**Training materials for seminars**

**Methods on molecular biotechnology**

**Seminar 1.** Main approaches and methods of molecular biotechnology

Whole blood samples are one of the main sources used to obtain DNA, and there are many different protocols available to perform nucleic acid extraction on such samples. These methods vary from very basic manual protocols to more sophisticated methods included in automated DNA extraction protocols.

Solution-based DNA extraction methods using organic solvents

Solution-based DNA extraction methods using salting out

Solid-phase DNA extraction methods

DNA extraction methods using silica and silica matrices

DNA extraction using anion exchange resins

DNA extraction methods using magnetic beads

Main types of DNA extraction methods from human whole blood samples

**Seminar 2** Methods of extraction of nucleic acids from different biological materials

DNA extraction protocols using organic solvents derived originally from a series of related RNA extraction methods. Some of the main steps used in these methods are:

1) cell lysis undertaken by adding a detergent/chaotropic-containing solution, including SDS or N-Lauroyl sarcosine;

2) inactivation of DNases and RNases, usually through the use of organic solvents;

3) purification of DNA and removal of RNA, lipids, and proteins;

4) resuspension of extracted nucleic acids.

**DNA extraction methods using silica and silica matrices**

Silica matrices have unique properties for DNA binding. They are positively charged and have high affinity toward the negative charge of the DNA backbone. High salt conditions and pH are achieved using sodium cations, which bind tightly to the negatively charged oxygen in the phosphate backbone of DNA. Contaminants are removed with a series of washing steps, followed by DNA elution under low ionic strength (pH ≥7) using TE buffer or sterile distilled water.A substance that contains high amounts of silica (up to 94%) known as kieselguhr, diatomite, or diatomaceous earth has also been used for DNA purification.

The traditional protocol involves initial cell disruption and digestion with SDS–proteinase K, followed by the addition of high concentrations of salts, usually 6 M sodium chloride. The mixture is then centrifuged to allow proteins to precipitate to the bottom, with the supernatant containing DNA then transferred to a new vial. DNA is then precipitated using ethanol or isopropanol in the same manner as described for organic solvent methods.

**DNA extraction using anion exchange resins**

Positively charged chemical substances able to bind to negatively charged nucleic acids or contaminants or enzymes, such as nucleases, are called anion exchange resins, and they have also been used as part of DNA extraction protocols from blood samples.Blood samples could be lysed using proteinase K and/or incubation at high temperature, and removal of contaminants is achieved by adding Chelex® 100 resin, which precipitates them. Single-stranded DNA is obtained and remains suspended in the supernatant, which can be immediately used in downstream application or can be transferred to a new vial for long-term storage.

*Reference:*

1.Tan SC, Yiap BC. DNA, RNA, and protein extraction: the past and the present. J Biomed Biotechnol. 2009;2009:574398.

2.Price CW, Leslie DC, Landers JP. Nucleic acid extraction techniques and application to the microchip. Lab Chip. 2009;9(17):2484–2494.

3.Arévalo-Herrera M, Soto L, Perlaza BL, et al. Antibody-mediated and cellular immune responses induced in naive volunteers by vaccination with long synthetic peptides derived from the plasmodium vivax circumsporozoite protein. Am J Trop Med Hyg. 2011;84(Suppl 2):35–42.

**Seminar 3** Analysis and Characterization of nucleic acids

***Detection Mechanisms.***

Detection of a positive allelic discrimination reaction is done by monitoring the light emitted by the products, measuring the mass of the products, or detecting a change in the electrical property when the products are formed. Numerous labels with various light-emitting properties have been utilized in detection methods based on light detection. In general, only one label with ordinary properties is needed in genotyping methods where the products are separated or purified from the excess starting reagents. For homogeneous reactions, where no separation or purification is needed, the property of the label has to be changed when a product is formed. This usually requires interaction of the label with another component of the reaction when a product is formed. A number of elegant genotyping methods are developed to take advantage of certain physical characteristics of the labels. Monitoring light emission is the most widely used detection modality in genotyping and there are many ways to do so. Luminescence, fluorescence, time-resolved fluorescence, fluorescence resonance energy transfer (FRET), and fluorescence polarization (FP) are useful properties of light utilized in a host of genotyping methods. Luminescence Detection Luminescence is emitted in an ATP-dependent luciferase reaction. When ATP production is coupled with a primer extension reaction, luminescence is observed every time a deoxyribosenucleoside is added in the primer extension reaction. Because the background is extremely low, luminescence has a very good signal to noise ratio. However, the additional enzymatic steps and substrates required complicate the experimental procedure and increase the cost of the assay. Fluorescence Detection Fluorescence detection is straightforward and easy to implement. Besides using it in a setting of capturing fluorescent labels on a solid support or separating the fluorescent product from the label by gel or capillary electrophoresis, fluorescence detection can be used to monitor the formation of double stranded DNA with a DNA intercalating dye that only fluoresces in the presence of double stranded DNA. Direct fluorescence detection is very versatile and can be done in multiplex to a certain extent. However, the need for product purification or separation when fluorescent labels are used and the interference by non-specific double stranded DNA species when intercalating dyes are used are some of the drawbacks of direct fluorescence detection. Time-Resolved Fluorescence Detection Time-resolved fluorescence as a detection approach is feasible when the emission half-life of the fluorescent dye is long. With this class of dyes (mostly compounds of rare earth elements such as Lanthanides), the fluorescence reading is done sufficiently long after excitation such that autofluorescence (which has a very short half-life) is not observed (Hansen et al., 1995; Kirschstein et al., 1999). The background in time-resolved fluorescence detection is almost non-existent so that this is a very sensitive detection modality. The drawback is that the lanthanides are inorganic compounds that cannot be used to label nucleic acids directly. An organic chelator conjugated to the probe must be used to bind the lanthanides in the reaction. Fluorescence Resonance Energy Transfer Fluorescence resonance energy transfer is a popular detection method in homogeneous genotyping assays. FRET occurs when two conditions are met. First, the emission spectrum of the fluorescent donor dye must overlap with the excitation wavelength of the acceptor dye. Second, the two dyes must be in close proximity to each other because energy transfer drops off quickly with distance. The proximity requirement is what makes FRET a good detection method for a number of allelic discrimination mechanisms. Basically, any reaction that brings together or separates two dyes can use FRET as the detection method. FRET detection has therefore been used in primer extension and ligation reactions where the two labels are brought into close proximity to each other. It has also been used in the 5'-nuclease reaction, the molecular beacon reaction, and the invasive cleavage reactions where the neighboring donor/acceptor pair is separated by cleavage or disruption of the stem-loop structure that holds them together (Livak, 1999; Kostrikis et al., 1998; Tyagi et al., 1998; Hall et al., 2000). The major drawback of the method is the cost of the labeled probes required in all the genotyping approaches with FRET detection. In the cleavage approaches, the probes are doubly labeled, increasing the cost of the probe synthesis even further. Fluorescence Polarization Fluorescence polarization (FP) has been used in clinical diagnosis and numerous binding assays for years but its use as a detection method for SNP genotyping has a very short history. This is because instruments sensitive enough for detecting small amounts of dyes are not available until recently. When a dye is excited by plane polarized light, the emitted fluorescence is also polarized. The degree of polarization is determined by temperature, viscosity of the solvent, and the molecular volume of the fluorescent molecule. All these factors affect molecular motion and, in general, the faster a molecule tumbles and rotates in solution, the less polarized is its fluorescence. Because molecular volume is proportional to molecular weight, fluorescence polarization is therefore a good method to detect changes molecular weight. In principle, any genotyping method in which the product of the allelic discrimination reaction is substantial larger or smaller than the starting fluorescent molecule can use FP as a detection method. Indeed, FP has been used as the detection method in the primer extension reaction where small fluorescent dye terminators are incorporated into a larger probe (Chen et al., 1999). Furthermore, FP has been shown recently that it is a good detection method in the 5'-nuclease reaction where small fluorescent molecules are formed when large fluorescent probes are cleaved in the reaction (Latif et al., 2001). FP can also be used as a detection method for the invasive cleavage reaction where the large fluorescent signal probe is cleaved, producing a small fluorescent tag (Hsu et al., 2001b). The advantages of the FP detection method include the much smaller amount of fluorescent dyes needed compared to FRET or direct fluorescence detection methods, cheaper probes used, and the potential for utilizing the full visible spectrum in multiplex reactions. The drawback is mainly that any non-specific products will increase the noise in the signal.

***Mass Spectrometry.***

Unlike all other detection methods that infer the identity of the products generated in the allelic discrimination reaction by monitoring the fate of some label, mass spectrometry (MS) measures the molecular weight of the products formed and is therefore the most direct method of detection. Because MS determines the fundamental property of the DNA molecule, no labels are needed. High resolution MS can easily distinguish between DNA molecules that differ by only one base (Ross et al., 1998; Liet al., 1999; Berlin and But, 1999; Buetow et al., 2001). A further advantage of MS is that it takes only milliseconds to analyze each sample so even though MS analyzes each sample serially, the throughput is still very high. Furthermore, by appropriately designing the probes, moderate multiplexing is possible (Ross et al., 1998). The main disadvantage of the MS detection method is the exquisite purity the analyte has to be for it to work. With further refinement of the product purification process, it may be possible to overcome this drawback. Electrical Detection A promising detection method is one that monitors a change in the electrical properties of the products of the allelic discrimination reaction. Currently, this is done on solid support where oligonucleotides are deposited on electrodes (Cornell et al., 1997; Wang et al., 1997). The electrical property of the probe is altered when the DNA complementary to the probe is annealed to it. This is exaggerated if a ferromagnetic label is used. Electrical detection combines semiconductor technology with biochemistry and eliminates the need for light detection or extensive product processing. This area is still in its infancy and there are still a number of biochemical and engineering obstacles to overcome before the throughput of genotyping methods based on this detection mechanism is high enough and the cost low enough for its wide acceptance.

**Seminar 4.** MicroRNA Cloning from Cells of the Immune System. Use of nucleases, ligase, exonuclease, restrictase in molecular biotechnology

***Ligation***

DNA ligase is highly specific in repairing nicks in the DNA molecule. When two adjacent oligonucleotides are annealed to a DNA template, they are ligated together only if the oligonucleotides perfectly match the template at the junction. Allele-specific oligonucleotides can therefore be made to interrogate the nature of the base at the polymorphic site. One can infer the allele(s) present in the target DNA by determining whether ligation has occurred or not. While ligation has the highest level of specificity and is easiest to optimize among all allelic discrimination mechanisms, it is the slowest reaction and requires the largest number of modified probes. However, ligation as a mechanism has the potential of genotyping without need for prior target amplification by PCR. This can be accomplished by either the ligation chain reaction (LCR) (Barany, 1991) or by the use of ligation (padlock) probes that circularize by DNA ligase followed by rolling circle signal amplification (Lizardi et al., 1998; Baner et al., 1998).

***Reaction Formats***

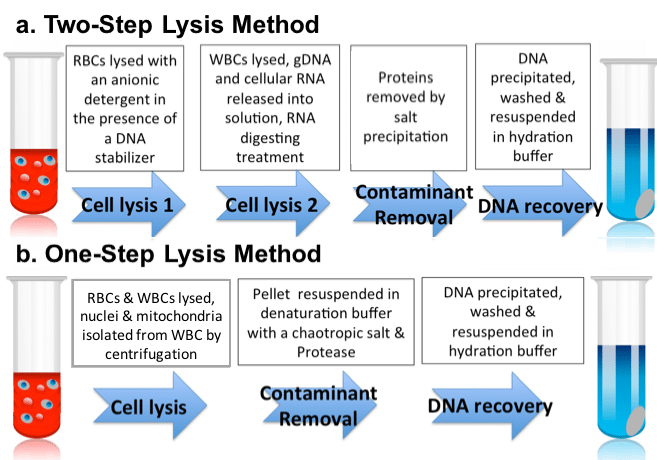
Starting with genomic DNA, each genotyping method undergoes a series of biochemical steps and a product detection step. The reaction format mostly reflects the requirements of the detection modality. In general, biochemical reactions are more robust in solution but capturing the reaction products on solid support allows for detection in parallel and increases the throughput substantially.

* **Seminar 5.** Analysis of different types nuclear acids.

***Homogeneous Reactions.***

A number of innovative genotyping methods are done in solution from beginning to end and are therefore designated as homogeneous reactions. Some of them require no further manipulations once the reaction is set up initially. Others call for a number of reagent addition steps but no separation or purification steps are needed. Homogeneous assays are usually robust, highly flexible and not labor intensive. The major drawback is the limited amount of multiplexing one can do with homogeneous assays. Reactions on Solid Support Solid supports used in genotyping can be a latex bead, a glass slide, a silicon chip, or just the walls of a microtiter well. In some cases, marker specific oligonucleotides are placed on the solid support and the allelic discrimination reaction is done on the support; in other cases, generic oligonucleotides are placed on the solid support and they are used to capture complementary sequence tags conjugated to marker specific probes. In the former strategy, the oligonucleotide arrays act as a collection of reactors where the target DNA molecules find their counterparts and the allelic discrimination step for numerous markers proceeds in parallel. In the latter, the arrayed oligonucleotides are used to sort the products of the allelic discrimination reactions (also done in parallel) performed in aqueous solution. In both cases, the identity of an oligonucleotide on a latex bead or at a particular location on the microarray (on a glass slide or silicon chip) is known and the genotypes are inferred by determining which immobilized oligonucleotide is associated with a positive signal. The major advantage of performing genotyping reactions on solid supports is that many markers can be interrogated at the same time. Besides savings in time and reagents, performing numerous reactions in parallel also decreases the probability of sample/results mix-ups. The drawback of performing genotyping reactions on solid support is that design of the oligonucleotide arrays and optimization of the multiplex reactions require substantial capital and time investment. With better algorithms for multiplex PCR design, this limitation may be alleviated in the near future.

* **Seminar 6.** DNA Separation Techniques for different types of DNA

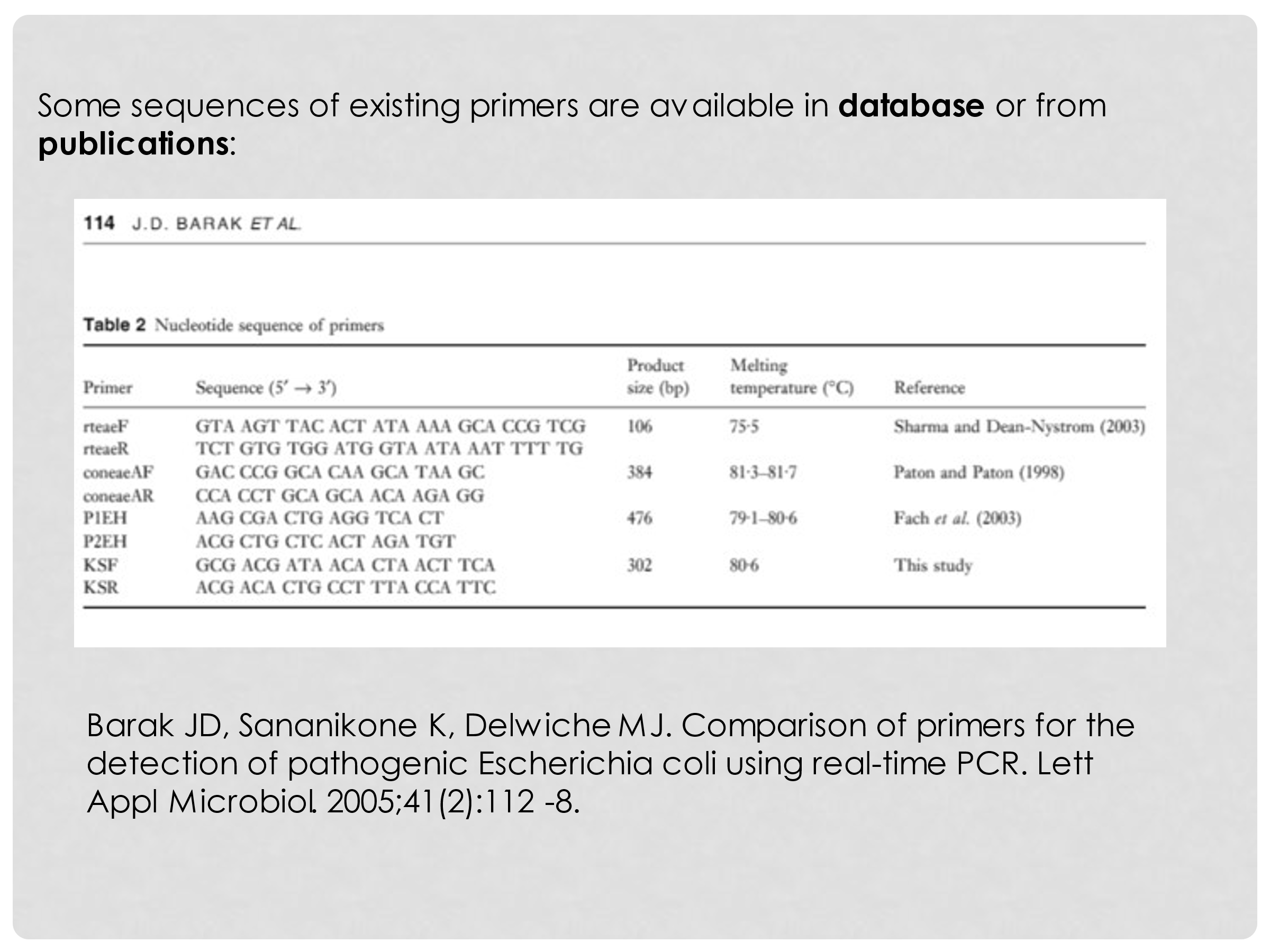
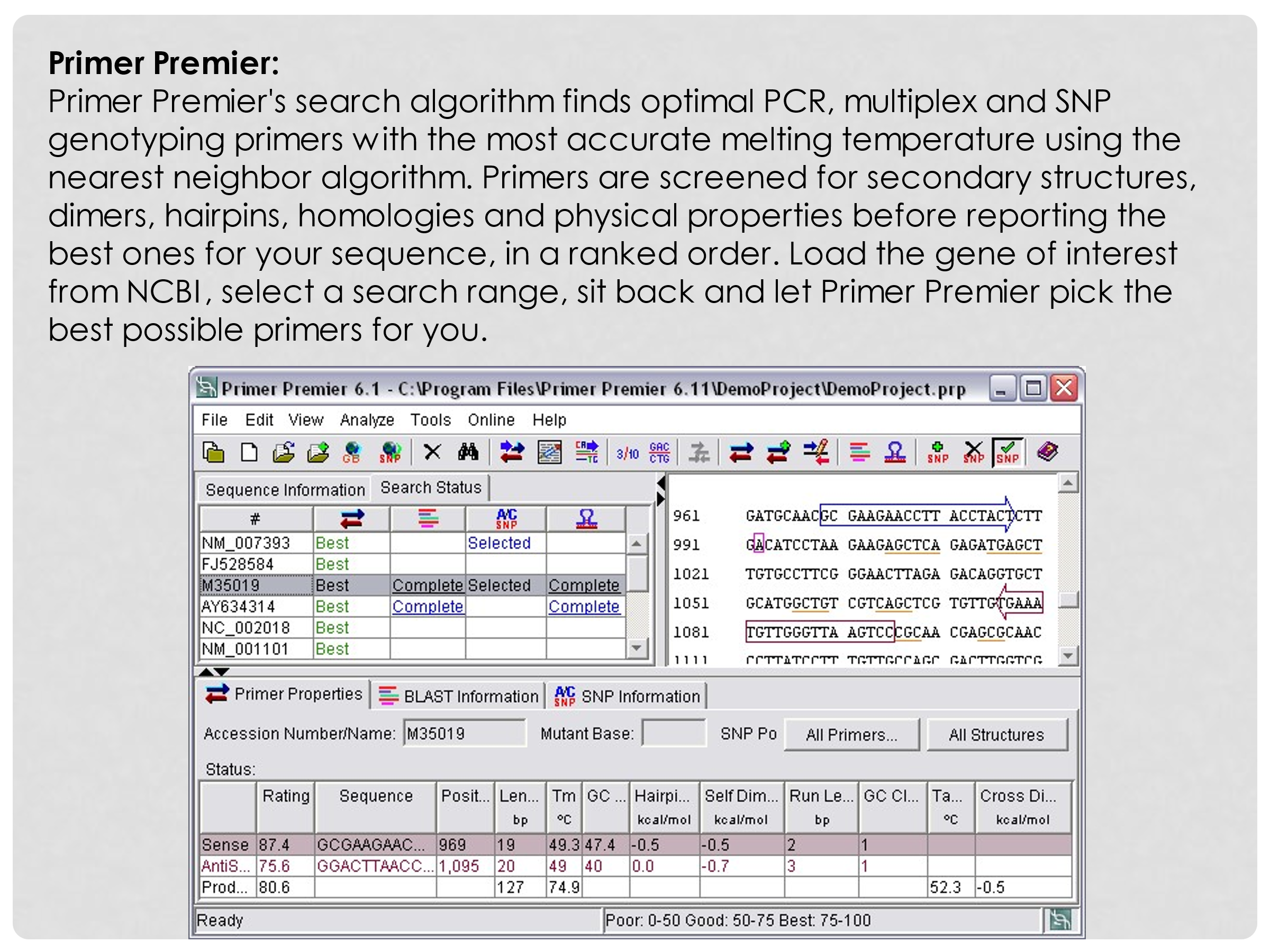


*Technologies For Genotyping Known SNPs*

The ideal genotyping method must be easily and reliably developed from sequence information, robust, inexpensive, flexible, automated, and produces data that are easily analyzed. Although no such ideal genotyping method exists, a number of promising SNP genotyping methods are currently available and further improvements in biochemistry, engineering, and analytical software will bring the existing methods closer to the ideal. The evolution of SNP genotyping technologies is characterized by improvements in the three aspects of the genotyping process: allelic discrimination, reaction format, and signal detection.

***Hybridization.***

With the hybridization approach, two allele-specific probes are designed to hybridize to the target sequence only when they match perfectly. Under optimized assay conditions, even one-base mismatches destabilize the hybridization sufficiently to prevent the allelic probe from annealing to the target sequence. Because no enzymes are involved in allelic discrimination, hybridization is the simplest mechanism for genotyping. The challenge lies in designing allele-specific probes that can distinguish one-base mismatches. With ever more sophisticated probe design algorithms and the use of hybridization enhancing moieties such as DNA minor groove binders, allele-specific probes can be designed with high success rate. When the allele-specific probes are immobilized on a solid support, labeled target DNA samples are captured and the hybridization event is visualized by detecting the label after the unbound targets are washed away. Knowing the location of the probe sequences on the solid support allows one to infer the genotype of the target DNA sample (Wang et al., 1998). Allele-specific hybridization is also the basis of several elegant homogeneous genotyping assays. These assays differ in the way they report the hybridization event. In the 5'-nuclease assay, a probe annealed to target DNA being amplified is cleaved during PCR. Monitoring the cleavage event is therefore a way to determine whether hybridization has occurred (Livak, 1999). With molecular beacon detection, hybridization to target DNA opens up the stem-loop structure. Determining the open-closed status of the stem-loop structure is therefore a way to figure out if hybridization has occurred (Kostrikis et al., 1998; Tyagi et al., 1998). With “light-up” probes, the thiazole orange derivative linked to a peptide nucleic acid (PNA) oligomer fluoresces only when the PNA oligomer hybridizes specifically to complementary nucleic acids (Svanvik et al., 2000). Fluorescence is therefore evidence for hybridization.

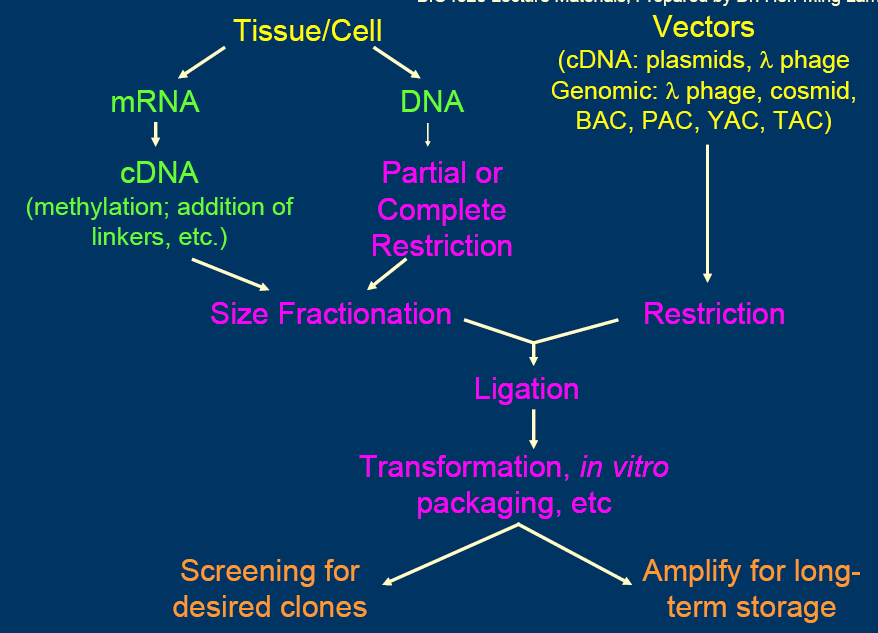
* **Antigen**: foreign molecules that generate antibodies or any substance that can be bound specifically by an antibody molecule
* **Proteins**, sugars, lipids or nucleic acids
* Natural or synthetic
* **Antibody**: molecules (protein) responsible for specific recognition and elimination (neutralization) of antigens
* Different structures (7-8 classes in mammals)
* Powefull research tools for basic research, clinical applications and drug design
* **Seminar 11.** Strategies for SNP detections strategies for arrays.
* Single nucleotide polymorphism (SNP) detection technologies are used to scan for new polymorphisms and to determine the allele(s) of a known polymorphism in target sequences. SNP detection technologies have evolved from labor intensive, time consuming, and expensive processes to some of the most highly automated, efficient, and relatively inexpensive methods. Driven by the Human Genome Project, these technologies are now maturing and robust strategies are found in both SNP discovery and genotyping areas. The nearly completed human genome sequence provides the reference against which all other sequencing data can be compared. Global SNP discovery is therefore only limited by the amount of funding available for the activity. Local, target, SNP discovery relies mostly on direct DNA sequencing or on denaturing high performance liquid chromatography (dHPLC). The number of SNP genotyping methods has exploded in recent years and many robust methods are currently available. The demand for SNP genotyping is great, however, and no one method is able to meet the needs of all studies using SNPs. Despite the considerable gains over the last decade, new approaches must be developed to lower the cost and increase the speed of SNP detection.
* Technologies For Global SNP Discovery The main issue with global SNP discovery is that on average, there is one SNP in every 1,000 bp of DNA when two human genomes are compared to each other (The International SNP Map Working Group, 2001). To maximize the chance of finding SNPs, one must be able to scan 1,000 bp pieces of DNA in a generic way. The first attempt to identify SNPs randomly in the human genome was to scan for alterations in restriction sites in the genome (Botstein et al., 1980). Although the actual sequence variation was not determined, restriction fragment length polymorphisms (RFLPs) were the first SNPs found in a random, global approach (Donis-Keller et al., 1987). Developed before the polymerase chain reaction (PCR) was even conceived, RFLP analysis was a powerful, but very laborious, strategy. High quality genomic DNA from multiple individuals were cut with a number of restriction enzymes, separated by gel electrophoresis and transferred to nylon filters. These Southern blots were then probed with random genomic clones to identify variations in the restriction fragment lengths. Because only small amounts of DNA were found on the filter, radioactive labels were used with the probes. Needless to say, this was a very labor intensive and expensive way to identify SNPs, and it was almost impossible to extract the sequence differences associated with the RFLPs. Furthermore, even with probes that were thousands of bases long, one was scanning only a small fraction of the bases at the restriction sites. The situation did not improve with the advent of PCR (Saiki et al., 1985). This was due to the fact that in order for PCR to work, DNA sequence data had to be obtained to design loci-specific PCR primers. To complicate things further, DNA sequencing and oligonucleotide synthesis were both costly activities until quite recently. Global scanning of SNPs was not seriously attempted in the early 1990’s because simple sequence repeat polymorphisms (SSRPs), the highly informative microsatellite markers, were much more easily developed by probing genomic libraries with synthetic oligonucleotides bearing the simple sequence repeat motifs and sequencing the positive clones (Weber and May, 1989; Litt and Luty, 1989). When physical mapping of the human genome began in earnest in the mid-1990’s, many mapped sequence-tagged-sites (STSs) and their PCR primers became available (Hudson et al., 1995). Using a resequencing approach, these STSs were scanned for SNPs (Kwok et al., 1996) and several thousand SNPs were found in the first “large-scale” SNP identification project (Wang et al., 1998). However, even after all the available STSs had been scanned, the number of SNPs found amounted only to a small fraction of the SNPs needed for complex genetic analysis. In the late 1990’s, several favorable factors came together to make global SNP discovery a reality. First, the human genome project sped up its production rate in response to the efforts of a new company to sequence the human genome by whole genome shotgun sequencing. This development meant that high quality sequences of some 90% of the human genome sequence became available for comparison (International Human Genome Sequencing Consortium, 2001; Venter et al., 2001). Second, the cost and efficiency of DNA sequencing improved significantly and one can obtain millions of 500 bp sequencing reads over a relatively short time. Third, bioinformatics tools were developed to compare DNA sequences and identify, with high degree of confidence, sequence variations (Nickerson et al., 1997; Marth et al., 1999). Fourth, the SNP consortium (TSC), a coalition formed by pharmaceutical companies, high technology companies, and the British charity, Wellcome Trust, enlisted the help of 4 academic centers to identify a large number of SNPs using a modified whole genome shotgun sequencing approach (Masood, 1999; Altshuler et al., 2000). The result of large-scale sequencing and comparing sequences deriving from different individuals (or chromosomes) at the same loci was that millions of SNPs were found in the human genome (The International SNP Map Working Group, 2001; Venter et al., 2001). The majority of the SNPs came from three sources. The first set came from analyzing overlapping large-insert clone sequences from the human genome project. When the overlapping bacterial artificial chromosome (BAC) clones came from different libraries, the overlapping sequences (most of them 20-30 kb in length) are from different individuals and many SNPs are found. Because the donors of the BAC libraries are diploid, even when the overlapping clones are from the same library, the two BACs are derived from different parental lineage 50% of the time and SNPs will be found in that situation (Taillon-Miller et al., 1998). Indeed, some 800,000 SNPs were found in the overlapping sequences by the end of year 2000. The second set came from the TSC with data derived from the sequencing of clones from “reduced representation libraries”. By reducing the complexity of the genome through cloning of sizeselected restriction fragments from a pooled DNA sample, the same loci were sequenced multiple times when a large number of clones from each library was sequenced (Altshuler et al., 2000). Even for the singleton clones, their sequences were useful for SNP discovery because they could be compared to the reference human genome sequence. Some 900,000 SNPs were deposited to the public database by the end of 2000. The third set came from the whole genome shotgun sequencing project by a private enterprise and the data were kept from the public. This large set of SNPs is available only to “subscribers” of the private database (Venter et al., 2001). In addition to these large-scale projects, smaller efforts using a variety of approaches also contributed to the global SNP discovery project (for example, Irizzary et al., 2000; Buetow et al., 2001). With the reference human genome sequence now close to completion, global SNP discovery is most efficient when one takes a shotgun sequencing approach and compares the sequencing data obtained against the reference sequence. The density of the markers is determined by the number of genomic clones sequenced. This approach is applicable to any organism as long as the reference sequence of that organism is available.
* **Technologies For Targeted SNP Discovery**
* Unlike global SNP discovery, where speed and achieving a certain density of marker are the most important factors, targeted SNP discovery focuses on finding most, if not all, of the SNPs in the regions of interest. Of course, DNA sequencing is the gold standard of SNP discovery and the evolution of DNA sequencing technology is described in detail in other reviews (Galas and McCormack, 2002). The breakthrough in DNA sequencing technology for SNP discovery came when direct sequencing of PCR products was perfected and the recognition that sequencing data from a heterozygote could be distinguished from those from a homozygote quite unambiguously (Kwok et al., 1994; Zakeri, 1998). However, DNA sequencing was quite labor intensive and costly until very recently and several highly successful polymorphism scanning methods were developed to scan DNA fragments for SNPs and mutations (Kwok and Chen, 1998). Most SNP scanning methods exploit the difference between the mismatched heteroduplex DNA from the perfectly matched homoduplex DNA. The earliest SNP scanning technologies distinguished the homoduplex from the heteroduplex in different ways. They were elegant methods developed without the benefit of PCR. PCR opens up other ways of looking for polymorphisms and a number of other more efficient methods have been developed for SNP discovery.
*  

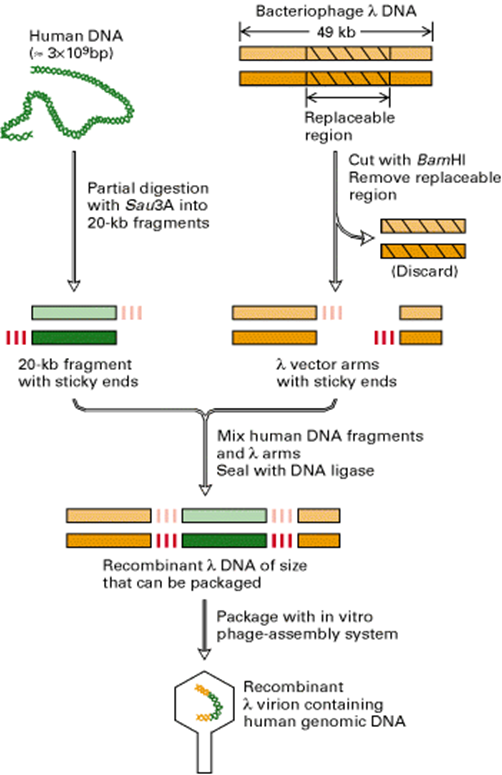
## Seminar 12. Types of DNA microarrays

The use of genetic information is a powerful tool that today is becoming more readily available to scientists.

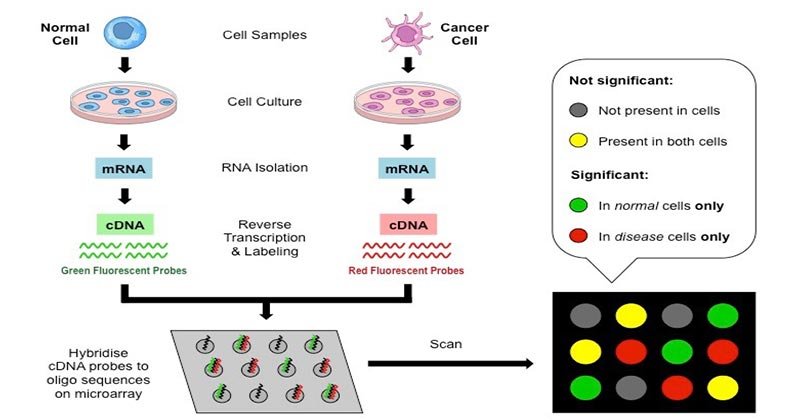
In order to use this powerful tool it necessary to know how to navigate throughout the entire genome. The human genome is about 3 x 10\*E9 bp.

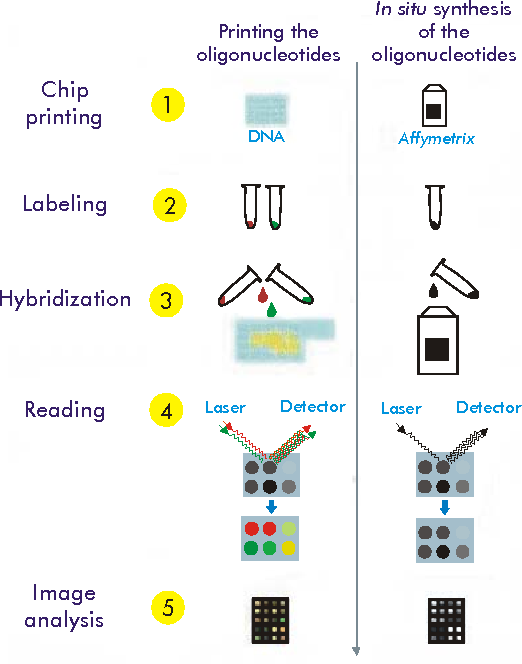
In humans this project is known as Human Genome Project.

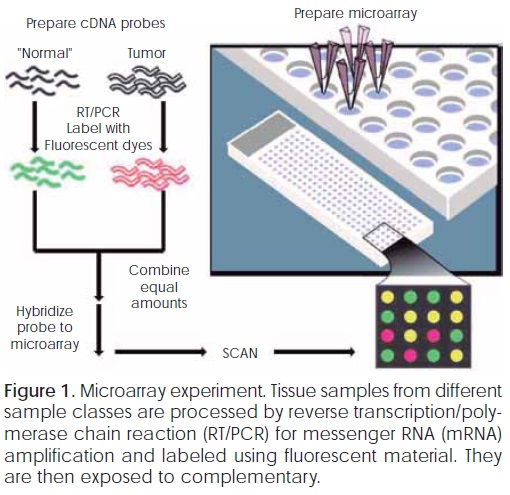
* The Future of DNA arrays. Data standards and data exchange. DNA microarrays for transcription factor binding analysis.



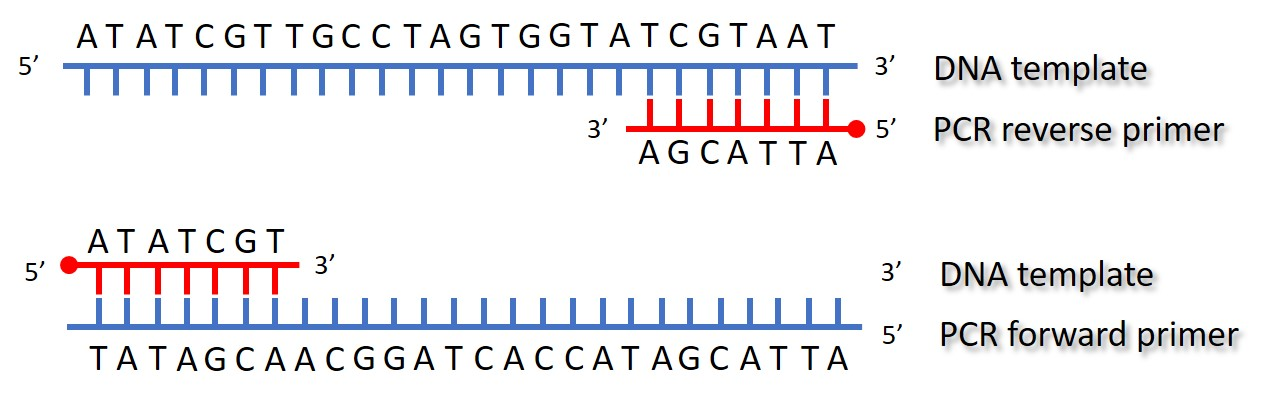
* **Seminar 13.** Preparation of DNA chip and the experiment

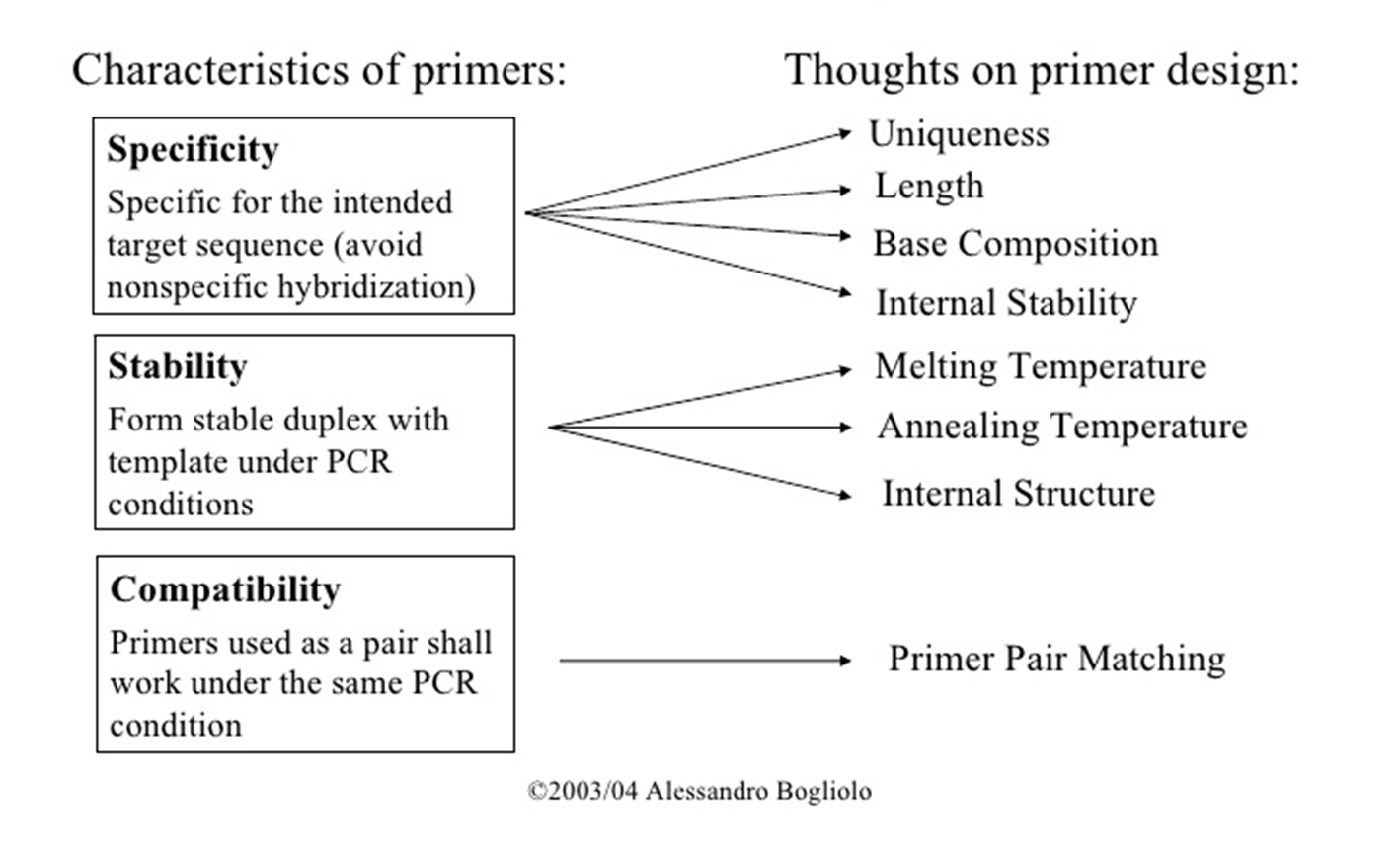
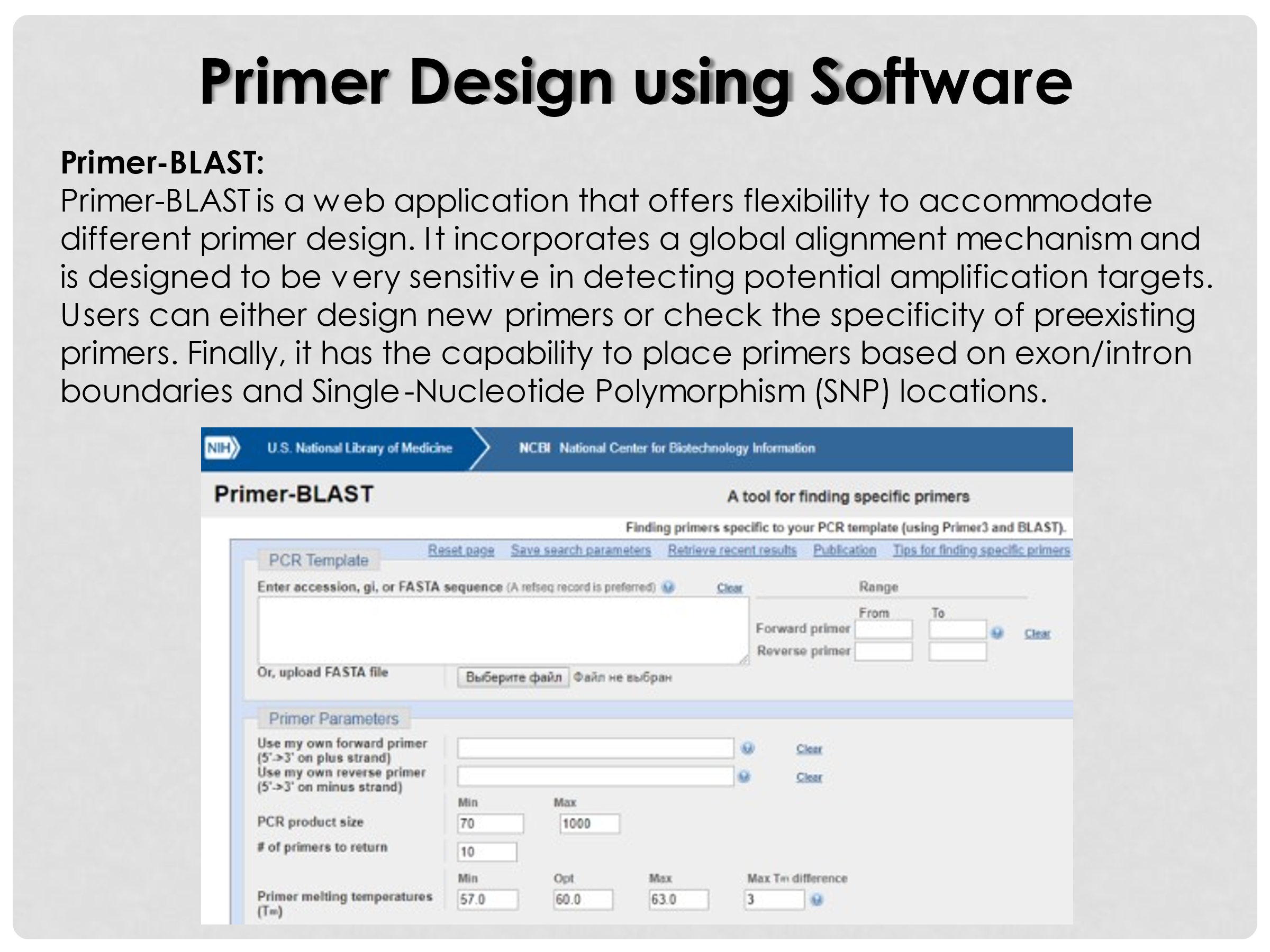






* **Seminar 14.** Applications of different types of molecular markers and PCR in molecular biotechnology
* A **primer** is a short single strand of RNA or DNA (generally about **18-22 bases**) that serves as a **starting point** for DNA synthesis. It is required for DNA replication because the enzymes that catalyze this process, DNA polymerases, can only add new nucleotides to an **existing strand** of DNA. The polymerase **starts** replication at the **3′-end of the primer**, and copies the opposite strand.
* There are **two types of primers** for the replication of each strand of DNA:
* **→ Forward primer**
* **← Reverse primer**



* 
* **Seminar 15. QTL applications in breeding**

Many agriculturally important traits such as productivity and

quality, tolerance to environmental stresses, and some of forms of disease resistance are quantitative, are controlled by polygenes which complicate the breeding

process since phenotypic performances only partially reflects the genetic values of

individuals. These complex traits are referred to as quantitative traits (also called as

polygenic or multifactorial traits) and the regions within genomes that contain genes

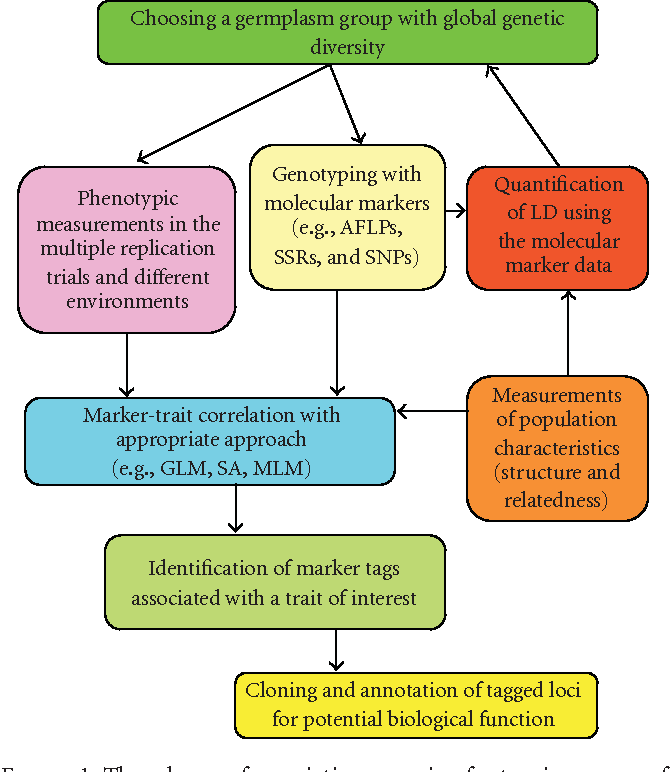
associated with a particular quantitative trait are known ***as quantitative trait loci***

*(QTLs).* (also called polygenic, continuous, multifactorial, or complex traits) in nature.

* The genetic variation of a quantitative trait is controlled by the collective

effects of numerous genes, known as quantitative trait loci (QTLs).

* is a region of DNA that is associated with a particular phenotypic trait.
* These QTLs are often found on different chromosomes.
* Knowing the number of QTLs that explains variation in the phenotypic trait tells us about the genetic architecture of a trait. It may tell us that plant height is controlled by *many genes* of small effect, or by a few genes of large effect.
* Another use of QTLs is to identify candidate genes underlying a trait.
* Once a region of DNA is identified as contributing to a phenotype, it can be sequenced. The DNA sequence of any genes in this region can then be compared to a database of DNA for genes whose function is already known.
* In a recent development, classical QTL analyses are combined with gene expression profiling i.e. by DNA microarrays.
* Such expression QTLs (eQTLs[)](https://en.wikipedia.org/wiki/Expression_quantitative_trait_loci) describe cis- and trans-controlling elements for the expression of often disease-associated genes. Observed epistatic effects have been found beneficial to identify the gene responsible by a cross-validation of genes within the interacting loci with metabolic pathway- and scientific literature databases

Figure 1: The scheme of association mapping for tagging a gene of interest using germplasm accessions. Note that the outlined scheme may vary based on population characteristics and methodology chosen for association

