

10 Evolution of Noncoding DNA in Eukaryotes

Josep M. Comeron
The University of Iowa, Iowa City, IA

- 1 **The C-value Paradox 326**
- 2 **Noncoding Genome Size: Passive versus Selective Trait 327**
- 3 **Deletion Bias (DB) Based on Small Indels 328**
 - 3.1 DB in *Drosophila* 328
 - 3.2 What is the True DB in *Drosophila*? 328
 - 3.3 DB in Other Eukaryotes 329
 - 3.4 Overall Rate of DNA Loss due to Small Indels and Genome Size 329
- 4 **Transposable Elements (TE) and DNA Gain 330**
- 5 **Variation across and among Genomes: The Influence of
Recombination 330**
- 6 **Noncoding DNA is not Free of Genetic Information or Evolutionary
Effect 331**
- 7 **Summary 332**
- Acknowledgment 333**
- Bibliography 333**
 - Book and Reviews 333
 - Primary Literature 333

Keywords

Deletion Bias

The study of small insertions and deletions (indels) in most eukaryotes shows an excess of deletions over insertions, forecasting DNA loss.

Effective Population Size (N_e)

A theoretical number of individuals that would explain the amount of genetic variation observed within a species in the absence of selection. N_e is somewhat related to, but always smaller than, the actual number of individuals. N_e is positively associated with the effectiveness of selection and varies among species as well as across genomes.

The amount of haploid nuclear DNA (or C-value) among eukaryotes varies more than 4 orders of magnitude, with brewer's yeast *Saccharomyces cerevisiae* (1.2×10^7 bp) or the parasitic microsporidium *Encephalitozoon intestinalis* (3×10^6 bp) as prototypical examples of the smallest genomes and several amoebae mentioned as the eukaryotes with largest genomes with $>6 \times 10^{11}$ bp (e.g. *Amoeba dubia* and *Chaos chaos*). Differences in the amount of noncoding DNA are also observed within a given taxonomic group, with algae varying more than 5000-fold, more than 1000-fold among flowering plants or invertebrates, more than 300-fold among vertebrates, and 100-fold among amphibians and insects. An apparent "paradox" arises when the size of the nuclear genome is compared to organismal complexity (morphological or developmental) if we assume that the amount of genetic material (genome size) should somehow correspond to the amount of genetic information. The so-called C-value paradox or enigma was coined to describe this lack of correspondence. Later findings revealed that coding sequences account for only a small proportion of the genomic DNA in most eukaryotes. This observation does not solve the C-value paradox, but shifts the debate from the number of genes responsible for a given degree of complexity (the G-value paradox) to the amount of noncoding DNA and the causes for its tremendous variation. In fact, recent studies based on fully sequenced genomes and expression data in model eukaryotes, including humans, have shown that complex gene regulation and the number of different transcripts – with a high fraction of genes having multiple transcripts – and not the number of genes, would better represent biological "complexity," hence explaining the G-value paradox.

1

The C-value Paradox

Many cause/consequence relationships have been proposed on the basis of phenotypic correlations. The variation in

noncoding genome size among eukaryotes has been associated with biological factors, both cellular and organismal: cell and nucleus size, cell division rate, metabolic rate, developmental rate, developmental complexity, and overall gene expression.

Indeed, eukaryotes with large genomes tend to have larger nucleus and cells, and species with high developmental rate to developmental complexity ratio tend to have reduced genomes; other associations are only observed in particular lineages.

However, an aspect not always considered is that many biological factors are associated with species or population parameters such as generation time, population size, and effective population size (N_e). This indirect association between cellular or developmental factors and N_e is pertinent in the analysis of any genomic characteristic (or any biological feature for that matter) because N_e is a main parameter influencing natural selection. Population genetic theory predicts and evolutionary analyses confirm that N_e is positively associated with the effectiveness of selection. (Increased effectiveness of selection will favor the fixation of advantageous mutations and the removal of deleterious mutations.) As a consequence, it is not direct to distinguish between cause, effect, and indirect or coincidental association; many proposed causal associations illustrate this difficulty. For instance, a causal association between fast developmental rate (and short generation time) and small genomes can be proposed. However, many species with short generation time also have large population sizes and N_e , and therefore the observed small genome is likely the result of more efficient selective constraints favoring small genomes while the negative relationship between developmental rate and genome size would be secondary.

Moreover, several cases of closely related species – with equivalent cellular and developmental parameters – exemplify that the size of noncoding DNA can change very rapidly in evolutionary terms independent of biological factors. For instance,

closely related *Drosophila* species show severalfold differences in noncoding DNA and there is significant variation even at intraspecific level. Another piece of information comes from genomic analyses that have shown that several features of noncoding DNA such as the amount of intergenic DNA (or gene density), and intron presence and size also show a heterogeneous distribution across genomes. Thus, an integral explanation to the presence and size of noncoding DNA should be able to explain differences between distant groups or taxa, differences between closely related species with equivalent cellular/developmental features, as well as differences across a given genome.

2

Noncoding Genome Size: Passive versus Selective Trait

One possible explanation to the size of noncoding DNA and genome size in eukaryotes is that the amount of noncoding DNA is the passive result of mutational tendencies only; the manifestation of a mutation-drift equilibrium between insertions and deletions. Under this mutational-only scenario, variation in genome size is explained by differences in the patterns of insertion and deletion of “selfish” DNA or due to variation in the intrinsic rates of DNA loss through small insertions and deletions (indels). Special attention has been paid recently to the latter process, with several eukaryotes showing an excess of small deletions over insertions (the so-called deletion bias; DB). In particular, it has been proposed that differences in the rate of DNA loss caused by small indels are a major factor influencing genome size.

Under a selective scenario, genome size represents the equilibrium of mutational

tendencies to either increase or decrease the amount of DNA counterbalanced by selective tendencies: a mutational selection–drift equilibrium. Mutational tendencies can certainly vary among species, with differences in the amount of repetitive DNA and its tendency to expand in the genome (mostly associated with transposable elements; TE), as well as variation in DB. It follows, however, that a mutational-only hypothesis – which assumes that the presence of noncoding DNA has no fitness costs – would predict either the collapse of noncoding DNA or the unlimited expansion of noncoding DNA. A more likely proposal assumes that both mutational tendencies (DNA expansion by TE amplification and DNA loss by DB) are balanced as a result of selective constraints.

The selective causes acting on genome size can be varied and be particular to groups or taxa according to specific cellular or developmental requirements (e.g. fast cell division), targeting replication costs, or biological effects of bulk DNA. As indicated earlier, the effectiveness of selection acting on noncoding DNA would also depend on parameters such as N_e and recombination (see the following) and so the outcome of the equilibrium would be species specific. Moreover, both N_e and recombination vary across genomes. Thus, a mutation selection–drift equilibrium is expected to be highly variable among taxa, between closely related species, and also vary across a given genome.

3

Deletion Bias (DB) Based on Small Indels

3.1

DB in *Drosophila*

The struggle to determine the mutational DB on the basis of small indels is best

exemplified in *Drosophila*, in which three different approaches give three different results. The first significant attempt to investigate DB in *Drosophila*, studied nonallelic copies of non long–terminal repeat retrotransposable elements, so-called “dead-on-arrival” (DOA) elements. The study of *Helena* DOA elements gives a DB of 8.7, suggesting a very high rate of DNA loss. The second approach investigated defective TEs, generating a DB of 3.6. The third approach is the study of indels at polymorphic level in DNA sequences likely free of selective constraints (intergenic regions and introns), with a DB that ranges between 1.4 and 1.9, evidencing a significant but moderate rate of DNA loss.

3.2

What is the True DB in *Drosophila*?

As indicated by B. Charlesworth, the study of nonallelic DOA elements (that are either fixed or segregate at high frequencies within a species) could bias the estimates of DB if selection favors deletions in recently inserted elements, hence causing an overestimation of DB. The rationale for arguing selection against inserted repetitive elements (favoring deletions and shorter versions of the elements) is based on the notion that shorter elements would reduce the probability of ectopic recombination and its deleterious effects. On the other hand, population genetic theory predicts that polymorphic studies will estimate mutational parameters that can only be closer to the true values compared to analyses that use mutations that are either fixed or segregate at high frequencies within a species. If indels are neutral, all approaches should give equivalent results, but if indels are under weak/moderate

selection, the study of indels at polymorphic level is least influenced by selection. Strong selective constraints on the sequences used to investigate polymorphic indels, however, could also reduce the number of deletions observed at polymorphic level, hence causing underestimates in DB.

Several observations suggest that the estimates of DB based on different polymorphic studies is close to the true DB. First, an *ad hoc* theoretical model presuming that long deletions in introns are under strong selection would forecast a lower DB, but even in this case it cannot explain the observed data. Second, the study of polymorphic indels in intergenic regions also produces a very low DB (1.7) and, therefore, the argument that the observed DB is biased because of selection is not likely to be correct. And third, a recent study on the frequency of several families of TEs in populations has generated a key result to understand the observed discrepancy in DB based on long repetitive elements. This study shows that there is a negative correlation between the length of TEs within a family and the frequency in the population, in agreement with the ectopic recombination model that claims that deletions would be selectively favored in long repetitive elements. The consequence of this observation is that estimates of DB based on long repetitive elements will be biased, showing an excess of deletions than those expected by mutational tendencies only.

3.3

DB in Other Eukaryotes

Measures of DB have been obtained in other eukaryotes also on the basis of different approaches. One of the first studies to detect an excess of small deletions over

insertions was applied to mammals by analyzing processed pseudogenes. Large-scale analyses of pseudogenes in humans and murids show an average DB of 2.74. The study of pseudogenes in the insect *Podisma pedestris* and in *Caenorhabditis elegans* shows DB of 2.7 and 3.8, respectively. The study of DOA *Lau1* elements in the Hawaiian crickets of the genus *Laupala* and *Maui* elements in two pufferfish species suggests a DB of 2.7 and 1.3–2.0, respectively.

3.4

Overall Rate of DNA Loss due to Small Indels and Genome Size

The rate of DNA loss depends on DB as well as the size of indels, and in most species deletions are longer than insertions. In order of increasing genome size, the overall rate of DNA loss observed in the different eukaryotes is the following: 1.8 in *Drosophila melanogaster* (based on polymorphic indels in intergenic regions and introns), 9 to 15 in Pufferfish, 3.8 in *Laupala*, 2.6 in rat/mouse, 1.7 in humans, and 3.6 in *Podisma*. Conversely, *C. elegans* shows a net gain of DNA based on small indels. As T.R. Gregory concluded recently, the use of updated estimates of DNA loss for mammals and *Drosophila* raise considerable doubts about the strength of any correlation between DNA loss rate by small indel bias and genome size. Therefore, although the mutational rate of DNA loss caused by small indels can vary evolutionarily, and might itself be the result of selection, it cannot explain the observed differences in noncoding DNA content and genome size between species. Moreover, this mutational explanation cannot account also for the observed differences across genomes.

4

Transposable Elements (TE) and DNA Gain

TE elements are widespread in eukaryotic genomes and comprise in many cases a substantial portion of euchromatic genome size, including humans where TEs explain 45% of intergenic and intronic sequences. The recent invasion/amplification of particular TE families causes rapid increases in genome size, temporarily altering any specific equilibrium among selective and long-term mutational tendencies. Two examples of such effect are maize, in which genome size has recently doubled due to TE invasion, and *Drosophila*, with substantial differences in TE presence between closely related species, as well as among populations of the same species. The opposite trend is also detected, with species that have recently lost repetitive DNA showing a sudden reduction in genome size albeit having similar mutational tendencies associated with small indels. For instance, tetraodontid (smooth) and diodontid (spiny) pufferfish have similar DB but smooth pufferfish shows a twofold reduction in genome size because of a recent reduction in repetitive elements. This latter case also illustrates that any study of mutational tendencies toward DNA gain or loss should include small indels (i.e. DB), as well as the rate of large insertions and TE amplification.

It is possible to argue that a generic mutational tendency of TE to replicate and increase genome size might have favored a mutational mechanism associated with DNA repair of small indels with a tendency toward DNA loss, as observed in most eukaryotes. In this regard, small differences in DB among taxa would reflect long-term tendencies associated with TE amplification in the genome.

The additional influence of selective forces that can be species specific and/or vary across genomes would generate the final outcome.

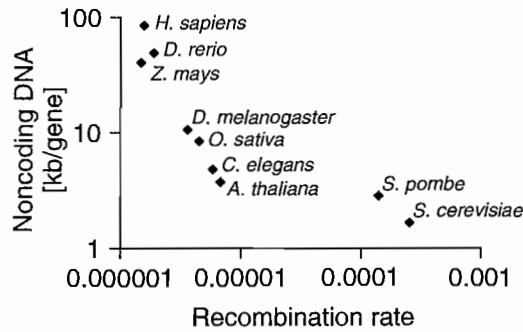
5

Variation across and among Genomes: The Influence of Recombination

Many genomic analyses have revealed heterogeneity of noncoding DNA presence across eukaryotic genomes. In *Drosophila*, the amount of intergenic DNA is higher in genomic regions with low recombination (and crossing-over) rates. Introns are also longer and more frequent in genes located in regions with severely reduced rates of crossing-over. An equivalent trend is observed in humans, with longer intergenic sequences, and introns in isochores are associated with reduced rates of crossing-over. *S. cerevisiae* also shows an influence of recombination on the presence of noncoding DNA. A single exception to this tendency is observed in *C. elegans*, in which an atypical pattern of crossing-over along chromosomes has been reported.

Interestingly, a negative relationship between recombination rates and the amount of noncoding DNA is observed among species. The study of nine distant eukaryotes with well-characterized genetic and physical maps (Fig. 1) reveals a significant negative relationship between recombination rate and a measure of the amount of noncoding DNA once the number of genes is taken into account. (Albeit the study is based on a small number of species, the observed relationship is not caused by any single species.) This observation suggests that similar, or at least overlapping, forces are shaping the size of noncoding DNA among species and

Fig. 1 Relationship between recombination rate and gene density among evolutionary distant model eukaryotes. The nine eukaryotes with well-characterized genetic and physical maps are: *Homo sapiens*, *Danio rerio* (zebra fish), *Zea mays* (maize), *Oryza sativa* (rice), *D. melanogaster*, *C. elegans*, *Arabidopsis thaliana*, *S. cerevisiae*, and *Schizosaccharomyces pombe* (fission yeast). Recombination rate indicates the average percentage of meiotic recombination (cM) between adjacent base pairs (total genetic map/total physical map), taking also into account the number of chromosomes to appraise the recombination caused by the random segregation of chromosomes. Although *A. thaliana* is a predominantly selfing plant, this feature has only recently been acquired, with all closely related species being outcrossing and, therefore, the recombination rate used for *A. thaliana* reflects the recent ancestral outcrossing condition. Nonparametric Spearman's rank correlation $R = -0.95$ ($p = 0.00009$). The least significant relationship after removing any one species is $R = -0.929$ ($p = 0.0009$). The removal of both distantly related yeasts (*S. cerevisiae* and *S. pombe*) from the analysis does not influence the outcome either ($R = -0.893$, $p = 0.007$).



across genomes. These equivalent findings across and among genomes could be explained if the mechanism of mismatch repair associated with recombination instigates DNA loss. Although this possibility needs to be investigated thoroughly in different species, in *D. melanogaster* there are similar patterns of indels in regions with normal and reduced rates of crossing-over.

An alternative explanation, invoking natural selection, emerges when we consider the influence that recombination has on the effectiveness of selection. Indeed, theoretical and experimental investigations predict that recombination will have an impact on the power or effectiveness of selection. In particular, most models of selection forecast that recombination will raise the effectiveness of selection. The total length of genetic map in a species as well as the number of chromosomes will play a role in the total meiotic recombination rate. Recombination rates and N_e (mentioned earlier) are two parameters that have been usually limited to population genetics studies but they should be included in any general explanation to

genome architecture, the amount of noncoding DNA and genome size due to their influence to the effectiveness of selection. Moreover, both recombination and N_e are influenced by species-specific factors as well as vary across genomes.

6 Noncoding DNA is not Free of Genetic Information or Evolutionary Effect

Noncoding DNA in eukaryotes has been often labeled as “junk” DNA, useless (with no genetic information) as well as with no evolutionary effect (neutral to selection). In fact, however, noncoding DNA can be evolutionarily relevant by containing genetic information as well by its only presence. The number of known noncoding genes and especially noncoding RNA genes (ncRNA) has recently increased rapidly. Many of these ncRNA function directly as RNA while others play an essential role in gene regulation, with antisense RNAs being an important class of these genes. Moreover, noncoding DNA,

intergenic or intronic, harbors most of the information on gene regulation.

On the other hand, the presence of DNA with no genetic information between genes or between exons should not be assumed to be devoid of evolutionary effect. For instance, the presence of introns and their length can have a direct fitness effect on transcription costs and be a trait under selection. However, the presence of DNA sequences with no direct fitness effects can also be a selective trait. Note that the significant parameter associated with recombination is the total recombination between genes or exons (i.e. clusters of sites under selection) and this is influenced by the rate per physical unit as well as by the physical distance. Therefore, under realistic conditions for most eukaryotes, an organism might increase the effectiveness of selection by increasing the amount of noncoding DNA, intergenic or intronic. Under this scenario, the presence of “junk” DNA can be viewed as a modifier of recombination with a definite evolutionary long-term effect and can itself be a selective trait. The long-term fitness effects of increasing the physical distance between clusters of sites under selection (genes, exons, or regulatory regions) and recombination, is expected to differ among species (according to the overall recombination rate) and across genomes (according to variation in recombination rates along chromosomes).

7

Summary

The study of noncoding DNA is no longer restricted to phenotypic correlations and all recent observations support the notion that noncoding genome size in eukaryotes

is the result of a highly dynamic balance between mutational and selective forces. Population parameters that influence levels and patterns of polymorphisms also influence the effectiveness of selection and have the potential to influence genomic features. Therefore, the evolution of noncoding DNA (intergenic and intron presence, and size) ought to be investigated from a population genetics perspective (i.e. the so-called *population genomics* approach). The presence of regulatory sequences and noncoding genes (e.g. ncRNA genes) would add genetic information required for organismal “complexity,” but it can explain only a fraction of noncoding DNA.

Mutational tendencies are strongly influenced by TE invasion and amplification, which can be species specific and cause sudden changes in genome size. This “one-way ticket to genome obesity” is somewhat counteracted by a moderate tendency toward DNA loss due to small indels observed in most eukaryotes. The widespread presence of TEs in eukaryotes might have caused a selective pressure on mechanisms of DNA repair favoring DNA loss. The study of small indels only cannot produce an accurate picture of mutational tendencies. On the other hand, selective forces on the presence of noncoding DNA can vary in different species and/or taxa and be associated with replication costs, nuclear architecture, or other biological effects of bulk DNA. Importantly, the actual impact of these selective tendencies is determined by recombination rates and N_e parameters that can vary across genomes, as well as among species and should be included in any general explanation to genome evolution. Future studies on variation of recombination rates, indel evolution, and N_e in many eukaryotes will give us a broader and more

precise picture on the selective and mutational forces influencing the evolution of noncoding DNA.

Acknowledgment

The author thanks Dr. A. Llopart for her critical review of the manuscript. Josep M. Comeron is funded by National Science Foundation (NSF), DEB03-44209.

Bibliography

Book and Reviews

- Cavalier-Smith, T. (1982) Skeletal DNA and the evolution of genome size, *Annu. Rev. Biophys. Bioeng.* **11**, 273–302.
- Cavalier-Smith, T. (1985) *The Evolution of Genome Size*, John Wiley & Sons, New York.
- Comeron, J.M. (2001) What controls the length of noncoding DNA? *Curr. Opin. Genet. Dev.* **11**, 652–659.
- Gregory, T.R. (2002) Genome size and developmental complexity, *Genetica* **115**, 131–146.
- Hartl, D.L. (2000) Molecular melodies in high and low C, *Nat. Rev. Genet.* **1**, 145–149.
- Ohno, S. (1972) So much 'junk' DNA in Our Genome, in: Smith, H.H. (Ed.) *Evolution of Genetic Systems*, Gordon and Breach, New York, pp. 366–370.

Primary Literature

- Awadalla, P. (2003) The evolutionary genomics of pathogen recombination, *Nat. Rev. Genet.* **4**, 50–60.
- Bennett, M.D. (1971) The duration of meiosis, *Proc. R. Soc. Lond., B, Biol. Sci.* **178**, 259–275.
- Bensasson, D., Petrov, D.A., Zhang, D.X., Hartl, D.L., Hewitt, G.M. (2001) Genomic gigantism: DNA loss is slow in mountain grasshoppers, *Mol. Biol. Evol.* **18**, 246–253.
- Biemont, C. (1999) Distribution of transposable elements in *Drosophila* species, *Genetica* **105**, 43–62.

- Blumenstiel, J.P., Hartl, D.L., Lozovsky, E.R. (2002) Patterns of insertion and deletion in contrasting chromatin domains, *Mol. Biol. Evol.* **19**, 2211–2225.
- Carvalho, A.B., Clark, A.G. (1999) Intron size and natural selection, *Nature* **401**, 344.
- Castillo-Davis, C.I., Mekhedov, S.L., Hartl, D.L., Koonin, E.V., Kondrashov, F.A. (2002) Selection for short introns in highly expressed genes, *Nat. Genet.* **31**, 415–418.
- Cavalier-Smith, T. (1978) Nuclear volume control by nucleoskeletal DNA, selection for cell volume and cell growth rate, and the solution of the DNA C-value paradox, *J. Cell Sci.* **34**, 247–278.
- Charlesworth, B. (1996) The changing sizes of genes, *Nature* **384**, 315–316.
- Claverie, J.M. (2000) What if there are only 30,000 human genes? *Science* **291**, 1255–1257.
- Comeron, J.M. (2004) Selective and mutational patterns associated with gene expression in humans: influences on synonymous composition and intron presence, *Genetics* **167**, 1293–1304.
- Comeron, J.M., Kreitman, M. (2000) The correlation between intron length and recombination in *Drosophila*. Dynamic equilibrium between mutational and selective forces, *Genetics* **156**, 1175–1190.
- Comeron, J.M., Kreitman, M. (2002) Population, evolutionary and genomic consequences of interference selection, *Genetics* **161**, 389–410.
- Comeron, J.M., Kreitman, M., Aguade, M. (1999) Natural selection on synonymous sites is correlated with gene length and recombination in *Drosophila*, *Genetics* **151**, 239–249.
- Crow, J.F., Kimura, K. (1970) *An Introduction to Population Genetics Theory*, Harper and Row, New York.
- Day, T., Gloor, G.B. (1997) Homology requirements for targeting heterologous sequences during P-induced gap repair in *Drosophila melanogaster*, *Genetics* **147**, 689–699.
- Devos, K.M., Brown, J.K.M., Bennetzen, J.L. (2002) Genome size reduction through illegitimate recombination counteracts genome expansion in *Arabidopsis*, *Genome Res.* **12**, 1075–1079.
- Doolittle, W.F., Sapienza, C. (1980) Selfish genes, the phenotype paradigm and genome evolution, *Nature* **284**, 601–603.

- Duret, L., Mouchiroud, D., Gautier, C. (1995) Statistical analysis of vertebrate sequences reveals that long genes are scarce in GC-rich isochores, *J. Mol. Evol.* **40**, 308–317.
- Evgen'ev, M., Zelentsova, H., Mnjoian, L., Loluectova, H., Kidwell, M.G. (2000) Invasion of *Drosophila virilis* by the Penelope transposable element, *Chromosoma* **109**, 350–357.
- Graur, D., Shuali, Y., Li, W.-H. (1989) Deletions in processed pseudogenes accumulate faster in rodents than in humans, *J. Mol. Evol.* **28**, 279–285.
- Gregory, T.R. (2004) Insertion-deletion biases and the evolution of genome size, *Gene* **324**, 15–34.
- Hey, J. (1998) Selfish genes, pleiotropy and the origin of recombination, *Genetics* **149**, 2089–2097.
- Hey, J., Kliman, R.M. (2002) Interactions between natural selection, recombination and gene density in the genes of *Drosophila*, *Genetics* **160**, 595–608.
- Hill, W.G., Robertson, A. (1966) The effect of linkage on limits to artificial selection, *Genet. Res.* **8**, 269–294.
- Kliman, R.M., Irving, N., Santiago, M. (2003) Selection conflicts, gene expression, and codon usage trends in yeast, *J. Mol. Evol.* **57**, 98–109.
- Lander, E.S., Linton, L.M., Birren, B., Nussbaum, C., Zody, M.C., et al. (2001) Initial sequencing and analysis of the human genome, *Nature* **409**, 860–921.
- Mirsky, A.E., Ris, H. (1951) The desoxyribonucleic acid content of animal cells and its evolutionary significance, *J. Gen. Physiol.* **34**, 451–462.
- Neafsey, D.E., Palumbi, S.R. (2003) Genome size evolution in pufferfish: a comparative analysis of diodontid and tetraodontid pufferfish genomes, *Genome Res.* **13**, 821–830.
- Parsch, J. (2003) Selective constraints on intron evolution in *Drosophila*, *Genetics* **165**, 1843–1851.
- Petrov, D.A. (2002a) DNA loss and evolution of genome size in *Drosophila*, *Genetica* **115**, 81–91.
- Petrov, D.A. (2002b) Mutational equilibrium model of genome size evolution, *Theor. Popul. Biol.* **61**, 531–544.
- Petrov, D.A., Hartl, D.L. (1998) High rate of DNA loss in the *Drosophila melanogaster* and *Drosophila virilis* species groups, *Mol. Biol. Evol.* **15**, 293–302.
- Petrov, D.A., Hartl, D.L. (2000) Pseudogene evolution and natural selection for a compact genome, *J. Hered.* **91**, 221–227.
- Petrov, D.A., Lozovskaya, E.R., Hartl, D.L. (1996) High intrinsic rate of DNA loss in *Drosophila*, *Nature* **384**, 346–349.
- Petrov, D.A., Chao, Y.C., Stephenson, E.C., Hartl, D.L. (1998) Pseudogene evolution in *Drosophila* suggests a high rate of DNA loss, *Mol. Biol. Evol.* **15**, 1562–1567.
- Petrov, D.A., Sangster, T.A., Johnston, J.S., Hartl, D.L., Shaw, K.L. (2000) Evidence for DNA loss as a determinant of genome size, *Science* **287**, 1060–1062.
- Petrov, D.A., Aminetzach, Y.T., Davis, J.C., Bensasson, D., Hirsh, A.E. (2003) Size matters: non-LTR retrotransposable elements and ectopic recombination in *Drosophila*, *Mol. Biol. Evol.* **20**, 880–892.
- Powell, J.R. (1997) *Progress and Prospects in Evolutionary Biology. The Drosophila Model*, Oxford University Press, New York.
- Ptak, S.E., Petrov, D.A. (2002) How intron splicing affects the deletion and insertion profile in *Drosophila melanogaster*, *Genetics* **162**, 1233–1244.
- Robertson, H.M. (2000) The large *srh* family of chemoreceptor genes in *Caenorhabditis* nematodes reveals processes of genome evolution involving large duplications and deletions and intron gains and losses, *Genome Res.* **10**, 192–203.
- San Miguel, P., Gaut, B.S., Tikhonov, A., Nakajima, Y., Bennetzen, J.L. (1998) The paleontology of intergene retrotransposons of maize, *Nat. Genet.* **20**, 43–45.
- Schaeffer, S.W. (2002) Molecular population genetics of sequence length diversity in the *Adh* region of *Drosophila pseudoobscura*, *Genet. Res.* **80**, 163–175.
- Smit, A.F. (1999) Interspersed repeats and other mementos of transposable elements in mammalian genomes, *Curr. Opin. Genet. Dev.* **9**, 657–663.
- Thomas, C.A. (1971) The genetic organization of chromosomes, *Annu. Rev. Genet.* **5**, 237–256.
- Vinogradov, A.E. (1995) Nucleotypic effect in homeotherms: body-mass-corrected basal metabolic rate of mammals is related to genome size, *Evolution* **49**, 1249–1259.
- Wright, S. (1931) Evolution in Mendelian populations, *Genetics* **16**, 97–159.

Zuckermandl, E. (1976) Gene control in eukaryotes and the c-value paradox "excess" DNA as an impediment to transcription of coding sequences, *J. Mol. Evol.* **9**, 73–104.

Zuckermandl, E. (2002) Why so many noncoding nucleotides? The eukaryote genome as an epigenetic machine, *Genetica* **115**, 105–129.