CONSTRUCTION, APPLICATION AND ANALYSIS OF THE OLIGONUCLEOTIDE DATABASE, VIROLIGO

By

KENJI ONODERA

Bachelor of Engineering
Akita University
Akita, Japan
1995

Master of Science
Oklahoma State University
Stillwater, Oklahoma
1999

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
DOCTOR OF PHILOSOPHY
December, 2002

CONSTRUCTION, APPLICATION AND ANALYSIS OF

THE OLIGONUCLEOTIDE DATABASE, VIROLIGO

Thesis Approved:

Ulrich Melcher

Thesis Adviser

Richard C Essenberg

Robert L Matts

Timothy J. Pettibone

Dean of the Graduate College

ACKNOWLEGMENTS

I would like to express my great appreciation to my major advisers, Dr. Ulrich Melcher who supported, advised, and lead directions of the research. I am thankful for his time spending on reading all chapters of the draft multiple times and provided insightful comments. I also thank my other committee members, Dr. Jean d'Offay, Dr. Richard Eberle, Dr. Richard Essenberg and Dr. Robert Matts for offering suggestions on my research. Dr. Alexander Lai and Dr. Srikumar Sengupta's help on providing experimental samples, results and suggestions were also greatly appreciated. Thanks also to the Department of Biochemistry and Molecular Biology and Noble Foundation (Ardmore, Oklahoma) for providing me with this research opportunity and financial support. My thanks also to Mr. Brad Mons and a group of biochemistry students who assisted me construct the VirOligo database. Finally, I would like to thank my parents Tadashi and Takiko Onodera, and my wife Shu-Fen Peng, who just gave birth to our first precious daughter, Naomi. Without their encouragement and support, I wouldn't have the strength to complete this degree.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

NOMENCLATURE

A	adenine
ASCII	American National Standard Code for Information Interchange
AIDS	acquired immune deficiency syndrome
AVS	Advertisement and Volunteer System
BDV	border disease virus
BHV	bovine herpesvirus
BRD	bovine respiratory disease
bp	base pair
BSA	bovine serum albumin
BVDV	bovine viral diarrhea virus
C	cytosine
CCD	charge coupling device
CDC	The Center for Disease Control and Prevention
cDNA	complementary DNA
Cy	cyanine
dCTP	deoxycytidine triphosphate
ddH_2O	double-distilled water
DMSO	dimethylsulfoxide

DNA	deoxyribonucleic acid
dNTP	deoxyribonucleoside triphosphate
dTTP	deoxythymidine triphosphate
dUTP	deoxyuridine triphosphate
EDTA	ethylenediaminetetraacetic acid
FMV	foot-and-mouth disease virus
g	gram
G	guanine
HA	hemagglutinin
HIV	human immunodeficiency virus
HK68	human influenza viruses, A/Hong Kong/68
ILL	Inter-Library Loan
kb	kilo base pair
KY98	equine influenza virus, A/Equine/Kentucky/98
L	liter
LMS	Literature Management System
M	molar
Mg^{2+}	magnesium ion
MI63	equine influenza virus, A/Equine/Maimi/98
M-MLV	moloney murine leukemia virus
NCBI	National Center for Biotechnology Information
NJ76	human influenza viruses, A/New Jersey/76
nt	nucleotide

PAN99	human influenza viruses, A/Panama/99
PCR	polymerase chain reaction
RID	request ID
RNA	ribonucleic acid
RT	reverse transcription
S	guanine or cytosine
s.d.	standard deviation
SDS	sodium dodecyl sulfate
SSC	saline-sodium citrate
T	thymine
TAE	tris-acetate-EDTA
Taq	*thermus aquaticus*
T_a	annealing temperature
$T_a$max	maximum annealing temperature
$T_a$min	minimum annealing temperature
T_d	dissociation temperature
TE	tris-EDTA
T_m	melting temperature
u	Unit
UV	ultraviolet
ViSA	Virus Signature Amplification
ViSH	Virus Signature Hybridization
W	adenine or thymine

Chapter 1: Introduction

Importance of virus identification

As of October 31, 2002, there were 5738 virus species listed in the NCBI Taxonomy Database, an increase of 882 virus species in year 2001 alone. The number of new virus species listed each year in the Taxonomy Database is increasing (Figure 1-1). In 2002, (as of October 31, 2002), 1033 new species have already been added. Since 1039 virus species were known in 1993 and five times more virus species have been registered in less than ten years, only a small part of the viruses on earth is probably recognized currently (Condit, 2001).

Newly identified diseases that cause public health problems locally or internationally are called emerging infectious diseases. Examples of emerging infectious disease associated with viruses are human immunodeficiency virus (HIV, isolated in 1983 and causing AIDS), nipah virus (isolated in 1999 and causing encephalitis), ebola virus (isolated in 1977 and causing ebola hemorrhagic fever), hepatitis C (isolated in 1989 and causing liver cirrhosis and/or liver cancer), sin nombre virus (isolated in 1993 and causing highly fatal respiratory disease), and influenza A (H5N1) virus (known pathogen in birds and first isolated from human case in 1997) (World Health Organization, WHO, http://www.who.int/inf-fs/en/fact097.html and http://www.who.int/inf-fs/en/fact262.html, viewed on Sept. 2, 2002). Since we have recently experienced outbreaks of viral diseases, it is certain that we will encounter more emerging infectious diseases and unidentified viruses that are potentially pathogenic to humans in the near future.

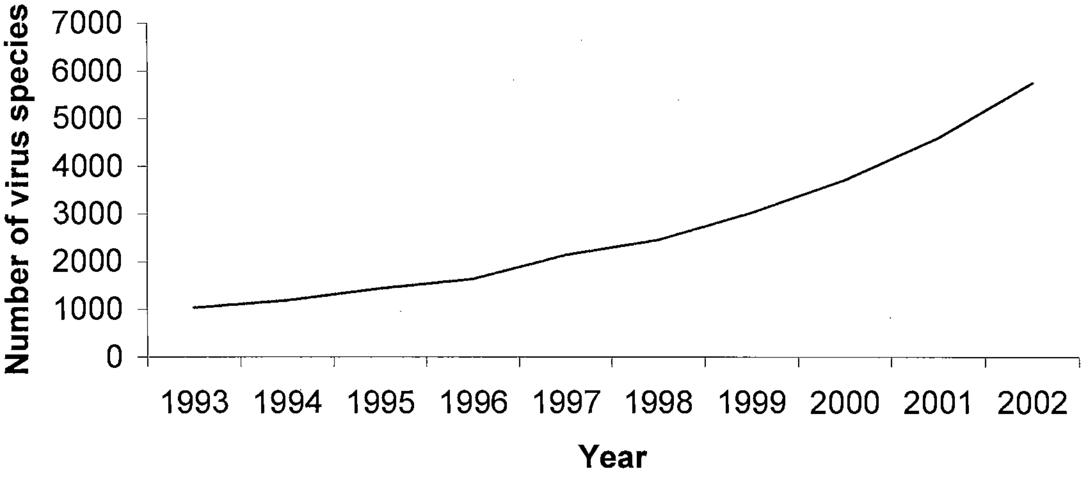


Figure 1–1, The number of virus species registered in NCBI taxonomy database from 1993 to 2002. The number in year 2002 was as of October 31, 2002.

The human pathogenic viruses are not the only viruses important to us. Canadian and British researchers mainly targeted pathogens of animals and crops during World War II, and the US bioweapons program from 1945 to 1969 weaponized plant pathogens (Hilleman, 2002). Animals and crops are the sources of our food supplies, and food shortages become significant weaknesses during war. Economic damages cannot be ignored, either. Foot-and-mouth disease hit the UK in 2001. More than six million animals were slaughtered, costing £4.8 billion for compensation to farmers as well as disinfection, veterinary support and slaughter costs (Stone, 2002).

Vaccination is used for prevention of some viral diseases. The development of a new vaccine takes 8 to 10 years and typically costs $ 500 million (Hilleman, 2002). Thus, it is not practical to expect vaccines for all human pathogens. Recently, fears of bioterrorism against civilians have come to the forefront. The Centers for Disease Control and Prevention (CDC) lists possible bioterrorism agents. Pathogens listed in category A, such as anthrax, smallpox and viral hemorrhagic fever, have the potential to cause the greatest impact on public health (Rotz et al., 2002). Preparation of vaccines for category A pathogens has not been completed. A vaccine against one of the viral hemorrhagic fever agents, yellow fever virus, has been licensed, but developments of vaccines for other viral hemorrhagic fever agents, such as ebola virus and lassa virus have not been completed. The current smallpox vaccination kills 2 persons in every million vaccinations (Hilleman, 2002).

At this moment, our ability to prevent virus outbreak is limited. However, fears of biowarfare and bioterrorism are increasing, and fears from emerging infectious diseases also exist. For example, ebola was probably a local disease but development of and human encroachment into rain forests and rapid international travel are responsible for the ebola now being a global threat (Larkin, 2000). In the midst of globalization, it is only to be expected that we will face additional threats from emerging infectious diseases like ebola. Early detection of outbreaks can facilitate initiation of effective counter-measurements, such as rapid and accurate diagnosis and quarantine of infected individuals, animals and/or plants. Thus, rapid virus detection methods that cover wide varieties of viruses are indispensable.

Current methods for diagnosing and identifying viruses

In my MS thesis (Onodera, 1999), I studied the detection of the viral agents of bovine respiratory disease (BRD). The common methods to diagnose viral infection for BRD are virus isolation in cell culture, immunofluorescent antibody tests (IFA), necropsy, and serologic tests (Rocha et al. 1999, Loken 1995). Those tests present some disadvantages. Virus isolation tests are not very sensitive because some viruses grow with difficulty, and the tests may take 1 to 3 weeks to complete as in the case of Border disease in sheep (Loken 1995). Virus identification by serology and necropsy require either recovery or fatality of a bovine, respectively. This delays initiation of treatment for other cattle in the herd to avoid spread of these viruses. The immunoperoxidase plate assay can handle large numbers of samples simultaneously (Brock 1995), but this method also takes 5 to 7 working days to complete.

PCR is the principal method of identifying virus genomes. Standard PCR cannot compare template concentrations because of the plateau effect (Kainz, 2000), which reduces amplification efficiency after PCR product concentrations reach too high levels.

The PCR amplification result is not known until all PCR cycles are completed. Recently, real-time PCR has become popular due to quantitation needs and the demand for faster diagnosis (Zubritsky 1999). There are many real-time PCR approaches. Two of them are to generate PCR products with two primers as in standard PCR with one fluorescent probe. These probes are exonuclease probes (TaqMan probes) and Molecular Beacon probes. The TaqMan probes are specific to an internal region of the PCR product and are labeled with a reporter fluorescent dye [FAM (6-carboxy-fluorescein)] and a quencher fluorescent dye [TAMRA (6-carboxy-tetramethyl-rhodamine)] (http://biochem.roche.com/lightcycler/lc_support/lc_faqs04.htm). A reporter and a quencher are located next to each other, so that the quencher absorbs energy from the excited reporter fluorophore. When the probe anneals to the PCR product, the quencher is removed by the exonuclease activity of Taq polymerase and the fluorescence from the reporter can be observed.

More PCR product will lead to more annealing of the probes and thus more fluorescence. By monitoring the reaction in real time, this procedure can monitor PCR product quantitatively. Another approach to real-time PCR is similar, using Molecular Beacon probe. A Molecular Beacon contains a PCR product-specific hairpin loop and its 3' and 5' ends are labeled by a quencher and a reporter fluorescent dye, respectively (Tyagi et al., 1996). This probe anneals to the PCR product, and straightens its hairpin region. When the probe looses the hairpin region, the distance between the quencher and the reporter is large enough to observe fluorescence.

Other real-time PCR probes similar to TaqMan and Molecular Beacon are the Hybridization Probes from Roche (http://www.lightcycler-online.com/lc_principles/lc_prin_pcr_mon01.htm, viewed on Sept. 4, 2002). Hybridization Probes use two sequence-specific probes, one labeled with fluorescein and the other with LC Red 640. When the two probes are hybridized onto a PCR product and the two dyes are in close proximity (1 to 5 nucleotides), the excitation energy of fluorescein is transferred to LC Red 640, which emits light at a longer wavelength. The Amplifluor Uniprimer system also uses a hairpin region. Uniprimer contains a 5' hairpin region with a fluorophore at the 5' end and a quencher at the 3' end of the hairpin region along with a 3' tail sequence (http://www.serologicals.com/products/int_prod/body_pcr_amplifluor.html, viewed on Sept. 4, 2002). One of two PCR primers contains a 5' tail sequence. PCR products with tail-attached primers anneal to the tail region of the Uniprimer, and produce 5'end Uniprimer-attached PCR products. The hairpin regions of the Uniprimers are unfolded in the next PCR cycle and fluorescence can be observed.

Another approach for real-time PCR uses an intercalating dye, such as SYBR Green I. SYBR Green I is a fluorescent probe and binds to the minor groove of the DNA double helix (http://www.lightcycler-online.com/lc_principles/lc_prin_dna_det01.htm, viewed on Sept. 4, 2002). Thus, the fluorescence level increases with the increase of PCR products even if PCR products are primer-dimers or primer-artifacts.

Real-time PCR is faster than soluble PCR, but it uses three to four types of oligonucleotides, two PCR primers and one or two fluorescent probes, or an intercalating dye. The intercalating dye cannot differentiate template-specific product from contaminants or artifacts. When we use more oligonucleotides with fluorescent probes, multiplex PCR specific problems arise. Multiplex PCR is used for simultaneous amplification of two or more DNA fragments in the same reaction (Henegariu et al. 1999, Burgart 1992), and is used for DNA testing, including analyses of deletions, mutations and polymorphisms (Henegariu et al. 1999). As compared to single primer-pair PCR, multiplex PCR can reduce the total number of tubes, the cost for the reagents, and possible contamination (Burgart 1992). However, multiplex PCR is not suitable for monitoring and detecting emerging viruses or a wide range of viruses. Multiplex PCR is expected to be difficult for three reasons. First, competition among the multiplex PCR products makes it difficult to amplify viruses which are present in small amounts in the sample or which have low amplification efficiency during PCR (Boles et al. 1998, Weissensteiner et al. 1996). Second, size separation and identification of PCR products by gel electrophoresis may be difficult. This is because there are many fragment sizes to identify on the gel and the same size products are not necessarily the same PCR products. The need to identify the fragment sizes puts restrictions on the primer design. Third, primer-dimers and primer-artifacts may be a problem (Boles et al. 1998). It is possible that such products might interact with PCR products and make PCR products of unexpected sizes.

Conventional PCR with gel electrophoresis is difficult to apply to virus monitoring systems due to multiplex PCR problems (Onodera 1999). Recently, microarray technology has become popular and many vendors are available. The most commonly used microarray technique is that a microscopic array of cDNA molecules is immobilized on a solid surface (i.e. nylon, nitrocellulose, and glass) as hybridization targets for biochemical analysis (Lemieux et al.1998, Schena et al. 1996). The cDNA array is printed on the solid surface as targets, and the mRNAs are labeled enzymatically as fluorescent cDNA probes. These cDNAs are hybridized to the microarray and unincorporated labeled probes are washed away. Thus, a fluorescently labeled probe hybridizes specifically to the cDNA target. The fluorescence is detected by confocal laser scanning and charge-coupled device (CCD) imaging (Lemieux et al.1998, Schena et al. 1996). This method was developed for accurate measurement of the expression of the specific genes (Schena et al. 1996). Array densities are 400 to 250,000 cDNAs per cm^2. This allows quantitative estimation of the expression of many genes in one single hybridization (Lemieux et al.1998). GSI Lumonics (Watertown, CA) developed a four-color fluorescence detection scheme that allows up to four kinds of probes to be used in one expression analysis. Today, fluorescently labeled DNA or RNA from biological sources is used as probe and a complementary sequence is used as target (Lemieux et al.1998). The applications of microarrays have been extended to expression analysis, polymorphism detection, DNA resequencing, and genotyping on a genome scale (Lemieux et al.1998).

Affymetrix (Santa Clara, CA) developed on-chip photolithographic synthesis, GeneChip technology. It can make miniaturized, high-density arrays of oligonucleotide probes by light-directed chemical synthesis. First, the protecting groups are removed to generate free hydroxyl groups at positions illuminated by UV light passing through the photolithographic mask. Then, 5'-protected phosphoramidite is added to deprotected sites, and oligonucleotides are synthesized by repeating this process (Ramsay 1998).

Microarrays are not commonly used for detection of trace amounts of nucleotide sequences because microarrays cannot amplify nucleotide sequences. Mosaic Technologies, Inc. devised bridge amplification technology to overcome this shortfall

(Abrams et al. 1998, Rehman et al. 1999, Boles et al. 1998). They immobilized acrylamide-modified oligonucleotides on many kinds of media (i.e. nylon beads, optical fibers, glass plates) and performed immobilized PCR. The "bridge" of bridge amplification technology comes from the shape when the extension product from one primer binds to another primer during the annealing step (Boles et al. 1998). Many primer sets are placed on the media separately. This procedure should make it possible to diagnose tens to hundreds of individual amplification reactions in a single reaction mixture and vessel (Boles et al. 1998). The Mosaic Technologies' primer attachment method is co-polymerization of acrylamide-modified oligonucleotide into a polyacrylamide co-polymer (Rehman et al. 1999). The benefit of this method is that it can avoid cross-contamination from amplified products since all amplified products are bound to the media. This system will be easier to automate and the use of real-time PCR is possible. Furthermore, this method costs less than a single template PCR for each primer since it can test many primers in a single assay (Boles et al. 1998).

Because of the increasing availability of microarray systems, hybridization has begun to be used for viral or bacterial nucleic acid detection using pathogen-specific oligonucleotides. Liu, et al. (2001) used 16S rRNA specific oligonucleotides on a microarray to detect the presence of bacteria and attempted to differentiate among members of the genus *Bacillus*. In the bacterial detection system, twenty oligonucleotides were designed by a probe design program, ARB (Strunk and Ludwig, Technical University of Munich, Germany). Only 25 % of oligonucleotides were specific only to the expected species of *Bacillus*. In virus detection, enteric viruses and hepatitis

B virus were detected by a small number of virus-specific oligonucleotides on microarray chips (Jeong et al., 2002; Park et al., 2002).

ViSA card and ViSH chip methods

In my M.S. research, I experimented with Virus Signature Amplification (ViSA) card and Virus Signature Hybridization (ViSH) chip methods (Onodera, 1999; Onodera et al., 2002). The ViSA card and the ViSH methods are an immobilized PCR and hybridization methods, respectively, on a medium (i.e. nylon membrane or glass slide), on which virus-specific oligonucleotides have been attached. Both devices were tested for multiplex PCR of bovine herpesvirus type 1 (BHV-1) and bovine viral diarrhea virus (BVDV). Both BHV-1 and BVDV are viral agents of BRD and they sometimes cause double infection (Fulton et al., 2000). Both viruses were successfully detected in multiplex PCR, but the amplified products in multiplex PCR were much lower in intensity than those in single PCR, and there were also PCR products of unexpected sizes. Non-specific products were probably due to unexpected interaction among primers, primer-dimers and PCR products. The interaction can be reduced using a ViSA card since all primer-pairs are immobilized on separated spots on a medium. In my M.S. thesis, primers were immobilized by UV light, which form a covalent bond between thymine and the amino group of the nylon membrane (http://www.stratagene.com/manuals/400072.pdf, viewed on Sept. 4, 2002). During the process of ViSA card modification, DMSO was found to inhibit PCR amplification on ViSA cards while DMSO was a required component for BHV-1 and BHV-2 amplification in soluble PCR with the primer sets used. BHV-1 and BHV-2 were detected without DMSO in ViSA card experiments as well as BVDV-1 and BVDV-2. There were two important factors in ViSA card experiments: primer and template concentrations. When one or both of their concentrations was high, the entire membrane reacted with the chemiluminescent substrate, which indicated the existence of PCR products, but the primer-spot was not identifiable due to a background problem. This was not a problem for the detection of single PCR products, but became a problem under multiplex detection since it was not possible to determine which primers produced the PCR product. Thus, template and primer concentrations were optimized for BHV-2 and BVDV-2 detections, and both viruses were successfully identified in a multiplex format in a ViSA card.

ViSH chip experiments were also performed. P1, P2 and TS2 are BVDV-2 specific primers used in PCR experiments. Those primers were immobilized on silylated glass slides along with BHV-1 specific primers and border disease virus (BDV) specific primers. BDV and BVDV are both pestiviruses, and the ViSH chip could distinguish between them. BVDV-2 was amplified with P1 and P2 primers and hybridized on ViSH chip. The ViSH chip detected hybridization with P1, P2 and TS2 primers. A negative control was prepared by amplification with P1 and P2 primers without template. It detected hybridization of PCR product only on P1 primer. A soluble PCR experiment also produced primer-dimers with the primer set used. Thus, this may indicate that the P1 primer produced the primer-dimers or that the P2 primer part of the primer-dimer caused a hairpin and cannot hybridize with the P2 primer.

Needs of virus specific oligonucleotides for ViSA card and ViSH chip

The ViSA card and ViSH chip results suggest that these methods can be candidates for universal virus detection systems. For further development, both methods require a number of virus-specific oligonucleotides as primers or probes. Two methods can identify virus-specific oligonucleotides for the ViSA card and ViSH chip methods. One approach was to use primer design programs, and another approach was to obtain such oligonucleotides from the published literature.

Many programs are available for oligonucleotide design. For example, Oligo (Molecular Biology Insights, Inc. Cascade, CO), OMIGA (Accelrys, Inc. San Diego, CA) and Primer Premier 5 (PREMIER Biosoft International, Palo Alto, CA) are available for the design of primers, and Array Designer (PREMIER Biosoft International, Palo Alto, CA) is available for the design of oligonucleotide probes for microarrays. However, to conduct a PCR, several reaction conditions such as temperatures and times of the cycles and concentrations of buffer components may need to be optimized. Primer design programs only provide an estimate of annealing temperature (T_a).

On the other hand, the published literature contains a number of virus-specific oligonucleotide sequences along with experimental conditions that have been tested. Even though DMSO was required in soluble PCR for BHV-1 and BHV-2 and ViSA card did not require DMSO, other PCR conditions were the same in ViSA card and soluble PCR. Thus, the PCR conditions in the published literature can be applied to ViSA card experiments, and oligonucleotides with the tested conditions from the published literature are more practical than those from design programs. The other imponderable problems of the oligonucleotide design programs are unknown specificity and annealing efficiency of the oligonucleotides since the oligonucleotides have not been tested. As mentioned above, most oligonucleotides designed for 16S rRNA were not specific only to the target species. The oligonucleotides in the VirOligo database have been used in either PCR or hybridization at least once before publication, and many oligonucleotides have also been tested for specificity to differentiate among species or strains of the viruses.

Publicly available oligonucleotide database

Since a large number of oligonucleotides with experimental conditions are required for the ViSA card, making a database of oligonucleotides is absolutely imperative to avoid overlooking collected data. There are four PCR primer or oligonucleotide databases. EBI (European Bioinformatic Institute) has hosted the PCR primers database since 1996 (Shomer, 1996). It contains a total of 60 primer pairs as of Dec. 2000. Molecular Probe Database has accumulated 4300 oligonucleotide sequences as of Dec. 2000 since its creation was first published in 1992 (Grasso et al. 1998). Data are mainly taken from the literature. UK Human Genome Mapping Project Resource Centre offered Primer Bank until 7 July 1999, but it was withdrawn and no data are shown currently. Oligonucleotide Probe Database (OPD) contains 96 PCR primers and probes with experimental conditions along with brief results and references for 16S 18S like SSU rRNA, 23S 28S like LSU rRNA, and ATPase8 from the literature (Alm et al. 1996). None of these databases were updated in the past four years, and virus-specific oligonucleotides are not available from any of the existing databases.

Needs for PCR optimization protocols

When PCR is performed for the first time with newly designed primers, optimization may be required (Boleda et al 1996), and the optimization is a time-consuming and elaborate task. The costs (i.e. polymerase and labor) and the volume of PCR template increase with the number of optimization steps.

PCR experiments without the optimization of PCR conditions usually generate fewer PCR products than the optimized experiments, and sometimes generate no PCR products (Sambrook and Russell, 2001). Optimization factors in PCR are T_a, length of each step (denaturation, annealing and extension) in a cycle, the number of PCR cycles, buffer concentrations, $MgCl_2$ concentration, concentration of dNTPs and sometimes the addition of additives such as DMSO, glycerol or bovine serum albumin (BSA) (Boleda, et al., 1996). Currently, we don't know which factors or whether all of these factors need to be optimized. PCR optimization kits are available from several manufactures, such as Invitrogen (Carlsbad, CA), Stratagene (La Jolla, CA) and Roche Diagnostics (Indianapolis, IN), but the kits are not sufficient for optimization due to the complex interactions of reaction factors (Newton, 1995). Thus, trial-and-error approaches or empirical testing is used in the optimization and primer design process (Elfath, et al., 2000; Boleda, et al., 1996).

Recommendations for primer design

Primer lengths

Recommended approaches to primer design and reaction conditions have been widely disseminated in review articles and books (Sambrook and Russell, 2001; Newton and Graham, 1997). However, those recommendations are inconsistent with each other. Primer length and G+C content are among the most mentioned considerations in PCR primer design. Some instructions recommend primers 18 nt or longer (Newton, 1995; Sambrook and Russell, 2001; Kidd and Ruano, 1995). The maximum recommended length of the primer varies: 30 nt (Newton, 1995; Taylor, 1991), 28 nt (Innis and Gelfand, 1990), 25 nt (Sharrocks, 1994), and 24 nt (Dieffenbach, et al., 1993). Longer primers provide higher specificity. They also allow using higher T_a to reduce temperature differences among denaturation, annealing, and extension and make PCR cycles shorter (Newton, 1995; Rychlik, 1995). However, increasing primer length also increases the chances of primer-dimer formation. Since the long primers do not require annealing of the entire primer to the templates, the specificity of the PCR cannot be increased by adding more nucleotides at the 5'end of primers of adequate length (Newton and Graham, 1997; Rychlik, 1995).

Primer % (G+C) contents and melting temperatures (T_m)

40 to 60 % (G+C) contents of primers are generally recommended (Sambrook and Russell, 2001; Mitsuhashi, 1996; Newton, 1995), but some references recommended narrower ranges of %(G+C), 45 to 55 % (Sharrocks, 1994) and 50 to 60 % (Innis and Gelfand, 1990). Another reference recommended 40 to 65 % (Hyndman, et al., 1996). According to Sharrocks (1994), %(G+C) contents of the primers are related to binding strengths and melting efficiencies between templates and primers. Binding strengths and melting efficiencies can also be defined by T_m, but T_m of primer is not well discussed in the recommendations probably due to difficulty of accurate T_m calculations. Innis and

Gelfand (1990) recommended 55 to 80 °C for primer T_m. The difference in T_m between a pair of primers should be close to zero to avoid non-specific binding of less stable primers in a primer-pair (Kidd and Ruano, 1995; Sharrocks, 1994; Taylor, 1991; Hyndman, et al., 1996).

The 3' end triplet codes of primers

Along with the size and the G+C content of the primer, the 3' end triplet of a primer is also one of the considerations mentioned frequently in primer design. Although many researchers recognize the importance of the 3' end triplet, the recommendations are contradictory. According to Mitsuhashi (1996), a high G+C content at the 3' end of a primer may not require annealing of an entire primer for priming to occur with consequent loss of specificity. A low G+C content at the 3' end may not be elongated by DNA polymerase (Mitsuhashi, 1996) or may increase the importance of complete annealing (Hyndman, et al., 1996) due to weak annealing. Recommended compositions of primers at the 3' end triplets differ: no T at the 3' end and at least one W at the 3' end triplets (Kidd and Ruano, 1995); S at the 3' end and no GC or CG due to formation of hairpins and primer-dimers (Sambrook and Russell, 2001); low G+C (Hyndman, et al., 1996); high G+C (Rychlik, 1995); 1 or 2 S (Sharrocks, 1994).

Product sizes

Less than 100 bp PCR products are difficult to differentiate from primer dimers by agarose gel electrophoresis, and large PCR products are generally difficult to amplify (Mitsuhashi, 1996). The recommended ranges for PCR products are 300 to 350 bp

(Mitsuhashi, 1996), 200 to 400 bp (Rychlik, 1995), 100 to 600 bp (Sharrocks, 1994), and 120 to 300 bp (Dieffenbach, et al., 1993).

Recommendations for PCR conditions

T_a calculations

Finding the right T_a is the most important thing in PCR optimization (Newton and Graham, 1997). At too high T_a, primers cannot anneal to template, and at too low T_a, primers may cause non-specific annealing to template (Newton and Graham, 1997). The ranges of recommended T_a also vary for different authors, 30-60 °C (Newton, 1995); 50-65 °C (Burke, 1996); 55-72 °C (Innis and Gelfand, 1990); 45-65 °C (Mitsuhashi, 1996). T_as in PCR experiments may be chosen based on melting temperature (T_m) of PCR primers. There are several recommendations for T_a: optimum T_a is three to five degrees below T_m of PCR primers (Newton and Graham, 1997; Sambrook and Russell, 2001); within a few degrees of T_m of PCR primers (Dieffenbach, et al., 1993); equal to T_m of PCR primers (Kidd and Ruano, 1995); T_a=0.7×[T_m of PCR product]+0.3×[T_m of primer]-25 (°C) (Wetmur, 1991).

T_m calculations

All recommendations listed above deduce the T_a using the T_m of PCR primers. There are several methods to estimate T_m.

For short oligonucleotides (18 nt or less; Ref, Howe, 1995):

T_m (°C) = 4× [the number of G and C] + 2× [the number of A and T] (Eq. 1; Thein et al., 1986).

For longer oligonucleotides (more than 19 nt; Howe, 1995):

$$Tm = 81.5(°C) + 16.6\log_{10}[Na^{+}] + 0.41 \times [\%(G+C)] - \frac{600}{[Size]}$$ (Eq. 2; Bolton et al., 1962)

where [%(G+C)] and [Size] are %(G+C) contents and length (bp) of PCR product, respectively.

For more precise calculation of T_m of short oligonucleotides, the nearest-neighbor method (SantaLucia, 1998) is used:

$$\Delta S([Na^{+}]) = \Delta S([1M]) + 0.368 \times N \times \ln[Na^{+}] \quad [\text{Eq. 3}]$$

$$Tm = \frac{\Delta H}{\Delta S + R\ln(C/4)} - 273.15(°C) \quad [\text{Eq. 4}]$$

where ΔH and ΔS are the enthalpy and entropy for helix formation, respectively, N is the primer length-1, $[Na^{+}]$ is salt equivalent concentration, R is the molar gas constant, and C is the concentration of oligonucleotide. ΔS values are available at 1M NaCl and need to be adjusted for the PCR condition.

The first two equations (Eq. 1 and 2) calculate T_m based on %(G+C) of the DNA sequence and the last one (nearest-neighbor) uses entropies and enthalpies of dinucleotide pairs. Salt equivalent concentration of 0.2 M (equivalent to 1.5 mM $MgCl_2$ and 50 mM KCl; Wetmur, 1991) was used in calculations using Eq.2 and 3, and ΔS is calculated by Eq.3 for use in Eq.4.

Other parameters in a thermal cycle

A thermal cycle consists of denaturation, annealing and extension. The T_a has been mentioned above. Recommended annealing time is not well discussed in review articles and books. Newton (1995) recommended 5 to 30 seconds to achieve maximum fidelity and 2 to 3 minutes to achieve a maximum yield.

Recommended denaturation temperatures and times are 92 to 95 °C for 5 to 30 seconds (Newton, 1995), 94 to 97 °C (Innis and Gelfand, 1990), 94 °C for 1 minute (Kidd and Ruano, 1995) and 92 to 100 °C (Newton and Graham, 1997). High temperatures and longer denaturation can lower polymerase activity (Innis and Gelfand, 1990) and damage template DNAs (Newton, 1995). In contrast, low denaturation temperatures cause incomplete denaturation and lower yield of PCR products (Innis and Gelfand, 1990).

Extension temperature is dependent on polymerase activities (Newton, 1995). For example, the optimum temperature for Taq polymerase is 72 °C (Newton and Graham, 1997). The length of extension depends on the product size, and usually 3 minutes (Newton, 1995) or 2 minutes (Newton and Graham, 1997) is sufficient.

Mg^{2+} concentrations

Mg^{2+} concentrations dramatically change the specificity and yield of PCR products (Newton and Graham, 1997). Mg^{2+} is a cofactor for the polymerase, and it simulates polymerase activity. However, the effects of Mg^{2+} concentrations are complex since dNTPs chelate Mg^{2+} (Kidd and Ruano, 1995), and changing dNTPs concentration requires re-optimization of Mg^{2+} concentrations. Low or high Mg^{2+} concentration may cause primer-dimers, weak or no products, or non-specific products (Newton and Graham, 1997). The range of Mg^{2+} concentrations recommended are wide; 0.5-2.5 mM

(Innis and Gelfand, 1990); 1.0-5.0 mM (Taylor, 1991); 0.5-2.0 mM (Mitsuhashi, 1996); 0.5-5.0 mM (Newton, 1995).

dNTPs concentrations

Recommended ranges of each dNTP concentration are 200 μM (Kidd and Ruano, 1995), 20 to 200 μM (Innis and Gelfand, 1990) and 10 to 100 μM (Newton and Graham, 1997). According to Promega (Madison, WI), PCR buffer (Part No. M190A and M190G) supplied with polymerase is optimized at 200 μM each of dNTPs.

Primer concentrations

High concentrations of primers cause the formations of the primer-dimers. Recommended ranges of primer concentrations are 0.1 to 1 μM (Newton, 1995), 0.1 to 0.5 μM (Kidd and Ruano, 1995; Innis and Gelfand, 1990) and up to 1 μM (Newton and Graham, 1997).

Other buffer components

PCR buffer is usually supplied with the polymerase. For example, the PCR buffer from Promega (Madison, WI; Part No. M190A and M190G) for Taq DNA polymerase consists of 50 mM KCl, 10 mM Tris-HCl pH 9.0 at 25 °C and 0.1 % Triton X-100. Authors of PCR tutorial books recommend lower pH value than Promega's PCR buffer, such as pH 8.3 to 8.8 (Newton, 1995), 10 to 50 mM Tris-HCl pH 8.3 (Kidd and Ruano, 1995) and 10 to 50 mM Tris-HCl pH 8.3 to 8.8 and up to 50mM KCl (Innis and Gelfand, 1990). High pH (9.0-9.5) is recommended for long PCR products. High KCl concentrations inhibit polymerase activities (Innis and Gelfand, 1990). Other effects from buffer conditions are not well known (Innis and Gelfand, 1990).

As the PCR buffer from Promega uses Triton X-100, other additives are sometimes mixed in the buffer. Triton X-100, Tween 20, Laureth 12, gelatin and bovine serum albumin (BSA) are added to stabilize polymerase (Kidd and Ruano, 1995; Innis and Gelfand, 1990). DMSO, glycerin, betaine, 7-deaza dGTP and dITP were added in the reaction mixture for the amplification of G+C rich templates (Henke, et al., 1997). However, Taq DNA polymerase activity drop approximately 50 % when 10 % DMSO is used in the reaction mixture (Innis and Gelfand, 1990).

Primer design programs

In the section above, three programs for the primer design (Oligo, OMIGA, Primer Premier) were introduced. At the default setting, primers are selected from18 to 22 nt in length (all three programs); 40 to 60 % (G+C) (OMIGA); T_m of 45 to 65 °C (OMIGA) or 50 to 60 °C (Primer Premier); 400 to 800 bp product size (OMIGA) or 100 to 500 (Primer Premier); T_m difference in a primer-pair of not more than 2 °C (OMIGA) or 3 °C (Primer Premier); primer length difference in a primer-pair of less than 4 nt; the 3' end GC clamp (all three programs); T_m is calculated by a nearest-neighbor method (all three programs); T_a is calculated by the Wetmur's equation (1991), T_a=0.7×[T_m of PCR product]+0.3×[T_m of primer]-25 (°C), in all three programs. In OMIGA, all primer-pairs are ranked according to the pre-determined score on each parameter.

Overview of the Dissertation

Two approaches to obtain the oligonucleotide sequences for the ViSA card and the ViSH chip were introduced in the section above. Since PCR conditions needed in the ViSA card method are listed and oligonucleotide specificity has been tested in published literature as explained above, I decided to collect oligonucleotide sequences from the published literature. The collected sequences were compiled in the VirOligo database along with PCR and hybridization conditions (Chapter 3). To examine the usefulness of the VirOligo database created, the influenza virus specific oligonucleotides from the VirOligo database were tested as probes in ViSH chip and compared to oligonucleotides generated by primer design software (Chapter 4). Many publications are available giving tutorials for primer design and PCR optimization, but the recommendations are not consistent. Using the data from the VirOligo database, the range of each condition used the most frequently was analyzed, and the PCR optimization methods were suggested. Then, one of the suggestions was tested experimentally (Chapter 5 for primer design and Chapter 6 for PCR optimization).

Chapter 2: Materials and Methods

Reagents

TaqPlus Long DNA polymerase was purchased from Stratagene, La Jolla, CA. SYTO 61 was purchased from Molecular Probes, Inc., Eugene, OR. Cy3 and Cy5 Mono-Reactive dye packs were purchased from Amersham Pharmacia Biotech, Arlington Heights, IL. 5-(3-aminoallyl)-2'-deoxyuridine 5'-triphosphate (aa-dUTP) and hydroxylamine (50% wt solution in H_2O) were purchased from Sigma-Aldrich Corp., St. Louis, MO. Taq DNA polymerase in Buffer B, 10x PCR buffer, and 25 mM $MgCl_2$ was purchased from Promega, Madison, WI. 10x PCR buffer contains 10 mM Tris-HCl, 50 mM KCl, and 0.1 % Triton X-100. Array it UniHyb was purchased from Telechem International, Inc., Sunnyvale, CA. Sodium borohydride and Tween 20 were purchased from Fisher Scientific, Pittsburgh, PA. M-MLV reverse transcriptase, RNaseOUT Ribonuclease Inhibitor, 5X first-strand buffer [250 mM Tris-HCl (pH 8.3), 375 mM KCl and 15 mM MgCl2], and 100 bp DNA ladder were purchased from Invitrogen Corporation, Carlsbad, CA.

TAE buffer consisted of 40 mM Tris-acetate and 1 mM EDTA (pH 8.2). TE buffer consisted of 10 mM Tris-HCl and 1 mM EDTA (pH 8.0). 20x SSC consisted of 3 M NaCl and 0.3 M sodium citrate. Stop reagent consisted of 50% glycerol, 10 mM EDTA (pH 8.0), and 0.025% bromphenol blue.

Materials

Aldehyde slide plates (Cat. # ALS-25) were purchased from TeleChem International, Inc., Sunnyvale, CA. Microcon YM-30 was purchased from Millipore, Bedford, MA. QIAquick PCR Purification Kit was purchased from Qiagen Inc., Valencia, CA. Hybridization Chamber (Cat. # 2551) was purchased from Corning, Inc., Corning, NY.

Machines

Fluorescent scanner used was ScanArray 3000 made by PerkinElmer Life Sciences Inc., Boston, MA. Arrayer used was PixSys 5500 made by Cartesian Technologies, Inc., Irvine, CA. PCR cyclers used were PTC-100 and PTC-200 made by MJ Research, Inc., Waltham, MA. PTZ-200 was a gradient thermal cycler and kindly provided by Dr. Muriana (Dept. of Animal Science, OSU). Gel Doc 1000 was made by Bio-Rad, Hercules, CA. Sonicator used was Fisher Model 60 Sonic Dismembrator made by Fisher Scientific, Pittsburgh, PA, and kindly provided by Dr. Longtine (Dept. of Biochemistry and Molecular Biology, OSU). Waterbath used was Fisher ISOTEMP 1016S made by Fisher Scientific, Pittsburgh, PA. Speed Vac (sc110A) was made by Thermo Savant, Holbrook, NY.

Templates

All templates were provided by Dr. Sengupta (Dept. of Biochemistry and Molecular Biology, OSU) and Dr. Lai (Dept. of Microbiology and Molecular Genetics, OSU).

Influenza virus was propagated in chicken egg allantoic fluid, and RNA was extracted using phenol/chloroform, ethanol precipitated and resuspended in RNase free water according to Lai and Chambers (1995).

The reverse transcription (RT) mixture (20 μL) consisted of 5 μL of the resuspended influenza RNA sample, 0.2 μg or 0.08 μg of Uni3 universal primer (Appendix 1 for primer sequence), 1X first strand buffer, 0.5 mM each dNTPs, and 3.2 u RNaseOUT. The RT mixture was incubated at 75°C for 3 minutes followed by 2 minutes on ice. Then 400 u M-MLV reverse transcriptase was added to the RT mixture and the mixture was further incubated at 42°C for 1 hour. By adding 80 μL sterile water, the total volume of the RT mixture was adjusted to 100 μL.

Preparation of Cy-labeled targets

Universal Amplification with aminoallyl-dUTP (aa-dUTP)

PCR procedure was modified from the published universal PCR protocol for influenza A virus (Offringa, et al., 2000). The PCR mixture totaled 25 μL and included 1.5 u TaqPlus Long polymerase, 1× PCR buffer, 1.5 mM $MgCl_2$, 0.1 μg or 0.04 μg of Uni3 universal primers, 0.1 μg Uni5 universal primers (Appendix 1 for primer sequences), 0.14 mM dATP, dCTP and dGTP, 0.08 mM dTTP, 0.06 mM aa-dUTP and 2 μL RT product. PCR was initiated with denaturation at 94 °C for 2 minute, followed by thirty cycles of denaturation at 94 °C for 1 minute, annealing at 40 °C for 2 minutes, and extension at 72 °C for 3 minutes. Final extension was allowed at 72 °C for 10 minutes.

Coupling of Cy dye and aa-dUTP

The PCR products were sonicated to less than 300 bp and small fragments less than 30,000 nominal molecular weight were removed by a Microcon centrifugal filter. Then, Cy dye was incubated with the PCR products for coupling and any un-reacted Cy dye was quenched by adding hydroxylamine according to Seidel's protocol (http://www.pangloss.com/seidel/Protocols/amino-allylRT.html; viewed on Aug. 20, 2002). Finally, quenched Cy dye was removed from the reaction mixture by QIAquick column, and the products were dried in a Speed Vac.

Selection of oligonucleotide probes

For oligonucleotide probes used in the ViSH chip, 463 influenza virus specific oligonucleotides were chosen from the VirOligo database and selected by omitting oligonucleotides less than 17 nt and more than 29 nt, and omitting those that had a perfect match or a 1 bp mismatch with an organism other than influenza virus in BLAST results. One 32 bp equine influenza virus-specific oligonucleotide was also included. Fourteen equine influenza hemagglutinin gene (A/Equine/Kentucky/98; Accession No. AF197241) specific oligonucleotide sequences were obtained using OMIGA software (Accelrys, San Diego) with OMIGA default settings, except that the number of primer pairs to be generated was set at 10.

Attachment of oligonucleotide probes on an aldehyde glass slide

Oligonucleotide probes (10 mg/mL, Appendix 1 for probe sequences) with 5' C6 amino modification in 5× SSC were spotted in quadruplicate on an aldehyde glass slide by an arrayer. Then, the slide was dried and washed with 0.2% SDS solution for 1 minute and with ddH$_2$O for 1 minute twice. Next, the slide was incubated in sodium borohydride solution for 5 minutes, and washed with 0.2 % SDS solution for 1 minute and with ddH_2O for 1 minute twice. This slide was kept in the dark and room temperature until used.

Detection of soluble PCR products by agarose gel electrophoresis

Soluble PCR products were analyzed by agarose gel electrophoresis. Agarose gels (1.0 %) in 1× TAE buffer contained 0.5 mg/mL ethidum bromide for detection of DNA. Samples loaded consisted of 5 μL PCR product, 4 μL stop reagent, and 10 μL ddH_2O. Electrophoresis was performed at 100 V (8 V/cm) in 1× TAE buffer for approximately 40 minutes. A 100 bp ladder was used to determine product sizes. DNA was detected by Gel-Doc 1000.

Hybridization and detection in ViSH chip

Hybridization of a target

Total 10 μL of hybridization solution (one volume of the resuspended Cy-labeled target and four volumes of UniHyb hybridization solution) was pre-heated at 65 °C for 3 minutes and shaken several times. Then, the solution was applied on a 2.2×2.2 mm #1 thickness glass cover. The glass cover was placed on the printed ViSH chip in a hybridization chamber with 20 μL sterile water for humidity. This cassette was placed in a heated water bath for one hour at 22 °C.

Detection of the hybridized target

The chips were washed in 2x SSC and 0.2% SDS solution for 15 seconds and in 0.5x SSC solution for 15 seconds in the dark. Then, they were dried by Kimwipes without touching the probe printed area of the slide surfaces. Presence of Cy3 and Cy5 was determined by a fluorescent scanner (ScanArray 3000) at 100 % laser and 95 % photomultiplier tube power.

Oligonucleotide detection by SYTO 61

The protocol for SYTO 61 was obtained from Microarray Core Facility's web page at OSU (http://opbs.okstate.edu/CORE/arrayer/files/SYTO-61.html; viewed on Aug. 23, 2002). The printed ViSH chip was soaked in 50 mL of 1 μM SYTO 61 in 1X TE buffer and 10% ethanol for 5 minutes at room temperature. The chip was washed with 0.1 % Tween 20 solution for 5 minutes and with sterile water for 5 minutes twice. Then, the chip was dried by spinning at 500 to 1000 rpm in a HL-4 rotor of a GLC-2 centrifuge. The hybrids of SYTO 61 and probes were detected by ScanArray 3000 (using channel 2).

PCR protocols for KY98 HA segment amplification

The PCR mixture totaled 25 μL and included 2.5 u Taq DNA polymerase, 1× PCR buffer, 1.5 mM $MgCl_2$, 4 μM each primer, 0.2 mM each dNTPs and 2 μL RT product. PCR was initiated with denaturation at 94 °C for 1 minute, followed by thirty cycles of denaturation at 94 °C for 1 minute, annealing at 42-72 °C for 1 minute, and extension at 72 °C for 1 minute. Final extension was performed at 72 °C for 10 minutes. Twelve different annealing temperatures were tested for each primer-pair by a gradient thermal cycler, PTZ-200.

Analysis of primers sequences and PCR conditions

Primer sequences and PCR conditions were obtained from the VirOligo database on February 6, 2002 for the analysis. On that date, the VirOligo database covered a total of 1685 articles from PubMed search results queried on December 19, 2001. The articles contained 3985 published virus-specific oligonucleotide sequences and 2300 PCR and hybridization conditions for detection of the viruses for alcelaphine herpesvirus, bovine adenovirus, bovine viral diarrhea virus, bovine herpesvirus (BHV), bovine respiratory syncytial virus, bovine rotavirus, bovine coronavirus, foot-and-mouth disease virus (FMV), variola (smallpox) virus, cowpox virus, and human adenovirus. The resulting list was filtered to eliminate redundancy.

Data assortment and analysis were performed by homemade perl and php scripts (Appendix 2) along with Microsoft Excel 2000 and statistical analysis tools provided by Excel.

Melting temperature

Melting temperatures (T_m) for oligonucleotides were calculated by the nearest-neighbor method using Eq. 3 and Eq. 4 in the Introduction. Because the nearest-neighbor method is only suitable to calculate T_ms of short oligonucleotides, T_m of PCR products were obtained using Eq. 2 in the Introduction.

Chapter 3: The VirOligo Database

The VirOligo database

VirOligo is an oligonucleotide database that contains PCR primers and hybridization probes from published literature that are used for detection of viral nucleotides along with experimental conditions for their use (Onodera and Melcher, 2002). As of September 2002, VirOligo contained 4318 oligonucleotide sequences and 3824 experimental conditions from 2874 articles. Currently, VirOligo covers all virus-specific oligonucleotide sequences available in the articles listed in PubMed search result queried on March 5, 2002 for bovine adenovirus, bovine parainfluenza, influenza virus, alcelaphine herpesvirus, bovine viral diarrhea virus, bovine herpesvirus, bovine respiratory syncytial virus, bovine rotavirus, bovine rhinovirus, bovine enterovirus, bovine coronavirus, foot-and-mouth disease virus, variola (smallpox) virus, cowpox virus, human adenovirus, and vaccinia virus. The coverage of VirOligo is being expanded to other viruses, and bovine parvovirus, canine parvovirus, porcine parvovirus, and rodent parvovirus are nearly completed. Simultaneously, VirOligo is being updated with newly published articles for the existing virus coverage, and almost all articles listed in PubMed search queried on March 25, 2002 are being added to VirOligo.

Evolution of VirOligo

A prototype of VirOligo was created in October 2000 to provide a large number of virus specific PCR primers for ViSA card experiments (Onodera, et al., 2002). That VirOligo was a flatfile database, and was constructed with FileMaker Pro, ver. 4

(FileMaker, Inc., a subsidiary of Apple Computer, Inc). PCR conditions and primer sequences were entered at the same time by an undergraduate student using an internet browser from a website created for VirOligo. The prototype contained the same entry fields as the current version of the VirOligo database, but expected the entry person to enter only one pair of oligonucleotides. Further, the same PCR conditions needed to be entered twice for each pair of primers since each entry took only one oligonucleotide sequence. Some types of PCR, such as nested PCR and multiplex PCR, use more than two primers.

The current version of the VirOligo entry system was created in December 2000. It allows entry of more than one pair of primers for a set of PCR conditions and will accept any virus-specific oligonucleotides. The system is a relational database consisting of two tables, Common data and Oligo data. The Oligo data table contains oligonucleotide sequences and the Common data table contains the experimental conditions. Since the Oligo data and the Common data are stored separately, the same oligonucleotide sequence can be used in several PCR or hybridization condition entries. Conversely, entries in the Common data table allow up to ten oligonucleotides to be linked to them and can thus be used for multiplex and nested PCRs. This version of the VirOligo entry system is well established and has been used for 22 months as of October 2002.

After the VirOligo database was launched, twelve graduate and undergraduate students, including freshmen, tested data entry methods. The number of students involved dictated that labor distribution and training needed to be optimized. We considered whether students should work together or individually.

The benefits of group work compared to individual work were that, since entry people can specialize in one job function, less knowledge was required of beginners and that more proofing occurs for each entry. The entry task was separated into four job functions: literature collection, filtering articles (removing articles without primers, or with primers used for PCR-mediated cloning experiments), oligonucleotide entry and experimental condition data entry. At the early stages of the project, filtering and literature collection were combined. The literature collection task included literature management to avoid missing articles from among the hundreds surveyed and keeping the database of articles updated.

It is critical that VirOligo provide oligonucleotide sequences without typographical errors. The oligonucleotide entry person entered the sequence and wrote the Oligo ID (identification number for each oligonucleotide entry) on the article. Then, the common data person received the article and found the Oligo ID. After entering the common data, the common data person linked the Oligo ID into the Common data and double-checked the sequence for correctness. Because the job functions and working schedules of the participants were different, stresses were occasionally induced against other job functions. These were resolved through weekly staff meetings.

In individual work, each person did the entire job from literature collection based on PubMed Identifier numbers (PMID) to data entries. Undergraduate students had to learn a lot, such as how to request articles by interlibrary loan, how to use BLAST and how to calculate concentrations. Every author has his/her own style of writing, and it was sometimes difficult to decipher what was meant. For example, some authors write the volume of dNTPs added in RT and PCR separately. The molarity calculation becomes complex for some undergraduate students if a part of a cDNA reaction was used in PCR.

The biggest benefit of individual work was that all workers understood the entire job and the database did not suffer a shortage of workers when one person quit. If there is only one person for one job shift in the group work, the entire job will be stopped by missing the job shift, but in individual work, there are no concerns from one missing shift.

The VirOligo database project started with the individual working plan. A two day training session covered the gamut from math and basic biology to PCR techniques. The individual work plan was not successful since most students had difficulty reading and understanding journal articles. The group-work plan was started to reduce the amount each individual had to learn. The group-work plan operated smoothly, and all personnel contributed well. The filtering of articles was assigned to the literature preparation shift, but the filtering job was harder for students than expected. The filtering person needed to read the articles and understand whether the oligonucleotides were virus-specific. Thus, the filtering task has become the common data entry person's task or one independent task.

During Fall 2001, the expansion of viruses covered in VirOligo made it difficult to track the articles, whether they needed to be obtained from the libraries or to be put on hold waiting for an interlibrary loan return. Overcoming the difficulty thus required improvement of literature collection and management tasks. As a result, a Literature Management System (LMS) was introduced (see below).

Due to concerns with computer stability with FileMaker Pro and the computer's processing speed, a Publishing Server for the VirOligo database was established. It uses the Linux operating system and Perl scripts, a MySQL database, an Apache web server and PHP web-scripts. It also is described in more detail below.

All preparations and modifications for VirOligo were made by the end of 2001, and the VirOligo database was finally opened for the public by announcing VirOligo in Nucleic Acids Research, Database issue 2002. VirOligo initially covered ten bovine viruses (bovine herpesvirus, bovine viral diarrhea virus, bovine adenovirus, bovine parainfluenza virus, bovine respiratory syncytial virus, bovine rhinovirus, bovine coronavirus, bovine reovirus, bovine enterovirus, and alcelaphine herpesvirus). As mentioned above, the number of viruses covered has expanded to 16 viruses, and has been updated for existing viruses.

The VirOligo System

The VirOligo database system consists of two servers, the Publishing Server and the Entry Server, and the Literature Management System. The servers are separated to achieve better security and performance.

A team of student workers was assembled to extract the details of oligonucleotides used in virus detection from research articles identified and managed by the Literature Management System. Entered via the Entry Server are the sequence of the oligonucleotide, the virus name, its NCBI taxonomy ID, reference GI number to the virus sequence in GenBank, and the position in that sequence that the oligonucleotide corresponds to. Additionally, if the oligonucleotide is part of a PCR primer pair, information on the PCR conditions (temperatures, buffer concentrations, divalent cation concentrations, etc.) are added. The database is searchable. Results are presented via the Publishing Server with links to a BLASTn search of GenBank, and to the relevant PubMed, GenBank and Taxonomy entries. Calculated information such as melting temperature, product length, oligonucleotide length and degeneracy are also presented.

To make VirOligo useful to the virological community as soon as possible, we elected to approach filling the database one virus or group of related viruses at a time. To date, we believe to have achieved complete coverage of oligonucleotides published in the literature that are related to the approximately dozen viruses associated with bovine respiratory disease, foot and mouth disease virus, human adenovirus, variola and cowpox viruses, and influenza virus. In addition, about a hundred other viruses are partially covered. For all viruses, as of January 2002, 1,232 PCRs were covered. The team examined 1,510 articles and entered 3,799 oligonucleotides. This number of oligonucleotides is substantially higher than in other available databases of oligonucleotides (Campi et al., 1998; Alm et al., 1996; Shomer, 1996).

The VirOligo Entry Server

The Entry Server is used solely for database entry. Our database group staff and students are the only users. The Entry Server is embedded in FileMaker Pro version 4 under Microsoft windows 98 OS and is dedicated to this purpose. Since other software, such as an E-mail client, is not running, there is little fear of computer virus infection. Most security breeches are the result of known security holes attacked with automated hacking software (http://www.cert.org/archive/pdf/attack_trends.pdf, CERT Coordination

Center, Carnegie Mellon University, viewed on May, 9, 2002). Since FileMaker software is not popular and the currently marketed version of FileMaker is 5.5, information about security holes of version 4 is not easily available. We believe that unpopular software is safer than popular software, and automated hacking software for FileMaker, especially for version 4, is less likely to be created and used. As an extra precaution, the database files are regularly transferred to another computer for backup.

The Entry Server contains seven tables: Common data, Oligo data, User data, Oligo user data, bulletin board, working hours, and Security database. The Security database possesses user ID and password information, and is used for user authentication when a user accesses the database to create or modify an entry. Information extracted from articles is entered into two tables. Oligonucleotide information is put into the Oligo data table and experimental conditions for PCR and hybridization are put into the Common data table. Each Oligo data entry consists of oligonucleotide sequence, target region, name of the oligonucleotide, type of usage (PCR primer, PCR probe, hybridization or other), notes and direction of the oligonucleotide (forward or reverse). Degeneracy and length are calculated by FileMaker automatically upon the entry of an oligonucleotide. Dissociation temperatures (T_d) of oligonucleotides are calculated by FileMaker scripts.

Initially, Common data were entered first, and then, the Oligo data were added to the Common data. This method did not allow entries beginning with the Oligo data. Oligonucleotide sequences are the most important in VirOligo and a typographical error will reduce the creditability of the database. Thus, the oligonucleotide sequences should be entered first and the sequences should be confirmed by another person during the

Common data entries. Thus, a function was added during Spring 2001 to enable the Oligo data entry before the Common data entry.

Currently, Oligo data personnel enter Oligo data first. On completion of data entry they write the Oligo IDs provided by the server on a copy of the article along with GI numbers. Then, the Common data personnel receive the articles and perform the Common data entry. Each Common data entry contains publication date of the reference, its PMID (PubMed unique identifier), GI (GenBank) of the target virus sequence, virus name, virus taxonomy ID (NCBI Taxonomy database), PCR cycle or hybridization temperature, buffer, dNTPs and $MgCl_2$ concentrations, polymerase used, PCR product size, note, type of hybridization or PCR (i.e. nested PCR) and a curator checking field. The latter is used by the curator to mark whether the entry was verified by the curator. If this field is marked as "Fix", the notification appears on the user's web page. The note field contains additional information, such as reverse transcription conditions or unique features of each entry. A nested PCR is either entered as one entry or separated into two or more entries. If the PCR conditions, in general, are different between first and second amplifications in the nested PCR, each amplification condition is entered separately. When a nested PCR is one entry, all Oligo data are listed in one Common data entry and information for primer-pairs are written in the note field. If a nested PCR is separated in two or more entries, each Common data entry contains one primer-pair, and VirOligo IDs for other nested PCRs are written in the note field in the VirOligo specific format.

The VirOligo database uses identifiers to link to other data sources. If one article has several entries (i.e. nested PCR), the identifier used is "[see *VirOligo ID*]". When the PHP script in the Publishing Server recognizes the "[see ******]" identifier, the script in the Publishing Server modifies the identifier as a clickable link to another entry. Further identifiers, "[PMID: ******]", "[GI: ******]", "[Taxonomy: ******]" link to the PubMed, GenBank and Taxonomy databases, respectively.

User data and Oligo user data tables are used for assigning VirOligo IDs and Oligo IDs, respectively. They store newly added data IDs for each user in the Entry Server. Every time users log in to the Entry Server, users are transferred to a user page or an Oligo user page depending on user-selection. Users can select whether they wish to modify the entry or create a new entry. If creating a new entry is selected, FileMaker assigns the latest data ID plus one as a data ID, and automatically enters it on the entry page.

At login and logout, users are required to use a punch clock and bulletin board web page. Punch clock data and bulletin board data are stored in a working hours table and a bulletin board table, respectively. Punch clock data are pasted to an Excel spread sheet file so that total working time for each user can be calculated automatically. The bulletin board web page is used for communication among participants. For example, a Common data entry person may ask an Oligo data entry person to fix an oligonucleotide sequence, or each user may report their accomplishments at the end of day.

All entries are stored in the Entry Server and backup is performed by MS-DOS scripts with the Window 98 embedded scheduled task program hourly from 9 AM to 7 PM for Common data and Oligo data files and every day at 4 AM for all database files. The files are not transferred to the Publishing Server until the curator verifies all entries. In addition, with each new virus, transfer does not occur until all articles for that virus are completed.

As mentioned above, oligonucleotide sequences are reconfirmed by Common data personnel, and all entries are checked by the curator. The curator verifies that reagent concentrations and PCR cycle values are reasonable. The curator web page contains automated links to PMID, GI and Taxonomy ID, and those are randomly checked by clicking the links and viewing the linked web sites. The curator confirms that oligonucleotide sequences are consistent with the lengths of oligonucleotides and the sizes of target regions. When the lengths and the sizes of target entered by an Oligo data person differ considerably from expectation, the curator checks all oligonucleotide sequences entered by that person by comparison with the original articles.

Additionally, a molarity calculator written by Java script has been made available from the Entry Server website. The calculator helps users who have difficulty in molarity calculations. It solves a final concentration in PCR reagent from initial concentrations and volumes added to PCR reagent and the total volume of PCR reagent. It also solves the final concentration when the same chemical is added separately at reverse transcription and at PCR.

The VirOligo Publishing Server

The Publishing Server runs under the Linux operating system with the popular web server program, Apache. Apache uses a structured query language, MySQL, for handling the database, and a script language, PHP, to create dynamic web pages, to search for related experimental conditions and to automatically add links to the NCBI databases. The set of programs for the Publishing Server were selected to provide users faster access to the database. FileMaker version 4 does not have enough functions to create dynamic web pages. Searching the database and formatting dynamic web pages requires more computational power than adding entries, and the processing speed of FileMaker version 4 was not sufficient for the Publishing Server. The Publishing Server was compromised once in December 2001 during the testing period. The cause was a security hole and a patch that had not been applied promptly. However, since all VirOligo data is stored in the Entry Server and the backup computer, the recovery process went smoothly. To prevent future security breeches, new security patches are applied within 24 hours of receipt of announcements from the Linux vendor using an email notification system offered by RedHat. At the same time, we also subscribe to CERT security alert to fix security holes before RedHat's security patches become available.

The Publishing Server uses MySQL as a database program and differs from the Entry Server. Thus, database files from the Entry Server need to be converted to a MySQL acceptable format. Both programs accept tab-separated files and comma-separated files, but those formats cannot be used since some entries contain tabs in the note field and the comma-separated file formats are slightly different between the two programs. In the VirOligo database system, database files are exported in HTML table format from the Entry Server and converted by a Perl script to a homemade format, which is similar to a comma-separated format. Converted files are transferred by secured shell (SSH) and added to the MySQL database in the Publishing Server.

The Publishing Server consists of two MySQL tables, Common data and Oligo data, and a PHP query retrieval server-side includes (SSI). The fields contained in the two tables are the same as those in the Entry Server except that the curator checking field used by the curator is omitted. Besides converting homemade identifiers to links to NCBI, the PHP script searches the database and creates dynamic web pages with summarized or detailed results. An oligonucleotide selection method was developed to provide user-friendly access to the oligonucleotide data. Users needing a common PCR cycle to detect several viruses can select oligonucleotides based on viral specificity and T_d. Users can compare oligonucleotides from a variety of fields, such as the target gene of viruses and oligonucleotide length. Search queries can be initiated with virus name, T_d, length of oligonucleotide, PMID or the VirOligo ID number. A summarized result is presented with virus name, oligonucleotide name and use, type of experiment (i.e. type of PCR), PCR cycle, T_d, publication date and PMID with the VirOligo ID number. By clicking the VirOligo ID number for each result, users can see the complete information for the identified entry. The detailed result has direct links inserted by the PHP script to BLAST for each oligonucleotide sequence, so that users can check up-to-date knowledge about the specificity of the oligonucleotide without typing the oligonucleotide sequences themselves.

The VirOligo database is able to accept volunteer data entries from colleagues. Since volunteers do not have User IDs needed by the Entry Server and the location of the Entry Server is not published due to security reasons, volunteer entries are handled in the Publishing Server. The Publishing Server possesses an empty set of the Common data and the Oligo data tables in addition to the one used for the VirOligo website. Volunteer entries are temporarily stored in these tables. Affiliations of volunteers are entered in the note field of the Common data. The website for volunteer entry has been available for four months, but no volunteer entry has yet been made. This lack is probably due to insufficient advertisement and the requirement for the entry of detailed information. As 58 primer sequences were entered by 15 volunteers in the Primers database in one and half years (Shomer, 1996), VirOligo could expect some volunteers if the volunteer entry asks only for oligonucleotide sequences. However, such a policy just brings uncertain oligonucleotide entries, and will be against the spirit of VirOligo.

The VirOligo Literature Management System

The expansion of viruses covered in the VirOligo database made literature collection tasks more complex. At early stages of the database, entry personnel printed the PubMed search results for each virus and marked articles off when they were obtained. Many articles are obtained through the Inter-Library Loan (ILL) system from which it takes up to months to receive articles. Finding the articles that were ILL requested and received in the PubMed print-outs became very time-consuming. As a result, a Literature Management System (LMS) was established. The idea of LMS derived from PubCrawler, a program to search newly added data in PubMed and GenBank (Hokamp & Wolfe, 1999). LSM obtains from PubMed a list of articles needed for the VirOligo processing and, upon user request, creates a list of articles along with the articles' physical locations. PubCrawler, written only in the script language Perl, was too complex for our application. Perl is a general purpose language, and can be used as the sole program for database and CGI in the web. However, the specialized programs such as MySQL for databases and PHP for a website program were easier to program for our purposes and sometimes operate faster. Thus, LMS was created in Perl borrowing the concepts from PubCrawler but using code sets that incorporated the use of MySQL and PHP scripts.

The LMS system consists of four MySQL tables (Journal Location, PMIDList, userdata and Reserved), three Perl scripts to obtain article information from PubMed and the Entry Server of VirOligo, and one PHP script to search the database and create a dynamic web page for users. Queries of articles being searched by LMS are stored in PMID.config in a PubMed query format of "*virus name* AND (PCR OR oligonucleotide)". Perl script 1 sends the queries to PubMed and compares returned PMIDs with PMIDs in a MySQL table, PMIDList. Then, Perl script 1 stores the unique PMIDs in PMIDList. Perl script 2 sends the newly added PMIDs, whose entries do not have VirOligo IDs in the Publishing Server, to the VirOligo Entry Server, and obtains the VirOligo IDs if the articles have been entered in the Entry Server. Perl script 3 obtains from PubMed and stores in the PMIDList detailed information (journal title, article title, authors, published date, volume, issue, and page numbers) of newly added articles without VirOligo IDs in either server.

When accessing the LMS, a user is asked for a password and username. Username and password are compared to a MySQL table, userdata, and verified by a PHP script. After verification, a web page asks for the number of articles requested, and selections of locations from Main Library, Vet Medicine Library, Online and ILL. After user selection, the PHP script obtains an article without a VirOligo ID from the PMIDList, and checks whether the PMID of the article is not listed in the 'Reserved' table and what the location of the article is. If the article is in the 'Reserved' table and the reservation has not expired, or the article location is different from the user's request, the PHP script searches for the next article. If the article reservation has expired or the article is absent from the 'Reserved' table and the article location matches the user's request, the PMID for the article is stored in the 'Reserved' table along with the user who requested it, the current time and expiration time. Then, detailed information about the article (PMID, article title, authors, journal title, volume, issue, page numbers, and call number assigned by the library or URL if the article is available online) appear on the web page. No searching of locations is required. The PHP script repeats the above process until the number of articles that the user requested is obtained. Since a call number or a URL is in the list, the person simply prints the list and finds the articles. The script also calculates estimated copying costs from the number of pages copied. Expiration time for the 'Reserved' table is set to assign the article to another person when a certain time period has passed. Expiration time is currently set at 20 minutes for online articles and 2 days for Main and Veterinary Medicine Libraries. If ILL is selected, the PHP script obtains personal information required by the ILL office from the userdata table, and creates an ILL request form for articles. Priority is set for the locations searched. If the article is available online, it does not appear in the list for Libraries or ILL. Only the articles that are not available elsewhere are requested from ILL.

Common mistakes in journal articles

In the process of the VirOligo data entry, a number of mistakes in journal articles were found. Although those were all kinds of careless mistakes made by authors, some mistakes were repeatedly found in the articles. Many common mistakes concerned oligonucleotide sequences and their target regions. Lengths of oligonucleotides were sometimes one nucleotide shorter or longer than lengths estimated from the target regions. For example, a target region was written as 200-220 when a 20 nt long oligonucleotide has a target region of 200-219. The length of an oligonucleotide with target region of 200-220 is 21 nt. Another common mistake in oligonucleotide sequences were that a pair of primers consists of two forward or two reverse primers based on BLAST searches. It was probably caused by the authors finding primer sequences from a genome sequence during writing a manuscript, and forgetting to write a complementary sequence. These mistakes can be easily avoided by BLAST search before writing a manuscript.

Units in the journal articles also commonly contained mistakes. One PCR tube usually contains 10 to 50 μL of PCR reagents. Instead of 10 to 50 μL, some articles wrote a PCR volume as 50 mL. The interchanges of units can be caused in computer programs. When word processor files containing "μ" are converted to text files or pasted into e-mail, computer programs change "μ" into "m" since ASCII text format does not have "μ" defined. Thus, authors should take care of some special characters when files are converted.

Current coverage in VirOligo

As of May 2002, the VirOligo database contained 16 completed viruses and is being expanded. The strategy adopted to fill the database was to obtain as near complete coverage as possible of one group of viruses at a time. For each new virus, articles are identified whose PubMed abstract contains the virus name and 'PCR' or 'oligonucleotide'. Since the PubMed searches retrieve all articles with a particular virus name, articles do contain oligonucleotides specific to viruses other than the searched viruses. Thus, after the completion of entries for sixteen viruses, the VirOligo Entry Server contained oligonucleotides for more than 100 viruses.

VirOligo was established as a means to create ViSH chips and ViSA cards so that one can monitor and survey the environment for the appearance of novel viruses or the disappearance of existing viruses. However, feedback from VirOligo users indicates that the database is found useful by diagnostic laboratories since it simplifies their searches for assays to detect particular viruses.

To assess the reputation of the database, access events were monitored. More than 1200 visitors accessed the index page of the VirOligo database in the first 4 months of 2002 (the number excludes any accesses made within one minute of each other from the same access point). Since some accesses directly obtain the database search results through links in web pages, such as Yahoo and Google, a larger number of visitors probably occurred. The number of searches made in the VirOligo database were more than 7400 in the four months ending April 30, 2002. A related database, MPDB, was published in Nucleic Acid Research six times, and 554 searches were made by its users in eight months from January 1997 to August 1997 (Campi et al., 1998). Thus, people's interest in the VirOligo database is significantly higher than in another oligonucleotide database. Science introduced the VirOligo database in their Net Watch column (Vol 295, Issue 5561).

The number of websites with links introducing the VirOligo database is increasing. Websites simply listing databases introduced in the Nucleic Acids Research Database issue were excluded from the following analysis. A wide variety of academic institutions display links to VirOligo. VirOligo is linked from professional organizations (ASM, Ohio Branch; The American Phytopathological Society), university libraries (University of Manitoba Libraries; Westfälische Wilhelms-Universität Münster, Germany), scientific institutes (Università di Genova, Department of Surgical Specialistiche Sciences, Anesthesiology and Transplant of Organs, Italy; Institut Pasteur, France; Institute of Biochemistry, University of Provence, France; Biological Research Information Center, The Pohang University of Science and Technology, Korea; Institute of Animal Health, UK), academic projects (Virology down under, The University of Queensland, Australia; Internet Scout Project, The NSDL Scout Report for Life Sciences, Computer Sciences Department, University of Wisconsin-Madison) and a museum (Exploratorium, the museum of science, art, and human perception at San Francisco). Moreover, the links are not only from US institutions but also foreign institutions. Thus, the VirOligo database has been globally recognized just four months after its first announcement.

Conclusions

The reputation of VirOligo is rapidly growing. The collection of virus-specific oligonucleotides is unique and can be used not only for finding proven PCR methods for listed viruses but also for finding optimum PCR conditions for all viruses and obtaining oligonucleotides to detect a wide variety of viruses in one trial by the microarray methods, such as ViSA and ViSH. The system of VirOligo is stable and well designed for continuous future usage. However, it is still a starting point, and VirOligo needs wider virus coverage. As of May 13, 2002, 39,479 articles of PubMed contain 'virus' and ‘PCR’ or “oligonucleotide’ in their abstracts, and VirOligo contains approximately six percent of the 39,479 articles. Collecting all the articles in PubMed will help researchers in many ways, and ViSH chips or ViSA cards that detect all viruses in one experiment will be possible. More accurate analysis of PCR conditions is also possible with more oligonucleotides. The VirOligo database is available from the publishing server at http://viroligo.okstate.edu/. All entries in the entry server are transferred immediately after curator’s verification.

Possible additional modules for VirOligo in the near future

Boosting data entry speed simply needs more manpower, but entry can be more efficient with additional functions listed in next sections.

BLAST for PCR primers

Currently, Oligo data entry requires GI numbers unless GI or Accession numbers are listed in the article. To obtain GI numbers for PCR primers, Oligo entry personnel perform BLAST for each primer. PCR usually uses a pair of primers, so that Oligo entry personnel obtain two BLAST results. In the case of nested PCR or multiplex PCR, the number of BLAST results is more than two. The number of perfect matches with the primer sequences in BLAST results sometimes exceeds fifty GenBank nucleotide entries, and each entry has its own GI number. The entry personnel need to compare more than two BLAST results and find one GI number matched with both (or more in the case of more complex PCRs) BLAST results. When BLAST results contain more than fifty perfect matches, more than one GI number often matches in both results, and the entry person selects one GI number by finding the virus name used in the article. If the entry person overlooks matched GI numbers or enters the wrong sequence, finding a GI number causes a lot of confusion. It sometimes took more than one hour to select a GI number. VirOligo uses NCBI BLAST search, but it is sometimes slow due to system overload. Thus, it would be beneficial if VirOligo could automatically process BLAST for PCR primers.

In a BLAST for PCR primers system, the Oligo entry personnel enter primer sequences first. The system consists of two parts. First, the system sends each sequence and obtains the request ID (RID) for each sequence from NCBI. After sending all sequences, the system sends each RID and obtains BLAST result for each RID. The BLAST results are stored in MySQL database. The procedure has been programmed by Perl script, and has obtained 100 BLAST results in 5 minutes.

Second, the system compares the BLAST results for primer pairs and shows users which GI numbers matched. Since the entry person has fewer lines to read in the results of BLAST for PCR primers and there is no need to compare several BLAST results, the work efficiency will be improved. Moreover, the entry person will not need to wait for BLAST processing, and the entry person keeps either entering the primer sequences or evaluating the result and creating the Oligo entries.

Advertisement and Volunteer system, Email to authors

Since the opening of the VirOligo database, no volunteer entry has been made yet by colleagues. The causes are probably insufficient advertisement of the database and that colleagues do not know that we need their entries. To announce our needs and the database, we propose to send emails to authors of articles. The Advertisement and Volunteer System (AVS) will search newly added articles whose abstracts contain 'virus' and 'PCR' or 'oligonucleotide' from Current Contents (Institute of Scientific Information, Philadelphia). Current Contents will be chosen because not all PubMed abstracts contain email addresses of authors while Current Contents does. Since LMS manages all entries by PMID, AVS will search PMIDs in PubMed from article title and author names obtained from Current Contents, and store them in MySQL database to avoid resending the same request. AVS will send an email to the author with PMID, author names, article title and journal names, and ask the author to enter the oligonucleotide data in the VirOligo database.

Linking with other databases

We are currently exploring the possibility of linking our contents with International Committee on Taxonomy of Viruses (ICTV). Linking with other databases will be beneficial by increasing the traffic to the VirOligo database, thereby increasing the reputation of VirOligo. NCBI also offers LinkOut to place direct links in PubMed, GenBank and Taxonomy database. The VirOligo database can be searched by PMID but not by GI or Taxonomy ID, so that, it would be necessary to add more search functions to the VirOligo database. After modification, we will prepare the files that NCBI requested.

Chapter 4: Application of the VirOligo Database for the ViSH Chip Method

Results

Total 480 influenza specific oligonucleotides, whose sequences were obtained from the VirOligo database and OMIGA primer design software, were synthesized and printed on microarray chips. Sequences of oligonucleotides are available in Appendix 1. Universal primers are specific to all segments of all influenza A viruses. Cy dye labeled PCR products were prepared as hybridization targets using universal primers and two equine influenza viruses, A/Equine/Kentucky/98 (KY98); A/Equine/Miami/63 (MI63), and three human influenza viruses, A/Hong Kong/68 (HK68); A/New Jersey/76 (NJ76); A/Panama/99 (PAN99). After hybridizations with targets, the amount of products in association with each oligonucleotide probe was determined by measuring the fluorescence intensity using a fluorescence scanner (Scanarray 3000, Packard BioScience, Meriden, CT). A scanned image with a KY98 target is shown in Figure 4-1 as an example.

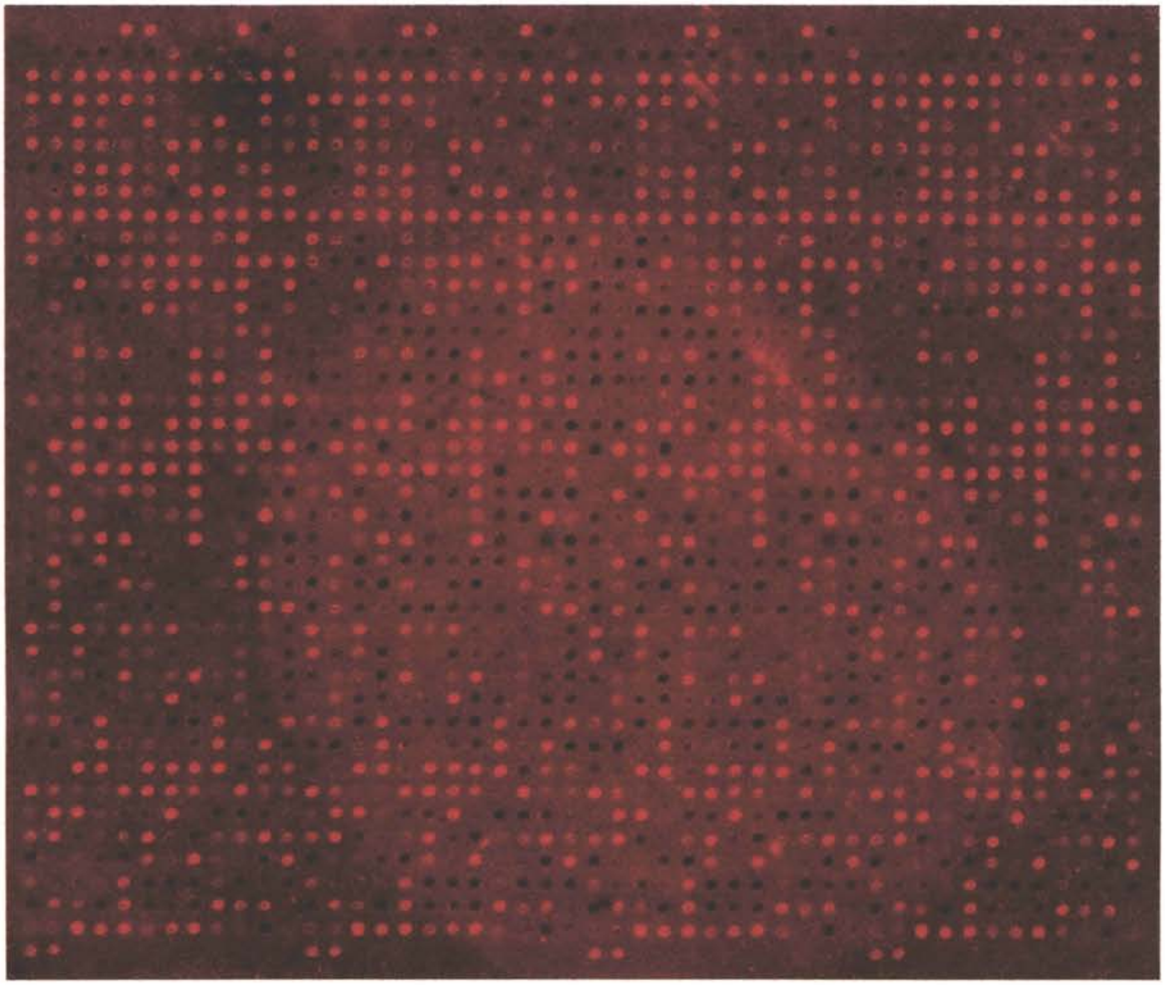


Figure 4–1, A scanned image sample of a ViSH chip with a KY98 target.

The amounts of hybridization for KY98 specific oligonucleotides were extracted from the scanned results by a data acquisition program, GenePix Pro 4 (Axon Instruments, Inc., Union City, CA). The extracted results contained intensities of hybridizations for thirteen oligonucleotides, whose sequences were obtained from the VirOligo database, fourteen oligonucleotides designed by OMIGA software, and two universal primers (Table 4-1). Negative values in the table indicate that the background intensity was stronger than the signal intensity. The range of theoretically possible intensities ranges from –65536 at no hybridization with very intense background to 65536 at very intense hybridization without background. However, it is rare that the background is substantially more intense than the hybridization spot. Thus, the intensity values for no hybridization are close to zero. Spots of printed probes on the ViSH chip had diameters of approximately 150 μm and were spotted at four positions for each probe. When hybridization intensity was more than 1200, fluorescence scanner images showed clear circles in all four positions for most spots. At intensities under 1200, spots could not be recognized easily at all four positions.

Table 4-1, Hybridization intensities of KY98 hemagglutinin segment specific probes and five targets, A/Equine/Kentucky/98 (KY98), A/Equine/Miami/63 (MI63), A/Hong Kong/68 (HK68), A/New Jersey/76 (NJ76), A/Panama/99 (PAN99).

	KY98	HK68	NJ76	PAN99	MI63
VirOligo database					
34	60895	61549	26949	20468	5584
62	-968	13308	7896	11376	13245
82	5664	3132	513	-173	-215
83	7146	61341	3723	3143	1775
84	5345	28380	8552	10360	4912
160	4748	1338	104	833	20587
181	9546	9196	8394	10512	9392
197	11980	92	2771	-157	39156
221	12293	36546	17078	28609	14142
274	21756	62313	8040	6956	840
386	3472	-1015	-515	-2848	-668
486	9103	1220	4291	348	236
591	11170	2367	426	6391	21971
OMIGA generated					
K1	33143	1766	4823	5494	2995
K10	5861	2751	5373	9140	2273
K2	70	-517	28	-1831	-28
K3	8899	23	185	-1251	477
K4	21	-954	12	-1894	-184
K5	1956	-650	181	-1455	277
K51	1358	-409	1	-1148	7976
K52	31268	4542	3788	4668	17973
K54	18993	4644	48173	44670	36960
K55	25110	10458	48941	45190	51961
K56	5296	1417	14757	13260	14864
K58	9565	5552	15551	14323	24523
K59	60396	30820	63412	62067	64681
K8	3815	3679	4213	2600	1446
Universal Primers					
Uni5	517	22453	282	-310	5185
Uni3	-2405	2929	-222	-2271	-215

Note: From top, sequences for first thirteen probes were obtained from the VirOligo database. Next fourteen probes were obtained from a primer design program, OMIGA, and last two probes were universal probes. Shaded cells indicate the specificity to the targets. For example, Oligo 34 is specific to KY98 and MI63, and Uni5 is specific to all targets.

Since all probes in Table 4-1 were specific to the KY98 HA segment, appreciable levels of hybridization were expected for the probes when the KY98 target was used in the ViSH chip. However, the intensities of hybridization between KY98 specific probes and KY98 targets were distributed from –968 (Oligo 62) to 60895 (Oligo 34). The hybridization result was similar for the MI63 target. Nineteen KY98 specific probes were also specific to MI63 because both KY98 and MI63 are equine influenza viruses and H3N8 subtypes. The hybridization intensities of MI63 specific probes with MI63 target ranged from –217 (Oligo 82) to 64681 (Oligo K59), and the intensities were scattered widely.

HK68, NJ76, and PAN99 targets were also tested for hybridization with KY98 specific probes. BLAST analysis (BLASTn, NCBI; results obtained at expect value 100) revealed no sequence similarity between KY98 specific probes and HK68 targets. Since complete sequences of all HK68 segments are available at GenBank and no probes were specific to HK68, no hybridization should occur at the KY98 specific probes. The distribution of hybridization intensities was scattered and ranged from negative to more than 60000 with HK68 target. The results indicated non-specific binding of the targets to some of the KY98 probes. The results for hybridization for PAN99 and NJ76 targets were similar to the KY98 target and distributions of hybridization intensities were scattered. Only a partial sequence of the HA segments is available for PAN99, and no sequence of the HA segment is available for NJ76 at GenBank or the Influenza Sequence Database (http://www.flu.lanl.gov/). Thus, expected reactivity with the KY98 probes was known for neither PAN99 nor NJ76. However, non-specific binding of the targets was suspected for PAN99 and NJ76 since both are human viruses and no matches were found with the partial sequence of PAN99 HA segment when pairwise BLAST searches (BLAST2, NCBI; results obtained at expect value 100) were performed with all KY98 specific probes.

The negative intensities with KY98 targets were not due to inappropriate hybridization temperature. Universal probes in Table 4-1 were 50 %(G+C) for Uni3 and 33 %(G+C) for Uni5 and both 12 nt length. Because the universal primers were shorter than KY98 specific probes, which were at least 18 nt length, the universal primers had lower T_ms than the KY98 specific probes. Two targets for Uni5 and one target for Uni3 generated hybridization intensities of at least 2900, and the HK68 target generated an intensity of 22453 with Uni5. Thus, the universal probes were hybridized at the hybridization temperature used, and the hybridization temperature should be low enough for all probes.

The universal probes were included in the array of the ViSH chips as standards. Targets were prepared by PCR. Amplification and labeling efficiencies, or PCR template concentrations vary in PCR, so that hybridization intensities in the scanning results probably differ among the targets due to differences in amounts of the labeled PCR products. Thus, the differences in the hybridization intensities due to the labeled PCR products need to be adjusted for comparisons of the hybridization intensities among targets. Since the universal probes were specific to all influenza A targets, the hybridization intensities in the universal probes were theoretically equal for all influenza A targets under the same experimental conditions. Adjustments of a ViSH result can be performed by calculating the coefficient to equalize the hybridization intensities in the universal probes between results for two targets, followed by multiplying all the hybridization intensities by the coefficient. In the ViSH results shown in Table 4-1, the hybridization intensities of Uni3 and Uni5 were also scattered and some were negative values, even though both universal probes share the same sequences with all targets. Since the hybridization intensities contained negative values, the universal probes could not be used for standards. The sequences of Uni3 and Uni5 were similar. Only three nucleotides were different. Thus, the 5' end of the target may self-anneal to the 3' end of the same target strand when the target was amplified with these universal primers, and high intensity values may be due to hybridizations between Uni3 and Uni5.

Even though the comparison of the hybridization intensities between targets is not appropriate due to variable target concentrations, some notable results can be seen in Table 4-1. The intensities for Oligo 386 were 3472 for the KY98 target and negative for all others. An intensity value of 3472 is usually associated with a visually identifiable spot, and negative values indicate that the background level is more intense than the signal and no or a trace amount of hybridization was detected. Thus, it was possible to determine that Oligo 386 detected the KY98 target and differentiated it from other targets although the hybridization intensity was low. On the other hand, the hybridization intensity of Oligo 221 was 12293 with KY98 target, but all other targets generated higher intensities with values from 14142 to 36546. Thus, Oligo 221 hybridized with KY98 but did not differentiate it from other targets.

Some probes, including Oligo 221 and 34, exhibited very high hybridization intensities while others had very low intensities. A correlation was sought between hybridization intensity and probe properties in length, T_m, %(G+C), target genome position, or number of mismatches. Thus, the correlations were examined (Table 4-2).

No significant correlation in probe length and hybridization intensity was found. For example, Oligo 62 was 19 nt long and generated a negative intensity but 20 nt (Oligo 34) and 18 nt (Oligo K52) probes generated intensities of more than 10000. Probes shorter or longer than 19 nt did not hybridize as poorly as 19 nt did. Additionally, no significant correlations were found between the hybridization intensity and target strand and region, T_m, or %(G+C). The first 17 nt of Oligo 82, 83, and 84 shared the same sequence. Oligo 84 was a perfect match to the KY98 target, and Oligo 82 and 83 had mismatches to the KY98 target at their 3' ends. The hybridization intensities of Oligo 82, 83, and 84 were 5664, 7146, and 5345, respectively, and no significant differences in the intensity from the 3' end mismatches were found. Oligo 274 and 591 had 5' end mismatches to the KY98 target, and the hybridization intensities for both probes were more than 10000. Oligo 591 and K2 were perfect matches to the KY98 target but have an internal mismatch with the MI63 target. Since Oligo 591 had high intensity values (11170 for the KY98 target and 21971 for the MI63 target) and Oligo K2 had low intensity values for both targets (70 for the KY98 target and –28 for the MI63 target), significant differences could not be observed between with and without an internal mismatch. Thus, no negative effect of mismatches on hybridization was found from mismatches on the ends and an internal mismatch.

Table 4-2, Probe lengths, T_ms, %(G+C) contents, and genome and matched positions along with the hybridization intensities with the KY98 target.

	Intensity				Genome position on		Matched region on	
	KY98	Length	Tm (°C)	%(G+C)	KY98	Strand	KY98	MI63
From VirOligo								
34	60895	20	47.5	35.0	360-379	+	PM	1-18
62	-968	19	50.6	47.4	679-661	-	PM	
82	5664	24	56.8	41.7	1-17	+	1-17	1-17
83	7146	22	55.0	45.5	1-17	+	1-17	1-17
84	5345	25	58.3	44.0	1-25	+	PM	PM
160	4748	20	58.5	60.0	843-824	-	PM	
181	9546	20	48.5	40.0	322-341	+	PM	1-16
197	11980	21	50.1	38.1	831-811	-	PM	5-21
221	12293	24	54.1	41.7	9-32	+	PM	PM
274	21756	20	56.7	55.0	1-17	+	4-20	4-20
386	3472	19	44.1	31.6	208-226	+	PM	
486	9103	20	54.6	55.0	286-305	+	PM	
591	11170	20	56.8	55.0	847-831	-	4-20	1MM
OMIGA generated								
K1	33143	19	49.0	47.4	71-89	+	PM	3-19
K10	5861	19	52.9	57.9	294-312	+	PM	
K2	70	22	52.5	45.5	326-347	+	PM	1MM
K3	8899	18	47.0	44.4	72-89	+	PM	2-18
K4	21	19	49.0	47.4	363-381	+	PM	
K5	1956	21	51.1	42.9	361-381	+	PM	1-17
K51	1358	18	47.5	44.4	539-522	-	PM	PM
K52	31268	18	51.2	55.6	753-736	-	PM	
K54	18993	18	48.2	50.0	764-747	-	PM	PM
K55	25110	20	51.3	50.0	766-747	-	PM	PM
K56	5296	19	49.5	47.4	765-747	-	PM	PM
K58	9565	22	56.5	50.0	768-747	-	PM	PM
K59	60396	21	56.4	57.1	760-740	-	PM	1-17
K8	3815	21	57.5	57.1	292-312	+	PM	
Universal Probes								
Uni5	516.5	12	21.7	33.3	5' end	-		
Uni3	-2405	12	31.9	50.0	3' end	+		

Note: In matched region, PM stands for perfect match, and MM stands for mismatches in the middle of probe (1MM = one mismatches). Range of matched region is shown for 5' or 3' end mismatch. T_m was calculated by nearest-neighbor method (SantaLucia, 1998).

As Oligo 82, 83 and 84 shared the same region of the KY98 target, the target regions for Oligo K52, K54, K55, K56, K58 and K59 were also close to each other as shown in Table 4-3. When two nucleotides from the 5' end of K58 were removed, the hybridization intensities increased from 9565 (Oligo K58) to 25110 (Oligo K55). However, the hybridization intensities of probes that had three and four nucleotides removed from the 5' end of K58 were 5296 (Oligo K56) and 18993 (Oligo K54), respectively. Since Oligo K54, K55, K56 and K58 contain the same sequence, the intensity value should be similar if there was no effect from 5' end sequences differences. However, the intensity values were scattered from 5296 to 25110. Thus, the differences in the intensities may be due to 5' end nucleotides.

The sequence alignments of Oligo K51 to K59 did not suggest a cause of the variable hybridization intensities, but free energies of hairpins in Oligo K52 to K59 were a function of the hybridization intensities (Table 4-4; correlation coefficient, 0.943). No hairpin loop was predicted in Oligo K59 and the hybridization intensity was 60396. As the free energies of the hairpin loops decreased, the hybridization intensities declined. The hybridization intensity was 31268 (Oligo K52) at the free energy of the hairpin –0.91 kcal/mol, 25110 (Oligo K55) and 18993 (Oligo K54) at –1.56 kcal/mol, and 9565 (Oligo K58) at –1.79 kcal/mol. Only the hybridization intensity of Oligo K56 did not follow the order of the free energies of the hairpin loops. Thus, all KY98 specific probes were examined to see whether the hairpin loops were correlated with the variable hybridization intensities (Table 4-5). However, no correlation was found between the hairpin free energies and the hybridization intensities for the entire set of KY98 specific probes (Correlation coefficient, 0.222). In addition to the hairpin, the free energies of dimer formations and base compositions of the probe sequences were examined but no correlations were found with the hybridization intensities (Correlation coefficients, 0.117 for dimer and -0.289 to 0.368 for base composition A to T).

Table 4-3, Sequence alignments along with the target regions and the hybridization intensities with the KY98 target.

Oligo ID	Target region	Intensity	Sequence
K58	768-747	9565	TGCTTATCCTGCCTGATTGACC
K55	766-747	25110	CTTATCCTGCCTGATTGACC
K56	765-747	5296	TTATCCTGCCTGATTGACC
K54	764-747	18993	TATCCTGCCTGATTGACC
K59	760-740	60396	CTGCCTGATTGACCCCTAACC
K52	753-736	31268	ATTGACCCCTAACCCACG

Table 4-4, Comparison of free energies of hairpin loops and the hybridization intensities with the KY98 target.

Oligo ID	Intensity	Hairpin (kcal/mol)
K59	60396	0
K52	31268	-0.91
K55	25110	-1.56
K54	18993	-1.56
K58	9565	-1.79
K56	5296	-1.56

Note: Free energies of hairpins were obtained from Oligo Analyzer (http://www.rnature.com/oligonucleotide.html) using nearest-neighbor method.

Table 4-5, Intensities with a fluorescent nucleic acid stain (SYTO 61), free energies of hairpin and dimer formations and base compositions of the probes along with the hybridization intensities with the KY98 target.

	Intensity				Base composition			
			Hairpin	Dimer				
	KY98	SYTO 61	(kcal/mol)	(kcal/mol)	A	C	G	T
From VirOligo								
34	60895	28749	-0.91	-4.2	7	5	2	6
62	-968	26152	-2.76	-2.76	3	5	4	7
82	5664	22043	-0.91	-1.79	12	3	7	2
83	7146	29488	-2.76	-2.76	8	4	6	4
84	5345	20711	-2.01	-2.01	8	4	7	6
160	4748	18174	0	-9.89	5	7	5	3
181	9546	14804	-0.91	-2.22	6	4	4	6
197	11980	22883	-2.06	-7.19	6	5	3	7
221	12293	9644	-2.01	-2.01	6	4	6	8
274	21756	10460	-1.79	-3.08	8	4	7	1
386	3472	18142	-3.06	-3.06	8	2	4	5
486	9103	19088	-1.36	-4.24	5	6	5	4
591	11170	11018	0	-9.89	4	6	5	5
OMIGA generated								
K1	33143	23821	-0.96	-1.42	8	8	1	2
K10	5861	12541	-0.96	-4.24	2	8	3	6
K2	70	8886	-3.08	-3.08	8	5	5	4
K3	8899	13968	-0.96	-1.42	8	7	1	2
K4	21	10607	-0.91	-4.2	5	7	2	5
K5	1956	12184	-0.91	-4.2	6	7	2	6
K51	1358	12163	-0.91	-2.32	4	4	4	6
K52	31268	13739	-0.91	-0.91	5	8	2	3
K54	18993	17912	-1.56	-1.79	3	6	3	6
K55	25110	16334	-1.56	-1.79	3	7	3	7
K56	5296	15164	-1.56	-1.79	3	6	3	7
K58	9565	12854	-1.79	-1.79	3	7	4	8
K59	60396	15764	0	-1.79	4	9	3	5
K8	3815	10660	-0.96	-4.24	3	9	3	6
Universal Probes								
Uni5	516.5	11615	0	-0.96	7	1	3	1
Uni3	-2405	11437	0	-1.79	6	2	4	0

Note: Free energy of hairpin and dimer was obtained from Oligo Analyzer (http://www.rnature.com/oligonucleotide.html) using nearest-neighbor method.

All probes were printed at four positions, and the hybridization intensities of the four positions may be scattered due to variable qualities of the slide surface. Thus, maximum and minimum hybridization intensities of KY98 specific probes were examined (Table 4-6). Although the ratios of maximum to minimum intensities were spread wider with probes with low hybridization intensities than with those with high intensities due to noise in the scanning image, the maximum difference between the maximum and the minimum intensities was 7931 (Oligo 274), and the scattered hybridization intensities of the KY98 specific probes could not be accounted for by this variation. Thus, the variation in intensities was not due to a poor quality of slide surfaces.

Table 4-6, The hybridization intensities of the KY98 specific probes and the KY98 target.

	Median	Max	Min	Min/Max
VirOligo database				
34	60895	61428	56159	0.91
62	-968	-752	-2037	2.71
82	5664	7182	1842	0.26
83	7146	7700	2308	0.30
84	5345	7426	1559	0.21
160	4748	6385	2640	0.41
181	9546	13745	5937	0.43
197	11980	15230	8586	0.56
221	12293	16445	11170	0.68
274	21756	24374	16443	0.67
386	3472	4198	122	0.03
486	9103	9412	8119	0.86
591	11170	11787	10612	0.90
OMIGA generated				
K1	33143	33783	29190	0.86
K10	5861	6975	4769	0.68
K2	70	966	-530	-0.55
K3	8899	10582	7503	0.71
K4	21	412	-537	-1.30
K5	1956	2520	719	0.29
K51	1358	4497	797	0.18
K52	31268	31908	28465	0.89
K54	18993	19689	16736	0.85
K55	25110	26946	21808	0.81
K56	5296	5814	3580	0.62
K58	9565	10553	8131	0.77
K59	60396	61412	59681	0.97
K8	3815	4499	2914	0.65
Universal Primers				
Uni5	517	1121	-558	-0.50
Uni3	-2405	-1846	-3373	1.83

Note: All probes were printed at four positions. Maximums, minimums and medians of the hybridization intensities from four replicas were listed.

Discussion

The purpose of the ViSH chip experiments was a comparison of probes obtained from VirOligo with those from a probe design program. However, the hybridization intensities for both groups of probes were scattered, and I could not determine a superiority of one method over another to obtain the probes for the ViSH chip. Thus, possible causes of the scattered intensities were examined in the Results. However, no significant facts to identify the cause were found in the analysis.

Hybridization temperature was sufficiently low for hybridization of all probes, but some probes did not generate detectable levels of hybridization with the KY98 target. All probes were specific to KY98 HA segments and the presence of the HA segments in all PCR products used as targets was confirmed in gel electrophoretic images (data not shown). All images contained 1.8 kb HA segments. Attachment of all probes on the chip was confirmed by a red fluorescent nucleic acid stain, SYTO 61 (Molecular Probes, Inc., Eugene, OR). Although the intensities with the stain for the KY98 probes varied up to 3.3 times, the intensities were not correlated with the hybridization intensities with the KY98 targets (Table 4-5). Thus, the causes of the scattered intensities were not due to the preparations of targets and printed slides.

The intensities of Oligo K52 to K59 probes were related to the formations of the hairpins. Even though the relation was not supported by the entire KY98 specific probe set, hairpins may play a role in hybridization efficiency as observed in Oligo K52 to K59. Thus, probes with hairpins should be avoided. Array designer 2 (PREMIER Biosoft International, Palo Alto, CA) is a program to design probes for the oligonucleotide microarray, and the program selects probes with fewer hairpins and dimers within user-selected ranges of probe T_m and length. A similar algorithm was used in OligoArray (Rouillard et al., 2002).

Another approach for probe selection in oligonucleotide microarray was introduced by Lockhart et al., (1996) and Relogio et al., (2002). The probe selection program using their approach was created by Shteynberg (the program, Featurama, is available at http://probepicker.sourceforge.net/; pages viewed on Aug. 16, 2002). In this approach, probes were excluded if the numbers of (A+T) or (C+G), the number of single-base repeats or the number of hairpins was exceeded. No KY98 specific probes were excluded by the standards for single-base repeat and hairpins set by Lockhart et al., (1996) and Relogio et al., (2002). Since the numbers of (A+T) or (C+G) were set for 20 nt (Lockhart et al., 1996) and 25-35 nt (Relogio et al., 2002), and the probes in the ViSH chip were 12 to 29 nt, the standards were not applicable to all the probes in the ViSH chip. Although the numbers of (A+T) or (C+G) correlate with %(G+C), no specific %(G+C) contents were found to generate weak or strong hybridization intensities in the KY98 specific probes (Table 4-2).

Cushman (Personal communication, 1999; University of Nevada at Reno) found that the hybridization buffer we used caused non-specific binding of targets with probes. Thus, the hybridization buffer may be a cause of hybridization between the KY98 specific probes and non-KY98 targets. Stringency of slide washing steps may not be enough to remove non-specific binding from non-KY98 targets, or too stringent and remove target specific probes from KY98 and MI63 targets. However, the binding strength between targets and probes is a function of the probe T_ms. Since correlations between T_m and the hybridization intensity were not found in all targets, the stringency of washing cannot explain why only some probes generated very low or negative hybridization intensities.

The aldehyde slide (Catalog ID: ALS-25) was used in the experiments. The slides were purchased from TeleChem International, Inc. The company offers two qualities of the aldehyde slides, the ones used in these experiments cost $40 per 25 slides and the others cost $250 per 25 slides (Catalog ID: SMABC). The ViSH chip methods were initiated in my MS thesis (Onodera, 1999). At that time, an array-printing machine (arrayer) was not used, and each probe was spotted manually by a pipette. Thus, each manually spotted area was considerably larger than arrayer-generated spots, and the quality of the slide was probably not important in such larger spots. The size of each spot in arrayer-generated slides had a diameter of 150 μm. When the spot is smaller, uniformity of the slide surface may affect the spot intensities. As mentioned in Results, maximum differences of the hybridization intensities were more than 7000, and the quality of slide surface may be one of the causes of the scattered intensities.

The significant causes of the failure could not be determined by the analysis. Similar results were obtained by Wang, et al. (2002). They tried to detect human intestinal bacteria in fecal samples by oligonucleotide microarray hybridization using aldehyde slides. Ten out of sixty probes did not hybridize to the target for unknown reasons, and the hybridization intensities were scattered. Therefore, the results obtained in the ViSH chip were not anomalous nor due to experimental failure. I assume that complex interactions among the hairpin loops, the hybridization buffer and the slide surface quality may interfere with the analysis of the ViSH chip results.

Many problems on the ViSH chip were discussed, but there was one positive result. At least one probe (Oligo 386) with low hybridization intensity differentiated KY98 from other influenza viruses, and another probe with a high hybridization intensity (Oligo 221) did not differentiate KY98 from other influenza viruses at all. For detection of viral nucleic acids, for each probe one needs to identify whether the target hybridized to the probe based on the hybridization intensity. Thus, a threshold value of the fluorescence intensity needs to be determined for the virus detection. The finding indicated that using only one threshold of the hybridization intensity to define positive or negative in hybridization is not valid since the hybridization efficiency is different for each probe. The oligonucleotide microarray experiments performed by Lie et al. (2001) also support the finding. In their study, universal probes were used to detect five closely related bacilli. The hybridization intensity to one probe was up to 11 times more intense than to another probe when the same *Bacillus* species were hybridized. Thus, each probe must be tested with positive (probe specific) targets, and the ViSH chip threshold value needs to be defined for each probe before testing an unknown target.

Chapter 5: Primer Design

Results

Primer lengths and G+C contents

Primer length and G+C content are among the most mentioned considerations in PCR primer design. To examine the ranges of primer length and G+C content that are used the most frequently, 2,137 virus-specific PCR primers were obtained from VirOligo. Most of primers used for PCR were between 18 and 22 nt long (65.8 %; Figure 5-1). Primers were most frequently 20 nt long (30.4 %). Thus, accumulated practice suggests that, if possible, one should select primers from 18 to 22 nt long. Although some sources recommend using primers of 18 to 30 nt (Newton, 1995; McPherson, et al., 1991), 16 and 17 nt long primers were used more frequently (Average 2.7 %) than primers between 27 and 30 nt long (Average 1.3 %). Thus, 16 or 17 nt long primer can be used with pure template since increasing selectivity using longer primers is not necessary with pure viral PCRs.

From 40 to 60 % is the generally recommended range for the G+C contents of PCR primers, and 78.0 % of PCR primers reported in the VirOligo database were in this range (Figure 5-2). The most frequent G+C content was 50 % (23.7 % of all primers). However, 22 % of primers have G+C contents outside of the range. BHV-1 specific primers have a different distribution of G+C contents (Figure 5-3). G+C contents of 65 % are the most frequent (23.9 %) among 134 BHV-1 specific primers and their distribution of G+C contents differed significantly from that of the entire collection of oligonucleotides (t-test, p=2.6E-32).

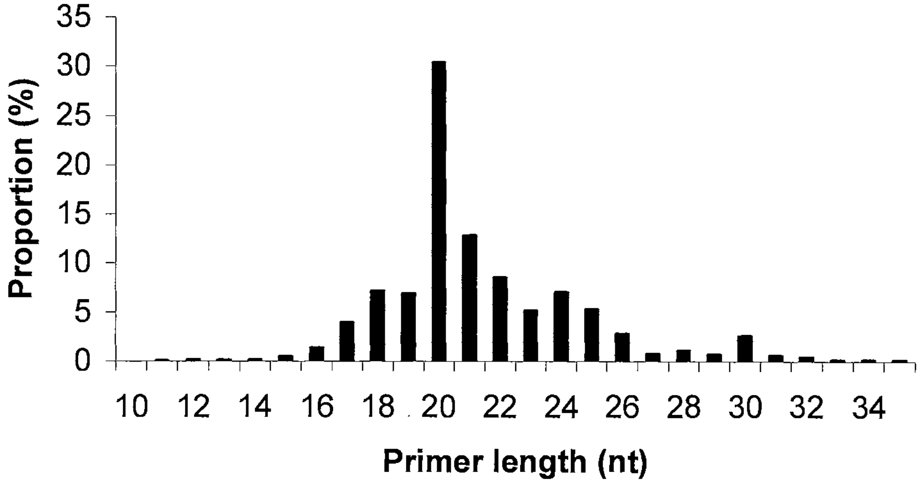


Figure 5–1, Distribution of lengths in entire primers.

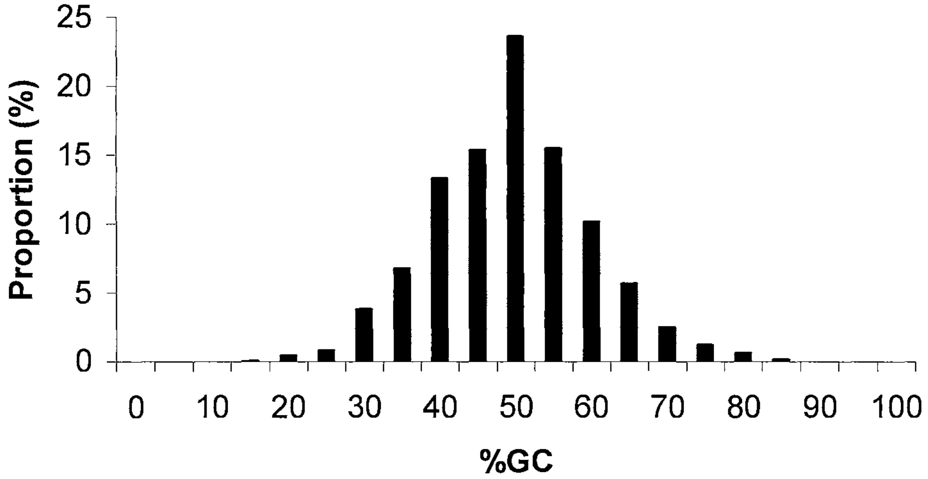


Figure 5–2, Distribution of %(G+C) contents in entire primers.

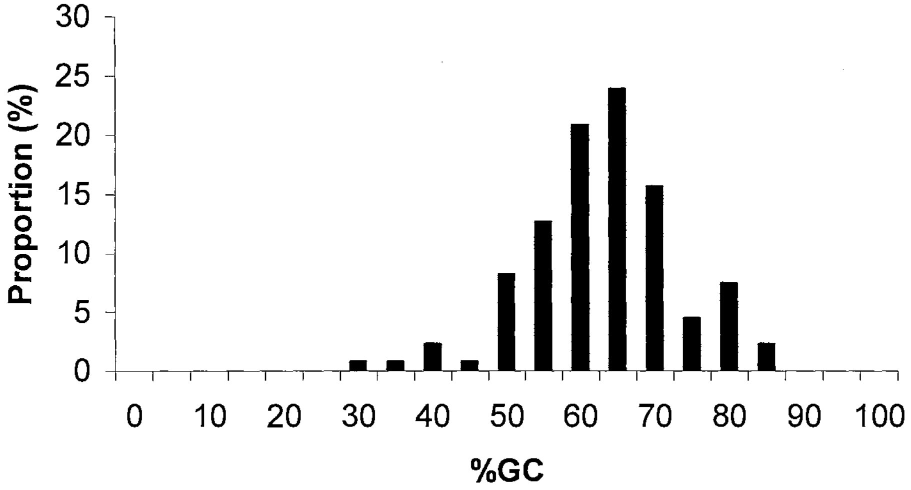


Figure 5–3, Distribution of %(G+C) contents in BHV-1 specific primers.

Primer melting temperatures (T_m)

PCR depends on the melting and reannealing characteristics of DNA. Since the distribution of G+C content for BHV-1 primers differed from that for the collection of all primers, the optimal T_m for PCR primers may be virus specific. T_m was calculated by the nearest-neighbor method (SantaLucia, 1998). To obtain the predicted T_m distribution of 20 bp fragments from the entire BHV-1 genome, 135,382 sequences were obtained from the entire BHV-1 genome (GenBank Accession No. AJ004801) by sliding a 20 bp frame from the 5' end to the 3' end one nucleotide at a time. Total 134 BHV-1 specific PCR primers were obtained from the VirOligo database. If T_ms of primers are not significant factors in PCR experiments, distributions of predicted T_ms from the entire genome and from primers reported in VirOligo should be indistinguishable. The distribution of primer T_m between 30 and 62 °C was not significantly different by χ^2 test between those predicted from the entire BHV-1 genome (20 bp) and primers reported (actual) in the VirOligo database (Figure 5-4; χ^2 test, p=0.083). χ^2 test value was low due to a spike at 58 °C in Figure 5-4. There was another spike at 64 °C. The causes of these spikes are unknown. Figure 5-4 indicates that primers were selected without regard to T_m as long as the T_m was less than 64 °C in BHV amplifications, and oligonucleotides with T_m s of 66 °C or more were less favored for PCR and such primers were used less frequently.

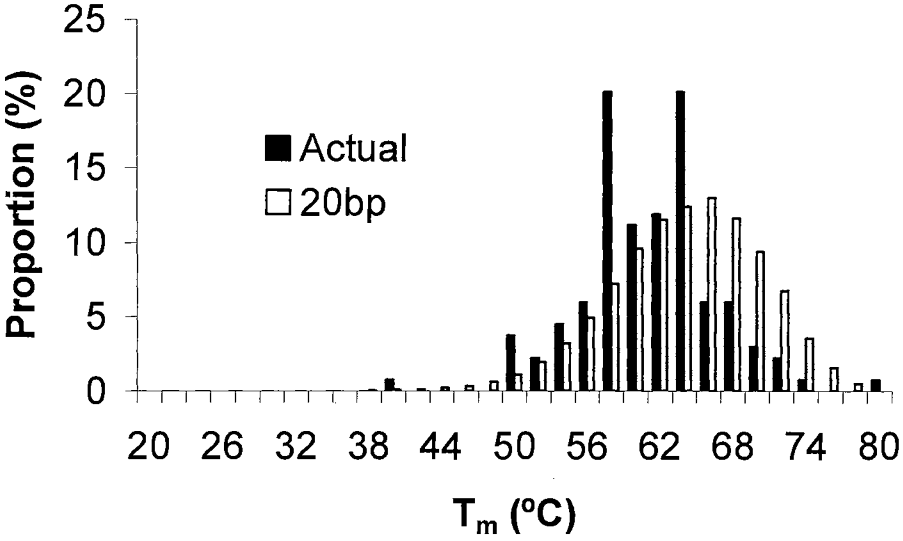


Figure 5–4, Distributions of T_ms for BHV-1 specific primers (Actual, closed bar) and predicted 20 bp sequences from entire BHV-1 genome (20 bp, open bar).

The T_m distribution of FMV specific primers was also tested. While The G+C content of BHV-1 genome is high at 72.4 %, the G+C content of the FMV genome is lower than that of BHV-1 at 53.1 % (GenBank Accession No. AF377945). For FMV, 7,794 sequences fragments of 20 bp size were obtained from the entire FMV genome for a distribution of predicted T_ms, and compared with 120 FMV specific PCR primers reported in the VirOligo database. The T_m distributions of the predicted T_ms from the entire FMV genome and from PCR primers in VirOligo database were similar except for one primer in the VirOligo database whose T_m was 70 °C. Its inclusion significantly changed the χ^2 test result (Figure 5-5; χ^2 test, p=0.484 when the 70 °C oligonucleotide was excluded). The predicted T_ms for FMV were distributed at lower temperatures than for BHV-1, and even at higher T_m, the predicted T_ms from the entire FMV genome and from primers in the VirOligo database were similar.

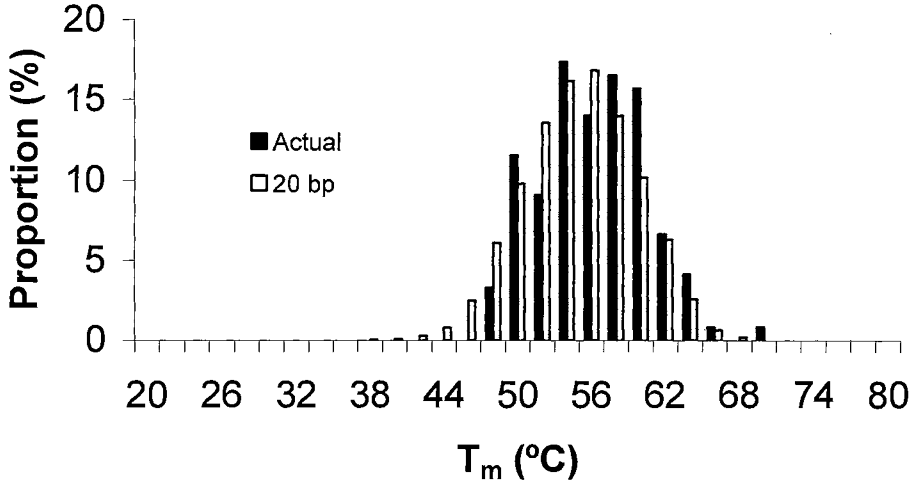


Figure 5–5, Distributions of T_ms for FMV specific primers (Actual, closed bar) and predicted 20 bp sequences from entire FMV genome (20 bp, open bar).

The % (G+C) content of the BVDV genome is 45.5 % (GenBank Accession No. AF220247.1), which is lower than those of FMV or BHV genomes. The T_m distribution of BVDV specific primers was obtained from 225 BVDV specific primers by the same methods used for BHV and FMV. T_ms lower than 50 °C were less frequent in PCR primers than expected in the BVDV genome, and T_ms higher than 60 °C were more frequent in PCR primers than expected (Figure 5-6). Since the high frequencies of high T_m primers in BVDV contrasted with the results obtained from BHV, the compositions of high T_m primers were examined. The mean of primer lengths with 60 °C or more T_m was significantly longer than primer lengths with less than 60 °C T_m. (27.8 nt for 60 °C or more T_m and 20.3 nt for less than 60 °C T_m; t test, p=5.38E-12). Only 11.3 % of primers with 60 °C or more T_m were tailed primers. Means between long primers and 20 bp fragments from the BVDV genome are different since longer primers have higher T_ms. Thus, the comparisons between long primers and 20 bp fragments from the genome are not meaningful and cannot indicate under or over-representations of primers T_ms. Such long primers should not be taken into consideration when representation of primer T_ms are obtained.

The results with BHV-1 primers indicated less frequent high T_m primers (at 66 °C or higher) and the results with BVDV indicated less frequent low T_m primers (at 50 °C or lower). Most FMV specific primers were T_ms between 50 and 64 °C, and primers were not under-represented in that range (95.0 %). Therefore, primers should be selected with T_m between 52 and 64 °C inclusive.

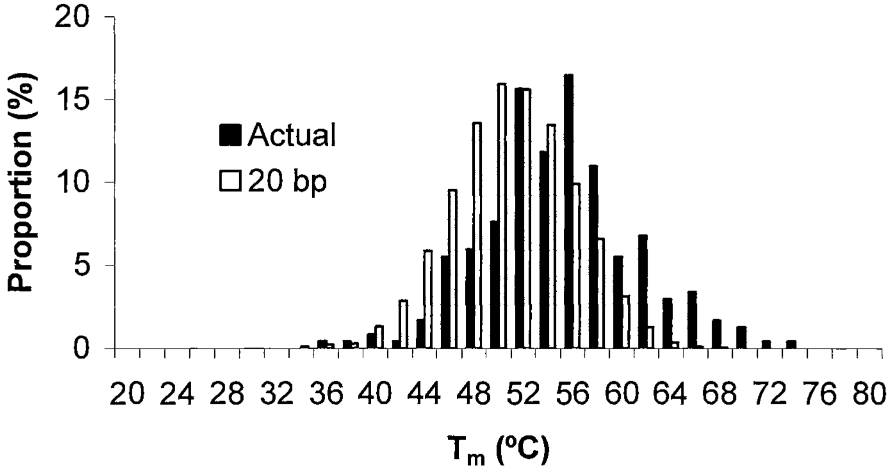


Figure 5–6, Distributions of T_ms for BVDV specific primers (Actual, closed bar) and predicted 20 bp sequences from entire BVDV genome (20 bp, open bar).

The 3' end triplet of primer

Along with the size and the G+C content of the primer, the 3' end triplet of a primer is also one of the considerations mentioned frequently in primer design. If the 3' end triplet does not affect PCR efficiency, all triplets at the 3' end should be distributed equally. Instead, the 3' end triplets were not uniformly distributed (Figure 5-7; χ^2 test, p=1.5E-39). The most popular triplet, AGG (3.27 %), was 7.8 times more frequently used than the least popular triplet, TTA (0.42 %). The mean and standard deviation (s.d.) for triplet distribution was 1.56 % and 0.63%, respectively. Most frequently reported triplets (frequency greater than mean plus s.d.) were of 11 types, AGG (3.27 %), TGG (2.95 %), CTG (2.85 %), TCC (2.76 %), ACC (2.76 %), CAG (2.71 %), AGC (2.57 %), TTC (2.48 %), GTG (2.48 %), CAC (2.38 %), and TGC (2.34 %). All but one of the 11 triplets (TTC) had 2 S and one W. Six of these 11 triplets were WSS sequences. Commercially available software, such as OMIGA (Accelrys, San Diego), searches primers with 3' WSS as the default setting. Consistent with this default, six out of eight possible combinations of WSS were in the top 11 types. However, two of eight WSS triplets, ACG (1.31 %) and TCG (1.08 %), were two to three times less frequently reported in the VirOligo database than other combinations of WSS. WGC was popular and in the top 11 combinations, but the proportion of WCG was less than the average. The under-representation of WCG suggests a high frequency of failure of PCRs using oligonucleotides ending in WCG since commercial software includes selection for WSS in the default setting. Four of the top 11 triplets were SWS triplets. All top 16 reported triplets had a 3' end G or C. Thus, the results suggest the following recommendation: 2 S and 1 W at the 3' end of primers with S at the 3' end. TTS was also frequently reported, and it was in the top 14 combinations, but TTA was the least reported triplet (0.42 %). The least frequently reported triplets (less than mean minus s.d.) were of 6 types, TTA (0.42 %), TAA (0.61 %), CGA (0.65 %), ATT (0.75 %), CGT (0.75 %), and GGG (0.84 %). Two of the lowest six types contained CGW. As mentioned above, WCG was less frequently reported compared to other combinations of WSS. Thus, CG with T or A at the 3' end seems less favored for PCR probably due to a weak binding of the 3' end. Three out of the lowest six types were triplets devoid of G or C, WWW. The most frequently reported types in WWW combinations were AAA (1.45 %) and AAT (1.22 %). Thus, it is advisable not to select low G+C contents at the 3' end of PCR primers to reduce the chances of failure in PCR.

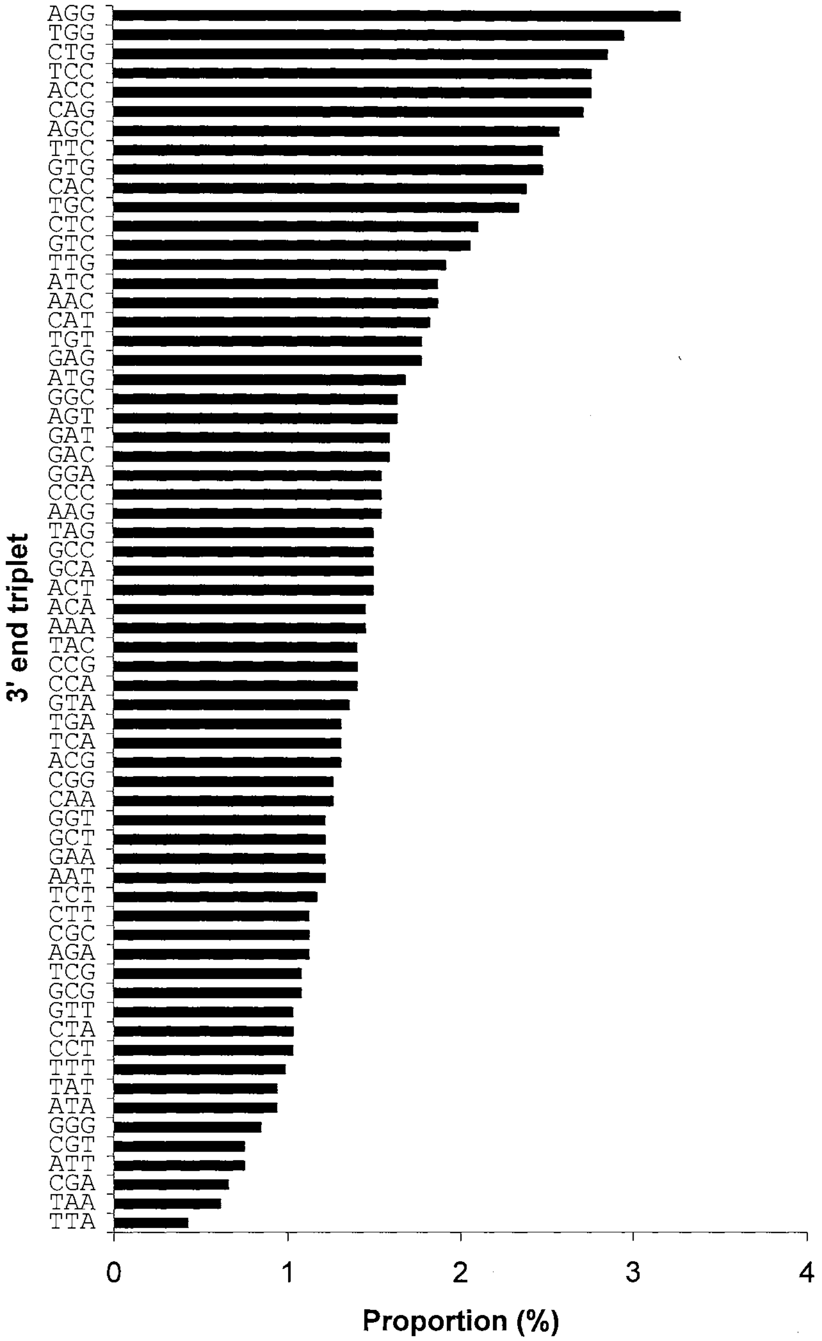


Figure 5–7, Distribution of the 3' end triplet.

When a triplet appears less frequently in a genome sequence, the triplet is generally used less frequently in primer sequences. Thus, the frequencies of the 3' end triplets of primers may be proportional to the frequencies of the triplets in virus genome sequences. Since the primer sequence data consisted of 134 BHV-1 primers, 237 BVDV primers and 121 FMV primers, these viruses were used to test whether genome compositions affected the 3' end triplet frequencies. Thus, triplet frequencies of the BHV-1 genome (GenBank Accession No. AJ004801) were subtracted from frequencies of 3' end triplets in BHV specific primers (Table 5-1). The same processes were performed for BVDV and FMV (GenBank Accession Nos. AF220247 and AF377945, respectively). Negative values indicate under-representation of triplets and thus those triplets were used less frequently in primers than in genome sequences. A value of zero indicates that the triplet's frequencies in a genome sequence and in primers as the 3' end triplet were equal. The ten most and least frequently used 3' end triplets, adjusted for triplet frequencies in BHV, BVDV and FMV genomes, are shown in Figure 5-8. Since the number of primer sequences were few compared to total number of combinations for triplets (64 combinations), adjusted triplet frequencies were rough references to determine the effects of genome compositions. Most of the ten most frequently used 3' end triplets were over-represented in genome sequences (Figure 5-8A), and most of the ten least frequently used 3' end triplets were under-represented in genome sequences (Figure 5-8B). Means of frequencies between top ten and least ten 3' end triplets were significantly different in all three viruses tested (t test, p=0.011 for BHV; p=0.008 for BVDV; p=0.047 for FMV). Thus, the results of 3' end triplet frequencies were not due to genome compositions.

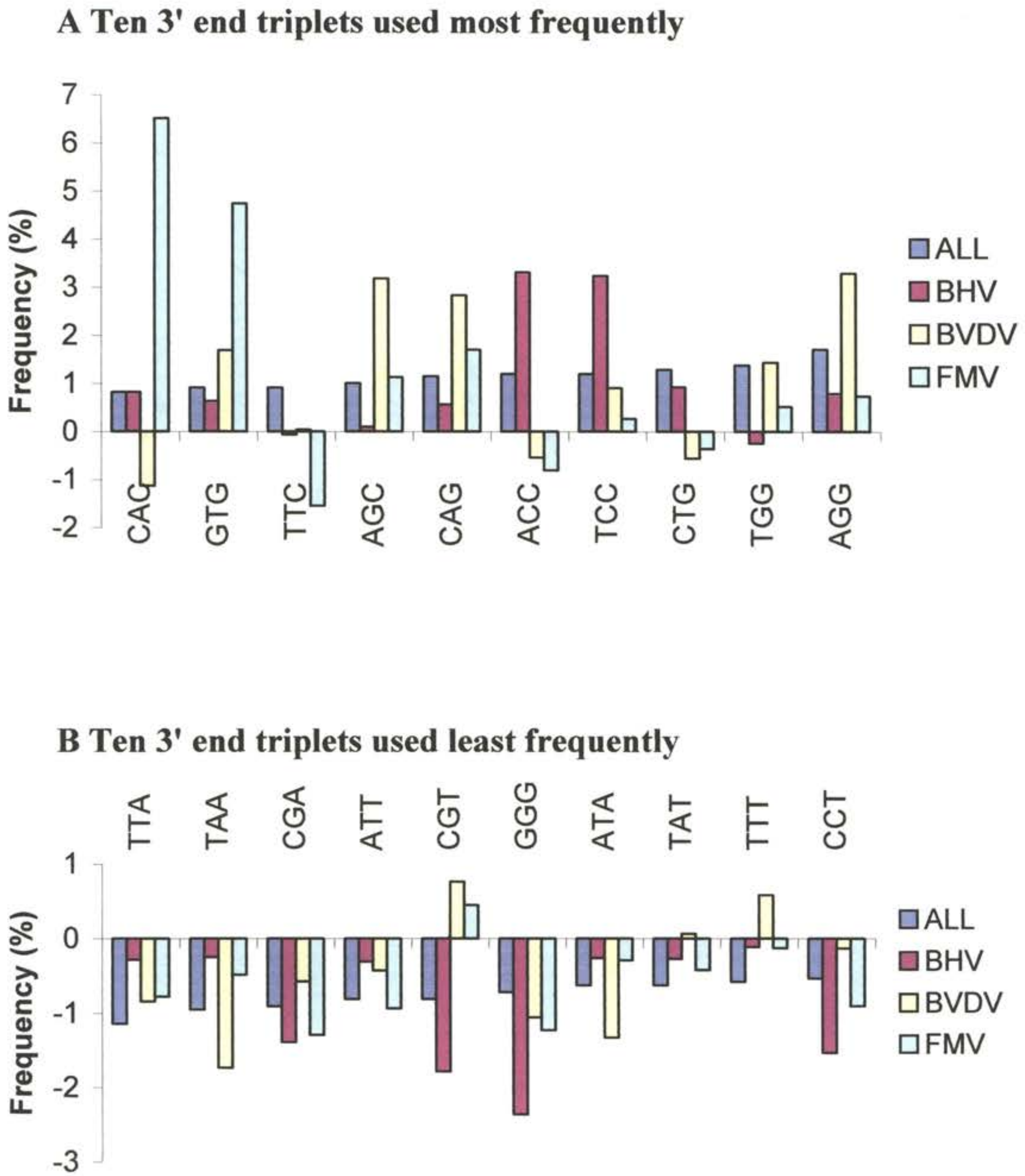
Table 5-1, Comparison among adjusted triplet frequencies of entire primers and BHV; BVDV; FMV specific primers.

	All viruses	BHV primers	BVDV primers	FMV primers
TTA	-1.14	-0.29	-0.84	-0.78
TAA	-0.95	-0.25	-1.73	-0.49
CGA	-0.91	-1.39	-0.57	-1.29
ATT	-0.81	-0.31	-0.43	-0.93
CGT	-0.81	-1.79	0.76	0.45
GGG	-0.72	-2.37	-1.06	-1.23
ATA	-0.63	-0.26	-1.33	-0.29
TAT	-0.63	-0.27	0.06	-0.42
TTT	-0.58	-0.11	0.58	-0.13
CCT	-0.53	-1.53	-0.13	-0.9
CTA	-0.53	-0.43	-0.72	0.11
GTT	-0.53	-0.68	-0.76	-0.75
GCG	-0.49	-2.69	-0.36	0.48
TCG	-0.49	1.01	0.8	0.62
AGA	-0.44	0.04	-1.63	-0.11
CGC	-0.44	-4.04	0.02	1.67
CTT	-0.44	1.26	0	-0.15
TCT	-0.39	-0.7	0.24	0.39
AAT	-0.35	-0.27	-1.91	0.06
GAA	-0.35	0.72	-2.54	0.55
GCT	-0.35	-2.46	0.09	-0.66
GGT	-0.35	1.05	-0.44	-0.02
CAA	-0.3	-0.78	-1.99	-0.93
CGG	-0.3	-3.9	0.63	1.71
ACG	-0.25	3.2	0.08	1.49
TCA	-0.25	0.24	0.68	0.76
TGA	-0.25	-0.6	-1.73	-1.66
GTA	-0.21	1.61	-0.58	-0.88
CCA	-0.16	0.79	2.29	-1.38
CCG	-0.16	-1.44	-0.11	2.43
TAC	-0.16	1.56	0.41	-1.41
AAA	-0.11	0.04	-3.11	-2.2
ACA	-0.11	0.8	-0.75	-0.84
ACT	-0.07	-0.56	-0.56	0.73
GCA	-0.07	0.37	2.47	-0.83

Table 5-1 (Cont.)

	All viruses	BHV primers	BVDV primers	FMV primers
GCC	-0.07	-2.01	-0.01	-0.97
TAG	-0.07	1.09	2.43	-0.36
AAG	-0.02	2.1	-1.5	-2.1
CCC	-0.02	-0.88	2.26	-0.99
GGA	-0.02	2.26	-1.14	1.37
GAC	0.03	-0.22	-0.31	0.9
GAT	0.03	0.92	-0.83	-0.34
AGT	0.08	0.98	-0.35	-1.18
GGC	0.08	-2.19	1.87	-0.04
ATG	0.12	0.08	0.64	-0.62
GAG	0.22	-1.11	0.73	-1.97
TGT	0.22	-0.71	0.34	-1.43
CAT	0.26	0.85	-0.51	0.35
AAC	0.31	0.85	-1.16	-2.14
ATC	0.31	0.21	-0.38	2.32
TTG	0.36	-0.87	-1.23	-0.33
GTC	0.5	0.94	0.8	2.97
CTC	0.54	0.45	0.91	-0.88
TGC	0.78	1.59	0.49	-0.61
CAC	0.82	0.82	-1.13	6.52
GTG	0.92	0.64	1.69	4.74
TTC	0.92	-0.06	0.04	-1.55
AGC	1.01	0.1	3.18	1.14
CAG	1.15	0.56	2.83	1.71
ACC	1.2	3.31	-0.54	-0.81
TCC	1.2	3.23	0.91	0.27
CTG	1.29	0.93	-0.56	-0.36
TGG	1.38	-0.24	1.43	0.52
AGG	1.71	0.79	3.29	0.74

Note: Data is sorted according to 3' end triplet frequencies. The adjusted 3' end triplet frequency values (All viruses) represent 3' end triplet frequencies (%) minus average. Values for BHV, BVDV and FMV specific primers represent 3' end triplet frequencies (%) minus triplet frequencies (%) in the genome sequences. Negative values in BHV, BVDV and FMV indicate under-representation of the triplets and positive values indicate over-representation of the triplets.



A Ten 3' end triplets used most frequently

B Ten 3' end triplets used least frequently

Figure 5–8, Comparison among adjusted triplet frequencies of all primers (ALL) and BHV; BVDV; FMV specific primers. The adjusted 3' end triplet frequency values for all primers represent 3' end triplet frequencies (%) minus average. Values for BHV, BVDV and FMV specific primers represent 3' end triplet frequencies (%) minus triplet frequencies (%) in the genome sequences. Negative values in BHV, BVDV and FMV indicate under-representation of the triplets and positive values indicate over-representation of the triplets.

Product sizes

There are certain limitations on the sizes of product that PCR can produce. As mentioned in the introduction, PCR products less than 100 bp are difficult to differentiate from primer dimers in agarose gel electrophoresis, and large PCR products are generally difficult to amplify. To examine the ranges of PCR product sizes used, the distribution of product sizes was obtained from the VirOligo database (Figure 5-9). The range of product sizes most frequently used was 200 to 299 bp (18.3%). Less than 100 bp was not used frequently, as expected (2.0 %). More than half of PCR products were less than 500 bp (57.8 %), and 82.3 % were less than 1 kb. The primer design program called OMIGA (Accelrys, Inc. San Diego, CA) selects 400 to 800 bp product in its default setting, but 400 to 800 bp size products accounted for only 28.1 % of all primer-pairs in the VirOligo database. Thus, primer design programs may not select the best primer pairs in their default settings.

Large PCR products of 5 kb or more in size accounted for 0.9 % of total PCR products. Some polymerases amplify long PCR products, and TaqPlus Long DNA polymerase (Stratagene, La Jolla, CA) can amplify up to 35 kb. However, the high proportion of larger PCR products was because many primer pairs were generated from sequences that were not in GenBank. When a primer target position in a genome sequence was not mentioned in the articles, NCBI BLAST was used to search for the target position. BLAST results may give wrong positions and thus the wrong PCR product size if the primers were designed from a sequence that is not registered in GenBank.

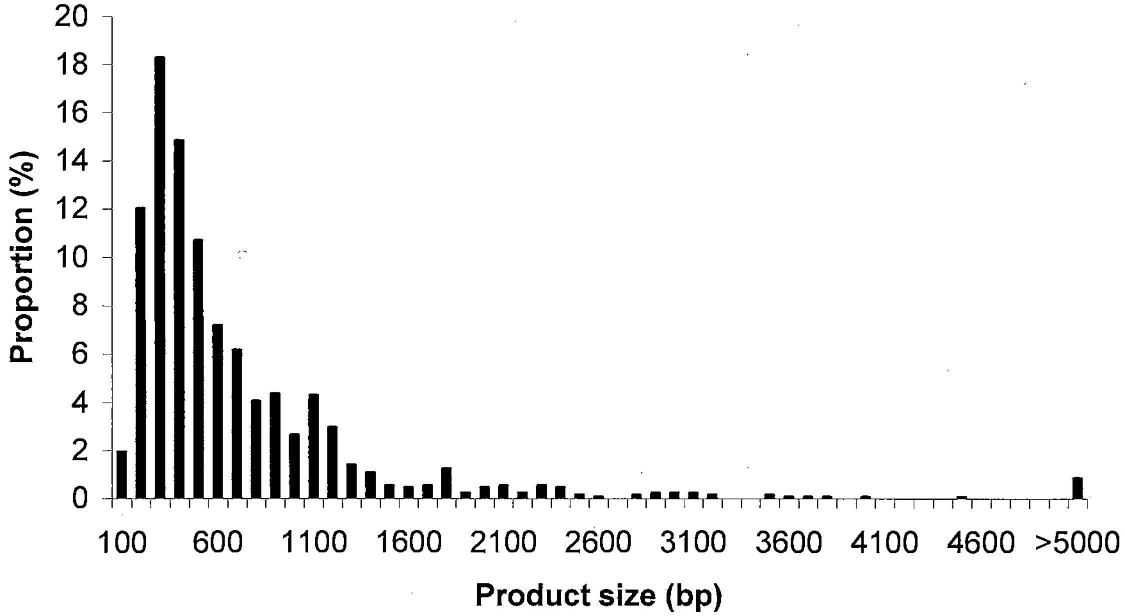


Figure 5–9, Distribution of product size for all primer pairs in the VirOligo database.

Conclusion

Proper primer design is a key for successful PCR. Primer design programs may not select the best primers with their default settings since the properties of primers used in virus detection differed from those recommended. Based on our analysis of primers actually used in PCR of viral genomes, we make the following recommendations: a primer length of 20 nt should be selected if possible, and shorter primers (16 nt or more) may not need to be avoided in the amplification of pure viral templates. Primers should be selected with T_ms between 52 and 64 °C. WSS or SWS at the 3' end is a good choice for PCR primers. However, CG in the 3' end triplet, WCG or CGW, should not be selected for primers. The PCR product size should be 100 to 500 bp or smaller size. By adding the above considerations into the design of primers, the success ratio of PCR design should be improved.

Chapter 6: Optimizing reaction conditions

Results

Range of each variable in PCR

A PCR mixture uses many components, and each component may be a target for optimization. Because PCR experiments have many variables, it is not practical to optimize all PCR components. As mentioned in the introduction, PCR tutorials usually list all possible variables, but they do not identify which variables are more important than others. Selecting the essential variables helps to reduce optimization processes by optimizing only the most critical factors. Determining the range of each variable is also helpful for PCR optimization. Analysis of PCR conditions in the VirOligo database reveals the popular ranges of PCR conditions. The factors that usually are not optimized probably do not need to be optimized since most experiments obtain amplified products without optimization of such factors. In similar fashion, the ranges of the variables obtained from analysis of the VirOligo database are probably the ranges that need to be optimized.

Annealing temperature (T_a) and time

The distributions of T_as and annealing times were acquired from 1070 PCR experiments registered in the VirOligo database to determine the range of T_a. Approximately 30 % of PCR experiments were performed at T_a 55 °C (29.3 %), and almost all PCR experiments used temperatures between 45 °C and 65 °C (91.8 %; Figure 6-1). Many experiments selected T_as from multiples of five, and 65.2 % were 45, 50, 55, 60 or 65 °C as T_as. For annealing time, 44.5 % of experiments used 60 seconds (Figure 6-2). 23.5 % used 30 second and 15.0 % used 120 second annealing time.

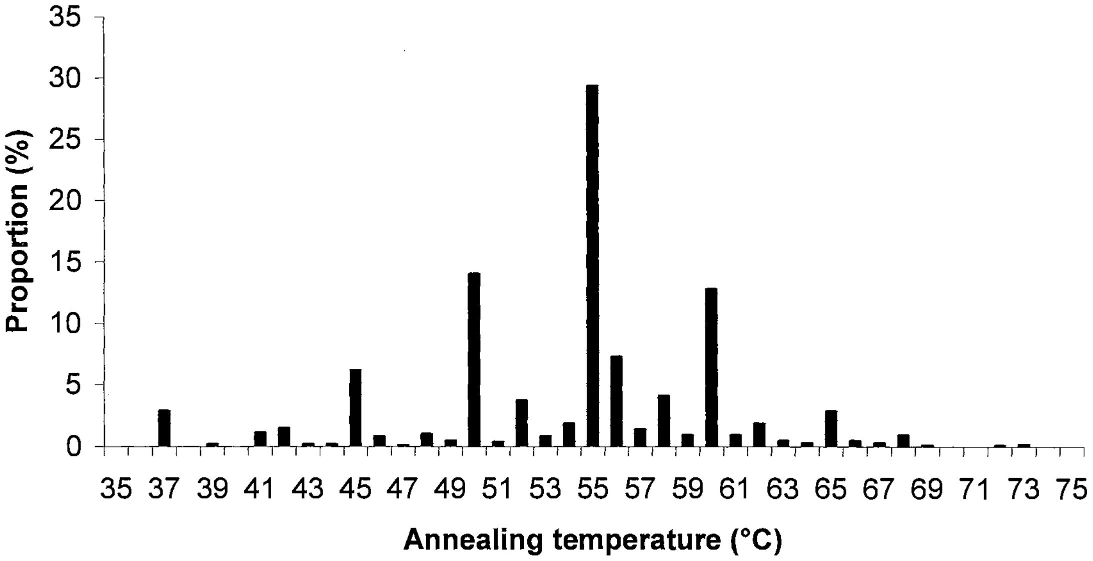


Figure 6–1, Distribution of annealing temperatures for 1070 PCR experiments obtained from the VirOligo database.

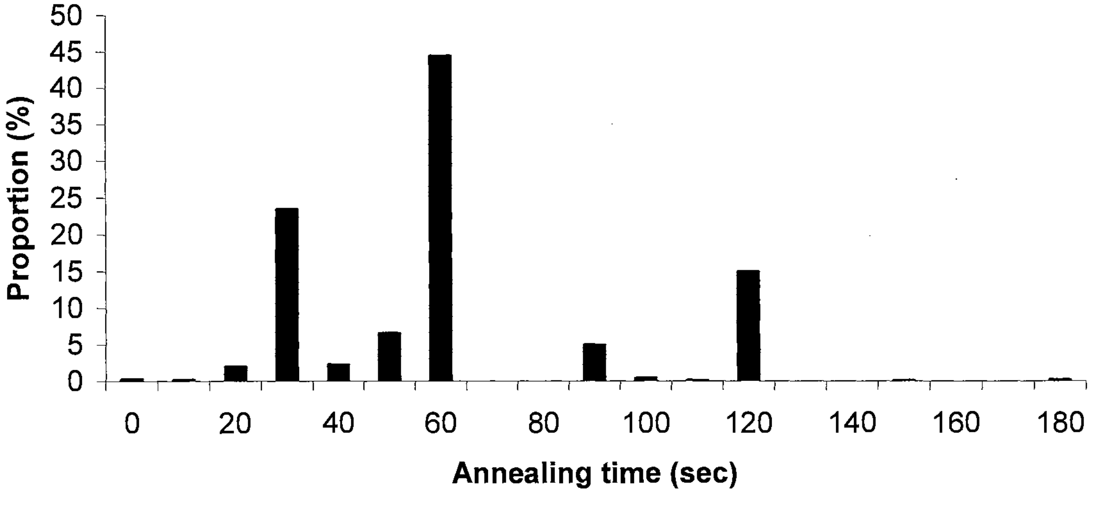


Figure 6–2, Distribution of annealing times for 1070 PCR experiments obtained from the VirOligo database.

Denaturation temperature and time

The distributions of denaturation temperatures and times were acquired from 1071 PCR experiments in the VirOligo database. Most experiments used either 94 °C or 95 °C for denaturation (72.1 % and 20.2 %, respectively; Figure 6-3). For denaturation time, most experiments used either 30 seconds or 60 seconds (23.4 % and 58.3 %, respectively; Figure 6-4).

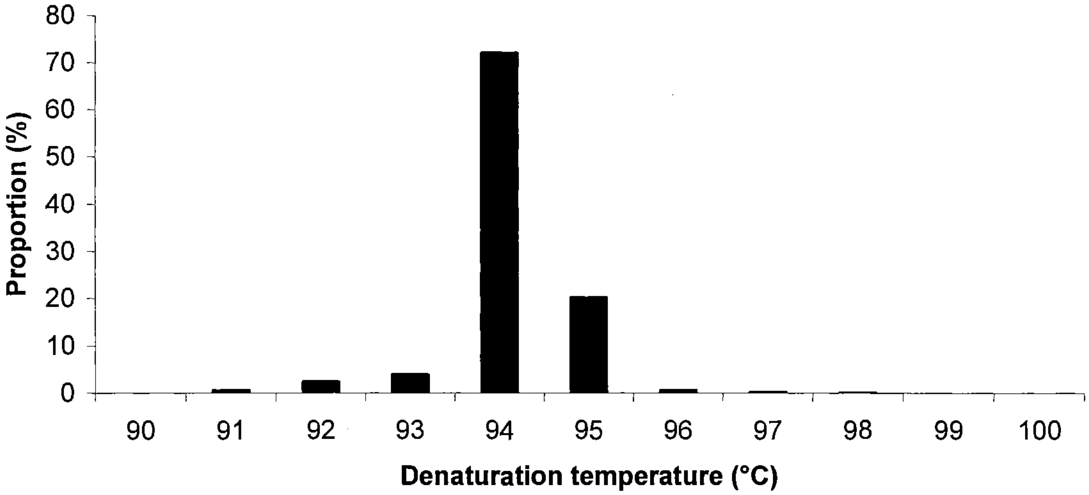


Figure 6–3, Distribution of denaturation temperatures for 1071 PCR experiments obtained from the VirOligo database.

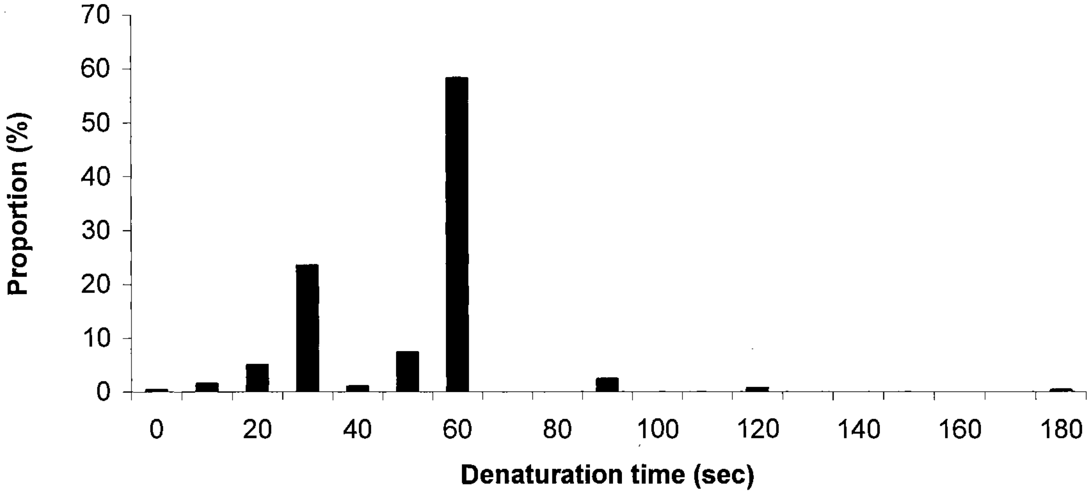


Figure 6–4, Distribution of denaturation times for 1071 PCR experiments obtained from the VirOligo database.

Extension temperature and time

The distributions of extension temperatures and times were acquired from 1062 PCR experiments in the VirOligo database. Most experiments used 72 °C extension temperature (85.5 %; Figure 6-5). Some experiments used lower temperature because annealing and extension were performed at the same temperature. For extension time, 97.5 % of experiments used three minutes or less (Figure 6-6), and 54.0 % used one minute or less. Many experiments used multiples of 30 seconds or 60 seconds as extension times. Extension time 60, 120 and 180 seconds were used in 67.3 % of experiments, and 86.1 % used a time among 30, 60, 90, 120 and 180 seconds.

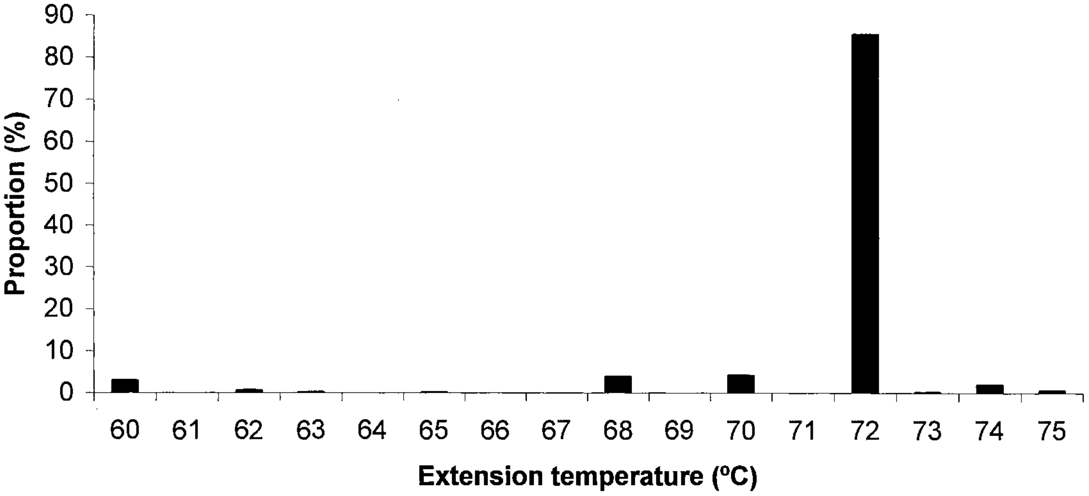


Figure 6–5, Distribution of extension temperatures for 1062 PCR experiments obtained from the VirOligo database.

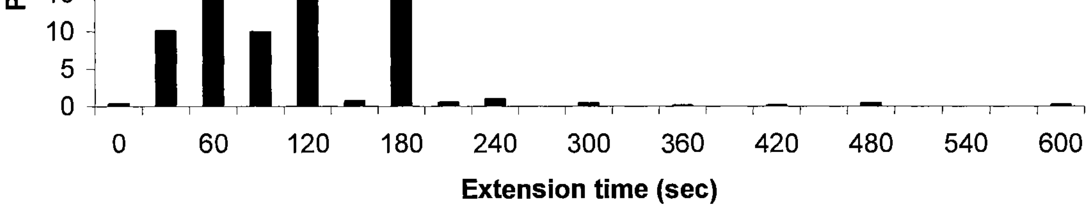
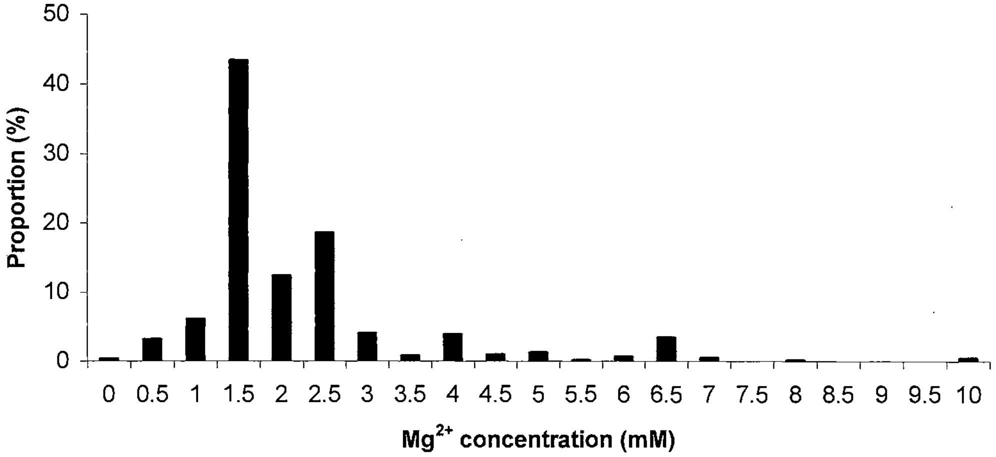


Figure 6–6, Distribution of extension times for 1062 PCR experiments obtained from the VirOligo database.

Magnesium ion (Mg^{2+}) concentration

The distribution of Mg^{2+} concentrations was acquired from 647 PCR experiments in the VirOligo database (Figure 6-7). Most experiments used less than 3.25 mM Mg^{2+} (87.5 %), and 65.2 % of experiments used less than 2.25 mM Mg^{2+}. Mg^{2+} concentration of 2.0 mM was less frequently used than that of 1.5 or 2.5 mM. Many experiments probably used 1.5 mM Mg^{2+} since it is the starting concentration of Mg^{2+} recommended by the Taq DNA polymerase manufacturer. Only when amplifications with 1.5 mM Mg^{2+} concentration were not satisfactory, were optimizations of Mg^{2+} concentration performed. If 2.0 mM Mg^{2+} is optimum for the PCR experiment, such an experiment probably produce PCR product using 1.5 mM Mg^{2+} since the difference between 1.5 mM and 2.0 mM is probably small enough for PCR experiment. Thus, 1.5 mM Mg^{2+} may be satisfactory for experiments with Mg^{2+} optimum at 2.0 mM and optimizations were not performed. If 2.5 mM Mg^{2+} is optimum for the PCR experiment, such an experiment may not produce PCR product using 1.5 mM Mg^{2+}. Thus, 1.5 mM Mg^{2+} may not produce PCR product, and therefore, optimizations were performed for such PCR experiments.



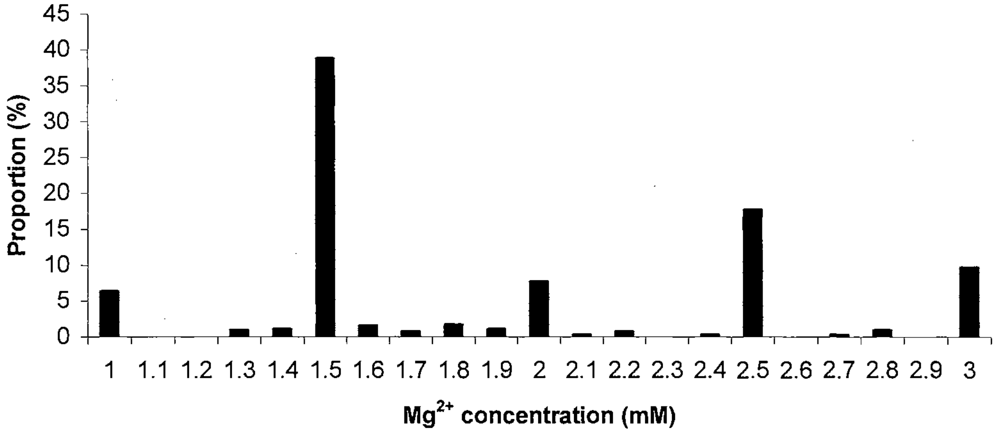


Figure 6–7, Distribution of Mg^{2+} concentrations in 647 PCR experiments obtained from the VirOligo database. The bottom histogram is a detail of the top histogram between 1.0 and 3.0 mM Mg^{2+} concentration.

dNTP concentration

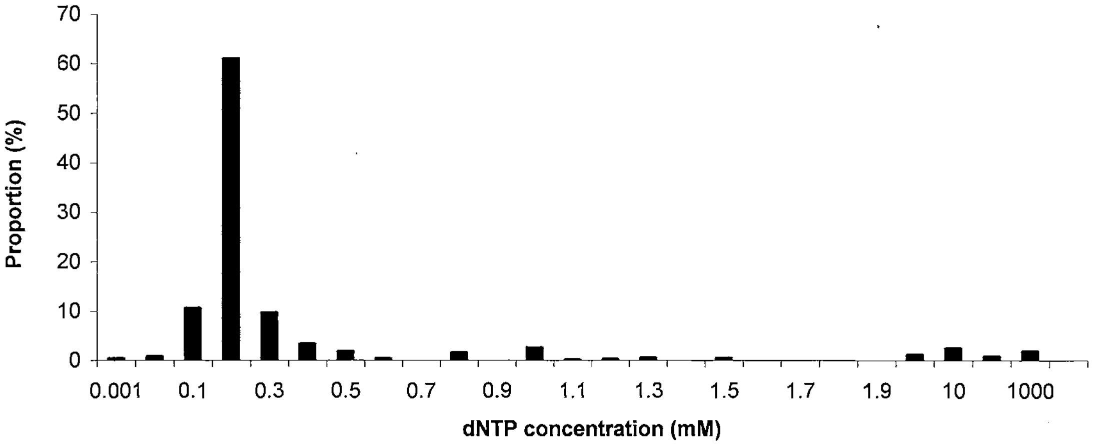
The distribution of dNTP concentrations was acquired from 773 PCR experiments in the VirOligo database (Figure 6-8). Most experiments used dNTPs concentrations between 0.10 and 0.30 mM (75.4 %), and 61.1 % used concentrations between 0.11 and 0.20 mM. The most frequent dNTP concentration was 0.20 mM (59.5 %).

Primer concentration

The distribution of primer concentrations was acquired from 510 PCR experiments in the VirOligo database (Figure 6-9). Primer concentrations were widely distributed in PCR experiments registered in the VirOligo database. While 62.2 % of experiments used primer concentrations between 0.11 and 1.0 μM, 5.9 % used concentrations between 11 and 100 μM. The four most frequent primer concentrations, 0.2; 0.4; 0.5; 1.0 μM, were about equivalent in frequency (18.0 to 22.5 %).

PCR Buffer

PCR buffer usually consists of KCl and Tris-HCl. Most of 487 PCR experiments in the VirOligo database used 10 mM KCl and 50 mM Tris-HCl (76.4% and 88.9 %, respectively). pHs of PCR buffers were acquired from 538 PCR experiments in the VirOligo database. Most experiments used pH 8.3 (63.8 %; Figure 6-10).



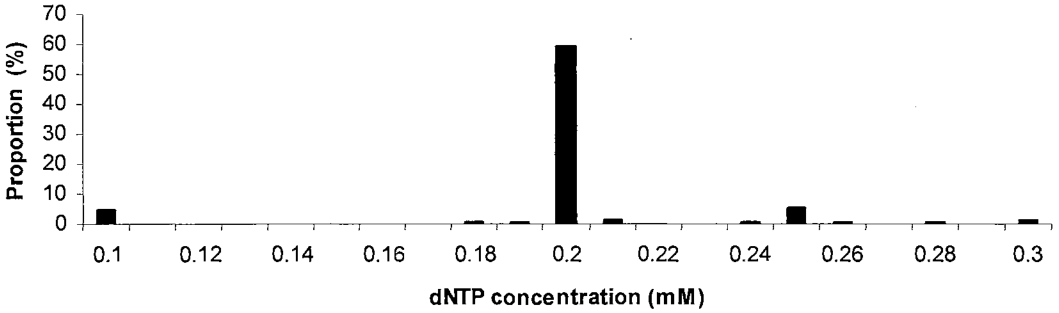


Figure 6–8, Distribution of dNTP concentrations in 773 PCR experiments obtained from the VirOligo database. The bottom histogram is a detail of the top histogram between 0.10 and 0.30 mM dNTP concentration.

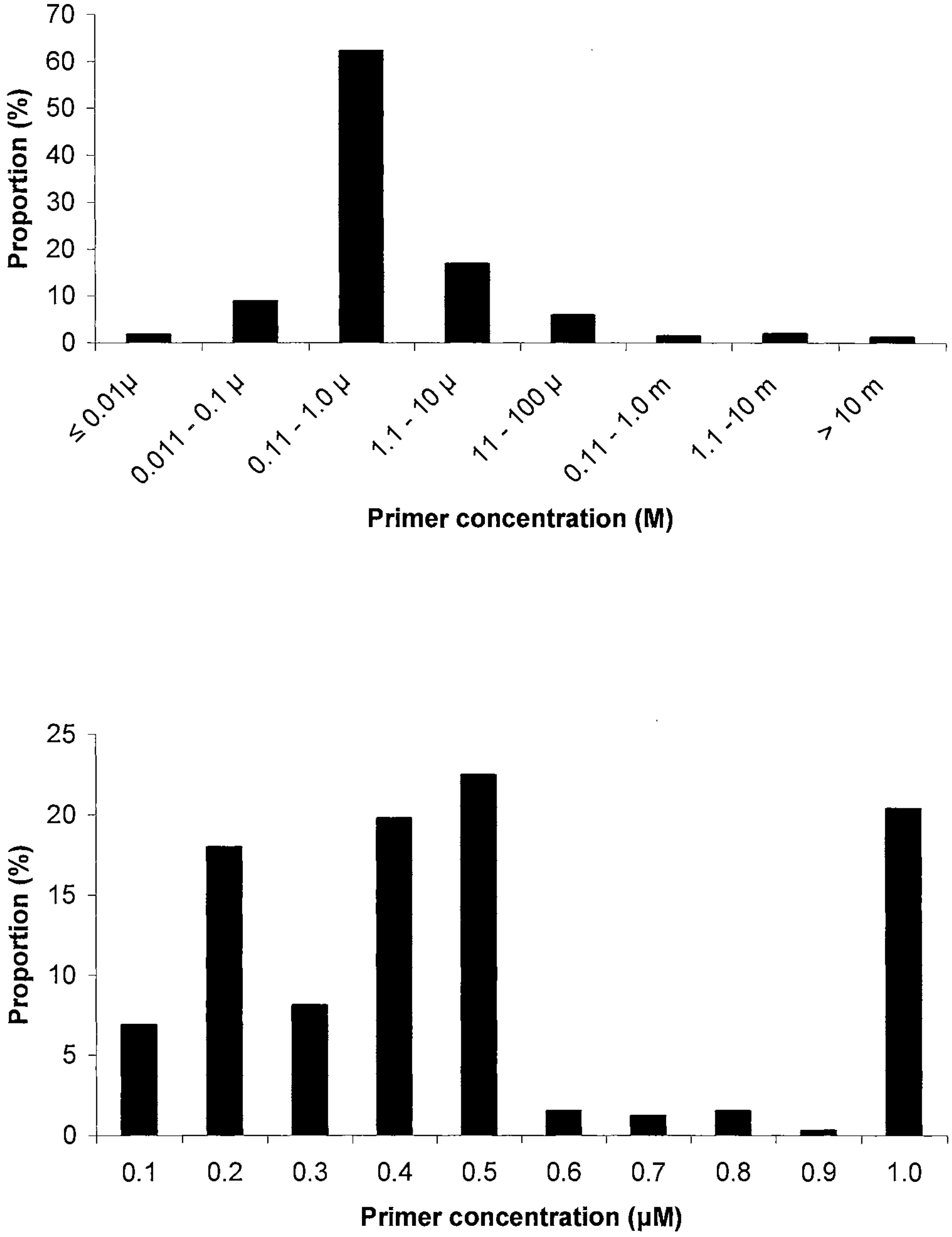


Figure 6–9, Distribution of primer concentrations in 510 PCR experiments obtained from the VirOligo database. The bottom histogram is a detail of the top histogram between 0.1 and 1.0 µM primer concentration.

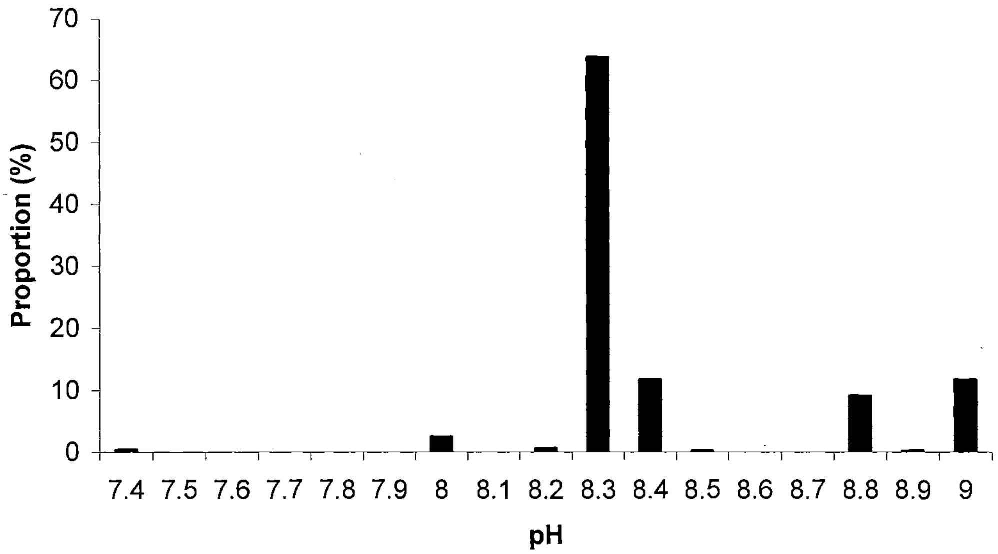


Figure 6–10, Distribution of pHs in 536 PCR experiments obtained from the VirOligo database.

Polymerase

Most of 863 experiments in the VirOligo database used Taq DNA polymerase (72.2 %; Figure 6-11). The ratio becomes 86.9 % when modified Taq polymerases (i.e. AmpliTaq) were included in above ratio. Next most popular polymerase was Tth polymerase (2.1 %). The distribution of Taq DNA polymerase concentrations was acquired from 449 PCR experiments in the VirOligo database (Figure 6-12). Taq DNA polymerase concentration of 0.03 U/µL were used most frequently in the PCR mixtures (39.6 %), and 81.1 % used between 0.01 and 0.05 U/µL.

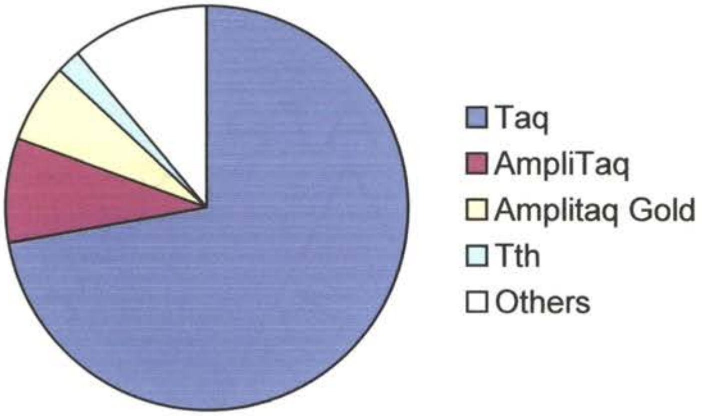


Figure 6–11, Usage of thermostable polymerases in 863 PCR experiments obtained from the VirOligo database.

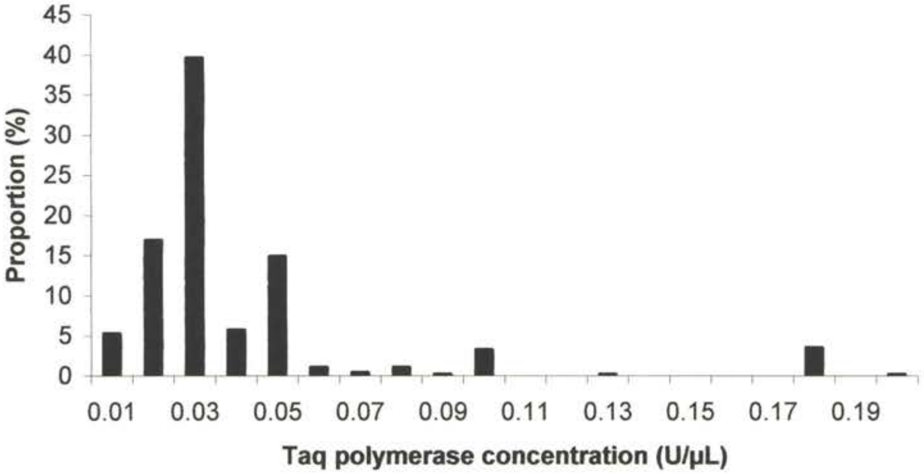


Figure 6–12, Distribution of Taq polymerase concentrations in 449 PCR experiments obtained from the VirOligo database.

Correlations among PCR conditions

While denaturation temperature and time, extension temperature, dNTP concentration, PCR buffer, thermo-stable polymerase and Taq polymerase concentration were the same or similar in most experiments in the VirOligo database, annealing temperature and time, extension time, Mg^{2+} concentration and primer concentration varied considerably. The variables that showed little variation cannot be used for correlation with other conditions. Thus, correlations among the more variable factors, annealing temperature and time; extension time; Mg^{2+} concentration; and primer concentration, were analyzed.

Primer concentration

Using the distribution of the primer concentrations (Figure 6-9) as a guide, the primer concentrations in 510 samples were divided into three ranges, 0.3 µM or less (A); more than 0.3 and less than 1.5 µM (B); 1.5 µM or more (C). For each range of primer concentrations, means of primer concentration and product T_{m}, T_a and Mg^{2+} concentration were obtained. The mean of T_a in the range C was 55.0 °C and it was significantly higher than the means of the ranges A (52.8 °C) and B (53.2 °C) using t tests (Table 6-1). The mean of product T_m in the range A was 88.3 °C and it was also significantly higher than the means of the ranges B (87.6 °C) and C (87.2 °C) using t tests (Table 6-1). Thus, primers in PCR experiments at higher primer concentrations had higher mean T_as and lower mean of product T_m although the differences in T_m and T_a between primer concentration ranges were small. No significant differences were observed by t tests between primer concentrations and the means of primer T_m or Mg^{2+} concentrations (Table 6-1).

Table 6-1, Effects of primer concentration on PCR conditions and results (p value) of statistical tests.

Product T_m

Primer Concentration	≤ 0.3 μM	0.3 - 1.5 μM	≥ 1.5 μM
No. of Samples	149	236	134
Median	88.0	87.8	87.4
Mean	88.3	87.6	87.2

p value	≤0.3 vs 0.3-1.5	0.3-1.5 vs ≥1.5	≤0.3 vs ≥1.5
t test	0.048	0.335	0.006
F test	0.169	0.074	0.669

Primer T_m (less stable primer)

Primer Concentration	≤ 0.3 μM	0.3 - 1.5 μM	≥ 1.5 μM
No. of Samples	166	203	123
Median	54.1	54.2	54.3
Mean	54.3	54.6	54.7

p value	≤0.3 vs 0.3-1.5	0.3-1.5 vs ≥1.5	≤0.3 vs ≥1.5
t test	0.639	0.932	0.623
F test	0.044	0.952	0.083

Primer T_m (more stable primer)

Primer Concentration	≤ 0.3 μM	0.3 - 1.5 μM	≥ 1.5 μM
No. of Samples	158	206	129
Median	55.1	54.7	54.3
Mean	55.2	55.0	54.5

p value	≤0.3 vs 0.3-1.5	0.3-1.5 vs ≥1.5	≤0.3 vs ≥1.5
t test	0.723	0.469	0.300
F test	0.266	0.165	0.738

Table 6-1 (Cont.)

T_a

Primer Concentration	**≤ 0.3 μM**	**0.3 - 1.5 μM**	**≥ 1.5 μM**
No. of Samples	165	223	130
Median	55.0	55.0	55.0
Mean	52.8	53.2	55.0

p value	≤0.3 vs 0.3-1.5	0.3-1.5 vs ≥1.5	≤0.3 vs ≥1.5
t test	0.471	0.005	< 0.001
F test	0.434	0.160	0.516

Mg^{2+} concentration

Primer Concentration	**≤ 0.3 μM**	**0.3 - 1.5 μM**	**≥ 1.5 μM**
No. of Samples	147	196	124
Median	1.50	1.50	1.75
Mean	2.33	2.11	9.37

p value	≤0.3 vs 0.3-1.5	0.3-1.5 vs ≥1.5	≤0.3 vs ≥1.5
t test	0.226	0.273	0.288
F test	0.004	< 0.001	< 0.001

Note: Three ranges, 0.3 μM or less (≤0.3 μM); 0.3 μM or more and 1.5 μM or less (0.3-1.5); 1.5 μM or more (≥1.5 μM), were compared by F test. When p value of F test was bigger than 0.05, "t-Test: Two-Sample Assuming Equal Variances" were used. When p value was 0.05 or less, "t-Test: Two-Sample Assuming Unequal Variances" were used.

Mg^{2+} concentration

Using the distribution of the Mg^{2+} concentration (Figure 6-7) as a guide, the Mg^{2+} concentrations in 647 samples were divided into two ranges, 1.5 mM or less (A) and more than 1.5 mM (B). For both ranges of the Mg^{2+} concentrations, means of primer T_m, product T_m and T_a were obtained. 324 entries were in the range A and 323 entries were in the range B. If Mg^{2+} concentration affects T_m or T_a in PCR experiments, the means should be significantly different between the ranges. However, the results in two groups of Mg^{2+} concentrations were not significantly different in primer T_m, product T_m and T_a (t test; Table 6-2).

Table 6-2, t test results (p value) and comparisons between Mg^{2+} concentrations 1.5 mM or less (≤1.5 mM) and more than 1.5 mM (>1.5 mM) for mean T_a, primer T_m and product T_m.

	Mean (°C)		**p value**
Mg^{2+} concentration	≤ 1.5 mM	> 1.5 mM	t test
T_m of Less Stable Primers	54.9	55.5	0.202
T_m of More Stable Primer	56.7	57.4	0.143
T_m of PCR Products	88.1	87.4	0.055
T_a	54.1	54.4	0.661

Additives

Gelatin and BSA were sometimes added in PCR mixtures to stabilize polymerase. Both were observed in PCR experiments registered in the VirOligo database. Gelatin and BSA were used in 9.7 % and 4.4 % of PCR experiments, respectively. However, the usages of gelatin and BSA were probably more frequent than those percentages since some PCR buffers contain gelatin (e.g. Applied Biosystems, Foster city, CA) or BSA (e.g. Fermentas, Inc., Hanover, MD), and most PCR experiments did not list all components of PCR buffers. Since it could not be determined which PCR experiments truly lacked gelatin or BSA, correlations among PCR conditions with and without gelatin or BSA were not analyzed.

DMSO was the only additive found in the VirOligo database not usually contained in PCR buffers or polymerase storage buffers. DMSO were used in 34 PCR experiments. Of these 27 experiments were for amplification of herpesvirus, and 11 PCRs were for BHV-1. DMSO is used for amplification of templates with high % (G+C) contents. Since the VirOligo database contained 48 PCR experiments amplifing BHV-1 with a clear description of the buffer used, the 11 PCR experiments with DMSO represented only 23 % of total BHV-1 amplification experiments. The differences in PCR conditions with and without DMSO were tested. There were no significant differences found among product %(G+C), product size, product $T_{m,}$ primer $T_{m,}$ Mg^{2+} concentration, primer concentration, denaturation time, extension temperature, extension time, T_a or annealing time (Table 6-3). Means of denaturation temperature were significantly different between with and without DMSO (t test, p = 0.0048), but the difference between the means was only 0.6 °C.

Table 6-3, Effects of DMSO on PCR conditions and results (p value) of statistical tests.

Mean	Product %(G+C)	Product size (bp)	Product T_m	More stable primer T_m	Less stable primer T_m	Mg^{2+} conc. (mM)	Primer conc. (microM)
w/ DMSO	68.0	378.0	96.3	61.2	54.4	2.2	1.0
w/o DMSO	71.1	455.4	97.1	60.6	60.5	1.6	0.7
p value							
t test	0.077	0.351	0.216	0.704	0.245	0.054	N/A
F test	0.084	< 0.001	0.080	0.413	< 0.001	0.142	N/A

Mean	Denaturation temp	Denaturation time	Extension temp	Extension time	T_a	Annealing time
w/ DMSO	94.3	49.1	72.2	60.0	57.4	60.0
w/o DMSO	94.9	57.2	72.1	85.9	59.1	62.8
p value						
t test	0.005	0.363	0.690	N/A	0.357	N/A
F test	0.166	< 0.001	0.468	N/A	0.540	N/A

Note: Means of PCR conditions were compared between with and without DMSO. When p value of F test was bigger than 0.05, "t-Test: Two-Sample Assuming Equal Variances" were used. When p value was 0.05 or less, "t-Test: Two-Sample Assuming Unequal Variances" were used.

Extension time

During the extension step in a PCR cycle, PCR products are elongated. Longer extension times allow polymerase to produce larger PCR products. Since 86.1 % of experiments used 30, 60, 90, 120 or 180 seconds as mentioned above, means of product size and product T_m were obtained to compare the effect of extension time (Table 6-4). Both 30 and 60 second extensions had similar means of product sizes (t test, $p = 0.45$). Extension times among 90, 120, and 180 second had similar means of product sizes, too (t tests, $p = 0.25$ for between 90 and 120 seconds; $p = 0.85$ for between 120 and 180 seconds). The means of product sizes were 0.4 kb for 30 and 60 seconds, and 0.7 to 0.8 kb for 90, 120 and 180 seconds. Thus, extension time was either 60 seconds or less or more than 60 seconds based on product size. No significant differences were found in the means of product T_m (t tests; Table 6-4).

Table 6-4, Effects of extension time on PCR conditions and results (p value) of statistical tests. Means of product sizes and T_m were obtained. Product sizes and T_m compared by F test. When p value of F test was bigger than 0.05, "t-Test: Two-Sample Assuming Equal Variances" were used. When p value was 0.05 or less, "t-Test: Two-Sample Assuming Unequal Variances" were used.

Extension time (sec)	**Product size (bp)**	**Product T_m (°C)**	**No. of samples**
30	422.0	87.7	96
60	459.2	88.0	364
90	831.7	87.8	103
120	727.1	87.8	172
180	742.9	87.5	173

Product size

	30 sec vs. 60 sec	**60 sec vs. 90 sec**	**90 sec vs. 120 sec**	**120 sec vs. 180 sec**
t test	0.454	< 0.001	0.250	0.850
F test	0.455	< 0.001	< 0.001	0.834

Product T_m

	30 sec vs. 60 sec	**60 sec vs. 90 sec**	**90 sec vs. 120 sec**	**120 sec vs. 180 sec**
t test	0.404	0.654	0.945	0.288
F test	< 0.001	0.051	0.496	0.021

Annealing time

Annealing time was 2 minutes or less in most experiments. Annealing time may correlate with annealing temperature since both affect annealing between template and PCR primer. Thus, correlation coefficient between annealing time and T_a was determined. Additionally, correlation coefficients between annealing time and product %(G+C), product size, product T_m or primer T_m were obtained. T_a and annealing time were not significantly correlated, but they were the most correlated of all parameters checked (correlation coefficient, -0.18; Table 6-5). The correlation between T_a and annealing time was examined further. The means of T_a were obtained for four categories of annealing time, 30 seconds or less; 60 seconds; 90 seconds; 120 seconds. Based on t test results, the mean of T_a in the category of 120 seconds was significantly lower than means in the other three categories by 3 to 5 °C (Table 6-6). The differences in T_a were small compared to the T_a distribution of 37 to 65 °C.

Table 6-5, Correlation coefficients between annealing time and other PCR conditions.

	Product %(G+C)	Product size	Product T_m	More stable primer T_m	Less stable primer T_m	T_a	Annealing time
Annealing time	-0.084	0.071	-0.032	-0.025	-0.066	-0.176	1

Table 6-6, Effects of annealing time on PCR conditions and results (p value) of statistical tests.

Annealing time (sec)	Mean of T_a	No . of samples
30 or less	55.0	277
60	54.1	479
90	53.0	53
120	50.4	160

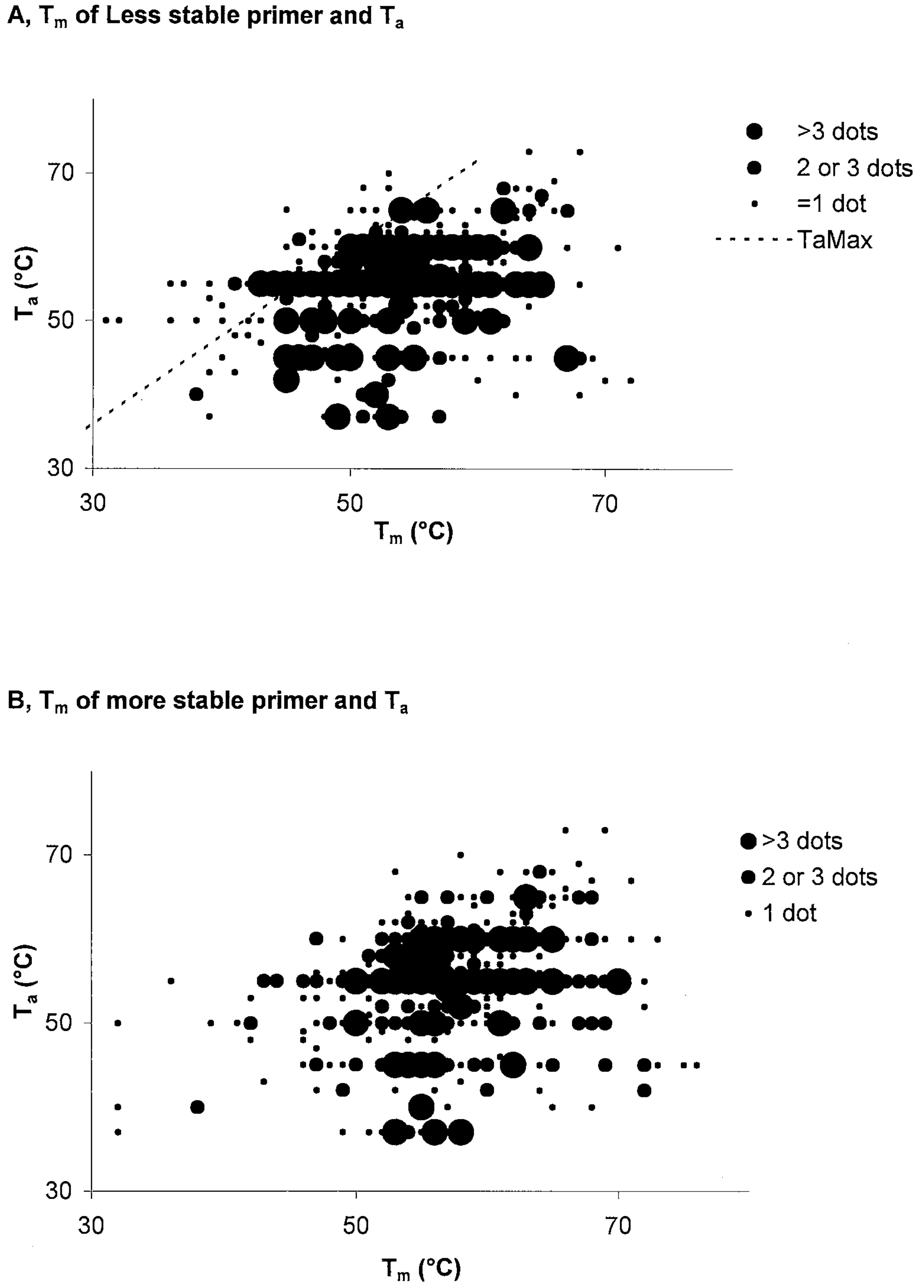
	30 sec or less vs. 60 sec	60 sec vs. 90 sec	90 sec vs. 120 sec
t test	0.059	0.302	0.016
F test	< 0.001	0.198	0.935

Note: Means of T_a were obtained and compared. When p value of F test was bigger than 0.05, "t-Test: Two-Sample Assuming Equal Variances" were used. When p value was 0.05 or less, "t-Test: Two-Sample Assuming Unequal Variances" were used.

Annealing temperature (T_a)

An estimation of T_a has been attempted (Newton and Graham, 1997; Sambrook and Russell, 2001; Dieffenbach, et al., 1993; Kidd and Ruano, 1995; Wetmur, 1991). Those attempts use only the T_m of primers or T_m of primers and product to estimate T_a. Thus, the correlation between T_a and the T_m of products or primers was tested for 647 PCR experiments registered in the VirOligo database. T_a did not correlate well with product T_m (correlation coefficient, 0.15). The correlations between T_a and primer T_m were more than between T_a and product T_m (correlation coefficients were 0.26 for less stable primers and 0.19 for more stable primers), but these were still not significantly correlated.

A certain range of T_a can be used to amplify the PCR product. Usage of a low T_a increases chances of artifacts and non-specific bindings of primers. One of the causes of artifacts and non-specific bindings is often contamination in templates. Thus, an estimation of the minimum T_a ($T_a$min) can be affected by template purity. A usage of low T_a also lengthens the total PCR time due to ramping times in a thermal cycler, and longer PCR time lowers polymerase activity. Thus, $T_a$min is affected by the performance of a thermal cycler and the polymerase. While it is probably difficult to estimate the $T_a$min due to involvement of many factors, the maximum T_a ($T_a$max) may possibly be estimated since annealing and melting are closely related and T_m can be calculated. An estimation of $T_a$max is useful since artifacts can be avoided by using the highest T_a still compatible with amplification. Thus, primer T_m and T_a were compared in scatter graphs. Dots in the scatter graphs formed roughly quadrilateral shapes (Figure 6-13A for T_m of less stable primer and Figure 6-13B for T_m of more stable primer). Since both distributions had similar shapes and the T_m of the less stable primer had a higher correlation coefficient than the T_m of the more stable primer, the graph for T_m of the less stable primer and T_a was used to obtain $T_a$max. T_ms rounded to the nearest five °C, and T_as at 95 percentile were calculated for primer T_ms, which have at least ten T_as. Then, a regression line was obtained from those T_as at 95 percentile and primer T_m. The line deduced was $T_a$max (°C)=1.2×[T_m of less stable primer] when T_m was between 35 and 55 °C. Primers with T_ms of 55 °C or more reached maximum $T_a$max of 65 °C since higher T_a made it difficult for primers to hybridize to template due to active melting of hybrids.



A, T_m of Less stable primer and T_a

B, T_m of more stable primer and T_a

Figure 6–13, Scatter diagrams of T_a reported in the VirOligo database against less stable primers (A), and more stable primers (B). An oblique line shown in Fig. A indicates maximum T_a ($T_a$max) against T_m of less stable primer. Three sizes of spots in the diagrams are used according to the number of dots at the given X-Y coordinates.

Confirmation of the relationship between T_m of primer and T_a

An equation to estimate $T_a$max was suggested above. To test the suggested equation in actual experiments, ten primer pairs (Table 6-7) that are specific to equine influenza virus (GenBank Accession number: AF197241) were designed by a PCR primer design program, OMIGA (Accelrys, San Diego). Twelve T_as between 42 and 66 °C for primer pairs 1-7, or 48 and 72 °C for primer pairs 8-10 were tested in PCR experiments. Then, the ranges of T_a that generated PCR products were identified by gel electrophoresis (Figure 6-14). Actual $T_a$maxs for all the ten primer pairs were slightly higher than estimated $T_a$maxs by 0.4 to 4.6 °C (Table 6-8 and Figure 6-15). This was probably because the equation was deduced from analysis of published PCR conditions, which may not all have been using $T_a$max. Thus, it is reasonable that all actual $T_a$maxs were higher than estimated $T_a$maxs. As seen in the gel electrophoresis results of PCR experiments (Figure 6-14), most PCR experiments performed at the actual $T_a$max produced less PCR products than experiments with lower T_a. Therefore, it is best to select T_a without exceeding the estimated $T_a$max.

Table 6-7, PCR primers.

Primer Pair	Sequence	Tm (°C)	Genome Position
1	CTACAGTCAAAACCCAACC	49.0	71-539
	TGTTAGCCAATTCAGTCG	47.5	
2	GGACCTCTTCATAGAAAGAAGC	52.5	236-753
	ATTGACCCCTAACCCACG	51.2	
3	TACAGTCAAAACCCAACC	47.0	72-539
	TGTTAGCCAATTCAGTCG	47.5	
4	TGCTACCCATATGACATCC	49.0	363-764
	TATCCTGCCTGATTGACC	48.2	
5	ATTGCTACCCATATGACATCC	51.1	361-766
	CTTATCCTGCCTGATTGACC	51.3	
6	TGCTACCCATATGACATCC	49.0	363-765
	TTATCCTGCCTGATTGACC	49.5	
7	ATTGCTACCCATATGACATCC	51.1	361-765
	TTATCCTGCCTGATTGACC	49.5	
8	ACCCCCACTGTGATGTCTTCC	57.5	292-768
	TGCTTATCCTGCCTGATTGACC	56.5	
9	ACCCCCACTGTGATGTCTTCC	57.5	292-760
	CTGCCTGATTGACCCCTAACC	56.4	
10	CCCCACTGTGATGTCTTCC	52.9	294-766
	CTTATCCTGCCTGATTGACC	51.3	

Note: Total fourteen primers were in ten primer-pairs. Some primers were used more than once. Genome positions were based on GenBank® accession no. AF197241.1.

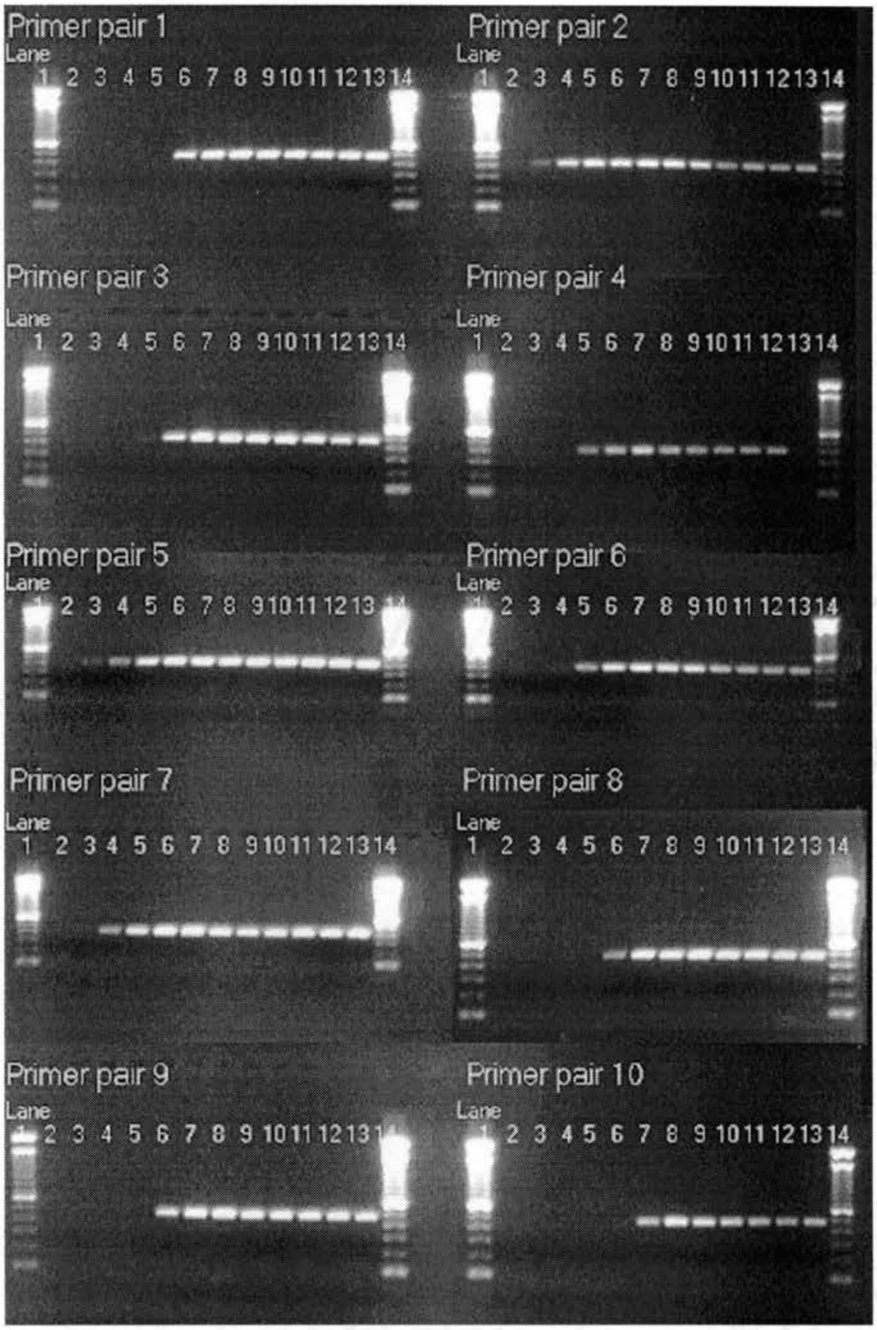


Figure 6–14, PCR results with ten primer pairs for amplification of an equine influenza virus. Twelve annealing temperatures were tested for each primer pair. Lane assignments for primer pair 1 to 7 were lanes 1 and 14, 100 bp ladder; lane 2, 66.0 °C; lane 3, 65.4 °C; lane 4, 64.0 °C; lane 5, 62.1 °C; lane 6, 59.4 °C; lane 7, 56.0 °C; lane 8, 52.0 °C; lane 9, 48.5 °C; lane 10, 45.8 °C; lane 11, 43.8 °C; lane 12, 42.6 °C; lane 13, 42.0 °C. Lane assignments for primer pair 8 to 10 were lanes 1 and 14, 100 bp ladder; lane 2, 72.0 °C; lane 3, 71.5 °C; lane 4, 70.0 °C; lane 5, 68.1 °C; lane 6, 65.4 °C; lane 7, 62.0 °C; lane 8, 58.0 °C; lane 9, 54.6 °C; lane 10, 41.8 °C; lane 11, 49.8 °C; lane 12, 48.6 °C; lane 13, 48.0 °C.

Table 6-8, Actual $T_a$maxs and estimated $T_a$maxs for ten primer pairs.

Primer-pair No.	Actual $T_a$max (°C)	T_m of less stable primer (°C)	T_m of more stable primer (°C)	Estimated $T_a$max (°C)
1	59.4	49.0	47.5	57.0
2	64.0	52.5	51.2	61.4
3	59.4	47.0	47.5	57.0
4	62.1	49.0	48.2	57.8
5	64.0	51.1	51.3	61.6
6	62.1	49.0	49.5	59.4
7	64.0	51.1	49.5	59.4
8	65.4	57.5	56.5	65.0
9	65.4	57.5	56.4	65.0
10	62.0	52.9	51.3	61.6

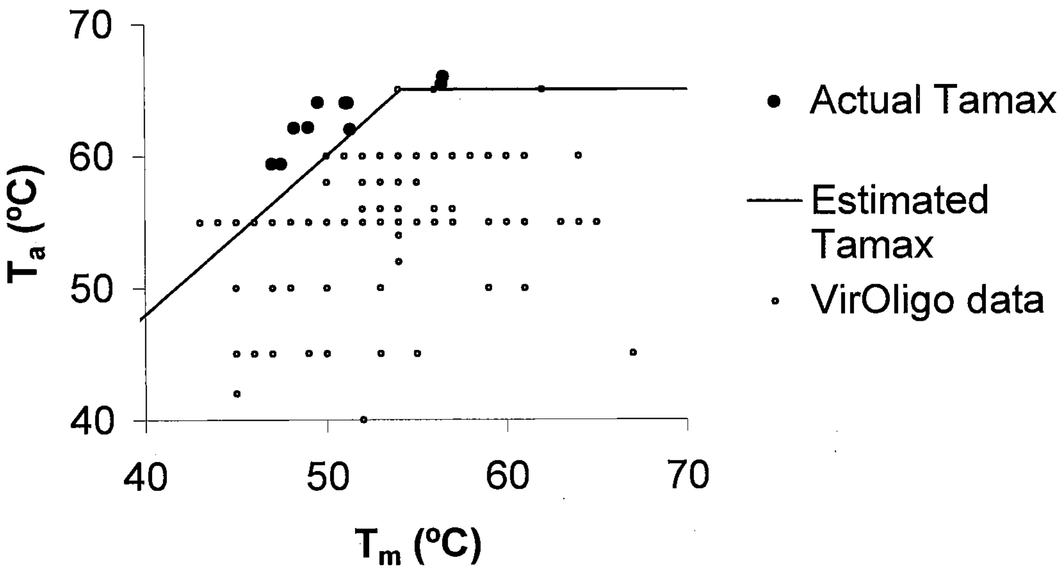


Figure 6–15, Actual $T_a$maxs obtained for ten primer-pairs and estimated $T_a$maxs against T_m of less stable primers. Those are overlaid on Fig. 6-13A with more than three dots in X-Y coordinates (VirOligo data).

Discussion

Since many variables exist and each variable interacts with other factors, the possible combinations of PCR conditions may exceed hundreds. However, only several factors were optimized in most PCR experiments.

One of the notable facts was high usage of Taq polymerase (86.9 %). The second most popular polymerase in the VirOligo database was Tth, but only 2.1 % of experiments used Tth, which was first used in 1991 (Myers and Gelfand, 1991). One quarter (24.9 %) of the PCR experiments used for the analysis were published after January 2000. Thus, high usages of Taq polymerase were not due to the age of the analyzed data, and Taq polymerase is simply still very popular even though so many polymerases are available on the market.

Denaturation temperature and time, extension temperature, dNTP concentration, polymerase and PCR buffer, were not adjusted in most experiments, as expected. The purpose of the denaturation in a PCR cycle is complete separation of double strands. Higher temperature and longer time are better for denaturation. However, the temperature and time are limited by half-life of the polymerase. Since most experiments used Taq polymerase, it is not surprising that most experiments used similar denaturation temperatures and times. Extension temperature depends on the temperature at which the polymerase has the best activity. PCR buffers are provided by polymerase manufactures. These buffers are mostly the same, and according to Promega, the buffer is optimized for 0.2 mM each dNTP. Thus, it is expected that most experiments did not adjust PCR buffer and dNTP concentration.

Annealing temperature (T_a) and time, extension time, Mg^{2+} concentration and primer concentration were variable factors. Although optimum Mg^{2+} concentration could not be estimated, 1.5 mM and 2.5 mM Mg^{2+} concentrations were found to be exclusively used in PCR experiments. Annealing and extension times were weakly correlated with T_a and product size, respectively. However, those distributions were significantly different only at one dividing point, between 90 seconds or less and 120 seconds or more for annealing time; between 60 seconds or less and 90 seconds or more for extension time. For example, no significant difference in T_a was found between 30 second annealing and 60 second annealing or between 90 second annealing and 120 second annealing. Since no continuous differences were found statistically among the distributions in 30, 60, 90 and 120 second annealing or extension, researchers may choose longer annealing or extension times when T_a is lower or the PCR product is larger even though longer annealing or extension times are actually not necessary for the PCR experiments. Extension rate for Taq polymerase is 2 to 4 kb per minute (Qiagen Inc., Valencia, CA). Thus, one-minute extension may be sufficient for most experiments.

Primer concentration was correlated with T_a and product T_m. When primer concentration was high, T_a was high and product T_m was low. Although mechanisms that could account for correlation between primer concentration and product T_m are not known, the correlation between primer concentration and T_a is explainable. When higher T_a is used in the PCR experiment, a lower percentage of primers anneals to templates. Thus, a higher concentration of primers in PCR mixture is required when a higher T_a is used. However, the effects of primer concentration are not worth considering during planning PCR experiments since the differences in means of T_a and product T_m were no more than 2.7 °C and 1.3 °C, respectively, when various primer concentrations were used.

The $T_a$max estimation equations were deduced from the PCR experimental data in the VirOligo database and confirmed in actual PCR experiments with ten primer pairs. Since higher T_a can reduce non-specific products (Sambrook and Russell, 2001), the $T_a$max estimation equations are useful, and T_a should be selected within estimated $T_a$max.

Many variables exist in PCR experiments, but factors that need to be optimized in most PCR experiments were only Mg^{2+} concentration and T_a. Estimation methods for the best T_a was suggested above and either 1.5 mM or 2.5 mM $MgCl_2$ was sufficient for most PCR experiments.

Recent developments of microarray technology and SNPs (Lashkari, et al., 1997; Buetow, et al., 2001) sometimes require preparations of thousands of PCR experiments. When a large number of PCR experiments are conducted, reductions of optimization steps are critical to reduce cost and labor as well as consumption of PCR templates. Therefore, the suggestions made from the analysis results will help such large scale PCR experiments.

Chapter 7: Conclusion

The aim of this study was preparing the fundamentals for the implementation of the universal virus detection system. The proposed methods of the universal virus detection system were ViSA card and ViSH chip methods. Those methods were first tested in my M.S. research. Then, this study was conducted for further development of ViSA card and ViSH chip methods. Thus, this study is a second step towards the universal detection system.

Construction of the VirOligo database was useful for the ViSA card and ViSH chip methods in three ways. First, VirOligo can help provide virus specific oligonucleotide sequences for use in both methods. Second, analysis of oligonucleotides in VirOligo can provide methods to design oligonucleotides for use in the methods. Third, analysis of reaction conditions of PCR experiments can help reduce the optimization processes of the ViSA card method.

Analysis of oligonucleotides revealed that oligonucleotide design programs were not always selecting the best oligonucleotides. Thus, VirOligo was an appropriate method to collect the oligonucleotides from journal articles for ViSA card and ViSH chip though the oligonucleotide sequences from VirOligo showed unexpected behaviors in influenza detection. Requirements for hybridization probe design and primer design may be different, and ViSH chip needs hybridization probes instead of PCR primers. VirOligo is a database of virus specific oligonucleotides, but most oligonucleotides listed were PCR primers. Thus, further expansions of VirOligo database collections may help ViSH chip. The expansion also helps analysis of hybridization probe requirements.

Recommendations for reaction conditions were proposed. Since ViSA card uses a number of primers, it is necessary to reduce optimization processes due to cost of ViSA cards and complex interactions of factors. Since ViSA card is a PCR based detection method, the recommendations will help optimization processes faster. Primers for ViSA card can be obtained from proposed primer design methods and the published primers in the VirOligo database.

The universal virus detection systems, ViSA and ViSH, are under development. However, I believe that this study brings us one step closer to the realization of the system. The universal virus detection system is necessary to improve our public health by diagnosing virus infections from the patients with chronic disease; rapid detection of virus outbreaks under biological terrorism attacks; and testing for food or water contamination. Recent developments of microarray technologies also help giving life to the universal virus detection system.

REFERENCES

Abrams, E.S., Audeh, M., Boles, C. (1998). Bridge amplification technology technical update, Mosaic technologies, Inc.

Alm, E.W., Oerther, D.B., Larsen, N., et al. (1996). The oligonucleotide probe database. *Appl. Environ. Microbiol.*, **62(10),** 3557-3559.

Boleda, M.D., Briones, P., Farres, J., et al. (1996). Experimental design: a useful tool for PCR Optimization. *Biotechniques*, **21(1),** 134-140.

Boles, T.C., Rehman, F.N., Audeh, M., et al. (1998). Bridge amplification: A solid phase PCR system for the amplification and detection of allelic differences in single copy genes. Mosaic technologies, Inc.

Bolton, E.T. and McCarthy, B.J. (1962). *PNAS*, **48**, 1390-1397.

Brock, K.V. (1995). Diagnosis of bovine viral diarrhea virus infections. *Vet. Clin. North. Am. Food. Anim. Pract.*, **11(3)**, 549-561.

Buetow, K.H., Edmonson, M., MacDonald, R., et al. (2001). High-throughput development and characterization of a genomewide collection of gene-based single nucleotide polymorphism markers by chip-based matrix-assisted laser desorption/ ionization time-of-flight mass spectrometry. *Proc. Natl. Acad. Sci. USA,* **98(2),** 581-584.

Burgart, L.J., Robinson, R.A., Heller, M.J., et al. (1992). Multiplex polymerase chain reaction. *Mod. Pathol.*, **5(3),** 320-323.

Campi, M.G., Romano, P., Milanesi, L., et al. (1998). Molecular Probe Data Base (MPDB). *Nucleic. Acids Res.*, **26(1)**, 145-147.

Condit, R. (2001). Principles of Virology. In Fields, B.N., Howley P.M.,Griffin, D. (eds). *Fields – Virology,* Lippincott Williams & Wilkins, Philadelphia, 19-52.

Dieffenbach, C.W., Lowe, T.M., Dveksler, G.S. (1993). General concepts for PCR primer design. *PCR Methods Appl.*, **3(3),** S30-S37.

Elfath, M.D., Whitley, P., Jacobson, M.S., et al. (2000). Evaluation of an automated system for the collection of packed RBCs, platelets, and plasma. *Transfusion*, **40(10),** 1214-1422.

Ely, J.J., Reeves-Daniel, A., Campbell, M.L., et al. (1998). Influence of magnesium ion concentration and PCR amplification conditions on cross-species PCR. *Biotechniques,* **25(1),**38-40, 42.

Fulton, R.W., Purdy, C.W., Confer, A.W., et al. (2000). Bovine viral diarrhea viral infections in feeder calves with respiratory disease: interactions with Pasteurella spp., parainfluenza-3 virus, and bovine respiratory syncytial virus. *Can. J. Vet. Res.*, **64(3),** 151-159.

Fulton, R.W., Saliki, J.T., Confer, A.W., et al. (2000). Bovine viral diarrhea virus cytopathic and noncytopathic biotypes and type 1 and 2 genotypes in diagnostic laboratory accessions: clinical and necropsy samples from cattle. *J. Vet. Diagn. Invest.*, **12(1),** 33-38.

Henegariu, O., Heerema, N.A., Dlouhy, S.R., et al. (1997). Multiplex PCR: critical parameters and step-by-step protocol. *Biotechniques*, **23(3)**, 504-511.

Henke, W., Herdel, K., Jung, K., et al. (1997). Betaine improves the PCR amplification of GC-rich DNA sequences. *Nucleic Acids Res.*, **25(19),** 3957-3958.

Hilleman, M. (2002). Overview: cause and prevention in biowarfare and bioterrorism. *Vaccine*, **20(25-26),** 3055-3067.

Hokamp, K. and Wolfe, K. (1999). What's new in the library? What's new in GenBank? let PubCrawler tell you. *Trends Genet.*, **15**, 471-472.

Howe, C. (1995), *Gene cloning and manipulation,* Cambridge University Press, New York, 88-90.

Hyndman, D., Cooper, A., Pruzinsky, S., et al. (1996). Software to determine optimal oligonucleotide sequences based on hybridization simulation data. *Biotechniques,* **20,** 1090-1094, 1096-1097.

Innis, M.A. and Gelfand, D.H. (1990). Optimization of PCRs. In: Innis, M.A., Gelfand, D.H., Sninsky, J.J. and White, T.J. (eds), *PCR protocols: A Guide to Methods and Applications,* Academic Press, San Diego, 3-12.

Jeong, J.H. and Park, T.G. (2001). Novel polymer-DNA hybrid polymeric micelles composed of hydrophobic poly (D,L-lactic-co-glycolic acid) and hydrophilic oligonucleotides. *Bioconjug Chem,* **12(6),** 917-923.

Kainz, P. (2000). The PCR plateau phase - towards an understanding of its limitations. *Biochim. Biophys. Acta.*, **1494(1-2),** 23-27.

Kidd, K.K. and Ruano, G. (1995). Optimizing PCR. In: McPherson, M.J., Hames, B.D. and Taylor, G.R. (eds). *PCR 2 : a practical approach,* IRL Press, Oxford 1-21.

Lai, A.C. and Chambers, T.M. (1995). Rapid protocol for sequencing RNA virus using delta Taq version 2.0 DNA polymerase. *Biotechniques*, **19(5),** 704-706.

Larkin, M. (2000). Hunting and logging linked to emerging infectious diseases. *Lancet*, **356(9236)**, 1173.

Lashkari, D.A., DeRisi, J.L., McCusker, J.H., et al. (1997). Yeast microarrays for genome wide parallel genetic and gene expression analysis. *Proc. Natl. Acad. Sci. USA,* **94(24),** 13057-13062.

Lemieux, B., Aharoni, A., Schena, M. (1998). Overview of DNA chip technology. *Mol. Breeding*, **4**, 277-289.

Liu, W.T., Mirzabekov, A.D., Stahl, D.A. (2001). Optimization of an oligonucleotide microchip for microbial identification studies: a non-equilibrium dissociation approach. *Environ. Microbiol.*, **10,** 619-629.

Lockhart, D.J., Dong, H., Byrne, M.C., et al. (1996) Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat. Biotechnol.*, **14(13),** 1675-1680.

Loken, T. (1995). Border Disease In Sheep. *Vet. Clin. North Am. Food Anim. Pract.*, **11(3)**, 579-595.

Mitsuhashi, M. (1996). Technical report: Part 2. Basic requirements for designing optimal PCR primers, *J. Clin. Lab. Anal.,* **10,** 285-293.

Myers, T.W. and Gelfand, D.H. (1991). Reverse transcription and DNA amplification by a Thermus thermophilus DNA polymerase. *Biochemistry*, **30(31),** 7661-7666.

Newton, C.R. and Graham, G,A. (1997). Instrumentation, Reagents and Consumables. *PCR,* 2nd Ed, BIOS Scientific Publishers, Oxford, 9-28.

Newton, C.R. (1995). Primers.. In: Newton, C.R. (ed). *PCR Essential Data,* Chichester, New York : J. Wiley & Sons, 49-56.

Offringa, D.P., Tyson-Medlock, V., Ye, Z., Levandowski, R.A. (2000). A comprehensive systematic approach to identification of influenza A virus genotype using RT-PCR and RFLP. *J. Virol. Methods*, **88(1),** 15-24.

Onodera, K. (1999). Detection of viruses causing bovine respiratory disease. Master Thesis, Oklahoma State University. Stillwater, OK.

Onodera, K. and Melcher, U. (2002). VirOligo: a database of virus-specific oligonucleotides. *Nucleic Acids Res.*, **30**, 203-204.

Onodera, K., d'Offay, J. and Melcher, U. (2002). Nylon membrane-immobilized PCR for detection of bovine viruses, *Biotechniques*, **32**, 74-80.

Park, K.J., Choi, S.H., Lee, S.Y., et al. (2002). Nonstructural 5A protein of hepatitis C virus modulates tumor necrosis factor alpha-stimulated nuclear factor kappa B activation. *J. Biol. Chem.*, **277(15),** 13122-13128.

Ramsay, G. (1998). DNA chips: state-of-the art. (Rev). *Nat. Biotechnol.*, **16(1),** 40-44.

Rehman, F.N., Audeh, M., Abrams, E.S., et al. (1999). Immobilization of acrylamide-modified oligonucleotides by co-polymerization. *Nucleic Acids Res.*, **27(2)**, 649-655.

Relogio, A., Schwager, C., Richter, A., et al. (2002). Optimization of oligonucleotide-based DNA microarrays. *Nucleic Acids Res.*, **30(11),** e51.

Rocha, M.A., Barbosa, E.F., Guedes, R.M., et al. (1999). Detection of BHV-1 in a naturally infected bovine fetus by a nested PCR assay. *Vet. Res. Commun.*, **23(2),**133-141.

Rouillard, J.M., Herbert, C.J., Zuker, M. (2002). OligoArray: genome-scale oligonucleotide design for microarrays. *Bioinformatics,* **18(3),** 486-487.

Rotz, L.D., Khan, A.S., Lillibridge, S.R., et al. (2002). Public health assessment of potential biological terrorism agents. *Emerg. Infect. Dis.*, **8(2),** 225-230.

Rychlik, W. (1995). Selection of primers for polymerase chain reaction, *Mol. Biotechnol.,* **3**, 129-134.

Sambrook, J. and Russell, D.W. (2001). In Vitro Amplification of DNA by the Polymerase Chain Reaction. *Molecular Cloning: A Laboratory Manual,* 3rd Ed., Cold Spring Harbor Laboratory Press, New York, 8.1-8.17.

SantaLucia, J. Jr. (1998). A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc. Natl. Acad. Sci. USA*, **95(4),** 1460-1465.

Schena, M., Shalon, D., Heller, R., et al. (1996). Parallel human genome analysis: microarray-based expression monitoring of 1000 genes. *Proc. Natl. Acad. Sci. USA,* **93(20)**, 10614-9.

Sharrocks, A.D. (1994). The design of Primers for PCR. In: Griffin, H.G. and Griffin, A.M. (eds), *PCR technology: current innovations*, Boca Raton, Florida: CRC Press, 5-12.

Shomer, B. (1996). The PCR primers database. *DNA Seq.*, **6(4),** 255-256.

Stone, R. (2002). FOOT-AND-MOUTH DISEASE: Report Urges U.K. to Vaccinate Herds. *Science*, **297**, 320-321.

Taylor, G.R. (1991). Polymerase chain reaction: basic principles and automation. In: McPherson, M.J., Quirke, P. and Taylor, G.R. (eds), *PCR, a practical approach*, IRL Press, Oxford, 1-14.

Thein, S.L. and Wallace, R.B. (1986). *Human Genetic Diseases: a practical approach,* IRL Press, Herndon, Virginia, 33-50.

Tyagi, S and Kramer, F.R. (1996). Molecular beacons: probes that fluoresce upon hybridization. *Nat. Biotechnol.*, **14(3),** 303-308.

Wang, R.F., Beggs, M.L., Robertson, L.H., Cerniglia, C.E. (2002). Design and evaluation of oligonucleotide-microarray method for the detection of human intestinal bacteria in fecal samples. *FEMS Microbiology Letters*, **10560,** 1-8.

Weissensteiner, T., Lanchbury, J.S. (1996). Strategy for controlling preferential amplification and avoiding false negatives in PCR typing. *Biotechniques*, **21(6),** 1102-1108.

Wetmur, J.G. (1991). DNA probes: applications of the principles of nucleic acid hybridization, *Crit. Rev. Biochem. Mol. Biol.*, **26**, 227-259.

Zubritsky E. (1999). SNP mining. The rush is on. *Anal. Chem.*, **71(19),** 683A-686A.

Appendixes

Appendix 1: Oligonucleotides Used in the ViSH Chip Experiments

Oligo ID	Sequence	Length (nt)
1	AAAAGCAGGGGATAATTCTATTAA	24
3	AAACAAACATCCCTAAGAT	19
4	AAACCAGCAATAGCTCCGAA	20
5	AAACCGGCAATGGCTCCAAA	20
8	AAAGCAGGGGAAAATAAAAACAACC	25
9	AAAGCAGGTACTGATCCG	18
10	AAATACGGTGGATTAAATAAAAGCAA	26
11	AAATCATGGTCCTACATTGCAGAAA	25
12	AAATGGTGCAGGCAATGAGAG	21
13	AACATGCCCATCATCATTCCAG	22
14	AACCCCTTATCGAACCCT	18
15	AACGAGGGTATGTCCACTCC	20
16	AACGGCATAATAACTGAAACC	21
18	AAGCAGGGGATAATTCTATT	20
19	AAGCATGGCTGCATGTTTGTG	21
20	AAGGGCTTTCACCGAAGAGG	20
21	AAGGGGTTTTCATACAGGTATGGT	24
22	AAGGTCCTGCATTCCAAGTG	20
23	AAGTGCACCAGCAGAATAACTGAG	24
24	AATAGATGTAAGGCTTGCAT	20
26	AATCCATCCACATAGGCT	18
27	AATCTAATGTCGCAGTCTC	19
28	AATCTTCTCAGAGGATATGA	20
29	AATCTTCTCAGAGGATATGAA	21
30	AATGATCGGAACTTCTGGAG	20
31	AATGCTGCTCCCACTAGTCCAG	22
34	AATTGCTACCCATATGACAT	20
35	AATTTGCTATGGCTGACGGA	20
36	AATTTTGATGCCTGAAACCGT	21
37	ACAAATTGAGGTGGGTCCG	19

38	ACACTTCCAACCCAATTTGG	20
39	ACAGACCCCTTACCCAGGGT	20
40	ACAGATTGCTGATTCCCAGC	20
41	ACATACAGTGGGATAAGAACC	21
43	ACCAAAATGAAGGCAAACC	19
44	ACCAAGCAACCGATTCAAAC	20
45	ACCAGCAATAGCTCCGAAGA	20
46	ACCAGCAATAGCTCCGAAGAAACC	24
47	ACCAGGATATCGAGGATAACAGGA	24
48	ACCAGGCGATCATGGATAAA	20
49	ACCAGGGTAGTCAAGGGCT	19
50	ACCGCTGGGATACAAATCAG	20
51	ACCGGCAATGGCTCCAAA	18
53	ACCTTCGGCAAAAGCTTCAATACTCCA	27
54	ACCTTGCATTAGAGAGCACA	20
55	ACGCATTCCGACTCCTGG	18
59	ACTATCATTGCTTTGAGC	18
60	ACTCAAAACACCCATTTCCG	20
61	ACTCCAATGGGGGCGATAAAC	21
62	ACTCGTCCTGATTCTTGGA	19
65	ACTGCAGCGTAGACGCTTTGTCC	23
67	ACTGTTACCCTTATGATGTG	20
68	AGACCAGAGGGAAACTATGCCC	22
69	AGAGCTCTTGTTCTCTGATAGGTG	24
73	AGCAAAAGCAGGAGTTTAAAATG	23
74	AGCAAAAGCAGGAGTTTAAAATGAA	25
75	AGCAAAAGCAGGAGTTTAAATGAAT	25
76	AGCAAAAGCAGGCAAACCAT	20
77	AGCAAAAGCAGGCAAACTATTTGAA	25
79	AGCAAAAGCAGGGGAAAATAA	21
80	AGCAAAAGCAGGGGAAAATAAAAAC	25
82	AGCAAAAGCAGGGGATACAAAATG	24
83	AGCAAAAGCAGGGGATACTTTC	22
84	AGCAAAAGCAGGGGATATTTCTGTC	25
85	AGCAAAAGCAGGGTAGATA	19

86	AGCAAAAGCAGGGTAGATAATCACT	25
88	AGCAAAAGCAGGGTGACAA	19
89	AGCAAAAGCAGGGTGACAAA	20
90	AGCAAAAGCAGGGTGACAAAAACAT	25
91	AGCAAAAGCAGGGTGACAAAGACA	24
92	AGCAAAAGCAGGTACTGATCCAAAA	25
93	AGCAAAAGCAGGTAGATATTG	21
94	AGCAAAAGCAGGTAGATATTGAAAG	25
95	AGCAAAAGCAGGTCAAATATATTCA	25
96	AGCAAAGCTTTCAGCAACTG	20
97	AGCAATGGCTCCAAACAGACC	21
98	AGCAATGGCTTCAAACAGACC	21
99	AGCACACATAACTGGAAACAATGC	24
101	AGCAGGGGTATAATCTGTCA	20
103	AGCATGAATTTTGCCAAGAG	20
105	AGCCTGGCCTCAAGCTCATTGAA	23
106	AGCCTTCTAACCGAGGTCGA	20
108	AGCGTAGACGCTTTGTC	17
109	AGCGTTCCTAGTTTTACTTGCAT	23
110	AGCGTTCCTAGTTTTACTTGCATTGA	26
112	AGGACAGAAGCCCTTATAGG	20
113	AGGCGATCAAGAATCCACAA	20
114	AGGGACAATTGGCACGGTTC	20
115	AGGGACTGAAAAGGGTTGGACT	22
117	AGGTTCATCACTATCATA	18
118	AGTACCATCTTTTCTATTGT	20
121	AGTAGAAACAAGGAGTTTTTT	21
122	AGTAGAAACAAGGAGTTTTTTTCGT	25
124	AGTAGAAACAAGGCATTTTTTCACG	25
125	AGTAGAAACAAGGCATTTTTTCAT	24
126	AGTAGAAACAAGGGTATT	18
127	AGTAGAAACAAGGGTATTTTTC	22
128	AGTAGAAACAAGGGTATTTTTCTTT	25
131	AGTAGAAACAAGGGTGTTTTT	21
132	AGTAGAAACAAGGGTGTTTTTAACT	25

133	AGTAGAAACAAGGGTGTTTTTAAT	24
134	AGTAGAAACAAGGGTGTTTTTT	22
135	AGTAGAAACAAGGGTGTTTTTTAT	24
136	AGTAGAAACAAGGTACTTTTTTGGAC	26
138	AGTAGAAACAAGGTAGTTTTT	21
139	AGTAGAAACAAGGTAGTTTTTT	22
140	AGTAGAAACAAGGTAGTTTTTTACT	25
141	AGTAGAAACAAGGTAGTTTTTTACTC	26
142	AGTAGAAACAAGGTCGTTTTTAAAC	25
144	AGTTCCATGCTATTCCCCAG	20
145	AGTTTTACTTGCATTGAATA	20
146	ATAAACGCCCCTGGTAATCC	20
147	ATAAAGATGCTAGTGGGTC	19
148	ATAAGATCCTTCATTACTCAT	21
149	ATAATGAGGTCAGATGCACC	20
150	ATAATGTTTTTCTCATTACT	20
151	ATACCAACCGTCTACCATTC	20
152	ATACCATCCATCTACCATTCC	21
153	ATAGGCTACCATGCGAACAA	20
154	ATAGTCACGTTCAGCGCTG	19
156	ATCAATCTGTCTGGTAGT	18
157	ATCAATTTATTCCCACTTC	19
158	ATCACTCACTGAGTGACATC	20
159	ATCATTCCAGTCCATCCCCCTTCAAT	26
160	ATCCCCGCGGTGCAACTAAG	20
163	ATCTCTTTGTCCCATCCGTG	20
166	ATGATGTGCCGGATTATGCC	20
167	ATGCCTGAAACCGTACCAAC	20
169	ATGGCCATCGGATCCTCAAC	20
170	ATGGCTGCTTGAGTGCTT	18
171	ATGGTAATGGTGTTTGGATAGGAAG	25
172	ATGGTCCAGCTCAAGTTGTCA	21
173	ATGTCGCTGTTTGGAGACAC	20
174	ATGTTCCTTAGTCCTGTAACCAT	23
176	ATTAAATAAGCTGAAACG	18

177	ATTACAGGATTTGCACCTTT	20
178	ATTAGAGCGGAGAAAGGTGG	20
179	ATTCCATCACCATTGTTCC	19
180	ATTGAGAGGGTCAGTTGCTC	20
181	ATTGGGACCTCTTCATAGAA	20
184	ATTTCGCCAACAATTGCTCC	20
186	ATTTCTTTGGCCCCATGGAATGT	23
187	ATTTGTTATCCAGGCAAATT	20
188	ATTTTCTATGAAACCTGCTATTGC	24
192	CAAAGGCCCAGCCTTTCACT	20
193	CAAATTCTCTATCCTCCTTTCCAA	24
194	CAACATGCACCCTGGAGTGC	20
195	CAACCATGAAGCCTCATCAGG	21
196	CAACGGCCTCAAACTGAGTGT	21
197	CAACTAAGTTGCCATTACTGT	21
198	CAAGCGAATCTCTGTAGAGT	20
199	CAATAGCTGGTTTTATAGAA	20
200	CAATCATTTAGTTCTGCAGATG	22
201	CAATGAAACCGGCAATGGCTCC	22
202	CAATTGCATCGGGGGAG	17
203	CACAACTTTCCCTTATACC	19
204	CACAAGCACTGCCTGCTGTA	20
205	CACAAGCACTGCCTGCTGTAC	21
207	CACATTTGGCCTTTGGTTCC	20
208	CACCCATATTGGGCAATTTCCTATGGC	27
209	CACGCTCACCGTGCCCAGTG	20
210	CACTGTCATACTTATTGTATTCGGA	25
211	CAGAGACTTGAAGATGTCTT	20
212	CAGAGACTTGAAGATGTCTTTGC	23
213	CAGAGACTTGAAGATGTCTTTGCTG	25
214	CAGAGGCCATGGATATTGCT	20
215	CAGATGCAGACACAATATGT	20
216	CAGATGCAGACACAATATGTATAGG	25
217	CAGATTGAAGTGACTAATGC	20
218	CAGCATTTTCTTGTGAGCTTCG	22

219	CAGCCATGCTTTTCCATCGTG	21
220	CAGCTCTATGCTGACAAA	18
221	CAGGGGATATTTCTGTCAATCATG	24
222	CAGTAGCATTAGTCACTTCA	20
223	CATAGATGTACCCATACAG	19
225	CATCCTGTTGTATATGAGGCCCAT	24
226	CATGGAATGGCTAAAGACAAGACC	24
227	CATGGACTCAACTGTCATTC	20
228	CATTCTGTTGTATATGAGGCCCAT	24
229	CATTGCAAGCATGAGAAGGA	20
231	CATTTTGCAAATCTCAAAGG	20
232	CCAATCTACAGTATCACTATTCAC	24
235	CCAGCAATAGCTCCGAAGAAA	21
236	CCAGTGACAGAGAACCTGCTAC	22
237	CCATAGACATCTTCGAAGC	19
238	CCATGCATTCATTGTCACACTTGTGG	26
240	CCATTGGGTCAATCTGTATGG	21
241	CCCAGGGATCATTAATTAG	19
242	CCCATTCTCATTACTGCTTC	20
245	CCCTAATGTCTGGAGAAGCCAC	22
246	CCCTCTTGTTGTTGCCGC	18
248	CCGAGATCGCACAGAGACTTGAAGAT	26
250	CCGAGATCGCGCAGAGACTTGAAGAT	26
251	CCGATGATTGTCGAGCTTGC	20
252	CCGTCAGGCCCCCTCAAAGC	20
253	CCGTCAGGCCCCCTCAAAGCCGA	23
254	CCGTCTACCATTCCCTCCCA	20
256	CCGTGCTCGGGTGGTGAACC	20
258	CCTATAATGCACGACAGAAC	20
259	CCTCCAGTTTTCTTAGGATC	20
260	CCTGAATTACAACCAGCAATGC	22
261	CCTGACAGCAACTCCCTCAT	20
262	CCTGAGCACACATAACTGGA	20
265	CCTGCGATTGCGCCGAAT	18
266	CCTGGGACATGCCCCATATG	20

267	CCTTCTGACCGAGGTCGAAAC	21
268	CGAACTGTGTTATCATTCCATTCAAG	26
269	CGAATCTGCACTGGGATAACATC	23
270	CGAGTCTGCGGACATGAGTA	20
273	CGATGCAATTGCAGGCACTT	20
274	CGCAGCAAAAGCAGGGGATA	20
275	CGCTTTGTGGTAGCCCTCCGT	21
276	CGGACAACATGACCACAAC	19
279	CGGGTTCGTTGCCTTTTCGTC	21
280	CGGTAGCCATAACGAATCCGC	21
281	CGGTATCAGAGTGGAACCCC	20
282	CGTCAGCCATAGCAAATTTC	20
283	CGTCTGAGCTCTTCAATGGT	20
284	CTACAGAGACATAAGCATTTC	21
285	CTATAAAGGCCCCATTAAAAC	21
286	CTATCATTGCTTTGAGC	17
287	CTATTGGGAGACCCTCATTGTGATGG	26
288	CTCAGTTATTCTGCTGGTGCACTTGCCA	28
289	CTCATGGAATGGCTAAAGACA	21
292	CTCATTCTCCTGTTCACAGC	20
293	CTCATTTCTTGATCTCCATG	20
294	CTCCAAACAGACCTCTTTTT	20
295	CTCCAGCTCTATGCTGACAAA	21
296	CTCCATAGCCTTAGCCGTAGT	21
297	CTCCCAGGAAACGACAACAG	20
299	CTCGAGAGCAAAAGCAGGGT	20
304	CTCTGAACCCACTGCTACAT	20
305	CTCTGCATTGTCTCCGAAG	19
306	CTCTGGCACTCCTTCCGTAG	20
308	CTCTTCGGTGAAAGCCCTTAG	21
309	CTCTTTAGATCTGCAGCTTGTCCTGTTCCTTC	32
310	CTGACATTGATTACAATTTG	20
311	CTGATGTCGCAGTCTCGCAC	20
312	CTGCAGCGTAGACGCTTTGTCCAAAATG	28
313	CTGCTATTGCGCTGAATAG	19

314	CTGGGACATCATGCAGTGCCAAAC	24
316	CTGTCAGTAAGTATGCTAGAGTCCC	25
317	CTGTCGTGCATTATAGGAAAGCAC	24
318	CTGTGGGAGCAAACCCG	17
319	CTGTTAAGGGACAATGAAGC	20
320	CTTAACAGTCCCATTTGA	18
321	CTTAGAATGAAATGGATGAT	20
322	CTTAGTCCTGTAACCATCCT	20
323	CTTCGTCCCTTTGATGAAGC	20
324	CTTCTAACCGAGGTCGAAACG	21
325	CTTGATCGGCTTCGCCGAGATCAG	24
326	CTTGTGAACTTCAAGTACCAAC	22
327	CTTTCCCAACGGCTTCGAATTG	22
328	GAAAAATTACACTGTTGGTTCGG	23
329	GAAAAATTACACTGTTGGTTCGGTG	25
331	GAAACAATGGGATTCAC	17
334	GAAGATGAGCATCTCTTGGC	20
335	GAAGCAGGGGTACTTTTCC	19
336	GAAGCATCCATTCCCTATGTCTAC	24
337	GAAGGCAAAGCAGAACTAGC	20
338	GAAGGCAATAATTGTACTACTCAT	24
339	GAAGTCTTCATTGATAAACTCCAG	24
340	GAATCCAGCACACAAGAGTC	20
341	GAATCTGCACTGGGATAACATC	22
342	GAATGGATGTCAATCCGACC	20
343	GACAAAGACATAATGGATCC	20
345	GACATTTGAGAAAGCTTGCC	20
346	GACCAGCACTGGAGCTAGGA	20
347	GACCAGCACTGGAGCTAGGG	20
348	GACCCCTTACCCAGGGTCTG	20
349	GACCTTCTGATATTGTGGGGAA	22
350	GACGATCAAGAATCCAC	17
352	GACTATCATTGCTTTGAGC	19
354	GAGCAAGGTAATAAAAGGGTCC	22
355	GAGCAATTGCAACACCCG	18

356	GAGCAGCTGAAACTGCGGTG	20
357	GAGCTGGTTCAGAGTTCCTC	20
358	GAGGCACTTAAAATGACCAT	20
361	GATAATCACTCACTGAGTGAC	21
362	GATCAGAAGTCCCTAAGAGGAAGAG	25
363	GATGCAGACACAATATGTATAGG	23
364	GATTCGCCCTATTGACGAAA	20
365	GATTGAAGTGACTAATGCTA	20
366	GATTGATGTCCGCTCCATCAG	21
367	GATTTGCCATAATGGATAC	19
368	GCAAAAGCAGGGTGACAAAAAC	22
369	GCAAAAGCAGGTAGATATTG	20
370	GCAAAAGCTTCAATACTCCAC	21
372	GCAAGAAACATTGTCTCTGGAGACC	25
374	GCAATCAAAGGAGGTGGAACT	21
375	GCACACTGATAGATGCTCTATTGGG	25
377	GCAGATCCTGCACTGCAGAG	20
378	GCAGGGGAAAATAAAAACAACC	22
379	GCAGGGGAAAATAAAAGCCAC	21
380	GCAGGGGAAACAATGCTATC	20
381	GCAGTTTCCTATAGCAATCC	20
383	GCATGCAATTCTGCTGCAT	19
385	GCATTGTCTCCGAAGAAATAAG	22
386	GCATTTCAATAGGGAAAAT	19
387	GCATTTGGATCATCAAGC	18
388	GCATTTTCTAATATCCACAA	20
389	GCCACTCCACAATATACACCC	21
390	GCCAGTAATACCAGCAATCTCG	22
391	GCCAGTTTTTGGACGTCTTC	20
394	GCGTATTTTGAAGTAACCCCG	21
395	GCGTCGCGTTACCTAACCGA	20
396	GCTACCTTCAACTATACAAACG	22
397	GCTACTGAGCTGGTTCAGAGTTC	23
398	GCTAGAATCAGGCCTTTCTT	20
399	GCTCCAAGCAACATAGCACC	20

400	GCTCTGTCCATGTTATTTGGAT	22
401	GCTCTGTCCATGTTATTTGGATC	23
402	GCTGAAACGAGAAAGTTCTT	20
403	GCTGCCCTGTCGGTGAAGCTCCGT	24
404	GCTGCTTGTCCTGTGCCCTC	20
405	GCTTCAAATGAGAACATGGA	20
406	GCTTCCATCATCTGGTCTGG	20
407	GCTTCCATTTGGAGTGATGC	20
409	GCTTTTGCCCTCAACAGAGAAAAT	24
412	GGACTGCAGCGTAGACGCTT	20
413	GGATTGTTTGGGGCAATAGC	20
414	GGCAAGTGCACCAGCAGAATAAC	23
415	GGCAAGTGCACCAGCAGAATAACT	24
417	GGCACCCTGGAGTTTATCAA	20
419	GGCTGTCAGTAAGTATGCTA	20
420	GGCTTCATACCCAACCATAGAG	22
421	GGGAAAGATCCTAAGAAAAC	20
422	GGGAATTGAACATATCGA	18
424	GGGAGATTGCTGTTTATAGATCC	23
426	GGGGCTTTCATAGCTCCAGATCGTGC	26
433	GGTAGATATTGAAAGATG	18
434	GGTGAAATCAAACATCTTCA	20
435	GGTGACGAGAGAACCTTATG	20
436	GGTGATGCCCCATTCCTTGA	20
437	GGTGCATCTGACCTCATTATTGAG	24
438	GGTGTGACGGCAGCATGCCCT	21
439	GGTGTTCATCACCCGTCTAACAT	23
440	GGTGTTTTTAACTACAATCTGG	22
442	GTAAACACCCACATTCCAAACG	22
443	GTAATCCCGTTAATGGCA	18
444	GTACCTGCTTCTCAGTTC	18
445	GTCAAGAGCACCGATTATCAC	21
447	GTCAGCATCCACAGCACTCTGCTGTTCC	28
448	GTCATACTCCTCTGCATTGT	20
451	GTCCTTCCTATCCAAACACC	20

453	GTGACTGGTGTGATACCACT	20
454	GTGACTGGTGTGATACCACTAAC	23
455	GTGCAAACTGCATCTTGTGG	20
457	GTGCCCAGTGAGCGAGGAC	19
458	GTGCGTGTAAGAGAACAGTG	20
459	GTGGAAACAGAGAAACAT	18
460	GTGGCAATAACTAATCGGTCA	21
461	GTGGGGTCATCAAAGTACAAC	21
462	GTGGTAGCCCTCCGTCTTCTG	21
463	GTGTTAGGAAGGAGTTGAAC	20
464	GTGTTTGACACTTCGCGTCACAT	23
465	GTTGAGGAGCTCTGGACTAG	20
466	GTTGGGAGAAGAGCAACAGC	20
467	GTTGTTTACATAGGACTTGCTCAG	24
468	GTTTAGTCACTGGCAAACGG	20
469	GTTTCTCTGGTACATTCCGC	20
470	GTTTCTCTGGTACATTCCGCA	21
472	TAAAGATATGCCTTCAAAAGCAA	23
473	TAATCCTCTGCTGTGTCCCTCC	22
474	TAATGCCTCATCAGGAGTGA	20
476	TACACCCAGTCACAATAGGAGAGTG	25
477	TACCAAATTGAGCGCTATGC	20
478	TACCATACCTGTATGAAAACC	21
481	TACGCTGCAGTCCTCGCTCA	20
482	TACTTGTCAATGGTGAACGG	20
483	TAGCAGAAGATTCACCCCAG	20
486	TAGGAGACCCCCACTGTGAT	20
487	TATAGGCTACCATGCGAACAA	21
489	TATGCCTATAAAATTGTCAAG	21
490	TATGGACCTGCCGTAGCC	18
494	TCACAATGAGGGTCTCCCA	19
495	TCACATATACTCCTTGCCA	19
499	TCACTTTGCCGGTATCAGGG	20
500	TCAGATTGAAGTGACTAATGCT	22
502	TCATCGCCGAACAATTCACC	20

503	TCATTGGGATCTTGCAC	17
504	TCCAGTTATGTGTGCTCAGG	20
505	TCCCCTGCTCGTTGCTATGGTGG	23
506	TCCCTTAGGTCACTAGTTGC	20
507	TCCGAATAGGAACAATCACAGCTTGGTT	28
508	TCCGGCACATCATAAGGG	18
514	TCCTCTGCATTGTCTCCG	18
515	TCCTGTCACCTCTGACTAAGGGGATTTTG	29
518	TCGGTAGCCATAACGAATCC	20
519	TCGTCAGCATCCACAGCA	18
520	TCTAATTGCCCCTTGGTATG	20
521	TCTAGTTTGTTTCTCTGGTA	20
522	TCTAGTTTGTTTTTCTGGTACAT	23
524	TCTCATTTTGCAAATCTCAAAGG	23
525	TCTCCTTGTGCATTTTGATGC	21
526	TCTCCTTGTGCATTTTGATGCC	22
528	TCTCTTGCTCCACTTCAAGC	20
529	TCTGCAATGTAGGACCATGA	20
530	TCTGCTTCACCAATTAAAGG	20
531	TCTGTCCATCCATTAGGATCC	21
533	TCTTCAATGGTGGAACAG	18
534	TCTTCGGTGAAGGCCCTTAGTAAT	24
535	TCTTTCTGGCACGGTCTG	18
536	TGAAAGATGAGCCTTCTAACCGA	23
538	TGAAGTGACTAATGCTACTG	20
540	TGAATGAGTTGGGTGTTCCATTTCA	25
541	TGACTGGGTGTACATTCTGG	20
542	TGAGGACCCAAGTCACGA	18
543	TGAGGGAGCAATTGAGCTCA	20
544	TGAGTTGGGTGTTCCATTTC	20
546	TGATCATCCTGACCAATTCCAT	22
547	TGATGTATGCCCCACATGAA	20
548	TGATTTGGGATCCTAATGGATGGACAG	27
549	TGCACCATGTAATCAACAACAA	22
551	TGCACTTTCCATCATCCTTA	20

552	TGCATGTTCTCCTGTGTAGTAAG	23
553	TGCCTCAAATATTATTGTGT	20
554	TGCCTCAAATATTATTGTGTC	21
555	TGCCTCAAATATTATTGTGTCCCCGGGT	28
556	TGCGATGCTCAGGACGTT	18
557	TGCGGGGAAGGATCCTAAGAAAAC	24
558	TGCTCCCTCTTCGGTGA	17
559	TGCTCTAAGCAAGCAAGCTG	20
560	TGCTGAGCACTTCCTGACAATGGGCT	26
561	TGCTGGGAGTCAGCAATCTG	20
562	TGGAAGATTTTGTGCGAC	18
563	TGGAGGCAATCTGCTTCACC	20
564	TGGATGGCATGCCACTCTGC	20
565	TGGCCTTCTGCTATTTCAAA	20
566	TGGCGCCAAGCGAACAATGGAGAAGA	26
567	TGGGCTGTCTCTGGTTATTC	20
568	TGGTACATTCCGCATCCCTG	20
569	TGTCAGCTATTATGGAGCTG	20
570	TGTCGCTGTTTGGAGACACA	20
573	TGTTCCTGTTACCCTCGATATCC	23
574	TGTTGCACCTAATGTTGCC	19
575	TGTTTTCACCCATATTGGGC	20
578	TTACACTGTTGGTTCGGTGG	20
579	TTACGAACAGATGGAGACT	19
580	TTACTCCAGCTCTATGCTGA	20
581	TTACTGTTAGACGGGTGAT	19
582	TTAGCTCAGGATGTTGAACG	20
583	TTATCATACAGATTCTTGAC	20
584	TTCCAACTCCTTTGACTGCAGCAC	24
585	TTCCATTGATCTGGTCGATGGCTG	24
587	TTCGCCCAAAAACTTCCCGG	20
588	TTCTACTAGACAGACCCCTTACCC	24
589	TTCTAGCTGAGAGAAAAATGAGAAGATGT	29
591	TTGAAATATCCCCGCGGTGC	20
592	TTGAACGCAGCAAAGCTTTCAGCA	24

593	TTGAATGGATGTCAATCCGA	20
596	TTGCCATCCTGGCAACTACTGT	22
597	TTGCTGGTTTTATTGAGGGG	20
598	TTGCTTGGTCAGCAAGTGCA	20
599	TTGGAACCTCAGGATCTTGCC	21
600	TTGGGATATTGCCCGACA	18
601	TTGGGGGGTTCACCACC	17
602	TTGGGTCAGTGAGGGAAAAC	20
603	TTGTCTCCGAAGAAATAAGA	20
604	TTGTTGAACGCAGCAAAGCT	20
606	TTTCACCACTGCCCTCTCTT	20
607	TTTCTAATATCCACAAAATGA	21
608	TTTCTAATATCCACAAAATGAAGGC	25
609	TTTGGGAAGCCACCAATCTG	20
K1	CTACAGTCAAAACCCAACC	19
K10	CCCCACTGTGATGTCTTCC	19
K3	TACAGTCAAAACCCAACC	18
K4	TGCTACCCATATGACATCC	19
K5	ATTGCTACCCATATGACATCC	21
K51	TGTTAGCCAATTCAGTCG	18
K52	ATTGACCCCTAACCCACG	18
K54	TATCCTGCCTGATTGACC	18
K55	CTTATCCTGCCTGATTGACC	20
K56	TTATCCTGCCTGATTGACC	19
K58	TGCTTATCCTGCCTGATTGACC	22
K59	CTGCCTGATTGACCCCTAACC	21
K8	ACCCCCACTGTGATGTCTTCC	21
K2	GGACCTCTTCATAGAAAGAAGC	22
Uni3	AGCAAAAGCAGG	12
Uni5	AGTAGAAACAAG	12
	"Total 480, Average"	21.1

Appendix 2: Homemade Scripts and Descriptions

LMS related files

LMS provides a list of literature whose entries need to be prepared for the VirOligo entries (See diagram A).

Configuration file name: PMID.config (text)

Description "This text file is LMS configuration file. Lists new and updated virus names."

Script name: 1_1st_VisitTo_NCBI (perl)

Description "This script accesses PubMed, downloads all PMIDs for queries in PMID.config file, and stores only unique PMIDs to MySQL database."

Script name: 2_Insert_FM_DataID (perl)

Description "This script accesses the VirOligo entry server and updates completed PMIDs in MySQL database."

Script name: 3_2nd_VisitTo_NCBI (perl)

Description "This script accesses PubMed, and retrieves author's name, journal title, volume number, page number, date published, and article title for each article, which has not been entered to the database. Then, the script stores the retrieved information in MySQL database."

Table name: PMID (MySQL)

Description “This table stores literature information from PubMed.”

Table name: Copied (MySQL)

Description “This table stores a list of waiting articles, for which Inter-Library loan is requested or searching and copying is in process.”

Table name: Ref (MySQL)

Description “This table stores journal locations, Call numbers or web addresses.”

Script name: PMID_request.php (PHP)

Description “This is a user interface for LMS. The script searches a MySQL table, PMID, according to the user selection and finds a location of the journal from a MySQL table, Ref. At the same time, the script excludes the articles other users requested within a certain time according to a MySQL table, Copied.”

Script name: unlisted.php (PHP)

Description “This script compares two MySQL tables, PMID and Ref, and lists journal names not listed in Ref. Journals not listed in Ref need call numbers or web addresses entered in Ref manually.”

Script name: unlistedvolume.php (PHP)

Description “This script compares two MySQL tables, PMID and Ref, and lists volumes of journal not listed in Ref. Library is not subscribing to all volumes of a journal. Thus, additional information for particular volumes or issues is required in Ref.”

VirOligo database related files for the publishing server program (Diagram B)

Table name: dat1 (MySQL)

Description “This table stores common data. dat1 and dat2 are the relational database”

Table name: dat2 (MySQL)

Description “This table stores primer data. dat1 and dat2 are the relational database”

Script name: main.php (PHP)

Description “This is a user interface for a VirOligo database search. The script searches dat1 for Data ID, virus names, or PMID, or searches dat2 for oligo length, Tm, or sequence according to user’s interest, and returns a summary of a search result. If a user selects a VirOligo ID, the script returns detailed information about the entry with the VirOligo ID. The script also look for GI, PMID, Data ID, and entries with the same oligonucleotide sequence and adds links to the result.”

Script name: chformat.pl (Perl)

Description "This script is used for changing a data format of dat1 and dat2. FMP database format is not acceptable for MySQL database. The script change FMP database HTM export format to MySQL "load" command acceptable format."

Script name: primer.php (PHP)

Description "This is a user interface for data entry. The script is only used for volunteer entries from anonymous people. The script receives a Common data entry and a Primer data entry, and returns a primer-entry-page or transfers to "thankyou.php". The entries are stored to MySQL database but these entries are not searchable by "main.php" until curator's approval."

Script name: thankyou.php (PHP)

Description "This script is used after finishing a volunteer entry and emails a notice to the curator of VirOligo database."

Data Analysis related scripts

Script name: tmcal.php (PHP)

Description "This script calculates a Tm by nearest-neighbor method. Only one sequence can be processed at a time. The sequence can contain numbers and spaces, and the script ignores these characters. Thus, the sequence can be directly pasted from NCBI Entrez page, which contains nucleotide position in each line and a space every 10 bp."

Script name: 2tmcal.php (PHP)

Description "This script calculates T_ms by nearest-neighbor method. Multiple sequences can be processed at a time. Sequences can be separated by spaces, tabs, and carriage returns. A Java script in "2tmcal.php" calculates salt concentration from magnesium and potassium concentrations."

Script name: 1-counter.pl (Perl)

Description "This script counts the number of unique sequences or IDs. Before counting, a file must be sorted by Excel program and exported by Tab-delimited text format."

Script name: 4-counterMakeResfile.pl (Perl)

Description "This script counts and creates a file with unique sequences or IDs. Before processing, a target file must be sorted by Excel program and exported by Tab-delimited text format."

Script name: 2-TmofProductbyGenBankResult (Perl)

Description "This script calculates %GC and Tm of PCR products from GI and target locations by accessing NCBI. The script consists of two processes. First, the script provides primer pair selection interface. Some entries contain more than one pair of primers and primer pairs need to be selected according to notes in the VirOligo database. If only two primers are registered in an entry, the script reads target locations of primers and determines a target location of PCR product. If there are more than 2 primers in an entry, the script shows primers and notes from the entry and asks a user to select primer pairs. After all entries are processed, the script generates a file with target locations of PCR products and GI numbers.

Then, the script accesses Genbank and obtains sequence data according to the GI numbers in the file generated above. The script obtains PCR product sequences from the sequence data and target location, and calculates %GC and T_m."

Script name: 8-splice (Perl)

Description "This script generates sequence fragments in a user-selected size from a FASTA format file. If a user selects 20-bp fragments for a 135,301-bp BHV-1 sequence, the script creates a 20-bp fragment from 5' end of BHV-1 sequence and makes more 20-bp fragments by shifting 1 bp to 3' end until a 20-bp fragment reaches 3' end of BHV-1 sequence. At the end, the script generates a file with 135,282 20-bp fragments."

ViSH chip Oligonucleotide Selection Related Files

Table name: blast2 (MySQL)

Description "This table stores BLAST results for oligonucleotides generated by primer design program."

Table name: blast3 (MySQL)

Description "This script stores BLAST results for influenza-specific oligonucleotides from VirOligo database."

Script name: blast2.php (PHP)

Description "This is a user interface for a MySQL table, blast2. The script shows all sequences in the blast2 table. If user selects a sequence, the script shows a BLAST result for the sequence from the blast2 table. The script also defines conserved sequences or non-conserved sequences according to the number of perfect-matches in BLAST results."

Script name: blast3.php (PHP)

Description "This is a user interface for a MySQL table, blast3. The script shows all sequences in the blast3 table. If user selects a sequence, the script shows a BLAST result for the sequence from the blast3 table. The script also can create a table with all sequences, length of sequences, matched stands, matched lengths, and first five matched organism names for each sequence."

Script name: mainProgForBLAST.pl (Perl)

Description "This script accesses NCBI BLAST and stores the BLAST results in MySQL database. The script consists of three steps for faster process. First, the script reads a sequence file, sends each sequence to BLAST, and obtains a RID for each sequence. The RID is the BLAST processing ID number from NCBI and is assigned to each BLAST request. Second, the script sends a last RID to NCBI every 60 seconds and checks whether the last BLAST request is completed. Third, the script sends RIDs, obtains BLAST results, and stores to MySQL database."

Script name: blast.php.pl (Perl)

Description "This script searches a MySQL table, blast3, and finds oligonucleotide length, matched length in BLAST, and strand matched. The script also selects organisms from BLAST results within a user-selected range of mismatches, and reports organism names other than influenza virus. The script generates HTML table format files for easier pasting to Excel spread sheet."

VirOligo database related files for the entry server

Diagram C shows the data entry process. In addition to the diagram, there are curator's options to check and fix entries; manage personnel; help molarity calculations. Due to security concerns, script names and functions of scripts are not listed here. All scripts are written by Java script, FMP script, or FMP web language (CDML).

Molarity calculation script (Java script)

Description "This script calculates final concentration from initial concentration, initial volume, and total volume. The script can handle additions of reagents at RT and/or PCR."

Table name: time.fp3 (FMP)

Description "This table is used as a punch-clock for entry persons"

Excel spread sheet: Timesheet.xlm

Description "This excel spread sheet functions calculates the number of hours worked in any given time or total hours worked for each person or a group. Data is pasted from a FMP table, time.fp3."

Script name: perform script all

Description: "This script calculates Tm for each oligonucleotide based on %GC method."

Script name: delete un-linked date

Description: "FMP creates an empty entry only with Data ID sometimes due to bugs in the FMP. The script removes such empty entries."

Diagram A: LMS scripts and MySQL relations

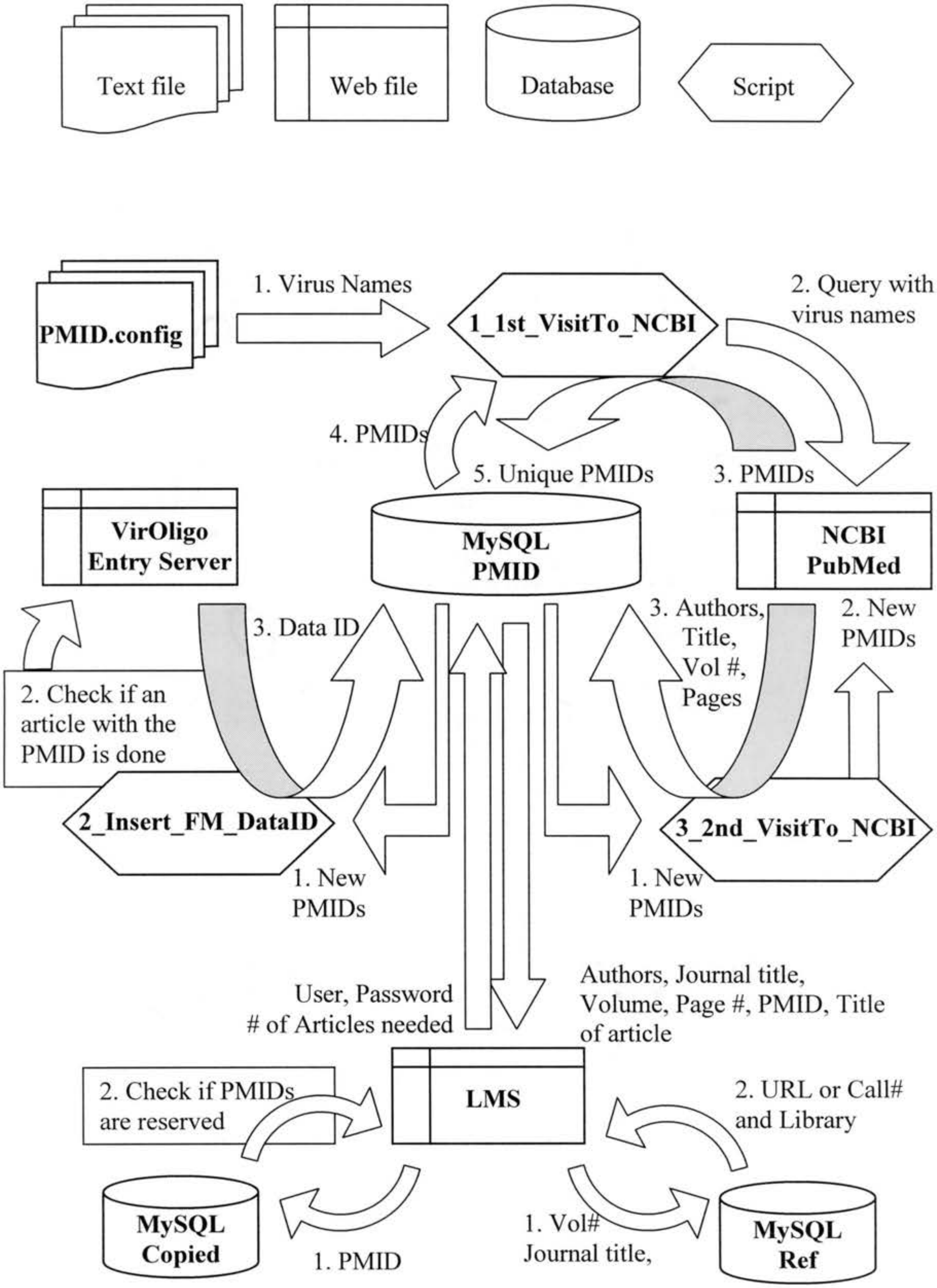


Diagram B: VirOligo Database Search Script and MySQL relations

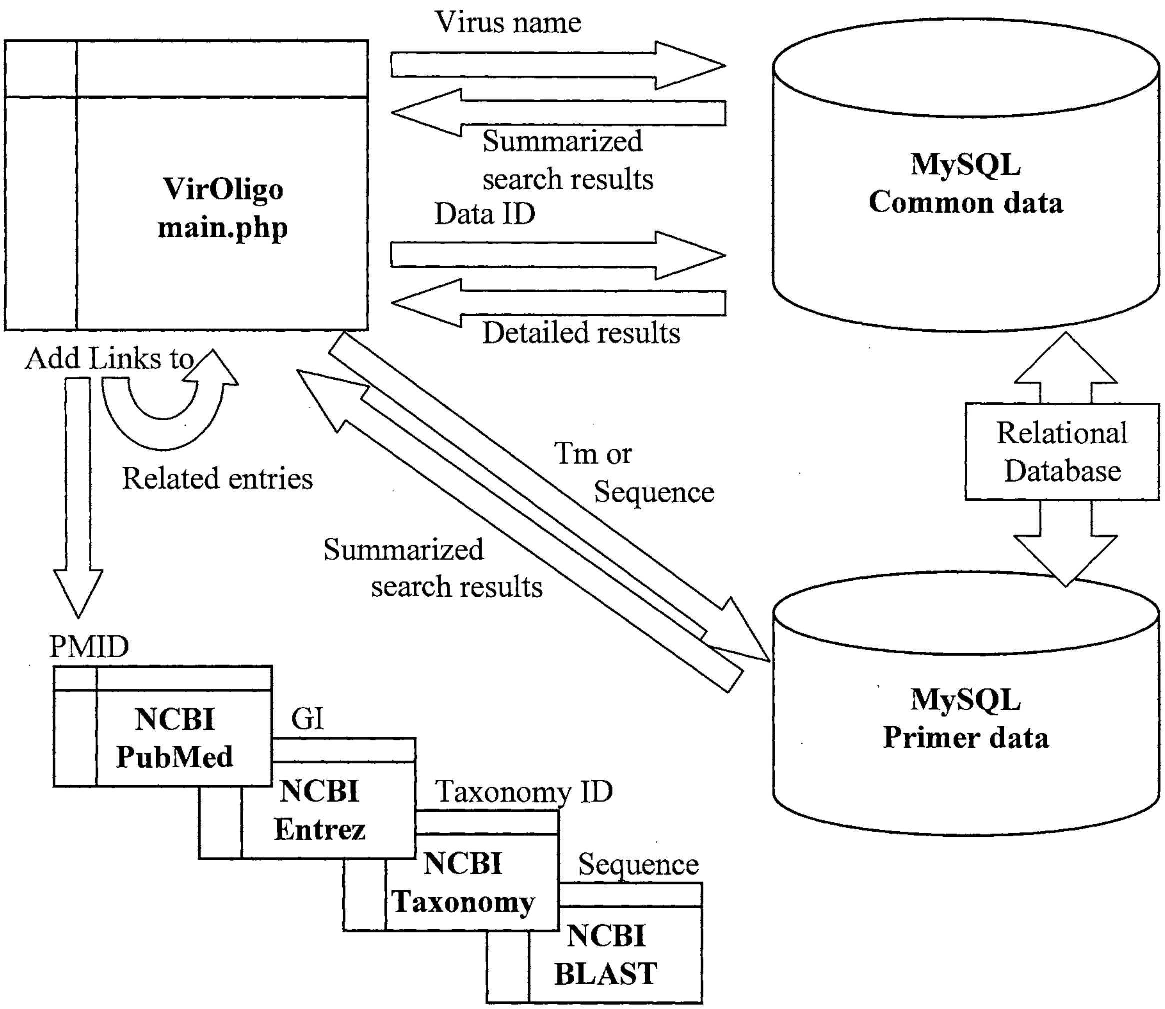
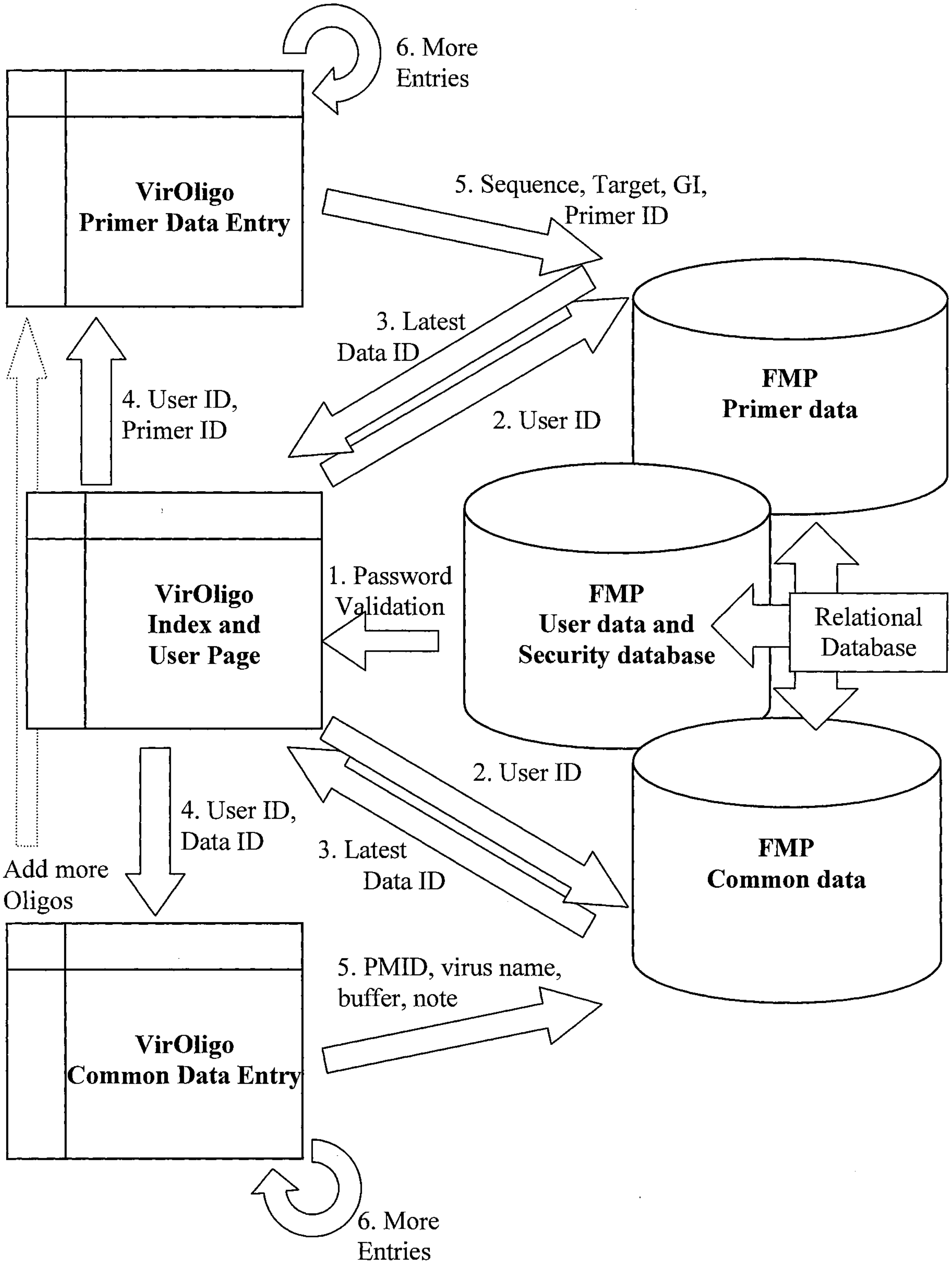
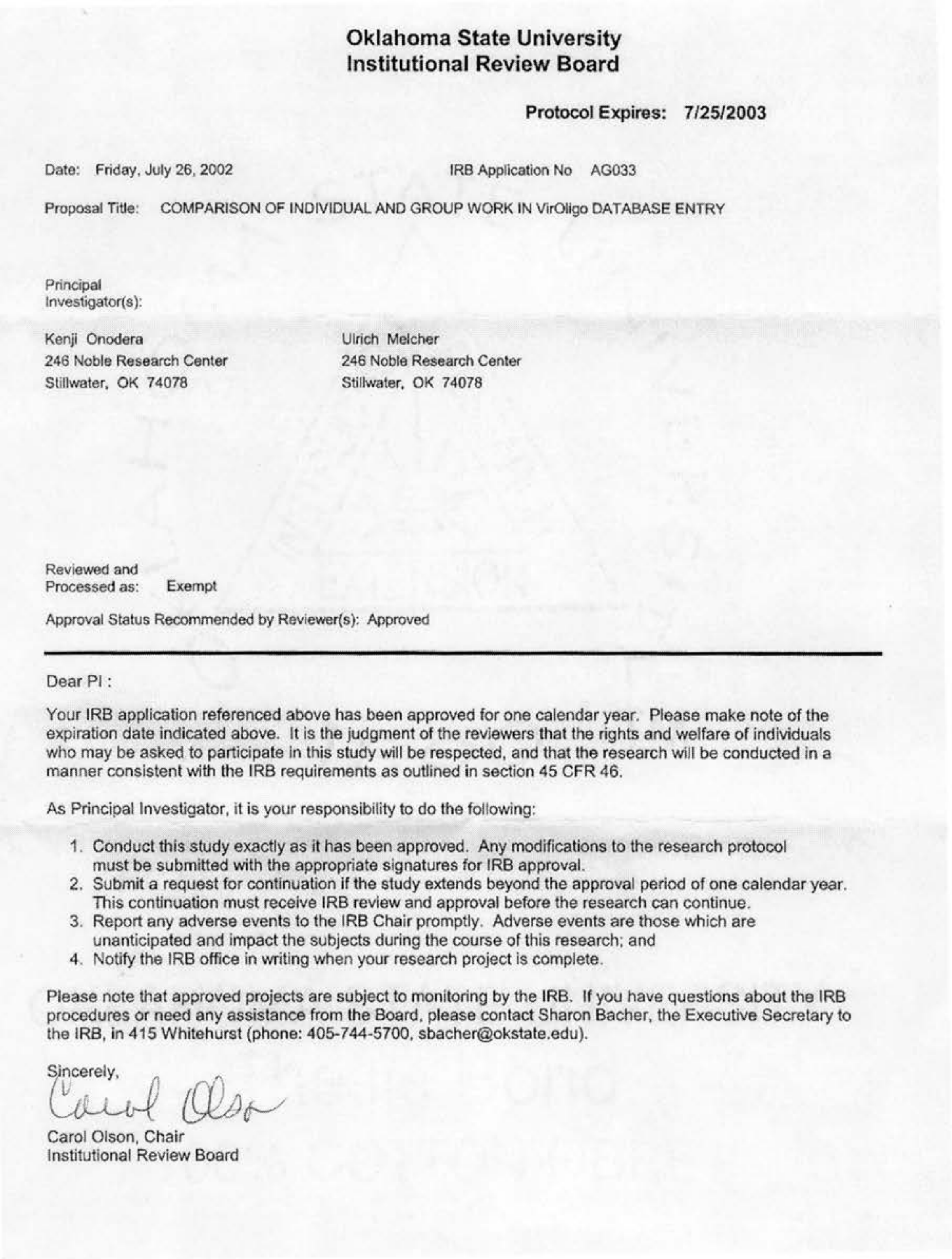


Diagram C: VirOligo Database Entry



Appendix 3: IRB Form



Oklahoma State University
Institutional Review Board

Protocol Expires: 7/25/2003

Date: Friday, July 26, 2002 IRB Application No AG033

Proposal Title: COMPARISON OF INDIVIDUAL AND GROUP WORK IN VirOligo DATABASE ENTRY

Principal
Investigator(s):

Kenji Onodera 246 Noble Research Center Stillwater, OK 74078	Ulrich Melcher 246 Noble Research Center Stillwater, OK 74078

Reviewed and
Processed as: Exempt

Approval Status Recommended by Reviewer(s): Approved

Dear PI :

Your IRB application referenced above has been approved for one calendar year. Please make note of the expiration date indicated above. It is the judgment of the reviewers that the rights and welfare of individuals who may be asked to participate in this study will be respected, and that the research will be conducted in a manner consistent with the IRB requirements as outlined in section 45 CFR 46.

As Principal Investigator, it is your responsibility to do the following:

1. Conduct this study exactly as it has been approved. Any modifications to the research protocol must be submitted with the appropriate signatures for IRB approval.
2. Submit a request for continuation if the study extends beyond the approval period of one calendar year. This continuation must receive IRB review and approval before the research can continue.
3. Report any adverse events to the IRB Chair promptly. Adverse events are those which are unanticipated and impact the subjects during the course of this research; and
4. Notify the IRB office in writing when your research project is complete.

Please note that approved projects are subject to monitoring by the IRB. If you have questions about the IRB procedures or need any assistance from the Board, please contact Sharon Bacher, the Executive Secretary to the IRB, in 415 Whitehurst (phone: 405-744-5700, sbacher@okstate.edu).

Sincerely,

Carol Olson

Carol Olson, Chair
Institutional Review Board

VITA

Kenji Onodera 2

Candidate for the Degree of

Doctor of Philosophy

Thesis: CONSTRUCTION, APPLICATION AND ANALYSIS OF THE OLIGONUCLEOTIDE DATABASE, VIROLIGO

Major Field: Biochemistry and Molecular Biology

Biographical:

Education: Graduated from Kita-Hiroshima High School, Kita-Hiroshima, Japan in March 1991; received Bachelor of Engineering degree in Electrical Engineering from Akita University, Akita, Japan in March 1995. Awarded Master of Science degree in Biochemistry and Molecular Biology at Oklahoma State University, Stillwater, Oklahoma in December 1999. Completed the requirements for the Doctor of Philosophy degree with a major in Biochemistry and Molecular Biology at Oklahoma State University in December 2002.

Professional Experience: Research Assistant, Department of Biochemistry and Molecular Biology, Oklahoma State University, October 1997 to Present.