



Stockholm, 28. November 2017

## Master Thesis Project Proposal

### Coding for DNA-Based Information Storage

**Motivation** – DNA is well known for its storage capability of biological information. DNA is a very rich storage medium that can be stored effectively over very long durations in time with very low maintenance costs. The fact that DNA sequencing is getting cheap has recently triggered investigations on DNA-based information storage system (see, e.g., [5] for a survey), and in proof-of-concept implementations (see, e.g., [1-4]) storage densities of 2 PB/g (petabytes per gram) have been reached. These features make DNA attractive for long-term archival systems. From a communications engineering perspective, very simple solutions have been adopted in [1-4]. And even though more rigorous approach based on coding and information theory was taken in, e.g., [6], it appears that coding for DNA-based information is new research area that offers many exiting research opportunities.

**Purpose** – The goal of this project is to study existing information theoretic models for the DNA storage channel and to develop new concepts (i.e., coding schemes) for the DNA storage channel. Depending on the competence of the candidate, the project can be purely theoretical or combine theoretical work with simulation studies based on the established models.

**Requirements** - The candidate is expected to have excellent theoretic knowledge within at least one of the fields of information, communication, and coding theory. Good programming skills and a high degree of independence are necessary.

**Contact**      **Ragnar Thobaben**  
Information Science and Engineering  
KTH – School of Electrical Engineering  
Email: [ragnart@kth.se](mailto:ragnart@kth.se)

#### References

- [1] G. M. Church, Y. Gao, and S. Kosuri, "Next-generation digital information storage in DNA," *Science*, vol. 337, no. 6102, p. 1628, 2012.
- [2] N. Goldman et al., "Towards practical, high-capacity, low-maintenance information storage in synthesized DNA," *Nature*, vol. 494, no. 7435, pp. 77–80, 2013.
- [3] R.N. Grass, R. Heckel, M. Puddu, D. Paunescu, and W.J. Stark, "Robust chemical preservation of digital information on DNA in silica with error-correcting codes," *Angew. Chem. Int. Ed.*, vol. 54, no. 8, pp. 2552–2555, 2015.
- [4] S. Yazdi, Y. Yuan, J. Ma, H. Zhao, and O. Milenkovic, "A rewritable, random-access DNA-based storage system," *Sci. Rep.*, vol. 5, no. 14138, 2015.
- [5] S. M. H. T. Yazdi, H. M. Kiah, E. Garcia-Ruiz, J. Ma, H. Zhao and O. Milenkovic, "DNA-Based Storage: Trends and Methods," in *IEEE Transactions on Molecular, Biological and Multi-Scale Communications*, vol. 1, no. 3, pp. 230-248, Sept. 2015.
- [6] H. M. Kiah, G. J. Puleo and O. Milenkovic, "Codes for DNA Sequence Profiles," in *IEEE Transactions on Information Theory*, vol. 62, no. 6, pp. 3125-3146, June 2016.