Our knowledge of the diversity and physiology of soil microorganisms has vastly improved since we have gained a better understanding of microbial genetics. Genetics is classically defined as the study of the heredity of individuals. Basically this means the study of why and how offspring are very similar to their parents but not identical. Therefore, the driving forces of genetics are the processes that conserve genetic information and those that contribute to changes. The beginning of the field of genetics would be attributed by most to the work of Gregor Mendel and Charles Darwin in the mid-1800s (Brock, 1990). The field did not really explode until the 20th century when the basis of genetics (DNA structure, replication, etc.) was discovered and the implementation of molecular genetic techniques occurred. This new century began in the era of genomics (the determination of the entire genome of organisms), which has made us realize we still have much more to learn and understand. There are now many fields, including soil science, that utilize the principals of genetics. In addition to the fundamental knowledge gained many methods have been developed from our knowledge of genetics that now allows us to study soil organisms from the level of whole communities to single cells. This chapter will cover the fundamental aspects of genetics and other chapters will demonstrate the implementation of this knowledge to different aspects of soil microbiology.

Nucleic Acids

Nucleic acids are the basis of inheritance and are found in all living organisms. The information is encoded in the sequence of nucleic acids, genotype, which gives rise to the traits expressed, phenotype, by organisms. There are two main categories of nucleic acids, DNA, deoxyribonucleic acid, and RNA, ribonucleic acid. The basic chemical structure of all nucleic acids is similar (Fig. 1). They are composed of nucleotides that consist of a sugar, a phosphate, and a purine or pyrimidine base. There are two purine bases, adenine (A) and guanine (G); and three pyrimidines, cytosine (C), thymine (T, only in DNA) and uracil (U, only in RNA). The subsequent product derived from the genetic information is determined by the order or sequence of different nucleotide bases joined together by phosphodiester bonds (a type of covalent bond). Laboratory methods have been developed to determine the sequence of nucleotides. Using these methods the entire genome (all the DNA) of many microorganisms have now been determined.

DNA is typically composed of two or double strands (ds) of nucleotides with a deoxyribonucleic acid as the sugar component and the bases A, C, G, or T. The various bonds of the two strands cause them to twist together forming a double helix. The two strands are joined by hydrogen bonds that form between complementary nucleotide bases (Fig. 1). The cytosine always complements (or pairs with) guanine due to the formation of three hydrogen bonds and the adenine complements thymine due to the formation of two hydrogen bonds. This specific bonding results in the maintenance of the nucleotide sequence during DNA replication. The weak nature of hydrogen bonds and nucleotide complementation are also important because many techniques used for molecular genetic analysis are possible because of these properties. Denaturation of DNA occurs when the weak hydrogen bonds are disrupted resulting in the separation of the two strands. Commonly used ways of disrupting these bonds in the laboratory include heating, urea or guanidine thiocyanate. Once the DNA has been denatured, the complementary strands can be separated with a process called hybridization.
is by applying heat or exposure to alkaline pHs. **Renaturation** or **reannealing** is the process in which the complementary strands are reformed by the formation of hydrogen bonds. For example, DNA that has been denatured by heating can be renatured by lowering the temperature.

### Nucleic Acid Hybridization

The most accurate and easily used methods to determine species evenness or abundance in a community without the inclusion of biases introduced using PCR are based on nucleic acid hybridization (Nakatsu and Forney, 1996). It is commonly accomplished by hybridization (or annealing) of a previously characterized nucleic acid sequence (probe) to its counterpart (target) within a mixture of DNA or RNA. Probe sequences are known that can differentiate community members at phylogenetic levels ranging from broad (all organisms, kingdom, classes) to narrow (species or strain). The intensity of hybridization signal gives an indication of the evenness of that group. In addition to species evenness these methods can be used to assess physical distribution by examining intact in cells using fluorescent in situ hybridization (FISH) (Amann et al., 2001). As we gain a better understanding of the diversity of microbial communities more informative methods based on hybridization such as microarrays are being developed to expand information obtained to include species richness and function. Microarrays are basically large-scale hybridization experiments in which thousands of targets can be probed simultaneously on a small platform in the size range of a microscope slide.

RNA is a single strand (ss) of nucleotides with a ribonucleic acid as the sugar component and the bases A, C, G, or U. There are three main types of RNA that serve different roles in the cell: messenger RNA (mRNA), transfer RNA (tRNA) and ribosomal RNA (rRNA). Messenger RNA is made from transcription of DNA and directs the order that amino acids are assembled in the ribosome during translation of the protein. Transfer RNA is the adapter molecule responsible for matching the 3 base-code on the mRNA to its specific amino acid. Ribosomal RNA is involved in the structure and catalytic activity of components of the ribosome. The ribosome is composed of two subunits, often referred to as small and large or by their size (determined by their sedimentation rate S). In procaryotes the small subunit contains 16S rRNA whereas in eukaryotes the subunit is slightly larger and contains 18S rRNA. The large subunit in prokaryotes contain both 23S and 5S rRNA and in eukaryotes they are 28S, 5.8S and 5S in size. In addition to the importance and necessity of rRNA genes for cellular function, the rRNA gene sequences have contributed to our understanding of microorganisms. Comparisons of nucleotide sequences have shown there are regions of rRNA sequences that are highly conserved between all organisms and other regions that vary to different degrees. The variability in these regions increases as the evolutionary distance between two organisms increases. In a very short period of time tremendously important information of microorganisms has been obtained based on these sequences.

### Replication, transcription and translation

A new strand of DNA, which is the exact duplicate of the parent molecule, is synthesized in a process called replication (Fig. 2). The exact sequence is maintained because of the basepair specificity discussed above. There are a number of different proteins required for this entire process but most of them will not be covered here. Each strand of DNA serves as a template and they replicate at the same time. The initiation of DNA replication requires the presence of a primer as a place for the nucleotides to be added. The nucleotides are added to the growing strand using an enzyme called DNA polymerase. DNA synthesis occurs only in one
direction because the polymerase only forms bonds between the 5' triphosphates and the free 3'-OH group in the sugar. Since replication is unidirectional the opposite strand must replicate in a discontinuous fashion and is referred to as the lagging strand. Knowledge of the basic process of replication was used to develop techniques such as the polymerase chain reaction (PCR).

The RNA can also be synthesized by using DNA as a template in a process called transcription (Fig. 2). The process differs from replication because the entire DNA strand is not transcribed into RNA and is instead limited to regions coding for a single gene or a group of associated genes called an operon. A gene is defined as the basic unit of inheritance, a sequence of nucleotides that codes for a single polypeptide (protein). The nucleotides are added to the strand by the enzyme RNA polymerase starting at a specific sequences before a gene called the promoter region and ending at a termination sequence or protein. The mRNA is the template used for protein synthesis, a process called translation (Fig. 2). Translation occurs at the ribosomes, structures composed of proteins and rRNA. The tRNA reads the nucleotide sequence of the mRNA in sequential groups of three called codons and brings in the correct amino acid to the ribosome. Of the 64 possible permutations of the four nucleotides in groups of three, 61 codons correspond to 20 amino acids and three to stop codons. The redundancy in the genetic code means that one amino acid can be represented by more that one codon. The amino acid typically indicating the start of a coding region is methionine (AUG) and on occasion valine.

**ribosomal RNA (rRNA).** Many direct and indirect analyses of communities and individuals in soils involve the use of ribosomal RNA (rRNA) sequences. The ribosome is composed of two subunits, referred to as small and large rRNA. In prokaryotes the small subunit contains 16S rRNA and the large subunit contains both 23S and 5S rRNA. In eukaryotes the small subunit is slightly larger and contains 18S rRNA and the large subunit has rRNA 28S, 5.8S and 5S in size. Comparison of rRNA gene nucleotide sequences has indicated that they are phylogenetically informative because some regions have been retained over time to conserve function but at the same time other regions of sequence have changed as the organisms evolved (Woese et al., 1983). The variability in these regions increases with evolutionary distance between two organisms. The 16S rRNA gene sequence has been used most extensively to study prokaryotic microorganisms. The intergenic region, which is the sequences between the genes, is often included in analyses of eukaryotic organisms and to some extent prokaryotes. The rRNA or rDNA can be isolated from soil and examined directly or after amplification using polymerase chain reaction (PCR).
Organization of Nucleic Acids in Cells

In prokaryotes, DNA is organized into either genetic elements called **chromosomes** or **plasmids**. Both are typically circular, supercoiled in order to fit into the cell, and self-replicating. The chromosome forms a compact unit called a **nucleoid** and carries all the genetic information required for the survival and maintenance of the cell. Plasmids are usually smaller than chromosomes. They have been studied extensively because they can carry genetic information that can be invaluable, such as, genes for antibiotic resistance, catabolic functions (e.g., degradation of pesticides), and metal tolerance. Plasmids have also become an invaluable tool used in molecular genetic methods. The small size and ability to self-replicate, has made it possible to use them as **vectors**, a means of transferring genetic information from one host to another. In eukaryotes, DNA is contained in a membrane bound structure called the **nucleus** and is also found in cell organelles called mitochondrion and chloroplast. In the nucleus DNA are organized as linear chromosomes that are held in formation by proteins. The DNA in mitochondria and chloroplasts are similar to those in prokaryotes both in organization and in the housekeeping genes encoded.

In prokaryotes the RNA is not contained within a cellular structure. This makes it possible for transcription and translation to occur in close association. In contrast the mRNA in eukaryotes is made in the nucleus then transported out for translations to occur.

Genetic Mutations

Mutations are the driving force of evolution. They are caused by a change in genotype compared to the wild type organism. The wild type strain is arbitrarily chosen; usually the strain that is first isolated and expresses a phenotype of interest. Mutations are random events, that is, the site of mutation in a gene cannot be predicted nor do they result as a response to selection conditions. Any change in the nucleotide sequence of DNA is considered to be mutations but not all mutations can be observed by a change in phenotype. Genetic changes underlying mutations can be point mutations (changes in a single nucleotide) or larger genetic rearrangements caused by duplications, deletions, and/or inversions of a length of sequence. Even a single nucleotide change can alter a single amino acid that may be responsible for the folding or functional features of a protein, thus making it non-functional. Other mutations will cause the protein to be only partially disruptive or have no detectable affect on the gene product. The maintenance of a genetic mutation is dependent on **natural selection**. Natural selection, survival of the fittest, occurs on traits that give the organism a competitive advantage over other organisms that share the same habitat.

Gene Transfer in the Environment

There are three basic mechanisms for gene transfer in the environment, transconjugation, transformation and transduction (Fig. 3) (Davison, 1999; Droge, 1999). They, especially transconjugation, are the closest equivalent to sexual reproduction that occurs between eukaryotic organisms. However, unlike sexual reproduction in eukaryotes, gene transfer between bacteria is often not limited to cells belonging to the same species. Genes can be transferred between phylogenetically different bacteria and even from bacteria to plants. The success of transfer of genetic information diverse hosts is limited by its incorporation into the genome or the ability of the plasmid to replicate in alternate hosts. All these mechanisms occur
naturally but they have also been used extensively in the laboratory to study basic and applied aspects of microbiology.

**Transconjugation** (also called conjugation or mating) is the transfer of genetic material between bacteria via cell-to-cell contact. The cell giving the genetic material is called the donor and the one receiving is the recipient. The genetic element typically transferred is a plasmid. Before transfer, in some cases, a quantity of chromosomal information can be incorporated into the plasmid. After transfer the plasmid may replicate independently or the whole plasmid or a portion of it may integrate into the chromosome of the new host. The frequency of transfer and the compatibility between individuals is donor and recipient dependent.

**Transformation** is the uptake of DNA into the cell without contact with another living cell. The DNA is extracellular and found in the environment as a result of cell death and lysis. Once the DNA is in the cell it must be incorporated into the genome unless it carries the genetic information for self-replication. Some but not all species of bacteria are naturally transformable. In the laboratory cells can be chemically or electrically modified for transformation. In this case the information to be transferred is usually carried on a plasmid.

**Transduction** is the phage (virus, see chapter 7) mediated transfer of bacterial DNA between two hosts. To infect a new host a virus packages its nucleic acids into a phage particle. If the viral DNA was incorporated into the bacterial DNA then during excision some of the host genome may be included. Therefore, when the virus infects a new host some bacterial DNA from the previous host is transferred and can become part of its genome.

For gene maintenance after transfer into a new host it must become integrated into the genome especially if it is not introduced on a plasmid or if the plasmid cannot self-replicate in the new host. Integration can occur by **recombination** or **transposition**. Integration by homologous recombination occurs between identical sequences of DNA in the newly acquired DNA and in the host genome. Elements that move by transposition are called **transposons**. There are many families of transposons and transposon-like elements. In general, transposons carry genes that allow it and any associated genes to be inserted into random sites in the genome. Both these events are not limited to the integration of exogenous DNA, they can also occur within the existing genome causing rearrangements of the genes that can affect their subsequent expression. They are an important source of genetic variation.

**Fundamental of the Polymerase Chain Reaction**

The polymerase chain reaction (PCR) method mechanizes the steps required for nucleic acid replication resulting in the amplification of a specific nucleotide sequence by about $10^6$ fold. This method is extremely useful when the sequence of interest: 1) has a low copy number, 2) presence is unknown, 3) location in the genome is unknown, or 4) is present in an organism that cannot be cultivated. This technology has many uses and one that has been used extensively in soil microbiology is for the characterization of microbial communities by the amplification of rRNA gene sequences. Similar to all other procedures based on nucleic acid replication the major components of a PCR are: template (DNA or RNA), all four nucleotides, primers, polymerase (either RNA or DNA depending on the template being used) and proper solution conditions (e.g., pH, buffer, salts). A specific or non-specific sequence of nucleic acid can be amplified by PCR depending on the primers that are chosen. Primers used for PCR are oligonucleotides (a short sequence of nucleotides) that have been synthesized in a laboratory. The primers must be able to anneal to complementary regions of opposite strands of the DNA molecule. The sequence can be based on a known target, typically a gene of interest. However,
sequences for unknown target areas can be used if you wish to amplify random regions of the genome for genetic fingerprinting. The polymerase use for PCR must remain active at elevated temperatures therefore they have been obtained from microorganisms living under extreme thermal conditions (e.g., hot springs or marine thermal vents).

The sequence of steps in PCR is controlled by changes to the temperature of the reaction mix, which determines the nucleic acid form (either double or single stranded) and polymerase enzyme activity (Fig. 4). There are basically three steps in the process. (1) The temperature of the solution is raised (usually 94 or 95°C) in order to denature the two strands of DNA making them single stranded. (2) The temperature of the solution is then lowered to cause the primer to anneal to the single stranded DNA. There is excess amount of primers in the solution and they are shorter therefore they preferentially anneal to the template DNA strands. The temperature must be such that specific annealing of the primer to template occurs (typical range is 50°C to 65°C). (3) The temperature is then raised to a temperature optimal for activity of the thermal tolerant polymerase (best from 75-80°C). During this stage the polymerase adds nucleotides to each strand starting from the primers. This process is repeated for 25-30 cycles resulting in the exponential amplification of the target sequence with each cycle.

In the first cycle random lengths of the template are amplified starting from each primer. In cycle 2 random lengths of the template are amplified in addition to the strands that were amplified in the first round. In cycle 3 the original template, random lengths of strands amplified in the first round in addition to the amplification products of cycle 2 that are the length between the two primers are amplified. The majority of strands are now being synthesized strands that are the length between the two primers. As the number of cycles continue the desired amplification product preferentially accumulates. The number of cycles used depends on the amount of PCR product desired and the life of the polymerase under the conditions used.

There are some factors to keep in mind when using any PCR-based methods. The primers chosen for PCR is based on sequence information available therefore there is a bias towards these known sequences. Other artifacts include: the formation of heteroduplexes, chimeric molecules, or sequence heterogeneities due to inaccurate polymerization. Nevertheless, the benefits of using PCR in soil microbiology presently far outweigh any limitations.

**Nucleotide Sequencing:**

There are a number of different methods used for determining nucleotide sequences. Presently, the more commonly used method is based on the dideoxynucleotide method developed by Sanger. First a primer is annealed to single stranded of DNA and then the complementary DNA strand is synthesized using DNA polymerase. In addition to the standard nucleotide bases, analogs called dideoxynucleoside triphosphates (ddNTP) are added to the mix. The ddNTP lack the 3’-OH group on the ribose moiety, making it impossible for a phosphodiester bond to form, thus terminating the addition of subsequent nucleotides. Four separate reactions are carried out, each containing all the dNTP’s but only one of the dideoxy analogs causing all the reactions in a tube to terminate at the same base. The incorporation of the ddNTP will occur randomly therefore different lengths of extension occur. The reactions are labeled using a fluorescent or radioactive molecule (e.g., α 32S-dATP) therefore the synthesized fragments can be separated on a polyacrylamide gel or capillary column and their sizes (i.e. nucleotide sequence) visualized either manually or with computer assistance.
**Nucleotide sequence analysis of ribosomal RNA genes.** Determination and comparison of nucleotide sequences of the rRNA genes based on nucleic acids extracted directly from soils has revealed the presence of many microorganisms important in soil functions that were previously unknown (Barns et al., 1999; Borneman and Triplett, 1997). First, the sequence is usually amplified by PCR using gene specific primers. After sequence determination comparisons to known sequences can be made using algorithms and databases available on the internet. Examples of two commonly used databanks are GenBank housed at the National Center for Biotechnology Information ([http://www.ncbi.nlm.nih.gov/](http://www.ncbi.nlm.nih.gov/)) (Wheeler et al., 2001) and the Ribosomal Database Project ([http://rdp.cme.msu.edu/html/](http://rdp.cme.msu.edu/html/)) (Maidak et al., 2001).

**Soil Microbial Community Analysis**

Recent applications of molecular genetics have provided tools to determine microbial presence and diversity in soil. These methods have enabled investigators to studying soil microbial communities and populations from a wide range of habitats and geographic locations. For example, using a molecular genetic approach over 4000 different bacterial genomes were estimated to be present in a forest soil compared to less than a hundred that would be obtained using a traditional cultivation technique (Torsvik et al., 2002). The high diversity of microorganisms in most soils has been demonstrated using molecular genetic based techniques that are independent of cultivation. Application of these methods has provided insight into microbial composition, richness, and evenness in soils. Some molecular genetic techniques that have been used are: total DNA isolation and characterization; DNA separation based on G + C content, rRNA gene sequence determination; PCR amplification of rDNA or rRNA; PCR amplification of functional genes; and in situ hybridization of fluorescently labeled rRNA oligonucleotide probes (Akkermans et al., 1995). Cultivation is still used when one wishes to conduct detailed genetic or physiological studies of a microorganism but researchers estimate that only 1% of the bacteria species have been cultivated (Torsvik et al., 2002). Alternatively, differences in microbial community structure can be determined by comparing genetic profiles or fingerprints based on differences in rDNA sequences. Nucleic acids can be extracted directly from soils employing a number of methods or kits specifically designed to minimize contamination of samples by organic acids while maximizing the cell types that are disrupted. Typically the rRNA genes from a community are first amplified by PCR then the mixed products are separated by gel electrophoresis. Differences between communities are determined by assessing the number of common bands found between genetic profiles. The number of fragments in a genetic fingerprint profile provides a crude indication of the community richness.

**Genetic fingerprinting of soil communities.** There are three methods that are being used to distinguish between PCR products of different populations present in a community. The traditional electrophoretic approach is to separate nucleic acids by size by applying an electrical current through a gel matrix (e.g., agarose or polyacrylamide) containing PCR products. The differences in PCR amplicon sizes of a rRNA gene is not significant therefore Ribosomal Intergenic Spacer Analysis (RISA) method is used in which the PCR product include the intergenic region between rRNA genes (Ranjard et al., 2001). Another approach that differentiates populations based on fragment sizes is Terminal-Restriction Fragment Length Polymorphism (T-RFLP). In this method PCR products are separated by size after restriction enzyme digestion of the PCR products (Liu et al., 1997). In order to differentiate the terminal ends each primer is labeled with a different colored fluorescent marker that can be recognized by
a scanner. In both methods the number and size of each fragment can be analyzed using an automatic sequencing system. The third method separates PCR amplicons based on their sequence composition instead of fragment size (Muyzer et al., 1996). A chemical gradient of the gel matrix (polyacrylamide) is made for denaturing gradient gel electrophoresis (DGGE) or a thermal gradient is applied (TGGE) (Muyzer, 1999). This method can separate community PCR products because of the different melting characteristics of DNA with different sequences. As explained earlier more energy is required to denature double stranded DNA with higher G+C content. Therefore the extent that strands separate under different denaturing conditions is dependent on their nucleotide sequence. As the double strands separate their migration becomes more impeded in the gel matrix. Although all three methods use different targets or means of separation they all provide a fingerprint pattern, which characterizes the community. Bands represent populations that are present in that community thus making it possible to assess the similarity between different communities.

![DGGE fingerprint pattern](image)

An example of DGGE fingerprint patterns illustrating the change in rhizosphere communities during early corn development is illustrated. Stages represented are germination (G), emergence (E), first leaf (V1), second leaf (V2), third leaf (V3) and sixth leaf (V6). Variation in band intensity can be seen. Examples of some bands common to all or almost all growth stages (<) and unique bands (*) are indicated.

**Summary**

The foundation of genetic analysis presented in this chapter has been the cornerstone for expanding our knowledge of soil microorganisms. The basic concepts covered: biochemistry of nucleic acids; process of replication, transcription and translation; organization of nucleic acids in cells; genetic mutations; and mechanisms of gene transfer have allowed scientists to develop
tools such as the polymerase chain reaction (PCR), nucleotide sequencing, cloning and hybridization. These tools have then been employed to expand our knowledge of soil microbial community analysis. Undoubtedly there is a wealth of knowledge buried in the soil ecosystem that is now becoming available to us by using these tools. Other chapters examine in greater detail microbial diversity and dynamics in the soil community that was not accessible to us in the past. The soil environment is far more diverse than ever expected and groups important for soil functions are now being discovered.

Cited References


**General References**


Molecular phylogeny

[http://genome.uc.edu/genome/HelpPages/seq-anal-tutorial/molecular_phylogeny.html](http://genome.uc.edu/genome/HelpPages/seq-anal-tutorial/molecular_phylogeny.html)