Preface

Genetic recombination is the process by which a strand of genetic material (usually DNA; but can also be RNA) is broken and then joined to a different DNA molecule. In eukaryotes recombination commonly occurs during meiosis as chromosomal crossover between paired chromosomes. This process leads to offsprings having different combinations of genes from their parents and can produce new chimeric alleles. In evolutionary biology this shuffling of genes is thought to have many advantages, including that of allowing sexually reproducing organisms to avoid Muller's ratchet. However, a recombination pathway in DNA is any way by which a broken DNA molecule is reconnected to form a whole DNA strand.

In molecular biology "recombination" can also refer to artificial and deliberate recombination of disparate pieces of DNA, often from different organisms, creating what is called recombinant DNA.

Enzymes called recombinases catalyze natural recombination reactions. RecA, the recombinase found in E. coli, is responsible for the repair of DNA double strand breaks (DSBs). In yeast and other eukaryotic organisms there are two recombinases required for repairing DSBs. The RAD51 protein is required for mitotic and meiotic recombination and the DMC1 protein is specific to meiotic recombination.

This new book presents the latest research in the field.

In response to DNA damage, the cell institutes an orchestrated and measured response. Central to this is the canonical DNA damage response pathway. The components of this pathway are many and include ATM, MRE11, RAD50, NBS1, CHEK2, p53, GADD45a, p21, BRCA1, BRCA2, FA proteins, BLM and WRN. Depending on cell cycle and damage type, different combinations of these components are utilized to affect an appropriate cellular response, including DNA repair. Absence of any one of these components is largely associated with clinical diseases, severe developmental defects, an increased propensity to develop cancer at an early age, and/or aging defects. Homologous recombination repair is a DNA repair mechanism that is capable of repairing a wide range of damages. It has been determined that homologous recombination repair is altered in the absence of almost every component of the damage response pathway in either a direct or indirect fashion or by functional implication. In Chapter 1 the authors will present the evidence for the intimate relationship of the damage response pathway, cell cycle and homologous recombination repair based on *in vivo* models and discuss the implications of their intertwining in mammalian systems.

Inducible gene expression systems offer great potential toward achieving a wide variety of basic and applied biomedical research goals. Current inducible gene technologies are based on binary transgenic systems in which the expression of a target gene is dependent upon the activity of a unique transcriptional activator in the presence or absence of its drug ligand.

Despite the conceptual simplicity, most gene inducible strategies suffer from a number of limitations, such as leaky basal transcriptional activity when uninduced and poor dynamic control of the system. Because of these shortcomings, regulated gene expression technologies are still being evaluated using stringent genetic studies and none are currently used in human gene therapy. To overcome some of the disadvantages of the current systems, the authors describe the development of a next-generation inducible gene expression system, which relies upon *multiple* heterologous transactivators and their respective ligands. Their system would provide negative/positive selection for transgene expression in its induced state that would create virtually no transgene expression when not induced (i.e., no leakage). In Chapter 2 the authors propose a strategy to engineer expression systems that include heterodimeric hormone receptors combined with other heterologous transactivators.

Genetic recombination, originating from the occurrences of crossovers during meiosis, may be seen as the foundation of genetic variation, i.e. underlying both population diversities and general evolution in itself. Moreover, it is the driving force behind the original formation of genetic material forming the genetic components corresponding to, human or other, diseases. The recombination process generally imply a low level of recombinations, given a fixed number of meiotic generations, with respect to closely located loci; being the basic mechanism facilitating gene mapping studies through using collected data in the form of genetic information on multigenerational families (pedigrees). A major role behind such mapping is played by a relevant set of statistical techniques, methods and algorithms; developed using tools and theory from, for instance, mathematics, computer science and genetics.

In linkage analysis or, in a wider sense, gene mapping one searches for disease loci along a genome. This is done by observing so called marker genotypes and phenotypes of a pedigree set, i.e. a set of pedigrees, in order to locate the loci corresponding to the underlying disease genes or, at least, to narrow down the interesting genome regions. In this context the key concept is the genetic inheritance of alleles and its correlation to phenotype outcomes; which is informative according to genetic recombination. A significant deviation from what is expected under random inheritance is taken as statistical evidence of existing genetic components suggested to be located at, or close to, the loci giving significant results.

Chapter 3 mainly serves as a (subjective) review on gene mapping in general and nonparametric linkage analysis in particular, but also includes some material on recent or new research. The authors will follow a path vaguely outlined as: Firstly, they begin by outlining the needed genetical foundation of statistical genetics as well as some basic concepts, for instance, noting on the process of allelic inheritance, the genetic model, the pedigree set, the inheritance vector and various types of genetic information. Secondly, they give an introduction to one-locus nonparametric linkage (NPL) analysis focusing on significance calculations of NPL scores. Corresponding approaches are based either on analytical approximations or Monte Carlo simulations. Thirdly, the authors make some comments on the generalizations to two-locus procedures; both on the unconditional (simultaneous) and conditional (sequential) search for two disease loci. Along the way they also at some length discuss, for instance, parametric linkage analysis, fuzzy significance, multiple testing and

evidential interpretation of significant findings. Fourthly, they very briefly discuss some competing and complementary subfields within the context of gene mapping-based statistical genetics.

Discovering the impact of biological processes, such as lateral gene transfer, has significantly transformed our understanding of microbial evolution and our views of the natural genetic relationships between diverse life forms. As the diversity of evolutionary processes are revealed, real natural connections appear to be much more complex than initially believed at the time traditional molecular phylogenetics assigned itself the task to reconstruct the universal Tree of life, aka the unique genealogy of species, based on the use of carefully selected genes. Both the notions of species and of a unique inclusive hierarchy of living beings needs in fact to be reevaluated. Importantly, such a reevaluation calls for a substantial renewal of our phylogenetic practices and focuses, opening new perspectives to the whole field of evolutionary biology. Here, these changes are justified by recalling an important recent lesson in the philosophy of biology: how the consideration of evolutionary processes, necessary for species definition, indicates that many incompatible but legitimate definitions of species taxa and taxonomies are more realistic when it comes to represent life's natural relationships. In Chapter 4 the authors discuss why, in addition, many other evolutionary units, smaller or larger than the usual species taxa, have also to be considered as real, because they too emerge from the evolutionary process. Thus, even though the trajectories of these multiple evolutionary units may conflict, they deserve equally to be fully investigated by phylogenetics. Instead of the use of a unique tree, the authors explain how the consideration of multiple databases could help properly systematize a portion of the real biological diversity. In addition, consideration of networks could partly help investigate the dynamics sustaining the natural genetic connections between all the evolutionary units. Such analyses go beyond the scope of traditional phylogenetics, yet they matter, because, while some evolutionary units are "closed", others are "open", vastly changing, thus presenting fuzzy rather than precisely definable boundaries. They argue why allowing such a distinction at all evolutionary levels has important bearing on our ability and ways to describe life's evolution, which can not be accounted for by a unique model. Finally, after encouraging the development of additional evolutionary metaphors to describe the true course of evolution, the authors briefly present implications for biotechnologies and conservation biology that makes this pluralistic phylogenetics particularly valuable and promising.

Chapter 5 describes the molecular mechanisms which underly joining of human and animal retroviral DNA to host cell DNA. The authors discuss contributions of the viral protein integrase, as well as cellular co-factors, to the efficiency of integration and selection of integration sites in the host cell genome. They also address the questions related to the treatment of HIV (human immunodeficiency virus) infection, since the HIV-1 integrase is an attractive target for the development of anti-HIV-1 therapeutics. Finally, the authors present opportunities as well as the broblems associated with the integration of retroviral vectors, which are used in gene therapy applications.

ϕC31 integrase is a site specific recombinase derived from the *Streptomyces* phage. In the phage lifecycle, the enzyme mediates lysogeny by mediating recombination between specific sequences termed *attB* (present in the bacterial DNA) and *attP* (present in the phage genome). Screening the enzyme activity in mammalian cells provided positive results and also showed that the enzyme retained its property of site specific recombination into mammalian genomes. Mammalian genomes have been shown to contain sequences that are similar to the wild type

attP sequence of the *Streptomyces* phage genome and experiments with the integrase in mammalian cells showed that it could mediate recombination and subsequent integration of any DNA bearing an *attB* site into these pseudo*attP* sites. These properties have made ϕC31 integrase emerge as a promising tool for nonviral gene therapy to achieve long-term gene expression in different tissues *in vitro* and *in vivo*. Chapter 6 introduces the different recombination mediating enzymes but focuses primarily on the activity of ϕC31 integrase in different tissues. Results of this enzyme system in the lung tissue and hematopoietic cells are also discussed. In murine and human lung cell lines, it could be shown that the ϕC31 integrase mediates specific recombination between *attB* and *attP* and subsequently long-term gene expression could be achieved without any selection pressure. The results were further validated *in vivo* in mice. However, when integration into a specific site "mpsL1" was investigated, it was revealed that this site was targeted only in 50% of the mice. This was in contrast to the reports from other tissues where all the mice showed integration at mpsL1. These results indicated that the activity of this enzyme may actually be tissue dependent.

Tissue specific efficiency of the ϕC31 integrase was confirmed in the human hematopoietic system. The activity of the ϕC31 integrase was extremely reduced in $CD34^+$ hematopoietic stem cells, primary T lymphocytes and T cell derived cell lines in comparison with mesenchymal stem cells and cell lines derived from lung -, liver - and cervix tissue. No enhanced long-term expression mediated by the ϕC31 integrase could be observed in T cell lines. A direct comparison of hematopoietic Jurkat T and A549 alveolar type-II lung cell lines indicated up to a 100-fold higher activity of the ϕC31 integrase in lung tissue compared to hematopoietic cells. Looking for possible mechanisms responsible for this discrepancy of ϕC31 integrase activity in different tissues, revealed that Jurkat cells contain significantly higher amounts of DAXX protein compared to A549 cells. As DAXX has been reported to interact with the ϕC31 integrase and inhibit recombination, higher levels in hematopoietic cells may be one reason for low activity of ϕC31 integrase in these cells.

These studies in lungs and the hematopoietic cells provide evidence for the tissue specificity of the ϕC31 integrase system and also raise the need for development of novel strategies for achieving recombination and long-term expression.

Recombination is a key evolutionary mechanism that should not be ignored. It is directly related with the amount of linkage disequilibrium, which is important in order to characterize populations from an evolutionary point of view and to localize genes in humans and other organisms. Recombination effects are not independent of other evolutionary forces, such as natural selection and genetic drift. Recently, new methods became available to infer positive selection in the presence of recombination and *viceversa*. These new methods are computationally very intensive and as data sets become larger, their analysis becomes unavoidable. Therefore, faster and efficient algorithms are needed to estimate recombination at the fine-scale and genomic level. Such approaches need also to disentangle recombination and natural selection signals on DNA. Additionally, it is usually difficult to evaluate the adequacy of the methods since there are few simulation tools capable to produce data under complex evolutionary models. Consequently, developing programs that allow simulating both, natural selection and recombination, is also necessary.

In what follows in Chapter 7 the authors will review recently developed algorithms to estimate recombination at the fine scale. In addition, they will also consider approaches that allow the simultaneous estimation of recombination and selection. The authors will explore

possible future directions for recombination and selection estimation research. They will stress the importance of incorporating recombination jointly with other genetic factors both in evolutionary and epidemiological contexts, e.g. to model drug resistance emergence. In this framework, they will present a new result concerning the faster evolution of resistance favoured by the minimum co-infection rate combined with the higher recombination value. Finally, the authors will consider simulation tools. They will briefly point out how to simulate DNA sequences under a specific nonsynonymous/synonymous (dN/dS) ratio and recombination rate (inter and intracodon level). In doing this they will introduce a new method of forward simulation and an efficient way of simulating different substitution models forward in time.

As explained in Chapter 8, direct observations of chiasmata showed that there were chromosome intervals where only single chiasmata appeared and other intervals where rather chiasmata appeared which were accompanied by other chiasmata elsewhere. Particularly, double chiasmata were involved in both chromosome ends, but single chiasmata only in one end. There is no model of a chiasma process developed so far which is appropriate to such events. Therefore, a new model was developed here to infer the phenomenon. It takes into account suppression interference and allows the determination of the chiasma number distribution and the distribution of the chiasma locations for each number of appearing chiasmata. It was shown with an example that the novel model explains the chiasma formation process much better than other models.

Chapter 9 presents the disease Fanconi anemia (FA) is a cancer predisposition disorder involving progressive anemia, caused by deficiency in any of 13 known (FANC) genes. At the cellular level FA is a classical chromosomal instability disease associated with sensitivity to DNA damage, particularly interstrand crosslinks (ICLs). Unlike classical DNA repair pathways, it is not understood how most of the FANC proteins function to maintain chromosome stability, but many of them are necessary for the monoubiquitylation of two of the FANC proteins, FANCD2 and FANCI, both posttranslational modifications that are considered necessary for a functional FA "pathway". Through genetic and biochemical studies, this pathway is variously implicated in a number of key S-phase events such prevention of replication fork breaks by promotion of translesion synthesis (TLS), and the repair of broken chromatids through both of the double-strand break (DSB) repair pathways: homologous recombination repair (HRR) and non-homologous end joining (NHEJ). However, many published results concerning the functional links between the FA pathway and the two mutually exclusive DSB repair processes are, not surprisingly, contradictory. Recent use of the mutagenesis model system of hamster CHO cells, with knock-out clones defective in one of the FANC proteins (FANCG), or an HRR protein (Rad51D), have been highly informative in distinguishing functional differences between the FA and HRR pathways. Importantly, the experimental precision of CHO cells has clarified seemingly conflicting mutagenesis measurements in human FA cells, and implies contribution of the FA pathway in promoting all three replication-associated damage-recovery processes: TLS, HRR, and NHEJ. FA cells, therefore, are not completely deficient in any one of these processes, but are less efficient in all of them, suggesting a role for the FA proteins in co-ordination or optimization of these genome stabilizing processes, rather than direct participation in any one of them.

Chapter 10 discusses recombination, which might be an evolutionary development as ancient as the origin of life. In spite of numerous data on the mechanisms of genetic

recombination in living organisms since the discovery of Rec mutations and of the existence of genetic exchanges in bacteria, some genetic mechanisms in these organisms remain not fully understood, notably in relation to the genetic diversity of bacteria and to their position with respect to the very first organisms having appeared on earth, ancestors of both eukaryotes and prokaryotes. Spontaneous zygogenesis (or Z-mating), recently discovered in *Escherichia coli*, is intriguing, as it resembles fusion of gametes in eukaryotes in that it involves complete genetic mixing. It is also question on how genetic exchanges can enable bacteria having undergone severe deletions or temporary chromosomal inactivation to survive and overcome what could be a selective disadvantage.

In the past, studying of the gene was very difficult. Genetic laboratory seems to be a complicated and mysterious field. However, the blooming of molecular biology leads to many simplified techniques for genetic studying. Finding the amino acid sequence of a gene can be easily performed by basic sequencing technique. At the end of the 20's century, the completion of the genome project lead to a new era of post genomics. The in silico laboratory is an advent in the post genomics era. Based on in silico techniques, manipulation on a genetic sequence is easy. This can help us to better understand the genetic recombination phenomenon. To study a genetic recombination, in silico mutating and docking can help create a genetic recombinant. In addition, in silico gene expression analysis such as gene ontology techniques can help identify the changes in phenotypic expression of a designed recombinant. In Chapter 11, principles and interesting examples of in silico genetic recombinant studies will be briefly discussed.

Pseudotyping lentiviral vector with other viral surface proteins could be applied for treating genetic anomalies in human skin. In Chapter 12, the modification of HIV vector tropism by pseudotyping with the envelope glycoprotein from vesicular stomatitis virus (VSV), the Zaire Ebola (EboZ) virus, murine leukemia virus (MuLV), lymphocytic choriomeningitis virus (LCMV), Rabies, or the rabies-related Mokola virus encoding *LacZ* as a reporter gene was evaluated qualitatively and quantitatively in human skin xenografts. High transgene expression was detected in dermal fibroblasts transduced with VSV-G-, EboZ-, or MuLV-pseudotyped HIV vector with tissue irregularities in the dermal compartments following repeated injections of EboZ- or LCMV-pseudotyped vectors. Four weeks after transduction, double-labeling immunofluorescence of β-galactosidase and involucrin or integrin β1 demonstrated that VSV-G-, EboZ-, or MuLV-pseudotyped HIV vector effectively targeted quiescent epidermal stem cells and their progenies, which were expressed dorsally and underwent terminal differentiation. Among the six different pseudotyped HIV-based vectors evaluated, VSV-G-pseudotyped vector was found to be the most efficient viral glycoprotein for cutaneous transduction as demonstrated by the highest level attained in β-galactosidase activity and genome copy number evaluated by *Taq*Man PCR.

The central nervous system comprises neuron networks, in which enormously diversified neurons connect and interact with each other. For decades, the diversity of such neurons has been hypothetically ascribed to a gene diversification mechanism similar to that of the antigen receptor genes in the immune system. Synaptogenesis, memory, and odoreceptor diversification have been raised as candidate representatives of the neuronal functions involved in DNA rearrangements mediated by the hypothetical somatic DNA recombination mechanism in the brain. Some reports have described the somatic DNA recombination activity and the possible rearranging gene loci in association with these neuronal functions in the brain. In spite of every effort to search for the associated gene loci for traces of gene

rearrangement, no physiologically functional gene rearrangement has yet been identified. Two rearranged genomic regions were, however, identified in the brain by a rearranged genomic region-oriented approach. A repetitive genomic region (LINE) and a non-repetitive genomic region (BC-1) are the only known examples to undergo genomic rearrangement in the brain, thus far. Both of the regions have been observed to undergo DNA rearrangement not only in the brain and but also in the lens, thus implying that these DNA rearrangements are associated with ectodermal development. The relationship between the neighboring gene function and the rearrangement of the non-repetitive genomic region is discussed in Chapter 13.

As explained Innocuous loss of intron or *in situ* loss of intron is common during evolution. Together with intron gain, loss of intron has re-shaped many genomes dramatically, and has changed many gene structures together with intron gain. Five modes of intron loss can be defined in a multiple-intron gene: complete loss of all introns, 3′-biased loss, concerted loss of several internal adjacent introns, intron exclusion and multiple intron exclusion. The cDMHR/DSBR model presented in Chapter 14, which the authors recently established, can accommodate all the five patterns of intron losses. In this model, cDNA undergoes homologous recombination (HR) with its parent intron-containing genomic copy. This cDNA recombination is facilitated by the HR repair machinery of DSB in the cell. This process can be triggered by a DSB in a specific intron, which results in the loss of the very intron suffering the DSB. The reverse transcriptase activity could be from retrotransposon such as the yeast Ty1 element and possibly the mammalian LINE. DSB in intron, retrotransposon-encoded reverse transcriptase and homologous repair machinery might be the long searched driving forces for *in situ* intron elimination. This model is strongly supported by the independent experimental data from yeast in which cDNA recombination with the corresponding genomic copy is directly demonstrated, and this process is dependent on the HR protein RAD52. The widely used gap-repair technique also supports this model.

Genetic recombination is an important phenomenon in gene medicine. This phenomenon can lead to a new genotype and phenotype. Genetic recombination can be either natural or artificial. Natural genetic recombination is an important contributing factor to genetic shift and drift. In medicine, the genetic recombination in pathogenesis of an existing disease is important. Considering three epidemiological determinants, genetic recombination of agent is easier than that of host and environment. For pathogens, natural genetic recombination can result in a change in virulence and susceptibility. Natural genetic recombination in the virus is well characterized in clinical microbiology. In Chapter 15, the author performed a literature review on the natural genetic recombination of pathogens. Examples of important genetic recombination of pathogens and their clinical correlations are presented.

Chapter 16 discusses a randomized Spruce Budworm model with Holling III Functional Response . The authors show that the positive solution of the associated stochastic differential equation does not explode to infinity in a finite time. The authors proof the existence and uniqueness of the positive solutions.In addition, Uniformly Continuous of solution is studied.