

# Three Dimensional Organization of Genome Might Have Guided the Dynamics of Gene Order Evolution in Eukaryotes

Meenakshi Bagadia, Arashdeep Singh, and Kuljeet Singh Sandhu\*

Department of Biological Sciences, Indian Institute of Science Education and Research (IISER)—Mohali, SAS Nagar 140306, India

\*Corresponding author: E-mail: sandhuks@iisermohali.ac.in.

Accepted: March 2, 2016

## Abstract

In eukaryotes, genes are nonrandomly organized into short gene-dense regions or “gene-clusters” interspersed by long gene-poor regions. How these gene-clusters have evolved is not entirely clear. Gene duplication may not account for all the gene-clusters since the genes in most of the clusters do not exhibit significant sequence similarity. In this study, using genome-wide data sets from budding yeast, fruit-fly, and human, we show that: 1) long-range evolutionary repositioning of genes strongly associate with their spatial proximity in the nucleus; 2) presence of evolutionary DNA break-points at involved loci hints at their susceptibility to undergo long-range genomic rearrangements; and 3) correlated epigenetic and transcriptional states of engaged genes highlight the underlying evolutionary constraints. The significance of observation 1, 2, and 3 are particularly stronger for the instances of inferred evolutionary gain, as compared with loss, of linear gene-clustering. These observations suggest that the long-range genomic rearrangements guided through 3D genome organization might have contributed to the evolution of gene order. We further hypothesize that the evolution of linear gene-clusters in eukaryotic genomes might have been mediated through spatial interactions among distant loci in order to optimize co-ordinated regulation of genes. We model this hypothesis through a heuristic model of gene-order evolution.

**Key words:** genome organization, gene-clusters, genomic rearrangements, chromatin interactions.

## Introduction

Multiple lines of evidence refute the presumption that eukaryotic genome organization is random and it is not plausible anymore to consider a gene as an autonomous transcriptional unit (Hurst et al. 2004; Kosak and Groudine 2004). Eukaryotic genome is often organized into short gene-rich regions or “gene-clusters” interrupted by long gene-poor regions (Lawrence 1999; Hurst et al. 2004). The linear clustering of genes is shown to be evolutionarily constrained in eukaryotic genomes (Blanco et al. 2008). What might have constrained such ordering of genes? Various studies on numerous organisms suggest that the genes in the gene clusters tend to coexpress (Cohen et al. 2000; Spellman and Rubin 2002), can be involved in the same metabolic pathway (Lee and Sonnhammer 2003) and may interact with each other at protein level (Teichmann and Veitia 2004). The present understanding is that the deposition of similar chromatin marks and concurrent opening and closing of chromatin might mediate coexpression of functionally similar genes and selection of such chromatin level coordination might have been favored

in the evolution. This model is also supported by the existence of large domains of distinct chromatin types, like actively transcribing, weakly transcribing, poised, and repressed chromatin domains (de Wit et al. 2008; Filion et al. 2010). This idea of domain organization of genome has important implications in developmental reprogramming, disease development, and pathogenicity due to position effect (Kleinjan and van Heyningen 2005; Kleinjan and Lettice 2008; van Heyningen and Bickmore 2013). Common *cis*-element requirement might impose another constraint that could favor the linear clustering of genes (Frasch et al. 1995; Flint et al. 2001; Ohlsson et al. 2001; Lomvardas et al. 2006; Splinter et al. 2006; Spitz and Duboule 2008; Sandhu et al. 2009; Vernimmen et al. 2009; Tena et al. 2011; Berlivet et al. 2013; Maeso et al. 2013). Some studies have also suggested that the constraints to minimize the expression noise by keeping the essential genes in the regions of open chromatin can guide the linear clustering of genes (Batada and Hurst 2007).

© The Author 2016. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact [journals.permissions@oup.com](mailto:journals.permissions@oup.com)

The mechanistic basis that governed the evolution of linear gene clustering is not entirely clear. Gene-family clusters, like *HIST*, *HOX*, *KRT*, *OR*, etc., wherein neighboring genes also share sequence similarity in addition to functional similarity, are argued to have evolved through duplication events (Ferrier and Holland 2001; Demuth et al. 2006). However, the genes in clusters other than gene-family do not generally show sequence similarity, suggesting that tandem duplication alone cannot explain their evolutionary convergence. Understanding the underlying mechanisms of gene order change, therefore, could provide some insights to evolution of gene clusters. One way the gene order can change is the segmental or whole-genome duplication followed by sequential loss of one copy duplicated genes (Fischer et al. 2001). Genomic rearrangements like inversions and translocations can also serve as potential mean to relocate genes and selection can then favor or disfavor the reconfigured gene order. Indeed, islands of genes conferring adaptation to changing environment have been proposed to be formed through genomic rearrangements (Yeaman 2013). However, there is no strong evidence supporting this model. Here, we present a few lines of evidence through analysis of 3D genome architectures, evolutionary break-points and functional data sets that support the role of 3D genome organization in mediating evolutionary dynamics of long range alterations in gene order.

## Materials and Methods

### Data Sets

The details of all the data sets used in the study are given in the [supplementary material, Supplementary Material online](#).

### Resampling Procedure

We generated random null models for each comparison by randomly picking the pairs of genomic loci having trans-interactions while preserving the chromosomal distribution and number of interactions of genomic loci same as in the original test set. This process was repeated 1,000 times to obtain a null distribution. *P* value was calculated as following:

$$p = \frac{1}{B} \left( 1 + \sum_{b=1}^B (1)_{\frac{k_b}{k'} > \frac{k}{k'}} \right)$$

where, *B* = number of resamplings (1,000)

*k<sub>b</sub>* = proportion of resampled pairs exhibiting *trans* chromatin interactions

*k* = proportion of *trans* chromatin interactions among the observed pairs ([supplementary fig. S1, Supplementary Material online](#))

*k'* = expected proportion of *trans* chromatin interactions in the genome ([supplementary fig. S1, Supplementary Material online](#)).

Similar approach was used for figures 2*b*, 3, and 4.

### Heuristic evolutionary model of gene clustering

A heuristic model was developed to test the idea that the spatial interactions of coregulated genes could lead to linear clustering of genes. We designed a hypothetical genome of two chromosomes having 50 genes each. The genes were equally spaced. Each possible gene-pair was randomly assigned a number between 0 and 1, which represented the coregulation of genes in the gene-pair. Closer the value to 1, greater the coregulation was. Spatial interactions among genes were generated in semirandom manner. Around 30% of interacting gene-pairs were considered to have moderate level of coregulation. We generated a homogenous population of 100 such genomes. Translocations were simulated by taking random break-point on each chromosome. We varied the frequency of genomic rearrangements as 0.01, 0.1, and 0.2. These rearrangements may decrease or increase the usual intergenic distance of rearranged genes. However, if it happened to decrease, then the probability that the neighboring genes at rearranged site would interact with each other would depend on the altered distance between them. This was calculated using the following function:

$$y = 0.98x^{-0.607}$$

where *y* is the probability of interaction and *x* is the distance between the genes. This equation was derived from Milele and Dekker (2009). Rearranged loci when interacting would also tend to share their interacting partners with each other. This was depicted in the [supplementary figure S2, Supplementary Material online](#). This step introduced clusters (triangles) in the chromatin interaction network and their frequency was calculated by measuring global clustering coefficient. The triangles in the chromatin interaction network were scored for the regulatory fitness score that was given by,

$$f_{abc} = \frac{1}{2} \sum_{i \neq j} x_{ij}, \forall i, j \in \{a, b, c\} \text{ and } x_{ij} = x_{ji}$$

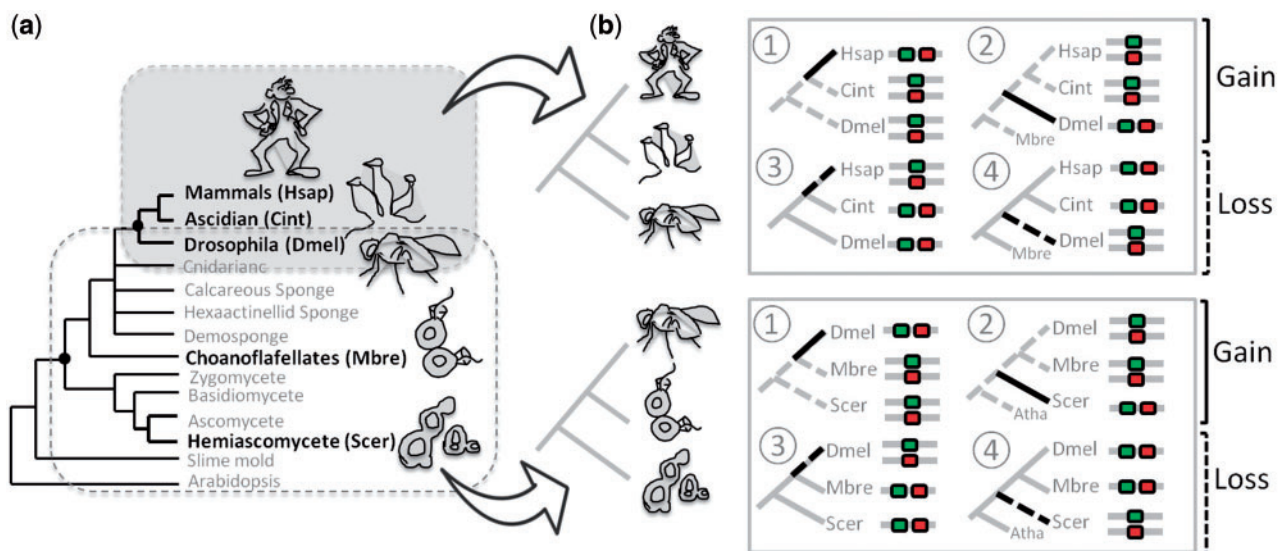
where *a*, *b*, and *c* were the genes and *x<sub>ij</sub>* was the coregulation between two genes *i* and *j*.

The population of reorganized genomes was then subjected to evolutionary selection using Roulette Wheel method, while keeping the population size constant (*n* = 100). The best configuration of genome was drifted directly to the next generation (elitism). After selection step, the population was again subjected to random rearrangement, rewiring of interactions, fitness scoring, selection, elitism, and the whole process was repeated for 3,000 generations. If the coregulated genes tended to cluster linearly, the correlation between linear distance of genes and their coregulatory score would show gradual decrease in correlation over time. All the processed data sets are provided in the [supplementary material, Supplementary Material online](#).

## Results

We exploited the linear and the 3D genome organization data of budding yeast (*Scer*), *Drosophila* (*Dmel*), and human (*Hsap*) for our analyses. To infer the gene order in the hypothetical common ancestor of yeast and *Drosophila*, we obtained the genome assembly of a choanoflagellate, *Monosiga brevicollis* (*Mbre*) (King et al. 2008). Choanoflagellates are strategically well placed between fungi and metazoans in the phylogenetic tree (King et al. 2008) (fig. 1a). Similarly, to infer the common ancestor of human and *Drosophila*, we used the genome of *Ciona intestinalis* (*Cint*), an ascidian that shares many properties of free-living ancestral chordate (Satoh et al. 2014). We then compiled four sets of gene-pairs as illustrated in figure 1b and [supplementary figure S1, Supplementary Material online](#): 1) genes on different chromosomes (“split” organization) in yeast and *Monosiga*, but within 1 Mb on the same chromosome (“clustered” organization) in *Drosophila*; 2) genes split in *Drosophila* and *Monosiga*, but clustered in yeast (<100kb); 3) genes split in *Drosophila*, but clustered in yeast and *Monosiga* (same scaffold); and 4) genes split in yeast, but clustered in *Drosophila* and *Monosiga*. We had the same four scenarios for *Drosophila*–*Ciona*–Human comparison. For scenario 1 and 3 in Yeast–*Monosiga*–*Drosophila* comparison, we inferred that if the gene order (clustered or split) of a candidate gene-pair in *Monosiga* genome was similar to that of yeast, the common ancestral genome of eumetazoan and fungal clades would have had similar gene order as illustrated in figure 1b (scenario 1 and 3). Similarly, in

*Drosophila*–*Ciona*–Human comparison, we inferred that if the gene order of a gene-pair in *Ciona* was similar to that of *Drosophila*, the common ancestor of Deuterostoma and Protostoma would also, most likely, have had similar gene order (fig. 1b). For scenario 2 and 4, inference of gene order in their hypothetical common ancestral genome needed some additional support because *Monosiga* is closer to metazoans than fungi as shown in the phylogenetic tree in the figure 1a. For example, if a candidate gene-pair was split in *Drosophila* and *Monosiga*, but clustered in yeast, we could not directly infer that the gene-pair was split or clustered in the hypothetical common ancestor of metazoans and fungi. However, if the gene-pair was also split in *Arabidopsis thaliana* (*Atha*) in addition to *Drosophila* and *Monosiga*, it could be inferred that the gene-pair was most likely split in the common ancestor of metazoans and fungi. Similar inferences were drawn for scenario 2 and 4 in *Drosophila*–*Ciona*–Human comparison by comparing gene order of human and *Ciona* with that of *Monosiga* genome. Implementing the above strategy, we showed that the 85% of the split gene-pairs of *Monosiga* and *Drosophila* in Yeast–*Monosiga*–*Drosophila* comparison, which could be mapped onto *Arabidopsis*, were split in *Arabidopsis* too, suggesting that the clustered organization in yeast was independently acquired in the fungal clade as shown in the figure 1b. Similarly, 95% of the split gene-pairs of *Ciona* and human in *Drosophila*–*Ciona*–Human comparison, which could be mapped onto *Monosiga*, were split in *Monosiga* too, suggesting that clustered organization in *Drosophila* was independently acquired in the protostomal clade. There was

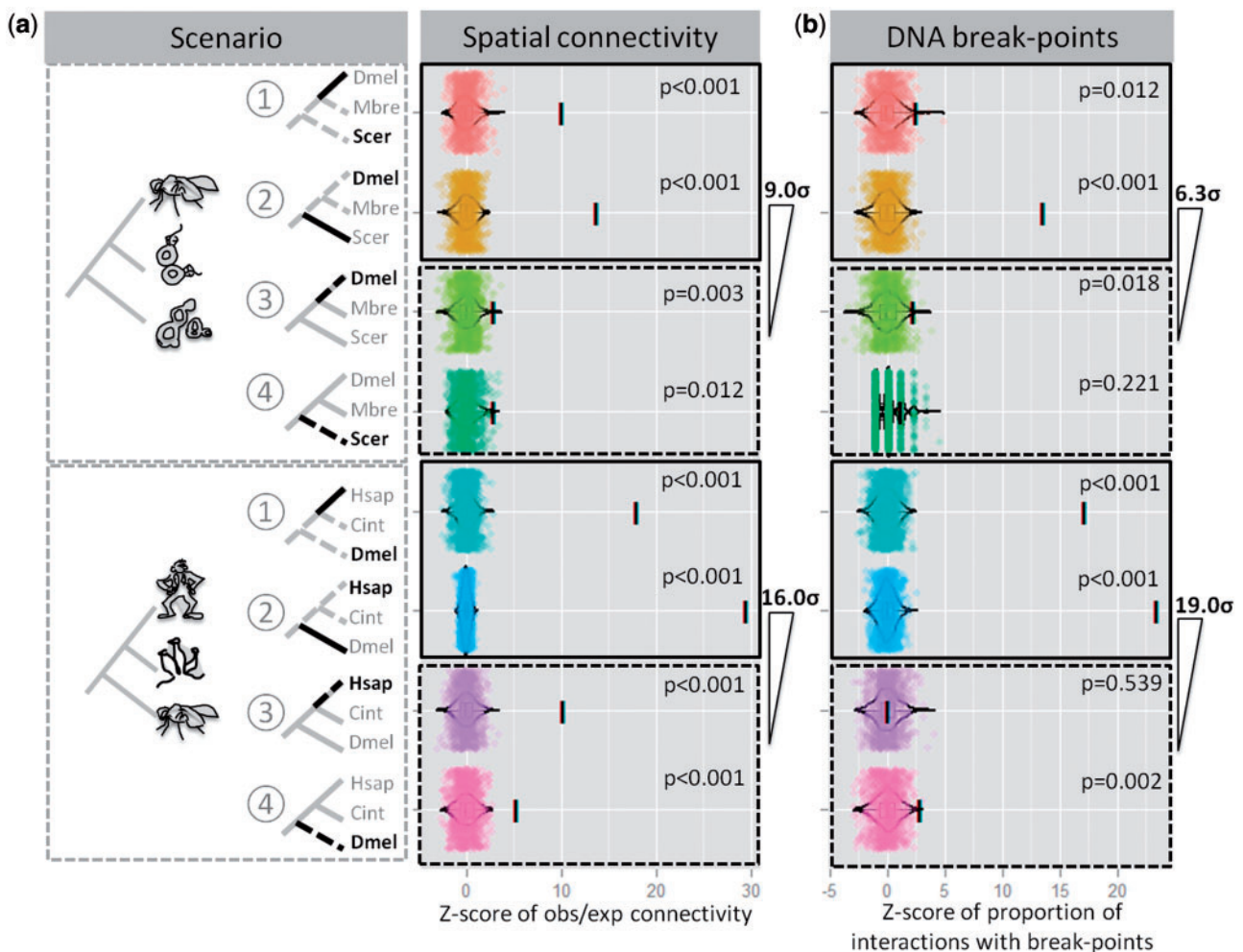


**FIG. 1.**—(a) Relative positioning of yeast, *Monosiga*, *Drosophila*, *Ciona*, and Human in the phylogenetic tree. Dark dots represent the common ancestors of metazoans/fungi and mammals/arthropods. (b) Schematic representation of different scenarios of gene clustering and splitting instances. Smooth and dashed lines represent clustered and split organization of gene-pairs. Black line represents the inferred gain or loss of gene clustering independently in one of the clades. Scenario 1 and 2 represent inferred gain of clustering and scenario 3 and 4 represent inferred loss of gene clustering along one of the clades.

insufficient mapping of gene pairs in scenario 4 to *Arabidopsis* and *Monosiga* genomes for Yeast–*Monosiga*–*Drosophila* and *Drosophila*–*Ciona*–Human comparisons, respectively, and we could not test precisely, for example, whether the clustered gene-pairs (196 only) in Dmel and Mbre genomes were also clustered in *Atha* genome. Nevertheless, by using the above criteria we were mostly able to demarcate the instances where clustering was independently acquired or lost in the analyzed clades.

We assessed the spatial connectivity of genes when they were distant in one of the species in all four scenario shown in the figure 1a. We had two key observations: 1) significant number of the gene-pairs in each comparison exhibited spatial connectivity in their split form (fig. 2a) and 2) the statistical significance (measured as Z-score) of spatial connectivity

of split gene-pairs were significantly greater (average of  $9\sigma$  and  $16\sigma$  increase in Yeast–*Monosiga*–*Drosophila* and *Drosophila*–*Ciona*–Human comparisons, respectively) in case of clustering as compared with splitting of genes, suggesting that significantly fewer number of gene-pairs remained spatially proximal after split as compared with the gene-pairs that got clustered (fig. 2a). These observations led us to hypothesize that the spatially proximal genes impinging from different chromosomes in the ancestral genome might have undergone repositioning through long-range genomic rearrangements like translocations. Since the prerequisite to translocation is DNA breakage event (Roukos and Misteli 2014), we tested if the interacting loci were enriched with DNA breakpoints. A comparison with the null distribution obtained from all the *trans* interactions suggested nonrandomly greater number

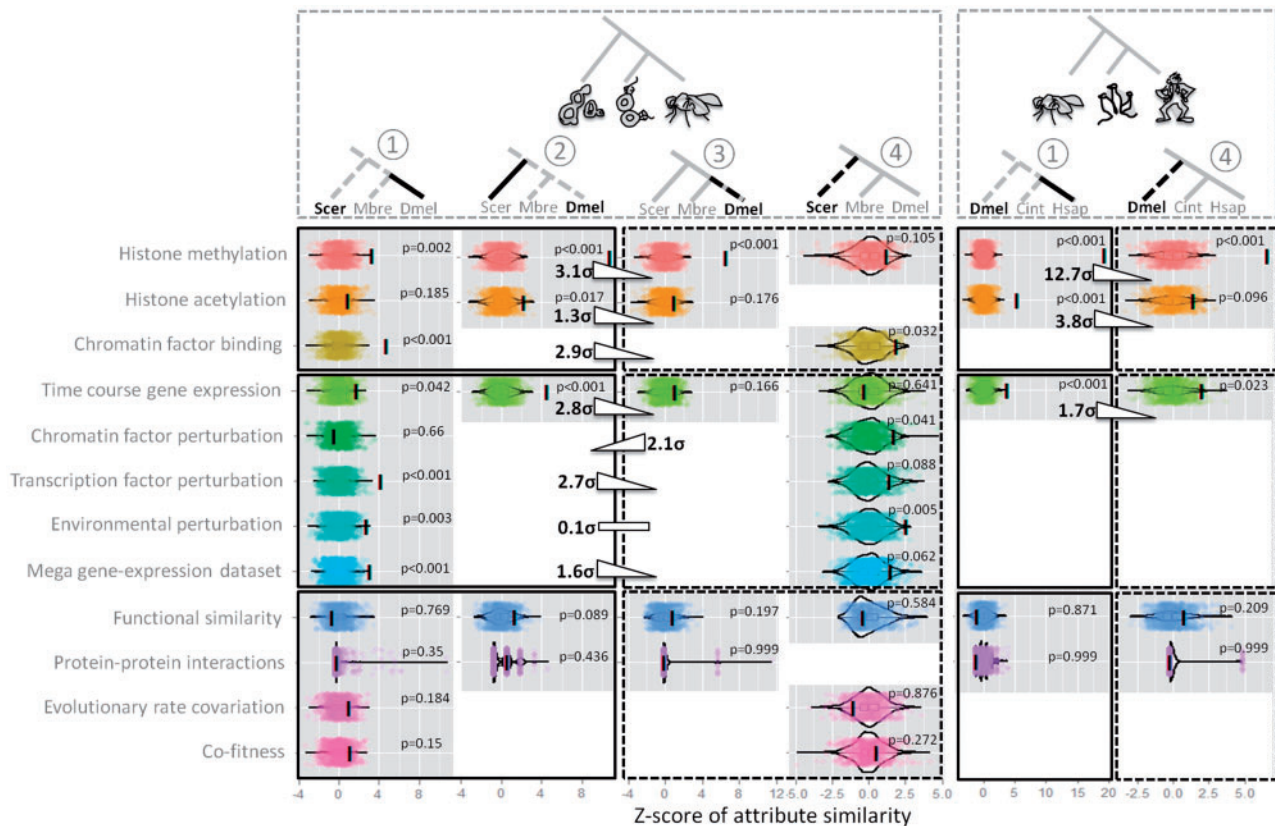


**FIG. 2.**—(a) Spatial connectivity of loci that independently acquired or lost gene clustering along one of clades. The species in which the spatial connectivity was assessed is marked in bold letters in each scenario. Observed values of spatial connectivity is represented by vertical bar overlaid upon the null distributions generated using the strategy given in the Materials and Methods section. (b) Number of interacting pairs of loci with DNA break-points overlaid upon null distribution for the same. Z-score is plotted for the spatial connectivity and enrichment of DNA break-points in order to compare relative values across different scenarios. Change shown in each comparison is the average of Z-score change (1)–(4) and (2)–(3).

of interacting loci with DNA break-points (on both loci) in most comparisons, highlighting the susceptibility of engaged loci to long-range genomic rearrangements (fig. 2b). Again, it was notable that the cases of clustering exhibited greater statistical significance as compared with the cases of splitting (average of  $6.3\sigma$  and  $19\sigma$  increase in Yeast–*Monosiga*–*Drosophila* and *Drosophila*–*Ciona*–Human comparisons, respectively; fig. 2b). These results were suggestive of an evolutionary mechanism that endowed long-range reordering of genes, which might have, in part, guided the evolution of gene-clustering.

What would have constrained the spatial colocalization of the analyzed gene-pairs? To address this, we used genome-wide multidimensional data sets associated with the chromatin states, transcription, and function of genes (Materials and Methods). For each data set, we obtained null distribution by randomly picking pairs of loci having interactions in *trans* and calculating Pearson’s correlation coefficient for those set of gene-pairs (Materials and Methods). As shown in the

figure 3, epigenetic and transcriptional profiles of spatially proximal genes, in general, had significantly greater Pearson’s correlations when compared with the null distributions. However, they did not exhibit functional association as inferred from semantic similarity of gene ontology terms, protein–protein interactions, etc, suggesting that the access to common chromatin/transcription factor foci for the coordinated transcriptional response, not necessarily the functional similarity, of engaged genes might be the common requirement of spatially proximal genes. Further, the splitting instances in the figure 3 showed relatively lower significance (often insignificant) of epigenetic and transcriptional attributes as compared with clustering events. We also confirmed that the genes remained transcriptionally correlated in their linearly clustered organization in gene-clustering instances, but not in gene-splitting instances, suggesting that the coregulation of engaged gene might have served as a constraint preferring the evolutionary selection of linearly clustered organization of rearranged loci, which were spatially proximal in the ancestor



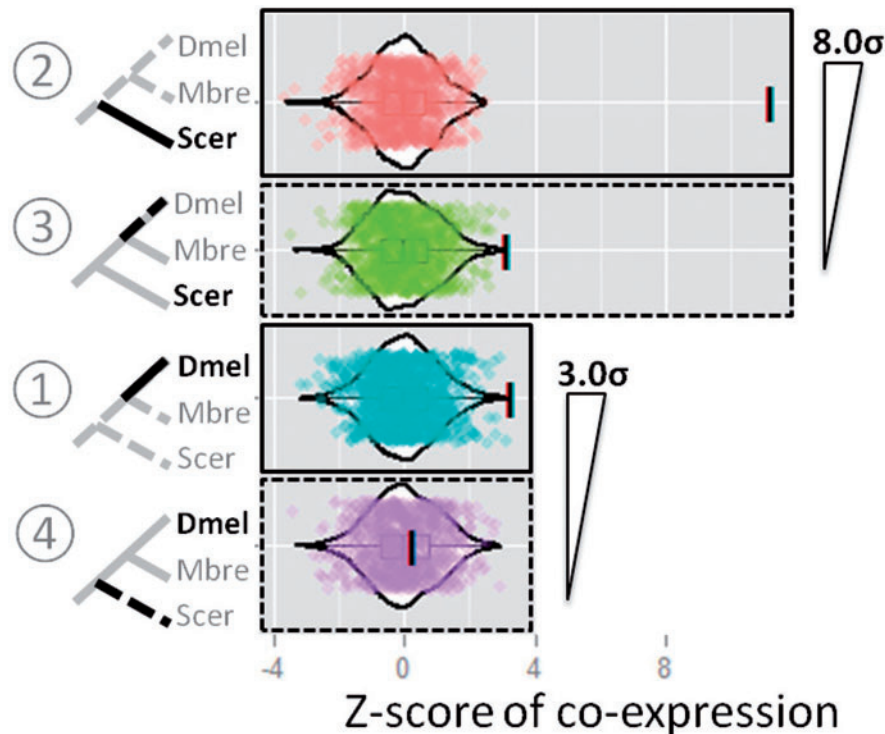
**FIG. 3.**—Epigenetic, transcriptional, and functional similarities of *trans*-interacting genes, which were on different chromosomes (split) in the highlighted (in bold letters) species in each comparison. The observed values of similarities are plotted as vertical bars overlaid upon the null distributions (colored dots superimposed over violin plots). Smooth and dashed lines in the cartoon of the phylogenetic tree represent clustered and split organization of gene-pairs respectively. Black colored smooth and dashed lines in the tree cartoon represent inferred gain and loss of gene-clustering, respectively. Three boxes in each vertical panel represent epigenetic, transcriptional and functional attributes, respectively. Change shown in each comparison is the average of Z-score change (1)–(4) and (2)–(3).

genome (fig. 4). These results hinted at selective evolutionary constraint favoring linear clustering of distant genes and not necessarily splitting of clustered genes.

Our observations suggested that the recurrent events of long-range chromosomal rearrangements at spatially proximal and epigenetically correlated genomic sites might have served as one of the mechanisms that guided the evolution of gene order. We further pressed upon a possibility whether aforementioned mechanism of gene order change could explain the formation of profound gene-clusters in eukaryotes from the ancestors that had relatively less profound clustering of genes. We simulated the evolutionary process computationally, details of which is given in the Materials and Methods section and in [supplementary figure S2, Supplementary Material online](#). Briefly, a population of 100 hypothetical genomes, each having two different chromosomes with equally spaced 50 genes, was subjected to the process of interchromosomal rearrangement (translocation), induced by spatial proximity of engaged loci, in a probabilistic manner. In figure 5, we applied the translocation frequency of 0.1, which was roughly equivalent to the maximal rate of gene order loss in yeast lineage (Fischer et al. 2006). Varying this frequency from 0.01 to 0.2 did not impact the overall

observations except that the convergence took more iterations for lower translocation frequencies ([supplementary fig. S3, Supplementary Material online](#)). The population of reconfigured genomes then underwent a probabilistic selection based on coregulation of spatially proximal genes. This process was iterated over 3,000 times. As shown in the figure 5, we observed a gradual decline in the correlation between genomic distance and the coregulation of neighboring genes, suggesting that if the maximization of coregulation was the evolutionary favored strategy of genome, then the genes tended to cluster linearly through long-range genomic rearrangements. Based on these observations, we hypothesized that our proposed mechanism of gene order change might account for the profound linear gene clustering in eukaryotes. Our results hinted that the convergence from spatial proximity to linear proximity might serve as one of the strategies to maximize the transcriptional coordination among genes, whereas the divergence of linearly clustered genes to distinct chromosomes might only occur for the gene-pairs which were not significantly constrained by their transcriptional coordination as illustrated in our results.

Role of spatial proximity in mutagenic processes in cancer genomes has been proposed earlier (Lin et al. 2009; Mathas

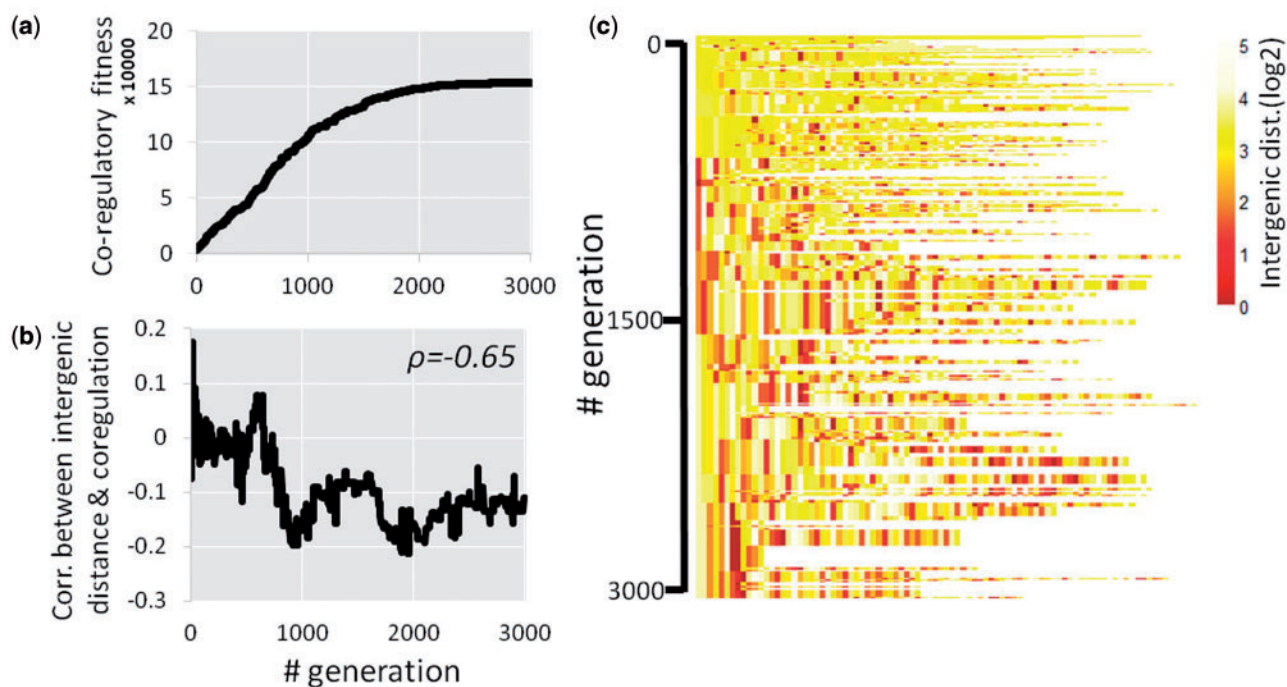


**FIG. 4.**—Average coexpression (correlation of expression profiles) of clustered gene-pairs superimposed over corresponding null distributions for different scenarios. The species in which we assessed the coexpression is highlighted in bold letters. Changes shown on the right side are the changes in Z-scores. The gene expression data for Scer and Dmel was “mega gene expression data set” and “time course embryonic development,” respectively. The interacting gene-pairs for the null distributions are taken from the same chromosome (within 1 Mb for Dmel and within 100 kb for Scer).

et al. 2009; Duan et al. 2010; Veron et al. 2011; Engreitz et al. 2012). Our observations suggested that the same mechanism might have been exploited in the evolution to alter the gene order and select the one that was beneficial to optimize certain genomic function. Given that the statistical significance levels for gene-clustering instances were consistently greater as compared with gene-splitting instances throughout our analyses, it can be inferred that the pronounced gene-clustering in eukaryotic genomes might have been evolved, to some extent, through long-range mechanisms of gene order change, a hypothesis that we simulated using a heuristic model. It is noteworthy that the percentage of instances where we observed spatial proximity among split genes that acquired clustering along one of the clades varied from 7% to 69%, clearly suggesting that the translocation events alone cannot explain all the instances of gene clustering in the evolution, neither we claim so. First of all, the mechanism of repositioning of genes might not necessarily be the translocation event. Long-range inversions can also give similar results for the genes impinging from distant positions on the same chromosome. Important point to be considered here is that inversion would also require physical proximity of distantly located genomic elements. Second, segmental or whole-genome duplications followed by sequential loss of one of the gene-copies serve as potent mediator of gene order change in the eukaryotic genome (Fischer et al. 2001). It is notoriously

difficult to map these events for distant species and appears untestable in our hands at present. Evolution of gene clusters through tandem duplication alone is out of context here because we tested the gene-pairs that were present, either distantly or proximally, in all three species analyzed in each comparison. Moreover, these gene-pairs did not share sequence homology and each gene in the pair belong to distinct gene family as observed through EPGD database.

We further extrapolate that the evolutionary dynamics of linear gene-clustering might have been consequently implicated in radial organization of gene clusters in the nuclear space based on relative gene-densities. The linear gene clusters would result in local attractors or “black-holes” sequestering most of the protein factors important for essential genomic functions like transcription and replication. As a consequence, the distal gene clusters need to be proximal in the nuclear space in order to access those factors and allow the efficient transcription/replication of genes. Therefore, if distinct gene clusters are considered analogous to “planets” and their affinity to bind to shared transcription factors is considered as “gravitational” attraction, the gene-clusters might naturally converge to “galaxy-like” structures, where gene-clusters with high gene density would converge interior of the nucleus, whereas the ones with low gene density would locate toward periphery. Though speculative at present, such



**FIG. 5.**—Results obtained from heuristic model of gene order evolution. (a) Average coregulatory fitness of interacting genes in a population for each generation. (b) Correlation between genomic distance and the coregulation between genes at each generation. (c) Chromosomal heatmaps depicting intergenic distances in a representative chromosome of a population at each generation.

a hypothesis can be tested using dynamical simulations in future.

## Conclusion

In summary, the study reports strong evidence supporting a rather underappreciated mechanism that could have guided the evolution of gene order in eukaryotes. Three dimensional organization of genome predisposes certain interacting loci to long-range genomic rearrangements and the rearranged linearly proximal loci that had correlated chromatin and transcriptional states would have been selected through evolution.

## Supplementary Material

Supplementary material is available at *Genome Biology and Evolution* online (<http://www.gbe.oxfordjournals.org/>).

## Acknowledgments

Authors acknowledge the financial support from Ministry of Human Resource and Development (MHRD), India. M.B. thanks Mr Keerthivasan Raanin Chandradoss for technical help.

## Literature Cited

- Batada NN, Hurst LD. 2007. Evolution of chromosome organization driven by selection for reduced gene expression noise. *Nat Genet.* 39:945–949.
- Berlivet S, et al. 2013. Clustering of tissue-specific sub-TADs accompanies the regulation of HoxA genes in developing limbs. *PLoS Genet.* 9:e1004018.
- Blanco E, et al. 2008. Conserved chromosomal clustering of genes governed by chromatin regulators in *Drosophila*. *Genome Biol.* 9:R134.
- Cohen BA, Mitra RD, Hughes JD, Church GM. 2000. A computational analysis of whole-genome expression data reveals chromosomal domains of gene expression. *Nat Genet.* 26:183–186.
- de Wit E, Braunschweig U, Greil F, Bussemaker HJ, van Steensel B. 2008. Global chromatin domain organization of the *Drosophila* genome. *PLoS Genet.* 4:e1000045.
- Demuth JP, De Bie T, Stajich JE, Cristianini N, Hahn MW. 2006. The evolution of mammalian gene families. *PLoS One* 1:e85.
- Duan Z, et al. 2010. A three-dimensional model of the yeast genome. *Nature* 465:363–367.
- Engreitz JM, Agarwala V, Mirny LA. 2012. Three-dimensional genome architecture influences partner selection for chromosomal translocations in human disease. *PLoS One* 7:e44196.
- Ferrier DE, Holland PW. 2001. Ancient origin of the Hox gene cluster. *Nat Rev Genet.* 2:33–38.
- Filion GJ, et al. 2010. Systematic protein location mapping reveals five principal chromatin types in *Drosophila* cells. *Cell* 143:212–224.
- Fischer G, Neuveglise C, Durrens P, Gaillardin C, Dujon B. 2001. Evolution of gene order in the genomes of two related yeast species. *Genome Res.* 11:2009–2019.
- Fischer G, Rocha EP, Brunet F, Vergassola M, Dujon B. 2006. Highly variable rates of genome rearrangements between hemiascomycetous yeast lineages. *PLoS Genet.* 2:e32.
- Flint J, et al. 2001. Comparative genome analysis delimits a chromosomal domain and identifies key regulatory elements in the alpha globin cluster. *Hum Mol Genet.* 10:371–382.
- Frasch M, Chen X, Lufkin T. 1995. Evolutionary-conserved enhancers direct region-specific expression of the murine Hoxa-1 and Hoxa-2 loci in both mice and *Drosophila*. *Development* 121:957–974.
- Hurst LD, Pal C, Lercher MJ. 2004. The evolutionary dynamics of eukaryotic gene order. *Nat Rev Genet.* 5:299–310.
- King N, et al. 2008. The genome of the choanoflagellate *Monosiga brevicollis* and the origin of metazoans. *Nature* 451:783–788.
- Kleinjan DA, Lettice LA. 2008. Long-range gene control and genetic disease. *Adv Genet.* 61:339–388.
- Kleinjan DA, van Heyningen V. 2005. Long-range control of gene expression: emerging mechanisms and disruption in disease. *Am J Hum Genet.* 76:8–32.
- Kosak ST, Groudine M. 2004. Form follows function: the genomic organization of cellular differentiation. *Genes Dev.* 18:1371–1384.
- Lawrence J. 1999. Selfish operons: the evolutionary impact of gene clustering in prokaryotes and eukaryotes. *Curr Opin Genet Dev.* 9:642–648.
- Lee JM, Sonnhammer EL. 2003. Genomic gene clustering analysis of pathways in eukaryotes. *Genome Res.* 13:875–882.
- Lin C, et al. 2009. Nuclear receptor-induced chromosomal proximity and DNA breaks underlie specific translocations in cancer. *Cell* 139:1069–1083.
- Lomvardas S, et al. 2006. Interchromosomal interactions and olfactory receptor choice. *Cell* 126:403–413.
- Maeso I, Irimia M, Tena JJ, Casares F, Gomez-Skarmeta JL. 2013. Deep conservation of cis-regulatory elements in metazoans. *Philos Trans R Soc Lond B Biol Sci.* 368:20130020.
- Mathas S, et al. 2009. Gene deregulation and spatial genome reorganization near breakpoints prior to formation of translocations in anaplastic large cell lymphoma. *Proc Natl Acad Sci U S A.* 106:5831–5836.
- Miele A, Dekker J. 2009. Mapping cis- and trans- chromatin interaction networks using chromosome conformation capture (3C). *Methods Mol Biol.* 464:105–121.
- Ohlsson R, Paldi A, Graves JA. 2001. Did genomic imprinting and X chromosome inactivation arise from stochastic expression? *Trends Genet.* 17:136–141.
- Roukos V, Misteli T. 2014. The biogenesis of chromosome translocations. *Nat Cell Biol.* 16:293–300.
- Sandhu KS, et al. 2009. Nonallelic transvection of multiple imprinted loci is organized by the H19 imprinting control region during germline development. *Genes Dev.* 23:2598–2603.
- Satoh N, Rokhsar D, Nishikawa T. 2014. Chordate evolution and the three-phylum system. *Proc Biol Sci.* 281:20141729.
- Spellman PT, Rubin GM. 2002. Evidence for large domains of similarly expressed genes in the *Drosophila* genome. *J Biol.* 1:5.
- Spitz F, Duboule D. 2008. Global control regions and regulatory landscapes in vertebrate development and evolution. *Adv Genet.* 61:175–205.
- Splinter E, et al. 2006. CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. *Genes Dev.* 20:2349–2354.
- Teichmann SA, Veitia RA. 2004. Genes encoding subunits of stable complexes are clustered on the yeast chromosomes: an interpretation from a dosage balance perspective. *Genetics* 167:2121–2125.
- Tena JJ, et al. 2011. An evolutionarily conserved three-dimensional structure in the vertebrate lrx clusters facilitates enhancer sharing and coregulation. *Nat Commun.* 2:310.
- van Heyningen V, Bickmore W. 2013. Regulation from a distance: long-range control of gene expression in development and disease. *Philos Trans R Soc Lond B Biol Sci.* 368:20120372.



Vernimmen D, et al. 2009. Chromosome looping at the human alpha-globin locus is mediated via the major upstream regulatory element (HS -40). *Blood* 114:4253–4260.

Veron AS, Lemaitre C, Gautier C, Lacroix V, Sagot MF. 2011. Close 3D proximity of evolutionary breakpoints argues for the notion of spatial synteny. *BMC Genomics* 12:303.

Yeaman S. 2013. Genomic rearrangements and the evolution of clusters of locally adaptive loci. *Proc Natl Acad Sci U S A*. 110:E1743–E1751.

**Associate editor:** Partha Majumder