

## Glossary

**Agar:** a polysaccharide complex extracted from algae found in seawater. Powdered agar is added to hot liquid growth media and the mixture is poured into petri dishes. Upon cooling, the agar mixture forms a semi-solid (Jell-O like) surface used for the growth of bacteria or other microorganisms.

**Agarose:** a highly purified form of agar, specifically used to make gels for electrophoresis.

**Alkaline lysis:** a method involving raising the pH to lyse bacterial cells to isolate plasmid DNA. This method is used in genomic sequencing projects as a way to get large quantities of plasmid DNA from the bacterial cells in liquid cultures.

**Analysis Core:** a core facility that is responsible for annotating (labeling specific regions of) the genomic sequence data that is assembled in Finishing Core. Analysis Core is also responsible for uploading the completed genomic sequence data from the Genome Sequencing Center to NCBI public databases.

**Antibiotic resistance:** a property of a cell that enables it to avoid the effect of an antibiotic that would otherwise kill or inhibit the growth of that cell. Genes that provide resistance to antibiotics are commonly found on plasmids, and have been used as selection markers in vector constructs.

**Antibiotics:** toxins produced by many simple organisms such as bacteria, yeasts, and molds. Some common antibiotics are ampicillin, kanamycin, penicillin, and tetracycline.

**Antibiotic selectivity:** a term referring to the use of antibiotic resistance to allow survival of only cells carrying antibiotic resistance genes. Antibiotics are often added to bacterial growth media in order to select for bacteria resistant to antibiotics. Genes that confer antibiotic resistance are often used as selection markers on vector constructs.

**Autoclave:** an apparatus that uses steam under pressure to sterilize instruments, glassware, and various types of growth media. It is similar to a giant pressure cooker.

**BAC (bacterial artificial chromosome):** a DNA vector designed to carry large segments of DNA (50,000 – 250,000 bp) from another species into a bacterial cell. Once a host bacterium has been transformed with a BAC, it will replicate the foreign DNA in the BAC along with its own DNA. BACs are used in genomic sequencing projects to carry large chunks of DNA from the genome of the organism being studied.

**Bacterial transformation:** the process by which bacterial cells take up foreign DNA molecules. Bacteria are generally treated with  $\text{CaCl}_2$  or subjected to an electric shock to open pores in the cell membrane, allowing DNA to enter the cell. Electroporation is the method commonly used in genomic sequencing projects.

**Bacteriophage:** a virus that attaches to, injects its DNA into, and multiplies inside bacteria. It is often abbreviated as “phage.” Bacteriophage lambda is commonly used as a vector with *E. coli* bacteria.

**Bioinformatics:** the use of computers to generate, store, and efficiently analyze large data sets of biological interest. In genome sequencing projects this includes acquisition and assembly of nucleotide sequence data that can be analyzed to identify genes, as well as the comparison of nucleotide sequence data to find sequence similarities with other genes or parts of genes.

**BLAST (basic local alignment search tool):** a computer tool used for searching and aligning nucleic acid or amino acid sequences to find identical or similar regions.

**Blunt ends:** the ends of a double-stranded DNA fragment in which both strands are broken at the same nucleotide position.

**Centrifugation:** the separation of molecules by size and density using a centrifugal force.

**Chimeric DNA molecule:** a DNA molecule containing nucleotide sequences from different organisms. A good example is a bacterial plasmid containing a fragment of human DNA. It is also known as recombinant DNA or genetically engineered DNA. See “Construct” below.

**Chromosome:** one molecule of DNA carrying a linear end-to-end, or sequential, arrangement of genes interspersed with other sequences. In prokaryotes the chromosome is often a circle of DNA, while in eukaryotes the chromosome is linear, extending from one end, a telomere, through the centromere, to the other telomere.

**Clone (a molecule):** to create copies of a specific molecule. See “Cloned DNA” below.

**Clone (an organism):** a group of individual organisms, or cells, produced from one individual cell through asexual reproduction, resulting in the clone members having identical genetic composition. The word “clone” may be used as a noun or a verb. For example, scientists have recently “cloned” many adult mammals.

**Cloned DNA:** DNA that has been isolated, inserted into a plasmid or other vector, and used to transform a host cell. By growing up large amounts of the host cell, one can recover significant amounts of the vector and hence significant amounts of the insert DNA sequence.

**Comparative genomics:** the study of genomes by comparisons among related organisms such as mice, chimpanzee, and humans. It is often abbreviated to “genomics.”

**Consed:** a computer program used to look at the assembled DNA sequence data from a genomic sequencing project. It is used to identify problems, including low-quality sequence data or high quality discrepancies in sequence data.

**Construct:** an artificially constructed, or genetically engineered, DNA molecule. It may also be referred to as a “cassette” or “transgene.” It may include gene promoter(s), leader sequence, termination codon, etc. A construct is often put into a vector for transformation into a host cell.

**Contig (contiguous sequence):** a group of overlapping DNA fragments that generate an uninterrupted sequence for one region of a genome.

**Deletion:** the loss of a DNA sequence. The size of the loss can be anywhere from just a few base pairs to a large segment thousands of base pairs long.

**DNA fingerprinting:** a technique that defines a unique identity for a given DNA molecule by breaking the molecule into a unique set of fragments based on its sequence, using various restriction enzymes.

**DNA ligase:** an enzyme that joins broken pieces of DNA by catalyzing the formation of a covalent bond between one nucleotide and another nucleotide. It is needed during DNA replication to join together the newly synthesized fragments of the lagging strand. It is used in genetic engineering to seal together the DNA sequences of vectors and inserts.

**DNA polymerase:** the enzyme capable of extending a strand of DNA from a primer by adding successive nucleotides in the order dictated by an associated template strand. Essentially, the enzyme extends the primer strand by one nucleotide each time it acts. DNA polymerase requires a template strand, which dictates the sequence of bases added, and a primer that provides the site to which nucleotides are added. *Taq* DNA polymerase is the enzyme used in the PCR-based sequencing reactions of genomic sequencing projects.

**Duplication:** the gain of a second, tandem copy of a particular DNA sequence. The size of a duplication can range anywhere from just a few base pairs to a DNA segment thousands of base pairs long.

**Electropherogram:** a pictorial representation of the results from a sequencing reaction. The series of colored peaks are used to represent verification of particular nucleotides in their correct sequence order.

**Electroporation:** a technique for transforming or introducing DNA vectors into cells. Electroporation uses the application of a high voltage electric pulse to open pores in the cell membrane, allowing the vector to be transferred into the cell. This transformation method is commonly used in genomic sequencing projects.

**Endonuclease:** an enzyme that cleaves bonds within a polynucleotide chain to create new ends. A particular endonuclease may be specific for single-stranded or double-stranded RNA or DNA. Restriction enzymes are endonucleases that cut DNA at a specific sequence, for example  $CGCG \rightarrow CG \quad CG$ .

**Ethidium bromide:** a fluorescent stain used for visualizing DNA under ultraviolet light. Electrophoresis gels are often stained with ethidium bromide so that the separation of different sized DNA molecules can be visualized.

**Exonuclease:** an enzyme that cleaves nucleotides one at a time from the end of a polynucleotide chain. A particular exonuclease may be specific for either the 5' or 3' end of either DNA or RNA.

**Finishing Core:** a core facility that is responsible for downloading sequence data off of computer servers and assembling the genomic sequence fragments in the correct order. The assembled data is then sent on to Analysis Core.

**Gel electrophoresis:** a technique used to separate molecules (or fragments of a molecule) according to size and electrical charge. Smaller molecules will move faster through a gel than larger molecules. This process is used to separate DNA fragments by size after DNA has been cut with restriction enzymes. It is also used to separate the DNA fragments generated in PCR-based sequencing reactions by size.

**Gel extraction:** the recovery of DNA fragments from the agarose of an electrophoresis gel. A DNA band must be physically cut out of the agarose gel and then chemically treated to remove all traces of agarose.

**Gene:** the entire contiguous nucleic acid sequence within a genome that is necessary and sufficient for the synthesis of a functional product, whether the product is a protein or a molecule of RNA.

**Genome:** all of the DNA in a particular organism, or one copy of the entire hereditary material in a particular cell. It should be noted that in addition to the DNA contained in a cell nucleus, an organism's cells contain DNA in other locations – bacteria have plasmid DNA, plants have plastid DNA, animals have mitochondrial DNA.

**Genomic sequencing:** determining the exact order of nucleotide bases in the genome of an organism.

**Genotype:** the total genetic, or hereditary, makeup that individuals receive from their parents.

**Golden Path:** a term that describes the collection of genomic sequence fragments, carried in BACs, which overlap to adequately represent an entire genome. It is this collection of BACs that will be sequenced at the Genome Sequencing Center.

**Growth media:** a mixture of nutrients required for the growth of microorganisms or cells. Depending on the preference of the organism, it can be used as a liquid or as a solid with the addition of agar.

**Host-vector system:** the combination of a particular vector and the host cell in which it can replicate. It is essential that the host cell can be transformed by the vector DNA at a reasonable frequency, and that the vector has an appropriate origin of replication to function in the host cell.

**Inoculation:** the transfer of microorganisms into or onto new growth media. This often refers to moving a single colony of bacteria into new liquid growth media to grow up a liquid culture.

**Inoculation Core:** a core facility that is responsible for using a portion of the liquid cultures in the 384-well plates from Picking Core for inoculation of cultures in 96-well plates that will go on to Prepping Core. The remaining liquid cultures are frozen and stored in archives at the Genome Sequencing Center.

**Library Core:** a core facility that is responsible for generating a library or collection of the specific genomic fragments from each BAC selected to be sequenced. Each genomic fragment that was carried on a BAC and selected for the Golden Path is broken down into smaller fragments of about 2000 bp each. These fragments are ligated into vectors and sent to Plating Core. The total collection of small fragments is called a genomic library.

**Ligase:** an enzyme used to catalyze the covalent joining of two adjacent nucleotides. See “DNA ligase” above.

**Ligation:** the formation of a covalent bond linking two adjacent bases that are separated by a nick in one strand of a double-stranded piece of DNA. Sticky ends that have hybridized, as well as blunt ends, can be joined covalently by ligase. The term is also used to refer to the joining of vector and insert DNA into a single recombinant DNA molecule.

**Loading Core:** a core facility that is responsible for placing the products of the sequencing reactions into sequencing machines. The sequencing reactions are done in 96-well plates and the sequencing machines at the Genome Sequencing Center use a 96-capillary tube electrophoresis system to separate the sequencing reaction products by size. The sequencing machines analyze and record DNA sequence data, which is then sent to computer servers in the Genome Sequencing Center.

**Mapping Core:** a core facility that is responsible for identifying and choosing the BACs that will go through the Production Pipeline at the Genome Sequencing Center. Each BAC carries a part of the genome that is to be sequenced. Mapping Core uses DNA fingerprinting gels to decide which BACs will make up the Golden Path, and therefore will be sent into the Production Pipeline.

**Materials Core:** a core facility that is responsible for ordering, receiving, and tracking all physical materials used at the Genome Sequencing Center.

**Media Core:** a core facility that is responsible for preparing all growth media used at the Genome Sequencing Center.

**NCBI (National Center for Biotechnology Information):** a national resource and repository for molecular biology information established in 1988. NCBI creates public databases, conducts research in computational biology, develops software tools for analyzing genome data, and provides access to biomedical information.

**Nucleotide:** the building block of DNA and RNA. A nucleotide is a hybrid molecule consisting of a purine or pyrimidine base covalently bonded to a sugar ring, which is covalently bonded to a phosphate group. Adenine (A) and guanine (G) are purines, and cytosine (C) and thymine (T) are pyrimidines. In RNA, thymine is replaced by uracil (U).

**Oligonucleotide primer:** a short chain of nucleotides that has been synthesized (made) by chemically linking together a number of specific nucleotides. “Oligos” are typically 20-24 bases long. They base pair with a complimentary template DNA strand and function as the starting point for the addition of nucleotides by DNA polymerase.

**Origin of replication (ori):** the specific sequence of DNA at which replication is initiated in living cells.

**PCR:** See “Polymerase Chain Reaction” below.

**PCR machine:** See “Thermocycler” below.

**Phenotype:** the outward appearance or other visible characteristics of an organism, which is determined by its DNA and by environmental factors.

**Phred and Phrap:** computer programs used to collect and represent high-quality DNA sequence data.

**Picking:** a technique for selecting specific bacterial colonies for inoculation into liquid growth media.

**Picking Core:** a core facility that is responsible for selecting transformed bacterial cells from the colonies grown up on large agar plates and then using these to inoculate liquid cultures in 384-well plates. Each Q-Pix machine in Picking Core uses a camera and computer to robotically pick appropriate bacterial colonies and inoculate the 384-well plates. These plates are sent on to Inoculation Core.

**Plasmid:** an independent, stable, self-replicating piece of double-stranded DNA found in bacterial cells. A plasmid is not part of the normal bacterial cell genome and never becomes integrated into the host chromosome. Plasmids are often used in recombinant DNA experiments as acceptors of foreign DNA.

**Plating:** a technique for spreading bacterial cells onto solid growth media, usually agar plates.

**Plating Core:** a core facility that is responsible for using electroporation to put the recombinant DNA molecules made up of vector and one 2000 bp fragment from the genomic library (generated in Library Core) into bacterial cells. The transformed bacterial cells are plated on large agar plates and the plates are sent on to Picking Core.

**Point mutation:** a change in a single base pair of a double-stranded DNA sequence.

**Polyacrylamide gel:** a type of gel used in the separation of sequencing reaction products by electrophoresis. In capillary system sequencing machines, polyacrylamide gel is the separation matrix found within the capillary tubes.

**Polymerase Chain Reaction (PCR):** a method for amplifying (making millions of copies of) a specific DNA sequence that otherwise could not be detected or studied. It can be thought of as similar to DNA replication. A PCR reaction requires template DNA, oligonucleotide primers, deoxyribonucleotides (As, Ts, Cs, and Gs), a suitable DNA polymerase (preferably *Taq* or another heat stable DNA polymerase), and buffer.

**Polymorphism:** the occurrence in a population of several phenotypic forms due to differences in gene sequences at particular alleles.

**Prepping Core:** a core facility that is responsible for isolating and purifying DNA from liquid cell cultures grown in 96-well plates, sent from Inoculation Core. Prepping Core uses an alkaline lysis solution and a magnetic bead prep to prepare clean DNA for Sequencing Core. The method is automated using the Plate Trak machine.

**Proteomics:** the scientific study of an organism's proteins and their role in an organism's structure, growth, health, disease (and/or the organism's resistance to growth, etc.).

**Recognition site:** see "Restriction site" below.

**Restriction enzymes:** endonucleases (enzymes) that cut DNA by recognizing a specific target nucleotide sequence of DNA and breaking the DNA chain at that site or target.

**Restriction map:** a pictorial representation of the position of specific restriction sites in a DNA molecule. Restriction sites are the specific sequence locations on a DNA molecule that may be cleaved, or cut, by a given restriction enzyme. The restriction map of a plasmid is circular.

**Restriction site:** a nucleotide sequence in a DNA molecule that is "recognized" and cleaved by a given restriction endonuclease. It may also be referred to as a "recognition site."

**Sequencing Core:** a core facility that is responsible for setting up the PCR-based dye-terminator sequencing reactions in the Production Pipeline at the Genome Sequencing Center. Clean template DNA for the sequencing reactions is received in 96-well plates from Prepping Core. The sequencing reactions are set up in 96-well plates and run in thermocyclers. Once the sequencing reactions are complete, the 96-well plates are sent on to Loading Core.

**Size markers:** a mixture of DNA fragments of known sizes, used for determining the size of other DNA fragments separated during gel electrophoresis.

**Sonication:** the use of high-pitched sound waves to break DNA into smaller fragments.

**Sterile technique:** the controlled procedures used to maintain sterile, or uncontaminated, conditions when performing experiments.

**Supernatant:** the liquid fraction of a sample after centrifugation or precipitation of insoluble solids.

**Taq DNA polymerase:** DNA polymerase from the organism *Thermus aquaticus*, that resists denaturation at high temperatures and is able to synthesize DNA at temperatures up to 72°C.

**Thermocycler:** an instrument used to perform polymerase chain reactions by cycling through three different temperatures. The PCR-based sequencing reactions used at the Genome Sequencing Center use 96°C to denature the DNA, 50°C to allow the primers to anneal, and 60°C to allow the *Taq* DNA polymerase to synthesize the new strand. Thermocyclers are also known as a PCR machines.

**Trace file:** a computer file of sequenced DNA data that is stored on a server where it can be accessed for analysis.

**Vector (or cloning vector):** a double-stranded circular DNA molecule that has the ability to carry foreign DNA sequences into a host cell. Once in the host cell, the vector can be replicated, or copied, and thus the foreign gene sequences carried in the vector will also be replicated. Basically, vectors are used to move specific gene sequences from one organism into another organism.

**Vortex:** an apparatus used to swirl or mix liquid solutions at very high speeds.

**Yeast artificial chromosome (YAC):** a vector able to accept a large insert of foreign DNA (~ 500,000 bp) and that can be propagated as a chromosome in yeast. In addition to the cloning site, the YAC contains selectable marker genes, as well as the yeast *ars* (autonomous replication sequence, or origin of replication), a centromere, and two telomeres.

## References

Elgin, Sarah C.R. *Biology 280: DNA Science for Non-Majors Glossary*. 2003.

Gonick, Larry and Mark Wheelis. *The Cartoon Guide to Genetics*, updated edition. 1991. Harper Collins Publishers. New York, NY.

Griffiths, A.J.F., Miller, J.H., Suzuki, D.T., Lewontin, R.C., and Gelbart, W.M. *An Introduction to Genetic Analysis*, 7th edition. 2000. W.H. Freeman and Company. New York, NY.

Lewin, B. *Genes VIII*. 2000. Oxford University Press. New York, NY.

Micklos, D.A., Freyer, G.A., Crotty, D.A. *DNA Science: A First Course*. 2003. Cold Spring Harbor Press. Cold Spring Harbor, NY.

Nill, Kimball. *Glossary of Biotechnology Terms*, 3rd edition. 2002. CRC Press. Boca Raton, FL.

Primrose, S.B. *Principles of Genome Analysis: A Guide to Mapping and Sequencing DNA from Different Organisms*. 1998. Blackwell Publishing. Malden, MA.

Steinberg, Mark L. and Sharon D. Cosloy. *Facts on File Dictionary of Biotechnology and Genetic Engineering*. 2000. Facts On File Inc. New York, NY.

