Physical Maps of Chromosomes

Q1 *Cassandra L. Smith,● Giang H. Nguyen, Denan Wang, and Nickolaev Bukanov*

Keywords

***Alu* Repeat**
A 300-bp DNA interspersed repeat sequence that occurs about once every 5000 bp. The most common repeat element in the human genome, it is CG rich and appears to occur preferentially in light Giemsa bands on condensed chromosomes.

Bacterial Artificial Chromosome (BAC)
An artificial chromosome cloning system in *E. coli* based on elements of the native circular chromosome.

Cloneless Library
A library of genomic fragments consisting of gel slices of electrophoretically fractionated DNA.

Comparative Genomic Hybridization (CGH)
A molecular cytogenetic method capable of detecting and locating relative genomic sequence copy number differences between pairs of DNA samples. Equal concentrations of differentially labeled DNAs are hybridized simultaneously to metaphase chromosome spreads. Regions with deleted or amplified DNA sequences are seen as changes in the ratio of the intensities of these two labels along the target chromosome.

Expressed Sequence Tag (EST)
A segment of a sequence from cDNA clone that corresponds to an mRNA. Mapping ESTs helps make maps and points directly to expressed genes.

Fluorescence *In Situ* Hybridization (FISH)
Hybridization of nucleic acid probes to chromosomes immobilized on microscope slides or filters.

Inter-*Alu* PCR
A polymerase chain reaction method that uses primers contained in *Alu* repeat element to amplify single-copy sequences between adjacent *Alu* elements.

Jumping Libraries
A library of clones containing DNA sequences from the ends of the same restriction fragment.

***Kpn* Repeat**
The second most commonly occurring repeat sequence in the human genome, occurring on average once every 50 000 bp. It is a relatively AT-rich, LINE repeat and occurs preferentially in dark Giemsa bands on condensed chromosomes.

Linking Libraries
Libraries of clones containing DNA sequences that span a selected restriction enzyme recognition site.

Locus
A region on a chromosome linked to a functional unit such as a gene, telomere, centromere, and replication origin.

Long Interspersed Repeat Elements (LINEs)
Long interspersed repeating sequences that appear to be similar to retroposons.

Multiplex Analysis
An analytical approach to increase throughput by collecting multiple data simultaneously.

P1-derived Artificial Chromosomes (PAC)
An *E. coli* bacteriophage cloning system used for cloning of fragments up to 150 kb in size.

Polonies
Isolation of DNA sequences on beads rather than in clones.

Polymerase Chain Reaction (PCR)
A method for amplifying DNA by alternatively denaturing double-stranded DNA, annealing pairs of primers located near each other on complementary strands, and synthesizing the DNA between the primers using DNA polymerase.

Polymorphism Link-up
A mapping approach that establishes continuity between restriction fragments by taking advantage of the naturally occurring polymorphism in different DNAs. In some cell lines, hybridized probes appear to identify different fragments, whereas in others they appear to identify the same fragment. The pattern of occurrence of these fragments can be used to assess whether the probes identify the same or adjacent fragments.

Pulsed Field Gel (PFG) Electrophoresis
A method of electrophoresis that exposes nucleic acids to alternating electrical fields. Fractionation is based on the speed at which the molecules can change directions.

Radiation Hybrid Maps
Radiation hybrid maps are created by analyzing randomly broken DNA cloned into hybrid cell lines. The breakage is done using X rays and the centiRay distance is a function of how often two loci remain together in a library of radiation hybrid cell lines.

RARE (RecA-assisted Restriction Endonuclease) Cleavage
A site-specific DNA cleavage method based on the ability of RecA protein from *E. coli* to pair an oligonucleotide to its homologous sequence in duplex DNA. This three-stranded DNA complex is protected from methylase. After methylation and removal of RecA, restriction endonuclease cleavage is limited to the site previously protected from methylation. If pairs of oligonucleotides are used, a specific fragment can be cleaved out of genomes.

Retroposons
A group of DNA sequence elements that appear to transpose through an RNA intermediate. Retroposon elements do not code for reverse transcriptase, do not have terminally redundant sequences, and do have a 39-poly(A)$_n$ stretch. A variable-sized target duplication occurs at the site of integration.

Sequence-tagged Restriction Site (STAR)
A short DNA sequence used to identify the DNA surrounding a restriction nuclease cleavage site.

Sequence-tagged Site (STS)
A short DNA sequence used to identify a DNA segment.

Short Interspersed Repeat Elements (SINEs)
Repeated sequences less than 500-bp long and present in a high copy number (10^5 per human genome). The *Alu* repeat element consists of SINEs.

Yeast Artificial Chromosome (YAC)
An artificial yeast chromosome constructed by cloning genomic fragments into vectors that can replicate in yeast. YACs have the following characteristics: a yeast centromere, two telomeric sequences, and a selectable marker.

This entry reviews the types of physical map that can be constructed, the methods used to construct them and their most likely current use. In particular, the emphasis is on the construction of genomic restriction maps and ordering overlapping libraries using top-down mapping approaches to enable *de novo* genome sequencing and for gap filling in sequencing projects. Also explored is the use of these approaches in functional studies.

1 Introduction

In the past, most genome studies were limited to organisms having well-developed genetic systems. Thus, a few model systems were studied intensely (e.g. bacteriophage lambda, *Eschericia coli*, Drosophila, mice) in a large number of different ways. The accumulation of diverse knowledge on these model organisms further promoted research on these model systems. Now, a number of molecular approaches have been developed that allow the construction of physical maps for virtually any chromosome. This means that analysis of an uncharacterized organism can begin with physical dissection including sequencing. Knowledge gained from this initial foray provides the foundation for functional studies. Hence, a new synergism in biology is provided by bottom-up genome studies.

Bacterial genomes are small and range from ~1 to 15 Mb in size, reflecting in part, the relative ability of these organisms to be free living. The entire genomic sequence of an increasing number of bacterial genomes is becoming available. These genomes are small enough so that sequencing of random clones is used to obtain the entire genome sequence. Hence, physical genomic maps for these organisms are used to study global gene expression or other global genomic activities.

In the past, the study of some lower eukaryotic genomes such as protozoa was hampered by the lack of well-developed genetic systems. Also, these chromosomes did not condense during cell division and could not be visualized microscopically. The application of pulsed field gel electrophoresis (PFG) to examine the genetic make up of these organisms was quite useful. PFG separates DNA chromosomal DNAs or large fragments that range up to ~10 Mb in size. Protozoan genomes range in size from the length that overlaps large bacterial genomes, up to about 100 Mb. Surprisingly, one parasitic protozoan genome, the *Giardia lamblia* genome, was found to be only about 12 Mb in size (i.e. smaller than the *Saccharomyces* genome).

Until recently, large, complex genomes, like the human genome (~3000 Mb in size) have been particularly recalcitrant to molecular dissection. Human and other higher eukaryotic chromosomes are 50 to 300 Mb in size, condense and are visible microscopically. Differential staining of the condensed chromosomes allows a finer division into regions (''bands'') estimated to be 5 to 30 Mb in size that appear to reflect regional differences in GC content. The banding patterns serve as anchors to a large amount of genomic data. Also, the division of the genome into chromosomes and chromosomal bands provides convenient pieces for top-down mapping approaches.

In top-down mapping approaches, the genome is divided into units to facilitate study. For instance, chromosomes naturally divide the genome, and chromosome-banding patterns represent another division. Further division will depend on the method of analysis. Conventional recombinant DNA approaches to genome analysis allowed the characterization of molecules up to only about 0.05 Mb. The ''resolution gap,'' 0.05 to 10 Mb was exactly the size range that was most amenable to study by PFG techniques and the entire size range of DNA molecules could be analyzed. PFG also promoted the extension of recombinant DNA methods to large cloning system such as yeast artificial chromosomes (YACs), P1 artificial

chromosomes (PACs), and bacterial artificial chromosomes (BACs).

2 Genomic Mapping Terms and Concepts

Many molecular techniques with varying degrees of resolution can be used to characterize chromosomes. Hence, there is no unifying concept of what constitutes a chromosomal physical map. Further, there are ambiguities in the literature with respect to the distinction of physical versus genetic maps.

Any map will consist of markers or objects. The order, and perhaps the distance, between pairs of objects will be known (some might argue that a map must consist of both order and distance). Object ordering along a chromosome should be maintained irrespective of the method used to construct the map, whereas map distances are method dependent. For instance, the amount of recombination along a chromosome is not constant. Hence, it is not surprising that a comparison of the physical and genetic distances along the long arm of chromosome 21 revealed at least a sixfold variation in the distances. It is quite clear that the ultimate map is the entire sequence of a chromosomal DNA. All maps and objects will be anchored to the DNA sequence once it is available. Some imprecision will always exist because map objects may be imprecisely defined, or have imprecisely defined locations by their very nature.

Classically, a genetic map was composed of chromosomal loci and the amount of recombination between the loci was the genetic distance. For eukaryotic organisms, genetic distance, expressed in centimorgans (cM), is a measure of the coinheritance of genetic markers. This type of analysis requires the examination of two loci in multiple generations. In bacteria like *E. coli*, recombination is measured as the time of transfer and integration of DNA from one cell into the chromosome of another cell.

In some instances, the placement of genes on various physical maps has been referred to as a *genetic map* (i.e. ''a gene map''), although only physical distances and locations were known. Cytogenetic maps lead to further confusion in terminology. In these maps, the presence or absence of a gene or map object is correlated with an observable genomic location (i.e. a chromosome band).

Another concept whose use results in some confusion is ''locus.'' Classically, this term defined the location genes (i.e. ''genetic locus''). More generally, however, a chromosomal locus, representing a location on a chromosome, can consist of objects other than genes. Examples include DNA sequences, sequenced tagged sites (STS), expressed sequenced tags (ESTs), probe sites, restriction enzyme sites, clone sites, centromeres, telomeres, and chromosomal breakpoints (such as those that occur naturally or are induced by ultraviolet breakage). To add to the confusion, new genetic markers based on anonymous DNA sequences now define genetic loci because they are used in genetic mapping experiments. In this entry, a *chromosomal locus* is any chromosomal location that has been identified in a distinctive manner.

Here, a physical map is considered to be any map consisting of objects that have been located by physical rather than by genetic methods. Thus, a physical map can consist of objects located along the chromosome, such as a chromosomal band, a breakpoint, a genomic restriction fragment, and the location of a clone

on a chromosome. Here, the ordering of restriction fragments and clones will be used. These maps are constructed using molecular methods.

3 Evolving Approaches for Physical Mapping

The term ''physical maps'' has been used to describe both genomic restriction maps and genomic clone libraries. Low-resolution whole-genome restriction maps are created by ordering fragments that have been fractionated by size electrophoretically. These maps have been created by hybridization experiments using cloned sequences as probes or by PCR (polymerase chain reaction) testing for the presence of specific STSs/ESTs.

Another physical type of map is a radiation hybrid map. In this case, randomly fragmented DNA is cloned to create a hybrid cell line library. Different fragments are present in each hybrid cell line that is tested for the presence of a library of loci. The distance loci are reflected in the frequency of co-occurrence of two loci in the same cell line.

Also, genome restriction maps may be created by analysis of restriction sites contained on overlapping clones. The resolution of such restriction maps depends on the restriction enzyme used, and the frequency at which its recognition site occurs within the genomic DNA sample. Furthermore, an overlapping library may in itself represent a map consisting of ordered objects whose size can be approximated but not stated with certainty.

More efficient ''cloneless'' library approaches use PCR to analyze gel slices of lanes containing specific fragments. Here, the gel slices represent the cloneless library units that can be used in place of cloned DNA. In the future, cloneless approaches will likely replace time-consuming clone library construction and ordering.

4 Genomic Restriction Maps

The first step in creating a genomic restriction map is choosing the DNA source. For many organisms or chromosomes, this is obvious, since there is a well-characterized isolate or a cell line that may be useful. For most eukaryotic genomes, if available, some complications associated with the analysis of polymorphic diploid DNA can be avoided by using DNA from hybrid cell lines containing a single chromosome.

Usually, the genomic DNA is extracted and purified intact, in agarose to prevent shear damage. Small, 10-Mb chromosomal DNAs may be sized directly by PFG. Other chromosomal DNAs must be cleaved with a restriction enzyme before they are subjected to PFG analysis. The largest size standard for PFG electrophoresis is 6 Mb. Thus, analytical results use this limit, although enhanced versions of the technique allow the fractionation of molecules greater than 6 Mb.

Usually, maps are constructed using restriction enzymes that have large recognition sequences or a recognition site that occurs infrequently in the genome of interest. In some cases, it has been useful to test a battery of enzymes. The usefulness of a particular enzyme may be estimated roughly from the size of the site or the GC content of the test organism. Even so, these predictions are somewhat inaccurate because the genomic DNA sequences do not occur at random, whereas calculated occurrences assume randomness. Furthermore, the frequency of occurrence

in known sequences may not be representative of the entire genome because in the past many molecular studies have focused on gene sequencing.

Besides restriction enzymes, a number of other enzymes or methods have been adapted for cutting genomic DNA into specific, large pieces. These methods usually depend on the formation of an unusual structure (e.g. a triplex or D-loop structure) having an associated single-stranded region. D-loops are formed with the aid of RecA protein at targeted sites. In one strategy, termed the *Achilles' heel* strategy, or RARE (RecA-assisted restriction endonuclease) cleavage, genomic DNA containing a D-loop is treated with DNA methylase. The methylase recognizes and modifies specific sequence except when the sites occur within the D-loop. The methylase is removed, and the DNA is then subjected to digestion with the cognate restriction enzyme. Only the restriction enzyme recognition site in the previously formed D-loop is now susceptible to cleavage by the restriction enzyme.

This type of approach allows for site-directed cleavage of genomic DNA at specific locations and has been used to map the end of human chromosomes, as only one RARE cleavage is required. The combination of two RARE reactions from nearby sites would produce an interstitial fragment whose length would be equivalent to the distance between the two sites. Furthermore, fragments smaller than 10 Mb in size could be purified from the remaining genomic DNA using PFG electrophoresis.

Genomic map construction is similar to putting a puzzle together. It is much easier if all the pieces are identified in advance. For small genomes, this is accomplished with a restriction enzyme that produces reasonable number of fragments resolvable by PFG analysis. In these experiments, all genomic DNA is visualized by simple ethidium bromide staining.

A method was developed for visualizing all the megabase restriction fragments for individual chromosomes from large genomes. This approach analyzes PFG-fractionated restriction fragments from genomic DNA that is obtained from monosomic hybrid cell lines. Here, a species-specific interspersed repetitive hybridization probe is used to identify the megabase restriction fragments. For example, the megabase restriction fragments from chromosome 20 are visualized with a human-specific *Alu* probe in Fig. 1. This approach works as long as the restriction enzyme digestion goes to completion. Otherwise, partial digestion products can obscure the complete and especially larger digestion fragments. For the human genome, the restriction enzymes *Not* I, *Sgr*A I, *Fse* I, and *Asc* I have this desired characteristic and cut the genome into fragments that appear to average 1 Mb in size. In higher eukaryotic organisms, partial cleavage by a restriction enzyme is due to partial CpG methylation of their recognition sites inhibiting cleavage.

Most organisms will have several interspersed repeats that are of different lengths. The most commonly occurring interspersed repeat in the human genome is the *Alu* repeat. This repeat is a short interspersed repeating element (SINE) that is estimated to occur about every 5000 bp. The second most commonly occurring interspersed human repeat is the *Kpn* repeat. The *Kpn* element, a LINE (long interspersed) repeat, is estimated to occur at a frequency tenfold less than that of the *Alu* repeat. Theoretically, these two repeats used as hybridization probes to hybrid cell line DNA cleaved with the appropriate enzyme should reveal all the

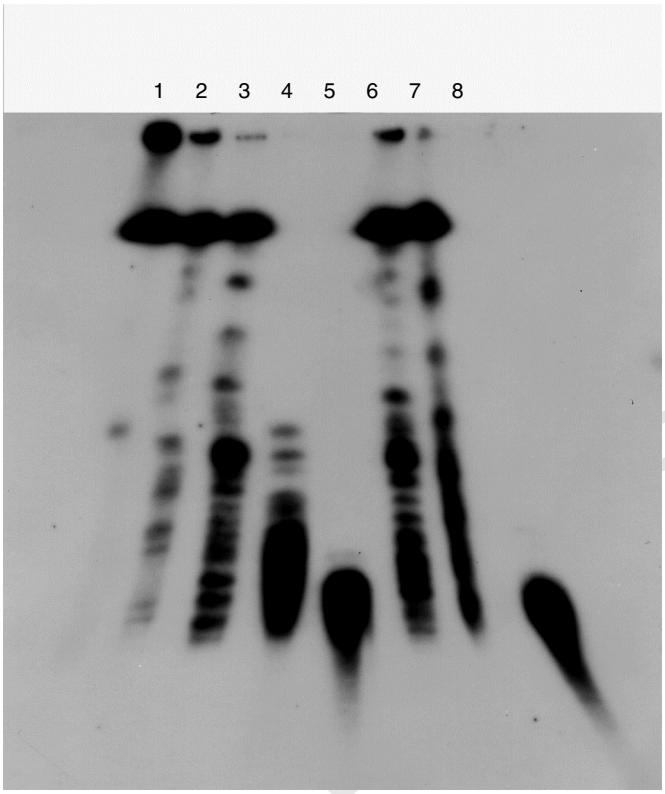

Fig. 1 Detection of PFG-fractionated restriction fragments of human chromosomes. DNA from a monosomic hybrid cell line containing chromosome 20 was digested to completion with different enzymes, PFG fractionated and hybridized to the human-specific *Alu* probe. The recognition sites for the enzymes used are: (1) *Sgf* I (GCGATCGC); (2) *Asc* I (GGCGCGCC); (3) *Fse* I (GGCCGG/CC); (4) *Pme* I (GTTTAAAC); (5) *Pac* I (TTAATTAA); (6) *Not* I (GCGGCCGC); (7) *Sgr*AI (CRCCGGYG); and (8) *Swa* I (ATTT/AAAT).

human megabase fragments. However, this approach may not distinguish two fragments of the same or very similar sizes even with high-resolution PFG fractionations.

The use of repeat sequences as hybridization probes can reveal information about the size and distribution of restriction fragments, but do not reveal order. Regional specific repetitive sequences, like those that occur at telomeric and centromeric regions may be used to identify the ends of physical map, as well as to provide an important anchor for regions of condensed chromosomes visualized microscopically. In human-hybrid cell lines, use of such sequences identifies ends of individual chromosomes.

In conventional mapping experiments, interstitial megabase fragments are linked to single-copy sequences by hybridization experiment or by the more efficient

PCR approach described in the following. Single-copy sequences, also called *STSs*, that have been located on genetic maps can be used for the regional assignment of restriction fragments and can serve as anchors between genetic and physical maps. The accuracy of fragment location will be a function of the resolution of the genetic map. When this approach is applied to a sample like that in Fig. 1, the cataloged megabase restriction fragments containing interspersed repeats are linked to specific single-copy sequences and, if known, genomic locations. ESTs are STS linked to mRNAs and their use would make a map of expressed genes. Note that any PCR primer pair can be used for mapping even if the genetic location of the specific sequence is unknown.

Complete genomic restriction maps do not identify neighboring fragments. At least one neighbor can be identified unambiguously using an STS to analyze partially digested DNA. The difficulty of interpreting the partial digestion data increases dramatically with the number of partial digestion products. For instance, it is important to consider that both the neighboring fragments may be of the same size, that is, the possibility that a single partial product band could represent two different products of the same size must always be borne in mind. This means that neighboring fragments must be confirmed in partial digestion experiments using STSs from the two complete digest fragments. Many times, the confusion associated with partial digest data can be sorted out by obtaining partial digest data on nearly adjacent fragments. In this approach, different sets of partial digest fragments are identified by sequences on different and complete digest fragments. The correct map interpretation is one that is consistent with all the data.

The analysis of partial products is simplified if the STS is from a telomere region or is situated on a small fragment located next to a very large fragment. In both cases, the partial digest information reflects the order of fragments in only one direction. Otherwise, the partial digest products results will represent bidirectional information. This approach also eliminates the problem of comigration of same-size products associated with bidirectional partial mapping data.

Megabase restriction fragments themselves may also be fingerprinted by cleavage with a second enzyme. The products are best analyzed using probes located at the end of the original megabase fragment. The partial digest product data will be different for the two probes located at different positions. The interpretation of the distribution of the cleavage sites of the second enzyme must be consistent with data obtained using the two interstitial (or terminal) probes. This approach is particularly useful for analyzing megabase fragments that are at the limit of the PFG technique, because the products of the second enzyme will be within the PFG limit.

There are several other approaches to proving that two restriction fragments are adjacent. By far the most powerful is the use of linking libraries – small-insert libraries that contain DNA segments from adjacent megabase restriction fragments. Hence, a *Not* I linking clone used as a hybridization probe to genomic DNA digested with *Not* I will identify adjacent *Not* I fragments. A complete linking library would suffice to construct, in the most efficient manner, a complete genomic restriction map. Without a complete linking library, partial digest strategies combined with polymorphism link-up and double restriction enzyme strategies can provide information on neighboring fragments. In

both approaches, parallel analysis of a region provides a regional signature.

The same polymorphism link-up refers to the results obtained when one probe is used to analyze restriction enzyme digested DNA from different cells. For instance, a restriction site may be missing in one cell line, either because there is a mutation at the restriction enzyme site or because methylation is interfering with restriction enzyme cleavage. When a number of DNA sources are examined, it is often possible to fingerprint the polymorphism of a restriction site (i.e. assuming that the site is polymorphic). The probes that are usually located on separate, but adjacent, fragments can, in some DNAs, be found on the same fragment. Many times this latter fragment will be equal to the total size of the two smaller fragments seen in other cell lines. In such a case, the pattern of occurrence of the two distinct smaller fragments, detected by the different probes, is self-consistent. This is very similar to the more familiar approach of fingerprinting a DNA using many different restriction enzymes.

5
Genomic Clone Libraries

The construction and ordering of an overlapping library begins with the selection of the genomic DNA sample as discussed earlier. The next decision entails choice of the type of vector that will be used. The most distinguishing feature of vector possibilities is the size of the DNA that can be cloned into them. It is helpful to begin whole-genome mapping projects using the largest DNA fragments possible, minimizing the number of clones to order. In the past, the largest cloning vectors were cosmids that contained 40-kb cloned sequences. Today, several systems are available for the cloning of larger segments. For instance, megabase fragments can be cloned into YACs (yeast artificial chromosomes) in yeast cells. These libraries have been replaced by the easier to use and more stable PAC (P1 artificial chromosomes) and BAC (bacterial articificial chromosomes) libraries that clone fragments of about 100 kb in size. Today, the construction and ordering of PAC/BAC libraries still remains a major effort.

Usually, today's top-down sequencing approaches begin by creating large clone BAC libraries. Then, random, small overlapping fragments of each BAC are subcloned into sequencing vectors. The large clone libraries are ordered using a variety of techniques (see the following), whereas sequencing is done on an unordered library. Ultimately, the small-insert library is ordered when overlaps are identified while the sequence is put together.

Clone Ordering. In the past, chromosome walking (or ''crawling'') experiments, involved hybridizing individual clones to an entire genomic library to detect overlapping clones. Although this approach is guaranteed to detect overlapping clones, it is slow and laborious.

Today, many efficient genomic approaches are used to order all the clones in a library. The genomic approaches analyze many samples at the same time in a ''multiplex'' approach and generally use methods that are easy to automate. For instance, clones are analyzed in pools to minimize the number of experiments that are needed to order the entire library. In some cases, a tiered pooling strategy is used, where the first testing is done on a small group of superpools, each composed of different pools. Only the pools that make

up a positive superpool needs to be further analyzed. Some designed pooling strategies have each clone present in multiple but distinct pools. The pattern of positive signals to such a set of pools will reveal the identity of the overlapping clones.

Most bottom-up, fingerprinting methods require that libraries be constructed using partial restriction enzyme digests to ensure that overlaps exist between different clones and that all or most of the genome is cloned. There are a number of ways to fingerprint clones to identify overlaps. Bottom-up strategies for library ordering usually involve testing individual clones to search for an overlapping restriction fragment or restriction site pattern. This type of fingerprinting is easy to automate. Clones may also be fingerprinted using interspersed repetitive sequences. Here, clones, or restriction fragments of the clones, are hybridized to a set of short oligonucleotides or longer interspersed repetitive sequences, and the patterns of hybridization are used to link clones.

A variation in this approach that has proven to be quite efficient is the use of genomic DNA restriction fragments as hybridization probes to order clone libraries. Megabase restriction fragment probes allow the assignment of clones to particular fragments. If the fragments are mapped, the clones can then be regionally assigned in the genome. Clones located in different regions are ordered simultaneously by further hybridization experiments using probes generated from a pool of restriction fragments. Here, the pools are created from gel slices containing small restriction fragments. Each pool consists of genomic restriction fragments of a particular size range from different regions of the genome. Overlapping clones will be located on the same megabase fragment and hybridize to the same pool of small fragments. This approach increases the efficiency of library ordering by eightfold.

Another ordering approach involves detecting overlaps by DNA sequencing. Here, the sequences may be chosen at random sites (e.g. STSs) or the sequences can be collected at specific locations originally called *sequence-tagged restriction sites* (STARs). The STAR approach may be used with both partial and complete digest libraries. DNA sequence information may be collected at specific restriction sites, including the ends of the cloned sequences. Clone linking is done using a linking clone library that spans the restriction enzyme site when the first library was created from completely digested DNA. Clones that generated from partial restriction enzyme digestions have ends that overlap. The later method eliminates the need for up-front clone ordering, but a highly representative library is required, and sequence read lengths must be long enough to span most small repeats.

6 Cloneless Genomic Libraries

Cloneless libraries have several advantages over clone libraries. These libraries can be made rapidly and inexpensively from any genomic source and all genomic DNA is present in the library. These libraries are made from agarose or acrylamide gels containing size fractionation DNA of low or high complexity. The gel slice containing the cloneless fraction is treated in the same manner as a cloned DNA sample.

In most of the discussion below, only complete digest cloneless libraries are discussed. However, partially digested or random broken DNAs could be used to

generate a cloneless library. In essence, the analysis and use of a cloneless library is very similar to that of a clone library.

The gel lane is carefully sliced into thin pieces using a clean coverslip for each cut to insure minimum cross-contamination. The DNA in each slice is used as a source for testing, for instance, with PCR. For small and medium size genome, the number of unique fragments within a slice will be small. For large genomes, there are many fragments in a single slice. In the case of samples from the gels, like those shown in Fig. 1, the only visible fragment is from the human DNA but other rodent fragments are present. The cloneless approach is a hybrid between genomic mapping and clone ordering.

One consideration with the cloneless approach is that DNA in agarose is believed to be difficult to manipulate enzymatically. However, others and we have developed robust and reliable simple methods for using DNA in agarose that avoids purification. Most methods were developed some time ago for PFG applications. Additional methods were developed more recently for PCR that involve the addition of preservatives to the stored fraction and the PCR.

Another potential problem with cloneless fractions is the low amount of DNA in each slice. There are a number of PCR methods that solve this problem by randomly amplifying the DNA in a fraction (or clone) to generate a large amount of additional material. Furthermore, these methods have been developed to minimize the introduction of mutations. Hence, the amplified DNA appears to be representative of the starting template.

Cloneless fractions can be used for a variety of experiments in addition to that of genomic sequencing. For instance, a recently developed method termed *comparative genome hybridization* (CGH) uses differentially labeled cDNA, or genomic DNA, from two samples. The labeled samples are mixed together in equal amounts and used as a hybridization probe. The target samples are metaphase chromosomes or clones. Quantitative analysis of the hybridization signal provides information on the relative amounts of DNA from different chromosomal regions. In summary, CGH experiments provide positional information of differences between samples.

An alternative to conventional CGH would utilize genome DNA arrays that have been generated from size-fractionated genomic restriction fragments. An array consisting of 2-mm slices from a 15-cm gel lane would have ~40-Mb resolution. Each fraction could be cleaved with a second enzyme and subjected to a second electrophoretic size fractionation. The ~1000 slices generated from the entire second fractionation would provide DNA at ~3-Mb resolution.

The cloneless, like a clone library approach does not provide positional genomic information directly. For instance, clones are linked to genomic position by FISH (fluorescence *in situ* hybridization) experiments to metaphase chromosomes. Similarly, the DNA in each slice can be used in a FISH experiment to metaphase chromosome and locate it to a specific genomic region. Genomic restriction maps, or ordered clone, or cloneless libraries can be used in place of metaphase chromosomes especially in those organisms, like protozoan, whose chromosomes do not condense during cell division.

The approaches used to order clone libraries may be applied to cloneless libraries. For instance, fraction can be linked to STSs/ESTs. Alternatively, hybridization experiments or other methods that were

described earlier can be used to detect overlaps with clone libraries or a second cloneless library generated with a second enzyme.

Cloneless libraries can be used in place of X-ray hybrid maps. X-ray hybrid maps are created by determining the frequency that two loci are separated by random breaks in DNA generated by X rays. In this approach, randomly broken DNA is cloned into a hybrid cell line and a library of different cell lines are created with different pieces of DNA. Then, each cell line is tested for the presence of a marker and a centiRay distance is calculated that reflects the frequency at which two loci occur in the same cell line. Like genetic mapping, the closer two loci are, the more likely they will be on the same randomly generate fragments. Here, instead of cell lines, randomly broken DNA fractionated by size could be used to create a cloneless library. Then each fraction could be tested for each loci and the distance between loci calculated as a function of the number of fractions that contain both loci.

7 Gap Closure

It is easy to start maps but difficult to finish them. Each gap in each map presents a unique problem. Gaps may be true gaps or pseudogaps. Pseudogaps contain repetitive sequences and require different strategies for closure than true gaps.

True gaps contain unique sequences and need to be treated individually. The best strategy will depend on the putative size of the gap, the amount of polymorphism in the region, the number of unassigned megabase fragments or clones. Gaps that arise from unclonable sequences may be filled using a cloneless library. For instance, STSs could be developed from cloneless fraction that overlap two regions as we have done for chromosome 21. For instance, single-copy human sequences located between *Alu* elements can be generated from inter-*Alu* PCR amplifications and used as probes to identify fragments in gap regions. Alternatively, the template contained within a fraction can be labeled and used as a hybridization probe to an array of samples at the borders of gaps. Ultimately, the DNA within a gap is identified in a cloneless library and directly sequenced using a direct genomic sequences approach.

8 Prospectus

What is the role of physical maps today and in the future? An increasing sequencing efficiency is coupled to a decreasing need to create and order expensive and time-consuming clone libraries. For small genomes, whole-genome sequencing projects use a bottom-up sequencing approach with unordered clone libraries. Soon a cloneless sequencing approach using ''polonies'' may replace the need for cloning of these small genomes. The polonies approach sequences pools of single DNA molecules that have been immobilized on beads. Sequencing is done on a library of immobilized DNAs simultaneously. Today, this method allows sequencing read of up to ~14 bases per fragment end, or ~28 bases per polony, and up to 1×10^9 beads/slide. Alternatively, imaging methods that directly view the sequence of individual DNA molecules may be used. Meanwhile, adaption of the cloneless approach for larger genome sequencing projects should considerably reduce *de novo* sequencing by eliminating

the need for creating clone libraries, as well as enable gap closure.

Bibliography

Books and Reviews

Cantor, C.R., Smith, C.L., Matthew, M.K. (1988) Pulsed field gel electrophoresis of very large DNA molecules, *Annu. Rev. Biophys. Biophys. Chem.* **17**, 287–304.

Ledbetter, S.A., Nelson, D.L. (1991) Genome Amplification Using Primers Directed to Interspersed Repetitive Sequences (Irs-Pcr), in: McPherson, M.J., Quirke, P., Taylor, G.R. (Eds.) *Polymerase Chain Reaction I: A Practical Approach*, IRL Press, Oxford.

Murphy, M.I., Hammond, H.A., Caskey, C.T. (1996) Solid-phase automated sequencing of Pcr-amplified genomic DNA, *Methods Mol. Biol.* **65**, 163–176.

Schalkwyk, L.C., Francis, F., Lehrach, H. (1995) Techniques in mammalian genome mapping, *Curr. Opin. Biotechnol.* **6**(1), 37–43.

Smith, C.L., Condemine, G. (1990) New approaches for physical mapping of small genomes, *J. Bacteriol.* **172**(3), 1167–1172.

Smith, C.L., Lawrance, S.K., Gillespie, G.A., Cantor, C.R., Weissman, S.M., Collins, F.S. (1987) Strategies for mapping and cloning macroregions of mammalian genomes, *Methods Enzymol.* **151**, 461–489.

Smith, C.L., Klco, S., Zhang, T.Y., Fang, H., Oliva, R., Fan, J.B., Bremer, M., Lawrence, S. (1993) Analysis of Megabase DNA Using Pulsed Field Gel Electrophoresis, in: Adolph, K.W. (Ed.) *Methods in Molecular Genetics*, Academic Press, San Diego, CA.

Van der ploeg, L.H.T., Gottesdiener, K.M., Korman, S.H., Weiden, M., Le Blancq, S. (1992) Protozoan Genomes, in: Ulanovsky, L., Burmeister, M. (Ed.) *Methods in Molecular Biology, Pulsed-Field Gel Electrophoresis. Protocols, Methods, and Theories*, Humana Press, Totowa, NJ.

Welsh, J., McClelland, M. (1994) Fingerprinting Using Arbitrarily Primed Pcr: Application to Genetic Mapping, Population Biology, Epidemiology, and Detection of Differentially Expressed Rnas, in: Mullis, K.B., Ferre, F., Gibbs, R.A. (Eds.) *The Polymerase Chain Reaction*, Birkhauser, Boston, MA..

Primary Literature

Albarghouthi, M.N., Barron, A.E. (2000) Polymeric matrices for DNA sequencing by capillary electrophoresis, *Electrophoresis* **21**(18), 4096–4111.

Barnes, W.M. (1994) Pcr amplification of up to 35-Kb DNA with high fidelity and high yield from lambda bacteriophage templates, *Proc. Natl. Acad. Sci. U.S.A.* **91**(6), 2216–2220.

Cai, W., Jing, J., Irvin, B., Ohler, L., Rose, E., Shizuya, H., Kim, U.J., Simon, M., Anantharaman, T., Mishra, B., Schwartz, D.C. (1998) High-resolution restriction maps of bacterial artificial chromosomes constructed by optical mapping, *Proc. Natl. Acad. Sci. U.S.A.* **95**(7), 3390–3395.

Dederich, D.A., Okwuonu, G., Garner, T., Denn, A., Sutton, A., Escotto, M., Martindale, A., Delgado, O., Muzny, D.M., Gibbs, R.A., Metzker, M.L. (2002) Glass bead purification of plasmid template DNA for high throughput sequencing of mammalian genomes, *Nucleic Acids Res.* **30**(7), e32.

Ferrin, L.J., Camerini-Otero, R.D. (1994) Long-range mapping of gaps and telomeres with Reca-assisted restriction endonuclease (Rare) cleavage, *Nat. Genet.* **6**(4), 379–383.

Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A., Kirkness, E.F., Kerlavage, A.R., Bult, C.J., Tomb, J.F., Dougherty, B.A., Merrick, J.M. et al. (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* rd, *Science* **269**(5223), 496–512.

Florijn, R.J., Bonden, L.A., Vrolijk, H., Wiegant, J., Vaandrager, J.W., Baas, F., den Dunnen, J.T., Tanke, H.J., van Ommen, G.J., Raap, A.K. (1995) High-resolution DNA fiber-fish for genomic DNA mapping and colour bar-coding of large genes, *Hum. Mol. Genet.* **4**(5), 831–836.

Fonstein, M., Haselkorn, R. (1995) Physical mapping of bacterial genomes, *J. Bacteriol.* **177**(12), 3361–3369.

Goldenberger, D., Perschil, I., Ritzler, M., Altwegg, M. (1995) A simple ''Universal'' DNA extraction procedure using Sds and proteinase K is compatible with direct Pcr amplification, *PCR Methods Appl.* **4**(6), 368–370.

Green, E.D. (2001) Strategies for the systematic sequencing of complex genomes, *Nat. Rev. Genet.* **2**(8), 573–583.

Hatada, I., Hayashizaki, Y., Hirotsune, S., Komatsubara, H., Mukai, T. (1991) A genomic scanning method for higher organisms using restriction sites as landmarks, *Proc. Natl. Acad. Sci. U.S.A.* **88**(21), 9523–9527.

Heiskanen, M., Kallioniemi, O., Palotie, A. (1996) Fiber-fish: experiences and a refined protocol, *Genet. Anal.* **12**(5–6), 179–184.

Hozier, J., Graham, R., Westfall, T., Siebert, P., Davis, L. (1994) Preparative in situ hybridization: selection of chromosome region-specific libraries on mitotic chromosomes, *Genomics* **19**(3), 441–447.

Kallioniemi, A., Kallioniemi, O.P., Sudar, D., Rutovitz, D., Gray, J.W., Waldman, F., Pinkel, D. (1992) Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors, *Science* **258**(5083), 818–821.

Kim, U.J., Birren, B.W., Slepak, T., Mancino, V., Boysen, C., Kang, H.L., Simon, M.I., Shizuya, H. (1996) Construction and characterization of a human bacterial artificial chromosome library, *Genomics* **34**(2), 213–218.

Kovacic, R.T., Comai, L., Bendich, A.J. (1995) Protection of megabase DNA from shearing, *Nucleic Acids Res.* **23**(19), 3999–4000.

Lawrence, J.B., Carter, K.C., Gerdes, M.J. (1992) Extending the capabilities of interphase chromatin mapping, *Nat. Genet.* **2**(3), 171–172.

Marra, M.A., Kucaba, T.A., Dietrich, N.L., Green, E.D., Brownstein, B., Wilson, R.K., McDonald, K.M., Hillier, L.W., McPherson, J.D., Waterston, R.H. (1997) High throughput fingerprint analysis of large-insert clones, *Genome Res.* **7**(11), 1072–1084.

Michalet, X., Ekong, R., Fougerousse, F., Rousseaux, S., Schurra, C., Hornigold, N., van Slegtenhorst, M., Wolfe, J., Povey, S., Beckmann, J.S., Bensimon, A. (1997) Dynamic molecular combing: stretching the whole human genome for high-resolution studies, *Science* **277**(5331), 1518–1523.

Newell, W., Beck, S., Lehrach, H., Lyall, A. (1998) Estimation of distances and map construction using radiation hybrids, *Genome Res.* **8**(5), 493–508.

Paez, J.G., Lin, M., Beroukhim, R., Lee, J.C., Zhao, X., Richter, D.J., Gabriel, S., Herman, P., Sasaki, H., Altshuler, D., Li, C., Meyerson, M., Sellers, W.R. (2004) Genome coverage and sequence fidelity of Phi29 polymerase-based multiple strand displacement whole genome amplification, *Nucleic Acids Res.* **32**(9), e71.

Ricke, D.O., Liu, Q., Gostout, B., Sommer, S.S. (1995) Nonrandom patterns of simple and cryptic triplet repeats in coding and noncoding sequences, *Genomics* **26**(3), 510–520.

Smith, C.L., Warburton, P., Gaal, A., Cantor, C.R. (1986) Analysis of genome organization and rearrangements by pulsed gradient gel electrophoresis, *Genet. Eng.* **8**, 45–70.

Wu, C.C., Nimmakayala, P., Santos, F.A., Springman, R., Scheuring, C., Meksem, K., Lightfoot, D.A., Zhang, H.B. (2004) Construction and characterization of a soybean bacterial artificial chromosome library and use of multiple complementary libraries for genome physical mapping, *Theor. Appl. Genet.* **109**(5), 1041–1050.

Zabarovska, V.I., Gizatullin, R.Z., Al-Amin, A.N., Podowski, R., Protopopov, A.I., Lofdahl, S., Wahlestedt, C., Winberg, G., Kashuba, V.I., Ernberg, I., Zabarovsky, E.R. (2002) A new approach to genome mapping and sequencing: slalom libraries, *Nucleic Acids Res.* **30**(2), E6.

500056

QUERIES TO BE ANSWERED BY AUTHOR (SEE MARGINAL MARKS Q..)

IMPORTANT NOTE: Please mark your corrections and answers to these queries directly onto the proof at the relevant place. Do NOT mark your corrections on this query sheet.

Q1. Please provide the affiliations of all the authors.