



Implication of the changing concept of genes on plant breeder's work

Kátia Ferreira Marques de Resende^{1*}, Fernanda Motta da Costa Santos¹, Marco Aurélio D. Dias² and Magno Antônio Patto Ramalho¹

Received 6 January 2011

Accepted 8 May 2011

ABSTRACT – *The recent genome sequencing of some species has accumulated evidence that for a large number of traits, the control and action of genes are far more complex than previously thought. This article discusses possible implications of new insights into the gene concept on the work of plant breeders. Apparently, the successful application of biotechnological techniques is not as simple as once assumed. The evident changes in the available concept of genes confirmed what the past experience had shown, i.e., selection should focus on the phenotype, under the same conditions as the plant is to be cultivated in. Advanced vocational training of plant breeders must be continuously maintained, focusing on phenotype-based selection in as accurate as possible experiments.*

Key words: *plant breeding, genetics, molecular marker, transgenic.*

INTRODUCTION

The overall concept of plant breeding is defined as “the science, art and business of improving plants for human benefit” (Bernardo 2002). This approach expressly declares that the breeder should have scientific knowledge, individual skills (art) to select plants and/or progenies, the aptness of a good manager and, above all, a holistic view of agriculture and its interactions with human society.

In terms of scientific knowledge, the concept of genes/alleles is of course, primarily essential for plant breeders. However, as will be shown, this concept has been changing over time, due to the accumulation of information about how gene regulation and action occur (Wain et al. 2002, Pearson 2006, Gerstein et al. 2007, Scherrer and Jost 2007a,b, Joaquim and El Hani 2010). These changes

became more far-reaching after the publication of studies of the genome of some species (AG Initiative 2000, Qing-Po and Qing-Zhong 2006, Orjuela et al. 2010). It is an open question how the gene concept will affect the work of breeders in the future.

Ever since the domestication of plants, selection has been based on the phenotype. With the advent of biotechnology, selection became possible directly in the genotype by means of molecular marker techniques and by gene introduction into unrelated species through recombinant DNA technology. Some success has been achieved, but much remains to be done and recent findings on gene action and regulation allow the identification of some difficulties, particularly with regard to biotechnological methods, to contribute effectively to the establishment of new cultivars.

¹ Universidade Federal de Lavras (UFLA), Departamento de Biologia, Campus Universitário, CP 3037, 37.200-000, Lavras, MG, Brazil. *E-mail: katia.lavras@ig.com.br

² UFLA, Departamento de Zootecnia, Campus Universitário, 37.200-000, Lavras, MG, Brazil

The purpose of this review is to approach the changing concept of genes and specifically discuss the implications of current insights into gene regulation and action on the work of plant breeders in the coming years.

The concept of genes/alleles prior to the discovery of the DNA structure

It is undeniable that the question how genetic information is passed on from father to son, or between cells of a body, must have attracted the attention of mankind ever since its origin. There are reports that the Greeks established some hypotheses trying to explain the transmission of genetic information. However, only in the late nineteenth century some information about the genetic control of traits was published with the work of Mendel. However, Mendel did not use the term gene, but “factor” instead. The term “genetics” for example, was coined by Bateson in 1902, in his book “Mendel’s principles of heredity: a defence”, who also proposed the definition of allele as “one of two or more dissimilar factors or genes which on account of their corresponding position in corresponding chromosomes are subject to alternative (Mendelian) inheritance” (Darlington and Mather 1950). The Danish Wilhelm Johannsen termed the Mendelian factors “genes”, in analogy to genetics, as “any particle to which the properties of a Mendelian factor may be attributed” (Darlington and Mather 1950)

The work of Thomas Hunt Morgan, in which he associated the gene to chromosomes, represented another major advance in the 20s of the twentieth century. He conceptualized a gene as “any particle in a chromosome which is distinguishable from other particles by either crossing-over or mutation” (Darlington and Mather 1950). This idea has triggered great advances in genetics and has not been challenged ever since.

The publications on genetics of this period barely ever mention the chemical composition and structure of a gene. In 1928, F. Griffith had already shown that DNA was the transforming principle in bacteria. The proof that in most organisms the gene was the DNA was only delivered in 1944, in studies of Avery, MacLeod and McCarthy (Griffiths et al. 2000, Ramalho et al. 2008).

It is noteworthy that in the time preceding the discovery of the DNA structure, substantial progress was made in the genetic improvement of cultivated plants, without any knowledge of the structure and chemical composition of genes. Heterosis became known in the early 20th century. In this period, the influence of the environment

on the phenotypic trait expression was also emphasized, requiring the evaluation of progenies in experiments with replications; i.e., experimental statistics were developed (Simmonds 2000).

Great importance should also be attached to the paper of Beadle and Tatum published in 1941, which associated gene mutations with metabolic pathways, leading to the establishment of the famous “one gene - one enzyme” hypothesis (Griffiths et al. 2000).

The gene/allele concept according to Watson and Crick

The DNA structure must have intrigued many researchers. One of the first reports in this regard, by Darlington and Mather (1950), reads as follows (pages 145 and 146): “The nucleic acid consists of an unlimited polymerization of desoxyribose nucleotides to form a column. The nucleotides are spaced at the same distance of 3.3 Angström units apart, as are the repeats in the stretched polypeptide chain. Desoxyribose nucleic acid is found only attached to chromosomes and in bacteria and certain animal and bacterial viruses. In all of these it must indicate the presence of genetically analogous structures.” Clearly, they hinted at a chromosome structure that is very similar to that proposed later for DNA, by Watson and Crick (1953). Interestingly, shortly after the publication of Watson and Crick, some researchers commented on the difficulty of conceptualizing what a gene is. Waddington (1956) commented that the definition of a gene is unsatisfactory, since there are cases where it does not apply, for example to the effects of position, which shows that the action of interchangeable genes on the development of an organism is not independent.

The period from 1961 to 1966 was particularly important for science, when the genetic code was deciphered (Ramalho et al. 2008). This code associated a triplet of mRNA bases, called codon, with one of the 20 essential amino acids. This evidenced the DNA-protein association even more clearly. The allele has been conceptualized as alternative form of the same gene that arises by changes in the sequence of one or few DNA bases. From then on, the concept of a gene has also began to be more explicitly cited in papers, as a DNA segment located in a specific position of the chromosome responsible for producing a polypeptide chain (Burns 1983, Elseth and Baumgardner 1984). However, exceptions to this gene concept began to appear, since not all transcribed RNAs are translated into polypeptides, e.g., rRNAs and tRNAs, as Waddington (1956) had commented. With the ongoing development of

cloning and sequencing techniques, unearthing details on how genes are organized and expressed, molecular biology was revolutionized (Gerstein et al. 2007).

The second half of the 20th century saw the greatest-ever population boom. The world population increased from 2.5 billion to 6.0 billion (USCB 2008). This rise required enormous efforts in food production to be able to meet such a high demand. By the use of fertilizers, irrigation and mechanization, along with genetic improvement, all the necessary food was produced (Miles and Pandey 2004, Tracy et al. 2004). It is worth emphasizing that improvement has been achieved so far without making use of the molecular genetic knowledge (Duvick et al. 2004).

The era of biotechnology at the end of the 20th century aroused great expectations with regard to the possibilities of selection of plants and animals directly by the genotype. The use of molecular markers was seen as a key strategy to continuous genetic progress. The recombinant DNA and transgenic technology showed that the passage of genes between species, hitherto unthought of, was possible. The challenges to satisfy all these expectations were huge and led to the need to sequence the entire genome of a species.

The of gene/allele concept in the era of genomics

In the late 90s of last century, research was intensified to sequence genomes. As of 1999, there was a massive effort to sequence the human genome. The development of new sequencing machines and the evolution of bioinformatics contributed such that the task that was expected to take 15 years or more was completed in little less than three years. In June 2000, the world was informed that the human DNA had been sequenced (Venter et al. 2001, Collins et al. 2003). The greatest challenge was yet to come, that is, establish an association of the DNA sequence with a particular phenotype (Houle et al. 2010).

Another ten years have passed and some new facts have emerged elucidating gene regulation and action of the gene itself. It became evident that there are more polypeptide chains than genes, so the association one gene - one protein is not an absolute truth. It was also shown that the DNA previously considered “junk”, i.e., untranslated DNA, is extremely important for gene regulation. The environmental effect on gene expression is also more pronounced than previously assumed. These facts and novel concepts, hitherto unimagined, some of which are listed in Table 1, cast doubt on the generalization of the gene concept outlined for the control of all

characters. The controversy was such that the Nature magazine, in 2006, issued an invitation to renowned geneticists to discuss the concept gene for two days. The conclusion was “a locatable region of genomic sequence, corresponding to a unit of inheritance, which is associated with regulatory regions, transcribed regions and/or other functional sequence regions” (Pearson 2006). However, even this broad concept is now being challenged. The ENCODE Project Consortium (ENCODE- ENCYClopedia Of DNA Elements), which unites a group of institutions working to develop knowledge on the human genome, makes restrictions to the previously accepted concept (The Encode Project Consortium 2007).

Thus, evidence accumulated that the gene consists of differently combined modular DNA segments to generate the product (Gerstein et al. 2007). In other words, the gene action depends on several DNA segments scattered throughout the genome, usually in regions distant from the structural part that will produce mRNA. These DNA segments are involved in the production of transcription factors, i.e., factors that participate in the induction or repression of transcription as well as in other processes, and in the production of several RNA types whose function is related to the regulation of gene action. In Brazil, this issue has attracted great interest, for obvious reasons (Joaquim and El Hani 2010). The main conclusion is that there is not always a ratio of 1:1:1, in other words, one gene – one polypeptide chain – one function, but it seems that this relationship is far more complex.

The changing concepts of gene and plant breeding

As stressed above, improvement is science as well as art. It is art, translated in the intuition of someone to see an opportunity in a plant, beginning with domestication and continued to the present, though less intense. It is science, evolved from the work of Mendel, passing on to experimental statistics, to cytogenetics and computational tools, among other areas. In the 20th century, science and art jointly resulted in the development of new cultivars, with optimized grain yields for the continuous and vertiginous growth of the human species.

The challenges in the 21st century will by no means be smaller. Certainly there will be some difficulty in opening up new areas for agriculture, even in countries as Brazil, where extensive farmland is still available. Probably there will not be an ever-increasing use of fertilizers, due to the gradual depletion of reserves. The more extensive use of irrigation is also repeatedly questioned due to the probable

Table 1. Phenomena that make it difficult to have a single gene concept applicable to all situations

Phenomenon	Description	Issue
<i>Gene location and structure</i>		
A gene within another, ie with overlapping reading	A DNA region can code for two or more different protein products in different parts of the readings.	Two genes can occupy the same place, so there is no 1:1 correspondence between DNA and protein sequence
Enhancers, silencers	Gene structural elements	DNA sequences determining expression can be widely separated from one another in the genome, there is no clear border between the different genes.
<i>Structural variation</i>		
Mobile elements	Genetic element appears at new locations over generations	A genetic element may not be constant at its location
Gene rearrangements/introns and exons can alter in different genic products	DNA rearrangement or splicing in somatic cells results in many alternative gene products	Gene structure is not hereditary, or structure may differ across individuals or cells/tissues
Copy-number variants	Copy number of genes/regulatory elements may differ between individuals	Genetic elements may differ in their number
<i>Epigenetics and chromosome structure</i>		
Epigenetic modifications, imprinting	Inherited information may not be only DNA-sequence based	Phenotype is not determined strictly by genotype
Effect of chromatin structure	Chromatin structure, which influences gene expression, is only loosely associated with particular DNA sequences	Gene expression depends on packing of DNA. DNA sequence is insufficient to predict gene product.
<i>Post-transcriptional events</i>		
Alternative splicing of RNA	One transcript can generate multiple mRNAs, resulting in different protein products	Multiple products from one genetic locus; information in DNA is not linearly related to that in protein
RNA editing	RNA is enzymatically modified	The information in the DNA is not encoded directly into RNA sequence
<i>Post-translational events</i>		
Protein splicing, viral polyproteins	Protein product self-cleaves and can generate multiple functional products.	Start and end sites of protein not determined by genetic code
Protein <i>trans</i> -splicing	Distinct proteins can be spliced together in the absence of a <i>trans</i> -spliced transcript	Start and end sites of protein not determined by genetic code
Protein modification	Protein is modified to alter structure and function of the final product	The information on the DNA is not encoded directly into protein sequence
<i>Retrogenes</i>		
Retrogenes	A retrogene is formed from reverse transcription of the mRNA of its parent gene and by insertion of the DNA product into a genome	RNA-to-DNA flow of information

Sources: adapted from Gerstein et al. (2007) and Joaquim and El-Hani (2010).

lack of drinking water and even the increasing cost of water in the 21st century. It is known that the expansion of environmental awareness will result in a reduced use of pesticides. Summing up, unless some new extraordinary

fact is discovered, the genetic improvement of plants will continue to be the main alternative to increase grain yields.

Biotechnology has emerged in the second half of the 20th century, as the main strategy available to overcome

the challenge of developing new cultivars in a short time, able to produce more and better food. The use of DNA molecular markers, by the possibility of selection of genotypes rather than phenotypes, was hailed as an innovation with great potential, without the least objection. In the last 30 years many types of markers have been developed, the latest marker always being touted as the most powerful and revolutionary (Guimarães et al. 2007, Bernardo 2008, Borém and Caixeta 2009). The number of publications on this topic was and is huge.

After the period of maturation of biotechnological techniques, scientists now question whether these means are being used efficiently in plant breeding. There is evidence that their application in improvement is effectively restricted to a few multinational seed producers (Eathington et al. 2007). But even in these cases there is no information on whether biotechnological techniques are indispensable in the generation of new cultivars. As discussed above, the gene is an entity distributed across the entire genome and its expression depends on the coordinated action of all these regions as well as on the environmental influence. So hardly a marker could or will be able to identify the DNA region which, under given conditions, is responsible for a better phenotypic expression. Considering "alternative splicing", that is, a same DNA sequence, structural region of a given gene, being rearranged to originate multiple gene products, the use of the marker is even more restrictive. There is evidence that alternative splicing is less frequent in plants than in animals, but still does occur (Griffiths et al. 2000). One should not forget that rearrangements of proteins also occur after translation. In this case, it is very unlikely that a DNA marker would be associated with the various arrangements of polypeptide chains. These are some of the arguments that would restrict the use of DNA molecular markers, for a large number of characters, based on the current concept of the gene. All this indicates that there is still a long way to go before the techniques of molecular markers can effectively contribute to the production of new cultivars superior to the existing.

The recombinant DNA technology for the production of genetically modified plants has been a huge success, especially in the case of herbicide resistance and insect pest control, by bt insertion. The efficiency was surprising. It is noteworthy that in all successful cases, a gene of

prokaryotic origin was used, which seems to have a simpler regulatory system. The current knowledge of the genome clearly indicates that for many characters the insertion of an exogenous gene and its subsequent expression would be a most unlikely event. Although there are publications emphasizing that a new generation of GM crops is nearly market-ready (Green 2009, James 2009), their market release has been delayed for some reason. Once again, the mode of action and regulation of eukaryotic genes shows that much remains to be done until other economically viable GM products will be available on the market.

The challenge is even greater when considering the quantitative traits. What are polygenes really like? Do they actually exist? Are they simply the products of one or few genes, whose action and regulation are distributed throughout the genome? The answers to these questions could explain the response to long-term selection observed in a number of situations (Dudley and Lambert 2004, Muir et al. 2004). That is, during selection the benefits add up in different genome regions that affect one or few genes involved in the trait expression.

In this scenario, the phenotypic selection most breeders use is not likely to change much. The reason is that even without directly manipulating the genotype, the breeder tries to identify plants/progenies with the best phenotypes for the target environment of the selection, exploiting the gene/allele action and the factors that regulate their expression. The experience of the past allows the conclusion that phenotypic selection is a method that will always be efficient and therefore always vital. Lee (2006) commented "another certainty is that the phenotypic era of plant breeding is endless and irreplaceable. The real challenge is how to enable phenotypic selection and to make it more effective". This author concludes by quoting the evolutionist Mayr (1997) "the development of the phenotype involves many stochastic processes that preclude a one-to-one relation between genotype and phenotype. This is, of course, precisely why we must accept the phenotype as the object of selection rather than the genotype." In short, the evidence for an evolution of the available concept of genes demonstrated what the experience in the past had shown, i.e., selection should be performed for the phenotype, under the cultivation conditions of the plant in the future.

RESUMO – Recentemente, com o sequenciamento do genoma de algumas espécies, acumularam-se evidências, que para um grande número de caracteres, a regulação e a ação dos genes é bem mais complexa do que se imaginava. O que se discute neste artigo são quais as implicações das informações atuais a respeito do conceito de gene no trabalho dos melhoristas de plantas. Ao que tudo indica o sucesso do emprego das técnicas biotecnológicas será mais difícil do que se supunha. As evidências da evolução do conceito de gene disponível comprovaram o que a experiência do passado vem mostrando, ou seja, a seleção deve ser realizada pelo fenótipo nas condições em que a planta será cultivada no futuro. É necessário continuar o treinamento dos melhoristas de plantas, tendo como foco a seleção fundamentada no fenótipo em experimentos com a maior precisão possível.

Palavras-chave: *melhoramento de plantas, genética, marcador molecular, transgênico.*

REFERENCES

- AG Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. **Nature** **408**: 796-815.
- Bernardo R (2002) **Breeding for quantitative traits in plants**. Stemma Press, Minnesota, 369p.
- Bernardo R (2008) Molecular markers and selection for complex traits in plants: learning from the last 20 years. **Crop Science** **48**: 1649-1664.
- Borém A and Caixeta ET (2009) **Marcadores moleculares**. Editora UFV, Viçosa, 532p.
- Burns GW (1983) **An introduction to heredity**. Macmillan Publishing, New York, 515p.
- Collins F, Grenn ED, Guttmacher AE and Guyer MS (2003) A vision for the future of genomics research. A blueprint for the genomic era. **Nature** **422**: 835-847.
- Darlington CD and Mather K (1950) **The elements of genetics**. George Allen, London, 446p.
- Dudley JW and Lambert RJ (2004) 100 generations of selection for oil and protein in corn. In Janick J (ed) **Plant breeding reviews**. John Wiley & Sons, New York, v. 24, part 1, p. 79-110.
- Duvick DN, Smith JSC and Cooper M (2004) Long-term selection in a commercial hybrid maize breeding program. In Janick J (ed) **Plant breeding reviews**. John Wiley & Sons, New York, v. 24, part 2, p. 109-152.
- Eathington SR, Crosbie TM, Edwards, MD, Reiter RS and Bull JK (2007) Molecular markers in a commercial breeding program. **Crop Science** **47**: 154-163.
- Elseth GD and Baumgardner KD (1984) **Genetics**. Addison-Wesley Publishing, Reading, 780p.
- Gerstein MB, Bruce C, Rozowsky, JS, Zheng, D, Du J, Korbel, JO, Emanuelsson O, Zhang ZD, Weissman S and Snyder M (2007) What is a gene, post-ENCODE? History and updated definition. **Genome Research** **17**: 669-681.
- Green JM (2009) Evolution of glyphosate-resistant crop technology. **Weed Science** **57**: 108-117.
- Griffiths AJF, Miller JH, Suzuki DT, Lewontin RC and Gelbart WM (2000) **An introduction to genetic analysis**. WH Freeman, New York, 860p.
- Guimarães EP, Ruane J, Scherf BD, Sonnino A and Dargie JD (2007) **Marker-assisted selection: current status and future perspectives in crops, livestock, forestry and fish**. FAO, Roma, 471p.
- Houle D, Govindaraju DR and Omholt S (2010) Phenomics: the next challenge. **Nature Reviews Genetics** **11**: 855-866.
- James C (2009) **Situação global das culturas biotecnológicas/GM comercializadas: 2009**. ISAAA Briefs, Ithaca, NY, 44p.
- Joaquim LM and El Hani CN (2010) A genética em transformação: crise e revisão do conceito de gene. **Scientiae Studia** **81**: 93-128.
- Lee M (2006) The phenotypic and genotypic eras of plant breeding. In Lamkey KR and Lee M **Plant breeding: the Arnel R. Hallauer international symposium**. Blackwell Publishing, Ames, 380p.
- Mayr E (1997) The objects of selection. **Proceedings of the National Academy of Sciences of the United States of America** **94**: 2091-2094.
- Miles JW and Pandey S (2004) Long-term selection in plants in the developing world. In Janick J (ed) **Plant breeding reviews**. John Wiley & Sons, New York, v. 24, part 2, p. 45-88.
- Muir WM, Miles D and Bell AE (2004) Long-term selection for pupal weight in *Tribolium castaneum*. In Janick J (ed) **Plant breeding reviews**. John Wiley & Sons, New York, v. 24, part 2, p. 214-224.
- Orjuela J, Garavito A, Bouniol M, Arbelaez JD, Moreno L, Kimball J, Wilson G, Rami JF, Tohme J, McCouch SR and Lorieux M (2010) A universal core genetic map for rice. **Theoretical and Applied Genetics** **120**: 563-572.
- Pearson H (2006) Genetics: what is a gene? **Nature** **441**: 398-401.
- Qing-Po L and Qing-Zhong X (2006) Genome sequencing and identification of gene function in rice. **Acta Genetica Sinica** **33**: 669-677.

- Ramalho MAP, Santos JB and Pinto CABP (2008) **Genética na agropecuária**. Editora UFLA, Lavras, 464p.
- Scherrer K and Jost J (2007a) The gene and the genon concept: a functional and information-theoretic analysis. **Molecular Systems Biology** 3: 1-11.
- Scherrer K and Jost J (2007b) Gene and genon concept: coding versus regulation. **Theory Bioscience** 126: 65-113.
- Simmonds NW (2000) An informal history of statistics. In Janick J (ed) **Plant breeding reviews**. John Wiley & Sons, New York, v. 17, p. 259-316.
- The Encode Project Consortium (2007) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. **Nature** 447: 799-816.
- Tracy W, Goldman IL, Tefenthaler AE and Schaber MA (2004) Trends in productivity of U.S. crops and long-term selection. In Janick J (ed) **Plant breeding reviews**. John Wiley & Sons, New York, v. 24, part 2, p. 89-108.
- USCB (2008) U.S. Census Bureau. Total midyear population for the world: 1950-2050. Available at <<http://www.census.gov/ipc/www/idb/worldpop.html>> Accessed on Aug. 22, 2009.
- Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, Ballew RM, Huson DH, Wortman JR, Zhang Q, Kodira CD, Zheng XH, Chen L, Skupski M, Subramanian G, Thomas PD, Zhang J, Miklos GLG, Nelson C, Broder S, Clark AG, Nadeau J, McKusick VA, Zinder N, Levine AJ, Roberts RJ, Simon M, Slayman C, Hunkapiller M, Bolanos R, Delcher A, Dew I, Fasulo D, Flanigan M, Florea L, Halpern A, Hannenhalli S, Kravitz S, Levy S, Mobarry C, Reinert K, Remington K, Abu-Threideh J, Beasley E, Biddick K, Bonazzi V, Brandon R, Cargill M, Chandramouliswaran I, Charlab R, Chaturvedi K, Deng Z, Di Francesco V, Dunn P, Eilbeck K, Evangelista C, Gabrielian AE, Gan W, Ge W, Gong F, Gu Z, Guan P, Heiman TJ, Higgins ME, Ji RR, Ke Z, Ketchum KA, Lai Z, Lei Y, Li Z, Li J, Liang Y, Lin X, Lu F, Merkulov GV, Milshina N, Moore HM, Naik AK, Narayan VA, Neelam B, Nusskern D, Rusch DB, Salzberg S, Shao W, Shue B, Sun Jingtao, Wang ZY, Wang A, Wang X, Wang J, Wei MH, Wides R, Xiao C, Yan C, Yao A, Ye J, Zhan M, Zhang W, Zhang H, Zhao Q, Zheng L, Zhong F, Zhong W, Zhu SC, Zhao S, Gilbert D, Baumhueter S, Spier G, Carter C, Cravchik A, Woodage T, Ali F, An H, Awe A, Baldwin D, Baden H, Barnstead M, Barrow I, Beeson K, Busam D, Carver A, Center A, Cheng ML, Curry L, Danaher S, Davenport L, Desilets R, Dietz S, Dodson K, Doup L, Ferreira S, Garg N, Gluecksmann A, Hart B, Haynes J, Haynes C, Heiner C, Hladun S, Hostin D, Houck J, Howland T, Ibegwam C, Johnson J, Kalush F, Kline L, Koduru S, Love A, Mann F, May D, McCawley S, McIntosh T, McMullen Ivy, Moy M, Moy L, Murphy B, Nelson K, Pfannkoch C, Pratts E, Puri V, Qureshi H, Reardon M, Rodriguez R, Rogers YH, Romblad D, Ruhfel B, Scott R, Sitter C, Smallwood M, Stewart E, Strong R, Suh E, Thomas R, Tint NN, Tse S, Vech C, Wang G, Wetter J, Williams S, Williams M, Windsor S, Deen EW, Wolfe K, Zaveri J, Zaveri K, Abril JF, Guigó R, Campbel MJ, Sjolander KV, Karlak B, Kejariwal A, Mi H, Lazareva B, Hatton T, Narechania A, Diemer K, Muruganujan A, Guo N, Sato S, Bafna V, Istrail S, Lippert R, Schwartz R, Walenz B, Yooseph S, Allen D, Basu A, Baxendale J, Blick L, Caminha M, Carnes-Stine J, Caulk P, Chiang YH, Coyne M, Dahlke C, Mays AD, Dombroski M, Donnelly M, Ely D, Esparham S, Fosler C, Gire H, Glanowski S, Glasser K, Glodek A, Gorokhov M, Graham K, Gropman B, Harris M, Heil J, Henderson S, Hoover J, Jennings D, Jordan C, Jordan J, Kasha J, Kagan L, Kraft C, Levitsky A, Lewis M, Liu X, Lopez J, Ma D, Majoros W, McDaniel J, Murphy S, Newman M, Nguyen T, Nguyen N, Nodell M, Pan S, Peck J, Peterson M, Rowe W, Sanders R, Scott J, Simpson M, Smith T, Sprague A, Stockwell T, Turner R, Venter E, Wang M, Wen M, Wu D, Wu M, Xia A, Zandieh A and Zhu X (2001) The sequence of the human genome. **Science** 291: 1304-1351.
- Waddington CH (1956) **An introduction to modern genetics**. George Allen, London, 483p.
- Wain HM, Bruford EA, Lovering RC, Lush MJ, Wright MW and Povey S (2002) Guidelines for Human Gene. **Genomics** 79: 464-470.
- Watson JD and Crick FHC (1953) Molecular structures of nucleic acids. A structure for Deoxyribose Nucleic Acid. **Nature** 171: 737-738.