Stephan Steigele:
Computer assisted analysis of RNA-based cellular regulation

The aim of this thesis was to develop and utilize different computational methods. They were used for the analysis and characterization of the abundance and evolution of natural antisense transcripts (NATs), the abundance of ncRNAs in yeasts and the reconstuction of repetitive elements from shotgun sequence datasets.

First, a method was developed to measure the abundance of natural antisense sequences in fungal genomes. The method was applied to open reading frames (ORFs) of four ascomycete and one microsporidian fungi. It explicitely allowed the detection of *cis-* and *trans-NATs*. We found similar numbers of *cis-NATs* in *A. gossypii*, *S. pombe* and *N. crassa*, but not in *S. cerevisiae* and *E. cuniculi*, where much higher numbers were found. *Trans-NATs* were also found and large differences in the number of involved ORFs were observed. Compared to other methods, the description of antisense relations was expanded to the analysis of antisense networks, illustrating the high interconnectivity of antisense transcripts. Here, we found significant groupings of specific functional classification terms, which are related to sub-networks of the antisense network.
In addition, a method was developed for the comparative analysis of antisense transcripts in human and other multicellular species. Results show, that many antisense transcripts are exclusively overlapping in human and that the number of conserved antisense overlaps is very low. It is concluded, that genomes in general contain unique sets of antisene transcripts with a definable functionality.

Regarding the abundance of complementary sequences in open reading frames, a method was developed considering three aspects: the genetic code, protein coding regions and antisense sequences. Based on the enumeration of peptide words in different sequences, a vocabulary of words was devised, that is responsible for potential antisense regions in protein coding sequences. Examination of the codon usage shows a clear relationship between the number of antisense transcripts and the codon usage.

A comparative search for conserved signals of RNA in whole genomes was performed. The methods focus was on the annotation and classification of sequences with RNA structure. The search in yeasts showed a large number of signals indicative for structured RNAs. Structured RNAs were found not only in intergenic regions, but also in coding regions and untranslated regions of coding sequences. Functional annotation of these CDS, provides significant clustering of functional terms. Further functional groups were found for coding sequences with structured regions in UTRs. Annotation of intergenic structured RNAs showed a limited homology to known elements and to sequences outside the yeast phylum.

As a last point, a methodology for the detailed reconstruction of repetitive elements was developed. It is able to reconstruct functional genes by the use of shotgun sequence data. The method was applied to sequences of a nematode species. A large number of repeated elements from known coding and noncoding genes was found. A detailed analysis of one repetitive element (*PpmaT1*)

was conducted. It showed, that the presented repeat reconstruction method is capable to reconstruct almost all features of an repetitive element.