UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

A COMPARISON OF THE STANDARD DNA SHOTGUN SEQUENCING APPROACH AND THE EXPONENTIAL DNA AMPLIFICATION METHOD FOR SEQUENCING BAC, FOSMID, PLASMID, AND MITOCHONDRIAL DNA

A Dissertation

SUBMITTED TO THE GRADUATE FACULTY

In partial fulfillment of the requirements for the

Degree of

Doctor of Philosophy

By

Jami Elaine Milam

Norman, Oklahoma

2003

INFORMATION TO USERS

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleed-through, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

A COMPARISON OF THE STANDARD DNA SHOTGUN SEQUENCING APPROACH AND THE EXPONENTIAL DNA AMPLIFICATION METHODE FOR SEQUENCING BAC, FOSMID, PLASMID, AND MITOCHONDRIAL DNA

A Dissertation APPROVED FOR THE

DEPARTMENT OF CHEMISTRY AND BIOCHEMISTRY

By

Bruce Roe, Ph.D.

Paul Cook, Ph.D.

Richard Broughton, Ph.D.

C. LeRoy Blank, Ph.D.

Clyde Snow, Ph.D.

Acknowledgements

I want to start by thanking the most important contributor and that is Hashem who has made this dissertation possible.

I wish to convey my gratitude to my advisory committee members, past and present, Drs. Roland Lehr, David McCarthy, Richard Broughton, C. LeRoy Blank, Paul Cook, Clyde Snow and Bruce Roe for their advice and support of my dissertation work. Specifically, I want to thank Dr. Lehr for kindly consenting to temporarily set on the committee for the generals' exam as well as Dr. Broughton for stepping in to replace Dr. McCarthy, when he departed for greener pastures. Further, I would like to thank Dr. Blank for his advice and urging of me to always read the rules. Furthermore, I would like to give Dr. Cook a special thank you for all the extra time he gave me my first semester to assist me in the transition from the more applied oriented medical based biochemistry to the more research oriented field of biochemistry. Another very special thank you, goes to Dr. Snow, who although is very busy, dealing with matters of international importance, yet has he spent some of his precious time setting on this committee as well as encourage a lowly Anthropologist to step out and not fear the deep end of the pool. Likewise, I want to give Dr. Roe a very special thanks too for all his patience, understanding and who continued to wholly stand behind me throughout all the trials and tribulations I have gone through as a graduate student.

I also want to thank all of my lab-mates at the Advanced Center for Genome Technology for all the things they have done to assist me in my endeavors. My

thanks also go to Steve Shaull and Dr. Fares Najar for all the little things they did to assist me in my research. Further, I want to thank Steve Kenton, who taught me on the "phrapisms" we encountered in the mitochondrial project as well as his continued assistance and patience with my inadequate computer skills on many other projects.

Another special thank you goes to my good friend Diana Hamilton of Dr. Blank's lab, who took time out on her research to assist me in the writing of my proposal.

Finally, I want to thank my Dad for his tireless assistance and to my Mother, who permitted it. I also want to thank my supportive brother, Nathaniel, and my sister, Mara. I, further, want to thank all my Christian friends such as Louise Burchett, Lottie Adams, Pastor Elmore and his family and Church as well as all the people at Hillel, Chabad, and Emanuel Synagogue for their prayers and support, which was a very vital ingredient to my research.

Table of Contents

List of Tables

Table of Figures

Abstract

It has been almost three decades since the initial reports of Sanger and Coulson (Sanger, Nicklen, and Coulson, 1977) and Maxam and Gilbert (1977) described the enzymatic and chemical DNA sequencing methods, respectively. During this time, refinements of these methods have resulted in an increase of several orders of magnitude in the rate of DNA sequencing. These improvements have come about because individual DNA sequencing groups have made a conscious effort to not become satisfied with the technology, but have continuously worked to improve the methodology, while increasing accuracy and reducing the cost. The rational behind this philosophy is the target for DNA sequencing studies that have rapidly progressed from a few kilobase plasmid or cDNA to a multimegabase eukaryote genome. The increasing promise of gaining significant information directly related to the genetic basis of many living systems, as well as the role of the genome in normal and abnormal cellular development, has been the catalyst.

Therefore, during the initial phase of my studies, I became well versed in the standard shotgun sequencing techniques and completed the sequence of six regions of the human and mouse genome totaling over 1 Mbp. Then, because of my interest in forensics and the wide spread use of mitochondrial analysis in this field, I sequenced the entire ~18 Kbp Zebrafish mitochondrial genome. However, it soon became apparent that one of the major troublesome areas in genomic studies is the ability to obtain highly purified DNAs that could serve as templates for further DNA sequencing. To investigate the idea of developing improved methods to obtain the desired, highly purified DNA needed to move the field forward, I therefore, investigated the possibility of amplifying DNA via the rolling circle amplification (Fire and Xu, 1995; Liu *et al.* 1996; Lizardi *et*

al. 1998). My initial studies employed the TempliPhi (Dean *et al.* 2001) whole genome amplification procedure that was developed for amplifying plasmid-based clones directly from lysed bacterial cells. Although this TempliPhi protocol was used successfully for plasmid and BAC amplification, it lacked the specificity needed to prevent unwanted host genomic DNA amplification along with the plasmid or BAC clone that had been transformed into the bacterial host. Since the Phi-29 enzyme remained active at 0°C numerous artifacts were observed. Therefore, in an effort to overcome these limitations, I investigated replacing the Phi-29 enzyme with the Klenow fragment of *Bacillus stearothermophilus* DNA polymerase I (Bst). The results of these experiments revealed that the Bst polymerase could amplify DNA template from one (1) nangram to approximately three (3) micrograms via isothermic rolling circle amplification, and allow for the use of BAC or plasmid specific primers since the enzyme was active at 65°C, and it is both economically priced and does not produce the low temperature associated false positives observed with the Phi-29 enzyme. The Bst, upon further investigation, demonstrated that the Bst DNA polymerase could successfully amplify both plasmid and BAC clones to provide a suitable template for both DNA sequencing and for chromosomal gap closure. Unfortunately, the reproducibility of Bst amplified BACs for end sequencing was extremely variable and could not be stabilized successfully. Thus, a modified Phi-29 amplification mix (GenomiPhi, TempliPhi with additional nucleotides) was investigated. Through these studies, I demonstrated that the GenomiPhi phi-29 mix could be used to successfully amplify plasmid, fosmid and BAC DNA such that they were suitable templates for end sequencing.

Introduction

DNA History

The history of DNA has two paths, one is the proof that DNA is the element through which heredity is carried and the second is the determination of the structure. In 1919, Phoebus Levene proposed that the DNA structure is a tetranucleotide, four bases arranged in a row, one after another. Later in 1938, Signer, Caspersson and Hammarsten determine the molecular weight of DNA to be between 500,000 and 1,000,000 daltons, expanding Levene's tetranucleotide to a polytetranucleotide structure. In 1949, Erwin Chargaff determined that the ratio purine bases, adenine and thymine, and pyrimidine bases, guanine and cytosine, is one to one. Then, in 1951 Rosalind Franklin determined two crystalline forms of DNA, the A form and the B form. Then, in 1952, Gosling and Franklin produce an X-ray diffraction pattern of the B form of DNA. This led to the Watson and Crick paper on the structure of DNA in 1953 (Olby, 2003).

The path toward proving that DNA was the carrier of heredity, started in 1869 when Fritz Miescher discovered that inside the nuclei of pus cells there was an acidic substance, that he called "nuclein," later it was learned that this nuclein is nucleic acid or DNA. Then in 1928, Frederick Griffith showed that these substances in heat-killed bacteria could be transmitted to live bacteria. Now, although it was known that DNA resided in the nucleus, it was still not accepted as the substance through which heredity was passed. It was believed that proteins, like histone, that also are found in the nucleus were the carriers of the heredity information. The major reason is, there are 20 different

amino acids that make up a protein and only four different bases that make up DNA. However, in 1944, Avery, MacLeod and McCarty showed through the use of proteases that the genetic material could not be a protein, since proteases did not prevent the transformation of genetic material, but nucleases could (Brown, 1989; Olby, 2003). Then later on in 1954, Gamow developed a possible DNA code for the synthesis of proteins and then, Crick proposed the central dogma in 1957. Lastly, in 1961, Marshall Nirenberg and Johann Matthaei showed that an amino acid could be encoded by a DNA sequence (Olby, 2003).

DNA

Deoxyribonucleic acid (DNA) is a polymer with an alternating sugar - phosphate backbone with bases or nucleotides attached to the sugar via a glycosidic bond (see figure below). There are four major bases in DNA and the linear arrangement of these bases results in the genetic blueprint that is passed from one cell to the next and ultimately from one generation to the next. The DNA in prokaryotes is found within the cytoplasm, while in eukaryotes, it is contained within the nucleus. In eukaryotes, the mitochondria of plants and animals, and the chloroplast of plants, contain their own genomic DNA. This mitochondrial DNA is a remnant of its symbiotic microbial origin. The DNA's double helix consists of two antiparallel DNA strands that are stabilized by the A-T and G-C, Watson and Crick pairs and by base stacking interactions. DNA is both flexible and stable as compared to RNA, because of the presence of the 2'deoxyribose rather than ribose sugar in its sugar-phosphate backbone.

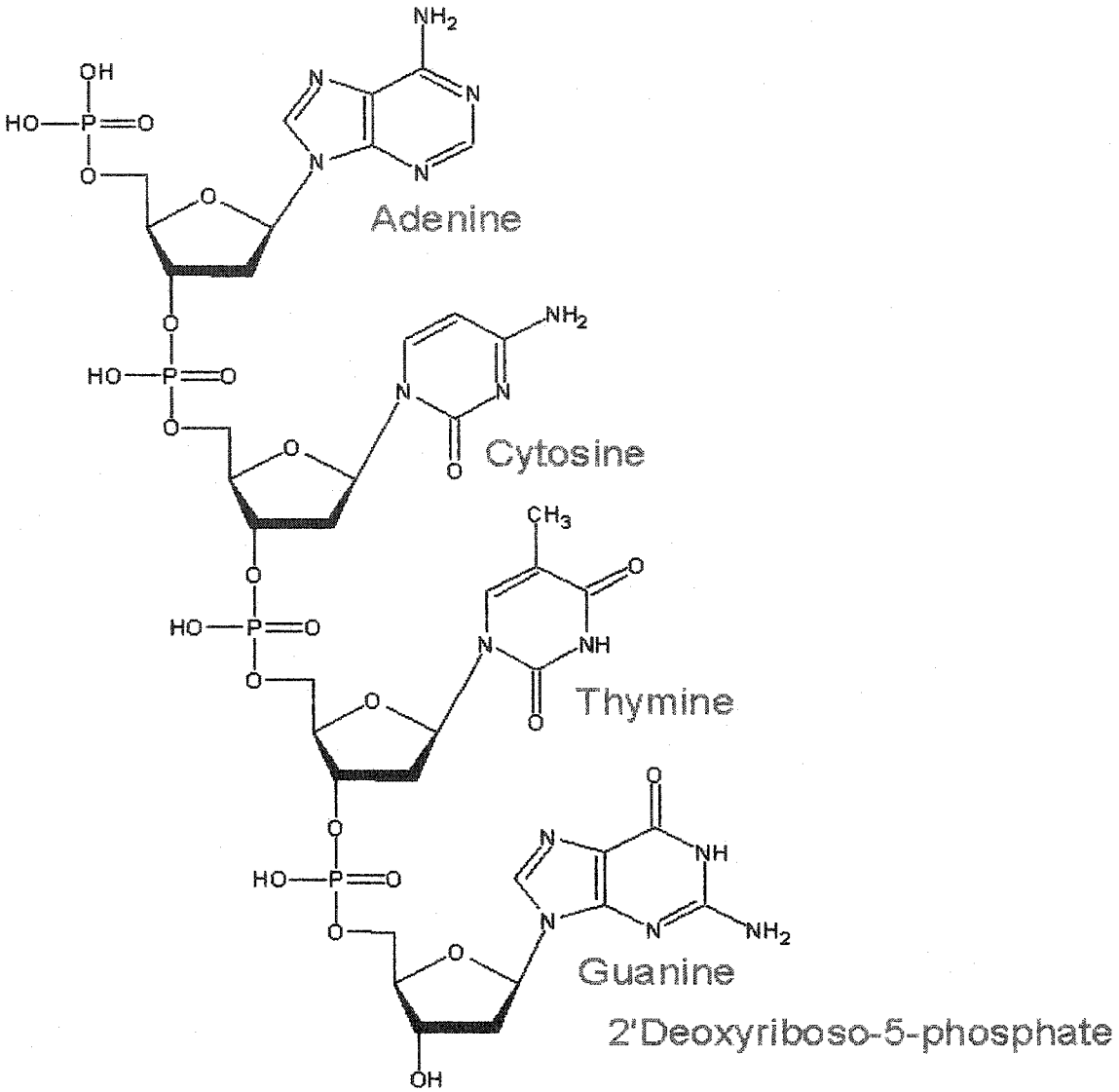


Figure 1 DNA Structures (ChemSketch)

The above pictures were drawn using ACD/ChemSketch.

The A, B, and Z forms of DNA

Accompanying the publication of the structure of the B-form of the double-stranded DNA helix in 1953 (Watson and Crick 1953), Wilkins also published a discussion of B-form DNA (Wilkins *et al.* 1953), and Franklin published additional information on the A-form (Franklin *et al.* 1953). Much later, Alex Rich (Mathews and Van Holde, 1996) reported the crystal structure of the Z-form and subsequently, stable crystals of the C, D, E, and T forms have been obtained (Brown, 1989).

DNA in nature is predominately in the B-form. The DNA-RNA hybrids found during replication, RNA-RNA hybrids and tRNAs are always found in the A form, because the addition of the ribose sugar and the steric hindrances produced. The Z-form of DNA, which is the only known helix that is left handed producing a zigzag pattern is favored when the 5'carbon of cytosine is methylated, which takes place shortly after replication (Mathews and Van Holde, 1996). It is also seen in strands of only CpG nucleotides (Wang *et al.*1981).

There are many additional forms of DNA structures that have been described (Ghosh and Manju, 2003). The C DNA is a slightly different form of B DNA and has been used in NMR studies to better understand the B DNA form (http://www.mhl.soton.ac.uk/research/presentations/2000/Firenze.pdf). The D-form of DNA has been seen in *Staphylococcus aureus.* During transcription, RNA polymerase recognizes and binds to an AT-rich site in which the DNA that becomes over wound and adopts the D conformation (Schumacher *et al.* 2001). The E-form of DNA is an intermediate between the B-form and the A-form. The E-form of DNA can be produced

by cytosine methylation or bromination (Vargeson *et al.* 2000). The T-form of DNA is seen for phage T2 DNA, it is geometrically the same as the D-form of DNA (Ghosh, A. and Manju, 2003). More information on the above DNA is shown in table 1 below.

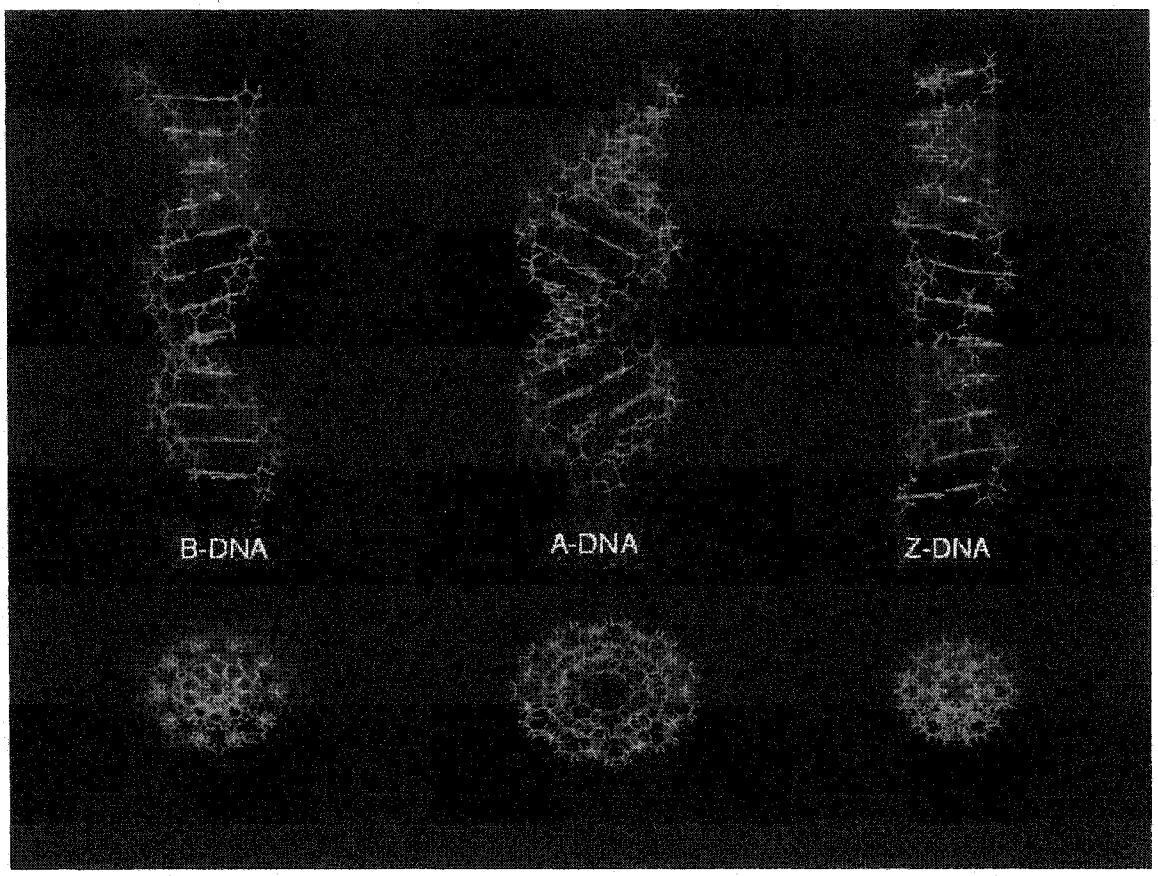


Figure 2 DNA forms: A, B, Z.

The drawing above was done using Rasmol.

	B-form DNA	A-form DNA	Z-form DNA	C-form DNA	D-form DNA	T-form DNA	A-form RNA
Helix Direction	Right	Right	Left	Right	Right	Right	Right
Helix Shape	standard	Wider than B	Zigzag	?	?	?	?
Humidity	92%	75%	High salt	66%	?	?	?
Base pairs per turn	10.5	11	12	9.3	8	8	11
Distance between base pairs (A)	3.4	2.6	3.7	3.3	?	?	?
Base pairs per repeat unit	1	1	2	?	?	?	?
Rise per Base pair	3.4	2.6	3.7	3.31	3.03	3.4	2.8
Helix Diameter (A)	19	23	18	19	?	?	?
Width of Major groove (A)	11.7	2.7	8.8	10.5	8.9	wide	
Width of Minor groove (A)	5.7	11.0	2.0	4.8	1.3	narrow	
Depth of Major groove (A)	8.8	13.5	3.7	7.5	5.8	shallow	
Depth of Minor groove (A)	7.5	2.8	13.8	7.9	5.7	deep	
Twist/bp	36	32.7	-9,-51	38.5	45	45	32.7
Base Tilt	-6	20	-7	-8	-16	-6	16-19
Suger Pucker	C2'-endo	C3'-endo	C3'-endo (syn)	C3'-exo	C3'-exo	C2'-endo	C3'-endo

Table 1 A Comparision of DNA and RNA Forms

(Wilkins *et al.* 1953; Franklin *et al.* 1953; Bowater 2003; Brown 1989)

Replication, Transcription and Translation

The central dogma of molecular biology, first described by Francis Crick (Crick 1970) describes the flow of genetic information from DNA to RNA and RNA to proteins through transcription and translation. In 1970, Temin and Baltimore (Temin 1972; Varmus 1987; Brown 1989) modified Cricks original hypothesis by describing reverse transcription, where DNA is synthesized from a viral RNA genome template prior to transcription and translation. DNA replication now is considered part of central dogma and the entire process is depicted in the figure 3 below.

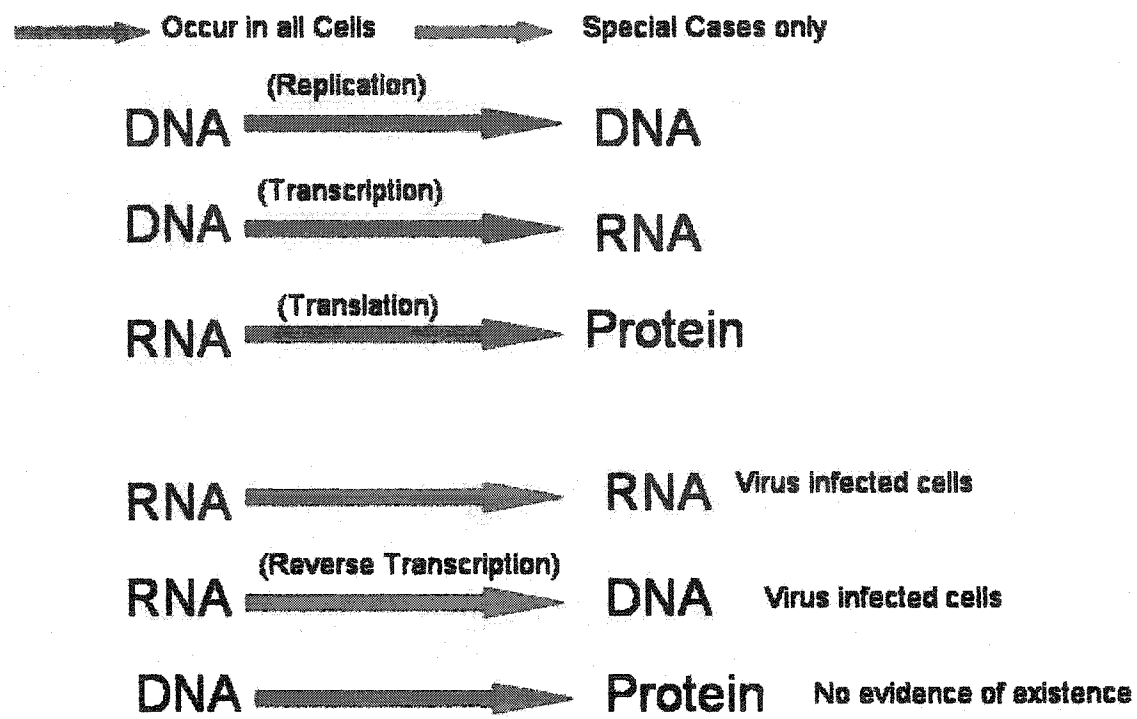


Figure 3 The Central Dogma.

DNA replication is semiconservative. The new strand of DNA is made from each template strand. DNA synthesis requires a polymerase and a template, the original stand of DNA to direct the deoxynucleotide triphosphates; substrates, necessary to make the new strand; and a primer to provide a 3' hydroxyl necessary for the enzymatic addition of the substrate to the primer chain to produce the new DNA strand.

The helicase unwinds the DNA to provide a single stranded template for replication. In front of the helicase is a topoisomerase that relieves both positive and negative supercoils as the replication fork moves forward. (Champe and Harvey, 1994) Polymerase III catalyzes the addition of deoxynucleotide triphosphates to the leading and lagging strands. The leading strand's 3' hydroxyl necessary for the enzymatic addition

follows the opening of the fork to provide a continuous stand addition. However, the opposite side has the 3' hydroxyl moving away from the fork opening requiring multiple segments (Okazaki fragments) of replication to make the lagging strand's side. Each new strand requires a new primer; the primase of the primosome makes these new RNA primers for the initiation of each new Okazaki fragment. After the amplification of the Okazaki fragment, Polymerase I removes these RNA primers and replaces them with DNA. The DNA segments are joined via DNA ligase, to make one continuous strand. The DNA replication process is shown schematically in figure 4 below (Davidson and Sittman, 1999).

Transcription is the process of constructing a messenger ribonucleic acid (mRNA) using the DNA as a template. RNA differs from DNA by the use of the sugar ribose and the base, uracil (Fig. 1). After the heterogeneous nuclear RNA, hnRNA, construction, and posttranscriptional modifications, the mRNA leaves the nuclease of Eukaryotes for translation. In prokaryotes, there isn't a membrane bound nuclease and the transcription and translation are coupled. After translation, protein synthesis occurs to produce the final protein product (Davidson and Sittman, 1999; Mathews and Van Holde, 1996).

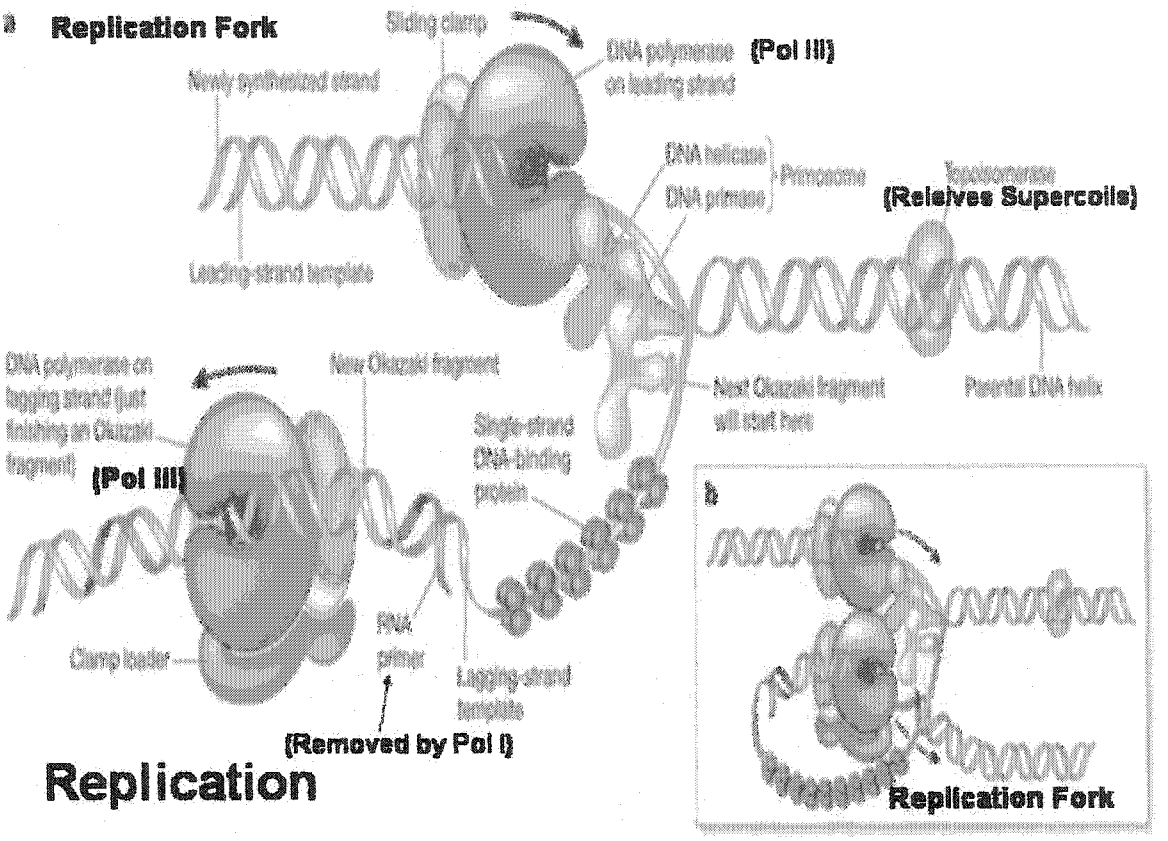


Figure 4 Replication

Redrawn from Nature 2003 (Originally: Alberts, 2002)

Amplification by PCR

Arther Kornberg was the first to synthesis DNA *in vitro* in 1960 (Kornberg, 1960). In 1983, Kary Mullis invented the technique called the Polymerase Chain Reaction (PCR). In PCR, the DNA helix is first denatured and separated by raising the temperature. The temperature then is lowered to allow the small complementary single stranded DNA pieces, primers, to anneal to the template at the target site. The temperature then is raised to the polymerase's optimum polymerizing temperature for a short period of time. The DNA is produced exponentially, however, the amplification is limited to short pieces and errors can be introduced (Innis *et al.* 1990; Dean *et al.* 2001).

The fidelity of an enzyme is its ability to determine the correct substrates to be matched to the master template or the accuracy of the polymerized template made by the enzyme. The lower the error rate, the greater the fidelity of the enzyme (Yang and Chatterjee, 1999).

When comparing the polymerases based on their error rates, due to different techniques used to assay them, only comparison of polymerases whose error rates were determined by the same author should be done. Part of the list below was collected by Eric First, and posted by Paul Hengen (First and Hengen, 2003) and is shown in table 2 below.

Name of DNA Polymerases	Abbrev.	Error Rate	References
Thermus Aquaticus	Taq	$1.1x10^{-4}$ (base substitutions/bps) $2.4x10^{-5}$ (framshift mutation/bp) $2.1x10^{-4}$ (errors/bp) 7.2×10^{-5} (errors/bp) 8.9×10^{-5} (errors/bp) 2.0×10^{-5} (errors/bp) 1.1×10^{-4} (errors/bp)	(Tindall and Kunel, 1988) (Tindall and Kunel, 1988) (Keohavang and Thilly, 1989) (Ling *et al.* 1991) (Cariello *et al.* 1991) (Lundberg *et al.* 1991) (Barnes, 1992)
Thermus aquaticus (Mutant: N-terminal deletion)	KlenTaq	5.1×10^{-5} (errors/bp)	(Barnes, 1992)
Thermoccus litoralis	Vent	2.4×10^{-5} (errors/bp) 4.5×10^{-5} (errors/bp) 5.7×10^{-5} (errors/bp)	(Cariello *et al.* 1991) (Ling *et al.* 1991) (Matilla *et al.* 1991)
Thermoccus litoralis (exo-)	Vent (exo-)	1.9×10^{-4} (errors/bp)	(Matilla *et al.* 1991)
Pyrococcus furiosus	Pfu	1.6×10^{-6} (errors/bp)	(Lundberg *et al.* 1991)
Thermus flavis	Replinase	1.03×10^{-4} (errors/bp)	(Matilla *et al.* 1991)
Bacteriophage Phi-29	Phi-29	1×10^{-9} (errors/bp) 1×10^{-12} (errors/bp) $1 \times 10^{4} - 10^{6}$ (insertion/ bps) $1 \times 10^{5} - 10^{6}$ – fold lower than a properly paired primer	(Drake, 1969:Esteban et 1993) (Cox, 1976: Esteban et 1993) (Esteban *et al.* 1993) (Esteban *et al.* 1993) (Blanco and Salas., 1996)
Bacillus stearothermophilus	Bst	$1 \times 10^{6} - 10^{7}$ (Base substitutions/bps)	(Roe, 2002)

Table 2 Polymerase Error Rates.

Rolling Circle Amplification

To understand rolling circle amplification, the non-synthetic version, rolling circle replication will be intoduced first.

Felix d'Herelle discovered the first bacteriophage (Mulligan, 2003), a virus that infects bacteria. Once the virus deposits its DNA into a host cell some or all of the bacterial host's cellular machinery is used to replicate the viral DNA during the lytic pathway that results in numerous mature viruses that eventually cause the cell to lyse. Alternatively, the viral DNA integrates into the host genome through the lysogenic pathway (Lwoff, 1953; Mayer, 2003). DNA replication has too mechanisms, one is called θ-replication in which the replication forks amplify in opposite directions. The other is called rolling circle replication. Here, one of the DNA strands is nicked and the DNA polymerase extends the DNA, while displacing the DNA strand in its path (Cairns, 1963; Kornberg and Baker, 1992; Brown, 1989).

Rolling Circle Amplification (Fig. 5) is an *in vitro* adaptation of the rolling circle replication. Here, the circular DNA template is amplified via a strand displacing isothermal polymerase with two specific primers, dNTPs, in an optimal pH and salt containing buffer solution (Fire and Xu, 1995; Liu *et al.* 1996; Lizardi *et al.* 1998). The use of just two primers produces the continual exponential, hyperbranched amplification (Dean *et al.* 2001).

Whole Genome Strand Displacement Amplification is similar (Fig. 6) to Rolling Circle Amplification (Fig.5), except a set of random primers are used, instead of just two specific primers (Lizardi, 2001). Whole Genome Strand Displacement Amplification has

one major advantage over Polymerase Chain Reaction, the strand displacement prevents re-annealing artifacts US patent 6,280,949 (Lizardi, 2001).

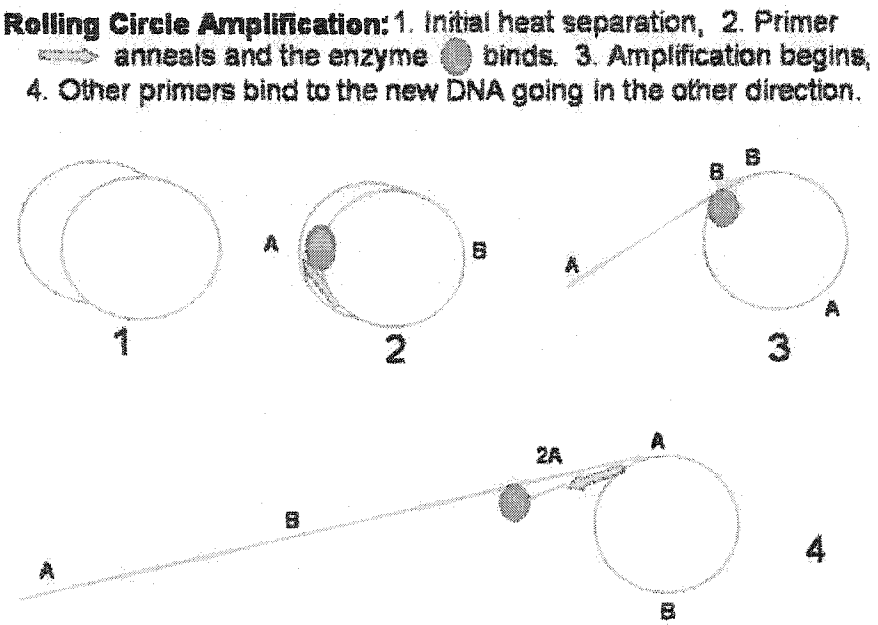


Figure 5 Rollin Circle Amplification

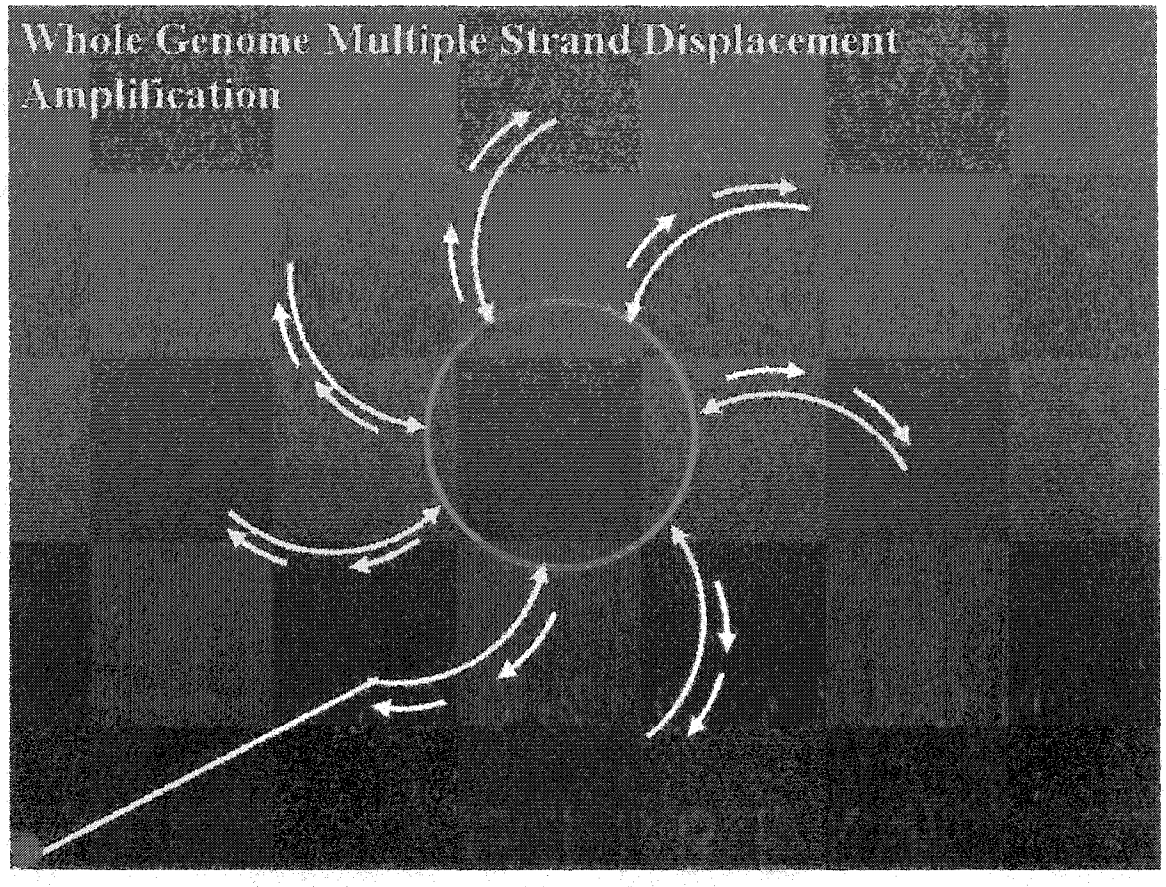


Figure 6 Whole Genome Multiple Strand Displacement Amplification.

It produces continous strands of DNA, similar to unrolling toilet paper, which once another priming site is revealled, it also produces this amplification in the opposite direction. The enzyme is strand displacing, therefore as soon as the amplification starts, another primer behind that location can anneal, allowing for many amplification forks.

Enzymes

Enzymes are proteins that catalyze the rate of reactions without being changed during the process. They are highly specific to their substrates based on their active site according to the lock and key model. They can increase the rate of a reaction from 10^3 to 10^8 times faster than an uncatalyzed reaction producing between 100 to 1000 molecules of product per second (Champ and Harvey, 1994).

Factors Affecting Reaction Rate

Normally, the rate of a reaction increases as the concentration of the substrate increases and does not plateau until the enzymes are saturated with substrate. This is shown in a hyperbolic curve by plotting the velocity vs. the substrate for enzymes that show Michaelis-Menten Kinetics, which are the majority. Michaelis-Menten Kinetics is based on a model, where the enzyme reversibly combines with its substrate to form an enzyme-substrate complex that produces product and regenerates the free enzyme (Champ and Harvey, 1994).

The temperature also affects the rate of the reaction. The rate of a reaction increases as the temperature increases until a velocity plateau is reached. If the temperature continues to increase beyond this plateau the enzyme will reach its' denaturing temperature and the reaction will discontinue soon. If the temperature is decreased from this plateau, the reaction rate will as well (Champ and Harvey, 1994).

The concentration of protons or pH, affects the rate of a reaction. The enzyme and the substrate both have chemical side chains that must be in an ionized or unionized state before the reaction can take place. Each enzyme, functions in a certain pH range and outside of this range will usually causes denaturation of the enzyme due to the change in the ionic state of the protein side chains that form the structure of the enzyme (Champ and Harvey, 1994).

DNA Polymerases

A polymerase is an enzyme that "polymerizes" either ribonucleotide subunits to form RNA or deoxyribonucleotides to form DNA. The first DNA polymerase (pol) was discovered in 1957 by Arthur Kornberg. He isolated it from *E. coli* and it is referred to as DNA pol I. He believed, it was this polymerase that was responsible for DNA replication in the cell, but through the study of mutants and the isolation of DNA pol III in 1972, it was agreed that pol III was most probably the main replication enzyme. It is now known that E. *coli's* Pol I is capable of both DNA amplification and repair, while pol II only does repair (Brown, 1989).

There are at least five types of DNA polymerases in eukaryotes: DNA Pol α (I), which replicates nuclear DNA from the discontinuous strand (Okazaki fragments); DNA Pol β, responsible for DNA repair; DNA Pol γ, that replicates mitochondrial DNA; DNA Pol δ (III), that replicates the nuclear DNA from the continuous strand; and lastly, DNA Pol ε (II), that is involved in DNA repair (Brown, 1989).

There are several DNA polymerases used for *in vitro* amplification and sequencing, e.g. Klenow fragment of *E.coli* DNA polymerase I, DNA dependent RNA polymerase also called reverse transcriptase, bacteriophage T7 DNA polymerase, and Thermus aquaticus DNA polymerase. In the following paragraphs each will briefly be discussed and compared.

Klenow fragment of E. coli DNA polymerase I

The Klenow fragment of E. coli DNA polymerase I is a low processive enzyme. Klenow does not require dideoxynucleotides, because it randomly dissociates from the

DNA template. Since, Klenow is unable to clutch/bind the DNA for very long; this means a shorter sequence, averaging between 250 and 350 nucleotides. Other disadvantages, Klenow does not work well with analogs or with homopolymer tracts, secondary structures, although, it is possible for Klenow to sequence through secondary structures, when the heat is increased to 55 °C. Klenow seems to be designed for sequencing small pieces of DNA, DNA strands under 250 bases. However, it is possible to increase this number of bases by starting with a small amount of dNTPs for the labeling and then increase the number later for the extension (Sambrook et al. 1989).

Bacteriophage T7 DNA polymerase

Sequenase is a commercial variation of bacteriophage T7 DNA polymerase that lacks 3' to 5' exonuclease activity (Sambrook *et al.*1989).

Thermus aquaticus DNA polymerase I

The *Thermus aquaticus* DNA polymerase, Taq polymerase, has a high level of processtivity and polymerizes optimally at 70 °C to 75 °C. These high temperatures enable the polymerase to read/polymerize through strong secondary structures caused by G+C rich regions (Sambrook *et al.* 1989).

Bacteriochage Phi-29 DNA polymerase

The Bacillus subtilis bacteriophage Phi-29 DNA polymerase is a 66 kilo Dalton monomeric enzyme (Blanco and Salas, 1996). Phi-29's characterization and purification are described by Blanco and Salsa (1984) and U.S. patent 5576204 (Blanco *et al.*1996). Phi-29 is a strand displacing isothermic enzyme, preferring 30 °C (Dean *et al.* 2001) that can be partially denatured at 40-42 °C. Its activities are DNA polymerization (pol I) (Blanco *et al.* 1991; Blanco *et al.* 1993; Blanco *et al.* 1996), pyrophosphorolysis, and 3'-

5'-exonuclease (Esteban *et al.* 1993). It has the strand displacing capability and is highly processive, greater than 70,000 bases (Blanco *et al.* 1989; Blanco and Salas, 1996). The Phi-29 enzyme can replicate continuously and without the necessity of Okazaki fragments (Blanco and Salas, 1996). In a ten minute period, a DNA product larger than a 23 kilobase marker can be produced, polymerizing at a rate of 53 nucleotides per second (Lizardi *et al.* 1998; Blanco *et al.* 1989). No slippage occurs with this enzyme due to constitutive strand-displacement (Viguera *et al.* 2001). Phi-29's processiveness combined with the lack of slippage and a circular DNA template allows the rolling circle amplification method to produce a 10,000 fold DNA amplification in only a few hours (Dean *et al.* 2001). Phi-29 has a high fidelity with a mismatch elongation efficiency of 10^5 to 10^6 fold lower than correctly paired primer terminus (Esteban *et al.* 1993). "Unlike PCR, this method does not appear to be limited by target length. Phi-29 DNA polymerase readily synthesizes DNA strands of ~0.5 Mb in length (Baner *et al.* 1998)" (Dean *et al.* 2001).

The Phi-29 use in amplification and sequencing methods have been patented: for "Strand Displacement Amplification" (Walker, 1998); "Whole Genome Strand Displacement Amplification" (Lizardi, 2001); "Unimolecular segment amplification and sequencing" (Lizardi, 2001); "Amplification Methods" (Zhang, 2002); and for "Sequencing" (Blanco *et al.*1991); (Blanco *et al.*1993).

Bacillus stearothermophilus DNA polymerase I

Bacillus stearothermophilus (Bst) DNA polymerase I (large fragment) (Fig.7) is a 76 kilo Dalton enzyme. Bst was first isolated and characterized by Stenesh and Roe (1972). Since then, the over-expression and purification has been patented, (Kong *et al.* 1998). The Bst (large fragment) has been crystallized (Kiefer, 1997). Bst is a isothermic enzyme preferring 65 °C for peak performance, the highest polymerization at the quickest rate (Kong *et al.* 1998). Bst has a published denature temperature of 80 °C in a 10 minute period (New England Biolabs, 2003), however, experimentally it appears to denature at 67 to 68 °C. Bst's only published activity is DNA polymerization (pol I). It has the strand displacing capability and is highly processive, with a Km (dissociation constant) of 9×10^{-5} M (Stenesh and Roe, 1972). No slippage occurs with this enzyme (Viguera *et al.* 2001). Bst has a high fidelity, with a mismatch elongation efficiency of 10^6 to 10^7 (Roe, 2002) and amplifies from nanogram amounts (Mead *et al.* 1991). Bst is the most preferred polymerase for Multiple Strand Displacement Amplification (Lizardi, 2001). Bst's uses in amplification, in sequencing and as a reverse transcriptase have been patented: for "Isothermal Strand Displacement Nucleaic Acid Amplification" (Dattagupt, 1995); "Strand Displacement Amplification" (Swaminathan and Wilkosz 1998); "Low-Temperature Cycle Extension of DNA with High Priming Specificity" (Hong *et al.* 2002); "Method of reducing non-specific amplification in PCR" (McLaughlin *et al.* 2003); "Strand Displacement Amplification" (Walker, 1998); "Whole Genome Strand Displacement Amplification" (Lizardi, 2001); "Unimolecular Segment Amplification and Sequencing" (Lizardi, 2001); "Amplification Methods" (Zhang, 2002); for "Sequencing" (Prockop *et al.* 1994); (Yang and Chatterjee, 1999) ; and with the addition of Mg^{+2} to form a reverse transcriptase, (Schanke, 2000).

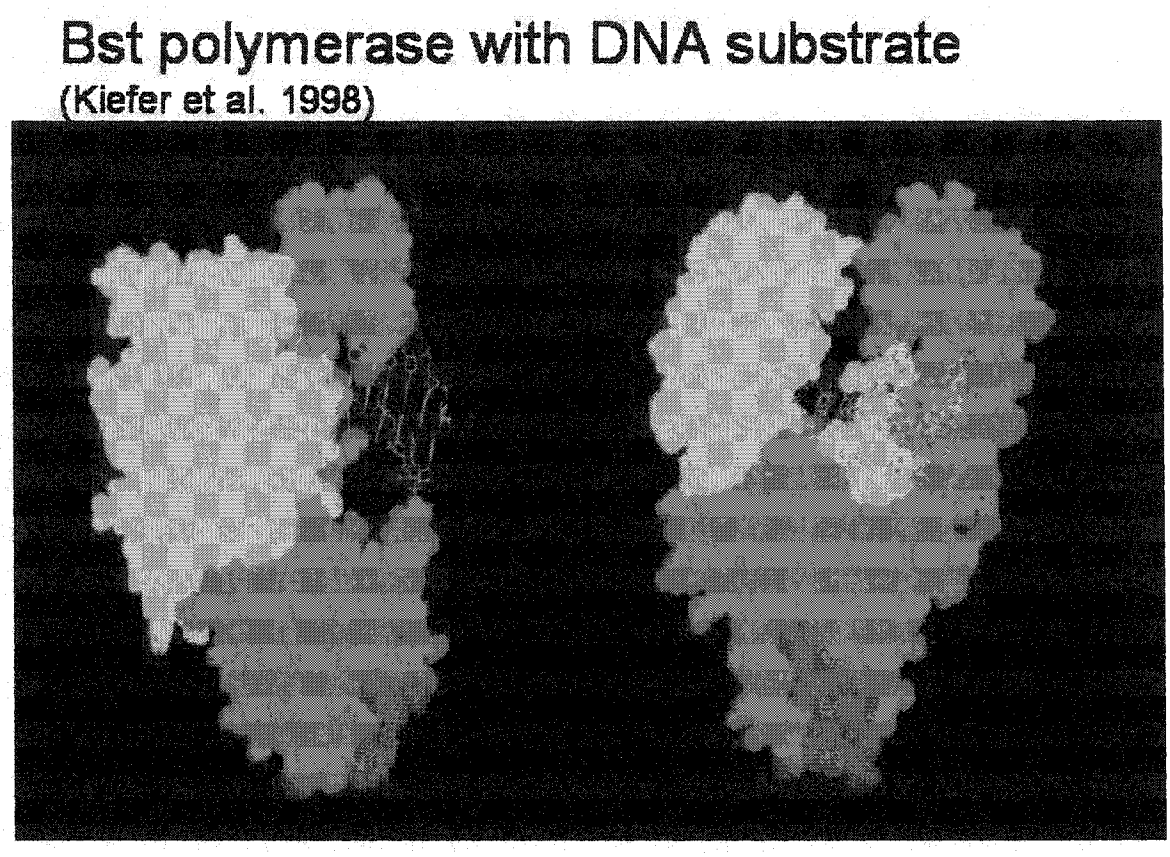


Figure 7 Bst polymerase holding DNA- 2 views

[The drawing above was done using Rasmol (using group colors) with coordinates from the Protein Data Base: Kiefer et al. 1998]

Hierarchical Shotgun Sequencing Overview

After the genomic DNA has been subcloned into a BAC (Bacterial Artificial Chromosome) (Shizuya *et al.* 1992) or PAC (P1 Artificial Chromosome) (Ioannou *et al.* 1994) and mapped. The BACs of interest are chosen and one must isolate the DNA out of the vector and clone it into a new smaller insert vector, multiply the DNA, and then isolate the DNA once more, and sequence the DNAas shown in Figure 8 below.

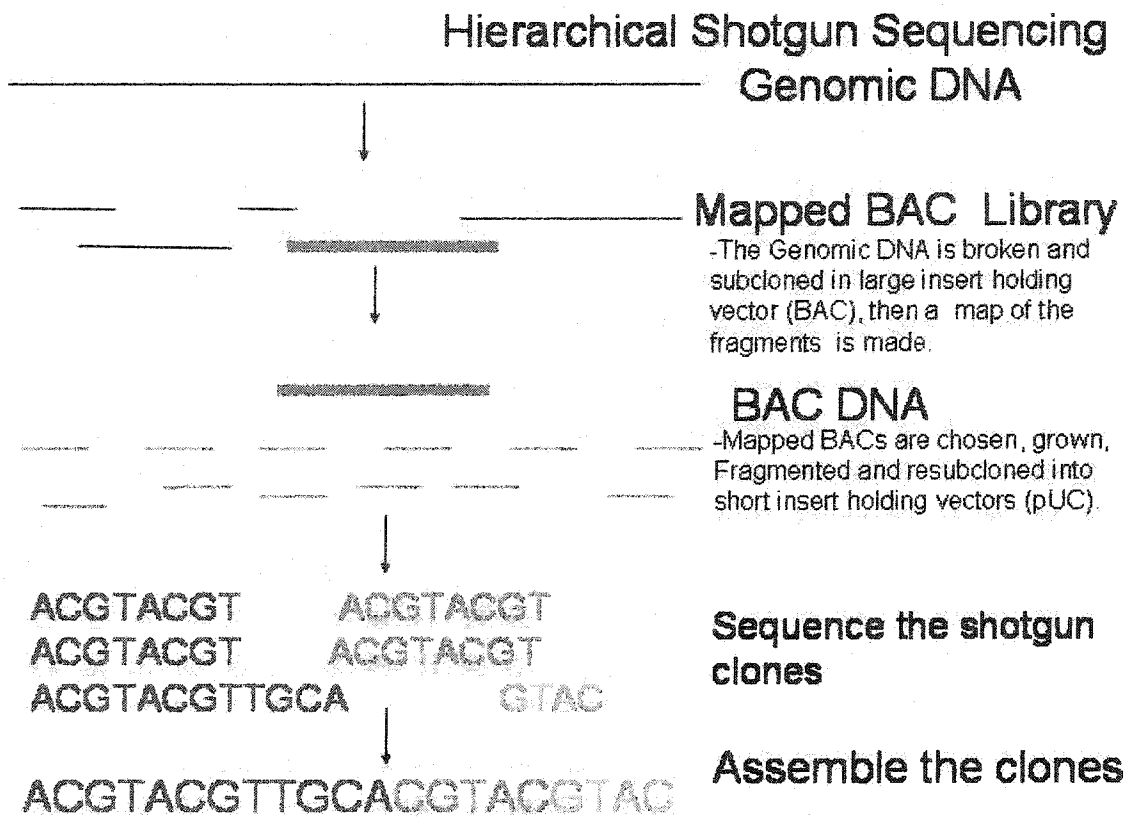


Figure 8 Hierarchical Shotgun Sequencing

Drawing based on information from (International Human Genome Sequencing Consortium, 2001)

DNA Sequencing Methods

The two sequencing methods that currently are used originated from work in the 1970's. These methods are the Maxam and Gilbert chemical degradation method, and the Sanger enzymatic method.

Maxam and Gilbert Method

The Maxam-Gilbert method typically used is for DNA fragments that are less than 250 nucleotides, due to the distribution of cleavage sites, the specific activity of the ^{32}P end-labeled DNA, and the limitations of the gel. In the Maxam and Gilbert method a

fragment of labeled DNA, is partially cleaved at individual bases by five separate base specific chemical reactions. The resulting nested fragment set then is separated on a polyacrylamide gel and autoradiographed to determine the location of bases in the original DNA. A modification of the Maxam-Gilbert method permits the analysis of DNA modifications as well as the study DNA secondary structure and of protein-DNA interactions (Sambrook *et al.* 1989).

Sanger Method

The Sanger chain-terminating sequencing method usually begins with the cloning of the target DNA into a vector. The vector specific primer hybridizes next to the cloning site and its 3' end pointing to the insert. The primer is extended using a DNA polymerase and base incorporation complementary to the DNA template. In a radioactive Sanger DNA sequencing, a reaction is done in four separate tubes based on the 4 types of bases using a radioactive dATP with each tube containing a different chain terminating, 2', 3' – dideoxy nucleotide. Since replication extends a DNA chain from the 3' hydroxyl, when the 3' hydroxyl is missing, a new phosphodiester bond can not be formed and thus the extension of the chain is prevented. Varying the ratio of the 2', 3'-dideoxy nucleotides to deoxy nucleotides results in a wide range of product sizes. If the ratio of the 2', 3'-dideoxy nucleotides is high, only short sequences are produced, but if the ratio of the 2', 3'-dideoxy nucleotides is low only long sequences are produced. Once the nested fragment set is produced, four-reaction mixtures are loaded in parallel on a 6% polyacrylamide gel containing 7M urea. The DNA fragments are separated according to their molecular weight by electrophoresis. Then, autoradiography produces bands on X-ray film that are read from the bottom 5'end to the top 3' end (Weaver, 1999).

In fluorescent Sanger DNA sequencing each chain terminator, contains different fluorescent tags and the reaction is performed with a single reaction vessel. The resulting nested fragment set then is electrophoresed to separate the DNA fragments according to size, the smallest reaching the bottom of the gel first. A laser beam, then passes through the gel and the fluorescent tags are visualized by a laser excitation and return to the ground state of a specific wave length that is read by the detector. The detector sends this information to a computer, which interprets the λmax and calls the bases accordingly (Weaver, 1999).

Dideoxynucleotide DNA Sequencing Reaction Components

The ABI BigDye mix contains the following: AmpliTaq DNA Polymerase, FS, carrying a thermally stable pyrophosphatase, $MgCl_2$; Tris-HCl buffer with a pH of 9.0; deoxynucleotide triphosphates (dATP, dCTP, dITP, dUTP); A-Dye Terminator labeled with dichloro[R6G]; C-Dye Terminator labeled with dichloro[ROX]; G-Dye Terminator labeled with dichloro[R110]; T-Dye Terminator labeled with dichloro[TAMRA] (Perkin-Elmer, 1998). While the Amersham ET Dye Terminator mix contains: Thermo Sequenase™ II DNA polymerase, similar to T7 Sequenace, and terminators labeled with fluorescein and one of four Rhodamine dyes, e.g., Rhodamine 110, Rhodamine-6-G, Tetramethyl Rhodamine, and Rhodamine-X (Amersham Biosciences 2002).

The Polymerase is a mutant of *Thermus aquaticus* DNA polymerase that carries two mutations. One is a point mutation that allows the polymerase to be less restrictive against dideoxynucleotides, to allow the peaks to be of a more even intensity in the gel

and produce less spectral overlap. The second mutation is in the amino terminal where the 5' to 3' exonuclease is deleted (Perkin-Elmer 1998).

Both the ABI BigDye and ET Terminators use dITP instead of dGTP to lessen the band compressions (Perkin-Elmer 1998: Amersham Biosciences 2002) and the ABI BigDye mix also contains dUTP instead of dTTP, to improve the addition of the ddT terminator, resulting in a more even T pattern (Perkin-Elmer 1998).

The ABI BigDyes and Amersham ET Terminators are more efficient in reducing chromatography artifacts than the dRhodamine dyes, because they use a fluorescein donor dye called 6-carboxyfluorescein (6-FAM), which is bound to one of the four dRhodamine dyes. The donor dye absorbs the energy from the laser and transfers it to the acceptor via a linker. Once the energy is transferred to the acceptor, it is released producing a light emission at a different wavelength specific for each base-dye (Perkin-Elmer 1998).

Special Chemistries

dGTP mix, 7-deaza-dGTP mix and dRhodamine

The ABI Big Dye mix or the Amersham ET mix is routinely used for the majority of shotgun sequencing. However, in areas where there are long stretches of repeats forming difficult to close gaps, dRhodamine dye mix and 7-deaza-dGTP dye mix are used.

dGTP mix is used to sequence areas with long stretches of poly G, GT, GC, or C. This mix has the dITP in the regular BigDye mix replaced by dGTP, which although it is considered better for preventing band compression, it causes short read lengths of poly G, GT, GC, or C sequence. An alternative is 7-deaza-dGTP mix, which also prevents band compression, while maintaining the advantages of straight dGTP mix (Perkin-Elmer 1999).

dRhodamine mixes are used with sequences that have long stretches of homopolymer T, A, and G (Perkin-Elmer 1999). A summary of the varying DNA sequencing reaction mixes used in typical DNA sequencing projects is given below in table 3.

	Rhodamine	**IdRhodamine**	**BigDye Terminator**	**dGTP Bigdye & 7deaza-dGTP BigDye**	**Fluorescein/ Rhodamine dye**	**BigDye Primer**
Polymerase	AmpliTaq DNA Polymerase, FS	AmpliTaq DNA Polymerase, FS	AmpliTaq DNA Polymerase, FS	AmpliTaq DNA Polymerase, FS	AmpliTaq DNA Polymerase, FS	AmpliTaq DNA Polymerase, FS
Pyro-phasphatase	rTth Pyro-phasphatase	rTth Pyro-phasphatase	rTth Pyro-phasphatase	rTth Pyro-phasphatase	rTth Pyro-phasphatase	rTth Pyro-phasphatase
$MgCl_2$	yes	yes	yes	yes	yes	Yes
Buffer	yes	yes	yes	yes	yes	yes
dNTPs	dATP, dTTP dCTP dITP for (dGTP)	dATP, dTTP dCTP dITP for (dGTP)	dATP, dUTP for(dTTP) dCTP dITP for (dGTP)	dATP, dTTP dCTP dGTP or 7-deaza-dGTP	dATP, dTTP dCTP 7-deaza-dGTP	dATP, dTTP dCTP dITP for (dGTP)
Terminator/ Primer	Terminator	Terminator	Terminator	Terminator	Primer	Primer
Dye label for A	R6G	dichloro R6G	dichloro R6G	dichloro R6G	JOE	dichloro R6G
Dye label for C	ROX	dichloro ROX	dichloro ROX	dichloro ROX	5-FAM	dichloro ROX
Dye label for G	R110	dichloro R110	dichloro R110	dichloro R110	TAMRA	dichloro R110
Dye label for T	TAMRA	dichloro TAMRA	dichloro TAMRA	dichloro TAMRA	ROX	dichloro TAMRA
Ethelene Oxide linker	N/A	ddG-EO-dR110 ddT-EO-dROX ddC-EO-dTAMRA only				
6-FAM donor dye			Addition of Carboxy-fluroescein to 1 of 4 terminators			
Why?	less band	less band	less band	N/A	N/A	less band

dITP for (dGTP)	compressions	compressions	compressions			compressions
Why? dUTP for dITP	N/A	N/A	Easier addition by polymerase = better T pattern			
Why? Terminators with dichloro	N/A	-more even peaks, narrower emission spectra = less spectra overlap				
Why? Ethelene Oxide linker	N/A	Easier terminator addition by polymerase				
Uses?	Original-not used anymore					
Why? dGTP	N/A	N/A		Use dGTP for better G, to prevent early signal loss of G		
Why? 6-FAM donor dye	N/A	N/A	Terminators are 2-3 x brighter			
Why? 7-deaza-dGTP				less band compressions with 7-deaza-dGTP	less band compressions with 7-deaza-dGTP	

Table 3 Dye Characteristics

(Perkin-Elmer 1999)

Automated DNA Sequencing Instraments

There are five common types of sequencers in use at present. They are the Perkin-Elmer's ABI PRISM ® 377 DNA sequencer, the Perkin-Elmer's ABI PRISM ® 3700 DNA Analyzer, the Molecular ® Dynamic Mega ™ Base 1000 DNA Sequencing System, the LI-COR ® LI-COR 4200 Series and the MJ Research's Base Station. These instruments are summarized below in table 4.

ABI PRISM ® 377 DNA Sequencer

The ABI 377 DNA Sequencer is used to analyze fluorescently labeled DNA fragments via electrophoresis. There are three gel lengths: 24, 36, and 48, the longer the gel the better the resolution. The DNA fragment is labeled with one of four different dyes and is loaded on one of the 36 lanes in an acrylamide gel. The DNA fragments

move through the gel separating according to molecular weight. "At the lower portion of the gel they [DNA fragments] pass through regions where a laser beam [argon] scans continuously across the gel. The laser excites the fluorescent dyes attached to the fragments, and the dyes emit light at a specific wavelength for each dye. The light is collected in 194 channels during each scan and separated according to wavelength by a spectrograph on to a cooled, charge-coupled device (CCD) camera, so all four types of fluorescent emissions can be detected with one pass of the laser. The data collection software collects the light intensities from the CCD at particular wavelength bands (virtual filters) and stores them on a Power Macintosh computer as digital signals for processing" (Perkin-Elmer, 1998). The Sequencing software included decides what to call/label the base (A, C, G, T) based on the intensity of the fluorescence (Perkin-Elmer, 1998).

The type of dye used with this machine is determined by the researcher. One can use dRhodamine, or energy transfer labeled dideoxynucleotide terminator, e.g., BigDye terminator or the Amersham ET Terminator.

ABI PRISM ® 3700 DNA Analyzer

The ABI 3700 sequencer, employs capillaries filled with 3700 POP-6 polymer, a matrix to resolve a fluorescent labeled DNA nested fragments which are loaded via a robotic arm transfers 96 samples to the injection wells. "The glass capillary arrays extend from the injection wells at the cathode end of the electrophoresis chamber to the detection cuvette located at the anode end of the electrophoresis chamber" (Perkin-Elmer 1999). The fluorescently labeled DNA fragments are electrophoresed into the column

and after the injectors are washed and the chamber is filled with buffer and electrophoresis begins. When the DNA fragments reach the end of the capillary, the DNA electrophoresis out of the capillary and into a polymer filled sheath flow where the laser beam excites the dyes and the resulting fluorescence is detected via a charge-coupled devise (CCD), whose silicon chip converts the fluorescent information into digital information (Perking-Elmer 1999).

This machine can detect BigDye terminators, BigDye Primers, dRhodamine terminators, and ET terminators (Perkin-Elmer 1999).

The capillary electrophoresis has several advantages over slab gel electrophoresis. First, one does not have to either pour a gel or load one, as both of these steps are automated. Second, the DNA fragments are separated faster and the capillaries automatically recycle themselves. Third, the removal of the most time consuming step on slab gels, i.e. manual tracking, which is no longer is necessary and no tracking errors occur using the capillaries (Perkin-Elmer 1999).

Mega ™ Bace 1000 DNA Sequencing System

The MegaBace basically is similar to the ABI 3700, except that the laser scans through the capillaries and there are two lasers instead of one. The "...argon-ion laser emits 499-nm (blue) light and a solid-state laser that emits 532-nm (green) light" (MegaBace, 1997). The researcher can use one laser or both alternating. The emitant is caught through an objective lens and is mirrored to a beam splitter, where an Achromatic

lens catches it and focuses it to a point. The emitant goes through an Aperture and is finally received by a photomultiplier tube (PMT) detector (MegaBace, 1997).

LI-COR 4200 Series

The Li-cor uses a slab gel with four wells similar to the manual radioactive method, but instead of using fluorescently tagged dideoxynucletide terminators, it uses fluorescently tagged primers and unlabeled dideoxynucleotides for chain termination. The Li-cor, is similar to the ABI 377, but it uses four wells, one per type of base, and one primer type per well. Li-cor is very sensitive and can detect samples containing as little as 10 attomoles of label. Again, the Li-cor employs electrophoresis to resolve the nested fragment set, but there is no need for autoradiography because it uses a, "...focusing fluorescence microscope containing a solid-state silicon avalanche photodiode scan back and forth across the width of the gel collecting data in real time" (LI-COR, 1998). The resulting "picture" is displayed on the computer screen and the software determines the bases (LI-COR 1998).

The Li-cor uses fluorescein/rhodamine dye primers: JOE for A terminators, FAM for C terminators, TAMARA for G terminators, and ROX for T terminators. The Li-cor mix includes the following: formamide dye, Thermo Sequenase DNA polymerase with a thermostable pyrophosphatase; Tris-HCI pH 9.5, Tween™20, Nonidet™ P-40, 2-mercaptoethanol, dATP, dCTP, 7-deaza-dGTP, dTTP, and one of the previous dNTP's would be a ddNTP and the gel is a polyacrylamide gel (Amersham Pharmacia, 1998).

Base Station

The Base station uses a thin slab gel, which takes about fifteen (15) minutes to solidify. The Base station loads ninety-six (96) samples/run and reads the sequencing samples in ~3.5 hours. Although the Base station is self loading, however, the individual sample lanes must be manually tracked (MJ Research, 2003).

Types of Sequencers

	377	3700	LI-COR	MegaBace	Base Station
Read length/ Number of samples	5-600 48	600 384 X 2 plates 96 X 4 plates	up to 1000 96	600 96 X 4 plates	1000 plus 96
Seq.time/plate	7-11 hr.	4 hr.	12 hr.	3 hr.	3.5 hr
Dye: (Normally used in our Lab)	BigDye1,2,3 Et Terminators dGmix dRhodamine	BigDye1,2,3 Et Terminators dGmix dRhodamine	Fluorescein/ Rhodamine dye	BigDye1,2,3 Et Terminators dGmix dRhodamine	BigDye1,2,3 Et Terminators dGmix dRhodamine
Enzyme:	AmpliTaq	AmpliTaq	Thermo Sequenase	AmpliTaq	AmpliTaq
Fluorescent	@ Terminator	@Terminator	@Primer	@ Terminator	@ Terminator
Labeling Site					
Gel Type:	Polyacrylamide	Pop-6 polymer	Polyacrylamide	Polyacrylamide	Polyacrylamide
Gel Position:	Slab gel	In capillary	Slab gel	In capillary	Slab gel
Loading:	Manual	Self loading In capillary	Manual	Self loading In capillary	Manual
Tracking	Manual Tracking	Self Tracking	Manual Tracking	Self Tracking	Manual Tracking
Shot gun:	Yes	Yes	No	Yes	No
Closure:	Yes	Yes	Only	Preferred	Preferred

Rating 1-5	3	5	1	2	4

1= Worst
5=Best

Table 4 Sequencer Characteristics

DNA Sequencing Effectors

When sequencing DNA, there are different elements to take into consideration that can affect the sequence. The elements that can make a difference are the quality of primers, the quality of the DNA template, DNA polymerases, labeled reagents, and deoxynucleotide analogs. Each will be discussed in the following paragraphs.

Primers

There are many types of primers, but most researchers use the "universal" primer that anneals to vector sequences such as M13 forward and reverse for pUC and the Sp6 and T7 primers for BACs, which also flank the target DNA insert site. Universal primers used for the sequencing of bacteriophage M13 recombinant clones are typically 15-29 nucleotides in length and anneal to the sequences immediately adjacent to one the *Hind*III site in the polycloning region of the bacteriophage M13mp18 and two the *Eco*RI site in the polycloning region of the bacteriophage M13mp19 (Sambrook *et al.* 1989). These primers also are used with the pUC plasmids that have an insert inserted into the multiple cloning site (Sambrook *et al.* 1989). Typically, synthesized primers are analyzed by

capillary electrophoresis prior to use to insure that a full length oligonucleotidehas been synthesized.

Sequencing Computer Programs

The DNA sequence is a long polymer composed of only four bases, which makes assembly of hundreds or thousands of DNA sequences by hand quite impossible. Therefore, computer programs are used to assemble the data. Phred is a program that reads the trace files produced by the sequencers to assign the bases, gives them quality values and writes these to output files (Ewing *et al.* 1998).

Phrap, **Phi**l's **r**apid **a**ssembly **p**rogram, is a computer program that uses the entire shotgun read, not just the highest quality, to assemble the reads. Crossmatch is a program that also uses swat, which is used in conjunction with Phrap to mask the vector sequence (Green 1998).

Consed is a program that is used to view the DNA sequences. Consed flags repeats and allows the viewer to view simultaneously different trace files and export the sequences (Gordon *et al.* 1998).

Once, the assembly is completed and the sequence gaps are revealed, PrimOU, a conversion of PRIMO (Li, P. *et al.* 1997), generates a primer list to close the gaps and improve the sequence quality. This program works in conjunction with Exgap, a computer program written by Axin Hua, "...for contig ordering, subclone and primer selection for primer walking and graphic display of relationships between clone read pairs in contigs" (Roe, 2003).

Annotation and Ensembl

Most sequences, except for bacterial, mitochondrial, and similar, have repeats in their sequences that must be masked before additional annotation can be completed. The annotation of the human, mouse, and other vertebrate genomes thus first entails masking the repeats using Repeatmasker (A.F.A. Smith and P.Green, unpublished; http://www.genome.washington.edu/UWGC/analysistools/RepeatMasker.cfm).

The next step is to use a program to scan for the genes, such as Gencan (Burge and Karlin, 1997), and then analyze the predicted genes using Blast to compare them to other sequences (Altschul, 1997). A 90% match of an unknown gene to a known genes gives an acceptable prediction.

Other programs, such as tRNA scan, which scans for tRNAs (Fichant and Burks 1991). NIP (nucleotide interpretation program for tRNA) (Staden, 1986) and Plotfold for rRNA use the nucleotide sequence to predict various structures (Zuker, 1981).

Other programs used to compare gene/possible gene sequences are Crossmatch (Green, P., 1998), ClustalW (http//www.ebi.ac.uk/clustalw/) and DOTTER, a dotplot analysis program (Sonnhammer, 1995).

Comparative Genomics

Genomics is the study of an organism's genome or DNA, the coding, noncoding, repeats and regulatory sequence. In eukaryotes, protein coding genes consist of exons (coding regions), introns (noncoding regions between exons), and regulatory (noncoding) sequence that eventually results in a protein product. Comparative genomics is the comparison of these genomic features between different organisms.

Frazer puts forth a multistep example of the process to do comparative sequence analysis (Frazer, 2003). First, one identifies the evolutionarily related genomic sequences: homologs, genes that come from a common ancestral gene; orthologs, homologous genes in different species that come from a common ancestor (gene from common ancestral species); and paralogs, homologous genes in same species (duplicated genes from a duplicated chromosomal segment). Second, one annotates the reference sequence: gene sequences, repetitive elements [long interspersed elements (LINES), and short interspersed elements (SINES). SINES tend to contain a recognition restriction site for Alu1 and are called Alu sequences (http://www.cquest.utoronto.ca/botany/bio349s/Lecture/Notes/Lec4.PDF) and the CpG islands, unmethylated CpG dinucleotides, normally methylated, found connected to the 5' end of regulated and/or house-keeping genes (Bird, 1986; Larsen *et al.* 1992). Third, align the genomic sequences; local and/or global alignment programs. Fourth, Identify

conserved sequences; percent identity and thresholds. Fifth, visualize the conserved sequences; using PipMaker and/or VISTA. Both will give a visualization of the orientation of the aligned sequences: exons, repeats genes and conserved non-coding regions (Frazer, 2003).

Pipmaker

Pipmaker is a local alignment tool. To compare the genes using Pipmaker, one uses the genes nucleotide sequences from the different organisms (~2 mb size, the time of implementation is limited) and the masked gene sequence of the base organism's gene as well as the base organisms' annotation file (exon locations) in Gff (General Feature Format) Specifications Document (http://www.sanger.ac.uk/Software/formats/GFF_Spec.shtml) and use Pipmaker (http://bio.cse.psu.edu) (Schwart *et al.* 2000), and/or Multipipmaker (Schwart *et al.* 2003) based on the number of sequences used. PipMaker/Multipipmaker produces a very detailed plot using dashes and dots to show the conservation level between two or more sequences: the X-axis = the reference sequence and the Y-axis = the % identity conserved (Schwart *et al.* 2000)(Schwart *et al.* 2003).

VISTA

Vista is a visualization tool for global alignments. To compare the genes one uses the gene nucleotide sequence from the different organisms (~4 mb size limit) and the masked gene sequence of the base organism's gene as well as the base organisms' annotation file (exon locations) in Gff (General Feature Format) Specifications Document (http://www.sanger.ac.uk/Software/formats/GFF_Spec.shtml) and Vista (http:www-

gsd.lbl.gov/vista) (Mayor *et al.* 2000) to visualize the conserved sequences. Vista produces an output of hills and valleys: the X-axis = the reference sequence; the Y-axis = the % identity conserved; the Exons – numbered boxes above; the Repeat area = other colors based on type (key on plot); the Conserved coding regions = purple; the Conserved Non-coding regions = peach (Dubchak I. *et al.* 2000; Mayor C., *et al.* 2000).

The Human Genome Project

The Human genome project began in 1990's and the first rough draft was completed October 7, 2000 (Cheng, 2003). The human genome project was a collaboration between the United States Department of Energy's Human Genome program and the National Institute of Health's National Human Genome Research Institute. The major goals of the project were to sequence all 3 billion DNA bases, to discover all 30,000 genes and to make the information accessible to the public for further research (HGMIS, 2003).

The Initial Sequencing and Analysis of the Human Genome was published in 2001 (International Human Genome Sequencing Consortium, 2001) and the Initial Sequence and Comparative Analysis of the mouse genome was published in 2002 (Mouse Genome Sequencing Consortium, 2002). There is approximately 75 million years between the human and the mouse, "... evolution has altered their genome sequences and caused them to diverge by nearly one substitution for every two nucleotides as well as by deletion and insertion" (Mouse Genome Sequencing Consortium, 2002). The human genome is 2.9 Giga bases verses the mouse is 2.5 Giga bases in length. The human and mouse genomes can be aligned to give 40% conservation at the nucleotide level and at the gene level, 90% of the genes are conserved segments (ordered genes).

Human chromosome 22 was the first chromosome to be sequenced and it was completed in 1999 (Dunham *et al.* 1999). Of the autosomes, chromosome 22 is the second smallest (21 is slightly smaller) chromosome approximately 1.6-1.8 % of the total Human genome sequence. Chromosome 22 is made of a short p-arm that encodes hundreds of full repeats and ribosomal RNA gene repeats. The long arm is the q-arm, that contains at least 800 genes, many of which have been implicated in various congenital disorders such as DiGeorge syndrome, schizophrenia and Cat eye syndrome (Dunham *et al.* 1999). A reevaluation of the gene annotation for human chromosome 22 was completed and published in 2003 (Collins *et al.* 2003) in which a 74 % increase of annotated regions were reported. The total protein coding genes found was 546, 234 pseudogenes, 32 partial/gene duplications, 31 nonprotein coding transcripts and 16 probably antisense RNAs (Collins *et al.* 2003).

Mitochondrial Genome

A human cell contains thousands of copies of mitochondrial DNA (mtDNA) that encode for several of the proteins involved in the mitochondrial electron transport system. The human mitochondrial DNA represents only 0.5 percentage of the total DNA in a nucleated cell, which is a double stranded circular genome and one strand is the heavy (H) strand, that is rich in guanines and the other strand is the light (L) strand, that is rich in cytosines.

Normally, during DNA replication, unwinding and synthesis are coupled, but this is not so of mitochondrial DNA. One strand is synthesized on a segment of the template strands, while the other is not initiated until later. Both heavy and light strands of

mammalian mt DNA are synthesized continuously as leading strands (Melcher, 2000). The only region of mtDNA that lacks coding sequences is the D loop (displacement loop). This region is transiently triple stranded, doubled plus a single strand, due to the presynthesis of a piece of the H strand that lays dormant for a long time period, before the actual synthesis resumes (Lae, 2000). Replication begins at two specific origins; the H origin is in the D loop, while the L origin only becomes exposed after replication of the H is 2/3 complete. The light strand is then synthesized in the opposite direction using the old Heavy strand as a template (Lae, 2000). Replication is controlled and RNA primers are used; therefore, replication and transcription are related (Melcher, 2000; Scheffler, 1999).

The mitochondrial DNA encodes for 37 genes, 28 on heavy strand, and 9 on light strand, including 2 rRNAs, 22 tRNAs, and 13 polypeptide encoding genes. The mitochondrial genome is extremely compact with about 93% coding (the nuclear genome is only 3%) and no introns. The genes are close together including two with overlapping reading frames (ATPases 6 and 8). The coding sequences of most genes are contiguous or separated by only one or two bases (Lae, 2000; Anderson, 1981; Anderson, 1982; Boore, 1999).

Materials and Methods

Preparation of the DNA for sequencing

After the genomic DNA has been subcloned into a BAC (Bacterial Artificial Chromosome) (Shizuya *et al.* 1992) or PAC (P1 Artificial Chromosome) (Ioannou *et al.* 1994) and mapped, the BAC's of interest are chosen and the BAC DNA is isolated. Then, the DNA is sheared, cloned into a plasmid vector, transformed into a bacterial host to colony purify and amplify the individual cloned fragments, that are then isolated and finally sequenced.

The following is a brief over view of the standard isolation protocols and techniques that were used in the initial phases of this research to sequence in part or all of the human, mouse, and mitochondrial genomes (Roe, 1997). These protocols are also found at (http://www.genome.ou.edu/proto.html).

Large Scale Large Insert DNA Isolation

Starting with a host that already carries the large insert containing vector of interest, BAC, PAC, or Fosmid, a smear of colonies is picked and inoculated into 3 ml of LB medium supplemented with the appropriate antibiotic, after incubation for 8 to 10 hours at 37 °C while shaking at 250 rpm, the 3 ml solution is transferred into 50 ml of LB medium supplemented with the appropriate antibiotic and incubated for another 8 to 10 hours at 37 °C while shaking at 250 rpm. Now, the next step depends on whether one

wishes to use the deep well block for isolation or the 500 ml bottles. If using the blocks, after the allotted time, transfer the 50 ml solution into 1L of LB medium supplemented with the appropriate antibiotic and incubate for another 8 to 10 hours at 37 °C while shaking at 250 rpm. If using the bottles, divide the cells in half place, placing half in each of the one liter flasks of LB medium supplemented with the appropriate antibiotic and incubate for another 8 to 10 hours at 37 °C while shaking at 250 rpm. If using the blocks, to harvest the cells, transfer 1.25 ml of culture in to each well of four 96 deep well blocks and centrifuge the blocks for 10 to 15 minutes at 3000 rpm, Beckman C56R centrifuge,, Beckman C56R, decant and freeze at -80 °C over night. If using the bottles, to harvest the cells, divide the media into two 500 ml bottles and centrifuge at 5000 rpm, Beckman C56R, Beckman C56R centrifuge, 15 minutes, decant and freeze at -80 °C over night.

After the allotted time, the cells need to be resuspended, if using the blocks, add 200 µl of 50:10 TE buffer (Tris-HCl pH 7.6 and disodiumethylenediame tetraacetate pH 8) shaking on a titer plate at a setting of 7 until the cells are totally resuspended. If using the bottles, thaw and add 60 ml of 10mM EDTA (disodiumethylenediame tetraacetate) pH 8. To remove lumps pipette up and down gently.

Next, an alkaline lysis solution (1 % SDS/0.2 N NaOH, sodium dodecyl sulfate and sodium hydroxide) is added to lyse the cells. If using the blocks, add 200 µl to each well then, shake on a titer plate at a setting of 5 for 5 minutes or until completely clear. If using the bottles add 80 ml and swirl the solution very gently, pipetting up and down very slowly to remove lumps until the solution is semitransparent.

Then, 3 M NaOAc pH 4.5 or 3 M KOAc pH 4.8 is added to lower the pH and to facilitate formation of the SDS-lipid complex. If using the blocks, add 200 µl of 3 M NaOAc pH 4.5 incubate at 37 °C while shaking at 350 rpm, Beckman C56R centrifuge, for 10 minutes. Then incubate for 1 hour at -80 °C. If using the bottles, add 60 ml cold 3 M KOAc pH 4.8. Then gentle swirl, then let the bottles set in an ice water bath for 5 minutes, then swirl again.

Next, the centrifuge pellets the cellular membrane, cell debris and proteins to separate them from the DNA. If using the blocks, thaw the blocks in a 37 °C water bath and centrifuge at 3200 rpm, Beckman C56R centrifuge, for 45 minutes. Next, using the Hydra, transfer the supernatant, 400 µl, to new/clean blocks. If using the bottles, centrifuge at 9,000 rpm, Beckman C56R centrifuge,, Beckman C56R, for 20 minutes, then transfer to new/clean 500 ml bottles through 4 layers of cheese cloth.

To precipitate the DNA, in the blocks, add 250 µl of isopropanol to each well, the cover the block with a foil sealer and invert ~3 times to mix the solutions, the remove the sealer and let stand for 5 minutes at room temperature. Centrifuge at 3000 rpm, Beckman C56R centrifuge, for 25 minutes to pellet the DNA, then decant and drain inverted on a paper towel. If using the bottles, add an equal volume of isopropanol and invert ~3 times slowly to mix the solutions. Then, centrifuge at 5,000 rpm, Beckman C56R centrifuge, for 20 minutes, decant the supernatant and drain the pellets by inverting the bottle on a paper towel.

Next, the DNA is resuspended in TE buffer containing RNase A and T1 RNase to hydrolyze any RNA is added to the blocks and later to the bottles. 7.5 M KOAc

is added to salt out the host *E. coli* DNA that has not been sheared to smaller than 200-300K fragments. If using the blocks, add 100 µl per well of 10:1 TE buffer (100 ml) supplemented with RNase A 250 µl of 20 mg/ml and 40 µl of T1 at 100 Uµl . Incubate at 37 °C with shaking at 350 rpm, Beckman C56R centrifuge, for 30 minutes. Next, add 50 µl of 7.5 M KOAc to each well and centrifuge up and down, then pool the solution from all 4 blocks into one block using the Hydra. Incubate at -80 °C until frozen. If using the bottles, resuspend the DNA in 18 ml of 10:50 TE and pipette up and down, gently, then add 9 ml of 7.5 M KOAc and incubate at -80 °C until frozen.

After centrifugation, to pellet the host DNA, the supernatant is transferred to a new container and the BAC DNA is precipitated with two volumes of 100% ethanol (or one volume of isopropanol) and then washed with 70 % ethanol to remove any salts. The DNA then is dissolved in sterile double distilled water since EDTA in TE buffer inhibits polymerases or in TE for long term storage. If using the blocks, thaw the blocks in a 37 °C water bath for ~10 and then centrifuge at 2500 rpm, Beckman C56R centrifuge, for 30 minutes. Transfer 550 µl of supernatant from each well to a new/clean block. Then, add 1.25 ml of ice cold 100% ethanol and incubate at -80 °C for 30 minutes. Centrifuge at 3000 rpm, Beckman C56R centrifuge, for 30 minutes and decant. Next, drain on a paper towel inverted, then cover with a kimwipe and either dry over night on the bench top or in a dryer for ~30 minutes. Lastly, add 40 µl of double distilled water, vortex and centrifuge up and down and incubate over night in cold room (4°C) or in 37 °C water bath for 15 minutes to resuspend the DNA. The pool into 4 Eppendorf tubes (~500 µl/tube). If using the bottles, thaw and centrifuge at 5000 rpm, Beckman C56R centrifuge,, Beckman C56R, for 20 minutes, then transfer the supernatant to two 50 ml

polypropylene tubes and centrifuge at 3200 rpm, Beckman C56R centrifuge, for 10 minutes (2^{nd} pelleting of the host DNA, produces cleaner supernatant). Transfer the supernatant to another 50 ml polypropylene tube and add an equal volume of isopropanol, invert ~3 times to mix. Centrifuge the solution 3200 rpm, Beckman C56R centrifuge, for 30 minutes, then decant and drain on a paper towel. Next, cover the tube end with a kimwipe and dry over night on the bench top (easier to resuspend the DNA) or in a dryer for ~2 hours. Finally, dissolve the DNA in 700 µl 50:50 TE, pipette up and down repeatedly to mix and store the solution at 4 °C overnight. Transfer the solution to 1.5 ml Eppendorf tubes and add 20 µl RNase A 250 µl of 20 mg/ml and 40 µl of T1 at 100 U/µl, incubate at 37 °C water bath for an hour.

Phenol/Ether Extraction

Since the DNA must be extremely pure, a phenol/ether (or chloroform) extraction typically is done. First, add 500 µl (one volume) of cold TE- saturated phenol (leave in cold room until needed), then vortex approx. 30 sec and centrifuge (preferably in the cold room 4 °C) for 5 minutes. Next, transfer the upper layer to a clean Eppendorf tube being careful not to pick up the white (protein) or phenol layers. If necessary, back extract by adding sterile double distilled water to the phenol layer, revortex, recentrifuge, and add the additional top layer to the new tube. Next, take the new Eppendorf tube with the DNA in it and centrifuge it for 5 minutes. Transfer, the DNA solution to another new Eppendorf tube being careful not to disturb or pick up the phenol/gel (after running a agrose gel for size selection, later) pellet at the bottom, repeat this until there is no pellet. Next, add 500 ul (1 volume) of sterile double distilled water saturated ether to each tube, vortex and centrifuge at 3000 rpm in the Beckman C56R centrifuge for 5 minutes. Then,

remove the upper ether layer as much as possible and place tubes in the Speedvac for 20 minutes to evaporate the rest.

Ethanol Sodium Acetate

Then, add 1ml (twice the volume) of ethanol sodium acetate to each tube and vortex well. Incubate the tubes on ice for 15 minutes and centrifuge at 3,000 rpm, Beckman C56R centrifuge, for 15 minutes at 4 °C. (in the cold room). Carefully, decant the ethanol-NaOAc and add 500ul (or 1 volume) of 70% ethanol to each tube, mix by inverting the tube about 8 times. Then, centrifuge at 12,000 rpm, Beckman C56R centrifuge, for 5 minutes. Carefully, decant and dry the tubes in the speedvac for 1 to 2 hours until dry. The tubes may have a small clear/white pellet at the bottom, when dry. Dissolve the DNA Resuspend the DNA in 500 µl of sterile double distilled water.

Shotgun Cloning

Nebulize/Hydroshearing the DNA

The point of this step is to break the DNA into random small clonable pieces. There are multiple methods available: "One class involves enzymatic digestion, either partial digestion with a restriction endonuclease (Maniatis *et al.* 1982) or controlled degradation with DNase I (Anderson 1981). Another class involves physical stresses induced by sonication (Deininger 1983), atomization (Cavalier and Rosenberg 1959), nebulization (Bodenteich *et al.* 1994), or point-sink shearing (Oefner *et al.* 1996)" (Thorsten *et al.* 1998). In my research either the nebuliser or the point-sink shearing "Hydrashear" by GeneMachine (Oefner *et al.* 1996; Thorsten *et al.* 1998) was used.

The nebulizer shears the DNA by forming small droplets that cause chain breakage at the droplet surface from the pressure of the DNA against the nebulizer wall, while in the hydroshear, "...breakage is caused by both shearing terms (when the fluid is inside the narrow tube or orifice) and the extensional strain terms (when the fluid approaches the orifice)" (Thorsten *et al.* 1998). Basically the narrowing of the tubing and increase in fluid speed produces a drag force that causes the DNA to stretch until it snaps. The "flow rate of the fluid and the size of the contraction determine the final DNA fragment sizes" (http://www.genemachine.com/hydroshear/hstech.html).

Normally, the point-sink shearing works by forcing DNA, typically 200-300 µl solution containing 50 ug of DNA, through a small diameter tubing using pressure with a HPLC pump. Here, a inexpensive syringe pump is used with a laser drilled hole, diameter between 0.0016-0.0030 of an inch, in a ruby. This syringe permits sample a small as 50 µl to be used. The sample is sheared smaller each time the solution is forced through the hole (Thorsten *et al.* 1998).

Using a nebulizer, cap the nebulizer cup, add 800 µl of the amplified DNA, 200 µl sterile double distilled-water and 1 ml of sterile glycerol (Use a new razor blade and cut the end of the pipette tip just enough to increase the opening, which will allow you to pipette the glycerol easier) and pipette up and down to mix. In an isopropanol-dry ice bath at a temperature of -20 °C, place the nebulizer for 5 minutes. Then, nebulize the solution for 2.5 minutes at 6-8 psi (depending on the size range you want). While nebulizing, tap the nebulizer against the bathing container, to keep the solution in the bottom of the nebulizer cup. Then the nebulizer is centrifuged briefly to collect the solution that is aliquated into 4 snap top 1.5 ml tubes.

Low Melt Gel

Electrophoresis on a low melting point agrose gel is used to separate the DNA. Here, the percent agrose determines the size of the holes in the matrix, the higher the percent the smaller the holes and the smaller the percent, the larger the holes. Typically, a 0.8 % SeaPlaque GTG agarose gel, over loaded with 0.1-1.0 ug of DNA solution of 50% glycerol/water with markers on either side of the samples, with an empty well between the sample and the marker, to prevent accidental contamination of sample when cutting. With the fan blowing on high across the electrophoreses machine, electrophorese at 100 mA for 30 minutes, then 150 mA for 1.5 hours. Excise the 2 to 4 kb region into snap top, Eppendorf, tubes. Incubate tubes at -80 °C for at least 30 minutes. After cutting the desired fragment length DNA, extract the DNA from the gel piece using the phenol/ether extraction and ethanol sodium acetate as described above and then dissolve the precipitated DNA in ~ 60 ul of sterile double distilled water.

Fill-in Kinase

The Klenow (fragment of *E.coli* polymerase I) fills in nucleotides in the 5' to 3' direction and removes unpaired bases in the 3' to 5' direction, while the T4 polynucleotide Kinase adds a phosphate to the 5' ends.

Make a pre-mix (for 4 tubes make for 5 reactions) in this order: 10x Kinase buffer (5 µl/reaction), 10 mM rATP (5 µl/rxn), 0.25mM dNTPs (7.5 µl/reaction), 5 U/µl Klenow (2 µl/reaction), and 3 U/µl T4 polynucleotide Kinase(1 µl/reaction). Then, pipette the pre-mix up and down 3-8 times as you add each item. Add the enzyme to the reaction pre-mix slowly, moving

from the bottom to the top in a circular motion pipette all up and down, gently, 3 times very, very briefly centrifuge down. Add ~20.5 to each of the 4 tubes, pipette up and down, centrifuge. Incubate in the 37 °C bath for 30-45 minutes. Then denature the enzyme by incubating the reactions in at least 80 °C for 15 minutes. Afterward, an ethanol sodium acetate clean up as described above is done. Then dissolve the DNA in 20 µl of sterile double distilled water.

Ligation

Ligase covalently bonds the 5' –phosphate and a 3' hydroxyl (Sambrook *et al.*1989). It is necessary to determine the concentration of DNA needed for efficient ligation. To small snap top Eppendorf tubes, add sterile double distilled water (DNA) in the amounts of 0.5, 1, 2, 4, and 6 along with the pUC, Alu I restriction endonuclease and cells only for controls as shown:

	6X	4X	2X	X	0.5X	Alu cleaved DNA	pUC	Cells-only
DNA	6	4	2	1	0.5	1	0	-
water	0	2	4	5	5.5	5	6	-

Table 5 Ligation Concentration Study

Then, on ice, collect the following items, and make a " mixture pot" (for 7 reactions include enough materials for 8 reactions) in this order: pUC (2 µl/reaction), Ligase Buffer (1 µl/reaction), and Ligase (1 µl/reaction). Pipette the pot up and down 3-8 times as you add each item. Add the enzyme to the reaction pot slowly, moving from the

bottom to the top in a circular motion, pipette all of the solution up and down, gently, 3 times. Then, very, very briefly centrifuge down and add 4 ul of the pot to each of the 7 tubes, pipette up and down, and centrifuge down. Next, let the ligation set at room temperature for 1 day, then move to the cold room for one day, before transformation.

Electrotransformation

A plate of cells with IPTG (isoprpylthio-β-D-galactoside) an inducer of B-galactosidase and X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactoside) is a chromogenic substrate. Neither the plasmid nor the host cells are Lac +, but can alpha-complement to form Lac +, obvious by blue coloring in presence of x-gal. This can be interrupted by the DNA being ligated in to the cloning site of the plasmid causing a white color in the colonies (Sambrook *et al.* 1989).

In ice bucket with ice place the cuvettes, competent cells and ligased DNA. Incubate YENB at 37 °C in a water bath to warm until time to transform. Label 5 ml culture tubes and set out antibiotic plates to dry the moisture. Set transformer to Ec2 in the window, the settings: bacteria light will be on. Wipe down work area and incubator with 10% bleach. Process: take 2 μl of ligation mix and add it to the competent cells, the XL1 Blue (MRF'). Then, transfer 40 ul, the cell + ligation mixture, to a cuvette adding it via a 45 degree angle. Tap the side of the cuvette to the bench to insure that the solution is spread evenly. Check for bubbles, if large bubble, tap cuvette hard, and if small bubble, many small tapes to remove them. Place the cuvette into the sliding arm and slide the arm into position, locking in place. Push the pulse button and the window will read PL5. The bacteria light will go off, until you hear a small beep and the bacteria light will come back on. If you hear a loud pop, the machine has just produced an arc (due to bubble or

high DNA conc.). After hearing the beeping noise and seeing the return of the bacteria light, remove cuvette and add 1 ml of YENB pump the pipette up and down 2 times to mix, then add the solution to the 5 ml culture tubes. Repeat until you have down 6 to 9 transformations, placing them in the 37 °C shaker container and placing them in this shaker for 30-40 minutes. Centrifuge at 3,000 rpm in the Beckman C56R centrifuge for 5 minutes, decant the supernatant and add 35 µl IPTG to the medium and 35µl Xgal (directly to the pellet) then flick the tube with your finger to mix and add 4 µl Amp. (Amplicillin) to insure plate antibiotic to each tube. Plate 100 µl the solution using a plastic bioloop, distribute the solution using the four corner technique. Let plate set to dry for 15 minutes, then place in the incubator for 16-20 hrs. After this time period, remove the plates from the incubator and survey the controls.

It is expected that the negative control, cells only, no colonies should grow on the Petri dish. However, if colonies do grow, then, possibly the competent cells or the antibiotic plates are contaminated, with an antibiotic resistant strain of bacteria (Sambrook *et al.* 1989). The other possibility is that the plates lack the antibiotic or it was added while the agar was too hot (Sambrook *et al.* 1989).

For the positive control plates, the pUC transformed cells should produce small blue colonies, confirming that the X-gal and IPTG were present. The Alu I cleaved DNA should produce numerous white colonies with maybe a few small blue colonies. If you have lots of colonies, the cells were intact, and both the ligase and ligase buffer were satisfactory. If no colonies for the project, then likely too little DNA was used or the fill-in kinase was not successful.

Finally, the non-Control colonies should be mostly white with a few blue colonies.

Next, incubate the plates in the cold room (4 °C) for 2 hours to intensify the blue color. Afterwards pick the colonies into 384-well flat bottom microtiter plate, NUNC plate, containing 80 µl TB + salt supplemented with 100 µg/ml ampicilin using the Flexys colony picker. Incubate the plates in a HiGro incubator (Gene Machines) for 22 hours at 37°C with shaking at 520 rpm, Beckman C56R centrifuge,. The oxygenated flow begins 3.5 hours after the shaking begins and flow setting at 0.5 second on and 0.5 minutes off. After the 22 hours are complete, replace the contents of wells with no cell growth with culture with growth cells obtained from an extra plate containing additional colonies picked into a partial plate. Centrifuge the 384 well plates at 3000 rpm, Beckman C56R centrifuge, in the Beckman CS-6R table top microtiter plate centrifuge for 10 minutes. Decant the supernatant by inverting the plates onto 3 layers of paper towels by centrifuging up and down to 300 rpm, Beckman C56R centrifuge,. Freeze the plates for at least 3 hours at -20 °C.

Shotgun Subclone DNA Sequencing Template Isolation

Similar to the large scale above, but with the use of the Zymark and the VPrep, robots as described at the URL: (http://www.genome.ou.edu/Zymark_384_well_isolation.html) for more specifics.

Remove the plates from the freezer and place them in a Zymark rack. The robotic arm picks four plates at a time and places them on four magnetic shakers inside the deck. A 384-channel pipettor head adds 23 µl of TE-RNase solution. The magnetic shakers

shake the plates for 10 minutes at 1000 rpm (a setting of 1). Then the robot head adds 23 µl of lysis buffer. The magnetic shakers shake the plates for another 10 minutes at a setting of 1. The 384 head adds 23 µl of 3 M NaOAc or 3 M KOAc. The shakers shake the plates for 10 minutes at a setting of 1. Then, the arm then moves the plates from the deck and places the in a storage rack. The rack is incubated at -80 °C freezer overnight. Next, thaw the plates and centrifuge at 3000 rpm, Beckman C56R centrifuge, for 45 minutes.

Using the VPrep, transfer 50 µl of the supernatant into a new 384-well plate and add 50 µl 100% isopropanol. Then 50 µl of air bubbles are used to mix the isopropanol. Centrifuge at 3000 rpm in the Beckman CS-6R centrifuge for 30 minutes. Decant the supernatant by inverting the plates onto 3 layers of paper towels by centrifuging up and down to 300 rpm, Beckman C56R centrifuge, and add 50 µl of 70% ethanol using the VPrep and centrifuge at 3000 rpm in the Beckman CS-6R centrifuge for 10 minutes. Decant the supernatant by inverting the plates onto 3 layers of paper towels by centrifuging up and down to 300 rpm in the Beckman CS-6R centrifuge and dry the pelleted DNA for 10 minutes in a vacuum and resuspend the DNA in 20 µl of water and shake on a bench-top shaker for 10 minutes at a setting of 8. This normally 5µul of this solution is equal to 200 ng, enough to sequence.

Sequencing

Normal sequencing reactions were done using ~80 ng of DNA from pUC (400 ng of Fosmid and 1600 ng of BAC) using forward or reverse primers at 6.5 µM and 1:20 dilution of either the ABI BigDye terminator DNA sequencing reaction mix v1.0, 2.0,

3.0 or Amersham ET terminator DNA sequencing reaction mix. Cycle sequencing reactions were ran on a Perkin Elmer GeneAmp PCR system 9600 thermal cycler : 95 °C hold for 5 minutes; 60-99 cycles of 95 °C for 30 seconds, 50 °C for 20 seconds, 60 °C for 4 minutes: 4 °C hold to infinity. Clean up was done using an ethanol precipitation, explained above or running the samples through a Sephadex G-50 microtiter plate centrifuge columns to remove unbound BigDye molecules.

PCR

In order to close gaps, primers were chosen off the sequence ends using PrimOU. PCR reactions included: 2µl (15 ng or less) of the target DNA, 3 µl (40 µM) of each primer, forward and reverse, 2 µl (5U) AmpliTaq DNA polymerase, 5 µl dNTP (2 mM of each dNTP), 5 µl 10x PCR Buffer (500 mM KCl, 100 mM Tris-HCL, pH 7.6), 3 µl $MgCl_2$, and 27 µl of sterile double distilled water. The thermocycling conditions: 95 °C hold for 2 minutes; 35 cycles of 94 °C for 1 minute, 50 °C for 2 minutes, 72 °C for 3 minutes: 72 °C hold for 10 minutes , 4 °C hold to infinity.

For clean up use 10 % enzyme (50 µl reaction use 5 µl of enzyme) of Shrimp Alkaline phosphatase (1U/µl), removes the 5' phosphates and Exonuclease III (10 U/µl), cleaves nucleotides. The enzymes are added straight to the reaction, no buffer. Incubate at 37 °C for 45 minutes and 80 °C for 15 minutes (to denature the enzymes) (Sambrook, 1989).

The PCR products are sequenced using special chemistries based on the type of gap, as given in refer to table 3. Normally, the PCR products were sequenced using 4 µl of DNA of the above, 2 µl of 6.5 uM primer and 1/8 BigDye or ET terminators

Non-Standard Isolation Methods

To investigate the of development improved methods for DNA amplification, the following enzymes were investigated; TempliPhi, phi-29 polymerase, (Dean *et al.* 2001), *Bacillus stearothermophilus* DNA polymerase I Klenow fragment, and a modified Phi-29 amplification mix (GenomiPhi, which is TempliPhi with additional nucleotides). The following protocols, seen below, were developed in this investigation. Originally, all investigations began with 1-10 nanograms of DNA, later colonies and finally glycerol stock was amplified. The DNA types included plasmid, fosmid and BAC libraries. Also, different types of plates were used for amplification, initially tubes, then 96 well plates and finally 384 well plates were used producing different protocols.

Starting with the original published protocol, that accompanied the enzyme mix (TempliPhi), for plasmid libraries and amending it as indicated below the following protocols (below) were developed.

In the beginning, it first was necessary to remove an unknown cellular inhibitor that was accompanying the DNA to the sequencing reactions and inhibiting the sequencing enzyme. This was done by first replacing the sample buffer that accompanied the enzyme and contained EDTA that inhibits polymerases, with sterile double distilled water. Second, by adding a freeze-thaw-centrifugation, a rediluting step that followed the colony picking after heating of the cellular solution, and the use of only 10 % of the volume in subsequent reactions. Third, it was necessary to clean the amplification product without substantial loses. Here, a simple 3 M sodium acetate ethanol precipitation was insufficient, it was necessary to replace the sodium ion with potassium and more than double the molarity of the salt solution. When working with BACs, it was

necessary to find a method to adequately break the DNA, a method of salt and heat shearing was used as well as liquid Nitrogen. Later, after the initial Phi-29 testing, with the Bst and the Genomiphi (phi-29 revisited) diluting the DNA, shearing it with vigorous vortexing and then passing it through a Sephadex column to produce DNA that was free of contaminants that inhibit the DNA sequencing reaction.

Finally, the amount of enzyme mix necessary to produce adequate amplification of DNA was optimized. To make the enzyme mix more economical, the enzyme used was greatly reduced from the amount in the protocol accompanying the enzyme mix. In addition, a new dilution buffer was used with additional dNTPs and equal volumes of DNA and enzyme mix were used.

Phi-29 (TempliPhi) Methods

Plasmid-Original: Phi-29, the original protocol from Amersham (Phi-29 for plasmid-based DNA amplification)

The Phi-29 mix needed one nanogram of isolated pUC + insert from in 1 µl of DNA, or one colony picked with the colony picker were to be added to 9 µl of buffer and heated 95 °C for three minutes, cooled and 10 µl of Phi-29 mix (5 U) was added to the reaction. The reaction was mixed and incubated at 30 °C for 6 hours. It should produce at least 1 µg (probably 2-3 µg). Then, after vortexing/mixing the amplified mixture, one needed 1-4 µl for each sequencing reaction.

Bst Methods

The Klenow fragment of Bacillus stearothermophilus DNA polymerase I (BST) comes from New England Biolabs, the Bst 8000 units, catalog #M0275L, $220.00 and the buffer, catalog# B9004s, $10.00. The buffer comes as a 10x concentration, and it is diluted down to 5X for use in the amplification reactions. (Reaction Buffer at 1X contains: 1X ThermoPol Buffer [20 mM Tris-HCl (pH 8.8, @ 25°C), 10 mM KCI, 10 mM $(NH_4)_2SO_4$, 2 mM $MgSO_4$, 0.1% Triton X-100]). (Kong *et al.* unpublished, product literature)

Phi-29 vs. Bst, 96 vs. 384 well thermocycler, Plasmid vs. BAC, Isolated vs. Colony vs. Gycerol Stock DNA Protocol Comparision

(The complete separate protocol for each of the above are in Appendix IV for the Phi-29 and Appendix V for the Bst.)

The following protocols are numbered and titled based on what is used, a whole protocol would include: 1, 2, 3, 4, and 5, e.g. say: a "1" using colonies, a "2" using the Bst mix, a "3" transfer DNA, a "4" Bst main protocol, and a "5" desalt and sequence protocol. If you were using the Phi-29 or isolated DNA, some of the numbers would change and some would not.

1-If Isolated 96 well thermocycler for Plasmid:

Using the Hydra, add 20 µl of sterile double distilled water is added to a 96 well thermocycler plate. To prevent evaporation, between all steps cover and keep plates on ice. Add 2 µl of isolated DNA (~1 ng), then centrifuge up and down. Next, cover plate with thermocycler sealer and heat at 95 °C for 3 minutes.

1-If Isolated 384 well thermocycler for Plasmid:

Using the Hydra, add 20 µl of sterile double distilled water is added to a 96 well thermocycler plate. To prevent evaporation, between all steps cover and keep plates on ice. Add 2 µl of isolated DNA (~1 ng), then centrifuge up and down. Next, cover plate with thermocycler sealer and heat at 95 °C for 3 minutes.

1-If Isolated 96 well thermocycler for BAC:

Using the Hydra, add 20 µl of sterile double distilled water is added to a 96 well thermocycler plate. To prevent evaporation, between all steps cover and keep plates on ice. Add 2 µl of isolated DNA (~10 ng), then centrifuge up and down. Next, cover plate with thermocycler sealer and heat at 95 °C for 3 minutes.

1-If Isolated 384 well thermocycler for BAC:

Using the Hydra, add 20 µl of sterile double distilled water is added to a 96 well thermocycler plate. To prevent evaporation, between all steps cover and keep plates on ice. Add 2 µl of isolated DNA (~10 ng), then centrifuge up and down. Next, cover plate with thermocycler sealer and heat at 95 °C for 3 minutes.

1-If Colony DNA from Plasmid:

Use the Hydra and add 20 µl of sterile double distilled water to a 96 well thermocycler plate. To prevent evaporation, between all steps cover and keep plates on ice.

If using colonies for the template: Use the colony picker (not by hand!) to pick colonies into a 384 well thermocycler (the Qpix II robot will not pick into a 96 well thermocycler plate) plate from a shotgun library petri dish or 12x12 square petri plate using the Qpix II robot (~3-5 dips). Cover plate with thermocycler cover and heat at 95 °C. for 3 minutes. Cover plates with aluminum cover and freeze the plates ~2-3 hr. in freezer until solid. Thaw ~10 min, until completely thawed, and centrifuge 45 minutes at 3200 rpm, Beckman C56R centrifuge.

1-If Colony DNA from BAC:

Single BAC colonies are obtained by diluting a smear of BAC colonies from the distribution Petri dish over the range 1:10 through 1:108. This is done by placing a bioloop swipe of the BAC colonies into 1000 µl LB+antibiotic in a 12x75 mm culture tube (VWR # 60818-565) and then adding 100 µl of the 1000 µl 900 µl LB+antibiotic and vortexing 1 sec. (for a 1:10 dilution in tube 1), and then repeating this 7 times to give 1:102 dilution in tube 2, give 1:103 dilution in tube 3, give 1:104 dilution in tube 4, give 1:105 dilution in tube 5. etc.

Plate 100 µl of tubes 5-8 into each of 3-4 plates/dilution, on LB plates containing the correct antibiotic, typically Kan or Cm (not Amp), and incubate 16-18 hrs. Use the Hydra and add 20µl of sterile double distilled-water to the 384 well thermocycler plate(s), to prevent evaporation, between all steps cover and keep plates on ice. Use the colony picker (not by hand!) to pick colonies into a 384 well thermocycler (the Qpix II robot will

not pick into a 96 well thermocycler plate) plate from a shotgun library petri dish or 12x12 square petri plate using the Qpix II robot (~3-5 dips).

1-If Gycerol Stock DNA:

Use the Hydra and add 20 µl of sterile double distilled water to a 96 well thermocycler plate. To prevent evaporation, between all steps cover and keep plates on ice.

If using the glycerol stocks for template: Use the 12 channel pipette and scrap the well lightly, just enough to get a small amount of ice on the tips of the 12 channel. Place the 12 channel in the well and pump the pipette 8 times. Cover plate with thermocycler cover and heat at 95 °C. for 3 minutes. Cover plates with aluminum cover and freeze the plates ~2-3 hr. in freezer until solid. Thaw ~10 min, until completely thawed, and centrifuge 45 minutes at 3200 rpm, Beckman C56R centrifuge.

1-If Glycerol Stock DNA- fresh:

This protocol was included because it worked very well, by staring with colonies, picking and growing the cell to make the glycerol stock fresh, the DNA amplified very well and sequenced very well.

To LB medium add the antibiotic [ex. subclone amp (5mg/ml) resistant: LB 1L-20 ml amp; LB 500ml -10 ml amp; LB 250- 5ml amp] To another sterile container add LB and Freezer mix [ex. LB 900ml - 100ml freezer mix; LB 450 ml- 50ml freezer mix; LB 225ml - 25ml freezer mix] LB medium [for 1 L : 10.0g NaCl, 10.0 g Difco (0123-01-1) Bacto-Typtone, 5.0 g Difco (0127-05-3)Bacto-Yeast Extract and autoclave at least 20 minutes at 121°C.] 10X Freezer mix, Dr. Doris Kupfers' recipe, to make 1 L :[62.7 g K_2HPO_4, 17.96 G KH_2PO_4, 5.0 G Sodium Citrate, 0.98 g $MgSO_4$-

$7H_2O$, 8.98 $(NH_4)2SO_4$, 440 ml of 44%Glycerol bring to 1 L with sterile double distilled water and filter through a 0.2um filter]

Add 80 µl of LB + antibiotic + freezer mix to a 384 microtiter plate by V- prep or 12 channel and use the colony picker (5 dips) to inoculate the plates. Grow the plates over night in the Higro ~22 hours. Remove the plates and put in them in the -80 degree C freezer until completely frozen ~3-4 hours. Remove the plates from the freezer and allow them to completely thaw ~ 1 hr. Transfer with the Hydra, 2 µl from the media plate and add it to 10 µl sterile double distilled water in a thermocycler plate (96 well). Cover and heat the plate in the thermocycler at 95 oC for 3 minutes. Cool to 4 oC and then, pull with the hydra, 2 µl of the media + water solution and add it to another thermocycler plate (96 well).

2-If using Phi-29 mix, no buffer dilution:

Remove the correct number of phi-29 mix plates, 96 well thermocycler plates from the -80 freezer that contained 2 µl/well of phi-29 mix or predispenced enzyme mix tubes with 100-500 µl/tube and place them on an ice water bath in the cold room (4 °C) to thaw ~10 minutes. This thawing time should not be exceeded, because if it is the enzyme will produce a significant level of nonspecific DNA.

2-If using Phi-29 mix, buffer B dilution:

In order to make the Phi-29 mix reactions more economical a buffer was developed and used to dilute the enzyme mix and custom synthetic primers were added as well as more dntps. The buffer worked so well, it could lower the enzyme needed to 1/80th or (2 µl needed = 1/40th pre reaction) the manufacture recommended amount

[Normally 5 U/10 µl mix/reaction-(Dean 2001)]. The buffer and the protocol of the initial experiment are listed below.

Phi-29 mixed Buffer B: 50 mM Tris-HCl pH 7.6, 10 mM MgCl2, 10% v/v Glycerol, 1 mM DDT, ~48.8 µM Universal Forward primer, ~48.8 µM Universal Reverse primer, 50 mM dntps (for each dNTP). (This same buffer was made with other primers as testing required)

Take one tube and add 200 µl (0.2 g/L) pGem (or other isolated DNA) and add 200 µl sterile double distilled water. Heat to 95 oC for 3 minute and cool down to below 30 oC. In the cold room on ice: Take another small tube and add 50 µl phi-29 enzyme mix (25 U) and 150 µl of the above buffer = 0.25 dilution. Take a third small tube and add 25 µl of phi-29 enzyme mix (12.5 U) and 175 µl of the above buffer = 0.125 dilution. Touch all 3 tubes to the vortex machine to mix and add 2 µl of the DNA solution to two 96 well thermocycler plates with a repeat pipette. Add 2 µl of the 0.25 enzyme dilution to one plate and 2 µl of the 0.125 enzyme dilution to the other plate with a repeat pipette

2-If using Phi-29 mix, buffer A dilution:

Phi-29 mixed Buffer A without primers: 50 mM Tris-HCl pH 7.6, 10 mM MgCl2, 10% v/v Glycerol, 1 mM DDT, 50mM dntps.

Next, prepare the Phi-29 enzyme-only reaction: 2 µl of Phi-29 enzyme (250ng) polymerase; 3 µl mix of forward (each primer 5 nM/40 µl); 3 µl mix of reverse primer (each primer 5nM/40 µl); 5 µl of 2mM DNTPs; 5 µl of above Phi-29 Buffer B~30 µl sterile double distilled water.

2- If using Bst mix, Plasmid:

Next, prepare the Bst mix: 1 µl of Bacillus stearothermophilus DNA polymerase large fragment is 8 units, use according to the dilution wanted (range: 0.5 µl = 4units, 0.25 µl = 2 units, 0.125 µl = 1 unit, 0.062 µl = 0.5 units, 0.031 = 0.25 units); 3 µl mix of forward (each primer 5nM/40ul); 3µl mix of reverse primer (each primer 5nM/40 µl); 5 µl of 2mM DNTPs; ~33 µl (based on dilution of enzyme) of 5x Reaction buffer (~3.5 X reaction buffer – total dilution).

Bst mix: 1 µl of Bacillus stearothermophilus DNA polymerase large fragment is 8 units, use according to the dilution wanted: range per 96 well plate: (add 2 µl per well/per reaction)]

Units of Bst pol /Reaction	Volume of Bst (µl)	Volume of 5X Buffer µl)	Total Vol. (µl)
8	110	110	220
4	55	165	220
2	27.5	192.5	220
1	13.7	206.3	220
0.5	6.8	213.2	220
0.25	3.4	216.6	220

Table 6 Bst Reaction Unit and Buffer Vol.

2- If using Bst mix, BAC:

Reaction mix contains: 3 µl of forward (primer 200 µM), 3 µl of reverse primer (primer 200 µM), 5 µl of 2 mM DNTPs, 35 µl 5x Reaction buffer (50 % 10x Buffer and 50% sterile double distilled water) (~3.5 X reaction buffer – total dilution).

Bst mix contains: 13.7 µl of Bst enzyme and 206.3 µl of 5x Reaction buffer. [1 µl of Bacillus stearothermophilus DNA polymerase large fragment is 8 units, use according to the dilution wanted: range per 96 well plate: (add 2µl per well/per reaction)] (Currently, the amount used is 1 unit per reaction to get whole BAC amplification.)

2-If using Bst Test control protocol for BAC-based DNA amplification and from isolated DNA

When experimentally working with Bst enzyme to confirm amplification, the volume of the reaction is 4 µl, the partial extraction would complicate the amount of dilution to be used later also after cleaning and diluting the DNA one can not see the DNA on a gel, If one could see the DNA, it would be too concentrated to sequence. Therefore, this control should be used.

Use Isolated DNA, (diluted) BAC, pUC and/or Fosmid DNA. Based on what is easier place test DNA in a few wells of a 96 well thermocycler plate or in a tube or tubes add 20 µl of sterile double distilled water To prevent evaporation, between all steps cover and keep on ice. Add 2ul of isolated DNA (~10 ng), then centrifuge, Beckman C56R, up and down. Next, cover plate with thermocycler sealer or snap top closed on tube and heat at 95 oC for 3 minutes.

Make a reaction pot using primers (vector specific) made for each type of DNA (example: BAC primers for BAC DNA) Reaction mix contains: 3 µl of primer mix (~24 pairs) forward primer (5nM/40µl), 3 µl of primer mix (~24 pairs) reverse primer (5nM/40µl), (5 µl of 2mM DNTPs, 35µul 5x Reaction buffer (50 % 10x Buffer and 50% sterile double distilled water) (~3.5 X reaction buffer – total dilution).

Next, add 2ul of the Bst (16 units) per well with a repeat pipette and cover with thermocycler sealer. Centrifuge for a 1000-rpm, Beckman C56R centrifuge, up and down. Then, place the plate in the thermocycler for X number of hours [depending on the amount of DNA you want, (for first time use, I recommend 12 to 18 hours to see what the amplified products look like and how well they amplify, then drop the time)] at 65 oC. with the volume setting of 50 µl.

After the set hours is finished, and then heat the thermocycler up to 95 oC for 5 minutes to denature the enzyme and cool to 4 oC before removal to prevent evaporation.

3-If using isolated DNA, Colony, Gycerol, addition to enzyme:

Use the Hydra and transfer 2 µl of the supernatant to the enzyme mix plate or place the 2µul DNA solution into a new 96 well thermocycler plate and add 2µl of the enzyme mix with a repeat pipette to the side and knock down, then cover with thermocycler sealer.

4-Main Phi-29 protocol:

Vortex the plate gently on a flat top vortex machine moving the plate in a circular motion for 8 seconds to mix. Next, place the DNA + enzyme plate in an ice water bath and transport to the centrifuge for a 1000 rpm, Beckman C56R centrifuge, for 1-2 seconds to concentrate the reaction mix in the bottom of the plate. Then, place the plate in the thermocycler for 12 hours at 30 °C with the volume setting of 5 µl. After the 12 hours, heat the thermocycler to 95 °C for 5 minutes.

4-Main Phi-29 protocol if for BAC Shotgun Cloning:

Cover plate with thermocycler cover and heat at 95 oC. for 3 minutes. Cover plates with aluminum cover and freeze plates ~2-3 hr. in a -20 freezer until solid. Thaw

~10 min, until completely thawed and centrifuge, Beckman C56R, the picked cell suspension for 45 min. in the Beckman GPR centrifuge, Beckman C56R, to pellet cell debris. While waiting, approx. 10 min. before centrifuge is finished, obtain the correct number of phi-29 plates from the -80 oC freezer (already made with 2 µl/well) or predispenced enzyme tubes with 500 µl/tube and place in ice water bath to thaw, ~ 5 min., to thaw the enzyme (Note 1: The enzyme is unstable if stored at 4 deg C for hours. Note 2: To prepare the microtiter plate with 2 µl/well, distribute 50 µl of the thawed, bulk, stock phi-29 mix into a "stock" microtiter plate using a repeat pipettor. This plate should be thawed on an ice water bath, takes ~30 minutes with periodic vortexing, and then the 2 µl aliquats are distributed to the microtiter plates using the hydra.) Once the 45 minutes of centrifugation is finished, immediately use the Hydra in the cold room and withdraw 2 µl of the upper portion of the supernatant from each centrifuged cell plate and add it to the enzyme mix plate or place the 2 µl cellular solution into a new 96 well thermocycler plate and add 2µl of the enzyme mix with a repeat pipette and knock down, cover with thermocycler sealer and vortex gently for 8 seconds. Put DNA + enzyme plate in ice water bath and transport to the centrifuge for a quick 1000 rpm, Beckman C56R centrifuge, up and down spin. Then, place the plate in the thermocycler for 12 hours at 30 oC. with the volume setting of 5 µl. After the 12 hrs is up, heat the thermocycler up to 95 oC for 5 min. Add 16 µl of sterile double distilled water, vortex for 30 seconds and centrifuge, Beckman C56R,up and down. Pool the first half of the plate (A1-12. B1-12, C1-12, and D1-12) into A1-12, pull 3 µl of DNA/well. Then run a 0.6% gel to check DNA. Pool A1-12 in to one Eppendorf tube (gives ~800 µl) and the remainder into a second Eppendorf.(~ 130-180 µg/plate). Then, renature the DNA by placing the

Eppendorf tubes in a beaker containing 95 oC water and allowing the water (and thus the solution in the Eppendorf) to cool slowly to room temperature (~1 hour). This step will allow the single stranded DNA produced by the Phi-29 enzyme via the rolling circle replication to renature the DNA into a large, double stranded structure.

4-Main Bst protocol:

Next, add 2 µl of the Bst mix per well with a repeat pipette and cover with thermocycler sealer. Centrifuge for a 1000-rpm, Beckman C56R centrifuge, up and down and place the plate in the thermocycler for 12 hours at 65 °C with the volume setting of 5 µl. After the 12 hours is finished, heat the thermocycler up to 95 °C for 5 minutes. Using the hydra, add 16 µl of sterile double distilled water (total volume 20 µl) to each well, then vortex for 30 seconds add 2 µl of 7.5 KOAc to each well with the Hydra or repeat pipette. Knock or centrifuge, Beckman C56R,the KOAc down and vortex for 30 seconds, then add with the Hydra, repeat pipette, or 12 channel, 53 µl of 100% ethanol, vortex for 30 seconds. Centrifuge, Beckman C56R, at 3200 for 45 minutes, and then invert on a paper towel and centrifuge up to 400 rpm, Beckman C56R centrifuge,. Immediately, dissolve the DNA in the appropriate amount of sterile double distilled water.

5- If Isolated 96 well theremocycler for Plasmid, Desalt and Sequence:

Using the hydra, add 16 µl of sterile double distilled water bringing the total volume to 20 µl/well. Then, vortex for 30 seconds and add 2 µl of 7.5 KOAc to each well with the Hydra or repeat pipette. Centrifuge the KOAc down and vortex for 30 seconds, then add with the Hydra or 12 channel pipette, add 53 µl of 100% ethanol and vortex for 30 seconds. Centrifuge at 3200 rpm, Beckman C56R centrifuge, for 45

minutes, and then invert on a paper towel and centrifuge for 1-2 seconds at 400 rpm, Beckman C56R centrifuge, to remove any residual ethanol. Immediately, without drying, dissolve the DNA in 20-50 µl of sterile double distilled water with the Hydra or 12 channel pipette and vortex for 30 seconds, and centrifuge, Beckman C56R centrifuge, to collect the solution at the bottom of the wells. Check the amplification by electrophoresing ~3 µl of DNA on a 0.6% agarose gel. To sequence, use 2 to 4 µl of DNA with the ABI Big Dye 3 or Amershams ET mixes diluted to 1/12 - 1/16.

5- If Isolated 384 well theremocycler for Plasmid, Desalt and Sequence:

After the 12 hours is up, heat the thermocycler up to 95 oC for 5 minutes next, add 6 µl of sterile double distilled water (total volume 10µl) to each well, then vortex for 30 seconds add 2 µl of 50% of 7.5 KOAc to each well with the Hydra. Centrifuge, Beckman C56R,the KOAc down and vortex for 30 seconds, then add with the Hydra 12 µl of 100% ethanol, vortex for 30 seconds. Centrifuge, Beckman C56R, at 3200 for 45 minutes, and then invert on a paper towel and centrifuge up to 400 rpm, Beckman C56R centrifuge,. Immediately, dissolve the DNA in 12 µl of sterile double distilled water with the Hydra or 12 channel and vortex for 30 seconds, and centrifuge, Beckman C56R,up and down. Confirm amplification by running ~3 µl of DNA on a 0.6% agarose gel. To sequence, use 2 to 4 µl of DNA with Big Dye 3 (1/12 - 1/16 dilution).

5- If for BAC: Desalting amplified DNA and Direct Sequencing:

To clean the Bst reactions, take the 50 µl reactions and add 50 µl of sterile double distilled water to each. Vortex the plate for 30 seconds and centrifuge for a 1000 rpm, Beckman C56R centrifuge, up and down. Next, remove the residual water from the columns of the prepared the Sephadex G-50 millipore filter plates via centrifugation, then

add 30 µl to the top of each Sephadex column in a row from each sample with a 12 channel pipette. Tape the Sephadex plate on top of a clean 96 well thermocycler plate and centrifuge, Beckman C56R, at 1300-1400 for 2.5 minutes. Pool the sample wells to one well per sample.

To sequence the Bst reactions do a (1:10) dilution (2 of DNA /20 sterile double distilled water) of the DNA. The sequencing reaction contains: 2 µl DNA from the (1:10) dilution, 2 µl sterile double distilled water, 2 µl primer 200µM, and 2 µl ET dye. Thermocycling conditions: Hold-95 °C for 2min; Cycle-95 °C for 30 sec, 50 °C for 20 sec, 60 °C for 4 minutes; Hold at 4° C for infinity. Clean the reactions using the standard sodium acetate ethanol clean up.

5- If for BAC: Shotgun Cloning, Desalting the amplified DNA:

To precipitate after, diluting the DNA and heat/cooling to reanneal the DNA: Take the 2mls and redistribute 333 µl into 6 tubes. Next, add 34µul of 7.5 M potassium acetate and 901 of 100% ethanol to each tube, then vortex well ~30 sec and incubate on ice for 15 minutes. Centrifuge at 12,000 rpm, Beckman C56R centrifuges for 45 minutes in the cold room (4 °C), then, carefully, decant the ethanol-KOAc. Next, add 500 µl of 70% ethanol to each tube, mix by inverting the tube ~8 times, then, centrifuge at 12,000 rpm, Beckman C56R centrifuge for 45 minutes. Decant and dry the tubes in the Speedvac (no heat) for 15 minutes (no longer) and immediately, dissolve the fragmented DNA in the appropriate volume of sterile double distilled water/tube, vortex 30 sec.

Using the standard shotgun sequencing methods go forward starting with the nebulization of the above DNA.

Vol. of Supernatant from cells in 20 µl water	Water	Phi-29 mix	DNA produced (ug) = Wells / nebulization (50 ug)
2 µl	0 µl	2 µl	1.3-1.8 µg = 48
4 µl	6 µl	µ	6-9 µg = 10-12
5 µl	15 µl		13-18 µg = 5-6

Table 7 Phi-29 BAC Reactions and Yield.

Salt and Heat Shearing of BAC and Genomic DNA

Salt protects DNA against depurination, by stabilizing the helix, and by preventing autocatalysis, DNA is a polyphosphoric acid. (Marguet and Forterre, 1998)

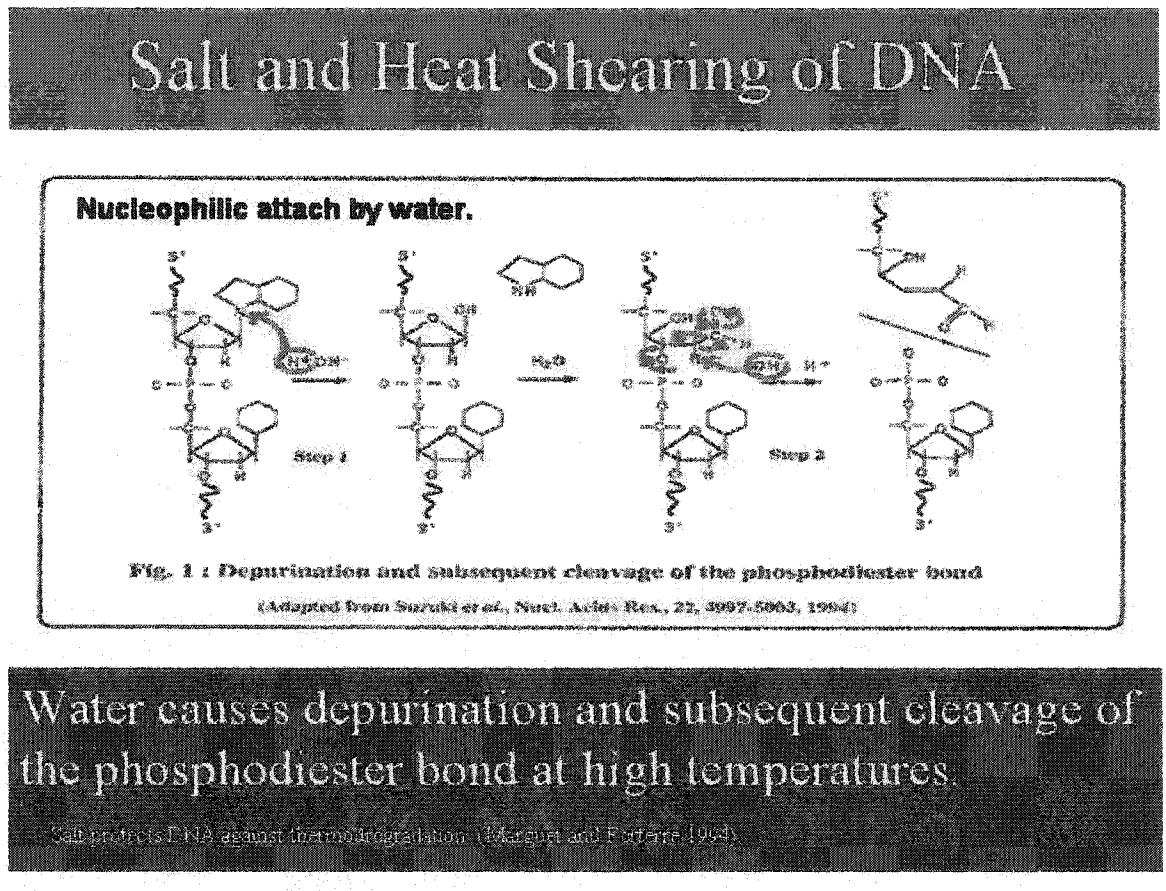


Figure 9 Salt and Heat Shearing of DNA

Take the amplified DNA and mix it with 0.1 M NaCl to make a 50% dilution, then heat the solution to 90 °C for 45 minutes to shear, and cool to 4 °C.

Shearing Genomic DNA and BAC DNA via Liquid Nitrogen Freezing

This application is the standard freeze and thaw technique for breakage. After amplifying the DNA, it is necessary to do the 95 °C denature for 5 minutes, because it assists or allows the breakage of DNA. After heating the samples and cooling to 4 °C, they should be transferred to 250µul Eppendorf tubes. Liquid nitrogen is then added to the tubes in a styrofoam cup ~3 ml just enough to cover the tubes; they should immediately freeze, and then place the tubes in the -80 °C freezer until ready to continue the experiment. When ready, place the tubes in an ice water bath in the cold room for ~1 hour to thaw slowly. This appears to promote the annealing of the single strands as well as the DNA breakage.

Phi-29 (GenomiPhi) Methods

Phi-29 (GenomiPhi)

The first experiments were done using the manufactures protocol even though most of their protocols, originated from this dissertation work. This is Amersham Biosciences. GenomiPhi DNA Amplification Kit Amersham Biosciences Corp. 2003 (http://www4.amershambiosciences.com/APTRIX/upp00919.nsf/Content/WD%3AGenomiPhi+DNA+A%28236845074-B500%29?OpenDocument&hometitle=WebDocs).

Phi-29 for plasmid, Fosmid, BAC based DNA amplification (Standard)

In a new 96 well thermocycler plate, mix 1 u of the DNA (10ng to just over this,

50 ng, if it is higher than this, dilute it down), pUC, Fosmid, BAC or whole Genome in 9 µl of sample buffer (contains random hexamers) and heat to 95 °C for 3 minutes to denature the sample and cool to 4 °C. Next, mix 1ul of the enzyme mix [Phi-29 polymerase, it only needs 2 minutes to thaw, don't retrieve it until ready to use and keep it in an ice water bath in the cold room (prepare the reaction in the cold room, once the enzyme is added to the plate keep the plate in an ice water bath for transport)] and 9ul of reaction buffer (salts and deoxynucleotides, adjusted pH.) and add to the DNA and sample buffer. Next, incubate the solution at 30 °C in a thermocycler for ~12 hrs, up to 18 hours won't hurt the DNA (don't use cycles for this time period use a 30 °C hold). After the allotted time, heat the samples at 65 °C to denature the enzyme. "10 ng of human genomic DNA should produce 4-7ug of DNA" (http://www4.amershambiosciences.com/APTRIX/upp00919.nsf/Content/WD%3AGenomiPhi+DNA+A%28236845074-B500%29?OpenDocument&hometitle=WebDocs).

Phi-29 for plasmid, Fosmid, BAC based DNA, cleaning and sequencing from an isolated template in a 96 well thermocycling plate

It was very difficult to produce a clean up procedure. This is because of the way the DNA is amplified, it is a knotted mess that has to be busted apart to accurately clean and read (OD) as well as sequence.

Dilute the DNA sample in 3 times the volume of sterile double distilled (sterile double distilled) water to each well. Next, pipette with a 12 channel each well up and down ~8 times, then vortex the plate for ~30 seconds on the highest setting (8) on a flat top vortexer moving the plate in a circle. Repeat (pipette with a 12 channel each well up and down ~8 times, then vortex the plate for another 30 seconds. Centrifuge up and

down to 1000 rpm, Beckman C56R centrifuge, cover the plate and heat in a thermocycler to 95 °C for 10 minutes and cool to 4 °C. Vortex the plate for 30 seconds and centrifuge up and down to 1000 rpm, Beckman C56R centrifuge. Next, remove the residual water from the columns of the prepared the Sephadex G-50 millipore filter plates via centrifugation, then add 30 µl to the top of each Sephadex column in a row per Sephadex plate [if 96 samples, then one well per plate (60 µl total of a diluted sample needs 2 Sephadex plates, 2 wells)] from each sample with a 12 channel pipette. Tape the Sephadex plate on top of a clean 96 well thermocycler plate and centrifuge, Beckman C56R, at 1300-1400 for 2.5 minutes. Next take an OD reading (already cleaned and busted apart, transfer 10 µl and add it to 1000 µl of sterile water (use the same for the reference) for the OD for a 100 dilution to determine the concentration of the DNA. To sequence the amount of DNA needed ~ [pUC (8,000 bp) = 80 ng ~1-4 µl; Fosmid (40,000 bp) = 400 ng ~2-6 µl; BAC (150-250, 000 bp) = 1600 ng ~8-14 µl)], 0.07-0.09 mM Primers, and 2 µl Et terminator dye mix. Thermocycling conditions: No hold, Cycle-95 °C for 30 sec, 50 °C for 20 sec, 60 °C for 4 minutes for 99 cycles; Hold at 4 °C for infinity. Clean the reactions using the standard sodium acetate ethanol clean up.

Phi-29 UV Clean–up

It was very difficult to produce a clean up procedure that produced accurate UV spectrophotometer readings. This is because the way the DNA is amplified, it is a knotted mess that has to be busted apart to accurately clean and read as well as sequence.

Pipette 10 µl of each amplified DNA to be read with the UV spectrophotometer and add 5 µl to 2 wells of a 96 well thermocycler plate, then 100 µl of water to each well.

Next, pipette with a 12 channel each well up and down ~8 times, then vortex the plate for ~30 seconds. Repeat (pipette with a 12 channel each well up and down ~8 times, then vortex the plate for ~30 seconds). Centrifuge up to 1000 rpm, Beckman C56R centrifuge, cover the plate and heat in a thermocycler to 95 °C for 10 minutes and cool to 4 °C. Vortex the plate for 30 seconds and centrifuge up to 1000 rpm, Beckman C56R centrifuge. Next, remove the residual water from the columns the prepared the Sephadex G-50 millipore filter plates via centrifugation, then add 30 µl to the top of each Sephadex column in a row from each sample with a 12 channel pipette. Tape the Sephadex plate on top of a clean 96 well thermocycler plate and centrifuge, Beckman C56R, at 1300-1400 for 2.5 minutes. Then, pool the wells to two sample wells and pool again, one sample/1.5 Eppendorf tube, then add water to the 1ml mark. Set the spectrophotometer to 260 nm and use sterile double distilled water for the reference. Use the zero time samples to subtract from the other samples (remove absorption of the enzyme etc).

Results and Discussion

During the initial phase of my studies, I completed the sequence of large insert clones containing genomic DNA from six regions of the human and mouse genome totaling over 1 Mbp as listed in table10. Since, this is only a minor part of this dissertation, only a brief look at the results based on the annotation programs and web options mentioned in the introduction will be done.

First, a Multipip overview to supply a visual comparison of the tabled information (explained in the next paragraph) will be given. Then, a table of the Human and Mouse clone, lastly, an interesting gene comparison from the comparative organisms will be shown using the visual output from Multipip or Vista.

Table 10 gives a list of the clones/genomes that were sequenced.

Clone	Vector	Source	Chr	Genbank Acc #	Insert Size	Consed_ Error/10 KB
p20k14	PAC	Human	22q11	AC006548	200,149	0.11
678g6	BAC	Human	22q11	AC007845	134,036	0.07
rp23-213m14	BAC	Mouse	16	AC090977	208,412	0.24
rp23-81d13b	BAC	Mouse	16	AC123977	212,251	0.01
rp23-7n16	BAC	Mouse	16	AC091002	217,939	0.04
rp23-46k8	BAC	Mouse	17	AC113265	143,758	0.02
Mit-1	Mito	*Danio Rerio*		AC024175	16,596	0.01
				Final Total =	1,133,141	

Figure 10 Sequenced DNA

The sequencing of the PAC p20k14 in figure 28 above was a confirmation of the original sequence done by Ray, L.A., Qi,S., Sloan,D., McDermid,H. and Roe,B.A.

PAC 20k14

Human chromosome 22q11 PAC 20k14 NIH accession # AC006548 contains two novel genes. Currently, there are no known comparative genes in other organisms for either gene, but a listing of these genes and their locations are given in Appendix I.

BAC 678g6

Human chromosome 22q11 BAC 678g6 NIH accession # AC007845 contains six novel genes, no gene number names exist at this time. Currently, there are no known comparative genes in other organisms for any of these genes, but the genes and their locations are given in the Appendix I.

BAC rp23-7n16

Mouse chromosome 16 BAC rp23-7n16 NIH accession # AC091002 contains seven genes that are listed in detail in Appendix I and include: Rtn4r, Ranbp1, Htf9c, Dgcr8, and three additional novel genes.

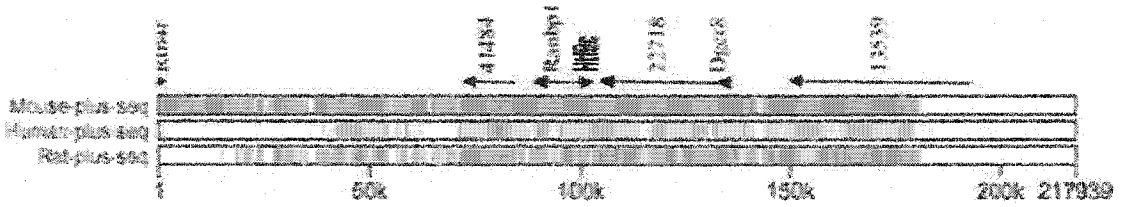


Figure 11 Pip Overview of BACrp23-7n16, based on the syntenic region of the Mouse genome.

The Mouse BAC is compared to the following in descending order as positioned on the Pip overview: Mouse-plus-sequence, the Human-plus-sequence, and the Rat-plus-sequence. The extra sequence (the whole gene not partial, 1353) was used encase of differences in gene position and to show where the BAC ends. The colors vary from green which represents highly conserved DNA to red which represents identical sequence. RESULTS: The rat sequence (red with a little green), is closer to the mouse sequence than to the human sequence (a little red with mostly green). Genes 41484 (nameless) and Ranbp1 appear to be the most conserved genes for all three organisms.

This mouse BAC also contains a partial gene, 13539 (nameless); therefore additional sequence was used beyond the BAC to compare these sequences. The BAC sequence ends where the red color ends on the Mouse–plus-sequence. The BAC sequence was the reference and is not listed. The rat sequence (red with a little green), is closer to the mouse sequence than to the human sequence (a little red with mostly green). Genes 41484 (nameless) and Ranbp1 appear to be the most conserved genes for all three organisms.

Gene Rtn4r is a reticulon 4 receptor precursor; it mediates the axonal growth inhibition and possibly regulates adult axional regeneration (Swiss Prot: Q99P18; Fournier *et al.* 2001).

Gene RANBP1 is a protein that binds to RAN (1) and activates GTPase (Swiss Prot: P34022; Coutavas *et al.* 1993).

Gene Htf9c produces HpaII. It is a SAM-dependent methyltransferase. Both RanBP1 and Htf9-c genes are regulated and share a bidirectional promoter (NM 008307; Guarguaglini *et al.* 1997; Di Matteo *et al.* 1995; Bressan *et al.* 1991).

Gene DGCR8 is a fragment of DiGeorge syndrome critical region 8 homolog (Swiss Prot: Q9EQM6; Gong and Yeh, 1999). The other genes without names are unknown genes.

Gene Htf9c

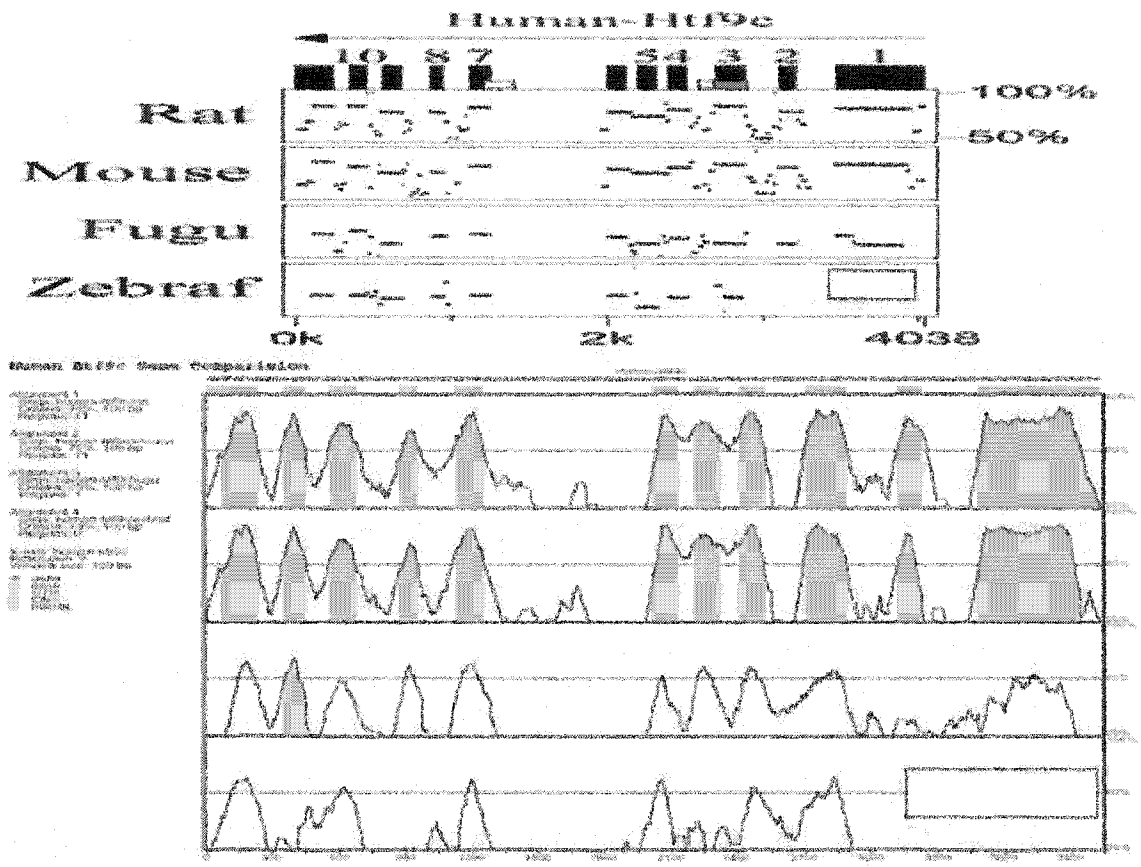


Figure 12 Pip (Top plot) Vista (Bottom plot) of Gene Htf9c, based on the syntenic region of the Human genome.

The Human Htf9c gene is compared to the following in descending order as positioned on the Pip and the Vista plots: Rat, Mouse, Fugu, and Zebrafish sequence. The Vista below is harder to read the sequences, but the positions are equal to the positions of the Pip above.
MULTIPIP (Top plot) KEY: The line with the arrow shows the gene and its direction. The black boxes with the numbers above the top line are exons, which correspond to the lines directly below in the boxes. Each box has a range of conservation to the human ranging from the 50% (bottom of box) to 100% (top of box).
VISTA (Bottom plot) KEY: The gray arrow shows the gene and its direction. The purple boxes above the top line are exons, which correspond to the lines, hills or valleys, directly below in the boxes. The purple color shown under the hills are conserved coding regions. Each box has a range of conservation to the human ranging from 20% (bottom of box) to 50% (middle line through the box) to 100% (top of box).
RESULTS: The Pip and Vista of the gene Htf9c above is based on the human gene. It is notable that all the exons (Pip: the box on top that is numbered is the exons, corresponding to the straight line below. The Vista: the purple box on top is the exons, corresponding to purple under the curve below.) are conserved in all the organisms except, exon 1 and 2 shown missing in position of the red box, which are missing in Zebrafish. It is also notable that in Fugu shows lower conservation of exon 1 (percent sequence agreement shown on the side based on position of the top (100%) or the bottom (50%).

The Pip and Vista of the gene Htf9c above is based on the human gene. It is notable that all the exons (Pip: the box on top that is numbered is the exons, corresponding to the straight line below. The Vista: the purple box on top is the exons, corresponding to purple under the curve below.) are conserved in all the organisms except, exon 1 and 2 shown missing in position of the red box, which are missing in Zebrafish. It is also notable that in Fugu shows lower conservation of exon 1 (percent sequence agreement shown on the side based on position of the top (100%) or the bottom (50%).

BAC rp23-213m14 and BAC rp23-81d13b

Mouse chromosome 16 BAC rp23-213m14 NIH accession # AC090977 and BAC rp23-81d13b NIH accession # AC123977 are over lapping and together contain four genes, Rik-1, Ephb3, and 2 novel genes. A list of these genes, the comparative organisms

(human and rat), their gene names and their locations is given in the Appendix I.

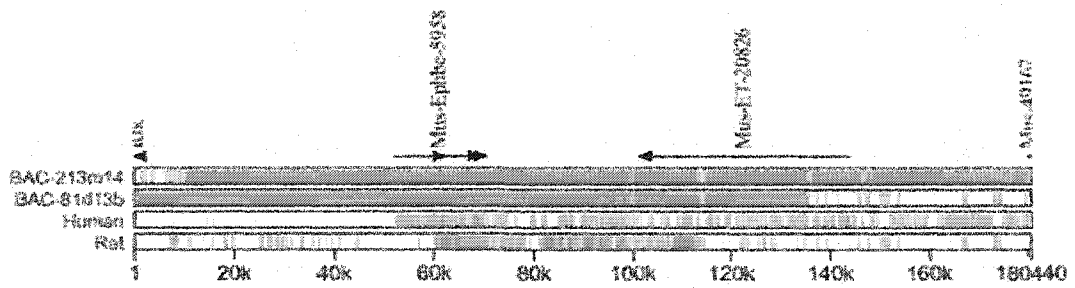


Figure 13 Pip Overview of BACrp23-213m14 and rp23-81d13b, based on the syntenic region of the Mouse genome.

The Mouse sequence containing the BAC s are compared to the following in descending order as positioned on the Pip overview: BAC-213m14, BAC-81d13b, Human, and Rat sequence. The first two are the BACs, which are overlapping, they are compared to the sequences encompassing this area in the human and the rat. This was used to show position of the BACs and the gene conservation. The colors vary from green which represents highly conserved DNA to red which represents identical sequence. The rat sequence (red) is closer in homology to the mouse sequence than the human sequence (green), and all 3 organisms show some conservation of the genes EPHBE-5958 and ET-208-26.

As can be seen from the Pip Overview in figure 13, the rat sequence (red) is closer in homology to the mouse sequence than the human sequence (green), and all 3 organisms show some conservation of the genes EPHBE-5958 and ET-208-26.

Gene EPHBE-5958 is an Ephrin type-B receptor precursor for a developmental kinase, SEK-4, and Tyrosine-protein kinase receptor MDK-5 (Swiss Prot: P54754; Ciossek *et al.* 1995; Becker *et al.* 1994; Imondi et al. 2000). The function of the other genes are unkown at this time.

Gene EPHB3

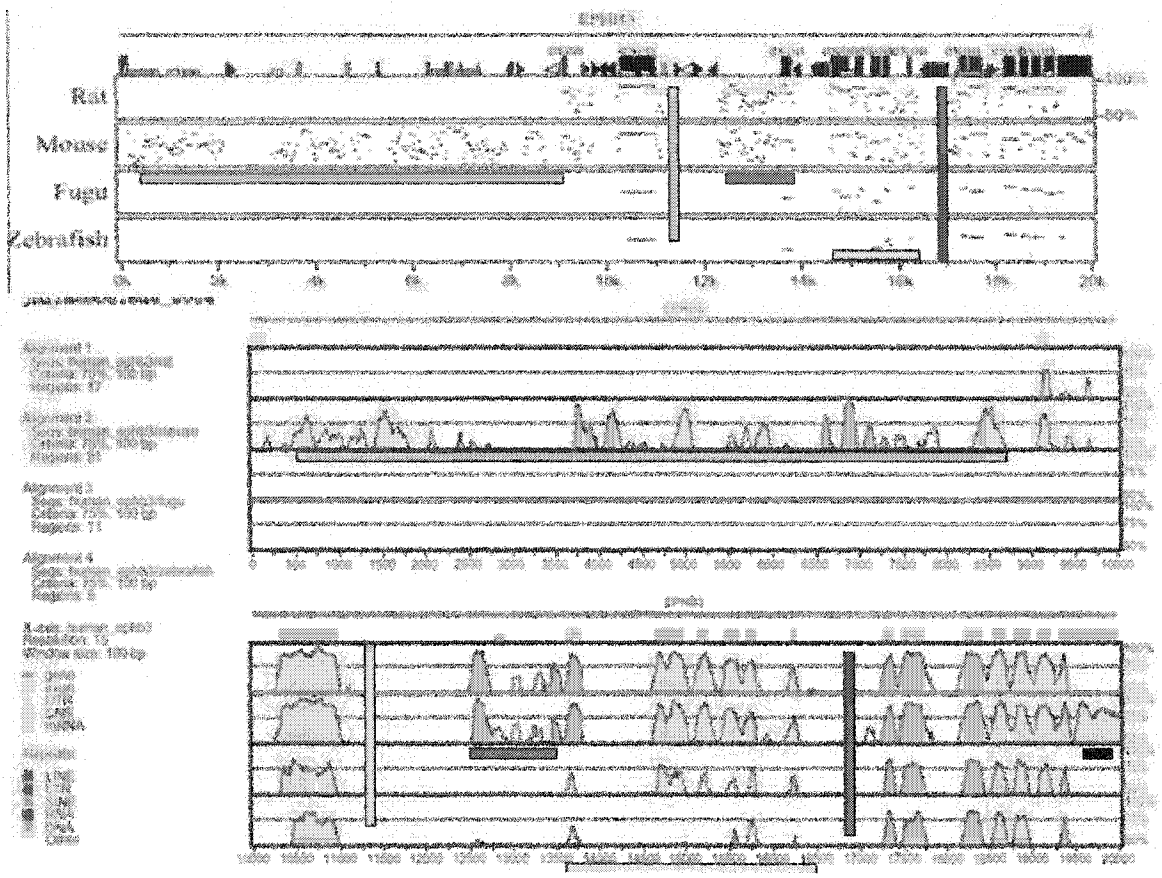


Figure 14 Gene EBH3 Pip (above) and Vista (below), based on the syntenic region of the Human genome.

The Human EPHB3 gene is compared to the following in descending order as positioned on the Pip and the Vista plots: Rat, Mouse, Fugu, and Zebrafish sequence. The Vista below is harder to read the sequences, but the positions are equal to the positions of the Pip above.
MULTIPIP (Top plot) KEY: The line with the arrow shows the gene and its direction. The black boxes with the numbers above the top line are exons, which correspond to the lines directly below in the boxes. The white pointed boxes (L1 repeats); Light gray triangles (SINEs), but not MIRs; Black triangles (MIRs); Black pointed boxes (LINE2s); Dark gray triangles and pointed boxes are other types of repeats, like LTRs and transposons; Short Dark gray boxes (CpG islands), where the CpG/GpC is greater than 0.75; Short White boxes (CpG islands), where the CpG/GpC is between 0.6 and 0.75. Each box has a range of conservation to the human ranging from the 50% (bottom of box) to 100% (top of box).
VISTA (Bottom plot) KEY: The gray arrow shows the gene and its direction. The purple boxes above the top line are exons, which correspond to the lines, hills or valleys, directly below in the boxes. The purple color shown under the hills are conserved coding regions. The peach color represents the conserved noncoding sequence. Each box has a range of conservation to the human ranging from 20% (bottom of box) to 50% (middle line through the box) to 100% (top of box).
RESULTS: The Pip and Vista of the gene EBH3 above is based on the human gene. In the beginning of this region, only the mouse shows conservation, because of the slight conservation in this non coding sequence (above pink boxes). However, both the mouse and rat show parallel conservation in the exon-containing sequence (above rust boxes). The third exon is conserved across all the

organisms, light green box and most of the last exons, behind the pink box. The very last human exon is only conserved in the mouse genome, above the black box.

The Pip and Vista of the gene EBH3 above is based on the human gene. In the beginning of this region, only the mouse shows conservation, because of the slight conservation in this non coding sequence (above pink boxes). However, both the mouse and rat show parallel conservation in the exon-containing sequence (above rust boxes). The third exon is conserved across all the organisms, light green box and most of the last exons, behind the pink box. The very last human exon is only conserved in the mouse genome, above the black box.

BAC rp23-46k8

Mouse chromosome 17 BAC rp23-46k8 NIH accession # AC113265 contains two genes, Park2 and a fragment from another gene. These genes, the comparative organisms (human, rat-1, rat-2, and fugu), their gene names and their locations are listed in Appendix I.

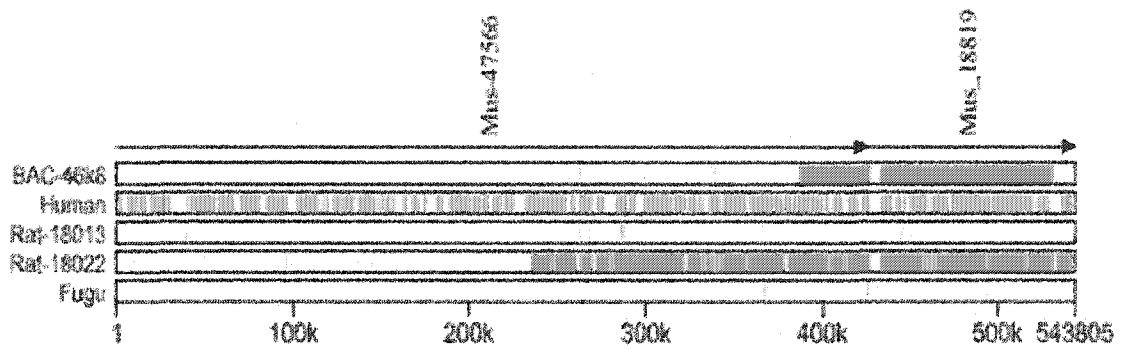


Figure 15 Pip Overview of BAC rp23-46k8, based on the syntenic region of the Mouse genome.

The Mouse sequence containing the BAC and all of the genes sequence, not the partial amount in the BAC is compared to the following in descending order as positioned on the Pip overview: BAC-46k8, Human, Rat-18113, Rat-18022, and Fugu sequence. The extra sequence (the whole gene not partial, Mus-47566) was used encase of differences in gene position and to show where the BAC ends. This was used to show position of the BAC and the gene conservation. The colors vary from green which represents highly conserved DNA to red which represents identical sequence.
RESULTS: The rat sequence (red) of course, is closer to the mouse sequence than the human sequence (green). The rp23-46k8 BAC portion is red. The rat gene18022 sequence is nearly identical for all of mouse gene 18819 and most of mouse gene 47566, while the rat gene 18013 appears to be only slightly conserved. The human is conserved throughout both genes. Fugu, the most distant organism shows only very slight conservation in the exonic regions.

The comparisons above are based on the mouse sequence for both whole genes. The rat sequence (red) of course, is closer to the mouse sequence than the human sequence (green). The rp23-46k8 BAC portion is red. The rat gene18022 sequence is nearly identical for all of mouse gene 18819 and most of mouse gene 47566, while the rat gene 18013 appears to be only slightly conserved. The human is conserved throughout both genes. Fugu, the most distant organism shows only very slight conservation in the exonic regions.

Gene 47566 is the parkin or park2 gene that degrades polyglutamine proteins and protects proteasome function. (NM 016694; Tsai *et al.* 2003) Mouse park2 is a homolog to the human Parkinson candidate gene (NM 016694; Tomac and Hoffer, 2001). The function of the second partial gene is unkown at this time.

Gene Park2

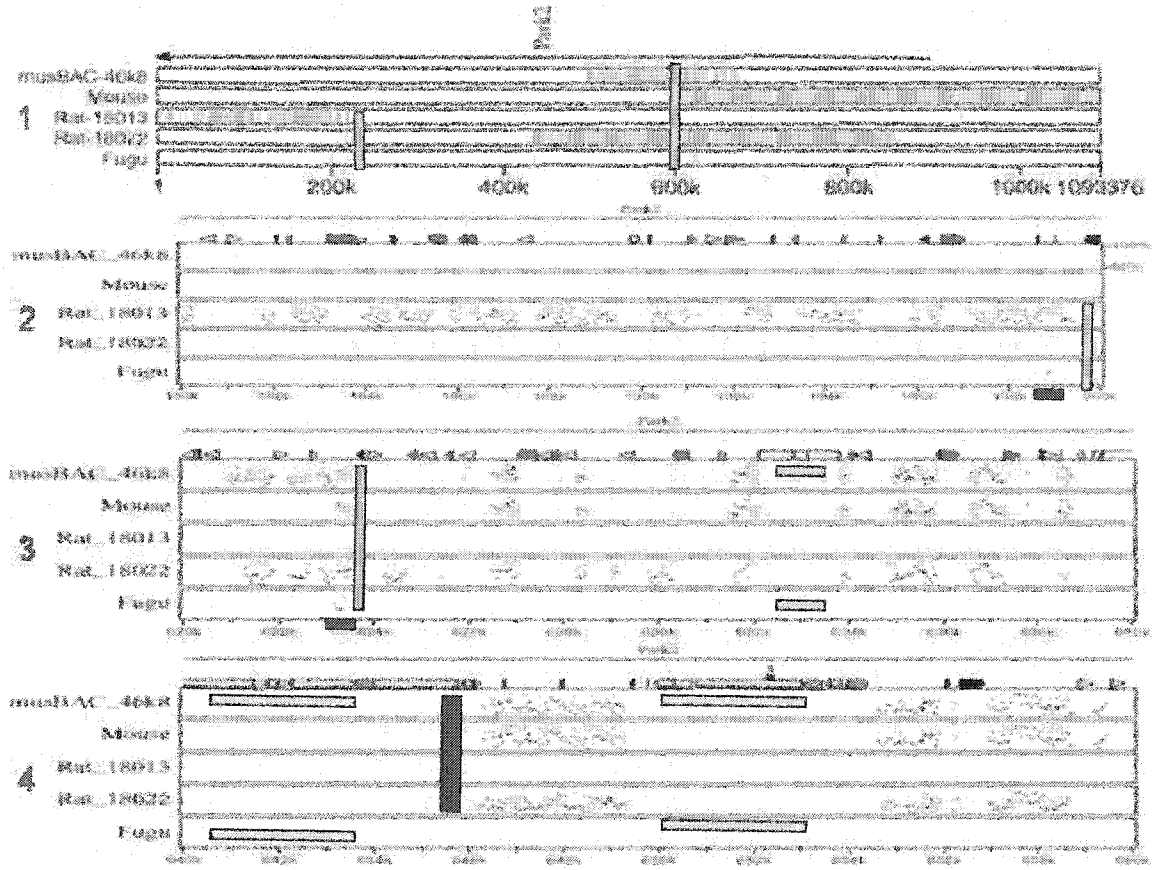


Figure 16 Park2 Gene: Pip Overview (1), Early Pip (2), Middle Pip (3) and Late Pip (4), based on the syntenic region of the Human genome.

The Human Park2 gene is compared to the following in descending order as positioned on the Pip plots: BAC-46k8, Human, Rat-18113, Rat-18022, and Fugu sequence.
PIP OVERVIEW (Top plot) KEY: The Mouse sequence containing the BAC and all of the genes sequence, not the partial amount in the BAC is compared to the following in descending order as positioned on the Pip overview: BAC-46k8, Human, Rat-18113, Rat-18022, and Fugu sequence. The extra sequence (the whole gene not partial, Mus-47566) was used encase of differences in gene position and to show where the BAC ends. This was used to show position of the BAC and the gene conservation. The colors vary from green which represents highly conserved DNA to red which represents identical sequence.
MULTIPIP(Bottom plot) KEY: The line with the arrow shows the gene and its direction. The black boxes with the numbers above the top line are exons, which correspond to the lines directly below in the boxes. The white pointed boxes (L1 repeats); Light gray triangles (SINEs), but not MIRs; Black triangles (MIRs); Black pointed boxes (LINE2s); Dark gray triangles and pointed boxes are other types of repeats, like LTRs and transposons; Short Dark gray boxes (CpG islands), where the CpG/GpC is greater than 0.75; Short White boxes (CpG islands), where the CpG/GpC is between 0.6 and 0.75. Each box has a range of conservation to the human ranging from the 50% (bottom of box) to 100% (top of box).
RESULTS: Rat 18013 gene sequence has sequence in the beginning that is similar to the Human. This is found in front of the light green box on Pip1 and 2. Rat 18022 gene has more sequence similar to what is found in the mouse. This is found behind the light green box on Pip 1 and 3. Fugu the most distant organism has a very small amount of sequence similar to the human in both areas.

This is found next to the light green box in Pip 1 and just above the red boxes in Pip's 2 and 3. It is notable, that no similar sequence in any of the organisms is found next to the L1 repeats, they are just above the light blue boxes. The majority of the sequence of course is found under the exons and CpG islands.

The Pip above shows the location of the Park2 gene incomparision to the Human. The mouse BAC 46K8 contained only a fragment of the Park2 gene the rest of the gene found below in the Pip labeled mouse. Rat 18013 gene sequence has sequence in the beginning that is similar to the Human. This is found in front of the light green box on Pip1 and 2. Rat 18022 gene has more sequence similar to what is found in the mouse. This is found behind the light green box on Pip 1 and 3. Fugu the most distant organism has a very small amount of sequence similar to the human in both areas. This is found next to the light green box in Pip 1 and just above the red boxes in Pip's 2 and 3. It is notable, that no similar sequence in any of the organisms is found next to the L1 repeats, they are just above the light blue boxes. The majority of the sequence of course is found under the exons and CpG islands as shown in figure 16.

The human sequences presented in this dissertation were published in Nature in 1999 (Dunham *et al.* 1999). The mouse sequences presented in this dissertation were used for the Initial Sequence and Comparative Analysis of the mouse genome that was published in 2002 (Mouse Genome Sequencing Consortium, 2002).

Mitochondria

The sequence of the maternally inherited human mitochondrial genome was completed 2 decades ago (Anderson, S. *et al.* 1981) and since then, numerous mitochondrial comparative sequencing studies have served as the basis for understanding

anthropology and evolution, as well as for forensic studies. An interest in forensics and the wide spread use of mitochondrial analysis in this field, led to the sequencing of an entire ~18 Kbp Zebrafish mitochondrial genome of the *euteleostei* taxon. This Zebrafish mitochondrial of *(Brachydanio rerio* - a model system for vertebrate developmental biology) was sequenced based on the whole genome PCR-shotgun approach [for more information on the method behind the isolation, PCR, sequencing and annotation, refer to (Broughton *et al.* 2001)]. The mitochondria was completed to a level of accuracy of greater than 99.999% (error rate of 0.01%) after closure and finishing using custom synthetic primer-based methods.

The annotation of the zebrafish mitochondria was done using a different technique as well as some of the standard ones. This annotation was based on the knowledge of the carp and the goldfish mitochondria, which are sister fishes to the zebrafish. These sequence annotations were compared using cross matching (Green, P. 1998). The beginning and ending of the coding regions of the zebrafish were determined by the single alignment of the zebrafish with each of its sister fishes and then all three sequences aligned via the Clustal W program version 1.80 multiple sequence alignment, (Thompson *et al.* 1994). The beginning and ending of the coding regions of the zebrafish were reconfirmed using tRNAscan and Blast. The tRNAscan program (Fichant and Burks, 1991) confirmed both the position and the orientation of all the tRNAs except tryptophan. The secondary structure was predicted on the tRNA sequences as were the rRNA sequences [NIP nucleotide interpretation program for tRNA (Staden, 1986)] and [Plotfold for rRNA use the suggested nucleotide sequence to predict their structures (Zuker, 1981)] and these structures are shown in Appendix III. The free energies of the

rRNA secondary structures are -335.6 kcal/mole for the 16S subunit and -220.3 kcal/mole for the 12S subunit (Broughton *et al.* 2001).

Next, each gene, based on the beginning and ending numbers of the coding regions predicted by the other two fish, were used to make a fasta file (Pearson and Lipman, 1988; Pearson, 1990) of the zebrafishes' DNA sequence using the Consed (Green, 1998) viewing program. These fasta files were then, "Blasted" (http://www.ncbi.nih.gov/blast/Blast.cgi) against the NIH database to confirm that it hit itself, the zebrafish mtDNA sequence genome item (for example, zebrafish mito NADH dehydrogenase subunit II) or the sister fish, other mtDNA from other organisms to confirm the sequence was not contamination and most importantly that each gene had Blast homology with the same gene from other organisms.

After these coding and D-loop regions were confirmed the annotation was formatted for upload to NIH Sequin program (http://www.ncbi.nlm.nih.gov/Sequin/sequin.hlp.html) that also assists in finding errors in the sequence annotation.

The zebrafish mitochondrial genome has 16,596 bases and encoded the typical 37 animal mitochondrial genes, 28 on the heavy strand and 9 on the light strand. The mitochondrial genome also contained the standard 2rRNA, 22 tRNAs, and 13 polypeptide encoding genes. It is 93% coding (the nuclear genome is only 3%) and has no introns. The mitochondrial genome has the standard ATP synthase F0 subunit 8 and subunit 6 overlapping as well as many others. The mitochondrial genome contains TAA stop codons, which are created through posttranslational polyadenylation (Ojala *et al.* 1981).

The D-loop (950 bp) is the noncoding control region that is the site of initiation for both the "heavystrand" replication and transcription (Clayton 1982; Clayton 1991; Shadel and Clayton 1997; Broughton *et al.* 2001). The D-loop contains the (O_L) or origin of light strand replication, which may form the perfect 11bp stem and a 14 bp loop secondary structures that act as initiation signals for replication (Wong and Clayton 1985; Broughton *et al.* 2001). The D-loop also contains the "Conserved sequence blocks (CSBs) 1-3, found in the 3' end of the control region, appear to be involved in positioning RNA polymerase both for transcription and for priming replication (Clayton 1991; Shadel and Clayton 1997)" (Broughton *et al.* 2001).

The zebrafish mitochondrial sequence was compared to other mitochondrial sequences using various programs. The ClustalW program (Thompson 1994) is a sequence alignment program that was used to show a mitochondrial comparison of one of the Cox's gene, see figure 17 below. The stars below the sequence show the nucleotides that are conserved. Genes' Cox I and Cox II are the most conserved of the different types of mitochondria.

ClustalW Example

Figure 17 ClustalW

The ClustalW program (Thompson 1994) is a sequence alignment program that was used to show a mitochondrial comparison of one of the Cox's gene.
RESULTS: The stars below the sequence show the nucleotides that are conserved. Genes' Cox I and Cox II are the most conserved of the different types of mitochondria.

The dotter (Sonnhammer, 1995) is a graphical dotplot program that helped to show the conservation between the different sequences and also helped assemble the Zebrafish mito DNA sequences, refer to figure below, which shows a comparison of the Zebrafish sequence against the Zebrafish sequence, Zebrafish sequence against the Carp sequence (sister fish), and Zebrafish sequence against the Coelacanth sequence (distant fish).

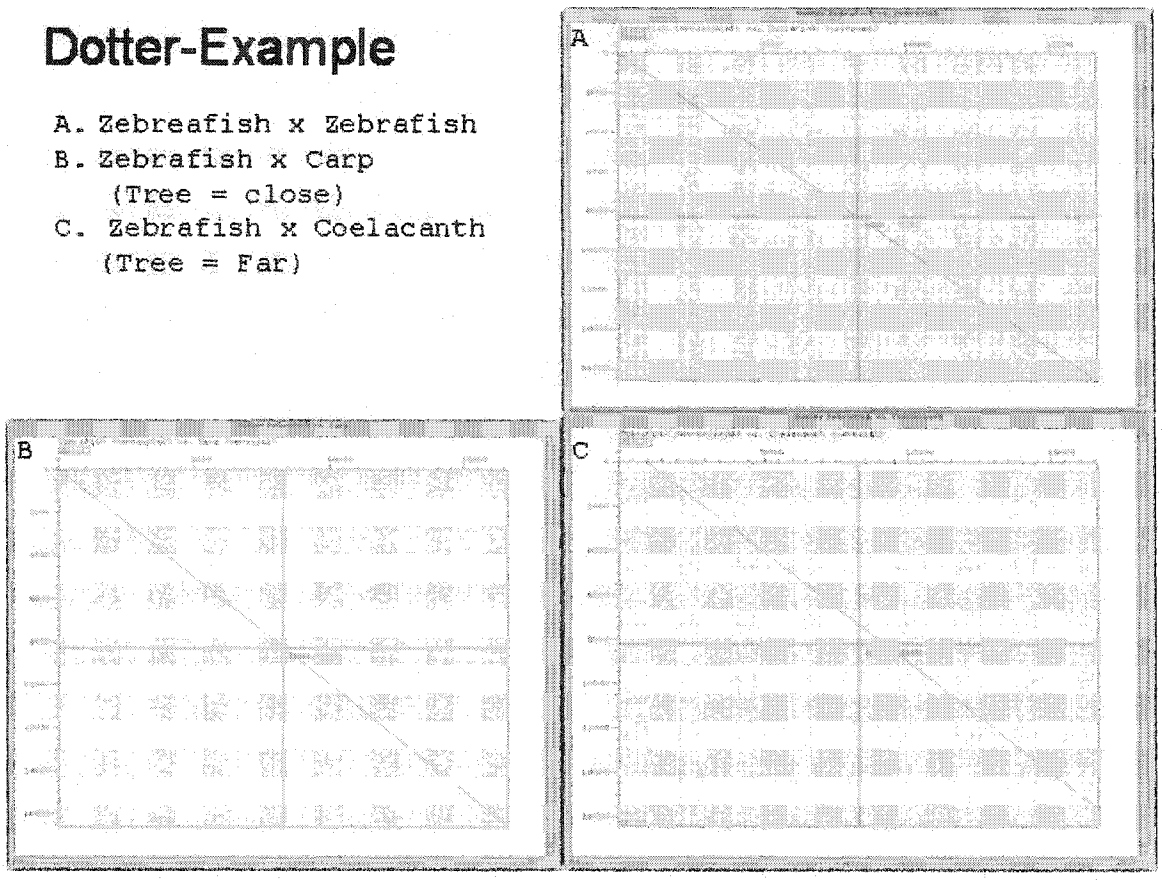


Figure 18 Dotter

Dotter (Sonnhammer, 1995) is a graphical program that compares the residuals from each sequence. One sequence runs along the x-axis, while the second runs along the y-axis. In areas, where the two sequences are very similar this is shown diagonally across the dot matrix. A comparison of the Zebrafish sequence against the Zebrafish sequence (control), Zebrafish mtDNA sequence against the Carp mtDNA sequence (sister fish), and Zebrafish sequence against the Coelacanth mtDNA sequence (distant fish).
RESULTS: The comparison of the Zebrafish mtDNA sequence against itself (A) produces a diagonal plot through the center, a dark line showing highly conservation. The comparison of the Zebrafish mtDNA sequence against the Carp mtDNA sequence (sister fish) plot (B) produces a diagonal plot through the center with a slight amount of line missing in the lowest quadrant, slightly less conserved than the (A) plot. The Zebrafish mtDNA sequence against the Coelacanth mtDNA sequence (distant fish) plot (C) produces a diagonal plot off center with a little amount of line missing in the lowest quadrant, less conserved than the (A) or (B) plots, which correlates with the tree distance.

The comparison of the Zebrafish mtDNA sequence against itself (A) produces a diagonal plot through the center, a dark line showing highly conservation. The comparison of the Zebrafish mtDNA sequence against the Carp mtDNA sequence (sister fish) plot (B) produces a diagonal plot through the center with a slight amount of line

missing in the lowest quadrant, slightly less conserved than the (A) plot. The Zebrafish mtDNA sequence against the Coelacanth mtDNA sequence (distant fish) plot (C) produces a diagonal plot off center with a little amount of line missing in the lowest quadrant, less conserved than the (A) or (B) plots, which correlates with the tree distance.

Miropeats (Parsons, 1995), a graphical program similar to a dotplot, figure 19 below, helped to show the repeated area, conserved regions between the two mitochondrias being compared as well as showing whether the mitochodrial sequence was cut to linerize the DNA sequence.

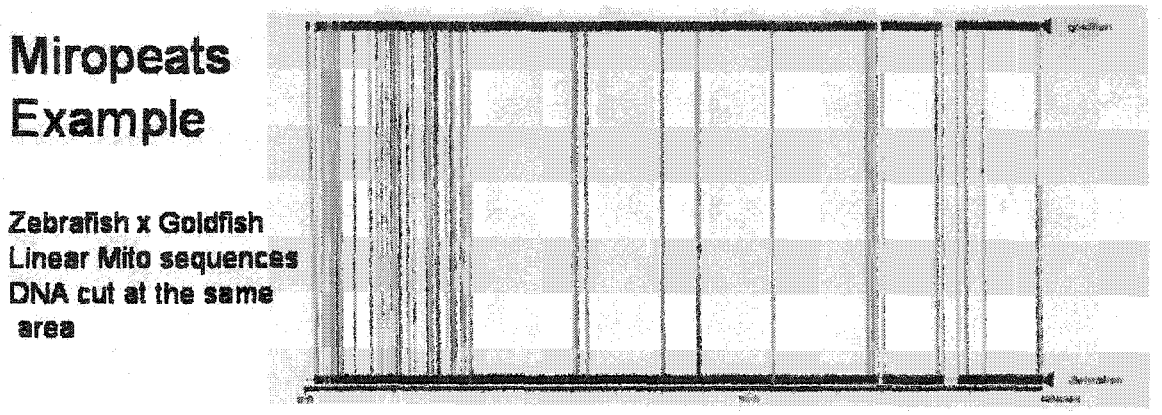


Figure 19 Miropeats

Miropeats (Parsons, 1995), a graphical program similar to a dotplot that finds the areas of similarity and shows them in a graphical form including tandem, inverted, and palindrome repeats. The Zebrafish mtDNA sequence is compared to the Goldfish mtDNA sequence (sister fish).
RESULTS: Zebrafish mtDNA sequence compared to the Goldfish mtDNA sequence (sister fish) of course is highly conserved. The right side of the plot where there are numerous lines is the D-loop, noncoding region.

Zebrafish mtDNA sequence compared to the Goldfish mtDNA sequence (sister fish) of course is highly conserved. The right side of the plot where there are numerous lines is the D-loop, noncoding region.

Comparative analysis was done using both Vista and MultiPip. In the MultiPip (Schwart et al. 2003), figure below, shown via the lines, that most of the sequences are

very conserved no matter how phylogenetically distant. The MultiPip shows some genes are more conserved than others, while on the Carp and Goldfish the D-loop (noncoding region) is only some what conserved and not at all for the non-sister fishes.

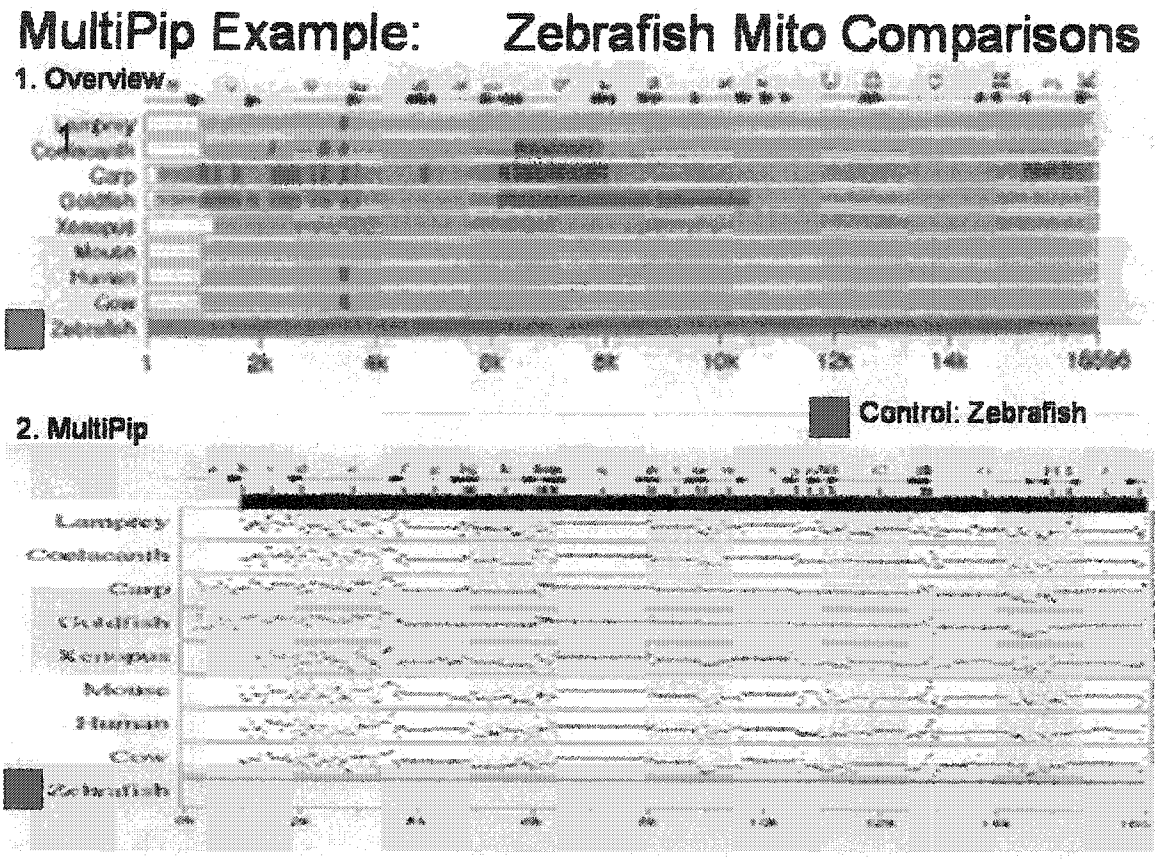


Figure 20 MultiPip

The Zebrafish sequence of the mtDNA genes are compared to the sequence of the mtDNA genes of the following in descending order as positioned on the Pip plots: Lamprey, Coelcanth, Carp, Goldfish, Xenopus, Mouse, Human, Cow, and the Zebrafish (control).
PIP OVERVIEW (Top plot) KEY: The colors vary from white which represents no sequence conservation to green which represents highly conserved DNA to red which represents identical sequence.
MULTIPIP(Bottom plot) KEY: The line with the arrow shows the gene and its direction. The black boxes with the numbers above the top line are exons, which correspond to the lines directly below in the boxes. No dotes or dashes represent no sequence conservation, a mixture of dotes and dashes show some similarity and straiht lines show conservation that is greatest at the top of the box. Each box has a range of conservation to the Zebrafish ranging from the 50% (bottom of box) to 100% (top of box).
RESULTS: In the overview above, the red stands for an exact match, while green appears to refer to a very conserved match. It can be seen that some genes are more conserved than others, while on the Carp and Goldfish the D-loop (noncoding region) is somewhat conserved.

The Vista (Mayor *et al.* 2000), figure below, also show that the sequences are highly conserved. This zebrafish mitochondria sequence and its' annotation as well as several interesting evolutionary patterns were published in Genome Research in 2001 (Broughton *et al.* 2001).

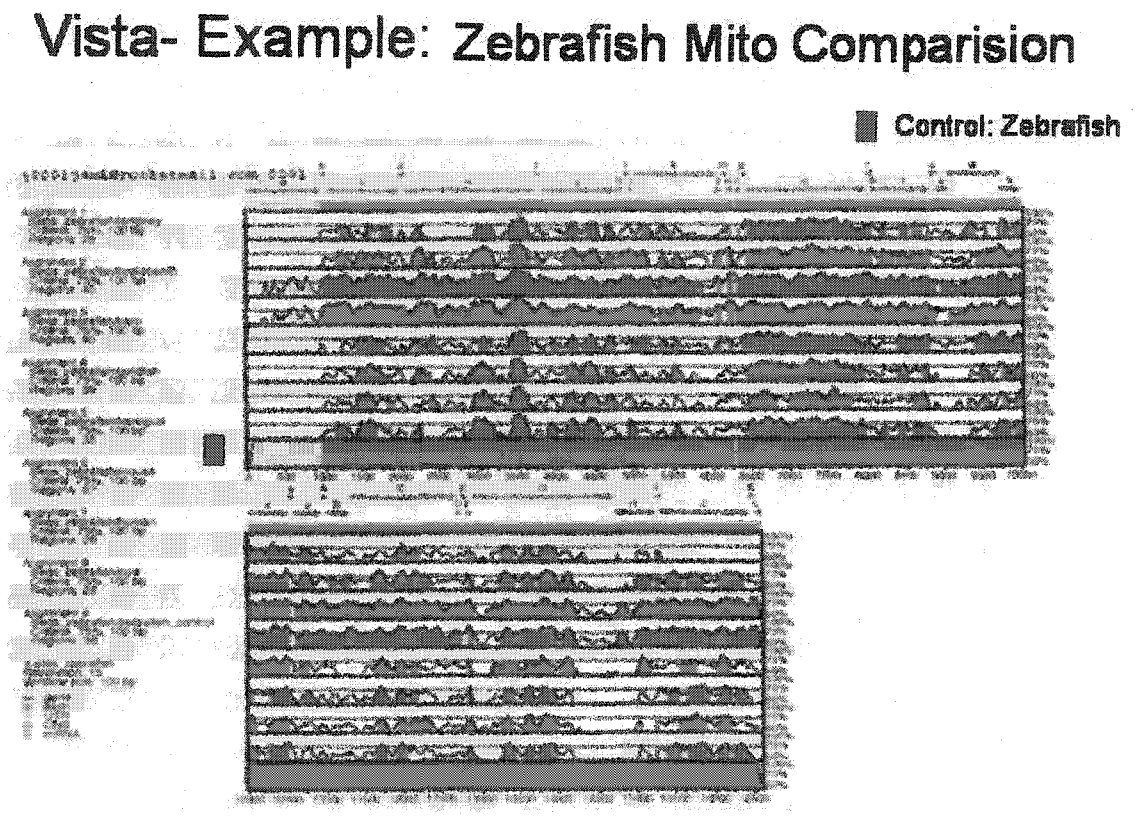


Figure 21 VISTA

The Zebrafish sequence of the mtDNA genes are compared to the sequence of the mtDNA genes of the following in descending order as positioned on the Pip plots: Lamprey, Coelcanth, Carp, Goldfish, Xenopus, Mouse, Human, Cow, and the Zebrafish (control).
VISTA KEY: The gray arrow shows the gene and its direction. The purple boxes above the top line are exons, which correspond to the lines, hills or valleys, directly below in the boxes. The purple color shown under the hills are conserved coding regions. The light blue color shown under the hills represents the UTR, Untranslated region. Each box has a range of conservation to the human ranging from 20% (bottom of box) to 50% (middle line through the box) to 100% (top of box).
RESULTS: A repeat of the same sequences used in the MultiPip in figure 20 for comparison. The purple clearly shows the high level of conservation between the mitochondrial sequences. It Notice, the light blue UTR, Untranslated region, for the Carp and Goldfish's conserved D-loop (noncoding region).

DNA Amplification

Phi-29 Experiments

Over the past few years, it has become increasingly apparent that previously unattainable genetic information would become available if robust procedures could be developed to amplify large amounts of DNA accurately starting from minute amounts. One initial goal of my research was to amplify several fish mitochondrial genomes in sufficient quantities that they could be directly sequenced. While searching for enhanced DNA amplification methods, I became aware of Pharmacias' research in DNA amplification using TempliPhi, a reaction mixture containing the Phi-29 DNA polymerase and hexamers. At that time, it was being tested to amplify plasmid-based DNA sequencing templates and we were fortunate to become a Beta test site for this enzyme test.

Phi-29 mix: P1: Amplification from pGem Gold (Isolated no insert) DNA

The first question to be answered was does the Phi-29 mix satisfactorily amplify DNA that can yield DNA sequence data comparable to template obtained via the existing cleared lysate isolation protocol. I, therefore, first confirmed that Phi-29 could amplify 1 ng of pUC DNA 1000-fold, shown in figure 22.

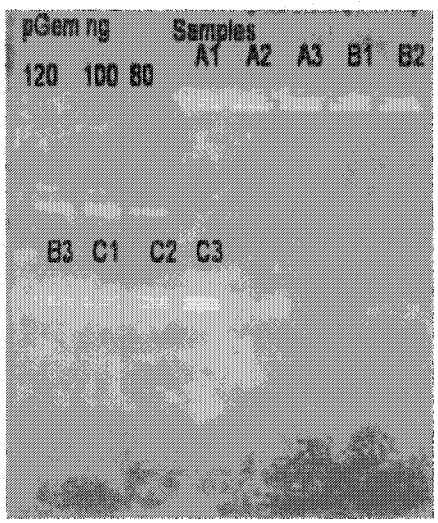


Figure 22 Amplification from pGem DNA

One nanogram of pUC from in 1ul of DNA, 9 different samples (A1,A2,A3,B1,B2,B3,C1,C2, andC3) was added to 9 µl of buffer and heated 95 °C for three minutes and 10 µl of Phi-29 mix (5 U) was added to the reaction. The solution, then, was incubated at 30 °C for 4 hours. 10% of the amplification product was loaded on to 0.6% agrose gel, electrophoresed and stained with ethidium bromide.
RESULTS: The DNA produced clearly is more than 100 ng/ lane and therefore, more than 1 µg was produced per reaction, confirming that the Amersham Phi-29 mix works satisfactorily.

The DNA produced clearly is more than 100 ng/ lane and therefore, more than 1 µg was produced per reaction, confirming that the Amersham Phi-29 mix works satisfactorily.

Phi-29 mix: P2: Amplification from Cells and Sequence

I, then tested, if the Phi-29 mix could amplify from shotgun clone containing colonies, and if the DNA was double stranded, i.e. were both strands amplified?

AS shown in figure 23, the DNA was successfully amplified from lysed cells and this amplified DNA could be cleaved, indicating that it must be double stranded.

To determine if this amplified DNA could be sequenced, it was diluted by 50% and 1ul, 2ul, and 4ul of each sample were sequenced using the forward and reverse primers with the ABI BigDye mix. As shown in figure 23, the DNA amplified from cells could be sequenced directly, but did not result in reproducible results.

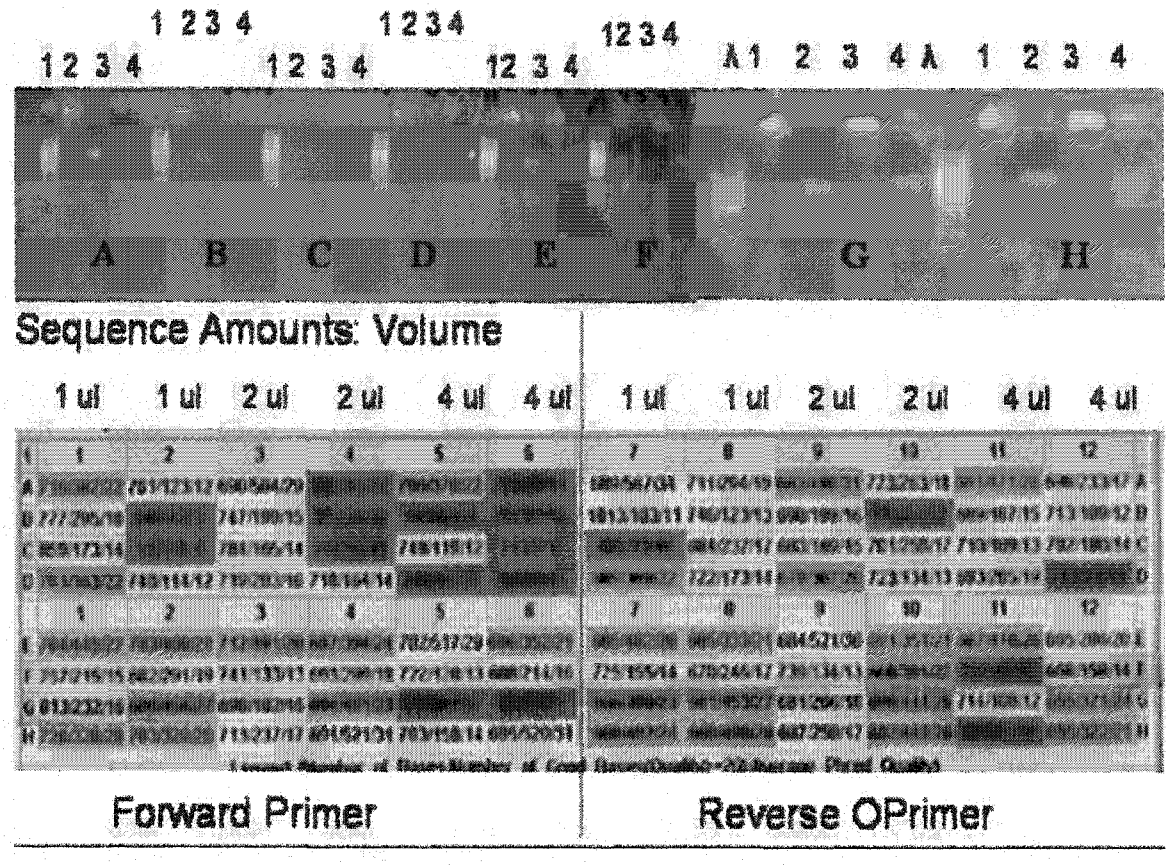


Figure 23 Amplification from Cells and Sequence

Lambda is the first well, lane 1 is the undigested DNA, lane 2 is the EcoR1 cleaved DNA, lane 3 is the undigested DNA, and lane 4 is EcoR1 cleaved DNA. Lanes 1 and 3 for each letter was sequenced using forward and reversed. Sample from letter A of the gel was sequenced in horizontal letter A on the Phred card at 1μL, 2 μL, and 4 μL.

EXPT: Tooth picks covered in aluminum foil were used to pick the colonies and dip into the buffer to lyse the cells. The released DNA was amplified as described in the methods section. After the amplification, 20 μl of sterile double distilled water was added to the 20 μl reaction mixture to dilute the DNA.

ECORI DIGEST (Top): An EcoR1 restriction digest was performed to confirm the DNA was not nonspecific and that it was double stranded. One unit of enzyme should digest one microgram of double stranded DNA in one hour (Roe ed. 1997) and the enzyme needs both strands to produce a recognition site in order to cut the DNA. EcoR1 reaction: 2 μl of the amplified DNA was used, 2μ of Buffer, 1.5 μl (1 μl = 10 U) of EcoR1 enzyme, and 14.5 μl sterile double distilled water. The reaction was incubated at 37 °C for 2 hours (standard parameters). The original DNA and the restriction digested DNA were electrophoresed on a 0.6% agrose gel at 150 mA for 15 minutes.

PHRED CARD (Bottom): Phred is a program that reads the trace files produced by the sequencers to assign the bases, their quality values and writes these to output files (Ewing et al., 1998). A Phred Card is a picture showing each well as a certain color based on the number of nucleotides sequenced that had a Phred quality of at least 20, the quality required for Phrap to assemble the sequence is shown in fig. 10: 0-100 bases = Red, 101-300 bases = Yellow, C. 301-500 bases = Green, D. 501 and above bases = Blue.

RESULTS: The DNA amplified from cells could be sequenced directly, but did not result in reproducible results.

Phi-29 mix: P3: Buffer vs. Water and Denatured vs. Non-Denatured on Isolated DNA.

Since the sequencing reaction results were reproducible, we questioned whether the parameters were the same for the isolated and non-isolated, cellular, DNA? It was hypothesized that since the buffer used to lyse the cells contained EDTA, a known polymerase inhibiter, was affecting the sequencing results. Therefore, it was questioned whether the buffer was necessary, if water could be used instead and if it was, would the denature step, 95 °C for 3 minutes, be necessary? We also tested if the DNA needed to be cleaned via ethanol precipitation before sequencing to remove the EDTA.

In order to answer the above questions an experiment was designed, comparing the sequence obtained from amplification reactions comparing water (no EDTA) and buffer (with EDTA), whether the water or the buffer's DNA needed denaturation (95° C for 3 minutes to insure DNA strand separation for primer annealing for amplification) before the amplification enzyme mix is added to the reaction. The reactions are repeated on the plate, so that half the reactions were cleaned via ethanol precipitation and the other half were not cleaned. The diagram below shows what is in each well, which correlates with the Phred card in figure 24.

DNA used	1	2	3	4	5	6	7	8	9	10	11	12
pGem~1ng	sterile double distilled Water W	WD	WD	WD	WD	pUC~1ng Control	**Denature** D	WD	WD	WD	WD	pUC~1ng Control
pGem~1ng	sterile double distilled Water W	W ND	W ND	W ND	W ND	pUC~1ng Control	**No Denature** ND	W ND	W ND	W ND	W ND	pUC~1ng Control
pGem~1ng	Buffer B	B D	B D	B D	B D	pUC~1ng Control	**Denature** D	B D	B D	B D	B D	pUC~1ng Control
pGem~1ng	Buffer B	B ND	B ND	B ND	B ND	pUC~1ng Control	**No Denature** ND	B ND	B ND	B ND	B ND	pUC~1ng Control
pUC~	sterile	WD	WD	WD	WD	WD	**Denature**	WD	WD	WD	WD	WD

Colony	double distilled Water W	C	C	C	C	C	D	C	C	C	C	C
pUC-Colony	sterile double distilled Water W	W ND C	W ND C	W ND C	W ND C	W ND C	No Denature ND	W ND C	W ND C	W ND C	W ND C	W ND C
pUC-Colony	Buffer B	B D C	B D C	B D C	B D C	B D C	Denature D	B D C	B D C	B D C	B D C	B D C
pUC-Colony	Buffer B	B ND C	B ND C	B ND C	B ND C	B ND C	No Denature ND	B ND C	B ND C	B ND C	B ND C	B ND C

Uncleaned---------- Uncleaned | **Etoh ppt**----------**Etoh ppt**

A Black "W" represents water, a Red "D" represents denatured sample, a green "ND" represents a non-denatured sample, a light blue "C" represents colony DNA, and a Bark blue "B" represents Buffered sample.

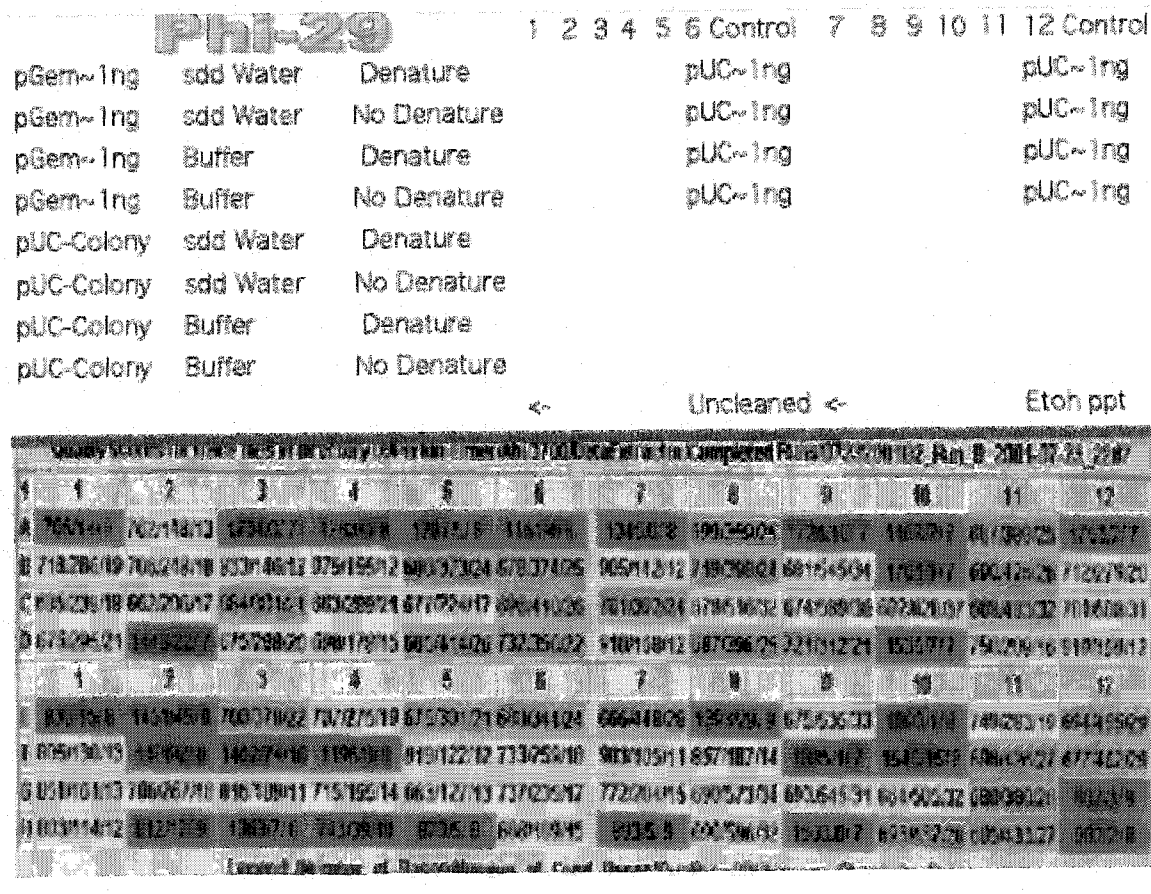


Figure 24 Phi-29 mix: Colony vs. Isolated, Buffer vs. Water, Denatured vs. Non-Denatured.

PHI-29 (Top): The two figures above the Phred Card are used to show what type of sample is in each well.

PHRED CARD (Bottom): Phred is a program that reads the trace files produced by the sequencers to assign the bases, their quality values and writes these to output files (Ewing et al., 1998). A Phred

Card is a picture showing each well as a certain color based on the number of nucleotides sequenced that had a Phred quality of at least 20, the quality required for Phrap to assemble the sequence is shown in fig. 10: 0-100 bases = Red, 101-300 bases = Yellow, C. 301-500 bases = Green, D. 501 and above bases = Blue.
RESULTS: The sequence produced by the experiment indicates that water can be used to lyse the cells and that when amplifying from previously isolated DNA, no denaturing step is necessary, while, when working with DNA from colonies, denaturation is necessary. Both types of DNA amplifications (colonies and isolated DNA) using water need to be cleaned with ethanol precipitation, to give satisfactory sequencing results.

The sequence produced by the experiment shown in figure 24 indicates that water can be used to lyse the cells and that when amplifying from previously isolated DNA, no denaturing step is necessary, while, when working with DNA from colonies, denaturation is necessary. Both types of DNA amplifications (colonies and isolated DNA) using water need to be cleaned with ethanol precipitation, to give satisfactory sequencing results.

Phi-29 mix: P4: Cellular DNA Cleaning for Amplification

Since the above results were successful, we, then, began to scale up the procedure for eventual automation using 96 samples/plate. To accomplish this, it was necessary to move from using aluminum foil covered tooth picks to pick colonies to the colony picker for mass production and greater precision in the amount of cells lysed. However, this resulted in even more reduced DNA sequence reproducibility. Problems with amplification began at this time. It was hypothesized that the some unknown cellular inhibitor's percentage had increased with the addition of the colony picker for picking the colonies. Various experiments were attempted to remove or lessen this cellular inhibitor and it was during this time that it was learned that after using the standard 3 M sodium ethanol precipitation for cleaning the amplification reactions that one could see the DNA on a gel before cleaning, but not after (rediluting in the same reaction volume).

After various experiments (many failures), the following was done: First, the colony picker picked into a thermocycler plate with 20 µl of sterile double distilled water.

These reactions were heated to 95 °C for 3 minutes in a thermocycler, then frozen until solid ~2 to 3 hours, thawed for ~10 minutes and centrifuged for 45 minutes at 3200 rpm, Beckman C56R centrifuge. The supernatant ~5 µl was transferred to another plate; this removed the majority of the cellular inhibitor. Then, 5 µl sterile double distilled water and 10 µl of Phi-29 reaction mix was added. The reaction was incubated at 30 °C for 12 hours, followed by denaturing the enzyme via a 95 °C hold for 5 minutes. The reactions were precipitated using 2 µl of 7.5 M KOAc and 53 µl of 100% ethanol. This solution was centrifuged for 3200 rpm, Beckman C56R centrifuge, for 45 minutes, then inverted on a paper towel and centrifuged to 400 rpm, Beckman C56R centrifuge. The wells were immediately rediluted in 20 µl of sterile double distilled water and 3 µl were electrophoresed on a 0.6 %. The results of these experiments are shown in figure 25.

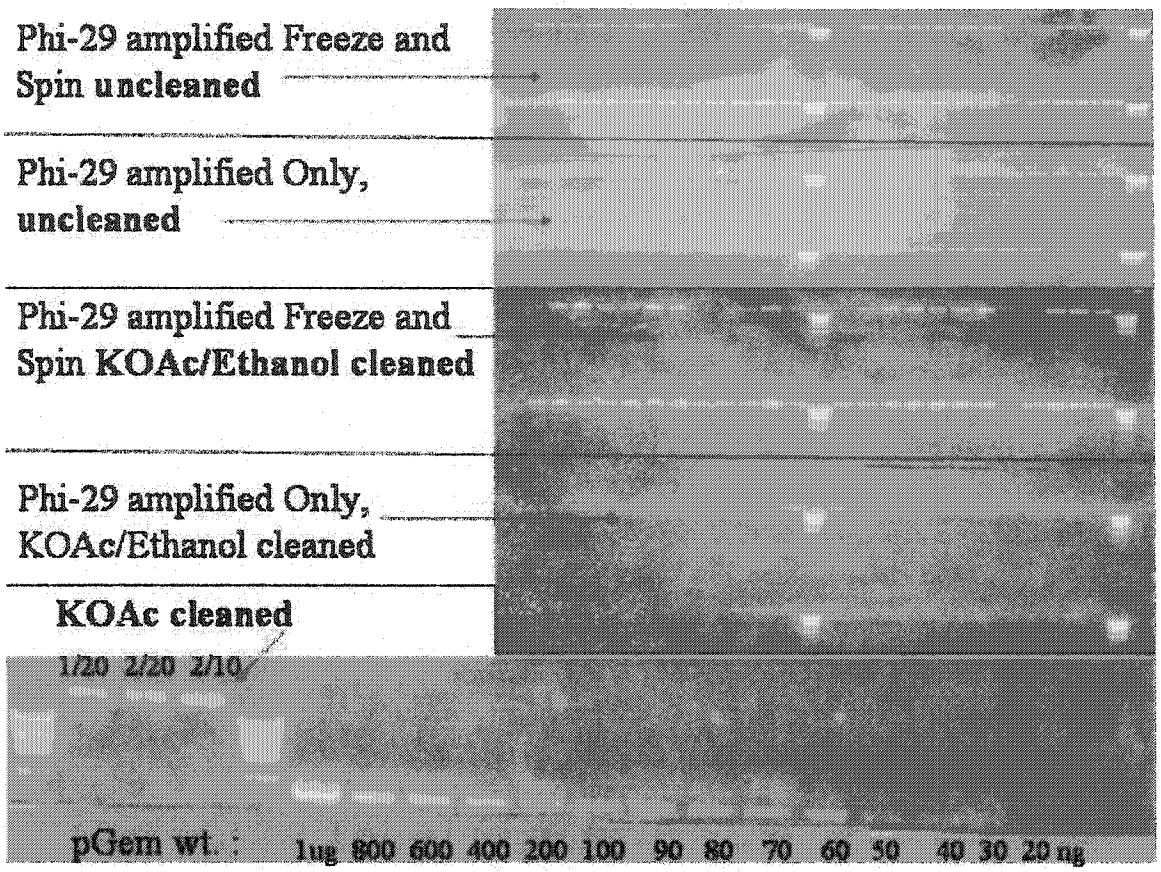


Figure 25 PHi-29 Amplification Cleaning Comparison

GEL (Top): The following sample types are listed in descending order as positioned on the gel: Phi-20 amplified DNA using the freeze and spin technique without KOAc/ ETOH precipitation, Phi-20 amplified DNA (control) without KOAc/ ETOH precipitation, Phi-20 amplified DNA using the freeze and spin technique with KOAc/ ETOH precipitation, and Phi-20 amplified DNA (control) with KOAc/ ETOH precipitation.
GEL (Bottom): Shows Phi-20 amplified DNA using the freeze and spin technique with KOAc/ ETOH precipitation 1µL and 2 µL out of 20 µL as well as 2 µL out of a total of 10 µL reaction compared to known pGem DNA amounts in nanagrams.
The gel above shows the amount of DNA recovered ranging from gray (invisible DNA, very small amount), to very bright white DNA (large recovery).
EXPT: The figure above is Phi-29 amplified DNA that was electrophoresed on a 0.6% agrose gel at 150 mA for 15 minutes.
RESULTS: The major result of this cleaning comparison experiment was that the addition of the freeze and centrifuge to pellet the cellular debris (inhibitor), successfully produced DNA that could be reproducibly sequenced.

The major result of this cleaning comparison experiment was that the addition of the freeze and centrifuge to pellet the cellular debris (inhibitor), successfully produced DNA that could be reproducibly sequenced.

Phi-29 mix: P5: Cellular DNA Concentration Study for Amplification

After finding a cleaning protocol for the DNA, it was then questioned exactly how diluted could the cellular DNA be for amplification and would a greater dilution produce better results? An experiment was set up using differing amounts of supernatant (5 µl, 2.5 µl, 1.25 µl, and 0.625 µl). This was done via a serial dilution starting with 5 µl volume of cellular DNA and maintaining that volume to be used for the amplification reaction and then a sample from each of these for sequence. After the reactions were amplified, they were KOAc/ethanol precipitated, rediluted in 20 µl water and 3 µl were used to check for DNA on a gel.

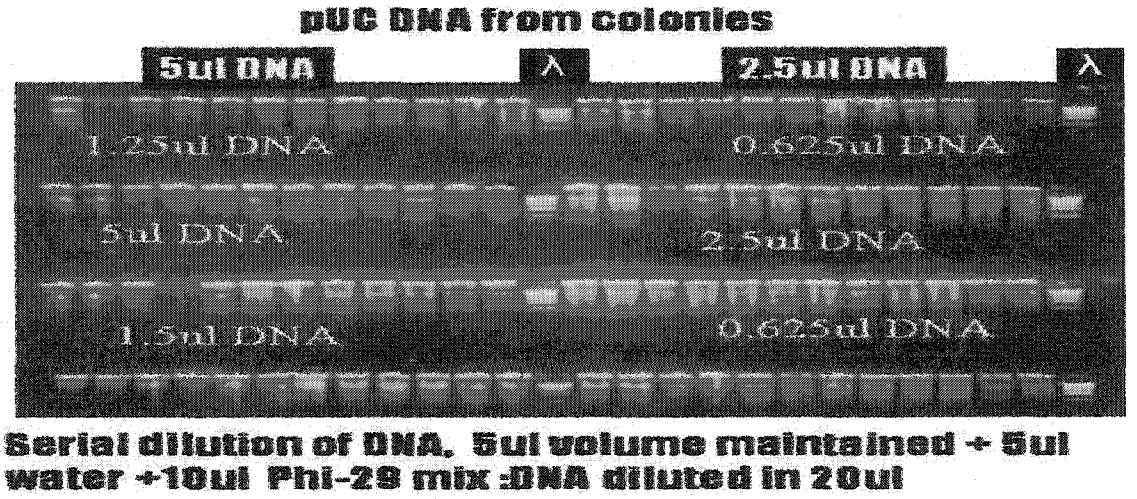


Figure 26 Phi-29 Amplification of Cellular DNA a concentration Study.

There are 12 samples between each lambda of varying amounts of supernatant used to amplify: 5 µL, 2.5 µL, 1.25 µL, and 0.625 µL.
EXPT: A comparison of the amplified reactions were done to determine if dilution of the cellular DNA would affect the amplification. After amplification the samples were KOAc/ethanol precipitated, rediluted in 20 µl water and 3 µl were used to check for DNA on a gel. The figure above is Phi-29 amplified DNA that was electrophoresed on a 0.6% agrose gel at 150 mA for 15 minutes.
RESULTS: As shown all of the samples produced amplified DNA, however, when 2 µl of this DNA was sequenced using BigDye I and BigDye III, it did not sequence very well, but when 5 µl of amplified DNA was sequenced with Big Dye I, it produced better sequence than when Big Dye III was used.

As shown in figure 26, all of the samples produced amplified DNA, however, when 2 µl of this DNA was sequenced using BigDye I and BigDye III, it did not sequence very well, but when 5 µl of amplified DNA was sequenced with Big Dye I, it produced better sequence than when Big Dye III was used.

Phi-29 mix: P6: Enzyme mix Concentration Study: Time versus rate of Amplification

After, getting the above amplification to work, I investigated if less enzyme could be used, since the Phi-29 mix was not economical for high throughput use. In order to produce a quick and dirty estimate of the amount of DNA produced based on the length of incubation time at 30 °C. In the initial experiment, a time study, was performed.

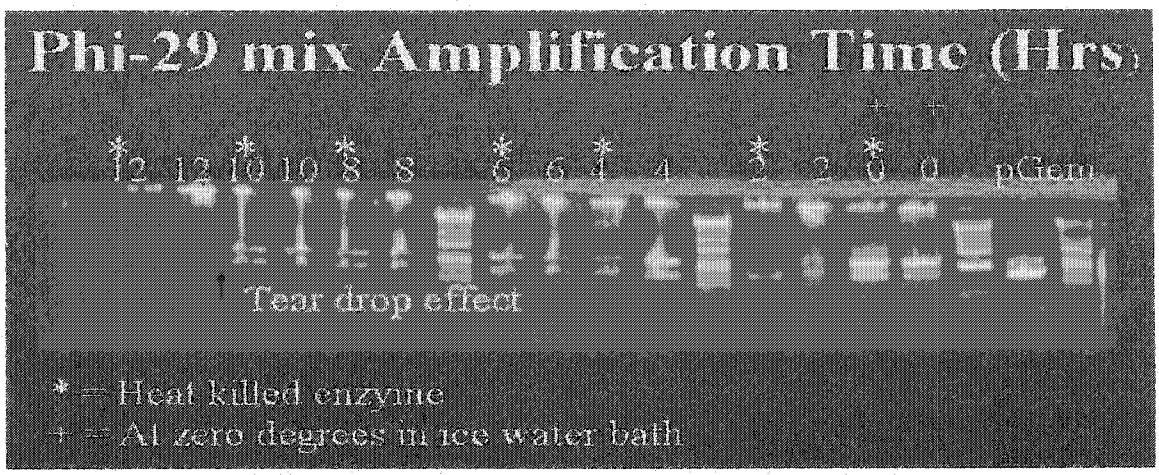


Figure 27 Phi-29 Enzyme mix Concentration Study: Time vs. Rate of Amplification

EXPT: The reactions were pipetted on ice and then incubated for 0, 2, 4, 6, 8, 10, and 12 hours. One sample from each time period would be heated to 95 °C for 5 minutes to kill the enzyme and the second would be replaced in the water bath until the end of the experiment, where both samples would be immediately electrophoresed on a 0.6 % agrose gel. The DNA used was pGem using 5 units of Phi-29 enzyme. The control DNA pGem at the far right and the numbers 0, 2, 4, 6, 8, 10, 12 represent the time the samples were incubated at 30 °C. An asterisk is above those samples that were heated to 95 °C to denature the enzyme and stop the reaction.
RESULTS: After 4 hours the DNA is so highly amplified that it produces a tear drop effect, indicating the DNA is not fully dissolved in the loading buffer, it contains a high salt concentration. It also is noticeable that a large amount of DNA is long enough that it remains in the well during electrophoresis. There appears to be less of a tear drop effect occurring when the DNA is heated to 95 ° and the amplification results from the reactions amplified and stored at 4 °C appears "blotchy" compared to the amplified and heat denatured reactions, indicating DNA amplification at 4 °C.

Supportingly, we observed that when the reaction mixture is incubated for only 2 to 5 minutes at 0 °C during pipetting, the enzymes catalyzes the production of amplified DNA. The major result of the experiment was that the DNA amplification continues after 2 hours, but it was difficult to annotate the results at this time for reasons that will be discussed later.

As can be seen in figure 27, after 4 hours the DNA is so highly amplified that it produces a tear drop effect, indicating the DNA is not fully dissolved in the loading buffer, it contains a high salt concentration. It also is noticeable that a large amount of DNA is long enough that it remains in the well during electrophoresis.

There appears to be less of a tear drop effect occurring when the DNA is heated to 95 ° and the amplification results from the reactions amplified and stored at 4 °C appears "blotchy" compared to the amplified and heat denatured reactions, indicating DNA amplification at 4 °C. Supportingly, we observed that when the reaction mixture is incubated for only 2 to 5 minutes at 0 °C during pipetting, the enzymes catalyzes the production of amplified DNA.

The major result of the experiment was that the DNA amplification continues after 2 hours, but it was difficult to annotate the results at this time for reasons that will be discussed later.

Phi-29 mix: P7: Enzyme mix Concentration Study for Amplification from Cellular DNA

Based on results of the Cellular DNA Concentration Study for Amplification experiment, additional work was necessary to use the amplification method followed by subsequent DNA sequencing with BigDye3. It was proposed that too much DNA was being added to the sequencing reactions. Instead of just diluting the DNA down further, it seemed more logical and cost efficient to lower the enzyme amount added to the amplification reactions.

In order to test the above, the amplification reactions were set up as follows: 2ul of Phi-29 enzyme mix was added to 2ul of cellular DNA, 4ul of enzyme mix was added to 4ul of cellular DNA, 5ul of enzyme mix was added to 5ul of cellular DNA, and the current standard reaction amount, 10ul of enzyme mix was added to 5ul of cellular DNA and 5ul sterile double distilled water. After amplifying and cleaning the reactions, they were diluted in varying amounts of sterile double distilled water. The results are shown in figures 28 and 29.

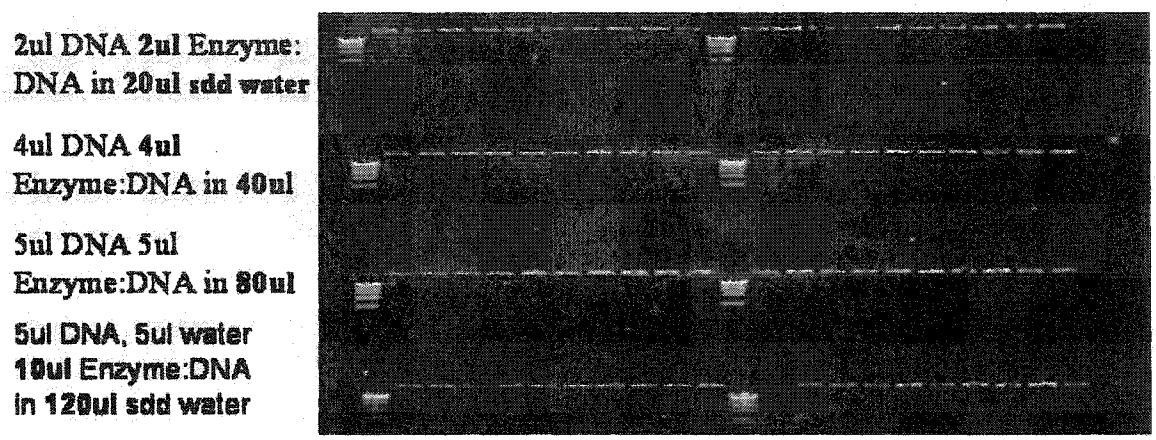


Figure 28 Phi-29 Enzyme Concentration study

EXPT: 2ul of enzyme mix and 2ul of cellular DNA were rediluted in 20 µl sterile double distilled water, 4ul of enzyme mix and 4ul of cellular DNA were rediluted in 40 µl sterile double distilled water, 5µl of enzyme mix and 5µl of cellular DNA were rediluted in 80 µl sterile double distilled water, and the current standard reaction amount, 10µl of enzyme mix and 5µl of cellular DNA and 5µl sterile double distilled water were rediluted in 120 µl sterile double distilled water. 3ul of DNA were electrophoresed on a 0.6% gel and 2µl from each of the above amplified and diluted samples were sequenced using BigDye 1 and Big Dye 3.
RESULTS: The above shows a lot of amplified DNA for the dilution.

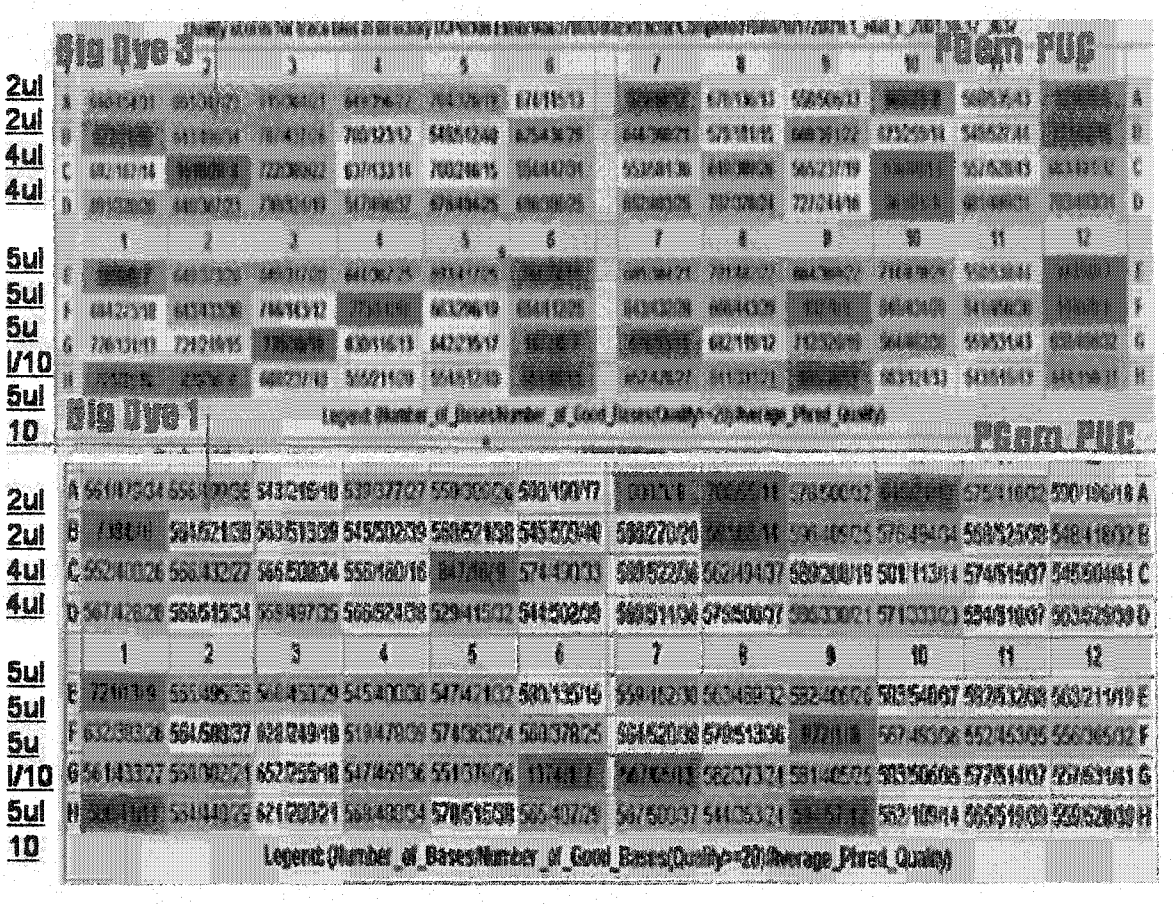


Figure 29 Sequence of the Phi-29 Enzyme Concentration study

From the previous figure, 2ul from the amplified and diluted samples were sequenced using BigDye 1 and Big Dye 3.

PHRED CARD (Bottom): Phred is a program that reads the trace files produced by the sequencers to assign the bases, their quality values and writes these to output files (Ewing et al., 1998). A Phred Card is a picture showing each well as a certain color based on the number of nucleotides sequenced that had a Phred quality of at least 20, the quality required for Phrap to assemble the sequence is shown in fig. 10: 0-100 bases = Red, 101-300 bases = Yellow, C. 301-500 bases = Green, D. 501 and above bases = Blue. Wells A-H of column 11 and 12 are controls pGem and pUC correspondingly. RESULTS: The above indicates that the amount of enzyme used in the amplification reaction, must be lowered to produce high quality DNA sequence.

The above figures, 28 and 29, indicate that the amount of enzyme used in the amplification reaction, must be lowered to produce high quality DNA sequence, the less DNA the better: 2 μL of cellular DNA using 2 μL of the Phi-29 enzyme amplifies and then produces the highest quality sequence.

Phi-29 mix: P8: Enzyme mix Concentration Study for Amplification from Glycerol Stocks

A similar study was done using colonies from glycerol stocks. Here, there seemed to be no significant difference between using a colony from a Petri dish or colonies directly from glycerol stocks, although if the glycerol stocks were made fresh, the resulting amplified DNA sequenced far better as shown in figure 30.

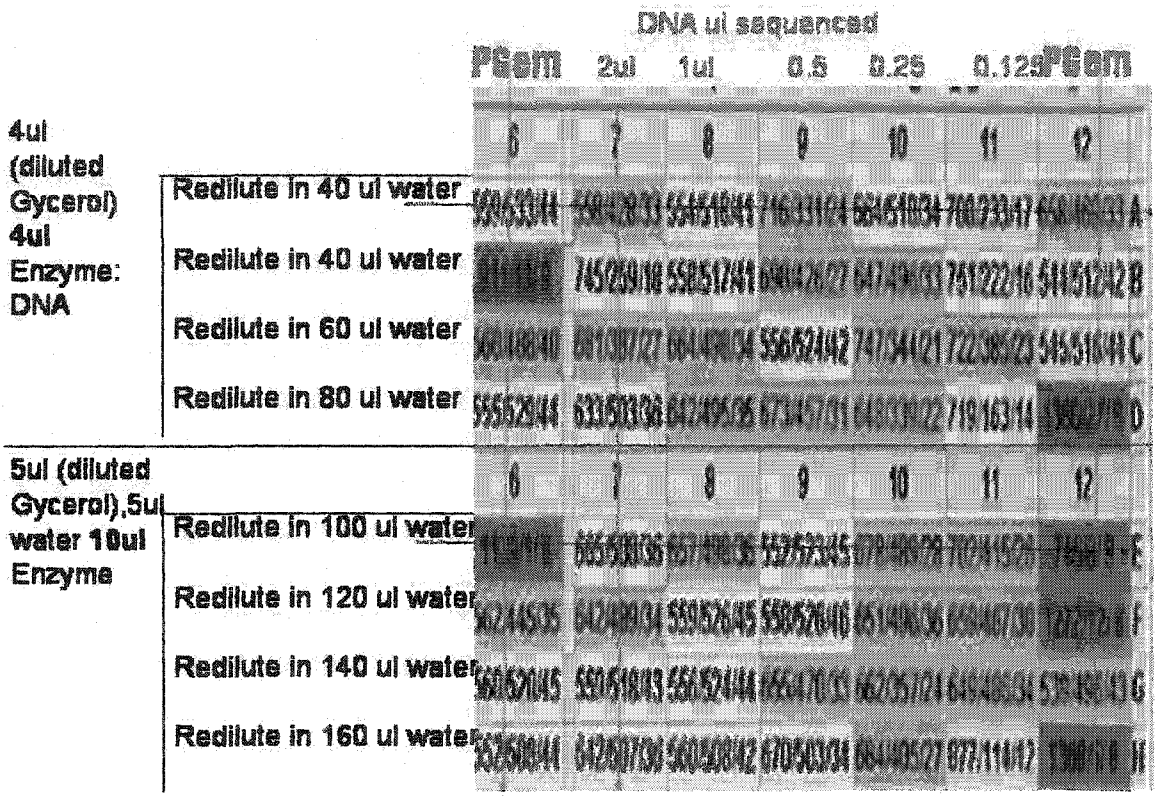


Figure 30 Sequencing Results from the Enzyme mix Concentration Study for Amplification from Glycerol Stocks.

EXPT: The amplification protocols used were similar to that discussed in the legend of figure29, but the amount of diluted glycerol stock and the amount of phi-29 enzyme mix were varied, along with the amount of sterile double distilled water added to the amplified DNA after cleaning. Each column represents a different DNA sample as indicated, except for the sequencing control DNA (pGem) all the DNA is the exactly the same, but each sample was amplified, diluted, and sequenced using different parameters as indicated in the figure.

PHRED CARD : Phred is a program that reads the trace files produced by the sequencers to assign the bases, their quality values and writes these to output files (Ewing et al., 1998). A Phred Card is a picture showing each well as a certain color based on the number of nucleotides sequenced that had a Phred quality of at least 20, the quality required for phrap to assemble the sequence is shown in fig. 10: 0-100 bases = Red, 101-300 bases = Yellow, C. 301-500 bases = Green, D. 501 and above bases = Blue. Wells A-H of column 6 and 12 are controls using isolated DNA, pGem. The DNA was sequenced using 2 µL, 1 µL, 0.5 µL, and 0.125 µL

RESULTS: When DNA was amplified from glycerol stocks, it could be diluted between 10 and 16 times and between 2 and 1/8 µl of DNA produced excellent sequencing results with plasmid DNA.

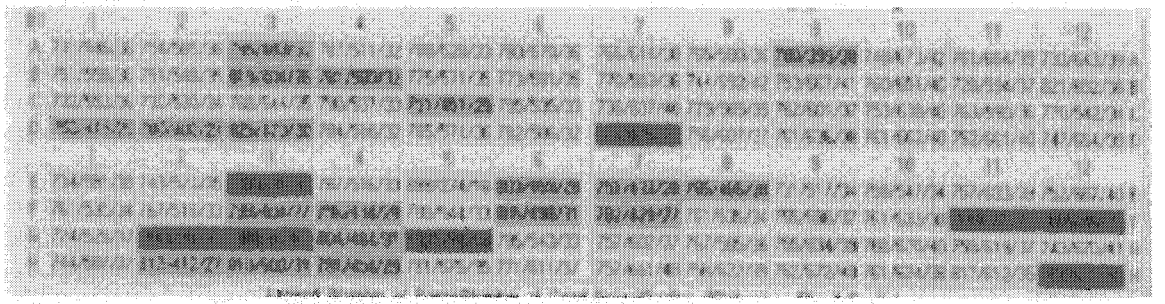
As shown in figure 30, when DNA was amplified from glycerol stocks, it could be diluted between 10 and 16 times and between 2 and 1/8 µl of DNA produced excellent sequencing results with plasmid DNA.

Phi-29 mix: P9: Enzyme mix in Buffer B, Concentration Study for Amplification and Sequence

To further investigate additional dilutions for DNA amplification and after reviewing the literature, a new dilution buffer was made to dilute the Phi-29 enzyme. This buffer contained: 50 mM Tris-HCl pH 7.6, 10 mM $MgCl_2$, 10% v/v Glycerol, 1 mM DDT, ~48.8 uM Universal Forward primer, ~48.8 uM Universal Reverse primer, 50 mM dNTPs (for each dNTP); this same buffer was made with other primers as testing required.

Amplification and sequencing was done with purified pGem DNA and varying the Phi-29 enzyme mix dilutions.

Phi-29 enzyme only pGEm DNA & BigDye 3
~0.25 ul (0.125 U) Enzyme diluted
w/ mixed Buffer B



Phi-29 enzyme only 1/80th, but 2ul per reaction =1/40, the recommended amount of enzyme pGEm & BigDye 3
~0.125 ul (0.0625 U)Enzyme diluted w/ mixed Buffer B

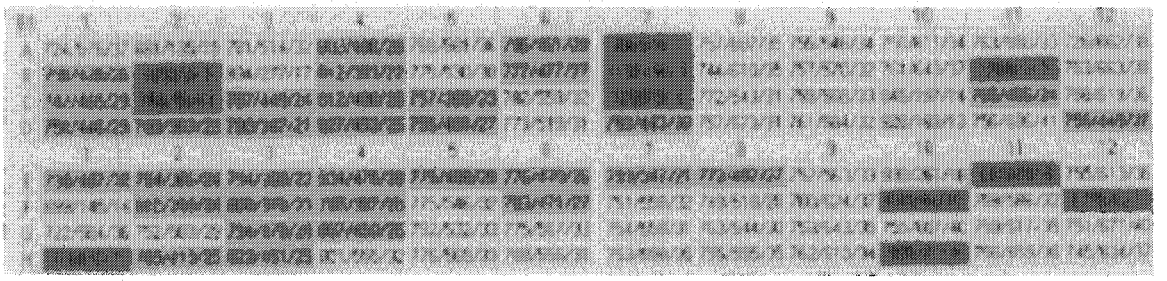


Figure 31 Sequencing results of the Phi-29 enzyme mix diluted in Buffer B, Concentration Study for Amplification and Sequence

EXPT: Amplification and sequencing was done with purified pGem DNA and varying the Phi-29 enzyme mix dilutions
Phi-29 Phred Card (Top): Amplification of pGem DNA using 0.125 units of Phi-29 diluted in Buffer B, sequencing with BigDye3.
Phi-29 Phred Card (Bottom): Amplification of pGem DNA using 0.0625 units of Phi-29 diluted in Buffer B, sequencing with BigDye3.
PHRED CARD: Ph red is a program that reads the trace files produced by the sequencers to assign the bases, their quality values and writes these to output files (Ewing et al., 1998). A Phred Card is a picture showing each well as a certain color based on the number of nucleotides sequenced that had a Phred quality of at least 20, the quality required for phrap to assemble the sequence is shown in fig. 10: 0-100 bases = Red, 101-300 bases = Yellow, C. 301-500 bases = Green, D. 501 and above bases = Blue.
RESULTS: The enzyme concentration could be successfully lowered to 0.25 µl (0.125 U) and 0.125 µl (0.0625 U) of enzyme for amplification using this new buffer. This new dilution buffer allowed the enzyme to be lowered 1/80th or (2 µl needed = 1/40th pre reaction) from the manufacture recommended amount [Normally 5 U/10 µl mix/reaction-(Dean 2001)].

As shown in figure 31, the enzyme concentration was successfully lowered to 0.25 µl (0.125 U) and 0.125 µl (0.0625 U) of enzyme for amplification using this new buffer. This new dilution buffer allowed the enzyme to be lowered 1/80th or (2 µl

needed = 1/40th pre reaction) from the manufacture recommended amount [Normally 5 U/10 μl mix/reaction-(Dean 2001)]. However, in order for the phi-29 reactions to be economically feasible, the cost per amplification reaction had to be less than or equal to a penny. Since, the amounts of the phi-29 necessary to reproducibly produce quality amplification for sequence could be lowered only to 1/40th per reaction, which was 1.5 cents/ reaction. This was slightly above the cost allowed and amplification of plasmid DNA with Phi-29, although successful was not continued. However, what was need in the lab was a robust BAC amplification protocol; therefore, I began investigating the possibility of BAC amplification for shotgun and gap closure using Phi-29.

Phi-29 for Gap Closure

Typically, polymerases dissociate from the template when regions of nucleotide repeats (i.e.. G-C, G-A) our encountered. The Phi-29 DNA polymerase is so tightly bound to its substrate that it does not dissociate from the DNA like other DNA polymerases do. The Phi-29 DNA polymerase also has the ability to displace the opposite strand (i.e. strand displacement) and is capable of reading through hairpin loops and long repeats. These properties might be useful for obtaining DNA sequencing template for DNA regions that are either unclonible or difficult to obtain and/or sequence. To test their possibility, I investigated the standard protocol developed for plasmid amplification, but instead used primers flanking the region to amplify for sequence.

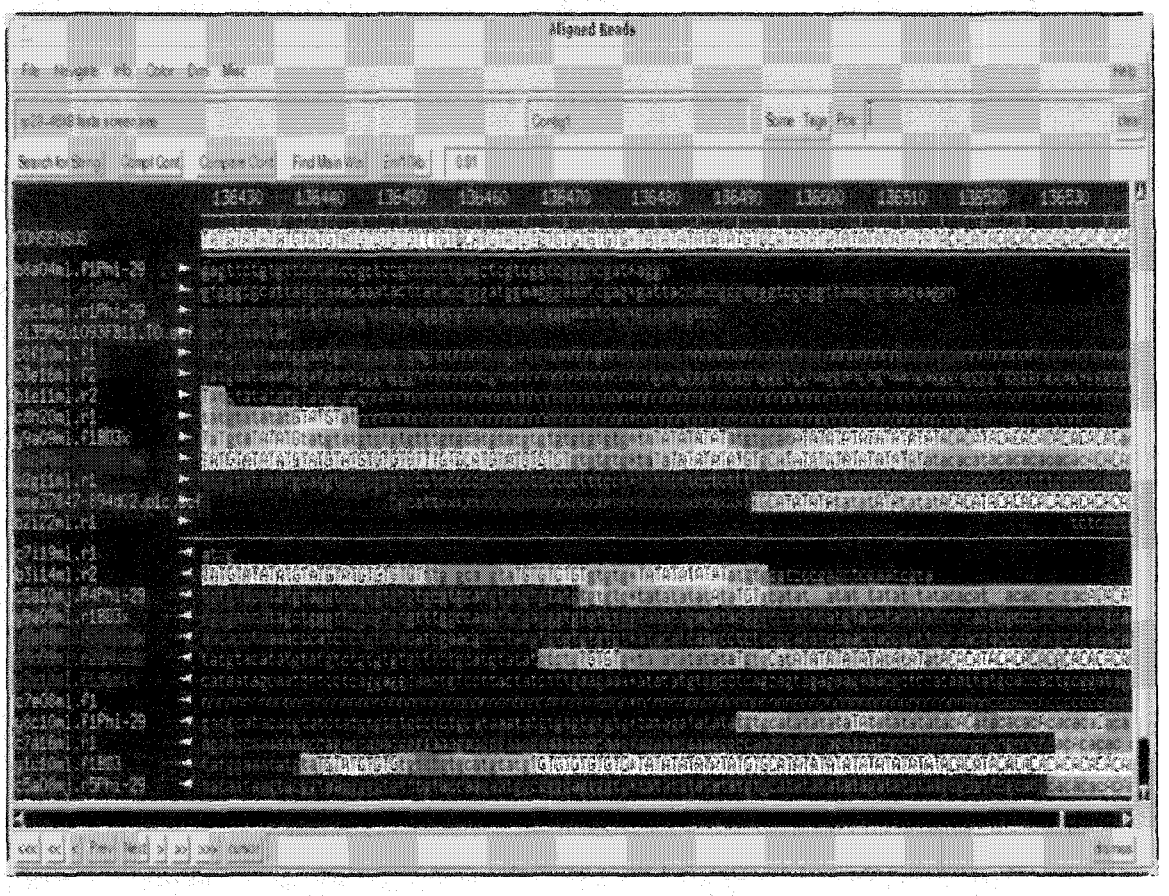


Figure 32 Phi-29 used for Gap Closure-1

A consed output for a difficult region to sequence in a region of a BAC sequencing project, the purple sequences represent phi-29 amplified DNA that was sequenced.
EXPT: To test their possibility, I investigated the standard protocol developed for plasmid amplification, but instead used primers flanking the region to amplify for sequence.
RESULTS: This approach worked exceptionally well and was able to amplify DNA that was difficult to obtain by other methods and this amplified DNA sequenced as well. It was used to close a gap in a plasmid shotgun clone that contained both "G-T and A-T" repeats that did not yield satisfactory sequencing results using typical DNA sequencing protocols.

As shown in figure 32, this approach worked exceptionally well and was able to amplify DNA that was difficult to obtain by other methods and this amplified DNA sequenced as well. It was used to close a gap in a plasmid shotgun clone that contained both "G-T and A-T" repeats that did not yield satisfactory sequencing results using typical DNA sequencing protocols.

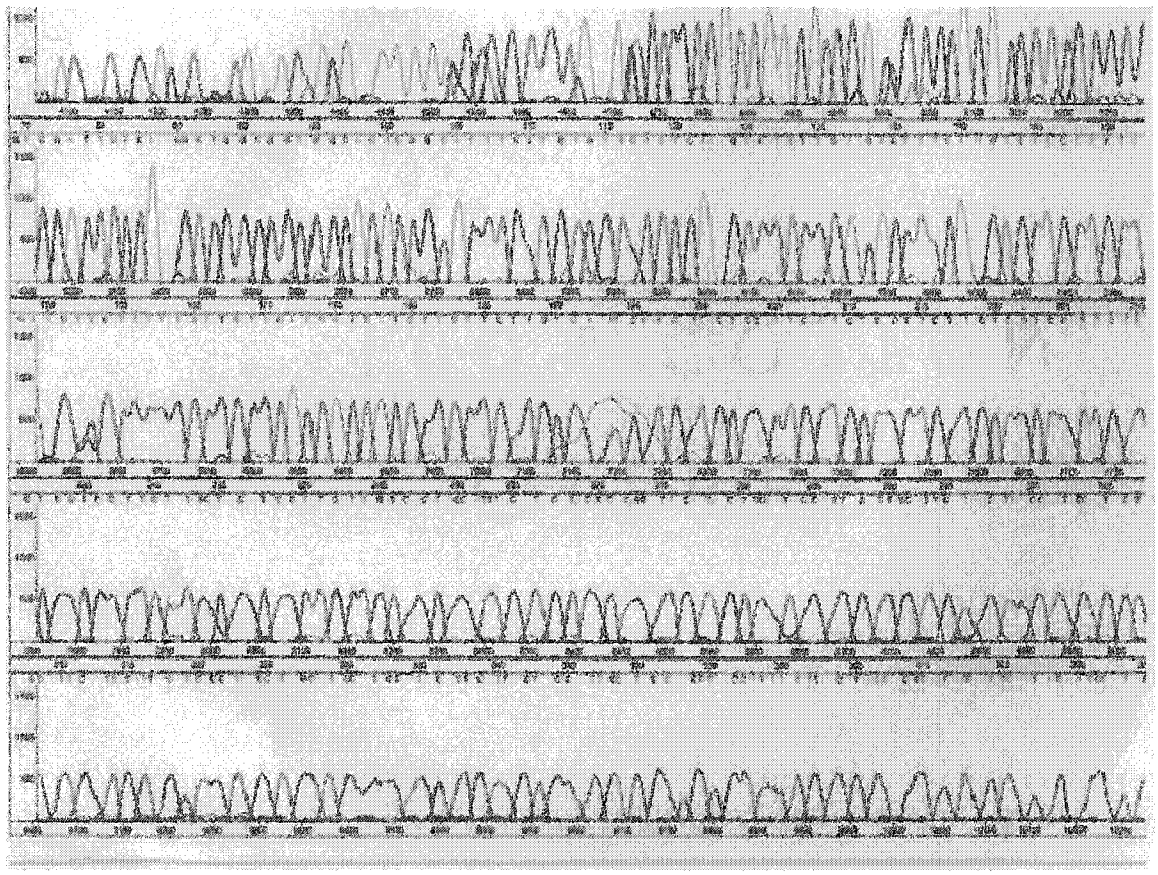


Figure 33 Phi-29 used for Gap Closure-3

The above is a trace file, obtained from the sequence of a large gap amplified by Phi-29 and sequenced on the Base station.
EXPT: To test their possibility, I investigated the standard protocol developed for plasmid amplification, but instead used primers flanking the region to amplify for sequence.
RESULTS: This approach worked exceptionally well and was able to amplify DNA that was difficult to obtain by other methods and this amplified DNA sequenced as well. It was used to close a gap in a plasmid shotgun clone that contained both "G-T and A-T" repeats that did not yield satisfactory sequencing results using typical DNA sequencing protocols.

As shown above, the Phi-29 protocol developed during the course of this research can take advantage of the high processity of the Phi-29 enzyme and produce amplified DNA from regions difficult to otherwise obtain, that, then can be used successfully for DNA sequencing.

Phi-29 BAC Experiments

After developing a robust protocol to amplify plasmid DNA using the Phi-29 mix, I then, began investigating using the Phi-29 to amplify a Mouse BAC, so that it could be PCR amplified further and then be used as a template for sequencing. This experiment worked successfully following the developed protocols (data not shown) and, therefore, I began investigating to produce templates for either "BAC-end" or shotgun sequencing instead of the labor intensive large scale BAC isolation procedure presently used. If this could be accomplished, it would be more cost efficient and produce a significant reduction in labor for this process. Such an amplification method would also increase the production rate by decreasing substantially the time requirements needed to grow and isolate the BAC DNA by one week. Therefore, I initially investigated modifying the parameters developed for the subclone, because of the 10-20 fold larger size of the insert containing BAC DNA. As a reference point, mitochondrial DNA, is about 18,000 bp, pUC subclones are about 4,000 to 6,000 bp, but a BAC insert clone can have a size range of between 130,000-250,000 bp.

Phi-29 mix: B1: Amplification Amounts needed for BAC Shotgun Sequencing

The first issue to be addressed was to determine how much cellular DNA and Phi-29 enzyme is necessary to produce a minimum of 50 µg of amplified BAC DNA, because this is the amount of DNA needed to nebulize for shotgun library construction.

Cellular DNA	Sterile double distilled Water	Phi-29 Mix	µg/well	Wells/Nebulization
2 µl	0 µl	2 µl	2.2 µg	~ 48
4 µl	6 µl	10 µl	6.3-7.3 µg	~12
5 µl	15 µl	20 µl	13.9-14.9 µg	~ 6

10 µl	10 µl	20 µl	Changes 8-33 µg	-------

Table 8 Results of the UV determination of the Phi-29 Amplified DNA for BAC Shotgun Sequencing Note: the amount of DNA and water must equal the amount of phi-29

Therefore, as shown in table 8, the DNA amplified from ~ 6 samples with varying amounts of enzyme in the typical amplification protocol was determined by measuring the absorbance of the amplified DNA. The results shown above indicate that only 2 to 5 µl of cellular DNA shall be used in the amplification reaction as when more is used inhibition is observed. If using only 2 µl of Phi-29 mix, then approximately 48 amplification wells would be necessary to produce 50 µg of DNA for nebulization. However, if the reaction mixture was increased 5 to 10 times, it would be possible to produce 50 µg of DNA for nebulization in a 6 tube reaction. This was done and the resulting DNA was nebulized and taken through the shotgun cloning protocol. However, the sequence produced was either zero length or was pUC alone without a large insert.

Phi-29 mix: B2: Using Specific Primers and Raising the Temperature for BAC-specific Amplification

To test the Phi-29 for the possibility that what was being amplified above was non-insert containing pUC or *E.coli* genomic DNA, an experiment was designed in which, the Phi-29 mix containing hexamers was incubated with pGem DNA(~ 200 ng) and universal forward and reverse primers at various temperatures.

As shown in figure 34, the Phi-29 mix that contained both hexamers and primers yielded 3 bands at 20 °C, 30 °C, 40 °C, only 2 bands at 50 °C, 60 °C, and only one band at 70 °C, 80 °C, while the reaction mixture without the universal primers produced 2 bands at 30 °C and 40 °C, the peak Phi-29 polymerizing temperature, only 1 band at 20 °C,

50°C, 60 °C , 70 °C and a smear at 80 °C. These results indicate that the Phi-29 enzyme appears to begin to denature at 70 °C and the hexamers do not bind as well as temperature increases. Therefore, at the elevated temperatures below the denaturation point, the universal primers likely were the only primers used.

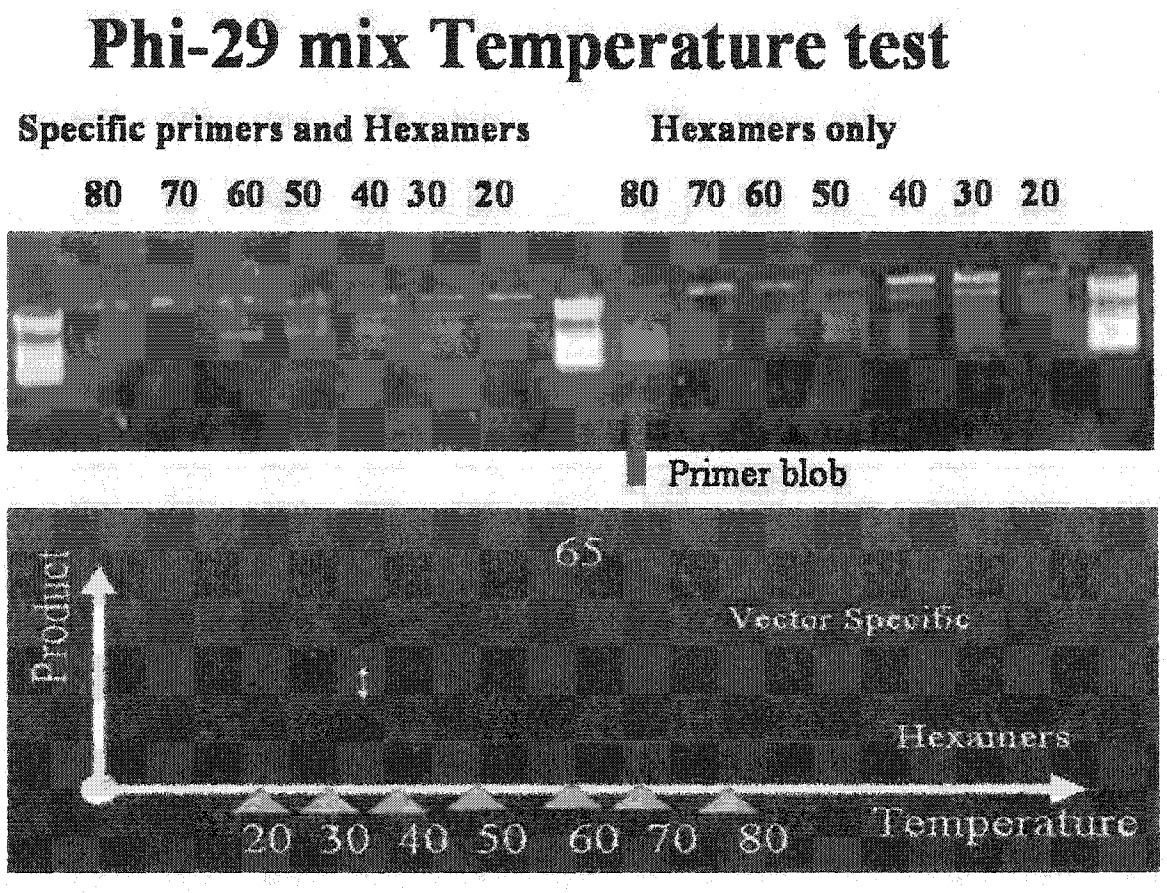


Figure 34 Results of the Phi-29, using Specific Primers and Raising the Temperature for BAC-specific Amplification

EXPT: The control, Phi-29 mix (hexamers only) was incubated with pGem DNA at 20 (room temp.), 30, 40, 50, 60, 70, and 80 °C. The DNA plus primers (6.5uM) volumes equaled the enzyme add ~2 µl of each)

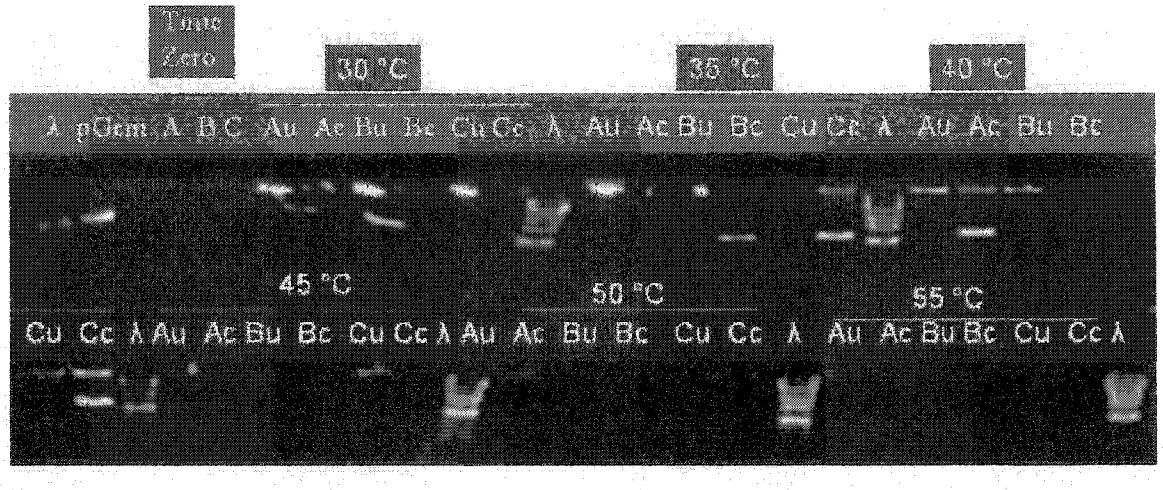
RESULTS: The Phi-29 mix that contained both hexamers and primers yielded 3 bands at 20 °C, 30 °C, 40 °C, only 2 bands at 50 °C, 60 °C, and only one band at 70 °C, 80 °C, while the reaction mixture without the universal primers produced 2 bands at 30 °C and 40 °C, the peak Phi-29 polymerizing temperature, only 1 band at 20 °C, 50°C, 60 °C , 70 °C and a smear at 80 °C. These results indicate that the Phi-29 enzyme appears to begin to denature at 70 °C and the hexamers do not bind as well as temperature increases. Therefore, at the elevated temperatures below the denaturation point, the universal primers likely were the only primers used.

Phi-29 mix: B3: Experiment Comparing the Amplification at 30 °C, 35°C, 40 °C, 45 °C, 50 °C, and 55 °C using Specific Primers.

Based on the above experimental results, another similar experiment comparing the Amplification at 30 °C, 35 °C , 40 °C , 45 °C , 50 °C , and 55 °C using specific primers, but with additional dNTPs, with and without buffer was performed. In addition, the amplified DNA from these reactions was checked for nonspecific amplification and the double strandedness via an EcoR1 restriction digestion. Three pre-mixes were used in this study and their components are given in table 9.

Label	C	B	A
DNA (pGem ~1-7 ng)	2 µl	2 µl	2 µl
Forward Primer ~ 6.5 uM	6 µl	6 µl	6 µl
Reverse Primer ~ 6.5uM	6 µl	6 µl	6 µl
DNTP s	10 µl	10 µl	0 µl
Buffer A (no primers)	76 µl	0 µl	µ
Water	0 µl	76 µl	
Total	100 µl	100 µl	100 µ

Table 9 Solutions for Phi-29 mix Experiment Comparing the Amplification at 30 °C, 35 °C , 40 °C , 45 °C , 50 °C , and 55 °C using Specific Primers.



Key:	A	B	C
Buffer	NO	NO	NO
Primer	YES	YES	YES
Dntps	NO	YES	YES

Key:	A	B	C
Uncut	Au	Bu	Cu
Cut	Ac	Bc	Cc

Figure 35 Results of the Phi-29 mix Experiment Comparing the Amplification at 30 °C, 35 °C, 40 °C, 45 °C, 50 °C, and 55 °C using Specific Primers.

EXPT: Ten µl of each of the pre-mixes was combined with ten µl of Phi-29 mixture and was incubated at the following temperatures 30 °C, 35 °C , 40 °C , 45 °C , 50 °C , and 55 °C for 12 hours. For the Ecor1 digestion, 2 µl of the amplified DNA was combined with: 2 µl of the amplified DNA was used, 2µl of Ecor1 Buffer, 1.5 µl (1ul = 10 U) of Ecor1 enzyme, and 14.5 µl sterile double distilled water. The reaction was incubated at 37 °C for 2 hours and the original DNA and the restriction digested DNA were electrophoresed on a 0.6% agrose gel at 150 mA for 15 minutes.
RESULTS: The results indicate that at a temperature below 30°C, the reactions products are as expected, the A, B, C Phi-29 reaction controls at time zero did not amplify and A, B, C Phi-29 control reactions at 30°C produce intact high molecular weight DNA that remains in the wells and the expected EcoR1 digested bands for all the conditions (Cc is hard to see in this picture, but was in the gel). At 35°C A, B, C Phi-29 reactions produce high molecular weight DNA that remains in the wells, and this DNA is EcoR1 digested under all the conditions (Ac is also hard to see here, but was in the gel). At 40°C, the A, B, C Phi-29 reactions also produce high molecular weight DNA that remains in the wells, that is EcoR1 digested, while at 45°C, 50°C, and 55°C, the A, B, C Phi-29 reactions produced little amplified DNA. The results indicate that for specific amplification using universal primers, the incubation temperature should be raised to 50-60°C, because at the higher temperature the hexamers do not efficiently anneal to the template. However, as we raise the temperature above 40°C, the Phi-29 enzyme begins to denature. Therefore another approach, such as switching to a thermophilic amplification enzyme was investigated.

The results shown in figure 35, indicate that at a temperature below 30°C, the reactions products are as expected, the A, B, C Phi-29 reaction controls at time zero did not amplify and A, B, C Phi-29 control reactions at 30°C produce intact high molecular weight DNA that remains in the wells and the expected EcoR1 digested bands for all the

conditions (Cc is hard to see in this picture, but was in the gel). At 35°C A, B, C Phi-29 reactions produce high molecular weight DNA that remains in the wells, and this DNA is EcoR1 digested under all the conditions (Ac is also hard to see here, but was in the gel). At 40°C, the A, B, C Phi-29 reactions also produce high molecular weight DNA that remains in the wells, that is EcoR1 digested, while at 45°C, 50°C, and 55°C, the A, B, C Phi-29 reactions produced little amplified DNA.

The results indicate that for specific amplification using universal primers, the incubation temperature should be raised to 50-60°C, because at the higher temperature the hexamers do not efficiently anneal to the template. However, as we raise the temperature above 40°C, the Phi-29 enzyme begins to denature. Therefore another approach, such as switching to a thermophilic amplification enzyme was investigated.

Bst Experiments

The Bst enzyme, large Klenow fragment contains the polymerization activity of *Bacillus stearothermophilus* DNA polymerase I. Bst can amplify nanogram amounts of template DNA (Mead 1991) via Strand Displacement Amplification (Kong *et al.* 1998). Bst has an optimal temperature of 65 °C (Stenesh and Roe, 1972) that would allow vector specific primers to be used. In addition, Bst was stable at 4°C to room temperature and was economically priced.

Bst pUC + Inserts Experiments

In the initial experiment the reaction condition for Bst amplification was studied. Here, the reagents normally found in a PCR reaction were used as a starting point. This was done, and a reaction mix for the Bst enzyme containing: 2 µl of Bacillus stearothermophilus DNA polymerase large fragment (16 units), 3 µl mix of forward (each primer 5nM/40ul); 3 µl mix of reverse primer (each primer 5 nM/40 u); 5 µl of 2 mM DNTPs; ~31 µl water, and 5 µl of 10x Reaction buffer was used to amplify the pGem vector.

Bst: P1: Amplification from pGem Gold (Isolated no insert) DNA

As shown in figure 36, ~1 µg of pGem that was heated to 95 °C for three minutes and cooled to 4 °C, then, combined with the enzyme mix containing, ~16 unites of Bst in a total volume of 50 µl and incubated for ~6 hours at 65 °C .

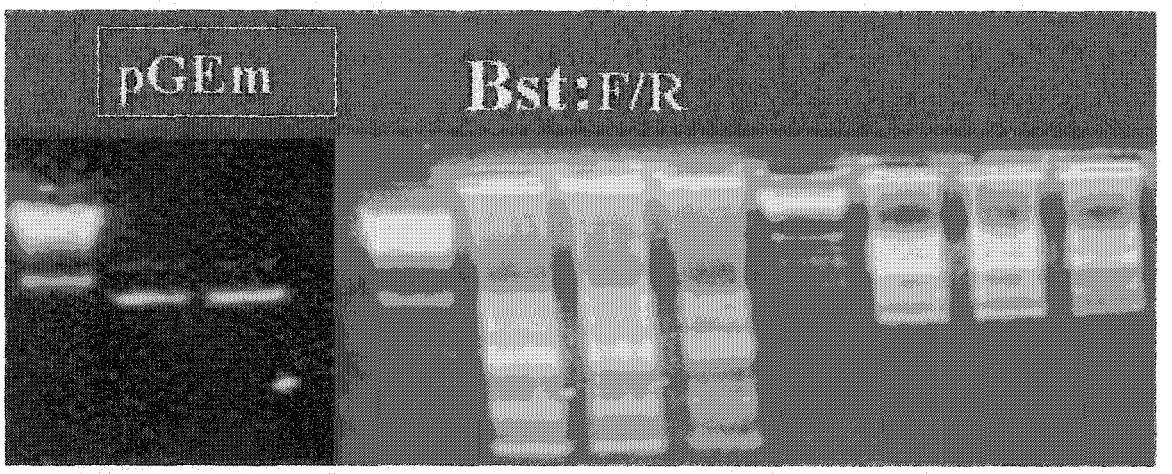


Figure 36 Results of the Bst Amplification from pGem Gold (Isolated no insert) DNA

EXPT: ~1 µg of pGem that was heated to 95 °C for three minutes and cooled to 4 °C, then, combined with the enzyme mix containing, ~16 unites of Bst in a total volume of 50 µl and incubated for ~6 hours at 65 °C .

RESULTS: Bst successfully amplified the plasmid DNA, producing high molecular weight DNA that remained in the wells. Even though, the amount of DNA and enzyme used was far in excess this initial experiment demonstrated that Bst could indeed produce high molecular weight DNA, but also produce a high level of short DNA fragments.

Bst successfully amplified the plasmid DNA, producing high molecular weight DNA that remained in the wells. Even though, the amount of DNA and enzyme used was far in excess this initial experiment demonstrated that Bst could indeed produce high molecular weight DNA, but also produce a high level of short DNA fragments.

Bst BAC Experiments

Bst: B1: Amplification of BAC DNA from Isolated DNA

Next we tested if Bst could amplify BAC DNA using the above reaction components. This was done by heating 15 ng of BAC DNA to 95 °C for three minutes, cooling to 4 °C, adding the enzyme mix, ~16 unites of Bst and incubating at 65 °C for 18 hours in a total reaction volume of 50 µl.

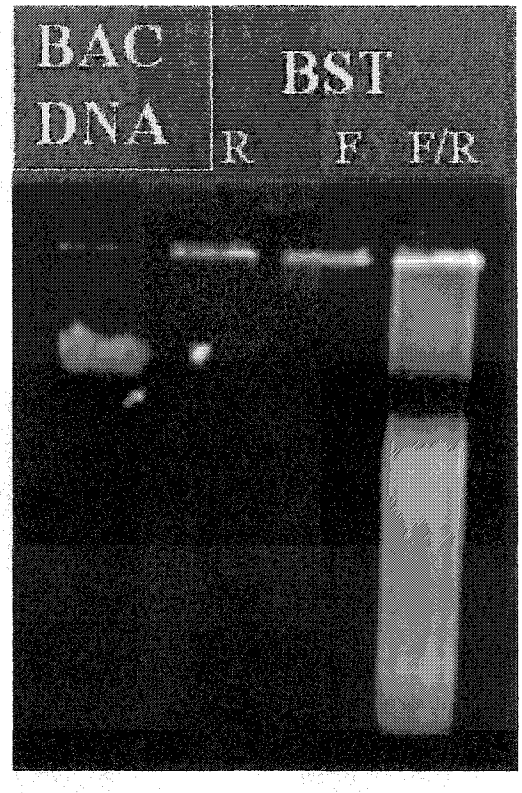


Figure 37 Results of the Bst Amplification of BAC DNA from Isolated DNA

EXPT: The Bst was tested to see if Bst could amplify BAC DNA using the above reaction components. This was done by heating 15 ng of BAC DNA to 95 °C for three minutes, cooling to 4 °C, adding the enzyme mix, ~16 unites of Bst and incubating at 65 °C for 18 hours in a total reaction volume of 50 µl.
RESULTS: it appears that when both forward and reverse primers are present the product contains a high level of short DNA fragment that produces a smear of DNA on the gel. However, when using only one primer only high molecular weight DNA is produced. When only one primer is used the DNA is replicated only off one strand. However, when two primers that prime and amplify in opposite directions are used simultaneously, replication occurs on both strands. Thus every time the amplification reaches the region complementary to the other primer, it is possible for replication to proceed in the opposite direction. This produces a cascade of DNA amplification in both directions and the amplification ultimately exhausts the available dNTPs with many fragments of amplified template of varying length.

As can be seen in figure 37, it appears that when both forward and reverse primers are present the product contains a high level of short DNA fragment that produces a smear of DNA on the gel. However, when using only one primer only high molecular weight DNA is produced. When only one primer is used the DNA is replicated only off one strand. However, when two primers that prime and amplify in opposite directions are

used simultaneously, replication occurs on both strands. Thus every time the amplification reaches the region complementary to the other primer, it is possible for replication to proceed in the opposite direction. This produces a cascade of DNA amplification in both directions and the amplification ultimately exhausts the available dNTPs with many fragments of amplified template of varying length.

Bst: B2: Amplification from Cells

The next issue to be addressed was to determine if the Bst could amplify from BAC colonies. Since the first step in this procedure entailed robotically pipetting colonies, we investigated the number of times the colony containing probe was dipped into the wells of the microtiter plate. Since, ten dips were normally used in the lab for regular growth of subclones, therefore, we began with this number. However, as shown in figure 38, this transferred too many cells and caused inhibition of the Bst amplification. However, when only three dips were used, the amplification reaction was successful.

We therefore, concluded that an inhibitor from the picked colonies was being added to the solution from which an liquate of lysed cells and the level of this inhibitor could be satisfactorily reduced by reducing the amount of cells transferred.

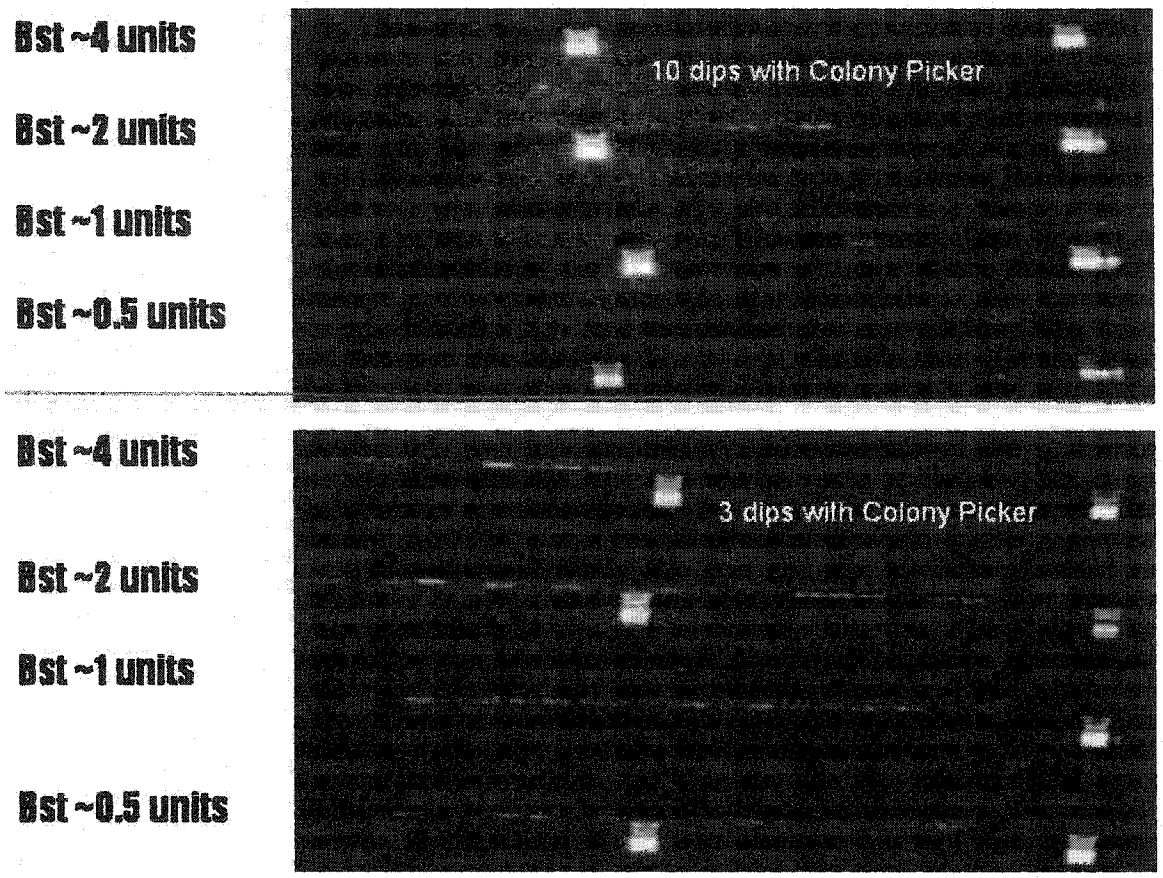


Figure 38 Bst Results from Amplification using differing numbers of dips from the Colony Picker.

EXPT: The figure above is Bst amplified DNA amplified according to the protocol in the Methods that was electrophoresed on a 0.6% agrose gel at 150 mA for 15 minutes.
GEL (Top): The Bst was amplified using cellular DNA from 10 dips of the colony picker with the Bst reactions in descending order as positioned across on the gel: using the Bst at 4 units, 2 units, 1unit, and 0.5 units.
GEL (Bottom): The Bst was amplified using cellular DNA from 3 dips of the colony picker with the Bst reactions in descending order as positioned across on the gel: using the Bst at 4 units, 2 units, 1unit, and 0.5 units.
RESULTS: Since, ten dips were normally used in the lab for regular growth of subclones, this was the starting point, however, this transferred too many cells and caused inhibition of the Bst amplification. When only three dips were used, the amplification reaction was successful, therefore, it was concluded that an inhibitor from the picked colonies was being added to the solution from which an liquate of lysed cells and the level of this inhibitor could be satisfactorily reduced by reducing the amount of cells transferred.

Bst: B3: Amplification Buffer

To test the effect of the buffer {The buffer comes as a 10x concentration, and it is diluted down, the reaction buffer at 1X contains: 1X ThermoPol Buffer (20 mM Tris-HCl pH 8.8, @ 25°C), 10 mM KCl, 10 mM $(NH_4)_2SO_4$, 2 mM $MgSO_4$, 0.1% Triton X-100}

(Kong *et al.* unpublished), and ionic strength on the Bst and subsequent amplification on a series of buffer dilutions were tested using 0.25x, 0.5x, and 1x buffers. As can be seen in figure 39, the Bst enzyme seemed to perform optimally, when in the presence of the 5x buffer, which gave a final reaction concentration of ~3.5x.

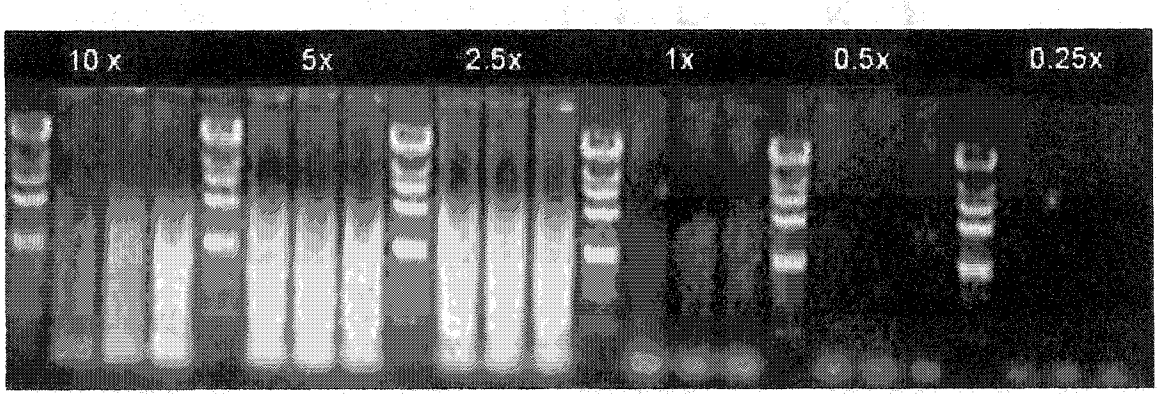


Figure 39 Bst Amplification Results from various Concentrations of Buffer using 0.25x, 0.5x, and 1x buffers.

EXPT: To test the effect of the buffer and ionic strength on the Bst and subsequent amplification on a series of buffer dilutions were tested using 0.25x, 0.5x, and 1x buffers. The figure above is Bst amplified DNA amplified according to the Methods protocol that was electrophoresed on a 0.6% agrose gel at 150 mA for 15 minutes.
RESULTS: The Bst enzyme seemed to perform optimally, when in the presence of the 5x buffer, which gave a final reaction concentration of ~3.5x.

Bst: B4: New Bst Reaction Protocol Confirmed on BACs from Cells and Glycerol Stock.

The protocol shown at the beginning with the buffer change (which is the protocol in the Methods section) was used to amplify from cellular DNA and from glycerol stocks as seen in figure 40. Their seemed to be no real difference between DNA amplified from colonies vs. DNA amplified from glycerol stocks.

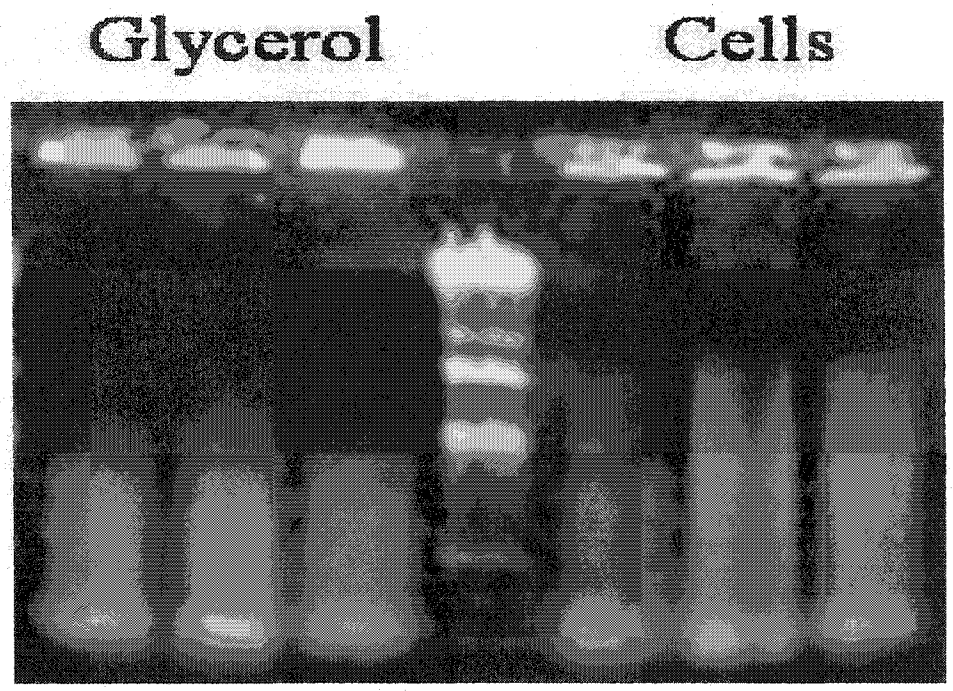


Figure 40 New Bst Reaction Protocol Confirmed on BACs from Cells and Glycerol Stock separated by HindIII cleaved λ DNA marker.

EXPT: The figure above is Bst amplified DNA amplified from glycerol stocks according to the Methods protocol that was electrophoresed on a 0.6% agrose gel at 150 mA for 15 minutes.
RESULTS: Their seemed to be no real difference between DNA amplified from colonies vs. DNA amplified from glycerol stocks.

Bst: B5: Amplification vs. Time

The time course of the Bst amplification of the BAC DNA then was investigated using the Standard Bst protocol for BAC amplification in a 96 well thermocycler plate, comparing the effects of adding additional dNTPs and including an additional 10 minute at 72°C incubation step.

The results of the Amplification are seen in the figure 41, where the amplification results were linear over the time period and both the additional dNTPs and dNTPs with added time yielded ~2-fold more amplified product than the standard protocol.

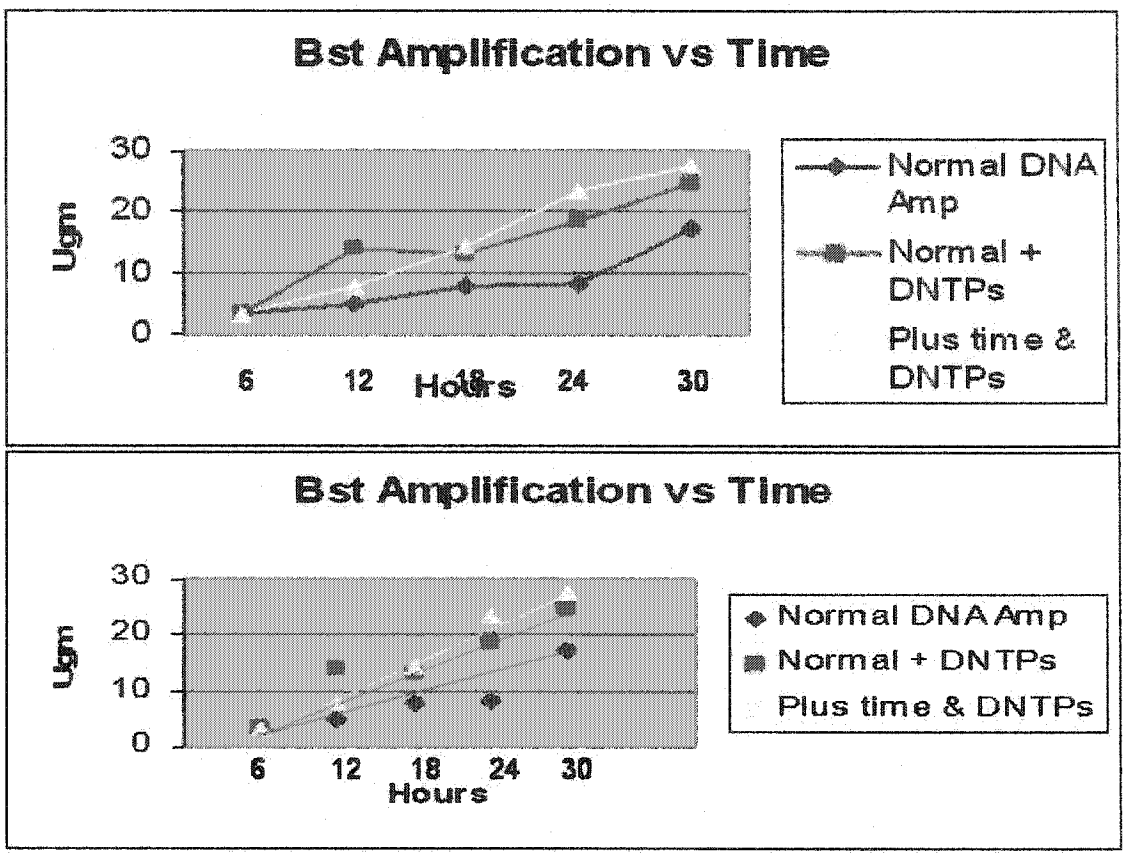


Figure 41 Bst Amplification vs. Time

EXPT: 3 µl of forward (primer 200uM), 3 µl of reverse primer (primer 200uM), 5 µl of 2mM DNTPs, 35 µl 5x Reaction buffer (50 % 10x Buffer and 50% sterile double distilled water) (~3.5 X reaction buffer – total dilution) were combined with 2ul of cellular BAC DNA and 1 unit of Bst polymerase and incubated in the thermocycler for 6, 12, 18, 24 and 30 hours at 65 °C with the volume setting of 50 µl. As the times indicated the temperature was reduced to 55 °C for two hours and then, heated to 95 °C for 5 minutes to denature the enzyme, then held at 4 °C before removal to prevent evaporation. Then, 50 µl of sterile double distilled water was added to each 50µl reaction and after mixing, ~30µl of the 1:2 diluted completed reactions were desalted by passage through a Sephadex G-50 plate, further diluted 100 fold and the A_{260} was measured.
RESULTS: The amplification results were linear over the time period and both the additional dNTPs and dNTPs with added time yielded ~2-fold more amplified product than the standard protocol.

Bst: B6: Sodium Shearing the Amplified BAC DNA for Shotgun

I next investigated amplifying DNA and then shotgun cloning it for sequencing. This was to determine if, indeed, the BAC DNA was amplified. However, because I had earlier shown that it was extremely difficult to dissolve ethanol precipitated Phi-29 or Bst

amplified DNA, I investigated an alternative approach to hydrolyze the DNA by heating it in an aqueous solution at high temperatures (Marguet and Forterre, 1994). It also is known that salt protects DNA against depurination, by stabilizing the helix, and by preventing autocatalysis (Marguet and Forterre, 1998). Therefore, the salt concentration study shown in figure 42 was performed to investigate the effect of a 90 °C heating of a salt solution to effectively dissolve and shear the amplified DNA test was done to find the correct amount of salt and water for DNA.

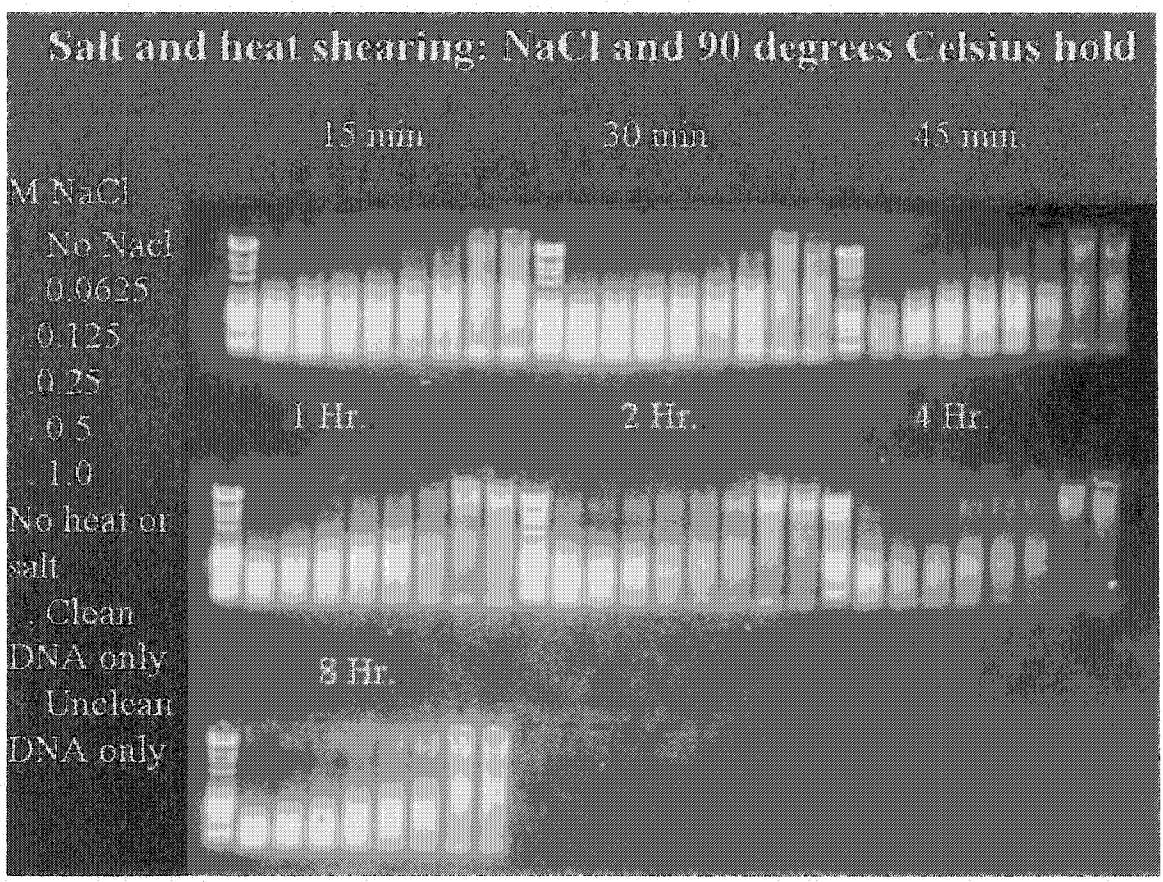


Figure 42 Bst test Salt and Heat Shearing.

EXPT: The salt concentration study shown in figure 43 was performed to investigate the effect of a 90 °C heating of a salt solution to effectively dissolve and shear the amplified DNA test was done to find the correct amount of salt and water for DNA.
At the indicated, aliquats of DNA samples incubated in water alone and in 0.125, 0.25, 0.5 M NaCl appeared equal and 0.1 M NaCl were electrophoresised on a 6% agrose-ethidium bromide gel, photographed under UV light.
RESULTS: 1.0 M NaCl for 45 minutes produced DNA fragments of the desired size range.

DNA degrades as the time increases, but this allowed for the DNA shearing at 1.0 M NaCl for 45 minutes to produce DNA fragments of the desired size range.

Since ~1.0 M NaCl for 45 minutes produced DNA fragments of the desired size range, we then, investigated heat sheared DNA was an acceptable substrate for the fill-in kinase reaction and if it was necessary to desalt the DNA prior to the fill-in step. Therefore the experiment outlined in table 10 was performed and after incubation for 45 minutes at 37 °C for the standard fill-in kinase reaction, the DNA was electrophoresed on a gel to confirm the size. See below. Next, the DNA was electrophoresed on a low melt gel, DNA in the 1-2 kb size range was excised from the gel and after ligation and transformation, the resulting colonies were counted to yield the results shown in table 11.

DNA Sample	Salt/ No Salt	Water at 90°C for 45 minutes	Clean for Fill-in Kinase Reaction ?No	Fill-in Kinase Reaction /No
DNA Sample 1-A	No	Water @ 90°C-45 min	No	Fill-in Kinase
DNA Sample 1-B	No	Water @ 90°C-45 min	No	No
DNA Sample 2-A	0.1M NaCl	Water @ 90°C-45 min	Clean	Fill-in Kinase
DNA Sample 2-B	0.1M NaCl	Water @ 90°C-45 min	Clean	No
DNA Sample 3-A	0.1M NaCl	Water @ 90°C-45 min	No	Fill-in Kinase
DNA Sample 3-B	0.1M NaCl	Water @ 90°C-45 min	No	No
DNA Sample 4-A	No	No	No	Fill-in Kinase
DNA Sample 4-B	No	No	No	No

Table 10 Bst Heat Shearing, Cleaning and Fill-in Kinase Reaction Experiment

DNA Sample	White Colonies
DNA Sample 1-A	6
DNA Sample 1-B	63
DNA Sample 2-A	347

DNA Sample 2-B	12
DNA Sample 3-A	12
DNA Sample 3-B	10
DNA Sample 4-A	76
DNA Sample 4-B	46

Table 11 Cloning efficiency of the amplified DNA described in table 10: Bst Heat Shearing, Cleaning and Fill-in Kinase Reaction Experiment

Based on the results shown in table 11, cells produced from the sample that were desalted prior to the kinase fill-in sample 2A was more likely to be subcloned when the DNA was heat sheared in water for 45 minutes with salt and ethanol precipitated to remove the salt before doing the fill-in kinase reaction, the shotgun cloned DNA from these cells was obtained and sequenced using the standard procedure used routinely for shotgun sequencing template isolation and over 90 % of the 384 sequencing reactions yielded an average Phred 20 read > 600 bases as shown in figure 43.

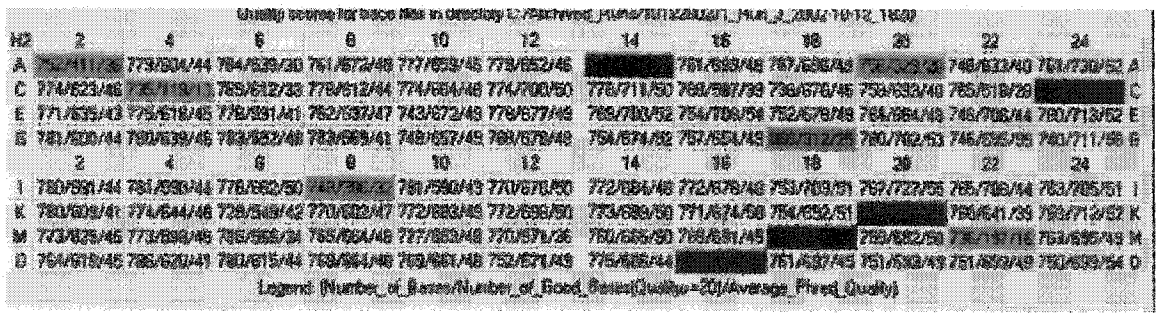


Figure 43 Sequencing Results from the Bst Heat Shearing, Cleaning and Fill-in Kinase Reaction
All the 384 plates were similar to the above of DNA sample 2A.

EXPT: The cells produced from the 2A sample were desalted prior to the kinase fill-in and subcloned when the DNA was heat sheared in water for 45 minutes with salt and ethanol precipitated to remove the salt before doing the fill-in kinase reaction, the shotgun cloned DNA from these cells was obtained and sequenced using the standard procedure used routinely for shotgun sequencing template isolation.
RESULTS: Over 90 % of the 384 sequencing reactions yielded an average Phred 20 read > 600.

After analyzing the reaction products as shown in figure 44, the major sequence product matched stuffer fragment from the BAC. The stuffer fragment should have been removed and replaced with the DNA insert, but based on these results, it is likely that the clones in the pBACe3.6 based library contain both BACs with insert and intact BAC containing the stuffer fragment, but without insert. Because the stuffer with the stuffer fragment is only ~ 12,000 bases vs. the vector with insert is ~140,000, the Bst polymerase preferentially amplified the smaller vector, refer to figure 47.

Bst amplified BAC and Zymark isolated subclone sequence

Sequence statistics:

Problem : % pUC was too high

	1	2	3
Avg length (For >300 nt)	666.2	3659.2	654.5
Avg length (For All)	607.72	619.5	567.2
Total traces	96	96	96
Zero length	7	5	9
< 300 nt	2	1	5
> = 300 nt	87	90	82
> pUC	65	68	59
> Ecoli	1	0	0

Figure 44 Sequencing Stats of the Results from the Bst Heat Shearing, Cleaning and Fill-in Kinase Reaction

EXPT: Refer to the information given in figure 43.
RESULTS: the major sequence product matched stuffer fragment from the BAC. The stuffer fragment should have been removed and replaced with the DNA insert, but based on these results, it

is likely that the clones in the pBACe3.6 based library contain both BACs with insert and intact BAC containing the stuffer fragment, but without insert. Because the stuffer with the stuffer fragment is only ~ 12,000 bases vs. the vector with insert is ~140,000, the Bst polymerase preferentially amplified the smaller vector,

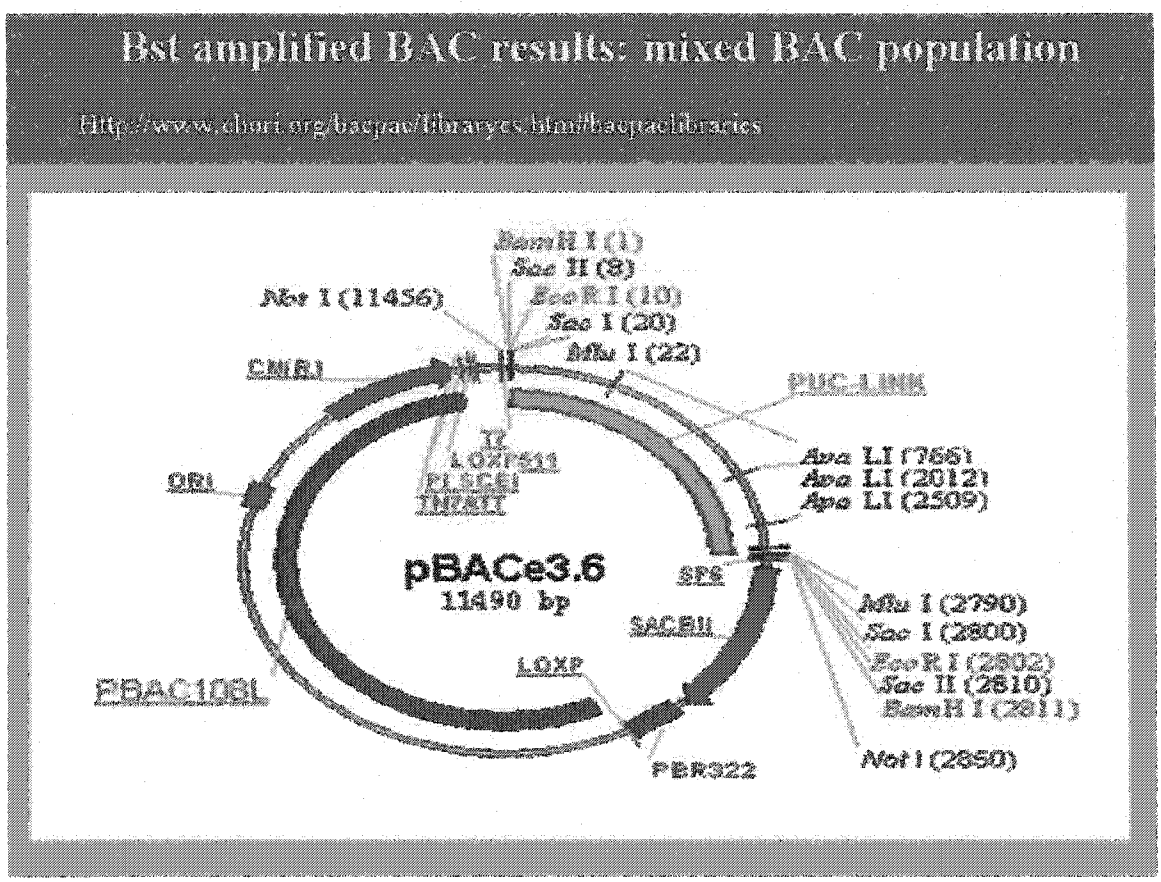


Figure 45 pBACe3.6 DNA: Vector + Stuffer

RESULTS: The stuffer fragment should have been removed and replaced with the DNA insert, but based on these results, it is likely that the clones in the pBACe3.6 based library contain both BACs with insert and intact BAC containing the stuffer fragment, but without insert. Because the stuffer with the stuffer fragment is only ~ 12,000 bases vs. the vector with insert is ~140,000, the Bst polymerase preferentially amplified the smaller vector,

Bst for Gap Closure

Because Bst is a highly processive enzyme (Hugh and Griffin 1994; McClary *et al.* 1991), it can efficiently replicate GC rich regions, hairpin loops and long repeats, producing useful templates for sequence through difficult regions permitting gaps to be

filled with real sequence rather than PCR estimated guess sequence. This hypothesis was tested by using the standard protocol for plasmid amplification to amplify a difficult to sequence region of a mouse BAC and then use the PCR products, gap flanking primers to PCR the amplified region and, then sequence.

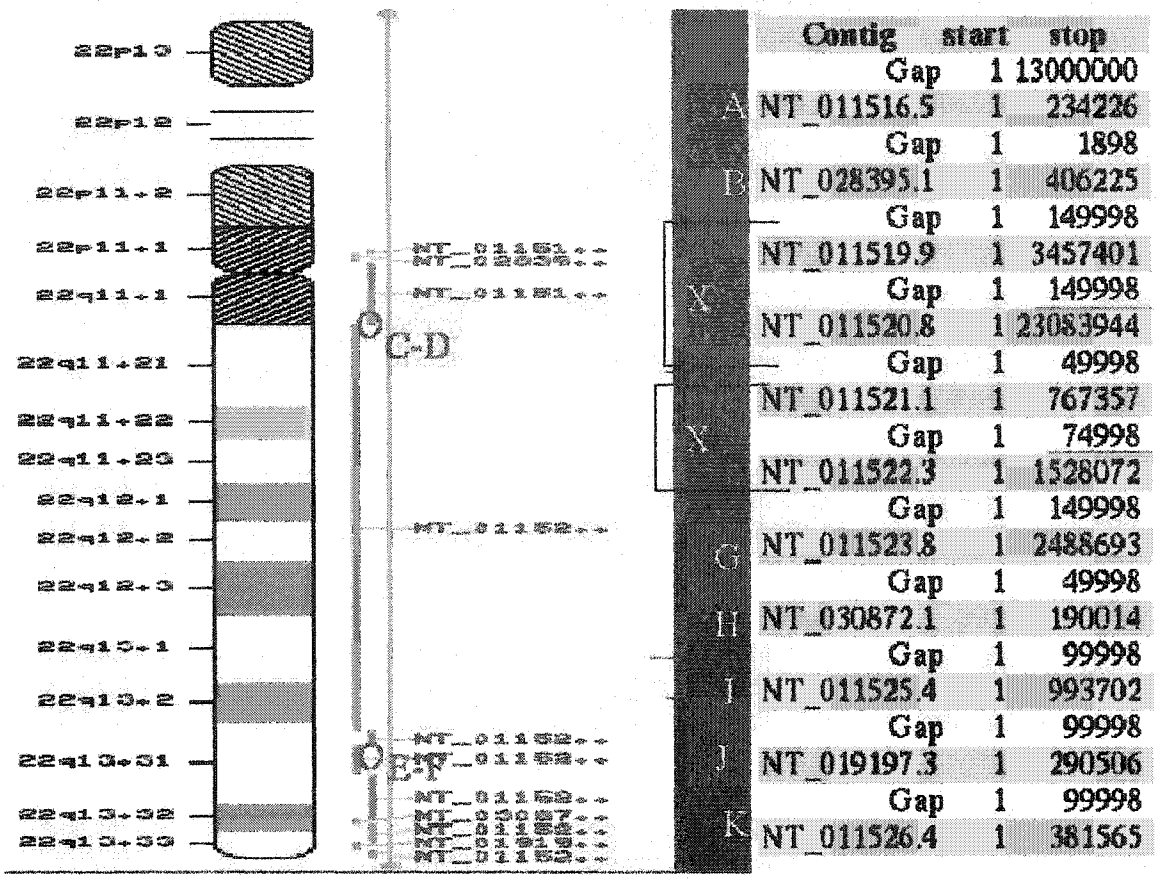
The results from an experiment are shown in figure 48, where the Bst was used as explained above to close a gap and lower the projects' percent error rate.

Figure 46 Bst used for GapClosure-1

The above is a Consed view, the reference reads ending in Bst and colored purple are the Bst gap closure sequences.
EXPT: The figure above is Bst amplified DNA amplified according to the protocol in the Methods for BAC amplification.
RESULTS: can efficiently replicate GC rich regions, hairpin loops and long repeats, producing useful templates for sequence through difficult regions permitting gaps to be filled with real sequence rather than PCR estimated guess sequence.

Bst and Human Chromosome 22 Experiments

It was believed that the technology did not exist to fill in the gaps at the time of the sequencing of the chromosome 22 (In October of 1999, Dunham, Senior Research Fellow of the Sanger Center in Cambridge, England, said, "the current technology would not allow the team to fill all the gaps..."Chromosome 22gaps; (http://wwwgrg.org/chrom22.htm.). It also was believed that the sequence in the gaps could not be cloned as at least seven remain, not for lack of trying, but because, these regions are unstable, making cloning them impossible (DeFrancesco, 2001). Based on the hypothesis just successfully tested, i.e. that the Bst could amplify difficult regions, chromosomal gap closure was attempted.



Contig	start	stop
Gap	1	13000000
NT_011516.5	1	234226
Gap	1	1898
NT_028395.1	1	406225
Gap	1	149998
NT_011519.9	1	3457401
Gap	1	149998
NT_011520.8	1	23083944
Gap	1	49998
NT_011521.1	1	767357
Gap	1	74998
NT_011522.3	1	1528072
Gap	1	149998
NT_011523.8	1	2488693
Gap	1	49998
NT_030872.1	1	190014
Gap	1	99998
NT_011525.4	1	993702
Gap	1	99998
NT_019197.3	1	290506
Gap	1	99998
NT_011526.4	1	381565

Figure 47 Human Chromosome 22 Contigs and Gaps

The above shows the gap size and human chromosome 22 position that was used for Bst amplification testing. (Red "O" represents gap position on the chromosome, yellow "X" represents the gap, to show the size of the gap between the red contig letter above and the red contig letter below.
(http://www.ncbi.nlm.nih.gov/mapview/maps.cgi?ORG=hum&MAPS=ideogr,est,loc&LINKS=ON&VERBOSE=ON&CHR=22)

I attempted to amplify the two gaps in the human chromosome 22 sequence. Therefore, primers were chosen on the gap ends outside of the repeats 12 forward and 12 reverse on each end approximately 200 bases a part totaling at least 2400 bases farthest from the end. One round of PCR confirmed that the primer pairs would amplify the

correct gap flanking region, as shown schematically in figure 48.

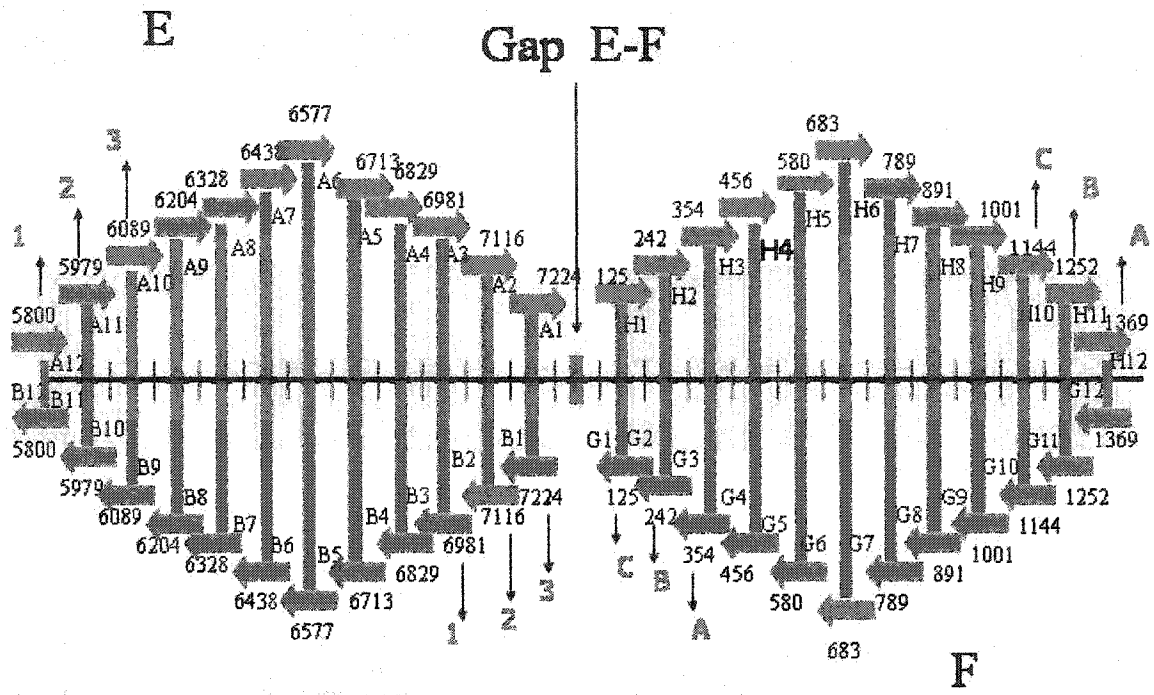


Figure 48 Chromosome 22 Primer Set-up Tool and Example

The # above or below the arrow (primer) represents the distance away from the gap contig end. X= the addition of two primers in a reaction.

After, the primers were confirmed to produce PCR product, they were used to amplify across the gap region via a nested PCR strategy. Here, the two primers farthest away from each other were used to amplify the gap region with Bst Then, after purification of the product on a low melting agrose gel, a second amplification reaction would be done using primers that were the second farthest from the ends. Then, after a gel purification the last nested amplified product, it was amplified a third time using primers that were the third farthest from the ends. This process is shown schematically in figure 51. Then, after snap freezing the DNA in liquid nitrogen and allow it to slowly

thaw to room temperature a on ice, a shotgun library of the resulting in small clonible sized double stranded DNA was produced, and after the DNA was isolate, it was sequenced.

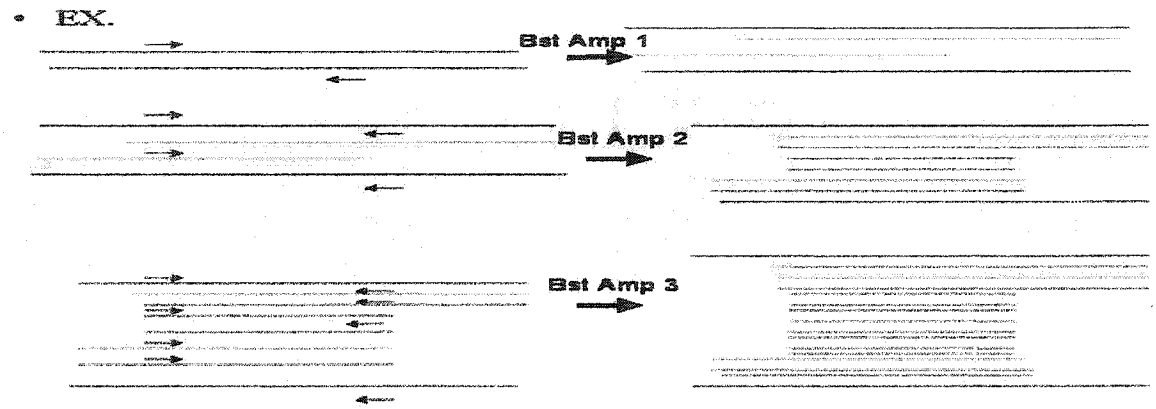


Figure 49 A schematic of Multiple rounds of nested amplification used for Chromosome 22 Gap Closure

The above is amplification using the Bst and nested primers. Bst doesn't PCR, it is very processive, but since the opposite primer should anneal to the amplified DNA, the template should shrink to only our desired DNA quickly.

Bst: Chr 22-1: The Flipped Contig Experiment

In order to sort of "PCR" the gap of interest using Bst and non-unique primers 12 primers were chosen to increase the odds of having the primers anneal at the desired spot. The Primers at ~200 bp intervals were picked using PrimOU and chromosome 22 DNA using a TM of ~60 ^{o}C with a GC content of 40-60. The gap used for this closure experiment was surrounded by contigs NT_011520.8 and contig NT_011521.1. The forward primers are numbered based on NIH's numbering, the list of forward primers were chosen starting from the gap and going reverse on contig NT_011520.8 and the reverse of this on contig NT_011521.1. Complementary primers were chosen to be used to make the controls, to sequence the end of the amplified DNA to confirm the

amplification of the desired DNA. The closest primer to the gap was used for sequencing the DNA, since the other primers would not anneal to all templates.

Thus, a PCR reaction confirmed that the desired DNA was amplified and a control to confirm the assembly of the chromosome at the gap edge. The PCR reaction primer sets (from above) are in the spreadsheet below the primer list labeled A-H with a picture below them to help explain the reactions. Those listed with "comp" means the primers is going in the opposite direction. The primer list is given in the Appendix II.

PCR primer list:

Reaction Letter	First Primer	Second Primer	Type of Reaction	~Distance between Primers (bases)
A	F23083825.comp	R000221.comp	FC X RC	X (shouldn't work)
B	F23082848	R001579	F X R	731
C	F23083561	R001579	F X R	1018
D	R001457	R000754.comp	R X RC	707
E	F23083825.comp	F23083210	FC X F	615
F	F23083825	F23083311.comp	F X FC	X (shouldn't work)
G	R000221.comp	R000754	RC X R	533
H	R000221	R000754.comp	R X RC	X (shouldn't work)

Table 12 Actual PCR primer list

Key: F= forward primer, FC = forward compliment primer, R = reverse primer, RC = reverse compliment primer, and X = PCR reaction with the two primer abbreviations listed.

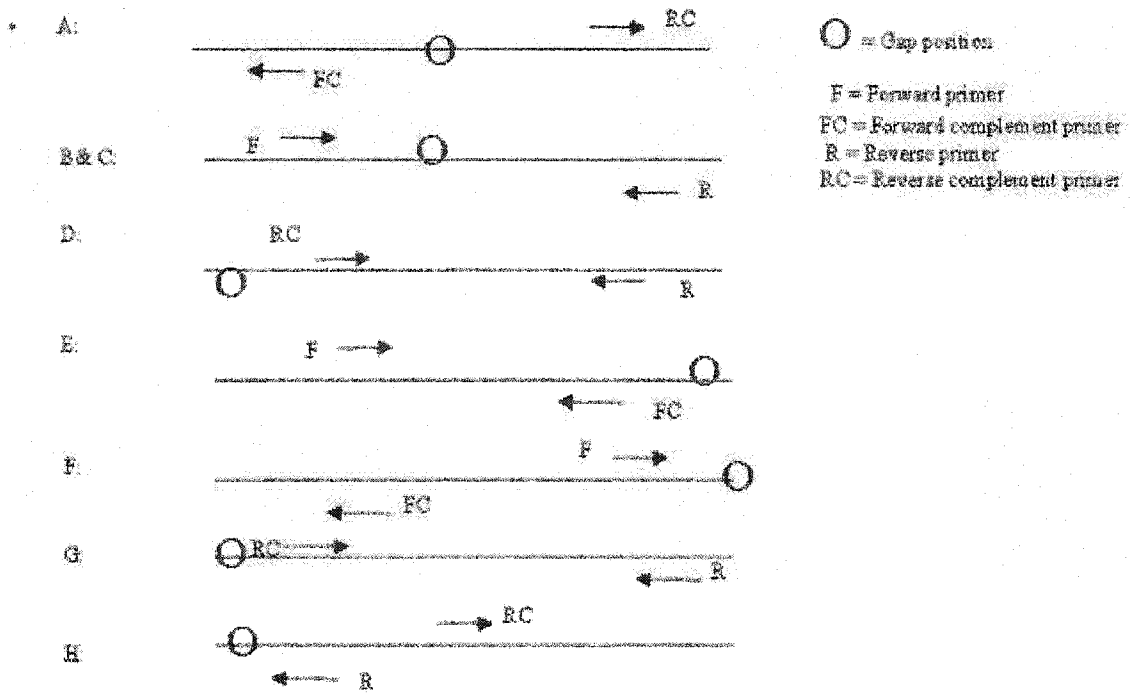


Figure 50 Chromosome 22-PCR Confirm Contigs Reactions

EXPT: refer to figure 49.
RESULTS: none of the above reactions produced PCR products except for reaction F. Thus, it is possible that the forward contig had an additional sequence on the end of the contig, where the primer anneals that is reversed and complimented.

Surprisingly, none of the above reactions produced PCR products except for reaction F. Thus, it is possible that the forward contig had an additional sequence on the end of the contig, where the primer anneals that is reversed and complimented, i.e. flipped, and thus miss-assembled. If part of the forward contig is flipped, then it would produce the following results: Reaction A: both primers going the same direction- no product; B and C: both primers going the same direction- no product; D : Concerning R and RC as long as there are no miss-assemblies on the reverse side these reactions should work, check primers and assembly; E: both primers going the same direction- no product; G: Concerning R and RC as long as there are no miss-assemblies on the reverse side these reactions should work, check primers and assembly; H: Not possible, the primers are going the opposite directions unless there is an assembly error.

The using PCR to check primers, primers for Gaps C-D and E-F were confirmed (not shown here) and the primers used for Gaps C (NT_011519.9) to D (NT_011520.8) and E (NT_011521.1) to F (NT_011522.3) are give in Appendix II.

Bst: Chr 22-2: Liquid Nitrogen

After using PCR to check primers (results are not shown here, the above was an example of how this was done) for primers for Gaps C-D and E-F were confirmed and experiment was done, after amplifying the DNA, liquid Nitrogen was used to snap freeze the DNA and shear the double stranded DNA into clonible pieces, followed by a slow reannealing step to help reannealing the sides.

Since heat cleaves DNA as well, after amplifying the DNA, the sample was divided into two portions, one of which was heated it to 95 °C for the denaturing of the enzyme. This also produces separation of the DNA strands. (This could be argued, due to the amplification temperature being 65 °C, however, the DNA temperature must reach 70 °C in order to completely denature, refer to the figure 51.) The other half of DNA was amplified at 65 °C, without denaturing of the enzyme; this temperature is not high enough to totally produce single stranded DNA.

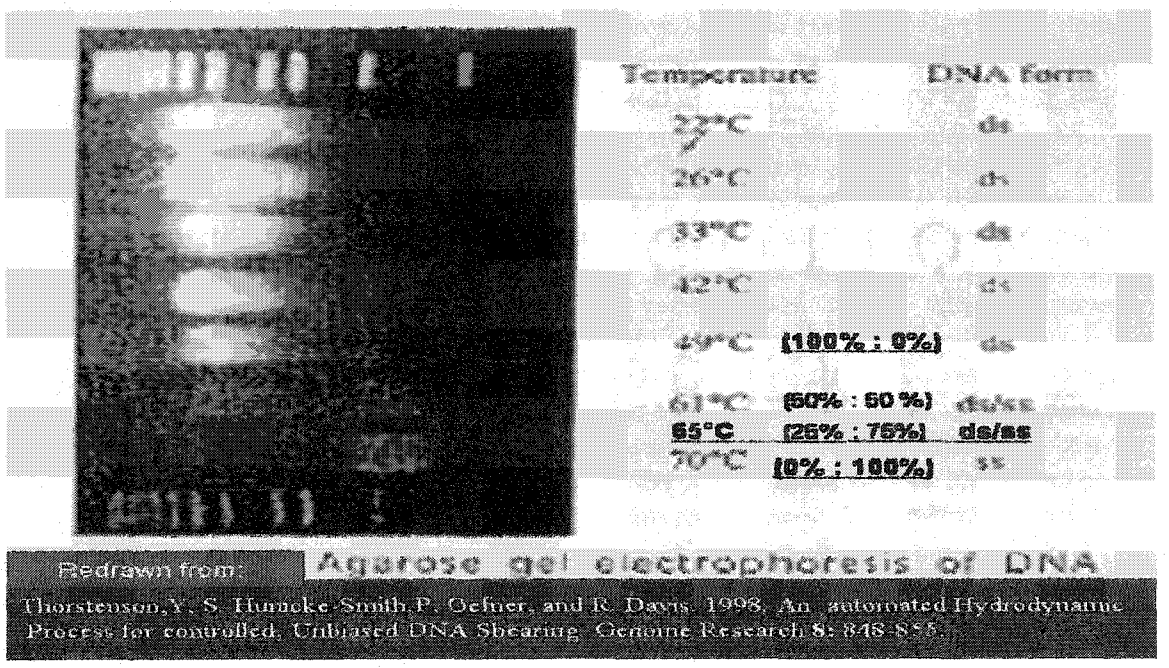


Figure 51 DNA at different Temperatures.

The amplification temperature was 65 °C (~75% single stranded DNA), however, the DNA temperature must reach 70 °C in order to completely denature.

A comparison between single stranded/double stranded, and double stranded-only DNA with snap freezing of the DNA with liquid nitrogen therefore would be the next logical experiment..

Therefore, a BAC (7n16 mouse), Human Gap C to D and E to F DNA were amplified using Bst in a 96 well thermocycler plate using the standard Bst protocol, but with only one pair of primers. Next, half the samples should be heated to 95 °C for 5 minutes, the other samples should be transferred to 250 µl Eppendorf tubes. After heating these samples they also were transferred to 250 µl Eppendorf tubes. A comparison between single stranded/double stranded, and double stranded only DNA, desalted vs. untreated DNA, before freezing with liquid nitrogen, then was made.

Liquid nitrogen was added to the tubes in a Styrofoam cup ~3 ml just enough to cover the tubes; they should immediately freeze, and then placed in the -180 freezer until ready

to continue the experiment. When ready, the tubes were placed in an ice water bath in the cold room for ~1 hour to thaw slowly. After thawing run the Bst amplified DNA on a 0.6% low melt gel using 1.5 volts and 100 mA for 30 minutes, then increase the amperage to 150 mA for 1.5 hours.

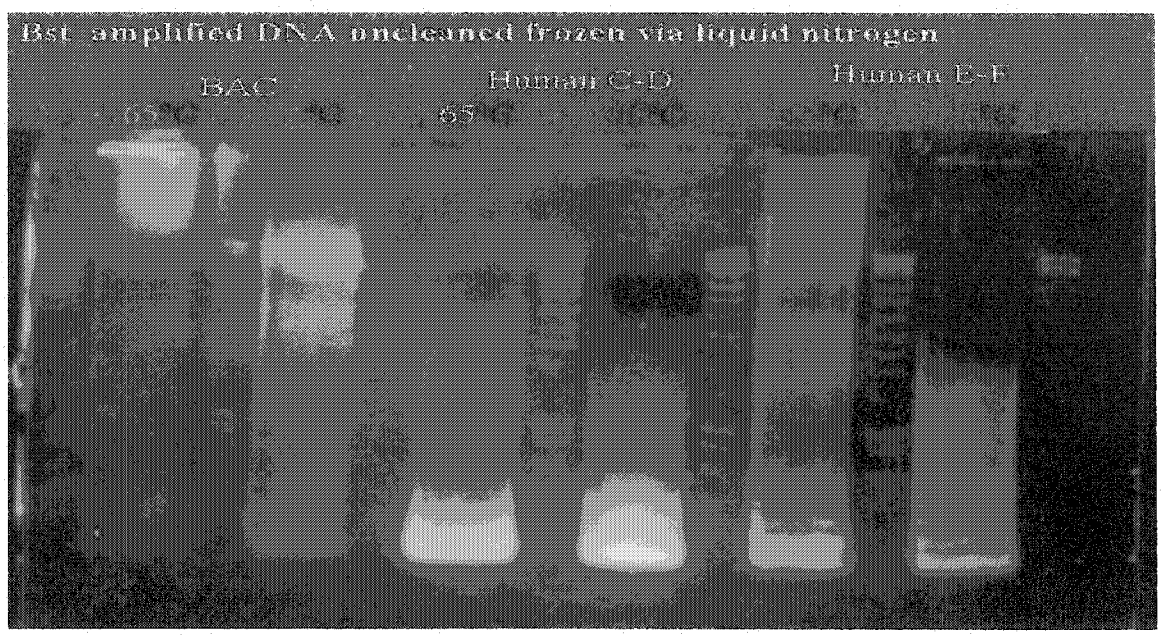


Figure 52 Bst Amplified DNA, Broken via Liquid Nitrogen

The above shows the above show the results of the snap freezing of the plus and minus the addition of 95 °C temperature to denature the enzyme. No denaturing was done for the 65 °C samples and the denatured DNA of course is the 95 °C samples.
EXPT: The figure above is Bst amplified DNA amplified from BAC/Chromosomal amplifications: BAC (7n16 mouse), Human Chr. 22 Gap C to D and Human Chr. 22 Gap E to F DNA were amplified using Bst in a 96 well thermocycler plate using the standard Bst protocol, but with only one pair of primers. Half the samples were heated to 95 °C for 5 minutes, the other half were stopped at 65°C and a comparison between single stranded/double stranded, and double stranded-only DNA, desalted vs. untreated DNA, before freezing with liquid nitrogen, then placed in an ice water bath in the cold room for ~1 hour to thaw slowly. After thawing run the Bst amplified DNA on a 0.6% low melt gel using 1.5 volts and 100 mA for 30 minutes, then increase the amperage to 150 mA for 1.5 hours.
RESULTS: What appears to be a double stranded band under the BAC DNA that was denatured with 95 °C can be observed, the denaturing step produces a better breaking of the DNA, and the BAC DNA that had been heated to 95°C had formed the ~9,000 bp band. Clearly liquid Nitrogen snap freezing shears the DNA and is reannealible if after the amplification process a 95°C denaturing has taken place.

The figure 52, shows the results of the snap freezing and plus and minus the addition of 95 °C temperature to denature the enzyme. No denaturing was done for the 65 °C samples

and denaturation was done at 95 °C. As shown in figure 52, what appears to be a double stranded band under the BAC DNA at 95 °C can be observed, the denaturing step produces a better breaking of the DNA, and the BAC DNA that had been heated to 95°C had formed the ~9,000 bp band. Clearly liquid Nitrogen snap freezing shears the DNA and is reannealible if after the amplification process a 95°C denaturing has taken place.

Next, the DNA was sheared to 1,000 to 4,000 base paired fragments for subcloning using the standard shotgun cloning procedure as described for isolated BAC DNA (Roe, ed.1997), sequenced with the standard Amersham ET protocol and analyzed by a comparison with the genbank NR database using blastn. Typical results of the blast output from this analysis is shown in figure 53 and the multiple sequence alignment is shown in figure 54. Both figures show that clearly, human DNA was sequenced from the Bst amplification (not primer-dimers) and that some of the amplification does indeed match the gap contig end sequence.

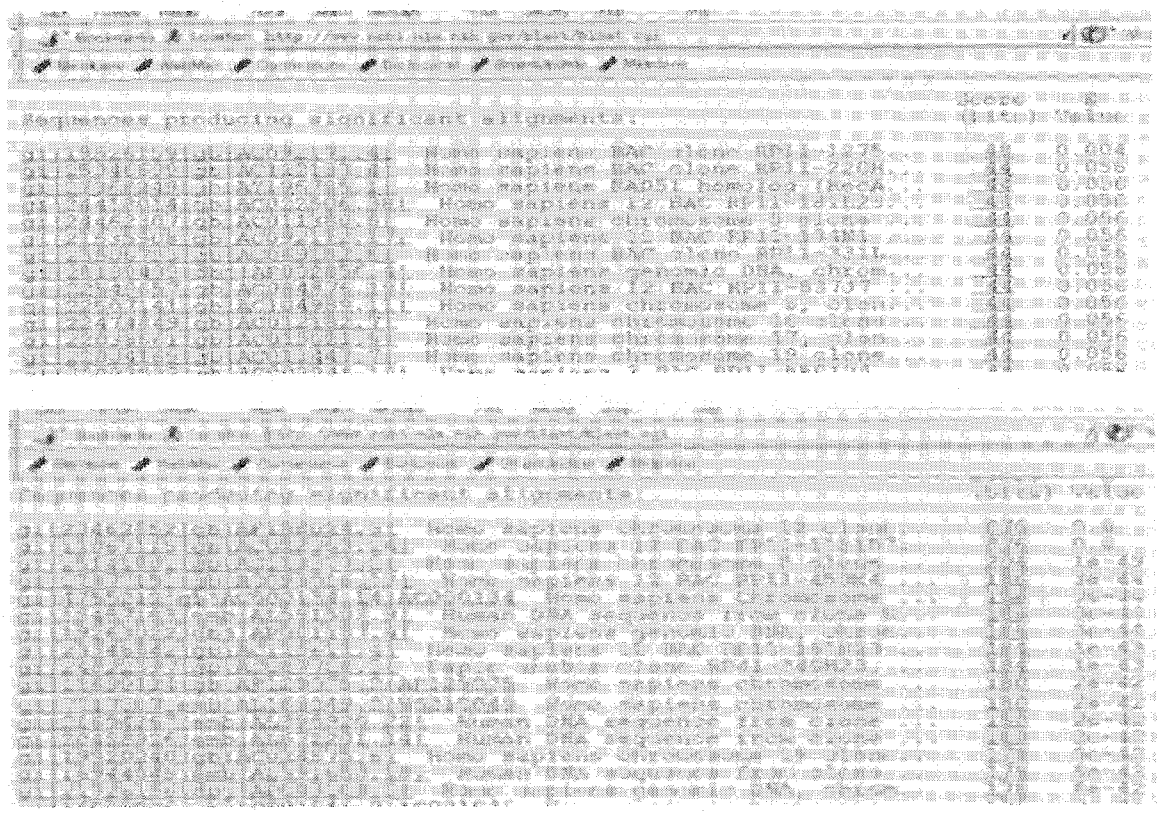


Figure 53 Blast analysis of Bst amplified, sheared, subcloned, and sequenced chromosome 22 gap region.

RESULTS: That human DNA was sequenced from the Bst amplification, not primer-dimers.

Figure 54 Multiple sequence alignment of a portion of the Bst amplified gap region Comparison of New Sequence to Current Chromosome 22Gap Ends region.

RESULTS: That some of the amplification sequence does indeed match the gap contig end sequence.

Before, blasting the gap sequence against the known human chromosome 22 sequence, the first ~10, 000 bases on the ends of the gaps that was used to pick primers was used to get a control of where they anneal when blasted against the Human Genome, where the sequence is similar to other sequence on the Human Genome. As shown in the figure 55, this gap contig ends' sequence has similarities to regions of several chromosomes, therefore even though the DNA anneals to another chromosome, it may still be from the amplified gap of interest on chromosome 22.

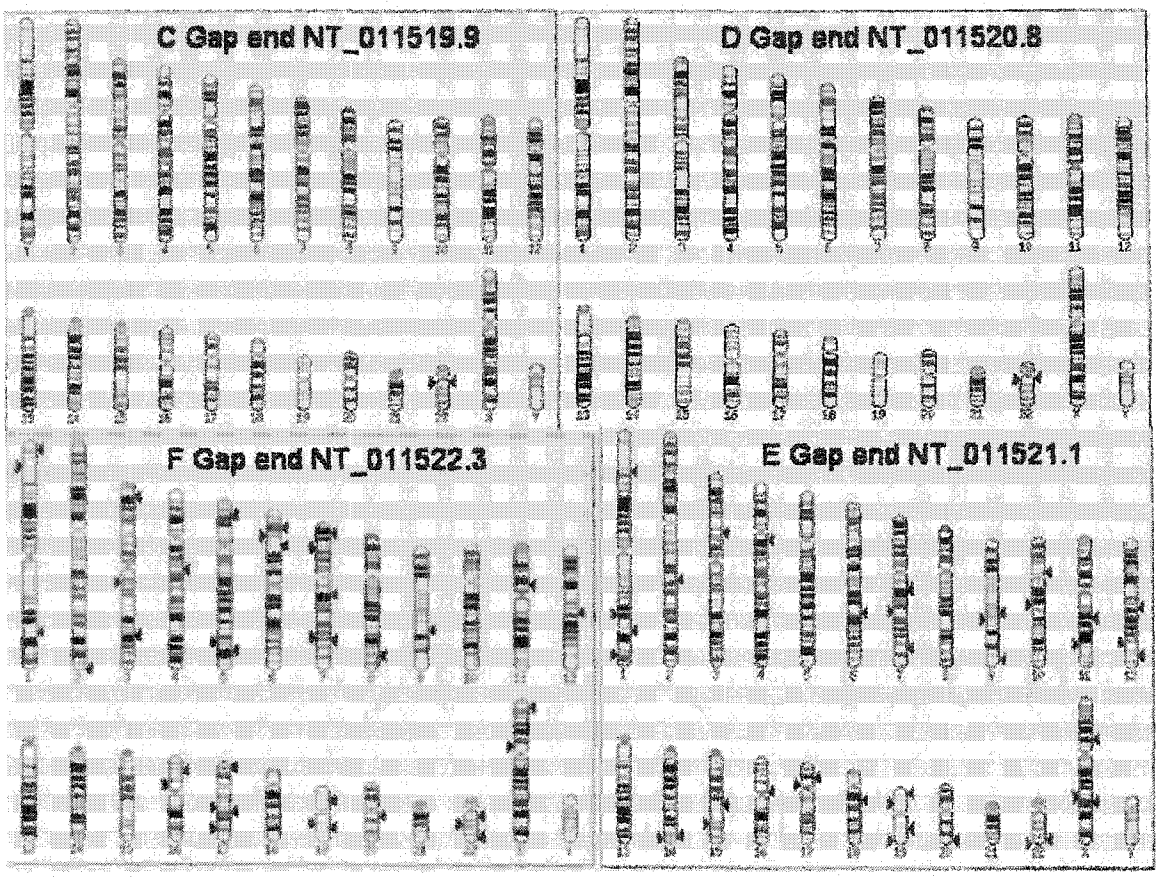


Figure 55 Genome under analysis of Bst amplified chromosome 22 gap sequences obtained via a comparison to the presently available human genome sequence.

C gap end NT_011519.9 sequence in the human genome is similar to D gap end NT_011520.8 for the C-D gap and the F gap end NT_011522.3 sequence and E gap end NT_011521.1 sequence in the human genome is similar to many chromosomes sequence. The results above are from the Human SSAHAView (http://www.ensembl.org/Homo_sapiens/ssahaview).
EXPT: Before, blasting the gap sequence against the known human chromosome 22 sequence, the first ~10, 000 bases on the ends of the gaps that was used to pick primers was used to get a control of where they anneal when blasted against the Human Genome, where the sequence is similar to other sequence on the Human Genome.
RESULTS: The gap contig ends' sequence has similarities to regions of several chromosomes, therefore even though the DNA anneals to another chromosome, it may still be from the amplified gap of interest on chromosome 22.

As can be seen in figure 55, the sequences of the contigs immediately flanking the C (NT_011519.9) and D (NT_011520.8) of the C-D gap were specific for gap region of human chromosome 22, while the F (NT_011522.3) and E (NT_011521.1) of the F-E gap flanking sequence were similar to sequence throughout the genome and therefore, nonspecific.

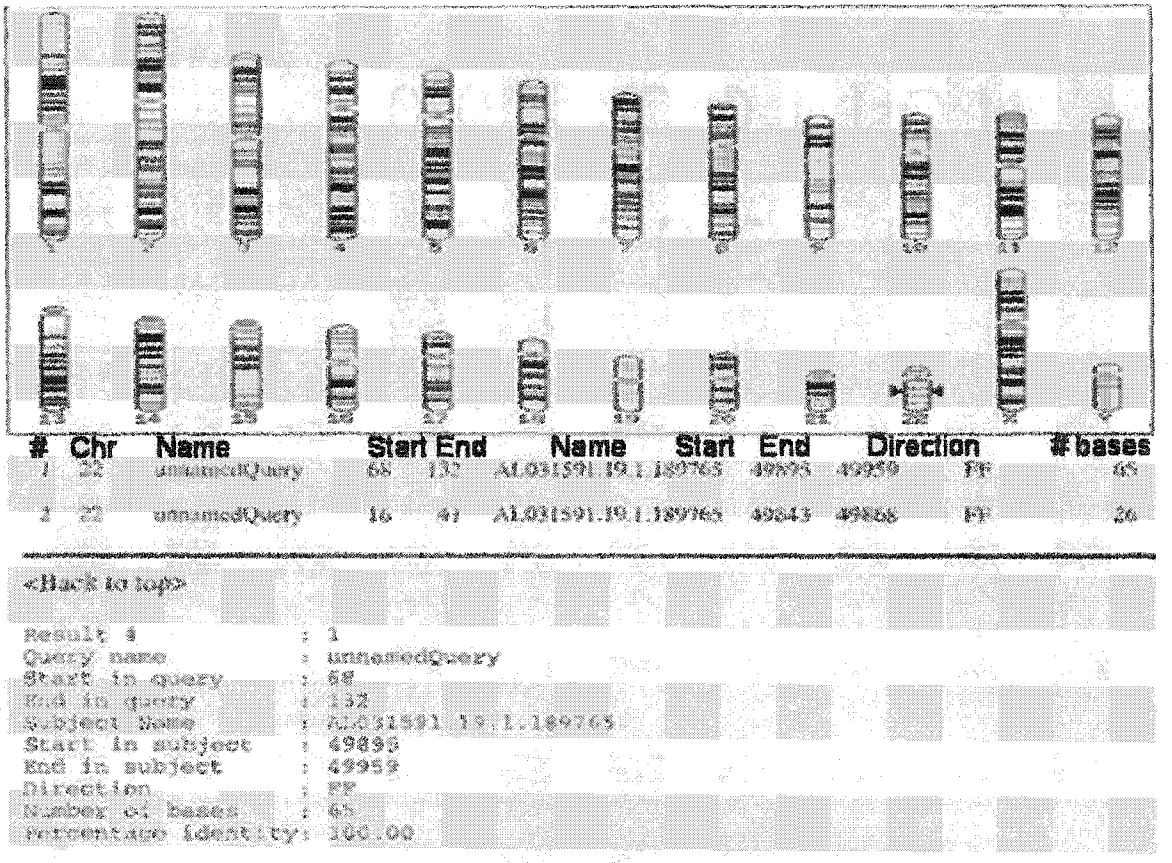


Figure 56 Sequence determined for the Bst amplified C-D gap correlating to human chromosome 22.

The above show DNA sequence form Bst amplified Human Chr. 22 sequence.
RESULTS: Clearly, the Bst was able to amplify the human chromosome 22 gaps, because sequence was produced that clearly are on the ends of the human chromosome 22 gaps and are clearly specific to human chromosome 22 only.

Based on the above results in figure 56, Bst was able to amplify the human chromosome 22 gaps, because sequence was produced that clearly are on the ends of the human chromosome 22 gaps and are clearly specific to human chromosome 22 only. However, based on the results shown in figure 57, the gap sequence results revealed highly repetitive DNA in the gap, assembly of the sequence was impossible using Phred/Phrap, even though assembly was done at a repeat stringency of 0.9998.

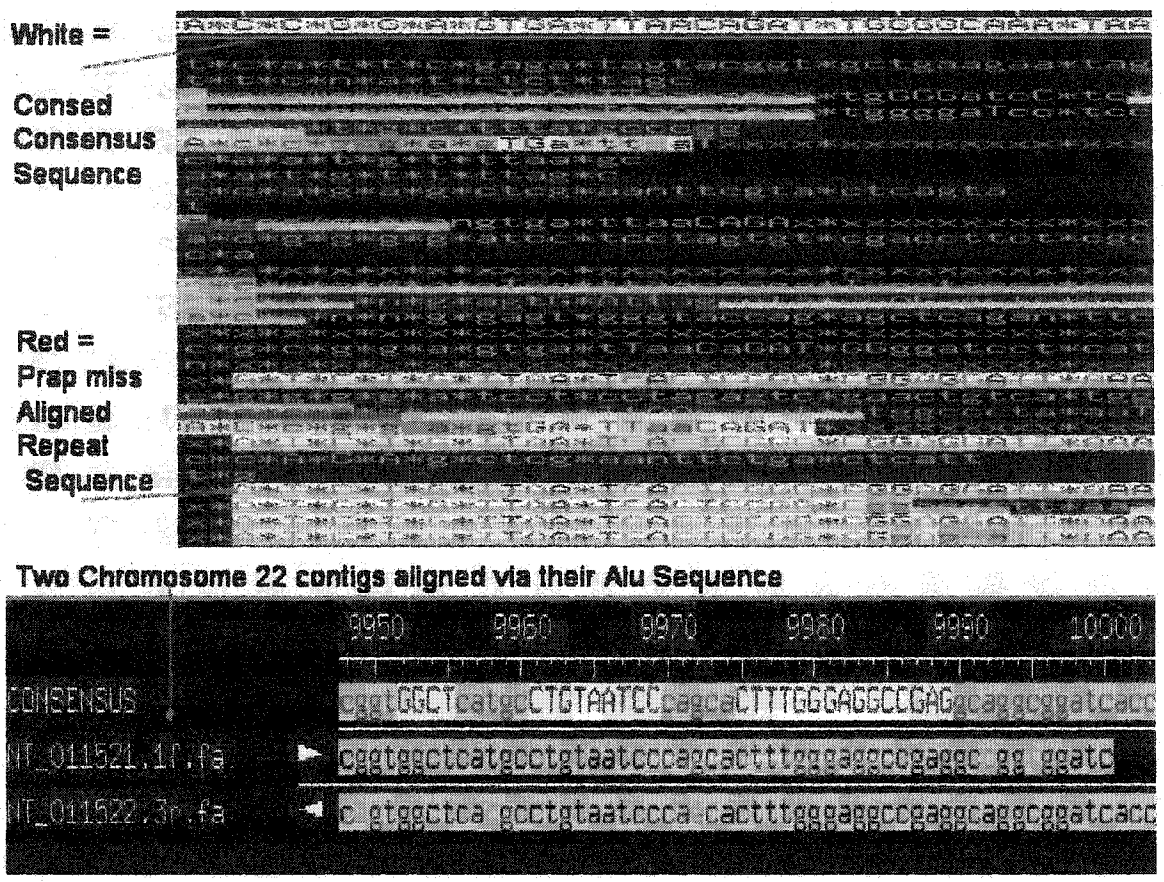


Figure 57 Chromosome 22 and Phrap Alu problems

RESULTS: A typical gap assembly from the amplified gap region, in the top portion of the figure, the numerous SNPs or miss- aligned sequence, while the bottom portion of the figure show two gap end contigs that are miss aligned because of the presence of alu sequences in both. The gap sequence results of highly repetitive DNA in the gap, made assembly of the sequence impossible using Phred/Phrap, even though assembly was done at a repeat stringency of 0.9998.

A typical gap assembly from the amplified gap region is shown in figure 57, in the top portion of the figure, the numerous SNPs or miss- aligned sequence, while the bottom portion of the figure show two gap end contigs that are miss aligned because of the presence of alu sequences in both.

Thus, because of these observations, it became obvious that it would be impossible to close the gaps remaining in human chromosome 22, even through this Bst

amplification approach, because the current programs for assembling the contigs were not capable of resolving the high repeat content of these gap regions.

Bst BAC-end Sequencing

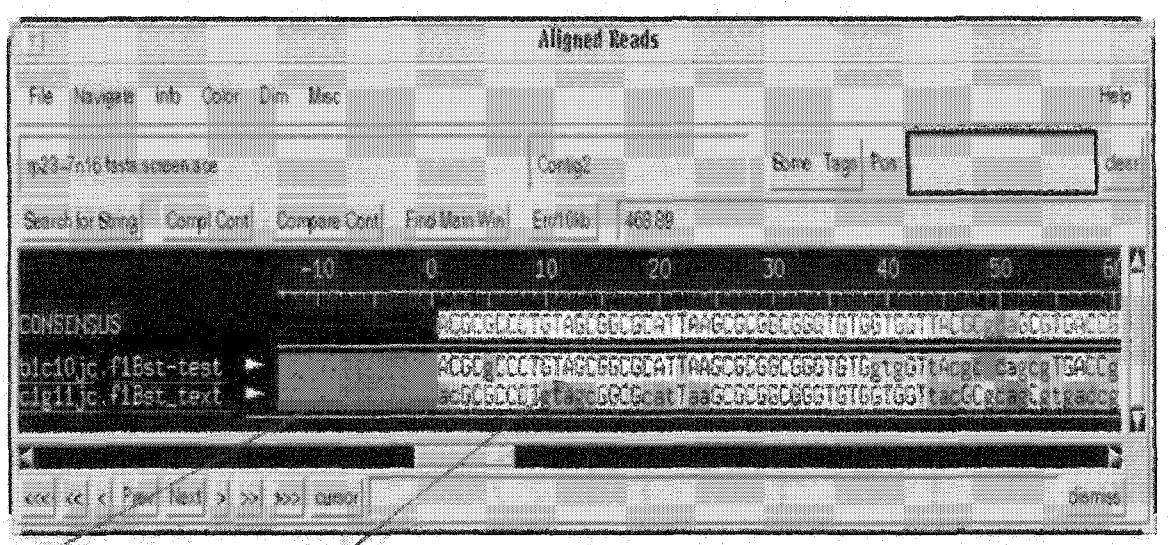
It was at this point that it was decided that to abandon the human chromosome 22 gap closure project and re-investigate amplification-based BAC end sequencing. An efficient reproducible protocol for BAC end sequencing still was needed because the present procedures were both expensive and marginally reproducible. The same protocols used to amplify the BAC using Bst for shotgun thus were used, with only two vector specific primers. Initially, we used the standard Sp6 and T7 primers, but later, when it became obvious that the primer sequence was causing primer-dimer problems, primers picked using PrimOU were used.

As shown in the figure 58, the Consed view, shows the Bst trace files produced BAC end vector sequence shown in gray, and high quality white insert sequence. The screenout view (Bottom), shows the sequenced files that were screened out (e.g. b1c10.f1Bst-test) due to conservation via crossmatch with the BAC end vector sequence (e.g. pBACe3.6), both are underlined in green.

Bst could successfully amplify and produce sequencible BAC DNA.

The Exgap output is shown in figure 59, where the dark blue circle represents the insert and the light blue region is the BAC vector. The top portion of figure 59, shows only one light blue vector region, while the bottom portion of the Exgap output shows both, the other end was produced, when the Bst amplified sequence data was added to the

project. The Exgap in figure 59 compares to the Bst trace files underlined in red to the BAC end vector , light blue, to prove the origin of the new vector sequence. Thus, sequencing the amplified DNA in this BAC clone revealed both BAC-vector-insert junction sequences.




```
 34  4.88 0.00 2.44  b1c10jc.f1Bst-test      71  111 (865)  C puc18    (2418)  268  228

148  1.32 0.00 0.00  b1c10jc.f1Bst-test     144  235 (481)  C pBACe3.6 (8891) 2721
2570

 55  0.00 0.00 0.00  c1g11jc.f1Bst_text      51  105 (679)  C puc18    (2403)  283  229
157  0.62 0.00 0.00  c1g11jc.f1Bst_text     127  288 (496)  C pBACe3.6 (8891) 2731 2570
 38  0.00 0.00 0.00  c1g11jc.f1Bst_text     745  784 (0)    C pBACe3.6 (9045) 2567 2528
```

Figure 58 Bst BAC end Sequencing: Mouse pBace3.6 rp23-7n16 Results.

Consed view (Top), shows the Bst trace files (under lined in red) produced BAC end vector sequence shown in gray, and high quality white insert sequence.
The screenout view (Bottom), shows the sequenced files that were screened out (e.g. b1c10.f1Bst-test) due to conservation via crossmatch with the BAC end vector sequence (e.g. pBACe3.6), both are underlined in green.
EXPT: The same protocols used to amplify the BAC using Bst for shotgun thus were used, with only two vector specific primers. Initially, the standard Sp6 and T7 primers were used, but later, when it became obvious that the primer sequence was causing primer-dimer problems, primers picked using PrimOU were used.
RESULTS: Bst could successfully amplify and produce sequencible BAC DNA.

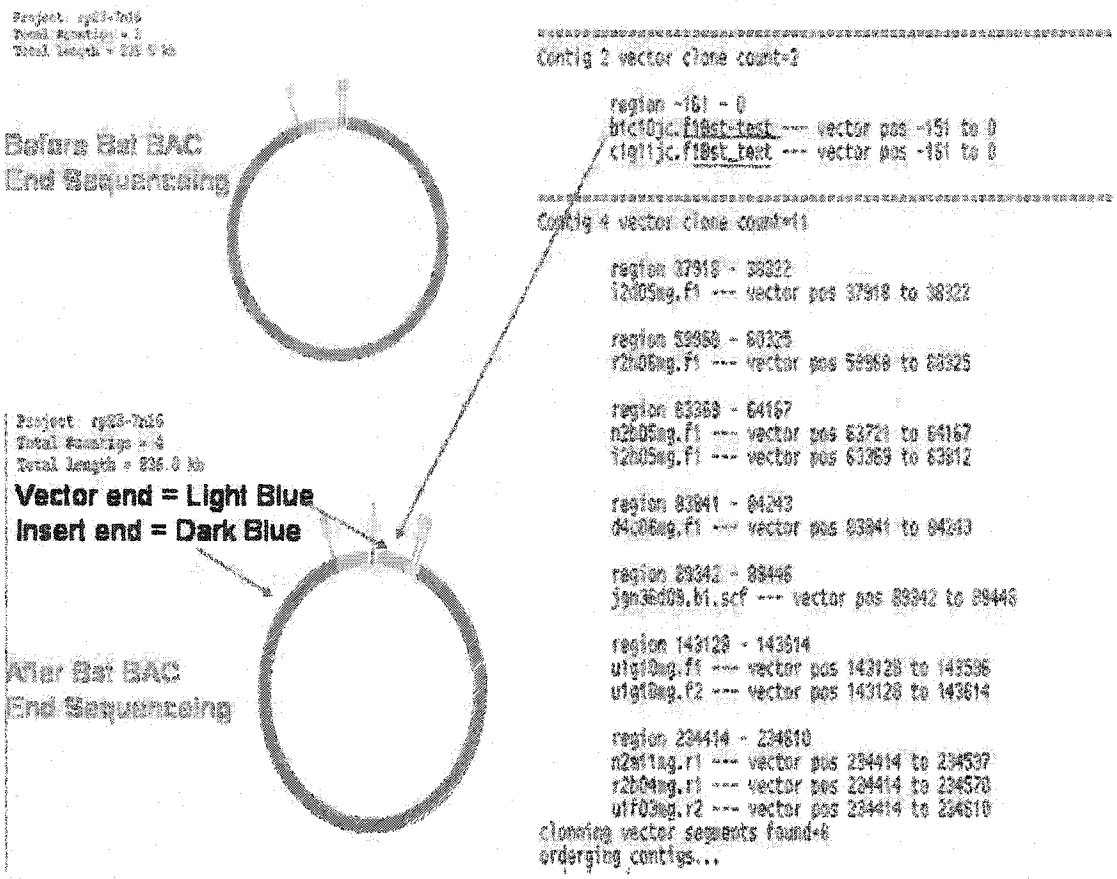


Figure 59 Bst BAC end Sequencing: Mouse pBace3.6 rp23-7n16 Results.

RESULTS: The Exgap output is shown above, the dark blue circle represents the insert and the light blue region is the BAC vector. The top portion, shows only one light blue vector region, while the bottom portion of the Exgap output shows both, the other end was produced, when the Bst amplified sequence data was added to the project. The Exgap at the right, compares to the Bst trace files underlined in red to the BAC end vector, light blue, to prove the origin of the new vector sequence. Thus, sequencing the amplified DNA in this BAC clone revealed both BAC-vector-insert junction sequences.

In conclusion, the Bst protocol and technique worked to amplify the BAC for BAC end sequencing.

Primer-Dimers

During this dissertation, it was found that when performing amplification, a nonspecific DNA amplification often was observed and termed a primer-dimer. Primer-

dimers usually are when the sequence of the primer contains a palindrome or similar sequence through which two primers hybridize and a dimer is formed

There does not appear to be much published at all concerning the primer-dimer phenomenon although it is well known that it exists (Vandesompele 2001). This likely is because primer-dimer formation is not well understood. However, during this dissertation research a significant amount of primer-dimer sequence was obtained, we decided to investigate this phenomenon further, figures 61 and 62, show sequence of two contigs that were amplified with the Bst polymerase that demonstrate this phenomenon.

The 1st primer-dymer contig:

```
>FZN_Bst_Contig1
TAGAGAGGGGCGCACCCCTTGAGAAAAGAGAGAGAGTTCACACCCCATTA
GAAAAGACAGACGGTTCACACCCCCTAGGAAAAGACAGAGGTTCACACCC
CGTACAAAAGACAGAGGTTCACACCCCGTAGAAAAGACAGAGGTTCACAC
CCCCTAAAAAAGACAGAGGTTCACACCCCCTAGAAAAGACAGAGGTTCAC
ACCCCCTAGAAAAGACAGAGGTTCACACCCCTAGAAAAGACAGAGGTTCA
CACCCCTAGAAAAGACAGAGGTTCACACCCCTAGAAAAGACAGAGGTTCA
CACCCCTAGAAAAGACAGAGGTTCACACCCCTAGAAAAGACAGAGGTTCA
CACCCCTAGAAAAGACAGAGGTTCACACCCCTAGAAAAGACAGAGGTTCA
CACCCCTAGAAAAGACAGAGGTTCACACCCCTAGAAAAGACAGAGGTTCA
CACCCCTAGAAAAGACAGAGGTTCACACCCCTAGAAAAGACAGAGGTTCA
CACCCCTAGAAAAGACAGAGGTTCACACCCCTAGAAAAGACAGAGGTTCA
CACCCCCTAGAAAAGACAGAGGTTCACACCCCTAGAAAAGAACAGAGGTT
TCACACCCCCTAGAAAAGAAAGAGGTTCACACCCCTTAAAAAGACGGAGG
GTCCACACCCCTAGAAAAAGAAAGAGGGTCCCCCCCCTAAAAAAAGACAG
AGGGTTCACACCCCCCTCAAAAAAGACAGAGGGGCCCACCCCCTAAAAAA
AGGAGAAGGGTGAAACCCCCCTTCAAACGGCAGGGGGGTCCCCCCCCCT
ACAAAAAGAAAGAGGGGTGTACCCCCCCTCCAAAAAGAAAGAGGGGGCCA
CACCCC
```

Figure 60 The 1st primer-dimer contig: FZN_Bst_Contig1

The 2nd primer-dymer contig:
Bst_ Contig2

```
TTACAGCATGGAAAGTACGGGTTGAAATGTGACTCACTACTAATGAGTTG  50
TATTACAGTAATACGACTCACTATAGTCGAGTCGTATTACAGTAATACGA 100
CTCACTATAATGAGTCGTATTACAGTAATACGACTCACTATAGTGAGTCG 150
TATTACATGTAATACGACTCACTATAGTGAGTCGTATTACAGTAATACGA 200
CTCACTATAGTGAGTCGTATTACAGTAATACGACTCACTATAGTGAGTCG 250
TATTACAGTAATACGACTCACTATAGTGAGTCGTATTACAGTAATACGAC 300
TCACTATAGTGAGTCGTATTACAGTAATACGACTCACTATAGTGAGTCGT 350
ATTACAGTAATACGACTCACTATAGTGAGTCGTATTACAGTAATACGACT 400
CACTATAGTGAGTCGTATTACAGTAATACGACTCACTATAGTGAGTCGTA 450
TTACAGTAATACGACTCACTATAGTGAGTCGTATTACAGTAATACGACTC 500
ACTATAGTGAGTCGTATTACAGTAATACGACTCACTATAGTGAGTCGTAT 550
TACAGTAATACGACTCACTATAGTGAGTCGTATTACAGTAATACGACTCA 600
CTATAGTGAGTCGTATTACAGTAATACGACTCACTATAGTGAGTCGTATT 650
ACAGTAATACGACCCACTATAGTGAGTCGTATTACAGTAATACGACCCAC 700
TATAGGGAGGTCGGTATTACGGTAATACCGACTCACTATAGTGAGGTCGT 750
CTCACCGGTAATCGGCTCCACTATGTGAGGTGGTATTACGGTAATACGGA 800
CCACTTTATTGAGTCGTATTTCGGTAAGACGACCCGCTATTGGGGGGTCG 850
TTGTCCCGGTAGTAGGACCCCCCCATGGGGGGGTGGGCTTCCCGGTGATA 900
```

Figure 61 The 2nd primer-dimer contig: Bst_ Contig 2

It is only after comparing the contig 2 against the amplifying primers and their compliments that the sequence pattern is seen and becomes a confirmed primer-dimer. Using the primer sequences below, the contig shows the position of the primers to form the primer-dimer contig. The red is the forward contig and the blue is the Reverse Complement Forward Primer. The black is the actual contig sequence and the reddish-purple seen at the beginning of the contig and at the end is the normal poor quality beginning and ending sequence. The light green sequence shows when a base in the sequence is different than the primer. This is due to either the poor quality beginning and ending sequences or the additional "A" base added after the primer sequence by the enzyme. This "A" addition is not unusual, the *Taq* polymerase has been known to add a non-template additional nucleotide to the 3' end of PCR products, which is primarily an "A" or adenosine (Clark, 1988; Brownstein 1996). This "A" is normally is preferentially added after a 3' terminal "C" base (Brownstein 1996).

*Forward primer = GTAATACGACTCACTATAG-OH

*Reverse Complement Forward Primer = HO-CTATAGTGAGTCGTATTAC

Bst_Contig1

```
*                          FWDGTAATACGACTCACTATAG
*TTACAGCATGGAAAGTACGGGTTGAAATGTGACTCACTACTAATGAGTTG 50
*                                    RCFWDCTATAGTGAGTCG
*
*         GTAATACGACTCACTATAG                    GTAATACGA
*TATTACAGTAATACGACTCACTATAGTCGAGTCGTATTACAGTAATACGA 100
*TATTAC                  CTATAGT GAGTCGTATTAC
*
*CTCACTATAG                GTAATACGACTCACTATAG
*CTCACTATAATGAGTCGTATTACAGTAATACGACTCACTATAGTGAGTCG 150
*     CTATAGTGAGTCGTATTAC                    CTATAGTGAGTCG


*         GTAATACGACTCACTATAG                    GTAATACGA
*TATTACATGTAATACGACTCACTATAGTGAGTCGTATTACAGTAATACGA  200
*TATTAC                  CTATAGTGAGTCGTATTAC
*

*CTCACTATAG                GTAATACGACTCACTATAG
*CTCACTATAGTGAGTCGTATTACAGTAATACGACTCACTATAGTGAGTCG 250
*     CTATAGTGAGTCGTATTAC                    CTATAGTGAGTCG

*         GTAATACGACTCACTATAG                    GTAATACGAC
*TATTACAGTAATACGACTCACTATAGTGAGTCGTATTACAGTAATACGAC 300
*TATTAC                  CTATAGTGAGTCGTATTAC

*TCACTATAGG                GTAATACGACTCACTATAG
TCACTATAGTGAGTCGTATTACAGTAATACGACTCACTATAGTGAGTCGT 350
*    CTATAGTGAGTCGTATTAC                    CTATAGTGAGTCGT

*        GTAATACGACTCACTATAG                    GTAATACGACT
*ATTACAGTAATACGACTCACTATAGTGAGTCGTATTACAGTAATACGACT 400
*ATTAC                   CTATAGTGAGTCGTATTAC

*CACTATAG                  GTAATACGACTCACTATAG
*CACTATAGTGAGTCGTATTACAGTAATACGACTCACTATAGTGAGTCGTA 450
*    CTATAGTGAGTCGTATTAC                    CTATAGTGAGTCGTA

*       GTAATACGACTCACTATAG                    GTAATACGACTC
*TTACAGTAATACGACTCACTATAGTGAGTCGTATTACAGTAATACGACTC 500
```

```
•TTAC                    CTATAGTGAGTCGTATTAC

•ACTATAG                 GTAATACGACTCACTATAG
•ACTATAGTGAGTCGTATTACAGTAATACGACTCACTATAGTGAGTCGTAT 550
•  CTATAGTGAGTCGTATTAC                 CTATAGTGAGTCGTAT

•     GTAATACGACTCACTATAG                 GTAATACGACTCA
•TACAGTAATACGACTCACTATAGTGAGTCGTATTACAGTAATACGACTCA 600
•TAC                  CTATAGTGAGTCG TATTAC

•  CTATAG                 GTAATACGACTCACTATAG
•  CTATAGTGAGTCGTATTACAGTAATACGACTCACTATAGTGAGTCGTATT 650
•  CTATAGTGAGTCGTATTAC                 CTATAGTGAGTCGTATT

•

•    GTAATACGACTCACTATAG                 GTAATACGACTCAC
•ACAGTAATACGACCCACTATAGTGAGTCGTATTACAGTAATACGACCCAC 700
•AC                CTATAGTGAGTCGTATTAC
•
•TATAG                  GTAATACGACTCACTATAG
•TATAGGGAGGTCGGTATTACGGTAATACCGACTCACTATAGTGAGGTCGT 750
•TATAGTGAGGTCG TATTAC                 CTATAGTGAGGTCGT

•         GTAATACGACTCACTATAG                 GTAATACGACTCACTATAG
•CTCACCGGTAATCGGCTCCACTATGTGAGGTGGTATTACGGTAATACGGA 800
•ATTAC                  CTATAGTGAGGTCGTATTAC

•CCACTTTATTGAGTCGTATTTCGGTAAGACGACCCGCTATTGGGGGGTCG 850

•TTGTCCCGGTAGTAGGACCCCCCCATGGGGGGGTGGGCTTCCCGGTGATA 900
```

•DNA sequence from alignment:

•

```
•         GTAATACGACTCACTATAG                 GTAATACGACT
•ATTACAGTAATACGACTCACTATAGTGAGTCGTATTACAGTAATACGACT 400
•ATTAC                  CTATAGTGAGTCGTATTAC
```

1. The Primer forms the Hair pin loop:

```
GTAATACGACTCACTATAG-OH ->

      T-A-G-OH
      | | |
      A-T-C-A-C-T-C-A-G-C-A-T-AA-T-G
```

2. The enzyme amplifies from the hair pin priming site:

```
T-A-G-T-G-A-G-T-C-G-T-A-T-T-A-C-OH
| | | | | | | | | | | | | | | |
A-T-C-A-C-T-C-A-G-C-A-T-A-A-T-G
```

3. The DNA breathes at 65 °C and therefore opens:

```
T-A-G-T-G-A-G-T-C-G-T-A-T-T-A-C-OH
| | | | | | | | | | | | | | | |          →
A-T-C-A-C-T-C-A-G-C-A-T-A-A-T-G

G-T-A-A-T-A-C-G-A-C-T-C-A-C-T-A-T-A-G-T-G-A-G-T-C-G-T-A-T-T-OH
```

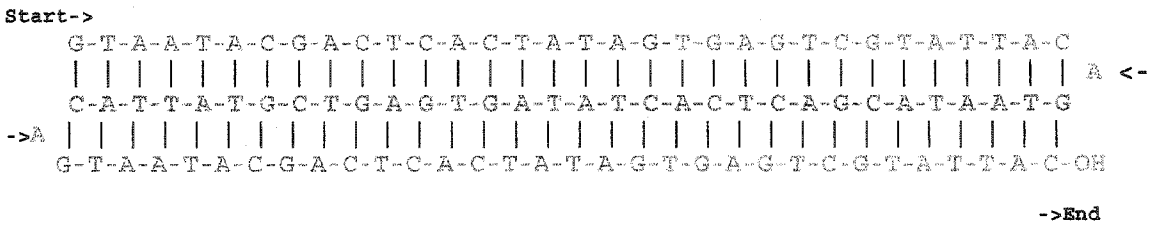
4. The enzyme for what ever reason adds an "A" then flips over and starts down the DNA.

```
  G-T-A-A-T-A-C-G-A-C-T-C-A-C-T-A-T-A-G-T-G-A-G-T-C-G-T-A-T-T-A-C-OH

            →

   G-T-A-A-T-A-C-G-A-C-T-C-A-C-T-A-T-A-G-T-G-A-G-T-C-G-T-A-T-T-A-C
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | A
HO-C-A-T-T-A-T-G-C-T-G-A-G-T-G-A-T-A-T-C-A-C-T-C-A-G-C-A-T-T-A-T-G
```

5. The DNA breathes at 65 °C again and therefore the DNA opens and the enzyme adds an "A" then flips over and starts polymerizing down the DNA again.

```
  G-T-A-A-T-A
                C-G-A-C-T-C-A-C-T-A-T-A-G-T-G-A-G-T-C-G-T-A-T-T-A-C
                | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | A
   C-A-T-T-A-T-G-C-T-G-A-G-T-G-A-T-A-T-C-A-C-T-C-A-G-C-A-T-A-A-T-G
A | | | | | |
  G T A A T-OH
```

6. Therefore the primer-dimer sequence is:



•DNA sequence from alignment used for confirmation:

ATTACAGTAATACGACTCACTATAGTGAGTCGTATTACAGTAATACGACT **400**
CACTATAGTGAGTCGTATTACAGTAATACGACTCACTATAGTGAGTCGTA **450**
TTACAGTAATACGACTCACTATAGTGAGTCGTATTACAGTAATACGACTC **500**

Compare: DNA sequence to above made sequence (black-actual sequence)

A-T-T-A-C-A-G-T-A-A-T-A-C-G-A-C-T-C-A-C-T-A-T-A-G-T-G-A-G-T-C-G-T-A-T-T-A-C-A-G-T-A-A-T-
G-T-A-A-T-A-C-G-A-C-T-C-A-C-T-A-T-A-G-T-G-A-G-T-C-G-T-A-T-T-A-C-A-G-T-A-A-T-

A-C-G-A-C-T **400** C-A-C-T-A-T-A-G-T-G-A-G-T-C-G-T-A-T-T-A-C-A-G-T-A-A-T-A-C-G-A-C-T-C-A-C-
A-C-G-A-C-T C-A-C-T-A-T-A-G-T-G-A-G-T-C-G-T-A-T-T-A-C-A-G-T-A-A-T-A-C-G-A-C-T-C-A-C-

T-A-T-A-G-T-G-A-G-T-C-G-T-A **450** T-T-A-C
T-A-T-A-G-T-G-A-G-T-C-G-T-A T-T-A-C-OH

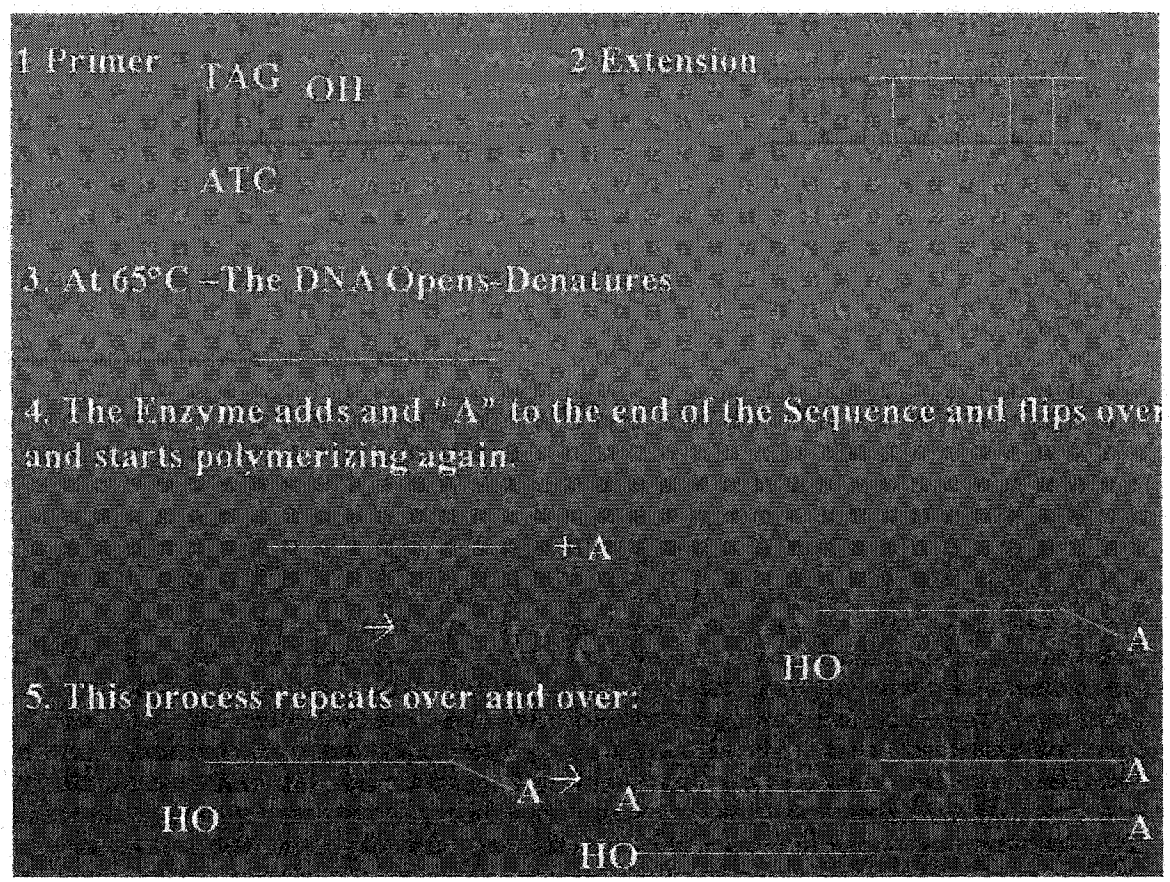


Figure 62 The Making of a single primer nonspecific amplification product (resembling a primer-dimer).

RESULTS: The figure above shows how the amplification, Bst isothermal amplification, of single primer-dimer is done.

•Full sequence comparison of Contig1 and amplification and sequence:

```
•TTACAGCATGGAAAGTACGGGTTGAAATGTGACTCACTACTAATGAGTTG 50

•TATTACAGTAATACGACTCACTATAGTCGAGTCGTATTACAGTAATACGA 100
 TATTACAGTAATACGACTCACTATAGTGAGTCGTATTACAGTAATACGA

•CTCACTATAATGAGTCGTATTACAGTAATACGACTCACTATAGTGAGTCG 150
•CTCACTATAGTGAGTCGTATTACAGTAATACGACTCACTATAGTGAGTCG
•

•TATTACATGTAATACGACTCACTATAGTGAGTCGTATTACAGTAATACGA 200
•TATTACA GTAATACGACTCACTATAGTGAGTCGTATTACAGTAATACGA
•

•CTCACTATAGTGAGTCGTATTACAGTAATACGACTCACTATAGTGAGTCG 250
•CTCACTATAGTGAGTCGTATTACAGTAATACGACTCACTATAGTGAGTCG
•
```

•TATTACAGTAATACGACTCACTATAGTGAGTCGTATTACAGTAATACGAC 300
•TATTACAGTAATACGACTCACTATAGTGAGTCGTATTACAGTAATACGAC
•

•TCACTATAGTGAGTCGTATTACAGTAATACGACTCACTATAGTGAGTCGT 350
•TCACTATAGTGAGTCGTATTACAGTAATACGACTCACTATAGTGAGTCGT
•

•ATTACAGTAATACGACTCACTATAGTGAGTCGTATTACAGTAATACGACT 400
ATTACAGTAATACGACTCACTATAGTGAGTCGTATTACAGTAATACGACT

•CACTATAGTGAGTCGTATTACAGTAATACGACTCACTATAGTGAGTCGTA 450
•CACTATAGTGAGTCGTATTACAGTAATACGACTCACTATAGTGAGTCGTA

TTACAGTAATACGACTCACTATAGTGAGTCGTATTACAGTAATACGACTC 500
TTACAGTAATACGACTCACTATAGTGAGTCGTATTACAGTAATACGACTC

•ACTATAGTGAGTCGTATTACAGTAATACGACTCACTATAGTGAGTCGTAT 550
ACTATAGTGAGTCGTATTACAGTAATACGACTCACTATAGTGAGTCGTAT
•

TACAGTAATACGACTCACTATAGTGAGTCGTATTACAGTAATACGACTCA 600
TACAGTAATACGACTCACTATAGTGAGTCGTATTACAGTAATACGACTCA

•CTATAGTGAGTCGTATTACAGTAATACGACTCACTATAGTGAGTCGTATT 650
CTATAGTGAGTCGTATTACAGTAATACGACTCACTATAGTGAGTCGTATT

•ACAGTAATACGACCCACTATAGTGAGTCGTATTACAGTAATACGACCCAC 700
ACAGTAATACGACTCACTATAGTGAGTCGTATTACAGTAATACGACTCAC
•

•TATAGGGAGGTCGGTATTACGGTAATACCGACTCACTATAGTGAGGTCGT 750
TATAGTGA GTCGTATTACA
•CTCACCGGTAATCGGCTCCACTATGTGAGGTGGTATTACGGTAATACGGA 800

•CCACTTTATTGAGTCGTATTTCGGTAAGACGACCCGCTATTGGGGGGTCG 850

•TTGTCCCGGTAGTAGGACCCCCCCATGGGGGGGTGGGCTTCCCGGTGATA 900

How it sequenced, the forward primer annealed to its compliment sequence:

1. H-O-C-A-T-T-A-T-G-C-T-G-A-G-T-G-A-T-A-T-C-A-C-T-A-G-C-A-T-A-A-T-G

+ Forward primer = GTAATACGACTCACTATAG-OH →

G-T-A-A-T-A-C-G-A-C-T-C-A-C-T-A-T-A-G-OH
H-O-C-A-T-T-A-T-G-C-T-G-A-G-T-G-A-T-A-T-C-A-C-T-A-G-C-A-T-A-A-T-G →

G-T-A-A-T-A-C-G-A-C-T-C-A-C-T-A-T-A-G-A-G-T-G-A-T-C-G-T-A-
H-O-C-A-T-T-A-T-G-C-T-G-A-G-T-G-A-T-A-T-C-A-C-T-A-G-C-A-T-

T-T-A-C-OH
A-A-T-G

As shown in the examples above, Bst can amplify via the adenylation of a primer addition to a hair pin loop. It is notable that the "A" is added after a 3' terminal "C" as Brownstein (1996) had noted in PCR products producing primer-dimers. It is also notable, based on the primer used in this amplification and the method used to form the amplification that a palindrome is not necessary to accomplish the primer-dimer amplification.

Phi-29 (GenomiPhi) Experiments

Since, the Bst and BAC end amplification and sequencing did not yield consistently reproducible results, it was decided that a revisit to the Phi-29 mix for BAC amplification and end sequencing should be done. Fortunately, Amersham had chosen us as a beta test site for their new GenomiPhi DNA Amplification Kit for whole genome amplification only. I then, investigated using their enzyme to reproducibly obtain amplified BAC DNA and obtain accurate sequence.

This final experiment was done using the manufactures' protocol for amplification, which was their implementation of the original, Phi-29 protocols, developed during this dissertation work.

The first question was if this mix allows the whole amplification of a million plus base organism that according to this protocol should take about 16-18 hours, then how long does one need to incubate the solution to amplify a pUC vector plus insert, ~8,000 bases, a fosmid plus insert, ~40,000 bases, and a BAC plus insert , ~150,000 bases. Therefore a test was set up using ~10 ng of each above from three different samples for all three different vectors subclones (9 samples). They all 9 samples, a set, were put on different 96 well theremocycler plates to be incubated 30 °C at the following times: 0, 1hr, 3 hr, 6 hr, 9 hr, 12 hr, 15 hr, 18 hr and then heated to 65 °C to denature the enzyme.

The DNA amplified so well that when trying to pipette it, the DNA would be pulled out of the pipette tip and back into the well. Next, the DNA was diluted, pipetted and vortexed multiple times and then heated (95 °C for 10 minutes) to dissolve and cleave the DNA to shorter fragments that then were desalted by passage through a Sephadex G-50 column to remove the all the unused primers and dNTPs prior to measuring the A_{260} of the amplified DNA.

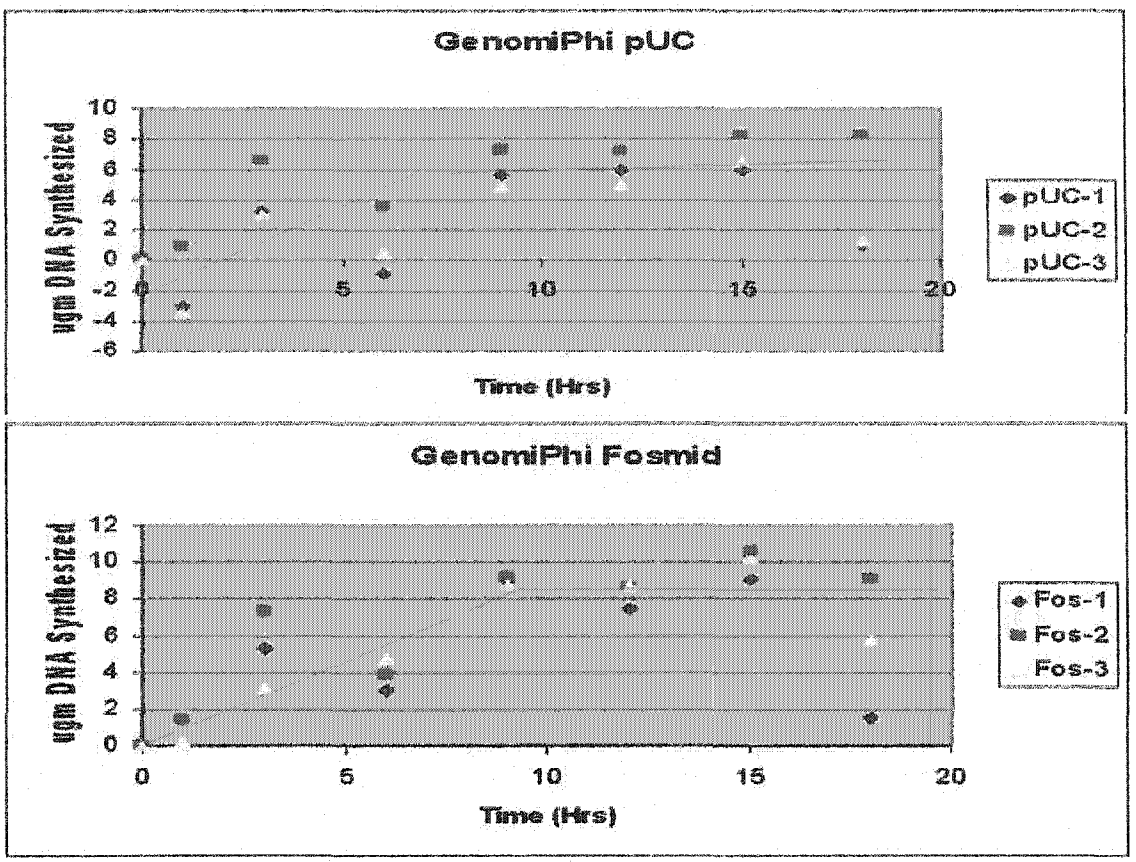


Figure 63 GenomiPhi pUC and Fosmid Amplication Time Study

EXPT: The first question was if this mix allows the whole amplification of a million plus base organism that according to this protocol should take about 16-18 hours, then how long does one need to incubate the solution to amplify a pUC vector plus insert, ~8,000 bases, a fosmid plus insert, ~40,000 bases, and a BAC plus insert , ~150,000 bases. Therefore a test was set up using ~10 ng of each above from three different samples for all three different vectors subclones (9 samples). They all 9 samples, a set, were put on different 96 well theremocycler plates to be incubated 30 °C at the following times: 0, 1hr, 3 hr, 6 hr, 9 hr, 12 hr, 15 hr, 18 hr and then heated to 65 °C to denature the enzyme. The DNA was diluted, pipetted and vortexed multiple times and then heated (95 °C for 10 minutes) to dissolve and cleave the DNA to shorter fragments that then were desalted by passage through a Sephadex G-50 column to remove the all the unused primers and dNTPs prior to measuring the A_{260} of the amplified DNA.
RESULTS: The amplified input DNAs reached a plateau at ~ 9 hours.

As seen in figures 63 and 64, show aliquats of the amplified input DNAs, which reached a plateau at ~ 9 hours. The 9 hours and above time points were subjected to DNA sequencing.

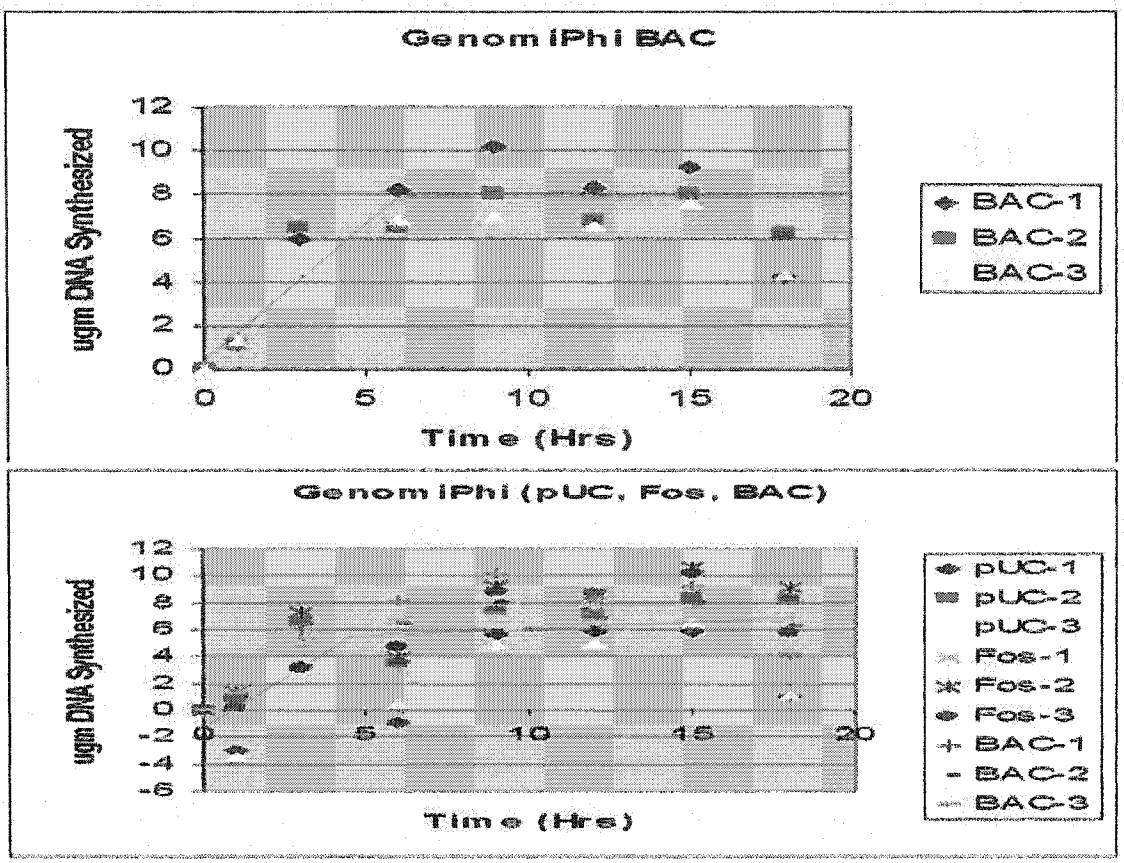


Figure 64 GenomiPhi BAC and all 3 Amplification Results

EXPT: The first question was if this mix allows the whole amplification of a million plus base organism that according to this protocol should take about 16-18 hours, then how long does one need to incubate the solution to amplify a pUC vector plus insert, ~8,000 bases, a fosmid plus insert, ~40,000 bases, and a BAC plus insert , ~150,000 bases. Therefore a test was set up using ~10 ng of each above from three different samples for all three different vectors subclones (9 samples). They all 9 samples, a set, were put on different 96 well theremocycler plates to be incubated 30 °C at the following times: 0, 1hr, 3 hr, 6 hr, 9 hr, 12 hr, 15 hr, 18 hr and then heated to 65 °C to denature the enzyme. The DNA was diluted, pipetted and vortexed multiple times and then heated (95 °C for 10 minutes) to dissolve and cleave the DNA to shorter fragments that then were desalted by passage through a Sephadex G-50 column to remove the all the unused primers and dNTPs prior to measuring the A_{260} of the amplified DNA.

RESULTS: The amplified input DNAs reached a plateau at ~ 9 hours.

DNA	need to seq	ng/µl	µl to seq	Total Vol in µl	1 µl	2 µl	4 xl	F/R pimers
pUC A	80ng	115	1ul	12	115	230	460	2 µl ET
pUC B	80ng	145	1ul	36	145	290	580	
pUC C	80ng	80	1ul	24	80	160	320	
			µl to seq		2 µl	4 µl	µl***	F/R primers
Fos A	400ng	120	4 µl	55	240	480	720	2 µl ET
Fos B	400ng	155	3 µl	55	310	620	930	
Fos C	400ng	160	3 µl	33	320	640	960	
			µl to seq		6 µl	12 µl	18 µl	just F
BAC A	1600ng	130	12--13	37	780	1560	2340	2 µl ET
BAC B	1600ng	185	8--9	50	1110	2220	3330	
BAC C	1600ng	135	11--12	65	810	1620	2430	

Table 13 GenomiPhi Sequencing Test

EXPT: The above is using 0.07-0.09 mM of primer and 2 µl of Et terminators to sequence using the following parameters: No hold, Cycle-95 °C for 30 sec, 50 °C for 20 sec, 60 °C for 4 minutes for 99 cycles; Hold at 4 °C for infinity and desalted using the standard sodium acetate ethanol precipitation.

	1	2	3	4	5	5	5	8	9	10	11	12
	pGem	puc-A-	puc-A-	puc-A-	puc-A-	puc-A-	puc-A-	pGem				
A	400ng	115	230	460	115	230	460	400ng				
		puc-B-	puc-B-	puc-B-	puc-B-	puc-B-	puc-B-					
B	200ng	145	290	580	145	290	580	200ng				
		puc-C-	puc-C-	puc-C-	puc-C-	puc-C-	puc-C-					
C	100ng	80	160	320	80	160	320	100ng				
		fos-A-	fos-A-	fos-A-	fos-A-	fos-A-	fos-A-					
D		240	480	720	240	480	720					
		fos-B-	fos-B-	fos-B-	fos-B-	fos-B-	fos-B-					
E		310	620	930	310	620	930					
		fos-C-	fos-C-	fos-C-	fos-C-	fos-C-	fos-C-					
F		320	640	960	320	640	960					
		bac-A-	bac-A	bac-A-	bac-B-	bac-B	bac-B-	bac-C-	bac-C-			
G		780	1560	2340	1110	2220	3330	810	1620	bac-C-2430		
H												

Table 14 Sequencing Amount of GenomiPhi Test

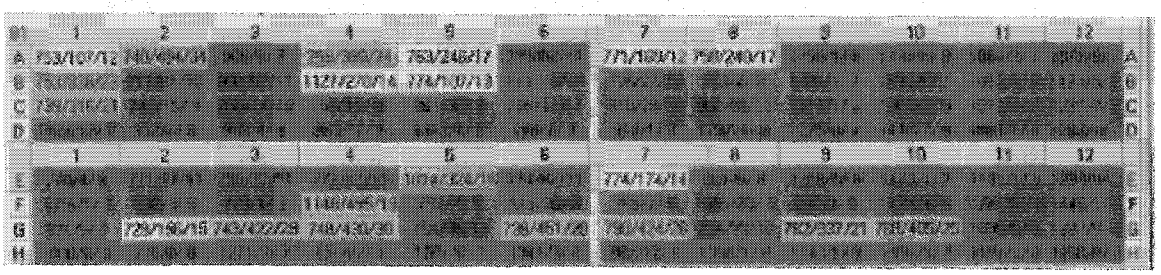


Figure 65 Sequence of GenomiPhi Test

The Top view shows thetype of DNA and the amount used per well to sequence correlating with the Bottom view of the Phred Card.
PHRED CARD: Ph red is a program that reads the trace files produced by the sequencers to assign the bases, their quality values and writes these to output files (Ewing et al., 1998). A Phred Card is a picture showing each well as a certain color based on the number of nucleotides sequenced that had a Phred quality of at least 20, the quality required for phrap to assemble the sequence is shown in fig. 10: 0-100 bases = Red, 101-300 bases = Yellow, C. 301-500 bases = Green, D. 501 and above bases = Blue.
RESULTS: The pUC vector plus insert, a fosmid plus insert, and a BAC plus insert all produced high quality DNA sequence.

As shown in figure 65, the pUC vector plus insert, a fosmid plus insert, and a BAC plus insert all produced high quality DNA sequences and a subsequent blast analysis revealed that the sequence correctly corresponded to the amplified input DNA, example figure 66 is shown below .

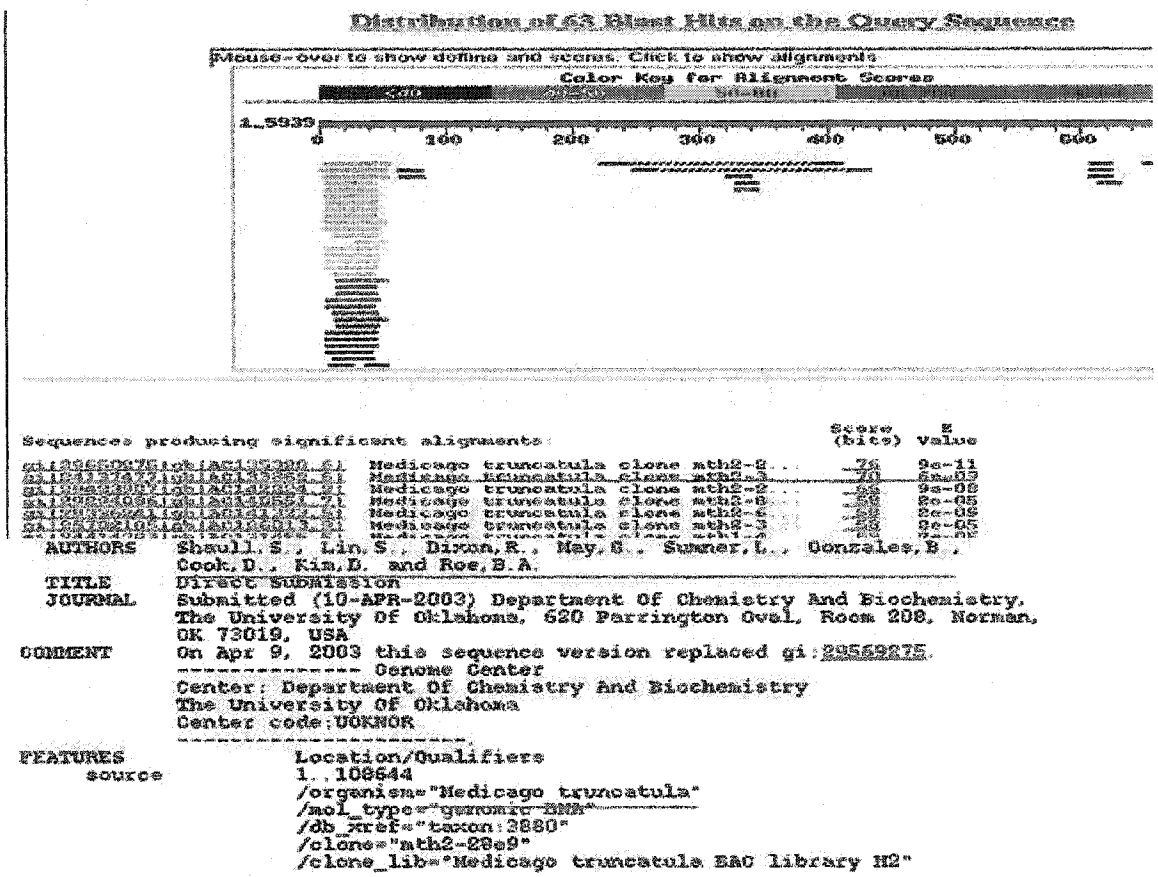


Figure 66 NCBI's Blast of BAC- end Sequence Contig (Based on BAC B, but all matched)

The above shows the results of a sequence Blast showing the actual clone from our lab as a Blast hit.
RESULTS: The pUC vector plus insert, a fosmid plus insert, and a BAC plus insert all produced high quality DNA sequence and a subsequent blast analysis revealed that the sequence correctly corresponded to the amplified input DNA.

Conclusion

During this dissertation research, DNA amplification protocols were developed that allow for picogram size DNA samples to be sequenced. In addition, the human and mouse genomic reference sequences of which ~1 million bases of mammalian genomic DNA were sequenced as part of this present research, will provide the basis for future genetic and forensic research. The published sequences of the mouse genome and the zebrafish mitochondria, both of which are important medical model systems, also will be valuable in both current and future medical research. Two very important disease homologs that emerged from this present research are the mouse homolog for the human park2 gene, a human Parkinson candidate gene, and the mouse homolog DGCR8 gene encoded in the human DiGeorge syndrome critical region. The alignments of all the above sequences were and will be used to compare and contrast other models for comparable genes, for possible functions, for possible anomalies, for evolutionary relationships as well as being possible targets for future drug therapy.

A surprise finding in this research was a sequence verified explanation for the primer-dimer phenomenon. This has been difficult to explain, but now it is clear that the primers are amplified by incorporating a dATP at the end of the primer hairpin loop plus the reverse complemented (primer) sequence. The DNA being at 65°C continuously denatures and reanneals, while the DNA is denatured, the enzyme repositions itself (flips-over) and begins polymerizing once more in the reverse complemented direction until it

reaches the end, when it incorporates another dATP and the process repeats itself cyclically.

Another major development of this research was the development of a closure technique for gaps in chromosomal DNA sequences. Before this research, it was considered impossible to accomplish this with the available technology, because the clones covering the gap regions were unstable. Now, by employing isothermal rolling circle amplification, obtaining the templates needed for sequencing access to the gaps may become a reality.

The Phi-29 and Bst amplification and sequencing protocols for plasmids, fosmids, BACs and chromosomal DNA, including gap closure, will enhance the field of forensics by allowing the accurate amplification of DNA from minute amounts of starting material.

During this research protocols were developed using Phi-29 (*Bacillus subtilis bacteriophage* Phi-29 DNA polymerase) and Bst (*Bacillus stearothermophilus* polymerase I) for DNA Isothermal Rolling Circle Amplification that resulted in the successful sequencing of BAC-based clones. Many of these protocols now have been adopted by Amersham, Inc. and incorporated into their isothermal amplification notes.

Although Bst originally was isolated by my principal professor during his doctorial research, I rediscovered it during patent literature research. The highly processible Phi-29 can amplify 70 kb without becoming dissociated from its substrate, but Bst is much more processive as it can amplify more than three times that amount without dissociating. In addition, Phi-29's optimal polymerizing temperature of 30 °C is low enough to allow nonspecific primers to anneal, while Bst's optimal polymerizing

temperature is 65 °C, which allows for longer specific primes to be used for directed amplification of specific target DNAs.

The Phi-29 and Bst amplification protocols developed during this research are quite robust and can significantly reduce the time and effort for obtaining the DNA needed as a template for DNA sequencing, while increasing the sensitivity to minute samples that the standard methods can not accomplish. Therefore, through this research the Isothermal Cascading Whole Genome Multiple Strand Displacement Amplification now has become a reality that in the future may supplant the PCR method for DNA amplification.

References

Altschul, S. F., T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman. (1997 *Nucleic Acids Res.* **25**, 3389-3402.

Amersham Biosciences. (2002) unpublished, product information, *DYEnamic ET Terminator Cycle Sequencing Kit*, Amersham Pharmacia Bioteck UK Limited, UK

Amersham Pharmacia biotech. (1998) unpublished, product information, *Thermo Sequenase fluorescent labeled primer cycle Sequencing kit with 7-deaza-dGTP*

Amersham Pharmacia Biotech (2001). unpublished, product information, Draft:TempliPhi™ DNA sequencing template amplification kit 25-6400-01·

Anderson, S. *et al.* (1981) *Nature* **290,** 457-465.

Anderson S. (1981) *Nucleic Acids Res.* **9,** 3015-3027

Anderson, S. *et al.* (1982) *Genomes.* **5,** 43

Barner, J., M., Nilsson, M., Mendel-Hartvig, and U., Landeregren. (1998) *Nucleic Acids Res.* **26,** 5073-5078.

Barnes, W.M. (1992) *Gene* **112,** 29-35.

Becker, N., Seitanidou,T., Murphy, P., Mattei, M., Topilko, P., Nieto,A., Wilkinson,D.G., Charnay,P., Gilardi,P. (1994) *Mech. Dev.* **47,**3-17.

Beckman Coulter. (2003) http://www.beckman.com/products/instrument/analytical/ce/pacemdq .

Bedford, E., TS., Tabor, and C. Richardson. (1999). Patent #: 5,972,603.

Blanco, L. and Salas, M. (1984) *Proc Natl Acad Sci.* **81,** 5325-5329.

Blanco, L., A. Bernad and M. Salsa. (1991)) U.S. Patent #: 5,001,050.

Blanco, L., A. Bernad and M. Salsa. (1993) U.S.Patent #: 5,198,543.

Blanco, L., A. Bernad and M. Salsa. (1996)) U.S.Patent #: 5576204.

Blanco, L. and Salas, M. (1996) J. *Biol. Chem.* **271,** 8509-8512.

Blanco, L., Bernard, A., Lazaro, J.M., Martin, G., Garmendia, C., and Salas, M. (1989) J. *Biol. Chem.* **264,** 8935-8940.

Boore, J. L. (1999). *Nucleic Acids Res.* **27**, 1767-1780.

Bowater, R. (2003). <http://www.ehgonline.net>.

Bressan,A. Somma,M.P. Lewis,J. Santolamazza,C. Copeland,N.G. Gilbert,D.J. Jenkins,N.A. Lavia,P. (1991) *Gene* **103,** 201-209.

Broughton, R., Milam, J. and Roe, B. (2001) *Genome Research* **1,** 1958-1967.

Brown, T.A. (1989) *Genetics: a molecular approach,* 2nd Ed., Chapman and Hall, London, UK

Brownstein, D., and Smith, J.(1996) *Biotechniques* **20,** 1004-1010.

Burge, C. and Karlin, S. (199) *J.Mol.Biol.* **268,** 78-94.

Cairns, J. (1963) *J. Mol. Biol.* **6,** 208-13.

Cariello, N.F., J.A., Swenberg, and T.R., Skopek, (199) *Nucleic Acids Res* **19,** 4193-4198.

Cheng, J. (2003)

< http://www.lbl.gov/Education/CSEE/cup/Su01/Sum01Web/Lam/Intro/intro.html>

Ciossek,T., Lerch,M.M., Ullrich,A. (1995) *Oncogene* **11,** 2085-2095

Clark, JM. (1988) *Nucleic Acids Res.* **20,** 9677-9686

Clayton, D.A. (1982) *Cell* **28,** 693-705.

Clayton, D.A. (1991) *Trends Biochem. Sci.* **16,** 107-111.

Collins, J., *et al.* (2003) *Genome Research* **13,** 27-36.

Coutavas,E., Ren,M., Oppenheim,J.D., D'Eustachio,P., Rush,M.G. (1993) *Nature* **366,** 585-587.

Cox, E.C. (1976) *Ann. Rev. Genet.* **10,** 135-156.

Crick, F. (970) *Nature* **227,** 561-563.

Dattagupt, N. (1995) European patent #: WO 95/25180. .

Dean, F.B., J. R. Nelson, T.L.Giesler, and R.S. Lasken. (200) *Genome Research.* **11,** 1095-1099.

DeFrancesco, L. (2001) *The Scientist* **15,** 23- 24

Di Matteo,G. Fuschi,P. Zerfass,K. Moretti,S. Ricordy,R. Cenciarelli,C. Tripodi,M. Jansen-Durr,P. Lavia,P. (1995) *Cell Growth Differ.* ***6,*** 1213-1224.

Drake, J. W. (1969) *Nature* **221,** 1132-1133.

Dunham *et al.* (1999) *Nature* **402,** 489-495

Eckert, K.A., and T.A., Kunkel, (1990) *Nucleic Acids Res.* **18,** 3739-3744.

Eckert, K.A., and T.A., Kunkel, (1991), *PCR Methods Appl* **1,** 17-24

Esteban, J.A., Salas, M. and Blanco, L (1993) *J. Biol. Chem.* **268,** 2719-2726

Ewing, B., L. Hillier, C. M. Wendl, and P. Green.(1998) *Genome Research* **8,** 175-185.

Ewing, B. and P. Green (1998) *Genome Research,* 186-194.

Fichant, G. A. & Burks, C. (1991) *J.Mol.Biol.* **220**, 659-671

Fire, A. and S. Q. Xu. (1995) *Proc. Natl. Acad. Sci.* **92,** 4641-4645.

First, E. and P. Hengen. (2003) <http://micro.nwfsc.noaa.gov/protocols/taq-errors.html>

Fournier, A.E., Grandpre, T., Strittmatter, S.M. (2001) *Nature* **409,** 341-346.

Franklin, R.E. and Gosling, R. G. (1953) *Nature* **171,** 740-741.

Frazer, K., L. Elnitski, D. M. Church, I. Dubchak, and R.C. Hardison. (2003) Genome Research **13,** 1-12

Georgia State University, Chemistry Department. (2003) <http://chemistry.gsu.edu/glactone/PDB/DNA_RNA/.>

Ghosh, A. and Manju, B. (2003) *Acta Cryst.* **D59**, 620-626.

Gong,L.,and Yeh,E.T.H. (1999) EMBL GenBank DDBJ database.

Gordon, D., C. Abajian, and P. Green (1998) *Genome Res.* **8,** 195-202.

Green, P. (1998) unpublished, product information, Cross_match, and Consed software beta version. Copyright (c) Phil Green and the University of Washington.

Green, P. and Ewing, B. (1998) unpublished, product information, Phred software beta version. Copyright (c) Phil Green, Brent Ewing and the University of Washington

Guarguaglini,G. Battistoni,A. Pittoggi,C. Di Matteo,G. Di, Fiore,B. Lavia,P. (1997) *Biochem. J.* **325,** 277-286.

Hong, G., Y. Yang, and J. Zhu. (2002) European patent #: WO 02/101004 A2.

Hubbard, T., D. Barker, E. Birney, G. Cameron, Y. Chen, L. Clark, T. Cox, J. Cuff, V. Curwen, T. Down, R. Durbin, E. Eyras, J. Gilbert, M. Hammond, L. Huminiecki, A. Kasprzyk, H. Lehvaslaiho, P. Lijnzaad, C. Melsopp, E. Mongin, R. Pettett, M. Pocock, S. Potter, A. Rust, E. Schmidt, S. Searle, G. Slater, J. Smith, W. Spooner, A. Stabenau, J. Stalker, E. Stupka, A. Ureta-Vidal, I. Vastrik, and M. Clamp (2002) *NucleicAcids Res.* **30,** 38-41.

Hugo T. Matern, Charles Yeaman, W. James Nelson, and Richard H. Scheller (2001) *PNAS* **98,** 9648-9653

Hugh, G. and Griffin, M. (1994) PCR Technology, **1**, 228-229.

Human Genome Management Information System (HGMIS). (2001) <http://www.ornl.gov/TechResources/Human_Genome/about.html>.

Human Genome Management Information System (HGMIS). (2003 <http://www.ornl.gov/TechResources/Human_Genome/project/about.html>.

Human Genome Sequencing Consortium. (2001) *Nature* **409,** 860-941.

Imondi,R., Wideman,C., Kaprielian,Z. (2000) *Development* **127,** 1397-1410

International Human Genome sequencing Consortium. (2001) *Nature* **409,** 860-921

Ioannou, P. A., Amemiya, C. T., Garnes, J., Kroisel, P. M., Shizuya, H., Chen, C., Batzer, M. A. & de Jong, P. J. (1994) *Nat. Genet.* **6**, 84–89.

Keohavong, P., and W.G., Thilly. (1989) *Proc Natl Acad Sci* **86,** 9253-9257

Kiefer, J.R., C. Moa and L. Beese. (1997) *Nature* **391,** 304.

Kiefer, J.R., C. Moa, C. Hansen, S. Basehore, H. Hogrefe, J. Braman and L. **Beese**. (1997). *Structure* **5,** 95-108.

Kong, H., Aliotta, J. and Pelletier, J. J., New England Biolabs, unpublished, product information.

Kong, H., R.B., Kucera, and W.E., Jack. (1993) *J Biol Chem* **268,** 1965-1975.

Kong, H., J. Pelletier and J. Aliotta. (1998) U.S. Patent #: 5,814,506

Kornberg, A. (1960) *Science* **131**, 1503-1508

Kornberg, A and T. A. Baker. (1992). *DNA replication*. W. H. Freeman and Company, San Francisco.

LAE, M. C. (2003) <http://www.ich.bpmf.ac.uk/cmgs/mito.htm>.

Li, P. *et al.* (1997) *Genomics* **40**, 476-485.

LI-COR®. (1998) unpublished, product information,. *LI-COR 4200 Series Sequencing Manual.*, LI-COR, Inc.,

Ling, L.L., P., Keohavong, C., Dias, and W.G., Thilly (1991) *PCR Methods Appl* **1,** 63-69

Liu, D., Daubendiek, S.L.,Zillman, M.A., Ryan, K., and E.T. Kool. (1996) *J. AM. Chem. Soc.* **118,** 1587-1594

Lizardi, P. M. (2001. U.S. Patent #: 6,280,949.

Lizardi, P. M. (2001U.S. Patent #: 6,329,150.

Lizardi, P. M. , X. Huang, Z. Zhu, P. Bray-Ward, D.C. Thomas, and D.C.Ward. (1998) *Nat. Genet.* **19,** 225-232.

Loh, M. T. *et al.* (2000) . http://infinity.gen.emory.edu/MITOMAP/seqtable.html .

Lundberg, K.S., D.D., Shoemaker, M.W., Adams, , J.M., Short, J.A., Sorge, and E.J. Mathur. (1991). *Gene* **108,** 1-6.

Lwoff, A. (1953) Bacteriol. Rev. **17,** 269-337.

Marguet E, and Forterre P. (1998) *Extremophiles* **2**, 115-22

Matilla, P., J., Korpela, T., Tenkanen, and K. Pitkanen. (1991) *Nucleic Acids Res* **19,** 4967-4973.

Mario H. Skiadopoulos, Leatrice Vogel, Jeffrey M. Riggs, Sonja R. Surman, Peter L. Collins, and Brian R. Murphy (2003) *Journal of Virology* **77,** 270-279.

Mathews, C.K. and K.E. Van Holde. (1996) *Biochemistry* 2nd Ed., Benjamin/Cummings Publishing Company, New York

Maxam, A. M. & Gilbert, W. (1977) P*roc.Natl.Acad.Sci.* **74**, 560-564.

Mead, D. A., McClary, J. A., Luckey, J. A., Witney, F. R., & Smith, LM. (1991) *BioTechniques* **11**, 76-87.

Melcher, U.. <http://opbs.okstate.edu/~Melcher/MG/MGW1/MG1375.html . 2000>.

McClary, J. *et al.* (1991) *J. DNA Sequencing and Mapping* **3,** 173-180.

McLaughlin, I., S. Coticone, and W. Bloch. (2003). U.S. patent #: 20030104395.

MJ Research. (2003) <http://www.basestation.com/overview/overview8.html>

Molecular® Dynamics. (1999) unpublished, produce information, Megabace, Instrument User's Guide: DRAFT. Molecular Dynamics, Inc.

Mouse Genome Sequencing Consortium.(2002) *Nature* **420,** 520-562.

Mulligan, Martin. (2003)

<http://www.mun.ca/biochem/courses/4103/topics/Lambda/Lambda.html>

Mullis, K.B. and F.A. Faloona. (1987) *Meth. Enzymol.* **155,** 335-350.

Mayer, Gene. (2003) "*MBIM Microbiology and Immunology on line.*",

<http://www.med.sc.edu:85/mayer/phage.htm.>

Mayor, C., *et al.* (2000). *Bioinformatics Applications* **16,** 1046-1047.

New England Biolabs. (2003).

<http://www.neb.com/neb/products/polymerases/275.html>.

Oefner, P., S. Hunicke-Smith, L. Chiang,F. Dietrich, J. Mulligan, and R. Davis. (1996)

Nucleaic Acids Res. **24,** 3879-3886.

Ojala, D., Montoya, J., and Attardi, G. (1981) *Nature* **290,** 470-474.

Olby, R. (2003) *Nature* **421,** 402-404.

Parsons, J. D. (1995). *Comput. Applic. Biosci.* **11,** 615-619.

Pearson, W. R. & D.J.Lipman. (1988) *PNAS* **85**, 2444-2448

Pearson, W. R. (1990) *Methods in Enzymol.* **183**, 98.

Perkin-Elmer Biosystems (1999) unpublished, produce information, *ABI Prism™ 3700 DNA ANAlyzer, DRAFT: Chemistry Guide*. CA: Perkin-Elmer Corp.

Perkin-Elmer Biosystems. (1999)) unpublished, produce information, *ABI Prism™ 377, DNA Sequencer: User's Manual.* CA: Perkin-Elmer Corp.

Perkin-Elmer Applied Biosystems. (1999)) unpublished, produce information, *Automated DNA Sequencing: Chemistry Guide*. CA: Perkin-Elmer Corp.

Perkin-Elmer Applied Biosystems. (1998)) unpublished, produce information, *ABI Prism™ BigDye™ Terminator Cycle Sequencing Ready Reaction Kit, with AmpiTaq® DNA Polymerase, FS, Protocol.*, Perkin-Elmer Corp, Ca.

Prockop, D., J. Early, G. Tromp, S. Kuivaniemi. (1994) European patent #: WO 94/16107.

Roe, B. (1997) *Protocols for Recombinant DNA Isolation, Cloning, and Sequencing.* Unpublished, ACGT laboratory manual.

Roe, B. (1996) unpublished, NCHGR Special Site Visit to Bruce Roe Lab.

Roe, B.A. (2002). unpublished, personal conversation.

Roe, B. (2003) . < http://www.genome.ou.edu/>.

Sambrook, J., Fritsch, B. Y., & ManiatisT. (1989) *Molecular Cloning: A Laboratory Manual,* 2nd Ed., Cold Spring HarborLaboratory, Cold Spring Harbor, NY

Sanger, P., Nicklen, S., & Coulson A. (1977) *Proc.Natl Acad.sci* **74**, 5463-5476

Schanke, J. (2000) European Patent #: WO 00/71739 A1.

Scheffler, I. E. (1999) *Mitochondria,* John Wiley & Sons, Inc., N.Y.

Schumacher, M., Hurlburt, B., and Brennan, R. (2001) *Nature* **409**: 215-219

Schwartz *et al.* (2000) *Genome Research* **10,** 577-586

Schwartz *et al.* (2003 *Nucleic Acids Res.* **31,** 3518-3524.

Shadel, G.S. and Clayton, D.A. (1997) *Annu. Rev. Biochem.* **66,** 409-435

Shizuya, H., Birren, B., Kim, U. J., Mancino, V., Slepak, T., Tachiiri, Y. & Simon, M. (1992) *Proc. Natl. Acad. Sci* **89,** 8794–8797

Sonnhammer, E. L. L. and R. Durbin. (199) *Gene* **167,** 1-10.

Staden, R. (1986) *Nucleic Acids Res.* **14**, 217-231.

Stansfield, William D. (1991) *Schaum's outline of theory and problems of genetics.* Third edition. McGraw-Hill, New York

Stenesh, J. and B. Roe. (1972) *Biochimica Et Biophysica Acta.* **272,** 156-166

Swaminathan, N. and R. Wilkosz. (1998) . U.S. patent #: 5,830,714.

Temin, H. (1972. *Scientific American* **226** (1), 24-31.

Staff. (2001). *The* Scientist **15**[17]

Thompson, J. D., Higgins, D. G., & Gibson, T. J. (1994) *Nucleic Acids Res.* **22**, 4673-4680.

Thorstenson, Y., S. Hunicke-Smith, P. Oefner, and R. Davis. (1998) *Genome Methods* 848-855.

Tindall, K.R., and T.A. Kunkel. (1988) *Biochemistry* **27**, 6008-6013.

Tomac, A.C. Hoffer,B.J. (2001) *Cytogenet. Cell Genet.* **95**, 120-121.

Tsai,Y.C. Fishman,P.S. Thakor,N.V. Oyler,G.A. (2003) *J. Biol. Chem.* **278**, 22044-22055.

Vandesompele, J., De Paepa, A., and Speleman, F. (2001) *Analytical Biochemistry* September 5, 2001.

Vargason, J., Eichman, B., and Shing Ho, P. (2000) *Nat. Struct. Biol.* **7**, 758-761.

Varmus, H. (1987. *Scientific American*, **257** (3), 48-54.

Viguera, E., D. Canceill and S.D. Ehrlich.(2001) *J. Mol. Biol.* **312**, 323-333.

Walker, G. (1998U.S. patent #: 5,712,124.

Wang, A., Quigley, G., Kolpak, F., van der Marel, G., van Boom, J., and Rich A. (1981) Left-handed double helical DNA: variations in the backbone conformation. *Science* 211 (4478), 171-176.

Watson, J. D. and F. H. Crick. (1953) *Nature* **171,** 737-738.

Weaver, R. F. (1999) *Molecular Biology,*McGraw- Hill, Boston.

Wilkins, M.H.F., Stokes, A. R. & Wilson, H. R. (1953) *Nature* **171,** 738-740.

Wong, T.W. and Clayton, D.A. (1985) *Cell* **42,** 951-958

Yang, S. and D. Chatterjee. (1999). #:WO 99/10366.

Zuker M, and P. Stiegler (198). *Nucleic Acids Res.* **9**(1), 133-148.

Zhang, David. (2002) . U.S. Patent #: 20020182598..

Appendix I

For each Human and Mouse clone, a table is shown that reveals the following: the organism (Organ.); the vector type, (BAC or PAC); the NIH Accession number; chromosome or chromosome scaffold number (Chr/ Chr Scaffold #), contigs; The strand direction of the gene (S= strand, - or + determines the side or direction of the gene, example: S-1); gene number; gene name (Novel = new); gene ensemble number; genes chromosomal position; transcript (A predicted gene with research to support this belief for example a mRNA that codes for this transcript.); Exon (DNA that cods for a gene) number; Exon chromosomal position; different exon locations based on different splicing and lastly all the organisms with comparative genes and their gene locations.

Gene Organization of the PAC 20k14

Table 15 Gene Organization of the PAC 20k14

Organ	PAC		Accession #			
Human	**20k14**		**AC006548**	Chr/	or Chr Scaffold	Hu
Organism	Genes	Name	Ensembl ID	#	Chromosome Location	#
Human	**Genes-1**	Novel	ENSG00000172963	22q11	15777569-15777617	
	Transcrpt-1	splice-2	ENST00000310115			S-1
	Exon-1				15777569-15777617	2
	Exon-2				15769723-15769994	1
	Transcrpt-2	splice-2	ENST00000327926			S-1
	Exon-1				15770375-15770423	2
	Exon-2				15769723-15769994	1

Human	**Genes-2**	Novel	ENSG00000138860	22q11	15818177 - 15863447	
	Transcrpt-1	splice-2	ENST00000266136			S-1
	Exon-1				15863385-15863443	9
	Exon-2				15847317-15847620	8
	Exon-3				15843404-15843611	7
	Exon-4				15824988-15825637	6
	Exon-5				15821544-15821804	5
	Exon-6				15820622-15820712	4
	Exon-7				15820210-15820306	3
	Exon-8				15819169-15819276	2
	Exon-9				15818177-15818356	1
Human	**Transcrpt-2**	splice-2	ENST00000330197			S-1
	Exon-1				15863385-15863447	10
	Exon-2				15847317-15847620	9
	Exon-3				15843404-15843611	8
	Exon-4				15825057-15825637	7
	Exon-5				15823742-15823827	6
	Exon-6				15821544-15821808	5
	Exon-7				15820622-15820712	4
	Exon-8				15820210-15820306	3
	Exon-9				15819169-15819276	2
	Exon-10				15818177-15818356	1

Gene Organization of the BAC 678g6

Table 16 Gene Organization of the BAC 678g6

Organism	**BAC**		Accession #			Hu
Human	**678g6**		**AC007845**	Chr/	or Chr Scaffold	
Organism	Genes	Name	Ensembl ID	#	Bac Location	#
Human	**Genes-1**	Novel		22q11	7315-18212	S-1
	Exon-1				7315-7248	8
	Exon-2				9503-9449	7
	Exon-3				9950-9738	6
	Exon-4				11206-11067	5
	Exon-5				15730-15520	4
	Exon-6				16246-16149	3
	Exon-7				16552-16300	2
	Exon-8				18229-18212	1
Human	**Genes-2**	Novel			58715-26101	S-1
	Exon-1				26101-25653	7
	Exon-2				46265-46064	6

	Exon-3		46875-46816	5
	Exon-4		47427-47329	4
	Exon-5		51651-51530	3
	Exon-6		55843-55820	2
	Exon-7		58715-58553	1
Human	**Genes-3**	Novel	62606-69474	S-1
	Exon-1		62606-62422	7
	Exon-2		62761-62688	6
	Exon-3		63038-62892	5
	Exon-4		64837-64717	4
	Exon-5		68184-68040	3
	Exon-6		68945-68804	2
	Exon-7		69601-69474	1
Human	**Genes-4**	Novel	73937-105347	S-1
	Exon-1		73937-73789	19
	Exon-2		75225-75175	18
	Exon-3		79768-79676	17
	Exon-4		80300-80207	16
	Exon-5		82166-82056	15
	Exon-6		82961-82803	14
	Exon-7		88137-88043	13
	Exon-8		90482-88664	12
	Exon-9		92922-92840	11
	Exon-10		93583-93443	10
	Exon-11		94473-94360	9
	Exon-12		96161-96106	8
	Exon-13		97205-96995	7
	Exon-14		99196-99059	6
	Exon-15		99814-99673	5
	Exon-16		101445-101137	4
	Exon-17		101569-101520	3
	Exon-18		103519-103266	2
	Exon-19		105411-105347	1
Human	**Genes-5**	Novel	112655-117652	S-1
	Exon-1		112655-108666	4
	Exon-2		113436-113241	3
	Exon-3		115198-114979	2
	Exon-4		117693-117652	1
Human	**Genes-6**	Novel	123068-132327	S+1
	Exon-1		123068-123078	1
	Exon-2		125829-125960	2
	Exon-3		128942-129043	3
	Exon-4		132222-132327	4

Gene Organization and Comparative Genetics of BAC rp23-7n16

Table 17 Gene Organization and Comparative Genetics of BAC rp23-7n16

Organism	BAC		Accession #			
Mouse	**rp23-7n16**		**AC091002**	Chr/ #	or Chr Scaffold Chromosome Location	mus #
Organism	Genes	Name	Ensembl ID			
Mouse	**Genes-1**	Rtn4r	ENSMUSG00000043811	16	17623863-17625263	S+1
	Exon-1				17623863-17625263	1
Human	**Genes-1**	Rtn4r	ENSG00000040608	22	18603788-18605188	S-1
	Exon-1				18603788-18605188	1
Rat	**Genes-1**	RT4R_RAT	ENSRNOG00000001882	11	2784830-2786231	S+1
	Exon-1			Scfold	2784830-2786231	1
Fugu	**Genes-1**	Q9BZR6	SINFRUG00000128826	67	144168-145481	S-1
	Exon-1				144168-145481	1
Mouse	**Genes-2**	Novel	ENSMUSG00000041484	16	17696338-17708142	
	Transcrpt-1	splice-2	ENSMUST00000045507			S-1
	Exon-1				17708039-17708142	12
	Exon-2				17701767-17701888	11
	Exon-3				17701446-17701603	10
	Exon-4				17701191-17701363	9
	Exon-5				17700908-17701010	8
	Exon-6				17700487-17700578	7
	Exon-7				17700185-17700320	6
	Exon-8				17699778-17699898	5
	Exon-9				1769959-17699685	4
	Exon-10				17697701-17698347	3
	Exon-11				17697350-17697577	2
	Exon-12				17696338-17696509	1
Mouse	**Transcrpt-2**	splice-2	ENSMUST00000058904	16		S-1
	Exon-1				17708039-17708142	11
	Exon-2				17701767-17701888	10
	Exon-3				17701446-17701603	9
	Exon-4				17701191-17701363	8
	Exon-5				17700908-17701010	7
	Exon-6				17700487-17700578	6
	Exon-7				17700185-17700320	5
	Exon-8				17699778-17699898	4
	Exon-9				17699591-17699700	3
	Exon-10				17697350-17698347	2
	Exon-11				17696338-17696509	1

Human	**Genes-2**	ZDHHC8	ENSG00000099904	22	18494025-18508527	S+1
	Exon-1				18494025-18494128	1
	Exon-2				18501271-18501392	2
	Exon-3				18501555-18501712	3
	Exon-4				18502694-18502785	4
	Exon-5				18502948-18503089	5
	Exon-6				18503294-18503414	6
	Exon-7				18503498-18503607	7
	Exon-8				18504833-18505833	8
	Exon-9				18507306-18508527	9
Rat	**Genes-2**	Novel	ENSRNOG00000001883	11	2865358-2865471	S-1
	Exon-1			Scfold	2865358-2865471	1
Fugu	**Genes-2**	Novel	SINFRUG00000149349	195	225635-253405	S+1
	Exon-1				225635-225738	1
	Exon-2				248141-248262	2
	Exon-3				249850-250007	3
	Exon-4				250087-250259	4
	Exon-5				250390-250492	5
	Exon-6				250793-250884	6
	Exon-7				252160-252322	7
	Exon-8			chr-frag	252395-253405	8
Zebraf.	**Genes-2**	Novel	ENSDARG00000018508	ctg	71515-84059	S-1
	Exon-1			1062983	925-84059	6
	Exon-2				80713-80824	5
	Exon-3				77696-77827	4
	Exon-4				74745-75370	3
	Exon-5				74390-74742	2
	Exon-6				71515-71841	1
Mouse	**Genes-3**	Ranbp1	ENSMUSG00000041484	16	17696338-17708142	
	Transcrpt-1	splice-2	ENSMUST00000045507			S-1
	Exon-1				17708039-17708142	7
	Exon-2				17720421-17720557	6
	Exon-3				17718333-17718490	5
	Exon-4				17714877-17715020	4
	Exon-5				17714026-17714091	3
	Exon-6				17712957-17713325	2
	Exon-7				17712924-17712943	1
	Transcrpt-2	splice-2	ENSMUST00000052325	22		S-1
	Exon-1				17721680-17721792	6
	Exon-2				17720421-17720557	5
	Exon-3				17718333-17718490	4
	Exon-4				17714877-17715005	3
	Exon-5				17714026-17714091	2
	Exon-6				17712917-17713325	1
Human	**Genes-3**	RANBP1	ENSG00000099901	22	18478015-18489258	S+1
	Exon-1				18479573-18479741	1
	Exon-2				18481090-18481226	2
	Exon-3				18484341-18484498	3

	Exon-4				18487385-18487513	4
	Exon-5				18488380-18488445	5
	Exon-6				18489029-18489258	6
Rat	**Genes-3**	Novel	ENSRNOG00000001884	11	2881303-2888224	S-1
	Exon-1				2888079-2888224	5
	Exon-2				2886021-2886178	4
	Exon-3				2883059-2883202	3
	Exon-4				2882221-2882286	2
	Exon-5				2881303-2881709	1
Fugu	**Genes-3**	RANG	SINFRUG00000149352	Scfold	257750-258692	S-1
	Exon-1	MOUSE		195	258574-258692	4
	Exon-2				258339-258496	3
	Exon-3				257910-258038	2
	Exon-4			chr-frag	257750-257799	1
Zebraf.	**Genes-3**	Novel	ENSDARG00000014817	ctg	25692-28113	S+1
	Exon-1			2571	25692-25848	1
	Exon-2				25935-26063	2
	Exon-3				27626-27688	3
	Exon-4				27928-28040	4
	Exon-5				28045-28113	5
Mouse	**Genes-4**	Htf9c	ENSMUSG00000022721	16	17721811-17726633	
	Transcrpt-1		ENSMUST00000014843			S+1
	Exon-1				17722410-17722661	1
	Exon-2				17722664-17722732	2
	Exon-3				17722735-17722983	3
	Exon-4				17722988-17723032	4
	Exon-5				17723035-17723067	5
	Exon-6				17723070-17723108	6
	Exon-7				17723111-17723169	7
	Exon-8				17723586-17723694	8
	Exon-9				17723952-17724133	9
	Exon-10				17724288-17724402	10
	Exon-11				17724491-17724606	11
	Exon-12				17724681-17724792	12
	Exon-13				17725258-17725380	13
	Exon-14				17725534-17725609	14
	Exon-15				17725851-17725967	15
	Exon-16				17726056-17726152	16
	Exon-17				17726232-17726633	17
Mouse	**Transcrpt-2**	splice-2	ENSMUST00000050704	22		S+1
	Exon-1	P70220			17721811-17722023	1
	Exon-2				17722026-17723162	2
	Exon-3				17723583-17723694	3
	Exon-4				17723952-17724133	4
	Exon-5				17724288-17724402	5
	Exon-6				17724491-17724606	6
	Exon-7				17724681-17724792	7

	Exon-8				17725258-17725380	8
	Exon-9				17725534-17725609	9
	Exon-10				17725851-17725967	10
	Exon-11				17726056-17726152	11
	Exon-12				17726232-17726633	12
Mouse	**Transcrpt-3**	splice-3	ENSMUST00000059696	22	17722137-17723165	S+1
	Exon-1				17723586-17723694	1
	Exon-2				17723952-17724133	2
	Exon-3				17724288-17724402	3
	Exon-4				17724491-17724606	4
	Exon-5				17724681-17724792	5
	Exon-6				17725258-17725380	6
	Exon-7				17725534-17725609	7
	Exon-8				17725534-17725609	8
	Exon-9				17725851-17725967	9
	Exon-10				17726056-17726152	10
	Exon-11				17726232-17726627	11
Human	**Genes-4**	NM_022727	ENSG00000099899	22	18474640-18478677	S-1
	Exon-1				18478115-18478677	11
	Exon-2				18477749-18477857	10
	Exon-3				18477343-18477524	9
	Exon-4				18477042-18477156	8
	Exon-5				18476838-18476953	7
	Exon-6				18476651-18476762	6
	Exon-7				18475763-18475885	5
	Exon-8				18475509-18475584	4
	Exon-9				18475195-18475311	3
	Exon-10				18474988-18475084	2
	Exon-11				18474640-18474871	1
Rat	**Genes-4**	Novel	ENSRNOG00000001885	11	2890154-2894289	S+1
	Exon-1				2890154-2890327	1
	Exon-2				2890330-2890392	2
	Exon-3				2890398-2890460	3
	Exon-4				2890463-2890714	4
	Exon-5				2890719-2890763	5
	Exon-6				2890766-2890798	6
	Exon-7				2890801-2890839	7
	Exon-8				2890842-2890900	8
	Exon-9				2891388-2891496	9
	Exon-10				2891830-2892011	10
	Exon-11				2892152-2892266	11
	Exon-12				2892362-2892477	12
	Exon-13				2892552-2892663	13
	Exon-14				2893128-2893250	14
	Exon-15				2893410-2893485	15
	Exon-16				2893685-2893801	16
	Exon-17				2893890-2893986	17

	Exon-18			Chr	2894067-2894289	18
Fugu	**Genes-4**	P70222	SINFRUG00000153624	Scfold	1988-4563	S+1
	Exon-1			10131	1988-2463	1
	Exon-2				2563-2671	2
	Exon-3				2787-2968	3
	Exon-4				3151-3265	4
	Exon-5				3342-3457	5
	Exon-6				3532-3643	6
	Exon-7				3729-3851	7
	Exon-8				3943-4018	8
	Exon-9				4091-4207	9
	Exon-10				4272-4368	10
	Exon-11			chr-frag	4443-4563	11
Zebraf.	**Genes-4**	Novel	ENSDARG00000002830	ctg	243974-250692	S+1
	Exon-1			9659	243974-244155	1
	Exon-2				244322-244436	2
	Exon-3				247392-247507	3
	Exon-4				247587-247698	4
	Exon-5				247835-247957	5
	Exon-6				248036-248111	6
	Exon-7				248508-248624	7
	Exon-8				248706-248802	8
	Exon-9				250385-250692	9
Mouse	**Genes-5**	Novel	ENSMUSG00000022718	16	17728654-17756415	S-1
	Exon-1				17756256-17756415	12
	Exon-2				17753746-17753888	11
	Exon-3				17753352-17753634	10
	Exon-4				17751451-17751648	9
	Exon-5				17751171-17751272	8
	Exon-6				17750298-17750396	7
	Exon-7				17745853-17745935	6
	Exon-8				17733011-17733111	5
	Exon-9				17732681-17732787	4
	Exon-10				17731326-17731453	3
	Exon-11				17729804-17729917	2
	Exon-12				17728654-17728737	1
Mouse	**Genes-6**	Dgcr8	ENSMUSG00000048552	16	17756817-17757548	S-1
	Exon-1				17756817-17757548	1
Human	**Genes-6**	Dgcr8	ENSG00000128191	22	18442368-18473949	S+1
	Exon-1				18442368-18442460	1
	Exon-2				18447764-18448760	2
	Exon-3				18449239-18449398	3
	Exon-4				18451746-18451888	4
	Exon-5				18452053-18452335	5
	Exon-6				18453512-18453709	6
	Exon-7				18453946-18454047	7
	Exon-8				18454886-18454984	8
	Exon-9				18456790-18456872	9
	Exon-10				18468254-18468354	10

	Exon-11				18468669-18468775	11
	Exon-12				18469348-18469475	12
	Exon-13				18470967-18471080	13
	Exon-14				18472102-18473949	14
Rat	**Genes-6**	Novel	ENSRNOG00000001886	11	2921531-2922259	S-1
	Exon-1			Chr	2921531 2922259	1
Fugu	**Genes-6**	NM_022775	SINFRUG00000153636	Scfold	6602-10393	S-1
	Exon-1			10139788-10393		11
	Exon-2				9202-9364	10
	Exon-3				8952-9094	9
	Exon-4				8474-8807	8
	Exon-5				8200-8397	7
	Exon-6				7953-8054	6
	Exon-7				7765-7863	5
	Exon-8				7591-7673	4
	Exon-9				7278-7378	3
	Exon-10				7068-7174	2
	Exon-11				6831-6995	1
	Exon-12			chr-frag	6602-6735	
Zebrafish	**Genes-6**	Novel	ENSDARG00000010858	ctg	252222-261614	S-1
	Exon-1			9659261009-261614		13
	Exon-2				260285-260447	12
	Exon-3				258410-258552	11
	Exon-4				257958-258294	10
	Exon-5				257538-257735	9
	Exon-6				257352-257453	8
	Exon-7				256614-256712	7
	Exon-8				256426-256508	6
	Exon-9				256240-256340	5
	Exon-10				255981-256087	4
	Exon-11				254246-254373	3
	Exon-12				254036-254149	2
	Exon-13				252222-252714	1
Mouse	**Genes-7**	D16H-22S680E	ENSMUSG00000013539	16	17773959-17817029	S-1
	Exon-1				17816931-17817029	9
	Exon-2				17797489-17797582	8
	Exon-3				17789778-17789866	7
	Exon-4				17785771-17785890	6
	Exon-5				17784001-17784115	5
	Exon-6				17781102-17781172	4
	Exon-7				17776815-17776968	3
	Exon-8				17775818-17775922	2
	Exon-9				17773959-17774684	1
Human	**Genes-7**	NM_152906	ENSG00000183597	22	18383191-18428003	S+1
	Exon-1				18399101-18399171	1
	Exon-2				18405432-18405520	2
	Exon-3				18414542-18414661	3

	Exon-4				18415514-18415628	4
	Exon-5				18418020-18418090	5
	Exon-6				18423607-18423760	6
	Exon-7				18425415-18425519	7
	Exon-8				18426619-18428003	8
Rat	**Genes-7**	Novel	ENSRNOG00000001887	11	2939849-2955741	S-1
	Exon-1				2955651-2955741	7
	Exon-2				2950174-2950293	6
	Exon-3				2948684-2948798	5
	Exon-4				2945704-2945774	4
	Exon-5				2942062-2942215	3
	Exon-6				2940501-2940605	2
	Exon-7			Chr	2939849-2939966	1
Fugu	**Genes-7**	Q96M16	SINFRUG00000153652	Scfold	12439-19759	S-1
	Exon-1			101	319672-19759	6
	Exon-2				18673-18792	5
	Exon-3				18426-18543	4
	Exon-4				13332-13517	3
	Exon-5				12810-12986	2
	Exon-6				12439-12556	1

Gene Organization and Comparative Genetics of BACs rp23-213m14 & rp23-81d13b

Table 18 Gene Organization and Comparative Genetics of BACs rp23-213m14 & rp23-81d13b

Organ	BAC (2)	overlap	Accession #			
Mouse	**rp23-213m14**		**AC090977**	Chr/	or Chr Scaffold	mus
	rp23-81d13b		**AC123977**			
Organism	Genes	Name	Ensembl ID	#	Chromosome Location	#
Mouse	**Genes-1**	rik-1	2310042E22Rik	16	20902529-20903484	S-1
	Exon-1	rp23-81d13b			20902642-20903484	1
Mouse	**Genes-2**	Ephb3	ENSMUSG00000005958	16	20954459-20972968	S+1
	Exon-1				20954459-20954928	1
	Exon-2				20962555-20962619	2
	Exon-3				20963990-20964662	3
	Exon-4				20966861-20967016	4
	Exon-5				20967680-20968030	5
	Exon-6				20968160-20968284	6
	Exon-7				20968459-20968618	7
	Exon-8				20968704-20968812	8
	Exon-9				20969120-20969172	9
	Exon-10				20970085-20970207	10
	Exon-11				20970312-20970559	11
	Exon-12				20970950-20971165	12
	Exon-13				20971292-20971441	13
	Exon-14				20971530-20971723	14
	Exon-15				20971814-20971969	15
	Exon-16				20972073-20972968	16
Human	**Genes-2**	EPHB3	ENSG00000182580		3185682024-185702223	S+1
	Exon-1				185682024-185682168	1
	Exon-2				185691133-185691197	2
	Exon-3				185692319-185692991	3
	Exon-4				185695645-185695800	4
	Exon-5				185696657-185696998	5
	Exon-6				185697158-185697282	6
	Exon-7				185697473-185697632	7
	Exon-8				185697713-185697821	8
	Exon-9				185698229-185698281	9
	Exon-10				185699292-185699414	10
	Exon-11				185699502-185699749	11
	Exon-12				185700217-185700432	12
	Exon-13				185700544-185700693	13

	Exon-14				185700787-185700980	14
	Exon-15				185701068-185701223	15
	Exon-16				185701329-185702223	16
Rat	**Genes-2**	Novel	ENSRNOG00000001760		115857994-5867748	S+1
	Exon-1				5857994-5858059	1
	Exon-2				5859391-5860063	2
	Exon-3				5862334-5862395	3
	Exon-4				5862398-5862488	4
	Exon-5				5863167-5863517	5
	Exon-6				5863645-5863769	6
	Exon-7				5863947-5864106	7
	Exon-8				5864195-5864303	8
	Exon-9				5864632-5864684	9
	Exon-10				5865597-5865719	10
	Exon-11				5865825-5866072	11
	Exon-12				5866514-5866729	12
	Exon-13				5866847-5866996	13
	Exon-14				5867087-5867280	14
	Exon-15				5867376-5867531	15
	Exon-16			Chr	5867643-5867748	16
Fugu	**Genes-2**	EPHB3	SINFRUG00000144679	Scfold	389439-407130	S+1
	Exon-1			45	389439-390111	1
	Exon-2				393116-393271	2
	Exon-3				394115-394450	3
	Exon-4				396384-396508	4
	Exon-5				396843-396999	5
	Exon-6				397261-397419	6
	Exon-7				398985-399107	7
	Exon-8				400869-401116	8
	Exon-9				401463-401678	9
	Exon-10				401863-402048	10
	Exon-11				402151-402350	11
	Exon-12				405396-405551	12
	Exon-13			chr-frag	407025-407130	13
Zebraf.	**Genes-2**	Novel	ENSDARG00000016114	ctg	199356-223110	S+1
	Exon-1			9398	199356-199893	
	Exon-2				204502-204657	
	Exon-3				213835-213901	
	Exon-4				214025-214112	
	Exon-5				214118-214378	
	Exon-6				214381-214528	
	Exon-7				216185-216293	
	Exon-8				217229-217281	
	Exon-9				217392-217514	
	Exon-10				218550-218797	
	Exon-11				219099-219314	
	Exon-12				219394-219543	
	Exon-13				219632-219825	
	Exon-14				222888-222939	

	Exon-15				222944-222994	
	Exon-16				222997-223110	
Mouse	**EST-1**	Gene-2?	ENSMUSESTG00000021608	16	20960189-20964506	S+1
	Exon-1				20960189-20960275	1
	Exon-2				20962555-20962619	2
	Exon-3				20963990-20964506	3
Mouse	**EST-2**	Gene-2?	ENSMUSESTG00000021609	16	20966880-20971631	S+1
	Exon-1				20966880-20967016	1
	Exon-2				20967680-20968030	2
	Exon-3				20968160-20968284	3
	Exon-4				20968459-20968618	4
	Exon-5				20968704-20968812	5
	Exon-6				20969120-20969172	6
	Exon-7				20970085-20970207	7
	Exon-8				20970312-20970559	8
	Exon-9				20970950-20971165	9
	Exon-10				20971292-20971441	10
	Exon-11				20971530-20971631	11
Mouse	**Genes-3**	**Novel/EST**	ENSMUSESTG00000020826	16	21004026-21046484	S-1
	Exon-1				21046383-21046484	4
	Exon-2				21032001-21032164	3
	Exon-3				21030210-21030494	2
	Exon-4				21004026-21004147	1
Mouse	**Genes-4**	Novel	ENSMUSG00000049167	16	21082584-21082968	S-1
	Exon-1				21082846-21082968	2
	Exon-2				21082584-21082844	1

Gene Organization and Comparative Genetics of BAC rp23-46k8

Table 19 Gene Organization and Comparative Genetics of BAC rp23-46k8

Organism	**BAC**		Accession #			
Mouse	**rp23-46k8**		**AC113265**	Chr/	or Chr Scaffold	mus
Organism	Genes	Name	Ensembl ID	#	Chromosome Location	#
Mouse	**Genes-1**	Park2	ENSMUSG00000047566	17	10191863 -10617531	S+1
	Exon-1				10191863-10192030	1
	Exon-2				10362196-10362436	2
	Exon-3				10428324-10428445	3
	Exon-4				10559284-10559367	4

	Exon-5				10617412-10617531	5
Human	**Genes-1**	Park2	ENSG00000185345	6	161604559-162697934	S-1
	Exon-1				162697770-162697934	9
	Exon-2				162516985-162517225	8
	Exon-3				162455591-162455712	7
	Exon-4				162308551-162308634	6
	Exon-5				162227762-162227877	5
	Exon-6				162040232-162040368	4
	Exon-7				161970535-161970591	3
	Exon-8				161823815-161823876	2
	Exon-9				161803296-161803463	1
Rat-1	**Genes-1**	Park2	ENSRNOG00000018022	1	36319391-36663031	S-1
	Exon-1				36662906-36663031	3
	Exon-2				36533324-36533407	2
	Exon-3				36467304-36467420	1
Rat-2	**Genes-1**	Q8K5C2	ENSRNOG00000018013	1	35900000-36112377	S-1
	Exon-1				36112315-36112377	4
	Exon-2				36096629-36096832	3
	Exon-3				35910896-35911052	2
	Exon-4			Chr	35900000-35900109	1
Fugu	**Genes-1**	Q8WW07	SINFRUG00000132227	Scfold	108073-110090	S+1
	Exon-1			289	108073-108237	1
	Exon-2				108307-108593	2
	Exon-3				108744-108843	3
	Exon-4				108986-109069	4
	Exon-5				109201-109316	5
	Exon-6				109458-109594	6
	Exon-7				109709-109770	7
	Exon-8				109893-110090	8
Mouse	**Genes-2**	**Novel/EST**	ENSMUSESTG00000018819	17	10617453-10735667	
	Exon-1	(not same			10617453-10617527	1
	Exon-2	as above)			10735331-10735667	2

Appendix II

Check Chromosome 22 Contig Assembly Primers

Table 20 Check Chromosome 22 Contig Assembly Primers

#	Contig	Primer #	Primer
A01	NT_011520.8	F23083825	ACACACACATCCACACGCTC
A02	NT_011520.8	F23083561	CACATGCACGTTCACACATG
A03	NT_011520.8	F23083459	GTGCTGCTTCTACAAACGCA
A04	NT_011520.8	F23083459	GTCTCCTTCCCGTACTGGCT
A05	NT_011520.8	F23083311	CATCTAGGGCAGGTGTGAGC
A06	NT_011520.8	F23083210	TATTGTGGAGACCTGCATGG
A07	NT_011520.8	F23083090	TTCTGTGACGGATTCGTCCT
A08	NT_011520.8	F23082961	GTAGACTTGGGCAGCTCCTG
A09	NT_011520.8	F23082848	GGATGATCTCCATTCCCATC
A10	NT_011520.8	F23082747	GTGAGGCAAAGACTTTCCCC
A11	NT_011520.8	F23082620	AGGACTATGTCGCCTGGTTG
A12	NT_011520.8	F23082515	TGTTTTCCTGAAGCCCTCTG
B01	NT_011520.8	F23083825.comp	GAGCGTGTGGATGTGTGTGT
B02	NT_011520.8	F23083662.comp	CATGTGTGAACGTGCATGTG
B03	NT_011520.8	F23083561.comp	TGCGTTTGTAGAAGCAGCAC
B04	NT_011520.8	F23083459.comp	AGCCAGTACGGGAAGGAGAC
B05	NT_011520.8	F23083311.comp	GCTCACACCTGCCCTAGATG
B06	NT_011521.1	R000221	GACACCAAAAGCACGAGCTA
B07	NT_011521.1	R000393	CAGCCTGTGACAGAGCAAGA
B08	NT_011521.1	R000495	GTGCACACCTGTAGTCCCAG
B09	NT_011521.1	R000598	ACATGACAAAACCCCACCTC
B10	NT_011521.1	R000754	AGGGGACGGGATCTTACTTC
B11	NT_011521.1	R000883	GTGGCCACACATGACAGAGA
B12	NT_011521.1	R000989	TTACAGGTGTGAGCGGGACT
C01	NT_011521.1	R001095	GGCCGGCTAATTTTGGTATT
C02	NT_011521.1	R001197	AGTGCAGTGACGTGATGTCG
C03	NT_011521.1	R001300	CAGGCACTCAGAGCATTTCA
C04	NT_011521.1	R001457	ACAGTGCTGTGTGTATGCCC
C05	NT_011521.1	R001579	GCTGGGGTAATCAAAAGGCT
C06	NT_011521.1	R000221.comp	TAGCTCGTGCTTTTGGTGTC
C07	NT_011521.1	R000393.comp	TCTTGCTCTGTCACAGGCTG
C08	NT_011521.1	R000495.comp	CTGGGACTACAGGTGTGCAC
C09	NT_011521.1	R000598.comp	GAGGTGGGGTTTTGTCATGT
C10	NT_011521.1	R000754.comp	GAAGTAAGATCCCGTCCCCT

PCR to check primers list and Amplification Primers

C (NT_011519.9) Primers:

>A01:NT_011519.9_000_F3457258 GAGAGTTTGTCTGGGCCAAG
>A02:NT_011519.9_000_F3457093 GTTTGGACACCTGCTCCTCT
>A03:NT_011519.9_000_F3456982 AGCCCCACGTGATGATATGT
>A04:NT_011519.9_000_F3456881 GATTAGCTGGGTGAAGACGC
>A05:NT_011519.9_000_F3456776 AGGTGGGACTCGAGATTGTG
>A06:NT_011519.9_000_F3456669 ATAATGACAAGCGCTCAGGC
>A07:NT_011519.9_000_F3456549 GCCAGCAAAGTGTCTGGTTT
>A08:NT_011519.9_000_F3456437 CCTGCTTTGACCAGGTTAGG
>A09:NT_011519.9_000_F3456296 GTGCATTACAGAGGCTGCTG
>A10:NT_011519.9_000_F3456165 CTTTAATAGGAGGGGCTCGG
>A11:NT_011519.9_000_F3456004 TAGAGCCTCTGGGTCAGGAC
>A12:NT_011519.9_000_F3455737 ACATTAAAGGCGAGACTCCG
>B01:NT_011519.9_000_F3457258.comp CTTGGCCCAGACAAACTCTC
>B02:NT_011519.9_000_F3457093.comp AGAGGAGCAGGTGTCCAAAC
>B03:NT_011519.9_000_F3456982.comp ACATATCATCACGTGGGGCT
>B04:NT_011519.9_000_F3456881.comp GCGTCTTCACCCAGCTAATC
>B05:NT_011519.9_000_F3456776.comp CACAATCTCGAGTCCCACCT
>B06:NT_011519.9_000_F3456669.comp GCCTGAGCGCTTGTCATTAT
>B07:NT_011519.9_000_F3456549.comp AAACCAGACACTTTGCTGGC
>B08:NT_011519.9_000_F3456437.comp CCTAACCTGGTCAAAGCAGG
>B09:NT_011519.9_000_F3456296.comp CAGCAGCCTCTGTAATGCAC
>B10:NT_011519.9_000_F3456165.comp CCGAGCCCCTCCTATTAAAG
>B11:NT_011519.9_000_F3456004.comp GTCCTGACCCAGAGGCTCTA
>B12:NT_011519.9_000_F3455737.comp CGGAGTCTCGCCTTTAATGT
>C01:NT_011519.9_000_R000135 CGATGATGATTCCATTCGAA
>C02:NT_011519.9_000_R000298 TTCCGTCGATTCTTTTCGAT
>C03:NT_011519.9_000_R000400 CCTTTCGGTTCCATTTGATG
>C04:NT_011519.9_000_R000528 TCCGTTTGATGATGATTCCA
>C05:NT_011519.9_000_R000646 CGATGATGATTCCGTTCCTT
>C06:NT_011519.9_000_R000752 CCATTCGATGATTCCATGCT
>C07:NT_011519.9_000_R000875 TCCATTCGATGATCATTCCA
>C08:NT_011519.9_000_R001014 TCCTTGATGATTCCATTCGA
>C09:NT_011519.9_000_R001134 CGATGATTCCATTCGATTCC
>C10:NT_011519.9_000_R001238 TCCATTCGATGATGATTCCA
>C11:NT_011519.9_000_R001341 TCCATTCCATTCGATGATGA
>C12:NT_011519.9_000_R001527 TCAATGAGGATTCCATTCGA
>D01:NT_011519.9_000_R000135.comp TTCGAATGGAATCATCATCG
>D02:NT_011519.9_000_R000298.comp ATCGAAAAGAATCGACGGAA
>D03:NT_011519.9_000_R000400.comp CATCAAATGGAACCGAAAGG
>D04:NT_011519.9_000_R000528.comp TGGAATCATCATCAAACGGA
>D05:NT_011519.9_000_R000646.comp AAGGAACGGAATCATCATCG
>D06:NT_011519.9_000_R000752.comp AGCATGGAATCATCGAATGG
>D07:NT_011519.9_000_R000875.comp TGGAATGATCATCGAATGGA
>D08:NT_011519.9_000_R001014.comp TCGAATGGAATCATCAAGGA
>D09:NT_011519.9_000_R001134.comp GGAATCGAATGGAATCATCG

>D10:NT_011519.9_000_R001238.comp TGGAATCATCATCGAATGGA
>D11:NT_011519.9_000_R001341.comp TCATCATCGAATGGAATGGA
>D12:NT_011519.9_000_R001527.comp TCGAATGGAATCCTCATTGA

D (NT_011520.8)
>E01:NT_011520.8_000_F23083825 ACACACACATCCACACGCTC
>E02:NT_011520.8_000_F23083662 CACATGCACGTTCACACATG
>E03:NT_011520.8_000_F23083561 GTGCTGCTTCTACAAACGCA
>E04:NT_011520.8_000_F23083459 GTCTCCTTCCCGTACTGGCT
>E05:NT_011520.8_000_F23083311 CATCTAGGGCAGGTGTGAGC
>E06:NT_011520.8_000_F23083210 TATTGTGGAGACCTGCATGG
>E07:NT_011520.8_000_F23083090 TTCTGTGACGGATTCGTCCT
>E08:NT_011520.8_000_F23082961 GTAGACTTGGGCAGCTCCTG
>E09:NT_011520.8_000_F23082848 GGATGATCTCCATTCCCATC
>E10:NT_011520.8_000_F23082747 GTGAGGCAAAGACTTTCCCC
>E11:NT_011520.8_000_F23082620 AGGACTATGTCGCCTGGTTG
>E12:NT_011520.8_000_F23082515 TGTTTTCCTGAAGCCCTCTG
>F01:NT_011520.8_000_F23083825.comp GAGCGTGTGGATGTGTGTGT
>F02:NT_011520.8_000_F23083662.comp CATGTGTGAACGTGCATGTG
>F03:NT_011520.8_000_F23083561.comp TGCGTTTGTAGAAGCAGCAC
>F04:NT_011520.8_000_F23083459.comp AGCCAGTACGGGAAGGAGAC
>F05:NT_011520.8_000_F23083311.comp GCTCACACCTGCCCTAGATG
>F06:NT_011520.8_000_F23083210.comp CCATGCAGGTCTCCACAATA
>F07:NT_011520.8_000_F23083090.comp AGGACGAATCCGTCACAGAA
>F08:NT_011520.8_000_F23082961.comp CAGGAGCTGCCCAAGTCTAC
>F09:NT_011520.8_000_F23082848.comp GATGGGAATGGAGATCATCC
>F10:NT_011520.8_000_F23082747.comp GGGGAAAGTCTTTGCCTCAC
>F11:NT_011520.8_000_F23082620.comp CAACCAGGCGACATAGTCCT
>F12:NT_011520.8_000_F23082515.comp CAGAGGGCTTCAGGAAAACA
>G01:NT_011520.8_000_R000120 GACCCAAGAGACCCCTCACT
>G02:NT_011520.8_000_R000266 TGAGGGTGCTGACATTCAGA
>G03:NT_011520.8_000_R000379 GTGGAGCATGCTTTCCTGTT
>G04:NT_011520.8_000_R000619 CTCTGCCCTCAGTTCTCTGC
>G05:NT_011520.8_000_R000720 GGGAACACAGCAGGATGACT
>G06:NT_011520.8_000_R000826 AGCAGAGCCCACTGAATCAA
>G07:NT_011520.8_000_R000994 GGTTGGTGTGCACAGCATAT
>G08:NT_011520.8_000_R001108 CTGAATACTCCCCGCAACTG
>G09:NT_011520.8_000_R001331 GCACCACTCTTTCAGGGAGA
>G10:NT_011520.8_000_R001446 GCCTCCTAAAGTGCTGGGAT
>G11:NT_011520.8_000_R001549 ACCACACCCGGCTAATTTTT
>G12:NT_011520.8_000_R001651 AGTGCAATGGCATGATCTTG
>H01:NT_011520.8_000_R000120.comp AGTGAGGGGTCTCTTGGGTC
>H02:NT_011520.8_000_R000266.comp TCTGAATGTCAGCACCCTCA
>H03:NT_011520.8_000_R000379.comp AACAGGAAAGCATGCTCCAC
>H04:NT_011520.8_000_R000619.comp GCAGAGAACTGAGGGCAGAG
>H05:NT_011520.8_000_R000720.comp AGTCATCCTGCTGTGTTCCC
>H06:NT_011520.8_000_R000826.comp TTGATTCAGTGGGCTCTGCT
>H07:NT_011520.8_000_R000994.comp ATATGCTGTGCACACCAACC
>H08:NT_011520.8_000_R001108.comp CAGTTGCGGGGAGTATTCAG
>H09:NT_011520.8_000_R001331.comp TCTCCCTGAAAGAGTGGTGC
>H10:NT_011520.8_000_R001446.comp ATCCCAGCACTTTAGGAGGC

>H11:NT_011520.8_000_R001549.comp AAAAATTAGCCGGGTGTGGT
>H12:NT_011520.8_000_R001651.comp CAAGATCATGCCATTGCACT

E (NT_011521.1)

>A01:NT_011521.1_000_F767224 CGCTCCTCACCACTATAGCA
>A02:NT_011521.1_000_F767116 TACGGATGGAAAACTGAGGC
>A03:NT_011521.1_000_F766981 ACCACTGGGCCAAGAAATAA
>A04:NT_011521.1_000_F766829 GTTTTTCTTCGGTGGTTCCC
>A05:NT_011521.1_000_F766713 CCGGGGTAGATGTCCATGTA
>A06:NT_011521.1_000_F766577 CCAGCGGCCTTATGTTAAAA
>A07:NT_011521.1_000_F766438 AGCTGTAATCCCACAGCTCC
>A08:NT_011521.1_000_F766328 GTTCTTGGCCTGAAGATGCT
>A09:NT_011521.1_000_F766204 TCTGTAGGAAGATGGGGGCT
>A10:NT_011521.1_000_F766089 AAAACCCTTGCCTGTCTCTG
>A11:NT_011521.1_000_F765979 GGGCAGCAGATCCTTTAAGA
>A12:NT_011521.1_000_F765800 ATCTGTTAATCACTCCGGCG
>B01:NT_011521.1_000_F767224.comp TGCTATAGTGGTGAGGAGCG
>B02:NT_011521.1_000_F767116.comp GCCTCAGTTTTCCATCCGTA
>B03:NT_011521.1_000_F766981.comp TTATTTCTTGGCCCAGTGGT
>B04:NT_011521.1_000_F766829.comp GGGAACCACCGAAGAAAAAC
>B05:NT_011521.1_000_F766713.comp TACATGGACATCTACCCCGG
>B06:NT_011521.1_000_F766577.comp TTTTAACATAAGGCCGCTGG
>B07:NT_011521.1_000_F766438.comp GGAGCTGTGGGATTACAGCT
>B08:NT_011521.1_000_F766328.comp AGCATCTTCAGGCCAAGAAC
>B09:NT_011521.1_000_F766204.comp AGCCCCCATCTTCCTACAGA
>B10:NT_011521.1_000_F766089.comp CAGAGACAGGCAAGGGTTTT
>B11:NT_011521.1_000_F765979.comp TCTTAAAGGATCTGCTGCCC
>B12:NT_011521.1_000_F765800.comp CGCCGGAGTGATTAACAGAT
>C01:NT_011521.1_000_R000221 GACACCAAAAGCACGAGCTA
>C02:NT_011521.1_000_R000393 CAGCCTGTGACAGAGCAAGA
>C03:NT_011521.1_000_R000495 GTGCACACCTGTAGTCCCAG
>C04:NT_011521.1_000_R000598 ACATGACAAAACCCCACCTC
>C05:NT_011521.1_000_R000754 AGGGGACGGGATCTTACTTC
>C06:NT_011521.1_000_R000883 GTGGCCACACATGACAGAGA
>C07:NT_011521.1_000_R000989 TTACAGGTGTGAGCGGGACT
>C08:NT_011521.1_000_R001095 GGCCGGCTAATTTTGGTATT
>C09:NT_011521.1_000_R001197 AGTGCAGTGACGTGATGTCG
>C10:NT_011521.1_000_R001300 CAGGCACTCAGAGCATTTCA
>C11:NT_011521.1_000_R001457 ACAGTGCTGTGTGTATGCCC
>C12:NT_011521.1_000_R001579 GCTGGGGTAATCAAAAGGCT
>D01:NT_011521.1_000_R000221.comp TAGCTCGTGCTTTTGGTGTC
>D02:NT_011521.1_000_R000393.comp TCTTGCTCTGTCACAGGCTG
>D03:NT_011521.1_000_R000495.comp CTGGGACTACAGGTGTGCAC
>D04:NT_011521.1_000_R000598.comp GAGGTGGGGTTTTGTCATGT
>D05:NT_011521.1_000_R000754.comp GAAGTAAGATCCCGTCCCCT
>D06:NT_011521.1_000_R000883.comp TCTCTGTCATGTGTGGCCAC
>D07:NT_011521.1_000_R000989.comp AGTCCCGCTCACACCTGTAA
>D08:NT_011521.1_000_R001095.comp AATACCAAAATTAGCCGGCC
>D09:NT_011521.1_000_R001197.comp CGACATCACGTCACTGCACT
>D10:NT_011521.1_000_R001300.comp TGAAATGCTCTGAGTGCCTG
>D11:NT_011521.1_000_R001457.comp GGGCATACACACAGCACTGT
>D12:NT_011521.1_000_R001579.comp AGCCTTTTGATTACCCCAGC

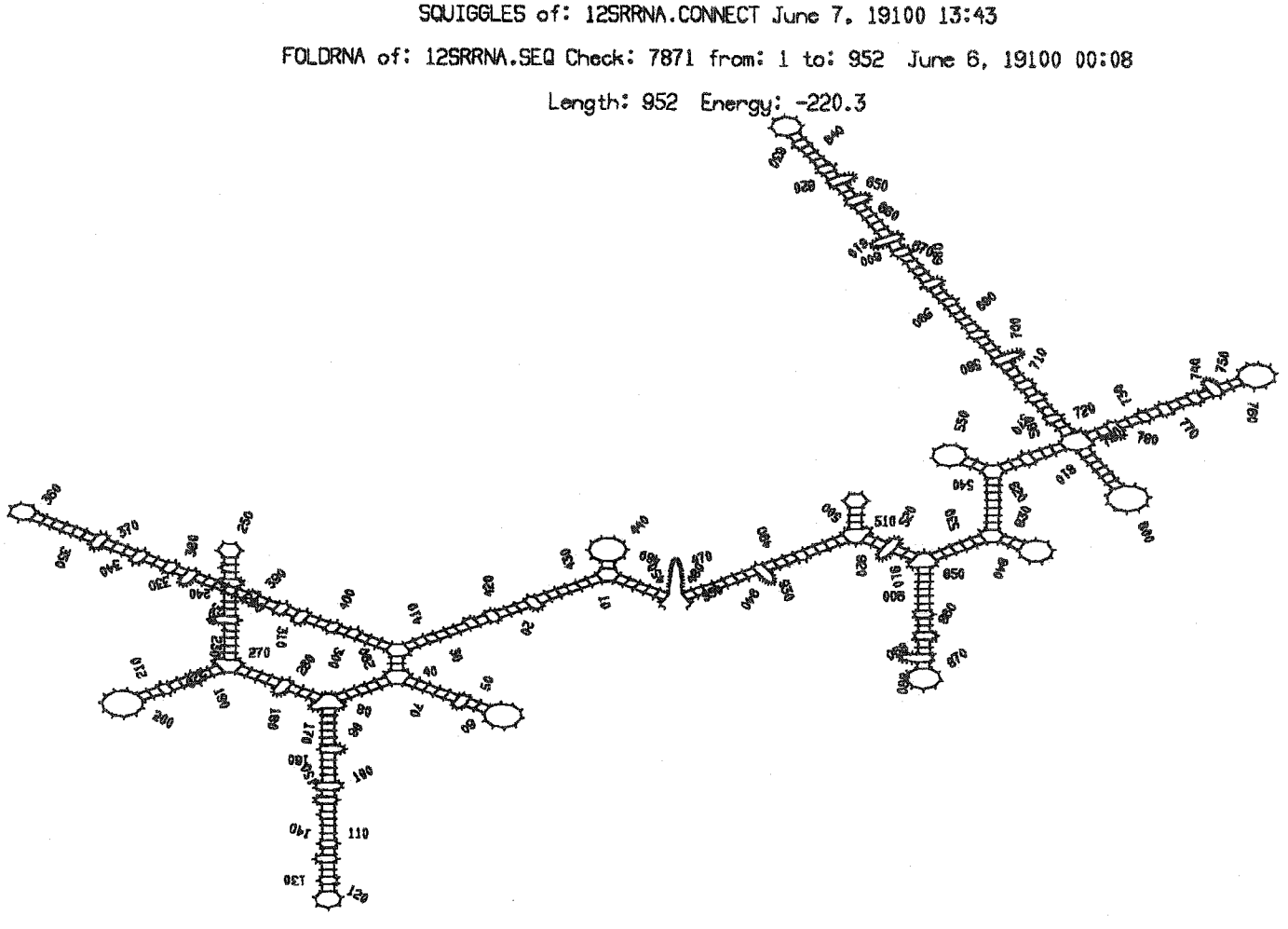
F (NT_011522.3)

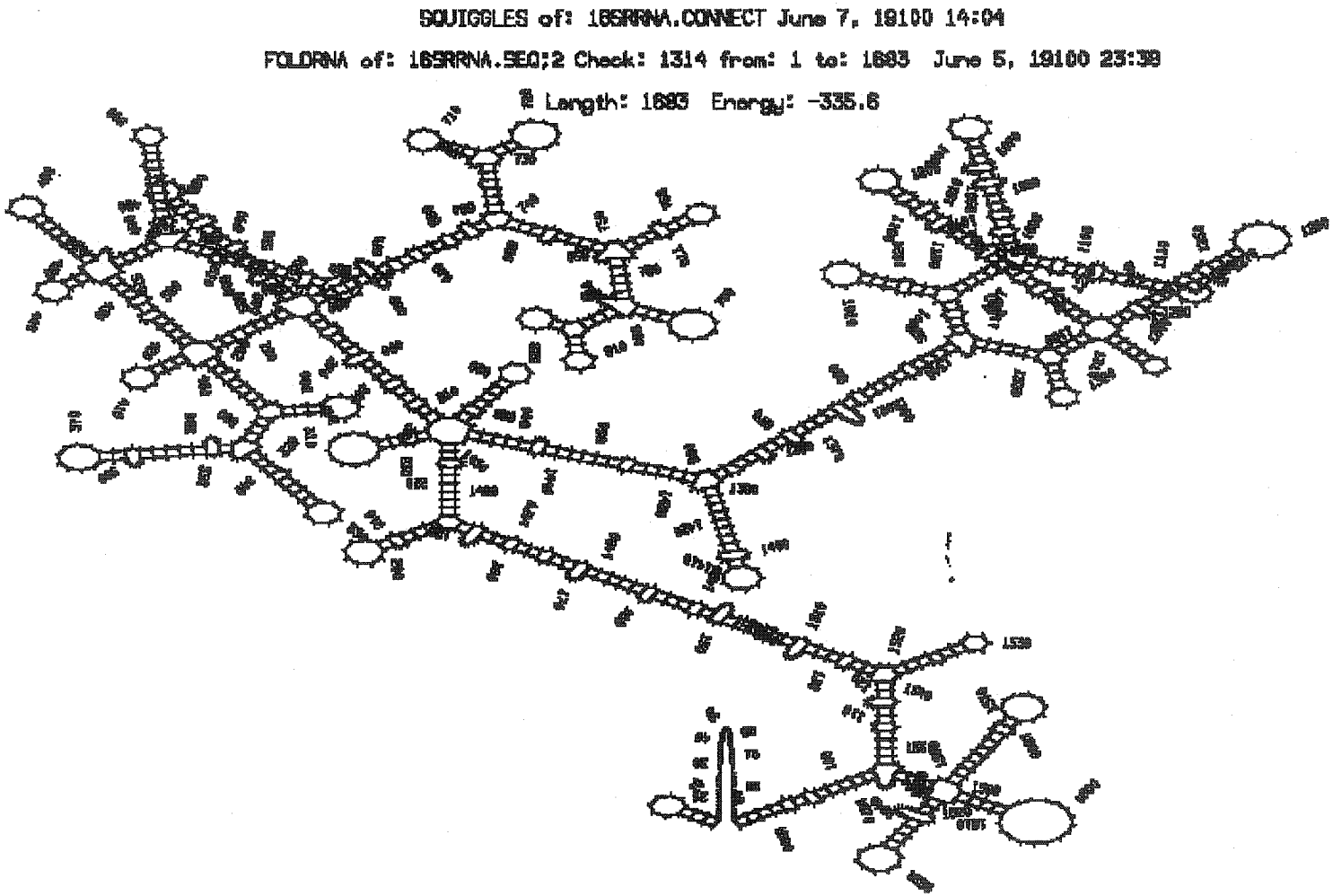
>E01:NT_011522.3_000_F1527875 CAGAGCAAGACCCTGTCTCA
>E02:NT_011522.3_000_F1527768 GGGGCTGCAATCTCAACTAC
>E03:NT_011522.3_000_F1527666 GCATGAGGATCACCTGAGGT
>E04:NT_011522.3_000_F1527556 GGCACAGTGACTCATGCTTG
>E05:NT_011522.3_000_F1527414 TGGGTGACACAGCAAGACTC
>E06:NT_011522.3_000_F1527310 TGCACCTGTAATCCCAGCTA
>E07:NT_011522.3_000_F1527207 GCAGGCAGATCACTTGAGGT
>E08:NT_011522.3_000_F1526940 ACTTAACCGCATCCCAGAAC
>E09:NT_011522.3_000_F1526819 GGGAGGGGAAGGTTGAGTAA
>E10:NT_011522.3_000_F1526667 CAGAGCGAAACTCCATCTCA
>E11:NT_011522.3_000_F1526565 TAGTCCCAGCTACTCGGGAG
>E12:NT_011522.3_000_F1526438 GGAGACAGAGCTTGCAGTGA
>F01:NT_011522.3_000_F1527875. comp TGAGACAGGGTCTTGCTCTG
>F02:NT_011522.3_000_F1527768.comp GTAGTTGAGATTGCAGCCCC
>F03:NT_011522.3_000_F1527666.comp ACCTCAGGTGATCCTCATGC
>F04:NT_011522.3_000_F1527556.comp CAAGCATGAGTCACTGTGCC
>F05:NT_011522.3_000_F1527414.comp GAGTCTTGCTGTGTCACCCA
>F06:NT_011522.3_000_F1527310.comp TAGCTGGGATTACAGGTGCA
>F07:NT_011522.3_000_F1527207.comp ACCTCAAGTGATCTGCCTGC
>F08:NT_011522.3_000_F1526940.comp GTTCTGGGATGCGGTTAAGT
>F09:NT_011522.3_000_F1526819.comp TTACTCAACCTTCCCCTCCC
>F10:NT_011522.3_000_F1526667.comp TGAGATGGAGTTTCGCTCTG
>F11:NT_011522.3_000_F1526565.comp CTCCCGAGTAGCTGGGACTA
>F12:NT_011522.3_000_F1526438.comp TCACTGCAAGCTCTGTCTCC
>G01:NT_011522.3_000_R000125 GCTCATGTGTGGCTTGGTAA
>G02:NT_011522.3_000_R000242 GAGACCAGGTTCCAGTGTGG
>G03:NT_011522.3_000_R000354 GAGGCCAATCTGAGCAGACT
>G04:NT_011522.3_000_R000456 CGGGGACTCTCAGTGTCTTT
>G05:NT_011522.3_000_R000580 CCCCCATCTACCATCTGTTC
>G06:NT_011522.3_000_R000683 CCATCCATCCACCAATTACC
>G07:NT_011522.3_000_R000789 GCCCACCCATTCACTTACTC
>G08:NT_011522.3_000_R000891 ACGTAGCCCTCCATCTACCC
>G09:NT_011522.3_000_R001001 CACCCATCATTCACCCAGTA
>G10:NT_011522.3_000_R001144 GCATTCTCCCCTGGAAGAGT
>G11:NT_011522.3_000_R001252 GTTTCATCAAGCATCTGCCC
>G12:NT_011522.3_000_R001369 GGTGGGGGCAAATAAAACAT
>H01:NT_011522.3_000_R000125.comp TTACCAAGCCACACATGAGC
>H02:NT_011522.3_000_R000242.comp CCACACTGGAACCTGGTCTC
>H03:NT_011522.3_000_R000354.comp AGTCTGCTCAGATTGGCCTC
>H04:NT_011522.3_000_R000456.comp AAAGACACTGAGAGTCCCCG
>H05:NT_011522.3_000_R000580.comp GAACAGATGGTAGATGGGGG
>H06:NT_011522.3_000_R000683.comp GGTAATTGGTGGATGGATGG
>H07:NT_011522.3_000_R000789.comp GAGTAAGTGAATGGGTGGGC
>H08:NT_011522.3_000_R000891.comp GGGTAGATGGAGGGCTACGT
>H09:NT_011522.3_000_R001001.comp TACTGGGTGAATGATGGGTG
>H10:NT_011522.3_000_R001144.comp ACTCTTCCAGGGGAGAATGC
>H11:NT_011522.3_000_R001252.comp GGGCAGATGCTTGATGAAAC
>H12:NT_011522.3_000_R001369.comp ATGTTTTATTTGCCCCCACC

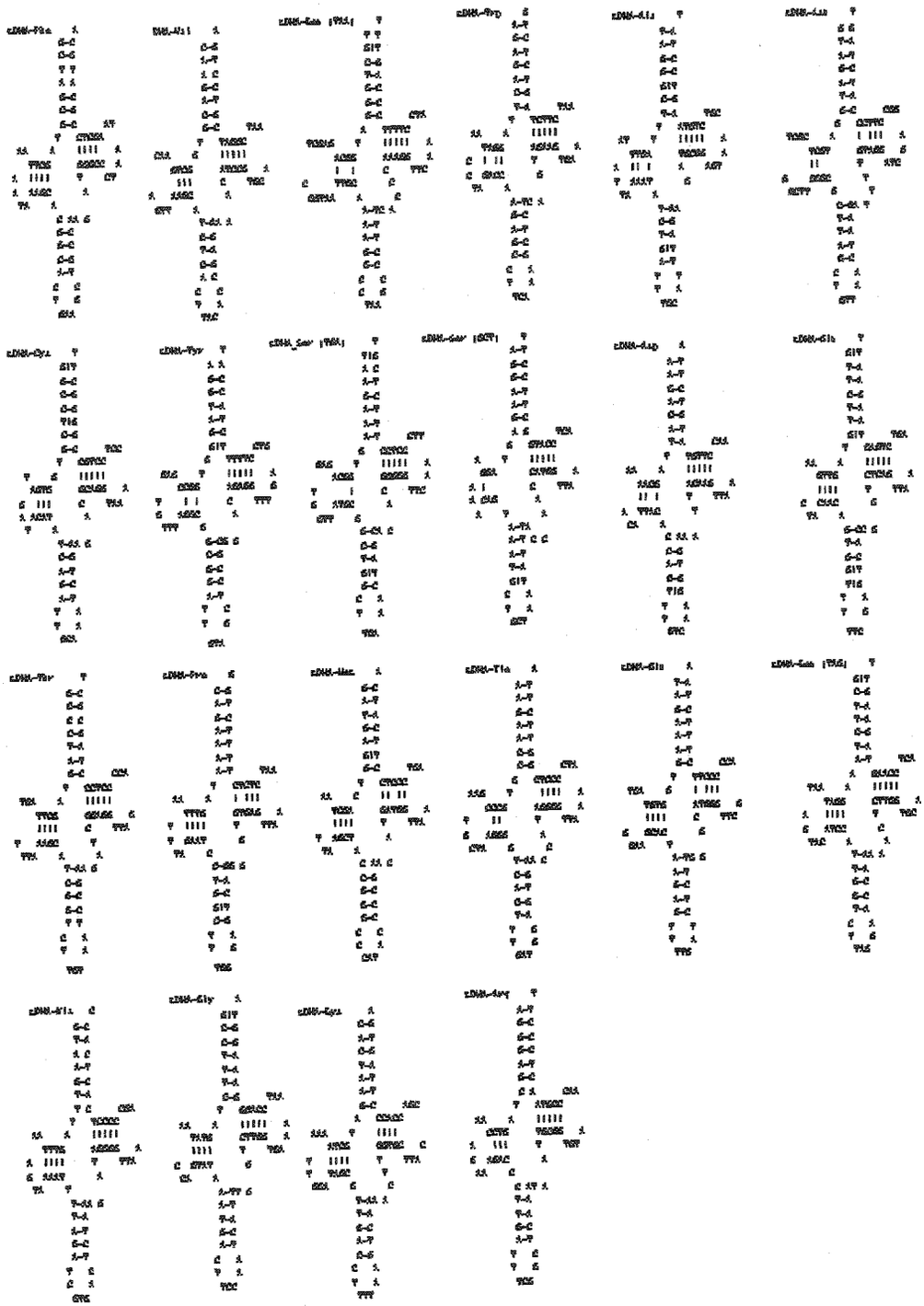
Appendix III

tRNA and RRNA Secondary Structures

The secondary structure was predicted on the tRNA sequences as were the rRNA sequences [NIP nucleotide interpretation program for tRNA (Staden, 1986)] and [Plotfold for rRNA use the suggested nucleotide sequence to predict their structures (Zuker, 1981)] and as shown below. The free energies of the rRNA secondary structures were -335.6 kcal/mole for the 16S subunit and -220.3 kcal/mole for the 12S subunit (Broughton *et al.* 2001).


SQUIGGLES of: 12SRRNA.CONNECT June 7, 19100 13:43
FOLDRNA of: 12SRRNA.SEQ Check: 7871 from: 1 to: 952 June 6, 19100 00:08
Length: 952 Energy: -220.3


SQUIGGLES of: 16SRRNA.CONNECT June 7, 19100 14:04
FOLDRNA of: 16SRRNA.SEQ;2 Check: 1314 from: 1 to: 1693 June 5, 19100 23:39
Length: 1693 Energy: -335.6



Appendix IV

Phi-29 (TempliPhi) Methods

(The unabbreviated set of Phi-29 protocols.)

P-1: Phi-29, the original protocol from Amersham (Phi-29 for plasmid-based DNA amplification)

The Phi-29 mix needed one nanagram of Isolated pUC + insert from in 1 µl of DNA, or one colony picked with the colony picker were to be added to 9 µl of buffer and heated 95 °C for three minutes, cooled and 10 µl of Phi-29 mix (5 U) was added to the reaction. The reaction was mixed and incubated at 30 °C for 6 hours. It should produce at least 1 ug (probably 2-3 ug). Then, after vortexing/mixing the amplified mixture, one needed 1-4 µl for each sequencing reaction.

P-2: Phi-29 for plasmid-based DNA amplification, cleaning and sequencing from an isolated template in a 96 well thermocycling plate

Using the Hydra, add 20 µl of sterile double distilled water is added to a 96 well thermocycler plate. To prevent evaporation, between all steps cover and keep plates on ice. Add 2 µl of isolated DNA (~1 ng), then centrifuge up and down. Next, cover plate with thermocycler sealer and heat at 95 °C for 3 minutes. Remove the correct number of phi-29 mix plates, 96 well thermocycler plates from the -80 freezer that contained 2 l/well of phi-29 mix or predispenced enzyme mix tubes with 100-500 µl/tube and place them on an ice water bath in the cold room (4 °C) to thaw ~10 minutes. This thawing time should not be exceed, because if it is the enzyme will produce a significant level of

nonspecific DNA. Use the Hydra and transfer 2 µl of the supernantant to the enzyme mix plate or place the 2 µl DNA solution into a new 96 well thermocycler plate and add 2 µl of the enzyme mix with a repeat pipette to the side and knock down, then cover with thermocycler sealer. Vortex the plate gently on a flat top vortex machine moving the plate in a circular motion for 8 seconds to mix. Next, place the DNA + enzyme plate in an ice water bath and transport to the centrifuge for a 1000 rpm, Beckman C56R centrifuge, for 1-2 seconds to concentrate the reaction mix in the bottom of the plate. Then, place the plate in the thermocycler for 12 hours at 30 °C with the volume setting of 5 µl. After the 12 hours, heat the thermocycler to 95 °C for 5 minutes. Using the hydra, add 16 µl of sterile double distilled water bringing the total volume to 20 µl/well. Then, vortex for 30 seconds and add 2 µl of 7.5 KOAc to each well with the Hydra or repeat pipette. Centrifuge the KOAc down and vortex for 30 seconds, then add with the Hydra or 12 channel pipette,add 53 µl of 100% ethanol and vortex for 30 seconds. Centrifuge at 3200 for 45 minutes, and then invert on a paper towel and centrifuge for 1-2 seconds at 400 rpm, Beckman C56R centrifuge, to remove any residgual ethanol. Immediately, without drying, dissolve the DNA in 20-50 µl of sterile double distilled water with the Hydra or 12 channel pipette and vortex for 30 seconds, and centrifuge, Beckman C56R centrifuge, to collect the solution at the bottom of the wells.up and down. Check the amplification by electrophoresing ~3 µl of DNA on a 0.6% agarose gel. To sequence, use 2 to 4 µl of DNA with the ABI Big Dye 3 or Amershams ET mixes diluted to 1/12 - 1/16.

P-3: Phi-29 for plasmid-based DNA amplification, cleaning and sequencing from an isolated template in a 384 well thermocycling plate

Using the Hydra, 20 µl of sterile double distilled water was added to a 384 well thermocycler plate. To prevent evaporation, between all steps cover and keep plates on ice. Add 2 µl of isolated DNA (~1ng), then centrifuge, Beckman C56R, up and down. Next, cover plate with thermocycler sealer and heat at 95 °C for 3 minutes. Remove the correct number of phi-29 mix plates, 384 well thermocycler plates from the -80 freezer (already made with 2 µl/well of phi-29 mix) or the predispenced enzyme tube with 100-500 µl/tube and place them on an ice water bath in the cold room (4 °C) to thaw ~10 minutes (no longer, the enzyme mix will amplify nonspecific DNA). Use the Hydra and transfer 200201 of the supernantant and transfer them to the enzyme plate or another new 384 well thermocycler plate and add 20 µl of the enzyme mix with a repeat pipette to the side and knock down. Vortex the plate gently on a flat top vortex machine moving the plate in a circular motion for 8 seconds to mix. Next, put DNA + enzyme plate in ice water bath and transport to the centrifuge for a 1000 rpm, Beckman C56R centrifuge, for an up and down spin. Then, place the plate in the thermocycler for 12 hours at 30 °C with the volume setting of 5 µl. After the 12 hours is up, heat the thermocycler up to 95 °C for 5 minutes next, add 6 µl of sterile double distilled water (total volume 10 µl) to each well, then vortex for 30 seconds add 2 µl of 50% of 7.5 KOAc to each well with the Hydra. Centrifuge, Beckman C56R,the KOAc down and vortex for 30 seconds, then add with the Hydra 12 µl of 100% ethanol, vortex for 30 seconds. Centrifuge, Beckman C56R, at 3200 for 45 minutes, and then invert on a paper towel and centrifuge up to 400 rpm, Beckman C56R centrifuge,. Immediately, dissolve the DNA in 12 µl of sterile double distilled water with the Hydra or 12 channel and vortex for 30 seconds, and

centrifuge, Beckman C56R,up and down. Confirm amplification by running ~3 µl of DNA on a 0.6% agarose gel. To sequence, use 2 to 4 µl of DNA with Big Dye 3 (1/12 - 1/16 dilution).

P-4: Phi-29 for plasmid-based DNA amplification, cleaning and sequencing from Colony and Glycerol Stocks template in a 96 well thermocycling plate

Use the Hydra and add 20 µl of sterile double distilled water to a 96 well thermocycler plate. To prevent evaporation, between all steps cover and keep plates on ice.

If using colonies for the template: Use the colony picker (not by hand!) to pick colonies into a 384 well thermocycler (the Qpix II robot will not pick into a 96 well thermocycler plate) plate from a shotgun library petri dish or 12x12 square petri plate using the Qpix II robot (~3-5 dips).

If using the glycerol stocks for template: Use the 12 channel and scrap the well lightly, just enough to get a small amount of ice on the tips of the 12 channel. Place the 12 channel in the well and pump the pipette 8 times.

Cover plate with thermocycler cover and heat at 95 °C. for 3 minutes. Cover plates with aluminum cover and freeze the plates ~2-3 hr. in freezer until solid. Thaw ~10 min, until completely thawed, and centrifuge 45 minutes at 3200 rpm, Beckman C56R centrifuge,s. While waiting, approx. 10 minutes before centrifuge is finished, remove correct number of phi-29 mix plates, 96 well thermocycler plates from the -80 freezer (already made with 2 µl/well of phi-29 mix) or predispenced enzyme mix tubes with 100-500 µl/tube and place them on an ice water bath in the cold room (4 °C) to thaw

~10 minutes (no longer, the enzyme mix will amplify nonspecific DNA). Once the 45 minutes of centrifugation is finished, immediately use the Hydra in the cold room, and remove 2 µl of the upper portion of the supernatant from each plate and place it into the enzyme mix plate, or place the 2 µl cellular solution into a new 96 well thermocycler plate and add 2 µl of the enzyme mix with a repeat pipette and knock down, then cover with thermocycler sealer. Vortex the plate gently for 8 seconds to mix. Put DNA + enzyme plate in ice water bath and transport to the centrifuge for a 1000 rpm, Beckman C56R centrifuge, up and down spin. Then, place the plate in the thermocycler for 12 hours at 30 °C with the volume setting of 5 µl. After the 12 hours is up, heat the thermocycler up to 95 °C. for 5 minutes Using the hydra add 16 µl of sterile double distilled water (total volume 20 µl) to each well, then vortex for 30 seconds add 2 µl of 7.5 KOAc to each well with the Hydra or repeat pipet. Knock or centrifuge, Beckman C56R,the KOAc down and vortex for 30 seconds, then add with the Hydra, repeat pipet, or 12 channel, 53 µl of 100% ethanol, vortex for 30 seconds. Centrifuge, Beckman C56R, at 3200 for 45 minutes, and then invert on a paper towel and centrifuge, Beckman C56R, up to 400. Immediately, dissolve the DNA in 20-50 µl of sterile double distilled water with the Hydra or 12 channel and vortex for 30 seconds, and centrifuge, Beckman C56R,up and down. Confirm amplification by running ~3 µl of DNA on a 0.6% agarose gel. To sequence, use 2 to 4 µl of DNA with Big Dye 3 (1/12 - 1/16 dilution).

P-5: Phi-29 for plasmid-based DNA amplification, cleaning and sequencing from Colony and Glycerol stocks template in a 384 well thermocycling plate

Use the Hydra and add 20 µl of sterile double distilled water to a 384 well thermocycler plate (viper). To prevent evaporation, between all steps cover and keep plates on ice.

If using colonies for the template: Use the colony picker (not by hand!) to pick colonies into a 384 well thermocycler plate from a shotgun library petri dish or 12x12 square petri plate using the Qpix II robot (~3-5 dips).

If using the glycerol stocks for template: Use the 12 channel and scrap the well lightly, just enough to get a small amount of ice on the tips of the 12 channel. Place the 12 channel in the well and pump the pipette 8 times.

Cover plate with thermocycler cover and heat at 95 °C. for 3 minutes. Cover plates with aluminum cover and freeze the plates ~2-3 hr. in freezer until solid. Thaw ~10 min, until completely thawed, and centrifuge 45 minutes at 3200 rpm, Beckman C56R centrifuge,s. While waiting, approx. 10 minutes before centrifuge is finished, remove correct number of phi-29 mix plates, 384 well thermocycler plates (viper) plates from the -80 freezer (already made with 2 µl/well of phi-29 mix) or predispenced enzyme mix tubes with 100-500 µl/tube and place them on an ice water bath in the cold room (4 °C) to thaw ~10 minutes (no longer, the enzyme mix will amplify nonspecific DNA). Once the 45 minutes of centrifugation is finished, immediately use the Hydra in the cold room, and remove 2 µl of the upper portion of the supernatant from each plate and place it into the enzyme mix plate, cover with thermocycler sealer. Vortex the plate gently for 8 seconds to mix. Put DNA + enzyme plate in ice water bath and transport to the

centrifuge for a 400 rpm, Beckman C56R centrifuge, up and down spin. Then, place the plate in the thermocycler for 12 hours at 30 °C with the volume setting of 5 µl. After the 12 hours is up, heat the thermocycler up to 95 °C. for 5 minutes Using the hydra add 6 µl of sterile double distilled water (total volume 10 µl) to each well, then vortex for 30 seconds add 2 µl of 50% of 7.5 KOAc to each well with the Hydra. Centrifuge, Beckman C56R,the KOAc down and vortex for 30 seconds, then add with the Hydra 12 µl of 100% ethanol, vortex for 30 seconds. Centrifuge, Beckman C56R, at 3200 for 45 minutes, and then invert on a paper towel and centrifuge, Beckman C56R, up to 400. Immediately, dissolve the DNA in 12 µl of sterile double distilled water with the Hydra or 12 channel and vortex for 30 seconds, and centrifuge, Beckman C56R,up and down. Confirm amplification by running ~3 µl of DNA on a 0.6% agarose gel. To sequence, use 2 to 4 µl of DNA with Big Dye 3 (1/12 - 1/16 dilution).

P-6: Phi-29 for plasmid-based DNA amplification, cleaning and sequencing from Fresh Freezer Media/Glycerol stocks template in a 384 well thermocycling plate

This protocol was included because it worked very well, by staring with colonies, picking and growing the cell to make the glycerol stock fresh, the DNA amplified very well and sequenced very well.

To LB medium add the antibiotic [ex. subclone amp (5mg/ml) resistent: LB 1L-20 ml amp; LB 500ml -10 ml amp; LB 250- 5ml amp] To another sterile container add LB and Freezer mix [ex. LB 900ml - 100ml freezer mix; LB 450 ml- 50ml freezer mix; LB 225ml - 25ml freezer mix] LB medium [for 1 L : 10.0g NaCl, 10.0 g Difco (0123-01-1) Bacto-Typtone, 5.0 g Difco (0127-05-3)Bacto-Yeast Extract and autoclave at least 20 minutes at 121 degresC.] 10X Freezer mix, Dr. Doris Kupfers'

recipe, to make 1 L :[62.7 g K_2HPO_4, 17.96 G KH_2PO_4, 5.0 G Sodium Citrate, 0.98 g $MgSO_4$-$7H_2O$, 8.98 $(NH_4)_2SO_4$, 440 ml of 44%Glycerol bring to 1 L with steril double distilled water and filter through a 0.2um filter]

Add 80 µl of LB + antibiotic + freezer mix to a 384 microtiter plate by V- prep or 12 channel and use the colony picker (5 dipes) to inoculate the plates. Grow the plates over night in the Higro ~22 hours. Remove the plates and put in them in the -80 degree C freezer until completely frozen ~3-4 hours. Remove the plates from the freezer and allow them to completely thaw ~ 1 hr. Transfer with the Hydra, 2 µl from the media plate and add it to 10 µl sterile double distilled water in a thermocycler plate (96 well). Cover and heat the plate in the thermocycler at 95 °C for 3 minutes. Cool to 4 °C and then, pull with the hydra, 2 µl of the media + water solution and add it to another thermocycler plate (96 well).

Thaw Phi-29 mix in ice water bath in the cold room, then add 2 µl of Phi-29 mix to each well. Centrifuge, Beckman C56R,up and down and the vortex the media/water +phi-29 mix on a flate top vortex machine for 30 seconds moving the plate in a circular direction. Put the plate in a thermocycler, volume setting of 5 µl on a 30 °C hold for 10-12 hours.

Then, heat the plate to 95 °C for 5 minutes. Using the hydra add sterile double distilled-water (total volume 20 µl) to each well, then vortex for 30 seconds, then, add 2 µl of 7.5 KOAc to each well with the Hydra or repeat pipette. Knock or centrifuge, Beckman C56R,the KOAc down and vortex for 8 seconds, then, add with the the hydra, repeat pipet, or 12 channel, 53 µl of 100% ethanol, vortex for 30 seconds. Centrifuge, Beckman C56R, at 3200 for 45 min, then invert on a paper towel and centrifuge, Beckman C56R,

up to 400. Immediately, dissolve the DNA in 50 µl of sterile double distilled-water with the hydra, repeat pipette, or 12 channel and vortex for 30 seconds, and centrifuge, Beckman C56R,up and down twice. Let the solution set in the cold room to help dissolve the DNA at least over night. To sequence, use 1 to 5 µl of DNA with Big Dye 3 (1:16) dilution or 1 to 7 µl with Big Dye 3 (1:12) dilution (with 5% DMSO). (Big Dye 3 (1:12) 0.75 µl x 220 = 165, Primer 1.0 µl x 220 = 220, 5 % DMSO 0.125µl x 220 = 27.5, 5x reaction Buffer 0.25 µl x 220 = 66, Total= 467.5/220= 2.125 => 2 µl/well) Preheat DNA 95 °C for 5 minutes, then add the Bigdye 3, then start cycles (60).

P-7: Phi-29 Enzyme lowering "Buffer B" protocol for Plasmid-based DNA amplification and cleaning for Isolated in a 96 well thermocycling plate

In order to make the Phi-29 mix reactions more economical a buffer was developed and used to dilute the enzyme mix and custom synthetic primers were added as well as more dntps. The buffer worked so well, it could lower the enzyme needed to $1/80^{th}$ or (2 µl needed = $1/40^{th}$ pre reaction) the manufacture recommended amount [Normally 5 U/10 µl mix/reaction-(Dean 2001)]. The buffer and the protocol of the initial experiment are listed below.

Phi-29 mixed Buffer B: 50 mM Tris-HCl pH 7.6, 10 mM $MgCl_2$, 10% v/v Glycerol, 1 mM DDT, ~48.8 uM Universal Forward primer, ~48.8 uM Universal Reverse primer, 50 mM dntps (for each dntp). (This same buffer was made with other primers as testing required)

Take one tube and add 200 µl (0.2 g/L) pGem and add 200 µl sterile double distilled water. Heat to 95 °C for 3 minute and cool down to below 30 °C. In the cold room on ice: Take another small tube and add 50 µl phi-29 enzyme mix (25 U) and 150

µl of the above buffer = 0.25 dilution. Take a third small tube and add 25 µl of phi-29 enzyme mix (12.5 U) and 175 µl of the above buffer = 0.125 dilution. Touch all 3 tubes to the vortex machine to mix and add 2 µl of the DNA solution to two 96 well thermocycler plates with a repeat pipette. Add 2 µl of the 0.25 enzyme dilution to one plate and 2 µl of the 0.125 enzyme dilution to the other plate with a repeat pipette. Cover each plate with a foil sealer and centrifuge, Beckman C56R, up and down. Place both plates in a 30 °C water bath for 12 hours. Using the hydra add 16 µl of sterile double distilled water (total volume 20 µl) to each well, then vortex for 30 seconds add 2 µl of 7.5 KOAc to each well with the Hydra or repeat pipette. Knock or centrifuge, Beckman C56R,the KOAc down and vortex for 8 seconds then add with the the hydra, repeat pipette, or 12 channels, 53 µl of 100% ethanol, vortex for 30 seconds. Centrifuge, Beckman C56R, at 3200 for 45 minutes, then invert on a paper towel and centrifuge, Beckman C56R, up to 400. Redilute the wells with 11 µl sterile double distilled water. Put 2 µl of the DNA to each plate with the Hydra and dry down, and then add 4 µl to each plate. Heat DNA for 5 minutes at 95 °C, cool and centrifuge, Beckman C56R, up and down. Add 2 µl of Big dye 3 mix (Forward) to each well, centrifuge, Beckman C56R, up and down, and place the reactions on the thermocycler, big dye 60 cycles, 5 µl volume.

P-8: Phi-29 mix and Phi-29-Only, "Buffer A with out primers" for amplification (for none mix Phi-29 enzyme and for pHi-29 dilution with other vectors: BACs and etc.) and Phi-29-Only reaction protocol.

Phi-29 mixed Buffer A witout primers: 50 mM Tris-HCl pH 7.6, 10 mM MgCl2, 10% v/v Glycerol, 1 mM DDT, 50mM dntps. Next, prepare the Phi-29 enzyme-only reaction: 2 µl of Phi-29 enzyme (250ng) polymerase; 3 µl mix of forward (each primer 5 nM/40 µl); 3 µl mix of reverse primer (each primer 5nM/40 µl); 5 µl of 2mM DNTPs; 5

µl of above Phi-29 Buffer B~30 µl sterile double distilled water. Incubate 12 hours at 30°C.

B-1: Phi-29 for BAC-based DNA amplification and cleaning for Shotgun sequencing from Colony template in a 96 well thermocycling plate

Single BAC colonies are obtained by diluting a smear of BAC colonies from the distribution Petri dish over the range 1:10 through 1:108. This is done by placing a bioloop swipe of the BAC colonies into 1000 µl LB+antibiotic in a 12x75 mm culture tube (VWR # 60818-565) and then adding 100 µl of the 1000 µl 900 µl LB+antibiotic and vortexing 1 sec. (for a 1:10 dilution in tube 1), and then repeating this 7 times to give 1:102 dilution in tube 2, give 1:103 dilution in tube 3, give 1:104 dilution in tube 4, give 1:105 dilution in tube 5. etc.

Plate 100 µl of tubes 5-8 into each of 3-4 plates/dilution, on LB plates containing the correct antibiotic, typically Kan or Cm (not Amp), and incubate 16-18 hrs. Use the Hydra and add 20 µl of sterile double distilled-water to the 384 well thermocycler plate(s), to prevent evaporation, between all steps cover and keep plates on ice. Use the colony picker (not by hand!) to pick colonies into a 384 well thermocycler (the Qpix II robot will not pick into a 96 well thermocycler plate) plate from a shotgun library petri dish or 12x12 square petri plate using the Qpix II robot (~3-5 dips).

Cover plate with thermocycler cover and heat at 95 °C. for 3 minutes. Cover plates with aluminum cover and freeze plates ~2-3 hr. in a -20 freezer until solid. Thaw ~10 min, until completely thawed and centrifuge, Beckman C56R, the picked cell suspension for 45 min. in the Beckman GPR centrifuge, Beckman C56R, to pellet cell debris. While waiting, approx. 10 min. before centrifuge is finished, obtain the correct

number of phi-29 plates from the -80 °C freezer (already made with 2 µl/well) or predispenced enzyme tubes with 500 µl/tube and place in ice water bath to thaw, ~ 5 min., to thaw the enzyme (Note 1: The enzyme is unstable if stored at 4 deg C for hours. Note 2: To prepare the microtiter plate with 2 µl/well, distribute 50 µl of the thawed, bulk, stock phi-29 mix into a "stock" microtiter plate using a repeat pipettor. This plate should be thawed on an ice water bath, takes ~30 minutes with periodic vortexing, and then the 2 µl alaquots are distributed to the microtiter plates using the hydra.) Once the 45 minutes of centrifugation is finished, immediately use the Hydra in the cold room and withdraw 2 µl of the upper portion of the supernatant from each centrifuged cell plate and add it to the enzyme mix plate or place the 2 µl cellular solution into a new 96 well thermocycler plate and add 2 µl of the enzyme mix with a repeat pipette and knock down, cover with thermocycler sealer and vortex gently for 8 seconds. Put DNA + enzyme plate in ice water bath and transport to the centrifuge for a quick 1000 rpm, Beckman C56R centrifuge, up and down spin. Then, place the plate in the thermocycler for 12 hours at 30 °C. with the volume setting of 5 µl. After the 12 hrs is up, heat the thermocycler up to 95 °C for 5 min. Add 16 µl of sterile double distilled water, vortex for 30 seconds and centrifuge, Beckman C56R,up and down. Pool the first half of the plate (A1-12. B1-12, C1-12, and D1-12) into A1-12, pull 3 µl of DNA/well. Then run a 0.6% gel to check DNA. Pool A1-12 in to one eppendorf tube (gives ~800 µl) and the remainder into a second eppendorf.(~ 130-180 ug/plate). Then, renature the DNA by placing the eppendorf tubes in a beaker containing 95 °C water and allowing the water (and thus the solution in the eppendorf) to cool slowly to room temperature (~1 hour). This step will

allow the single stranded DNA produced by the Phi-29 enzyme via the rolling circle replication to renature the DNA into a large, double stranded structure.

B2: Ph-29 Ethanol precipitate (2ml)

To precipitate after, diluting the DNA and heat/cooling to reanneal the DNA: Take the 2mls and redistribute 333 µl into 6 tubes. Next, add 34 µl of 7.5 M potassium acetate and 901 of 100% ethanol to each tube, then vortex well ~30 sec and incubate on ice for 15 minutes. Centrifuge at 12,000 rpm, Beckman C56R centrifuge for 45 minutes in the cold room (4 °C), then, carefully, decant the ethanol-KOAc. Next, add 500 µl of 70% ethanol to each tube, mix by inverting the tube ~8 times, then, centrifuge at 12,000 rpm, Beckman C56R centrifuge for 45 minutes. Decant and dry the tubes in the Speedvac (no heat) for 15 minutes (no longer) and immediately, dissolve the fragmented DNA in the appropriate volume of sterile double distilled water/tube, vortex 30 sec.

Using the standard shotgun sequencing methods go forward starting with the nebulization of the above DNA.

Vol. of Supernatant from cells in 20 µl water	**Water**	**Phi-29 mix**	**DNA produced (ug) = Wells / nebulization (50 ug)**
2 µl	0 µl	2 µl	1.3-1.8 ug = 48
4 µl	6 µl	10 µl	6-9 ug = 10-12
5 µl	15 µl	20 µl	13-18 ug = 5-6

Table 21 Phi-29 BAC Reactions and Yield

(Note: the amount of DNA and water must equal the amount of phi-29 mix.

Appendix V

Bst Methods

(The unabbreviated set of Bst protocols.)

The Klenow fragment of Bacillus stearothermophilus DNA polymerase I (BST) comes from New England Biolabs, the Bst 8000 units, catalog #M0275L, $220.00 and the buffer, catalog# B9004s, $10.00. The buffer comes as a 10x concentration, and it is diluted down to 5X for use in the amplification reactions. (Reaction Buffer at 1X contains: 1X ThermoPol Buffer [20 mM Tris-HCl (pH 8.8, @ 25°C), 10 mM KCl, 10 mM $(NH_4)_2SO_4$, 2 mM $MgSO_4$, 0.1% Triton X-100]). (Kong *et al.* unpublished)

P-1: Bst for plasmid-based DNA amplification, cleaning of the DNA from an isolated template in a 96 well thermocycling plate

In a 96 well thermocycler plate add 2ul of isolated DNA (~1ng), then centrifuge, Beckman C56R, up and down. Cover the plate with thermocycler cover and heat at 95 °C for 3 minutes.

Next, prepare the Bst mix: 1 µl of Bacillus stearothermophilus DNA polymerase large fragment is 8 units, use according to the dilution wanted (range: 0.5 µl = 4units, 0.25 µl = 2 units, 0.125 µl = 1 unit, 0.062 µl = 0.5 units, 0.031 = 0.25 units); 3 µl mix of forward (each primer 5nM/40ul); 3 µl mix of reverse primer (each primer 5nM/40 µl); 5 µl of 2mM DNTPs; ~33µl (based on dilution of enzyme) of 5x Reaction buffer (~3.5 X reaction buffer – total dilution).

Next, add 2 µl of the Bst mix per well with a repeat pipette and cover with thermocycler sealer. Centrifuge for a 1000-rpm, Beckman C56R centrifuge, up and down and place the plate in the thermocycler for 12 hours at 65 °C. with the volume setting of 5 µl. After the 12 hours is finished, heat the thermocycler up to 95 °C for 5 minutes.

Using the hydra, add 16 µl of sterile double distilled water (total volume 20 µl) to each well, then vortex for 30 seconds add 2 µl of 7.5 KOAc to each well with the Hydra or repeat pipette. Knock or centrifuge, Beckman C56R,the KOAc down and vortex for 30 seconds, then add with the Hydra, repeat pipette, or 12 channel, 53 µl of 100% ethanol, vortex for 30 seconds. Centrifuge, Beckman C56R, at 3200 for 45 minutes, and then invert on a paper towel and centrifuge up to 400 rpm, Beckman C56R centrifuge,. Immediately, dissolve the DNA in the appropriate amount of sterile double distilled water.

P-2: Bst for plasmid-based DNA amplification, cleaning from an isolated template in a 384 well thermocycling plate

Use the Hydra and add 20 µl of sterile double distilled water to a 384 well thermocycler plate (viper). To prevent evaporation, between all steps cover and keep plates on ice. Next, add 2 µl of isolated DNA (~1ng), then centrifuge, Beckman C56R, up and down. Cover the plate with thermocycler cover and heat at 95 °C for 3 minutes.

Next, prepare the Bst mix: 1 µl of Bacillus stearothermophilus DNA polymerase large fragment is 8 units, use according to the dilution wanted (range: 0.5µl = 4units, 0.25 µl = 2 units, 0.125 µl = 1 unit, 0.062 µl = 0.5 units, 0.031 = 0.25 units); 3 µl mix of forward (each primer 5nM/40 µl); 3 µl mix of reverse primer (each primer 5nM/40 µl); 5 µl of 2mM DNTPs; ~33 µl (based on dilution of enzyme) of 5x Reaction buffer (~3.5 X reaction buffer – total dilution).

Next, add 2 µl of the Bst mix per well with a repeat pipette and cover with thermocycler sealer. Centrifuge for a 1000 rpm, Beckman C56R centrifuge, up and down and place the plate in the thermocycler for 12 hours at 65 °C with the volume setting of 5 µl. After the 12 hours is finished, heat the thermocycler up to 95 °C for 5 minutes.

Using the hydra add 16 µl of sterile double distilled water (total volume 20µl) to each well, then vortex for 30 seconds and centrifuge, Beckman C56R, up and down. Next, use the Hydra and move the samples from the 384 well thermocycler plate to four 96 well thermocycler plates. To each well of the 96 well thermocycler plates, add 2 µl of 7.5 KOAc to each well with the Hydra or repeat pipette. Knock or centrifuge, Beckman C56R,the KOAc down and vortex for 30 seconds, then add with the Hydra, repeat pipette, or 12 channel, 53 µl of 100% ethanol, vortex for 30 seconds. Centrifuge, Beckman C56R, at 3200 for 45 minutes, and then invert on a paper towel and centrifuge

up to 400 , Beckman C56R centrifuge,. Immediately, dissolve the DNA in the appropriate amount of sterile double distilled water.

P-3: Bst for plasmid-based DNA amplification, cleaning from Colony and Glycerol Stocks template in a 96 well thermocycling plate

Use the Hydra and add 20 µl of sterile double distilled water to a 96 well thermocycler plate. To prevent evaporation, between all steps cover and keep plates on ice.

If using colonies for the template: Use the colony picker (not by hand!) to pick colonies into a 384 well thermocycler (the Qpix II robot will not pick into a 96 well thermocycler plate) plate from a shotgun library petri dish or 12x12 square petri plate using the Qpix II robot (~3-5 dips).

If using the glycerol stocks for template: Use the 12 channel and scrap the well lightly, just enough to get a small amount of ice on the tips of the 12 channel. Place the 12 channel in the well and pump the pipette 8 times.

Cover plate with thermocycler cover and heat at 95 °C. for 3 minutes. Cover plates with aluminum cover and freeze the plates ~2-3 hr. in freezer until solid. Thaw ~10 min, until completely thawed, and centrifuge 45 minutes at 3200 rpm, Beckman C56R centrifuge.

Next, prepare the Bst mix: 1 µl of Bacillus stearothermophilus DNA polymerase large fragment is 8 units, use according to the dilution wanted (range: 0.5 µl = 4units, 0.25 µl = 2 units, 0.125 µl = 1 unit, 0.062µl = 0.5 units, 0.031 = 0.25 units); 3 µl mix of forward (each primer 5nM/40µl); 3 µl mix of reverse primer (each primer 5nM/ 40

µl); 5 µl of 2mM DNTPs; ~33µl (based on dilution of enzyme) of 5x Reaction buffer (~3.5 X reaction buffer – total dilution).

Once the 45 minutes of centrifugation is finished, immediately use the Hydra remove 2 µl of cellular DNA from the bottom of the plate and place in a new 96 well thermocycler plate.

Next, add 2 µl of the Bst mix per well with a repeat pipette and cover with thermocycler sealer. Centrifuge for a 1000 rpm, Beckman C56R centrifuge, up and down and place the plate in the thermocycler for 12 hours at 65 °C with the volume setting of 5 µl. After the 12 hours is finished, heat the thermocycler up to 95 °C for 5 minutes.

Using the hydra, add 16 µl of sterile double distilled water (total volume 20 µl) to each well, then vortex for 30 seconds add 2 µl of 7.5 KOAc to each well with the Hydra or repeat pipette. Knock or centrifuge, Beckman C56R,the KOAc down and vortex for 30 seconds, then add with the Hydra, repeat pipette, or 12 channel, 53 µl of 100% ethanol, vortex for 30 seconds. Centrifuge, Beckman C56R, at 3200 for 45 minutes, and then invert on a paper towel and centrifuge up to 400 rpm, Beckman C56R centrifuge,. Immediately, dissolve the DNA in appropriate amount of sterile double distilled water.

P-4: Bst for plasmid-based DNA amplification, cleaning from Colony and Glycerol stocks template in a 384 well thermocycling plate

Use the Hydra and add 20 µl of sterile double distilled water to a 384 well thermocycler plate (viper). To prevent evaporation, between all steps cover and keep plates on ice.

If using colonies for the template: Use the colony picker (not by hand!) to pick colonies into a 384 well thermocycler plate from a shotgun library petri dish or 12x12 square petri plate using the Qpix II robot (~3-5 dips).

If using the glycerol stocks for template: Use the 12 channel and scrap the well lightly, just enough to get a small amount of ice on the tips of the 12 channel. Place the 12 channel in the well and pump the pipette 8 times.

Cover plate with thermocycler cover and heat at 95 °C for 3 minutes. Cover plates with aluminum cover and freeze the plates ~2-3 hr. in freezer until solid. Thaw ~10 min, until completely thawed, and centrifuge 45 minutes at 3200 rpm, Beckman C56R centrifuge.

Next, prepare the Bst mix: 1 µl of Bacillus stearothermophilus DNA polymerase large fragment is 8 units, use according to the dilution wanted (range: 0.5 µl = 4units, 0.25 µl = 2 units, 0.125 µl = 1 unit, 0.062 µl = 0.5 units, 0.031 = 0.25 units); 3 µl mix of forward (each primer 5nM/40 µl); 3 µl mix of reverse primer (each primer 5nM/40 µl); 5 µl of 2mM DNTPs; ~33 µl (based on dilution of enzyme) of 5x Reaction buffer (~3.5 X reaction buffer – total dilution).

Once the 45 minutes of centrifugation is finished, immediately use the Hydra remove 2 µl of cellular DNA from the bottom of the plate and place in a new 96 well thermocycler plate.

Next, add 2 µl of the Bst mix per well with a repeat pipette and cover with thermocycler sealer. Centrifuge for a 1000-rpm, Beckman C56R centrifuge, up and down

and place the plate in the thermocycler for 12 hours at 65 °C. with the volume setting of 5 µl. After the 12 hours is finished, heat the thermocycler up to 95 °C for 5 minutes.

Using the hydra add 16 µl of sterile double distilled water (total volume 20 µl) to each well, then vortex for 30 seconds and centrifuge, Beckman C56R, up and down. Next, use the Hydra and move the samples from the 384 well thermocycler plate to four 96 well thermocycler plates. To each well of the 96 well thermocycler plates, add 2 µl of 7.5 KOAc to each well with the Hydra or repeat pipette. Knock or centrifuge, Beckman C56R,the KOAc down and vortex for 30 seconds, then add with the Hydra, repeat pipette, or 12 channel, 53 µl of 100% ethanol, vortex for 30 seconds. Centrifuge, Beckman C56R, at 3200 for 45 minutes, and then invert on a paper towel and centrifuge, Beckman C56R, up to 400 rpm, Beckman C56R centrifuge,. Immediately, dissolve the DNA in the appropriate amount of water.

B-1: Bst for BAC-based DNA amplification, cleaning from an isolated template in a 96 well thermocycling plate

In a 96 well thermocycler plate add 2 µl of isolated DNA (~1-10ng), then centrifuge, Beckman C56R, up and down. Cover the plate with thermocycler cover and heat at 95 °C for 3 minutes.

Next, prepare the Bst mix: 1 µl of Bacillus stearothermophilus DNA polymerase large fragment is 8 units, use according to the dilution wanted (range: 0.5 µl = 4units, 0.25 µl = 2 units, 0.125 µl = 1 unit, 0.062 µl = 0.5 units, 0.031 = 0.25 units); 3 µl mix of forward (each primer 5nM/ 40 µl); 3 µl mix of reverse primer (each primer 5nM/40µl); 5 µl of 2mM DNTPs; ~33 µl (based on dilution of enzyme) of 5x Reaction buffer (~3.5 X reaction buffer – total dilution).

Next, add 2 µl of the Bst mix per well with a repeat pipette and cover with thermocycler sealer. Centrifuge for a 1000-rpm, Beckman C56R centrifuge, up and down and place the plate in the thermocycler for 12 hours at 65 °C with the volume setting of 5 µl. After the 12 hours is finished, heat the thermocycler up to 95 °C for 5 minutes.

Using the hydra, add 16µl of sterile double distilled water (total volume 20 µl) to each well, then vortex for 30 seconds add 2 µl of 7.5 KOAc to each well with the Hydra or repeat pipette. Knock or centrifuge, Beckman C56R,the KOAc down and vortex for 30 seconds, then add with the Hydra, repeat pipette, or 12 channel, 53 µl of 100% ethanol, vortex for 30 seconds. Centrifuge, Beckman C56R, at 3200 for 45 minutes, and then invert on a paper towel and centrifuge up to 400 rpm, Beckman C56R centrifuge,. Immediately, dissolve the DNA in the appropriate amount of water.

B-2: Bst for BAC-based DNA amplification, cleaning from an isolated template in a 384 well thermocycling plate

Use the Hydra and add 20 µl of sterile double distilled water to a 384 well thermocycler plate (viper). To prevent evaporation, between all steps cover and keep plates on ice. Next, add 2 µl of isolated DNA (~1-10 ng), then centrifuge, Beckman C56R, up and down. Cover the plate with thermocycler cover and heat at 95 °C. for 3 minutes.

Next, prepare the Bst mix: 1µl of Bacillus stearothermophilus DNA polymerase large fragment is 8 units, use according to the dilution wanted (range: 0.5 µl = 4units, 0.25 µl = 2 units, 0.125 µl = 1 unit, 0.062 µl = 0.5 units, 0.031 = 0.25 units); 3 µl mix of forward (each primer 5nM/40 µl); 3 µl mix of reverse primer (each primer 5nM/40 µl); 5 µl of 2 mM DNTPs; ~33 µl (based on dilution of enzyme) of 5x Reaction buffer (~3.5 X reaction buffer – total dilution).

Next, add 2 µl of the Bst mix per well with a repeat pipette and cover with thermocycler sealer. Centrifuge for a 1000-rpm, Beckman C56R centrifuge, up and down and place the plate in the thermocycler for 12 hours at 65 °C. with the volume setting of 5 µl. After the 12 hours is finished, heat the thermocycler up to 95 °C for 5 minutes.

Using the hydra add 16 µl of sterile double distilled water (total volume 20µl) to each well, then vortex for 30 seconds and centrifuge, Beckman C56R, up and down. Next, use the Hydra and move the samples from the 384 well thermocycler plate to four 96 well thermocycler plates. To each well of the 96 well thermocycler plates, add 2µl of 7.5 KOAc to each well with the Hydra or repeat pipette. Knock or centrifuge, Beckman C56R,the KOAc down and vortex for 30 seconds, then add with the Hydra, repeat pipette, or 12 channel, 53 µl of 100% ethanol, vortex for 30 seconds. Centrifuge, Beckman C56R, at 3200 for 45 minutes, and then invert on a paper towel and centrifuge, Beckman C56R, up to 400 rpm, Beckman C56R centrifuge,. Immediately, dissolve the DNA in the appropriate amount of water.

B-3: Bst for BAC-based DNA amplification and cleaning for Shotgun sequencing from Colony template in a 96 well thermocycling plate

Single BAC colonies are obtained by diluting a smear of BAC colonies from the distribution Petri dish over the range 1:10 through 1:108. This is done by placing a bioloop swipe of the BAC colonies into 1000 µl LB+antibiotic in a 12x75 mm culture tube (VWR # 60818-565) and then adding 100µl of the 1000 µl 900 µl LB+antibiotic and vortexing 1 sec. (for a 1:10 dilution in tube 1), and then repeating this 7 times to give 1:102 dilution in tube 2, give 1:103 dilution in tube 3, give 1:104 dilution in tube 4, give 1:105 dilution in tube 5. etc.

Plate 100 µl of tubes 5-8 into each of 3-4 plates/dilution, on LB plates containing the correct antibiotic, typically Kan or Cm (not Amp), and incubate 16-18 hrs. Use the Hydra and add 20 µl of sterile double distilled-water to the 384 well thermocycler plate(s), to prevent evaporation, between all steps cover and keep plates on ice. Use the colony picker (not by hand!) to pick colonies into a 384 well thermocycler (the Qpix II robot will not pick into a 96 well thermocycler plate) plate from a shotgun library petri dish or 12x12 square petri plate using the Qpix II robot (~3-5 dips).

Cover plate with thermocycler cover and heat at 95 °C for 3 minutes. Cover plates with aluminum cover and freeze plates ~2-3 hr. in a -20 freezer until solid. Thaw ~10 min, until completely thawed and centrifuge the picked cell suspension for 45 min. in the Beckman GPR centrifuge to pellet cell debris.

While waiting make the reaction mix and the enzyme mix (the reactions work better if the enzyme is added directly to the well, the enzyme is particular about being mixed and in large volumes this is hard to do)

Reaction mix contains: 3 µl of forward (primer 200uM), 3 µl of reverse primer (primer 200uM), 5 µl of 2 mM DNTPs, 35 µl 5x Reaction buffer (50 % 10x Buffer and 50% sterile double distilled water) (~3.5 X reaction buffer – total dilution).

Bst mix contains: 13.7 µl of Bst enzyme and 206.3 µl of 5x Reaction buffer. [1 µl of Bacillus stearothermophilus DNA polymerase large fragment is 8 units, use according to the dilution wanted: range per 96 well plate: (add 2 µl per well/per reaction)]

Once the 45 minutes of centrifugation is finished, immediately use the Hydra and remove 2 µl of cellular DNA from the bottom of the plate and place in a new 96 well

thermocycler plate and add 46 µl of the reaction mix to each well. Then add 2µl of Bst mix per well with and a repeat pipette and cover with thermocycler sealer. Centrifuge for a 1000 rpm, Beckman C56R centrifuge, up and down. Next, place the plate in the thermocycler for 18 hours at 65 °C with the volume setting of 50µl. After the allotted is finished, then heat the thermocycler up to 95 °C for 5 minutes to denature the enzyme and cool to 4 °C before removing to prevent evaporation. Let the DNA set in the cold room for at least 24-48 hours for the DNA to dissolve.

To clean the Bst reactions, take the 50 µl reactions and add 50 µl of sterile double distilled water to each. Vortex the plate for 30 seconds and centrifuge for a 1000 rpm, Beckman C56R centrifuge, up and down. Next, remove the residual water from the columns of the prepared the Sephadex G-50 millipore filter plates via centrifugation, then add 30 µl to the top of each Sephadex column in a row from each sample with a 12 channel pipette. Tape the Sephadex plate on top of a clean 96 well thermocycler plate and centrifuge, Beckman C56R, at 1300-1400 for 2.5 minutes. Pool the sample wells to one well per sample.

To sequence the Bst reactions do a (1:10) dilution (2 of DNA /20 sterile double distilled water) of the DNA. The sequencing reaction contains: 2 µl DNA from the (1:10) dilution, 2 µl sterile double distilled water, 2 µl primer 200 µM, and 2 µl ET dye. Thermocycling conditions: Hold-95 °C for 2min; Cycle-95 °C for 30 sec, 50 °C for 20 sec, 60 °C for 4 minutes; Hold at 4 °C for infinity. Clean the reactions using the standard sodium acetate ethanol clean up.

B-4: Bst for BAC-based DNA amplification and cleaning for Shotgun sequencing from Glycerol Stock template in a 96 well thermocycling plate

Use the Hydra and add 20 µl of sterile double distilled-water to the 96 well thermocycler plate(s), to prevent evaporation, between all steps cover and keep plates on ice. Use the 12 channel and scrap the Glycerol well lightly, just enough to get a small amount of ice on the tips of the 12 channel, then dip the tips into the water and pump up/down 8 times.

Cover plate with thermocycler cover and heat at 95 °C for 3 minutes. Cover plates with aluminum cover and freeze plates ~2-3 hr. in a -20 freezer until solid. Thaw ~10 min, until completely thawed and centrifuge the picked cell suspension for 45 min. in the Beckman GPR centrifuge to pellet cell debris.

While waiting make the reaction mix and the enzyme mix (the reactions work better if the enzyme is added directly to the well, the enzyme is particular about being mixed and in large volumes this is hard to do)

Reaction mix contains: 3 µl of forward (primer 20µM), 3 µl of reverse primer (primer 200µM), 5ul of 2 mM DNTPs, 35 µl 5x Reaction buffer (50 % 10x Buffer and 50% sterile double distilled water) (~3.5 X reaction buffer – total dilution).

Bst mix contains: 13.7 µl of Bst enzyme and 206.3 µl of 5x Reaction buffer. [1 µl of Bacillus stearothermophilus DNA polymerase large fragment is 8 units, use according to the dilution wanted: range per 96 well plate: (add 2 µl per well/ reaction)]

Once the 45 minutes of centrifugation is finished, immediately use the Hydra and remove 2 µl of cellular DNA from the bottom of the plate and place in a new 96 well

thermocycler plate. Next, add 46 µl of the reaction mix to each well. Then add 2 µl of Bst mix per well with and a repeat pipette and cover with thermocycler sealer. Centrifuge for a 1000 rpm, Beckman C56R centrifuge, up and down. Next, place the plate in the thermocycler for 18 hours at 65 °C with the volume setting of 50 µl. After the allotted is finished, then heat the thermocycler up to 95 °C for 5 minutes to denature the enzyme and cool to 4 °C before removing to prevent evaporation.

To clean the Bst reactions, add ~45 µl of sterile double distilled water (½ dilution, you will have a ~5 µl evaporation from sitting in the cold room). Vortex the plate for 30 seconds and centrifuge for a 1000 rpm, Beckman C56R centrifuge, up and down. Next, remove the residual water from the columns of the prepared the Sephadex G-50 millipore filter plates via centrifugation, then add 30 µl to the top of each Sephadex column per plate from each sample with a 12 channel pipette (3 plates/96 reactions). Tape the Sephadex plate on top of a clean 96 well thermocycler plate and centrifuge, Beckman C56R, at 1300-1400 for 2.5 minutes. Pool the sample wells (3 different plates) to one well per sample/ plate.

To sequence the Bst reactions, after cleaning the DNA through the Sephadex, use the Hydra, and pull 5 µl of DNA and add to another clean 96 well thermocycler plate and add 45 µl of sterile double distilled water (1/10 dilution) to the plate with 5ul of DNA. Next, mix reactions by pumping the DNA solution up/down with a 12 channel ~8 times.

The sequencing reaction contains: 4 µl DNA from the (1:10) dilution, 2 µl sterile double distilled water, 2 µl primer 100 µM, and 4 µl ET dye. Thermocycling conditions: Hold-95 °C for 2min; Cycle-95 °C for 30 sec, 50 °C for 20 sec, 60 °C for 4 minutes; Hold

at 4 °C for infinity. Clean the reactions using the standard sodium acetate ethanol clean up.

B-5: Bst Testing protocol for BAC-based DNA amplification and cleaning for Shotgun sequencing from Colony template in a 96 well thermocycling plate

Single BAC colonies are obtained by diluting a smear of BAC colonies from the distribution Petri dish over the range 1:10 through 1:108. This is done by placing a bioloop swipe of the BAC colonies into 1000 µl LB+antibiotic in a 12x75 mm culture tube (VWR # 60818-565) and then adding 100 µl of the 1000 µl 900 µl LB+antibiotic and vortexing 1 sec. (for a 1:10 dilution in tube 1), and then repeating this 7 times to give 1:102 dilution in tube 2, give 1:103 dilution in tube 3, give 1:104 dilution in tube 4, give 1:105 dilution in tube 5. etc.

Plate 100 µl of tubes 5-8 into each of 3-4 plates/dilution, on LB plates containing the correct antibiotic, typically Kan or Cm (not Amp), and incubate 16-18 hrs. Use the Hydra and add 20 µl of sterile double distilled-water to the 384 well thermocycler plate(s), to prevent evaporation, between all steps cover and keep plates on ice. Use the colony picker (not by hand!) to pick colonies into a 384 well thermocycler (the Qpix II robot will not pick into a 96 well thermocycler plate) plate from a shotgun library petri dish or 12x12 square petri plate using the Qpix II robot (~3-5 dips).

Cover plate with thermocycler cover and heat at 95 °C for 3 minutes. Cover plates with aluminum cover and freeze plates ~2-3 hr. in a -20 freezer until solid. Thaw ~10 min, until completely thawed and centrifuge the picked cell suspension for 45 min. in the Beckman GPR centrifuge to pellet cell debris.

While waiting make the reaction mix and the enzyme mix (the reactions work better if the enzyme is added directly to the well, the enzyme is particular about being mixed and in large volumes this is hard to do)

Reaction mix contains: 3 µl of forward (primer 100-200 µM), 3 µl of reverse primer (primer 100-200µM), 5 µl of 2 mM DNTPs, 35 µl 5x Reaction buffer (50 % 10x Buffer and 50% sterile double distilled water) (~3.5 X reaction buffer – total dilution). Or Reaction mix contains: 3ul of primer mix (~24 pairs) forward primer (5nM/40µl), 3 µl of primer mix (~24 pairs) reverse primer (5nM/40 µl), (5µl of 2mM DNTPs, 35µl 5x Reaction buffer (50 % 10x Buffer and 50% sterile double distilled water) (~3.5 X reaction buffer – total dilution).

Bst mix: 1 µl of Bacillus stearothermophilus DNA polymerase large fragment is 8 units, use according to the dilution wanted: range per 96 well plate: (add 2 µl per well/per reaction)]

Units of Bst pol /Reaction	Volume of Bst (µl)	Volume of 5X Buffer (µl)	Total Vol. (µl)
8	110	110	220
4	55	165	220
2	27.5	192.5	220
1	13.7	206.3	220
0.5	6.8	213.2	220
0.25	3.4	216.6	220

Table 22 Bst Reaction Unit and Buffer Vol.

(Currently, the amount used is 1 unit per reaction to get whole BAC amplification.)

Once the 45 minutes of centrifugation is finished, immediately use the Hydra and remove 2 µl of cellular DNA from the bottom of the plate and place in a new 96 well thermocycler plate and add 46 µl of the reaction mix to each well. Then add 2 µl of Bst mix per well with and a repeat pipette and cover with thermocycler sealer. Centrifuge for a 1000 rpm, Beckman C56R centrifuge, up and down. Next, place the plate in the thermocycler for 3, 6, 12, 18, 24, 30 hours (depending on the amount of DNA you want) at 65 °C. with the volume setting of 50 µl. After the allotted is finished, (for further extension allow the temperature to drop to 55 °C. for 2 hours) then heat the thermocycler up to 72 °C. for 10 minutes or 95 °C to denature the enzyme and cool to 4 °C before remove to prevent evaporation. Let the DNA set in the cold room for at least 24-48 hours for the DNA to dissolve.

To clean the Bst reactions, take the 50 µl reactions and add 50 µl of sterile double distilled water to each. Vortex the plate for 30 seconds and centrifuge for a 1000 rpm, Beckman C56R centrifuge, up and down. Next, remove the residual water from the columns of the prepared the Sephadex G-50 millipore filter plates via centrifugation, then add 30 µl to the top of each Sephadex column in a row from each sample with a 12 channel pipette. Tape the Sephadex plate on top of a clean 96 well thermocycler plate and centrifuge, Beckman C56R, at 1300-1400 for 2.5 minutes. Pool the sample wells to one well per sample.

To sequence the Bst reactions do a (1:10) dilution (2 of DNA/20 sterile double distilled water) of the DNA. The sequencing reaction contains: 2 µl DNA from the (1:10) dilution, 2 µl sterile double distilled water, 2 µl primer 200 µM, and 2 µl ET dye.

Thermocycling conditions: Hold-95 °C for 2min; Cycle-95 °C for 30 sec, 50 °C for 20 sec, 60 °C for 4 minutes; Hold at 4 °C for infinity. Clean the reactions using the standard sodium acetate ethanol clean up.

B-6: Bst Testing CONTROL protocol for BAC-based DNA amplification and cleaning for Shotgun sequencing from isolated template in a 96 well thermocycling plate

When experimentally working with Bst enzyme to confirm amplification, the volume of the reaction is 4 µl, the partial extraction would complicate the amount of dilution to be used later also after cleaning and diluting the DNA one can not see the DNA on a gel, If one could see the DNA, it would be too concentrated to sequence. Therefore, this control should be used.

Use Isolated DNA, (diluted) BAC, pUC and/or Fosmid DNA. Based on what is easier place test DNA in a few wells of a 96 well thermocycler plate or in a tube or tubes add 20 µl of sterile double distilled water To prevent evaporation, between all steps cover and keep on ice. Add 2 µl of isolated DNA (~10 ng), then centrifuge, Beckman C56R, up and down. Next, cover plate with thermocycler sealer or snap top closed on tube and heat at 95 °C for 3 minutes.

Make a reaction pot using primers (vector specific) made for each type of DNA (example: BAC primers for BAC DNA) Reaction mix contains: 3 µl of primer mix (~24 pairs) forward primer (5nM/40µl), 3 l of primer mix (~24 pairs) reverse primer (5nM/40µl), (5µl of 2mM DNTPs, 35 µl 5x Reaction buffer (50 % 10x Buffer and 50% sterile double distilled water) (~3.5 X reaction buffer – total dilution).

Next, add 2ul of the Bst (16 units) per well with a repeat pipette and cover with thermocycler sealer. Centrifuge for a 1000-rpm, Beckman C56R centrifuge, up and down. Then, place the plate in the thermocycler for X number of hours [depending on the amount of DNA you want, (for first time use, I recommend 12 to 18 hours to see what the amplified products look like and how well they amplify, then drop the time)] at 65 °C. with the volume setting of 50 µl. After the set hours is finished, and then heat the thermocycler up to 95 oC for 5 minutes to denature the enzyme and cool to 4 oC before removal to prevent evaporation

B-7: Salt and Heat Shearing of BAC and Genomic DNA

Salt protects DNA against depurination, by stabilizing the helix, and by preventing autocatalysis, DNA is a polyphosphoric acid. (Marguet and Forterre, 1998)

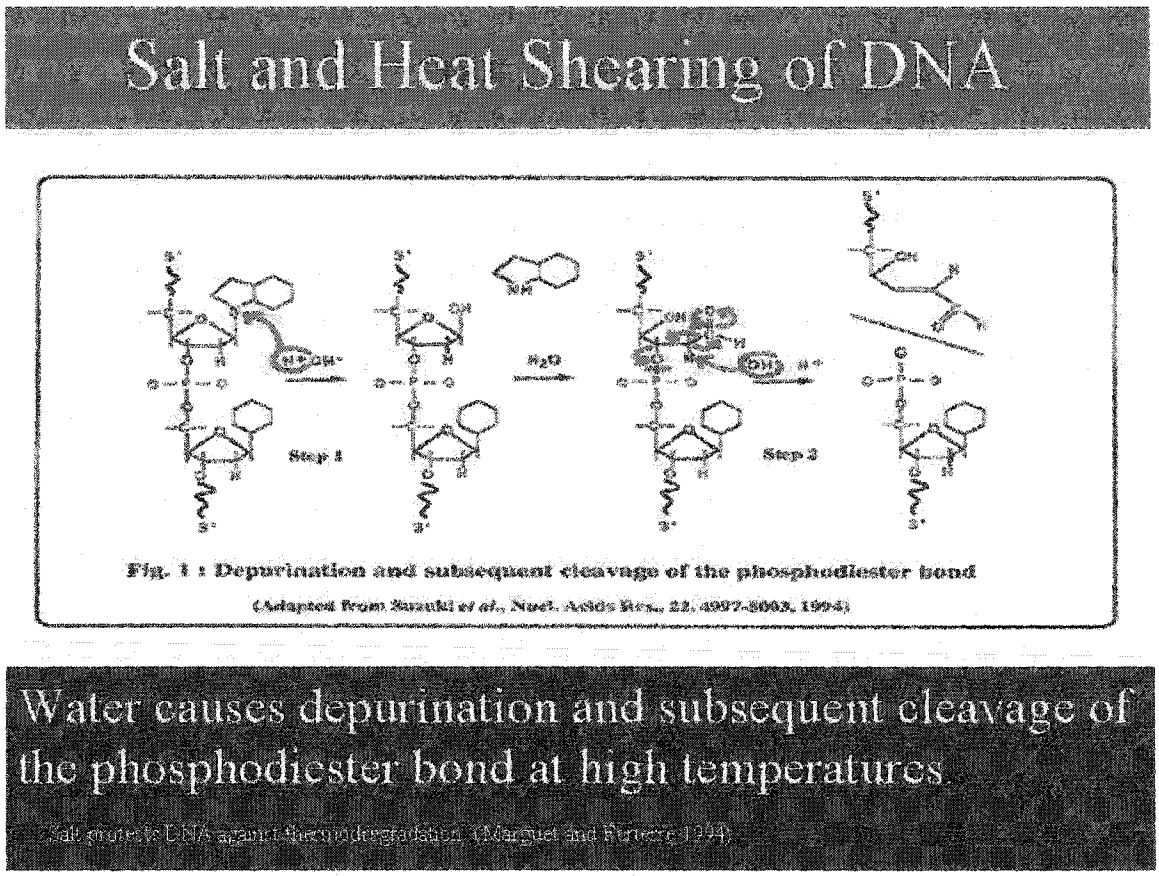


Figure 67 Salt and Heat Shearing of DNA

(http://www-archbac.u-psud.fr/Meetings/athensgr97/AthenGr97Poster.html)

Tale you amplified DNA and mix it with 0.1 M NaCl to make a 50% dilution, then heat the solution to 90 °C for 45 minutes to shear, and cool to 4 °C.

B-8: Shearing Genomic DNA and BAC DNA via Liquid Nitrogen Freezing

This application is the standard freeze and thaw technique for breakage. After amplifying the DNA, it is necessary to do the 95 °C denature for 5 minutes, because it assists or allows the breakage of DNA. After heating the samples and cooling to 4 °C, they should be transferred to 250 µl eppendorf tubes. Liquid nitrogen is then added to the tubes in a styrophoam cup ~3 ml just enough to cover the tubes; they should immediately freeze, then place the tubes in the -80 °C freezer until ready to continue the experiment.

When ready, place the tubes in an ice water bath in the cold room for ~1 hour to thaw slowly. This seems to perpetuate the single strands in finding there mate as well as the DNA breakage.