

## Standard Review

# Analysis and manipulation of the genome dynamic structure

Valentina Tosato\* and Carlo V. Bruschi

Yeast Molecular Genetics Laboratory, International Centre for Genetic Engineering and Biotechnology, AREA Science Park – W, Padriciano 99, IT-34012 Trieste, Italy.

Accepted 7 May, 2007

After the genomic era, during which DNA sequencing revealed genes and the post-genomic era, in which their functional analysis was implemented, the notion of a dynamic genome has become convincing. Indeed, since the early days of DNA transposition, new evidence has accumulated indicating a high level of intrinsic structural plasticity characterizing the genome. An ensemble of gross chromosomal rearrangements has been reported, together with their biological effects, some of which correlate to major cellular pathologies such as cancer. From microorganisms to human cells, the convoluted architecture of chromosomes has gained relevance not only from a descriptive point of view, but also as a potential multi-layered storage mechanism of genetic information. The higher-order structure of DNA, including hairpin turns, bending and curvature, as well as precise chromatin topology, could provide the metadata on super-information needed to explain the low number of inferred mammalian genes, perhaps conferring new scientific dignity to the infamous „junk DNA“. In this view, genome dynamics, including the clustering of essential genes and the synteny of others, appears as the paramount cellular response to varying environmental conditions, resulting in massive differential regulation of gene expression. A deeper understanding of the various orders of complexity of genomic DNA structure has allowed the design of more sophisticated biochemical and biophysical tools for its analysis and manipulation, which, in turn, has yielded a better knowledge of the genome itself. This creative cycle is providing new generations of diagnostics and intelligent drugs of pharmacogenomic origin that exploit the ever changing, yet stable, genome dynamic structure.

**Key words:** Genome dynamics, secondary DNA structure, gross chromosomal rearrangements

## Table of contents

Abstract  
Introduction  
Mathematical models  
New biological tools for genome analysis and manipulation  
Artificial Chromosomes  
Chromosome knockout  
GCRs  
Gene inactivation  
Genome microarray  
From bacteria to eukaryotes  
Homologous linear DNA molecules and GCRs  
Genomics and vaccine development  
Conclusions  
References

## INTRODUCTION

### Historical facts

In 1912, the William Braggs father and son discovered the atomic structure of crystals, by X-ray diffraction patt-

ern, which provided the key for Watson, Crick, Franklin and Wilkins to visualize the DNA architecture just 40 years later. However, there was a then fashionable belief that the molecular structure of this macromolecule was a

relatively stiff, static construction. Barbara Mc Clintock (Nobel Prize) had a prescient recognition of genomic plasticity prior to 1952, which took years to be generally accepted by the wider scientific community. By 1970 when Hamilton Smith isolated the first restriction enzyme, the idea that the genomic DNA could be manipulated and recombined *in vitro* began to be considered and assessed. In the eighties the plasticity of the genome was notable largely because it constituted a drawback in the taxonomic definition of the species of kinetoplastidae, such as *Trypanosoma* and *Crithidia*, the hyper-variability of few genetic loci impairing their classification by karyotyping. High frequency of genome rearrangements, due to repeated sequence elements, had already been described in archaea and other microbes like *Rhizobium* (Martinez et al., 1990; Flores et al., 2005). *Helicobacter pylori*, for another instance, is known to have such a high plasticity resulting in important differentiation among pathogenic strains. Real insight into the importance of genomic rearrangements however, came as a consequence of the enormous success of genome sequencing and functional analysis projects. Today, many researchers believe that DNA winds and unwinds or undergoes frequent melting of its double helical structure, generating single-strandedness at particular regions, then re-annealing to its normal status, an entropic phenomenon nicknamed "breathing". Among other things, this dynamic structure of the genome allows exogenous and endogenous DNA fragments to be integrated, regulates gene transcription by a topological control of the double-helix, and may lead to differentiation of undifferentiated cells, as well as to dedifferentiation (cancer) through gross chromosomal rearrangements (GCRs).

### **Mathematical models provide *in silico* predictions for chromatin organization, bendability and secondary structures**

One of the most fascinating and fast-moving fields in genome dynamics studies is the mathematical prediction and interpretation of the biological significance of DNA sequences. Nonlinear models have been proposed to explain the complex dynamics of DNA bubbles occurring in denaturation, transcription and genetic recombination events (Komarova and Soffer, 2005). Particularly during recombination events, the rate of the movement of the crossover junction has been estimated by branch migration assays (Karymov et al., 2005; McKinney et al., 2005). Also, a stochastic differential equation has been applied to the dynamic feature of the B-Z DNA transition (Lim and Feng, 2005). The free energy spent by topoisomerase to bind and bend DNA has been calculated, computing the twist and the writhe of supercoiling (Barbi et

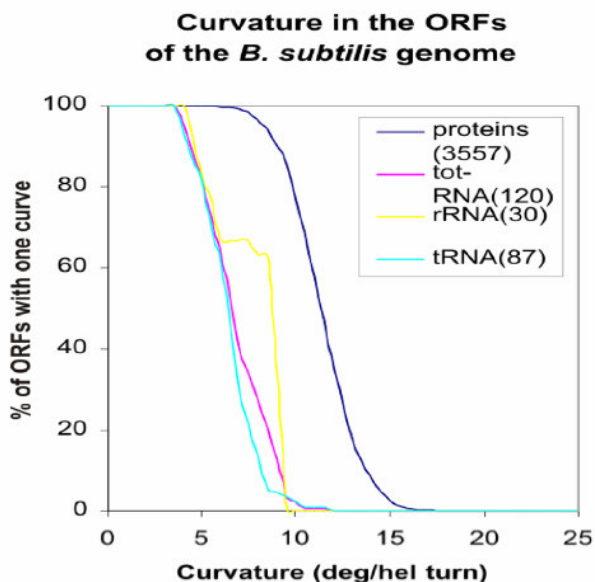
al., 2005; Charvin et al., 2005). These models should help to elucidate how the genome movements following its entropic breathing may interact with the complex regulation of replication, repair and transcription mechanisms.

Here, we will focus on the detailed *in silico* analyses that have been performed on chromatin remodeling, DNA bending and genomic secondary structures. It is not yet completely clear, for example, how the repair mechanism (either homologous recombination - HRS - or non-homologous end joining - NHEJ) gains access to DNA within chromatin and how chromatin structure is restored after repair. Perhaps, physical disruptions of the double helix such as double strand breaks (DSBs) are detected through an altered chromatin topology, which transmits a signal to be sensed by the cell at a checkpoint. The checkpoint system then arrests the cell cycle until the break is repaired (for a review see Ehrenhofer-Murray 2004). Effective modification of histone H2A following a double strand break formed by the HO endonuclease was recently characterized at the molecular level in yeast (Shroff et al., 2004). This analysis suggests that a break of the chromosomal DNA leads to rearrangements of chromatin and re-positioning of the repair protein complex. Moreover, chromatin-remodeling factors are responsible for ploidy control. Indeed, a mutation in these genes results in an increase of genomic instability and it is usually associated with cancer induction (Vries et al., 2005). The majority of these experimental evidences are predicted by exhaustive comparative *in silico* analysis of chromatin remodeling. Genome-wide mapping predicts that chromatin accessibility and gene expression are controlled by histone modifications of regulatory elements such as enhancers, locus control regions (LCRs) and insulators. In favor of this prediction, islands of acetylation have been identified within promoters and highly transcribed regions (Roh et al., 2005). Unconventional chromatin assays (digital analysis of chromatin structure-DACS) coupled with sophisticated algorithms allow precise predictions of regulatory sequences and the accessibility of their chromatin (Sabo et al., 2004).

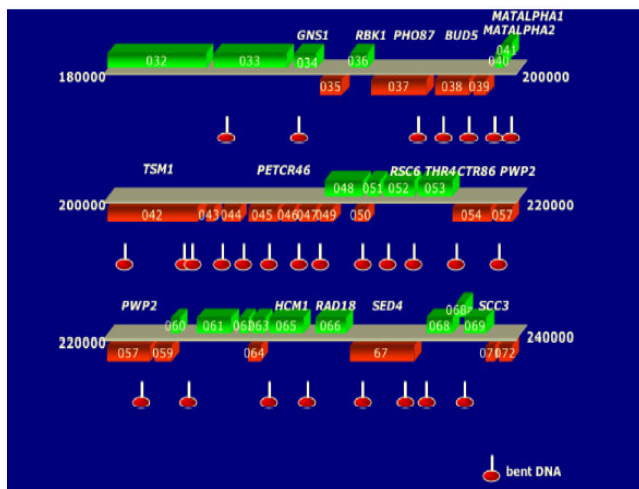
The *in silico* analysis of scaffold/matrix attachment sites (S/MARs), of the nucleosome-formation propensity, and of the presence of repeats within regions located upstream in human genes, brings into sight how much chromatin architecture may modify the transcription activity of the genome (Ganapathi et al. 2005). The analysis of chromatin features allows discrimination between house-keeping and tissue-specific genes, attributing to chromatin a super-informative function with respect to the genetic level. This super-information provided by chromatin is consistent with the effective low number of regulatory and specialized genes in the human genome. Unlike the case of polymorphism of the double helix (A, B or Z) structure, which is quite rare *in vivo*, the genomic DNA usually shows micro-polymorphisms in the B helical structure. The concept that DNA may be bent, especially in A-rich tracts, was proposed initially by Trifonov and

---

\*Corresponding author. E-mail: [tosato@icgeb.org](mailto:tosato@icgeb.org). Tel: +39-040-375-7318. Fax: +39-040-375-7343.



**Figure 1** Percentage. of ORFs overlapping with curved motifs in the *B. subtilis* genome. The total number of protein, RNA, rRNA and tRNA ORFs is given in parentheses.



**Figure 2.** Position of significant (>14 degrees/ helical turn) curvature peaks along 60,000 nucleotides of the DNA sequence of yeast *Saccharomyces cerevisiae* chromosome III. The green and the red numbered boxes represent ORFs.

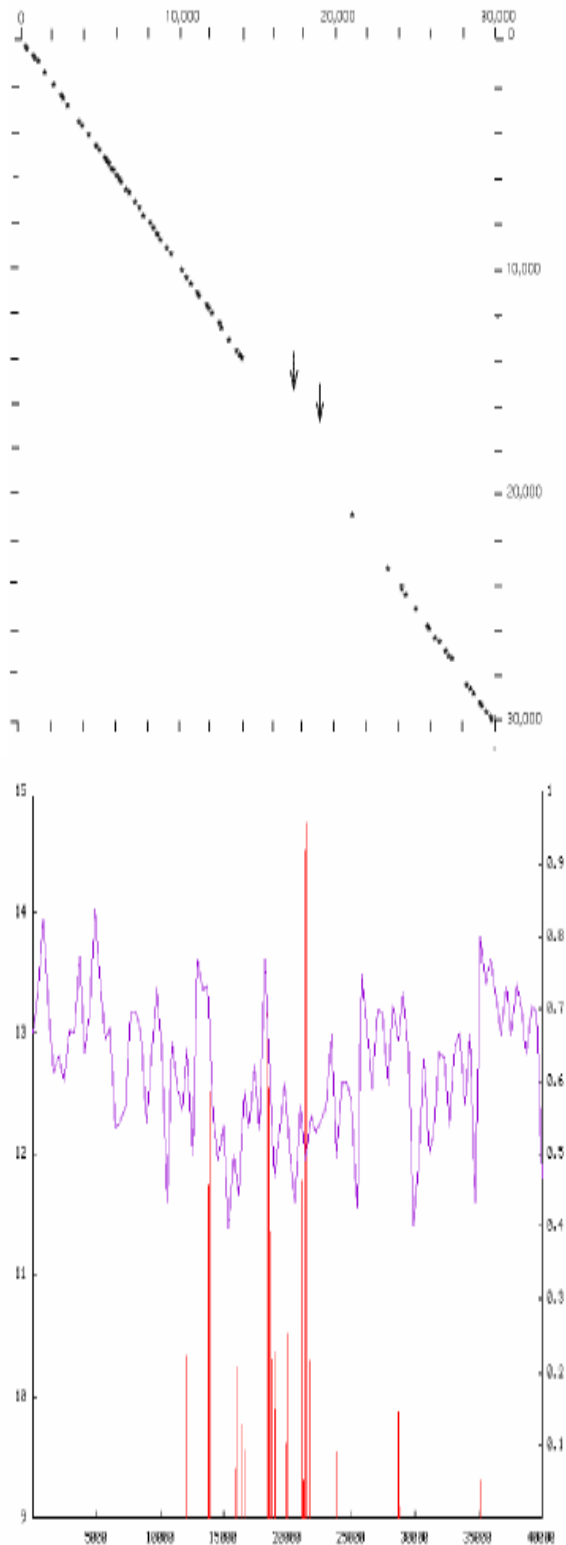
Sussman (1980) who observed that the natural anisotropy of the DNA molecule facilitates its smooth folding into chromatin. The curvature of DNA is measured in terms of angle of deflection between adjacent base pairs (degrees/helical turn). Several servers running bendability prediction programs have been developed to be applied to experimental systems, looking for a correlation between curvature and other features of the genomes (Vlahovicek et al., 2005).

As mentioned, AT-rich stretches in the “B” form usually correlate with high curvature. This is the case for all 16 centromeric DNA regions of *Saccharomyces cerevisiae* (Bechert et al., 1999). The structure of the centromere may correlate with the binding, assembly and the function of the chromosome determining element (CDE) protein complex, which regulates centromeric activity. To examine the effect of the A-T tracts or curvature, it is possible to simulate the molecular dynamics of the “B”-structure (McConnell and Beveridge 2001). Furthermore, there is a preference for curved segments to be distributed with respect to start and stop codons. Usually, intergenic regions such as promoters and terminators show a propensity to be curved. An example of that is reported in Figure 1 where only 6.2% of all ORFs contain one significant curvature (>14 degrees/helical turn) in the model bacterium *Bacillus subtilis* (Tosato et al., 2003).

An example of distribution of curvature around open reading frames in chromosome III of yeast is shown in Figure 2 where the exact DNA sequence and its corresponding coding and non-coding regions were initially identified following the genomic sequencing of the first eukaryotic chromosome from yeast (Oliver et al., 1992).

This non-uniform distribution of curvature correlates with a different propensity of specific regions of genomic DNA to spontaneously recombine. The promoter and terminator sequence are usually more prone to accept DNA integration by recombination events (Gjuracic et al., 2004). This phenomenon is significantly parallel to the nonrandom distribution of DSB regions along the chromosomes (Baudat and Nicolas, 1997). Hotspots for recombination via DSBs are known in all the chromosomes of yeast (Gerton et al., 2000) and many of them contain palindromic sequences (Nasar et al., 2000). The presence of secondary structures as palindromes may influence the recombination efficiency of a specific genomic region that thus becomes a fragile site in terms of breaks and recombination.

A decreasing gradient of large DNA hairpins from the origin towards the terC end of chromosomal replication as well as a high curvature in the intergenic regions characterizes the genome of *B. subtilis* (Tosato et al. 2003). Indeed, the lack of secondary structures around terC (with the exception of the two hairpins of the replication terminus) is in agreement with the low level of homologous recombination detected in this region (Chedin et al., 1998). This observation supports the idea that DNA secondary structures may trigger recombination (Lobachev et al., 1998). Higher order structures may also contain a different kind of super-information in pathogenic species. *Kinetoplastida* parasites, for example, are responsible for diseases severely affecting human health and retarding agriculture development in third world countries. *Kinetoplastida* species cause sleeping sickness (*Trypanosoma brucei*), Chagas (*Trypanosoma cruzi*) and leishmaniasis (*Leishmania* spp.). More than 400 million of people around the world are affected by



**Figure 3.** Distribution of hairpins, Top, and bendability, bottom, of the DNA sequence encompassing the switching point of *Leishmania major* chromosome I. In the top panel, the arrows indicate the extremes of the switching point DNA and the dots the position of the hairpins. In the bottom panel, the wavy, violet line represents the G+C content while the vertical red bars indicate the predicted DNA curvature.

these pathologies, especially in the developing Countries. The parasites share a unique organization of protein-coding genes into long, strand-specific, polycistronic clusters and a lack of general transcription factors (Ivens et al., 2005). Large surface antigen families occur at non-syntenic chromosome-internal and sub-telomeric regions.

The regions between poly-cistronic clusters, where the direction of transcription switches from one DNA strand to the other, have been hypothesized to contain origins of replication and possibly also centromeres and promoters. The base skews of divergent strand-switches of *Trypanosoma* species show patterns analogous to those for bacterial origins of replication (Nilsson and Andersson, 2005), but they differ from those of *Leishmania major*. In *Leishmania* chromosomes, for example, there are several coding strand switch regions that show peculiar characteristics. An *in silico* analysis suggests that there is a trend towards decreased number and stability of hairpins in the direction of the switching points. Moreover, within a few kilobases on either side of the coding-strand switch locus, the predicted secondary structures disappear (Tosato et al., 2001). Some of the results are summarized in Figure 3. Other features of the switch regions are a high AT content and a strong intrinsic curvature. The functional meaning of these regions as bi-directional promoters has been later confirmed experimentally (Martinez-Calvillo et al., 2003). Whether the switching point coincides with a centromeric region has not yet been resolved. Using a telomere-associated chromosome fragmentation approach, it was shown that the region required for the mitotic stability of chromosome 3 of *T. cruzi* encompasses a transcriptional switching point domain constituted of a GC-rich island. This region contains several retrotransposon-like insertions but, atypically, lacks the arrays of satellite repeats normally associated with centromeric regions (Obado et al., 1995).

Switching points are an example of junk DNA in lower eukaryotes, confirming the idea that the genetic code is only one of the keys we need to decipher the genome meaning. The super-information that is packed within higher-order structures, such as secondary DNA structures, bendability and chromatin topology, may contribute to the as yet unknown significance of this junk DNA.

### **New biochemical and biophysical tools for the analysis and manipulation of genomic DNA**

The extremely rapid recent development of a large assortment of new genomic tools has frustrated their cataloguing and renders incomplete any review of the field. Nevertheless, we can summarize the most important new tools as examples of the technical sophistication reached by genomic analytical methodology in this field. In doing this, we must recognize that the notion of the dynamic genome has not yet influenced, in general, the strategies for discovery of new tools, neither their

biochemical, or their biophysical characteristics.

### Artificial chromosomes

Yeast and Bacterial Artificial Chromosome (YAC and BAC)-based technology has been extensively used in the subcloning, of large genomic DNA for the construction of genomic libraries essential to sequencing projects (Zhang and Win, 1997; Bruschi et al., 2006). Together with advanced gene-targeting techniques, the construction of "YAC mega-cassettes (YMCs)" has proven efficient for the disruption of entire mouse genetic loci, and their replacement with equivalent human DNA regions in murine embryonic stem (ES) cells. This was performed to create transgenic mice with specific modifications of genes, enabling them to produce only antigen-specific humanized antibodies (Ledermann, 2000). Although this approach offered new experimental opportunities, there are considerable technical limitations that make it difficult for routine use. The major obstacle is generally the low frequency of gene targeting in mammalian cell lines, ranging up to  $10^{-6}$ /cell for most loci studied. The main reasons for this gene targeting inefficiency lay in: 1) the inherently refractory behaviour of the mammalian genome to mitotic homologous recombination; 2) the high frequency of competing repair pathways; and 3) the random DNA integration into chromosomal sites by illegitimate recombination, which outnumbers gene targeting by about 1000:1. Some enrichment in gene-targeted events could be obtained by using targeting vectors with both positive and negative selection markers or with "trap-vectors" containing modified selection markers which are expressed only in the case of vector integration into particular DNA sequences (promoter- and polyadenylation-trapping; Bradley et al. 1992). Another approach to increase the gene targeting frequency is based on the transient expression of the endogenous or foreign recombinational enzymes (Yanez and Porter, 1999; Scherbakova et al., 2000), resulting in a significant increase of targeted events.

However, retroviral targeting vectors have another limiting factor, namely the length of the DNA fragment homologous to the target locus. Due to the size limit that viral vectors can accommodate, the ideal length of DNA inserts ranges from 5 - 10 kb. This relatively small size of the homology between vector and target DNA influences not only the targeting frequency but also the size of the disruption/replacement that can be obtained (several tens of kb; Tsuda et al., 1997).

An alternative to retroviral targeting vectors could be provided by bacteriophage  $\lambda$  and P1 vectors, which can carry up to 90 kb of inserted DNA (Nehls et al. 1994; Brüggemann and Neuberger 1996). However, their cloning potential is still far below the size of many mammalian genes, which limits their use. Accommodation of even larger DNA fragments and their stable pro-

pagation in yeast *S. cerevisiae* was achieved by construction of yeast artificial chromosomes with megabase-long DNA inserts from several YAC clones bearing overlapping DNA inserts, through the exploitation of the efficient and accurate yeast recombination machinery (Silverman et al., 1990, Deunen et al., 1992). This characteristic of yeast also allows an introduction of subtle changes within cloned DNA and introduction of new selectable markers (retrofitting) for transfer of the modified YAC into mammalian cell lines. Many human loci have been transferred so far to murine ES cells using modified YAC vectors, and have been subsequently efficiently expressed in the transgenic mice tissues (Mendez et al.; Peterson, 1997). However, YACs were not used for improvement of gene targeting, but rather as vectors for the introduction of large human genomic DNA fragments into ES cell lines. To avoid heterogeneous expression of transgenic and endogenous loci, the cell line employed must have endogenous DNA region(s) silenced or, alternatively, the YAC-derived transgenic mouse need to be subsequently crossed with appropriate knock out strains. The main disadvantage of this approach lies in time-consuming process to obtain the desired genetic change in the mouse and in the necessarily prolonged culture passages that reduce dramatically the pluripotency of the ES cells, thus minimizing chances to produce transgenic animals. Within the category of artificial chromosomes, Mammalian Artificial Chromosomes (MACs) did not fare as well in the past, due to the difficulty to clone and perpetuate very large DNA molecules containing human centromeric regions that, *per se*, can reach the size of 1 megabase. These biotechnological instruments, therefore, were used only in particular, specific cases in which the genetic elements to be manipulated were already known and isolated from the whole genome. More recently though, mammalian artificial chromosome technology has been further improved and rendered almost as handy as BACs and YACs. Indeed, one class of MACs, the mammalian satellite DNA-based Artificial Chromosome Expression (ACE) systems can be replicated *de novo* in cell lines of different species and efficiently separated from the host chromosomes (Bunnell et al., 2005).

### Chromosome knockout

A major manipulation of the genome of a eukaryotic cell is, without any doubt, the elimination of one of its chromosomes. This event dramatically changes genome homeostasis as well as overall chromosome dynamics during meiotic division and mitotic segregation. Historically, chromosome loss has been achieved in eukaryotes by treating cells with chemical and physical agents that interfere with spindle formation and dynamics, with the result that this kind of loss is random and occurs with additional toxic cellular effects (Esposito et al., 1982;

Howlett and Schiestl, 2000). Monosomic yeast strains have been obtained in cell division cycle (*cdc*) mutant strains defective in gene products playing an essential role in DNA synthesis and chromosome segregation (Bueno and Russel, 1992; Bruschi et al., 1995; Storici et al., 1995). Moreover, a specific chromosome loss achieved by 2- $\mu$  DNA plasmid integration was reported many years ago (Falco and Botstein, 1983). In another report, the *GAL1* promoter was integrated immediately upstream of a centromere. Transcription of *GAL1* interfered with the centromere function, thus inducing chromosome instability (Guacci and Kaback, 1991). However, these methods are very elaborate and yield contradictory results. The establishment of randomly monosomic mammalian cell lines obtained through the inactivation of DNA topoisomerase II, which plays an important role in mitotic chromosome disjunction (Clarke and Gimenez-Albian, 1999) has highlighted the importance of DNA topology and dynamics in the stability of the whole genomic architecture.

More recently, selective chromosome V loss by the deletion of the corresponding centromere in yeast diploid strain, using the standard EUROFAN knockout technology, has been reported (Zang et al., 2002). The experimental approach was to substitute the centromere region of interest with a linear, double-stranded DNA cassette containing a specific yeast selectable marker (the *LEU2* gene in that case) flanked by two 40-bp DNA sequences homologous to the target peri-centromeric regions. In this way, the substituted chromosome is replicated during DNA synthesis, but it is transferred casually to the daughter cell at each cell division, instead of systematically, because of the lack of the kinetochore-anchoring complex for the mitotic spindle fibers. The result is the segregation of a cell line lacking one chromosome. However, it has further been demonstrated that, either immediately or after a certain number of cell generations, the cells endoreduplicated the remaining homologous chromosome. This restores the normal euploidy, but fixes the Loss of Heterozygosity (LOH) for the markers previously in heterozygous configuration on the pair of chromosomes involved in the knock out.

### Gross chromosomal rearrangements (GCRs)

The discovery of mutants with increased genomic instability suggested that chromosomal rearrangements normally are actively suppressed, probably by specific checkpoint genes (Kolodner et al. 2002). Transformation of yeast cells with a chromosomal fragmentation vector (CFV) resulted in the gain of a chromosomal fragment (CF) with or without the loss of the targeted chromosome, following DSB processing by break-copy duplication (Morrow et al., 1997). Later, it was demonstrated that a chromosomal DSB produced by the HO endonuclease could be repaired by break-induced-replication (BIR),

leading to non-homologous end joining (NHEJ)-mediated non-reciprocal translocation (Bosco and Haber, 1998). Recently, a *cre* site-specific recombination-based system has been developed to produce reciprocal translocations at pre-engineered *loxP* sites (Delneri et al., 2003). Finally, the HO system has been utilized for the production of a DSB on two chromosomes, resulting in reciprocal translocations by NHEJ (Yu and Gabriel, 2004). However, in these experimental yeast systems, cells needed to be previously engineered to obtain translocations at their modified chromosomal sites, and could not generate non-reciprocal translocations. Therefore, the technical difficulty of generating this type of gross chromosomal rearrangement (GCR) *in vivo* hinders the understanding of genome alterations and dynamics.

An experimental production of selectable translocants generated at desired chromosomal locations in wild-type yeast strains was developed (Tosato et al., 2005). Cells have been transformed with the *Kan<sup>R</sup>* linear DNA cassette having the selectable marker flanked by two DNA sequences homologous to two different chromosomes. Using this BIT (Bridge Induced Translocation) system, induction of targeted non-reciprocal translocations in mitosis was achieved (Tosato et al., 2005). This *in vivo* approach to generate specific chromosomal translocations is the first step into the manipulative understanding of genome dynamics, to mimic the GCRs alterations responsible for many genetic diseases, including cancer.

### Gene inactivation, DNA deletion, and integration technologies

Sophisticated genome manipulation requires the possibility of modifying any inter- or intragenic DNA sequence at will, without leaving large amounts of undesired vector DNA at the site of alteration. To this end, a long series of sophisticated vectors has been developed from the previous gene knockout plasmid systems, for example to integrate non-selectable foreign DNA at any desired genomic location in yeast, with a minimum amount of residual plasmid DNA. Some of these vectors have two mutated *F<sub>1</sub>p* recognition targets sequences (*FRTs*) of the endogenous 2-micron DNA plasmid for site-specific excision of the flanking *KanMX4* kanamycin-resistance gene. Outside the *FRT* boundaries, the plasmid carries multiple sites for subcloning of the DNA fragment to be integrated (Storici et al., 1999). Within the recyclable selection marker methodology, several other systems were available for gene disruption and replacement (Toh-e, 1995; Akada et al., 2002), as well as for epitope tagging of chromosomal genes (De Antoni and Gallwitz, 2000; Knop and Schiebel, 1997). However, none is capable of integrating at a specific locus, a desired DNA sequence having no directly detectable phenotype. The *FLP/FRT* system has now

been improved by implementing a new advanced strategy for *in vivo* genomic DNA alterations. The new system, called STIK (Specific Targeted Integration of Kanamycin resistance-associated non-selectable DNA), allows the integration, at any genomic location, of DNA sequences that express no directly selectable phenotype, such as spacers, tags, nuclear localization signal sequence and any intergenic or otherwise heterologous sequence (Waghmare et al., 2003). The STIK system accomplishes this task by exploiting the temporary integration of the recyclable, positively selectable *KanMX4* marker. This selectable marker can be recycled by *Fip* site-specific excision between two identical *FRTs*, for the integration of further DNA fragments. This chaperone-like plasmid system provides for a new molecular tool to "stik" (integrate) any DNA fragment at any genome location in yeast strains. Moreover, the system can be extrapolated to other eukaryotic cells in which the *FLP/FRT* system functions efficiently.

### Genome microarray imaging

One of the latest technologies in the field of genome analysis is the development of bioinformatics programs able to perform the manipulation of microarray images and the identification of known biological relationships among sets of genes (Zimmer et al., 2004). This technology lends itself to online consultation and interactive utilization for the re-ordering of microarray map data through genome image alignment programs that can be run or downloaded from <http://coli.berkeley.edu/genomeimages/> and other URL sites.

### From bacteria to eukaryotes: the distribution of genes as a consequence of evolution

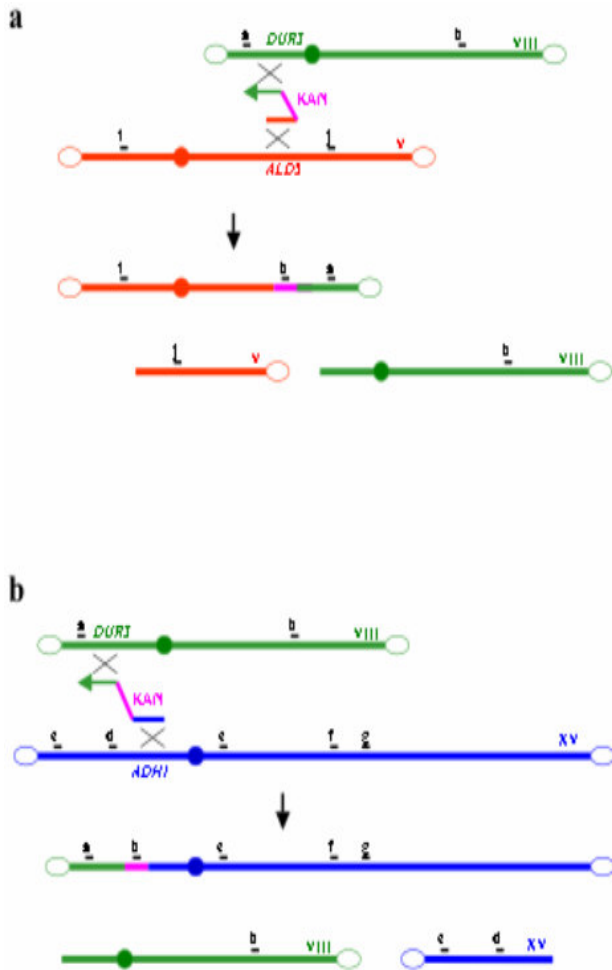
Despite the idea that the highly expressed genes are positioned on the leading DNA strand to allow faster replication and transcription, today it is believed that, at least in bacteria, gene essentiality drives gene-strand bias, (Rocha and Danchin, 2003), while lethality can be provoked not only by deletion, but also by strand switching or scrambling of essential genes. In this view, the genetic organization in bacteria may represent a kind of chromosomal hyper-structure that regulates cellular segregation (Rocha et al., 2003). These signs of chromosomal organization are in apparent contrast to the loss of genetic order due to gross chromosomal rearrangements (Rocha, 2004). The idea that scrambling may be useful for adaptation and evolution partially justifies this possible conflict. In eukaryotes, frequent recombination events leading to massive chromosomal rearrangements, may be necessary in meiosis to generate differentiations, but represent a mistake in mitosis. Gene distribution and genomic evolution are driven by spontaneous segmental duplication or aneuploidization.

The size of the putative segments that undergo duplications varies from 41 kb to 268 kb (Koszul et al., 2006).

One of the proposed molecular mechanisms used to explain duplication, is a fork collapse-generated DSB followed by intra or inter-molecular BIR. These spontaneous events are then fixed in a cellular population by positive selection. Adaptive sweeps are associated with the fixation of these duplicated loci (Moore and Purugganan, 2003). Around 15% of genes in the human genome, and up to 20% in *Drosophila*, *S. cerevisiae* and *C. elegans* are believed to arise from duplication events. For these organisms, the average half-life of a duplicate gene is approximately 4 million years. A conservative estimate of the average rate of origin of new gene duplicates is on the order of 0.01 per gene per million years, with rates in different species ranging from about 0.02 down to 0.002. (Lynch and Conery, 2000). Genomic duplication is now accepted as the major evolutionary mechanism providing a dispensable copy of any gene upon which selection can act, shaping the modifications compatible with life by tests that do not eliminate the failing mutations. Moreover, the knowledge of the existence of tandem chromosomal duplications in pathogens (Brosh et al., 2000) will help in the monitoring the altered immunogenicity of few pathogenic strains and their use as vaccines while new strategies of genomic manipulations such as insertion-duplication mutagenesis may contribute to the identification of potential therapeutic targets (Opperman et al., 2003).

### Homologous linear DNA molecules are bridging life to death via GCRs

Recombination is an essential step in meiosis to generate genomic variability, and genomic variability is necessary for speciation. Chromosomal rearrangements may have contributed to some of the speciation processes along the human and mouse lineages (Marques-Bonet and Navarro, 2005). Nevertheless, massive chromosomal rearrangements in meiosis may also lead to genetic diseases. In mitosis, where strict checkpoint mechanisms avoid excessive recombination among repeated elements and paralogs, the rearrangements are an occasional accident caused by unexpected Double Strand Breaks (DSBs). A replication fork barrier can generate the DSBs, following a chemical or physical damage (irradiation, methyl methanesulfonate-MMS exposure) or following an integration of an exo/endogenous DNA. In all these cases, the break should be immediately repaired; if not, the cell undergoes apoptosis. Different molecular sentinels sense the break and activate variegated salvage pathways like Homologous Recombination (HR), Non Homologous End Joining (NHEJ), Single Strand Annealing (SSA) and others. In all of these pathways, two distinct classes of genes are involved: DNA repair genes and checkpoint genes. Mutations in the first class of genes give rise to severe diseases such as the *Xeroder-*



**Figure 4.** BIT generation of translocant chromosomes with two *Kan<sup>R</sup>* DNA cassettes in the yeast *Saccharomyces cerevisiae*. a) selectable cassette with two ends homologous to the *ald5* and *dur3* loci on chromosomes V and VIII, respectively. b) Selectable DNA cassette with two ends homologous to the *dur3* and *adh1* loci on chromosome VIII and XV, respectively. Black letters indicate the positions of the strategic probes used for mapping the translocation by Southern hybridization. a = *apm2*, b = *crp1*, c = *cdc33*, d = *msh2*, e = *alg6*, f = *ade2*, g = *elg1*, h = kanamycin<sup>R</sup>, i = *bud16*, j = *rad51*.

*ma pigmentosum* and the Hereditary-non-polyposis colorectal cancer. Alterations in the second class may lead, for instance, to Ataxia-telangiectasia, Nijmegen breakage syndrome, Bloom, Werner and Rothmund-Thomson syndromes. The two major pathways of repair are HR (which is more used in lower eukaryotes) and NHEJ (very frequent in mammals, Kanaar et al., 1998). NHEJ occurs in mammals with a frequency of one event/ $10^2$ - $10^4$  cells vs  $10^5$ - $10^7$  of HR. The NHEJ therefore represents a barrier to targeted integration in higher eukaryotes. It is believed that both pathways are active in the regulation of Gross Chromosomal Rearrangements, which are a natural consequence of DNA repair, since in the attempt to repair a DSB the cell generates random GCRs.

When a eukaryotic model cell such as yeast captures linear exogenous DNA with ends homologous to the genome, it integrates it by homologous recombination. If the homologous loci belong to two different chromosomes, a non-reciprocal translocation is generated with a frequency that depends upon the extension of the homology itself. For a tail of 40 nt, the frequency of integration is around 2%, while increasing the homology up to 70 nt, the frequency doubles becoming approximately 4.8%. This observation implies that an exogenous linear DNA may create chromosomal translocations exploiting the endogenous HR system of the cell. Thus, this bridge-induced translocation (BIT) event is responsible for the loss of heterozygosity that is usually associated with many types of cancer (Tosato et al. 2005, Figure 4). It has to be assessed whether these phenomena may also play a role during evolution. When a linear molecule of DNA enters into the nucleus, even if it shows homologous ends, generates a complex scenario of events which consists not only of targeted translocations, but also of ectopic integrations, non-specific translocations and intra-chromosomal deletions. Therefore, it can be assumed that the two main pathways of DNA repair, homologous recombination and the non-homologous end joining, are co-operating to avoid a persisting DNA break and that, in this way, their activity may cause severe genomic rearrangements.

### The legacy of the genomic era is essential for vaccine development

Many infectious diseases that are endemic in developing Countries cannot be treated with an efficient vaccine. Drugs are usually available, but resistance is a recurrent event, especially in infections caused by protozoan parasites. The key obtained by deciphering the microbial genomes has also provided the possibility of developing new vaccines and drugs. Such genomic pioneering opened the curtain on a panorama of incredibly powerful technologies such as microarray, SAGE, cell-based drug design (CBDD) systems and more. When applied to parasites, these technologies highlighted new genes, relevant to the diseases, whose expression differs between the insect-borne stage and the human infectious stage (Duncan et al., 2004). Genome dynamics was observed by subtractive hybridization and microarray analysis also in *Yersinia pestis* (Wang et al., 2006; Zhou et al., 2004) and *Vibrio cholerae* (Dziejman et al., 2002) leading to the identification of the genes that correlate with bubonic/pneumonic plague and cholera. It was found that a surprising dynamics of the genome characterizes *Yersinia*. This is due to horizontal gene acquisition and genetic loss explaining the continuous bacterial evolution to deadly pathogenic lineages. By contrast, *V. cholerae* shows a high degree of conservation among different strains.

The advantage in terms of early diagnosis coming from genomic studies is well demonstrated for leishmaniasis. Knowledge of the genome (Ivens et al., 2005) allowed the development of a new PCR technique (gp63 PCR-RFLP) to discriminate among different *Leishmania* populations that are very close phylogenetically but not epidemiologically.

The elucidation of the dynamic of the microbial genomes helps in understanding the local sequences variations and the polymorphisms characterizing the different isolates. Therefore it is essential to characterize these genomic variations to design an appropriate vaccine. In rural areas of Tanzania the efficacy of different vaccines was tested on different patients bringing three subtypes of the HIV-1 strain (Arroyo et al., 2005). High-risk populations can provide an opportunity for the virus evolution, its recombination, and its adaptation to the host-specific genetic background (Herbinger et al., 2006). Sometimes the treatment of a disease with a vaccine may induce genomic changes and may help the persistence of the disease. One example is the wide spreading and reemergence of the endemic pertussis despite half a century of vaccination. Vaccination shifted the competitive balance between strains (Mooi et al., 2001). In particular the genes coding for two virulence factors, pertactin and pertussis toxin, varied with the adaptation of the microorganism to less react, acellular vaccines composed of purified *B. pertussis* proteins. DNA polymorphisms allowed the identification of three pertussis toxin variants. Five multi-locus sequence types (MLSTs) were found in different isolates of *B. pertussis* and their significant changes were observed after vaccination. The polymorphism in the pertussis toxin genes allowed the characterization of strains from widely separated geographic areas and their correlation with epidemics (van Loo et al., 2002).

The ability to manipulate genomes may be exploited to generate mutant collections without a revertible phenotype to be utilized as live attenuated vaccines. The manipulation of the genomic material bypasses by far the power of gamma-irradiation, long-term culture or chemical mutagenesis techniques. Technologies have been developed (Denise et al. 2004) to selectively delete a certain locus without leaving marker genes that mediate undesired antibiotic resistance to putative useful drugs. To understand the genome dynamics and the variations in gene expression of pathogens (also correlated with different stages of the parasites as for *Leishmania* spp.), several expression profilings were performed (Palacios et al., 2007; Leifso et al., 2007; Vora et al., 2005). They allowed the identification of strategies used by the pathogens for survival in the human host and the focusing on new candidate antigens for vaccine development. Few families that encode surface proteins in *Plasmodium falciparum* 3D7 (Daily et al., 2005) and several genes found in genomic regions with altered percentage of GC content in *Y. pestis* (Lawson et al., 2006) are only few

among a multitude of putative targets suitable for vaccine development which have been highlighted through expression arrays. Therefore, from the epidemiological, diagnostic and technological point of view, exploitation of the genome dynamics potential helps in reading its informational message and understanding its meaning to a much deeper extent.

## Conclusions

The enzymatic pathways that the cell activates in order to fix DNA injuries are not yet well understood, but represent the underlying molecular basis of genome plasticity. It is not completely clear, for example, if a crossing-over event of recombination or break-induced replication (BIR) machinery is the main mechanism responsible for the integration of a linear molecule into the genome. Because linear viral DNA integration as well as DNA transposition to naturally manipulate the genome had arisen during evolution, the elucidation of the molecular mechanism underlying this phenomenon will shed more light on how speciation occurred, and will foster the development of new drugs and vaccines. In this view, the dynamics of the genome structure represents still a crucial, unknown area of genetic information in which secondary DNA architecture warrants future investigations to better define its biological role.

## REFERENCES

- Akada R, Hiroswawa I, Kawahata M, Hoshida H, Nishizawa Y (2002). Sets of integrating plasmids and gene disruption cassettes containing improved counter-selection markers designed for repeated use in budding yeast. *Yeast* 19: 393-402.
- Arroyo MA, Hoelscher M, Sateren W, Samky E, Maboko L, Hoffmann O, Kijak G, Robb M, Birx DL, McCutchan FE (2005). HIV-1 diversity and prevalence differ between urban and rural areas in the Mbeya region of Tanzania. *AIDS* 19: 1517-1524.
- Barbi M, Mozziconacci J, Victor JM (2005). How the chromatin fiber deals with topological constraints. *Phys Rev E Stat Nonlin Soft Matter Phys.* 71(3 Pt 1): 031910.
- Baudat F, Nicolas A (1997). Clustering of meiotic double-strand breaks on yeast chromosome III. *Proc Natl Acad Sci U S A* 94: 5213-5218.
- Bechert T, Heck S, Fleig U, Diekmann S, Hegemann JH (1999). All 16 centromere DNAs from *Saccharomyces cerevisiae* show DNA curvature. *Nucleic Acids Res* 27: 1444-1449.
- Bosco G, Haber JE (1998). Chromosome break-induced DNA replication leads to nonreciprocal translocations and telomere capture. *Genetics* 150: 1037-1047.
- Bradley A, Hasty P, Davis A, Ramirez-Solis R (1992). Modifying the mouse: design and desire. *Biotechnology* 10: 534-539.
- Brosch R, Gordon SV, Buchrieser C, Pym AS, Garnier T, Cole ST (2000). Comparative genomics uncovers large tandem chromosomal duplications in *Mycobacterium bovis* BCG Pasteur. *Yeast* 17: 111-123.
- Brüggemann M, Neuberger MS (1996). Strategies for expressing human antibody repertoires in transgenic mice. *Immunol Today* 17: 391-397.
- Bruschi CV, Mcmilan JN, Cogliavina M, Esposito M (1995). The genomic instability of yeast cdc6-1/cdc6-1 mutants involves chromosome structure and recombination. *Mol. Gen. Genet* 249: 8-18.
- Bruschi CV, Gjuracic K, Tosato V. (2006). Yeast Artificial Chromosomes. In: *Encyclopedia of Life Sciences*. John Wiley & Sons, Ltd: Chichester, UK <http://www.els.net/>: doi:10.1038/npg.els.0006101.

- Bueno A, Russell P (1992). Dual function of *CDC6*: a yeast protein required for DNA replication also inhibits nuclear division. *EMBO J.* 11: 2167-2176.
- Bunnell BA, Izadpanah R, Lederbur Jr. HC, Perez CF (2005). Development of mammalian artificial chromosomes for the treatment of genetic diseases: Sandhoff and Krabbe diseases. *Exp. Opin. Biol. Ther.* 5:195-206.
- Charvin G, Strick T, Bensimon D, Croquette V (2005). Topoisomerase IV bends and overtwists DNA upon binding. *Biophys J.* 89: 384-392.
- Chedin F, Noirot P, Biaudet V, Ehrlich SD (1998). A five-nucleotide sequence protects DNA from exonucleolytic degradation by AddAB, the RecBCD analogue of *Bacillus subtilis*. *Mol. Microbiol.* 29: 1369-1377.
- Clarke DJ, Gimenez-Abian JF (1999). Check points controlling mitosis. *BioEssays* 22:351-363.
- Daily JP, Le Roch KG, Sarr O, Ndiaye D, Lukens A, Zhou Y, Ndir O, Mboup S, Sultan A, Winzeler EA, Wirth DF (2005). *In vivo* transcriptome of *Plasmodium falciparum* reveals overexpression of transcripts that encode surface proteins. *J. Infect. Dis* 191: 1196-203.
- De Antoni A, Gallwitz D (2000). A novel multi-purpose cassette for repeated integrative epitope tagging of genes in *Saccharomyces cerevisiae*. *Gene* 246: 179-185.
- Delneri D, Colson I, Grammenoudi S, Roberts IN, Louis EJ, Oliver SG (2003). Engineering evolution to study speciation in yeasts. *Nature* 422: 68-72.
- Den Dunnen JT, Grootsholten PM, Dauwerse JD, Monaco AP, Walker AP, Butler R, Anand R, Coffey AJ, Bentley DR, Steensma HY, Van Ommen GJB (1992). Reconstruction of the 2.4 Mb human DMD-gene by homologous YAC recombination. *Hum. Mol. Genet.* 1: 19-28.
- Denise H, Coombs GH, Mottram JC (2004). Generation of *Leishmania* mutants lacking antibiotic resistance genes using a versatile hit-and-run targeting strategy. *FEMS Microbiol Lett* 235: 89-94.
- Duncan RC, Salotra P, Goyal N, Akopyants NS, Beverley SM, Nakhasi HL (2004). The application of gene expression microarray technology to kinetoplastid research. *Curr. Mol. Med.* 4: 611-621.
- Dziejman M, Balon E, Boyd D, Fraser CM, Heidelberg JF, Mekalanos JJ (2002). Comparative genomic analysis of *Vibrio cholerae*: genes that correlate with cholera endemic and pandemic disease. *Proc. Natl. Acad. Sci. U.S.A.* 99: 1556-1561.
- Ehrenhofer-Murray AE (2004). Chromatin dynamics at DNA replication, transcription and repair. *Eur. J. Biochem.* 271: 2335-2349.
- Esposito MS, Maleas DT, Bjornstad KA, Bruschi CV (1982). Simultaneous detection of changes in chromosome number, gene conversion and intergenic recombination during mitosis of *S. cerevisiae*: Spontaneous and ultraviolet induced events. *Curr. Genet.* 6: 5-11.
- Falco SC, Botstein D (1983). A rapid chromosome-mapping method for cloned fragments of yeast DNA. *Genetics* 105: 857-872.
- Flores M, Morales L, Avila A, Gonzalez V, Bustos P, Garcia D, Mora Y, Guo X, Collado-Vides J, Pinero D, Davila G, Mora J, Palacios R. (2005). Diversification of DNA sequences in the symbiotic genome of *Rhizobium etli*. *J. Bacteriol.* 187: 7185-92.
- Ganapathi M, Srivastava P, Das Sutar SK, Kumar K, Dasgupta D, Singh GP, Brahmachari V, Brahmachari SK (2005). Comparative analysis of chromatin landscape in regulatory regions of human housekeeping and tissue specific genes. *BMC Bioinform.* 6: 126.
- Gerton JL, DeRisi J, Shroff R, Lichten M, Brown PO, Petes TD (2000). Inaugural article: global mapping of meiotic recombination hotspots and coldspots in the yeast *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. USA.* 97: 11383-1390.
- Gjuracic K, Pivetta E, Bruschi CV (2004). Targeted DNA integration within different functional gene domains in yeast reveals ORF sequences as recombinational cold-spots. *Mol. Genet. Genomics* 271: 437-446.
- Guacci V, Kaback DB (1991). Distributive disjunction of authentic chromosomes in *Saccharomyces cerevisiae*. *Genetics* 127: 475-488.
- Herbinger KH, Gerhardt M, Piyasirisilp S, Mloka D, Arroyo MA, Hoffmann O, Maboko L, Bix DL, Mmbando D, McCutchan FE, Hoelscher M. (2006) Frequency of HIV type 1 dual infection and HIV diversity: analysis of low- and high-risk populations in Mbeya Region, Tanzania. *AIDS Res Hum Retroviruses* 22: 599-606.
- Howlett NG, Schlest RH (2000). Simultaneous measurement of the frequencies of intrachromosomal recombination and chromosome gain using the yeast DEL assay. *Mutat. Res.* 454: 53-62.
- Ivens AC, Peacock CS, Worthey EA, Murphy L, Aggarwal G, Berriman M, Sisk E, Rajandream MA, Adlem E, Aert R, Anupama A, Apostolou Z, Attipoe P, Bason N, Bauser C, Beck A, Beverley SM, Bianchettin G, Borzym K, Bothe G, Bruschi CV, Collins M, Cadag E, Ciarloni L, Clayton C, Coulson RM, Cronin A, Cruz AK, Davies RM, De Gaudenzi J, Dobson DE, Duesterhoeft A, Fazelina G, Fosker N, Frasch AC, Fraser A, Fuchs M, Gabel C, Goble A, Goffeau A, Harris D, Hertz-Fowler C, Hilbert H, Horn D, Huang Y, Klages S, Knights A, Kube M, Larke N, Litvin L, Lord A, Louie T, Marra M, Masuy D, Matthews K, Michaeli S, Mottram JC, Muller-Auer S, Munden H, Nelson S, Norbertczak H, Oliver K, O'neil S, Pentony M, Pohl TM, Price C, Purnelle B, Quail MA, Rabinowitz E, Reinhardt R, Rieger M, Rinta J, Robben J, Robertson L, Ruiz JC, Rutter S, Saunders D, Schafer M, Schein J, Schwartz DC, Seeger K, Seyler A, Sharp S, Shin H, Sivam D, Squares R, Squares S, Tosato V, Vogt C, Volckaert G, Wambutt R, Warren T, Wedler H, Woodward J, Zhou S, Zimmermann W, Smith DF, Blackwell JM, Stuart KD, Barrell B, Myler PJ (2005). The genome of the kinetoplastid parasite, *Leishmania major*. *Science* 309: 436-442.
- Kanaar R, Hoeijmakers JH, van Gent DC. Molecular mechanisms of DNA double strand break repair. *Trends Cell Biol.* 1998 8: 483-489.
- Karymov M, Daniel D, Sankey OF, Lyubchenko YL (2005). Holliday junction dynamics and branch migration: Single-molecule analysis. *Proc Natl Acad Sci USA* 102: 8186-8191.
- Knop M, Schiebel E (1997). Spc98p and Spc97p of the yeast gamma-tubulin complex mediate binding to the spindle pole body via their interaction with Spc110p. *EMBO J.* 16: 6985-6995.
- Kolodner RD, Putnam CD, Myung K (2002). Maintenance of genome stability in *Saccharomyces cerevisiae*. *Science* 297: 552-557.
- Komarova NL, Soffer A (2005). Nonlinear waves in double-stranded DNA. *Bull. Math. Biol.* 67: 701-718.
- Koszul R, Dujon B, Fischer G (2006). Stability of large segmental duplications in the yeast genome. *Genetics* 172: 2211-2222.
- Lawson JN, Lyons CR, Johnston SA (2006). Expression profiling of *Yersinia pestis* during mouse pulmonary infection. *DNA Cell Biol.* 25: 608-816.
- Ledermann B (2000). Embryonic stem cells and gene targeting. *Exp. Physiol.* 85: 603-613.
- Leifso K, Cohen-Freue G, Dogra N, Murray A, McMaster WR (2007) Genomic and proteomic expression analysis of *Leishmania promastigote* and amastigote life stages: the *Leishmania* genome is constitutively expressed. *Mol. Biochem. Parasitol.* 152: 35-46.
- Lim W, Feng YP (2005). Applying the stochastic difference equation to DNA conformational transitions: A study of B-Z and B-A DNA transitions. *Biopolymers* 78:107-20.
- Lobachev KS, Shor BM, Tran HT, Taylor W, Keen JD, Resnick MA, Gordenin DA (1998) Factors affecting inverted repeat stimulation of recombination and deletion in *Saccharomyces cerevisiae*. *Genetics* 148: 1507-1524.
- Lynch M, Conery JS (2000). The evolutionary fate and consequences of duplicate genes. *Science* 290: 1151-1151.
- Marques-Bonet T, Navarro A (2005). Chromosomal rearrangements are associated with higher rates of molecular evolution in mammals. *Gene* 353:147-154.
- Martinez E, Romero D, Palacios R (1990) The *Rhizobium* genome. *Crit. Rev. Plant Sci.* 9: 59-93.
- Martinez-Calvillo S, Yan S, Nguyen D, Fox M, Stuart K, Myler PJ (2003) Transcription of *Leishmania major* Friedlin chromosome 1 initiates in both directions within a single region. *Mol Cell* 11:1291-1299.
- McConnell KJ, Beveridge DL (2001). Molecular dynamics simulations of B'-DNA: sequence effects on A-tract-induced bending and flexibility. *J. Mol. Biol.* 314:23-40.
- McKinney SA, Freeman AD, Lilley DM, Ha T (2005). Observing spontaneous branch migration of Holliday junctions one step at a time. *Proc Natl Acad Sci U S A* 102:5715-5720.
- Mendez MJ, Green LL, Corvalan JRF, Jia XC, Maynard-Currie CE, Yang XD, Gallo ML, Louie DM, Lee DV, Erickson KL, Luna J, Roy CM, Abderrahim H, Kirschenbaum F, Noguchi M, Smith DH, Fukushima A, Hales JF, Finer MH, Davis CG, Zsebo KM, Jakobovits A (1997). Functional transplant of megabase human immunoglobulin

- loci recapitulates human antibody response in mice. *Nature Genetics* 15: 146-156.
- Mooi FR, van Loo IH, King AJ (2001) Adaptation of *Bordetella pertussis* to vaccination: a cause for its reemergence? *Emerg Infect Dis* 7: 526-528.
- Moore RC, Purugganan MD (2003). The early stages of duplicate gene evolution. *Proc Natl Acad Sci U S A* 100:15682-15687.
- Morrow DM, Connelly C, Hieter P (1997). Break copy, duplication: a model for chromosome fragment formation in *Saccharomyces cerevisiae*. *Genetics* 147: 371-382.
- Nasar F, Jankowski C, Nag DK (2000). Long palindromic sequences induce double-strand breaks during meiosis in yeast. *Mol. Cell. Biol.* 20: 3449-3458.
- Nehls M, Messlerle M, Sirulnik A, Smith AJH, Boehm T (1994). Two large Insert vectors,  $\lambda$ PS and  $\lambda$ KO, facilitate rapid mapping and targeted disruption of mammalian genes. *Biotechniques* 17: 770-775.
- Nilsson D, Andersson B (2005) Strand asymmetry patterns in trypanosomatid parasites. *Exp. Parasitol.* 109: 143-149.
- Obado SO, Taylor MC, Wilkinson SR, Bromley EV, Kelly JM (2005). Functional mapping of a trypanosome centromere by chromosome fragmentation identifies a 16-kb GC-rich transcriptional "strand-switch" domain as a major feature. *Genome Res.* 15:36-43.
- Oliver SG, van der Aart QJ, Agostoni-Carbone ML, Aigle M, Alberghina L, Alexandraki D, Antoine G, Anwar R, Ballesta JP, Benit P, et al (1992). The complete DNA sequence of yeast chromosome III. *Nature* 357: 38-46.
- Opperman T, Ling LL, Moir DT (2003). Microbial pathogen genomes – new strategies for identifying therapeutic and vaccine targets. *Expert Opin Ther Targets* 7: 469-473.
- Palacios G, Quan PL, Jabado OJ, Conlan S, Hirschberg DL, Liu Y, Zhai J, Renwick N, Hui J, Hegyi H, Grolla A, Strong JE, Towner JS, Geisbert TW, Jahrling PB, Buchen Osmond C, Ellerbrok H, Sanchez-Seco MP, Lussier Y, Formenty P, Nichol MS, Feldmann H, Briese T, Lipkin WI (2007). Panmicrobial oligonucleotide array for diagnosis of infectious diseases. *Emerg Infect Dis* 13:73-81.
- Peterson KR (1997). Production of transgenic mice with yeast artificial chromosomes. *TIG* 13: 61-66.
- Rocha EP, Danchin A (2003) Essentiality, not expressiveness, drives gene-strand bias in bacteria. *Nat Genet* 34:377-378.
- Rocha EP, Fralick J, Vedyappan G, Danchin A, Norris V. (2003) A strand-specific model for chromosome segregation in bacteria. *Mol. Microbiol.* 49: 895-903.
- Rocha EP (2004). Order and disorder in bacterial genomes. *Curr. Opin. Microbiol* 7: 519-527.
- Roh TY, Cuddapah S, Zhao K (2005). Active chromatin domains are defined by acetylation islands revealed by genome-wide mapping. *Genes Dev.* 19: 542-552.
- Sabo PJ, Hawrylycz M, Wallace JC, Humbert R, Yu M, Shafer A, Kawamoto J, Hall R, Mack J, Dorschner MO, McArthur M, Stamatoyannopoulos JA (2004). Discovery of functional noncoding elements by digital analysis of chromatin structure. *Proc. Natl. Acad. Sci. USA* 101: 16837-16842.
- Scherbakova OG, Lanzov, VA, Ogawa H, Filatov MV (2000). Overexpression of bacterial *RecA* protein in somatic mammalian cells increases the frequency of gene targeting. *Mutat. Res.* 459: 65-71.
- Shroff R, Arbel-Eden A, Pilch D, Ira G, Bonner WM, Petrini JH, Haber JE, Lichten M (2004). Distribution and dynamics of chromatin modification induced by a defined DNA double-strand break. *Curr. Biol.* 14:1703-1711.
- Silverman GA, Green ED, Young RL, Jockel JI, Domer PH and Korsmeyer SJ (1990). Meiotic recombination between yeast artificial chromosomes yields a single clone containing the entire BCL2 protooncogene. *Proc. Natl. Acad. Sci. USA.* 87: 9913-9917.
- Storici F, Oberto J, Bruschi CV (1995). The *CDC6* gene is required for centromeric, episomal, and 2- $\mu$ m plasmid stability in yeast *Saccharomyces cerevisiae*. *Plasmid* 34: 184-197.
- Storici F, Coglievina M, Bruschi CV (1999). A 2-micron DNA-based marker recycling system for multiple gene disruption in the yeast *Saccharomyces cerevisiae*. *Yeast* 5: 271-283.
- Toh-e A (1995). Construction of a marker gene cassette which is repeatedly usable for gene disruption in yeast. *Curr. Genet.* 27:293-297.
- Tosato V, Ciarloni L, Ivens AC, Rajandream MA, Barrell BG, Bruschi CV (2001). Secondary DNA structure analysis of the coding strand switch regions of five *Leishmania major* Friedlin chromosomes. *Curr. Genet.* 40: 186-94.
- Tosato V, Gjuracic K, Vlahovicek K, Pongor S, Danchin A, Bruschi CV (2003). The DNA secondary structure of the *Bacillus subtilis* genome. *FEMS Microbiol. Lett.* 218: 23-30.
- Tosato V, Waghmare SK, Bruschi CV (2005). Non-reciprocal chromosomal bridge-induced translocation (BIT) by targeted DNA integration in yeast. *Chromosoma* 114: 15-27.
- Trifonov EN, Sussman JL (1980). The pitch of chromatin DNA is reflected in its nucleotide sequence. *Proc Natl Acad Sci U S A.* 77:3816-3820.
- Tsuda H, Maynard-Currie CE, Reid LH, Yoshida T, Edamura K, Maeda N, Smithies O, Jakobovits A (1997). Inactivation of the mouse HPRT locus by a 203-bp retroposon insertion and a 55-kb gene-targeted deletion: establishment of new HPRT-deficient mouse embryonic stem cell lines. *Genomics* 42: 413-421.
- van Loo IH, Heuvelman KJ, King AJ, Mooi FR (2002). Multilocus sequence typing of *Bordetella pertussis* based on surface protein genes. *J Clin Microbiol* 40: 1994-2001.
- Vlahovicek K, Kaján L, Szabó G Jr, Tosato V, Bruschi CV, Pongor S (2005). Web servers for the prediction of curvature as well as other characteristics from DNA sequence. In Arseni Markoff (ed) *Analytical Tools for Genes, Genomes and Sequence: The DNA Level*, DNA Press, Eagleville, PA, pp. 215-236.
- Vora GJ, Meador CE, Bird MM, Bopp CA, Andreadis JD, Stenger DA (2005). Microarray-based detection of genetic heterogeneity, antimicrobial resistance, and the viable but nonculturable state in human pathogenic *Vibrio spp.* *Proc Natl Acad Sci U S A* 102: 19109-19114.
- Vries RG, Bezrookove V, Zuijderdijn LM, Kia SK, Houweling A, Oruetxebarria I, Raap AK, Verrijzer CP (2005). Cancer-associated mutations in chromatin remodeler hSNF5 promote chromosomal instability by compromising the mitotic checkpoint. *Genes Dev.* 19: 665-670.
- Waghmare SK, Caputo V, Radovic S, Bruschi CV (2003). Specific targeted integration of kanamycin resistance-associated non-selectable DNA in the genome of the yeast *Saccharomyces cerevisiae*. *BioTechniques* 34: 1024-1033.
- Wang X, Zhou D, Qin L, Dai E, Zhang J, Han Y, Guo Z, Song Y, Du Z, Wang J, Wang J, Yang R (2006). Genomic comparison of *Yersinia pestis* and *Yersinia pseudotuberculosis* by combination of suppression subtractive hybridization and DNA microarray. *Arch Microbiol* 186: 151-159.
- Yanez RJ, Porter AC (1999). Gene targeting is enhanced in human cells overexpressing *hRAD51*. *Gene Therapy* 6: 1282-1290.
- Yu X, Gabriel A (2004) Reciprocal translocations in *Saccharomyces cerevisiae* formed by nonhomologous end joining. *Genetics* 166: 741-751.
- Zang Y, Garre` M, Gjuracic K, Bruschi CV (2002). Chromosome V loss due to centromere knockout or *MAD2*-deletion is immediately followed by restitution of homozygous diploidy in *Saccharomyces cerevisiae*. *Yeast* 19: 1-12.
- Zhang HB, Wing RA (1997). Physical mapping of the rice genome with BACs. *Plant Mol. Biol.* 35: 115-127.
- Zhou D, Han Y, Song Y, Tong Z, Wang J, Guo Z, Pei D, Pang X, Zhai J, Li M, Cui B, Qi Z, Jin L, Dai R, Du Z, Bao J, Zhang X, Yu J, Wang J, Huang P, Yang R (2004). DNA microarray analysis of genome dynamics in *Yersinia pestis*: insights into bacterial genome microevolution and niche adaptation. *J. Bacteriol.* 186: 5138-5146.
- Zimmer DP, Paliv O, Thomas B, Gyaneshwar P, Kustu S (2004). Genome image programs: visualization and interpretation of *Escherichia coli* microarray experiments. *Genetics* 167: 2111-2119.