

Research Signpost
37/661 (2), Fort P.O., Trivandrum-695 023, Kerala, India



Dynamical Genetics, 2004: 49-78 ISBN: 81-7736-231-3
Editors: Valerio Parisi, Valeria De Fonzo and Filippo Aluffi-Pentini

3

Evolutionary analyses of genetic recombination

Nicole Lewis-Rogers¹, Keith A. Crandall¹ and David Posada²

¹Department of Microbiology and Molecular Biology, Brigham Young University, Provo UT 84602, USA; ²Departamento de Bioquímica, Genética e Inmunología, Facultad de Biología Universidad de Vigo, Vigo 36200, Spain

Abstract

Recombination is one of the key evolutionary processes shaping the architecture of genomes. Quantifying the effect of recombination is crucial to our understanding of how genetic diversity is generated and maintained in populations, for the design and analysis of studies aimed at uncovering the genetic basis for disease, and for studying the evolution of virulence and pathogenicity in bacteria and virus. Ignoring the occurrence of recombination may influence the analysis of genetic data and the conclusions derived from it. Here we describe the evolutionary significance of recombination, its detection and estimation, and its effect on evolutionary analysis. We further characterize the significance of recombination in four biological systems: bacteria, virus, mitochondria, and the human genome.

Introduction

Natural selection is typically viewed as the most powerful evolutionary force; however, it only selects among variation that already exists in a population. Natural selection does not create new genetic variants. The most important sources of new genetic variants come from mutation and the novel combination of existing alleles through recombination. A mutation is a heritable change in the genetic material of an organism. Mutations can alter a gene thereby creating a new allele or affect the structure or expression profile (depending on the location of the mutation). However, not all mutations will influence the genetic diversity of a population. In sexually reproducing eukaryotes, only when mutations affect gametic cells will the change be passed onto successive generations. Furthermore, for both eukaryotes and prokaryotes, a mutation will affect the genetic diversity of a population only if it results in a nondeleterious nucleotide substitution and increases in frequency due to the influences of genetic drift or positive selection.

Recombination involves the rearrangement of genetic material; it is any process that joins previously unassociated DNA fragments, and is often accomplished by breaking and rejoining DNA strands. Recombination is usually initiated by a single or double strand break, and is a necessary process for DNA repair and disjunction of chromosomes during meiosis. There are four common types of recombination: homologous recombination, site-specific recombination, transposition, and copy choice or strand transfer (Figure 1).

Homologous recombination is the reciprocal exchange of genetic material between two homologous (similarity due to common descent) sequences of DNA that can align, join, cross over, and exchange genetic information. This is the type of recombination that is required during meiotic crossing over, for bacteriophage recombination, for recombination following bacterial conjugation, and during the formation of plasmid multimers. Site-specific recombination involves crossing-over between short recombination signal sequences, which are specific double-stranded DNA sequences that allow insertion of a foreign DNA from a lysogenic virus or transposon into a host chromosome. Transposition is the process whereby a copy of a transposable element (transposon) is inserted into a new location. Transposable elements are DNA sequences that move from one chromosomal location to another. Transposons and retroviruses can move within and between chromosomes by making copies of themselves. The elements that move as DNA are called transposons whereas; elements that use reverse transcriptase to propagate are referred to as retrotransposons. Hence, transposition is a nonreciprocal exchange of DNA among chromosomes. RNA viruses use another type of recombination called copy choice or strand transfer. In this case the polymerase switches from one template strand of DNA to the other so that the resulting copy is a mixture of parental copies of DNA [1]. These definitions of recombination describe processes in which sequences from different regions of the genome or different taxa (species) can be combined into a single sequence. An interesting quality of recombination is that it functions in the dual capacity of both potentially creating (increasing) allelic variation by uniting previously unassociated sequences together and removing (decreasing) variation by separating unique sequence combinations.

Attention to evolutionary processes and studying how organisms are related to one another is an area of interest that is being applied in more and diverse biological

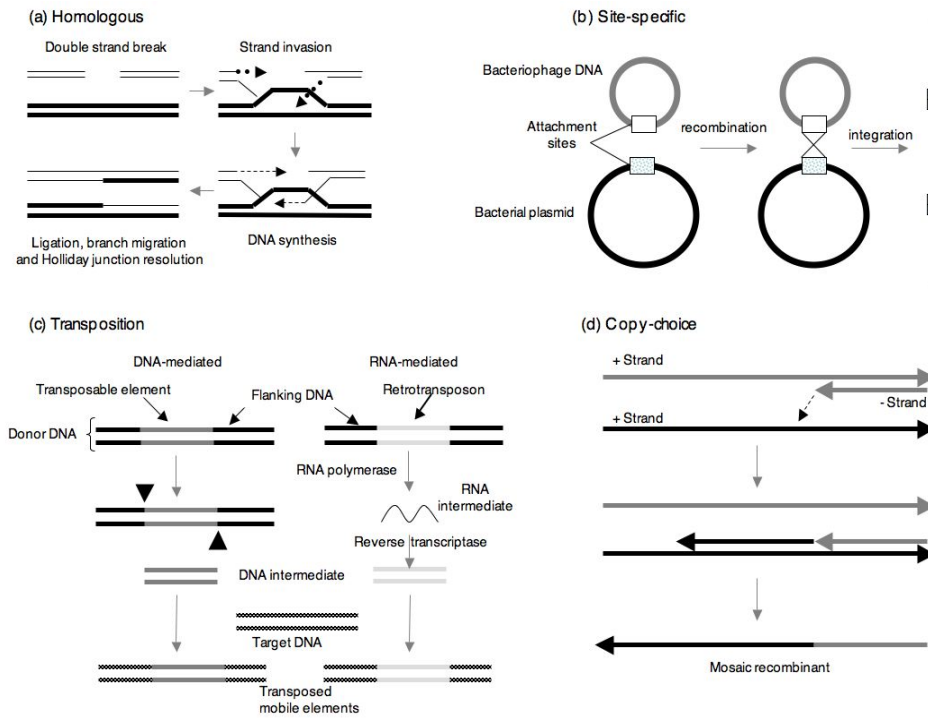


Figure 1. (a) Homologous recombination. Using the sequence information from the homologous regions of the homologous chromosomes (bolded DNA strands), the strands invade and are exchanged. Finally, DNA synthesis ensue in order to restore the original structure of the helix before the double strand break occurred. (b) Site-specific recombination between bacterial DNA and phage DNA. This recombination occurs between two nucleotide sequences that may not be homologous. Recombination is initiated at specific DNA sequences (attachment sites) which allows the two DNA molecules to attach together by base pairing. Once attached, enzymes catalyze the reaction resulting in the integration of the phage DNA within the bacterial genome. This process can result in inversions, deletions, or insertions. (c) DNA-mediated and RNA-mediated transposition. Transposition is the movement of DNA segments in the absence of any homology. Transposons are the DNA segments which can move within or between chromosomes; they carry a selectable marker (dark gray bands). Black wedges indicate where transposable element is duplicated. Retrotransposons are DNA segments which move around via a RNA intermediate (light gray bands). (d) Copy-choice recombination. During this process, the viral RNA-dependent RNA polymerase jumps from one RNA template to the other during replication creating a mosaic recombinant that is an amalgamation of both parental strands.

disciplines such as epidemiology, conservation biology, and genetic counseling. The information gained through evolutionary studies influences public policies regarding vaccine development, environmental standards, and has significantly altered the way scientists and lay people alike perceive and study biology. How does recombination affect the way we investigate the evolution of organisms and why is so much research being dedicated to answering this question? Recombination can seriously confound results obtained through molecular evolution studies [2, 3]. Evolutionary relationships

among organisms are studied using tools such as phylogenetic trees to examine the origin and diversification of species over geographical space and geological time. Phylogenetic trees are a mean to visually depict ancestor-descendent relationships among a group of organisms. A bifurcating, phylogenetic tree is made up of nodes and branches. Nodes represent ancestors of the taxonomic units (species, populations, individuals) at the tips of the branches. Branches represent the relationship between ancestor and descendants (Figure 2). The branching pattern or topology of the tree characterizes the relationships among all the taxa examined in the study. The process of recombination mixes sequences between DNA strands generating reticulate relationships which, if the parental sequences have different evolutionary histories, results in evolutionarily mosaic recombinants. When recombination goes undetected it can confound phylogenetic reconstruction efforts because traditional phylogenetic methods, based on a bifurcating tree, are unable to adequately depict the montage of evolutionary histories represented in a single DNA sequence.

Quantifying the effect of recombination is crucial to our understanding of how genetic diversity is generated and maintained in populations, for the design and analysis of studies aimed at uncovering the genetic basis for disease, and for studying the evolution of virulence and pathogenicity in bacteria and viruses for example [4]. Therefore, to better understand the influence of recombination on genomic evolution, it is imperative to evaluate the relative rates of its occurrence in lineages and identify recombinant sequences and recombination breakpoints in DNA alignments. In this chapter, recombination in bacteria, viruses, mitochondria, and in humans in general will be discussed. In eukaryotes recombination involves the process of reciprocal exchange of genetic material between homologous chromosomes. In prokaryotes and viruses recombination results in the addition of DNA, homologous or nonhomologous, to another's genome rather than a reciprocal exchange of homologous sequences. Regardless of the kind of recombination the result is the same: the creation of mosaic

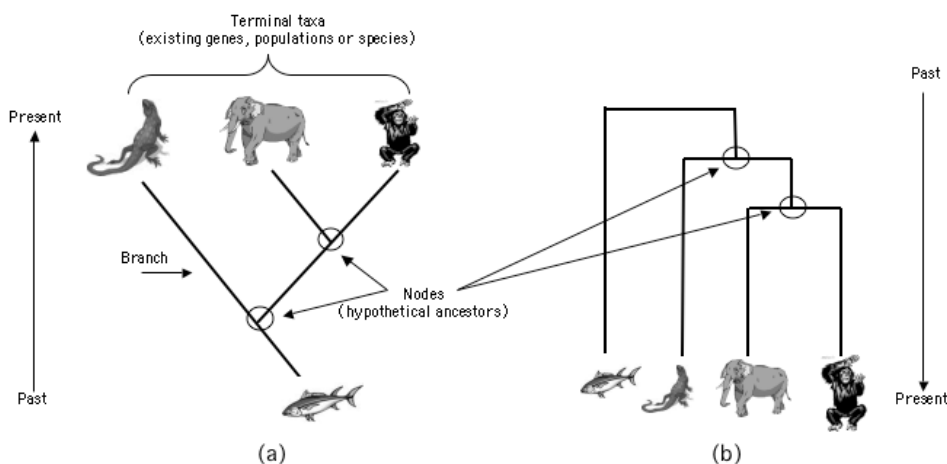


Figure 2. Phylogenetic trees. Both trees in this figure illustrate the same evolutionary relationships (a) This tree a is read from the bottom, starting with the most ancestral (fish) upward to the more derived taxa. (b) This tree is read from the top downward.

genomes. If recombination goes undetected or unaddressed, then the conclusions drawn from any evolutionary study can have serious faults and may have no biological meaning.

Detecting recombination

Detecting recombination is not a simple procedure and is dependent upon a whole suite of factors including the amount of divergence between the sequences undergoing recombination, where recombination occurs along a sequence and how frequently, effective population size, mutation rate, selection pressures, and the methods used to measure recombination [5, 6]. Identifying recombinational exchange also is dependant upon whether the resulting recombinant sequence is substantially different from either of the two parental sequences. For example, recombination events will go undetected when exchange occurs between identical or nearly identical parental sequences. When variation between sequences is low, it is very difficult to determine whether the differences are due to recombination or point mutations. There must be a series or run of mutations on one DNA strand that are different from those on the homologous strand in order for the resulting recombinant to be detected (Figure 3a) [6, 7]. Where recombination occurs along the sequence also will influence whether the event is detected. Suppose the sequences under study are variable but recombination occurs at the end of the variable region so only one variable site is exchanged; the result would be impossible to differentiate from the process of mutation (Figure 3b,c) [7, 8]. Furthermore, the nearly neutral theory of evolution predicts that most recombinants, even those that result in minor sequence alterations, will be deleterious and ultimately removed by purifying selection. The consequence of these limitations is that even the best estimate for the number of recombination events recovered will always underestimate the actual value.

Genetic diversity can refer to differences within individuals, between individuals, and between populations. With regard to detecting recombination we are interested in all three kinds. In general, genetic diversity refers to the amount of nucleotide variability as a function of the number of different alleles present in individuals or populations. There are several different ways to measure diversity. Theta (θ), also referred to as the population mutation parameter, (θ) = $N_e\mu$, is one way to define genetic diversity. N_e is the effective population size and μ is the per locus mutation rate per generation. The effective population size, which can be thought of as the average number of individuals in a population that contribute alleles to the next generation, affects the degree to which genetic drift influences allele frequencies. The larger the effective population size, the greater the number of alleles being transmitted each generation and the longer it will take for alleles to drift to fixation (Prob[fixation] = $1/N_e$). When N_e is smaller, there will be fewer alleles in the population and drift will randomly fix or lose alleles at a faster pace; which ultimately reduces variation within the population but can also make a population unique in comparison to others thereby increasing diversity among populations. Likewise, the rate of mutation influences the amount of genetic variability. With increasing N_e mutation will have a larger impact on genetic diversity due to the multiplicative effect. The mutation pattern at segregating sites can indirectly provide recombination information. Following the assumptions of the infinite-sites mutation model, where mutation events are assumed to result in unique nucleotide substitutions,

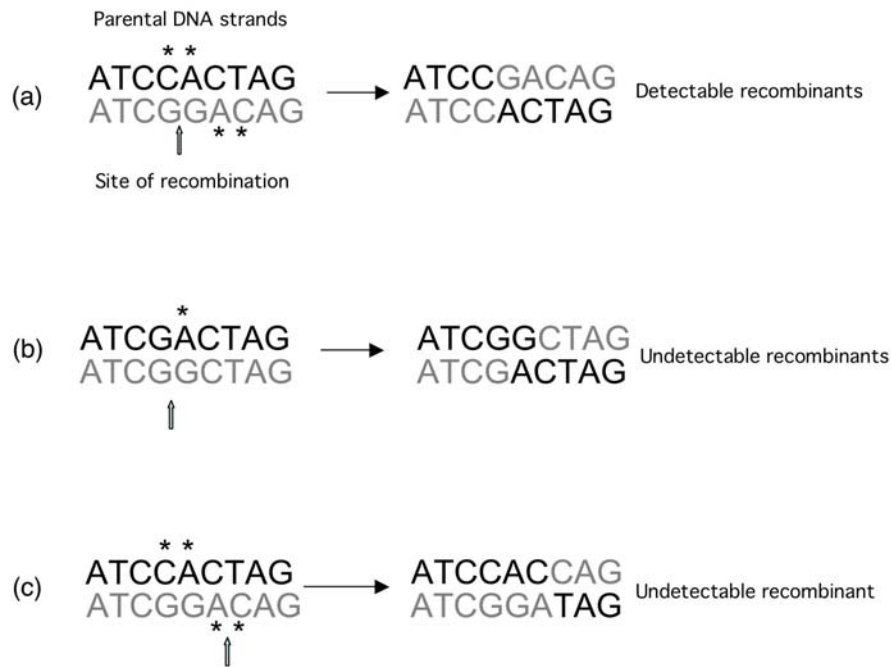


Figure 3. Detectable and undetectable recombination events. Asterisks (*) indicate variable sites. (a) When recombination occurs in the middle of the variable region between DNA strands with several nucleotide differences, the resulting recombinants are detectable. (b) When recombination occurs between sequences with one nucleotide difference, the resulting recombinants are indistinguishable from the parental DNA strands. (c) Although the parental strands present nucleotide differences, crossover occurs at the end of the variable region and recombination cannot be differentiated from point mutation.

past recombination events can be localized. This is because recombination events will shift groups of nucleotides but not alter specific nucleotide sites. Therefore, as genetic diversity in a population increases, the probability that recombination will be detected increases because the frequency of recombination occurring between identical or nearly identical sequences will be lower [7].

Quantifying the amount of genetic variation attributable to recombination within a population involves identifying and partitioning the contribution made by mutation. The population recombination parameter is described by the equation $\rho = 4N_e r$; N_e is the effective population, and r is the per locus recombination rate per generation. As we mentioned before, the population mutation parameter is described by $\theta = 4N_e \mu$. Variation in a population is shaped by both mutation and recombination; mutation creates new variation and recombination rearranges the nucleotides into unique alleles. If both ρ and θ can be accurately estimated, then we can define the relative rate of recombination compared to mutation $\varepsilon = \rho / \theta = r / \mu$. These population parameters are generally based on the assumptions of a constant sized, panmictic population evolving under neutrality, and the infinite-sites mutation model; although, models have been

developed that can independently estimate most of these parameters [9, 10]. Currently, phylogenetic and comparative methods for detecting recombination are not exceptionally powerful. Methods differ considerably in their relative abilities to detect recombination depending on the overall amount of genetic variation [5]. Simulation studies have demonstrated that recombination goes undetected even for relatively high mutation and recombination rates depending on the methods used [5, 7, 11]. For example, when rate heterogeneity and divergence was low ($\theta = 10$) the Homoplasy Test [12] can perform well however, when sequence divergence is high ($\theta = 200$) and recombination was exceedingly frequent ($\rho = 64$) the Homoplasy Test may perform poorly (Posada and Crandall, 2001).

Detection methods infer recombination events indirectly from the data. These methods are generally based on the coalescence model of evolution (described later) and make no *a priori* assumptions regarding the distribution of recombination. Different detection methods quantify the degree of similarity based on pairwise sequence comparisons and finding similarities among phylogenetic gene or haplotype trees [6, 13]. If congruence among sequence data or the phylogenetic tree topologies for several genes present the same evolutionary relationships, then two conclusions can be drawn. Either recombination has not occurred or it has occurred but not enough markers were exchanged to detect the event [14]. Inconsistencies among gene or haplotype trees can be a sign of past recombination. However, incongruent tree topologies also can be indicative of differences in population structure such as changes in population size and admixture or relative rates of gene evolution among the genes used in the analysis, for example. When the data do not support a single unique tree, phylogenetic networking methods [15] can depict relationships either as a reticulated tree, for evolutionary mosaic sequences, or as a bifurcating tree, for ideal sequences with one underlying phylogenetic history. The degree to which the tree is represented as a network indicates where, between which lineages recombination has occurred [16]. Another way to identify incompatible regions is to survey the data for regions of similar codon usage and base composition. By identifying regions that differ more than would be expected by chance recombination can be detected.

The extent and patterns of linkage disequilibrium (LD) [see for example 17, 18, 19] present in a population is another method to indirectly detect the presence of recombination. Genetic sequences at specific loci that are non-randomly associated and inherited together with a frequency greater than what is expected due to chance are said to be in linkage disequilibrium. When sequences exhibit LD the assumption is the loci under study are linked on the same chromosome and recombination has not occurred. However, a greater than expected value of LD cannot always be attributed to physical linkage or a lack of recombination. For example, LD can occur when recombination rate is low [14]. Population sub-structuring can also influence the way LD is interpreted. When samples are collected from diverse populations, lack of gene flow can cause some sites to appear to be linked even if there is no association between linkage and genetic distance [20]. There are two general statistics for measuring the amount of disequilibrium in a population: D , the difference between observed and expected frequencies of a gamete or haplotype [21] and r^2 , the square of the correlation of allele frequencies [22]. When recombination occurs between two bi-allelic loci there may be four potentially recombinants (Figure 4). If not all four recombinants are present in the

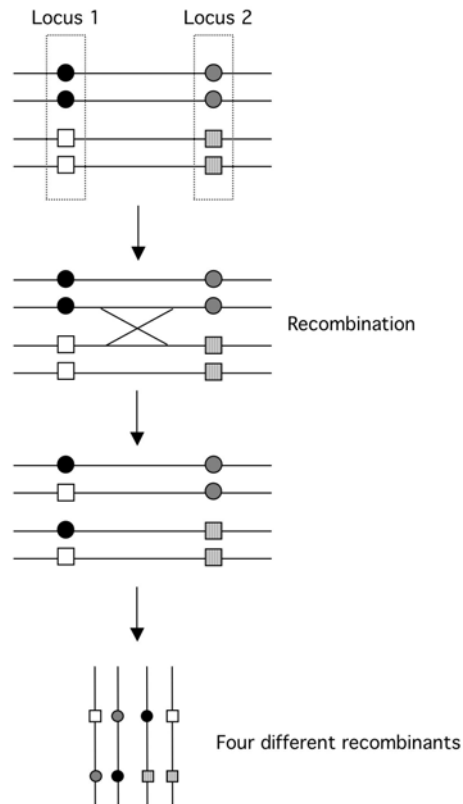


Figure 4. Recombination between biallelic loci. When recombination occurs between two biallelic loci there will be four potentially different recombinant types.

population $|D^*| = 1$ whereas, r^2 can give a low LD value depending on allele frequency and sample size. There are several limitations with recombination detection methods available today. The primary criticisms are they infer recombination indirectly from the data and second, they can only answer the question of whether recombination has occurred. To statistically compare the level of recombination across genomic regions, between species or between datasets methods that estimate recombination rate must be employed.

Estimating recombination rates

Both mutation and recombination contribute to the generation of unique genetic variants. Recombination also contributes by breaking down linkage disequilibrium, where alleles occur together more frequently than can be accounted for by chance, and in conjunction with natural selection, can function as an important evolutionary force. To quantify the contribution made by recombination we need to be able to distinguish between new genetic variants generated by mutation and those created by exchange of existing genetic material [8, 23, 24]. Many of the methods used to estimate

recombination rate and the population mutation (θ) and recombination (ρ) parameters make statistical assumptions based on neutral coalescent theory [25] with recombination [26-30].

The term coalescence refers to the process of looking backwards in time, starting with the existing sequence variation among samples, and estimating how they merge at times of common ancestry; referred to as the coalescent. This model is a statistical description of the evolutionary relationships among alleles, haplotypes or sequences that are randomly sampled from a panmictic population with no selection, and no recombination. Because mutations are assumed to be neutral, the fate of an individual within the population is not influenced by the mutations it may have, therefore the genealogical processes can be modeled independently of the mutation process. Given that mutations are independent of genealogy, mutations are assumed to occur randomly at a rate proportional to the product of the mutation rate and time to coalescence. Additional assumptions included in most analyses stipulate that lineages coalesce independently, no two lineages coalesce in a single generation for large population sizes, and the number of generations is so large it can be modeled as continuous time [25]. Nonetheless, there are models where each one of these assumptions are relaxed [see 31].

A coalescent event merges two lineages into one, whereas recombination joins lineages into a network of relationships. Adding recombination to the neutral coalescent model transforms the coalescence trees into a graph of lineages called the ancestral recombination graph (ARG) (Figure 5) [27, 32, 33]. Neutral coalescence with recombination contains all the assumptions of the coalescent model. The reason we need a model to estimate recombination rates is because we are not able to directly count the total number of recombination events due to the fact that most leave no trace of their occurrence. Even if we could count the number of recombination events, we still would have to know the number of generations our sample has undergone to be able to calculate the per generation recombination rate. Given that this information is often unavailable, neutral coalescence with recombination is the framework commonly used to estimate recombination rates.

The neutral coalescent model with recombination provides a method to distinguish between genetic variability generated by recombination as opposed to point mutation. There are two general approaches to estimating ρ , the population recombination parameter, utilizing the coalescent: quantifying recombination as a summary statistic or estimating recombination rates by considering all available data. Methods that quantify recombination as a summary statistic estimate the rate of recombination by identifying the number of nucleotide differences between sequences by pairwise comparison [34, 35] or by delimiting the minimum number of recombinations required [7, 39] to explain the data. These methods ignore all other evolutionary forces that could mimic the effects of recombination such as multiple mutations at a single site and demographic history. The benefit of using summary statistics is they are easy to calculate and are not computationally intensive. However, they are known to overestimate the population recombination rate, to have large confidence limits, in addition to only using part of the available information [4]. Thus, their effectiveness as estimators of recombination rates is generally regarded as inadequate [36]. One reason for their ineffectiveness is due to the strong correlation between the sequences used to calculate the summary statistic and the genealogical relationships among the sequences. Consequently if the sequences

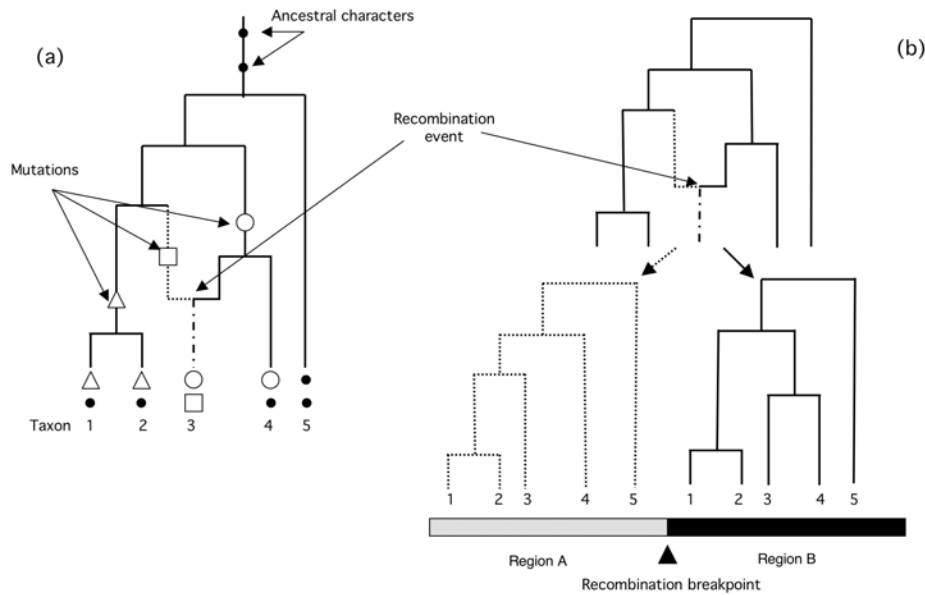


Figure 5. Recombination and phylogeny. (a) The symbol “•” represents the ancestral allele at, for example, two regions A and B within a gene (or two genes A and B). The symbols “△□○” represent mutations altering the ancestral sequences at those regions. Recombination results in taxon 3 having a combination of the “□” and “○” sequences. (b) Illustrates the problem of ignoring recombination when estimating evolutionary relationships. If region A is used, the gene tree on the left is recovered, where taxon 3 clusters with 1 and 2. If region B is used, the gene tree on the right is recovered, where taxon 3 clusters with 4. Neither tree alone adequately describes the evolution of these set of taxa.

are highly similar, then increasing the sample size will not effectively increase the amount of available information. Instead, utilizing full-likelihood and approximate or composite-likelihood approaches that incorporate all the data, including the underlying genealogical structure of the population, have proven to be more effective [36].

Many methods have been developed for estimating the frequency of recombination that use all empirical data rather than just a summary statistic. These methods employ both full and composite-likelihood methods to approximate joint-likelihood estimates for mutation and recombination rates by estimating the probability of observing a particular data set given different parameter values for θ and ρ . The advantage of these methods is greater analytical accuracy. However, full-likelihood approaches are extremely computationally intensive and become more analytically complicated as the size of the data set increases [37, 38]. Composite-likelihood methods have been developed that divide the dataset into computationally workable subsets. Therefore instead of a single dataset with many different parameters to analyze at once, there are several summary statistics that are used to characterize the dataset [39]. The likelihood for each subset is calculated independently and then all likelihood values are multiplied to obtain the composite-likelihood [36]. Employing this method significantly reduces computational time although some of the available information from the dataset is ignored [37]. The

most prominent assumptions applied by likelihood methods are constant population sizes, independence of nucleotide sites, neutral evolution, and most methods use the infinite-sites mutation model. These methods differ with regard to the manner of population sampling, level of nucleotide polymorphism, number and kind of nucleotide positions examined and the consistency with which they converge on the correct recombination rate in simulation studies[39].

Performance of estimators of recombination rate

At present, the performance of recombination rate estimators depends largely on the amount of genetic diversity present in the sampled population, with most methods performing poorly at low levels because of our inability to detect recombination events between identical and nearly identical sequences[39]. There are two approaches for comparing methods and evaluating their accuracy and consistency for converging on the correct value in simulated studies: how accurately methods approximate the likelihood surface for θ and ρ or evaluate the properties used to estimate θ and ρ [36]. The likelihood surface can be visualized as three dimensional space, composed of all the likelihood estimates for a particular parameter, from all of the models of evolution that are considered. The likelihood estimate that occupies the highest region in this space is the best parameter estimate.

It has been proposed that evaluating the accuracy of the likelihood surface is the best approach because it is based on all information contained within the sequence data. By evaluating the likelihood surface instead of relying on a single point estimate not only are better estimates obtained but reasonable confidence intervals can also be estimated for the parameter estimate. Therefore, if the likelihood surface can be estimated with greater accuracy, then estimation of θ and ρ should be more exact [36]. Although establishing the estimate of the recombination rate on the likelihood surface of the whole data rather than on a summary statistic is statistically more robust, there is uncertainty as to whether current methods are capable of estimating the likelihood surface with sufficient accuracy to approximate the maximum-likelihood estimate for θ and ρ [36]. Another problem is that a variety of different parameter values for ρ can produce the same phylogenetic tree. Therefore, several different values for ρ can, with equal likelihood, explain the genealogical relationships among the taxa examined [39]. In order for comparisons among methods to be meaningful, studies should use the same methods and range of parameters. As developing methods for efficiently and accurately estimating recombination rates have only recently received attention, few comparisons can be made at this time. Consequently, more research is needed in this field and additional research is needed to assess how vigorous these estimators are to violations of the neutral coalescent model.

Evolutionary implications of recombination

Effect on phylogenetic trees

Many scientific disciplines rely on phylogenetic trees to estimate rate heterogeneity, mutation bias, selection pressures, time of divergence between species, as well as sister group relationships. To their detriment, phylogenetic studies typically overlook the possibility of recombination, assuming the resulting phylogenetic trees accurately depict

the evolutionary associations of the organisms under study. The mosaic quality of recombinant sequences defies traditional phylogenetic reconstruction efforts because a single, non-reticulate, bifurcating tree cannot summarize their evolutionary relationships. When comparing sequences, multiple substitutions at a single nucleotide site can make it difficult or impossible to discern which character represents the ancestral state. Recurring mutations generate homoplasious character states (similarities that are due to chance and not to common descent). These factors inhibit our ability to correctly infer phylogenetic relationships [40]. As discussed, the effect of recombination can be difficult to distinguish from point mutations. When the rate of mutation is high, the probability of homoplasy occurring in the data set is higher. Additionally, recombinational exchange moves fragments of DNA therefore; some genetic regions will appear to have experienced greater mutational bias than they really have. Recombination generated homoplasies can complicate efforts to accurately estimate codon usage and adaptive evolution, mutation rate, selection pressures, and to test for the presence of a molecular clock [2-4, 6, 41-43].

Under the supposition of neutrality and constant population size the terminal branches of a phylogenetic tree are expected to be relatively similar in length. Increased branch length, in relation to internal branch lengths, is known to correlate with a recent expansion in population size, increased mutation rate, and selection. Recombination has the same effect on the length of terminal branches. The reasons why recombination causes branch lengths to increase and the shape of the tree to change is related to the effects recombination has on genetic diversity. When recombination breaks down linkage disequilibrium, disrupting rare sequence associations, genomic regions become more homogenous thus, the tree shape becomes more star or bush-like. Conversely, recombination can introduce nucleotide differences between sequences making one region appear to be more mutable or under greater selective pressure. Either case will cause the terminal branches to be longer in comparison to those lineages that have evolved under the neutral model of evolution. Recombination events can be placed into three categories based on their effect on phylogenetic tree topology: (1) events that do not change the branch length (recombination between identical sequences) (2) events that do change the branch length (recombination between sequences that have experienced mutations independent of one another) but do not change tree topology and (3) events that change the tree topology (recombination between very different sequences or hybridization events) [44].

Effect on population parameter estimates

Accurately estimating the rate of recombination has many implications for improving our ability to identify patterns of linkage disequilibrium for tracking the genetic basis of fitness and phenotypic traits [37]. The ability to understand the genetic components of phenotypic variation depends on our understanding of how different regions of the genome are linked. When recombination breaks apart linkage disequilibrium, these loci become independently segregating regions that will follow separate evolutionary trajectories. For example, an allele under positive selection will increase in frequency and eventually become fixed in the population. Neutral loci linked to this allele will also increase in frequency. When recombination separates these genomic regions, allowing loci to segregate independently, the neutral alleles are

released from this selective pressure and are no longer hitchhiking toward fixation. Thus, recombination will influence how accurately population genetic parameters can be estimated. If recombination goes undetected or is ignored the rate of mutation will be over estimated [4]; this is due to regions appearing to be more mutable than they truly are. The effective population size also will be over estimated because recombination creates new allelic varieties, increasing genetic diversity thereby making the population appear to larger than it is. Identifying which selective pressures are acting across loci will be confounded because neutral genes that were hitchhiking with selectively advantageous loci are now segregating independently and appear to have recently lost their selective advantage. The greater our understanding of linkage disequilibrium and the role of recombination as an evolutionary force, the more accurate the design and analysis of our studies will be for investigating the genetic basis of disease and diversity [45].

Genetic recombination in bacteria

The majority of bacteria have both chromosomal and extrachromosomal DNA. The chromosomal genome is a circle of double-stranded DNA floating free within the cytoplasm rather than being bound by a nuclear membrane. There are also smaller circular pieces of DNA called plasmids that are responsible for a significant proportion of the bacteria's genetic diversity. These extra chromosomal DNAs shape the evolution of bacteria by contributing genomic material that can influence pathogenicity and genetic diversity for example. In bacteria, recombination occurs through unidirectional transfer of genetic material from the donor to the recipient cell by three different processes; transformation, conjugation, and transduction. Transformation is the process by which bacteria can acquire new genes by taking up exogenous DNA molecules (i.e., plasmids or lineage fragments) from their surrounding environment. These DNA molecules can originate either from living or dead cells. Once inside the bacterium, the foreign DNA can be integrated into the host's chromosome by homologous recombination. The ability to deliberately transform bacteria, like *Escherichia coli*, has made possible the cloning of many genes, including human genes, and the development of the biotechnology industry. For bacteria that are naturally competent (cell can bind and internalize exogenous DNA molecules, thereby allowing transformation) like *Neisseria* sp., transformation significantly influences the frequency of recombination.

Bacteria also can receive DNA (bacterial or from other organisms) by infection from a bacteriophage via a process called transduction. A phage is a virus that attacks bacteria. Like most viruses, bacteriophages lack ribosomes, enzymes, and the ability to synthesize ATP; all of which are necessary for cellular function and reproduction. Hence, these obligate parasites can only proliferate by using the bacterial host's translational machinery. When the phage attacks, it injects its DNA into the bacterium. The viral DNA can be incorporated into the host's DNA by recombination between homologous regions. Prophages also can integrate their entire genome with the bacteria's. Inaccurate excision of the viral DNA from the host can remove a piece of the host DNA which is then packaged into the phage and transferred to other bacteria. The use of viruses as the vehicle to transmit genes is an important component to the success of gene therapy. Bacteriophages are also used extensively for DNA cloning in molecular biology. The

extent to which bacteriophages mediate recombination in natural populations has not been well studied and therefore, is not well document.

Conjugation is the process by which two bacteria come into physical contact and directly exchange DNA. The exchange of DNA is unidirectional from donor to recipient. In gram-negative bacteria conjugation begins when the donor cell initiates contact by creating an out pocket of its cell wall, a tube like structure called a sex pilus that connects the two bacterial cells. In gram-positive bacteria, conjugation does not involve pil. Prior to conjugation the recipient cell secretes substances that induces potential donors to produce proteins, called clumping factors, which function to draw bacterial cells together [46]. In both cases, the donor replicates its plasmid DNA and transfers it through a channel or pore on the cell surface. The connection between the two cells is tenuous and easily broken; hence it is rare for the donor to transfer its entire DNA to the recipient. Conjugation has been used extensively in genomic mapping studies facilitating our progress in locating the genetic components of disease. Conjugation does occur in many natural settings including in water, soil, and in various plants and animals.

Bacterial recombination differs from recombination in eukaryotes in both mechanism and its influence on recovering phylogenetic relationships [47]. The alternative pathways for conducting recombination in bacteria are generally referred to as horizontal or lateral gene transfer, in contrast to vertical transmission which is the passage of genetic material by mitosis or meiosis. The nonreciprocal exchange of DNA among bacteria means that a segment of DNA from one genome is replaced by another, which can result in gene conversion (Figure 6).

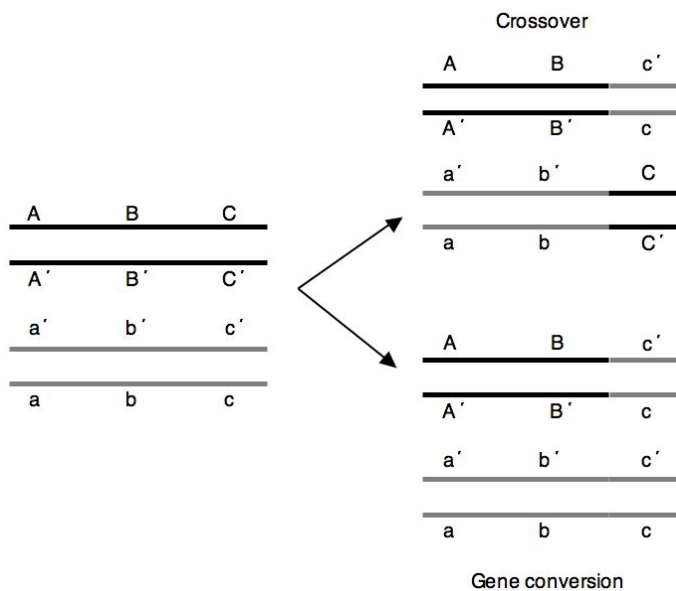


Figure 6. Illustration of the difference between DNA crossover and gene conversion. During crossover the two DNAs reciprocally exchange part of their genetic information (c-c' and C-C'). Gene conversion is a nonreciprocal transfer of genetic information –the gray DNA “donates” part of its genetic information (c-c' region) to the black DNA'.

Assessing the extent to which recombination occurs in bacteria has presented unique difficulties. Recombination is not a fundamental process in every generation as it is in higher organisms that undergo meiosis. The frequency of recombination also depends on the efficiency of genetic exchange and on a variety of biological and ecological factors [14]. Because the mechanisms involved in bacterial recombination result in nonreciprocal exchange of DNA, for recombination to occur and result in genetic variation strains of bacteria of the same or similar species must be present in the same environment [14, 48]. In addition, successful recombination requires that the donor DNA fragment be homologous to the DNA within the recipient cell.

Despite these general requirements for recombination, genetic exchanges have occurred among different bacterial species [49-51] with greater levels of sequence divergence than have been recorded for higher organisms [52]. Interspecific recombination has been documented to occur between sequences that were 25-30% divergent [14]. The evolutionary benefits of being able to exchange DNA with evolutionarily diverse species is evidenced by the number of bacteria that have acquired antibiotic resistant genes from multiple sources generating multi-antibiotic resistant strains [53-56]. Hence, the ability to estimate the frequency and location of recombination is essential for understanding the dynamic process of bacterial evolution.

Pervasiveness of recombination in bacteria

The prevalence of horizontal transfer and recombination in bacteria ranges from extensive in some species, including *Neisseria gonorrhoeae*, *N. meningitidis*, *Streptococcus pneumoniae*, *Staphylococcus aureus*, and *Helicobacter pylori* [14, 57, 58] to infrequent, as in *Escherichia coli* and *Salmonella enterica*, that are characterized by high levels of linkage disequilibrium and believed to be largely clonal [57]. Because bacteria reproduce by binary fission, it is not uncommon for their population structure to be characterized by clones and linkage disequilibrium even if recombination is common [14]. As discussed in detecting recombination, the amount of linkage disequilibrium does not necessarily correlate with frequency of recombination. Identifying phylogenetic incongruence among several gene or haplotype trees is one of the most common ways to detect recombination. Mutation and recombination rates can be highly variable therefore; molecular typing methods have been developed specifically for bacteria. Multilocus sequence typing (MLST) is the newest typing method developed for identifying bacterial isolates and characterizing population genetic parameters like recombination, genetic diversity and selection [59, 60]. MLST characterizes the alleles present at many loci, generally from a series of housekeeping genes, by sequencing approximately 600 bp from each gene. Every unique sequence for each locus is identified as a separate allele. And for each bacterial isolate the variation among alleles at all loci defines an allelic profile or sequence type (ST). Because housekeeping genes are involved in maintaining general cellular function, they are assumed to be evolutionarily conserved (under strong stabilizing or purifying selection). Therefore, detectable variation within a locus is most likely to be selectively neutral [61]. Although variation at each housekeeping gene is limited, simultaneous analysis of many loci greatly increases the power to discriminate among bacterial strains [62].

Characterizing the mode and tempo of recombination in bacterial evolution has become increasingly important, especially for pathogenic strains. Empirical data for

naturally transformable (*N. meningitidis* and *S. pneumoniae*) and transducible (*S. aureus*) species suggest that recombination was more likely than point mutations to generate genetic variability at neutral loci [14]. The implications of these findings are considerable. Recombination facilitates the spread of novel alleles across different genetic backgrounds. The high rate of recombination also has consequences for phylogenetic studies in bacteria, potentially obscuring evolutionary relationships, and impeding our ability to reconstruct patterns of bacterial evolution.

Recombination in viruses

Viruses are obligate, intracellular parasites dependent on their host's cellular machinery for self reproduction. The term virus usually refers to those particles that infect eukaryotes whereas bacteriophage or phage is used to describe those that invade bacteria. Viral genomes consist of either DNA or RNA encapsulated in a protein coat. Viruses are classified based upon the properties of their nucleic acid molecules; whether double or single stranded, linear, circular or linear but in two or more discrete segments. A DNA virus has DNA as its genetic material and does not make use of an RNA intermediate during replication. DNA viruses encode their genes on a single double-stranded DNA molecule (dsDNA), for example small pox, or on a single-stranded DNA (ssDNA) molecule, for example adeno-associated viruses (Table 1). Viruses that have RNA as its genetic material or use RNA as an intermediate during replication are classified as RNA viruses. RNA viruses exist as double stranded RNA (dsRNA), positive-sense single stranded in which the RNA is acting as the mRNA, negative-sense single stranded where the RNA is used as a template for mRNA synthesis, positive-sense single stranded RNA with a DNA intermediate in replication, and double stranded DNA with an RNA intermediate in replication (Table 1). Due to the different genomic

Table 1. Comparison of the seven different classes of viruses and examples of viral families and the disease they cause. The nucleic acid molecules are either double stranded (ds) or single stranded (ss); positive-sense strand (+) or negative-sense strand (-); and the chromosome (Chr.) is either linear (l), circular (c), or linear segmented into two or more chromosomes (ls).

Class	Nucleic Acid Types	Chr.	Virus Family	Example
I	dsDNA	l	Adenoviridae	Adenovirus (common cold)
			Poxviridae	Smallpox
II	ssDNA	l	Parvoviridae	Adeno-associated viruses
			Microviridae	Phi X174 (bacteriophage)
III	dsRNA	ls	Reoviridae	Simian rotavirus
IV	(+) ssRNA	l	Coronaviridae	Human Coronavirus (Severe Acute Respiratory Syndrome)
			Picornaviridae	Poliovirus
V	(-) ssRNA	ls	Orthomyxoviridae	Influenza A virus
			Arenaviridae	Sabia (Brazilian hemorrhagic fever)
VI	(+) RNA (reverse transcribed)	ls	Retroviridae	Human immunodeficiency virus
VII	DNA (reverse transcribed)	c	Hepadnaviridae	Hepatitis B virus

properties of these viruses, recombination in DNA and RNA viruses involve different processes. Recombination in DNA viruses is believed to proceed primarily by the same mechanisms operating in other DNA genomes; which entails enzyme-mediated breakage of covalent bonds, exchange of genetic information, and reuniting of covalent bonds. This classical model of recombination also occurs in RNA viruses which have a DNA intermediate, (retroviruses) copy-choice recombination is also employed by DNA viruses [63-65].

Recombination in RNA viruses occurs by two different methods; reassortment and template switching or copy-choice replication. Recombination by the process of reassortment is limited to viruses with segmented genomes such as influenza and rotaviruses [66]. Reassortment occurs when two or more strains infect the same cell and exchange genomic segments during viral replication (Figure 7). This mechanism has been well studied for influenza A and is postulated to account for the emergence of antigenically and genetically novel viruses that enable microbes to evade the immune response and persist in the host's body [see 8, 67]. The second method, copy-choice replication, can be utilized by either segmented or unsegmented viruses. Copy-choice is a process whereby the viral RNA-dependent RNA polymerase jumps from one RNA template to the other during replication creating a chimeric recombinant that is an amalgamation of both parental strands [64, 67] (Figure 1d). RNA viruses can initiate three different types of recombination all of which utilize copy-choice replication. Homologous recombination, which is the most common, aberrant homologous

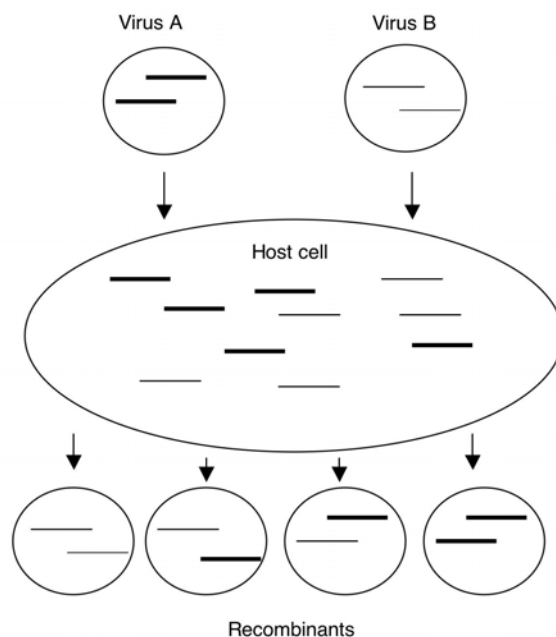


Figure 7. Genomic reassortment. If a virus has a segmented genome and if two variants of that virus infect the same host cell, then progeny virions (recombinants) can result with some segments from virus A, some from virus B.

recombination where related viruses exchange nucleotides without maintaining strict sequence alignment resulting in sequence duplication, insertion or deletion, and nonhomologous or illegitimate recombination which involves the exchange of genetic material between unrelated viruses [66].

Evidence for recombination in viruses is detected by identifying phylogenetic incongruence; utilizing the same methods as for bacteria and humans. Recombination has been documented to occur between related and unrelated viruses as well as between DNA and RNA viruses [8, 68, 69]. However, not all RNA viruses undergo recombination with similar frequency, in fact rates vary extensively. Homologous crossing over in retroviruses [70, 71] and reassortment in multi-segmented viruses occurs frequently [72]. Whereas recombination in positive-sense RNA viruses in animals, plants and bacteria appear to occur periodically and sporadically in negative-sense RNA viruses [73]. Differences in rates of exchange of genomic sequences have also been observed for viruses within the same family. For example, GB-virus C (hepatitis G virus) is a distant relative of hepatitis C both of which belong to the viral family Flaviviridae. Hepatitis C is a fairly clonal virus whereas GB-virus C recombines freely both within and between subtypes [74].

There are two hypotheses for explaining why recombination rates in RNA viruses are so variable. The recombination rate may be influenced by selection constraints on the size and stability of the genome and the necessity for increased genetic diversity [73]. In this framework, recombination functions to purge the genome of accumulated deleterious mutations as well as to increase diversity by dispersing beneficial combinations of mutations. For example large viral genomes, like coronaviruses, that experience high rates of mutation need a way to purge accumulated deleterious mutations from their genome [75]. Recombination provides genetic stability by removing mutations potentially enabling defective strains to regain function and influences diversity by creating and spreading novel genotypes. Alternatively, it is possible that recombination rates are primarily influenced by mechanistic or ecological constraints. Therefore, rates reflect the limitations imposed by genome structure and/or ecological needs of the virus. To date, the highest rates of recombination have been documented for retroviruses which package two copies of RNA into one virus particle or virion, greatly facilitating access and ultimately crossing over. At the other end the spectrum are single stranded positive-sense RNA viruses where recombinants are rarely detected and negative-sense RNA viruses whose genomes are organized into filamentous ribonucleoproteins which inhibit their ability to recombine. In addition to the physical requirements necessary for recombination to take place, multiple strains must be able to enter and infect the same host cell. Differing rates of recombination can be attributed to viral strains being ecologically or geographically isolated, the host's immune response preventing simultaneous infection, or lack of enough sequence homology between viral strains producing deleterious recombinants that are removed by purifying selection [73].

Although recombination can result in nonfunctional recombinants, there is substantial evidence that recombinants can achieve greater fitness through crossing over. For example, merging the drug resistant genes from two different HIV strains or generating virulent polio viruses. The oral poliovirus vaccine (OPV) consists of three live attenuated Sabie poliovirus strains. The vaccine confers immunity when the OPV strains multiply in the digestive tract [76]. Vaccine-vaccine and vaccine-nonvaccine

recombinants that have enhanced neurovirulence have been recovered from patients exhibiting vaccine-associated poliomyelitis. Thus, the potential for recombination to produce genomic modifications that influences virulence and pathogenicity is real and needs to be seriously evaluated when developing RNA vaccines and studying the evolution of these organisms [146].

Given the importance of recombination in the molecular evolution of viruses, it is crucial to determine the frequency with which it occurs and the mechanistic and ecological features that control its rate. Being able to predict the emergence of new viruses and improve therapeutic procedures is dependent upon the ability to identify the constraints to viral evolution. Determining the relative rate of mutation in comparison to recombination and if they correlate with adaptability, virulence, and geographic distribution are important areas for future research.

Recombination in mitochondria

Mitochondria are semiautonomous, self-reproducing organelles found in the cytoplasm of most eukaryotes. The mitochondrial genome consists of a small, circular DNA duplex, and there are normally five to 10 copies per mitochondrion. There can be up to 500-1000 mitochondria per animal cell [77]. Mitochondrial DNA (mtDNA) has become an important phylogenetic tool, in both applied and basic research, because of several important features such as its high copy number, high rate of mutation, and the assumptions of its exclusive maternal transmission, no recombination, and clonal inheritance. However, there is direct evidence for homologous recombination in plants, fungi and protists [78]. Recombination is thought to be extremely rare or absent in animals because recombinant haplotypes are rarely detected however, the occurrence or frequency of recombination is a different question from whether it can be detected. Detection requires creation of new recombinant haplotypes which necessitates heteroplasmy (the presence of more than one kind of mtDNA molecule within an individual). Heteroplasmy is uncommon in animals, with the notable exception of some male mussels where heteroplasmy is prevalent and recombination occurs frequently [79].

There are two requirements to be satisfied for recombination to occur in mitochondria. Mitochondrial DNA molecules must come into physical contact to exchange DNA and enzymes to catalyze the recombination reaction need to be present. In somatic cell hybrid studies, mtDNA molecules appear to cluster by haplotype, in this manner preventing the physical contact necessary for recombination [80]. Fusion of mitochondria has been documented in *Drosophila* [81-83]. Homologs for one of the genes (*fuzzy onions*) believed to mediate the fusion have been found in humans and yeast [84, 85]. These findings suggest that eukaryotes have some of the biological machinery necessary to achieve fusion although, evidence supporting mitochondrial fusion as a general process is contradictory. For example, studies of mitochondrial fusion in yeast [86], rat liver cells [83] and in human cells [81] indicate fusion is an efficient process. In contrast, experiments testing for complementation between different mtDNA molecules after fusion revealed that fusion and exchange of DNA is rare and not a general property of mitochondria in mammalian cells [87]. The presence of homologous DNA recombination in mitochondria has been proposed to function primarily as part of the replication and repair system [88]. The second requirement for recombination in mitochondria is the presence of enzymes needed to catalyze the reaction and to import

recombinant DNA or RNA from the cytoplasm. There is evidence, from *in vitro* and *in vivo* studies, that mammalian mitochondria can catalyze homologous [89] and non-homologous [90] recombination as well as intramolecular (within a single mtDNA molecule) recombination [91, 92].

Finally for homologous recombination to occur and result in non-clonal mitochondria, heteroplasmy must exist within the cell. Heteroplasmy can be achieved in a variety of ways through hybridization, paternal inheritance, double uniparental inheritance, and the presence of sequences of mitochondrial origin in the nuclear genome. Heteroplasmy, as a result of hybridization events, has been documented for a number of vertebrates [93-95]. The rate of recombination and the extent that it will produce new haplotypes in a population depends on the frequency of the co-occurrence of different mtDNA molecules within a cell and the amount of sequence divergence. In many species, paternal transmission of mtDNA is inhibited from persisting beyond fertilization. For example in mammals, upon entrance into the oocyte, sperm mitochondria and mtDNA are tagged with ubiquitin and consequently eliminated by proteolysis [88]. In intraspecific hybridizations, this process of locating and subsequent elimination of paternal mtDNA is believed to prevent paternal transmission the majority of the time. However, ubiquitination does not seem to interfere with paternal inheritance of mtDNA in interspecific hybridization crosses. The mechanisms involved in screening the zygote for sperm mtDNA appears to be species-specific [95]. Interestingly, in mice, it seems that mammalian sperm mitochondria have specific factors that guarantee their removal from the zygote whereas mtDNA derived from other tissue types like the liver were not selectively degraded [96]. Regardless of how stringent the processes are to prevent paternal leakage, the persistence of paternal mtDNA through embryogenesis has been observed in a variety of invertebrates and vertebrates including humans [93, 97-100]. Although crosses between closely related species can result in mitochondrial heteroplasmy and paternal leakage does occur, recombinant haplotypes have not been recovered in either circumstance.

Doubly uniparental inheritance of mtDNA is a special condition known to occur in two bivalve families, Mytilidae (sea mussels) and Unionidae (freshwater mussels). In these lineages, sperm mtDNA is eliminated from the zygote in female offspring but is retained in males which become 'homoplasmic mosaics' with the predominant mtDNA being paternally derived in their gonads and maternal mtDNA being predominant in their somatic tissue. Occasionally, the maternal genome invades the gonadal tissue and is transmitted to male offspring presenting the opportunity for recombination; which occurs with high frequency [79].

The second pathway by which the mtDNA could become non-clonal is by recombining with mitochondrial sequences present in the nuclear genome. Many studies on different organisms have demonstrated that genes [101, 102], pseudogenes [103-105] and introns [106] of mitochondrial origin have been incorporated into the nuclear genome. In humans it has been estimated that there are up to a 1000 mtDNA pseudo sequences in the nucleus [107]. Recombination would require that mitochondria have the ability to import DNA or RNA from the cytoplasm. Studies in yeast have documented that mtDNA can be exported from the mitochondria into the nucleus at a rate of $\sim 2 \times 10^{-5}$ (per cell per generation) however; DNA was not detected entering mitochondrial cells

[108]. Additionally, mitochondrial importation of an assortment of RNAs including viral RNAs have been documented for a variety of eukaryotes including humans [109].

Recombination in humans mtDNA

Recent publications suggesting that inheritance of human mtDNA might not be strictly clonal have reinvigorated the debate regarding how recombination is detected. Homoplasmy is a characteristic of all sequence data and in mtDNA it is particularly prevalent in the D-loop and certain coding regions. The presence of homoplasmy can confound attempts to identify the ancestral state of a nucleotide site but it is also informative with regard to estimating the neutral rate of mutation within a sequence. It is not clear how much rate heterogeneity exists in the human mitochondrial genome and if mutation alone can account for the amount of homoplasmy detected. Likewise, the process of fixation and the role of positive selection in mtDNA are not well characterized. There is evidence to support the hypothesis of non-independence of sites and hypermutability (sites evolve at a much faster rate than average) in the control region which might account for the high rate of mutation and heteroplasmy among individuals [110, 111]. The two mechanisms that cause homoplasmy are repeated mutations at the same site across the phylogenetic tree in different sequences and recombination. An important question is how to differentiate the effects generated by the two.

Three seminal papers published in 1999, [112-114] announced they had found evidence for recombination in human mtDNA. These articles proposed that the level of observed homoplasmy could not be explained by mutational processes alone, therefore, invoking recombination as the causative mechanism. Hagelberg *et al.* [114] found a rare point mutation in three distinct mtDNA lineages among western Pacific islanders. Their findings have been retracted due to a sequence alignment error [115]. Eyre-Walker *et al.* [113] found excessive homoplasmy at polymorphic sites concluding that mutation alone could not account for the observed mutational load. Their study used phylogenetic methods of tree reconstruction to reveal homoplasies. Posada [6] empirically demonstrated that substitution and compatibility methods were, in general, more powerful than phylogenetic methods for inferring recombination. Awadalla *et al.* [112] identified a correlation between reduced linkage disequilibrium (LD) with increased distance between segregating nucleotide sites in humans and chimpanzees. Critics have questioned the quality of their data [116], use of the measure r^2 for detecting recombination suggesting this LD statistic lacks statistical power because of its sensitivity to allele frequencies, and for inferring recombination without direct empirical evidence [117-120]. When the Awadalla *et al.* datasets were reanalyzed using D' no correlation between LD and distance was found; instead data were observed to be consistent with mutation. Furthermore, subsequent studies of other populations have failed to replicate the association between declining LD with increased distance between polymorphic sites [121]. Differences in the ability of D' and r^2 to infer recombination has been attributed to the way the two statistics use the data. D' will fail to detect recombination if all four haplotypes are not present whereas the r^2 statistic does not need all four haplotypes to infer recombination.

An important outcome of these recent studies is the realization of how difficult it is to detect recombination. Our detection methods are heavily influenced by the

assumptions included in the models used. While mitochondrial recombination in plants, fungi and protists appears to be fairly common, the evidence for recombination in animals and humans is equivocal. A broad assortment of patterns have been described for mutation and substitutions in the mitochondrial genome of animals such as codon-usage bias, substitution rate heterogeneity across the molecule, and lineage specific rates and patterns of substitution [122-125]. Determining whether these patterns can also be detected in human mtDNA and if they can explain the observed level of linkage disequilibrium and distribution of homoplasmy will need to be addressed.

Recombinational hotspots and human genetic diversity

There is an abundance of evidence suggesting meiotic homologous recombination events are not distributed evenly throughout prokaryotic or eukaryotic genomes [126-128]. Rather, there are regions of elevated recombination known as recombination hotspots that are interleaved with regions that experience little or no recombination, referred to as warm and coldspots. Descriptions of linkage disequilibrium (LD) patterns have focused on identifying regions or blocks of low and high haplotype diversity. Regions exhibiting LD and thus lower diversity are assumed to represent nucleotide blocks that have historically undergone recombination infrequently or not at all; whereas, regions with little or no LD and higher diversity are speculated to represent hotspots for recombination.

Ideally, LD analyses should be based on empirical rather than inferential data. One way to attain direct evidence is by identifying haplotype variation using single nucleotide polymorphisms or SNPs (pronounced “snips”). Single nucleotide polymorphisms are DNA sequence variations that occur at a single nucleotide. Variations must occur in at least 1% of the population for them to be categorized as SNPs [129]. They have become a popular and useful molecular tool because they represent approximately 90% of all human genetic variation and are distributed roughly every 1000-2000 nucleotides [130]. SNPs occur in both coding and noncoding regions, the majority of which do not appear to negatively affect cellular function, although some are believed to predispose people to disease, influence their response to environmental stressors (such as viruses and bacteria) and therapeutic drugs (Human Genome Project Information web site). SNPs are also believed to be evolutionarily stable therefore; they are ideal markers for population level studies and for constructing chromosome maps and localizing regions of historical recombination.

Linkage studies and analysis of LD patterns have led to the discovery of crossover hotspots in a number of regions in the human genome [131-140]. The most intensely studied region has been the major histocompatibility complex (MHC) class II region [135, 136, 141-143]. Although a number of hotspots have been identified, few have been extensively characterized. Uncovering the mechanistic basis and distribution of recombination has been hampered by the expansive distance over which recombination occurs in the genome, approximately 1% per one million bases [141]. To obtain greater map resolution and answer questions regarding the distribution and rate of recombination, analysis of large numbers of recombination events over smaller regions of DNA are essential. To obtain the sample size necessary large numbers of meioses need to be studied. Developments in single-sperm typing have greatly enhanced the precision of genetic maps produced from LD analyses. A single semen sample contains

hundreds of millions of sperm; each representing a unique meiotic crossing over event. Typically, patterns of SNP genotypes are used to find regions of LD (haplotype blocks). Then variations among these haplotype blocks are used to target crossover hotspots within this narrowly defined region. One of the limitations of sperm cross over analyses is only paternal meiotic information is acquired. To obtain whole genomic data, recombination within the maternal lineage must also be studied. Linkage analyses or pedigree studies is the only method presently available to study recombination contributed through the maternal lineage. The resolution of these maps is fairly low due to limited number of meioses that can be obtained from pedigree studies [141].

There are other mechanisms that can cause regions to appear like hotspots for crossing over [144]. The effects of selection can mimic the patterns and characteristics of recombination hotspots. For example, in coronaviruses that had been serially passaged through tissue culture, recombination hotspots were detected [75]. However once the selective pressure was removed, and RT-PCR was used to detect crossover events, recombination breakpoints appeared to be randomly distributed. Thus virus selection, especially laboratory induced selection, can falsely inflate the estimated rate of recombination.

Although recombination hotspots have not been extensively characterized in the human genome, there are some emerging properties: 1) the majority of recombination events are simple, involving a single site of exchange, 2) exchange between haplotypes is typically reciprocal, 3) the distribution of exchange points are generally symmetrical across hotspots thus allowing the epicenters to be localized with considerable accuracy (within ± 30 bp), 4) typically hotspots are 1-2kb, flanked by regions that are less active in recombination (1cM/Mb which is the genome average) [141]. Within MHC recombination rates vary significantly among individuals however, similar rates were observed between identical and fraternal twins [143]. This evidence suggests there might be a genetic component influencing crossover frequency. Rate heterogeneity has also been observed between males and females in the TAP2 gene [141] but not for other genes examine thus far within the class II region [143]. There does not appear to be a particular chromosomal region that is predisposed to increase cross over activity, although there is evidence for the presence of recombination promoting DNA motifs, (short specific sequences).

Understanding the role recombination plays in creating regions of LD, specifically how hotspots influence the rate and distribution of recombination across the genome will have a profound impact on the search for genes involved in inherited diseases [145] and provide insight into how genetic diversity is produced and maintained. The challenge is to develop methods for characterizing the distribution of crossovers along chromosomes, identify what mechanisms and processes operate during recombination, and how these factors influence diversity.

References

1. Lewin, B., *Genes VI*. 1997, Oxford: Oxford University Press. 1260.
2. Schierup, M.H. and J. Hein, *Consequences of recombination on traditional phylogenetic analysis*. *Genetics*, 2000. **156**: p. 879-891.
3. Schierup, M.H. and J. Hein, *Recombination and the molecular clock*. *Mol. Biol. Evol.*, 2000. **17**(10): p. 1578-1579.

4. Awadalla, P., *The evolutionary genomics of pathogen recombination*. Nat. Rev. Genet., 2003. **4**: p. 50-60.
5. Posada, D. and K.A. Crandall, *Evaluation of methods for detecting recombination from DNA sequences: computer simulations*. Proc. Natl. Acad. Sci. USA, 2001. **98**(24): p. 13757-13762.
6. Posada, D., *Evaluation of methods for detecting recombination from DNA sequences: empirical data*. Mol. Biol. Evol., 2002. **19**(5): p. 708-17.
7. Hudson, R.R. and N.L. Kaplan, *Statistical properties of the number of recombination events in the history of a sample of DNA sequences*. Genetics, 1985. **111**: p. 147-164.
8. Posada, D., K.A. Crandall, and E.C. Holmes, *Recombination in evolutionary genomics*. Annu. Rev. Genet., 2002. **36**: p. 75-97.
9. Hey, J., *A multi-dimensional coalescent process applied to multi-allelic selection models and migration models*. Theor. Popul. Biol., 1991. **39**(1): p. 30-48.
10. Sano, A., A. Shimizu, and M. Izuka, *Coalescent process with fluctuating population size and its effective size*. Theor. Popul. Biol., 2004. **65**(1): p. 39-48.
11. Stephens, J.C., *On the frequency of undetectable recombination events*. Genetics, 1986. **112**: p. 923-926.
12. Maynard Smith, J. and N.H. Smith, *Detecting recombination from gene trees*. Mol. Biol. Evol., 1998. **15**: p. 590-599.
13. Crandall, K.A. and A.R. Templeton, *Statistical methods for detecting recombination, in The Evolution of HIV*, K.A. Crandall, Editor. 1999, Johns Hopkins Univ. Press: Baltimore, MD. p. 153-76.
14. Feil, E.J. and B.G. Spratt, *Recombination and the population structures of bacterial pathogens*. Annu. Rev. Microbiol., 2001. **55**: p. 561-590.
15. Posada, D. and K.A. Crandall, *Intraspecific gene genealogies: trees grafting into networks*. Trends Ecol. Evol., 2001. **16**(1): p. 37-45.
16. Hudson, R.R., *Splits tree: a program for analyzing and visualizing evolutionary data*. Bioinformatics, 1998. **14**: p. 68-73.
17. Lewontin, R.C., *On measures of gametic disequilibrium*. Genetics, 1988. **120**: p. 849-852.
18. Hill, W.G., *Estimation of linkage disequilibrium in randomly mating populations*. Heredity, 1974. **33**(2): p. 229-239.
19. Hedrick, P. and S. Kumar, *Mutation and linkage disequilibrium in human mtDNA*. Eur. J. Hum. Genet., 2001. **9**: p. 969-972.
20. Meunier, J. and A. Eyre-Walker, *The correlation between linkage disequilibrium and distance: implications for recombination in hominid mitochondria*. Mol. Biol. Evol., 2001. **18**(11): p. 2132-2135.
21. Lewontin, R.C. and K. Kojima, *The evolutionary dynamics of complex polymorphisms*. Evolution, 1960. **14**: p. 450-472.
22. Hill, W.G. and A. Robertson, *Linkage disequilibrium in finite populations*. Theor. Appl. Genet., 1968. **38**: p. 226-231.
23. Felsenstein, J., *The evolutionary advantage of recombination*. Genetics, 1974. **78**: p. 737-756.
24. Hill, W.G. and A. Robertson, *The effect of linkage on limits to artificial selection*. Genet. Res., 1966. **8**: p. 269-294.
25. Kingman, J.F.C., *The coalescent*. Stochastic Processes and Their Applications, 1982. **13**: p. 235-248.
26. Griffiths, R.C. and S. Tavaré, *Ancestral inference in population genetics*. Stat. Sci., 1994. **9**: p. 307-319.
27. Hudson, R.R., *Properties of the neutral allele model with intergenic recombination*. Theor. Popul. Biol., 1983. **23**: p. 183-201.
28. Wiuf, C. and J. Hein, *Recombination as a point process along sequences*. Theor. Popul. Biol., 1999. **55**: p. 248-259.

29. Wiuf, C. and J. Hein, *The ancestry of a sample of sequences subject to recombination*. Genetics, 1999. **151**: p. 1217-1228.
30. Wiuf, C. and D. Posada, *A coalescent model of recombination hotspots*. Genetics, 2003. **164**(1): p. 407-17.
31. Rosenberg, N.A. and M. Nordborg, *Genealogical trees, coalescent theory and the analysis of genetic polymorphisms*. Nat. Rev. Genet., 2002. **3**(5): p. 380-90.
32. Griffiths, R.C. and P. Marjoram, *Ancestral inference from samples of DNA sequences with recombination*. J. Comp. Biol., 1996. **3**(4): p. 479-502.
33. Griffiths, R.C. and P. Marjoram, *An ancestral recombination graph*, in *Progress in population genetics and human evolution*, P. Donnelly and S. Tavaré, Editors. 1997, Springer-Verlag: Berlin. p. 257-270.
34. Hudson, R.R., *Estimating the recombination parameter of a finite population model without selection*. Genet. Res., 1987. **50**(3): p. 245-250.
35. Wakeley, J., *Using the variance of pairwise differences to estimate the recombination rate*. Genet. Res., 1997. **69**: p. 45-48.
36. Fearnhead, P. and P. Donnelly, *Estimating recombination rates from population genetic data*. Genetics, 2001. **159**: p. 1299-1318.
37. Stumpf, M.P. and G.A.T. McVean, *Estimating recombination rates from population-genetic data*. Nat. Rev. Genet., 2003. **4**: p. 959-968.
38. McVean, G.A.T., P. Awadalla, and P. Fearnhead, *A coalescent-based method for detecting and estimating recombination from gene sequences*. Genetics, 2002. **160**: p. 1231-1241.
39. Wall, J.D., *A comparison of estimators of the population recombination rate*. Mol. Biol. Evol., 2000. **17**: p. 156-163.
40. Posada, D. and K.A. Crandall, *The effect of recombination on the accuracy of phylogeny estimation*. J. Mol. Evol., 2002. **54**(3): p. 396-402.
41. Posada, D., *Unveiling the molecular clock in the presence of recombination*. Mol. Biol. Evol., 2001. **18**(10): p. 1976-8.
42. Shriener, D., et al., *Potential impact of recombination on sitewise approaches for detecting positive natural selection*. Genet. Res., 2003. **81**: p. 115-121.
43. Anisimova, M., R. Nielsen, and Z. Yang, *Effect of Recombination on the Accuracy of the Likelihood Method for Detecting Positive Selection at Amino Acid Sites*. Genetics, 2003. **164**(3): p. 1229-1236.
44. Wiuf, C., T. Christensen, and J. Hein, *A simulation study of the reliability of recombination detection methods*. Mol. Biol. Evol., 2001. **18**: p. 1929-1939.
45. Fearnhead, P., *Consistency of estimators of the population-scaled recombination rate*. Theor. Popul. Biol., 2003. **64**: p. 67-79.
46. Miller, R.V., *Bacterial gene swapping in nature*. Scientific America, 1998. **1**: p. 46.
47. DuBose, R.F., D.E. Dykhuizen, and D.L. Hartl, *Genetic exchange among natural isolates of bacteria: recombination within the phoA gene of Escherichia coli*. Proc. Natl. Acad. Sci. USA, 1998. **85**: p. 7036-7040.
48. Smith, J.M., C.G. Dowson, and B.G. Spratt, *Localised sex in bacteria*. Nature, 1991. **349**: p. 29-31.
49. Bowler, L.D., et al., *Inter-species recombination between the penA genes of Neisseria meningitidis and commensal Neisseria species during the emergence of penicillin resistance in N. meningitidis: natural events and laboratory simulation*. J. Bacteriol., 1994. **176**: p. 333-337.
50. Dowson, C.G., et al., *Penicillin-resistant viridans streptococci have obtained altered penicillin-binding protein genes from penicillin-resistant strains of Streptococcus pneumoniae*. Proc. Natl. Acad. Sci. USA, 1990. **87**: p. 5858-5862.
51. Reeves, P.R., *Evolution of Salmonella O antigen variation by interspecific gene transfer on a large scale*. Trends Genet., 1993. **9**: p. 17-22.

52. Matic, I., F. Taddei, and M. Radman, *Genetic barriers among bacteria*. Trends Microbiol., 1996. **4**: p. 69-73.
53. Davies, J.E., *Inactivation of antibiotics and the dissemination of resistance genes*. Science, 1994. **267**: p. 375-382.
54. Musser, J.M., et al., *Clonal analysis of methicillin-resistant Staphylococcus aureus strains from intercontinental sources: association of the mec gene with divergent phylogenetic lineages implies dissemination by horizontal gene transfer and recombination*. J. Clin. Microbiol., 1988. **30**: p. 2058-2063.
55. Spratt, B.G., *Resistance to antibiotics mediated by target alterations*. Science, 1994. **264**: p. 388-393.
56. Dzidic, S. and V. Bedekovic, *Horizontal gene transfer-emerging multidrug resistance in hospital bacteria*. Acta Pharmacol. Sin., 2003. **24**(6): p. 519-526.
57. Smith, J.M., et al., *How clonal are bacteria?* Proc. Natl. Acad. Sci. USA, 1993. **90**: p. 4384-4388.
58. Go, M.F., et al., *Population genetic analysis of Helicobacter pylori by multilocus enzyme electrophoresis-extensive allelic diversity and recombinational population structure*. J. Bacteriol., 1996. **178**: p. 3934-3938.
59. Feil, E.J., et al., *Estimating recombinational parameters in Streptococcus pneumoniae from multilocus sequence typing data*. Genetics, 2000. **154**: p. 1439-1450.
60. Smith, J.M., E.J. Feil, and N.H. Smith, *Population structure and evolutionary dynamics of pathogenic bacteria*. Bioessays, 2000. **22**(12): p. 1115-1122.
61. Feil, E., et al., *The relative contributions of recombination and mutation to the divergence of clones of Neisseria meningitidis*. Mol. Biol. Evol., 1999. **16**(11): p. 1496-1502.
62. Viscidi, R.P. and J.C. Demma, *Genetic diversity of Neisseria gonorrhoeae housekeeping genes*. J. Clin. Microbiol., 2003. **41**(1): p. 197-204.
63. d'Alencon, E., et al., *Copy-choice illegitimate DNA recombination revisited*. The European Molecular Biology Organization Journal, 1994. **13**(11): p. 2725-2734.
64. Negroni, M. and H. Buc, *Mechanisms of retroviral recombination*. Annu. Rev. Genet., 2001. **35**: p. 275-302.
65. Brunier, D., B. Michel, and S.D. Ehrlick, *Copy choice illegitimate DNA recombination*. Cell, 1988. **52**: p. 883-892.
66. Lai, M.M., *RNA recombination in animal and plant viruses*. Microbiological Reviews, 1992. **56**(1): p. 61-79.
67. Worobey, M. and E.C. Holmes, *Evolutionary aspects of recombination in RNA viruses*. J. Gen. Virol., 1999. **80**: p. 2535-2543.
68. Gibbs, M.J. and G.F. Weiller, *Evidence that a plant virus switched hosts to infect a vertebrate and then recombined with a vertebrate-infecting virus*. Proc. Natl. Acad. Sci. USA, 1999. **96**: p. 8022-8027.
69. Worobey, M., *Extensive homologous recombination among widely divergent TT viruses*. J. Virol., 2000. **74**(16): p. 7666-7670.
70. Hu, W.S. and H.M. Temin, *Retroviral recombination and reverse transcription*. Science, 1990. **250**: p. 1227-1233.
71. Jung, A., et al., *Multiply infected spleen cells in HIV patients*. Nature, 2002. **418**: p. 144.
72. Gorman, O.T., W.J. Bean, and R.G. Webster, *Evolutionary processes in influenza viruses: divergence, rapid evolution, and stasis*. Curr. Topic. Microbiol. Immunol., 2001.
73. Chare, E.R., E.A. Gould, and E.C. Holmes, *Phylogenetic analysis reveals a low rate of homologous recombination in negative-sense RNA viruses*. J. Gen. Virol., 2003. **84**: p. 2691-2703.
74. Worobey, M. and E.C. Holmes, *Homologous recombination in GB virus C/hepatitis G virus*. Mol. Biol. Evol., 2001. **18**(2): p. 254-261.

75. Lai, M.M.C., *Recombination in large RNA viruses: coronaviruses*. Semin. Virol., 1996. **7**(6): p. 381-388.
76. Georgescu, M.M., et al., *High diversity of poliovirus strains isolated from the central nervous system from patients with vaccine-associated paralytic poliomyelitis*. J. Virol., 1994. **68**(12): p. 8089-8101.
77. Graur, D. and W.-H. Li, *Fundamentals of molecular evolution*. second ed. 2000, Sunderland, Massachusetts: Sinauer Associates, Inc. 481.
78. Gray, M.W., *Origin and evolution of mitochondrial DNA*. Annu. Rev. Cell. Biol., 1989. **5**: p. 25-50.
79. Ladoukakis, E.D. and E. Zouros, *Direct evidence for homologous recombination in mussel (*Mytilus galloprovincialis*) mitochondrial DNA*. Mol. Biol. Evol., 2001. **18**(7): p. 1168-1175.
80. Satoh, M. and T. Kuroiwa, *Organization of multiple nucleoids and DNA molecules in mitochondria of a human cell*. Experimental Cell Research, 1991. **196**: p. 137-140.
81. Legros, F., et al., *Mitochondrial fusion in human cells is efficient, requires the inner membrane potential, and is mediated by mitofusion*. Mol. Biol. Cell, 2002. **13**(2): p. 4343-4354.
82. Yaffe, M.P., *The machinery of mitochondrial inheritance and behavior*. Science, 1999. **283**: p. 1493-1497.
83. Cortese, J.D., *Rat liver GTP-binding proteins mediate changes in mitochondrial membrane potential and organelle fusion*. Am. J. Physiol. Cell. Physiol., 1999. **276**(3): p. C611-C620.
84. Santel, A. and M.T. Fuller, *Control of mitochondrial morphology by a human mitofusion*. J. Cell. Sci., 2001. **114**(5): p. 867-874.
85. Rokas, A., E. Ladoukakis, and E. Zouros, *Animal mitochondrial DNA recombination revisited*. Trends Ecol. Evol., 2003. **18**(8): p. 411-417.
86. Azpiroz, P. and R.A. Butow, *Patterns of mitochondrial sorting in yeast zygotes*. Mol. Biol. Cell, 1993. **4**(1): p. 21-36.
87. Enriquez, J.A., et al., *Very rare complementation between mitochondria carrying different mitochondrial DNA mutations points to intrinsic genetic autonomy of the organelles in cultured human cells*. J. Biol. Chem., 2000. **275**(15): p. 11207-11215.
88. Schwartz, M. and J. Vissing, *New patterns of inheritance in mitochondrial disease*. Biochem. Biophys. Res. Commun., 2003. **310**: p. 247-251.
89. Thyagarajan, B., R.A. Padua, and C. Campbell, *Mammalian mitochondria possess homologous DNA recombination activity*. J. Biol. Chem., 1996. **271**(44): p. 27536-27543.
90. Lakshmipathy, U. and C. Campbell, *Double strand break rejoining by mammalian mitochondrial extracts*. Nucleic Acids Res., 1999. **27**(4): p. 1198-1204.
91. Holt, I.J., D.R. Dunbar, and H.T. Jacobs, *Behavior of a population of partially duplicated mitochondrial DNA molecules in cell culture: segregation, maintenance and recombination dependent upon nuclear background*. Hum. Mol. Genet., 1997. **6**(8): p. 1251-1260.
92. Tang, Y., et al., *Maintenance of human rearranged mitochondrial DNAs in long-term cultured transmitochondrial cell lines*. Mol. Biol. Cell, 2000. **11**(7): p. 2349-2359.
93. Kvist, L., et al., *Paternal leakage of mitochondrial DNA in the great tit (*Parus major*)*. Mol. Biol. Evol., 2003. **20**: p. 243-247.
94. Sutovsky, P., et al., *Ubiquitinated sperm mitochondria, selective proteolysis, and the regulation of mitochondrial inheritance in mammalian embryos*. Biol. Reprod., 2000. **63**: p. 582-590.
95. Kaneda, H., et al., *Elimination of paternal mitochondrial DNA in intraspecific crosses during early mouse embryogenesis*. Proc. Natl. Acad. Sci. USA, 1995. **92**: p. 4542-4546.
96. Shitara, H., et al., *Selective and continuous elimination of mitochondria microinjected into mouse eggs from spermatids, but not from liver cells, occurs throughout embryogenesis*. Genetics, 2000. **156**: p. 1277-1284.

97. Kondo, N., et al., *Incomplete maternal transmission of mitochondrial DNA in Drosophila*. Genetics, 1990. **126**(3): p. 657-663.
98. Gyllensten, U., et al., *Paternal inheritance of mitochondrial DNA in mice*. Nature, 1991. **352**: p. 255-257.
99. Magoulas, A. and E. Zouros, *Restriction-site heteroplasmy in anchovy (Engraulis encrasicolus) indicates incidental biparental inheritance of mitochondrial DNA*. Mol. Biol. Evol., 1993. **10**: p. 319-325.
100. Meusel, M.S. and R.F. Moritz, *Transfer of paternal mitochondrial DNA during fertilization of honeybee (Apis mellifera L.) eggs*. Curr. Genet., 1993. **24**: p. 539-543.
101. Nugent, J.M. and J.D. Palmer, *RNA-mediated transfer of the gene coxII from the mitochondrion to the nucleus during flowering plant evolution*. Cell, 1991. **66**: p. 476-481.
102. Van Den Boogaart, P., P. Samallo, and E. Agsteribbe, *Similar genes for a mitochondrial ATPase subunit in the nuclear and mitochondrial genomes of Neurospora crassa*. Nature, 1982. **298**: p. 137-189.
103. Farrelly, F. and R.A. Butow, *Rearranged mitochondrial genes in the yeast nuclear genome*. Nature, 1983. **301**: p. 296-301.
104. Gellissen, G., et al., *Mitochondrial DNA sequences in the nuclear genome of a locust*. Nature, 1983. **301**(631-634).
105. Jacobs, H.T., et al., *Mitochondrial DNA sequences in the nuclear genome of Stronglyocentrotus purpuratus*. J. Mol. Biol., 1983. **165**: p. 609-632.
106. Louis, E.J. and J.E. Haber, *Evolutionarily recent transfer of a group I mitochondrial intron to telemere regions in Saccharomyces cerevisiae*. Curr. Genet., 1991. **20**: p. 411-415.
107. Kamimura, N., et al., *Three separate mitochondrial DNA sequences are contiguous in human genomic DNA*. J. Mol. Biol., 1989. **210**(4): p. 703-7.
108. Thorsness, P.E. and T.D. Fox, *Escape of DNA from mitochondria to the nucleus in Saccharomyces cerevisiae*. Nature, 1990. **346**: p. 376-379.
109. Doersen, C.J., et al., *Characterization of an RNase P activity from HeLa cell mitochondria. Comparison with the cytosol RNase P activity*. J Biol Chem, 1985. **260**(10): p. 5942-9.
110. Stoneking, M., *Hypervariable sites in the mtDNA control region are mutational hotspots*. Am. J. Hum. Genet., 2000. **67**: p. 1029-1032.
111. Howell, N. and C. Bogolin Smejkal, *Persistent heteroplasmy of a mutation in the human mtDNA control region: hypermutation as an apparent consequence of simple-repeat expansion/contraction*. Am. J. Hum. Genet., 2000. **66**: p. 1589-1598.
112. Awadalla, P., A. Eyre-Walker, and J.M. Smith, *Linkage disequilibrium and recombination in hominid mitochondrial DNA*. Science, 1999. **286**: p. 2524-2525.
113. Eyre-Walker, A., N.H. Smith, and J.M. Smith, *How clonal are human mitochondria?* Proceedings of the Royal Society of London Series B-Biological Sciences, 1999. **266**: p. 477-483.
114. Hagelberg, E., et al., *Evidence for mitochondrial recombination in a human population of island Melanesia*. Proceedings of the Royal Society of London Series B-Biological Sciences, 1999. **266**: p. 485-492.
115. Hagelberg, E., et al., *Evidence for mitochondrial DNA recombination in a human population of island Melanesia: correction*. Proceedings of the Royal Society of London Series B-Biological Sciences, 2000. **267**(1452): p. 1595-1596.
116. Macaulay, V., M. Richards, and B. Sykes, *Mitochondrial DNA recombination-no need to panic*. Proceedings of the Royal Society of London Series B-Biological Sciences, 1999. **266**: p. 2037-2039.
117. Jorde, L.B. and M. Bamshad, *Questioning evidence for recombination in human mitochondrial DNA*. Science, 2000. **288**: p. 1931.
118. Kivisild, T. and R. Villems, *Questioning evidence for recombination in human mitochondrial DNA*. Science, 2000. **288**(1931).

119. Kumar, S., et al., *Questioning evidence for recombination in human mitochondrial DNA*. Science, 2000. **288**: p. 1931.
120. Parsons, T.J. and J.A. Irwin, *Questioning evidence for recombination in human mitochondrial DNA*. Science, 2000. **288**: p. 1931.
121. Ingman, M., et al., *Mitochondrial genome variation and the origin of modern humans*. Nature, 2000. **408**: p. 708-713.
122. Pesole, G., et al., *Nucleotide substitution rate of mammalian mitochondrial genomes*. J. Mol. Evol., 1999. **48**(4): p. 427-434.
123. Ballard, J.W.O., *Comparative genomics of mitochondrial DNA in members of the Drosophila melanogaster subgroup*. J. Mol. Evol., 2000. **51**(1): p. 48-63.
124. Yang, Z.H., et al., *Condon-substitution models for heterogeneous selection pressure at amino acid sites*. Genetics, 2000. **155**(1): p. 431-449.
125. Perna, N.T. and T.D. Kocher, *Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes*. J. Mol. Evol., 1995. **41**(3): p. 353-358.
126. Smith, G.R., *Hotspots of homologous recombination*. Experientia, 1994. **50**(3): p. 234-241.
127. Guillon, H. and B. de Massy, *An initiation site for meiotic crossingover and gene conversion in the mouse*. J. Immunol., 1998. **160**: p. 266-272.
128. Kohli, A., *Molecular characterization of transforming plasmid rearrangement in transgenic rice reveals a recombination hotspot in the CaMV promoter and confirms the predominance of microhomology mediated recombination*. Plant J., 1999. **17**(6): p. 591-601.
129. Jiang, R., et al., *Genome-wide evaluation of the public SNP databases*. Pharmacogenomics, 2003. **4**(6): p. 779-789.
130. Stoneking, M., *Single nucleotide polymorphisms: from the evolutionary past*. Nature, 2001. **409**: p. 821-822.
131. Chakravarti, A., et al., *Nonuniform recombination within the human B-globin gene cluster*. Am. J. Hum. Genet., 1984. **36**: p. 1239-1258.
132. Smith, R.A., et al., *Recombination breakpoints in the human B-globin gene cluster*. Blood, 1998. **92**: p. 4415-4421.
133. Yip, S.P., et al., *Mapping recombination hotspots in human phosphoglucomutase (PGM1)*. Hum. Mol. Genet., 1999. **9**: p. 1699-1706.
134. Templeton, A.R., et al., *Recombinational and mutational hotspots within the human lipoprotein lipase gene*. Am. J. Hum. Genet., 2000. **66**: p. 69-83.
135. Cullen, M., et al., *Molecular mapping of a recombination hotspot located in the second intron of the human TAP2 locus*. Am. J. Hum. Genet., 1995. **56**: p. 1350-1358.
136. Cullen, M., et al., *Characterization of recombination in the HLA class II region*. Am. J. Hum. Genet., 1997. **60**: p. 397-407.
137. Gabriel, S.B., et al., *The structure of haplotype blocks in the human genome*. Science, 2002. **296**(5576): p. 2225-9.
138. Daly, M.J., et al., *High-resolution haplotype structure in the human genome*. Nat. Genet., 2001. **29**(2): p. 229-32.
139. Phillips, M.S., et al., *Chromosome-wide distribution of haplotype blocks and the role of recombination hot spots*. Nat. Genet., 2003. **33**(3): p. 382-7.
140. Wall, J.D. and J.K. Pritchard, *Haplotype blocks and linkage disequilibrium in the human genome*. Nat. Rev. Genet., 2003. **4**(8): p. 587-97.
141. Jeffreys, A.J., et al., *Meiotic recombination hot spots and human DNA diversity*. Philos. Trans. R. Soc. Lond. B Biol. Sci., 2004. **359**(1441): p. 141-152.
142. Jeffreys, A.J., A. Ritchie, and R. Neumann, *High resolution analysis of haplotype diversity and meiotic crossover in the human TAP2 recombination hotspot*. Hum. Mol. Genet., 2000. **9**(5): p. 725-733.
143. Cullen, M., et al., *High-resolution patterns of meiotic recombination across the human major histocompatibility complex*. Am. J. Hum. Genet., 2002. **71**: p. 759-776.

-
144. Tishkoff, S.A. and B.C. Verrelli, *Role of evolutionary history on haplotype block structure in the human genome: implications for disease mapping*. *Current Opinion in Genetics and Development*, 2003. **13**(6): p. 569-75.
145. Cardon, L.R. and G.R. Abecasis, *Using haplotype blocks to map human complex trait loci*. *Trends Genet.*, 2003. **19**(3): p. 135-40.
146. Colina, R., et al., *Evidence of intratypic recombination in natural populations of hepatitis C virus*. *J Gen Virol*, 2004, **85**: p. 31-37.