Primer Selection for Polymerase Chain Reactions

Overview

Goal:
Discover previously unknown genes

Strategy:
Design PCR primers for large set of known gene family members

Unknown genes will (hopefully) be amplified

Problem Formulation

Optimal Primer Selection Problem:
Input: set of DNA sequences
Output: optimal set of primers

Theorem: NP-complete
Proof: reduction from set cover
Non-approximability Result:
Can not do better than \( \log(n) \) sequences) \( \Omega(\text{OPT}) \) within polynomial time

Approaches

Find minimal primer cover:

Exact algorithms:
Exhaustive brute-force
Branch-and-bound

Provably-good heuristics:
Solution quality: \( \log(n) \) sequences) \( \Omega(\text{OPT}) \)
(\text{Best possible within polynomial time})

Extension: Inexact Primers

Observation:
Mutually competing objectives
(e.g. melting temperature, primer composition, duplicated subsequences)

Three-phrase approach:
Primer formation
Cover construction
Weight minimization

Gene family:
Genes derived from common ancestors

Common regions among family members
Common primers

Primer group:

Basic Observations

Polymerase Chain Reaction:
An effective method to amplify DNA sequence

Primer:
DNA sequences of length 15-20
Reverse complementary to target regions

Sample Output

Solution:
2 primer groups cover 8 sequences
The first primer is
\( \text{start position} \quad \text{ACCCGATCTCTAC} \)
\#mismatches:
flyNPY (48)
ratBK2 (48)
humNPY (48)
pyNPY (48)
mosP2u (48)
ratRKY (48)

The second primer is
\( \text{start position} \quad \text{CCCCGATCTCTAC} \)
\#mismatches:
dogR1 (24)
humR1 (24)
ratR1 (24)

Resulting Primer Groups

Primer Selection: Experimental Results