UNIVERSITY OF CAMBRIDGE

4TH YEAR PROJECT

A Systematic Approach to Promoter Discovery for the Automated Compilation of a Parts Registry

Author:
Haydn KING

Supervisor: Dr. Jorge
GONÇALVES

A thesis submitted in fulfilment of the requirements
for the degree of MEng

in the

Department of Engineering

September 2012

Declaration of Authorship

I, Haydn KING, declare that this thesis titled, 'A Systematic Approach to Promoter Discovery for the Automated Compilation of a Parts Registry' and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.

- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.

- Where I have consulted the published work of others, this is always clearly attributed.

- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.

- I have acknowledged all main sources of help.

- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

_____

Date:

_____

UNIVERSITY OF CAMBRIDGE

Abstract

Department of Engineering

MEng

**A Systematic Approach to Promoter Discovery for the Automated
Compilation of a Parts Registry**

by Haydn KING

The Thesis Abstract is written here (and usually kept to just this page). The page is
kept centered vertically so can expand into the blank space above the title too. . .

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1   Molecular Biology

An introduction into the *central dogma* of Molecular biology, aimed at the engineer. Covers a basic introduction into the flow of information within the cell (genome -¿ mRNAs -¿ proteins) etc.

1.2   Synthetic Biology

An introduction to SynBio and the philosophy behind it as an engineering discipline. Include information about what is meant by a part and how they can be used, paying particular attention to the promoter.

1.2.1   *Marchantia Polymorpha*

A description of the liverwort *Marchantia Polymorpha*, and an explanation of why it is an exciting potential new vehicle for Synthetic Biology with potential applications beyond those of conventional model organisms.

Chapter 2

Hidden Markov Models

2.1 Mathematical Description

A description of the Hidden Markov Model (HMM) - a Markov Model whose internal state is unknown, but which emits a symbol from an alphabet with a distribution dependant on the state.

2.2 Use for Gene Prediction

A description of how HMMs have been successfully used to predict genes in a number of studies.

2.3 Use for Promoter Discovery

How HMMs can also be used to score putative promoter regions, and how this is distinct from the case with genes.

2.4 Implementation

A description of the implementation and the algorithms required for HMMs, including:

- Training the model transition and emission probabilities from a data set

- Efficiently representing an HMM in memory

- Calculating a score for a sequence given a set of transition and emission probabilities

- Choosing a decision threshold for promoter regions

2.5   Results

Results of the HMM implementation, both in *Marchantia* and on test genomes (where promoter annotation data is already available).

Chapter 3

Experimental Verification of Predicted Promoters

3.1   Designing and Building DNA Constructs

A description of the workflow for building a DNA construct to test a particular promoter (or a set of promoters).

3.2   Protocol for Transforming Marchantia Chloroplasts

Description of the protocol for transforming the chloroplasts, including any improvements made to this relatively new protocol.

3.3   Imaging Techniques

Description of how promoter activity is to be measured using microscopy.

3.4   Results

Results of the tests (for the most likely promoter regions).

Chapter 4

Conclusions and Further Work

4.1 Summary of the Work

Summary of everything that was done as a whole, including the key contributions to the field.

4.2 Future Work and Directions

A discussion of what else needs doing and what can be done to improve the characterisation of the promoters and to improve the usefulness of the software written during the project.