MuPlex: multi-objective multiplex PCR assay design

John Rachlin[1,*], Chunming Ding[2], Charles Cantor[1,3,4,6] and Simon Kasif[1,3,5]

[1]Bioinformatics Program, Boston University, Boston, MA 02215, USA, [2]Centre for Emerging Infectious Diseases, The Chinese University of Hong Kong, Prince of Wales Hospital, Shatin, New Territories, Hong Kong Special Administrative Region, China, [3]Department of Biomedical Engineering, [4]Center for Advanced Biotechnology, [5]Center for Advanced Genomic Technologies, Boston University, Boston, MA 02215, USA and [6]SEQUENOM, Inc., San Diego, CA 92121-1331, USA

ABSTRACT

We have developed a web-enabled system called MuPlex that aids researchers in the design of multiplex PCR assays. Multiplex PCR is a key technology for an endless list of applications, including detecting infectious microorganisms, whole-genome sequencing and closure, forensic analysis and for enabling flexible yet low-cost genotyping. However, the design of a multiplex PCR assays is computationally challenging because it involves tradeoffs among competing objectives, and extensive computational analysis is required in order to screen out primer-pair cross interactions. With MuPlex, users specify a set of DNA sequences along with primer selection criteria, interaction parameters and the target multiplexing level. MuPlex designs a set of multiplex PCR assays designed to cover as many of the input sequences as possible. MuPlex provides multiple solution alternatives that reveal tradeoffs among competing objectives. MuPlex is uniquely designed for large-scale multiplex PCR assay design in an automated high-throughput environment, where high coverage of potentially thousands of single nucleotide polymorphisms is required. The server is available at http://genomics14.bu.edu:8080/MuPlex/MuPlex.html.

INTRODUCTION

MuPlex (http://genomics14.bu.edu:8080/MuPlex/MuPlex.html) is a web-enabled system for designing multiplex PCR assays. A multiplex PCR solution specifies a forward and reverse primer for each single nucleotide polymorphism (SNP) and assigns each primer pair to one of a finite set of tubes. In partitioning SNP primers into individual tubes, care must be taken to ensure that all primers within a tube are mutually compatible, i.e. that they do not form primer-dimers through cross-hybridization, which would otherwise reduce target product yield. The multiplex PCR problem is equivalent to partitioning a graph $G(V, E)$ into a set of disjoint cliques, where nodes represent SNPs, edges connect two SNPs whose associated primers are tube-compatible and resulting cliques constitute valid multiplex PCR tubes. The problem of partitioning a graph into $k \leq K$ disjoint cliques is NP-complete (1). The MuPlex system is unique in that it provides multiple design alternatives that reveal inherent tradeoffs with respect to multiple competing objectives, such as average tube size, tube size uniformity and overall SNP coverage.

Multiplex PCR is a core enabling technology for high-throughput SNP genotyping, serving as a foundation for applications in forensic analysis, including human identification and paternity testing (2), the diagnosis of infectious diseases (3,4), whole-genome sequencing (5), and pharmacogenomic studies aimed at understanding the connection between individual genetic traits, drug response and disease susceptibility (6).

For example, in the hME assay (7), genomic sequences containing the SNPs of interest are first amplified by PCR. After shrimp alkaline phosphatase digestion of excess dNTPs, a primer extension reaction is carried out to interrogate the SNPs. The primer extension products (often oligonucleotides ~18 to 25 bases long) are then detected by matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) mass spectrometry. Given the large molecular weight window (4500–9000 Da) and the high resolution of the mass spectrometry, 20 or more SNPs can be easily and simultaneously genotyped. Thus, the throughput-limiting step is often the PCR plex level. In a 384-well format with 20-plex PCR, the per-SNP cost can be reduced to just a few cents while a single MALDI-TOF mass spectrometry can be used to genotype 76 800 SNPs by a single operator in 1 day.

MuPlex features

Given a set of DNA sequences and a SNP location at each, the system aims at designing (i) a set of pair forward and reverse

---

*To whom correspondence should be addressed. Tel: +1 617 921 9669; Fax: +1 617 353 4814; Email: rachlin@bu.edu

primers for each sequence; (ii) a placement of these primers into maximal size tubes such that the coverage (number of sequences) included in the PCR assay is maximized. Figure 1 presents the MuPlex home page where specific problems may be submitted. (No user registration or fee is required.) The user provides a set of SNPs and associated flanking sequences in the standard FASTA format. These sequences may be entered manually or uploaded from a file. To improve primer specificity, users may instruct the MuPlex server to filter resulting primer candidates by aligning them against the human genome using BLAT (8). In addition to the SNP sequences, the user specifies primer selection criteria, including length, GC content, positional constraints, and melting temperature T_m constraints for individual primer oligos as well as interaction parameters (maximum local alignment score, 3′ tail ΔG), and a maximum T_m range for all primer pairs within a single multiplex assay.

MuPlex then solves the dual problem of selecting primer pairs for amplifying the flanking sequence of each SNP and partitions these primer pairs into multiplex compatible sets each corresponding to a single multiplex PCR tube reaction. As noted above, MuPlex generates multiple solutions alternative each corresponding to a set of multiplex PCR tubes.

Each solution is evaluated with respect to the following objectives:

 (i) Total number of tubes required.
 (ii) Minimum, average and maximum tube size (multiplexing level).
(iii) Number of unique tube sizes.
(iv) Total SNP coverage measured both in terms of the percentage of SNPs (associated primer pairs) assigned

to maximum-sized tubes, as well as the percentage of SNPs assigned to tubes of any size.

For example, some solutions may achieve higher overall multiplexing levels but at the expense of lower coverage, i.e. by excluding some SNPs from the solution. In addition, MuPlex tries to minimize the number of unique tube sizes in order to facilitate automation in a high-throughput genomics environment. Resulting solutions are emailed to the user. The email contains a solution summary allowing quick comparison of each alternative, and details for each solution including the selected primers, their individual properties and assigned tube.

MuPlex employs a number of heuristic algorithms and allows new algorithms to be added over time. Solution time depends on the number of solution alternatives requested, the number of SNPs and the target multiplexing level. For typical problems involving <100 SNPs, multiple solutions can often be generated in 5–10 min.

ARCHITECTURE

MuPlex is written entirely in Java (j2sdk1.4.2_05) and employs the Apache Jakarta Tomcat server (http://jakarta. apache.org/tomcat) connected to a backend mySQL database (http://www.mysql.com). Individual solvers operate asynchronously on a network of workstations running a customized distribution of the Linux operating system based on Fedora Core 3 (http://fedora.redhat.com). These solvers monitor the MuPlex database for the arrival of new problems. New problems are assigned to the first available solver. As depicted
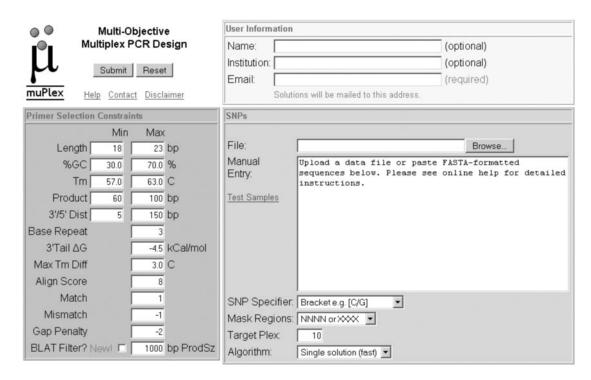

Figure 1. The MuPlex homepage. Users specify primer selection criteria and provide a collection of SNPs in the FASTA format. The system emails to the user one or more solution alternatives revealing key design tradeoffs.
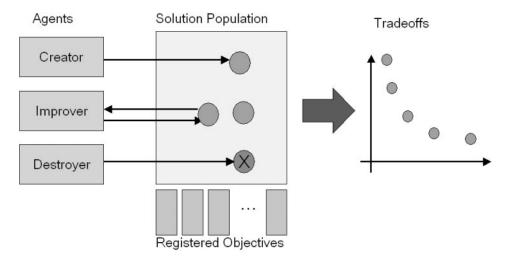
Figure 2. The MuPlex Optimizer. Once a problem is submitted and validated, it is assigned to one of the several solvers distributed across the network. Each solver instantiates one or more agents (algorithms) that either create new solutions from scratch, attempt to improve an existing solution candidate or eliminate unpromising solutions from further consideration. The collaboration of algorithms in this manner enables the system to produce multiple Multiplex PCR solutions that reveal intrinsic design tradeoffs.

in Figure 2, the solver maintains a population of candidate solutions. Each solution is evaluated with respect to a set of registered objectives. Agents encapsulating specific algorithms either create new solutions from scratch, improve or modify existing solutions, or remove unpromising solutions from further consideration. For example, one 'creator' algorithm is based on a best-fit methodology that iteratively assigns SNPs to the largest open compatible tube. When the tube size reaches the target multiplexing level specified by the user, it is closed, and no further additions or modifications to that tube are made. Each unassigned SNP is assigned a new primer pair candidate and the process repeats. One improver algorithm eliminates partial tubes in order to reduce the number of unique tube sizes but while incurring reduced coverage, while another attempts at reformulating partial tubes in an effort to identify additional full tubes. Efficiency is enhanced during the optimization process by periodically culling unpromising solutions from the population of candidates. The architecture is scaleable in the sense that new algorithms can be readily plugged-in over time, and it is robust in that it does not depend on a single algorithm to generate every viable alternative, and because system load is balanced across a distributed collection of solvers.

Within a given solution, there is no guarantee that a SNP will be assigned, and the results depend on the random order in which SNPs and primers are processed. Resulting coverage critically depends on the number of SNPs and the target level of multiplexing desired (J. Rachlin, C. M. Ding, C. Cantor and S. Kasif, manuscript submitted).

CONCLUSIONS AND FUTURE WORK

The MuPlex server allows scientists to design Multiplex PCR assays while explicitly considering intrinsic design tradeoffs. The consideration of competing alternatives has played a key role in the development of optimization and decision-support technologies in complex domains such as manufacturing and transportation logistics (9,10). Here, we have demonstrated the viability of such approaches to the optimization of multiplex PCR assays. Future efforts will focus on the development of new algorithms and on allowing users to impose dynamic feedback constraints in an effort to further guide the design optimization process towards solutions that more closely meet the scientist's particular design objectives. We also plan to develop a distributed version that will run on our 128-processor Linux cluster.

REFERENCES

1. Garey,M.R. and Johnson,D.S. (1979) *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W.H. Freeman, San Francisco, CA.
2. Inagaki,S., Yamamoto,Y., Doi,Y., Takata,T., Ishikawa,T., Imabayashi,K., Yoshitome,K., Miyaishi,S. and Ishizu,H. (2004) A new 39-plex analysis method for SNPs including 15 blood group loci. *Forensic Sci. Int*., **144**, 45–57.
3. Elnifro,E., Ashshi,A., Cooper,R. and Klapper,P. (2000) Multiplex PCR: optimization and application in diagnostic virology. *Clin. Microbiol. Rev*., **13**, 559–570.
4. Pinar,A., Bozdemir,N., Kocagoz,T. and Alacam,R. (2004) Rapid detection of bacterial atypical pneumonia agents by multiplex PCR. *Cent. Eur. J. Public Health*, **12**, 3–5.
5. Tettelin,H., Radune,D., Kasif,S., Khouri,H. and Salzberg,S. (1999) Optimized multiplex PCR: efficiently closing a whole-genome shotgun sequencing project. *Genomics*, **62**, 500–507.

6. Shi,M., Bleavins,M. and de la Iglesia,F. (1999) Technologies for detecting genetic polymorphisms in pharmacogenomics. *Mol. Diagn.*, **4**, 343–351.

7. Tang,K., Fu,D.J., Julien,D., Braun,A., Cantor,C.R. and Koster,H. (1999) Chip-based genotyping by mass spectrometry. *Proc. Natl Acad. Sci. USA*, **96**, 10016–10020.

8. Kent,W.J. (2002) BLAT: the BLAST-like alignment tool. *Genome Res.*, **12**, 656–664.

9. Murthy,S., Rachlin,J., Akkiraju,R. and Wu,F. (1997) Agent-Based Cooperative Scheduling. In: *Constraints & Agents*. Technical Report WS 1997-97-05.

10. Rachlin,J., Goodwin,R., Murthy,S., Akkiraju,R., Wu,F., Kumaran,S. and Das,R. (1999) A-Teams: An Agent Architecture for Optimization and Decision-Support. Lecture Notes in Artificial Intelligence. **1555**.