Restriction Endonucleases

Contents

Introduction

Restriction endonucleases are enzymes that cleave the sugar-phosphate backbone of DNA strands. The vast majority of these enzymes have been isolated from bacteria, where they carry out a host-defense function for the cell. These enzymes recognize a specific DNA base sequence and cleave both strands of a double-stranded DNA molecule at or near the recognition site. All restriction enzymes fall into one of three classes, based upon their molecular structure and need for specific co-factors. Class I endonucleases have a molecular weight around 300,000 Daltons, are composed of non-identical sub-units, and require Mg^{2+}, ATP (adenosine triphosphate), and SAM (S-adenosyl-methionine) as cofactors for activity. Class II enzymes are much smaller, with molecular weights in the range of 20,000 to 100,000 Daltons. They have identical sub-units and require only Mg^{2+} as a cofactor [1]. The Class III enzyme is a large molecule, with a molecular weight of around 200,000 Daltons, composed of non-identical sub-units. These enzymes differ from enzymes of the other two classes in that they require both Mg^{2+} and ATP but not SAM as co-factors. Class III endonucleases are the rarest of the three types.

The Discovery of Restriction Endonucleases

Prior to 1968 the existence of restriction enzymes was unknown. However, the phenomenon of restriction was well known. Restriction was the term given to the ability of bacteria to recognize and attack any foreign DNA source whether it came from a virus or from another strain of bacteria. In 1968 Matthew Meselson and Robert Yuan reported that they had identified an enzyme in the bacterium *Escherichia coli*, strain K-12, that appeared to be able to recognize and digest foreign DNAs [2]. This enzyme, they concluded, could be the agent responsible for restriction. They coined the term **restriction endonuclease** to refer to this enzyme. They further determined that such

enzymes would be ubiquitous among bacteria and that they would recognize and digest any double-stranded DNA that was not protected by a specific pattern of DNA base methylation [2]. Methylation of DNA involves adding a methyl-group (CH_3) to the DNA base such that the restriction enzyme will not recognize it. The process of methylation has been shown to be carried out by DNA sequence-specific methyltransferase enzymes [3, 4]. In plants and animals the primary methylated base is 5-methylcytosine (m5C) while in bacteria the major methylated base is N^6-methyladenine (mA) but N^4-methylcytosine (mC) is also found. Examples of these methylated DNA bases are shown in Figure 1. The restriction endonucleases found by Meselson and Yuan in *E. coli* required the presence of Mg^{2+}, SAM, and ATP for it to carry out its function. Thus, the first restriction enzyme to be identified was a Class I enzyme.

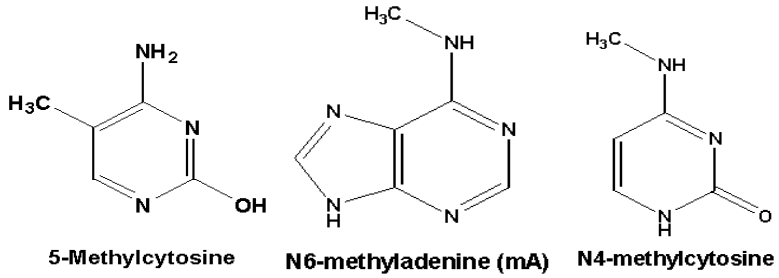

Figure 1. Structures of the three primary methylated DNA bases in prokaryotes and eukaryotes.

The report by Meselson and Yuan was quickly followed by two papers describing a similar enzyme in the bacterium *Haemophilus influenzae*, strain Rd [5, 6]. Like the *E. coli* enzyme, the *H. influenzae* endonucleases was inactive in the presence of native DNA but did recognize and digest foreign DNAs. Unlike the *E. coli* enzyme, however, the *H. influenzae* endonuclease only required the presence of Mg^{2+} for activity. The cleavage pattern of both enzymes was limited and consistently reproducible, suggesting that there was a specific DNA sequence that was recognized by the enzymes and that the enzyme would bind to this sequence prior to cleavage [6]. In the companion paper, Kelly and Smith offered evidence that the recognition site of their enzyme was a run of six specific nucleotides in the form,

5'… G T Py | Pu A C … 3'
3'… C A Pu | Py T G … 5'

where Py refers to either pyrimidine (T or C), Pu refers to either purine (A or G), and the vertical line indicates the cleavage site of the enzyme. Note that the symmetry of this recognition sequence is in the form of a **palindrome**, a nucleotide sequence in which the 5' to 3' sequence of one strand of a segment of DNA is the same as that of its complementary strand. This feature did not escape notice, "It is unlikely that the symmetry of this sequence is fortuitous, since the number of possible asymmetric sequences of this type

is about 30 times the number of possible symmetric sequences…" [5]. They concluded that symmetry in the recognition sequence was a basic feature of the action of the enzyme.

Properties of Restriction Enzymes

Within a few years the basic outlines of restriction endonuclease action were understood. The seminal papers of 1968 and 1970 opened the floodgates on isolation and identification of restriction enzymes in bacteria. In 1975, Nathans and Smith were able to review the state of knowledge and present a consistent, standard nomenclature for the rapidly growing numbers of known enzymes. The name given to each new enzyme would convey both the genus and the species of the bacterium from which it was isolated, the strain number, and the order in series in which the enzyme was found. Thus, the restriction enzyme designated *Bam HI* was the first enzyme found in the bacterium *Bacillus amyloliquifaciens*, strain H while the restriction enzyme *Hae III* was the third enzyme found in *Haemophilus aegyptius*. A list of some of the thousands of currently known restriction endonucleases is presented in Table 1. As the list of restriction enzymes grew and their recognition sequences were identified, it was found in some cases that more than one enzyme could recognize the same sequence. RJ Roberts conferred the term **isoschizomer** (same cutter) on restriction enzymes that recognized the same DNA sequence [1].

The search for new and unusual restriction enzymes continued apace so that, by 1982, a list of 357 identified restriction enzymes recognizing 90 different DNA sequences was published [7]. Most restriction enzyme recognition sequences are from four to eight bases long and most are palindromic (Table 1). Additional diversity was found among the isoschizomers. For example, the enzymes *Sma I* and *Xma I* both recognize the six base sequence CCCGGG but give different fragments with the former cutting CCC|GGG and the latter cutting C|CCGGG. Similarly, the isoschizomeric pair *Hha I* and *Hin PI* both recognize the sequence GCGC but the former cuts GCG|C and the latter G|CGC. Further differences were found in relation to sensitivity to methylation. Both *Mbo I* and *Sau 3A* cut GA|TC but when the sequence is methylated as GA*TC, *Mbo I* fails to cut while *Sau 3A* is not affected. Conversely, in the case GATC*, the situation is reversed. This phenomenon was put to good use in the case of the restriction enzyme pair *Hpa II* and *Msp I*. Both enzymes recognize the sequence CCGG but when methylated as CC*GG, *Mbo I* cuts the sequence and *Hpa II* does not. This pair of enzymes has proved to be extremely useful in identification of the so-called cpG "islands" that lie near protein coding genes [8, 9].

Table 1
Examples of Restriction Enzymes

Enzyme	Microorganism	Recognition Sequence	Isoschizomers
Alu I	Arthrobacter luteus	AG|CT	
Apa I	Acetobacter pasteurianus	GGGCC|C	Bsp120 I, PspOM I
Bam HI	Bacillus amiloliquifaciens	G|GATCC	
Bgl II	Bacillus globigii	A|GATCT	
Cla I	Caryophanon latum L	AT|CGAT	Bsp DI, Bsc I, BspX I
Dde I	Desulfovibrio desulfuricans	C|TNAG	BstDE I
Dra I	Deinococcus radiophilus	TTT|AAA	
Eco RI	Escherichia coli RY13	G|AATTC	
Eco RV	Escherichia coli J62	GAT|ATC	Eco32 I
Fnu4H I	Fusobacterium nucleatum 4H	GC|NGC	Fsp4H I, Ita I
Hae III	Haemophilus aegyptius	GG|CC	Bsh I, BsuR I, Pal I
Hind II	Haemophilus influenzae Rd	A|AGCTT	
Hinf I	Haemophilus influenzae Rf	G|ANTC	
Kpn I	Klebsiella pnumoniae OK8	GGTAC|C	Acc65 I, Asp718 I
Mbo I	Moraxella bovis	|GATC	Dpn II, Nde II, Sau3A I
Msp I	Morazella sp.	C|CGG	BsiS I, Hap II, Hpa II
Nde I	Neisseria dentrificans	CA|TATG	FauND I
Not I	Nocardia otitidis-caviarum	GC|GGCCGC	CciN I
Pst I	Providencia stuartii 164	CTGCA|G	
Pvu II	Proteus vulgaris	CAG|CTG	
Rsa I	Rhodopseudomonas sphaeroides	GT|AC	
Sma I	Serratio marcescens S	CCC|GGG	Cfr9 I, Psp A I, Xma I
Taq I	Thermus aquaticus YT1	T|CGA	TtaHB8 I
Xba I	Xanthomonas badrii	T|CTAGA	
Xho I	Xanthomonas holcicola	C|TCGAG	PaeR7 I, Sfr274 I, Tli I

As can be seen in Table 1, most restriction enzymes recognize CG-rich DNA sequences but some AT-only recognition sequences are known. The enzymes *Dra I* (TTTAAA), *Ssp I* (AATATT), and *Pac I* (TTAATTAA) are but three of these "AT-cutters." Further, most restriction enzymes will cleave the DNA inside the recognition site but there are several that do not. The enzyme *Mnl I* recognizes the non-palindromic sequence CCTC but cleaves the DNA seven bases downstream (i.e., CCTC 7/7). Other examples include *Bbv I* (GCAGC 8/12) and *Hga I* (GACGC 5/10). Another variation on the basic theme of restriction enzyme recognition sites are the so-called "degenerate" sequences. The enzyme *Acc I* recognizes the sequences GTATAC and GTCGAC. This can also be written as GTMKAC, referring to the standard nomenclature for degenerate sites (see Supplemental Materials). The five-base cutter *Hinf I* recognizes four different sequences denoted GANTC (again, see Supplemental Materials). This last form of recognition sequence degeneracy can lead to come very long recognition sites. The enzyme *Hgi EII* , for example, recognizes the sequence $ACCN_6GGT$ and the enzyme *Xcm I* recognizes the sequence $CCAN_9TGG$. Perhaps one of the longest recognition and activity sequences belongs to the enzyme *Bpl I* with (8/13)$GAGN_5CTC$(13/8) which is an eleven base recognition site plus cleavage sites a further 8 bases upstream and 13 bases downstream. Finally, the term **non-palindromic** was noted above. There are a number of enzymes that recognize DNA sequences of various lengths that do not form palindromes. Among these are *AarI* (GACCTGC), *Bsp MI* (ACCTGC (4/8)), *Fok I* (GGATG (9/13), and *Mbo II* (GAAGA (8/7)).

Mechanism of Action of Restriction Enzymes

The action of restriction enzymes is in many respects as varied as the enzymes themselves. In general, however, the process is one of recognition of the binding site, binding of the enzyme dimer to the DNA, cleavage of the DNA , and enzyme release (Figure 2). Pingoud and Jeltsch note that this scheme is a minimal scheme due to the complex variations in enzyme action that have been observed [10]. To begin, all restriction endonucleases will bind DNA specifically and, with much less strength, non-specifically. This is a characteristic of many proteins that interact with DNA. It is probable that even non-specific DNA binding will induce a conformational change in the restriction enzyme dimer that will result in the protein adapting to the surface of the DNA strands [11]. These changes are not the same as those that occur when the dimer binds to the recognition site though. As the dimer slides along the DNA strands, it searches for recognition elements and, when these are encountered, an interaction between the protein and the DNA ensues in which the non-specific complex is converted into a specific complex. This requires significant conformational changes in both the protein and the DNA as well as expulsion of water molecules from the protein/DNA interface so that more intimate contacts can be established [12]. In general, intimate contact is held by 15 – 20 hydrogen bonds that form between the protein and the DNA bases in the recognition site. These bonds are shown to be mediated through specific amino acids, primarily ASP and GLU, held in a proper three-dimensional configuration. There are differences among restriction enzymes with respect to how much water is expelled but, in all cases, it is a substantially greater amount than is expelled during non-specific binding.

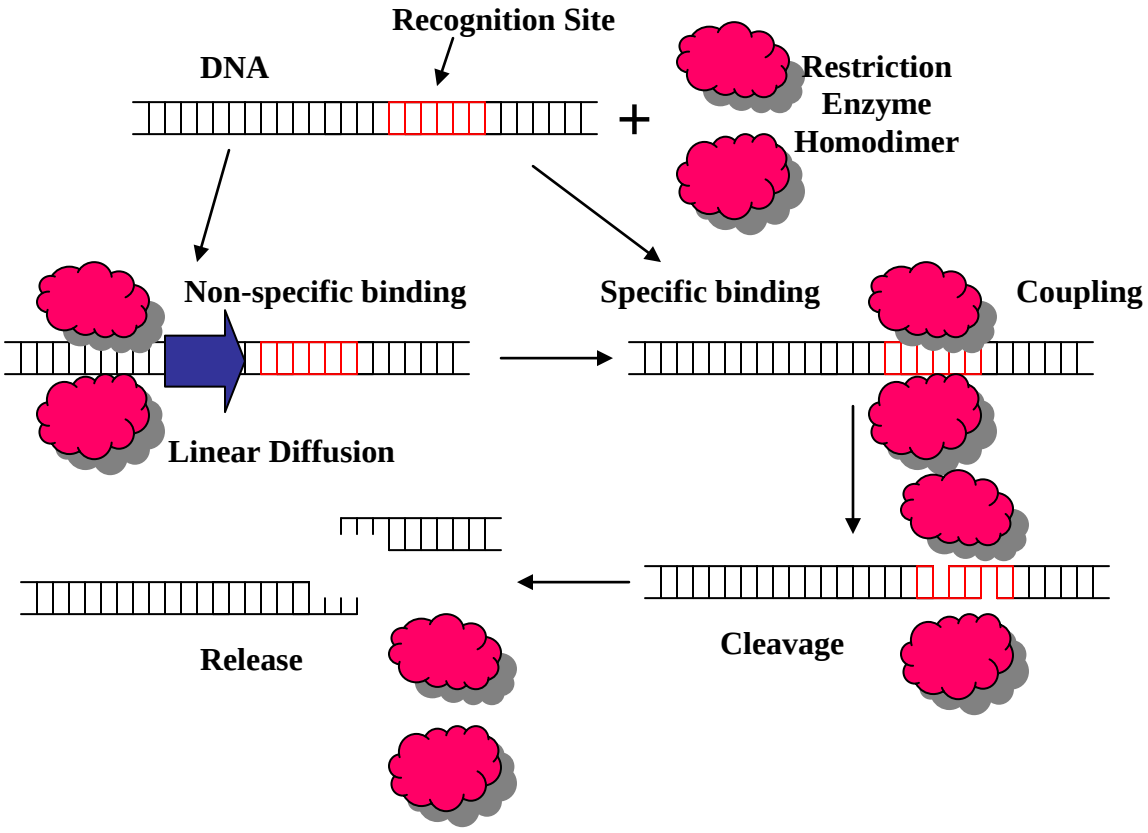

Figure 2. Simplified scheme of the mechanism of Type II restriction enzyme digestion. The homodimer will either bind directly to the recognition site (Specific Binding) or nearby (non-specific Binding). In the case of non-specific binding, if the recognition site is not too far away the enzyme will move along the DNA strand until it hits the recognition site. Once the enzyme locates the recognition site it will couple and then hydrolyze the sugar phosphate bonds of the DNA. Finally, the enzyme will release leaving the cleaved DNA molecule behind.

Once the enzyme is specifically bound to its cognate DNA sequence there are more differences in the cleavage reaction. To begin, some restriction enzymes will bind one magnesium divalent ion whereas some will bind two. Moreover, other metal ions may or may not be present as well. Interestingly, to date, the actual mechanism by which a restriction enzyme cuts the DNA to which it is bound has not been demonstrated. The conventional wisdom holds that hydrolysis mediated by metal ion binding is the paradigm [13].

Molecular Weight Markers and Polymorphisms

Even in the absence of a clear understanding of their mechanism of action, it was apparent from the outset that enzymes producing predictable and reproducible cleavage products in unprotected DNA would be very useful as laboratory reagents. One of the very first applications was the use of restriction enzymes to create reliable DNA molecular weight markers for gel electrophoresis. One of the first of these was the venerable marker produced by the complete digestion of the 48,502 base pair genome of the *E. coli* bacteriophage lambda with the restriction enzyme *Hind III*. Using the specific lambda phage c1857*ind*1 *Sam* 7, *Hind III* digestion yields eight constant restriction fragments. These are 23,130bp (weight 15.00 x 10^6 Daltons); 9,416bp (weight 6.12 x 10^6 Daltons); 6,557bp (weight 4.26 x 10^6 Daltons); 4,361bp (weight 2.83 x 10^6 Daltons); 2,322bp (weight 1.51 x 10^6 Daltons); 2,027bp (weight 1.32 x 10^6 Daltons); 564bp (weight 0.37 x 10^6 Daltons); and 125bp (weight 0.08 x 10^6 Daltons) [14, 15]. Since that time many combinations of viral genomic DNA and restriction enzymes have been used to produce molecular weight markers though only a few have achieved the level of universal acceptance and use that Lambda/*Hind III* has.

Another of the first uses of restriction enzymes as laboratory reagents was to produce a restriction site map of the rabbit β-globin gene [16]. This work led directly to the discovery of the RFLP (restriction fragment length polymorphism). When Jeffreys applied the restriction mapping technique to the human β-globin gene, he discovered that three of the restriction sites in a 300bp segment of the gene were polymorphic (Figure 3). That is, some individuals had the sites and others did not. Recognizing this as a potential new source of information on genetic variation, Jeffreys observed, “There is, however, almost no information on how much variation exists at the level of DNA sequences in man, and on what types of DNA sequence variants might occur in human

populations." [17]. The answer was, as is now known, that there is an enormous reservoir of DNA sequence variation in the human genome.

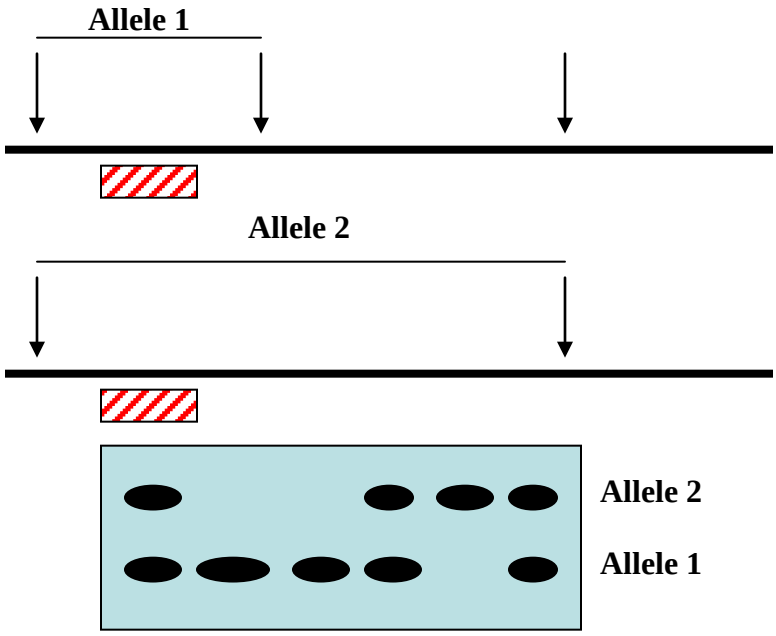

Figure 3. A graphic representation of a classic presence-absence restriction fragment length polymorphism. The probe (cross hatched) is assumed to be a random DNA clone and the wild-type state of the polymorphic restriction enzyme recognition site is unknown since there is no way to show whether the mutation that occurred in the recognition site created it or destroyed it. The "blot" is what might typically have been seen when probing a restriction enzyme digest of DNAs from six unrelated individuals. For information about "blots" see Southern [18].

Soon after the first human RFLP DNA sequence variants were discovered, a revolutionary idea emerged in genetics. If these new DNA sequence variants occurred throughout the human genome and in sufficient quantity, they could be used to make a map of the entire human genome [19]. At the time of this paper there were precious few known RFLPs, all of them were associated with genes, and none of them was particularly informative in the way that would be needed to produce a human genetic map. Undaunted by this fact and by the fact that there was absolutely no evidence that RFLPs existed in the quantity needed nor in the roughly evenly spaced full genome coverage that would be required, Botstein et al. boldly suggested, "The advent of recombinant DNA technology has suggested a theoretically possible way to define an arbitrarily large number of arbitrarily polymorphic marker loci." [19].

The crucial event came at roughly the same time as the Botstein et al. paper. Wyman and White, working with randomly selected clones from a human DNA library (a set of human DNA sequences that, together, cover the entire genome), discovered a clone called pAW101 [20]. This clone is perhaps the most historically important piece of DNA ever found because it simultaneously proved two of the critical assertions of the Botstein et al. paper. First, it was random. The clone was not associated with any

particular gene or gene product as it was picked at random from the human DNA library. Second, it was very polymorphic with eight different restriction fragment lengths observed immediately. Thus, it was possible that highly informative, anonymous DNA markers could be found. By the time of the Ninth International Congress on Human Gene Mapping (HGM9) in 1987, more than 1,000 human RFLPs had been found and by HGM10, just two years later, the number had more than doubled. Also during these years two new types of human DNA sequence polymorphism had been discovered. These were the VNTR (variable number of tandem repeat) polymorphism and the PCR-based microsatellite polymorphism [21, 22]. Together, these markers produced the human genome map that Botstein et al. suggested and served as reference points of the human genome sequence that followed.

References and Resources

*Much of the material in this tutorial was adapted from Devor EJ (1992) Introduction: A brief history of the RFLP. In: Devor EJ (ed.) Molecular Applications in Biological Anthropology. Cambridge: Cambridge University Press, 1–18.

1. Nathans D and Smith HO. (1975) Restriction endonucleases in the analysis and restructuring of DNA molecules. Annual Review of Biochemistry, 44: 273–293.

2. Meselson M and Yuan R. (1968) DNA restriction enzyme from *E. coli*. Nature, 217: 1110–1114.

3. McClelland M. (1981) The effect of sequence specific DNA methylation on restriction endonuclease cleavage. Nucleic Acids Research, 9: 5859–5866.

4. McCLelland M, Nelson M, and Raschke E. (1994) Effect of site-specific modification on restriction endonucleases and DNA modification methyltransferases. Nucleic Acids Research, 22: 3640–3659.

5. Kelly TJ and Smith HO. (1970) A restriction enzyme from *Heomophilus influenzae*. II. Base sequence of the recognition site. Journal of Molecular Biology, 51: 393–409.

6. Smith HO and Wilcox KW. (1970) A restriction enzyme from *Haemophilus influenzae*. I. Purification and general properties. Journal of Molecular Biology, 51: 379–391.

7. Linn SM and Roberts RJ. (1982) Nucleases. Cold Spring Harbor Monograph Series, No. 14. Cold Spring Harbor, New York: Cold Spring harbor Laboratory Press.

8. Bird AP. (1986) CpG-rich islands and the function of DNA methylation. Nature, 321: 209–213.

9. Bird AP. (1989) Two classes of observed frequency for rare-cutter sites in CpG islands. Nucleic Acids Research, 17: 9485.

10. Pingoud A and Jeltsch A. (2001) Structure and function of type II restriction endonucleases. Nucleic Acids Research, 29: 3705–3727.

11. Viadiu H and Aggarwal AK. (2000) Structure of *BamH I* bound to nonspecific DNA: A model for DNA sliding. Molecular Cell, 5: 889–895.

12. Robinson CR and Sligar SG. (1998) Changes in salvation during DNA binding and cleavage are critical to altered specificity of the *EcoR I* endonuclease. Proceedings of the National Academy of Sciences USA, 95: 2186–2191.

13. Galburt EA and Stoddard BL. (2000) Restriction endonucleases: one of these things is not like the others. Nature Structural Biology, 7: 89–91.

14. Daniels DL, de Wet JR, and Blattner FR. (1980) A new map of bacteriophage lambda. Journal of Virology, 33: 390–400.

15. Daniels DL, Schroeder JL, et al. (1983) Appendix II: Complete annotated lambda sequence. In: Hendrix RW, Roberts JW, et al. (eds.) Lambda II. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press. 519–674.

16. Jeffreys AJ and Flavell RA. (1977) A physical map of the DNA regions flanking the rabbit β-globin gene. Cell, 12: 429–439.

17. Jeffreys AJ. (1979) DNA sequence variants in the Gγ, Aγ-, δ, and β-globin genes in man. Cell, 18: 1–10.

18. Southern EM. (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. Journal of Molecular Biology, 98: 503–517.

19. Botstein D, White R, et al. (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. American Journal of Human Genetics, 35: 795–815.

20. Wyman AR and White R. (1980) A highly polymorphic locus in human DNA. Proceedings of the National Academy of Sciences USA, 77: 6754–6758.

21. Nakamura Y, Leppert M, et al. (1987) Variable number of tandem repeat (VNTR) markers for human gene mapping. Science, 235: 1616–1622.

22. Weber JL and May PE. (1989) Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. American Journal of Human Genetics, 44: 388–396.

Animation

http://www.cat.cc.md.us/courses/bio141/lecguide/unit4/genetics/recombination/recombinant/enuc.html

http://www.fhcrc.org/education/hutchlab/lessons/animate/ecorv.html

Supplemental Material: Standard Nucleotide Degeneracy Code

The standard nucleotide coding system is:

A, C, G, T, U
R (G or A) puRine
Y (T or C) pYrimidine
K (G or T) Keto
M (A or C) aMino
S (G or C) Strong
W (A or T) Weak
B (G,T,C) (not A)
D (G,A,T) (not C)
H (A,C,T) (not G)
V (G,C,A) (not T or U)
N (all)