



# Gene expression: sizing it all up

Jenna Lynn Woody<sup>1</sup> and Randy C. Shoemaker<sup>2\*</sup>

<sup>1</sup> Interdepartmental Genetics, Iowa State University, Ames, IA, USA

<sup>2</sup> Corn Insects and Crop Genetics Research Unit, Agricultural Research Service, United States Department of Agriculture, Ames, IA, USA

## Edited by:

Jun Yu, Beijing Institute of Genomics, China

## Reviewed by:

Barbara Hohn, Friedrich Miescher Institute for Biomedical Research, Switzerland

Paula Casati, Centro de Estudios Fotosintéticos – CONICET, Argentina

## \*Correspondence:

Randy C. Shoemaker, Corn Insects and Crop Genetics Research Unit, Agricultural Research Service, United States Department of Agriculture, G401 Agronomy Hall, Ames, IA 50010, USA.

e-mail: randy.shoemaker@ars.usda.gov

Genomic architecture appears to be a largely unexplored component of gene expression. That architecture can be related to chromatin domains, transposable element neighborhoods, epigenetic modifications of the genome, and more. Although surely not the end of the story, we are learning that when it comes to gene expression, size is also important. We have been surprised to find that certain patterns of expression, tissue specific versus constitutive, or high expression versus low expression, are often associated with physical attributes of the gene and genome. Multiple studies have shown an inverse relationship between gene expression patterns and various physical parameters of the genome such as intron size, exon size, intron number, and size of intergenic regions. An increase in expression level and breadth often correlates with a decrease in the size of physical attributes of the gene. Three models have been proposed to explain these relationships. Contradictory results were found in several organisms when expression level and expression breadth were analyzed independently. However, when both factors were combined in a single study a novel relationship was revealed. At low levels of expression, an increase in expression breadth correlated with an increase in genic, intergenic, and intragenic sizes. Contrastingly, at high levels of expression, an increase in expression breadth inversely correlated with the size of the gene. In this article we explore the several hypotheses regarding genome physical parameters and gene expression.

**Keywords:** expression level, expression breadth, selection, evolution

## INTRODUCTION

Ever since Beadle and Tatum conducted simple but elegant experiments that led to a basic understanding that “genes act by regulating definite chemical events” (Beadle and Tatum, 1941) we have known that mutations can influence the fate of an organism. This profound finding led to their receiving the Nobel Prize in Physiology or Medicine in 1958. We now know that the regulation of expression of genes is more complex. Expression is no longer thought to be controlled solely by the “strength” of a promoter, but is modulated by transcription factors, small RNAs, parachromatin, as well as by all of the components that make up epigenetics (Jorgensen, 2011).

Identifying the internal cues that regulate gene expression can help in deciphering the form and function of living organisms. With the surge in whole-genome sequencing, exploring the uncharted territories and complex evolutionary constraints is now possible. Until recently, genic properties such as exon size and intron size have been assumed to evolve under stochastic processes. In the last 10 years, a correlation between transcriptional demands and genic properties has been identified. Each gene has an individual profile varying in the level of transcription and the number of tissues in which it is expressed. As the transcriptional demands of a gene increase, the genic size tends to decrease. Proposals for the explanation of this relationship have focused on selection for economy, a regional mutational bias, or genomic organization. While this relationship is seemingly constant in animals, in plants many contradicting results have been found (Ren et al., 2006; Camiolo

et al., 2009; Yang, 2009; Woody et al., 2011). It is apparent that different selective forces are acting on the plant genomes than what has been previously thought.

## THE MODELS

“Selection for economy” proponents base their argument on the fact that transcription and translation are both time-consuming and costly (Urrutia and Hurst, 2003; Seoighe et al., 2005). To transcribe one nucleotide, two adenine triphosphate molecules and roughly 0.05 s are required (Carmel and Koonin, 2009) thus it would be advantageous to the organism to reduce the cost of those genes ubiquitously and highly transcribed and translated. As might be apparent, within the selection for economy argument, there are two sub-arguments; the energetic cost hypothesis and the time cost hypothesis. The energetic cost hypothesis states that selection is influenced by a drive to minimize the energetic cost of transcription. Alternatively, in the time cost hypothesis, shorter introns and shorter exons are selected when limited time periods are required to transcribe large amounts of mRNA (Rao et al., 2010). The common thread is that the decrease in genic size is a result of selected mutations with the purpose to decrease the demands of highly transcribed genes.

If indeed, selected mutations occur that result in decreased gene sizes and increased transcription one has to wonder when and how does this take place. Selection for gene reduction based on economical reasons could occur at two stages, transcription and translation. An equal decrease in intron and exon size would

suggest selection is occurring at the transcription stage while a decrease solely in the exon size would point to selection at the point of translation. To make this even more complex, selection could be occurring at both stages. For this reason, it appears that there are two facets to the argument for selection for economy, is it occurring and if so, is it in transcription or translation?

While the selection for economy hypothesis is reasonable, it does not explain the shortening of non-coding regions in genes that are highly and/or broadly transcribed. Vinogradov (2004) suggested that broadly expressed genes required simple regulation and therefore less regulatory elements. Conversely, tissue specific genes contain more functional domains and are associated with more complex protein architecture (Vinogradov, 2004) resulting in larger gene “spaces.” The genome complexity model postulates that the functional properties of a gene determine the length of the physical genic properties (Eisenberg and Levanon, 2003; Vinogradov, 2006). Intron and intergenic regions are hypothesized to be involved in chromatin-mediated suppression and higher order regulation thus introns and intergenic regions are increased when genes are transcribed at a low level or in a tissue specific manner.

The mutational model focuses on transcription-associated non-adaptive deletion bias, the idea that highly expressed regions are in chromosomal regions with high deletion rates resulting in the bias (Urrutia and Hurst, 2003; Comeron, 2004). The selection for economy model and the mutational bias model share a lot of overlap but the underlying concept is different for the two. The selection for economy model refers to the strain an individual gene’s transcription and translation puts on the cell. At a larger level, the mutational bias model suggests that the “neighborhood” of the gene is the cause for selection. Highly expressed genes tend to cluster in the chromosomes (Caron et al., 2001) and it is hypothesized that this clustering might result in local mutational bias.

Eukaryotic genomes are composed with a myriad of distinct regions of varying GC content. Genomic regions containing many genes tend to be GC rich (Urrutia and Hurst, 2003) and thus are also regions of high recombination rates (Fullerton et al., 2001). It is possible that the increase in recombination imposes a mutational bias on these highly expressed genes (Seoighe et al., 2005). However, the mutational bias model has also been suggested at the individual gene level. As a gene is transcribed more it is more disposed to retroposition and reverse transcription (Mourier and Jeffares, 2003).

In chicken, (Rao et al., 2010) gene size, CDS length, first intron length, average intron length, and total intron length are negatively correlated with expression level and expression breadth. In humans, (Eisenberg and Levanon, 2003) 575 constitutively expressed genes were analyzed and were found to have shorter introns, untranslated regions, and coding sequences than tissue specific genes. These studies add support to the selection for economy model as the regions that are transcribed are decreasing in size as expression increased. They also found that the difference in genic size between tissue specific and housekeeping genes was larger for the introns than for the exons and proposed that the coding sequences and UTRs would be less susceptible to change based on selection. Another study in humans and *Caenorhabditis elegans* identified a significant decrease in the intron size of highly expressed genes and this decrease was much larger than the

decrease in coding region size suggesting that the reduction is not functional but a result of natural selection (Castillo-Davis et al., 2002).

It is readily apparent that the models allow for conceptual overlap. A reduction in intron size could also support the genome complexity model. An increase in expression correlates with a decrease in regulatory elements and thus a decrease in intron and intergenic size according to the model. However, Li et al. (2007) analyzed genes with high functional/regulatory complexity in *M. musculus*, human, and *Arabidopsis thaliana* and found that these genes did not have longer introns or longer proteins. In addition, they did not find that housekeeping genes were more compact than tissue specific genes expressed at similar expression levels. And so, the controversy grows.

## THE “CONTROVERSY”

A controversy has emerged regarding expression and the structure of plant genomes. In a contradiction to the models, Ren et al. (2006) studied both *Oryza sativa* and *A. thaliana* and found that highly expressed genes contained more and longer introns and a produced a larger primary transcript than genes expressed at a low level. The genic parameters also increased as the expression breadth increased which is different than what had been found in animals. However, in a subsequent study in *Arabidopsis* both the non-coding and coding regions of the genes decreased as the expression level increased (Camiolo et al., 2009).

In accordance with the previous study, another study in *Arabidopsis* found that expression breadth positively correlated with the non-coding structural parameters (Yang, 2009), e.g., non-coding regions got larger as expression breadth increased. However, in the same study expression breadth was negatively correlated with the coding regions, e.g., coding regions got smaller. It is possible that plant genomes are under a different selection pressure than animals and that different methods are needed to decipher the evolutionary process.

Using a “primitive” plant, Stenøien (2007) studied the possible effect of selection on genome organization in the haploid moss *Physcomitrella patens*. They found that total intron length, the number of introns, and the total length of genes are negatively correlated with the level of expression. They suggest that if animals and plants have followed separate evolutionary pathways then this difference must have occurred after the split between vascular and non-vascular plants (250 mya, Palmer et al., 2004). One suggested explanation for this difference is that plants tend to have much smaller introns. *Arabidopsis* has an average intron length per gene of 152, 387 bp in rice (Ren et al., 2006) compared to 5.5 kbp in humans (Sakharkar et al., 2004). A much larger transcriptional demand on the introns of humans seems plausible. However, *P. patens*’ average intron length is 252 bp, not significantly different from *Arabidopsis* and smaller than rice (Rensing et al., 2005). Subsequent expression studies done in *Arabidopsis* and other plant species revealed different results. Colinas et al. (2008) found that the size of the introns and exons negatively correlated with expression levels. This seemingly nullified the argument that vascular and non-vascular plants are evolving under different constraints.

Interestingly, it is not just in plants that opposing correlations have been discovered.

In several yeasts and other unicellular organisms, highly expressed genes have longer introns than genes expressed at a low level (Vinogradov, 2001). In the unicellular green alga *Ostreococcus lucimarinus*, intron number and intron density are positively correlated with expression level (Lanier et al., 2008). Even in animals, as in the mouse example above, controversy has occurred. In chicken, ubiquitously expressed genes were compared with narrowly expressed genes and they found that ubiquitously expressed genes were larger (Rao et al., 2010). However, they found that gene size, CDS length, first intron length, average intron length, and total intron length all negatively correlated with expression level. Throughout the dispute, it is unclear as to whether the source of the contradictions is expression level or expression breadth.

An important consideration when evaluating the contradictions is the quantification and characterization of expression and genic properties both within and across species. Can an ancient polyploid with a large genome such as soybean be compared to a genome such as rice? Both have experienced dramatically different evolutionary trajectories. Can the evolutionary processes of plants be analyzed and compared with animals? Even within a species experiments vary. Expression breadth is relative to the tissue and time points analyzed in the study. This is not to say that we cannot compare across studies but this should be contemplated when making generalizations. A similar conflict occurs when analyzing genic properties. Each individual property (exon length, intron length, intergenic region, individual exon lengths) can tell us a different story to complement the fluid movements of the whole gene. Understanding the evolutionary differences between intron and exon length can give us a wealth of information on what may be occurring during transcription compared to translation.

### A NOVEL DICHOTOMY IN HIGHLY EXPRESSED GENES COMPARED TO LOWLY EXPRESSED GENES

A recent study in soybean took a unique approach and partitioned the genes first into categories of expression level (low, mid, high) and then into categories of expression breadth (Woody et al., 2011). A unique division was observed; genes that were expressed at high levels decreased in size as the expression breadth increased while genes that were expressed at low levels increased in size as the expression breadth increased. This leads to the hypothesis that multiple divergent evolutionary paths may be present. Those genes at a low level of expression may be under a different model of selection than those at a high level of expression. In humans, Zhu et al. (2008) looked at 17,288 RefSeq loci across 18 tissues and found that, on average, highly expressed genes are more compact but that genes expressed at a low level show a lot more variation. They suggested that highly expressed genes could be the only genes under an economical selection pressure (selection for economy). In *Arabidopsis* and rice, it was found that housekeeping genes, compared to tissue specific genes, are under stronger selective constraints and that weakly expressed genes, compared to highly expressed genes also are under stronger selective constraints (Mukhopadhyay et al., 2008). When analyzed further they found that highly expressed housekeeping genes had a lower synonymous substitution rate than lowly expressed housekeeping genes. Berg and Martelius (1995) suggested that a lower synonymous substitution rate was due to a transcriptional selection for economy.

Mukhopadhyay et al. (2008) found that by analyzing preferred codon usage, highly expressed genes that were broadly expressed were under selection for economy through tRNA copy number that was used to optimize the synonymous codon usage. Lowly expressed genes are under a stronger selective pressure than highly expressed genes but highly expressed housekeeping genes are also under a selective pressure and this can be localized to a codon usage bias.

Selection for economy may explain the evolution of highly expressed genes but other selective forces, potentially stronger forces, are acting upon weakly expressed genes. This selection appears to increase as the expression breadth increases. In Woody et al. (2011) it was observed that tissue specific genes did not display a large difference in genic size between low, mid, and highly expressed genes, although the physical parameters of highly expressed tissue specific genes were always slightly larger than lowly expressed tissue specific genes. It was postulated that the genes expressed at a low level of expression are selected upon by the demands of being polytypic (genes involved in alternative splicing events). Genes that are lowly expressed, with an increasing breadth of expression share many properties with polytypic genes. Genes expressed at a low level increased in total genic length by increasing the number of exons, not the size of exons and this is dissimilar to highly expressed genes. In humans, an increase in exons and larger transcripts were shown to correlate with polytypic genes expressed at a low level.

What properties of alternative splicing lead to a selection for an increase in exon number? Exon-exon junction complexes are placed on mRNAs during splicing. These complexes result in a post-transcriptional effect in that the size of the transcript and the efficiency of translation are both increased (Camiolo et al., 2009). In a previous study on alternative isoforms in humans, it was found that many gene isoforms of alternative splice genes contained premature termination codons and were subject to non-sense mediated decay and subsequently decreases the transcription level (Hillman et al., 2004). Thus, a selection for economy could be suggested in the highly expressed genes but the lowly expressed genes have a different method of evolutionary selection that possibly rises from the demands of being polytypic.

### SELECTION ON THE INDIVIDUAL GENE OR ON AN ENTIRE REGION?

If weakly expressed genes evolve under the umbrella of alternative splicing demands, it would appear evident that selection would be at an individual level. However, if nature was selecting for an economical purpose, it is reasonable to question whether entire neighborhoods are under specific selection. Clustering of highly expressed genes has been established and several physical genomic properties have been associated with these regions. In a study that combines transposable elements, gene length, and gene expression Jjingo et al. (2011) found that all three of those factors are closely related. Combined together, transposable elements and gene length account for 78% of the variation in expression level, 76% of the variation in expression breadth, and 66% of the variation in tissue specificity. The authors proposed a role for selection for economy but suggested that the removal of transposable elements may be a stronger mechanism of selection than

reduction of gene length. In a study done in rice (Tian et al., 2009) retrotransposons, genetic recombination, and gene density were all correlated and they suggested this relationship helped shape the makeup of the rice genome.

In rice, transposable element families were differentially distributed across the genomes in areas of varying methylation patterns (Takata et al., 2007). Kim et al. (2004) found that the expression breadth of a gene is highly correlated with Alu elements and expression level is highly correlated with L1 densities in human. Confirmed by Eller et al. (2007), highly and broadly expressed genes are enriched with Alu elements and depleted in L1 elements. This suggests that rather than gene expression or transposable element insertion accounting for a variation in genic level, epigenetics may be influencing the entire genetic region. Isochores, large regions within the genome that are homogeneous in their GC content have been characterized and analyzed since 1976 (Macaya et al., 1976). Gene density, gene expression, insertion of transposable elements and density of transposable elements are only a few of the basic biological properties associated with isochores (Bernardi, 2004). It is possible that these properties act as a unit and isochores are the homes for these interactions.

If different gene sizes and transposable element densities change across isochore families and these properties have a large influence on expression, it follows that expression profiles are also influenced by these homogeneous structures. Two questions would arise if this was the case: what is the relationship between these characteristics in the homogeneous regions and do heterogeneous regions have different sets of characteristics with their own distinguishing features. This brings us back to the cost of transcription and translation, the nucleosome formation potential, related to homogeneity and heterogeneity, could influence both the chromatin domain and the size of the gene.

Another variable to consider when studying the evolution of individual components and their relationship with expression level at a whole-genome level is replication timing. Replication timing and expression profiles do not directly influence each other but both seem to be regulated through a mediator (Gilbert, 2002; MacAlpine and Bell, 2005; Gilbert and Gasser, 2006; Hiratani et al., 2008; Farkash-Amar and Simon, 2009; Schwaiger et al., 2009; Ryba et al., 2010). There are two main stages in replication, early and late. If a replication domain changes timing, the chromatin state

usually changes and transcriptional activation or suppression usually follows. Replication timing correlates with isochore structure as well suggesting overarching domains.

Could chromatin domains be the top order of regulation? Chromatin domains have been well studied in many higher eukaryotes although *Arabidopsis* is the only plant with extensive research done. Replication domains in *Arabidopsis* are correlated with chromatin conformation and sequence content (Lee et al., 2010). Co-expression can be coordinated by the sharing of a promoter in neighboring genes. However, co-expressed domains at large distances have also been identified (Chen et al., 2010). It is known that epigenetics helps regulate transcription but its effects in whole-genome view are still unclear. Are the replication domains determining the chromatin domains which in turn regulate gene expression? Does the sequence composition, the isochore family, enrich these determinants or are they the determinants for the replication domain?

A circular debate seems inevitable if we try to account for the actions of one biological property such as gene size acting on another property such as presence or absence of transposable elements. It is becoming clear that we need to consider gene expression in a more holistic manner. A complex array of neighborhoods appears to be covering the genome. Jorgensen (2011) described the genome as comprised of two types of chromatin, “orthochromatin” which is the stable, constant function of the chromatin and “parachromatin,” a dynamic and reactive chromatin. Parachromatin could provide a large but dynamic and flexible cloud over the active properties within the genome. Each element, transcriptional demands, transposable element insertion, small RNAs, etc., impact the other but survival is not possible unless the elements are fit to live under the epigenetic cloud.

## ACKNOWLEDGMENTS

The authors gratefully acknowledge the financial support of the North Central Soybean Research Program and the USDA Agricultural Research Service. Names are necessary to report factually on the available data; however, the USDA neither guarantees nor warrants the standard of the product, and the use of the name by the USDA implies no approval of the product to the exclusion of others that may also be suitable.

## REFERENCES

- Beadle, G. W., and Tatum, E. L. (1941). Genetic control of biochemical reactions in *Neurospora*. *Proc. Natl. Acad. Sci. U.S.A.* 27, 499–506.
- Berg, O. G., and Martelius, M. (1995). Synonymous substitution-rate constants in *Escherichia coli* and *Salmonella typhimurium* and their relationship to gene expression and selection pressure. *J. Mol. Evol.* 41, 449–456.
- Bernardi, G. (2004). *Structural and Evolutionary Genomics. Natural Selection in Genome Evolution*. Amsterdam: Elsevier.
- Camiolo, S., Domenico, R., and Porceddu, A. (2009). Mutational biases and selective forces shaping the structure of *Arabidopsis* genes. *PLoS ONE* 4, e6356. doi:10.1371/journal.pone.0006356
- Carmel, L., and Koonin, E. V. (2009). A universal nonmonotonic relationship between gene compactness and expression levels in multicellular eukaryotes. *Genome Biol. Evol.* 1, 382–390.
- Caron, H., van Schaik, B., van der Mee, M., Baas, F., Riggins, G., van Sluis, P., Hermus, M., van Asperen, R., Boon, K., Voute, P. A., Heisterkamp, S., van Kampen, A., and Versteeg, R. (2001). The human transcriptome map: clustering of highly expressed genes in chromosomal domains. *Science* 291, 1289–1292.
- Castillo-Davis, C. I., Mekhedov, S. L., Hartl, D. L., Koonin, E. V., and Kondrashov, F. A. (2002). Selection for short introns in highly expressed genes. *Nat. Genet.* 31, 415–418.
- Chen, W., Meaux, J., and Lercher, M. J. (2010). Co-expression of neighbouring genes in *Arabidopsis*: separating chromatin effects from direct interactions. *BMC Genomics* 11, 178. doi:10.1186/1471-2164-11-178
- Colinas, J., Schmidler, S. C., Bohrer, G., Iordanov, B., and Benfey, P. N. (2008). Intergenic and genic sequence lengths have opposite relationships with respect to gene expression. *PLoS ONE* 3, e3670. doi:10.1371/journal.pone.0003670
- Comeron, J. M. (2004). Selective and mutational patterns associated with gene expression in humans: influences on synonymous composition and intron presence. *Genetics* 67, 1293–1304.
- Eisenberg, E., and Levanon, E. Y. (2003). Human housekeeping genes are compact. *Trends Genet.* 19, 362–365.
- Eller, C. D., Regelson, M., Merriam, B., Nelson, S., Horvath, S., and Marahrens, Y. (2007). Repetitive sequence environment distinguishes housekeeping genes. *Gene* 390, 153–165.

- Farkash-Amar, S., and Simon, I. (2009). Genome-wide analysis of the replication program in mammals. *Chromosome Res.* 18, 115–125.
- Fullerton, S. M., Bernardo Carvalho, A., and Clark, A. G. (2001). Local rates of recombination are positively correlated with GC content in the human genome. *Mol. Biol. Evol.* 18, 1139–1142.
- Gilbert, D. M. (2002). Replication timing and transcriptional control: beyond cause and effect. *Curr. Opin. Cell Biol.* 14, 377–383.
- Gilbert, D. M., and Gasser, S. M. (2006). “Nuclear structure and DNA replication,” in *DNA Replication and Human Disease*, ed. M. L. DePamphilis (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press), 175–196.
- Hillman, R. T., Green, R. E., and Brenner, S. E. (2004). An unappreciated role for RNA surveillance. *Genome Biol.* 5, R8.
- Hiratani, I., Ryba, T., Itoh, M., Yoochi, T., Schwaiger, M., Chang, C. W., Lyou, Y., Townes, T. M., Schubeler, D., and Gilbert, D. M. (2008). Global reorganization of replication domains during embryonic stem cell differentiation. *PLoS Biol.* 6, e245. doi:10.1371/journal.pbio.0060245
- Jjingo, D., Huda, A., Gundapuneni, M., Marino-Ramirez, L., and Jordan, I. K. (2011). Effect of the transposable element environment of human genes on gene length and expression. *Genome Biol. Evol.* 3, 259–271.
- Jorgensen, R. A. (2011). Epigenetics: biology’s quantum mechanics. *Front. Plant Sci.* 2:10. doi:2289/fpls.2011.00010
- Kim, T. M., Yung, Y. C., and Rhyu, M. G. (2004). Alu and L1 retroelements are correlated with the tissue extent and peak rate of gene expression, respectively. *J. Korean Med. Sci.* 19, 783–792.
- Lanier, W., Moustafa, A., Bhattacharya, D., and Comeron, J. M. (2008). EST Analysis of *Ostreococcus lucimarinus*, the most compact eukaryotic genome, shows an excess of introns in highly expressed genes. *PLoS ONE* 3, e2171. doi:10.1371/journal.pone.0002171
- Lee, T., Pascuzzi, P. E., Settlege, S. B., Shultz, R. W., Tanurdzic, M., Rabinowicz, P. D., Menges, M., Zheng, P., Main, D., Murray, J. A. H., Sosinski, B., Allen, G. C., Martienssen, R. A., Hanley-Bowdoin, L., Vaughn, M. W., and Thompson, W. F. (2010). *Arabidopsis thaliana* chromosome 4 replicates in two phases that correlate with chromatin state. *PLoS Genet.* 6, 6. doi:10.1371/journal.pgen.1000982
- Li, S. W., Feng, L., and Niu, D. K. (2007). Selection for the miniaturization of highly expressed genes. *Biochem. Biophys. Res. Commun.* 360, 586–592.
- MacAlpine, D. M., and Bell, S. P. (2005). A genomic view of eukaryotic DNA replication. *Chromosome Res.* 13, 309–326.
- Macaya, G., Thiery, J. P., and Bernardi, G. (1976). An approach to the organization of eukaryotic genomes at a macromolecular level. *J. Mol. Biol.* 108, 237–254.
- Mourier, T., and Jeffares, D. C. (2003). Eukaryotic intron loss. *Science* 300, 1393.
- Mukhopadhyay, P., Basak, S., and Ghosh, T. C. (2008). Differential selective constraints shaping codon usage pattern of housekeeping and tissue-specific homologous genes of rice and *Arabidopsis*. *DNA Res.* 15, 347–356.
- Palmer, J. D., Soltis, D. E., and Chase, M. W. (2004). The plant tree of life: an overview and some points of view. *Am. J. Bot.* 91, 1437–1445.
- Rao, Y. S., Zhang, F. W., Chai, X. W., Wu, G. Z., Zhou, M., Nie, Q. H., and Zhang, X. Q. (2010). Selection for the compactness of highly expressed genes in *Gallus gallus*. *Biol. Direct* 5, 35.
- Ren, X., Vorst, O., Fiers, M., Stiekema, W., and Nap, J. (2006). In plants, highly expressed genes are the least compact. *Trends Genet.* 22, 528–532.
- Rensing, S. A., Fritzowsky, D., Lang, D., and Reski, R. (2005). Protein encoding genes in an ancient plant: analysis of codon usage, retained genes and splice sites in a moss, *Physcomitrella patens*. *BMC Genomics* 6, 43. doi:10.1186/1471-2164-6-43
- Ryba, T., Hiratani, I., Lu, J., Itoh, J., Kulik, M., Zhang, J., Schulz, T. C., Robins, A. J., Dalton, S., and Gilbert, D. M. (2010). Evolutionarily conserved replication timing profiles predict long-range chromatin interactions and distinguish closely related cell types. *Genome Res.* 20, 761–770.
- Sakharkar, M. K., Chow, V. T., and Kanguane, P. (2004). Distributions of exons and introns in the human genome. *In silico Biol. (Gedrukt)* 4, 387–393.
- Schwaiger, M., Stadler, M. B., Bell, O., Kohler, H., Oakeley, E. J., and Schubeler, D. (2009). Chromatin state marks cell-type- and gender-specific replication of the *Drosophila* genome. *Genes Dev.* 23, 5890601.
- Seoighe, C., Gehring, C., and Hurst, L. D. (2005). Gametophytic selection in *Arabidopsis thaliana* supports the selective model of intron length reduction. *PLoS Genet.* 1, e13. doi:10.1371/journal.pgen.0010013
- Stenøien, H. K. (2007). Compact genes are highly expressed in the moss *Physcomitrella patens*. *J. Evol. Biol.* 20, 1223–1229.
- Takata, M., Kiyohara, A., Takasu, A., Kishima, Y., Ohtsubo, H., and Sano, Y. (2007). Rice transposable elements are characterized by various methylation environments in the genome. *BMC Genomics* 8, 469. doi:10.1186/1471-2164-8-469
- Tian, Z., Rizzon, C., Du, J., Zhu, L., Bennetzen, J. L., Jackson, S. A., Gaut, B. S. and Ma, J. (2009). Do genetic recombination and gene density shape the pattern of DNA elimination in rice long terminal repeat retrotransposons. *Genome Res.* 19, 2221–2230.
- Urrutia, A. O., and Hurst, L. D. (2003). The signature of selection mediated by expression on human genes. *Genome Res.* 13, 2260–2264.
- Vinogradov, A. E. (2001). Intron length and codon usage. *J. Mol. Evol.* 52, 2–5.
- Vinogradov, A. E. (2004). Compactness of human housekeeping genes: selection for economy or genomic design?. *Trends Genet.* 20, 248–253.
- Vinogradov, A. E. (2006). “Genome design” model: evidence from conserved intronic sequence in human-mouse comparison. *Genome Res.* 16, 347–354.
- Woody, J. L., Severin, A. J., Bolon, Y., Joseph, B., Diers, B. W., Farmer, A. D., Weeks, N., Muehlbauer, G. J., Nelson, R. T., Grant, D., Specht, J. E., Graham, M. A., Cannon, S. B., May, G. D., Vance, C. P., and Shoemaker, R. C. (2011). Gene expression patterns are correlated with genomic and genic structure in soybean. *Genome* 54, 10–18.
- Yang, H. (2009). In plants, expression breadth and expression level distinctly and non-linearly correlate with gene structure. *Biol. Direct* 4, 45; discussion 45.
- Zhu, J., He, F. H., Hu, S. N., and Yu, J. (2008). On the nature of human housekeeping genes. *Trends Genet.* 24, 481–484.

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 11 July 2011; accepted: 29 September 2011; published online: 17 October 2011.

Citation: Woody JL and Shoemaker RC (2011) Gene expression: sizing it all up. *Front. Gene.* 2:70. doi: 10.3389/fgene.2011.00070

This article was submitted to *Frontiers in Plant Genetics and Genomics*, a specialty of *Frontiers in Genetics*.

Copyright © 2011 Woody and Shoemaker. This is an open-access article subject to a non-exclusive license between the authors and Frontiers Media SA, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and other Frontiers conditions are complied with.