

Functional significance of genetic polymorphisms

Yingqing LU (✉)

State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China

© Higher Education Press and Springer-Verlag 2009

Abstract Natural genetic polymorphisms are gifts from nature and sources of variations at all levels. The post-genomic era permits new perspectives on interpreting genetic polymorphisms and also poses challenges for scientists to answer system questions. Rather than a comprehensive coverage of genetic polymorphisms up-to-date, this review attempts to present some of the results in a somewhat coherent manner to highlight the needs and potentials of pertinent studies.

Keywords omics, genetic polymorphism, metabolic pathway, functional significance, neutral status

1 Introduction

Developments in genomics have been creating new paradigms in biological investigations. Two marching fronts—technical advances in various omics, and functional discoveries and confirmations of individual genes—have moved towards each other, generating new insights and exciting syntheses in biology. One important expansion in the new syntheses is on searching the functional significance of genetic polymorphisms. Although the expansion is still at its infancy, it holds great promise of integrating many branches of biology due to its potential impact on biochemistry, physiology, ecology, and evolutionary biology. Genetic polymorphisms encompass all kinds of variations present among homological DNA sequences, taking the forms of single nucleotide substitution polymorphisms (SNPs), nucleotide deletions or insertions (indels), varying tandem repeats, gene duplications, nucleotide rearrangements, and the presence or absence of transposable elements, *etc.* The proportion of genomes having these variations may not be large, but much of the functional diversity and adaptation may well be founded on them. This review will summarize some of the advances made in recent years.

2 Prevalent polymorphisms and functional consequences

Genetic polymorphisms are a common feature of genomes. Over one million SNPs were found among the 20 accessions of the *Arabidopsis thaliana* genome (Clark et al., 2007), showing a level of natural polymorphisms that is likely not the extreme on the natural polymorphic spectrum of a species. Also in *A. thaliana*, a cDNA survey of single-feature-polymorphism (SFP) exhibits an uneven distribution with a significant elevation of SFP around centromeres and the heterochromatic knob region (Borevitz et al., 2007). At individual loci, genetic polymorphisms span all kinds of systems but are particularly prevalent in various recognition systems (Hedrick, 2006; Clark et al., 2007; Emerson et al., 2008). For instance, the extreme allelic polymorphism has been seen at the self-incompatibility locus (the *S* locus) in plants (Richman, 2000; Lu, 2002) and the major historical compatibility loci (MHC) in animals (Klein et al., 2007; Solberg et al., 2008). Although these extreme polymorphisms have been typically ascribed to balancing selection, additional influencing factors have been implicated by empirical data. For instance, in two self-incompatible plants, *Solanum carolinense* and *Physalis longifolia*, the former species maintains fewer and more divergent alleles at the *S* locus than the latter (Richman et al., 1995; Lu, 2002). It was shown that in addition to the balancing selection, seed dispersal mechanisms (Lu, 2006) and an ancient bottleneck (Paape et al., 2008) may further create variations in the allele number and allelic divergence between species. The functional significance of polymorphisms has been explored more with the MHC system in animals, particularly in humans (Milinski, 2006; Traherne, 2008). Heterozygous or specific MHC class II alleles may increase pathogen resistance (reviewed by Bernatchez and Landry, 2003). Most likely surviving three-spined stickleback fish during heat stress are those with a moderate number of MHC class II alleles (Wegner et al., 2008), implying that selective processes additional to balancing selection may modify genetic polymorphism at the recognition loci.

Despite various mechanisms related to genetic polymorphisms, data (Smith and Eyre-Walker, 2002) also suggest that some polymorphisms may have experienced temporal windows to maintain their neutral status after emergences. The functional significance of these tentatively neutral polymorphisms may be exploited under certain circumstances to benefit their carriers, hence increasing their chances of being fixed in the species. This kind of scenario (Fig. 1) might explain the following cases.

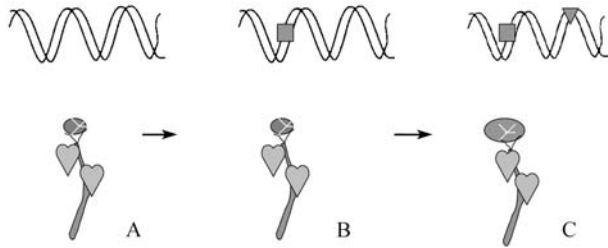


Fig. 1 A scenario that suggests different effects of genetic polymorphisms. A polymorphism (filled square) in DNA sequence has no effect on individual B compared to the wild type A, while an additional polymorphism (filled triangle) has an effect on individual C when compared to the wild type A.

The DNA repair protein O⁶-methylguanine DNA-methyltransferase (MGMT) has polymorphic variants of at least nine amino-acid substitutions in human populations, but the effects of some of the polymorphisms (e.g., Leu84Phe or Ile143Val) are subtle (Bugni et al., 2007). Nonetheless, others of the MGMT polymorphisms (Ile143Val and Lys178Arg) may differ at the expression level (Margison et al., 2005) and, further, patients carrying Lys178Arg polymorphism show a reduced risk for lung cancer (Crosbie et al., 2008), suggesting that individuals carrying different MGMT polymorphisms develop cancers at different rates. The data highlighted the point that not all polymorphisms have the same functional effect, and some may maintain more or less a neutral status.

Besides the expression level, the spatial expression pattern may also be altered by polymorphisms. During photosynthesis, the ultimate CO₂ fixation takes place in the mesophyll cells of C₃ plants, while C₄ plants first capture CO₂ in the mesophyll cells and then fix it at a high concentration in the bundle sheath cells to improve the photosynthetic rate (Fig. 2). This divergence in CO₂ fixation is assisted by the different gene expressions between mesophyll and bundle sheath cells (Nelson and Dengler, 1992).

The mesophyll-specific expression of phosphoenolpyruvate carboxylase is determined by the gene's promoter module, mesophyll expression module 1 (MEM1), in *Flaveria*. It has been shown that exchanges of the two MEM1 polymorphisms between C₄ plant *F. trinervia* and C₃ plant *F. pringlei* were necessary and sufficient for

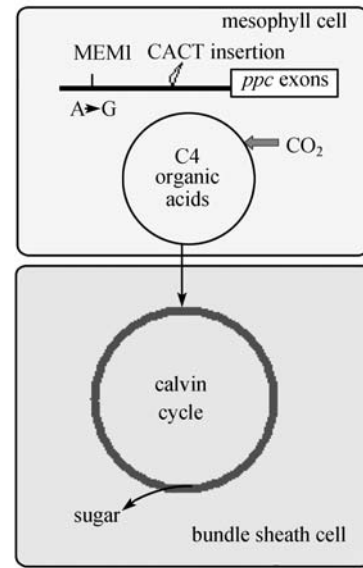


Fig. 2 Photosynthesis of C₄ pathway showing two promoter polymorphisms (a 4-nt insertion and an A-G transition) of the phosphoenolpyruvate carboxylase gene (*ppc*), which are capable of switching C₃ into C₄ pathway in *Flaveria*. MEM1: mesophyll expression module 1.

switching the photosynthetic phenotypes between the species (Akyildiz et al., 2007). This suggests that at least one of the two polymorphisms—an A-to-G substitution and a 4-nucleotide insertion—might not cause functional changes until the next polymorphism occurred.

Single-locus DNA polymorphisms are also known to cause about 200 diseases in common livestock (Ibeagha-Awemu et al., 2008), providing numerous cases of how genetic polymorphism affects cellular functions to varying degrees. Since functional assessments of genetic polymorphisms add an important dimension to the understanding of the natural distribution of genetic polymorphisms, more cases are expected to be documented in the future, providing a clearer view of how genetic polymorphisms fare in various molecular systems.

3 Pathway gene polymorphisms and patterns

Polymorphisms in the context of pathways or networks are more revealing. Like the internet transmitting digital information across the web, pathways carry substance flows within and among molecular systems encoded by genome. A metabolic pathway is, in a broad sense, a series of steps involving the generating and breaking-down of substances needed for cellular activities, including those for signal transduction and regulatory circuits. In a narrow sense, metabolic pathways represent those involved in primary and secondary metabolism, providing basic flows

of substances such as sugars, amino acids, lipids, and mesmerizing arrays of secondary compounds. The regulation of substance flow is intuitively a major task for genome to cope with environmental fluctuations as seen in *Pyrococcus furiosus* (Trauger et al., 2008).

Polymorphisms of multiple genes in the glycolytic pathway have been well-known in *Drosophila melanogaster* (Eanes, 1999; Flowers et al., 2007), and some of the enzyme activities are correlated across the pathway (Pecsénye et al., 2004). Also in *Drosophila*, RNAi pathway genes (*R2D2*, *Ago2*, and *Dcr2*) participating in antiviral function displayed high levels of genetic polymorphism (Obbard et al., 2006). Polymorphisms across pathway genes are also present in plants (Hanson et al., 1996; Aguade, 2001; Whitt et al., 2002; Rausher et al., 2008). Still, relevant data are insufficient to draw a general conclusion. We recently surveyed all major genes of the anthocyanin pathway in *Ipomoea purpurea* and found polymorphisms across all of the ten loci, including the regulatory and structural genes of the pathway. Further analysis of the functional significance of the polymorphisms is underway. Until now, how combinations of multiple polymorphisms affect the carriers remains obscure.

As in the cases of SNPs, human systems have provided better-studied examples so far for understanding the functionality of pathway polymorphisms. Insulin-like growth factor-I (IGF-I) is a polypeptide hormone promoting cellular growth and normal development. The insulin/IGF-I signal response pathway may influence life longevity (Tatar et al., 2003). In humans, polymorphisms of the insulin-like growth factor type 1 receptor (IGF-IR) and phosphoinositide 3-kinase genes (PI3KCB) affect the plasma levels of IGF-I (Bonafè et al., 2003). The analysis of a G-to-A transition in the IGF-IR coding region and a T-to-C transition in the promoter region of PI3KCB showed that individuals bearing at least one A allele at the IGF-IR locus (IGF-IR A +) had lower plasma IGF-I levels than the rest of the population as seen in people over 85 years of age. Furthermore, the proportion of IGF-IR/PI3KCB-A +/T+ carriers is significantly higher among long-lived people (Franceschi et al., 2005). Women of IGF-IR A + homozygote have about half the risk of getting low bone mineral density after menopause (Lee et al., 2008). Knowing the functional consequences of human genetic polymorphisms may help develop unprecedented personal health care that will revolutionize modern medicinal practices and eventually benefit human beings.

Investigations of genetic diversity in pathway genes may lay an important foundation for revealing and understanding how substance flows are directed and redirected in cells under various circumstances, particularly during the evolutionary processes of species, to respond to the ever-changing environment. Depending on the context of a pathway, genetic diversity among genes on the pathway may be modified by selection and constraints of various

intensities. For instance, the genetic diversity of the starch pathway of maize may have been significantly reduced by human selection (Whitt et al., 2002). As changes in one pathway may affect related pathways in the same genome and cause perturbations across transcriptome and metabolome (Dauwe et al., 2007), pathway analysis may link different domains of genome and allow interrogations of genomic regulations in response to exterior environmental changes.

4 Genome-wide polymorphisms via omics approaches

The rapid developments in omics approaches appear to hold great potential for biological discoveries so far. With liquid chromatography-time of flight mass spectrometry (LC-QTOF MS), extensive variations in metabolites have been found in the seedlings of *A. thaliana* accessions, and about 75% of the detected mass peaks could be assigned by genomic loci (Keurentjes et al., 2006). The genome-wide approach may help to uncover unknown pathways and potential regulators of the pathways by correlation analysis, as exemplified for the aliphatic glucosinolate formation. A genome-wide approach for a specific tissue or organ may also clarify the relationships between metabolites and phenotypic traits such as yield and harvest index as shown for tomato (Schauer et al., 2006).

The potential of combining genomic data with those of metabolomics (Fiehn, 2002) is only about to emerge because a great number of cases not resolvable with the traditional methods can now be tackled with the omics approaches. For example, the impact of transgenic crops on human health can be evaluated *via* these approaches. Elevated metabolites with known health hazards may then be singled out to ensure food safety. Food nutrition may also be monitored with detailed analyses and high accuracy (Wishart et al., 2008). Many omics approaches have the hope of answering questions on biological developments (Hennig, 2007) after sampling and inherent problems are corrected (Lay et al., 2006). More standard statistical techniques need to be developed to ameliorate errors and improve compatibility across different types of data.

Three major categories of techniques have been in use to identify sequence polymorphisms. One category is sequencing-based techniques including both Sanger sequencing and pyrosequencing (Ronaghi et al., 1996; Hudson, 2008; Morozova and Marra, 2008). The second category is based on microarray techniques that use oligonucleotide labeling and various tiling strategies (Mockler and Ecker, 2005; Miller et al., 2007; Gregory et al., 2008) to get the whole genome arrays (WGAs). For instance, Zhang et al. (2008) combined approaches of WGAs and enzyme methylome to detect polymorphic GC methylation sites across the *Arabidopsis* genome and documented a strong relationship between methylation and

enzyme effect. The third category involves matrix-assisted laser-desorption ionization time-of-flight mass spectrometry (MALDI-TOF MS), which directly measures the physical masses of DNA molecules under size 100 bp but with a high throughput capacity (Corona and Toffoli, 2004; Ragoussis et al., 2006). The effort put into cost reduction and efficiency improvement of the sequencing techniques will harvest the greatest return when genome-level polymorphisms can be routinely monitored at individual levels.

5 Outlook

The rapid developments in high-throughput techniques and data-generating abilities have been challenging biologists, statisticians, and computer scientists to create new paradigms to present and interpret their investigations. Collaborations across disciplines and large research centers will continue to dominate the front scene of biological explorations on genetic polymorphisms and underpin some major discoveries in the future. Individual laboratories may be better positioned in developing insights on small-scale systems that are manageable for one group while waiting for more portable and accessible equipments to become available in the future. Despite various challenges facing scientists, we are currently at the stage of asking more fundamental questions than ever. Newly obtained details, for instance, allow us to explore the meaning of character evolution, while the wholeness of genome makes one wonder whether genome has been evolving merely as a collection of integrated characters (systems) or more than the sum. Knowing the complexity, regularity, and evolution of genome requires a balanced analysis of local and global patterns of a genome. Systems biology, hopefully, will bring in new developments for a deeper understanding of how living beings act and evolve in the ever-changing environment. Such knowledge will potentially promote the welfare of the ecosystems as well as human societies.

Acknowledgements The study was supported by grants from the Chinese Academy of Sciences (No. 171846). The author thanks the reviewers for constructive comments.

References

- Aguade M (2001). Nucleotide sequence variation at two genes of the phenylpropanoid pathway, the FAH1 and F3H genes, in *Arabidopsis thaliana*. *Mol Biol Evol*, 18: 1–9
- Akyildiz M, Gowik U, Engelmann S, Koczor M, Streubel M, Westhoff P (2007). Evolution and function of a cis-regulatory module for mesophyll-specific gene expression in the C-4 dicot *Flaveria trinervia*. *Plant Cell*, 19: 3391–3402
- Bernatchez L, Landry C (2003). MHC studies in nonmodel vertebrates: what have we learned about natural selection in 15 years? *J Evol Biol*, 16: 363–377
- Bonafè M, Barbieri M, Marchegiani F, Olivieri F, Ragno E, Giampieri C, Mugianesi E, Centurelli M, Franceschi C, Paolisso G (2003). Polymorphic variants of insulin-like growth factor I (IGF-I) receptor and phosphoinositide 3-kinase genes affect IGF-I plasma levels and human longevity: cues for an evolutionarily conserved mechanism of life span control. *J Clin Endocrinol Metab*, 88: 3299–3304
- Borevitz J O, Hazen S P, Michael T P, Morris G P, Baxter I R, Hu T T, Chen H, Werner J D, Nordborg M, Salt D E, Kay S A, Chory J, Weigel D, Jones J D G, Ecker J R (2007). Genome-wide patterns of single-feature polymorphism in *Arabidopsis thaliana*. *Proc Natl Acad Sci USA*, 104: 12057–12062
- Bugni J M, Han J, Tsai M, Hunter D J, Samson L D (2007). Genetic association and functional studies of major polymorphic variants of MGMT. *DNA Repair*, 6: 1116–1126
- Clark R M, Schweikert G, Toomajian C, Ossowski S, Zeller G, Shinn P, Warthmann N, Hu T T, Fu G, Hinds D A, Chen H, Frazer K A, Huson D H, Schölkopf B, Nordborg M, Rättsch G, Ecker J R, Weigel D (2007). Common sequence polymorphisms shaping genetic diversity in *Arabidopsis thaliana*. *Science*, 317: 338–342
- Corona G, Toffoli G (2004). High throughput screening of genetic polymorphisms by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. *Comb Chem High Throughput Screen*, 7: 707–725
- Crosbie P A, McGown G, Thorncroft M R, O'Donnell P N, Barber P V, Lewis S J, Harrison K L, Agius R M, Santibáñez-Koref M F, Margison G P, Povey A C (2008). Association between lung cancer risk and single nucleotide polymorphisms in the first intron and codon 178 of the DNA repair gene, O-6-alkylguanine-DNA alkyltransferase. *Int J Cancer*, 122: 791–795
- Dauwe R, Morreel K, Goeminne G, Gielen B, Rohde A, Van Beeumen J, Ralph J, Boudet A M, Kopka J, Rochange S F, Halpin C, Messens E, Boerjan W (2007). Molecular phenotyping of lignin-modified tobacco reveals associated changes in cell-wall metabolism, primary metabolism, stress metabolism and photorespiration. *Plant J*, 52: 263–285
- Eanes W F (1999). Analysis of selection on enzyme polymorphisms. *Annu Rev Ecol Syst*, 30: 301–326
- Emerson J J, Cardoso-Moreira M, Borevitz J O, Long M (2008). Natural selection shapes genome-wide patterns of copy-number polymorphism in *Drosophila melanogaster*. *Science*, 320: 1629–1631
- Fiehn O (2002). Metabolomics-the link between genotypes and phenotypes. *Plant Mol Biol*, 48: 155–171
- Flowers J M, Sezgin E, Kumagai S, Duvernell D D, Matzkin L M, Schmidt P S, Eanes W F (2007). Adaptive evolution of metabolic pathways in *Drosophila*. *Mol Biol Evol*, 24: 1347–1354
- Franceschi C, Olivieri F, Marchegiani F, Cardelli M, Cavallone L, Capri M, Salvioli S, Valensin S, Benedictis G D, Iorio A D, Caruso C, Paolisso G, Monti D (2005). Genes involved in immune response/inflammation, IGF1/insulin pathway and response to oxidative stress play a major role in the genetics of human longevity: the lesson of centenarians. *Mech Ageing Dev*, 126: 351–361
- Gregory B D, Yazaki J, Ecker J R (2008). Utilizing tiling microarrays for whole-genome analysis in plants. *Plant J*, 53: 636–644
- Hanson M A, Gaut B S, Stec A O, Fuerstenberg S I, Goodman M M, Coe

- E H, Doebley J F (1996). Evolution of anthocyanin biosynthesis in maize kernels: The role of regulatory and enzymatic loci. *Genetics*, 143: 1395–1407
- Hedrick P W (2006). Genetic polymorphism in heterogeneous environments: the age of genomics. *Annu Rev Ecol Evol Syst*, 37: 69–93
- Hennig L (2007). Patterns of beauty-omics meets plant development. *Trends Plant Sci*, 12: 287–293
- Hudson M E (2008). Sequencing breakthroughs for genomic ecology and evolutionary biology. *Mol Ecol Res*, 8: 3–17
- Ibeagha-Awemu E M, Kgwatalala P, Ibeagha A E, Zhao X (2008). A critical analysis of disease-associated DNA polymorphisms in the genes of cattle, goat, sheep, and pig. *Mamm Genome*, 19: 226–245
- Keurentjes J J, Koornneef M, Vreugdenhil D (2008). Quantitative genetics in the age of omics. *Curr Opin Plant Biol*, 11: 123–128
- Keurentjes J J B, Fu J, de Vos C H R, Lommen A, Hall R D, Bino R J, Plas L H W, Jansen R C, Vreugdenhil D, Koornneef M (2006). The genetics of plant metabolism. *Nature Genet*, 38: 842–849
- Klein J, Sato A, Nikolaidis N (2007). MHC, TSP, and the origin of species: from immunogenetics to evolutionary genetics. *Annu Rev Genet*, 41: 281–304
- Lay J O, Borgmann S, Liyanage R, Wilkins C L (2006). Problems with the “omics”. *Trends Anal Chem*, 25: 1046–1056
- Lee D O, Jee B C, Ku S-Y, Suh C S, Kim S H, Choi Y M, Moon S Y, Kim J G (2008). Relationships between the insulin-like growth factor I (IGF-I) receptor gene G3174A polymorphism, serum IGF-I levels, and bone mineral density in postmenopausal Korean women. *J Bone Miner Metab*, 26: 42–46
- Lu Y (2002). Molecular Evolution at the Self-Incompatibility Locus of *Physalis longifolia* (Solanaceae). *J Mol Evol*, 54: 784–793
- Lu Y (2006). Historical events and allelic polymorphism at the gametophytic self-incompatibility locus in Solanaceae. *Heredity*, 96: 22–28
- Margison G P, Heighway J, Pearson S, McGown G, Thornicroft M R, Watson A J, Harrison K L, Lewis S J, Rohde K, Barber P V, O'Donnell P, Povey A C, Santibáñez-Koref M F (2005). Quantitative trait locus analysis reveals two intragenic sites that influence O6-alkylguanine-DNA alkyltransferase activity in peripheral blood mononuclear cells. *Carcinogenesis*, 26: 1473–1480
- Milinski M (2006). The major histocompatibility complex, sexual selection, and mate choice. *Annu Rev Ecol Evol Syst*, 37: 159–186
- Miller M R, Dunham J P, Amores A, Cresko W A, Johnson E A (2007). Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Res*, 17: 240–248
- Mockler T C, Ecker J R (2005). Applications of DNA tiling arrays for whole-genome analysis. *Genomics*, 85: 1–15
- Morozova O, Marra M A (2008). Applications of next-generation sequencing technologies in functional genomics. *Genomics*, 92: 255–264
- Nelson T, Dengler N G (1992). Photosynthetic tissue differentiation in C4 plants. *Int J Plant Sci*, 153: S93–S105
- Obbard D J, Jiggins F M, Halligan D L, Little T J (2006). Natural selection drives extremely rapid evolution in antiviral RNAi genes. *Curr Biol*, 16: 580–585
- Paape T, Iqic B, Smith S D, Olmstead R, Bohs L, Kohn J R (2008). A 15-myr-old genetic bottleneck. *Mol Biol Evol*, 25: 655–663
- Pecsenye K, Komlósi I, Saura A (2004). Heritabilities and additive genetic variances of the activities of some enzymes in *Drosophila melanogaster* populations living in different habitats. *Heredity*, 93: 215–221
- Ragoussis J, Elvidge G P, Kaur K, Colella S (2006). Matrix-assisted laser desorption/ionisation, time-of-flight mass spectrometry in genomics research. *PLoS Genet*, 2: e100
- Rausher M D, Lu Y, Meyer K (2008). Variation in constraint versus positive selection as an explanation for evolutionary rate variation among anthocyanin genes. *J Mol Evol*, 67: 137–144
- Richman A (2000). Evolution of balanced genetic polymorphism. *Mol Ecol*, 9: 1953–1963
- Richman A D, Kao T H, Schaeffer S W, Uyenoyama M K (1995). S-allele sequence diversity in natural populations of *Solanum carolinense* (Horsenettle). *Heredity*, 75: 405–415
- Ronaghi M, Karamohamed S, Pettersson B, Uhlén M, Nyren P (1996). Real-time DNA sequencing using detection of pyrophosphate release. *Anal Biochem*, 242: 84–89
- Schauer N, Semel Y, Roessner U, Gur A, Balbo I, Carrari F, Pleban T, Perez-Melis A, Bruedigam C, Kopka J, Willmitzer L, Zamir D, Fernie A R (2006). Comprehensive metabolic profiling and phenotyping of interspecific introgression lines for tomato improvement. *Nat Biotechnol*, 24: 447–454
- Smith N G C, Eyre-Walker A (2002). Adaptive protein evolution in *Drosophila*. *Nature*, 415: 1022–1024
- Solberg O D, Mack S J, Lancaster A K, Single R M, Tsai Y, Sanchez-Mazas A, Thomson G (2008). Balancing selection and heterogeneity across the classical human leukocyte antigen loci: A meta-analytic review of 497 population studies. *Human Immun*, 69: 443–464
- Tatar M, Bartke A, Antebi A (2003). The endocrine regulation of aging by insulin-like signals. *Science*, 299: 1346–1351
- Traherne J A (2008). Human MHC architecture and evolution: implications for disease association studies. *Int J Immunogenet*, 35: 179–192
- Trauger S A, Kalisak E, Kalisiak J, Morita H, Weinberg M V, Menon A L, Poole F L 2nd, Adams M W, Siuzdak G (2008). Correlating the transcriptome, proteome, and metabolome in the environmental adaptation of a hyperthermophile. *J Proteome Res*, 7: 1027–1035
- Wegner K M, Kalbe M, Milinski M, Reusch T B (2008). Mortality selection during the 2003 European heat wave in three-spined sticklebacks: effects of parasites and MHC genotype. *BMC Evol Biol*, 8: 124
- Whitt S R, Wilson L M, Tenailon M I, Gaut B S, Buckler E S 4th (2002). Genetic diversity and selection in the maize starch pathway. *Proc Natl Acad Sci USA*, 99: 12959–12962
- Wishart D S (2008). Metabolomics: applications to food science and nutrition research. *Trends Food Sci & Tech*, 19: 482–493
- Zhang X, Shiu S, Cal A, Borevitz J O (2008). Global analysis of genetic, epigenetic and transcriptional polymorphisms in *Arabidopsis thaliana*. *PLoS Genet*, 4: e100032