Genotyping SNPs Associated With Dyslexia

Ann Yezerski

King's College, Biology Department, 133 North River St, Wilkes-Barre PA 18711 USA
(**annyezerski@kings.edu**)

This exercise is designed to teach every aspect of a genotyping project by using a Single Nucleotide Polymorphism (SNP) in the KIAA0319 gene that is known to be associated with dyslexia. Initially, students extract their own DNA and then set-up subsequent PCR reactions and restriction digests in order to determine their genotype. Concurrent exercises in Bioinformatics demonstrate how a genotyping protocol is designed, as well as how phenotype can vary due to a single base pair change. As a result, this exercise demonstrates the multitude techniques that a research scientist uses to investigate how SNPs contribute to human phenotypes.

Keywords SNP, dyslexia, bioinformatics, genotyping, PCR, proteomics

Introduction

The use of Single Nucleotide Polymorphism (SNP) genotyping is becoming prevalent in the studies of human disease as it becomes more obvious that a single base pair change in the sequence of DNA can lead to drastic phenotypic change. Additionally, it is becoming more apparent that the interaction between genes, or epistasis, plays an integral role in the determination of phenotype. Dyslexia, a learning disability that affects many reading and writing skills, is an excellent example of a human condition that has several SNPs located on several chromosomes that contribute to the presence and severity of the trait (Williams and O'Donovan 2006). Out of the dozen or so loci claimed to be associated with dyslexia, a gene on chromosome 6 known as KIAA3019 has the strongest support in the literature (Cope, Harold et al. 2005; Paracchini, Thomas et al. 2006). Within this gene, two SNPs have been identified as being highly associated with a dyslexia diagnosis, especially when considered epistatically (Cope, Harold et al. 2005).

This laboratory exercise is based upon a genotyping protocol designed as part of our undergraduate research program. This protocol uses PCR to amplify one of the two important SNPs in the KIAA3019 gene followed by a restriction enzyme digest to determine genotype. While a there are several kits on the market that use similar genotyping techniques, this exercise is unique in that the SNP being examined directly relates to a human phenotype (see Ethical Issues in the Notes for Instructor). Furthermore, the laboratory is not limited to the genotyping exercise. Worksheets included in this manuscript include an exercise in how a genotyping protocol is designed using bioinformatics. Another worksheet demonstrates how proteomics can be used to determine that this particular SNP causes the resulting protein to be unable to insert itself into the cell membrane in the cerebellum of the brain, thus possibly causing in some of the symptoms of dyslexia.

The complete exercise is designed to take four weeks of laboratory time with the aforementioned worksheets designed to be used during the waiting times typical of molecular biology labs. Portions of the lab can be eliminated depending on the level of the student; freshman could simply genotype themselves while upperclassmen could complete the full exercise.

Student Outline

Student Handout 1

Genotyping Single Nucleotide Polymorphisms (SNPs) Associated with Dyslexia

Human beings are said to be about 99.9% similar to each other in their DNA. Although it may seem that we do not differ significantly from each other, the human genome has close to 3 billion base pairs of DNA contained in 23 chromosomes. Therefore, there are about 300,000 variable base pairs in the genome, or, on average, a sequence variation every 1000 base pairs of the DNA sequence. When the population of the human species varies at individual base pair sites, they are known as Single Nucleotide Polymorphisms (or SNPs). In other words, the majority of people may have a "G" at one site in the sequence where the minority may have an "A." These slight differences in the DNA sequence (genotype) are unlikely to have a significant effect on a person's phenotype since it is estimated that there only about 20,000 to 30,000 genes (or less than 1.5% of the genome) that actually code for proteins. Occasionally though, one single base pair can significantly affect a person's health, such as in the case of sickle cell anemia or cystic fibrosis.

Diseases or conditions caused by one SNP are rare. Most human traits are the result of a summation of the effects of several different genes, each of which could contain several different SNPs. Dyslexia, a learning disability that affects reading and spelling, is an example of such a polygenic trait. Scientific research has found over a dozen areas, or loci, within the human genome have an affect on the occurrence of dyslexia. Today we are examining only one, a site within a gene known as KIAA3019 on chromosome 6. This gene is strongly associated with dyslexia and seems to have an effect in the cerebellum of the brain, which is important to control of motor activity.

One of the classic ways of determining genotype when a sequence varies by only one base pair is to use a method known as Restriction Fragment Length Polymorphism (RFLP). In the modern version, scientists couple the Polymerase Chain Reaction (PCR), which makes billions of copies of one specific region of DNA, with a Restriction Digest that cuts the DNA only where an enzyme recognizes a very specific sequence. The result, when the DNA is separated by electrophoresis on an agarose gel, is a banding pattern that can differentiate between a person who is heterozygous or homozygous in their genotype.

In this laboratory exercise you will extract your own DNA from a cheek swab, use the PCR reaction to amplify a segment of the KIAA0319 gene, and then digest the amplified DNA with the restriction enzyme HhaI to determine your genotype for this dyslexia-associated locus. Although the result does not predict whether you do or will have dyslexia, it will give you insight not only about the process of genotyping SNPs but also a bit of knowledge about a tiny piece of your own DNA.

Procedure

DNA Extraction (WEAR GLOVES!)
- Obtain a collection swab for your instructor and label the casing with your name and date.
- Rinse your mouth with water thoroughly
- Swab the inside of EACH cheek twenty times.
- Immediately put the swab into the extraction solution in the screw-capped tube. Swish the swab in the solution at least five times and squeeze against the side before withdrawing the swab. DO NOT TOUCH the swab with your hands.
- Tightly close the tube, vortex for 10 seconds.
- Incubate for 10 minutes at 65°C.
- Vortex again for 15 seconds.
- Heat the tube in the 98°C for 2 minutes. Vortex for 15 more seconds.
- Quantify the DNA extraction using spectrophotometry with your instructor's help.

PCR I
- In a 0.2 ml PCR tube, mix the following:

Sterile water	9.5 μl
GoTaq Master Mix	12.5 μl
Upstream primer (1:10)	0.5 μl
Downstream primer (1:10)	0.5 μl
DNA template	2 μl

- Give your tube to your instructor to be put in the thermal cycler.

PCR II
- Repeat the PCR mix as above except use the PCR I product as the DNA template.

Flash Gel Check
- In a fresh 0.2 ml tube mix 10 μl of your PCR sample with 2 μl of the loading dye.
- Load 10 μl of this mix into the Lonza FlashGel cartridge.
- Your instructor will start and photograph the gel.

Restriction Enzyme Digest
- In a 0.2 ml PCR tube, mix the following:

Sterile water	17 μl
10x FastDigest Buffer	2 μl
PCR product	10 μl
HhaI enzyme	1 μl

- Mix gently and spin down.
- Incubate in the 37°C water block for 5 minutes.

Agarose Gel Electrophoresis
- Create a 1.5% gel by mixing 1.5 g of agarose for every 100 ml of 1x TBE
- Microwave the mix to boiling, let cool and then add 2 μl Ethidium bromide.
  - **N.B. EtBr is an EXTREMELY toxic mutagen. Wear gloves and handle very carefully!**
- Make sure the gel mold is set-up and the proper comb is added. Pour the agarose slowly into the mold and let set until solid.
- Mix 10 μl of your restriction digest with 2 μl of the loading dye.
- Load 10 μl of this mix into the agarose gel.
- Run the gel for 30-40 minutes @ 120 V.
- Document the gel with a UV light.
- Refer to the Genomics worksheet for expected results and score your genotype.

Analyses
- Summarize the class findings of genotypes in a table.
- Use the principles of Hardy-Weinberg to calculate the allele frequencies of our population.

> According to HapMap, the database of SNP genotypes, this SNP has the following frequencies in the United States population:
>
> | | |
> |---|---|
> | +/+ | 0.386 |
> | +/mutant | 0.456 |
> | mutant/mutant | 0.158 |
>
> What are the allele frequencies in the general population?
>
> Use a Chi-square test to see if our population differs from HapMap's population in both allele frequencies and genotypic frequencies.
>
> Explain your results. Keep in mind that our population is college biology students in their sophomore year.

Student Handout 2

Introduction to Bioinformatics

Bioinformatics is an emerging field that combines knowledge of biology with the power of computer analysis. Modern biological research generates very large databases that contain so much information that they cannot be fully analyzed with traditional mathematics and statistics. One of the prime examples of this is that of DNA sequence data. An organism's genome could be millions, billions, or even trillions of base pairs of information. Finding the important part of the sequences (such as the part that codes for proteins), comparing individuals' sequences, or even comparing various organisms to each other requires tremendous computational power. Over the last few decades there has been an effort to compile all DNA sequence information in one gigantic database known as GenBank. Additionally, there are many software programs designed specifically to search this database and then perform the many analyses necessary to better understand information locked in DNA sequences from molecular function to evolution.

In this introduction to Bioinformatics, you will be to play with some of these analyses and start to discover how modern biologists use the information stored in GenBank. You will then use the programs to design a method for genotyping people for a gene that is associated with dyslexia.

I. Searches
  a. In the DNAStar folder, open the program EditSeq.

  b. Under Net Search, choose, New Text Search.

  c. In the box where it says New Term, type the word "human."

      How many "hits" did you get? _____

  d. Close the search results, and start a new search as above. Type the name of any organism.

      How many "hits" did you get? _____

  e. Start as new search and again type "human" and now use the "+" to add the term "dyslexia" (make sure to use "AND" and not "OR")

      Now, how many hits do you get? _____

      Are all of the results human genes? Explain why not.

  f. Of the human results, which chromosomes are mentioned? _____

  g. While in your results window, under Net Search select "Search These Results." Use the keyword "Chromosome 6." Double click on the search result #134304839. This opens an Internet window with the details of this entry. This is the gene of interest. The file includes all publications referencing this entry, a translation of the gene, notes on areas of interest and, most importantly the sequence itself.

      From this window, what did you learn about the KIAA0319 gene?
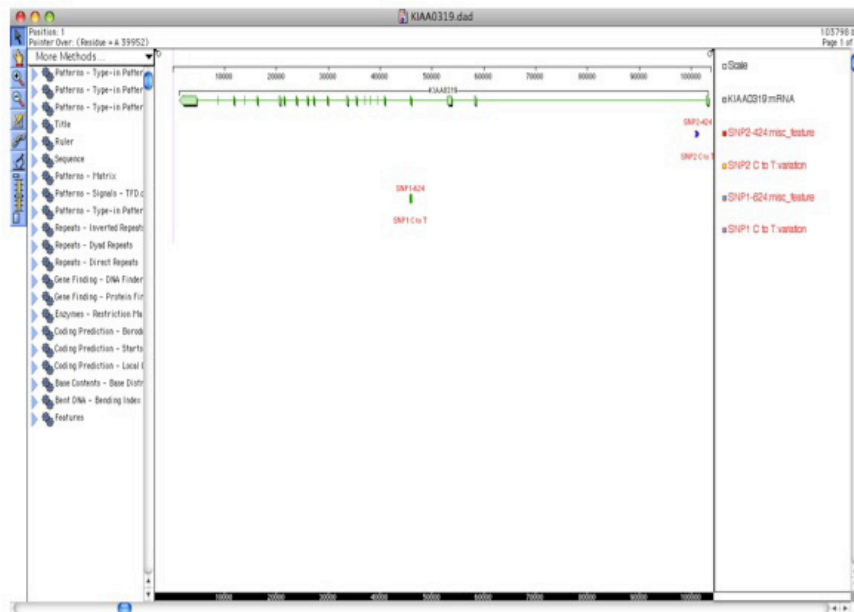
Figure 1. An example of how the KIAA0319 gene should look on GeneQuest.

II. Finding important areas

a. Within the DNAStar folder, open the program "GeneQuest."

b. Open the file "KIAA0319.dad" You will see a window similar to the one above, which represents the sequence for the KIAA0319 gene. The center part of the window is the main workspace, known as the assay surface. The left part lists various things you can do to your sequence. The right part of the window shows which methods have been applied to the sequence.

c. CONTROLS

1. ZOOM: Click on the magnifying glass with the "+" sign in the upper left, and then click on the main assay surface. To reverse, click on the "-" magnifying glass.

2. SCROLL: Use the arrows at the bottom and right sides of the screen to move the assay surface similar to any Office file.

3. APPLYING METHODS: Choose a method from the list on the left (clicking on the triangles opens multiple methods under that category). Click and drag the method to the assay surface to apply it to the sequence. Click on the method name in the right window to move or delete the method.

4. ANALYSES: Under the Analysis menu you can get information such as base composition and RNA folding. If you highlight a potion of the sequence, the analysis will only be done on that portion.

d. Work with your instructor to play with the controls and familiarize yourself with the program, then answer the following questions:

1, How big is the gene in its entirety? _____bp

2. What percentage of the sequence are G's? _____

3. Which amino acid is coded for most often? _____

4. Find the sequence CGCTCTGCGCGGGGC, at which base pair does it start? _____

III. SNPs

e. There are two SNPs labeled on this gene from previous dyslexia studies: SNP1 and SNP2.

1. Compare the areas of the two SNPs to the mRNA (transcribed) sequence. Based on this, are the SNPs in an intron or an exon? Explain how you know and what this means.

2. How could a polymorphism in an intron have an effect on the final protein?

f. Use the ZOOM feature to concentrate just on the SNP1 region around 45,000 bp. This area is the subject of our SNP lab. We will be using PCR to amplify the 1038 bp region and then nested PCR to amplify the 624 bp region.

g. Separately, the one bp difference, a change from a "C" to a "T" is labeled. In other words, most people have a "C" here while some individuals have a "T." We will find out which allele(s) you have.

h. Primers have been created specifically to amplify these regions. Primers have the complementary sequence to the DNA template. If the primers were made to be 20 bp long, what would their sequences be to amplify the 624 bp region?

Upstream primer _____

Downstream primer _____

IV. The program can also determine Open Reading Frames, or potential areas of transcription. Under the Method "Starts Stops ORFs", choose all of the six ORFs options and drag them to the assay surface.

a. How many potential ORFs are completely contained within the SNP1-624 bp region? _____

b. Tough one: Why are there six possible ways to define an ORF?

V. Genotyping

c. Genotyping of this SNP will be done using the RFLP method. After using PCR to amplify the 624 bp region, a specific restriction enzyme, HhaI will be used to find and cut the product at the sequence "GCGC."

d. Under "Enzymes – Restriction Map" choose the enzyme HhaI and drag it to the assay surface.

1. If we digested the whole gene instead of just the 624 bp sequence, how many pieces would we get? _____

2. Instead, we are just cutting the 624 bp product. How many pieces and what sizes do you get with the RFLP method?

_____

3. The mutation to a "T" at this spot prevents a cut from happening. Draw a hypothetical agarose gel showing the results. Start with a ladder that has a band every 100 base pairs from 100-1000 bp in Lane 1. In Lane 2 show a homozygous wild type, in Lane 3 show a heterozygote, and in Lane 4 show a homozygous mutant.

Student Handout 3

Dyslexia SNP transcription and translation

Below is the part of the sequence for the KIAA0319 gene. This portion of the sequence has been shown to be transcribed and then translated as part of the final protein product.

1
TTAAAAGTTA CCTGTCCTGG GAGCAGTGGT AGGAGATATG GGTAGCTCAG

61   *
ATGGGGTGGA CTCAGAGGGG GCTG**C**GCTAG TGGGAGGTGT TGGGATGCTG

111
TGCTCTGTAC TCCCCGGGGT GACTGTGAGC ACTGGGCTTT TCTCCACGGT

161
GACTGAGCTG AGCTCCAGGC TTGCCGGAGG AAGACTATGG GAAGGCATTA

211
GAACC

Remember that there are six (6) possible ways that this DNA sequence could be transcribed. Only one is the correct one. In this case the correct version has a single Open Reading Frame (ORF) that takes up almost the whole DNA sequence shown here.

1) At which base pair does the correct ORF start? End? In which direction is it read (left to right OR right to left)?

2) Once you have determined the correct ORF, give the entire transcription of the segment below:

3) Starting at the start codon, translate the mRNA sequence until the stop codon using the table in Figure 6.7. Give the order of the amino acids <u>using the single letter abbreviation for each amino acid</u> (Figure 6.7 gives both the three letter and one letter abbreviations).

4) The SNP is shown in the sequence bold and underlined with an asterisk. (You can't get much more obvious). Highlight the affected amino acid above (#3). In the "mutant" version of the gene, this C is a T. How does that change the translation?

5) Based on what you know about the side groups, how do you think this SNP could affect the folding final protein?

6) As it turns out, this protein product is known *to insert itself into the cell membrane*, especially in the cerebellum of the brain. How could this change of amino acids have an effect on the ability of this protein to do its job?

Second letter

	U	C	A	G	
U	UUU Phe (F) UUC UUA Leu UUG (L)	UCU UCC Ser UCA (S) UCG	UAU Tyr (Y) UAC UAA Stop UAG Stop	UGU Cys (C) UGC UGA Stop UGG Trp (W)	U C A G
C	CUU CUC Leu CUA (L) CUG	CCU CCC Pro CCA (P) CCG	CAU His (H) CAC CAA Gln CAG (Q)	CGU CGC Arg CGA (R) CGG	U C A G
A	AUU AUC Ile (I) AUA AUG Met (M)	ACU ACC Thr ACA (T) ACG	AAU Asn (N) AAC AAA Lys AAG (K)	AGU Ser (S) AGC AGA Arg AGG (R)	U C A G
G	GUU GUC Val GUA (V) GUG	GCU GCC Ala GCA (A) GCG	GAU Asp (D) GAC GAA Glu GAG (E)	GGU GGC Gly GGA (G) GGG	U C A G

First letter (left side), Third letter (right side)

🟥 = Chain termination codon (stop)
🟩 = Initiation codon

© 2010 Pearson Education, Inc.

Figure 2. (Taken from Figure 6.7 from iGenetics: A Molecular Approach, 3rd Edition by Peter J. Russell). The genetic code table showing which amino acids is coded for by which codon

Materials

Lab subsection	*S or E?	Description	Supplier	Qty. for 25	Catalog #
Bioinformatics	E	LaserGene software suite. Educational version	DNAStar	Depends	N/A
DNA extraction	S	BuccalAmp extraction kits	Epicenter	1 kit	BQ0908SCR
	E	Vortexers	N/A	N/A	N/A
	E	Heat blocks (65 and 98°C)	N/A	N/A	N/A
	E	UV-Spectrophotometer	N/A	N/A	N/A
Nested PCR	S	Primers SNP2-1038	IDT	10 nMoles	Custom order
		5'-GGCATCCCTGGACCCTGTTAGTGG-3'			
		5'-TGGCGAGGGTTTCTTGATTTATAGGTAGTC-3'			
	S	Primers SNP2-624	IDT	10 nMoles	Custom order
		5'-GTGGAAGGAGGTGTGGGGCAGAAA-3'			
		5'-CTGAAAAATGTGCCTGGAGGGAATGAGTAA-3'			
	S	1.5 ml tubes	USA Scientific	25	1615-5500
	S	0.2 ml tubes	USA Scientific	25	1402-8120
	E	Various micropipettors with tips	N/A	N/A	N/A
	S	GoTaq Master Mix	Promega	100 rxns.	M7133
	S	Sterile water	Sigma		95284
Restriction Digest	S	FastDigest Hha I w/ buffer	Fermentas	200 rxns.	FD1854
	S	0.2 ml tubes	USA Scientific	25	1402-8120
Gels	S	Agarose	Sigma	Varies	A4718
	S	1x TBE (from 10x)	Sigma	Varies	T4415
	S	Loading dye	Sigma	~ 500 µl	G2526
	E	UV light source	N/A	N/A	N/A
	S	FlashGels	Lonza	2-3 gels	57067
	E	Gel rigs	N/A	N/A	N/A
	S	FlashGel Ladder	Lonza	1	57033

* Equipment or Supply?

Notes for the Instructor

Ethical issues

Since you are extracting student DNA and you are genotyping a gene relevant to a human condition, there can be ethical issues associated with this process. Institutional Review Boards do not require approval for laboratory work used for educational purposes, under which this process could fall. However, it is important to contact the IRB at your institution in order to ensure that you are fulfilling the requirements of human scientific use on your campus. Any DNA collected from the students should be given an ID number in order to dissociate the student names from their results. Additionally, it is critical to make it clear to the students that the results of this genotyping process are NOT diagnostic. Any results obtained will not in any way predict the students' ability in coursework nor suggest a medical diagnosis of dyslexia.

Introductory PowerPoint

A notated PowerPoint presentation of all of the background material for this gene and this protocol is available upon request (**annyezerski@kings.edu**).

Bioinformatics

The LaserGene suite can be obtained for free in yearlong license increments for educational purposes. Contact:

> DNASTAR Inc.
> 3801 Regent Street
> Madison, WI 53705
> Direct Phone Line: 608-237-3054
> Fax: 608-258-7439

An answer sheet will be available for all exercises.

A pre-prepared data file for the KIAA3019 gene for GeneQuest is available upon request.

Free Bioinformatics services are also available at the NCBI website: **http://www.ncbi.nlm.nih.gov/** and this exercise could certainly be done with these tools. However, a new worksheet would have to be designed for the Genomics portion of the exercise.

DNA Extraction

It is very important that the students understand the concepts and results of contamination, especially increase in apparent heterozygotes that result from cross-contamination.

DNA can be frozen for future use. If frozen at -70°C it is useful almost indefinitely with this procedure.

Nested PCR

The PCR program takes about 1.5 hours on an MJ Research Thermocycler.

A standard (not nested) PCR reaction has also been tested to work effectively, but a nested PCR will give optimal results.

You can use FlashGels at any time to check the progress of the PCRs, but keep in mind that they are much more expensive per gel than a standard agarose gel. (They are also "flashy" and impressive). **www.lonza.com**

Restriction Digest

Although you can use standard HhaI enzyme, the FastDigest version from Fermentas not only takes just five minutes and it results in extremely consistent banding patterns. Addition of loading buffer is enough to stop the reaction.

Gels

FlashGels are fast but expensive. A 2.0% agarose gel is best for scoring. Gels can be stained with any standard stain, in the case of the Nested-PCR procedure you have more than enough DNA to visualize.

Scoring

I prefer to give all students the complete data set. That number is around 100 for our classes.

If you do this exercise, I would ask that you send your results to me to include in our database if, and only if, you have gotten IRB approval for the exercise and provide a signed informed consent form for each sample.

Acknowledgements

Literature Cited

Cope, N., D. Harold, et al. (2005). Strong evidence that KIAA0319 on chromosome 6p is a susceptibility gene for developmental dyslexia. *American Journal of Human Genetics* 76(4): 581-591.

Paracchini, S., A. Thomas, et al. (2006). The chromosome 6p22 haplotype associated with dyslexia reduces the expression of KIAA0319, a novel gene involved in neuronal migration. *Human Molecular Genetics* 15(10): 1659-1666.

Williams, J. and M. C. O'Donovan (2006). The genetics of developmental dyslexia. *European Journal of Human Genetics* 14(6): 681-689.

About the Author

Ann Yezerski, Ph.D. received both a M.S. and a Ph.D. in Molecular Genetics from the Univeristy of Vermont. She has been a faculty member at King's College since 1999 and has been the chairperson of the Biology Department since 2009. Although her research has historically used *Tribolium* beetles, Dr. Yezerski has recently been applying these same molecular genetics methods to human conditions. Her interest in the topic of dyslexia has come about from a combination of a neuroscience colleague's interest in the genetic basis of the trait along with the discovery that Dr. Yezerski's own twin boys have a learning disability and are homozygous for the mutant version of the SNP explored in this exercise. This exercise was designed to be used in King's College sophomore level Genetics course which Dr. Yezerski teaches along with Bioinformatics, Physiology and Introductory courses.

Appendix A: Sample results
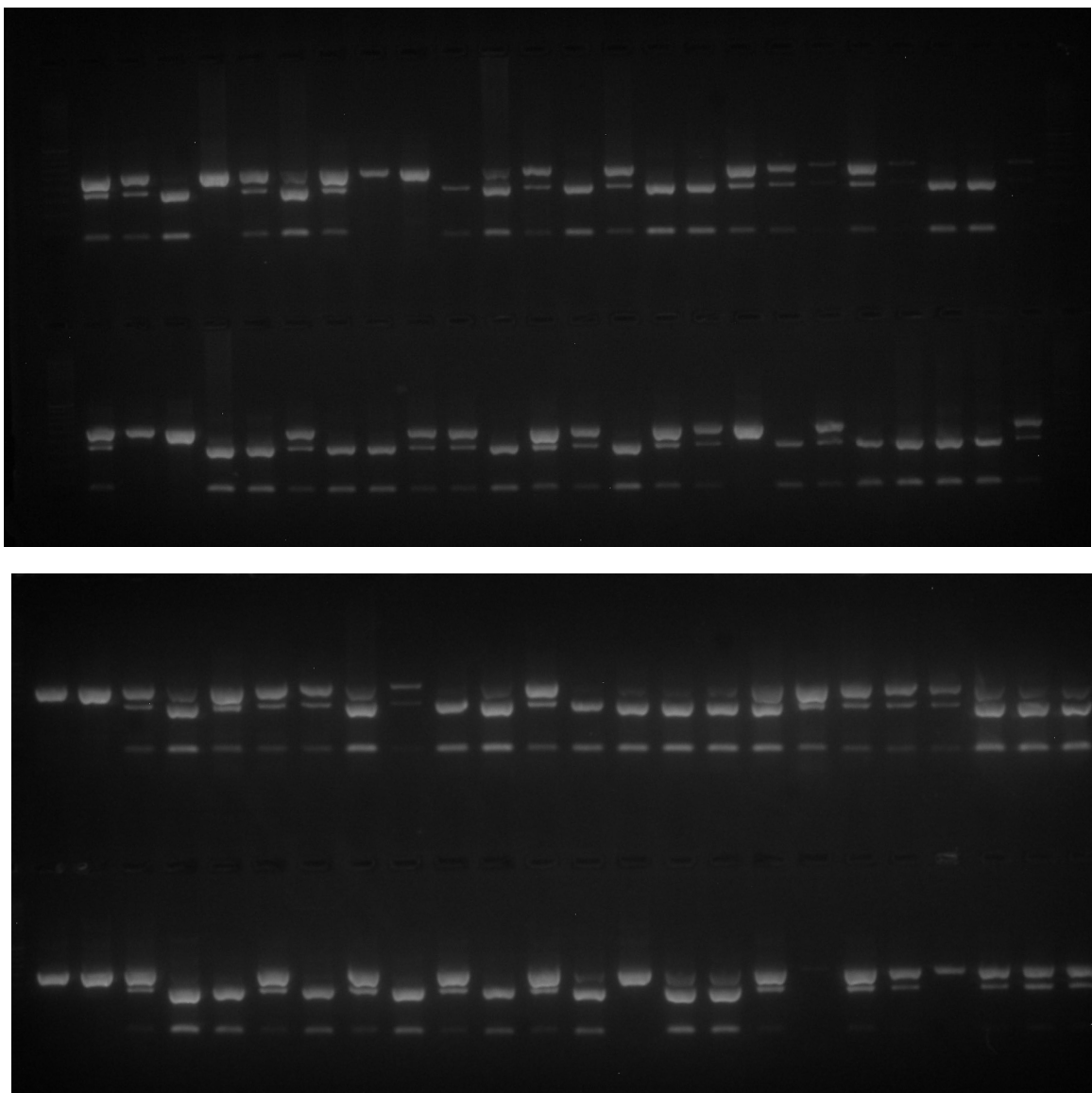
Figure 3. Example results actually obtained from a King's College genetics class. The single band pattern represents a homozygous mutant, the double-banded pattern is homozygous wild type and the three-banded pattern is a heterozygote.

Mission, Review Process & Disclaimer

The Association for Biology Laboratory Education (ABLE) was founded in 1979 to promote information exchange among university and college educators actively concerned with teaching biology in a laboratory setting. The focus of ABLE is to improve the undergraduate biology laboratory experience by promoting the development and dissemination of interesting, innovative, and reliable laboratory exercises. For more information about ABLE, please visit **http://www.ableweb.org/**

Papers published in *Tested Studies for Laboratory Teaching: Proceedings of the Conference of the Association for Biology Laboratory Education* are evaluated and selected by a committee prior to presentation at the conference, peer-reviewed by participants at the conference, and edited by members of the ABLE Editorial Board.

Although the laboratory exercises in this proceedings volume have been tested and due consideration has been given to safety, individuals performing these exercises must assume all responsibilities for risk. ABLE disclaims any liability with regards to safety in connection with the use of the exercises in this volume.

Citing This Article

Yezerski, A. 2011. Genotyping SNPs Associated With Dyslexia. Pages 225-236, in *Tested Studies for Laboratory Teaching,* Volume 32 (K. McMahon, Editor). Proceedings of the 32nd Conference of the Association for Biology Laboratory Education (ABLE), 445 pages. **http://**www.ableweb.org/volumes/vol-32/?art=18