

# A higher resolution multiplexed microsatellite marker set for reliable, cost effective genotyping

Robert J. Kolts, Elliott G. Richards, Sam Ashby, Will D. Brubaker, Aaron C. Ferguson, Paul Fjeldsted, Derek Hatch, and Edward Wilcox\*

Department of Integrative Biology  
401 WIDB

Brigham Young University  
Provo UT, 84602

\*Contact: edward\_wilcox@byu.edu

Key Words: genomics, microsatellites, genetics

---

## Summary

We are creating a new microsatellite marker set for public use. Because many projects involving genetic disorders in humans depend on microsatellite markers, there is an existing need for a reliable and cost-effective marker set that provides higher-quality statistical data while still being affordable. Our proposed set will consist of approximately 800 markers and currently has a rough spacing of one marker every 4.1 cM. The average heterozygosity is currently 0.85. To further reduce costs to users, we are arranging markers and corresponding PCR primers into ~178 multiplexed groups consisting of 4-5 markers each. These multiplexed groups will allow research teams to perform genomewide scans at minimal cost.

## Introduction

Microsatellite marker sets for genomewide linkage scans have been widely used in Mendelian disorders<sup>1</sup>. Current marker sets, however, are often unsuccessful in their demonstration of linkage of diseases and traits. A survey of 101 linkage studies of human diseases found that “most” (66.3%) of the studies did not show “significant” evidence of linkage. Furthermore, it found that studies of the same disease were often inconsistent in their results<sup>2</sup>, demonstrating the need for a higher-quality and more reliable marker set.

High quality microsatellite marker sets for linkage studies are available from several commercial sources for a price, effectively eliminating their use for cost-conscious research groups. Public microsatellite marker sets, with primer sequences freely available, have several shortcomings. First, some sets have a marker density of roughly 10 cM

or more. With such widely spaced markers it is often difficult to determine co-segregation of a specific marker allele and phenotype. At a 10 cM density level, the inheritance information content obtained is only ~70% when parental genotypes are available and drops to ~30% when parental genotypes are unavailable<sup>3</sup>.

Current microsatellite maps also contain numerous markers with lower heterozygosities, with many studies reporting average heterozygosities of ~0.72. Using markers with high heterozygosities is an important aspect of genotyping and linkage studies. The heterozygosity of a marker is directly proportional to the amount of inheritance information obtained<sup>3</sup>.

Other areas of concern with current public sets are that many existing primers fall within repetitive sequences, were often designed from poor quality sequence data, and were not designed by current standards in PCR primer design software.

## Methods

Established markers were derived from both the Japan Biological Information Research Center<sup>4, A</sup> (JBIRC) database and the Mammalian Genotyping Service of the Marshfield Clinic Research Foundation<sup>5, B</sup>. Markers were then arranged by heterozygosity. The JBIRC database yielded 62,505 micro-satellite markers, with 30,630 showing no displayed heterozygosity and an additional 26,969 having a heterozygosity below our minimum criteria of 0.82. The remaining 4,906 markers had an acceptable heterozygosity (>0.82). The Marshfield database yielded 8,306 microsatellite markers with 89 showing no displayed heterozygosity, an additional 6,961 having a heterozygosity of <0.82, and the

remaining 1,256 having an acceptable heterozygosity.

After removing duplicates, acceptable Marshfield markers were merged with the ~5,000 acceptable markers from the JBIRC database. Markers were then organized by physical and meiotic position as derived from the University of California-Santa Cruz (UCSC) Genome Browser database<sup>6, C</sup>. A set was then selected utilizing one marker every 4-5 cM based on the deCode recombination map<sup>7</sup>, with the most common marker featuring a dinucleotide repeat. Any gaps between established markers with high heterozygosities are filled with a candidate simple tandem repeat (STR) at a 5 cM density level or less.

To create new markers, STR tables were downloaded using the UCSC Genome Browser table retrieval tool<sup>8</sup>. STRs were chosen for markers by algorithms we developed, which rank the STRs based on repeat length, match %, and other factors. All work was based on the March 2006 (build 36.1, NCBI) build of the human genome.

### **Current Work**

While designing marker sets and corresponding PCR primers into multiplexed reactions is tedious and labor intensive, multiplex PCR offers several advantages over “singleplexed” PCR reactions to the end user. First, the number of individual reactions, and thus time required in set-up and analysis, is greatly reduced. Secondly, the amount of reagents consumed will be lower, reducing costs significantly. By multiplexing PCR reactions we are increasing the usability and functionality of this marker set.

Multiplexed sets of primers, including markers utilized from current public databases, are being designed and tested from several PCR primer design programs: Visual OMP v5.0 (DNA Software, Ann Arbor, MI), muPlex: Multi-Objective Multiplex PCR Design v2.2<sup>9, D</sup> and Primer3 on the World Wide Web v0.2<sup>10, E</sup>. Multiplexed sets typically contain 4-5 markers and their corresponding PCR primers. The multiplex primer design software select primer groups based on user selected criteria including target Tm, mis-hybridization, duplex formation, amplicon size, and cross-hybridization. All primers are blasted against the human genome to ensure specificity using internal software blast capabilities as well as the NCBI online blast<sup>F</sup>.

Primer oligos are ordered from Integrated DNA Technologies, Inc (Coralville, IA). Fluorescent labels are incorporated on to PCR products as documented by Markus Schuelke<sup>11</sup>. Markers are then run using standard fragment analysis methods on an ABI 3100 Genetic Analyzer.

Heterozygosities for the uncharacterized STR markers will be determined on a panel of approximately 100 random normal control DNA samples. Those not meeting acceptable heterozygosities will be replaced with a neighboring STR.

### **Difficulties**

Although the software we are using screen for potential problems with multiplex primer design, several obstacles are still present. First, VisualOMP, software intended primarily for multiplex primer design, does not take into account the 3' end stability of newly created primers, allowing a polymerase to amplify—albeit less efficiently—non-target DNA. Preliminary fragment analysis tests of primers have yielded less-than-specific peaks, also suggesting the blast filter criteria need to be adjusted. VisualOMP does, however, predict overall stability of primers and also avoids sequence areas with high amounts of secondary structure.

Primer3 presents a related problem. Primer3, which does not currently support multiplex primer design, seems to have an issue with its blast filtering. Although it claims to blast potential primers against a human mispriming library to avoid placing primers in repetitive sequences, alternate blasts using the NCBI online blast indicate poor overall primer specificity.

The muPlex server seems to overcome many of these difficulties, though it does not take into account DNA secondary structure. Preliminary fragment analysis tests of primers designed using muPlex have yielded specific peaks. muPlex also offers the benefit of placing primers designed from user-inputted sequences automatically into “tubes” of a pre-selected size. This greatly reduces man-hours spent on organizing multiplex groups.

### **Intended Use**

Our research group's current focus is to create a suitable set of markers to perform segregation analysis of nonsyndromic cleft lip with or without

cleft palate (CL/P) in large, three and four generation, consanguineous Pakistani families. Although there have been a number of previous genomewide scans for CL/P, heterogeneity seems to be an important factor<sup>12</sup> and has represented a serious obstacle towards progress in this field.

## Discussion

This microsatellite marker set will be tested in our project on DNA samples from Pakistan. The markers, however, will have wide applicability and high heterozygosities in other populations as well. Many cultures place an emphasis on marriage outside of one's immediate family, increasing genetic diversity. For recessive diseases, however, the power to detect linkage more than doubles when working with samples from consanguineous families such as those found in Pakistan, where half of all marriages are to first-degree relatives. Often, each family represents a unique genetic isolate and hence simplifies immensely the enormity of such studies. The impact of being able to simplify the study of inherited traits (be they simple highly heterogeneous Mendelian traits or more complex traits) can be demonstrated by the progress deCODE Corporation has made in mapping many different inherited traits in the Icelandic population. A deCODE paper published in this year's March issue of *Nature Genetics*<sup>13</sup> includes data on a specific mutation accounting for some 21% of attributable risk for diabetes type II in European populations, based on first looking for such a gene among the Icelandic people and then applying the information discovered to other populations.

With the advent of single-nucleotide polymorphism (SNP) maps becoming increasingly prominent in linkage and association studies, a question may arise concerning the focus we are placing on microsatellite maps. Individual microsatellite markers are more polymorphic than individual SNPs and as a result are more informative<sup>14, 15</sup>. Grant, et al (2006)<sup>13</sup>, found that to increase the power of a linkage study, it is effective to utilize microsatellite markers to determine linkage and then saturate possible linkage areas with high-quality SNPs and STRs. The aforementioned study found that no SNP "demonstrated stronger association" than an equally well placed STR.

## Conclusion

By creating a higher density marker set, we hope to improve the statistical analysis data of our

research, as well as provide a suitable set of markers to other research groups involved in mapping genetic traits. There is a premium to be paid for the "latest technology" that we and many others can little afford. Great strides in understanding the inheritance of many human genetic disorders can be made from improvements of current technology.

## Internet Resources

- A <http://www.jbirc.aist.go.jp/gdbs>
- B <http://research.marshfieldclinic.org/genetics/>
- C <http://genome.ucsc.edu/>
- D <http://genomics14.bu.edu:8080/MuPlex/MuPlex.html>
- E [http://fokker.wi.mit.edu/cgibin/primer3/primer3\\_www.cgi](http://fokker.wi.mit.edu/cgibin/primer3/primer3_www.cgi)
- F <http://www.ncbi.nlm.nih.gov/blast/index.shtml>

## References

- 1 Korstanje R, Paigen B. From QTL to gene: the harvest begins. *Nature Genet.* 2002;31:235-236.
- 2 Altmuller J, et al. Genomewide scans of complex human diseases: true linkage is hard to find. *Am J Hum Genet.* 2001 Nov; 69(5): 936-950.
- 3 Evans DM, Cardon LR. Guidelines for genotyping in genomewide linkage studies: single-nucleotide-polymorphism maps versus microsatellite maps. *Am J Hum Genet.* 2004 Oct; 75(4): 687-692.
- 4 Tamiya G, et al. Whole genome association study of rheumatoid arthritis using 27, 039 microsatellites. *Hum Molec Genet.* 2005; 14(16): 22305-2321.
- 5 Ghebranious N, et al. STRP screening sets for the human genome at 5 cM density. *BMC Genomics.* 2003;4: 6.;
- 6 Kent WJ, et al. The Human Genome Browser at UCSC. *Genome Res.* 2002; 12(6): 996-1006.
- 7 Kong A, et al. A high-resolution recombination map of the human genome. *Nat Genet.* 2002; 31(3): 225-226.
- 8 Karolchik D, et al. The UCSC Table Browser data retrieval tool. *Nucl. Acids Res.* 2004; 32(Suppl 1), D493-D496.

- 
- <sup>9</sup> Rachlin J, et al. muPlex: A Multi-Objective Approach to Multiplex PCR Assay Design. *Nucleic Acids Research*. 2005; 33(Web Server Issue):W544-W547.
- <sup>10</sup> Rozen S, Skaletsky H. (2000). Primer3 on the WWW for general users and for biologist programmers In: Krawetz S, Misener S (eds) *Bioinformatics Methods and Protocols: Methods in Molecular Biology*. Humana Press, Totowa, NJ, pp 365-386.
- <sup>11</sup> Schuelke M. An economic method for the fluorescent labeling of PCR fragments. *Nat Biotech*. 2000; 18(2): 233-234.
- <sup>12</sup> Marazita ML, et al. Meta analysis of 13 genome scans reveals multiple cleft lip/palate genes with novel loci on 9q21 and 2q32-35. *Am J Hum Genet*. 2004 Aug; 75(2):161-173.
- <sup>13</sup> Grant SFA, et al. Variant of transcription factor 7-like 2 (*TCF7L2*) gene confers risk of type 2 diabetes. *Nat Genet*. 2006; 38(3):320-323.
- <sup>14</sup> Daw EW, Heath SC, Lu Y. Single-nucleotide polymorphism versus microsatellite markers in a combined linkage and segregation analysis of a quantitative trait. *BMC Genetics* 2005, 6(Suppl 1):S32.
- <sup>15</sup> Kawashima M, et al. Genomewide association analysis of human narcolepsy and a new resistance gene. *Am J Hum Genet*. 2006 Aug; 79(2):252-263.