The Limits of Multiplex PCR

John Rachlin[1], Chunming Ding[2], Charles Cantor[3], Simon Kasif[4]

Keywords: multiplex PCR, phase transitions, graph theory

1 Introduction.

Multiplex PCR is an essential cost-saving technique for large scale genotyping with significant scientific, clinical, and commercial applications including gene expression [1], whole-genome sequencing [2,3], forensic analysis, including human identification and paternity testing [4] and to facilitate the diagnosis of infectious diseases [5]. In recent years, multiplex PCR has emerged as a core enabling technology for high-throughput SNP genotyping [6]. Designing multiplex assays is equivalent to finding a set of disjoint cliques in a novel graph structure called a multi-node graph. We demonstrate that the degree of coverage is subject to a computational phase transition such that designing high-coverage high-multiplexing assays becomes dramatically more difficult as pairwise SNP compatibility falls below a certain threshold.

2 Multi-Node Graphs.

Multiplex PCR assay design is a multiobjective optimization problem. Primers selected for each locus or SNP are designed to minimize potential interactions, while impacting the number of primer candidates. The number of SNPs per assay should be maximized and it is advantageous to ensure that all tubes are uniformly multiplexed to facilitate automation in high-throughput environments. Additionally, the number of SNPs assigned or "covered" to a full tube should be maximized. Achieving high (>95%) coverage becomes increasingly important when the focus is on a relatively small (10^2-10^3) set of SNPs each of which is suspected of having some biological impact.

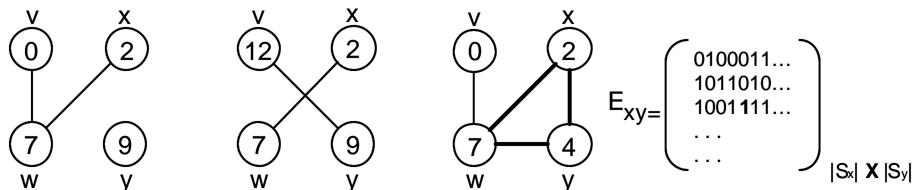

Figure 1. A multi-node graph is a graph in which individual nodes can take on one or more states. An edge between two nodes, X and Y, is determined by the state of the two nodes, or more specifically, an edge matrix $\mathbf{E}_{XY}$ connecting nodes X and Y. There is no restriction on the number of states per node, and each node may contain a different number of states. [4]

Designing one or more multiplex assays for SNPS with preselected primers is equivalent to finding a clique in a graph G where nodes are SNPS and edges denote pairwise multiplex compatibility. Theoretical bounds for covering a graph with disjoint cliques can be found in [7]. The multiplex PCR problem is more general in that we impact the graph topology by choice of primers. We use the term "multi-node graph" (Figure 1) to denote a graph whose nodes have multiple states. In a multi-node graph, an edge matrix E_{uv} is attached to each pair of nodes, (u,v). If node u (in state i) is multiplex compatible with node v (in state j) then $E_{uv}[i][j]= 1$. Otherwise, $E_{uv}[i][j]=0$. As illustrated above, nodes W, X, and Y are pair-wise compatible when in certain states and incompatible in other states.

[1] Boston University,44 Cummington Ave., Boston,MA,USA. Email: rachlin@bu.edu
[2] Boston University, 44 Cummington Ave., Boston,MA,USA. Email: cmding@bu.edu
[3] SEQUENOM, Inc. San Diego,CA,USA. Email: ccantor@sequenom.com
[4] Boston University, 44 Cummington Ave., Boston,MA,USA. Email: kasif@bu.edu

The nodes W,X,Y form a 3-clique (3-plex) when in states 7, 2, and 4 respectively. The multiplex PCR design problem is equivalent to covering as many nodes as possible with disjoint cliques of size M in a multi-node graph.

3 Phase Transitions and Multiplex Scalability.

In recent years, it has been shown that for broad classes of problems, there exist certain phase-transition boundaries across which the nature of the solutions and the computational effort needed to identify a solution changes dramatically [8]. When attempting to design multiplex PCR assays with high coverage, we observe a similar computational behavior. We induce a random multi-node graph by setting $E_{uv}[i][j]=1$ with probability P, for all node pairs u and v, in states i and j respectively. Using a best-fit greedy algorithm, we find that our ability to achieve high (>95%) coverage for randomly generated multi-node graphs critically depends on the compatibility probability, P, and the target level of multiplexing, M, as shown in Figure 2.

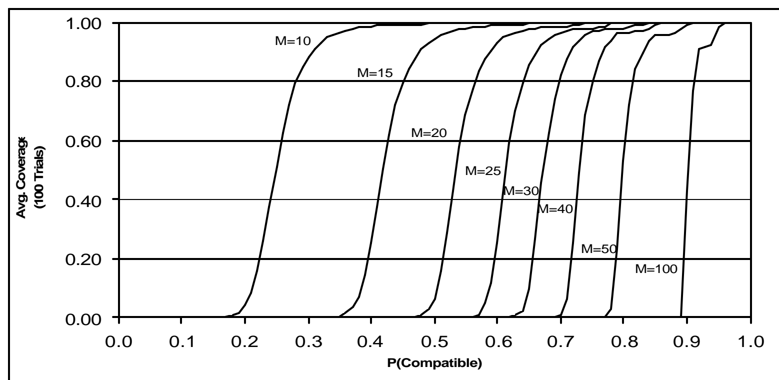

Figure 2. Phase transitions in M-coverage as a function of the SNP -SNP compatibility probability (primer interaction test stringency) for varying multiplexing targets 10 to 100.

References

[7] Alon, N. and J.H. Spencer, 2000. The Probabilistic Method. New York:Wiley.

[2] Beigel R., A.N., Apaydin S.M., Fortnow L., and Kasif S. 2001. An Optimal Multiplex PCR Protocol for Closing Gaps in Whole Genomes (*RECOMB 01)*.

[1] Ding, C. and C.R. Cantor, 2003. A high-throughput gene expression analysis technique using competitive PCR and matrix-assisted laser desorption ionization time-of-flight MS*Proc Natl Acad Sci* 100,3059-64.

[5] Elnifro, E.M., Ashshi, A.M., Cooper, R.J. & Klapper, P.E. 2000. Multiplex PCR: optimization and application in diagnostic virology. *Clin Microbiol Rev* 13,559-70.

[8] Hogg T., Huberman B.A., Williams C. 1996. Phase Transitions and the Search Problem. *Artificial Intelligence* 81,1-15.

[4] Inagaki, S., et al., 2004. A new 39-plex analysis method for SNPs including 15 blood group loci. Forensic Sci Int 144,45-57.

[6] Shi, M.M., 2001. Enabling large-scale pharmacogenetic studies by high-throughput mutation detection and genotyping technologies. *Clinical Chemistry*, 47,164-72.

[3] Tettelin, H., Radune, D., Kasif, S., Khouri, H. & Salzberg, S.L. 1999. Optimized multiplex PCR: efficiently closing a whole-genome shotgun sequencing project. Genomics 62,500-7.