A modular two-step and highly multiplexed amplicon design for Illumina sequencing

and the python DBC_amplicons package

Matt Settles
Institute for Bioinformatics and Evolutionary Studies,
University of Idaho,
Moscow,ID,
msettles@uidaho.edu

January 21, 2014

Introduction

Polymerase chain reaction (PCR) amplicon sequencing is an important tool used to query genetic variation and structure in individual biological samples and ecological communities. Applications range from determining the taxon community structure in microbial, fungal and other community types to determining mutation frequencies in a set of genes across many individuals, or even complete resequencing of small genomes. Common practice is to add a barcoded DNA sequencing adapter to the template specific primer, generating a PCR product that can be sequenced on an Illumina machine. The barcode can then be used post-sequencing to determine the sample the PCR product came from. As sequencing throughput continuing to increase the need for flexible methods that can maximize the sequencing effort are needed. For amplicons this implies sequencing more individual and/or amplicons in the same sequencing run.

An amplicon is an amplified molecule, usually generated via PCR, of a single type and is an exact replicate of the original DNA template. A pair of PCR primers are designed to uniquely target a particular region of DNA. In order to sequence a DNA fragment using the Illumina platform, certain sequences are necessary for the fragment to bind to the Illumina flowcell, amplify and then initiate sequencing. Typically, paired target specific PCR primers are designed to include the extra oligonucleotides necessary for sequencing (60bp to each primer). PCR amplicon sequencing in this manner then requires the researcher to purchase a unique pair of 80bp primer for every target region and sample (including barcode) in the experiment. A technique that is neither modular nor cost effective.

The increase in sequencing density offers an opportunity to sequence many loci across hundreds or even thousands of samples at significant depth of coverage. New techniques are needed however to both multiplex amplicons and samples in the same sequencing reaction.

Contents

Chapter 1

Amplicon Primer Design

chapter on design

Chapter 2

The dbcAmplicons python package

the package

2.1 Input Files

inputs

Barcode File

the barcode file

Appendix A

An appendix

and appendix