

The Revolution of the Biology of the Genome

WOLFGANG HENNIG

*Shanghai Institutes for Biological Sciences, DAAD Laboratory, Chinese Academy of Sciences,
320 Yue Yang Road, Shanghai 200031, P.R. China
E-mail: whennig@sibs.ac.cn*

ABSTRACT

Sequence data of entire eukaryotic genomes and their detailed comparison have provided new evidence on genome evolution. The major mechanisms involved in the increase of genome sizes are polyploidization and gene duplication. Subsequent gene silencing or mutations, preferentially in regulatory sequences of genes, modify the genome and permit the development of genes with new properties. Mechanisms such as lateral gene transfer, exon shuffling or the creation of new genes by transposition contribute to the evolution of a genome, but remain relatively restricted relevance. Mechanisms to decrease genome sizes and, in particular, to remove specific DNA sequences, such as blocks of satellite DNAs, appear to involve the action of RNA interference (RNAi). RNAi mechanisms have been proven to be involved in chromatin packaging related with gene inactivation as well as in DNA excision during the macronucleus development in ciliates.

Key words: Genome structure, gene duplication, RNAi, gene silencing, chromatin

INTRODUCTION

It is a general conception that the discovery of the structure of DNA and the understanding of its basic functions in the cell, as established in the second half of the 20th century, has been the central achievement in biology of the last 5 decades: This knowledge has taught us the basic principles of inheritance. Only recently we start to realize that the increasing knowledge on DNA has much more fundamental consequences for understanding biology. The results of the sequencing of an increasing number of complete genomes and the comparative evaluation of information in the genomic DNA sequences has opened new and unexpected aspects, which influence all fields of biology.

The two probably most important of these aspects emerging from the genomic sequencing data concern the evolution of genomes and the functional organization of genomes. The recent sequencing data of genomes reveal also that our picture on the genome changes considerably if high accuracy sequencing data are compared with the conclusions drawn from earlier shotgun sequencing approaches [1-2]. The availability of databases on DNA sequences represented at the RNA level - such as EST and cDNA databases - and the comparisons of these data with the sequences of the complete genome has as well yielded unexpected views on the extent of transcription of the genome. Together with the discovery of RNAi this opens revolutionary views on eukaryotic genomes - their evolution as well as their function. What are these new views and what are their consequences for future research in biology?

The analysis of the genomic sequences has also stimulated the development of new techniques which will allow the functional evaluation of the genome. Considerable parts of prior laboratory work have been shifted to the computer. Comparisons of nucleic acid sequences, which could earlier only be done by nucleic acid hybridization experiments, can now be carried out at much larger scales, with better resolution and with exactly determined precision by sequence searches and by the alignments of nucleic acid sequences as they are available in databases. As a consequence new concepts for handling of biological data are required. At present, such concepts are, however, hardly in their initial phases of development. Future research in biology will be substantially based on tools of bioinformatics but we are not yet sufficiently prepared for this in terms of the state of bioinformatics.

On the experimental side, new laboratory techniques concern especially three fields of biology. The investigation of the transcriptional aspects of genomes - summarized under the term transcriptome - has yielded possibilities to compare gene expression at the RNA level at a large scale by the development of *microarray techniques* for nucleic acid sequences. Current approaches are directed towards the registration of gene activities and their changing patterns at all cellular and developmental levels. These approaches are complemented by studies of RNA expression by in situ hybridization. Such studies are systematically performed, for example, on *Drosophila* embryos by collecting the expression patterns of all available transcripts using whole mount in situ hybridization on embryos. The data are deposited in databases and are freely available to researchers.

A comparable approach is made at the protein level. Proteomics tries to collect expression patterns of proteins in a way similar to the recording of transcripts by microarrays. The techniques of proteomics are still less elaborate than those of RNA investigations, but new developments like protein microarrays will soon enhance the power of this field. Even though this approach appears not very different from that towards the transcriptome, proteomics is faced with the fact of a much higher complexity of its targets: Proteins exist in many modifications (for example in phosphorylated, acetylated, methylated forms or in many other modifications) and in multiple splicing variants. It becomes evident that it will be necessary to register and functionally evaluate millions of protein molecules in their cellular context and in their respective relationship. The development of bioinformatic tools for handling such information and for the use of it in multiple biological contexts will be one of the most intriguing and demanding tasks in future biology.

A third field, which has developed rapidly in recent years, are the diverse microscopy techniques such as confocal microscopy, laser scanning microscopy and various techniques in electron microscopy. An example is the atomic force microscope (Binnig et al. 1986 [2]). The potential to localize macromolecules with very high sensitivity and precision in the cell or even in subcellular compartments and to analyze their localization and dynamics - which can, for example, be achieved by photo bleaching - in the cell in a three-dimensional fashion has provided new insights in cellular processes and will be essential for research in cellular and developmental biology in future. Added to these techniques are new genetic technologies which allow to introduce and express molecules, marked for example with fluorescent label (like Green Fluorescent Protein, GFP, and its variants), in a controlled fashion and with increasing precision and efficiency to manipulate single genes. This permits to establish the precise time of expression of molecules, the subcellular localization and the developmental pattern of distinct molecules, even if present in small amounts in single cells. Such techniques will be essential for integrating our knowledge and to transfer it onto a cellular or tissue level.

In the subsequent part of this review I shall concentrate on two aspects of the recent developments in molecular biology, the changing views on evolution of the eukaryotic genome, based on genome research, and on the consequences of RNAi mechanisms for understanding regulatory mechanisms at the genome level.

Genomes: A new world

To Charles Darwin probably hardly anything in modern biology would have been more exciting than our new insights in the evolution of genomes. Though biologists (except the few political indoctrinated cranks like those having promoted Lyssenkoism) during the last century have not questioned the descent theory as the basis for understanding biological evolution, the analysis of DNA sequencing data has yielded many unexpected facts in additional support of Darwin's descent theory. These facts can very simply be summarized by the statement that probably most - if not all - of the major genetics traits have evolved very early on an evolutionary scale and have been maintained throughout the history of all phyla. This holds not only true for fundamental cellular mechanisms such as DNA replication and repair, cell cycle control and the basic mechanisms of mitosis and meiosis but also, for example, for major principles of the early embryonic development of organisms. Such principles are conserved as is known since years from the study of genes involved in early *Drosophila* development and segmentation [3], review: [4]. The amazing maintenance of genetic tools in evolution reaches, however, much further: The development of entire organs is also based on very old principles as has first been demonstrated for the development of the eye [5]. While classical comparative anatomy taught us that the ingenious invention of the eye has by nature been achieved several times independently in analogous ways (in insects, cephalopods, mammals), this dogma has now been proven wrong. The principle genes required for the construction of an eye are the same and have been designed early in evolution, even though the details of their involvement in the construction of eyes in different phyla might be different. But even more surprising: The key gene for eye development recovered from the mouse genome (*Aniridia*) can be used to create an insect eye in *Drosophila* in ectopic positions or even replace the respective *Drosophila* gene (*eyeless*) [6]. Similar conclusions on evolutionary conservation of developmental pathways have been made for other gene products, as for example heart myosin, or for key genes in development like *engrailed* [7] and have recently been proven on a more universal level by general expression studies of groups of related genes [8].

One might argue that the evolutionary conservation of cellular processes such as cell cycle regulation, chromosome separation or DNA repair is not so unexpected as they define the basis of the function of a cell. But that the development of morphologically widely different organs has a similar ancient basis, is probably one of the most intriguing findings in modern biology. Such conclusions will have substantial consequences in practical and theoretical terms: The investigation of simple model organisms can supply us with the basis of knowledge which can be applied with little or no changes to mammals or even men [9], where experiments are complicated, time consuming, expensive or not even be possible [10]. There are consequences for our judgment on the probability that diseases, such as, for example, induced by viruses, can be transferred between organisms: An example is provided by the recent corona virus problems ("SARS") [43]. Not at least, we may be induced to raise ethical questions and questions on

medical security for the use of primates for experimentation, or on the implications of gene transfer with retroviruses as vectors or organ transplantations between mammals.

Genome complexity

One of the puzzling features in the comparison of eukaryotic genomes has always been their extremely variable size [11]. How is this compatible with the idea that the principle genetic systems are very similar throughout eukaryotes? How can we understand the increasing complexity of organisms if we assume a high degree of genetic stability in terms of equipment with similar genes?

First ideas with regard to the question of maintenance of principle genetic information have been obtained long ago after it was recognized that eukaryotes possess a considerable part of their DNA sequences in multiple copies, the so-called repetitive DNA fraction (review: [11]). Initially it was assumed that repetitive DNA is mainly found in a portion of the genome called heterochromatin, which is considered genetically reactively inert (for discussion see [12]). But soon it was recognized that repetitive DNA is found all over the genome and includes genes as well. This together with the analysis of the globin genes of vertebrates had soon led to the proposal that gene duplications play an important role in evolution [13]. The analysis of complete genome sequences has substantiated this assumption and has proven that considerable proportions of the genome can be considered as duplicated DNA sequences. The rate of gene duplications in the human genome is, according to recent estimates, 0.5-1% in 1 million years [14]. Consequently, the genome would be duplicated in approximately 200 Million years, a period short in terms of evolution. Obviously, not all duplications will be fixed in the genome, but they nevertheless provide ample possibilities for selection. Duplicated genes might remain identical in their products or they might diverge giving rise to proteins of different functions. Classic examples for the latter kind of genes are the globin genes which have diverged and are expressed in distinct developmental patterns (see [15-16]) and, more recently, the HOX genes. The evolutionary history of the Hox genes is particularly well investigated and allows to identify their duplication patterns within vertebrates in detail [16]. Gene duplications and the possible mechanisms for divergence were discussed for the eye crystalline genes by Piatogorsky and Wistom [17]. Examples of genes retained in an identical or very similar sequence are the genes coding for histone proteins or for ribosomal RNAs.

The detailed analysis of genes derived from duplications has revealed remarkable features [16][18]: The divergence of duplicated genes occurs much more rapidly in regulatory regions than within the protein-coding part of the gene. This may have different consequences, dependent on the kind of mutations: If the regulatory sequence becomes nonfunctional, the gene is functionally eliminated and will degenerate by mutation. A modified regulatory region may, however, also lead to a change in the gene expression pattern and initiate the evolution of a new, specific task for this gene product. This, in turn, may be accompanied or followed by divergence of the coding DNA sequence as well [19]. The networks formed by regulatory sequences appear of a major importance for the evolutionary fate of a duplicated gene. The existence of such regulatory networks have been proposed very early by Britten and Davidson [20].

An additional mechanism contributes to duplication of (parts of) genomes: In many organisms, polyploidization has occurred [21]. Polyploidization appears often to be accompanied by a rapid evolution of the duplicated parts of the genome [22]. Some of the polyploid genes

become inactive by silencing effects (see below) while other diverge into new function, comparable to the situation described for single-gene duplications [23-25]. Polyploidization events can hence explain the surprisingly large differences in the genome sizes of closely related organisms [24]. That ploidization steps have occurred can, however, often only be recognized by a detailed analysis of the complete genome sequence.

These two genomic mechanisms - gene duplication and ploidization - seem to account for much of the evolution of genomes. There exists other possibilities affecting gene evolution which one might consider as means for a „fine adjustment“. The fragmentation of eukaryotic genes into exons and introns appears to be important not only from the viewpoint that it offers a wide range for regulatory possibilities - for example by placing enhancers and silencers into the introns - but also by the opportunity to create varieties of gene products by various combinations of exons by differential splicing. Thus, in principle one might envisage this situation as a possibility to combine different „genes“ into one transcription unit [25]. Although our evidence in this regard is scarce, this situation may be functionally important since it guarantees that the amino acid sequences of different products in the recombined exons remains identical. Hence, interactions with other molecules may remain guaranteed for a particular part of a protein even if other exons are added and enforce different functions.

The exon-intron structure of genes has also been considered in the context of recombination between the exons of different genes which may create new genes with new functions, for example by combining a DNA-binding domain in one exon with a protein-interacting domain of a different exon into a new combination. The evidence for such exon „re-shuffling“ events is limited and it may have been of limited significance in the evolution of the genome but has certainly contributed to it ([26], [27],[28]).

The question arises how a long-distance recombinations of exons might occur [29]. One possibility is gene conversion induced by restricted nucleotide similarities inducing regional pairing and subsequent recombination of DNA sequences in the context of repair processes after single- or double-stranded breaks in DNA. Another - probably more frequent - mechanism is that of transposition of DNA induced by mobile elements. The evidence of the importance of mobile DNA elements for the genome structure and hence for evolution has amply increased during the past years ([30],[31],[32],[33]) and even functional interactions between different transposons have been demonstrated [34]. Moreover, insertion of transposons can modify the expression patterns of genes [35].

Another possibility of introducing novel functions into a genome is still rather speculative. While it is clear that lateral gene transfer plays a role in prokaryotes, it is still unclear whether it can be considered as relevant for eukaryotic genome evolution. Some cases of lateral gene transfer have been documented [36] and the analysis of the complete genome sequence of *Ciona* has provided some additional indirect evidence that such events may have taken place [37]. The mechanisms of such lateral transfer of DNA sequences are more or less open to speculation [38]. But clearly, the transfer of genomic sequence integrated into (retro)viral genomes seems a realistic possibility to explain lateral gene transfer. The argument of the host specificity of viruses has become rather questionable after it seems evident that relative minor changes in viral genomes may induce the possibility of infecting new hosts. This has for example been assumed to be the case for the infection of humans by corona virus originating from animals [39] leading to SARS epidemics. Also other viruses, for example hematorhagic viruses, may evolve in similar ways and change their hosts after relatively insignificant mutations.

The examples given before emphasize the major mechanisms for genome evolution as we see them today. It is sensible to ask whether there are other ways to generate completely new genes. An answer is difficult as one cannot exclude that a gene considered as „new“ is in fact a strongly diverged duplicated gene. More likely than the accidental de novo-evolution of a gene is that new genes might evolve from the combined action of various mechanisms. For example, a protein coding DNA sequences might be transposed into a genome region which is not functional. This could lead to the inclusion of nonfunctional DNA into a protein sequence if an open reading frame exists. The resulting protein might take over novel functions in the respective organisms. An example for genes which might be derived from such a process are the antifreeze glykoprotein (AFGP) genes of *Dissostichus mawsoni* (an Antarctic teleost fish) [40]. In general, the frequency of such events would, however, to be considered low and it is not expected that such events substantially contribute to the evolution of genomes.

Genomes do not only develop by the creation of new genes but one has to assume that genes can be lost. Otherwise the genome would continuously increase in size. The comparison of the number of genes expressed in the brain of *Drosophila*, the brain of the honey bee (*Apis*) and the human brain has shown that of the 3000 conserved genes found expressed in human and *Apis* brains 100 genes are not present in the *Drosophila* genome [41]. With other words, the *Drosophila* genome seems to have lost these 100 genes. If one accepts a linear extrapolation of this number to the entire genome, one would have to conclude that the *Drosophila* genome has in total lost some 500 genes, which were present in ancestors and which are still maintained in other phylogenetic groups up to man. One might derive that genomes in one phylogenetic group can discard sets of genes which in other phylogenetic groups remain conserved. The simultaneous inactivation of the genes of a regulatory gene networks - as have been discussed before in the context of the evolution of duplicated genes - would allow such processes. We can expect to obtain deeper insight into such processes from the comparisons of further complete genome sequences. This, however, still does not reveal any clues on the molecular mechanisms which might be responsible to discard genes or even sets of genes from a genome.

Regulation of the genome size by RNAi ?

A key to an answer of the question how genes or sets of genes can be eliminated from a genome may be found in the recent development of our knowledge on the transcriptome. Two important new insights on transcription emerged over the past few years, which appear be related to one another. First, the analysis of EST and cDNA databases assembling transcripts from the mouse genome shows that at least 60% of the mouse genome is represented in transcripts [42]. Since long it was know that a considerable proportion of newly synthesized RNA never leaves the nucleus and becomes rapidly degraded. However, the extent of transcription of the genome, even though these transcripts are accumulated from the many different cell types, is unexpectedly high. Until now we do not have a clue as to the significance of this in biological terms.

A second, most intriguing new information on transcription concerns the discovery that a mechanisms called RNA interference (RNAi) plays a general role, not only in the posttranscriptional regulation of RNA levels but also in chromatin assembly processes [43],[44],[45]. RNAi involves the activity of small RNA molecules derived from larger

transcripts by controlled degradation (siRNAs: small interfering RNAs) as well as small RNA molecules specifically synthesized with the aim of controlling mRNA levels (miRNA: micro-RNAs) (review: [46]). Both types of RNA represent molecules of a length between 19 and 25 nucleotides, but they differ in their origin:

- **miRNA** is synthesized from non-protein coding DNA and are metabolized from transcripts accommodating inverted repeats. The double-stranded RNA formed by foldback is processed by an RNase III-like enzyme, highly conserved through evolution from yeast to man and higher plants, called Dicer in animals or Dicer-like in plants. The nuclease cleaves within the double-stranded RNA (dsRNA) region and releases dsRNAs of 22-25 nucleotide pairs [47]. These molecules can interact with the 3'-UTRs of transcripts and inhibit translation.

- **siRNA** is cleaved from larger RNA molecules, a process which also involves Dicer. The processing products associate into nucleoprotein complexes, including the RNA-induced silencing complex (RISC), and then binds to mRNA and cause its degradation. This process leads to a fine tuning of mRNA levels. siRNA is however not only formed from mRNAs but can also be formed from non-protein coding transcripts such as centromeric DNA transcripts and, in particular, from LTRs of transposons. In these cases siRNA becomes a constituent of chromatin. It is essential for chromatin packaging and - related to this - gene silencing processes as has recently been shown [48] for centromeric chromosome regions of *Schizosaccharomyces*.

These experiments imply that siRNA is able to control the assembly of the chromatin. The molecular mechanism is still unknown. One could assume that it is based on a recognition of the centromeric DNA sequences corresponding to the siRNA sequence as siRNA generally acts sequence-specifically. Over the past years it has been established that transcriptionally inactive chromosomal regions, such as telomere regions, and silenced genes, such as the mating type locus, have a closely similar composition of proteins (summarized in [49]). Specific DNA signals are required to identify chromosome regions to be packaged in a way assuring gene inactivation. It is an obvious possibility that siRNA molecules identify by homology those chromosomal sites where packaging into inactive chromatin is required. These findings also shed a new light on the character of heterochromatin as a functional part of the genome.

In general, RNAi mechanism have apparently developed because of the need to identify nucleic sequences for regulatory purposes. Initially there were probably defense mechanism against viral infections [50]. Such tasks may be closely related to another activity, ascribed to the multiple functions RNAi: the guided excision of genomic DNA sequences in ciliates. Ciliates are highly specialized single-cellular eukaryotes, distinguished by two nuclei - a generative diploid micronucleus and a vegetative, polyploid macronucleus. The peculiarity of the macronucleus is that only a minor part of the micronuclear genome is ploidyized while the major part of the micronuclear genome becomes discarded [51]. By Yao et al. [52] it has been shown that the excision of DNA sequences from the genome is directed by small RNA molecules, synthesized in the micronucleus during conjugation and early macronuclear development. The exact mechanism is still unknown, but it is obvious that one can expect a close relationship of this mechanism with the DNA packaging events described earlier.

The observations of Schramke and Allshire [52] and Yao et al. [56] open a way to understand, what has been a puzzle to geneticists for a long time: How can a genome remove DNA in a directed way and limited to specified DNA sequences? The most commonly known case is that of satellite DNAs. The high degree of variability in satellite DNA sequences between closely related species has for a long time been a question of interest without answer: How can distinct satellite DNA sequences simultaneously and quickly be removed from a genome and

replaced by a totally different sequence type (see, for example: [53])? The observations on ciliates, indicating that small RNAs are involved in sequence-directed DNA excisions, now open a way to answer this question. Not only other cases of chromatin elimination, like those in *Ascaris* and in Crustaceans (see [53]), might make use of related mechanisms: The same mechanisms could be applied to the simultaneous removal of blocks of repetitive DNA sequences, especially from centromeres, in germ cells. The genome is widely transcribed during the meiotic prophase. This makes suitable transcripts available which could be used to excise blocks of satellite DNAs or other repeated DNA sequences (see [54],[55]). Such transcripts might also be important for the tight packaging of the DNA with basic proteins at the end of the primary spermatocyte stage as a means of inactivating of the genome, similar as described by Schramke and Allshire for centromeric DNA. Under special conditions these transcripts might induce or facilitate DNA excision events comparable to those during the macronucleus in ciliates.

Conclusions

The developments in genome research during the past years have contributed fundamentally new insights in the structure and evolution of eukaryotic genomes. Even though we still do not have clear concepts for the details of the mechanisms controlling genome sizes, we at least can imagine how genome sizes can increase rapidly in evolution: gene duplication events and, in particular, ploidization are the prominent ways to enhance the amount of genomic DNA. This demands mechanisms for the controlled reduction of the genome size. RNAi mechanisms may provide means to decrease genome sizes by a sequence-specific removal of DNA sequences from the genome.

One of the most intriguing new aspects of genome research is the unexpected conservation of genetic pathways and, in particular, of regulatory key genes throughout evolution. This has consequences at many different levels. Basically it implies that organisms are more close in their genetic properties than was expected earlier. The possibility of transgression of barriers between species appears easier under such conditions and they are - opposite to more conventional views - not distinguished by very fundamental genetic differences. Minor genetic changes might determine their sexual isolation from one another. Consequences of this might become important on an applied level, especially in considering medical aspects. For example, we do not know how endogenous viruses or transposable genetic elements behave in the case of transplantation of organs. Can viruses or transposons activated in a foreign environment? If so, can they invade other cells of the host with all the possible consequences of inserting in more or less random genomic positions? One might even see the potential for lateral gene transfer in such cases as mobilized viruses or transposons might carry donor DNA sequences into germ cells of the host. The recognition that animal viruses can easily invade humans, as it appears the case of the Severe Acute Respiratory Syndrome (SARS), requests for increased caution in food production and animal breeding.

Consequences must also be drawn for bioethics: We can no further ignore that primates are genetically very similar. This will have implications for the use of primates as experimental subjects. On the other hands, in early days of gene technology it has been made a strong point from non-biologists that species are considered strictly separated in nature and that it is unethical to transgress natural barriers by transformation experiments. This arguments is obviously not a convincing arguments as the separation of species might be caused by minor genetic differences.

Even if the separation of species should be more substantial it can be argued that the major genetic processes are similar enough to doubt the justification of such arguments against gene transfer.

Considering the fast changes of our views on genome evolution during the past few years we might still expect new and unforeseen insights within the next years when more data for comparative studies become available.

REFERENCES

- [1] Yu J et al. A Draft Sequence of the Rice Genome (*Oryza sativa* L. ssp. Indica). *Science* 2002; 296:79-92.
- [2] Goff SA et al. A draft sequence of the rice genome (*Oryza sativa* L. ssp. Japonica). *Science* 2002; 296:92-100.
- [3] Binning G, Quate CF, Gerber C. Atomic force microscope. *Phys Rev Lett* 1986; 56:950-953.
- [4] Nüsslein-Volhard C, Wieschaus E. Mutation affecting segment number and polarity in *Drosophila*. *Nature* 1980; 287:795-80
- [5] Pourquie O. The segmentation clock: converting embryonic time into spatial pattern. *Science* 2003; 301:328-330.
- [6] Halder G, Callaerts P, Gehring WJ. Induction of ectopic eyes by targeted expression of the eyeless gene in *Drosophila*. *Science* 1995; 267:1788-1792
- [7] Gehring W. The master control gene for morphogenesis and evolution of the eye. *Genes-Cells* 1997; 1:11-15
- [8] Gibert JM. The evolution of engrailed genes after duplication and speciation events. *Dev Genes Evol.* 2002; 212:307-318.
- [9] Stuart JM, Segal E, Koller D, Kim SK. A Gene-Coexpression Network for Global Discovery of Conserved Genetic Modules. *Science* 2003; 302:249-255
- [10] Hennig W. *Drosophila* spermatogenesis as a model system. Symposiums-Lecture Hilscher-Symposium, Neuss 1991. *Andrologia* 1992; 24:21-26.
- [11] Pagliarini RA, Xu, A. Genetic Screen in *Drosophila* for Metastatic Behavior. *Science* 2003; DOI: 10884741-0..
- [12] Britten RJ, Kohne DE. Repeated sequences in the DNA. *Science* 1968; 161:529-540.
- [13] Pardue ML, Hennig W. Heterochromatin: Junk or Collectors Item? *Chromosoma* 1990; 100:3-7.
- [14] Ohno S, Wolf U, Atkin NB. Evolution from fish to mammals by gene duplication. *Hereditas* 1968; 59:169-187.
- [15] Lynch M, Conery JS The evolutionary fate and consequences of duplicate genes. *Science* 2000; 290:1151-1155.
- [16] Dayhoff MO. Atlas of protein sequence and structure. Vol. 5. 2000; National Biomed Res Foundation, Washington
- [17] Efstratiatis A, Posakony JE, Manitis T et al. The structure and evolution of the human β -globin family. *Cell* 1980; 21:653-668.
- [18] Prince VE, Pickett FB. Splitting pairs: the diverging fates of duplicated genes. *Nat Rev Genet.* (2002) 3:827-37.
- [19] Piatigorsky J and Wistom G. The recruitment of crystallins: new functions precede gene duplication. *Science* 1991; 252:1076-1079.
- [20] Eichler EE, Sankoff D. Structural dynamics of eukaryotic chromosome evolution. *Science.* 2003; 301:793-797.
- [21] Ohta T. Evolution by gene duplication revisited: differentiation of regulatory elements versus proteins. *Genetica* 2003; 118:209-216.
- [22] Britten RJ, Davidson EH. Gene regulation for higher cells: A theory. *Science* 1968; 165:349-357.

-
- [23] Wolfe KH. Yesterday's polyploids and the mystery of diploidization. *Nature Rev Genet* 2001; 2:333-341.
- [24] McLysaght A, Hokamp K, Wolfe KH. Extensive genome duplication during early chordate evolution. *Nat Genet* 2002; 31:200-204.
- [25] Comai L et al. Phenotypic instability and rapid gene silencing in newly formed Arabidopsis allotetraploids. *Plant Cell* 2000; 12:1551-1568.
- [26] Lee HS, Chen ZJ Protein-coding genes are epigenetically regulated in Arabidopsis polyploids. *Proc Natl Acad Sci USA* 2001; 98:6753-6758.
- [27] Kashkush K, Feldman M, Levy AA. Gene loss, silencing and activation in a newly synthesized wheat allotetraploid. *Genetics* 2002; 160:1651-1659.
- [28] Abi-Rached L, Gilles A, Shiina T, Pontarotti P, Ionoko H. Evidence of en bloc duplication in vertebrate genomes. *Nat Genet* 2002; 31:100-105.
- [29] Kubo N, Harada K, Hirai A, Kadowaki K. A single nuclear transcript encoding mitochondrial RPS14 and SDHB of rice is processed by alternative splicing: common use of the same mitochondrial targeting signal for different proteins. *Proc Natl Acad Sci USA* 1999; 96:9207-9211.
- [30] Long M, Langley CH. Natural selection and the origin of jungwei, a chimeric processed functional gene in *Drosophila*. *Science* 1993; 260:91-95.
- [31] Moran JV, DeBerardinis RJ, Kazazian HH Jr. Abstract Exon shuffling by L1 retrotransposition. *Science* 1999; 283:1530-1534.
- [32] Wang W, Brunet FG, Nevo E, Long M. Origin of sphinx, a young chimeric RNA gene in *Drosophila melanogaster*. *Proc Natl Acad Sci USA* 2002; 99:4448-4453.
- [33] van Rijk A, Bloemendal H. Molecular mechanisms of exon shuffling: illegitimate recombination. *Genetica* 2003; 118:245-249.
- [34] Rubin GM, Kidwell MG, Bingham PM The molecular basis of hybrid dysgenesis: The nature of induced mutations. *Cell* 1983; 29:987-994.
- [35] Schwarz-Sommer Z, Leclercq L, Saedler H. *Cin4*, an insert altering the structure of the *A1* gene in *Zea mays*, exhibits properties of nonviral retrotransposons. *EMBO J* 1987; 6:3878-3880.
- [36] Fedoroff N. Transposons and genome evolution in plants. *Proc Natl Acad Sci U S A*. 2000; 97:7002-7007.
- [37] Dewannieux M, Esnault C, Heidmann T. LINE mediated retrotransposition of marked Alu sequences. *Nature Genetics* 2003; 35:41-48.
- [38] Kashkush K, Feldman M, Levy AA. Transcriptional activation of retrotransposons alters the expression of adjacent genes in wheat. *Nature Genetics* 2003; 33, 102 - 106.
- [39] Loreto EL, Valente VL, Zaha A, Silva JC, Kidwell MG. *Drosophila mediopunctata* P elements: a new example of horizontal transfer. *J Hered.* 2001; 92:375-81.
- [40] Dehal P et al. The draft genome of *Ciona intestinalis*: insights into chordate and vertebrate origins. *Science* 2002; 298:2157-2167.
- [41] Brown JR. Ancient horizontal gene transfer. *Nature Rev Genet* 2003; 4:121-132.
- [42] Cyranoski D, Abbott A. Virus detectives seek source of SARS in China's wild animals. *Nature* 2003; 423, 467.
- [43] Chen L, deVries AL, Cheng CH. Evolution of antifreeze glycoprotein gene from a trypsinogen gene in Antarctic notothenioid fish. *Proc Natl Acad Sci USA* 1997; 94:3811-3816.

-
- [44] Whitfield CW, Band MR, Bonaldo MF, Kumar CG, Liu L, Pardinias JR, Robertson HM, Soares MB, Robinson GE. Abstract Annotated expressed sequence tags and cDNA microarrays for studies of brain and behavior in the honey bee. *Genome Res.* 2002;12:555-66.
- [45] Okazaki Y, Furuno M, et al. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. *Nature* 2002; 420:563 - 573.
- [46] Lee RC, Feinbaum RL, Ambros V. The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell* 1993; 75:843-854.
- [47] Lee RC, Ambros V. An extensive class of small RNAs in *Caenorhabditis elegans*. *Science* 2001; 294:862-864.
- [48] Reinhart BJ, Bartel DP. Small RNAs correspond to centromere heterochromatic repeats. *Science* 2002; 297:1831.
- [49] Carrington JC, Ambros V. Role of microRNAs in plant and animal development. *Science* 2003; 301:336-338.
- [50] Lau NC, Lim LP, Weinstein EG, Bartel DP. An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science* 2001; 294:858-62.
- [51] Schramke V, Allshire R. Hairpin RNAs and retrotransposon LTRs effect RNAi and chromatin-based gene silencing. *Science* 2003; 301:1069-1074.
- [52] Hennig W. Heterochromatin. *Chromosoma* 1999; 108:1-9.
- [53] Yao MC, Fuller P, Xi X. Programmed DNA deletion as an RNA-guided system of genome defense. *Science* 2003; 300:1581-1584.
- [54] Ammermann, Steinbrück G, von Berger L., Hennig W. Differential replication during the macronucleus formation of the ciliated protozoan *Stylonychia mytilus*. *Chromosoma* 1974; 45, 401-429.
- [55] Hennig W, Walker PMB. Variations an the DNA from two rodent families (Cricetidae and Muridae). *Nature* 1970; 225, 915-919.
- [56] Lankenau DH, Huijser P, Jansen E, Miedema K, Hennig W. Microopia: a retro-transposon of *Drosophila* combines structural features of DNA-viruses, retroviruses, and non viral transposable elements. *J Mol Biol* 1988; 204:233-246.
- [57] Huijser P, Beckers L, Top B, Hermans M, Sinke R, Hennig W. Poly[d(C-A)]- poly[d(G-T)] is highly transcribed in testes of *Drosophila hydei*. *Chromosoma* 1990; 100:48-55.