene ^f	PFam A ^f	Count ^g
1,130	cr.02052	TTCGAAAAGTTGCACAGCGCC	+	1	5,106,348	Intron	Cre01.g035500	PI3K_C2, PI3Ka, PI3_P14_kinase	631
2,675a	miR1174.1, miR9897-5p	TACCGGGCGTGGGAGGGCAGG	—	10	3,399,959	3'-UTR	Cre10.g444300		896,822
2,675b	miR1174.2, miR9897-3p	TTACGGCTCTCTTATCGGC	+	10	3,399,983	Intergenic			647,884
3,318		TCAATTGCTCTTCTTGATG	—	10	3,399,872	3'-UTR	Cre10.g444300		3,165
3,704*	pre-miR1148 ^h	TGCACGTTGGTAAACATGCC	—	11	6,282,547	3'-UTR	Cre10.g465000	Pkinase	20,465
3,708	miR1147.1	TAGTGCCAGCGCTCTCGGC	—	11	1,442,258	3'-UTR	g11660		3,004
3,772*		TGAAAAATCGAGGCCGCTGGC	+	11	1,448,250	Intergenic			6,521
					1,713,373	5'-UTR	g11713		
					1,713,298	Intergenic			
3,911		TGGATGTCGAAACCCGCTACA	—	11	2,529,518	3'-UTR	Cre11.g476000		1,595
3,965*		TGGAGACTGTACGAGAGGCCA	—	11	2,856,089	3'-UTR	Cre11.g477700		1,237
5,094		TCAATAACGGACCTGGAGGAC	+	12	3,183,096	Intergenic			1,181
					3,183,213	3'-UTR	g12598		
5,858	miR1157-3p, cr.02116	TTCAGGTAGCGGGACCAAGTG	—	12	6,402,228	Intron	g13222	zf-CW, PWMWP, Chromo, SNF2_N, Helicase_C	130,801
6,589*		TCTGACCGTTCGGTCTGCCG	+	12	9,542,589	Intron	g13796		4,758
7,085a		TGACTCTCACTCTACTCGGC	+	13	2,001,158	Intergenic			37,202
					2,001,092	3'-UTR	g14171		
7,085b		TTCCACTGACTGCTTGCTGA	—	13	2,001,069	3'-UTR	g14171		12,380
8,319*	cr.01824	TCCACGTTCTCCCGCGCAGG	+	14	1,950,136	Intergenic			6,989
8,563a		TGCGGGGCCCGAGGTAGTGC	—	14	321,886	Intron	Cre14.g629200		704
8,563b	miR910	AGCAGCTCGGGCTCGACCGC	—	14	3,218,785	Intron	Cre14.g629200		625
9,303	miR1169-3p	TGGATGTTGCTGCTGGATAC	+	16	185,156	Intergenic			183,618
					185,089	3'-UTR	Cre16.g694950	Pkinase	
9,770a*		TTGATGCCGCTGGGAGCTGG	+	16	2,231,294	3'-UTR	Cre16.g658782		7,159
					2,231,401	Intergenic			
9,770b*		TGCGAGCTGCGGAATGGCGAC	—	16	2,231,221	3'-UTR	Cre16.g658782		1,753
9,770c*		TGCGTAGGACGTGGTAGGCGC	+	16	2,231,112	Intergenic			1,908
11,846	pre-miR1164 ⁱ	TGAGTGGTGGTCGACGCCAAT	+	17	3,137,574	3'-UTR	Cre17.g721553		3,013
12,364		TCGGAGAAGCGGGTAGCTGA	+	17	5,152,874	Intergenic			122,990
					5,152,806	5'-UTR	g17790		
12,551a		TAAACAGACAAGGCGACCGACA	—	17	6,144,115	3'-UTR	g17937		18,839
12,551b		TCGCCTTGCTGTTTATGTGG	+	17	6,144,121	Intergenic			9,949
					6,144,182	3'-UTR	g17937		
12,615*		TGCAAAATCAAGACCGGGGGA	+	17	6,674,056	Intergenic			1,243
					6,674,164	Intergenic			
12,905*	miR911*, cr.02798	TGGCCGTACTACTATTGTC	—	2	739,369	3'-UTR	Cre02.g078500		5,494
14,705*		TCGGAAAGGAATCCAACGGCC	+	2	9,101,230	Intron	g2826		14,785
					9,101,131	Intergenic			
14,711*		TGAACACGGCGAGCATCAGCGG	+	2	9,121,266	3'-UTR	g2830		1,231
14,712		TGCTTGCGCCCTCTAGCCGTC	+	2	9,129,575	3'-UTR	g2831		8,955
					9,129,507	Intergenic			
14,719*		TCGTGCTCGTCATCCCTCTG	+	2	9,173,370	3'-UTR	g2836		8,513
					9,173,305	Intergenic			
16,411*		TGCGGTGAATGTGAATGATGG	—	3	6,573,968	3'-UTR	Cre03.g195950		4,117
					6,573,895	Intergenic			

(continued)

Table 1, continued

Cluster ^a	Alias ^b	miRNA	O ^c	Chr ^d	Start ^e	Location ^f	Gene ^f	PFam A ^f	Count ^g
17,284*		TGACCACCCCTGCAGCTGACGC	—	4	1,804,340	Intron	g4741	Exostosin	2,415
17,620	miR1153-5p, cr.02052	TGGGCCATCGTATTACTATCAG	+	4	3,100,623	Intron	Cre04.g225700		377,890
17,755*	antisense-pre-miR1154/	TAACCTAGTCGTACAAAGGCG	+	4	3,694,795	Intergenic			1,588
18,100a	pre-miR919*	TCGGTCAGCATTTCTTTGG	+	5	935,314	Intergenic			91,516
18,100b		TACCCGGAGTGGACGCTTTC	+	5	935,397	Intergenic			15,493
19,166		TTGGCGGCGTTTGAAGATT	+	6	2,201,623	Intergenic			86,172
			—		2,201,668	Intron	g5972	WD40	
19,538	miR1162-3p, cr.02386	TGTTGTAGTAGTTTACGCCCTGC	+	6	3,067,434	Intron	Cre06.g274550	PKinase, PKinase_Tyr	307,549
20,399		TAGAGCTCGAAGAACTTGGGA	+	6	6,776,173	Intergenic			362,932
20,462	miR1152, cr.01969	TAAGAAGGTGCGCTGCTTGA	—	6	7,063,794	Intron	g6871	UBA_e1_thiolCys, UBACT, ThiF, UAE_Ubl	1,701
21,642		TCGACGCGGTGATGGGCCTGG	+	7	2,713,524	CDS	Cre07.g331114		2,702
22,587		TGGCTTTCGTCGGTCTAGG	+	8	121,861	Intergenic			26,418
			—		121,923	3'-UTR	g8351		
25,047	miR912	TGGATTGATCCAGCCAGGCG	—	9	6,365,908	Intron	Cre09.g407000	cobW	80,553
25,074		TTGGACGCGGACCCGGCGCAG	+	9	6,518,859	Intron	g10105	TENA_THI-4	26,633
			—		6,519,021	Intergenic			

^a Asterisks in the cluster number indicate that no miRNA* was identified in the small RNA library. Letters added to the cluster number indicate the existence of more than one predominant, candidate miRNA product.

^b Names for previously identified miRNAs are taken from miRBase, Molnar *et al.* (2007), Zhao *et al.* (2007), Shu and Hu (2012), and Yamasaki *et al.* (2013).

^c Strand orientation of the mapped reads. In cases where the miRNA is processed from a perfectly complementary region of the hairpin, the miRNA will map to both strands within the same locus. In these cases, the coordinates and region are shown for both strands (see Figure S4).

^d Chromosome where the cluster is located.

^e Start location for the miRNA (5' nucleotide), relative to the plus strand. For reads that map to the minus strand, this represents the position of the miRNA 3' nucleotide.

^f If the miRNA maps to a gene, the location of the miRNA within the gene is shown along with information about the gene itself, including conserved Pfam A domains.

^g Read counts for each miRNA are shown in absolute numbers and are not normalized by library size.

^h Cluster 3,704 corresponds to the predicted precursor RNA of miR1148 but the identified predominant read differs from the annotated miR1148.1 and miR1148.2 sequences.

ⁱ Cluster 11,846 corresponds to the predicted precursor RNA of miR1164 but the identified predominant read differs from the annotated miR1164 sequence.

^j Cluster 17,755 corresponds to the complementary strand of the miR1154 precursor RNA. miR1154 reads were found at very low abundance in the library of AGO3-associated small RNAs (Table S4).

^k Cluster 18,100 corresponds to the predicted precursor RNA of miR919 but the identified predominant reads differ from the annotated miR919.1 and miR919.2 sequences.

Figure 3 Predicted precursor structure for cluster 19,166 and detection of *Chlamydomonas* miRNAs by Northern hybridization. (A) Secondary structure plot of the predicted precursor miRNA for cluster 19,166 and frequency of reads from the *Chlamydomonas* AGO3-associated sRNA library. The mature miRNA and miRNA* sequences are indicated in boldface type. The putative pre-miRNA structure, shown in diagrammatic form on the right, includes nucleotides 2,201,614–2,201,689 of the minus strand of chromosome 6. (B) Northern blot analyses of small RNAs isolated from the indicated strains and detected with probes specific for each miRNA. The same filters were re-probed with the U6 small nuclear RNA sequence as a control for equivalent loading of the lanes. CC-124, wild type strain; Maa7-IR44s, CC-124 transformed with an IR transgene designed to induce RNAi of MAA7 (encoding tryptophan synthase β -subunit); and Mut-20, *TSN1* deletion mutant.

through cleavage and those likely to be regulated through translational repression. Cleavage targets were set to require perfect matching between a transcript and nucleotides 2–12 of the miRNA, encompassing the seed region (nucleotides 2–8; see Figure S1) and the catalytic or center region (nucleotides 9–12, where AGO-mediated cleavage of the target occurs; see Figure S1). In addition, they could have no more than three mismatches and three GU wobbles in the remainder of the miRNA/target hybrid or a gap of >1 nt (Figure 1B).

Based on the above criteria, we identified 74 target transcripts but none corresponded to a perfect miRNA:mRNA match (Table S1). To provide some validation for these findings, we examined transcript abundance of the putative targets in an RNAi-defective mutant, Mut-20, and in its parental strain, Maa7-IR44s (Table S2). Mut-20 contains a deletion of the gene coding for Tudor Staphylococcal Nuclease 1 (TSN1) (Figure 4), an evolutionarily conserved protein with apparently diverse functions in different eukaryotes (Caudy *et al.* 2003; Li *et al.* 2008; Frei dit Frey *et al.* 2010; S. Liu *et al.* 2010; Phetrungnapha *et al.* 2013; Gao *et al.* 2014). Tudor-SN is characterized by four N-terminal tandem repeats of the staphylococcal nuclease-like (SN) motif and a C-terminal TSN (Tudor-SN5) domain (Figure 4B) (Li *et al.* 2008; K. Liu *et al.* 2010; Gao *et al.* 2014). In land plants, Tudor-SN appears to associate with the cytoskeleton and play a role in stress responses and seed germination (Frei dit Frey *et al.* 2010; S. Liu *et al.* 2010). In metazoans, among other functions, Tudor-SN was found associated with stress granules and with RISC (Caudy *et al.* 2003; K. Liu *et al.* 2010;

Phetrungnapha *et al.* 2013; Gao *et al.* 2014), although its exact molecular role in RNA interference remains undefined. Intriguingly, in *Trypanosoma brucei* a protein (TbRIF5) containing tandem staphylococcal nuclease domains physically associates with a Dicer-like protein (Barnes *et al.* 2012), suggesting a possible involvement of proteins with SN motifs in sRNA biogenesis (Burroughs *et al.* 2014). Notably, *Chlamydomonas* Mut-20 shows greatly reduced sRNA/miRNA amounts (Figure 3B and Figure 4C) (Yamasaki *et al.* 2013) and, thus, transcripts targeted for direct cleavage by RISC would be expected to display increased accumulation in this mutant background. Yet only 6 of the predicted targets showed at least a twofold increase in transcript abundance in Mut-20 relative to the wild-type control (Table S1), and these changes were not statistically significant (adjusted *P*-values >0.05). The overall deficiency in sRNAs in *Chlamydomonas* Mut-20 would affect their association with any of the argonaute proteins (AGO1, AGO2, or AGO3), whose genes are not altered in this mutant background, and therefore compensatory loading of miRNAs into different AGOs would not explain the observed limited effect on target transcript stability.

Putative translational repression targets were set to require perfect matching to the seed region and no more than three mismatches and three GU wobbles in the 3' end of the hybrid duplex. However, unlike the pairing interaction for cleavage targets, they were allowed to have up to three mismatches or GU wobble pairs in the catalytic region (Figure 1B). These criteria are more stringent than what has been experimentally demonstrated as necessary for a reduction in target protein

Figure 4 Mut-20 contains a deletion encompassing the *Tudor Staphylococcal Nuclease 1 (TSN1)* gene. (A) Diagram of the *TSN1* (g11863) gene and surrounding chromosomal regions. The solid horizontal bar indicates the extent of the deletion in Mut-20. PCR analyses of genomic DNA in Mut-20 and in the parental strain, Maa7-IR44s, are shown below the diagram. (B) Drawing of the *TSN1* protein indicating the staphylococcal nuclease-like (SN) motifs and the Tudor domain. (C) Column-purified small RNAs from the indicated strains were separated in a 15% denaturing polyacrylamide gel and visualized by SYBR-Gold staining.

levels in *C. reinhardtii*, based on tests with reporter miRNA-target constructs (Yamasaki *et al.* 2013). Despite these stringent requirements, we identified 488 transcripts potentially repressed at the translational level (Table S3) or, on average, >10 putative targets per candidate miRNA. The abundance of most of these transcripts did not change significantly in the Mut-20 strain relative to the wild-type control (Table S3). Still, it remains uncertain how many of these RNAs, if any, are genuine miRNA targets possibly regulated via translation repression (see below).

sRNA-mediated endogenous gene regulation

To gain insight into how miRNAs and related sRNAs may regulate endogenous genes in *C. reinhardtii*, we examined mRNA and protein abundance for putative targets of a few sRNAs in Mut-20 and in the Maa7-IR44s parental strain. We chose to analyze miR B, even though it has no identifiable precursor RNA encoded in the *Chlamydomonas* genome (see above), because it is the only sRNA for which we could find computationally an unambiguous, perfectly complementary target transcript (Figure 5A), albeit coding for a protein of unknown function (Cre17.g697550). As expected for RISC-mediated target cleavage, 5'-RACE assays on RNA samples isolated from the strain with a wild-type background (Maa7-IR44s) detected

Cre17.g697550 transcripts predominantly truncated in between the residues hybridizing to nucleotides 10 and 11 of miR B (Figure 5A). Moreover, as anticipated for a mRNA regulated by miRNA-directed cleavage, the Cre17.g697550 transcript abundance increased substantially in Mut-20 (Figure 5B and Table S2). This result suggested that certain *Chlamydomonas* genes might be controlled by miRNAs/sRNAs through RISC-mediated mRNA cleavage.

To examine whether *Chlamydomonas* sRNAs can regulate endogenous transcripts by translation repression, we chose to test putative targets for cluster 14,711 (*i.e.*, Cre14.g629050, encoding a predicted copper-binding protein), cluster 19,166 (*i.e.*, Cre16.g659300, encoding a predicted cytochrome b5 isoform), and miR C (*i.e.*, Cre16.g683650, encoding a predicted protein kinase). These targets were selected because they code for predicted proteins of known functions and their miRNA-binding sites (Figure 6A and Table S3) display features similar to those characterized as involved in miRNA-mediated translation inhibition in animal systems (Bartel 2009; Tarver *et al.* 2012; Brümmer and Hausser 2014). Yet for the putative targets of clusters 14,711 and 19,166 no appreciable changes were detected in transcript levels (Table S3) or protein abundance (data not shown) in the Mut-20 strain relative to those in the Maa7-IR44s parental strain, suggesting that their

expression is not controlled by miRNAs under standard laboratory conditions.

As already mentioned, miR C maps next to a gap in the genome and, as a consequence, cannot be folded into a canonical hairpin with the miRNA sequence contained within the double-stranded RNA stem, although flanking sequences do form a long hairpin structure (see Figure S5). As previously proposed for this kind of *Chlamydomonas* sRNAs mapping next to sequence gaps (Molnar *et al.* 2007), miR C could correspond to a genuine miRNA or to a small interfering RNA (siRNA) processed from a very long double-stranded RNA precursor. The miR C binding site in *Cre16.g683650* is located in the 3'-UTR, partly overlapping the stop codon, and it contains a mismatch with nucleotide 10 of the sRNA, which would hinder AGO-mediated cleavage (Wee *et al.* 2012) (Figure 6A). Interestingly, the transcript abundance of the miR C target, *Cre16.g683650*, was reduced in Mut-20 relative to the wild type (Figure 6B and Table S2). In contrast, the protein abundance, detected by immunoblotting, was slightly increased in the mutant background (Figure 6C). Similar results were obtained with another *Chlamydomonas* mutant, defective in DCL3 (g8048), which shows greatly reduced levels of miR C (see Figure S6). Given that the *Cre16.g683650* transcript abundance is higher in the wild type without a concomitant increase in protein amount relative to Mut-20 or Dcl-3, these observations suggested that *Cre16.g683650* is translationally repressed, albeit at modest levels, in wild-type *Chlamydomonas*. This repression is partly relieved in the mutants, as expected for a sRNA/miRNA-regulated transcript (see Discussion for details).

Discussion

AGO3-associated miRNA population in *C. reinhardtii*

The sRNAs associated with AGO3 show a strong bias for a length of ~21 nt and the presence of uracil as the 5' nucleotide. The same trends were reported in the original studies describing miRNAs in *C. reinhardtii*, which characterized total cellular sRNA populations (Molnar *et al.* 2007; Zhao *et al.* 2007), but are much more pronounced in the AGO3-associated sRNAs. In particular, uracil was the most prevalent 5' nucleotide in previous reports, but it was present in <50% of the unique reads and a considerable number of sRNAs started with other nucleotides, most notably adenine (Molnar *et al.* 2007; Zhao *et al.* 2007). Since 80% of the AGO3-associated unique sequences (98% of total mapped reads) begin with a uracil (Figure 2E), it seems likely that the more variable sRNAs sequenced in studies of total cellular sRNA populations represent those associated with all three *Chlamydomonas* AGO proteins, implying that AGO1 and/or AGO2 may have different preferences than AGO3 for the first nucleotide of a sRNA. Alternatively, as suggested for animal systems (Schug *et al.* 2013; Flores *et al.* 2014; Thomson *et al.* 2014), only a fraction of mature sRNAs might be AGO associated in *Chlamydomonas* and total cellular sRNA populations may also include processed sRNAs that are not ordinarily incorporated into RISC. As

a methodological drawback, the 5'-nucleotide bias had a significant impact on the identification of miRNAs* in libraries of AGO3-associated sRNAs, as those miRNAs* that begin with a nucleotide other than U are much less likely to be represented in the libraries. Because of this technical issue, the presence of matching miRNA* sequences in our libraries was not used as a criterion for selecting canonical miRNAs.

By clustering all of the sRNAs associated with AGO3 based on their genomic locations, we were able to identify precursor RNA hairpins and classify the clusters as generating canonical miRNAs or other types of sRNAs (Coruh *et al.* 2014), taking into account hairpin folding and the apparent 5' processing of dominant reads. Using the criteria set forth by Tarver *et al.* (2012), we identified 32 novel miRNAs and confirmed 13 previously identified miRNAs, which together account for ~28% of all genome-mapped sRNAs associated with AGO3 (Figure 2C). In principle, this suggests that the majority of the AGO3-bound sRNAs may not be miRNAs, including many derived from very long hairpins that give rise to multiple products (Table S4). However, from a functional perspective, most of these non-miRNA sequences may have limited relevance. Recent findings in metazoans suggest that only strongly expressed sRNAs, above a certain threshold level, may lead to functionally significant target suppression (Mullokandov *et al.* 2012; Flores *et al.* 2014; Thomson *et al.* 2014). In *Chlamydomonas*, ~70% of the genome-mapped sRNAs associated with AGO3 (Figure 2C) correspond to nearly 100,000 distinct (unique) sequences (Figure 2D) with each being present on average at fairly low, possibly ineffective abundance.

Another caveat is that the *Chlamydomonas* nuclear genome is not fully sequenced and frequently sRNA sequences appear to lie next to sequence gaps (this work and Molnar *et al.* 2007). Thus, for many sRNAs it cannot be unambiguously ascertained whether they may represent miRNAs or siRNAs (e.g., miR C; see Figure S5). Moreover, almost 40% of the high-quality reads did not have a perfect match in the *Chlamydomonas* nuclear genome or a predicted precursor hairpin structure and were not analyzed further. Yet some of these sequences are very abundant and may represent functional sRNAs, as suggested for miR B (Figure 5), possibly encoded in unsequenced regions of the genome. This technical problem currently prevents a definitive analysis of sRNA populations in *C. reinhardtii*.

Interestingly, the clusters that meet the criteria for canonical miRNAs tend to have fewer mismatches between the miRNA and miRNA* than observed for plants or animals, which can have up to four or six mismatches within the duplex, respectively (Tarver *et al.* 2012). Indeed, many of the miRNA-encoding duplexes in *C. reinhardtii* show perfect complementarity (Figure S4). As a methodological difficulty, perfect complementarity in the stem of a hairpin encoding miRNA and miRNA* molecules results in both sequences mapping to both strands of the genome in a given cluster. However, mismatches in other parts of the hairpin stem often allow the identification of one strand with significantly more reads mapping to it, suggesting that that strand corresponds to the pre-miRNA (Figure S4). Based on these analyses, the 45 identified

Figure 5 Analysis of the miR B target *Cre17.g697550*. (A) Diagram of the *Cre17.g697550* gene (encoding a predicted protein of unknown function) indicating the miR B-binding site within the coding sequence. Hybridization of miR B to the target site is shown below. The arrows and numbers indicate the position and frequency, respectively, of 5' termini of truncated target mRNAs examined by 5' RACE. (B) Quantitative RT-PCR analysis of *Cre17.g697550* transcript levels in the indicated strains. Values, normalized to those in the wild-type strain, are means $\pm$ SD of three independent experiments. Samples are significantly different ($P < 0.05$) in a two-tailed Student's *t*-test.

canonical miRNAs are encoded almost in equal proportions in intergenic regions of the *Chlamydomonas* nuclear genome or in introns or the 3'-UTR of predicted (many hypothetical) protein-coding genes (Table 1).

miRNA targets and mechanism(s) of endogenous gene regulation

Despite apparent similarities in the RNAi machineries of *C. reinhardtii* and higher plants (Casas-Mollano *et al.* 2008; Cerutti *et al.* 2011), *Chlamydomonas* miRNAs do not seem to regulate targets primarily through the plant model of high complementarity between miRNA and transcript, resulting in mRNA cleavage (Liu *et al.* 2014). While we identified 74 target transcripts that had perfect complementarity to the corresponding miRNAs from nucleotides 2 to 12 (encompassing both the seed and catalytic regions; see Figure S1) and no more than three mismatches and three GU wobbles in the remaining region, only 6 of the predicted targets (8.1%) showed at least a twofold increase in transcript abundance in the miRNA-deficient Mut-20 (Table S1). The *Chlamydomonas* RNAi machinery does have the capability to operate by target transcript cleavage, as demonstrated with artificial miRNA constructs and reporters (Molnar *et al.* 2009; Zhao *et al.* 2009; Yamasaki *et al.* 2013) and suggested by the analysis of the miR B predicted endogenous target *Cre17.g697550* (Figure 5). However, with the caveat that some predictions may represent false positives (see below), the majority of the candidate endogenous targets do not appear to be subject to mRNA cleavage and degradation, since the anticipated increase in transcript abundance in Mut-20, which is almost devoid of sRNAs, was not observed. Interestingly, the Mut-20 strain did show elevated RNA levels (Table S2) for most genes predicted to encode miRNAs in their coding or untranslated regions (Table 1), consistent

with a role of the deleted *TSN1* in sRNA biogenesis. In contrast, when a miRNA was predicted to be encoded in an intron (Table 1), the corresponding transcript abundance did not change in a coherent manner in Mut-20 (Table S2). Yet, given the multiple functions of Tudor-SN proteins in eukaryotes (Freidit Frey *et al.* 2010; S. Liu *et al.* 2010; Phetrungnapha *et al.* 2013; Gao *et al.* 2014), we cannot rule out that some changes in transcript levels in the mutant background may be unrelated to an RNAi defect. Additionally, changes in mRNA abundance in Mut-20 may also be due to downstream effects of misregulated sRNA-targeted genes.

When considering the possibility of miRNA-mediated translational repression in *Chlamydomonas*, even using complementarity criteria far more strict than necessary for reduction of protein levels in miRNA-target reporter experiments (Yamasaki *et al.* 2013), we predicted on average >10 targets per miRNA (488 target transcripts for 45 miRNAs). However, most predicted translation repression targets did not show appreciable changes in transcript abundance in Mut-20 (Table S3) and it is unclear what fraction of these genes might correspond to genuine sRNA-regulated targets. Again, the *Chlamydomonas* RNAi machinery does have the capability to cause translation repression of target transcripts, as demonstrated with transgenic constructs (Ma *et al.* 2013; Yamasaki *et al.* 2013) and suggested by the analysis of the miR C candidate target, *Cre16.g683650* (Figure 6 and Figure S6). Yet only one of three examined putative endogenous targets showed modest changes in transcript and protein levels in Mut-20 consistent with sRNA-mediated translation repression.

MicroRNA:mRNA interactions are generally thought to result in gene silencing by reducing the stability and/or the translation of the RISC-associated transcripts. However, several reports suggest that this is not always the case, with miRNAs also being capable of protecting or even increasing

Figure 6 Analysis of the miR C target *Cre16.g683650*. (A) Diagram of the *Cre16.g683650* gene (encoding a predicted protein kinase) indicating the miR C-binding site overlapping the stop codon. Hybridization of miR C to the target site is shown below. The mismatch to nucleotide 10 of the miRNA would prevent cleavage of the target mRNA. (B) Quantitative RT-PCR analysis of *Cre16.g683650* transcript levels in the indicated strains. Values, normalized to those in the wild-type strain, are means $\pm$ SD of three independent experiments. Samples are significantly different ($P < 0.05$) in a two-tailed Student's *t*-test. (C) Immunoblot analyses of *Cre16.g683650* and histone H3 (loading control) protein levels. The graph represents densitometric quantitation of the *Cre16.g683650* immunoblot signals (normalized to H3). Values are means $\pm$ SD of three independent experiments.

the steady-state levels of target transcripts (Murphy *et al.* 2010; Carroll *et al.* 2012). Interestingly, the *Cre16.g683650* mRNA abundance was reduced in Mut-20 (Figure 6B), as well as in another RNAi mutant defective in DCL3 (see Figure S6), suggesting that the transcript may be stabilized by the presence of a functional RNAi machinery. We have previously demonstrated that mRNAs translationally repressed by sRNA-mediated mechanisms in *Chlamydomonas* remain associated with polyribosomes (Ma *et al.* 2013), which may perhaps protect against the action of ribonucleases. Another possibility is that some RISC-associated mRNAs may be sequestered within intracellular compartments and sheltered from degradation (Carroll *et al.* 2012). Despite a decrease in *Cre16.g683650* transcript amount to ~ 40 – 65% of wild-type levels, the corresponding protein was ~ 20 – 50% more abundant in Mut-20 or Dcl-3 (Figure 6 and Figure S6), implying that *Cre16.g683650* may indeed be translationally repressed in the wild-type parental strain. Yet if this case is representative of endogenous miRNA targets, net changes in protein levels between mutant and wild-type strains are fairly small, making it very difficult to validate predicted translation repression targets by high-throughput proteomic approaches.

For both higher plants and animals, target prediction uses evolutionary conservation of the miRNA-binding site to reduce the number of potential false positives (Willmann and Poethig 2007; Thiebaut *et al.* 2012; Allmer 2014). However, the recent profiling of sRNAs and miRNAs in the green alga *Volvox carterii*, the closest relative to *C. reinhardtii* with a published genome, indicated that there is very little, if any, conservation of miRNAs

between these two species (data not shown and Li *et al.* 2014). Other features that are commonly used to reduce the number of false positive targets are the accessibility of the miRNA target site and the change of free energy for binding (Kertesz *et al.* 2007), the nucleotide composition of the miRNA and target site analyzed with machine-learning-based classifiers (Yang *et al.* 2008; Mitra and Bandyopadhyay 2011; Kurubanjerdjit *et al.* 2013), and the presence of certain RNA-binding sites for protein modules overlapping the miRNA-binding site (Didiano and Hobert 2006, 2008; Grimson *et al.* 2007; Kedde *et al.* 2007). Minimum free energy hybridization criteria, incorporated into RNAhybrid (Kruger and Rehmsmeier 2006), were used in our searches for targets but implementing other approaches would require a large set of confirmed targets to apply machine-learning methods and/or knowledge of RNA-binding sites for specific protein domains. Nevertheless, our results suggest, as recently described for reporter constructs (Yamasaki *et al.* 2013), that miRNA regulation of endogenous transcript expression in *Chlamydomonas* may not involve the extensive complementarity that is typical of land plants (Liu *et al.* 2014). For certain targets, it may operate by translation repression although with modulatory, rather modest effects on protein abundance. Yet, given the methodological limitations outlined above, it remains uncertain what proportion of predicted *Chlamydomonas* targets are genuine.

Evolutionary role of miRNAs in green algae

In multicellular plants and animals, miRNAs play key roles in maintaining cell/tissue differentiation, and disruption in

miRNA production leads to abnormal growth and/or cancer or is fatal to the organisms (Fahlgren *et al.* 2010; Lenz *et al.* 2011; Sheng and Previti 2011). However, the lack of miRNA conservation among examined green algae [as well as with any other organism (Lenz *et al.* 2011; Nozawa *et al.* 2012)], the identification of few predicted targets with known function(s) (Table S1 and Table S3), and the modulatory nature of the regulation of some targets (~20–50% change in protein levels between wild type and RNAi-defective mutants) raise questions about the biological role(s) of miRNAs in *C. reinhardtii*. Furthermore, the miRNA/sRNA-deficient mutants, Mut-20 and Dcl-3, have no obvious abnormal phenotypes when grown under standard laboratory conditions (data not shown), implying that small RNAs are not essential for cell growth and survival. This is consistent with the observation that many unicellular eukaryotes, particularly those with small nuclear genomes, seem to have lost entirely the RNAi machinery or have retained only a basic set of RNAi components (Cerutti and Casas-Mollano 2006; Nakayashiki *et al.* 2006; Berezikov 2011; Drinnenberg *et al.* 2011). The elusive role of RNA interference in controlling endogenous gene expression in unicellular eukaryotes has also been recognized in the fission yeast *Schizosaccharomyces pombe* (Smialowska *et al.* 2014). Whereas RNAi does play a role in transposon silencing in *Chlamydomonas* (Casas-Mollano *et al.* 2008) and possibly in defense responses against (currently unidentified) viruses, defining the function of miRNAs will require additional work. For instance, miRNAs may have only a modulatory role in controlling gene expression under optimal laboratory conditions but they may become important regulators under certain stress conditions.

The lack of miRNA conservation with *V. carteri* also raises questions about how these sRNAs evolve in green algae. The two main possibilities are that miRNAs are transient, with a very fast evolutionary turnover, or that all of the miRNAs evolved independently in the two lineages after their divergence. MicroRNAs as accidental products could explain a narrow functional role and justify a fast turnover rate in *C. reinhardtii*. As proposed in plants and animals (Fahlgren *et al.* 2010; Cuperus *et al.* 2011; Lenz *et al.* 2011; Nozawa *et al.* 2012), hairpins may randomly arise both in introns and in transcribed intergenic regions as the genome evolves. The RNAi machinery would process transcribed RNA hairpins, generating endogenous sRNAs and potential miRNAs. However, these hairpins would evolve neutrally unless a processed sRNA confers a selective advantage or disadvantage to the cell. As a result, miRNA precursors could arise, evolve accumulating mutations over time, and eventually be lost from the genome. That is a pattern of rapid birth-and-death evolution (Cuperus *et al.* 2011; Lenz *et al.* 2011; Nozawa *et al.* 2012). Despite *V. carteri* being the closest relative of *C. reinhardtii* with a sequenced genome, these two species diverged ~200 million years ago (Chang and Liao 2013), possibly enough time for fast-evolving miRNAs to be diverged and completely turned over. Interestingly, while the miRNAs in higher plants that maintain tissue differentiation are highly conserved, those that are species specific often appear to follow this rapid turnover

model of evolution (Fahlgren *et al.* 2010; Cuperus *et al.* 2011; Lenz *et al.* 2011; Nozawa *et al.* 2012).

Conclusion

Much additional work will be needed to understand what role(s) miRNAs play in *C. reinhardtii*. Our results suggest that most AGO3-associated sRNAs are not canonical miRNAs although the latter appear to correspond to the most abundant, possibly functionally relevant, sequences. Regulation of *Chlamydomonas* protein-coding genes by miRNA-mediated mRNA cleavage seems to be rare. In addition, miRNAs predicted to operate by translational repression appear to have, with some exceptions, little effect on targets' transcript amounts. Some of these small RNAs may primarily fine-tune protein levels, as suggested by examination of a miR C target. However, *C. reinhardtii* miRNAs/sRNAs do not seem to regulate processes essential for cell growth and survival under standard laboratory conditions. It is tempting to speculate that the RNAi machinery initially evolved, in the last common ancestor of eukaryotes, as a defense response against viruses and other genomic parasites (Cerutti and Casas-Mollano 2006) and that regulatory miRNAs evolved at a later time mostly to control newly arisen processes such as cell differentiation and development in multicellular organisms. In unicellular eukaryotes, miRNAs arising as accidental products of random genome evolution may have no role in regulating ancestral, core cellular functions but may possibly be involved in more recently evolved responses to specific abiotic and/or biotic stresses.

Acknowledgments

This work was supported in part by a grant from the National Science Foundation (to H.C.). We also acknowledge the support of the Nebraska Experimental Program to Stimulate Competitive Research.

Literature Cited

- Allmer, J., 2014 Computational and bioinformatics methods for microRNA gene prediction. *Methods Mol. Biol.* 1107: 157–175.
- Anders, S., and W. Huber, 2010 Differential expression analysis for sequence count data. *Genome Biol.* 11: R106.
- Axtell, M. J., 2013 Classification and comparison of small RNAs from plants. *Annu. Rev. Plant Biol.* 64: 137–159.
- Barnes, R. L., H. Shi, N. G. Kolev, C. Tschudi, and E. Ullu, 2012 Comparative genomics reveals two novel RNAi factors in *Trypanosoma brucei* and provides insight into the core machinery. *PLoS Pathog.* 8: e1002678.
- Bartel, D. P., 2009 MicroRNAs: target recognition and regulatory functions. *Cell* 136: 215–233.
- Beauchair, L., A. Yu, and N. Bouche, 2010 microRNA-directed cleavage and translational repression of the copper chaperone for superoxide dismutase mRNA in *Arabidopsis*. *Plant J.* 62: 454–462.
- Berezikov, E., 2011 Evolution of microRNA diversity and regulation in animals. *Nat. Rev. Genet.* 12: 846–860.
- Brueggeman, A. J., D. S. Gangadharai, M. F. Cserhati, D. Casero, D. P. Weeks *et al.*, 2012 Activation of the carbon concentrating mechanism by CO₂ deprivation coincides with massive transcriptional

- restructuring in *Chlamydomonas reinhardtii*. *Plant Cell* 24: 1860–1875.
- Brümmer, A., and J. Hausser, 2014 MicroRNA binding sites in the coding region of mRNAs: extending the repertoire of post-transcriptional gene regulation. *BioEssays* 36: 617–626.
- Burroughs, A. M., Y. Ando, and L. Aravind, 2014 New perspectives on the diversification of the RNA interference system: insights from comparative genomics and small RNA sequencing. *Wiley Interdiscip. Rev. RNA* 5: 141–181.
- Carroll, A. P., N. Tran, P. A. Tooney, and M. J. Cairns, 2012 Alternative mRNA fates identified in microRNA-associated transcriptome analysis. *BMC Genomics* 13: 561.
- Carthew, R. W., and E. J. Sontheimer, 2009 Origins and mechanisms of miRNAs and siRNAs. *Cell* 136: 642–655.
- Casas-Mollano, J. A., J. Rohr, E. J. Kim, E. Balassa, K. van Dijk *et al.*, 2008 Diversification of the core RNA interference machinery in *Chlamydomonas reinhardtii* and the role of DCL1 in transposon silencing. *Genetics* 179: 69–81.
- Caudy, A. A., R. F. Ketting, S. M. Hammond, A. M. Denli, A. M. Bathoorn *et al.*, 2003 A micrococcal nuclease homologue in RNAi effector complexes. *Nature* 425: 411–414.
- Cerutti, H., and J. A. Casas-Mollano, 2006 On the origin and functions of RNA-mediated silencing: from protists to man. *Curr. Genet.* 50: 81–99.
- Cerutti, H., X. Ma, J. Msanne, and T. Repas, 2011 RNA-mediated silencing in algae: biological roles and tools for analysis of gene function. *Eukaryot. Cell* 10: 1164–1172.
- Chang, T. Y., and B. Y. Liao, 2013 Flagellated algae protein evolution suggests the prevalence of lineage-specific rules governing evolutionary rates of eukaryotic proteins. *Genome Biol. Evol.* 5: 913–922.
- Coruh, C., S. Shahid, and M. A. Axtell, 2014 Seeing the forest for the trees: annotating small RNA producing genes in plants. *Curr. Opin. Plant Biol.* 18: 87–95.
- Crooks, G. E., G. Hon, J. M. Chandonia, and S. E. Brenner, 2004 WebLogo: a sequence logo generator. *Genome Res.* 14: 1188–1190.
- Cuperus, J. T., N. Fahlgren, and J. C. Carrington, 2011 Evolution and functional diversification of miRNA genes. *Plant Cell* 23: 431–442.
- Didiano, D., and O. Hobert, 2006 Perfect seed pairing is not a generally reliable predictor for miRNA-target interactions. *Nat. Struct. Mol. Biol.* 13: 849–851.
- Didiano, D., and O. Hobert, 2008 Molecular architecture of a miRNA-regulated 3' UTR. *RNA* 14: 1297–1317.
- Djuranovic, S., A. Nahvi, and R. Green, 2012 miRNA-mediated gene silencing by translational repression followed by mRNA deadenylation and decay. *Science* 336: 237–240.
- Drinneberg, I. A., G. R. Fink, and D. P. Bartel, 2011 Compatibility with killer explains the rise of RNAi-deficient fungi. *Science* 333: 1592.
- Einhauser, A., and A. Jungbauer, 2001 The FLAG peptide, a versatile fusion tag for the purification of recombinant proteins. *J. Biochem. Biophys. Methods* 49: 455–465.
- Endo, Y., H. O. Iwakawa, and Y. Tomari, 2013 Arabidopsis ARGONAUTE7 selects miR390 through multiple checkpoints during RISC assembly. *EMBO Rep.* 14: 652–658.
- Fahlgren, N., S. Jogdeo, K. D. Kasschau, C. M. Sullivan, E. J. Chapman *et al.*, 2010 MicroRNA gene evolution in Arabidopsis lyrata and Arabidopsis thaliana. *Plant Cell* 22: 1074–1089.
- Finn, R. D., A. Bateman, J. Clements, P. Coghill, R. Y. Eberhardt *et al.*, 2014 Pfam: the protein families database. *Nucleic Acids Res.* 42: D222–D230.
- Fischer, N., and J. D. Rochaix, 2001 The flanking regions of PsAD drive efficient gene expression in the nucleus of the green alga *Chlamydomonas reinhardtii*. *Mol. Genet. Genomics* 265: 888–894.
- Flores, O., E. M. Kennedy, R. L. Skalsky, and B. R. Cullen, 2014 Differential RISC association of endogenous human microRNAs predicts their inhibitory potential. *Nucleic Acids Res.* 42: 4629–4639.
- Frank, F., N. Sonenberg, and B. Nagar, 2010 Structural basis for 5'-nucleotide base-specific recognition of guide RNA by human AGO2. *Nature* 465: 818–822.
- Frank, F., J. Hauver, N. Sonenberg, and B. Nagar, 2012 Arabidopsis Argonaute MID domains use their nucleotide specificity loop to sort small RNAs. *EMBO J.* 31: 3588–3595.
- Frei dit Frey, N., P. Muller, F. Jammes, D. Kizis, J. Leung *et al.*, 2010 The RNA binding protein Tudor-SN is essential for stress tolerance and stabilizes levels of stress-responsive mRNAs encoding secreted proteins in Arabidopsis. *Plant Cell* 22: 1575–1591.
- Fujiwara, T., and T. Yada, 2013 miRNA-target prediction based on transcriptional regulation. *BMC Genomics* 14(Suppl. 2): S3.
- Gao, X., X. Shi, X. Fu, L. Ge, Y. Zhang *et al.*, 2014 Human Tudor staphylococcal nuclease (Tudor-SN) protein modulates the kinetics of AGTR1-3'UTR granule formation. *FEBS Lett.* 588: 2154–2161.
- Goodstein, D. M., S. Shu, R. Howson, R. Neupane, R. D. Hayes *et al.*, 2012 Phytozone: a comparative platform for green plant genomics. *Nucleic Acids Res.* 40: D1178–D1186.
- Grimson, A., K. K. Farh, W. K. Johnston, P. Garrett-Engele, L. P. Lim *et al.*, 2007 MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol. Cell* 27: 91–105.
- Ibrahim, F., L. A. Rymarquis, E. J. Kim, J. Becker, E. Balassa *et al.*, 2010 Uridylation of mature miRNAs and siRNAs by the MUT68 nucleotidyltransferase promotes their degradation in *Chlamydomonas*. *Proc. Natl. Acad. Sci. USA* 107: 3906–3911.
- Janas, M. M., B. Wang, A. S. Harris, M. Aguiar, J. M. Shaffer *et al.*, 2012 Alternative RISC assembly: binding and repression of microRNA-mRNA duplexes by human Ago proteins. *RNA* 18: 2041–2055.
- Kapitonov, V. V., and J. Jurka, 2008 A universal classification of eukaryotic transposable elements implemented in Repbase. *Nat. Rev. Genet.* 9: 411–412, author reply 414.
- Kedde, M., M. J. Strasser, B. Boldajipour, J. A. Oude Vrielink, K. Slanchev *et al.*, 2007 RNA-binding protein Dnd1 inhibits microRNA access to target mRNA. *Cell* 131: 1273–1286.
- Kertesz, M., N. Iovino, U. Unnerstall, U. Gaul, and E. Segal, 2007 The role of site accessibility in microRNA target recognition. *Nat. Genet.* 39: 1278–1284.
- Kim, V. N., 2005 MicroRNA biogenesis: coordinated cropping and dicing. *Nat. Rev. Mol. Cell Biol.* 6: 376–385.
- Kozomara, A., and S. Griffiths-Jones, 2014 miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.* 42(Database issue): D68–D73.
- Kruger, J., and M. Rehmsmeier, 2006 RNAhybrid: microRNA target prediction easy, fast and flexible. *Nucleic Acids Res.* 34: W451–W454.
- Kurubanjerdjit, N., C. H. Huang, Y. L. Lee, J. J. Tsai, and K. L. Ng, 2013 Prediction of microRNA-regulated protein interaction pathways in Arabidopsis using machine learning algorithms. *Comput. Biol. Med.* 43: 1645–1652.
- Lanet, E., E. Delannoy, R. Sormani, M. Floris, P. Brodersen *et al.*, 2009 Biochemical evidence for translational repression by Arabidopsis microRNAs. *Plant Cell* 21: 1762–1768.
- Langmead, B., and S. L. Salzberg, 2012 Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9: 357–359.
- Lenz, D., P. May, and D. Walthers, 2011 Comparative analysis of miRNAs and their targets across four plant species. *BMC Res. Notes* 4: 483.
- Li, C. L., W. Z. Yang, Y. P. Chen, and H. S. Yuan, 2008 Structural and functional insights into human Tudor-SN, a key component linking RNA interference and editing. *Nucleic Acids Res.* 36: 3579–3589.

- Li, J., Y. Wu, and Y. Qi, 2014 MicroRNAs in a multicellular green alga *Volvox carter*. *Sci. China Life Sci.* 57: 36–45.
- Li, S., L. Liu, X. Zhuang, Y. Yu, X. Liu *et al.*, 2013 MicroRNAs inhibit the translation of target mRNAs on the endoplasmic reticulum in *Arabidopsis*. *Cell* 153: 562–574.
- Liu, K., C. Chen, Y. Guo, R. Lam, C. Bian *et al.*, 2010 Structural basis for recognition of arginine methylated Piwi proteins by the extended Tudor domain. *Proc. Natl. Acad. Sci. USA* 107: 18398–18403.
- Liu, Q., F. Wang, and M. J. Axtell, 2014 Analysis of complementarity requirements for plant microRNA targeting using a *Nicotiana benthamiana* quantitative transient assay. *Plant Cell* 26: 741–753.
- Liu, S., J. Jia, Y. Gao, B. Zhang, and Y. Han, 2010 The AtTudor2, a protein with SN-Tudor domains, is involved in control of seed germination in *Arabidopsis*. *Planta* 232: 197–207.
- Lopez, D., D. Casero, S. J. Cokus, S. S. Merchant, and M. Pellegrini, 2011 Algal Functional Annotation Tool: a web-based analysis suite to functionally interpret large gene lists using integrated annotation and expression data. *BMC Bioinformatics* 12: 282.
- Lorenz, R., S. H. Bernhart, C. Honer Zu Siederdisen, H. Tafer, C. Flamm *et al.*, 2011 ViennaRNA package 2.0. *Algorithms Mol. Biol.* 6: 26.
- Lv, H., G. Qu, X. Qi, L. Lu, C. Tian *et al.*, 2013 Transcriptome analysis of *Chlamydomonas reinhardtii* during the process of lipid accumulation. *Genomics* 101: 229–237.
- Ma, X., E. J. Kim, I. Kook, F. Ma, A. Voshall *et al.*, 2013 Small interfering RNA-mediated translation repression alters ribosome sensitivity to inhibition by cycloheximide in *Chlamydomonas reinhardtii*. *Plant Cell* 25: 985–998.
- Merchant, S. S., S. E. Prochnik, O. Vallon, E. H. Harris, S. J. Karpowicz *et al.*, 2007 The *Chlamydomonas* genome reveals the evolution of key animal and plant functions. *Science* 318: 245–250.
- Mi, S., T. Cai, Y. Hu, Y. Chen, E. Hodges *et al.*, 2008 Sorting of small RNAs into *Arabidopsis* argonaute complexes is directed by the 5' terminal nucleotide. *Cell* 133: 116–127.
- Mitra, R., and S. Bandyopadhyay, 2011 MultiMiTar: a novel multi objective optimization based miRNA-target prediction method. *PLoS ONE* 6: e24583.
- Molnar, A., F. Schwach, D. J. Studholme, E. C. Thuenemann, and D. C. Baulcombe, 2007 miRNAs control gene expression in the single-cell alga *Chlamydomonas reinhardtii*. *Nature* 447: 1126–1129.
- Molnar, A., A. Bassett, E. Thuenemann, F. Schwach, S. Karkare *et al.*, 2009 Highly specific gene silencing by artificial microRNAs in the unicellular alga *Chlamydomonas reinhardtii*. *Plant J.* 58: 165–174.
- Mortazavi, A., B. A. Williams, K. McCue, L. Schaeffer, and B. Wold, 2008 Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* 5: 621–628.
- Mullokandov, G., A. Baccarini, A. Ruzo, A. D. Jayaprakash, N. Tung *et al.*, 2012 High-throughput assessment of microRNA activity and function using microRNA sensor and decoy libraries. *Nat. Methods* 9: 840–846.
- Murphy, A. J., P. M. Guyre, and P. A. Pioli, 2010 Estradiol suppresses NF-kappa B activation through coordinated regulation of let-7a and miR-125b in primary human macrophages. *J. Immunol.* 184: 5029–5037.
- Nakayashiki, H., N. Kadotani, and S. Mayama, 2006 Evolution and diversification of RNA silencing proteins in fungi. *J. Mol. Evol.* 63: 127–135.
- Nozawa, M., S. Miura, and M. Nei, 2012 Origins and evolution of microRNA genes in plant species. *Genome Biol. Evol.* 4: 230–239.
- Phetrungnapha, A., S. Panyim, and C. Ongvarrasopone, 2013 *Penaeus monodon* Tudor staphylococcal nuclease preferentially interacts with N-terminal domain of Argonaute-1. *Fish Shellfish Immunol.* 34: 875–884.
- Rogers, K., and X. Chen, 2013 Biogenesis, turnover, and mode of action of plant microRNAs. *Plant Cell* 25: 2383–2399.
- Schug, J., L. B. McKenna, G. Walton, N. Hand, S. Mukherjee *et al.*, 2013 Dynamic recruitment of microRNAs to their mRNA targets in the regenerating liver. *BMC Genomics* 14: 264.
- Shen, Y., Z. Jiang, S. Lu, H. Lin, S. Gao *et al.*, 2013 Combined small RNA and degradome sequencing reveals microRNA regulation during immature maize embryo dedifferentiation. *Biochem. Biophys. Res. Commun.* 441: 425–430.
- Sheng, G., H. Zhao, J. Wang, Y. Rao, W. Tian *et al.*, 2014 Structure-based cleavage mechanism of *Thermus thermophilus* Argonaute DNA guide strand-mediated DNA target cleavage. *Proc. Natl. Acad. Sci. USA* 111: 652–657.
- Sheng, Y., and C. Previti, 2011 Genomic features and computational identification of human microRNAs under long-range developmental regulation. *BMC Genomics* 12: 270.
- Shu, L., and Z. Hu, 2012 Characterization and differential expression of microRNAs elicited by sulfur deprivation in *Chlamydomonas reinhardtii*. *BMC Genomics* 13: 108.
- Smialowska, A., I. Djupedal, J. Wang, P. Kylsten, P. Swoboda *et al.*, 2014 RNAi mediates post-transcriptional repression of gene expression in fission yeast *Schizosaccharomyces pombe*. *Biochem. Biophys. Res. Commun.* 444: 254–259.
- Stalder, L., W. Heusermann, L. Sokol, D. Trojer, J. Wirz *et al.*, 2013 The rough endoplasmic reticulum is a central nucleation site of siRNA-mediated RNA silencing. *EMBO J.* 32: 1115–1127.
- Stevens, D. R., J. D. Rochaix, and S. Purton, 1996 The bacterial phleomycin resistance gene *ble* as a dominant selectable marker in *Chlamydomonas*. *Mol. Gen. Genet.* 24: 23–30.
- Tarver, J. E., P. C. Donoghue, and K. J. Peterson, 2012 Do miRNAs have a deep evolutionary history? *BioEssays* 34: 857–866.
- Thiebaut, F., C. Grativol, M. Carnavale-Bottino, C. A. Rojas, M. Tanurdzic *et al.*, 2012 Computational identification and analysis of novel sugarcane microRNAs. *BMC Genomics* 13: 290.
- Thomson, D. W., K. A. Pillman, M. L. Anderson, D. M. Lawrence, J. Toubia *et al.*, 2015 Assessing the gene regulatory properties of Argonaute-bound small RNAs of diverse genomic origin. *Nucleic Acids Res.* 43: 470–481.
- van Dijk, K., K. E. Marley, B. R. Jeong, J. Xu, J. Hesson *et al.*, 2005 Monomethyl histone H3 lysine 4 as an epigenetic mark for silenced euchromatin in *Chlamydomonas*. *Plant Cell* 17: 2439–2453.
- Voinnet, O., 2009 Origin, biogenesis, and activity of plant microRNAs. *Cell* 136: 669–687.
- Wang, Y., S. Juranek, H. Li, G. Sheng, T. Tuschl *et al.*, 2008a Structure of an argonaute silencing complex with a seed-containing guide DNA and target RNA duplex. *Nature* 456: 921–926.
- Wang, Y., G. Sheng, S. Juranek, T. Tuschl, and D. J. Patel, 2008b Structure of the guide-strand-containing argonaute silencing complex. *Nature* 456: 209–213.
- Wee, L. M., C. F. Flores-Jasso, W. E. Salomon, and P. D. Zamore, 2012 Argonaute divides its RNA guide into domains with distinct functions and RNA-binding properties. *Cell* 151: 1055–1067.
- Willmann, M. R., and R. S. Poethig, 2007 Conservation and evolution of miRNA regulatory programs in plant development. *Curr. Opin. Plant Biol.* 10: 503–511.
- Winter, J., and S. Diederichs, 2011 Argonaute proteins regulate microRNA stability: increased microRNA abundance by Argonaute proteins is due to microRNA stabilization. *RNA Biol.* 8: 1149–1157.
- Yamasaki, T., A. Voshall, E. J. Kim, E. Moriyama, H. Cerutti *et al.*, 2013 Complementarity to an miRNA seed region is sufficient to induce moderate repression of a target transcript in the

- unicellular green alga *Chlamydomonas reinhardtii*. *Plant J.* 76: 1045–1056.
- Yang, L., G. Wu, and R. S. Poethig, 2012 Mutations in the GW-repeat protein SUO reveal a developmental function for microRNA-mediated translational repression in *Arabidopsis*. *Proc. Natl. Acad. Sci. USA* 109: 315–320.
- Yang, Y., Y. P. Wang, and K. B. Li, 2008 MiRTif: a support vector machine-based microRNA target interaction filter. *BMC Bioinformatics* 9(Suppl. 12): S4.
- Zha, X., Q. Xia, and Y. A. Yuan, 2012 Structural insights into small RNA sorting and mRNA target binding by *Arabidopsis* Argonaute Mid domains. *FEBS Lett.* 586: 3200–3207.
- Zhao, T., G. Li, S. Mi, S. Li, G. J. Hannon *et al.*, 2007 A complex system of small RNAs in the unicellular green alga *Chlamydomonas reinhardtii*. *Genes Dev.* 21: 1190–1203.
- Zhao, T., W. Wang, X. Bai, and Y. Qi, 2009 Gene silencing by artificial microRNAs in *Chlamydomonas*. *Plant J.* 58: 157–164.
- Zheng, H., R. Fu, J. T. Wang, Q. Liu, H. Chen *et al.*, 2013 Advances in the techniques for the prediction of microRNA targets. *Int. J. Mol. Sci.* 14: 8179–8187.
- Zhou, M., and H. Luo, 2013 MicroRNA-mediated gene regulation: potential applications for plant genetic engineering. *Plant Mol. Biol.* 83: 59–75.
- Zhu, H., Y. Zhou, C. Castillo-Gonzalez, A. Lu, C. Ge *et al.*, 2013 Bidirectional processing of pri-miRNAs with branched terminal loops by *Arabidopsis* Dicer-like1. *Nat. Struct. Mol. Biol.* 20: 1106–1115.

Communicating editor: P. K. Geyer

GENETICS

Supporting Information

<http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.115.174797/-/DC1>

Identification of AGO3-Associated miRNAs and Computational Prediction of Their Targets in the Green Alga *Chlamydomonas reinhardtii*

Adam Voshall, Eun-Jeong Kim, Xinrong Ma, Etsuko N. Moriyama, and Heriberto Cerutti

Figure S1 Diagrammatic representation of the interaction between a miRNA and its target RNA sequence within an argonaute (AGO) protein. The miRNA is shown in red and is oriented 3' to 5' with the 5' most nucleotide bound to the Mid domain of AGO. The target sequence is shown in black, with the cleavage site indicated between nucleotides 10 and 11 of the miRNA and located within the AGO PIWI domain. The catalytic (or center) and seed regions are indicated by surrounding blue and red boxes, respectively.

Figure S2 Immunofluorescence localization of FLAG-tagged AGO3 in *Chlamydomonas reinhardtii*. The panels indicate DAPI staining of nuclear and organellar DNA (DAPI, blue), immunolocalization of epitope-tagged AGO3, detected with an antibody conjugated to Alexa Fluor 488 (AGO3, green), the merged images (Merged), and phase contrast images of the cells (Phase). Representative images are shown with the location of the nucleus indicated by n. AGO3 displayed mainly cytoplasmic localization with a punctuated appearance. Maa7-IR44s-(FLAG-AGO3), transgenic strain expressing a FLAG-tagged AGO3 under the control of the *PsaD* promoter. Maa7-IR44s, parental strain (used as a negative control).

Figure S3 FLAG-tagged AGO3 associates with the *MAA7* transcript, which is translationally repressed by small RNAs in the *Chlamydomonas* strain Maa7-IR44s (Ma *et al.* 2013). (A) Immunoprecipitations were performed with an anti-FLAG antibody on the indicated strains. Total RNA associated with the immunoprecipitated proteins was purified and used as template in RT-PCR reactions to detect the *MAA7* and *ACT1* transcripts (right panels). Total RNA isolated from the cell lysates (prior to immunoprecipitation) was used as template in RT-PCR reactions to determine input transcript levels (left panels). The *MAA7* transcript co-immunoprecipitated with FLAG-tagged AGO3, at levels well above those detected as background in immunoprecipitations with a FLAG-tagged BLE protein. This suggested that the introduced FLAG-tagged AGO3 is functional and associates with a small RNA-repressed mRNA. In contrast, the *ACTIN1* (*ACT1*) transcript, not targeted for regulation by small RNAs, was only detected at background levels in the immunoprecipitates. Maa7-IR44s-(FLAG-AGO3), transgenic strain expressing a FLAG-tagged AGO3 under the control of the *PsaD* promoter. Maa7-IR44s-(FLAG-BLE), transgenic strain expressing a FLAG-tagged Ble protein (conferring resistance to bleomycin). (B) Immunoblot analysis of the immunoprecipitated FLAG-tagged proteins, detected with an anti-AGO3 antibody (Ibrahim *et al.* 2010), in the indicated strains. Similar loading of the lanes was verified by Coomassie-blue staining of an equivalent gel (not shown).

Cluster: 1130+ chromosome_1 Start: 5106342 End: 5106482

1
3
2
1
5
103
131
12
134
631
11
1
1
1
2
1
1

--A T ACA
TTCCATTTCGAAAG TGCACAGCGCCCTGCCCCCTCCCTTGTCCCTGTTG CCCCTGGCC G
AAGGTAAGCTTTTC ACGTGTTCGCGGGGGACGGGGGAGGGGAACAGGGACAAC GGGGACCGG /
CCCA C G GGG

10
20
1
1
2
2
1

Cluster: 2675+ chromosome_10 Start: 3399712 End: 3400011

4
2
1
519
1
2
2
17
2
1
1
1
11
1
189
19
44
79
2
2
8
1
11

---T C CT
GTGCCGATAAGAAGGAGCCGTAAGGTACCGGGCG GGGGAGGGCAGGGGCAGGGACGGGGATCAG G
TACCGGTATTCTTCTCTGGCATTCATGGCCCGC CCCCTCCCGTCCCGTCCCTGCCCCCTAGTC C
GCACC A CC

5
45
146
3
26
1
5
1
12
321
12
9
1489
11
1
453
37

-----ACGGCTATTCTTCTCGGCAT-----307
-----CGGCTATTCTTCTCGGCAT-----13052
-----TACGGCTATTCTTCTCGGCAT-----93
-----GGCTATTCTTCTCGGCATT-----27662
-----TACGGCTATTCTTCTCGGCATT-----4745
-----CTATTCTTCTCGGCATT-----111
-----GCTATTCTTCTCGGCATT-----3877
-----CTACGGCTATTCTTCTCGGCATT-----4
-----ACGGCTATTCTTCTCGGCATT-----5244
-----TATTCTTCTCGGCATT-----4
-----CGGCTATTCTTCTCGGCATT-----647884
-----GGCTATTCTTCTCGGCATT-----22
-----ACGGCTATTCTTCTCGGCATT-----2
-----CGGCTATTCTTCTCGGCATT-----408
-----GCTATTCTTCTCGGCATT-----3
-----TACGGCTATTCTTCTCGGCATT-----3
-----ATTCTTCTCGGCATTCC-----1
-----GCTATTCTTCTCGGCATTCC-----1
-----ATTCTTCTCGGCATTCCA-----1
-----TATTCTTCTCGGCATTCCA-----2
-----ATTCTTCTCGGCATTCCAT-----101
-----CTATTCTTCTCGGCATTCCAT-----2
-----TATTCTTCTCGGCATTCCAT-----62
-----CCATGGCCCGACCCCTCCCGT-----3
-----CCCGTCCCCGTCCCTGCCCT-----2
-----TCCCGTCCCCGTCCCTGCCCT-----1

Cluster: 2675- chromosome_10 Start: 3399743 End: 3400006
-----CGATAAGAAGGAGCCGTAAGG-----11
-----GATAAGAAGGAGCCGTAAGGT-----1
-----ATAAGAAGGAGCCGTAAGGT-----8
-----ATAAGAAGGAGCCGTAAGGTA-----2
-----TAAGAAGGAGCCGTAAGGTAC-----11
-----TAAGAAGGAGCCGTAAGG-----79
-----TAAGAAGGAGCCGTAAGGTACC-----1
-----TAAGAAGGAGCCGTAAGGT-----189
-----TAAGAAGGAGCCGTAAG-----44
-----TAAGAAGGAGCCGTAAGGTA-----19
-----AAGAAGGAGCCGTAAGGT-----1
-----AAGAAGGAGCCGTAAGGTA-----1
-----CCGTAAGGTACCGGGCGTGGGGA-----1
-----TAAGGTACCGGGCGTGGGGA-----8
-----TAAGGTACCGGGCGTGGGGAG-----312
-----TAAGGTACCGGGCGTGGGG-----33
-----TAAGGTACCGGGCGTGGGGAGG-----1806
-----TAAGGTACCGGGCGTGGG-----3
-----TAAGGTACCGGGCGTGG-----3
-----TAAGGTACCGGGCGTGGGGAGGGC-----1
-----TAAGGTACCGGGCGTGGGGAGGG-----2
-----TAAGGTACCGGGCGTGGGGAGGGCA-----1
-----AAGGTACCGGGCGTGGGGAG-----1
-----AAGGTACCGGGCGTGGGGAGG-----11
-----AGGTACCGGGCGTGGGGAGGGCAG-----1
-----AGGTACCGGGCGTGGGGAG-----1
-----AGGTACCGGGCGTGGGGAGG-----6
-----GTACCGGGCGTGGGGAGGGCAGG-----5
-----TACCGGGCGTGGGGAGGGCAGG-----896822
-----TACCGGGCGTGGGGA-----1002
-----TACCGGGCGTGGGGAGGGCAGGGG-----21
-----TACCGGGCGTGGGGAG-----2215
-----TACCGGGCGTGGGGAGGGCAGGG-----1950
-----TACCGGGCGTGGGGAGGGCAGGGGC-----2
-----TACCGGGCGTGGGGAGGGC-----669
-----TACCGGGCGTGGGGAGG-----2049
-----TACCGGGCGTGGGGAGGGCA-----1127
-----TACCGGGCGTGGGGAGGGCAG-----24150
-----TACCGGGCGTGGGGAGGG-----486
-----ACCGGGCGTGGGGAGGGCAGGG-----24
-----ACCGGGCGTGGGGAGGGCAG-----208
-----ACCGGGCGTGGGGAGGGCA-----9
-----ACCGGGCGTGGGGAGGGCAGG-----7843
-----ACCGGGCGTGGGGAGGGC-----4
-----CCGGGCGTGGGGAGGGCAG-----35
-----CCGGGCGTGGGGAGGGC-----1
-----CCGGGCGTGGGGAGGGCA-----1
-----CCGGGCGTGGGGAGGGCAGGG-----4
-----CCGGGCGTGGGGAGGGCAGG-----1191

-----CGGGCGTGGGGAGGGCAGG----- 17
-----GGGCGTGGGGAGGGCAGG----- 1

--A
TGCCGATAAGAAGGAGCCGTAAAGGTACCCGGGCGTGGGGAGGGCAGGGCAGGGACGGGGATCAG GG
ACGGGTATTCTTCTCTCGGCATTCCATGGCCCGCGCCCTCCCGTCCCCGTCCCTGCCCTAGTC G
GGAC GG

-----CCCGTCCCCGTCCCTGCCCT----- 2
-----TCCCGTCCCCGTCCCTGCCCT----- 1
-----GCCCTCCCGTCCCCGTCCCT----- 6
-----GGCCCCGCGCCCTCCCGT----- 3
-----TTCCATGGCCCCGCGCCCTCCCGT----- 6
-----ATGGCCCCGCGCCCTCCCGT----- 9
-----CCATGGCCCCGCGCCCTCCCGT----- 874
-----CATGGCCCCGCGCCCTCCCGT----- 364
-----TGGCCCCGCGCCCTCCCGT----- 2
-----TCCATGGCCCCGCGCCCTCCCGT----- 48
-----CCATGGCCCCGCGCCCTCCCG----- 7
-----CATGGCCCCGCGCCCTCCCG----- 3
-----TCCATGGCCCCGCGCCCTCCCG----- 1
-----CCATGGCCCCGCGCCCTCC----- 3
-----CATTCATGGCCCCGCGCCCT----- 1
-----TATTCTTCTCTCGGCATTCCAT----- 62
-----CTATTCTTCTCTCGGCATTCCAT----- 2
-----ATTCTTCTCTCGGCATTCCAT----- 101
-----TATTCTTCTCTCGGCATTCCA----- 2
-----ATTCTTCTCTCGGCATTCCA----- 1
-----ATTCTTCTCTCGGCATTCC----- 1
-----GCTATTCTTCTCTCGGCATTCC----- 1
-----CGGCTATTCTTCTCTCGGCATT----- 408
-----GCTATTCTTCTCTCGGCATT----- 3
-----ACGGCTATTCTTCTCTCGGCATT----- 2
-----GGCTATTCTTCTCTCGGCATT----- 22
-----CTATTCTTCTCTCGGCATT----- 111
-----ACGGCTATTCTTCTCTCGGCATT----- 5244
-----ACACGGCTATTCTTCTCTCGGCATT----- 2
-----GCTATTCTTCTCTCGGCATT----- 3877
-----CGGCTATTCTTCTCTCGGCATT----- 647884
-----CACGGCTATTCTTCTCTCGGCATT----- 8
-----GACACGGCTATTCTTCTCTCGGCATT----- 12
-----TATTCTTCTCTCGGCATT----- 4
-----GGCTATTCTTCTCTCGGCATT----- 27662
-----ACGGCTATTCTTCTCTCGGCAT----- 307
-----GCTATTCTTCTCTCGGCAT----- 37
-----GGCTATTCTTCTCTCGGCAT----- 453
-----CTATTCTTCTCTCGGCAT----- 1
-----CGGCTATTCTTCTCTCGGCAT----- 13052
-----GGCTATTCTTCTCTCGGCA----- 321
-----ACGGCTATTCTTCTCTCGGCA----- 11
-----CGGCTATTCTTCTCTCGGCA----- 1489
-----GCTATTCTTCTCTCGGCA----- 9
-----GGCTATTCTTCTCTCGGC----- 12
-----ACGGCTATTCTTCTCTCGGC----- 1
-----CGGCTATTCTTCTCTCGGC----- 5
-----GGCTATTCTTCTCTCGG----- 26
-----CGGCTATTCTTCTCTCGG----- 3
-----ACGGCTATTCTTCTCTCGG----- 1
-----GGCTATTCTTCTCTCG----- 146
-----GGCTATTCTTCTCTC----- 45
-----CGGCTATTCTTCTCTC----- 5

Cluster: 3318- chromosome_10 Start: 6282475 End: 6282731
-----TC-CGGTCATCAATTGCTCTTTC----- 1
-----TC-CGGTCATCAATTGCTCTTTCTGAT----- 1
-----CGGTCATCAATTGCTCTTTCTG----- 6
-----GGTCATCAATTGCTCTTTCTG----- 1
-----TCATCAATTGCTCTTTCTGATG----- 17
-----ATCAATTGCTCTTTCTGATG----- 5
-----TCAATTGCTCTTTCTGATGA----- 325
-----TCAATTGCTCTTTCTGAT----- 3
-----TCAATTGCTCTTTCTGA----- 3
-----TCAATTGCTCTTTCTGATGAG----- 29
-----TCAATTGCTCTTTCTGATG----- 3165
-----CAATTGCTCTTTCTGATG----- 72
-----CAATTGCTCTTTCTGATGA----- 13
-----AATTGCTCTTTCTGATG----- 24

```
-----AATTGCTCTTTCTGATGA----- 5
CA G - T CCG
CG CTC CGGTCATCAATTGCTCT TCTGATGAGCGATGCCGTGGCCT C
GC GAG GCCAGTAGTTAACGAGA AGACTACTCGCTACGGCACCGGA C
-- G T T CCC

-----AGTTAACGAGATAGACTACT----- 3
-----CCAGTAGTTAACGAGATAGACT----- 17
-----GCCAGTAGTTAACGAGATAGACT----- 1
-----GTAGTTAACGAGATAGACT----- 300
-----TAGTTAACGAGATAGACT----- 1
-----AGTAGTTAACGAGATAGACT----- 85
-----CAGTAGTTAACGAGATAGACT----- 214
-----GTAGTTAACGAGATAGAC----- 1
-----CAGTAGTTAACGAGATAGA----- 1
-----CCAGTAGTTAACGAGATAGA----- 1
-----GTAGTTAACGAGATAGA----- 2
-----AGTGCCAGTAGTTAACGAGATAG----- 2
-----GTGCCAGTAGTTAACGAGATAG----- 5
-----GTGCCAGTAGTTAACGAGAT----- 7
-----AGTGCCAGTAGTTAACGAGAT----- 1

Cluster: 3704- chromosome 11 Start: 1442246 End: 1442394
-----CTGTTTACCAACGTGCAGGGGG----- 1
-----CTGTTTACCAACGTGCAGGGG----- 2
-----TGTTTACCAACGTGCAGGGGG----- 103
-----TGTTTACCAACGTGCAGGG----- 1
-----TGTTTACCAACGTGCAGGGG----- 50
-----TTTACCAACGTGCAGGGGG----- 2
-----TTTACCAACGTGCAGGGGGACT----- 4
-----TTTACCAACGTGCAGG----- 1
-----TTTACCAACGTGCAGGGGGGAC----- 25
-----TTTACCAACGTGCAGGGGGGA----- 11
-----TTACCAACGTGCAGGGGGACTTAG----- 1
-----TTACCAACGTGCAGGGGGGA----- 2
-----TTACCAACGTGCAGGGGGGACT----- 6
-----TACCAACGTGCAGGGGGGACT----- 3
-----CCAACGTGCAGGGGGGACTTAG----- 1
-----AACGTGCAGGGGGGACTTAGTGG----- 6
-----TGCAGGGGGGACTTAGTGGAG----- 1623
-----TGCAGGGGGGACTTAGTGGAGATCC----- 1
-----TGCAGGGGGGACTTAGTGGAGA----- 4691
-----TGCAGGGGGGACTTAGTGG----- 18
-----TGCAGGGGGGACTTAG----- 1
-----TGCAGGGGGGACTTAGTGGGA----- 165
-----TGCAGGGGGGACTTAGTGGAGAT----- 21
-----TGCAGGGGGGACTTAGTGGAGATC----- 2
-----GCAGGGGGGACTTAGTGGAG----- 12
-----GCAGGGGGGACTTAGTGGAGAT----- 1
-----GCAGGGGGGACTTAGTGGAGA----- 29
-----CAGGGGGGACTTAGTGGAGATC----- 17
-----CAGGGGGGACTTAGTGGAGA----- 1
-----AGGGGGGACTTAGTGGAGA----- 1
-----AGGGGGGACTTAGTGGAG----- 1
-----GGGGACTTAGTGGAGA----- 2
-----TAGTGGAGATCCTCCTGTCCGG----- 2
-----TAGTGGAGATCCTCCTGTCCG----- 19
-----TAGTGGAGATCCTCCTGTCCGGC----- 1
-----TGGAGATCCTCCTGTCCGGC----- 46
-----TGGAGATCCTCCTGTCCGGCT--ACCTGG----- 1
-----TGGAGATCCTCCTGTCCGGCT----- 29
-----TGGAGATCCTCCTGTCCGG----- 1
-----TCCTCCTGTCCGGCT--ACCTGG----- 1
-----TGTCGGCT--ACCTGGAGCGCAAG----- 1

-----TTA GGA G AA
GCTGTTTACCAACGTGCAGGGGGGACTTAGTGGAGATCCTCCTGTCCGGCT ACCT GC C C
TGACAAATGGTTGCACGTCCCCCTGAATCACCTCTAGGAGGACAGGCCGA TGGA CG G /
CGAACGCGAGGTCCC TTT GAG G TT

-----GGACAGGCCGATTTTGGAGAGCG----- 3
-----GGACAGGCCGATTTTGGAGAG----- 2
-----GGACAGGCCGATTTTGGAG----- 1
-----CTAGGAGGACAGGCCGATTTT----- 1
-----AGGAGGACAGGCCGATTT----- 1
-----TCACCTCTAGGAGGACAGGCCGATT----- 1
```

```
-----CCTCTAGGAGGACAGGCCGAT-----2
-----TCACCTCTAGGAGGACAGGCCGAT-----3
-----CGTCCCCCTGAATCACCTCT-----11
-----ACGTCCCCCTGAATCACCTCT-----115
-----CACGTCCCCCTGAATCACCT-----1
-----GCACGTCCCCCTGAATCACCT-----57
-----TGCACGTCCCCCTGAATCACCT-----1
-----GTCCCCCTGAATCACCT-----1
-----GCACGTCCCCCTGAATCAC-----1
-----TGGTTGCACGTCCCCCTGAAT-----8
-----GGTTGCACGTCCCCCTGAAT-----16
-----GACAAATGGTTGCACGTCCCCCT-----6
-----TGACAAATGGTTGCACGTCCCCCT-----1
-----CCCTGACAAATGGTTGCACGTC-----8
-----TGACAAATGGTTGCACGT-----5
-----CCTGACAAATGGTTGCACGT-----151
-----TCCCTGACAAATGGTTGCACGT-----20
-----GACAAATGGTTGCACGT-----1
-----CTGACAAATGGTTGCACGT-----9
-----CGAGGTCCTGACAAATGGTTGCACGT-----1
-----GTCCCTGACAAATGGTTGCACGT-----56
-----CCCTGACAAATGGTTGCACGT-----20465
-----GAGGTCCTGACAAATGGTTGCACGT-----3
-----ACAAATGGTTGCACGT-----2
-----GGTCCCTGACAAATGGTTGCACGT-----52
-----GTCCCTGACAAATGGTTGCACG-----1
-----CCCTGACAAATGGTTGCACG-----68
-----CCCTGACAAATGGTTGCAC-----1
-----CCCTGACAAATGGTTGCA-----4
-----GAGGTCCTGACAAATGGTTGC-----8
-----CGCGAGGTCCTGACAAATGGTT-----1
-----CGAGGTCCTGACAAATGGTT-----87
-----GAGGTCCTGACAAATGGTT-----32
-----GCGAGGTCCTGACAAATGGTT-----9
-----CGAGGTCCTGACAAATGGT-----1
-----GCGAGGTCCTGACAAAT-----1
```

```
Cluster: 3708-      chromosome_11      Start: 1448180      End: 1448276
-----TTAGTGCCAGCGCGCTCTCGG-----53
-----TTAGTGCCAGCGCGCTCTCGGC-----4
-----TAGTGCCAGCGCGCTCTCGGCC-----33
-----TAGTGCCAGCGCGCTCTCGGC-----3004
-----TAGTGCCAGCGCGCTCTCGG-----679
-----AGTGCCAGCGCGCTCTCGGC-----114
-----AGTGCCAGCGCGCTCTCGGCC-----1
-----AGTGCCAGCGCGCTCTCGG-----4
-----GTGCCAGCGCGCTCTCGGCC-----1
-----GTGCCAGCGCGCTCTCGGC-----12
-----GTGCCAGCGCGCTCTCGG-----1
-----TGCCAGCGCGCTCTCGGCC-----35
-----TGCCAGCGCGCTCTCGGCCA-----6
-----TGCCAGCGCGCTCTCGGCCAAG-----63
-----GCCAGCGCGCTCTCGGCCAAG-----1
-----GCCAGCGCGCTCTCGGCC-----1
-----CTCTCGGCCAAGTCTGGCAGA-----4
-----TCTCGGCCAAGTCTGGCAGAAGC-----2
-----TCTCGGCCAAGTCTGGCAGAAGCGGA-----3
-----TCTCGGCCAAGTCTGGCAGA-----4
-----TCTCGGCCAAGTCTGGCAGAAGCGGAG-----19
-----TCTCGGCCAAGTCTGGCAGAAGCGG-----2
-----CTCGGCCAAGTCTGGCAG-----1
-----TCGGCCAAGTCTGGCAGAAGCGGAG-----33
-----TCGGCCAAGTCTGGCAGAAG-----43
-----TCGGCCAAGTCTGGCAGAAGCG-----11
-----TCGGCCAAGTCTGGCAGAAGCGGA-----8
-----TCGGCCAAGTCTGGCAGAAGC-----96
-----TCGGCCAAGTCTGGCAGAAGCGG-----5
--A      C      AA      AG
TAATTAGTGCCAGCGG TCTCGGCCAAGTCTGGCAG GCGG A
GTTAATCACGGTCGCGC AGAGCCGTTTACAGCCGTC CGCC /
CGGG      A      GC      GG
-----GGTTCAGACCGTCGCCGCCG-----1
-----GCAAGAGCCGTTTACAGCCGT-----5
-----CGCGCAAGAGCCGTTTACAGAC-----3
```

-----TCGCGCAAGAGCCGGTTCAGA-----	1
-----CGCGCAAGAGCCGGTTCAGA-----	12
-----CGGTCGCGCAAGAGCCGGT-----	3
-----TCACGGTCGCGCAAGAGCCGGT-----	42
-----ACGGTCGCGCAAGAGCCGGT-----	2
-----TAATCACGGTCGCGCAAGAGCCGGT-----	1
-----CACGGTCGCGCAAGAGCCGGT-----	281
-----CGGTCGCGCAAGAGCCGGT-----	4
-----GGTCGCGCAAGAGCCGGT-----	3
-----CGGTCGCGCAAGAGCCGG-----	1
-----CACGGTCGCGCAAGAGCCGG-----	4
-----TCACGGTCGCGCAAGAGCC-----	1
-----TAATCACGGTCGCGCAAGAGC-----	9
-----TCACGGTCGCGCAAGAG-----	1
-----GTTAATCACGGTCGCGCAAGA-----	1

Cluster: 3772+ chromosome_11 Start: 1713197 End: 1713520

-----TTTCCAGGTGCTCCAGAAC-GG-----	1
-----TTTTCCAGGTGCTCCAGAAC-GGC-----	1
-----TCGATTTTCCAGGTGCTCCAG-----	1
-----TCGATTTTCCAGGTGCTCCAGA-----	2
-----TCGATTTTCCAGGTGCTCC-----	2
-----TATTGCAGAGCCAGCGGCCTCG-----	1
-----TGTATTGCAGAGCCAGCGG-----	1
-----ACTACACTGTTATTGCAGAGC-----	1
-----AGGACTACACTGTTATTGCAG-----	1
-----AAGGACTACACTGTTATTGCA-----	2
-----GGGAACAAGGACTACACTG-----	3
-----CGGGAACAAGGACTACACTGT-----	1
-----CGGGAACAAGGACTACACTG-----	3
-----TCGGGAACAAGGACTACA-----	1
-----TCGGGAACAAGGACTACACTGT-----	206
-----TCGGGAACAAGGACTACACTG-----	649
-----TCGGGAACAAGGACTACACT-----	13
-----TCGGGAACAAGGACTACAC-----	1
-----TCACGCTTGGTCGGGAACAAGG-----	14
-----TCACGCTTGGTCGGGAACAAG-----	2
-----AGCGGCTCACGCTTGGTCGGG-----	1
-----GGACTGGAGCGGCTCACGCT-----	3
-----TGGACTGGAGCGGCTCACG-----	2
-----TGGACTGGAGCGGCTCACGC-----	5
-----TGGACTGGAGCGGCTCAC-----	1
-----TGGACTGGAGCGGCTCACGCT-----	183
-----CACACGGTGGACTGGAGCGG-----	1
-----CGGCACACCGGTGGACTGGAG-----	1
-----TTCGGCAGGACGGCACACCGG-----	2
-----TACGGAGGTTTCGGCAGGACGGC-----	2
-----TACGGAGGTTTCGGCAGGACG-----	1
-----TACGGAGGTTTCGGCAGGACGG-----	3
-----TTACGGAGGTTTCGGCAGGACG-----	5
-----TAATTACGGAGGTTTCGGCAGG-----	28
-----TGGTAATTACGGAGGTTTCGG-----	5
-----TGGTAATTACGGAGGTTTCGGC-----	38
-----ATGGTAATTACGGAGGTTTCGG-----	3
-----TCACGATGGTAATTACGGAGG-----	3

-----	T		C		AC-	AAAG	AG
GGCATCACGATGGTAAT	ACGGAGGTTTCGGCAGGACGGCACACCGGTGGACTGGAGCGGCTCACGCTTGGTCGGGAACAAGGA	TACACTGTTATTGCAGAGCCAGCGGCCTCGATTTTCCAGGTGCTCCAGA	GGCA	TGA	C		
CCGTAGTGCTACCATTA	TGCCTCCAAGCCGTCTCTGCCGTGTGGCCACCTGACCTCGCCGAGTGCGAACCAACCCCTTGTTCCT	ATGTGACAATAACGTCTCGGTGCGCGGAGCTAAAAGGTCACGAGGTCT	CTGT	GCT	A		
CCACCGAAGTCGAACGTTGTGCGGT	T	A	GAA	CAA-	AA		

-----GAAGTCGAACGTTGTGCGGT-----	1
-----CGAAGTCGAACGTTGTGCGGT-----	1
-----GAACGTTGTGCGGTCCGTAGT-----	1
-----CGTTGTGCGGTCCGTAGTGCT-----	7
-----GCGTCCGTAGTGCTACCA-----	1
-----CGCGTCCGTAGTGCTACCA-----	1
-----GTCGCGTCCGTAGTGCTACCAT-----	1
-----TCGCGTCCGTAGTGCTACCAT-----	6
-----GCGTCCGTAGTGCTACCATT-----	2
-----TCGCGTCCGTAGTGCTACCATT-----	1
-----CGCGTCCGTAGTGCTACCATT-----	46
-----GCGTCCGTAGTGCTACCATTA-----	6
-----CGTCCGTAGTGCTACCATTA-----	1
-----CGCGTCCGTAGTGCTACCATTA-----	1
-----CGTCCGTAGTGCTACCATTA-----	89
-----GTCCGTAGTGCTACCATTA-----	1

8
1
6
3
3
1
1
1
48
2
31
2
1
1
3
2
10
267
20
1
4
7
1
1
90
2
6
6521
143
3
13
5
2

1
2
1
1
183
5
2
1
3
1
2
14
1
13
50
1
1
1
2
2

CT	A	TTC-	TT
A	TG	CAG	GA T
T	GC	GTT	CT T
T-	C	TTCA	TT

5
13
143
3
6521
6
90
2
1
1
1
1
10
267
20
3

-----CGAGTGCGAACCAGCCCTTGT-----2
-----GAGTGCGAACCAGCCCTTG-----1
-----GCCGAGTGCGAACCAGCCCTT-----2
-----CGCCGAGTGCGAACCAGCCCTT-----1
-----CGCCGAGTGCGAACCAGCCCT-----31
-----TGGCCACCTGACCTCGCCGAGT-----2
-----GGCCACCTGACCTCGCCGAGT-----48
-----TGCCGTGTGGCCACCTGACCT-----1
-----CGTCCTGCCGTGTGGCCACCT-----1
-----CCTCCAAGCCGTCCTGCCGTGT-----1
-----CTCCAAGCCGTCCTGCCGTGT-----3
-----GCCTCCAAGCCGTCCTGCCGT-----1
-----CCTCCAAGCCGTCCTGCCGT-----6
-----CGTCCGTAGTGCTACCATTA-----3
-----GCGTCCGTAGTGCTACCATTA-----1
-----CGCGTCCGTAGTGCTACCATT-----6
-----TCGCGTCCGTAGTGCTACCATT-----46
-----GCGTCCGTAGTGCTACCATT-----1
-----TCGCGTCCGTAGTGCTACCAT-----2
-----GTCGCGTCCGTAGTGCTACCAT-----6
-----CGCGTCCGTAGTGCTACCA-----1
-----GCGTCCGTAGTGCTACCA-----1

Cluster: 3911- chromosome 11 Start: 2529187 End: 2529546
-----TCTGGATGGTCGAAACCGCG-----23
-----TCTGGATGGTCGAAACCGC-----9
-----TCTGGATGGTCGAAACCGGT-----2
-----TCTGGATGGTCGAAACCGGTAC-----31
-----TGGATGGTCGAAACCGGTAC-----125
-----TGGATGGTCGAAACCGGTACA-----1595
-----TGGATGGTCGAAACCGGTACAGC-----3
-----TGGATGGTCGAAACCGGTACAG-----1
-----GGATGGTCGAAACCGGTAC-----1
-----GGATGGTCGAAACCGGTACA-----11
-----GATGGTCGAAACCGGTACA-----1
-----TACAGCACTCCCGCGTCGCTC-----14
-----ACAGCACTCCCGCGTCGCTC-----1
-----TCCCGCGTCGCTCTGCGCCTT-----1
-----TTCCTGGCGGGAACCTGGGCC-----1
-----TGGCGGGAACCTGGGCCTTGT-----1

-----G C GG GTCCCCGCCCTTCTGGGC GC
AAGGTC**TGGATGGTCGAAACC** **CGTACAGCACTCCCGCGTC** CTCTGCGCCTTC T CGGGAACCTGGGCCTTTGTCA**T** CC T
TTCC**AGACTTACCAGCTTTGG** **GCAT**GTCTGAGGGCGGCAG GAGACGCGGAAG A GCCCTTGGACCCGGAACGGTA GG A
CTTCGAC - A T AA GT-----AA

-----AGACGCGGAAGTAAAGCCCTT-----10
-----GCGGCAGAGAGACGCGGAAGT-----7
-----CCAGCTTTGG-GCATGTCGTGA-----2
-----ACCAGCTTTGG-GCATGTCGT-----1
-----TACCAGCTTTGG-GCATGTCGT-----3
-----ACCAGCTTTGG-GCATGTCG-----1
-----**AGACTTACCAGCTTTGG-GCAT**-----13
-----CGACTTCCAGACTTACCAGCT-----8

Cluster: 3965- chromosome 11 Start: 2856059 End: 2856444
-----TAATAAGCAGCAGGATGTAAGTAG-----3
-----TAAGCAGCAGGATGTAAGTAG-----236
-----TAAGCAGCAGGATGTAAGTA-----2
-----AAGCAGCAGGATGTAAGTAG-----1
-----CAGCAGGATGTAAGTAGCAGG-----1
-----AGCAGGATGTAAGTAGCAGGA-----7
-----CAGGATGTAAGTAGCAGGATG-----2
-----AGGATGTAAGTAGCAGGATG-----6
-----GGATGTAAGTAGCAGGATGCT-----1
-----TGTAAGTAGCAGGATGCTC-----1
-----TAAGTAGCAGGATGCTCGCGC-----3
-----TAGCAGGATGCTCGCGCTGTGC-----5
-----TAGCAGGATGCTCGCGCTGT-----2
-----TAGCAGGATGCTCGCGCTGTGCAT-----1
-----TAGCAGGATGCTCGCGCTGTGCA-----2
-----TAGCAGGATGCTCGCGCTGTG-----77
-----TGCTCGCGCTGTGCATCTAAG-----8
-----TGCTCGCGCTGTGCATCTAAGG-----334
-----TCGCGCTGTGCATCTAAGGACAC-----1
-----TCGCGCTGTGCATCTAAGGACACC-----3

-----TCGCGCTGTGCATCTAAGGACA-----	12
-----TCGCGCTGTGCATCTAAGGAC-----	4
-----CGCTGTGCATCTAAGGAC-----	1
-----CGCTGTGCATCTAAGGACACCTG-----	1
-----TGTGCATCTAAGGACACCTGTGT-----	27
-----TGTGCATCTAAGGACACCTGTGTT-----	2
-----TGTGCATCTAAGGACACCTGTG-----	1
-----TGTGCATCTAAGGACACCTGT-----	3
-----TGCATCTAAGGACACCTGTGTTATT-----	1
-----GCATCTAAGGACACCTGTGTT-----	1
-----TGTTATTAATGGAGATGGCGGGGTT-----	1
-----TGTTATTAATGGAGATGGCGG-----	7
-----TTATTAATGGAGATGGCGGG-----	1
-----TATTAATGGAGATGGCGGGGT-----	1
-----TATTAATGGAGATGGCGGGGTT-----	2

-----	CAAAGATTACCTCTCCCGCCC-----	1
-----	-T- CCTGTG- ACACAAAGATTACCT-	1
-----	-CCTGTG- ACACAAAGATTACCT-	2
-----	-CGATACGTAGGT- CCTGTG- ACA-	4
-----	-TACGAGCAGGATACGTAGGT- CCT-	4
-----	-CGAGCAGGATACGTAGGT- CCT-	25
-----	CTACGAGCAGGATACGTAGGT-	15
-----	-CGAGCAGGATACGTAGGT-	1
-----	-TACGAGCAGGATACGTAGGT-	1
-----	-ACGAGCAGGATACGTAGGT-	25
-----	-CGTCCTACGAGCAGGATACGT-	1
-----	-CGTCCTACATTATCATCGTCCT-	19
-----	-TCGTCCTACATTATCATCGTCCT-	10
-----	-CGTCGTCCTACATTATCATCGT-	13
-----	-TCGTCGTCCTACATTATCATCGT-	10
-----	-GTCGTCCTACATTATCATCGT-	1
-----	GGAT-TCGTCGTCCTACATTCA-	2
-----	GAT-TCGTCGTCCTACATTCA-	1
-----	-T-TCGTCGTCCTACATTCA-	1
-----	GGAT-TCGTCGTCCTACATTCA-	1
-----	-CCTGGAT-TCGTCGTCCTACATT-	3
-----	-CTGGAT-TCGTCGTCCTACATT-	1
-----	-CCTGGAT-TCGTCGTCCTACAT-	2

CTCTCT
ACGGCAGCGCCCTGCCTCTGTACTTACTGCTGGCCCTCTCGT CAGTCTCCATCATCGTGGCG A- CTCCA- T
TGCTGTGCGGGACGGAGACAATGAATGACGACCGGAGAGCA GTCAGAGGTAGGTAGCACGCG TCGT AG T
----- T CGCA TC A
CG CACAAA A
-----GCATGTCAGAGGTAGGTAGCA----- 4
-----CGGAGCATGTCAGAGGTAGGT----- 2
-----GAGAGCATGTCAGAGGTAGGT----- 5
-----GGAGAGCATGTCAGAGGT----- 1

-----	CGGAGAGCATGTCAGAGGT-----	128
-----	CCGGAGAGCATGTCAGAGGT-----	66
-----	GACCGGAGAGCATGTCAGAGGT-----	2
-----	ACCGGAGAGCATGTCAGAGGT-----	1237
-----	CCGGAGAGCATGTCAGAGG-----	1
-----	ACCGGAGAGCATGTCAGAGG-----	13
-----	GAATGACGACCGGAGAGCATGT-----	9
-----	GCGGGACGGAGACAATGAATG-----	1
-----	GTGCGGGACGGAGACAATGAAT-----	1
-----	TGCGGGACGGAGACAATGAAT-----	2
Cluster: 5094+ chromosome_12 Start: 3183047 End: 3183300		
-----	TAACGGACCTGGAGGACAGGCGGCG-----	1
-----	TAACGGACCTGGAGGACAGGC-----	1
-----	TAACGGACCTGGAGGACAGG-----	39
-----	AATAACGGACCTGGAGGACA-----	2
-----	AATAACGGACCTGGAGGAC-----	3
-----	CAATAACGGACCTGGAGGACA-----	4
-----	CAATAACGGACCTGGAGGAC-----	6
-----	TCAATAACGGACCTGGAGGA-----	7
-----	TCAATAACGGACCTGGAG-----	1
-----	TCAATAACGGACCTGGAGGACA-----	377
-----	TCAATAACGGACCTGG-----	1
-----	TCAATAACGGACCTGGA-----	2
-----	TCAATAACGGACCTGGAGGAC-----	1181
-----	TTCAATAACGGACCTGGAGG-----	8
-----	TTCAATAACGGACCTGGAGGA-----	36
-----	TTCAATAACGGACCTGGAGGAC-----	17
-----	TTCAATAACGGACCTGGAGGACA-----	2
-----	AACTTCAATAACGGACCTGG-----	1
-----	TCAACTTCAATAACGGACCTGG-----	26
-----	TCAACTTCAATAACGGACC-----	9
-----	TCAACTTCAATAACGGACCTG-----	3
-----	TGCATCAACTTCAATAACGGA-----	13
-----	TGCATCAACTTCAATAACGGACC-----	6
-----	TGCATCAACTTCAATAACGG-----	1
-----	TTGCATCAACTTCAATAACGG-----	1
-----	TTTGCATCAACTTCAATAACGG-----	3
-----	TTTTCATCAACTTCAATAACGG-----	1
-----	TCCAAGGCCGGAAGCGGTGCC-----	1

-----	GGCGTCCAAGGCCGGAAGCGGTGCCAGCATACGTGGTTTTCATCAACTTCAATAACGGACCTGGAGGACAGGCGGCG-----	ATGACGGAGATGCCATAACTCATCAGSTCCTCGTCCCCA CA CCACA- CT
-----	CCGCAGGTTCCGGCCTTCGCCACGGTCGTATGCACCAAAACGTAGTTGAAGTTATTGCCTGGACCTCCTGTCCGCGCG-----	GCTGC CT TC A
-----	TTCTCCCGGGCCGTGCG-----	CGACG GA AG /
-----	-----	GAA----- AC CCGCAC AA
-----	CCCGGGGCCGTGCGCGCAGGT-----	5
-----	CGGGGCCGTGCGCGCAGGT-----	1
-----	CCGGGGCCGTGCGCGCAGGTT-----	5
-----	CGGGGCCGTGCGCGCAGGTT-----	2
-----	GGTTCCGGCCTTCGCCACGG-----	1
-----	GGTTCCGGCCTTCGCCACGGT-----	48
-----	CGGCCTTCGCCACGGTCGTAT-----	1
-----	GGCCTTCGCCACGGTCGTAT-----	2
-----	CGGTCGTATGCACCAAAACGT-----	4
-----	TCGTATGCACCAAAACGTAG-----	1
-----	GGTCGTATGCACCAAAACGTAGT-----	1
-----	CGTATGCACCAAAACGTAGTT-----	1
-----	CCAAAACGTAGTTGAAGTTAT-----	1
-----	CAAAACGTAGTTGAAGTTATT-----	1
-----	ACGTAGTTGAAGTTATTGCCT-----	2
-----	CGTAGTTGAAGTTATTGCCT-----	5
-----	TGAAGTTATTGCCTGGACCT-----	2
-----	GTTGAAGTTATTGCCTGGACCT-----	7
-----	GAAGTTATTGCCTGGACCTC-----	4
-----	GAAGTTATTGCCTGGACCTCCT-----	137
-----	AGTTATTGCCTGGACCTCCT-----	1
-----	TGAAGTTATTGCCTGGACCTCCT-----	1
-----	AGTTATTGCCTGGACCTCCTGT-----	1
-----	GTTATTGCCTGGACCTCCTGT-----	2
Cluster: 5094- chromosome_12 Start: 3183020 End: 3183283		
-----	TCCAAGGCCGGAAGCGGTGCC-----	1
-----	TTTTCATCAACTTCAATAACGG-----	1
-----	TTTTCATCAACTTCAATAACGG-----	3
-----	TTTCATCAACTTCAATAACGG-----	1

[illegible]

	-AGTTATTGCCTGGACCTCCTGT-	1
	-GTTATTGCCTGGACCTCCTGT-	2
	-AGTTATTGCCTGGACCTCCT-	1
	-TGAAGTTATTGCCTGGACCTCCT-	1
	-GAAGTTATTGCCTGGACCTCCT-	137
	-GAAGTTATTGCCTGGACCTC-	4
	-TGAAGTTATTGCCTGGACCT-	2
	-GTTGAAGTTATTGCCTGGACCT-	7
	-ACGTAGTTGAAGTTATTGCCT-	2
	-CGTAGTTGAAGTTATTGCCT-	5
	-CAAAACGTAGTTGAAGTTATT-	1
	-CCAAAACGTAGTTGAAGTTAT-	1
	-CGTATGCCACAAAACGTAGTT-	1
	-GTCGTATGCACAAAACGTAGT-	1
	-TCGTATGCACAAAACGTAG-	1
	-CGTCGTATGCACAAAACGT-	4
	-CGGCCTTCGCCACGTCGTAT-	2
	-GCGCCTTCGCCACGTCGTAT-	1
	-GGTTCCGCGCTTCGCCACGCT-	48
	-GGTTCCGCGCTTCGCCACGG-	1
	-CCGGGGCCGTCGCCACGAGTT-	5
	-CCGGGGCCGTCGCCACGAGTT-	2
	-CCGGGGCCGTCGCCACGAGT-	5
	-CCGGGGCCGTCGCCACGAGT-	1
	-GGTGGGCTCTCCGGGGCCGT-	2

-----CTG G G
GTACACCTGGTCCCGCTATTGGAATCGCTGATCG CACCAT G
CATGTGGACCAAGGCCGATGGACTTCGCGACTAGT GTGTGG G
AGGACCGTCACAAGGCCGACGTC TC G G

-	-GGACCAGGGCGATGGACTT-	125
-	-GACCAGGGCGATGGACT-	11
-	-TGGACCAGGGCGATGGACT-	2
-	-GGACCAGGGCGATGGACT-	22
-	-ATGTGGACCAGGGCGATGGACT-	12
-	-TGTGGACCAGGGCGATGGACT-	2949
-	-GTGGACCAGGGCGATGGACT-	2580
-	-CCAGGGCGATGGACT-	1
-	-TGTGGACCAGGGCGATGGAC-	33
-	-GTGGACCAGGGCGATGGAC-	243
-	-GACCAGGGCGATGGAC-	1
-	-GTGGACCAGGGCGATGGA-	292
-	-TGTGGACCAGGGCGATGGA-	15
-	-CGTCATGTGGACCAGGGCGAT-	1
-	-CGTCACAAGGCCGACGTCTAT-	2
-	-TCACAAGGCCGACGTCTAT-	1
-	-GTCAACAAGGCCGACGTCTAT-	3
-	-CCGTCACAAGGCCGACGTCTAT-	1

[illegible][illegible]

```
-----TAGGAGTGAGAGTCAG-----15
-----TAGGAGTGAGAGTCAGAGGG-----1
-----TAGGAGTGAGAGTCAGA-----1
-----TAGGAGTGAGAGTCAGAGGGG-----23
-----GTAGGAGTGAGAGTCAG-----1
-----CCGAGTAGGAGTGAGAGTCAG-----324
-----CCGAGTAGGAGTGAGAGTCAGA-----1

-----A      A      G      ACG
AGCGCCGAGTAGGAGTGAGAGTCAG GGGGTAT TAAGCG CT      G
TCGCGGCTCATCCTCACTCTCAGTC CCCCATA ATTTCG      GA      /
GATTCAGTCGTTTCGTAGTCCATCT      A      C      A      AAA

-----AGTCGTTTCGTAGTCC-----1
-----AGTCGTTTCGTAGTCCATCT-----5
-----CAGTCGTTTCGTAGTCCATCT-----2
-----GTGTTTCGTAGTCCATCTT-----1
-----AGTCGTTTCGTAGTCCATCTT-----27
-----CAGTCGTTTCGTAGTCCATCTT-----1
-----CGTCAGTCCATCTTCGGGCT-----1
-----TCGGGGCTCATCCTCACT-----1
-----GCGGGCTCATCCTCACTCT-----4
-----CGGGCTCATCCTCACTCT-----25
-----TTCGGGGCTCATCCTCACTCT-----1
-----CGGGCTCATCCTCACTCTCA-----4
-----GCGGGCTCATCCTCACTCTCA-----1
-----CGGGCTCATCCTCACTCTCA-----112
-----GCGGGCTCATCCTCACTCTCAG-----16
-----CGGGCTCATCCTCACTCTCAG-----572
-----CGGGCTCATCCTCACTCTCAG-----27
-----TCGGGGCTCATCCTCACTCTCAGT-----7
-----CGGGCTCATCCTCACTCTCAGT-----503
-----GCTCATCCTCACTCTCAGT-----2
-----GCGGGCTCATCCTCACTCTCAGT-----275
-----GGCTCATCCTCACTCTCAGT-----176
-----CGGGCTCATCCTCACTCTCAGT-----37202
-----CGGGCTCATCCTCACTCTCAGTC-----54
-----CGGGCTCATCCTCACTCTCAGTC-----4

Cluster: 7085-      chromosome_13      Start: 2001058      End: 2001210
-----TAAGTCAGCAAGCAGTCAGG-----1
-----AGTCAGCAAGCAGTCAGGTAG-----1
-----AGTCAGCAAGCAGTCAGGTAGA-----4
-----TCAGCAAGCAGTCAGGTAGAAGCG-----1
-----TCAGCAAGCAGTCAGGTAGAAG-----41
-----TCAGCAAGCAGTCAGGTAGA-----342
-----TCAGCAAGCAGTCAGGTAG-----3
-----TCAGCAAGCAGTCAGGTAGAA-----80
-----AGCAAGCAGTCAGGTAGAAGCG-----1
-----AGCAAGCAGTCAGGTAGAA-----1
-----GCAAGCAGTCAGGTAGAAGCG-----5
-----CAAGCAGTCAGGTAGAAGCG-----4
-----AAGCAGTCAGGTAGAAGCGC-----1
-----AAGCAGTCAGGTAGAAGCG-----1
-----TAGAAGCGCCGAGTAGGAG-----1
-----TAGAAGCGCCGAGTAGGAGTGAG-----1
-----TAGAAGCGCCGAGTAGGAGTGA-----1
-----CCGAGTAGGAGTGAGAGTCAG-----324
-----GTAGGAGTGAGAGTCAG-----1
-----TAGGAGTGAGAGTCAG-----15
-----TAGGAGTGAGAGTCAGTGGGG-----116
-----TAGGAGTGAGAGTCAGTGGG-----7
-----TAGGAGTGAGAGTCAGTGG-----77
-----TAGGAGTGAGAGTCAGT-----3
-----TAGGAGTGAGAGTCAGTGGGGT-----2
-----TAGGAGTGAGAGTCAGTGGGGTAT-----1
-----TAGGAGTGAGAGTCAGTG-----3
-----AGGAGTGAGAGTCAGTGGGG-----4
-----TGAGAGTCAGTGGGGTATG-----2
-----TGGGGTATGTAAGCGTCTTG-----6
-----TGGGGTATGTAAGCGTCTTGC-----272
-----TGGGGTATGTAAGCGTCT-----3
-----GGGGTATGTAAGCGTCTTGC-----3

--CGG T      A      T      T      TGC
C AAGTCAGCAAGCAGTCAGGT GAAGCGCCGAGTAGGAGTGAGAGTCAG GGGGTATGTAAGCG CT      C
G TTCAGTCGTTTCGTAGTCCA CTTTCGCGGCTCATCCTCACTCTCAGTC CCCCATATATTCG      GA      /
```

ATTGTA C	C	T	C TTT	
-----CTCCCCATATATTCGCCGATG-----				1
-----TCTCCCCATATATTCGCCGAT-----				2
-----CCTCACTCTCAGTCTCCCCAT-----				1
-----CGGCTCATCCTCACTCTCAGTCT-----				5
-----GGCTCATCCTCACTCTCAGTCT-----				17
-----GCTCATCCTCACTCTCAGTCT-----				12
-----CGGCTCATCCTCACTCTCAGTC-----				54
-----CGGGCTCATCCTCACTCTCAGTC-----				4
-----CGGGCTCATCCTCACTCTCAGT-----				503
-----GGCTCATCCTCACTCTCAGT-----				176
-----GCGGCTCATCCTCACTCTCAGT-----				275
-----CGGCTCATCCTCACTCTCAGT-----				37202
-----TCGCGGCTCATCCTCACTCTCAGT-----				7
-----GCTCATCCTCACTCTCAGT-----				2
-----GCGGCTCATCCTCACTCTCAG-----				16
-----CGCGGCTCATCCTCACTCTCAG-----				27
-----CGGCTCATCCTCACTCTCAG-----				572
-----CGGCTCATCCTCACTCTCA-----				112
-----CGCGGCTCATCCTCACTCTCA-----				4
-----GCGGCTCATCCTCACTCTCA-----				1
-----GCGGCTCATCCTCACTCT-----				4
-----TTCGCGGCTCATCCTCACTCT-----				1
-----CGCGGCTCATCCTCACTCT-----				25
-----TCGCGGCTCATCCTCACT-----				1
-----CCTTCGCGGCTCATCCTCACT-----				1
-----TCAGTCCACCTTCGCGGCTCAT-----				1
-----CAGTCCACCTTCGCGGCTCAT-----				9
-----AGTCCACCTTCGCGGCTCAT-----				5
-----CGTCAGTCCACCTTCGCGGCT-----				103
-----CGTCAGTCCACCTTCGCGGC-----				1
-----AGTCGTTCGTCAGTCCACCTTC-----				8
-----GTCGTTCGTCAGTCCACCTTC-----				60
-----CGTTCGTCAGTCCACCTTC-----				19
-----CAGTCGTTCGTCAGTCCACCTTC-----				7
-----TCGTTCGTCAGTCCACCTTC-----				7
-----GTCGTTCGTCAGTCCACCTTC-----				625
-----GTTTCGTCAGTCCACCTTC-----				1
-----AGTCGTTCGTCAGTCCACCTTC-----				12380
-----TCAGTCGTTCGTCAGTCCACCTTC-----				1
-----AGTCGTTCGTCAGTCCACCTTC-----				219
-----CAGTCGTTCGTCAGTCCACCTTC-----				8
-----CGTTCGTCAGTCCACCTTC-----				1
-----GTCGTTCGTCAGTCCACCTTC-----				43
-----TCAGTCGTTCGTCAGTCCACCTTC-----				2
-----AGTCGTTCGTCAGTCCACCTTC-----				2
-----CTTCAGTCGTTCGTCAGTCCAC-----				18
-----CAGTCGTTCGTCAGTCCAC-----				1
-----AGTCGTTCGTCAGTCCAC-----				21
-----GTCGTTCGTCAGTCCAC-----				1
-----AGTCGTTCGTCAGTCC-----				1
-----AGCTTCAGTCGTTCGTCAGT-----				6
-----TAGCTTCAGTCGTTCGTCAGT-----				1
Cluster: 8319+	chromosome_14	Start: 1950047	End: 1950359	
-----AGGCGCGGAGCTCATCAAGC-----				1
-----TAGGCGCGGAGCTCATCAAGC-----				32
-----TAGGCGCGGAGCTCATCAAG-----				3
-----TAGGCGCGGAGCTCATCAAGCG-----				1
-----TGCAGGGGCATCTCGATGTCC-----				1
-----TACTGCAGGGGCATCTC-----				3
-----TACTGCAGGGGCATCT-----				1
-----TCGCCGCGCAGGTACTGCAGG-----				18
-----TTCTCGCCGCGCAGGTACTGC-----				11
-----ACGTTTCTCGCCGCGCAGG-----				1
-----CACGTTTCTCGCCGCGCAGG-----				202
-----TCCACGTTTCTCGCCGCGCAGG-----				6989
-----TCCACGTTTCTCGCCGCGCAGGT-----				15
-----TCCACGTTTCTCGCCGCGCAGGTACT-----				1
-----TCCACGTTTCTCGCCGC-----				2
-----TCCACGTTTCTCGCCGCG-----				1
-----TCCACGTTTCTCGCCGCGCAG-----				20
-----CTCCACGTTTCTCGCCGCGCAGG-----				5
-----TTCTCCACGTTTCTCGCCGCGC-----				5
-----TTCTCCACGTTTCTCGCCGCGCA-----				3
-----TTCTCCGTCACCTTCTCCACG-----				7

																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																					</
--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	----

-----CGCCAGCTCGGGCTGCGACGAC-----
-----GCCAGCTCGGGCTGCGACGAC-----
-----CCAGCTCGGGCTGCGACGA-----
-----CGCCAGCTCGGGCTGCGACGA-----
-----ACGCCAGCTCGGGCTGCGACGA-----
-----GCCAGCTCGGGCTGCGACGA-----
-----CGCCAGCTCGGGCTGCGACG-----
-----GACCCGACGCCCTGGGCGT-----
-----ACGTCGGCCAGTCCACGTCAGCCATC-----
-----CCAGTCCACGTCAGCCAT-----
-----CGTCGGCCAGTCCACGTCAGCCAT-----
-----ACGTCGGCCAGTCCACGTCAGCC-----
-----TCGGCCAGTCCACGTCGGCCAGT-----
-----CAGGGGCAGGGACGTCGGCCAGT-----
-----AGGGGCAGGGACGTCGGCCAGT-----

13
1
1
625
2
1
9
1
3
2
1
1
2
14
1

Cluster: 9303+ chromosome 16 Start: 185086 End: 185184

-----TAAGCCGGGTTGGTAG----- 295
-----TAAGCCGGGTTGGTAGTG----- 2
-----TAAGCCGGGTTGGTA----- 3
-----TAAGCCGGGTTGGTAGA----- 10
-----TAAGCCGGGTTGGT----- 1
-----TAAGCCGGGTTGG----- 1
-----TCCACATAAGCCGGGTTGG----- 21
-----CAGCAAGCAACATCCACA----- 1
-----CAGCAAGCAACATCCACATA----- 1
-----TCCAGCAAGCAACATCCAC----- 4
-----TCCAGCAAGCAACATCCACA----- 103
-----TCCAGCAAGCAACATCCACATA----- 16
-----TCCAGCAAGCAACATCCACAT----- 30
-----ATCCAGCAAGCAACATCCACA----- 2
-----TATCCAGCAAGCAACATCCAC----- 6
-----TATCCAGCAAGCAACATCCACA----- 91
-----TATCCAGCAAGCAACATCCACAT----- 1

-----ATG
ACGGTATCCAGCAAGCAACATCCACATAAGCCGGGTTGGTAG T
TGCCATAGGTCGTCGTGTAGGTGTATTGGGCCCAACCATC G
GATA GGG

-----ATAGGTCGTTCGTTGT----- 1
-----GCCATAGGTCGTTCGTTGT----- 2
-----CCATAGGTCGTTCGTTGTAGG----- 241
-----ATAGGTCGTTCGTTGTAGG----- 628
-----CATAGGTCGTTCGTTGTAGG----- 2116
-----AGGTCGTTCGTTGTAGG----- 1
-----TAGGTCGTTCGTTGTAGG----- 10
-----GGTCGTTCGTTGTAGG----- 2
-----TAGGTCGTTCGTTGTAGGT----- 1324
-----CGTTCGTTGTAGGT----- 1
-----GTCGTTCGTTGTAGGT----- 8
-----ATAGGTCGTTCGTTGTAGGT----- 95326
-----TGCCATAGGTCGTTCGTTGTAGGT----- 10
-----CATAGGTCGTTCGTTGTAGGT----- 183618
-----CCATAGGTCGTTCGTTGTAGGT----- 31651
-----GCCATAGGTCGTTCGTTGTAGGT----- 49
-----GGTCGTTCGTTGTAGGT----- 131
-----TCGTTCTGTTGTAGGT----- 1
-----AGGTCGTTCTGTTGTAGGT----- 108
-----ATAGGTCGTTCTGTTGTAGGTG----- 36
-----CATAGGTCGTTCTGTTGTAGGTG----- 1
-----TAGGTCGTTCTGTTGTAGGTG----- 464
-----AGGTCGTTCTGTTGTAGGTG----- 74
-----GCCATAGGTCGTTCTGTTGTAGGTG----- 3
-----TAGGTCGTTCTGTTGTAGGTGT----- 75230
-----CATAGGTCGTTCTGTTGTAGGTGT----- 24
-----GTCGTTCTGTTGTAGGTGT----- 20
-----CCATAGGTCGTTCTGTTGTAGGTGT----- 10
-----TCGTTCTGTTGTAGGTGT----- 8
-----CGTTCGTTGTAGGTGT----- 5
-----AGGTCGTTCTGTTGTAGGTGT----- 8552
-----ATAGGTCGTTCTGTTGTAGGTGT----- 6226
-----GGTCGTTCTGTTGTAGGTGT----- 286
-----TAGGTCGTTCTGTTGTAGGTGTAT----- 2
-----AGGTCGTTCTGTTGTAGGTGTAT----- 8
-----GGTCGTTCTGTTGTAGGTGTAT----- 139
-----GTCGTTCTGTTGTAGGTGTATT----- 1

-----GTATTCTGGCCCCAACCATC----- 1

Cluster: 9303- chromosome 16 Start: 184972 End: 185187

-----TATACGGGTATCCAGCAAGCAACATC----- 1
-----TATACGGGTATCCAGCAAGC----- 1
-----ATACGGGTATCCAGCAA----- 1
-----TACGGGTATCCAGCAAGCAACATCC----- 1
-----TACGGGTATCCAGCAAGCAAC----- 1
-----TATCCAGCAAGCAACATCCACAT----- 1
-----TATCCAGCAAGCAACATCCACA----- 91
-----TATCCAGCAAGCAACATCCAC----- 6
-----ATCCAGCAAGCAACATCCACA----- 2
-----TCCAGCAAGCAACATCCACAT----- 30
-----TCCAGCAAGCAACATCCAC----- 4
-----TCCAGCAAGCAACATCCACATA----- 16
-----TCCAGCAAGCAACATCCACA----- 103
-----CAGCAAGCAACATCCACA----- 1
-----CAGCAAGCAACATCCACA----- 1
-----TCCACATAAGCCGGGGTTGG----- 21
-----TAAGCCGGGGTTGGTAGC----- 1406
-----TAAGCCGGGGTTGGTAG----- 295
-----TAAGCCGGGGTTGGT----- 1
-----TAAGCCGGGGTTGG----- 1
-----TAAGCCGGGGTTGGTAGCA----- 189
-----TAAGCCGGGGTTGGTA----- 3
-----TAAGCCGGGGTTGGTAGCAA----- 2087
-----AAGCCGGGGTTGGTAGCAA----- 5
-----AGCCGGGGTTGGTAGCAA----- 1

-GGGCTA CAA
TACGGTATCCAGCAAGCAACATCCACA TAAGCCGGGGTTGGTAG C
GTGCCATAGGTCTGTCTGTAGGTGTATTCTGGCCCCAACCATC T
TGGTGTA TTT

-----GTATTCTGGCCCCAACCATCT----- 4
-----GTATTCTGGCCCCAACCATC----- 1
-----GTCGTTTCGTTGTAGGTGTATT----- 1
-----AGGTTCGTTTCGTTGTAGGTGTAT----- 8
-----TAGGTTCGTTTCGTTGTAGGTGTAT----- 2
-----GGTCGTTTCGTTGTAGGTGTAT----- 139
-----CATAGGTTCGTTTCGTTGTAGGTGT----- 24
-----TAGGTTCGTTTCGTTGTAGGTGT----- 75230
-----GCCATAGGTTCGTTTCGTTGTAGGTGT----- 3
-----AGGTTCGTTTCGTTGTAGGTGT----- 8552
-----CCATAGGTTCGTTTCGTTGTAGGTGT----- 10
-----CGTTCGTTGTAGGTGT----- 5
-----ATAGGTTCGTTTCGTTGTAGGTGT----- 6226
-----TCGTTTCGTTGTAGGTGT----- 8
-----GGTCGTTTCGTTGTAGGTGT----- 286
-----TAGTGCCATAGGTTCGTTTCGTTGTAGGTGT----- 2
-----GTCGTTTCGTTGTAGGTGT----- 20
-----TAGGTTCGTTTCGTTGTAGGTG----- 464
-----CATAGGTTCGTTTCGTTGTAGGTG----- 1
-----ATAGGTTCGTTTCGTTGTAGGTG----- 36
-----AGGTTCGTTTCGTTGTAGGTG----- 74
-----TGCCATAGGTTCGTTTCGTTGTAGGT----- 10
-----ATAGGTTCGTTTCGTTGTAGGT----- 95326
-----CCATAGGTTCGTTTCGTTGTAGGT----- 31651
-----CATAGGTCTGTCTGTAGGT----- 183618
-----GGTCGTTTCGTTGTAGGT----- 131
-----CGTTCGTTGTAGGT----- 1
-----GTGTAGTGCCATAGGTTCGTTTCGTTGTAGGT----- 1
-----TCGTTTCGTTGTAGGT----- 1
-----TAGGTTCGTTTCGTTGTAGGT----- 1324
-----TAGTGCCATAGGTTCGTTTCGTTGTAGGT----- 2
-----GCCATAGGTTCGTTTCGTTGTAGGT----- 49
-----GTGCCATAGGTTCGTTTCGTTGTAGGT----- 6
-----GTCGTTTCGTTGTAGGT----- 8
-----AGTGCCATAGGTTCGTTTCGTTGTAGGT----- 14
-----AGGTTCGTTTCGTTGTAGGT----- 108
-----GTAGTGCCATAGGTTCGTTTCGTTGTAGGT----- 3
-----ATAGGTTCGTTTCGTTGTAGG----- 628
-----GGTCGTTTCGTTGTAGG----- 2
-----TAGGTTCGTTTCGTTGTAGG----- 10
-----AGGTTCGTTTCGTTGTAGG----- 1
-----CATAGGTTCGTTTCGTTGTAGG----- 2116
-----CCATAGGTTCGTTTCGTTGTAGG----- 241

```

-----GCCATAGGTCGTTGTTGT-----2
-----AGTGCCATAGGTCGTTGTTGT-----2
-----ATAGGTCGTTGTTGT-----1
-----TAGTGCCATAGGTCGTTGTT-----1
-----GTAGTGCCATAGGTCGTTGTT-----5
-----AGTGCCATAGGTCGTTGTT-----2
-----AGTGCCATAGGTCGTTGTT-----1

```

```

CGGCTTGGTACGACGCGTCCGT
GGCTTGGTACGACGCTCCGT
  TACGACGCGTCCGTGCAGGC
  TACGACGCGTCCGTGCAGGCA
    CGCGTCCGTGCAGGCAGCGGT
    GTCCGTGCAGGCAGCGGT
    CGGTCCGTGCAGGCAGCGGT
    CGTCCGTGCAGGCAGCGTAA
      CGTGCAGGCAGCGTAAGGCG
      GTGCAGGCAGCGTAAGGCGT
      CCGTGCAGGCAGCGTAAGGCGT
      CGTGCAGGCAGCGTAAGGCGT
        GCAGGCAGCGTAAGGCGTCG
        CAGCGTAAGGCGTCGAGCG
        GCAGCGTAAGGCGTCGAGCG
          GGTAAAGCGTCGAGCGT
          CGGTAAAGCGTCGAGCGT
            CAGCGTAAGGCGTCGAGCGT
            CAGCGTAAGGCGTCGAGCGT
              AGCGTAAGGCGTCGAGCGT
              GGCAGCGTAAGGCGTCGAGCGT
                GCGTAAGGCGTCGAGCGTGC
                AAGCGTCGAGCGTCGAGCT
                GCGTCGAGCGTCGAGCTGT
                  CGTGCAGCTGTTGCCAGGA
                  TCGAGCTGTTGCCAGGA CGA
                    GTTGCCAGGA CGACACGCGCA
                    TGCAGGA CGACACGCGCAT
                    TTGCCAGGA CGACACGCGCAT
                    GTTGCCAGGA CGACACGCGCAT
                    TGTGCCAGGA CGACACGCGCAT
                    GCTGTTGCCAGGA CGACACGCGCAT
                      GCCAGGA CGACACGCGCATG
                      GGA CGACACGCGCATGGT
                      CCAGGA CGACACGCGCATGGT
                      CAGGA CGACACGCGCATGGT
                        CGCGCATGGTGCACGATGCGT
                        GCAGATGCGTTCGCCAACCGGT
                        CAGCATGCGTTCGCCAACCGGT

```

-----GGTTCCGGCAGGCGGCGGAC-----1
-----TGGTTCCGGCAGGCGGCGGAC-----471
-----TGGTTCCGGCAGGCGGCGGA-----17
-----TGGTTCCGGCAGGCGGCGACG-----104
-----GGGAGCTGGTTCCGGCAGGCG-----1
-----TGGGAGCTGGTTCCGGCAGG-----21
-----TGGGAGCTGGTTCCGGCAGGCG-----14
-----TGGGAGCTGGTTCCGGCAGGCGG-----2
-----TGGGAGCTGGTTCCGGCAGGCG-----63
-----CGCTGGGAGCTGGTTCCGGC-----1
-----CGCTGGGAGCTGGTTCCGG-----2
-----TCCGCTGGGAGCTGGTTCCGG-----422
-----TCCGCTGGGAGCTGGTTCCG-----2
-----TCCGCTGGGAGCTGGTTCCGGC-----10
-----TGTCGCTGGGAGCTGGTTC-----1
-----TGTCGCTGGGAGCTGGTTCG-----2
-----GATGTCCGCTGGGAGCTGG-----10
-----TGATGTCCGCTGGGAGC-----1
-----TGATGTCCGCTGGGAGCTGG-----143
-----TGATGTCCGCTGGGAGCTGGT-----20
-----TTGATGTCCGCTGGGAGCTG-----7
-----TTGATGTCCGCTGGGA-----2
-----TTGATGTCCGCTGGGAGCTGGT-----1
-----TTGATGTCCGCTGGGAGCTGGT-----31
-----TTGATGTCCGCTGGGAGCTGG-----7159
-----TTGATGTCCGCTGGGAGC-----10
-----TTGATGTCCGCTGGGAGCT-----1
-----TTGATGTCCGCTGGGAGCTGGTTCG-----1
-----CTTGATGTCCGCTGGGAGCTGG-----5
-----GCCTTGATGTCCGCTGGGAGC-----1

A GC TC G G -- CCAT A C C
CGCC GCGCCTTGATGTCCGCTGGGAGCTGGT CGGCAGGCGGC GAC GT GC GGACGT GG TGTG C
GTGG TCGGAACTACAGGCGACCCTCGACCA GCCGTCCGCCG CTG CG CG CCTGCA CC GCAC G
- - - GA G G AA CAT- - T G
-----TGCGAACTACAGGCGACCCCT-----15
-----GCGGAACTACAGGCGACCCCT-----3
-----GCGACCCCTCGACCAGAGCCGT-----1
-----CGACCAGAGCCGTCCGCCGGCT-----3

Cluster: 9770- chromosome_16 Start: 2231045 End: 2231429
-----GCCTTGATGTCCGCTGGGAGC-----1
-----CTTGATGTCCGCTGGGAGCTGG-----5
-----TTGATGTCCGCTGGGA-----2
-----TTGATGTCCGCTGGGAGC-----10
-----TTGATGTCCGCTGGGAGCTGG-----7159
-----TTGATGTCCGCTGGGAGCTG-----7
-----TTGATGTCCGCTGGGAGCTGGT-----31
-----TTGATGTCCGCTGGGAGCT-----1
-----TGATGTCCGCTGGGAGCTGGT-----20
-----TGATGTCCGCTGGGAGC-----1
-----TGATGTCCGCTGGGAGCTGG-----143
-----GATGTCCGCTGGGAGCTGG-----10
-----CCA CT C C CT G - A - C
CGCCTTGATGTCCGCTGGGAGCTGGT CGGCAGGCGGC GAC G TGC TA GGACGTGG C GTG T
GCGGAACTACAGGCGACCCTCGACCA GCCGTCCGCCG CTG C ACG GT CCTGCATC G CAC G
GCCGTGCGTTGCGGTGGTGCGCCG AG C C -- G A C A G
-----GCGGAACTACAGGCGACCCCT-----3

Cluster: 9770- chromosome_16 Start: 2231045 End: 2231429
-----TGCTGCGCAGGCACGTCCGTC-----1
-----TGCTGCGCAGGCACGTCCG-----6
-----T-----CGACAACGGTCCT-GCTGTGC-----1
-----TGCTACGCAAGCGGTTGGCCATG-----2
AAC --- AACG - G G G TGC G C
CATGCTGCGCAGGCACGTCCGTCGCCATTCCGCAGCTCGCACGCT CGAC GTCCT GCTGTGCGC TACCACGT CTACGCAAGCGGTTG CCATGTA GC G
GTACGACGCGTCCGTGCAGGCGGCGGTGGGGCGTCCGGCGTGTGA GCTG CAGGA CGACACGCG ATGGTGCA GATGCGTTCGCTAGC GSTACAT CG G
GTA ACC A-- A G G G TA-- G
-----GCAGGATGCGTTCGCTAGC-GGT-----39
-----CAGGATGCGTTCGCTAGC-GGT-----4

	-GCGGATGGTGCAGGATGCGT-	8
	-CGCGGATGGTGCAGGATGCGT-	1908
	-GCGGATGGTGCAGGATGCG-	1
	-CGCGGATGGTGCAGGATGCG-	3
	-GACACGCGGATGGTGCAGGAT-	1
	-CGACACGCGGATGGTGCAGGAT-	1
	-CACGCGGATGGTGCAGGAT-	9
	-CGCGGATGGTGCAGGAT-	1
	-ACACGCGGATGGTGCAGGAT-	1
	CAGGAACGACACGCGGATGGT-	4
A-	CAGGAACGACACGCGGATGGT-	4
	GCGGTGGGGCGTCGGGCCGTG-	2
	-GCGGTGGGGCGTCGGGCCGT-	1
	-CGGTGGGGCGTCGGGCCGT-	2
	-GCAGGCGGCGGTGGGGCGTCGGGCCGT-	1
	-CGGCGGTGGGGCGTCGGGCCGT-	135
	-GGCGGTGGGGCGTCGGGCCGT-	15
	-GCGGCGGTGGGGCGTCGGGCCGT-	5
TACGACGCGTCCGTGCAGGC-		1

Cluster: 11846+	chromosome_17	Start: 3137537	End: 3137719	
			TTGAGTCTTACTTGCTT--GGCTTAC--	1
			-----TGCTTCAGTCTTACTTGCTT--G--	3
			-----TGCTTCAGTCTTACTTGCTT--GG--	2
			-----TGGTCGACGGCAATTGCTTCAGT	1
			-----TGGTCGACGGCAATTGCTTCAG--	1
			-----TGGTCGACGGCAATTGCTTCA--	1
			-----TGGTGGTTCGACGGCAAT--	1
			-----GTGGTGGTTCGACGGCAATTGC	1
			-----AGTGGTGGTTCGACGGCAATTGC	1
			-----GAGTGGTGGTTCGACGGCAAT--	3
			-----TGAGTGGTGGTTCGACGGCAA--	10
			-----TGAGTGGTGGTTCGACGGCAATTG--	1
			-----TGAGTGGTGGTTCGACGGCAAT	3013
			-----TGAGTGGTGGTTCGACGGCA--	4
			-----TGAGTGGTGGTTCGACGG	1
			-----TGAGTGGTGGTTCGACGGCAATT	60
			-----TTGAGTGGTGGTTCGACGGCAAT	12
			-----TTGAGTGGTGGTTCGACGGCAA--	4
			-----TTTGAGTGGTGGTTCGACGGCAAT	13
			-----TTTGAGTGGTGGTTCGACGGCAA--	8
			-----TGGTTTGAGTGGTGG--	1
			-----TGGTTTGAGTGGTGGTTCG--	16
			-----TGGTTTGAGTGGTGGTTCGACG--	4
			-----TGGTTTGAGTGGTGGTTCGACGGCA--	1
			-----TGGTTTGAGTGGTGGTTCGA	1
			-----GGTGCAACAGGCCAGTGGTT--	1
			-----TGGTGCAACAGGCCAGTGGT--	1
			-----TGGTGCAACAGGCCAGTGGTT--	25
TA--	GGTGCTCTTCTGGTGCAAC			1

- - -	TA - - -	T		G	A		- -	T	G
GCA	GGTGCTCT	CTGTGGTCAACAGGCCAGTGGT	TGA	TGGTGGTG	CGGCAAT	TGCTTCAGTCTTACTTGCTT	GGC	TAC	T
CGT	TCACGAGA	GACCACTGTGTCGGGTCA	AACT	GCCACAGC	GTCGTT	TAACGAAGTCAGAATGAACGAA	CCG	ATG	C
CGC	CCCAAA	C		A	C		CT	C	C

[illegible]

-----GCCGTCGTTAACGAAGTCAGAAT----- 9

```
Cluster: 12364+      chromosome_17      Start: 5152778      End: 5152924
-----TCTCCGAGCTAGCTTGACGC----- 1
-----CCCGCTTCTCCGAGCTAGC----- 2
-----ACCCGCTTCTCCGAGCTAG----- 1
-----TACCCGCTTCTCCGAGCT----- 1
-----TACCCGCTTCTCCGAGCTAGC----- 51
-----TACCCGCTTCTCCGAGCTAG----- 255
-----TACCCGCTTCTCCGAGCTA----- 3
-----CTACCCGCTTCTCCGAGCTAG----- 1
-----AGCTACCCGCTTCTCCGAGCT----- 1
-----AGCTACCCGCTTCTCCGAG----- 1
-----AGCTACCCGCTTCTCCGAGC----- 8
-----CAGCTACCCGCTTCTCCGAG----- 1
-----CAGCTACCCGCTTCTCCGAGC----- 10
-----TCAGTACCCGCTTCTCCG----- 6
-----TCAGTACCCGCTTCTCC----- 2
-----TCAGTACCCGCTTCTCCGAGC----- 2030
-----TCAGTACCCGCTTCTCCGAG----- 193
-----TCAGTACCCGCTTCTCCGAGCT----- 44
-----TCAGTACCCGCTTCTCCGA----- 2
-----CTCAGTACCCGCTTCTCCGAGCT----- 1
-----TACCTCAGTACCCGCTTCTCCG----- 6
-----TCTACCTCAGTACCCGCTTC----- 1
-----TCCTCTACCTCAGTACCCGC----- 1
-----TCTCCTCTACCTCAGC----- 1
-----CTTCTCCTCTACCTCAGC----- 1
-----TCTTCTCCTCTACCTCAGC----- 2
-----AATCTTCTCCTCTACCTCAG----- 3
-----TAATCTTCTCCTCTACCTCA----- 14
-----TAATCTTCTCCTCTACCTCAGC----- 55
-----TAATCTTCTCCTCTACCTCAG----- 72
-----TAATCTTCTCCTCTACCTC----- 6
-----TAATCTTCTCCTCTACC----- 1
-----TTAATCTTCTCCTCTACC----- 1
-----TTAATCTTCTCCTCTACCTC----- 52
-----TTAATCTTCTCCTCTACCTCA----- 363
-----TTAATCTTCTCCTCTACCTCAGC----- 101
-----TTAATCTTCTCCTCTACCTCAG----- 4
-----TTAATCTTCTCCTCTACCT----- 2
-----TTTAATCTTCTCCTCTACCTC----- 2316
-----TTTAATCTTCTCCTCTACC----- 81
-----TTTAATCTTCTCCTCTAC----- 3
-----TTTAATCTTCTCCTCTACCT----- 128
-----TTTAATCTTCTCCTCTACCTCA----- 83
-----TTTAATCTTCTCCTCTACCTCAGC----- 4
-----GTTTAATCTTCTCCTCTACCTC----- 1
-----GTTTAATCTTCTCCTCTACC----- 3
-----GTTTAATCTTCTCCTCTACCT----- 2
-----TGTTTAATCTTCTCCTCTACC----- 496
-----TGTTTAATCTTCTCCTC----- 1
-----TGTTTAATCTTCTCCTCTA----- 1
-----TGTTTAATCTTCTCCTCTACCTC----- 97
-----TGTTTAATCTTCTCCTCTAC----- 1
-----TGTTTAATCTTCTCCTCTACCT----- 97
-----TGTGTTTAATCTTCTCCTCTA----- 5
-----TGTGTTTAATCTTCTCCTCT----- 2
-----TTGTGTTTAATCTTCTCCTCT----- 140
-----TTGTGTTTAATCTTCTCCTC----- 6
-----TTGTGTTTAATCTTCTCCTCTA----- 14

-- G AT
GACATTGTGTTTAATCTTCTCCTCTACCTCAGCTACCCGCTTCTCCGAGCTAGCTTGACGCTCTCT AG G
CTGTAAACACAAATTAGAAGAGGAGATGGAGTCGATGGGCGAAGAGGCTCGATCGAACTGCGAGAGA TC G
CG G GG

-----ACACAAATTAGAAGAGGAG----- 1
-----AACACAAATTAGAAGAGGAGA----- 4
-----ACACAAATTAGAAGAGGAGA----- 13
-----GTAACACAAATTAGAAGAGGAGAT----- 1
-----ACACAAATTAGAAGAGGAGAT----- 781
-----AACACAAATTAGAAGAGGAGAT----- 129
-----TAACACAAATTAGAAGAGGAGAT----- 20
-----ACAAATTAGAAGAGGAGAT----- 17
-----CACAAATTAGAAGAGGAGAT----- 5
-----CAAATTAGAAGAGGAGAT----- 1
```

-----ACAAATTAGAAGAGGAGATGG-----	1
-----AATTAGAAGAGGAGATGGA-----	7
-----AAATTAGAAGAGGAGATGGA-----	2
-----AATTAGAAGAGGAGATGGAG-----	64
-----ATTAGAAGAGGAGATGGAG-----	8
-----AAATTAGAAGAGGAGATGGAG-----	5
-----CAAATTAGAAGAGGAGATGGAGT-----	5
-----ATTAGAAGAGGAGATGGAGT-----	2004
-----AGAAGAGGAGATGGAGT-----	4
-----GAAGAGGAGATGGAGT-----	1
-----TAGAAGAGGAGATGGAGT-----	2
-----AGAGGAGATGGAGT-----	1
-----AAATTAGAAGAGGAGATGGAGT-----	251
-----ACAAATTAGAAGAGGAGATGGAGT-----	2
-----TTAGAAGAGGAGATGGAGT-----	40
-----AATTAGAAGAGGAGATGGAGT-----	9982
-----AATTAGAAGAGGAGATGGAGTC-----	2
-----AGAAGAGGAGATGGAGTCGAT-----	3
-----GAAGAGGAGATGGAGTCGAT-----	11
-----GATGGAGTCGATGGGCGAAGA-----	1
-----GATGGAGTCGATGGGCGAAGAG-----	2
-----GGAGTCGATGGGCGAAGAGG-----	14
-----AGTCGATGGGCGAAGAGG-----	370
-----GAGTCGATGGGCGAAGAGG-----	6
-----GGAGTCGATGGGCGAAGAGGC-----	13
-----GAGTCGATGGGCGAAGAGGC-----	16
-----AGTCGATGGGCGAAGAGGC-----	294
-----GGAGTCGATGGGCGAAGAGGCT-----	5468
-----ATGGGCGAAGAGGCT-----	3
-----GATGGGCGAAGAGGCT-----	1
-----TGGAGTCGATGGGCGAAGAGGCT-----	11
-----TCGATGGGCGAAGAGGCT-----	2
-----GTCGATGGGCGAAGAGGCT-----	112
-----CGATGGGCGAAGAGGCT-----	1
-----AGTCGATGGGCGAAGAGGCT-----	122990
-----GAGTCGATGGGCGAAGAGGCT-----	6500
-----AGTCGATGGGCGAAGAGGCTC-----	168
-----GTCGATGGGCGAAGAGGCTC-----	3
-----GAGTCGATGGGCGAAGAGGCTC-----	4
-----GGAGTCGATGGGCGAAGAGGCTC-----	4
-----CGATGGGCGAAGAGGCTCGA-----	5
-----CGATGGGCGAAGAGGCTCGAT-----	672
-----GATGGGCGAAGAGGCTCGAT-----	7

Cluster: 12364- chromosome_17 Start: 5152766 End: 5153070	
--CTGGCCGTGCGACATTGTGTT-----	1
--TGGCCGTGCGACATTGTG-----	2
--TGGCCGTGCGACATTGTGTTA-----	1
--TGGCCGTGCGACATTGTGT-----	7
--TGGCCGTGCGACATTGTGTT-----	34
--TGGCCGTGCGACATTGTGTT-----	5
-----TTGTGTTTAATCTTCTCCTCTA-----	14
-----TTGTGTTTAATCTTCTCCTCT-----	140
-----TTGTGTTTAATCTTCTCCTC-----	6
-----TGTGTTTAATCTTCTCCTCTA-----	5
-----TGTGTTTAATCTTCTCCTCT-----	2
-----TGTTTAATCTTCTCCTCTA-----	1
-----TGTTTAATCTTCTCCTCTACCT-----	97
-----TGTTTAATCTTCTCCTC-----	1
-----TGTTTAATCTTCTCCTCTAC-----	1
-----TGTTTAATCTTCTCCTCTACCC-----	496
-----TGTTTAATCTTCTCCTCTACCTC-----	97
-----GTTTAATCTTCTCCTCTACCTC-----	1
-----GTTTAATCTTCTCCTCTACCC-----	3
-----GTTTAATCTTCTCCTCTACCT-----	2
-----TTTAATCTTCTCCTCTACCTCAGC-----	4
-----TTTAATCTTCTCCTCTACCT-----	128
-----TTTAATCTTCTCCTCTAC-----	3
-----TTTAATCTTCTCCTCTACCTCA-----	83
-----TTTAATCTTCTCCTCTACCTC-----	2316
-----TTTAATCTTCTCCTCTACCC-----	81
-----TTAATCTTCTCCTCTACCT-----	2
-----TTAATCTTCTCCTCTACCTCAG-----	4
-----TTAATCTTCTCCTCTACCC-----	1
-----TTAATCTTCTCCTCTACCTC-----	52
-----TTAATCTTCTCCTCTACCTCA-----	363
-----TTAATCTTCTCCTCTACCTCAGC-----	101

```

-----TAATCTTCTCCTCTACCTCAG-----72
-----TAATCTTCTCCTCTACC-----1
-----TAATCTTCTCCTCTACCTCA-----14
-----TAATCTTCTCCTCTACCTCAGC-----55
-----TAATCTTCTCCTCTACCTC-----6
-----AATCTTCTCCTCTACCTCAG-----3
-----TCTTCTCCTCTACCTCAGC-----2
-----CTTCTCCTCTACCTCAGC-----1
-----TCTCCTCTACCTCAGC-----1
-----TCCTCTACCTCAGCTACCCGC-----1
-----TCTACCTCAGCTACCCGCTTC-----1
-----TACCTCAGCTACCCGCTTCTCCG-----6
-----CTCAGCTACCCGCTTCTCCGAGCT-----1
-----TCAGCTACCCGCTTCTCCG-----6
-----TCAGCTACCCGCTTCTCCGAGCT-----44
-----TCAGCTACCCGCTTCTCCGAG-----193
-----TCAGCTACCCGCTTCTCCGAGC-----2030
-----TCAGCTACCCGCTTCTCCGA-----2
-----TCAGCTACCCGCTTCTCC-----2
-----CAGCTACCCGCTTCTCCGAGC-----10
-----CAGCTACCCGCTTCTCCGAG-----1
-----AGCTACCCGCTTCTCCGAGC-----8
-----AGCTACCCGCTTCTCCGAG-----1
-----AGCTACCCGCTTCTCCGAGCT-----1
-----CTACCCGCTTCTCCGAGCTAG-----1
-----TACCCGCTTCTCCGAGCT-----1
-----TACCCGCTTCTCCGAGCTA-----3
-----TACCCGCTTCTCCGAGCTAG-----255
-----TACCCGCTTCTCCGAGCTAGC-----51
-----ACCCGCTTCTCCGAGCTAG-----1
-----CCCGCTTCTCCGAGCTAGC-----2
-----TCTCCGAGCTAGCTTGACGC-----1
-----TAGCTTGACGCTCTCTCAGCA-----3
-----TAGCTTGACGCTCTCTCAGC-----15
-----AGCTTGACGCTCTCTCAGC-----2

- C C CA
ACTGGC GTGCGACATTGTGTTTAATCTTCTCCTCTACCTCAGCTACCCGCTTCTCCGAGCTAGCTTGACGCTCTCT AG T
TGACCG CACGCTGTAAACACAAATTAGAAGAGGAGATGGAAGTCGATGGGCGAAGAGGCTCGATCGAACTGCGAGAGA TC T
G A C TT

-----GATCGAACTGCGAGAGACTCT-----52
-----CGATCGAACTGCGAGAGACTCT-----19
-----GATGGGCGAAGAGGCTCGAT-----7
-----CGATGGGCGAAGAGGCTCGAT-----672
-----CGATGGGCGAAGAGGCTCGA-----5
-----AGTCGATGGGCGAAGAGGCTC-----168
-----GAGTCGATGGGCGAAGAGGCTC-----4
-----GGAGTCGATGGGCGAAGAGGCTC-----4
-----GTCGATGGGCGAAGAGGCTC-----3
-----TGGAGTCGATGGGCGAAGAGGCT-----11
-----CGATGGGCGAAGAGGCT-----1
-----GGAGTCGATGGGCGAAGAGGCT-----5468
-----ATGGGCGAAGAGGCT-----3
-----AGTCGATGGGCGAAGAGGCT-----122990
-----TCGATGGGCGAAGAGGCT-----2
-----GATGGGCGAAGAGGCT-----1
-----GAGTCGATGGGCGAAGAGGCT-----6500
-----GTCGATGGGCGAAGAGGCT-----112
-----GAGTCGATGGGCGAAGAGGC-----16
-----AGTCGATGGGCGAAGAGGC-----294
-----GGAGTCGATGGGCGAAGAGGC-----13
-----GGAGTCGATGGGCGAAGAGG-----14
-----AGTCGATGGGCGAAGAGG-----370
-----GAGTCGATGGGCGAAGAGG-----6
-----GATGGAGTCGATGGGCGAAGAG-----2
-----GATGGAGTCGATGGGCGAAGA-----1
-----AGAAGAGGAGATGGAGTCGAT-----3
-----GAAGAGGAGATGGAGTCGAT-----11
-----AATTAGAAGAGGAGATGGAGTC-----2
-----AAATTAGAAGAGGAGATGGAGT-----251
-----CAAATTAGAAGAGGAGATGGAGT-----5
-----GAAGAGGAGATGGAGT-----1
-----ATTAGAAGAGGAGATGGAGT-----2004
-----AGAGGAGATGGAGT-----1
-----AATTAGAAGAGGAGATGGAGT-----9982
-----TTAGAAGAGGAGATGGAGT-----40
-----TAGAAGAGGAGATGGAGT-----2

```

-----AGAAGAGGAGATGGAGT-----4
-----ACAAATTAGAAGAGGAGATGGAGT-----2
-----AATTAGAAGAGGAGATGGAG-----64
-----ATTAGAAGAGGAGATGGAG-----8
-----AAATTAGAAGAGGAGATGGAG-----5
-----AATTAGAAGAGGAGATGGA-----7
-----AAATTAGAAGAGGAGATGGA-----2
-----ACAAATTAGAAGAGGAGATGG-----1
-----GTAACACAAATTAGAAGAGGAGAT-----1
-----CACAAATTAGAAGAGGAGAT-----5
-----CAAAATTAGAAGAGGAGAT-----1
-----ACAAATTAGAAGAGGAGAT-----17
-----ACACAAATTAGAAGAGGAGAT-----781
-----TAACACAAATTAGAAGAGGAGAT-----20
-----AACACAAATTAGAAGAGGAGAT-----129
-----ACACAAATTAGAAGAGGAGA-----13
-----AACACAAATTAGAAGAGGAGA-----4
-----ACACAAATTAGAAGAGGAG-----1
-----CGACACGCTGTAACACAAATT-----14
-----CCGACACGCTGTAACACAAATT-----1
-----CCGACACGCTGTAACACAAAT-----4

Cluster: 12551+ chromosome 17 Start: 6144113 End: 6144214

-----TGGAGCAGGGCAGGCCGCAA--1
-----TGGAGCAGGGCAGGCCGCAA--1
-----TGGAGCAGGGCAGGCCGCA--1
-----TGGAGCAGGGCAGGCCGCAAT-3
-----TGGAGCAGGGCAGGCCGC--44
-----TTTATGTGGAGCAGGGCAGG--1
-----TCTGTTTATGTGGAGCGGC--4
-----TTGTCTGTTTATGTGGAGC--14
-----TTGTCTGTTTATGTGG--2
-----CTGTCTGTTTATGTGGAGC--1
-----GCCTTGTCTGTTTATG--1
-----GCCTTGTCTGTTTATGTGG--35
-----GCCTTGTCTGTTTATGTG--5
-----CGCCTTGTCTGTTTATG--1
-----CGCCTTGTCTGTTTATGTGG--154
-----CGCCTTGTCTGTTTATGTG--14
-----CGCCTTGTCTGTTTATGTGGA--6
-----TCGCCTTGTCTGTTTATGTGG--9949
-----TCGCCTTGTCTGTTTATG--9
-----TCGCCTTGTCTGTTTATGTGGA--156
-----TCGCCTTGTCTGTTTATGTGGAG--2
-----TCGCCTTGTCTGTTTATGTG--869
-----TCGCCTTGTCTGTTTATGTGGAGC--1
-----GTCGCCTTGTCTGTTTATGTG--1
-----GTCGCCTTGTCTGTTTATG--1
-----GGTCGCCTTGTCTGTTTATG--236
-----CGGTCGCCTTGTCTGTTTATG--345
-----CGGTCGCCTTGTCTGTTTATGTG--1
-----TCGGTCGCCTTGTCTGTTTAT--1
-----TCGGTCGCCTTGTCTGTTTATG--441

--CA G AAAT
T TCGGTTCGCCTTGTCTGTTTATGTGGAGCAGGGCAGGCCG G
A AGCCAGCGGAACAGACAAATACACCTCGTCCGTCGGCG G
GCAGC G GGGG

-----AGCCAGCGGAACAGACAA-----6
-----AGCCAGCGGAACAGACAAA-----17
-----GCCAGCGGAACAGACAAA-----1
-----AGAGCCAGCGGAACAGACAAA-----4
-----AGCCAGCGGAACAGACAAAT-----1571
-----GCCAGCGGAACAGACAAAT-----15
-----AGAGCCAGCGGAACAGACAAAT-----114
-----GAGCCAGCGGAACAGACAAAT-----2
-----CAGAGCCAGCGGAACAGACAAAT-----1
-----CCAGCGGAACAGACAAATACA-----1
-----AGCGGAACAGACAAATACACCT-----1
-----GCGGAACAGACAAATACACCT-----30

Cluster: 12551- chromosome 17 Start: 6144109 End: 6144230

-----TCGTGTGCCTGCCGTCGTCTCG-----1
-----TGTGCCTGCCGTCGTCTCGGT-----6
-----TGTGCCTGCCGTCGTCTCGG-----14

-----GTCGCGTCGCCTTGTCTGTT-----	1	
-----TCTCGGTGCGCCTTGTCTGTT-----	1	
-----CTCGGTGCGCCTTGTCTGTTTATG-----	2	
-----TCGGTCGCGCCTTGTCTGTTTATG-----	441	
-----TCGGTCGCGCCTTGTCTGTTTAT-----	1	
-----CGGTGCGCCTTGTCTGTTTATGTG-----	1	
-----CGGTGCGCCTTGTCTGTTTATG-----	345	
-----GGTCGCGCCTTGTCTGTTTATG-----	236	
-----GTGCGCCTTGTCTGTTTATG-----	1	
-----GTGCGCCTTGTCTGTTTATGTG-----	1	
-----TCGCGCTTGTCTGTTTATGTGG-----	9949	c12551b
-----TCGCGCTTGTCTGTTTATGTGGAG-----	2	
-----TCGCGCTTGTCTGTTTATG-----	9	
-----TCGCGCTTGTCTGTTTATGTGGAGC-----	1	
-----TCGCGCTTGTCTGTTTATGTGGA-----	156	
-----TCGCGCTTGTCTGTTTATGTG-----	869	
-----CGCCTTGTCTGTTTATG-----	1	
-----CGCCTTGTCTGTTTATGTGGA-----	6	
-----CGCCTTGTCTGTTTATGTGG-----	154	
-----CGCCTTGTCTGTTTATGTG-----	14	
-----GCCTTGTCTGTTTATG-----	1	
-----GCCTTGTCTGTTTATGTG-----	5	
-----GCCTTGTCTGTTTATGTGG-----	35	
-----CTTGTCTGTTTATGTGGAGC-----	1	
-----TTGTCTGTTTATGTGGAGC-----	14	
-----TTGTCTGTTTATGTGG-----	2	
-----TCTGTTTATGTGGAGCAGGGC-----	4	
-----TTTATGTGGAGCAGGGCAGG-----	1	
-----TGGAGCAGGGCAGGCCGCCA-----	26	
-----TGGAGCAGGGCAGGCCGCG-----	44	
-----TGGAGCAGGGCAGGCCGCC-----	159	
-----TGGAGCAGGGCAGGCCGCCAA-----	88	
-----TGGAGCAGGGCAGGCCGCCAATG-----	33	
-----TGGAGCAGGGCAGGCCGCCAAT-----	87	
CAGTCTGTTGCCTG C C TG		
CCGTCGT TCGGTGCGCCTTGTCTGTTTATGTGGAGCAGGGCAGGCCGC AA A		
GGCAGTA AGCCAGCGGAACAGACAAATACACCTCGTCCCGTCCGGCG TT /		
----- C - TT		
-----CCTCGTCCCGTCCGGCG-TTT-----	10	
-----ACCTCGTCCCGTCCGGCG-TTT-----	1	
-----ACACCTCGTCCCGTCCGGCG-T-----	3	
-----GCGGAACAGACAAATACACCT-----	30	
-----AGCGGAACAGACAAATACACCT-----	1	
-----CCAGCGGAACAGACAAATACA-----	1	
-----TACAGCCAGCGGAACAGACAAAT-----	30	
-----ACAGCCAGCGGAACAGACAAAT-----	18839	c12551a
-----GTACAGCCAGCGGAACAGACAAAT-----	2	
-----GCCAGCGGAACAGACAAAT-----	15	
-----AGCCAGCGGAACAGACAAAT-----	1571	
-----CAGCCAGCGGAACAGACAAAT-----	855	
-----ACAGCCAGCGGAACAGACAAA-----	280	
-----TACAGCCAGCGGAACAGACAAA-----	2	
-----CAGCCAGCGGAACAGACAAA-----	9	
-----AGCCAGCGGAACAGACAAA-----	17	
-----GCCAGCGGAACAGACAAA-----	1	
-----ACAGCCAGCGGAACAGACAA-----	56	
-----AGCCAGCGGAACAGACAA-----	6	
-----CAGCCAGCGGAACAGACAA-----	2	
-----ACAGCCAGCGGAACAG-----	2	
Cluster: 12615+ chromosome_17 Start: 6673966 End: 6674281		
-----TCAAGACCGGGGATGGAAG-----	35	
-----TCAAGACCGGGGATGGAAG-----	6	
-----ATCAAGACCGGGGATGGAAG-----	1	
-----GCAAAATCAAGACCGGGGGA-----	30	
-----TGCAAAATCAAGACCGGGG-----	1	
-----TGCAAAATCAAGACCGGGG-----	10	
-----TGCAAAATCAAGACCGGGGAT-----	3	
-----TGCAAAATCAAGACCGGGGGA-----	1243	
-----TTGCAAAATCAAGACCGGGGGA-----	3	
-----TTGCAAAATCAAGACCGGGG-----	2	
-----TTGCAAAATCAAGACCGGGG-----	18	
-----TTTGCAAAATCAAGACCGGG-----	4	
-----TTTGCAAAATCAAGACCGG-----	3	
-----TTTGCAAAATCAAGACCGGG-----	95	

-----	-----TTTGCAAAATCAAGACCGGGG-----	22
-----	-----TAAGGCGTTGATGCATTC-----	1
-----	-----TGACAGTAAGGCGTTGATG-----	20
-----	-----TGACAGTAAGGCGTTGATGC-----	15
-----	-----AATGACAGTAAGGCGTTGA-----	1
-----	-----CAAATGACAGTAAGGCGTGT-----	2
-----	-----TGGGACCCGCACATGCATG-----	1
-----	-----TTGGGACCCGCACATGCATG-----	6
-----	-----TTGGGACCCGCACATGCATGT-----	1
-----	-----TTGCTTGGGACCCGCACATG-----	8
-----	-----TTGCTTGGGACCCGCACATGC-----	3
-----	-----TGACTTGCTTGGGACCCGC-----	12
-----	-----TGACTTGCTTGGGACCCGCA-----	2
-----	-----TGACTTGCTTGGGACCCGCAC-----	1
-----	-----TGACTTGCTTGGGACCCGCACA-----	4
-----	-----TGAACAATGACTTGCTTGGGACC-----	20
-----	-----TGAACAATGACTTGCTTGGGA-----	15
-----	-----TGAACAATGACTTGCTTGG-----	2
-----	-----TGAACAATGACTTGCTTGGGAC-----	47
-----	-----TGAACAATGACTTGCTTGGG-----	61
-----	-----TGGGACGTGAACAATGACTTGCT-----	1
-----	-----TGGGACGTGAACAATGACTT-----	1
-----	-----TGGGACGTGAACAATGACTTG-----	7
-----	-----TGGGACGTGAACAATGACTTGC-----	3
-----	-----TGGGACGTGAACAATGAC-----	1
-----	-----TTGGGACGTGAACAATGAC-----	276
-----	-----TTGGGACGTGAACAATGACT-----	7
-----	-----TGCCTTGGGACGTGAACAATG-----	2
-----	-----TTGCGTTGGGACGTGAACAATG-----	3
-----	-----	
-----	-----C-----	
CGGCTTGCCTTGGGACGTGAACAATGACTTGCTTGGGACCCGCACATGCATGT	ACAAATGACAGTAAGGCGTGTGATGCATTAGATTGCAAAATCAAGACCGGGGATGGAAGG	TA TGGTGGAGGGGAGAAGGCCACCCGTCGGCACGCGCCCAACCCATACATCC G
GCCGAACGCAACCTGCACTTGTTACTGAACGAACCTGGGCGTGTACGTACA	TGTTTACTGTCATTCCGCACAACACTACGTAAGTCTAACCGTTTATGTTCTGGCCCCCTACCTTCC	TG ACGC T
CCGTGG	A	AC TGC G A
-----	-----	-----A-----
-----	-----	
GGCCCGAACGCAACCTTGCACT-----	-----	1
CCGAACGCAACCTTGCACTT-----	-----	1
GGCCGAACGCAACCTTGCACTT-----	-----	1
CGAACGCAACCTTGCACTTGT-----	-----	1
GAACGCAACCTTGCACTTGTT-----	-----	20
ACGCAACCTTGCACTTGTT-----	-----	1
CGAACGCAACCTTGCACTTGTT-----	-----	1
GCAACCTTGCACTTGTTACT-----	-----	1
GTTACTGAACGAACCTTGGGCGT-----	-----	1
ACTGAACGAACCTTGGGCGTGT-----	-----	1
CTGAACGAACCTTGGGCGTGT-----	-----	9
CGAACCTTGGGCGGTACGT-----	-----	1
GGGCGGTACGTACAATGTTT-----	-----	1
GTTTACTGTCATTCCGCACAA-----	-----	2
TTACTGTCATTCCGCACAAC-----	-----	1
TGTCATTCCGCACAACACTACGT-----	-----	16
ATTCCGCACAACACTACGTAAGT-----	-----	1
CATTCCGCACAACACTACGTAAGT-----	-----	3
CCGCACAACACTACGTAAGTCT-----	-----	9
TTCCGCACAACACTACGTAAGTCT-----	-----	1
TCCGCACAACACTACGTAAGTCT-----	-----	62
CGCACAACTACGTAAGTCTAAA-----	-----	1
CTACGTAAGTCTAAACGTT-----	-----	1
CGTAAGTCTAAACGTTTITAG-----	-----	1
ACGTAAGTCTAAACGTTTITAGT-----	-----	5
CGTAAGTCTAAACGTTTITAGT-----	-----	70
ACGTTTITAGTCTGGCCCCCT-----	-----	1
-----	-----	
Cluster: 12615- chromosome_17 Start: 6673955 End: 6674297	-----	
TGCTGCACCGCTGGCACCCGGC-----	-----	1
TTGCGTTGGGACGTGAACAATG-----	-----	3
TGCGTTGGGACGTGAACAATG-----	-----	2
TTGGGACGTGAACAATGAC-----	-----	276
TTGGGACGTGAACAATGACT-----	-----	7
TGGGACGTGAACAATGACTTGCT-----	-----	1
TGGGACGTGAACAATGACTTG-----	-----	7
TGGGACGTGAACAATGACTTGC-----	-----	3
TGGGACGTGAACAATGAC-----	-----	1
TGGGACGTGAACAATGACTT-----	-----	1
TGAACAATGACTTGCTTGGGAC-----	-----	47
TGAACAATGACTTGCTTGGGA-----	-----	15
TGAACAATGACTTGCTTGGG-----	-----	61

-----TGAACAATGACTTGCTTGGGACC-----20
-----TGAACAATGACTTGCTTGG-----2
-----TGACTTGCTTGGGACCCGCA-----2
-----TGACTTGCTTGGGACCCGCAC-----1
-----TGACTTGCTTGGGACCCGC-----12
-----TGACTTGCTTGGGACCCGACA-----4
-----TTGCTTGGGACCCGCACATG-----8
-----TTGCTTGGGACCCGCACATGC-----3
-----TTGGGACCCGCACATGCATGT-----1
-----TTGGGACCCGCACATGCATG-----6
-----TGGGACCCGCACATGCATG-----1
TGTTACAAATGACAGTAAGGCG-----14
TTACAAATGACAGTAAGGCGTG-----3
CAAATGACAGTAAGGCGTGT-----2
AATGACAGTAAGGCGTGTGA-----1
TGACAGTAAGGCGTGTGATGC-----15
TGACAGTAAGGCGTGTGATG-----20
TAAGGCGTGTGATGCATTC-----1
TTTGCAAAATCAAGACCGG-----3
TTTGCAAAATCAAGACCGGG-----4
TTTGCAAAATCAAGACCGGGG-----22
TTTGCAAAATCAAGACCGGGG-----95
TTTGCAAAATCAAGACCGGGG-----2
TTTGCAAAATCAAGACCGGGG-----18
TTTGCAAAATCAAGACCGGGGA-----3
TGCAAAATCAAGACCGGGGA-----1243
TGCAAAATCAAGACCGGGG-----10
TGCAAAATCAAGACCGGG-----1
TGCAAAATCAAGACCGGGGGAT-----3
GCAAAATCAAGACCGGGGA-----30
ATCAAGACCGGGGGATGGAAG-----1
TCAAGACCGGGGGATGGAAG-----6
TCAAGACCGGGGATGGAAG-----35
TGTGTGCTGCA C G
CCG T GCACCCGGCTTCGTTGGGACGTGAACAATGACTTGCTTGGGACCCGCACATGCATGTTACAAATGACAGTAAGGCGTGTGATGCATTAGATTGTGAGACGCTTACGC A TAT TT- - G G
GGC G CGTGGGCCGAACGCAACCTGCACCTTGTACTGAACGAACCTGGGCGTGTACGTACAGTGTTTACTGTCTATCCGCACAACCTACGTAAGTCTAAGCGTTTAGTTCTGGCCCTACCTTCC GTGG TG GGG GGC GC CGT C
----- A G CACC AC CCC CCG TG GCA G
ATA----- - CTC TCTT G G G
-----ACGTTTTAGTTCTGGCCCCCT-----1
-----ACGTAAGTCTAAACGTTTTAGT-----5
-----CGTAAGTCTAAACGTTTTAGT-----70
-----CGTAAGTCTAAACGTTTTAG-----1
-----CTACGTAAGTCTAAACGTT-----1
-----CGCACAACTACGTAAGTCTAAA-----1
-----TTCCGCACAACCTACGTAAGTCT-----1
-----TCCGCACAACCTACGTAAGTCT-----62
-----CCGCACAACCTACGTAAGTCT-----9
-----ATTCCGCACAACCTACGTAAGT-----1
-----CATTCGCACAACCTACGTAAGT-----3
-----TGTCATTCCGCACAACCTACGT-----16
-----TTACTGTCTATCCGCACAACCT-----1
GTTTACTGTCTATCCGCACAA-----2
-----TGGGCGGTGTACGTACAGTGTTT-----2
-----GGGCGGTGTACGTACAGTGTTT-----32
-----GGCGTGTACGTACAGTGTTT-----1
-----GGGCGGTGTACGTACAGTGTT-----15
-----TGGGCGGTGTACGTACAGTGTT-----26
-----TGGGCGGTGTACGTACAGTGT-----1
CGAACCCCTGGGCGGTGTACGT-----1
ACTGAACGAACCCCTGGGCGTGT-----1
CTGAACGAACCCCTGGGCGTGT-----9
GTTACTGAACGAACCCCTGGGCGT-----1
GCAACCCCTGCACTTGTACT-----1
ACGCAACCCCTGCACTTGT-----1
GAACGCAACCCCTGCACTTGT-----20
CGAACGCAACCCCTGCACTTGT-----1
CGAACGCAACCCCTGCACTTGT-----1
GGCCGAACGCAACCCCTGCACTT-----1
CCGAACGCAACCCCTGCACTT-----1
GGGCCGAACGCAACCCCTGCACT-----1
GGCGTGGGCCGAACGCAACCCCT-----1
Cluster: 12905- chromosome_2 Start: 739300 End: 739394
-----AATGGCCGTACTACTATGTGCT-----1
-----ATGGCCGTACTACTATGTGCT-----1
-----TGCCCGTACTACTATT-----1


```
-----TGGCCGTACTACTATIGTCAT-----7
-----TGGCCGTACTACTATIGT-----61
-----TGGCCGTACTACTATIGTCATG-----4
-----TGGCCGTACTACTATIGT-----5494
-----TGGCCGTACTACTATIG-----25
-----TGGCCGTACTACTATIGTCA-----463
-----TGGCCGTACTACTATIGTCATGT-----1
-----GGCCGTACTACTATIGT-----1
-----GGCCGTACTACTATIGTC-----144
-----GGCCGTACTACTATIGTCA-----23
-----CCGTACTACTATIGTC-----5
-----TCATGTCTGACTGCATCATGT-----1
-----TGTCTGACTGCATCATGT-----1

----GGT-----T
GAAAATGGCCGTACTACTATIGTCAATGTCTGACTGCATCATGT
CTTTACCGGCATGATGGTAACAGTACAGACTGACGTAGTAC/
AACTT TA-

-----GGTAACAGTACAGACTGACGT-----11
-----GATGGTAACAGTACAGACT-----1
-----TTTACCGGCATGATGGTAACAGT-----1
-----TTACCGGCATGATGGTAACAGT-----3
-----ACCGGCATGATGGTAACAGT-----1
---TCTATTACCGGCATGAT-----1

Cluster: 14705+ chromosome_2 Start: 9101099 End: 9101289
-----TCTCAGTACGAGCGACATAGG-----1
-----TTCCGACGCAGCGGCAGTGT-----23
-----TTTCCGACGCAGCGGCAGTGT-----30
-----TTTCCGACGCAGCGGCAGTGT-----1
-----TCCTTTCCGACGCAGCGGCAGT-----2
-----TCCTTTCCGACGCAGCGGCAG-----129
-----TTCCCTTTCCGACGCAGCGGCAG-----13
-----TTCCTTTCCGACGCAGCGGC-----1
-----TGGATTCCCTTTCCGACGCAGC-----5
-----TTGGATTCCCTTTCCGACGC-----2
-----AGGCCGTTGGATTCCCTTTCCG-----26
-----AGGCCGTTGGATTCCCTTTCCGA-----1
-----GAGGCCGTTGGATTCCCTTTCCG-----1
-----TGGAGGCCGTTGGATTCCCTTTCC-----1
-----TGGAGGCCGTTGGATTCCCTTT-----5
-----CATGGAGGCCGTTGGATTCC-----2
-----TCATGGAGGCCGTTGGATTCC-----1
-----TCATGGAGGCCGTTGGATTCC-----311
-----TGGACGCAAACCACACCAG-----2
-----TGGACGCAAACCACACCAGCG-----9
-----TGGACGCAAACCACACCAGCTC-----1
-----TGGACGCAAACCACACCAGC-----28

-----
AACATGGACGCAAACCACACCAGCGTCATGGAGGCCGTTGGATTCCCTTTCCGA CTA C A
TTGTACCTGCGTTTGGTGTGGTTCGAGTACCTCCGGCAACCTAAGGAAAGGCT CGTCGCCGTACAGAGTCATGCTCGCTGT CCG G A
CTCGCC C TC A A

---CCTTGTAACCTGCGTTTGGTGT-----1
---CCTGCGTTTGGTGTGGTTCGAGT-----1
-----TCCGGCAACCTAAGGAAAGG-----1
-----CCGGCAACCTAAGGAAAGG-----28
-----CCGGCAACCTAAGGAAAGGC-----120
-----TCCGGCAACCTAAGGAAAGGC-----3
-----CCTCCGGCAACCTAAGGAAAGGCT-----1
-----CTCCGGCAACCTAAGGAAAGGCT-----3
-----CCGGCAACCTAAGGAAAGGCT-----14785
-----GGCAACCTAAGGAAAGGCT-----4
-----GCAACCTAAGGAAAGGCT-----1
-----CGGCAACCTAAGGAAAGGCT-----73
-----TCCGGCAACCTAAGGAAAGGCT-----450
-----CCGGCAACCTAAGGAAAGGCTC-----3
-----TCCGGCAACCTAAGGAAAGGCTC-----39
-----CGGCAACCTAAGGAAAGGCTC-----2
-----GGCAACCTAAGGAAAGGCTCCGT-----1
-----GCAACCTAAGGAAAGGCTCCGT-----2
-----CAACCTAAGGAAAGGCTCCGT-----9
```

Cluster: 14705- chromosome_2 Start: 9101093 End: 9101283

-----TGGACGCAAACCACACCAG-----2
-----TGGACGCAAACCACACCAGCG-----9
-----TGGACGCAAACCACACCAGC-----28
-----TGGACGCAAACCACACCAGCGTC-----1
-----TCATGGAGGCCGTTGGATTCC-----311
-----TCATGGAGGCCGTTGGATTCCCT-----1
-----CATGGAGGCCGTTGGATTCC-----2
-----TGGAGGCCGTTGGATTCCCTTT-----5
-----TGGAGGCCGTTGGATTCCCTTCC-----1
-----GAGGCCGTTGGATTCCCTTCCG-----1
-----AGGCCGTTGGATTCCCTTCCGA-----1
-----AGGCCGTTGGATTCCCTTCCG-----26

-----A T
AACATGGACGCAAACCACACCAGCGTCATGGAGGCCGTTGGATTCCCTTCCGA G CAGCGGCAGTGTCTCAGTACGAGCGACA GGGCTC A
TTGTACCTGCGTTTGGTGTGGTTCGAGTACCTCCGGCAACCTAAGGAAAGGCT CGTCGCCGTCACAGAGTCATGCTCGCTGT TCCGGG T
CTCGCC G A T

-----CCGGCAACCTAAGGAAAGGCT-----14785
-----GCAACCTAAGGAAAGGCT-----1
-----GGCAACCTAAGGAAAGGCT-----4
-----CTCCGGCAACCTAAGGAAAGGCT-----3
-----CGGCAACCTAAGGAAAGGCT-----73
-----TCCGGCAACCTAAGGAAAGGCT-----450
-----CCTCCGGCAACCTAAGGAAAGGCT-----1
-----CCGGCAACCTAAGGAAAGGC-----120
-----TCCGGCAACCTAAGGAAAGGC-----3
-----TCCGGCAACCTAAGGAAAGG-----1
-----CCGGCAACCTAAGGAAAGG-----28
-----CCTGCGTTTGGTGTGGTTCGAGT-----1
-----CCTTGTAACCTGCGTTTGGTGT-----1

Cluster: 14711+ chromosome_2 Start: 9121113 End: 9121295
-----GCTCGCTGGGTTTATCATC-----1
-----TGCTCGCTGGGTTTATCATCC-----6
-----TGCTCGCTGGGTTTATCATC-----4
-----TGATGCTCGCTGGGTTTATCA-----1
-----TGGCTGGAGTGCAGCGCAGAGC-----8

GGGCTGGCTGGAGTGCAGCGCAGAGCTGGGCG C G AT - CCA T G A- G A-- - G G
GCCGT TCCGCTGATGCTCGCTG GTTTATC CCGCT TTGGC TGAT G TC CAAC CCGG G G CTGC T
CGGCA AGGCGACTACGAGCGGC CAAGTAG GCGCA AGCCG ATTA C AG GTTG GGCC C C GACG C
-----A A CG T --- - G AC - GTA G G C

-----GGCGACTACGAGCGGCACAAGT-----1231
-----GCGACTACGAGCGGCACAAGT-----784
-----AGGCGACTACGAGCGGCACAAGT-----7
-----AAAGGCGACTACGAGCGGCACAAGT-----1
-----CGACTACGAGCGGCACAAGT-----19
-----GCGGCACAAGTAGCGGGCGAT-----2

Cluster: 14712+ chromosome_2 Start: 9129505 End: 9129606
-----GAGGGCGCAAGCACGCAAG-----1
-----AGAGGGCGCAAGCACGCAAG-----9
-----AGAGGGCGCAAGCACGCA-----6
-----TAGAGGGCGCAAGCAC-----1
-----TAGAGGGCGCAAGCACGCAAG-----647
-----TAGAGGGCGCAAGCACGC-----142
-----TAGAGGGCGCAAGCACGCA-----798
-----TAGAGGGCGCAAGCACGCAAGG-----12
-----TAGAGGGCGCAAGCACGCAA-----35
-----CTAGAGGGCGCAAGCACGCA-----2
-----GCTAGAGGGCGCAAGCACGCA-----1
-----CGGCTAGAGGGCGCAAGCACG-----3
-----CGGCTAGAGGGCGCAAGCACGC-----1

-----G AC
AGACGGCTAGAGGGCGCAAGCACGCAAGGAGGGGATGACGAGC A
TCTGCCGATCTCCCGGTTCTGTCGTTTCTCCCTACTGCTCG /
GACTGCCGA TT

-----CGATCTGCCGATCTCCCGGT-----1
-----GCCGATCTGCCGATCTCCCGGT-----1
-----GCCGATCTCCCGGTTTCG-----125
-----CTGCCGATCTCCCGGTTTCG-----125
-----CTGCCGATCTCCCGGTTCTG-----8955

-----TCGCCGATCTCCCGCGTTCGT----- 22
-----CCGATCTCCCGCGTTCGT----- 4
-----CGATCTCCCGCGTTCGT----- 6
-----TGCCGATCTCCCGCGTTCGT----- 46
-----GATCTCCCGCGTTCGT----- 1
-----GCCGATCTCCCGCGTTCGT----- 235
-----GCCGATCTCCCGCGTTCGT----- 13
-----GCCGATCTCCCGCGTTCGTGC----- 91
-----CGATCTCCCGCGTTCGTGCG----- 2
-----CGATCTCCCGCGTTCGTGCGT----- 45
-----CGATCTCCCGCGTTCGTGCGTT----- 12
-----GATCTCCCGCGTTCGTGCGTT----- 3
-----GCGTTCGTGCGTTCCTCCCCT----- 3
-----CGTTCGTGCGTTCCTCCCCT----- 4
-----TCGTGCGTTCCTCCCCTACT----- 4

Cluster: 14712- chromosome_2 Start: 9129502 End: 9129598
-----CGGCTAGAGGGCGCAAGCACGC----- 1
-----CGGCTAGAGGGCGCAAGCACG----- 3
-----GCTAGAGGGCGCAAGCACGCA----- 1
-----CTAGAGGGCGCAAGCACGCA----- 2
-----TAGAGGGCGCAAGCACGCAAGG----- 12
-----TAGAGGGCGCAAGCACGCAAG----- 647
-----TAGAGGGCGCAAGCACGC----- 142
-----TAGAGGGCGCAAGCACGCAA----- 35
-----TAGAGGGCGCAAGCAC----- 1
-----TAGAGGGCGCAAGCACGCA----- 798
-----AGAGGGCGCAAGCACGCA----- 6
-----AGAGGGCGCAAGCACGCAAG----- 9
-----GAGGGCGCAAGCACGCAAG----- 1

--T AT
AGACGGCTAGAGGGCGCAAGCACGCAAGGAGGGGATGACGAGC G
TCTGCCGATCTCCCGCGTTCGTGCGTTCCTCCCCTACTGCTCG /
TGGC TT

-----TCGTGCGTTCCTCCCCTACT----- 4
-----GCGTTCGTGCGTTCCTCCCCT----- 3
-----CGTTCGTGCGTTCCTCCCCT----- 4
-----CGATCTCCCGCGTTCGTGCGTT----- 12
-----GATCTCCCGCGTTCGTGCGTT----- 3
-----CGATCTCCCGCGTTCGTGCGT----- 45
-----CGATCTCCCGCGTTCGTGCG----- 2
-----GCCGATCTCCCGCGTTCGTGC----- 91
-----GCCGATCTCCCGCGTTCGTG----- 13
-----TGCCGATCTCCCGCGTTCGT----- 46
-----CTGCCGATCTCCCGCGTTCGT----- 8955
-----CCGATCTCCCGCGTTCGT----- 4
-----GCCGATCTCCCGCGTTCGT----- 235
-----GATCTCCCGCGTTCGT----- 1
-----TCTGCCGATCTCCCGCGTTCGT----- 22
-----CGATCTCCCGCGTTCGT----- 6
-----CTGCCGATCTCCCGCGTTCG----- 125
-----GCCGATCTCCCGCGTTCG----- 1

Cluster: 14719+ chromosome_2 Start: 9173113 End: 9173421
-----GGAGGGCGCAAGCACGCAAC----- 6
-----GGAGGGCGCAAGCACGCAACG----- 6
-----TGAGGGCGCAAGCACGCAACGA----- 1
-----TGAGGGCGCAAGCACGCA----- 1
-----TGAGGGCGCAAGCACGCAAC----- 613
-----TGAGGGCGCAAGCACGCAA----- 2
-----TGAGGGCGCAAGCACGCAACG----- 974
-----TAGACGGCTGAGGGCGCAAG----- 5

A C CCT TC
GCTAGACGGCTGGAGGGCGCAAGCACGCAACGAGGGGATGACGAGCACGATG GGGCCTGATACAG CG C
CGATCTGCCGACCTCCCGCGTTCGTGCGTTGCTCCCCTACTGCTCGTAC CCCGACTATGTC GC /
C A C-- CC

---TCTGCCGACCTCCCGGTTTCGT----- 2
---TGCCGACCTCCCGGTTTCGT----- 1
---CTGCCGACCTCCCGGTTTCGT----- 20
---CGACCTCCCGGTTTCGTGCGT----- 58
---CGACCTCCCGGTTTCGTGCGTT----- 405
---ACCTCCCGGTTTCGTGCGTT----- 12

```
-----CCTCCCGGTTCGTGCGTT-----4
-----GACCTCCCGGTTCGTGCGTT-----8
-----GCGTTCGTGCGTTGCTCCCCT-----5
-----CGTTCGTGCGTTGCTCCCCT-----3
-----GTTTCGTGCGTTGCTCCCCTACT-----2
-----CGTTCGTGCTCCCCTACT-----1
-----TCGTGCGTTGCTCCCCTACTGCT-----1
-----CGTTCGTGCTCCCCTACTGCT-----1
-----GTGCGTTGCTCCCCTACTGCT-----6
-----GTTGCTCCCCTACTGCTCG-----3
-----GCGTTGCTCCCCTACTGCTCG-----3
-----CGTTGCTCCCCTACTGCTCG-----3
-----CGTTGCTCCCCTACTGCTCGT-----85
-----GTTGCTCCCCTACTGCTCGT-----45
-----GCGTTGCTCCCCTACTGCTCGT-----115
-----TTGCTCCCCTACTGCTCGT-----1
-----TGCTCCCCTACTGCTCGTG-----1
-----GCTCCCCTACTGCTCGTG-----20
-----GCTCCCCTACTGCTCGTGC-----57
-----GCTCCCCTACTGCTCGTGCT-----8513
-----TGCTCCCCTACTGCTCGTGCT-----279
```

```
Cluster: 14719-      chromosome_2      Start: 9173274      End: 9173422
---TAGACGGCTGGAGGGCGCAAG-----5
---TGGAGGGCGCAAGCACGCA-----1
---TGGAGGGCGCAAGCACGCAACG-----974
---TGGAGGGCGCAAGCACGCAACGA-----1
---TGGAGGGCGCAAGCACGCAAC-----613
---TGGAGGGCGCAAGCACGCAA-----2
---GGAGGGCGCAAGCACGCAAC-----6
---GGAGGGCGCAAGCACGCAACG-----6
```

```
C      G G- G
GGCTAGACGGCTGGAGGGCGCAAGCACGCAACGAGGGGATGACGAGCACGATGTGGGCCTGATACAG C G C
TCGATCTGCCGACCTCCCGGTTTCGTGCGTTGCTCCCCTACTGCTCGTGCTACGCCCGGACTATGTC G C A
-      G AG A
```

```
-----GCTCCCCTACTGCTCGTGCT-----8513
-----TGCTCCCCTACTGCTCGTGCT-----279
-----GCTCCCCTACTGCTCGTGC-----57
-----GCTCCCCTACTGCTCGTG-----20
-----TGCTCCCCTACTGCTCGTG-----1
-----TTGCTCCCCTACTGCTCGT-----1
-----CGTTGCTCCCCTACTGCTCGT-----85
-----GTTGCTCCCCTACTGCTCGT-----45
-----GCGTTGCTCCCCTACTGCTCGT-----115
-----GCGTTGCTCCCCTACTGCTCG-----3
-----GTTGCTCCCCTACTGCTCG-----3
-----CGTTGCTCCCCTACTGCTCG-----3
-----CGTTCGTGCTCCCCTACTGCT-----1
-----GTGCGTTGCTCCCCTACTGCT-----6
-----TCGTGCGTTGCTCCCCTACTGCT-----1
-----CGTTCGTGCTCCCCTACT-----1
-----GTTTCGTGCGTTGCTCCCCTACT-----2
-----GCGTTTCGTGCGTTGCTCCCCT-----5
-----CGTTTCGTGCGTTGCTCCCCT-----3
-----ACCTCCCGGTTTCGTGCGTT-----12
-----CGACCTCCCGGTTTCGTGCGTT-----405
-----GACCTCCCGGTTTCGTGCGTT-----8
-----CCTCCCGGTTTCGTGCGTT-----4
-----CGACCTCCCGGTTTCGTGCGT-----58
-----TCTGCCGACCTCCCGGTTTCGT-----2
-----CTGCCGACCTCCCGGTTTCGT-----20
-----TGCCGACCTCCCGGTTTCGT-----1
```

```
Cluster: 16411+      chromosome_3      Start: 6573885      End: 6573999
-----TCACCGCCGGAACCAACACGA---1
-----TCACCGCCGGAACCAACACGAC--2
-----TCACCGCCGGAACCAACACG---65
-----TGTCACCGCCGGAACCAACACG---1
-----TTCACCGCATGTACCCGCCG---5
-----TCACATTACCGCATGTCAAC---46
-----TCACATTACCGCATGTCAAC---1
-----TTCACATTACCGCATGTCAAC---4
-----TTCACATTACCGCATGTCAAC---12
-----ATTACATTACCGCATGTCA---1
```

-----ATTACATTACCGCATG-----	1
-----TCATTACATTACCGCATG-----	5
-----TCATTACATTACCGCATGTC-----	1
---TACACGCCATCATTACATT-----	2
T	
AAGTACACGCCATCATTACATTACCGCATGTCACCGCCGGAACCAACACGAC	A
TTCATGTGCGGTAGTAGTAAGTGAAGTGGCGTACAGTGGCGGCCTTGGTTGTGCTG	T
C	
-----TGCGGTAGTAAGTGTAAGT-----	1
-----TGTGCGGTAGTAAGTGTAAGT-----	185
-----GTGCGGTAGTAAGTGTAAGT-----	35
-----GCGGTAGTAAGTGTAAGT-----	2
-----ATGTGCGGTAGTAAGTGTAAGT-----	3
-----GCGGTAGTAAGTGTAAGTGGC-----	5
-----GGTAGTAAGTGTAAGTGGC-----	1
-----GGTAGTAAGTGTAAGTGGCG-----	11
-----GTAGTAAGTGTAAGTGGCG-----	4
-----GTAGTAAGTGTAAGTGGCGT-----	937
-----GCGGTAGTAAGTGTAAGTGGCGT-----	1
-----CGGTAGTAAGTGTAAGTGGCGT-----	4
-----GGTAGTAAGTGTAAGTGGCGT-----	4117
-----AGTAAGTGTAAGTGGCGT-----	1
-----GTAAGTGTAAGTGGCGTA-----	1
-----GTAGTAAGTGTAAGTGGCGTAC-----	1
-----AGTAAGTGTAAGTGGCGTAC-----	2
-----AGTAAGTGTAAGTGGCGTACA-----	1
-----TAAGTGTAAGTGGCGTACAG-----	7
-----AAGTGTAAGTGGCGTACAG-----	1
-----TGTAAAGTGGCGTACAGT-----	3
-----TAAGTGTAAGTGGCGTACAGT-----	833
-----TAGTAAGTGTAAGTGGCGTACAGT-----	7
-----AAGTGTAAGTGGCGTACAGT-----	145
-----AGTAAGTGTAAGTGGCGTACAGT-----	1
-----GTAAGTGGCGTACAGT-----	2
-----AGTGTAAGTGGCGTACAGT-----	11
-----GTAAGTGTAAGTGGCGTACAGT-----	20
-----GGCGTACAGTGGCGGCCTT-----	1
-----GCGTACAGTGGCGGCCTTGGT-----	3
-----GGCGTACAGTGGCGGCCTTGGT-----	5
-----GTACAGTGGCGGCCTTGGTT-----	1
-----CGTACAGTGGCGGCCTTGGTT-----	1
-----TACAGTGGCGGCCTTGGTTGT-----	2
-----ACAGTGGCGGCCTTGGTTGT-----	7
-----AGTGGCGGCCTTGGTTGTGCT-----	6
-----GTGGCGGCCTTGGTTGTGCT-----	24
Cluster: 16411- chromosome_3 Start: 6573885 End: 6573999	
---TACACGCCATCATTACATT-----	2
-----TCATTACATTACCGCATGTC-----	1
-----TCATTACATTACCGCATG-----	5
-----ATTACATTACCGCATG-----	1
-----ATTACATTACCGCATGTCA-----	1
-----TTCACATTACCGCATGTAC-----	4
-----TTCACATTACCGCATGTACCC-----	12
-----TCACATTACCGCATGTAC-----	1
-----TCACATTACCGCATGTACCC-----	46
-----TTCACCGCATGTACCGCCGG-----	5
-----TGTCACCGCCGGAACCAACACG-----	1
-----TCACCGCCGGAACCAACACGAC-----	2
-----TCACCGCCGGAACCAACACG-----	65
-----TCACCGCCGGAACCAACACGA-----	1
G	
AAGTACACGCCATCATTACATTACCGCATGTCACCGCCGGAACCAACACGAC	A
TTCATGTGCGGTAGTAGTAAGTGAAGTGGCGTACAGTGGCGGCCTTGGTTGTGCTG	C
A	
T	
-----AGTGGCGGCCTTGGTTGTGCT-----	6
-----GTGGCGGCCTTGGTTGTGCT-----	24
-----TACAGTGGCGGCCTTGGTTGT-----	2
-----ACAGTGGCGGCCTTGGTTGT-----	7
-----CGTACAGTGGCGGCCTTGGTT-----	1
-----GTACAGTGGCGGCCTTGGTT-----	1
-----GGCGTACAGTGGCGGCCTTGGT-----	5
-----GCGTACAGTGGCGGCCTTGGT-----	3

-----GGCGTACAGTGGCGGCCTT----- 1
-----GTAAGTGTAAGTGGCGTACAGT----- 20
-----TAAGTGTAAGTGGCGTACAGT----- 833
-----TAGTAAGTGTAAGTGGCGTACAGT----- 7
-----AGTAAGTGTAAGTGGCGTACAGT----- 1
-----AGTGTAAGTGGCGTACAGT----- 11
-----GTAAGTGGCGTACAGT----- 2
-----AAGTGTAAGTGGCGTACAGT----- 145
-----TGTAAAGTGGCGTACAGT----- 3
-----AAGTGTAAGTGGCGTACAG----- 1
-----TAAGTGTAAGTGGCGTACAG----- 7
-----AGTAAGTGTAAGTGGCGTACA----- 1
-----GTAGTAAGTGTAAGTGGCGTAC----- 1
-----AGTAAGTGTAAGTGGCGTAC----- 2
-----GTAAGTGTAAGTGGCGTA----- 1
-----CGGTAGTAAGTGTAAGTGGCGT----- 4
-----GTAGTAAGTGTAAGTGGCGT----- 937
-----GGTAGTAAGTGTAAGTGGCGT----- 4117
-----GCGGTAGTAAGTGTAAGTGGCGT----- 1
-----AGTAAGTGTAAGTGGCGT----- 1
-----GTAGTAAGTGTAAGTGGCG----- 4
-----GGTAGTAAGTGTAAGTGGCG----- 11
-----GGTAGTAAGTGTAAGTGGC----- 1
-----GCGGTAGTAAGTGTAAGTGGC----- 5
-----GTGCGGTAGTAAGTGTAAGT----- 35
-----ATGTGCGGTAGTAAGTGTAAGT----- 3
-----GCGGTAGTAAGTGTAAGT----- 2
-----TGTGCGGTAGTAAGTGTAAGT----- 185
-----TGCGGTAGTAAGTGTAAGT----- 1

Cluster: 17284- chromosome_4 Start: 1804230 End: 1804569

-----TGCCATCCCACCTACGCTGT----- 1
-----TCCCACCTACGCTGTCGGCT----- 1
-----TACGCTGTCGGCTGAGCTCACC----- 1
-----TCGGCTGAGCTCACC GCGG----- 17
-----TCGGCTGAGCTCACC GCGGGT----- 3
-----CGGCTGAGCTCACC GCGGGTC----- 1
-----TCATGTTCTCGATCAGCTGCG----- 14
-----TCATGTTCTCGATCAGCTGCGT----- 11
-----TTCTCGATCAGCTGCGTAGC----- 4
-----TTCTCGATCAGCTGCGTAGCC----- 183
-----TTCTCGATCAGCTGCGTAGCCT----- 11
-----TTCTCGATCAGCTGCGTAGCCTT----- 4
-----TCTCGATCAGCTGCGTAGCC----- 2
-----TCGATCAGCTGCGTAGCCTTG----- 1
-----TCGATCAGCTGCGTAGCCTTGG----- 4
-----TCAGCTGCGTAGCCTTGGGAGC----- 11
-----TGCGTAGCCTTGGGAGCCGGCA----- 4
-----TGCGTAGCCTTGGGAGCCGGC----- 56
-----GCGTAGCCTTGGGAGCCGGCA----- 1
-----GCGTAGCCTTGGGAGCCGGC----- 2
-----TAGCCTTGGGAGCCGGCAGGG----- 6
-----TAGCCTTGGGAGCCGGCAGGGC----- 1
-----TAGCCTTGGGAGCCGGCAGG----- 1
-----TTGGGAGCCGGCAGGGCCAGC----- 51
-----TGGGAGCCGGCAGGGCCAGCG----- 2
-----TGAGCTGCAGGGTGGTCATGT----- 4
-----TGAGCTGCAGGGTGGTCATGTGG----- 3
-----TGAGCTGCAGGGTGGTCATG----- 18
-----GAGCTGCAGGGTGGTCATGTGG----- 1
-----AGCTGCAGGGTGGTCATGTGG----- 1
-----TGCAGGGTGGTCATGTGGGACA----- 5
-----TGCAGGGTGGTCATGTGGG----- 2
-----TGCAGGGTGGTCATGTGGGA----- 4
-----TGCAGGGTGGTCATGTGGGAC----- 341
-----GCAGGGTGGTCATGTGGGAC----- 3
-----CAGGGTGGTCATGTGGGACAG----- 1
-----CAGGGTGGTCATGTGGGACA----- 1
-----TGGTCATGTGGGACAGATCCCC----- 2
-----TGGTCATGTGGGACAGATCCC----- 21
-----GGTCATGTGGGACAGATCCC----- 1
-----TGTGGGACAGATCCCCAACGG----- 4
-----TGGGACAGATCCCCAACGGCT-G----- 1
-----TGGGACAGATCCCCAACGGCT----- 15
-----TGGGACAGATCCCCAACGGC----- 2
-----GGGACAGATCCCCAACGGCT----- 1

-----C C TAC C A T TTC C T C - T ATT A CG
CCA TGCCAT CCACCC GCTGT GGCTG GCTCACCGCGGGTCAGGCGGG CATG TCGATCAGCTG GTAGCCT GGGAGCCGGCAGGGCCAGCGT AGCTGCAGGGTGGTCATGTGGGACAGATCCCCAA GGCT GCTG GCGG GGCCT GG G
GGT ACGGTA GGTGGG CGACA CCGAC CGATGGGCGCCAGTCCGCC GTAC AGTTAGTCGAC CATCGGA CCTCGGCCGTCCCGGTGCGCA TCGACGTCCCAACCAGTACACCCGTGTCTAGGGGTT CCGG CGGC GCGC CGGGG CC /
ACGCACGGCGCG A A CGT A G T CTA C C A T - AC- - AA

-----GACGTCCCACCAGTACACCCT-----3
-----CGACGTCCCACCAGTACACCCT-----1
-----TCGCAGTCGACGTCCCACCAGT-----16
-----GCAGTCGACGTCCCACCAGT-----59
-----CGCAGTCGACGTCCCACCAGT-----2415
-----GTCGCAGTCGACGTCCCACCAGT-----1
-----CGCAGTCGACGTCCCACCAG-----2
-----CGCAGTCGACGTCCCACCA-----1
-----CGCCCTGTACCTAAGTTAGT-----5
-----CGCCCTGTACCTAAGTTAG-----1
-----CCAGTCCGCCCTGTACCTAAGT-----1
-----CAACCGACGCGAGTGGCGCCCAGT-----1
-----CGACGCGAGTGGCGCCCAGT-----1
-----ACGGCGCGGGTAACGGTAAGGT-----1
-----CGGCGCGGGTAACGGTAAGGT-----1

Cluster: 17620+ chromosome_4 Start: 3100617 End: 3100762

-----TACAAGGCGGGCATGA-----1
-----TACAAGGCGGGCATGAAGGTCG-----16
-----TGCTAGCAACTACAAGG-----1
-----GCTGCTAGCAACTACAAG-----3
-----GCTGCTAGCAACTACAAGGCG-----6
-----GCTGCTAGCAACTACAAGGC-----173
-----GCTGCTAGCAACTACAAGG-----213
-----TGCTGCTAGCAACTACAAG-----30
-----TGCTGCTAGCAACTACAAGGCG-----247
-----TGCTGCTAGCAACTACAAGG-----8368
-----TGCTGCTAGCAACTACAAGGC-----5702
-----GTGCTGCTAGCAACTACAAGG-----1
-----AGTGCTGCTAGCAACTACAAG-----2
-----AGTGCTGCTAGCAACTACAAGGCG-----1
-----AGTGCTGCTAGCAACTACAAGGC-----1
-----AGTGCTGCTAGCAACTACAAGG-----10
-----CAGTGCTGCTAGCAACTACAAGG-----2
-----TTACTATCAGTGCTGCTAGC-----3
-----CGTATTACTATCAG-----1
-----TCGTATTACTATCAG-----1
-----TCGTATTACTATCAGTGCTG-----1
-----ATCGTATTACTATCAG-----1
-----CATCGTATTACTATC-----1
-----CATCGTATTACTATCAG-----1
-----CCATCGTATTACTATCAG-----7
-----GGCCATCGTATTACTATCA-----10
-----GGCCATCGTATTACTATC-----17
-----GGCCATCGTATTACTATCAG-----212
-----GGGCCATCGTATTACTATCAGT-----21
-----GGGCCATCGTATTACTAT-----5
-----GGGCCATCGTATTACTA-----1
-----GGGCCATCGTATTACTATC-----1023
-----GGGCCATCGTATTACTATCAG-----5992
-----GGGCCATCGTATTACTATCA-----283
-----GGGCCATCGTATTACT-----1
-----TGGGCCATCGTATTACTATCAG-----377890
-----TGGGCCATCGTATTACTAT-----382
-----TGGGCCATCGTATTAC-----7
-----TGGGCCATCGTATTACTA-----18
-----TGGGCCATCGTATTACTATC-----51461
-----TGGGCCATCGTATTACTATCAGT-----1018
-----TGGGCCATCGTATTACTATCA-----14287
-----TGGGCCATCGTATTACT-----42
-----TGGGCCATCGTATTACTATCAGTGC-----1
-----ATGGGCCATCGTATTACTATC-----1
-----AATGGGCCATCGTATTACTAT-----1

-----T T TA C G TG
GCAAA TGGGCCATCGTATTAC ATCAGTGCTGC GCAA TACAAGGCGG CATGAAGTTCGATT A
CGT T TACTCGGTAGCGTAATG TAGTCACGATG CGTT ATGTTCCGCC GTACTTCCAGCTAG C
GGAAAGCG T GC A A CC

---AGCGCGTTTACTCGGTAGCGT-----1
-----TTACTCGGTAGCGTAATG-----6
-----TTACTCGGTAGCGTAATGT-----22

-----CGTTTACTCGGTAGCGTAATGT----- 1
-----TACTCGGTAGCGTAATGTT----- 29
-----TTTACTCGGTAGCGTAATGTT----- 55
-----GTTTACTCGGTAGCGTAATGTT----- 1
-----TTACTCGGTAGCGTAATGTT----- 1405
-----TTACTCGGTAGCGTAATGTTA----- 3
-----TTACTCGGTAGCGTAATGTTAG----- 3
-----ACTCGGTAGCGTAATGTTAG----- 3
-----TACTCGGTAGCGTAATGTTAG----- 1
-----ACTCGGTAGCGTAATGTTAGT----- 194
-----TACTCGGTAGCGTAATGTTAGT----- 31
-----GGTAGCGTAATGTTAGT----- 1
-----CTCGGTAGCGTAATGTTAGT----- 1
-----TTACTCGGTAGCGTAATGTTAGT----- 115
-----ACTCGGTAGCGTAATGTTAGT----- 1
-----TAATGTTAGTCACGATGGCCGT----- 2
-----AATGTTAGTCACGATGGCCGT----- 23
-----TGTTAGTCACGATGGCCGTT----- 1
-----AATGTTAGTCACGATGGCCGTT----- 3
-----T TAGTCACGATGGCCGTTA----- 1
-----TTAGTCACGATGGCCGTTAA----- 1
-----TGTTAGTCACGATGGCCGTTAAT----- 2
-----TTAGTCACGATGGCCGTTAAT----- 6
-----AGTCACGATGGCCGTTAAT----- 2
-----TCACGATGGCCGTTAAT----- 1
-----ACGATGGCCGTTAATG----- 1
-----GTCACGATGGCCGTTAATG----- 1
-----TAGTCACGATGGCCGTTAATG----- 8
-----TTAGTCACGATGGCCGTTAATG----- 2
-----AGTCACGATGGCCGTTAATG----- 43
-----TCACGATGGCCGTTAATG----- 4
-----CGATGGCCGTTAATGT----- 8
-----TCACGATGGCCGTTAATGT----- 68
-----TTAGTCACGATGGCCGTTAATGT----- 105
-----CACGATGGCCGTTAATGT----- 1
-----GTCACGATGGCCGTTAATGT----- 41
-----ATGGCCGTTAATGT----- 1
-----TAGTCACGATGGCCGTTAATGT----- 248
-----AGTCACGATGGCCGTTAATGT----- 2894
-----GATGGCCGTTAATGT----- 1
-----TTAGTCACGATGGCCGTTAATGTT----- 24
-----GTCACGATGGCCGTTAATGTT----- 1419
-----CGATGGCCGTTAATGTT----- 112
-----AGTCACGATGGCCGTTAATGTT----- 2530
-----GATGGCCGTTAATGTT----- 7
-----CACGATGGCCGTTAATGTT----- 64
-----TAGTCACGATGGCCGTTAATGTT----- 115
-----ACGATGGCCGTTAATGTT----- 20
-----TCACGATGGCCGTTAATGTT----- 6892
-----TCACGATGGCCGTTAATGTTC----- 19
-----TTAATGTTCCGCCAGTACTT----- 1
-----GTTAATGTTCCGCCAGTACTT----- 10
-----TCCGCCAGTACTTCCAGCTAGC----- 2
-----TCCGCCAGTACTTCCAGCTAGCC----- 1

Cluster: 17755+ chromosome_4 Start: 3694773 End: 3694881
-----TATCACCCGTGTTACAGG----- 2
-----TTAGTCGTCACAAGGCGTG----- 3
-----ACTTAGTCGTCACAAGGCG----- 3
-----AACTTAGTCGTCACAAGGCG----- 23
-----TAACTTAGTCGTCACAAGGCGTG----- 11
-----TAACTTAGTCGTCACAAGGC----- 1
-----TAACTTAGTCGTCACAAGGCGT----- 1
-----TAACTTAGTCGTCACAAGGCG----- 1588
-----TGGTAACTTAGTCGTCACAAG----- 1
-----TGGTAACTTAGTCGTCACAAG----- 7
-----TGTAACGAGGGAAAAATGGTAA----- 2
-----TGTAACGAGGGAAAAATGGTA----- 34
GTCCTGTAACGAGGGAAAAATG A A T
GTAACCTTAGTCGTC CAAGGCGTGTATCACCCGTGTTAC GG C
CATTGAATCAGTAG GTTCCGCACATAGTGGGTACAATG CC C
----- G - C
-----GAATCAGTAGGGTTCCGCACAT----- 1

40 SI

-----TCTGCGCCGCACATGTCTCAGG-----	2
-----TCTGCGCCGCACATGTCTCAG-----	3
-----TCTGCGCCGCACATGTCTC-----	2

-----GACGTCCTGCGCCGCACATGTCTCAGGAGGA ATCG CC TA TAGCAAGCT CT AGGGTATCTCGGT CAGCATTTCG T TTGGCAAGATGTCCGCTTC GGTAC CCGCCGCTCA GCGATGA C AG	
-----CTGCAGACGTGGCGTGTACAGAGTCCTTCT TAGC GG AT ATCGTTCGA GA TCCCAT AGAGCCAGTCGTAGAGC AACCGTTCTGCAGGTGAGG CCATG GCGGCGAGT CGCTACT /	
-----GCCCGTTATGCT C A C- C C CC T C C A GG	
-----TATGCTCTGCAGACGTGGCGT-----	14
-----TTATGCTCTGCAGACGTGGCGT-----	29
-----TGCTCTGCAGACGTGGCGT-----	6
-----GTTATGCTCTGCAGACGTGGCGT-----	1
-----ATGCTCTGCAGACGTGGCGT-----	40
-----ATGCTCTGCAGACGTGGCGTG-----	2
-----GCTCTGCAGACGTGGCGTG-----	1
-----TTATGCTCTGCAGACGTGGCGTGT-----	33
-----ATGCTCTGCAGACGTGGCGTGT-----	1053
-----CTCTGCAGACGTGGCGTGT-----	3
-----TATGCTCTGCAGACGTGGCGTGT-----	12
-----GCTCTGCAGACGTGGCGTGT-----	26
-----TGCAGACGTGGCGTGT-----	1
-----GTTATGCTCTGCAGACGTGGCGTGT-----	1
-----TGCTCTGCAGACGTGGCGTGT-----	79
-----GGTTATGCTCTGCAGACGTGGCGTGT-----	1
-----TCTGCAGACGTGGCGTGT-----	1
-----AGACGTGGCGGTACAGAGT-----	21
-----GCAGACGTGGCGGTACAGAGT-----	17
-----CGTGGCGGTACAGAGTCCT-----	1
-----TGGCGGTACAGAGTCCTTCT-----	55
-----CGTGGCGGTACAGAGTCCTTCT-----	20
-----GTGGCGGTACAGAGTCCTTCT-----	17
-----GGCGGTACAGAGTCCTTCTCT-----	8
-----GTACAGAGTCCTTCTCTAGCAGGC-----	3
-----GAGTCCTTCTCTAGCAGGC-----	1
-----ACAGAGTCCTTCTCTAGCAGGC-----	20
-----CAGAGTCCTTCTCTAGCAGGC A-----	4
-----GAGTCCTTCTCTAGCAGGC A-----	6
-----ACAGAGTCCTTCTCTAGCAGGC A-----	1
-----GTCCTTCTCTAGCAGGC AT-----	1
-----CTTCTCTAGCAGGC AT-----	1
-----GAGTCCTTCTCTAGCAGGC AT-----	611
-----CCTTCTCTAGCAGGC AT-----	3
-----TCCTTCTCTAGCAGGC AT-----	1
-----AGTCCTTCTCTAGCAGGC AT-----	22
-----ACAGAGTCCTTCTCTAGCAGGC AT-----	115
-----AGAGTCCTTCTCTAGCAGGC AT-----	147
-----CAGAGTCCTTCTCTAGCAGGC AT-----	143
-----TACAGAGTCCTTCTCTAGCAGGC AT-----	3
-----CCTTCTCTAGCAGGC ATCATCGT-----	5
-----TTCTCTAGCAGGC ATCATCGT-----	1
-----CTCTAGCAGGC ATCATCGTT-----	2
-----TCTCTAGCAGGC ATCATCGTT-----	22
-----CCTTCTCTAGCAGGC ATCATCGTT-----	1
-----CATCGTTCGACGACCTC-----	1
-----CATCGTTCGACGACCTCCCA-----	2
-----TCATCGTTCGACGACCTCCCA-----	1
-----TCATCGTTCGACGACCTCCCAT-----	7
-----CATCGTTCGACGACCTCCCAT-----	632
-----C ATCATCGTTCGACGACCTCCCAT-----	1
-----TCGTTTCGACGACCTCCCAT-----	1
-----GGC ATCATCGTTCGACGACCTCCCAT-----	1
-----ATCATCGTTCGACGACCTCCCAT-----	1
-----CGACCTCCCATAGAGCCAGT-----	1
-----CGACGACCTCCCATAGAGCCAGT-----	2
-----GACGACCTCCCATAGAGCCAGT-----	1
-----ACGACCTCCCATAGAGCCAGT-----	3
-----ACCTCCCATAGAGCCAGTCGT-----	3
-----GACCTCCCATAGAGCCAGTCGT-----	1
-----TCCCATAGAGCCAGTCGTAGA-----	8
-----CATAGAGCCAGTCGTAGAGCT-----	3
-----GAGCCAGTCGTAGAGCTAA-----	1
----- AGAGCCAGTCGTAGAGCTAA -----	3
-----TAGAGCCAGTCGTAGAGCTAA-----	4
-----AGCCAGTCGTAGAGCTAACCG-----	2
-----GCCAGTCGTAGAGCTAACCG-----	2
-----AGCCAGTCGTAGAGCTAACCGT-----	9
-----GCCAGTCGTAGAGCTAACCGT-----	100

-----AGTCGTAGAGCTAACCGTTCT-----4
-----AGAGCTAACCGTTCTGCAGGT-----1
-----TAAACCGTTCTGCAGGTGAGGC-----1
-----CTAACCGTTCTGCAGGTGAGGC-----2
-----GCTAACCGTTCTGCAGGTGAGGC-----2
-----CGTTCTGCAGGTGAGGCC-----1
-----CCGTTCTGCAGGTGAGGCC-----1
-----CGTTCTGCAGGTGAGGCC-----25
-----CCGTTCTGCAGGTGAGGCC-----19
-----CGTTCTGCAGGTGAGGCCA-----120
-----CCGTTCTGCAGGTGAGGCCA-----112
-----GTTCTGCAGGTGAGGCCAT-----78
-----TCTGCAGGTGAGGCCAT-----1
-----ACCGTTCTGCAGGTGAGGCCAT-----44
-----GCAGGTGAGGCCAT-----1
-----CCGTTCTGCAGGTGAGGCCAT-----12535
-----CGTTCTGCAGGTGAGGCCAT-----15493
-----CTGCAGGTGAGGCCAT-----1
-----TTCTGCAGGTGAGGCCAT-----4

Cluster: 19166+ chromosome 6 Start: 2201619 End: 2201697
-----TAAGATTGCGGTAA--AGAT 2
-----TGTAAGATTGCGGTAA--AGAT 15
-----TTGTAAGATTGCGGTAA--AGA- 5
-----GGGCGGCGTTGTAAGATT-----62
-----TGGGCGGCGTTGTAAGATTGC-----1
-----TGGGCGGCGTTGTAAGAT-----2
-----TGGGCGGCGTTGTAAGATT-----533
-----TGGGCGGCGTTGTAAGATTG-----5
-----TTGGGCGGCGTTGTAA-----7
-----TTGGGCGGCGTTGTAAGAT-----337
-----TTGGGCGGCGTTGTAAGATTG-----104
-----TTGGGCGGCGTTGTAAGATT-----86172
-----TTGGGCGGCGTTGTAAAG-----69
-----TTGGGCGGCGTTGTAAAG-----45

-----A---A
CGAGTTGGGCGGCGTTGTAAGATTGCGGTAAGT
GCTCAACCCGCGCAACATTCTAACGCCATTCG
CCGTCGGTG G

---CGCTCAACCCGCCGCAA-----1
---CGCTCAACCCGCCGCAACAT-----7
---GCTCAACCCGCCGCAACATT-----1
---TCAACCCGCCGCAACATT-----1
---CGCTCAACCCGCCGCAACATT-----23
---TCAACCCGCCGCAACATT-----4
---CAACCCGCCGCAACATT-----1
---CGCTCAACCCGCCGCAACATTCT-----1
---GCTCAACCCGCCGCAACATTCT-----1
---TCAACCCGCCGCAACATTCT-----1053
---CAACCCGCCGCAACATTCT-----17
---CTCAACCCGCCGCAACATTCT-----13

Cluster: 19166- chromosome 6 Start: 2201614 End: 2201692
-----TTGGGCGGCGTTGTAA-----7
-----TTGGGCGGCGTTGTAAAG-----69
-----TTGGGCGGCGTTGTAAAGATT-----86172
-----TTGGGCGGCGTTGTAAAGAT-----337
-----TTGGGCGGCGTTGTAAAG-----45
-----TTGGGCGGCGTTGTAAAGATTG-----104
-----TGGGCGGCGTTGTAAAGATT-----533
-----TGGGCGGCGTTGTAAAGAT-----2
-----TGGGCGGCGTTGTAAAGATTG-----5
-----TGGGCGGCGTTGTAAAGATTGC-----1
-----GGGCGGCGTTGTAAAGATT-----62
-----CCACAG
CGAGTTGGGCGGCGTTGTAAGATTGCGGTA C
GCTCAACCCGCGCAACATTCTAACGCCAT /
CCGTC TTTTT
-----GCTCAACCCGCCGCAACATTCT-----1
-----TCAACCCGCCGCAACATTCT-----1053
-----CTCAACCCGCCGCAACATTCT-----13
-----CAACCCGCCGCAACATTCT-----17

---CGCTCAACCCGCCGCAACATTCT---	1
---CAACCCGCCGCAACATTCC---	1
---TCAACCCGCCGCAACATTCC---	4
---GCTCAACCCGCCGCAACATT---	1
---CGCTCAACCCGCCGCAACATT---	23
---TCAACCCGCCGCAACATT---	1
---CGCTCAACCCGCCGCAACAT---	7
---CGCTCAACCCGCCGCAA---	1

[illegible]

GGACTGGGTTGGCGCTGGCG	1
TGGACTGGGGTGGCGCTGGCGT	2
GGACTGGGGTGGCGCTGGCGT	66
GC	33
GC	10
GC	1
GC	1
GC	11
GC	4
GC	1
GC	1
GC	7
GC	9
GC	7
GC	2
GC	7
GC	2
GC	36
GC	61
GC	580
GC	3
GC	1
GC	208
GC	74
GC	70
GC	26
GC	1
GC	1
GC	8943
GC	307549
GC	2
GC	3
GC	3
GC	3

Cluster: 20399+ chromosome 6 Start: 6776059 End: 6776320

44 SI

-----AAGAAGCTCGAGAT-----3
-----GGAGGAAGGGTTCAAGAAGCTCGAGAT-----2
-----GAGGAAGGGTTCAAGAAGCTCGAGAT-----13
-----GTTCAAGAAGCTCGAGAT-----28
-----AAGGGTTCAAGAAGCTCGAGAT-----1095
-----GAAGGGTTCAAGAAGCTCGAGAT-----122429
-----AGGGTTCAAGAAGCTCGAGAT-----362932
-----CAAGAAGCTCGAGAT-----10
-----GGTTCAAGAAGCTCGAGAT-----1565
-----GGAAGGGTTCAAGAAGCTCGAGATG-----1
-----GGTTCAAGAAGCTCGAGATGG-----10
-----CAAGAAGCTCGAGATGGGCT-----1
-----TCAAGAAGCTCGAGATGGGCT-----1
-----TCAAGAAGCTCGAGATGGGCTT-----8
-----CAAGAAGCTCGAGATGGGCTT-----8
-----GCTCGAGATGGGCTTGCAGCT-----5
-----AGATGGGCTTGCAGCTCAAGT-----1

Cluster: 20462- chromosome_6 Start: 7063790 End: 7063888
-----AAGACGGCGCACCTTCTTAAC-----1

-----CAC
--CGTCAAGACGGCGCACCTTCTTAACGTCCTTCGGTGCTTGCTTG A
GCAGTTCTGTCGCGTGGAAGAATTGCGGGAAGCCACGGACGAAC T
TG TTT

---GTTCTGTCGCGTGGAAGAATT-----5
---TCTGTCGCGTGGAAGAATT-----2
---AGTTCTGTCGCGTGGAAGAATT-----1
---GTTCTGTCGCGTGGAAGAAT-----45
---AGTTCTGTCGCGTGGAAGAAT 1701
---AGTTCTGTCGCGTGGAAGAA-----14
---AGTTCTGTCGCGTGGAAGA-----5
---GTTCTGTCGCGTGG-----1
---AGTTCTGTCGCGTGG-----3

Cluster: 21642+ chromosome_7 Start: 2713436 End: 2713688
-----GACGCGGTGATGGGCCTGG-----10
-----CGACGCGGTGATGGGCCTGG-----9
-----TCGACGCGGTGATGGGCCTGGT-----49
-----TCGACGCGGTGATGGGCCTGG 2702
-----TTCGACGCGGTGATGGGCCTGG-----67
-----TGCTGCGAGATGCAGCAGGCCAGG-----2
-----ACTCCCATGTGCTGCGAGATG-----2
-----TGGAACTCCCATGTGCTGCGAG-----1
-----TATGGAACTCCCATGTGCTGC-----28
-----TCGACGATGCGGTATGGAAC-----19
-----TCGACGATGCGGTATGGAATC-----25
CTTGTCGACG A A GCA AT-- A C A-- CGCA
ATGCGGT TGGAATCCCATGTGCTGCG GATGCA GGCCAGGGCCATT GACAC AGTAACGATAT CATTACTTCGACGCGGTGATGGGCCTGGTCCGTCACAGGAC GGG C
TACGCTA GTCTTGAGGGTACACGACGC CTACGT CCGTCCCGGTAA TTGTG TCGTTGCTGTA GTAATGAAGCTGCGCCATTATCCGGACCAGGCAGTGTCTTG CCC /
----- A G --- ATTC A C GCA AAAA
-----CTAAGTCTTGAGGGTACACGA-----2
-----CTTGAGGGTACACGACGCGCT-----20
-----GGTCCCGGTAAATTCTTGTGAT-----1
-----GTGATCGTTGCTGTACGTAA-----1
-----GTGATCGTTGCTGTACGTAAT-----36
-----GAAGCTGCGCCATTATCCGGAG-----1

Cluster: 22587+ chromosome_8 Start: 121853 End: 121958
-----TCGGTCCTAGGCGGCGCGTG-----3
-----TCGGTCCTAGGCGGCGCGTGT-----1
-----TTTCGTGCGTCCTAGGCGGC-----7
-----TTTCGTGCGTCCTAGG-----1
-----GGCTTTTCGTGCGTCCTAGGC-----1
-----GGCTTTTCGTGCGTCCTAG-----1
-----GGCTTTTCGTGCGTCCTAGG-----127
-----TGCGTTTCGTGCGTCCTAGG 26418
-----TGGCTTTTCGTGCGTCCTAG-----221
-----TGGCTTTTCGTGCGTCCTAGGC-----397
-----AGTGGCTTTTCGTGCGTCCTAG-----3
-----TAGTGGCTTTTCGTGCGTCCTAGG-----15
-----TAGTGGCTTTTCGTGCGTCCTA-----3

```
-----TAGTGGCTTTCGTCGGT-----1
-----TAGTGGCTTTCGTCGGTCCTAG-----1
-----TTAGTGGCTTTCGTCGGTCCTAG-----6
-----TTAGTGGCTTTCGTCGGT-----1
-----TTAGTGGCTTTCGTCGGTCCTAGG-----4
-----TTAGTGGCTTTCGTCGGTCCTA-----46
-----TTAGTGGCTTTCGTCGGTCCT-----13

-----GAC
CAGCTTAGTGGCTTTCGTCGGTCCTAGGCGGCGCGTGTGCGCGG G
GTCGAATCACCgAAAGCAGCCAGGATCCGCCGCGCACACGCGCGC A
AGCACTGAAA

-----TCACCGAAAGCAGCCAGG-----6
-----ATCACCGAAAGCAGCCAGG-----4
-----CACCGAAAGCAGCCAGGA-----1
-----AATCACCGAAAGCAGCCAGGA-----2
-----CGAAAGCAGCCAGGA-----2
-----TCACCGAAAGCAGCCAGGA-----119
-----ATCACCGAAAGCAGCCAGGA-----32
-----GGTCGAATCACCGAAAGCAGCCAGGAT-----3
-----ATCACCGAAAGCAGCCAGGAT-----1513
-----TCACCGAAAGCAGCCAGGAT-----8566
-----GAATCACCGAAAGCAGCCAGGAT-----8
-----AATCACCGAAAGCAGCCAGGAT-----69
-----TCGAATCACCGAAAGCAGCCAGGAT-----2
-----CACCGAAAGCAGCCAGGAT-----62
-----CGAAAGCAGCCAGGAT-----33
-----GTCTGAATCACCGAAAGCAGCCAGGAT-----1
-----TGGTCGAATCACCGAAAGCAGCCAGGAT-----2
-----CTGGTCGAATCACCGAAAGCAGCCAGGAT-----1
-----CGAATCACCGAAAGCAGCCAGGAT-----4
-----AAAGCAGCCAGGAT-----7
-----CCGAAAGCAGCCAGGAT-----48
-----ACCGAAAGCAGCCAGGAT-----1
-----GAAAGCAGCCAGGAT-----6
-----AATCACCGAAAGCAGCCAGGATC-----1
-----TCACCGAAAGCAGCCAGGATC-----4

Cluster: 22587- chromosome 8 Start: 121846 End: 121967
---TGATGTCGTGACCACTTAGTGGCTTTC-----1
---GTCGTGACCACTTAGTGGCTTTC-----2
-----TTAGTGGCTTTCGTCGGTCCTAG-----6
-----TTAGTGGCTTTCGTCGGTCCTAGG-----4
-----TTAGTGGCTTTCGTCGGTCCT-----13
-----TTAGTGGCTTTCGTCGGT-----1
-----TTAGTGGCTTTCGTCGGTCCTA-----46
-----TAGTGGCTTTCGTCGGTCCTAG-----1
-----TAGTGGCTTTCGTCGGTCCTA-----3
-----TAGTGGCTTTCGTCGGT-----1
-----TAGTGGCTTTCGTCGGTCCTAGG-----15
-----AGTGGCTTTCGTCGGTCCTAG-----3
-----TGGCTTTCGTCGGTCCTAGG-----26418
-----TGGCTTTCGTCGGTCCTAG-----221
-----TGGCTTTCGTCGGTCCTAGGC-----397
-----GGCTTTCGTCGGTCCTAGGC-----1
-----GGCTTTCGTCGGTCCTAG-----1
-----GGCTTTCGTCGGTCCTAGG-----127
-----TTCGTCGGTCCTAGG-----1
-----TTTCGTCGGTCCTAGGCGGC-----7
-----TCGGTCCTAGGCGGCGCGTGT-----1
-----TCGGTCCTAGGCGGCGCGTG-----3

GGCTTGATGTC TCC
GTGACCACTTAGTGGCTTTCGTCGGTCCTAGGCGGCGGTGTGCGCGG G
CACTGTCGAATCACCgAAAGCAGCCAGGATCCGCCGCGCACACGCGCGC C
-----CA CCC

-----AATCACCGAAAGCAGCCAGGATC-----1
-----TCACCGAAAGCAGCCAGGATC-----4
-----GTCGAATCACCGAAAGCAGCCAGGAT-----1
-----ACCGAAAGCAGCCAGGAT-----1
-----AAAGCAGCCAGGAT-----7
-----CGAATCACCGAAAGCAGCCAGGAT-----4
-----GAAAGCAGCCAGGAT-----6
-----CTGGTCGAATCACCGAAAGCAGCCAGGAT-----1
-----AATCACCGAAAGCAGCCAGGAT-----69
```


-----CGAAAGCAGCCAGGAT-----48
-----TCAATCACCAGAAAGCAGCCAGGAT-----2
-----ATCACCAGAAAGCAGCCAGGAT-----1513
-----GAATCACCAGAAAGCAGCCAGGAT-----8
-----TGGTCGAATCACCAGAAAGCAGCCAGGAT-----2
-----GGTCGAATCACCAGAAAGCAGCCAGGAT-----3
-----CACCAGAAAGCAGCCAGGAT-----62
-----CGAAAGCAGCCAGGAT-----33
-----TCACCAGAAAGCAGCCAGGAT-----8566
-----TCACCAGAAAGCAGCCAGGA-----119
-----CACCAGAAAGCAGCCAGGA-----1
-----AATCACCAGAAAGCAGCCAGGA-----2
-----CGAAAGCAGCCAGGA-----2
-----ATCACCAGAAAGCAGCCAGGA-----32
-----ATCACCAGAAAGCAGCCAGG-----4
-----TCACCAGAAAGCAGCCAGG-----6

Cluster: 25047- chromosome_9 Start: 6365819 End: 6366048

-----TTCTAGTGCCTGGCTGGGATC-----13
-----TCTAGTGCCTGGCTGGGATC-----1
-----TAGTGCCTGGCTGGGATCAATGC-----6
-----TAGTGCCTGGCTGGGATCAA-----3
-----TAGTGCCTGGCTGGGATCAAT-----31
-----TAGTGCCTGGCTGGGATCAATG-----1
-----TAGTGCCTGGCTGGGATC-----1
-----TAGTGCCTGGCTGGG-----1
-----TGCCTGGCTGGGATCAATGCAAG-----5
-----CCTGGCTGGGATCAATGCAAG-----1
-----CCTGGCTGGGATCAATGCAAGA-----2
-----TGGCTGGGATCAATGCAAGATGAT-----24
-----TGGCTGGGATCAATGCAAGAT-----5
-----TGGCTGGGATCAATGCAAGA-----7
-----TGGCTGGGATCAATGCAAGATGA-----11
-----TGGCTGGGATCAATGCAAG-----5
-----TGGCTGGGATCAATGCAAGATG-----38
-----GGCTGGGATCAATGCAAGATG-----1
-----TGGGATCAATGCAAGA-----1
-----TGGGATCAATGCAAGATGATA-----66
-----TGGGATCAATGCAAGATGAT-----20
-----TGGGATCAATGCAAGATGATAG-----1
-----TCAATGCAAGATGATAGTTGG-----1
-----AATGCAAGATGATAGTTGGATC-----1
-----TGCAAGATGATAGTTGGATCC-----1893
-----TGCAAGATGATAGTTGGATC-----14
-----TGCAAGATGATAGTTGGATCCT-----9
-----GCAAGATGATAGTTGGATCC-----2
-----ATGATAGTTGGATCCTGGA-----1
-----ATGATAGTTGGATCCTGGACG-----1
-----ATGATAGTTGGATCCTGGACG-----1759
-----TGATAGTTGGATCCTGGACG-----3113
-----TGATAGTTGGATCCTGGACG-----32
-----TGATAGTTGGATCCTGG-----1
-----TGATAGTTGGATCCTGGA-----1
-----TGATAGTTGGATCCTGGAC-----1
-----TGATAGTTGGATCCTGGACGCC-----5
-----GATAGTTGGATCCTGGACG-----58
-----ATAGTTGGATCCTGGACGCCG-----2
-----ATAGTTGGATCCTGGACG-----1
-----TAGTTGGATCCTGGACG-----399
-----TAGTTGGATCCTGGAC-----1
-----TAGTTGGATCCTGGACGCC-----2
-----TAGTTGGATCCTGGACGCCG-----1
-----TAGTTGGATCCTGGACG-----3
-----AGTTGGATCCTGGACG-----1
-----TTGGATCCTGGACGCCGAAGGC-----7
-----TGGATCCTGGACGCCGAAGGC-----88
-----TGGATCCTGGACGCCGAAGGCC-----3
-----GGATCCTGGACGCCGAAGGC-----1
-----TGGACGCCGAAGGCCATC-----1
-----TGGACGCCGAAGGCCATCATG-----85
-----TGGACGCCGAAGGCCATCATGCG-----1
-----TGGACGCCGAAGGCCATCATGC-----6

-----T-----G-----A-----A-----G-----
CATTTTC AGTGCTGGCTGGGATCAAT G CAAGATGATAGTTGGATCCTGGACGCCGA GGCC TCATGCGGT T
GTAGAGG TC GCGGACCGACCCTAGTTA GTTCTGCTGTCGATCTAGGACCTGCGGTT CCGG AGTACGCCA A
GGGACCTTCGGTTGCTGCCACCACGGTAGCAGCGGCCAGACGAGCCTTGGGTCGGTTGTTCCCTCCCTTCTGGCACCCAG T G C A

A. Voshall *et al.*

47 SI

-----	-----GGACCTGCGGTTGCCGGCAGT-----	20
-----	-----GTCGATCTAGGACCTGCGGTT-----	1
-----	-----TTCTGCTGTCGATCTAGGACCT-----	27
-----	-----CTGTCGATCTAGGACCT-----	2
-----	-----GTTCTGCTGTCGATCTAGGACCT-----	1
-----	-----TGCTGTCGATCTAGGACCT-----	2
-----	-----TCTGCTGTCGATCTAGGACCT-----	4656
-----	-----CTGCTGTCGATCTAGGACCT-----	210
-----	-----GTCGATCTAGGACCT-----	1
-----	-----GCTGTCGATCTAGGACCT-----	4
-----	-----TCTGCTGTCGATCTAGGAC-----	18
-----	-----TCTGCTGTCGATCTAGGA-----	1
-----	-----TCTGCTGTCGATCTAGG-----	1
-----	-----AGTTAGGTTCTGCTGTCGATCT-----	7
-----	-----ACCCTAGTTAGGTTCTGCTGT-----	1
-----	-----CGACCCTAGTTAGGTTCTGCT-----	195
-----	-----CGACCCTAGTTAGGTTCTGC-----	1
-----	-----GACCGACCCTAGTTAGGTTCT-----	1
-----	-----GGACCGACCCTAGTTAGGTTCT-----	51
-----	-----GACCGACCCTAGTTAGGTTCT-----	1
-----	-----GGACCGACCCTAGTTAGGTTCT-----	2
-----	-----CGGACCGACCCTAGTTAGGTT-----	2
-----	-----GGACCGACCCTAGTTAGGTT-----	10
-----	-----GCGGACCGACCCTAGTTAGGTT-----	2
-----	-----CGACCCTAGTTAGGT-----	8
-----	-----GGTTCGCGGACCGACCCTAGTTAGGT-----	9
-----	-----GCGGACCGACCCTAGTTAGGT-----	80553
-----	-----GACCCTAGTTAGGT-----	1
-----	-----CGGACCGACCCTAGTTAGGT-----	11
-----	-----GACCGACCCTAGTTAGGT-----	7
-----	-----GGACCGACCCTAGTTAGGT-----	48
-----	-----CGCGGACCGACCCTAGTTAGGT-----	10
-----	-----GGACCGACCCTAGTTAGG-----	1
-----	-----GCGGACCGACCCTAGTTAGG-----	1072
-----	-----GTTTCGCGGACCGACCCTAGTT-----	1
-----	-----CCTCCCTTCTGGCACCCAGGT-----	1
-----	-----CCCTCCCTTCTGGCACCCAGGT-----	1
-----	-----GGTCCGTTGTTCCCTCCCTT-----	1
-----	-----GGGTCCGTTGTTCCCTCCCTT-----	1
-----	-----CCTCGGTTGCTGCCACCACGGT-----	39
-----	-----CCTCGGTTGCTGCCACCACG-----	2

Cluster: 25074+	chromosome_9	Start: 6518807	End: 6519095	-----TTAGAGCCTGTGGGGCTGACCTG-----	1
-----	-----	-----	-----	-----TGCGCACAGCCCTTAGAGCCT-----	3
-----	-----	-----	-----	-----TGCGCACAGCCCTTAGAGCCTG-----	3
-----	-----	-----	-----	-----TGCGCACAGCCCTTAGAGC-----	1
-----	-----	-----	-----	-----TGCGCACAGCCCTTAGAGCC-----	7
-----	-----	-----	-----	-----GGACGCGGACCCGGCGCA-----	2
-----	-----	-----	-----	-----GGACGCGGACCCGGCGCAGCG-----	1
-----	-----	-----	-----	-----GGACGCGGACCCGGCGCAGC-----	12
-----	-----	-----	-----	-----GGACGCGGACCCGGCGCAGCG-----	3
-----	-----	-----	-----	-----GGACGCGGACCCGGCGCAG-----	35
-----	-----	-----	-----	-----TGGACGCGGACCCGGCGCA-----	9
-----	-----	-----	-----	-----TGGACGCGGACCCGGCGCAGC-----	1329
-----	-----	-----	-----	-----TGGACGCGGACCCGGCGCAGCG-----	275
-----	-----	-----	-----	-----TGGACGCGGACCCGGCGCAGCGCG-----	3
-----	-----	-----	-----	-----TGGACGCGGACCCGGC-----	6
-----	-----	-----	-----	-----TGGACGCGGACCCGGCGC-----	9
-----	-----	-----	-----	-----TGGACGCGGACCCGGCGCAGCGGC-----	1
-----	-----	-----	-----	-----TGGACGCGGACCCGGCGCAGCGG-----	155
-----	-----	-----	-----	-----TGGACGCGGACCCGGCGCAG-----	2561
-----	-----	-----	-----	-----TTGGACGCGGACCCGGCGCA-----	537
-----	-----	-----	-----	-----TTGGACGCGGACCCG-----	4
-----	-----	-----	-----	-----TTGGACGCGGACCCGGC-----	51
-----	-----	-----	-----	-----TTGGACGCGGACCCGGCGCAG-----	26633
-----	-----	-----	-----	-----TTGGACGCGGACCCGGCGCAGCGGC-----	1
-----	-----	-----	-----	-----TTGGACGCGGACCCGGCG-----	24
-----	-----	-----	-----	-----TTGGACGCGGACCCGGCGCAGC-----	267
-----	-----	-----	-----	-----TTGGACGCGGACCCGG-----	3
-----	-----	-----	-----	-----TTGGACGCGGACCCGGCGCAGCG-----	17
-----	-----	-----	-----	-----TTGGACGCGGACCCGGCGC-----	186
-----	-----	-----	-----	-----TTGGACGCGGACCCGGCGCAGCGG-----	8
-----	-----	-----	-----	-----AAGTTGGACGCGGACCCGGCG-----	2
-----	-----	-----	-----	-----AGGCACCGCAACGCACCAAGT-----	1
-----	-----	-----	-----	-----AGGCACCGCAACGCACCAAG-----	4

49 SI

Figure S4 Secondary structure plots of predicted miRNA precursor loci and frequency of reads from the Chlamydomonas AGO3-associated sRNA library. Mature miRNAs are indicated in red and predicted passenger strands (miRNA*) are indicated in blue. Each plot is annotated with the locus ID (cluster), the strand (+ or -), the chromosome, and the chromosomal position (start and end). For loci where the miRNA is processed from a region of the hairpin with perfectly complementary arms, the folded structure for each strand of the locus is shown. Loci with long hairpins may code for more than one candidate miRNA. Additionally, some loci code for more than one hairpin in a tandem arrangement (clustered miRNA precursor loci) and, in these cases, the folding of each hairpin is shown separately.

Figure S6 Analysis of the miR C target *Cre16.g683650* in a mutant (Dcl-3) defective in the *g8048* gene, encoding *Chlamydomonas* Dicer-like 3. (A) Diagram of the *Cre16.g683650* gene indicating the miR C-binding site overlapping the stop codon. Hybridization of miR C to the target site is shown below. The mismatch to nucleotide 10 of the miRNA would prevent cleavage of the target mRNA. (B) Quantitative RT-PCR analysis of *Cre16.g683650* transcript levels in the indicated strains. Control refers to the parental strain of the Dcl-3 mutant. Values, normalized to those in the control strain, are means $\pm$ SD of three independent experiments. Samples are significantly different ($P < 0.05$) in a two tailed Student's *t*-test. (C) Immunoblot analyses of *Cre16.g683650* and histone H3 (loading control) protein levels. The graph represents densitometric quantitation of the *Cre16.g683650* immunoblot signals (normalized to H3). Values are means $\pm$ SD of three independent experiments. Samples are significantly different ($P < 0.05$) in a two tailed Student's *t*-test. (D) Northern blot analysis of small RNAs isolated from the indicated strains and detected with a probe specific for miR C. The same filter was reprobbed with the U6 small nuclear RNA sequence as a control for equivalent loading of the lanes.

Tables S1-S4

Available for download as Excel files at <http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.115.174797/-/DC1>.

Table S1 MicroRNA binding sites and expression of transcripts predicted to be regulated by cleavage.

Table S2 Differential expression analysis comparing an RNAi defective mutant (Mut-20) and its parental strain (Maa7-IR44s)

Table S3 MicroRNA binding sites and expression of transcripts predicted to be regulated by translational repression.

Table S4 AGO3-associated small RNAs, matching to the Chlamydomonas nuclear genome, clustered by genomic location