Preface

Research into gluten sensitivity has never been more popular nor more exciting. Thus a call for a new book, *Celiac Disease: Methods and Protocols*, devoted entirely to techniques and technology seemed a most appropriate undertaking. I am therefore grateful to Professor J. M. Walker for inviting me to complete this task for Humana Press. To do this would have been impossible without the contributions of friends and colleagues from around the world who have devoted so much interest to the project. It has also been necessary for them to master the unique chapter-writing skills required of every manuscript published in this series of laboratory monographs.

With regard to gluten sensitivity we are in a period of great change, occasioned by the introduction of reproducible methods for cloning lymphocytes, the application of physical methods to identify gluten sequences as T-cell antigens, the study of peptide responses in vitro and in vivo by either jejunal or rectal challenge, elucidating the locations of other genes concerned in pathogenesis, or the use of elegant immunohistocytochemical and mRNA probing techniques for analyzing the finer points of the mucosal inflammatory response to gluten. Never was a detailed laboratory handbook, of the quality assembled here through its contributing experts, so necessary as at this time!

It has been a privilege to put together *Celiac Disease: Methods and Protocols* and it is offered in the hope that its pages will contain the necessary information for any newcomer to this field to get himself or herself organized quickly with the least technical hindrance.

Michael N. Marsh, MD, DSc, FRCP

2

Genotyping Methodologies

Stephen Bevan and Richard S. Houlston

1. Introduction

This chapter details DNA extraction through polymerase chain reaction (PCR) amplification, gel running, allele assignment, and data management so that the genotyping data produced is suitable for use in linkage analysis programs.

1.1. DNA Extraction

The most convenient source of genomic DNA is via EDTA blood samples, which after collection can be frozen and stored at –70°C for long periods. Since only white blood cells (WBCs) contain DNA, the first process in the extraction protocol is to separate the red blood cells (RBCs) and WBCs either by centrifugation or by lysis of the RBCs in a hypotonic solution.

1.2. Considerations Before PCR

A genomewide search is typically based on between 250 and 400 markers to give 10–20 cm separation across the genome. Before embarking on a genomewide search, several factors need to be considered. These include detection of PCR products, what label should be used, and, given the large number of results that will be generated, which system will maximize throughput.

1.3. Detection of PCR Products

Fluorescent labeling and radioactive labeling are the two main methods of detecting PCR products with the resolution required for allele calling. Both methods have advantages and disadvantages, primarily in terms of cost and the laboratory equipment needed to detect them.

From: *Methods in Molecular Medicine, Vol. 41: Celiac Disease: Methods and Protocols*
Edited by: M. N. Marsh

The simplest way of labeling a PCR product is to label the primer before the PCR begins. In the case of fluorescence, the labeled primer is usually acquired from a commercial source, whereas with radioactivity, the primer can be labeled with ^{32}P on the bench. The two main advantages of using fluorescent primers are that they are nonhazardous and that they can be multiplexed to speed up analysis. The disadvantages of this approach are the expense and the requirement for specialized equipment such as an ABI 377 DNA sequencer (Applied Biosystems, Foster City, CA). Detection of fluorescently labeled PCR products works by electrophoresing the products through denaturing polyacrylamide gels along with a labeled size marker. The products migrate through the path of an argon laser beam and emit fluorescence, which is then detected. Four colors can be detected, allowing multiple samples to be loaded into a single lane. Furthermore, products of different sizes migrate at different rates, so more than one sample can be loaded with the same colored marker. This allows up to nine markers to be simultaneously loaded in a single lane. For radioactive markers, it is only possible to load a single marker in any one lane of the gel because there is only one type of output signal and the resolution is not as high as that for fluorescent markers. The main advantages of radioactively labeled markers over fluorescent ones is that they are relatively cheap, easy to generate and detect, and no expensive detection equipment is required. However, the allele numbers have to be scored manually, which can be time-consuming. Use of fluorescent primers and an ABI 377 means that the data are stored digitally and can be analyzed by computer programs such as genotyper or Genetic Analysis System (GAS) software.

1.4. Design of Fluorescent Marker Panels

Fluorescent markers are available either individually or in panels. The panels consist of a range of markers designed to be run together in a single gel lane to give maximal throughput. The main disadvantage of these panels is that they have been designed with a genomewide search in mind, and, as such, markers from a single chromosome are randomly distributed through the panels depending on the size of product they produce. Microsatellite markers, the size of the product they amplify, and their position in the genome can be found at several Internet sites, which are given in **Table 1**. Note, however, that these sites tend to use their own maps, and the distances quoted will vary from site to site. Thus, markers should be chosen from only one map rather than several. Also provided for each marker is a heterogeneity score ranging from 0 to 1. This is a measure of how informative the marker is for linkage; the higher the number the more informative the marker.

Table 1
Contact Addresses for Information on Fluorescent Primers and Related Technical Data

Marshfield Centre for Medical Genetics
http://www.marshmed.org/genetics/
Genethon
http://www.genethon.fr/genethon-en.html
Cooperative Human Linkage Centre
http://www.chlc.org/
Genetic Location Database
http://cedar.genetics.soton.ac.uk/public_html/
Genome Database
http://www.hgmp.mrc.ac.uk/gdb/gdbtop.html
CEPH Genotyping Database
http://www.cephb.fr/cephdb/
National Centre for Biotechnology Information
http://www.ncbi.nlm.nih.gov/
Perkin-Elmer Applied Biosystems
http://www.perkin-elmer.com/ab
Helena Bioscience
e-mail: info@helena.co.uk

1.5. Radioactive Primers

The best alternative to fluorescent primers is radioactively labeled primers. The two main methods of radioactively labeling are either to label the primer before performing PCR or to use a radioactively labeled dNTP for incorporation during PCR. However, the latter method provides a lower level of resolution, making radioactively labeled primers the method of choice. Endlabeling of primers relies on the use of T4 polynucleotide kinase to catalyze the transfer and exchange of phosphate from adenosine triphosphate to the 5' hydroxyl terminus of polynucleotides.

1.6. PCR

Standard PCR protocols can be used for both fluorescent and radioactive primers, and the commercial suppliers of fluorescently labeled primers will usually provide the PCR conditions suitable for their primers.

1.7. Allele Calling from Fluorescent Primers

Multiple PCR reactions can be mixed before genescan gel loading to enhance sample throughput. Ideally, approx 5–10 ng of DNA of each sample should be

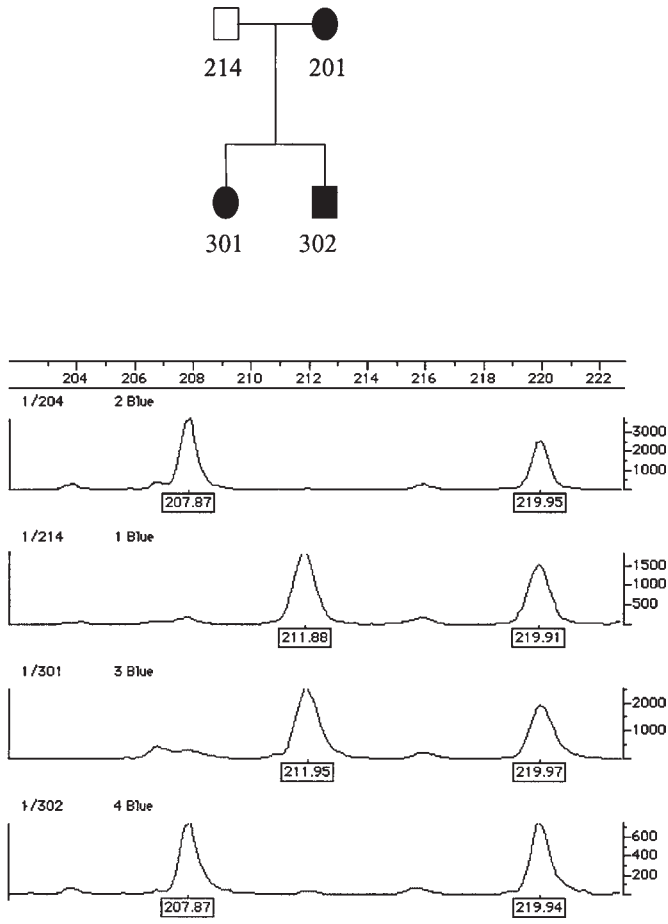


Fig. 1. Pedigree and corresponding Genotyper output for a single marker D10S1677.

loaded onto each lane of the gel. During electrophoresis the fluorescent data are collected and stored using the GeneScan Collection Software (Perkin-Elmer), and analyzed by the GeneScan Analysis Software at the end of the run. These programs come with the ABI 377 DNA sequencer. The gel file can then be downloaded to a computer (typically an Apple Macintosh) and alleles scored using the Genotyper software. This software produces a plot of size in base pairs against fluorescence intensity. A PCR product will produce a peak in fluorescence corresponding to its size. The allele number can then be scored either manually or automatically. **Figure 1** shows the pedigree of a small family and the corresponding Genotyper output for marker D10S677. This marker is a tetranucleotide repeat and has a predicted size range of between 197 and 225 bp.

From the Genotyper output in **Fig. 1** it can be seen that alleles should be labeled at 207, 211, 215, 219 bp, and so on. With a larger number of families, the number of different alleles will increase and accurate allele frequencies can be calculated. Because the expected size of D10S677 is 197–225 bp, 199 bp should be labeled allele 1, 203 bp allele 2, and so on. If an allele does not occur, then it can be given a frequency of 0 in subsequent linkage analysis. The family in **Fig. 1** would be scored as 3 6, 4 6, 4 6, 3 6 for person 1/201, 1/214, 1/301, and 1/302 respectively.

1.8. Detection of Radioactive PCR Products

Radioactive PCR products are run on urea denaturing gels, and in our laboratory, they are set up on a standard vertical gel electrophoresis apparatus (Model S2, Gibco BRL, Paisley, UK). The gels are 30 × 40 cm and can accommodate SO samples at any one time.

On radioactive gels, alleles are generally scored from top to bottom, assigning the highest band as allele 1, the next as allele 2, and so on. If two gels are being run with the same marker, it is essential to run a duplicate sample on both gels to ensure uniformity in calling alleles. This is important when calculating frequencies of alleles for linkage analysis.

1.9. Data Management

Following the assignment of alleles to individual DNA samples, the inheritance of these alleles should be examined for Mendelian transmission as a prelude to linkage analysis. This can be done by eye from a sheet of paper, but the process can become quite complex in families with many markers. A suitable program for displaying pedigrees and markers is the commercial software package Cyrillic (Cherwell Scientific, Oxford, UK).

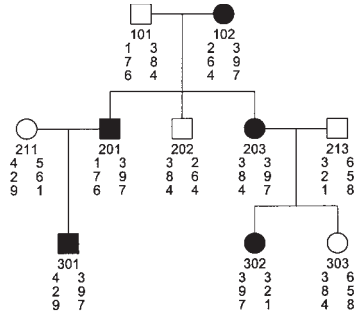
1.10. Cyrillic

In Cyrillic each pedigree can be drawn, along with the relevant individual's phenotype and marker alleles. The benefit of this is that a family can be associated with more than one disease and data for each disease kept separate. Cyrillic will also haplotype families automatically. Cyrillic has an export function, enabling marker data to be transferred out to analysis packages such as MLINK and FASTLINK. However, the program is inflexible since pedigrees cannot be automatically drawn by importing allele data from other programs.

1.11. Data Format

For linkage analysis, data must be written into or imported/read into the analysis software in a specific format. This is usually set out as follows:

Family ID, PID, FID, MID, Sex, Affection Status, Marker Typings



1	101	0	0	1	1	1	3	7	8	6	4
1	102	0	0	2	2	2	3	6	9	4	7
1	201	101	102	1	2	1	3	7	9	6	7
1	202	101	102	1	1	3	2	8	6	4	4
1	203	101	102	2	2	3	3	8	9	4	7
1	211	0	0	2	1	4	5	2	6	9	1
1	213	0	0	1	1	3	6	2	5	1	8
1	301	201	211	1	2	4	3	2	9	9	7
1	302	213	203	2	2	3	3	9	2	7	1
1	303	213	203	2	1	3	6	8	5	4	8

Fig. 2. A family pedigree typed with three markers and the output file for the pedigree ready for analysis by linkage analysis software.

where Family ID is the family number; PID is the person number; FID is the number of the person's father (0 = unknown); MID is the number of the person's mother (0 = unknown); Sex is 1 for male and 2 for female; Affection Status is 0 for unknown, 1 for unaffected, and 2 for affected; and the Marker Typings are the alleles produced from the genescan analysis or radioactive gel runnings. **Figure 2** shows a family (given the family ID of 1) typed for three markers and the format in which this family's data would be arranged for linkage analysis.

Cyrillic can be used to create and export output files from pedigrees, but these have to be drawn first, which is time-consuming. We have found it easier to create a database using Microsoft Access, which allows the marker alleles to be typed into a table. By setting up a table containing the standard family pedigree information (columns 1–6 in **Fig. 1**), marker alleles can be merged for analysis. No graphic representation is required, and as few or as many markers as required can be merged for analysis at any one time. The major advantage of Access is that tables can be linked together, so that inputting data into one table automatically adds the same data to other tables. This has allowed us to input data into tables on a panel-by-panel basis, automatically allocating them to

chromosome-specific tables. The chromosomal table is then merged with the pedigree table ready for linkage analysis.

Although Access does have an export function, exporting changes the spacing of the fields. A simple way to overcome this problem is to cut the table data you want to analyze (by highlighting the text and using the Edit, Cut command), opening Microsoft Word and using the Edit, Paste Special command. This gives two options: to paste either as formatted text (rich text format [RTF]) or as unformatted text. The data must be pasted as unformatted text and then saved as a text only file (*.txt). This maintains the spacing of the fields, and the text file can be read directly by linkage analysis software. Transfer between PC and a UNIX machine running linkage software can be conveniently performed by a file transfer protocol (ftp).

Genotyping data can also be managed by a suite of programs collectively called GAS. This program has the advantage that it will automatically put raw data into a format suitable for linkage analysis, and it also has analysis software. The GAS program, manual, and example files are available from ftp.well.ox.ac.uk by anonymous ftp and are available for IBM-PC, Vax UMS, DEC Ultrix, DEC Alpha, Sun solaris and Sun os. When logging in, your user name should be anonymous and your password your e-mail address.

2. Materials

1. Sucrose lysis mix: 218 g of sucrose, 20 mL of 1 *M* Tris (pH 7.5), 2 g of $MgCl_2$, 20 mL of Triton X-100. Make up to 2 L with dH_2O to provide enough solution for forty 10-mL blood samples.
2. Resuspension buffer: 2.6 mL of 5 *M* NaCl, 0.84 mL of 0.5 *M* EDTA (pH 8.0), 15 mL of 10% sodium dodecyl sulfate. Make up to 175 mL with dH_2O to provide enough solution for forty 10-mL blood samples.
3. GTB buffer (20X): 432 g of Tris, 144 g of taurine, 8 g of EDTA. Make up to 2 L with dH_2O and stir until dissolved.

3. Methods

3.1. Sucrose Lysis DNA Extraction

1. Decant 10 mL of blood into a 50-mL Falcon tube and add 40 mL of ice-cold dH_2O.
2. Invert the tube five times to mix the solutions gently, and then centrifuge at 500*g* for 20 min at 4°C. This lyses the RBCs and pellets the remaining WBCs.
3. Remove the supernatant and keep the pellet on ice. Add 25 mL of ice-cold sucrose lysis solution to the pellet and resuspend by moderate manual shaking to lyse the WBCs.
4. Centrifuge at 500*g* for 20 min at 4°C to pellet the released genomic DNA.
5. Discard the supernatant and resuspend the pellet in 3.5 mL of resuspension buffer supplemented with 20 mg/mL of proteinase K (0.5 mL of 20 mg/mL of proteinase K should be added to 175 mL of resuspension buffer immediately prior to use).

6. Following gentle resuspension, incubate overnight at 37°C or for 3 h at 60°C to allow protein digestion.
7. Add 1.2 mL of 5 *M* NaCl to the tube and shake vigorously for 20 s to precipitate digested protein. Centrifuge at room temperature for 30 min at 3000*g* to pellet the protein.
8. Transfer the supernatant to a 15-mL tube and add 2 vol of 100% ethanol. Then invert gently to precipitate the DNA. If necessary the sample can be left at –20°C for 30 min to enhance precipitation. If the DNA is visible, it can be removed with a pipet to a separate tube and dried before resuspending in Tris-EDTA (TE). If it is not visible, centrifuge at 2800*g* for 30 min to pellet the DNA, remove the supernatant, dry the pellet, and then resuspend in TE (200–500 µL depending on the size of the pellet).

3.2. Endlabeling of Primers with ^{32}P

1. Add to a 1.5-mL microfuge tube 20 µL of primer (at 5 outer diameter [OD] conc.), 2.5 µL of 10X kinase buffer, 1 µL of T4 polynucleotide kinase, 1 µL of ^{32}P, and make up to 25 µL with dH_2O.
2. Incubate at 37°C for 40 min to allow addition of the ^{32}P to the primer. Then add to the PCR stock mix ready for PCR.

3.3. PCR Protocol

All of the volumes in this protocol are applicable to a 96-well plate used on a Biomek 1000 robot (Beckman Coulter, Fullerton, CA), i.e., for one hundred 15-µL PCR reactions.

1. In a 1.5-mL microfuge tube, mix 530 µL of dH_2O, 150 µL of reaction buffer, 150 µL of dNTP mix (10 m*M*), 90 µL of 25 m*M* $MgCl_2$ (1.5 m*M* final), 15 µL of 10 mg/mL of bovine serum albumin, 30 µL of each primer (40 pmol), and 6 µL of *Taq* (5 U/µL).
2. Add 10 µL of the stock mix to 5 µL of DNA (at 2.5 ng/µL) and cover with 20 µL of mineral oil in a 96-well plate by the robot. (The PCR machine can be set to hot lid if required, dispensing with the need for oil, but the presence of oil helps reduce evaporation and condensation after PCR when the samples are stored at 4°C.)
3. Then run the PCR at conditions specific to the primer pair in question. For example, a primer pair with an annealing temperature of 55°C would have 30 cycles of 55°C for 1 min, 72°C for 2 min, and 94°C for 1 min before a final cycle of 55°C for 1 min followed by 72°C for 7 min to ensure full extension and maximization of double-stranded PCR product.

3.4. Electrophoresis of Radioactive Markers

1. Clean both gel plates with soapy water and then with 100% EtOH, and coat the small plate in a silane solution such as sigmacote (Sigma Aldrich, St. Louis, MO) to prevent the gel from sticking to it.
2. Once dry, add the spacers and tape the plates ready for gel pouring.

3. Make the gel by mixing 40 g of urea, 4 mL of 20X GTB buffer and 31 mL of dH_2O. Swirl gently until most of the mix has dissolved, and then heat for 20 s on full power in a microwave. Swirl gently until completely dissolved.
4. Add 12 mL of 40% acrylamide (6% final), 300 µL of adenosine 5'-phosphosulfate (APS), and 24 µL of TEMED; mix gently; and pour. Polymerization should occur within 30 min to 1 h.
5. To your 15-µL PCR reaction add 20 µL of running dye (200 µg of bromophenol blue, 200 µg of xylene cyanol in 100 mL of formamide) and immediately before loading heat to 94°C for 5 min. Then place on ice and load. This procedure is done to denature all double-stranded molecules and anneal them slowly to reduce to a minimum the amount of nonspecific binding, which leads to false bands on the gel.
6. Run at 80 W for as long as necessary to electrophorese the product into the bottom third of the plate (the longer the better since the further the products travel the better the separation).
7. Once the gel has run, remove the plates from the gel apparatus, separate the plates, and transfer the gel to filter paper (Whatman 3MM paper or similar) by laying the paper onto the gel and applying gentle pressure before peeling up from one corner, being careful to mark the orientation of the gel. Place a piece of Saran wrap over the gel and dry under vacuum at 80°C on a gel dryer for 40–60 min.
8. Check the activity of the gel with a Geiger counter and then expose to X-ray film in an autoradiography cassette for as long as required. Develop the gel in the usual manner and score the alleles.

4. Notes

1. Several commercial kits are available for the extraction of DNA from blood and solid tissues, but these are generally quite expensive—particularly when a large number of samples are to be extracted. For this reason, most laboratories have adopted the sucrose lysis method of genomic DNA extraction from whole blood. This method uses water to lyse the RBCs and a sucrose solution to burst the WBCs, allowing the genomic DNA to be precipitated following incubation with proteinase K to remove any contaminating protein.
2. Ten milliliters of fresh whole blood should yield between 200 and 1000 µg of genomic DNA. Following resuspension in TE, the DNA concentration can be determined by calculating A_{26O}, with an OD of 1.0 corresponding to 50 µg of DNA. The $A_{260/280}$ ratio can also be calculated to determine the protein level in the sample. Clean DNA should have a ratio of approx 1.6; a higher ratio implies contaminating protein and a lower ratio implies contaminating RNA. If contaminating protein is present, the sample can be reincubated with proteinase K, and contaminating RNA can be removed by incubation with RNase. Once the DNA purity is satisfactory, a 50 ng/µL working stock should be made ready for PCR.
3. Fluorescent markers can be purchased from a commercial supplier such as Genset (distributed by Helena Bioscience, Sunderland, UK) or Perkin-Elmer (Warrington, Cheshire, UK).

4. This is equivalent to taking 2.5 μL of an average PCR reaction into a final volume of 50 μL; i.e., for a genescan mix of 5 markers, take 2.5 μL each and add to 37.5 μL of dH_2O. Genescan mixes should ideally be made the day before the genescan is to be run, to allow adequate mixing of the samples. Mixes can be made just before running, but in our experience the quality of the genescan output is inferior.
5. The end-labeling protocol provides enough labeled primer for approx one hundred 15-μL PCR reactions—enough for one 96-well plate if automation is used.
6. If a large number of PCR reactions are to be performed, as is the case in a genomewide search, it is highly advantageous to consider some form of automation, either by simple multichannel pipetting or full automation on a robot such as the Biomek 1000. Both allow 96-well plates to be used, giving a total PCR reaction volume of 15 μL plus oil (oil is not necessary with heated-lid PCR machines, further speeding up sample preparation time). With automation only one stock mix per plate is required, incorporating everything except DNA. This is provided from a separate 96-deep-well tray, which means that every 96-well plate has the same template, thus decreasing the chance of the pipetting errors associated with manual handling.
7. Following PCR the samples are ready for gel electrophoresis. For radioactive PCR reactions, simply add the labeled primer in place of the fluorescent primer in the PCR protocol detailed in **Subheading 3.3.**
8. Following PCR, samples should be checked by agarose gel electrophoresis (run 5 μL of the 15-μL reaction on a 2% agarose gel) to estimate DNA concentration. A 15-μL PCR reaction will typically yield 25–200 ng/μL of DNA. These estimates are helpful since loading too much DNA onto a genescan invariably leads to "bleeding" of one colored dye into another, thus restricting the number of samples that can be analyzed at any one time.
9. The genescan mixes are run on denaturing urea gels, and plate cleaning, gel casting, and sample running should be carried out in accordance with the user's manual. However, we have found that the following gel recipe appears to give slightly better results than that detailed in the user's manual:
   a. Add 18 g of urea, 18 mL of dH_2O, and 5.7 mL of 40% (29:1) acrylamide/bisacrylamide solution to a clean glass beaker and stir until dissolved.
   b. Make up to 50 mL with dH_2O and add 250 μL of 10% APS and 35 μL of TEMED to polymerize the gel. Mix gently and pour.

Acknowledgment

We thank the Coeliac Society for granting a fellowship.