Send Orders of Reprints at reprints@benthamscience.org

On the Structural Plasticity of the Human Genome: Chromosomal Inversions Revisited

Joao M. Alves[1,2,3,*], Alexandra M. Lopes[2], Lounès Chikhi[3,4] and António Amorim[2,5]

[1]*Doctoral Program in Areas of Basic and Applied Biology (GABBA), University of Porto, Portugal;* [2]*IPATIMUP - Instituto de Patologia e Imunologia Molecular da Universidade do Porto, Porto, Portugal;* [3]*Instituto Gulbenkian de Ciência (IGC), Oeiras, Portugal;* [4]*CNRS (Centre National de la Recherche Scientifique), Université Paul Sabatier, Ecole Nationale de Formation Agronomique, Unité Mixte de Recherche 5174 EDB (Laboratoire Évolution & Diversité Biologique), F-31062 Toulouse, France;* [5]*Faculdade de Ciências da Universidade do Porto, Porto, Portugal*

Abstract: With the aid of novel and powerful molecular biology techniques, recent years have witnessed a dramatic increase in the number of studies reporting the involvement of complex structural variants in several genomic disorders. In fact, with the discovery of Copy Number Variants (CNVs) and other forms of unbalanced structural variation, much attention has been directed to the detection and characterization of such rearrangements, as well as the identification of the mechanisms involved in their formation. However, it has long been appreciated that chromosomes can undergo other forms of structural changes - balanced rearrangements - that do not involve quantitative variation of genetic material. Indeed, a particular subtype of balanced rearrangement – inversions – was recently found to be far more common than had been predicted from traditional cytogenetics. Chromosomal inversions alter the orientation of a specific genomic sequence and, unless involving breaks in coding or regulatory regions (and, disregarding complex *trans* effects, in their close vicinity), appear to be phenotypically silent. Such a surprising finding, which is difficult to reconcile with the classical interpretation of inversions as a mechanism causing subfertility (and ultimately reproductive isolation), motivated a new series of theoretical and empirical studies dedicated to understand their role in human genome evolution and to explore their possible association to complex genetic disorders. With this review, we attempt to describe the latest methodological improvements to inversions detection at a genome wide level, while exploring some of the possible implications of inversion rearrangements on the evolution of the human genome.

Received on: February 21, 2012- Revised on: September 23, 2012- Accepted on: September 24, 2012

Keywords: Chromosomal inversions, Genome architecture, Demographic history, Human evolution, Non-allelic homologous recombination, Segmental duplications.

INTRODUCTION

Over the last years, a growing number of geneticists and evolutionary biologists are shifting their attention from single nucleotide polymorphisms (SNPs) towards bigger and more complex alterations in the architecture of eukaryotic genomes thus going back to some of the oldest genetic markers (e.g. [1, 2]). With the aid of novel and powerful molecular biology techniques (e.g. high-throughput sequencing platforms, array-Comparative Genomic Hybridization and SNP microarrays (see [3] for a review)), the study of the structural plasticity of the genome has gained momentum. Indeed, we are currently witnessing major advances in the field of molecular and computational genomics with increasingly high quality whole-genome data accumulating for several species and fast improvements in computational and statistical tools that allow the extraction of reliable information from these sources.

*Address correspondence to this author at the Population Genetics Group, IPATIMUP - Instituto de Patologia e Imunologia Molecular da Universidade do Porto. Rua Dr. Roberto Frias, s/n 4200-465 – Porto, Portugal;
Tel: 225570700; Fax: 225570799;
E-mail: jalves@ipatimup.pt; jmalves@igc.gulbenkian.pt

This has led to the discovery, validation and characterization of a whole set of different types of structural variants (SVs) and it is now evident that genomic variation is far more complex than previously thought [4]. SVs can be defined as a wide variety of balanced and unbalanced genomic rearrangements of different sizes. They range from Copy Number Variants (CNVs) such as insertions, deletions, and duplications, all being unbalanced, to chromosomal inversions (balanced) and translocations (unbalanced or balanced). Biomedical and clinically oriented research became particularly focused in genomic imbalances, and architectural changes, with genome-wide association studies (GWAS) regularly highlighting the involvement of SVs in several genomic disorders [5-7].

At present, much attention is being directed to the identification of the mechanisms and processes involved in their formation, however uncertainty remains regarding the contribution of these heteromorphisms to phenotype differences between individuals, since most variants described have been found in healthy individuals [8-10].

In this review we consider a particular subtype of rearrangement – chromosomal inversions – that has been increasingly recognized as a relatively common source of

1875-5488/12 $58.00+.00 ©2012 Bentham Science Publishers

variation contrary to early predictions from classical cytogenetics [1].

Inversions alter the orientation of a specific genomic sequence and, for decades, they have been interpreted as a potential mechanical cause of subfertility (and ultimately reproductive isolation) since cross-over events (i.e. recombination) between inverted and non-inverted segments could result in unbalanced, and generally abortive, gametes [11, 12]. From an evolutionary point of view, inversions became recognized as privileged systems to study major processes (e.g. selection) [13] under the generalist idea often held but not always well defined that they could protect chromosomal regions from gene flow, and therefore act as an initial step towards genomic divergence [14]. Indeed, studies in chromosomal evolution have repeatedly attributed important evolutionary roles to these structural rearrangements, with several lines of evidence suggesting their involvement in phenotypic variability [15, 16], adaptive divergence within species [17, 18], and in the origin and evolution of sex chromosomes in mammals [1].

In humans, however, the role of inversions in disease or genome evolution remains unclear [8, 19]. At this stage, more than 1000 inversions have been deposited in the Database of Genomic Variants [20], involving all 22 autosomes, but the fact that only two inversion polymorphisms have been fully characterized at the population level [21-24] clearly illustrates the necessity of studying inversion polymorphisms at a larger scale [19].

Here, our aim is to review the latest theoretical and empirical work dedicated to chromosomal inversions in the human genome, either as disease-associated variants or as segregating polymorphisms in human populations. We discuss the recent advances made in the structural and genetic characterization of inversion polymorphisms - highlighting the major drawbacks in the current strategies - and important issues that have as yet received little attention. Furthermore, we will explore some of the evolutionary and demographic scenarios that have been invoked to explain the presence, maintenance and apparent rise in frequency of particular inversions in different human populations.

INVERSIONS – ON THE DETECTION OF BALANCED STRUCTURAL VARIANTS

The detection of inversions was traditionally limited to large-scale microscopically visible rearrangements via karyotype analysis using classical G-banding techniques [25-27]. With the implementation of improved comparative genomic strategies, both at population and species level, an extraordinary amount of previously unknown inversions were identified in recent years [28, 29]. While most experimental techniques (e.g. FISH, PGFE, Fusion-PCR) remain laborious and target-based [30], where one can only test the presence of a predicted inversion in a specific genomic location, new computational approaches have been recently introduced to identify or predict the location of inversions, from SNP array data and next-generation sequencing (NGS) data, at a genome-wide level [31-34].

For example, Bansal *et al.* [31] developed a statistical method to detect large polymorphic chromosomal segments (> 200 Kb) that are inverted in the majority of the chromosomes in a population, with respect to the human reference sequence and applied it to HapMap data. Even with limited statistical power to detect polymorphisms at frequencies lower than 0.25 (with respect to the human reference), a list of 176 candidate inversions was generated using this model, which overlapped with several previously known inversion polymorphisms. However, since the model uses patterns of strong, long range linkage disequilibrium (LD) to access putative sites of inversion rearrangements, some predicted inversions might be artifacts and may just represent regions of high LD due to low recombination or recent selective sweeps, as noted by the authors.

More recently, Sindi & Raphael [32] applied a probabilistic model, using differences in haplotype block structure, to identify inversion polymorphisms from this type of data. In opposition to [31], their method was able to predict inversion frequencies and detect inversions that are the minor allele in the population (i.e. where most individuals had the reference "non-inverted" haplotype). Furthermore, they generated a set of 355 putative inversion polymorphisms using SNP data from 4 populations (CEU, YRI, CHB+JPT), overlapping with several inversion polymorphisms that have already been validated by others, or for which direct evidence exists [3].

While it was possible to identify known inversion polymorphisms in both studies, hence validating the methods used, there are still several limitations that need to be considered when predicting inversion rearrangements from SNP-haplotype data. The proposed computational models rely on the assumption that (i) SNP haplotypes can be used as a proxy of the inversion status and (ii) strong LD is expected in regions harboring inversion rearrangements. As a consequence, only ancient inversions which have accumulated divergent mutations are likely to be captured. Another issue is that both models implicitly assume a single origin but multiple independent events might have given rise to the presence of a given inversion in different haplotypic backgrounds. Indeed, in the attempt to characterize 6 human disease-associated inversion polymorphisms, Antonacci *et al.* [19] showed, with the exception of one inversion (i.e. 17q21.31), no remarkable correlation between SNP-based haplotypes and the inversion structure. The authors concluded that each of these inversions may have occurred multiple times in the human lineage, on different haplotype backgrounds, providing evidence of recurrence. Similar results were later observed for the 17q21.31 inversion, in a cytogenetically-based study [35], where some individuals homozygous for the SNP-defined haplotype, previously thought to be completely associated with the inversion (i.e. H2), were in fact heterokaryotic (i.e. inversion heterozygotes). In the latter case, it shows that the "inverted haplotype" is sometimes oriented like the "non-inverted" haplotype.

In summary, identifying inversions by means of high density SNP data is promising but far from a trivial task, and in those cases where an inversion has arisen independently on at least 2 distinct haplotype backgrounds, genotyping methods based on SNP data are prone to artifacts (e.g. false negatives).

Alternatively, sequence-based computational approaches have been recently introduced to detect SVs (including in-

versions) making use of specific sequence data signatures (see [3]). Among others, paired-end mapping (PEM) algorithms are showing promising results in genome wide detection of inversion rearrangements as they are able to assess the orientation of paired-end reads, therefore allowing the identification of discordant mapping to a reference genome [36, 37]. A series of recent publications [33, 34, 38] have applied this new method to identify structural variants in the human genome, and while 56 inversions were found using a single individual [33], Kidd *et al.* [34] analyzing 8 genomes, identified a total of 217 inversions (but see [8] for a more comprehensive review).

Considerable technological improvements have boosted our ability to assay inversion variants in the human genome. NGS is becoming a routinely used tool in many biological fields [39-43], and has already contributed (and is still contributing) to a better understanding of the architecture of the human genome. Nevertheless, such technologies still represent a challenge to present-day research [39, 44]. For instance, inversion breakpoints are generally enriched in runs of duplicated segments of DNA (e.g. segmental duplications (SDs)), which greatly limits the ability to unambiguously map breakpoint regions [29]. Also, upon discovery, independent validation methods are still required to confirm the orientation of a specific chromosome segment.

Ultimately, validation studies that simultaneously take into account the limitations of the computational and molecular tools and experimental procedures are crucially needed to estimate the error rates of SNP- or NGS- inferred inversion rearrangements. Indeed, a recent review [44] explored the main limitations of the current approaches to discovering structural variants, highlighting the importance of designing algorithms that incorporate multiple methodologies to improve power, robustness, sensitivity and specificity.

INVERSIONS – ON THEIR IMPACT OF GENOME EVOLUTION

Molecular Effects of Inversions

As balanced rearrangements, inversions do not involve quantitative alteration in the content of cellular DNA (at least no significant change in theory), but the reorganization of a genomic segment induces an alteration of the original genetic background which may have several repercussions. Although much uncertainty remains regarding the direct effects of inversions at the molecular level (e.g. gene expression patterns), it has been shown that some inversions can result in major phenotypic alterations. For instance, the split of the mammalian *Hoxd* gene cluster into two independent pieces, using an experimental technique (STRING) that induced an inversion rearrangement [45], was responsible for the loss of expression of *Hoxd* genes during limb development. One likely explanation to this observation is that the artificial repositioning of the genes within the inverted region, relatively to flanking regulatory elements, led to changes in patterns of gene activity [29, 46]. Inversions exert some of their effects indirectly, by imposing new regimens of molecular evolution on the DNA sequences encompassed by them. This is due to a reduction, or even suppression of recombination within these segments in *heterokaryotypes*. As subtle as it may seem, such effect can have drastic consequences since, by acting as a genetic barrier, an inversion may "freeze" an alternative allelic/haplotypic sequence in a population [47].

Indeed, ever since their first identification in the 1920s [48], inversions have been particularly investigated for their putative role in population divergence and speciation phenomena [14, 49-52]. While classic models (e.g. hybrid dysfunction model of speciation such as the Bateson-Dobzhansky-Muller [53]) often rely on the idea of fertility cost to hybrids, overlooking the mechanisms by which rearrangements become established in the first place, new inversion-based speciation models [14, 49-51] have been proposed in recent years invoking the suppression of recombination as a major process for genetic diversification and speciation.

Recombination is regarded as one of the major evolutionary processes since it is responsible for the genetic shuffling and introduction of new allelic combinations, upon which selection can act [52]. Once an inversion arises in a population, recombination in that region becomes suppressed between chromosomes with different orientations (with the exception of double cross-overs within large inverted regions). Virtually all "suppressed-recombination" models explicitly suggest that such rearrangements provide a window of opportunity for the accumulation of differences between the two chromosomal configurations that could culminate in the evolution of reproductive isolation [14, 49-52].

At present, observations supporting these new models have been reported for several species, including birds [54], mammals [55], insects [15] and plants [17]. In primates much controversy has been building up in the last years over the contribution of suppressed recombination to the divergence of ancestral populations of humans and chimpanzees and, in spite of many efforts, accelerated evolution in rearranged versus collinear chromosomes between the two species has not been definitely proven [56-58]. However, since the scope of this review falls exclusively on human polymorphic inversions, we will not explore further the role of inversions on speciation, but instead we will focus on the possible mechanisms and processes by which inversions rise in frequency and may become established in populations.

From Genomic Novelties to Established Polymorphisms

The spread of these rearrangements can result from a combination of several factors, largely influenced by populations' demography, ecology and evolutionary history. It has been argued [18, 49] that an inversion could rise in frequency because it brings together locally adapted genes that become "protected" from introgression, due to a local reduction in recombination. According to this scenario, the selective advantage is not directly related to the new chromosomal structure but to its favorable genetic (i.e. haplotypic) composition [59]. As a consequence, the distribution of such inversions may display clines related with local adaptation [18]. Non-ecological processes, such as meiotic drive (i.e. a process in which an allele is over-transmitted in gametes during meiosis), might also influence the frequency and distribution of an inversion polymorphism by distorting its segregation [1]. However, while this is theoretically possible

such processes do not appear to be general features in establishing inversions in human populations, since most rearrangements seem to segregate normally.

As any other type of mutation, inversions are affected by evolutionary forces. On this basis, random genetic drift, selection and gene flow (i.e. migration) can play major roles in shaping their distribution and frequencies across populations. For example, Spirito *et al.* [60], using a multi-deme model of local extinction and recolonization, observed that even underdominant inversions could, by chance, persist or rise to fixation in populations. However, the authors noted that this scenario is only achieved in cases of small effective population size, where drift causes the maintenance or rise in frequency of the rearrangement albeit the systematic pressure of selection. In contrast, if the rearrangement offers a selective advantage to the carriers, its fixation is more likely, due to the expected advantage of the inversion homozygotes [52, 59, 60].

In humans, numerous inversion variants of different sizes segregate in populations [8, 31, 46]. Although the vast majority falls within the 10 to 100kb size interval, there are several inversion polymorphisms with sizes greater than 1Mb in length [8]. Such findings are not necessarily surprising as, in theory, the impact of an inversion is primarily related with its breakpoints location [8] and if no gene is disrupted, even large inversions may be neutral and, thus, spread within and between populations through stochastic processes.

However, in the absence of a robust high-throughput method to genotype balanced rearrangements, much uncertainty remains regarding the incidence of inversions in humans, how they are distributed throughout populations and their frequency as polymorphic variants.

HUMAN POLYMORPHIC INVERSIONS – WORLDWIDE DISTRIBUTION AND EVOLUTIONARY TRAJECTORIES

Aside from a small number of examples that come from indirect studies focusing on human diseases [19, 61, 62], only a couple of inversions have been extensively characterized at the population level [21-24]. Namely, (i) the 8p23.1 inversion that spans a 4.5 Mb region and is considered the largest polymorphic inversion known in the human genome [24], and (ii) the smaller but still very large 900 Kb inversion at 17q21.31 which attains relatively high frequencies in several European populations.

The 8p23.1 Inversion (*8p23.1-inv*)

Initial studies [19, 63] have made clear that this particular segment presents a very complex genomic architecture mainly due to the two large blocks of segmental duplications (SDs) it contains. Although considered a neutral polymorphism [24], it has been repeatedly argued [64] that, due to the presence of these highly identical structures, subsequent rearrangements via non allelic homologous recombination (NAHR – i.e. a mechanism of illegitimate recombination between sequences of high identity) can cause syndromic phenotypes (e.g. microdeletion syndromes) in the offspring of heterozygous mothers. However, the exact molecular mechanisms leading to disease phenotypes remain to be elucidated (but see below).

Another important aspect of the *8p23.1-inv* is the number of genes encompassed. The region contains at least 50 genes [63], among which the BLK - *B lymphocyte kinase* - gene that has been associated with systemic lupus erythematosus (SLE), rheumatoid arthritis (RA) and other autoimmune diseases [65]. Interestingly, it has been suggested that the risk alleles are specific to the non-inverted configuration [24].

In order to characterize its worldwide distribution, Salm *et al.* have recently applied an innovative approach to diploid SNP-genotype data [24]. Taking into consideration the limitations of most SNP-based tagging methods to identify inversions, as we noted above, the authors have designed a new and powerful multidimensional scaling (MDS) algorithm called PFIDO (Phase-Free Inversion Detection Operator) to efficiently categorize almost 2000 individuals from 56 populations by inversion status.

According to their results, this inversion polymorphism displays a worldwide clinal distribution with frequencies reaching 79% in a Mozabite sample (Algeria), 63% in an Italian sample and 25% in a "Manchu" sample (North-East Asia), which, the authors claimed would be consistent with demographic models of early human expansions out of Africa. However, since no single SNP was perfectly correlated with the inversion status, the 8p23.1 inversion may not act as an absolute recombination barrier and low levels of gene flow may have occurred throughout its evolution. This is not necessarily surprising given the size of the inversion, which may allow for some double cross-over events.

Based on these results, the authors concluded that the *8p23-inv* appears to have evolved neutrally (or under very weak selective pressure) in humans. Moreover, given the correlation between the genetic substructure and the inversion status, they suggested that recurrent events were also infrequent across this region in the *Homo* lineage.

The 17q21.31 Inversion (*17q21.31-inv*)

Another relatively common inversion polymorphism that became the focus of intense research in the last years is located at 17q21.31. In contrast to the *8p23-inv*, early studies suggested [21] that the 900 kb inversion polymorphism is undergoing selection in Europeans. After analyzing more than 29,000 Icelandic individuals, Stefansson *et al.* [21] observed that females carrying either one or two copies had more children, and, applying coalescent simulations, concluded that positive selection is likely acting on the rearrangement.

More recently, Zody *et al.* [22] analyzed the evolutionary history of the same inverted region, using data from several non-human primates. According to their results, this particular segment was prone to multiple recurrent events throughout primate evolution, which contributed to the complex duplicated architecture of the region. Moreover, they highlighted the emergence of directly oriented blocks of segmental duplications (SDs) in the human H2 haplotype (inversion-associated haplotype). SDs can act as substrates of non-allelic homologous recombination (NAHR) that can result in microdeletions and microduplications events, often associated with disease [22, 66, 67]. On this basis, Zody *et al.* [22] proposed that, due to the negative selection against the H2

haplotype, the H1 "chromosome" rose to high frequencies in humans. However, the high frequency of the H2 chromosome in some European populations (between 5 and 35%) was explained by founder effects during the peopling of Europe following the Out-of-Africa human colonization of the continent.

Similar demographic interpretations were subsequently given by Donnelly *et al.* [23] after analyzing a more detailed global distribution of the 17q21.31 haplotypes, using SNPs and short tandem repeats (STRs) polymorphisms. They found low frequencies of the H2 haplotype in most of the 63 non-European populations. Based on these observations, their model favored a complete fixation of the H1 haplotype followed by a *de novo* occurrence in the *Homo* line, hence explaining its patchy distribution. Donnelly *et al.* [23] also concluded that the Neolithic transition, rather than the first out of Africa wave, might be responsible for its present-day distribution in Europe.

Interestingly, two new and independent studies have focused on the duplicated architecture of the 17q21.31 region to further investigate its evolutionary trajectory [68, 69]. Using NGS data from more than 800 individuals and applying a strategy that combined BAC-based assemblies, read depth-base copy number estimates, BAC pool sequencing and FISH, Steinberg *et al.* [68] have identified distinct copy number polymorphisms (CNPs), including a short (CNP155) and long duplication (CNP205) exclusively associated with the H2 and H1 haplotypes, respectively. On the basis of these architectural differences, the authors were able to define four main structural haplotypes classified according to the inversion status and copy-number status. Furthermore, the frequency of the 17q21.31-inv in the African continent was reassessed by surveying a large collection of new population samples from different sources (e.g. 1000Genomes). Remarkably, it was reported that the different inversion-associated haplotypes (namely H2' and H2D) were segregating at fairly high frequencies (e.g. 7% in Maasai population) in several African ancestry groups, in opposition to earlier observations [23].

In light of these new results, Steinberg *et al.* [68] proposed a new model where an ancestral H2 haplotype arose in central or eastern Africa and spread to southern regions before the emergence of anatomically modern humans. Approximately 2.3 Million years ago the region (re-)inverted back to the direct orientation and the resulting genomic configuration (H1) spread throughout the *Homo* lineage becoming the predominant haplotype. The authors also note that the complex duplicated architecture of extant haplotypes (H2D and H1D) represents younger evolutionary events, as the duplications in the two major clades (H1 and H2) have occurred independently.

Another important conclusion from this study was finding that only one haplotype (H2D) predisposes to the syndromic 17q21.31 microdeletion, via NAHR. This configuration is characterized by the presence of directly oriented homologous SDs flanking the disease-critical region and it is associated with a duplication of the *KANSL1* locus. Intriguingly, this chromosomal variant appears to be enriched in some European populations, with frequencies reaching 25%, and with virtually no genetic variation between carriers.

Similar conclusions were reached in a parallel study by Boettger *et al.* [69], where two duplications of the *KANSL1* locus, one in each genomic background (H1 and H2), have also been reported. According to the authors, these architectural changes lead to a similar alteration at the molecular level creating a new transcript of the *KANSL* gene which may have an impact on female fertility, as demonstrated in a Drosophila mutant. [70], strengthening the initial idea of selection [21].

In summary, the (i) contradictory hypotheses raised to explain the high genetic divergence observed between the inverted and non-inverted configuration in modern humans, and (ii) the conflicting scenarios proposed to explain the expansion of inversion-carrying haplotype across populations, highlight two very important features of genetic data. First, complex spatial phenomena (e.g. human demographic expansions, contractions, and admixture events) can produce selection-like signatures in the genome [71, 72]. And secondly, species-specific characteristics, such as migration rates, population size, etc., are crucial when modeling genetic data. It is well known that human populations have gone through massive changes in size and distribution in the past, including expansions, bottlenecks, and admixture events, which resulted in distinct genetic diversity patterns among populations. However, quantifying the contribution of past events to the genetic pool of present-day populations remains a difficult task [73] in which new modeling approaches are needed.

Due to the complexity of the 17q21.31 region, the evolutionary history of this inversion remains a debated issue [68]. Although one cannot rule out the possibility of selection (nor a possible contribution of the *Homo neanderthalensis* [74]), it is quite likely that different demographic histories could produce the same patterns of variation with or without selection. Identifying the scenarios that best explain these patterns is a challenge that may be overcome with some recent advances in population genetics inference.

Simulation and Inferential Tools

One important question is whether there is an appropriate statistical framework which would allow us to choose among a set of currently proposed scenarios the most appropriate. Recent advances in population genetics modeling suggest that it may be possible thanks to improved simulation programs and to Approximate Bayesian Computation (ABC), which may provide part of the answer. In a few words, the ABC framework relies on the use of very large numbers of simulations under one or several models. The observed (or real) genetic data are summarized by several summary statistics such as the number of alleles or the expected heterozygosity. The simulated data are also summarized and compared to the observed data. The scenarios or parameter values that produce simulated data that are closest to the observed data are then considered to be the most likely ([75, 76] for a review). The ABC methodology relies on the ability to simulate genetic data very efficiently and rapidly, which was made possible thanks to the development of the coalescent theory [77]. In the last ten years the ABC framework has gained momentum and has been widely applied. It is the focus of intense research [78-81] which suggests that it

is a very flexible approach to model choice and parameter estimation. In the case of genomic data and inversions, one of the main constraints is the limitation in terms of simulating tools. While simulating large numbers of loci under the coalescent is relatively straightforward [76], even at a genome-wide scale [82], the simulation of inversions has unfortunately received little attention with few exceptions [83].

To our knowledge, invertFREGENE [83] is the first (and probably the only) software allowing the introduction of a single inversion polymorphism of specific length into a population. The authors ingeniously modified a version of a previously published software [84] to incorporate the possibility of modeling neutral inversion rearrangements under a finite sites mutation model.

The invertFREGENE software provides the possibility of simulating very large inversions, and to account for complex demographic scenarios to study the fate of inversions. Several features like the incorporation of population substructure, instantaneous expansions and contractions, are also allowed. However, there are several limitations which make it difficult for statistical inference. Indeed, invertFREGENE allows the simulation of inversions by specifying a "target" frequency (for instance the observed frequency today) but, since the number of simulations that actually took place in order to reach this target frequency is not kept, it is difficult to identify the parameter values most likely to produce the observed data. In other words, each run only gives the output for one successful inversion that reached the given target frequency. However, given that the code is freely available it should be possible to modify it so as to circumvent this limitation.

By using its core simulation engine, one could in principle develop an ABC approach that would allow us to identify models of recent human evolution with and without selection that best explain the current distribution of inversions in human populations. Recent simulation work by Li and Jakobsson [85] has for instance shown that the use of between several hundreds and a couple of thousands of SNPs, provides major improvements in the estimation of parameters. They did not explore the issue of model choice but other studies have done it with smaller number of loci [81, 86] For instance Fagundes *et al.* [86], were able to identify which model of human evolution was best supported using only 50 independent DNA sequences. With the arrival of genomic data, one could potentially determine how different regions of the genome are best explained by models with or without selection. Inverted regions could easily be typed for hundreds of SNPs and their demographic history compared to that of other regions. The general ABC framework has its limits. For instance, using forward-in-time simulators, such as invertFREGENE, could prove computationally very demanding. However, it is currently one of the most flexible and powerful approaches to explore the properties of genomic data, including inversions.

HUMAN INVERSIONS - OVERLOOKED ISSUES AND FUTURE PERSPECTIVES

Disease Associated Inversions

That the inversion of a DNA segment could interfere with gene function by disrupting its reading frame or rearranging the position of promoters, enhancers and other regulatory elements, should not be surprising. However, with the exception of a recurrent inversion located on the X chromosome [87], most human inversions do not appear to be directly linked to disease.

More often inversion rearrangements are associated with complex genomic disorders, as recently reviewed in [8]. In fact, due to the characteristic duplicated architecture of inversion breakpoints, they apparently increase the probability of disease through the occurrence of unbalanced rearrangements in the offspring [8, 22, 66, 67].

As seen above, if the duplicated copies present the same orientation they may lead to deletion or duplication events, as a result of NAHR [66, 67]. Theoretically, both events are expected to occur in equal proportions [88], however duplication-associated syndromes are rarely reported [89].

Several interpretations can be given for this observation. For instance, it has been suggested that, due to the phenotype variability observed in patients with NAHR-mediated duplications [88, 89], mild effects tend to be underdiagnosed, generating an ascertainment bias in evaluating the frequency of duplication-associated disorders.

Alternatively, one might also hypothesize that, depending on the size and the number of genes located on the duplicated region, extremely severe outcomes (i.e. abortion) might also result (e.g. as a consequence of gene dosage) from such events. However, our current interpretations rely on estimates of transmitted chromosomes, potentially generating a bias against negatively selected gametes. As a consequence, these suggestions remain merely speculative.

Lastly, the mechanistic details by which inversions contribute to complex genomic disorders are still unclear. Even with emerging technologies allowing the characterization of inversion breakpoints, SDs vary extensively in copy number [4] and understanding how these polymorphic features can simultaneously contribute (i) as a source of genetic variation, and (ii) to the establishment of human disorders, remains an important challenge to human genetics research.

Inversion Hotspots

From an evolutionary perspective, the presence of almost identical duplicated sequences in inversion breakpoints is also intriguing. Consider, for instance, the whole-genome comparative study by Murphy *et al.* [90] where the genome organization of 8 mammalian species was analyzed in order to identify patterns of chromosome evolution. Using homologous synteny blocks (HSBs) they have identified several regions of chromosome breakage that apparently have been reused throughout evolution (i.e. independent breaks occurring at the same chromosomal sites). Interestingly, the authors have also observed that most of primate-specific breaks involve inversions that have been generated via NAHR between duplicated HSBs.

Further support was later provided by Caceres *et al.* [91] who identified another example of long-term breakpoint reuse throughout mammalian evolution in a genomic segment containing a polymorphic inversion on the human X chromosome. By sequence comparison between 28 placental mammals, the authors have suggested that at least 10 inde-

pendent recurrent events must be considered to accommodate the present-day genomic structures observed in different species. In addition, recurrent events within multiple primate lineages have also been proposed for the 17q21.31 region [22].

Overall, these results appear to suggest that some genomic locations might exhibit greater rearrangement activity than others. One interesting possibility is that some regions represent conserved inversion hotspots that could have been maintained due to important functional or regulatory properties associated with the duplications [91]. Indeed, after analyzing a specific class of duplicated structures, defined as inverted repeats (IRs), Warburton *et al.* [92] hypothesized that their maintenance during primate evolution could be linked to important regulatory mechanisms controlling deleterious gene expression on sex-chromosomes.

In conclusion, future work is still needed in order to determine the distribution of these apparently non-randomly distributed break sites, as studies analyzing at depth the population genetics of inversions are scarce in the literature.

Inversions and Recombination Rate

Many authors have also overlooked the effect of chromosomal inversions on the overall recombination rate, despite the vital role of crossing over during meiosis for proper chromosome segregation [93, 94].

In humans, as in many other organisms [95, 96], recombination is affected by several genomic features, such as location (e.g. lower rates near centromeres and higher near telomeres), and gene density (but see [97] for a more detailed review). Interestingly, it has also been shown that most recombination events (approximately 80%) are concentrated in small genomic regions of 1-2 kb, known as recombination hotspots [98-100].

The PRDM9 gene was recently described [101, 102] as a major regulator of human recombination hotspots, with allelic variants of this gene influencing the differential usage of recombination hotspots. However, one might hypothesize that if an inversion happens to encompass an active hotspot, recombination will likely become inhibited in that particular region, disturbing the overall recombination rate by possibly de-localizing crossing-over events to different locations. For instance, it has been argued that, in *Drosophila* species, inversions significantly increase the recombination rate throughout the rest of the genome [94]. Interestingly, it has also been consistently reported that polymorphisms on the H2 (inverted) haplotype in 17q21.31 are associated with an increase of the genome-wide recombination rate in heterozygous females [21, 103]. Have these inversions trapped specific variants of more active recombination hotspot determinants? That is an intriguing possibility; however, the recombination machinery might be extremely different between these species, since no recombination hotspots were ever reported in *Drosophila* [97].

On an evolutionary time-scale, inversions may lead to new stabilizing points of the map of recombination events within the affected chromosome, as has been recurrently observed in the establishment of dimorphic sex chromosomes of mammals and other distantly related vertebrate taxa [104-106] as well as in plants [107]. In fact, incipient heteromorphic sex chromosomes (Y and Z chromosomes) often differentiate via the accumulation of inversion rearrangements that prevent recombination over increasingly large regions with their homologues. Nevertheless, recombination and successful disjunction are maintained and therefore the recombination machinery may be more labile than would be expect *a priori*.

Moreover, since current estimates suggest that approximately 25,000 putative hotspots exist in the human genome [100] understanding how inversion rearrangements might affect or contribute to differential hotspot usage will be a challenging task.

CONCLUSIONS

Given the increased interest on chromosomal rearrangements, scientists are now beginning to recognize inversions as important players shaping genetic variation. Over the last decade, fundamental questions began to emerge focusing on their molecular properties [66, 67], on the mechanisms responsible for their origin [108], on their evolutionary significance [1, 47, 109] and on their role in speciation [14, 47, 48, 50-52].

In humans, extensive sequencing efforts have revealed a somewhat surprising abundance of inversions segregating as polymorphisms [29]. This observation is in sharp contrast with previous expectations that suggested a direct impact of inversions on fertility [109]. However, as seen above, it is evident that such impact might be influenced by a combination of multiple processes [110, 111].

As genomic information continues to accumulate in publicly available databases, new *in silico* approaches combined with evidences of human demographic history – based on archaeological and linguistic theories - might prove useful when exploring the role of inversion polymorphisms as evolutionary significant elements. Nevertheless, genetic data should be used with extreme caution as different plausible scenarios might fit the observed patterns of present day diversity [72, 73].

In our opinion, due to its flexibility, robustness and efficiency, ABC (or other genomic inference) strategies should be considered in future studies, as these approaches allow us to quantify the relative contribution of ancient and recent factors, including selection, in shaping the genetic structure of present-day populations. Even if ABC modeling only represents an approximation, they surely constitute a promising statistical inferential framework to reconstruct important aspects of the evolutionary history of populations.

In addition, another limitation in most inversion-based evolutionary studies is that most authors only consider the evolutionary effects of a single genomic inversion. However, several inversions might operate simultaneously on an individuals' genome, and, no matter how accurate the methodology used, some confounding variables might create an apparent association or mask a real one. Fortunately, we are at a stage where comparative data might enable us to answer some of these questions.

In conclusion, future studies merging evolutionary and molecular perspectives will allow us to understand the impli-

cations of this specific type of structural variability to complex diseases, and how selective factors could have influenced their evolution.

CONFLICT OF INTEREST

The author(s) confirm that this article content has no conflicts of interest.

ACKNOWLEDGEMENTS

We would like to thank Gabriel Marais for a critical review of the manuscript, and for additional comments and suggestions. IPATIMUP is an Associate Laboratory of the Portuguese Ministry Education and Science and is partially supported by FCT, the Portuguese Foundation for Science and Technology. JMA is the recipient of a doctoral fellowship from FCT (SFRH/BD/51004/2010). AML is the recipient of a postdoctoral fellowship from FCT (SFRH/BPD/ 73366 /2010). LC is partly funded by the Laboratoire d'Ex cellence (LABEX) entitled TULIP (ANR-10-LABX-41).

REFERENCES

[1] Kirkpatrick, M. How and why chromosome inversions evolve. *PLoS biology*, **2010**, *8* (9).

[2] Dobzhansky, T.*Genetics and the origin of species*; Columbia University Press: New York, 1951.

[3] Alkan, C.; Coe, B.P.; Eichler, E.E.Genome structural variation discovery and genotyping. *Nature reviews. Genetics*, **2011,** *12*(5), 363-76.

[4] Alkan, C.; Kidd, J.M.; Marques-Bonet, T.; Aksay, G.; Antonacci, F.; Hormozdiari, F.; Kitzman, J.O.; *et al.* Personalized copy number and segmental duplication maps using next-generation sequencing. *Nature genetics*, **2009**, *41*(10), 1061-7.

[5] Craddock, N.; Hurles, M.E.; Cardin, N.; Pearson, R.D.; Plagnol, V.; Robson, S.; Vukcevic, D.; *et al.* Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. *Nature*, **2010**, *464*(7289), 713-20.

[6] Gonzalez, E; Kulkarni, H.; Bolivar, H.; Mangano, A.; Sanchez, R.; Catano, G.; Nibbs, R.J.; Freedman, B.I.; Quinones, M.P.; Bamshad, M.J.; *et al.* The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. *Science,* **2005***, 307*, 1434-1440.

[7] Fanciulli, M.; Norsworthy, P.J.; Petretto, E.; Dong, R.; Harper, L.; Kamesh, L.; Heward, J.M.; Gough, S.C.; de Smith, A.; Blakemore, A.I.; *et al.* FCGR3B copy number variation is associated with susceptibility to systemic, but not organ-specific, autoimmunity. *Nature genetics,* **2007***, 39*, 721-723.

[8] Feuk, L. Inversion variants in the human genome: role in disease and genome architecture. *Genome medicine*, **2007***, 2*(2), 11.

[9] Bailey, J.A.; Eichler, E.E. Primate segmental duplications: Crucibles of evolution, diversity, and disease. *Nature reviews,* **2006***, 7*, 552–564.

[10] Conrad, D.F.; *et al.* Origins and functional impact of copy number variation in the human genome. *Nature,* **2010**, *464*, 704–712.

[11] King, M. *Species Evolution: The Role of Chromosome Change*; Cambridge University Press: Cambridge, New York, 1993.

[12] White, M.J.D. *Modes of Speciation;* WH Freeman and Company: San Francisco, California, 1978.

[13] Navarro, A.; Barbadilla, A.; Ruiz, A. Effect of inversion polymorphism on the neutral nucleotide variability of linked chromosomal regions in Drosophila. *Genetics,* **2000***, 155(2)*, 685-698.

[14] Rieseberg, L.H. Chromosomal rearrangements and speciation. *Trends in ecology & evolution (Personal edition),* **2001***, 16*(7), 351-358.

[15] Joron, M.; Frezal, L.; Jones, R.T.; *et al.* Chromosomal rearrangements maintain a polymorphic supergene controlling butterfly mimicry. *Nature,* **2011***, 477*, 203–206.

[16] Fragata, I.; Balanyà, J.; Rego, C.; *et al.* Contrasting patterns of phenotypic variation linked to chromosomal inversions in native and colonizing populations in Drosophila subobscura. *Journal of Evolutionary Biology*, **2009***, 23(1)*, 112-123.

[17] Lowry, D.B.; Willis, J.H. A widespread chromosomal inversion polymorphism contributes to a major life-history transition, local adaptation, and reproductive isolation. *PLoS Biology,* **2010***, 8(9)*, e1000500.

[18] Ayala, D. ; Fontaine, M.C.; Cohuet, A.; Fontenille, D.; Vitalis, R.; Simard, F. Chromosomal inversions, natural selection and adaptation in the malaria vector Anopheles funestus. *Molecular biology and evolution*, **2010***, 28*(1), 745-758.

[19] Antonacci, F.; Kidd, J.M.; Marques-Bonet, T.; *et al.* Characterization of six human disease-associated inversion polymorphisms. *Human Molecular Genetics,* **2009***, 18(14)*, 2555-2566.

[20] Iafrate, A.J.; Feuk, L.; Rivera, M.N.; Listewnik, M.L.; Donahoe, P.K.; Qi, Y.; Scherer, S.W.; Lee, C. Detection of large-scale variation in the human genome. *Nature genetics,* **2004***, 36(9)*, 949-51.

[21] Stefansson, H.; Helgason, A.; Thorleifsson, G.; Steinthorsdottir, V.; Masson, G.; Barnard, J.; Baker, A.; *et al.* A common inversion under selection in Europeans. *Nature genetics*, **2005***, 37*(2), 129-37.

[22] Zody, M.C.; Jiang, Z.; Fung, H.C.; Antonacci, F.; Hillier, L.W.; Cardone, M.F.; Graves, T.A; *et al.* Evolutionary toggling of the MAPT 17q21.31 inversion region. *Nature genetics*, **2008***, 40*(9), 1076-83.

[23] Donnelly, M.P.; Paschou, P.; Grigorenko, E.; Gurwitz, D.; Mehdi, S.Q.; Kajuna, S.L.B.; Barta, C.; *et al.* The distribution and most recent common ancestor of the 17q21 inversion in humans. *American journal of human genetics*, **2010***, 86*(2), 161-71.

[24] Salm, M.P.A.; Horswell, S.D.; Hutchison, C.E.; *et al.* The origin, global distribution, and functional impact of the human 8p23 inversion polymorphism. *Genome Research,* **2012**, *22(6),* 1144-1153.

[25] Wilson, M.G.; Towner, J.W.; Coffin, G.S.; Forsman, I. Inherited pericentric inversion of chromosome no. 4. *American Journal of Human Genetics,* **1970***, 22*, 679-690.

[26] de la Chapelle, A.; Schroder, J.; Stenstrand, K.; Fellman, J.; Herva, R.; Saarni, M.; Anttolainen, I.; Tallila, I.; *et al.* Pericentric inversions of human chromosomes 9 and 10. *American Journal of Human Genetics,***1974**, *26*, 746-766.

[27] O'Neill, R.J.; Eldridge, M.D.B.; Metcalfe, C.J. Centromere dynamics and chromosome evolution in marsupials. *The Journal of heredity*, **2004**, *95*(5), 375-81.

[28] Hara, Y.; Imanishi T. Abundance of Ultramicro Inversions within Local Alignments between Human and Chimpanzee Genomes. *BMC evolutionary biology*, **2011**, *11*(1), 308.

[29] Feuk, L.; Macdonald, J.R.; Tang, T.; Carson, A.R.; Li, M.; Rao, G.; Khaja, R.; Scherer, S.W. Discovery of human inversion polymorphisms by comparative analysis of human and chimpanzee DNA sequence assemblies. *PLoS Genetics*, **2005**, *1(4)*, e56.

[30] Turner, D.J.; Shendure, J.; Porreca, G.; Church, G.; Green, P.; Tyler-Smith, C.; Hurles M.E. Assaying chromosomal inversions by single-molecule haplotyping. *Nature Methods,* **2006***, 3*, 439-445.

[31] Bansal, V.; Bashir, A.; Bafna, V. Evidence for large inversion polymorphisms in the human genome from HapMap data. *Genome research*, **2007**, *17*(2), 219-30.

[32] Sindi, S.S.; Raphael, B.J. Identification and frequency estimation of inversion polymorphisms from haplotype data. *Journal of computational biology: a journal of computational molecular cell biology*, **2010**, *17*(3), 517-31.

[33] Tuzun, E.; Sharp, A.; Bailey, J.; Kaul, R.; Morrison, V.; Pertz, L.; Haugen, E.; Hayden, H.; Albertson, D.; Pinkel, D.; *et al.* Fine-scale structural variation of the human genome. *Nature Genetics,* **2005**, *37*, 727–732.

[34] Kidd, J.M.; Cooper, G.M.; Donahue, W.F.; Hayden, H.S.; Sampas, N.; Graves, T.; Hansen, N.; *et al.* Mapping and sequencing of structural variation from eight human genomes. *Nature*, **2008**, *453*(7191), 56-64.

[35] Rao, P.N.; Li, W.; Vissers, L.E.L.M.; Veltman, J.A.; Ophoff, R.A. Recurrent Inversion Events at 17q21.31 Microdeletion Locus Are Linked to the MAPT H2 Haplotype. *Cytogenetic and genome research*, **2010**, *90095*, 275-279.

[36] Korbel, J.O.; Urban, A.E.; Affourtit, J.P.; Godwin, B.; Grubert, F.; Simons, J.F.; Kim, P.M.; Palejev, D.; Carriero, N.J.; *et al.* Paired-end mapping reveals extensive structural variation in the human genome. *Science,* **2007**, *318*, 420– 426.

[37] Medvedev, P.; Stanciu, M.; Brudno, M. Computational methods for discovering structural variation with next-generation sequencing. *Nature Methods,* **2009**, *6,* S13–S20.

[38] Levy, S.; Sutton, G.; Ng, P.C.; Feuk, L.; Halpern, A.L.; Walenz, B.P.; Axelrod, N.; Huang, J.; Kirkness, E.F.; *et al.* The diploid ge-

nome sequence of an individual human. *PLoS Biology,* **2007**, *5*, e254.

[39] Nothnagel, M.; Herrmann, A.; Wolf, A.; Schreiber, S.; Platzer, M.; Siebert, R.; Krawczak, M.; *et al.* Technology-specific error signatures in the 1000 Genomes Project data. *Human genetics*, **2011**, 505-516.

[40] Baird, N.A.; Etter, P.D.; Atwood, T.S.; Currey, M.C.; Shiver, A.L.; Lewis, Z.A.; Selker, E.U.; Cresko, W.A.; Johnson, E.A. Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers. *PLoS One,* **2008**, *3,* 110-115.

[41] Davey, J.W.; Hohenlohe, P.A.; Etter, P.D.; Boone, J.Q.;Catchen, J.M.; Blaxter, M.L. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Review Genetics,* **2011**, *12*, 499.

[42] Garvin, M.R.; Saitoh, K.; Gharrett, A.J. Application of single nucleotide polymorphisms to non-model species: a technical review. *Molecular Ecology Resources,* **2010**, *10(6)*, 915-934.

[43] Hohenlohe, P.A.; Bassham, S.; Etter, P.D.; Stiffler, N.; Johnson, E.A.; Cresko, W.A. Population Genomics of Parallel Adaptation in Threespine Stickleback using Sequenced RAD Tags. *PLoS Genetics,* **2010**, *6(2)*, e1000862.

[44] Alkan, C.; Sajjadian, S.; Eichler, E.E. Limitations of next-generation genome sequence assembly. *Nature Methods,* **2010**, *8*, 61–65.

[45] Spitz, F.; Herkenne, C.; Morris, M.A.; Duboule, D. Inversion-induced disruption of the Hoxd cluster leads to the partition of regulatory landscapes. *Nature Genetics,* **2005**, *37*, 889–893.

[46] Hoffmann, A.A.; Rieseberg, L.H. Revisiting the Impact of Inversions in Evolution: From Population Genetic Markers to Drivers of Adaptive Shifts and Speciation? *Annual review of ecology, evolution, and systematics*, **2008**, *39*, 21-42.

[47] Hey, J. Speciation and inversions: chimps and humans. *BioEssays*, **2003**, 825-828.

[48] Sturtevant, A.H. A case of rearrangement of genes in *Drosophila*. *Proceedings of the National Academy of Sciences,* **1921**, *7*, 235–23.

[49] Kirkpatrick, M.; Barton, N. Chromosome inversions, local adaptation and speciation. *Genetics*, **2006**, *173*(1), 419-34.

[50] Navarro, A.; Barton, N.H. Accumulating postzygotic isolation genes in parapatry: a new twist on chromosomal speciation. *Evolution; international journal of organic evolution*, **2003**, *57*(3), 447-59.

[51] Noor, M.A.F.; Grams, K.L.; Bertucci, L.A.; Reiland, J. Chromosomal inversions and the reproductive isolation of species. *Proceedings of the National Academy of Sciences*, *98*, **2001**, 12084–12088.

[52] Faria, R.; Neto, S.; Navarro, A. Role of Natural Selection in Chromosomal Speciation; In *Encyclopedia Life Sciences*. Chichester, UK, 2011.

[53] Orr, A. Dobzhansky, Bateson, and the Genetics of Speciation. *Genetics,* **1996**, *144*, 1331-1335.

[54] Huynh, L.Y.; Maney, D.L.; Thomas, J.W. Chromosome-wide linkage disequilibrium caused by an inversion polymorphism in the white-throated sparrow (Zonotrichia albicollis). *Heredity,* **2011**, *106*, 537–546.

[55] Bardhan, A.; Sharma, T. Meiosis and speciation: a study in a speciating Mus terricolor complex. *Journal of Genetics,* **2000**, *79*, 105–11.

[56] Navarro, A.; Barton, N.H. Chromosomal speciation and molecular divergence-accelerated evolution in rearranged chromosomes". *Science,* **2003**, *300(5617),* 321-324.

[57] Lu, J.; Li, W.H.; Wu, C.I. Comment on "Chromosomal speciation and molecular divergence-accelerated evolution in rearranged chromosomes". *Science,* **2003**, *302,* 988.

[58] Zang, J.; Wang, X.; Podlaha, O. Testing the chromosomal speciation hypothesis for humans and chimpanzees. *Genome Research,* **2004**, *14,* 845-851.

[59] Spirito, F. *Endless Forms: Species and Speciation*; Oxford University Press: Oxford, New York, 1998.

[60] Spirito, F.; Rizzoni, M.; Rossi, C. The establishment of underdominant chromosomal rearrangements in multi-deme systems with local extinction and colonization. *Theoretical Population Biology,* **1993**, *44*, 80-94.

[61] Bugge, M.; Bruun-Petersen, G.; Brøndum-Nielsen, K.; *et al.* Disease associated balanced chromosome rearrangements: a resource for large scale genotype-phenotype delineation in man. *Journal of Medical Genetics,* **2000**, *37,* 858-865.

[62] Shaw, C.J.; Lupski. J.R. Implications of human genome architecture for rearrangement-based disorders: the genomic basis of disease. *Human Molecular Genetics,* **2004**, *13(1),* 57-64.

[63] Bosch, N.; Morell, M.; Ponsa, I.; *et al.* Nucleotide, Cytogenetic and Expression Impact of the Human Chromosome 8p23.1 Inversion Polymorphism. *PLoS One*, **2009**, *4(12)*: e8269.

[64] Hollox, E.J.; Barber, J.C.K.; Brookes, A.J.; Armour, J.A.L. Defensins and the dynamic genome: What we can learn from structural variation at human chromosome band 8p23.1. *Genome Research,* **2008**, *18*, 1686-1697.

[65] Simpfendorfer, K.R.; Olsson, L.M.; Manjarrez Orduno, N.; *et al.* The autoimmunity-associated BLK haplotype exhibits cis-regulatory effects on mRNA and protein expression that are prominently observed in B cells early development. *Human Molecular Genetics,* **2012**, *21(17),* 3918-3925.

[66] Flores, M.; Morales, L.; Gonzaga-Jauregui, C.; Domínguez-Vidaña, R.; Zepeda, C.; Yañez, O.; Gutiérrez, M.; *et al.* Recurrent DNA inversion rearrangements in the human genome. *Proceedings of the National Academy of Sciences of the United States of America*, **2007**, *104*(15), 6099-106.

[67] Gu, W.; Zhang, F.; Lupski, J.R. Mechanisms for human genomic rearrangements. *Pathogenetics,* **2008**, *1(4)*.

[68] Steinberg, K.M.; Antonacci, F.; Sudmant, P.H.; *et al.* Structural diversity and African origin of the 17q21.31 inversion polymorphism. *Nature genetics,* **2012**, *44(8),* 872–880.

[69] Boettger, L.M.; Handsaker, R.E.; Zody, M.C.; McCarroll, S.A. Structural haplotypes and recent evolution of the human 17q21.31 region. *Nature genetics,* **2012**, *44(8)*, 881-885.

[70] Yu, L.; Song, Y.; Wharton, R.P. E(nos)/CG4699 required for nanos function in the female germ line of Drosophila. *Genesis,* **2010**, *48,* 161–170.

[71] Klopfstein, S.; Currat, M.; Excoffier, L. The fate of mutations surfing on the wave of a range expansion. *Molecular Biology and Evolution,* **2005**, *23*, 482–490.

[72] Currat, M.; Excoffier, L.; Maddison, W.; Otto, S.P.; Ray, N.; *et al.* Comment on ''ongoing adaptive evolution of ASPM, a brain size determinant in homo sapiens'' and ''microcephalin, a gene regulating brain size, continues to evolve adaptively in humans''. *Science,* **2006**, *313*, 172.

[73] Chikhi, L. Update to Chikhi *et al.*'s "Clinal Variation in the Nuclear DNA of Europeans" (1998): Genetic Data and Storytelling—From Archaeogenetics to Astrologenetics? *Human Biology,* **2009**, *81(5-6)*, 639-643.

[74] Hardy, J.; Pittman, A.; Myers, A.; Gwinn-Hardy, K.; Fung, H.C.; de Silva, R.; Hutton, M.; Duckworth, J. Evidence suggesting that Homo neanderthalensis contributed the H2 MAPT haplotype to Homo sapiens. *Biochemical Society Transactions,* **2005**, *33(4)*, 582–585.

[75] Beaumont, M.A.; Zhang, W.; Balding, D.J. Approximate Bayesian computation in population genetics. *Journal of Genetics,* **2002**,*162*, 2025–35.

[76] Beaumont, M.A. Approximate Bayesian Computation in Evolution and Ecology. *Annual Review of Ecology, Evolution, and Systematics*, **2010**, *41*(1), 379-406.

[77] Hudson, R. *Gene genealogies and the coalescent process*; in Oxford Surveys in Evolutionary Biology. Oxford University Press: Oxford, New York, 1990.

[78] Wegmann, D.; Leuenberger, C.; Excoffier, L. Efficient Approximate Bayesian Computation coupled with Markov chain Monte Carlo without likelihood. *Journal of Genetics,* **2009**, *182*, 1207–1218.

[79] Blum, M.B. Approximate Bayesian Computation: A Nonparametric Perspective. *Journal of the American Statistical Association,* **2010**, *105(491)*, 1178-1187.

[80] Sousa, V.; Fritz, M.; Beaumont, M.A.; Chikhi, L. Approximate Bayesian Computation without summary statistics: the case of admixture. *Journal of Genetics,* **2009**, *181*, 187-197.

[81] Sousa, V.; Beaumont, M.A.; Fernandes, P.; Coelho, M.M.; Chikhi, L. Population divergence with or without admixture: selecting models using an ABC approach. *Heredity*, **2011**, *108,* 521-530.

[82] Carvajal-Rodríguez, A. Simulation of Genes and Genomes Forward in Time. *Current Genomics,* **2010**, *11*, 58-61.

[83] O'Reilly, P.F.; Coin, L.J.M.; Hoggart, C.J. invertFREGENE: software for simulating inversions in population genetic data. *Bioinformatics,* **2010**, *26(6)*, 838-840.

[84] Chadeau-Hyam, M.; Hoggart, C.J.; O'Reilly, P.F.; Whittaker, J.C.; De Iorio, M.; Balding, D.J. Fregene: simulation of realistic sequence-level data in populations and ascertained samples. *BMC Bioinformatics,* **2008**, *9*, 364.

[85] Li, S.; Jakobsson, M. Estimating demographic parameters from large scale population genomic data using Approximate Bayesian Computation. *BMC Genetics,* **2012**, *13,* 22.

[86] Fagundes, N.J.R.; Ray, M.; Beaumont, M.; *et al.* Statistical evaluation of alternative models of human evolution. *Proceedings of the National Academy of Sciences,* 2007, *104,* 17,614–17,619.

[87] Antonarakis, J.P.; Rossiter, J.P.; Young, M.; Horst, J.; de Moerloose, P.; *et al.* Factor VIII gene inversions in severe hemophilia A: results of an international consortium study. *Blood,* **1995**, *86*, 2206–2212.

[88] Ensenauer, R.E.; Adeyinka, A.; Flynn, H.C.; Michels, V.V.; Lindor, N.M.; Dawson, D.B.; Thorland, E.C.; Lorentz, C.P.; Goldstein, J.L.; McDonald, M.T.; *et al.* Microduplication 22q11.2, an emerging syndrome: clinical, cytogenetic, and molecular analysis of thirteen patients. *American Journal of Human Genetics,* **2003**, *73*, 1027–1040.

[89] Grisart, B.; Willatt, L.; Destrée, A.; Fryns, J-P.; Rack, K.; de Ravel, T.; Rosenfeld, J.; *et al.* 17q21.31 microduplication patients are characterised by behavioural problems and poor social interaction. *Journal of Medical Genetics,* **2009**, *46*, 524-530.

[90] Murphy, W.J.; Larkin, D.M.; Everts-van der Wind, A.; Bourque, G.; Tesler, G.; Auvil, L.; Beever, J.E.; Chowdhary, B.P.; Galibert, F.; Gatzke, L.; *et al.* Dynamics of mammalian chromosome evolution inferred from multispecies comparative maps. *Science,* **2005**, *309*, 613-617.

[91] Caceres, M.; NIHISCCSP; Sullivan, R.T.; Thomas, J.W. A recurrent inversion on the eutherian X chromosome. *Proceedings of the National Academy of Sciences,* **2007**, *104(47)*, 18571-18576.

[92] Warburton, P.E.; Giordano, J.; Cheung, F.; *et al.* Inverted repeat structure of the human genome: the X-chromosome contains a preponderance of large, highly homologous inverted repeats that contain testes genes. *Genome Research,* **2004**, *14,* 1861–1869.

[93] Fledel-alon, A.; Leffler, E.M.; Guan, Y.; Stephens, M.; Coop, G.; Przeworski, M. Variation in Human Recombination Rates and Its Genetic Determinants. *Plos One,* **2011**, *6(6)*, e20321.

[94] Stevison, L.S.; Hoehn, K.B.; Noor, M.A.F. Effects of inversions on within- and between-species recombination and divergence. *Genome Biology and Evolution*, **2011**, *3*, 830-841.

[95] Petes, T.D. Meiotic recombination hot spots and cold spots. *Nature Review Genetics,* **2001**, *2*, 360–369.

[96] Jensen-Seaman, M.I.; Furey, T.S.; Payseur, B.A.; Lu, Y.; Roskin, K.M.; Chen, C-F.; Thomas, M.A.; Haussler, D.; Jacob, H.J. Comparative recombination rates in the rat, mouse, and human genomes. *Genome Research,* **2004**, *14*, 528–538.

[97] Coop, G.; Przeworski, M. An evolutionary view of human recombination. *Nature reviews Genetics*, **2007**, *8(1)*, 23-34.

[98] Goldstein, D.B.; Chikhi, L. Human migrations and population structure : what we know and why it matters. *Annual Review of Genomics and Human Genetics,* **2002**, *3*, 129-152.

[99] Goldstein, D.B.; Weale, M.E. Population genomics: Linkage disequilibrium holds the key. *Current Biology,* **2001**, *11*, 576–579.

[100] Myers, S.; Bottolo, L.; Freeman, C.; McVean, G.; Donnelly, P. A fine-scale map of recombination rates and hotspots across the human genome. *Science,* **2005**, *310*, 321–324.

[101] Baudat, F.; Buard, J.; Grey, C.; Fledel-Alon, A.; Ober, C.; Przeworski, M.; Coop, G.; de Massy, B. PRDM9 Is a major determinant of meiotic recombination hotspots in humans and mice. *Science,* **2010**, *237*, 836-840.

[102] Berg IL, Neumann R, Lam K-WG, Sarbajna S, Odenthal-Hesse L, *et al.* (2010). PRDM9 variation strongly influences recombination hot-spot activity and meiotic instability in humans. *Nature Genetics, 42*, 859–863.

[103] Chowdhury, R.; Bois, P.R.; Feingold, E.; Sherman, S.L.; Cheung, V.G. Genetic analysis of variation in human meiotic recombination. *PLoS Genetics,* **2009**, *5*, e1000648.

[104] Lahn, B.T.; Page, D.C. Four evolutionary strata on the human X chromosome. *Science,* **1999**, *286(5441)*, 964-7.

[105] Skaletsky, H.; Kuroda-Kawaguchi, T.; Minx, P.J.; Cordum, H.S.; Hillier, L.; Brown, L.G.; Repping, S.; Pyntikova, T.; Ali, J.; *et al.* The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature,* **2003**, *423(6942)*, 825-37.

[106] Ross, M.T.; Grafham, D.V.; Coffey, A.J.; Scherer, S.; McLay, K.; Muzny, D.; Platzer, M.; Howell, G.R.; Burrows, C.; Bird, C.P.; *et al.* The DNA sequence of the human X chromosome. *Nature,* **2005**, *434*, 325–337.

[107] Matsunaga, S. Sex chromosome-linked genes in plants. *Genes & Genetic Systems,* **2006**, *81*, 219–226.

[108] Lee, J.A.; Carvalho, C.M.; Lupski, J.R. A DNA replication mechanism for generating non-recurrent rearrangements associated with genomic disorders. *Cell,* **2007**, *131*, 1235–47.

[109] Brown, J.D.; O'Neill, R.J. Chromosomes, Conflict, and Epigenetics: Chromosomal Speciation Revisited. *Annual review of genomics and human genetics*, **2010**, 291-316.

[110] Nosil, P.; Feder, J.L. Genomic divergence during speciation: causes and consequences. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **2011**, *367*(1587), 332-342.

[111] Guerrero, R.F.; Rousset, F.; Kirkpatrick, M. Coalescent patterns for chromosomal inversions in divergent populations. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, **2012**, *367*(1587), 430-438.