Independent Gene Discovery and Testing

Vrushalee Palsule, Dijana Coric, Russell Delancy, Heather Dunham, Caleb Melancon, Dennis Thompson, Jamie Toms, Ashley White and Jeffry Shultz*

School of Biological Sciences, 1 Adams Blvd., Louisiana Tech University, Ruston, LA 71272

*Email: jlshultz@latech.edu

Abstract: A clear understanding of basic gene structure is critical when teaching molecular genetics, the central dogma and the biological sciences. We sought to create a gene-based teaching project to improve students' understanding of gene structure and to integrate this into a research project that can be implemented by instructors at the secondary level or higher. We modified a previous project performed within a genetics course in which students designed and tested polymerase chain reaction (PCR) primers based on expressed gene products from specific biochemical pathways. Students used these skills to investigate a gene that they hypothesized would be related to the multiple sclerosis phenotype in humans. They were each required to identify the gene structure (exons and introns) of their selected gene and to design primers around these elements. A total of seven genes hypothesized to interact with the Multiple Sclerosis (MS) phenotype were used to design 154 primers (77 forward and 77 reverse). Based on the structure of each gene, these primers were combined as 83 unique pairs and tested by seven students. A total of 25 pairs were visibly polymorphic and 58 pairs were monomorphic. All materials, including primer design protocols, modifiable gene graphics files, screening preparation documents and assessment information are presented to allow replication of this project.

Keywords: Undergraduate research, Gene structure, PCR, Primer design, Independent project

INTRODUCTION

Goals, Expected Outcomes and Background

The central dogma was first proposed by Francis Crick in 1958 (Crick, 1970; Crick, 1958). It proposed a relationship between DNA, RNA and polypeptides. The transfer of information between these molecules is vital to living organisms and can be used to understand sources of biological data and how they are obtained. It is also very useful when describing how the polymerase chain reaction relates to the replication of DNA *in-vivo*.

An understanding of gene structure is critical when describing the central dogma, transcription, post-transcriptional modification of mRNA and translation. In order to fully illustrate this concept, we sought to introduce students to several key research resources, including online biochemical pathways at the Kyoto Encyclopedia of Genes and Genomes website, expressed sequences attributed to the genes in these pathways, bioinformatics in the form of BLAST at the National Center for Biotechnology Information (NCBI) and designing primers for the polymerase chain reaction (PCR) using the Primer3 program (Rozen and Skaletsky, 2000).

Expressed sequences are used extensively to create libraries of genes for different organisms. Of primary importance is the fact that these genes have been through the transcription process within the organism. Comparing these transcribed and edited sequences to known gene sequences (even those from another organism) allows their identification as an enzyme in a particular pathway.

A drawback of these sequences is that they represent mature mRNA product, from which introns have been removed. This makes an understanding of the actual chromosomal structure difficult to discern without further analysis.

A comparison between the mature mRNA product and chromosomal DNA using the BLAST utility at NCBI can be used to determine the chromosomal location (in base pairs) of the gene and its introns and exons (NCBI, 2009). This information can be used to generate an ideogram of the gene and to illustrate how a gene can be post-transcriptionally modified.

The polymerase chain reaction exponentially replicates DNA with specificity derived from short oligonucleotides (primers) designed from the target DNA. In order to illustrate this procedure, students can design and test primers based on information obtained from a comparison of expressed vs genomic DNA.

Objectives

We sought to develop a procedure that integrates several aspects of scientific research within an overall educational goal of understanding how genomic sequence, pre-mRNA, introns and exons relate to each other. The introduction of students to as complete a research experience as possible within the confines of an academic year was an equal goal. In order to accomplish these tasks, we executed the

project in four steps, which included planning, grant writing, wet-lab research and reporting results.

Planning included the identification a gene of interest within a biochemical pathway, followed by the creation of an ideogram delineating exons and introns within the genomic region. After a gene was selected and described, students were expected to perform limited grant writing using a template document, indicating reasons for selecting their gene and incorporating their gene ideograms and primers designed into the grant proposal. If approved for funding, these primers were to be ordered and tested for polymorphism. Successful project completion was recognized after presentation of their research results at a college research symposium.

MATERIALS AND METHODS

Equipment and Consumables Required

The following equipment is required to perform this project: A computer with Internet access, thermocycler, horizontal electrophoresis gel rig and power supply, mini-centrifuge, heat block/water bath/incubator, pipettors and documentation capability. Consumables include: PCR tubes, Pipette tips, agarose, TBE Buffer, Taq Master Mix, TE buffer and micro-centrifuge tubes. Students taped their gel pictures into a combined lab notebook as they were generated.

Source of Human DNA

The DNA used for this study was provided by students participating in undergraduate genetics labs at Louisiana Tech University. All DNA was extracted using the included protocol and information such as height, weight, eye and hair color, birthplace and family history of MS or Cystic Fibrosis was obtained. As described in Supplemental File 1, students performed an oral saline rinse, followed by extraction using the BioRad Instagene Matrix (cat #732-6030). This project was approved by the Human Use Committee of Louisiana Tech University. All samples were re-labeled to maintain anonymity and all participants signed a Human Subjects Consent Form.

Gene Selection

Students were presented with basic information about the human multiple sclerosis phenotype (Supplemental File 2) which included hyperlinks to proposed pathways. After reading this brief synopsis, each student was encouraged to investigate a gene in a pathway that they felt might be a contributing factor to the MS phenotype. Once a gene was identified, students used the KEGG (2009) database to obtain expressed sequences attributed to that gene. Sequences were compared with the human genome using the BLAST utility on NCBI (2009), comparing against the Human genomic plus transcript (Human G+T) sequences, allowing the creation of gene-specific figures that show exon and intron boundaries (the template for gene figures is also included in Supplemental File 2). The genomic sequence related to each gene was saved and used to design PCR primers.

Primer Design

Primers were designed using the genomic sequences for each gene. All primers were created with the Primer 3 program (Rozen and Skaletsky, 2000) using the default settings. A detailed primer design and rehydration protocol is provided as Supplemental File #3. All primers were ordered from Integrated DNA Technology (IDT; Coralville, IA) at the 25 nmole scale.

PCR Conditions

Components in each reaction included 10 µl 2X PCR mix (cat # M7122, Promega, Madison, WI), 3 µl H_2O, 5 µl DNA (5-10 ng/µl) and 1 µl each of 10 mM forward and reverse primers. PCR conditions varied, based on procedure (testing for function or optimization) with lower annealing Tms used for non-functioning primer pairs and higher annealing Tms used for products that produced multiple or unclear banding patterns. The standard conditions were an initial 94^{o} C denaturation for 5 min, followed by 36 cycles of denaturing (94^{o} C for 60 s), annealing (50^{o} C for 75 s) and extension (72^{o} C for 45 s). After a final extension of 5 minutes at 72^{o} C, PCR product was loaded into a 1.5% agarose (cat # BP160, Fisher Scientific, Waltham, MA) prepared with ethidium bromide (0.1 µl:20ml gel matrix). Electrophoresis was performed @ 4v/cm for 3-4 hours in 0.5X TBE (TRIS:Boric Acid:EDTA) buffer and documented with a UVP Gel documentation system (Upland, CA).

RESULTS AND DISCUSSION

A total of seven students each investigated a single gene that they selected from a pathway, identifying introns and exons within the IL-7, CD28, HSDB3B1, HSD11B1, IL7R, NMUR1, PKA and TSHR genes using genomic sequence and the BLAST program. Examples of two student-designed gene ideograms (HSD11B1 and IL7R) are shown in Figure 1. Students designed PCR primers based on data shown in these figures (example primers from the two genes in Figure 1 are shown in Table 1).

A total of 154 gene-based primers were designed and tested by seven students. Of the 83 tested combinations, 25 were visibly polymorphic and 58 were monomorphic (Figure 2). Although designing primers based on mature mRNA sequence is effective (Shultz, 2008), the likelihood of PCR not working because a genomic primer spans exon boundaries (panel A, Figure 3) or amplifies unknown lengths of

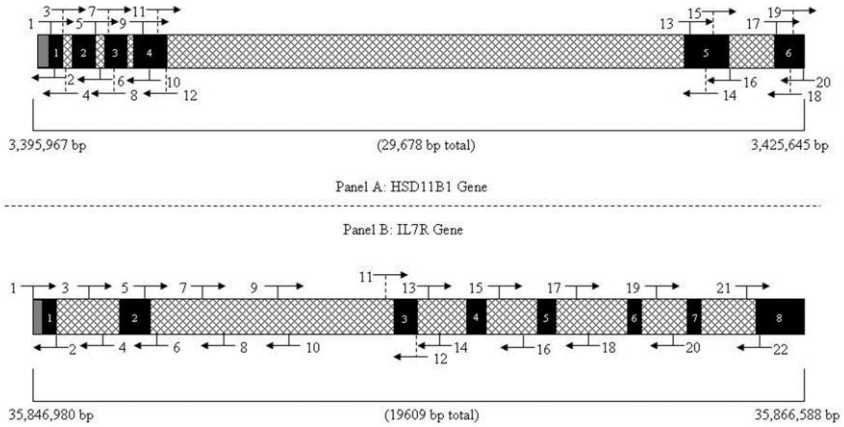

Fig. 1. Graphic representation of 29,678 bp of the human HSD11B1 gene located on chromosome 1 (panel A) and 19,609 bp of the human IL7R gene located on chromosome 5 (panel B). Exonic sequence is numbered and represented by solid black fill, intronic sequence is "hashed" and the promoter region (-100 bp upstream) is grey.

intronic DNA (panel B, Figure 3) can be eliminated by determining the structure of the gene prior to design.

There are two ways to look at the data generated from this research. Multiple products from each sample can be useful because they increase the chance of identifying polymorphism. Multiple products, however, are difficult to replicate using agarose detection methods and must be considered transient in nature. If a polymorphism is observed within multiple bands, it should be excised from the gel and stored. By storing the detected band you make it possible to sequence the product and design primers that amplify that fragment only. Single amplification products are easy to replicate, but generate fewer polymorphisms. If polymorphic between samples, these single products require no further processing to be useful as molecular markers. If a single amplification product is monomorphic, it is still useful in SNP genotyping as a pre-detection step (Lee et al., 2004; Shultz et al., 2008).

This procedure can be performed as several steps, each of which entails increasing educational benefit and cost. The initial identification of exon and intron boundaries is a free process requiring computers with access to the Internet. An additional step of identifying alternatively spliced mRNA can also be performed without cost. Although the design of primers is free, there would be limited instructional usefulness to this step without ordering the oligonucleotides. Initial testing of the primers can be performed at a limited cost (approx $0.20 per reaction), but requires the hardware resources listed in the materials and methods. The primers themselves represent a cost of $0.20 – 0.30 per base (we typically estimate $5 each). Additional testing can be performed on multiple DNA samples, creating a population genetics study but also increasing the cost as primers are used on an expanded DNA set. Additional work can be based on primers found to be polymorphic, such as sequencing of products and/or detection of single nucleotide polymorphisms (SNPs), but this expansion of the protocol requires significant equipment expenditures and/or vastly increased consumables cost, and would not typically be appropriate for course-based undergraduate research

In this study, the instructor decides what phenotype will be the focus of the investigation. This is supported by specific information in Supplemental File #2 for multiple sclerosis. A strength of this procedure is that by modifying only this file, an instructor can focus on their own phenotype of interest. The difficulty level of this project can be increased substantially by requiring students to create their own summary file, then act on their own collected information to select a gene of interest.

Because this research involves consumables such as agarose, pipette tips, buffers and other miscellaneous supplies, a faculty mini-grant equivalent to four student mini-grants (~$500) was also applied for and received. These funds were available for use within the project and included additional funds for primers in case fine mapping of a polymorphism was necessary. In this study, one of the student mini-grants was declined due to a technicality. A portion of the faculty mini-grant was used to cover this particular student's costs.

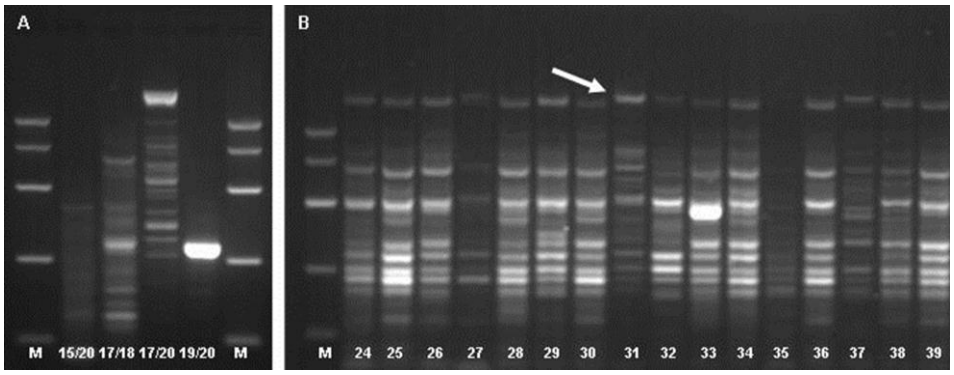

Fig. 2. Panel A) Initial test of primer function for 4 primer pairs within the HSD11B1gene. Primer pairs tested based on proximity to each other predicted by the configuration shown in Figure 1. Panel B) Test of primer pairs from the HSD11B1 gene on members of a population of 46 samples. A polymorphic product is indicated by a white arrow for primers 17 and 20 (subjects 24-39). The "M" annotation indicates a size standard of 100, 250, 500, 750 and 1000 bp.

Student Background and Assessment

The students involved in this project had two, 10 week terms of exposure to the required lab procedures (DeCaire et al., 2008; Shultz, 2008). Each student initially enrolled in an upper-class Genetics course. Students were selected from this course based on class participation and interest in the subject. These students then enrolled in one quarter of lab assisting for the same Genetics course. This process ensured that all students had a good working knowledge of pipetting, knew the location of materials within the lab and were familiar with the general procedure. This level of student preparedness is not necessary, but an instructor could expect an inverse relationship between the time required for training and the background of each student.

Table 1. Primers designed from Human HSD11B1 and IL7R genes. Primer numbers correspond to indicator arrows within the gene ideograms shown in Figure 1.

Primer Number	Sequence	Distance from 5' end	Primer Number	Sequence	Distance from 5' end
HSD11B1_1	AGGAGTCTTCAGGCCAGCTC	100	IL7R_1	CCTCCCTCCCTTCCTCTTAC	1
HSD11B1_2	TCCTCAGGAACACTCAAGCA	187	IL7R_2	TCACCATTTTGAGCATAGCC	186
HSD11B1_3	AAACCTCCCACTTCAGAGCA	800	IL7R_3	TCAGGTCCCAGTGGGTGTAT	3029
HSD11B1_4	GCCCCTGTGACAATCACTTT	964	IL7R_4	AGCAGGATGCTGGAAGAAAG	3236
HSD11B1_5	TGGGTATGGTCCTCACTTCC	964	IL7R_5	CCAGGAGACTTGGAAGATGC	3971
HSD11B1_6	TGCGATATGCTAGCTTCTGTG	1114	IL7R_6	AACCACCATCCCTCACCATA	4129
HSD11B1_7	CCCCAAAAATCTGCAGCTAA	1864	IL7R_7	TTGGGGCCTGTAATTGGATA	5889
HSD11B1_8	CCTCCTGCAAGAGATGGCTA	2014	IL7R_8	GGACAAACATCAGGTGGAAAG	6102
HSD11B1_9	TCCTCTGAACTTTGCCCTTG	2099	IL7R_9	TGTAGTGCTTTTTGTATCCTTGC	7972
HSD11B1_10	ACTTCTCTGCATGGAAACTGA	2344	IL7R_10	TGACGAGTTAATCGGTGCAG	8194
HSD11B1_11	TCAGTTTCCATGCAGAGAAGTA	2400	IL7R_11	TCAGGTTCAATTGCATCTAATTCT	10197
HSD11B1_12	TCATGGAGGAAGCAAAGTCC	2900	IL7R_12	GTGGAGGCATGTTCAGGAAT	10303
HSD11B1_13	CCACCCTATGCCTTCGATAC	27450	IL7R_13	TGGTTGTCACTTTTGGGCTA	10610
HSD11B1_14	CATTGACCCTGGACACTGAA	27650	IL7R_14	CCAGCTCCAGTTAGCCACTT	10850
HSD11B1_15	TTGCTTTGGATGGGTTCTTC	27700	IL7R_15	CTTGGCTGCCCTTTAGACAG	14135
HSD11B1_16	CCACAATGGTAAGGCAGACA	28000	IL7R_16	CCGTCCATTTGTTTTCATCC	14337
HSD11B1_17	GGCTGAGTCTCACCACATTC	28250	IL7R_17	TGCATGGCTACTGAATGCTC	17527
HSD11B1_18	TGGGTCTCGGTCTCTCTCTC	28800	IL7R_18	TGACCAACAGAGCGACAGAG	17637
HSD11B1_19	CCCTACTCTTCCCTTGTCATTC	29459	IL7R_19	CTGTGGTCTCTGGTCCAACC	18426
HSD11B1_20	AACAGTCCCAAAATCCCTCA	29678	IL7R_20	GGAGACTGGGCCATACGATA	18663
			IL7R_21	ACATGCTGGCAATTCTGTGA	18984
			IL7R_22	TCTGGCAGTCCAGGAAACTT	19151

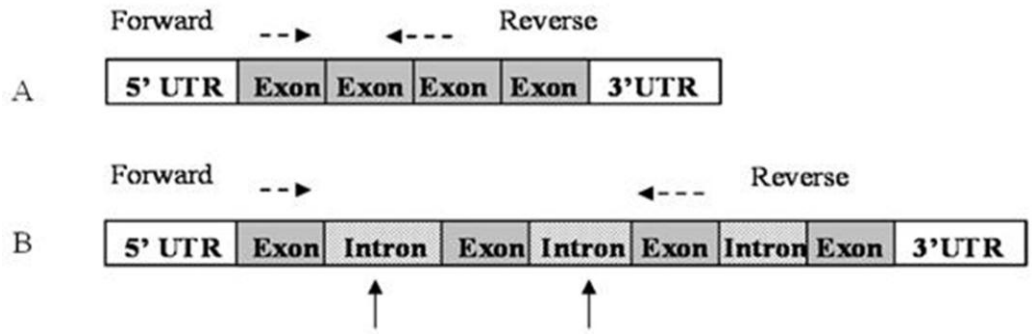

Fig. 3. Segment A shows primer design based on mature mRNA sequence only, with the reverse primer spanning 2 exons. This PCR reaction will fail because the reverse primer will not be able to anneal (bind) to the target (genomic) DNA. Segment B shows primers designed in exons, but with two introns of unknown length in between. If the total distance between the forward and reverse primers is too long, the reaction will likely fail because the target genomic DNA is too long for standard PCR.

An effective implementation of this course may also include the creation of a multi-instructor course, including technical communication for grant and poster preparation and a separate, hands-on laboratory section. In this form, the course would provide excellent preparation for graduate or professional school.

This procedure could also be performed in a standard laboratory course (i.e. not as an independent-study course). Factors that would increase the chance of successful execution include the utilization of students from the previous term as lab assistants, (which decreases the student/teacher ratio) and the use of available instructional materials (see supplemental files). We have found that most students can easily perform this procedure working independently with minimal instructor input. Obvious modifications for a standard course would be the removal of the grant writing component and replacement with a research report or lab notebook for assessment.

Assessment of students was conducted at three points during this independent study project. Completion of the schematic figure and the creation of a primer list were required in order to complete College mini-grant requests during the first quarter. Primer screening, optimization and completion of record-keeping were required for completion of the second quarter. Finally, the preparation and presentation of a poster at a University symposium during the third quarter completed the project.

Technical Discussion

Although we used gradient PCR to optimize the primers, in a large class setting it is entirely appropriate to set up a single thermal cycler at an annealing temperature of 55º C. If the product is still non-specific, the temperature can be raised to 60º C. If the products are faint, the temperature can be decreased to 50º C. The complexity of obtaining the genomic DNA sequence can also be circumvented by simply identifying the exonic boundaries, then designing primers that only amplify within a known exon.

An important concept to remember is that the expressed genes identified on the KEGG database may exist in several forms due to alternative splicing of pre-mRNA. These alternative splices may cause the misidentification of exons as introns. At this time, there are no plans for accounting for this possibility within this procedure, but it could be pursued by comparing multiple copies of the cDNA attributed to the same gene.

CONCLUSION

We present a procedure in which students independently identify a gene's structure, design primers based on this information, test these primers and present their findings. This project demonstrates how an integrated approach to teaching and research can be beneficial to both students and instructors. All materials used in this study are presented within the text or as supplemental material.

ABBREVIATIONS

BLASTBasic local alignment search tool
cDNAComplementary DNA
KEGGKyoto encyclopedia of genes and genomes
mRNA Mature RNA
MSMultiple sclerosis
NCBINational center for biotechnology information
PCRpolymerase chain reaction
SNPSingle nucleotide polymorphism
TBETris, Boric acid and EDTA buffer
TETris and EDTA buffer
TmMelting temperature

ACKNOWLEDGEMENTS

We would like to thank the College of Applied and Natural Sciences for the provision of Mini-grant funds for DC, HD, CM, DT, JT, AW and JS. We would also like to thank the Louisiana Tech Student Technology Fee Board (2007) for the provision of four thermocyclers.

AUTHOR CONTRIBUTIONS

VP acted as a resource to DC-AW and wrote portions of the manuscript. DC-AW performed the in-lab portions of the study and are listed in alphabetic order. JS conceived of the work, created the support documents, wrote portions of the manuscript and supervised the study.

REFERENCES

DECAIRE, J., BRUMETT, S., PAYNE, J., OAKLEY, J., RAY, J., *et al*. 2008. Gene duplication in soybean and polymorphism in the steroid biosynthesis pathway indicate highly conserved gene sequences. *Soybean Genetics Newsletter* 35:1-10.

KEGG. 2009. Kegg Database. Accessed from http://www.genome.jp/ in October, 2009.

LEE, S., WALKER, D., CREGAN, P., and H. BOERMA. 2004. Comparison of four flow cytometric SNP detection assays and their use in plant improvement. *Theor appl genet* 110:167-174.

NCBI. 2009. National Center for Biotechnology Information. Accessed from http://www.ncbi.nlm.nih.gov/ in October 2009.

ROZEN, S., and H. SKALETSKY. 2000. *Primer3 on the WWW for general users and for biologist programmers*. Humana Press, Totowa, NJ.

SHULTZ, J. 2008. Gene Polymorphism Studies in a Teaching Laboratory. *Jrnl of Sci Ed and Tech*. 18:48-59.

SHULTZ, J.L., RAY, J.D., and J. SMITH. 2008. Mapping two genes in the purine metabolism pathway of soybean. *DNA Seq.* 19:264-9.