Romanian Biotechnological Letters *Vol. 14, No. 1, 2009, pp. 4087-4094*

REVIEW

Model organisms – a journey from the dawn of biological research to the post-genomic era

Received for publication, October 10, 2008
Accepted, November 20, 2008

DANIEL I. PĂCURAR
Umeå Plant Science Centre, Department of Forest Genetics and Plant Physiology, Swedish University of Agricultural Sciences, 90133 Umeå, Sweden
University of Agricultural Sciences and Veterinary Medicine, 400372 Cluj Napoca, Romania
Tel: +46 (0)90 786 7888, Fax: +46 (0)90 786 8165
e-mail: Daniel.Pacurar@genfys.slu.se

Abstract

The study of models organisms has been the core of biological research for many decades. The amount of information collected from the research conducted on these species in the laboratories across the world is enormous, and the progress made in all fields of biology is highly tributary to them. Historically, the big breakthroughs that represent milestones in biological research go hand in hand with the discovery and development of different model organism species. The aim of this review is to underline the most important contributions brought to science by the 'classical models' and, further, to draw attention to newcomers in the field, that are expected to fill up the gaps and answer the most specific question we face in biology in the post-genomic era.

Keywords: model organism, model system, history, biology, post-genomic era

Introduction

In the eighteenth and nineteenth century, fascinated with diversity, biologists were engaged in studying a huge variety of organisms. It was the time when the Linnaean classification was proposed and when early knowledge about life cycles, evolution by natural selection, cytogenetics, embryology, and the cell theory arose. In the latter half of the nineteenth century, experimentation in plants and animals brought us rudimentary knowledge of metabolism, the mechanisms of embryogenesis, animal and plant physiology, photosynthesis, to name just a few [1]. In the early days of molecular biology (during the 1940´s and 1950´s) the only way of approaching the enormous number of questions about how cells work at the molecular level was to apply the tested tools of reductionism. Biologists reduced the complexity of the tasks in two ways: first, by focussing on a few central molecular mechanisms and second, by choosing the simplest organisms in which to conduct the research [2].

Defining a model system

Although many earlier approaches to biology emphasized taxonomy and the diversity of species, studies of specific biological systems have taken a central role in modern biology [3]. Species that were easy to maintain and manipulate under artificial conditions were selected and came to be viewed as models for broad categories of taxa and sometimes for biological systems in general. Molecular biology studies tended to accentuate the use of model organisms, and many findings from model organisms have frequently been successfully generalized to a surprisingly broad group of species [4]. Thus a model organism represents its

near relatives with respect to many attributes. It might even show facts or principles that are relevant to an entire domain of life.

But how we define a model organism? In the past decades, the term 'model organism' has been narrowly applied to species such as mouse, *Drosophila*, *Saccharomyces cerevisiae*, *Caenorhabditis elegans* or Arabidopsis, which are representative for species of medical, industrial, or agricultural interest. Model organisms are often small, have a short generation time, are easy to work with in large numbers and cheap to maintain, which facilitates experimental laboratory research and high-throughput manipulation and screening. In most cases, economics had an influence on the choice of an organism to study, such as agriculturally important species (e.g. rice) or those related to human health (e.g. the malarial parasite *Plasmodium*). All these species are receiving extraordinary attention from the research community and fall under the broad definition of 'model organism' [5].

Established and emerging model systems across the kingdoms

Model organisms are found among prokaryotes, protists, fungi, plants and animals, and even though they represent only a small fraction of the biodiversity existing on Earth, the data resulting from their study forms the core of our biological knowledge to date [5].

Among the **prokaryotes**, *Escherichia coli* is the classical model of molecular biology, the favourite organism to study transcription, translation, recombination, DNA repair and the regulation of gene activity, a super-model [1].

The human parasite *Plasmodium*, the causative agent of malaria, and the genetic model organisms such as *Dictyostelium* and *Volvox* are maybe the most prominent in the **protists**.

In the **fungi** group, which is thought to include millions of living fungal species, from which only ~80.000 have been described [6], are several models of great importance for genetics like *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe* and *Neurospora crassa*. Baker's yeast (*S. cerevisiae*) was the first eukaryote genome to be fully sequenced [7], and fission yeast (*S. pombe*) was the second fungal genome to be completed [8]. Yeasts, which are unicellular ascomycotan fungi, have proved to be extremely useful organisms for the study of basic eukaryotic phenomena. The fact that fungi are more related to animals than to plants emphasizes the value of these organisms as favorable models for human cells [9].

E. coli and *S. cerevisiae* soon eclipsed the filamentous fungi in many research areas. *S. cerevisiae* became the eukaryotic counterpart to *E. coli*, particularly with the introduction of transformation and genetic engineering techniques. It made strong progress towards solving the problems of genetic recombination, macromolecular synthesis, metabolic regulation, cell cycle and general cell biology [10].

There are ~300.000 known species of land **plants**, but genomic models are only found in a restricted number of species and families. These include *Arabidipsis thaliana*, the classical model of plant genetics and development; grasses such as rice (*Oryza sativa*), corn (*Zea mays*), wheat (*Triticum aestivum*) and *Brachypodium dystachyon*, which is being promoted as an alternative model system to rice, for the temperate cereals and forage grasses [11]; and other economically important crop plants, such as cotton (*Gossypium hirsutum*), soybean (*Glycine max*), tobacco (*Nicotiana tabaccum*). Beside these, which are well known plant models, other plant species have been proposed as models: *Antirrhinum majus* has been used as a model plant for the molecular analysis of transposons [12], asymmetric floral development [13] and floral pigmentation [14]; Micro-Tom, a small tomato cultivar has been proposed as a model plant for tomato genetics [15]; *Nemesia strumosa* Benth, has many characteristics that make it a potential model plant for the study of asymmetric floral development [16]; *Nicotiana attenuata*, has been proposed as a model ecological expression

system [17]; *Craterostigma plantagineum* has contributed significantly to our knowledge of the molecular regulation and physiology of dehydration tolerance in mature leaf tissue [18]; *Medicago truncatula* and *Lotus japonicus* have been selected for molecular genetic analyses of legume biology [19]; *Solanum nigrum* L., a Solanaceous relative of potato and tomato for which many genomic tools are being developed, is presented as a model ecological expression system [20]; *Physcomitrella patens,* in which homologous recombination occurs naturally, has been used as a model system in mosses, for biotechnological aspects [21], and the list can continue.

The most prominent **animal** model organisms are fruit flies (*Drosophila melanogaster*), mouse (*Mus musculus*), and the nematode *Caenorhabditis elegans*. T.H. Morgan chose *Drosophila* to study evolution. Together with his group he showed that the mutations observed in the studied populations obeyed Mendel's rules. Other discoveries made at the time, such as sex linkage, crossing over, and the behaviour of the attached-X chromosomes, led to the final consolidation of the chromosome theory of inheritance [22]. In the 1930s, cytological studies showed the giant chromosomes and banding in larval cells that were used soon after in cytogenetics work and studies of natural populations [23]. *Drosophila* illustrates that a model system, once developed, offers a rich context in which to interpret both old and new findings.

Before 1900, research on mouse breeding provided many traits for Mendelian analysis. Mouse was seen to become a model on which to study many aspects of human biology and medicine [1].

C. elegans was chosen for its simplicity and experimental tractability, with the initial interest in the development of the nervous system [24]. The deliberate choose of *C. elegans*, as a model was driven by the expectation that work with it can quickly consolidate and extend the knowledge that has previously been obtained from several non-model organisms.

The research on animal models has contributed with nearly all we know about cell biology. On the other hand, the choice of model organisms was not always crowned with success. Over time some organisms have fallen from grace. It is fair mentioning in here the rat (*Rattus norvegicus*) which was often an animal of choice two or three decades ago, but nowadays is less used because its genome cannot tolerate the insertion of foreign DNA near the extent of the mouse genome [2].

A brief timeline

We can distinguish three phases of model organisms' emergence in the last century. In the first phase, (1910-1937) mouse, corn and *Drosophila* are the most prominent. The research community's attention turned in the second phase (1938-1952) to microorganisms such as *Neurospora*, *A. nidulans*, *S. cerevisiae* and *E. coli*. After 1960, geneticists showed interest in animal and plant viruses, cultured mammalian cells and multicellular organisms. The most prominent in the last category are *C. elegans* and *A. thaliana* [1].

Arabidopsis became a model plant for molecular studies after a longer period of domestication, and once a stronger interest in multicellular life forms developed. Arabidopsis, like *C. elegans*, was developed as a model system in two steps. The first faze of its acceptance by the research community was promoted by its small size, the variability of natural isolates, easy genetics, a prolific seed set and the promise of developing mutagenic techniques, mutant collections and genetic maps [25]. Although the early work with Arabidopsis failed to raise interest for many people, its definitive acceptance came with the recognition of its small genome [26]. This feature raised the hope that molecular techniques could be more easily applied to Arabidopsis than to most of the other plants, which have much larger genomes,

with considerable DNA duplication and non-coding DNA. Early work on Arabidopsis was done with mutants that affected development and flowering [27]. The discovery that Arabidopsis could be transformed with DNA delivered by *Agrobacterium tumefaciens* [28], prepared it for molecular studies. When its complete genome sequence was published in 2000 [29] many of the genes could be classified with respect to their homology to those of other organisms. Having this valuable information in hand the research community could start complex studies of metabolism, environmental interactions, development, signalling and regulatory pathways, light responses, membrane transport, DNA synthesis and repair. Many of these directions of research either took on new life or now could begin. Apparently a latecomer like Arabidopsis could fully profit from previously research on other model organisms and quickly became a sophisticated model itself for the plant kingdom [1].

New challenges

Now, an increasingly important and practical issue to address is how we can use these model organisms, for instance, to study and quantify the impact of global warming on biodiversity and evolution. Many emerging model systems (e.g. *Nematostella*, *Saccoglossus*, *Populus* and *Aquilegia*, among others) have the potential to serve as 'sentinel' or indicator species because their sensitive biology can provide feedback on changing habitats. They can be used to monitor these areas through their presence or absence, thus indicating certain environmental conditions such as soil type, high levels of pollution, or, in rivers, low levels of dissolved oxygen. *Populus*, for example, is a useful model to study responses to climate change. In addition, with the recent completion of the *Populus* genome sequence [30] it might be useful for such applications as carbon sequestration, bioremediation and even biofuel [31].

On the way to humans as the ultimate model

Over time the number of species considered models have risen and fallen and is practically impossible to refer in here to all aspects of this extremely vast subject. From case to case there has been consensus and controversy about their importance. Still with all the enormous progress made in the field, scientists believe that we need maybe 20-30 years to solve all the fundamental pathways, structures and mechanisms of the lower model organisms [32]. Therefore it is certain that the current model organisms, and several new others, will populate laboratories, even if the list of their application changes. The need for model organisms will only diminish when most of the fundamental mechanisms of biology have been solved. One might argue that with the growing maturity of *in silico* and stem-cell-based techniques their role and importance will soon decrease.

The regenerative capacity of stem cells offers unprecedented opportunities for developing medical therapies for debilitating diseases and represents a source of new, healthy tissue to treat or replace diseased or injured human organs [33]. They are an excellent model that will help us to understand how they transform into the myriad of specialized cells that make us what we are. Their potential in medicine is enormous, and the advances made in the field, since the first human embryonic stem cell culture was developed in 1998 [34], have exploded in the recent years being reflected in a huge number of publications. There are dozens of excellent reviews covering all aspects of this fast growing field, from therapeutic potential and scientific achievements to ethical issues and legal aspects [e.g. 33, 35, 36, 37, 38, 39], that will be not further discussed here. Spite, beside other limitations [reviewed in e.g. 33], yet stem cells and human tissue are not sufficient to model more complex organ systems, such for instance immune system. We need to decipher fundamental processes first in the simplest

models like yeast, worms, flies, etc, then in the mouse, and only when we will know enough we can start to fully decipher them in humans [2].

A significant amount of research conducted on described models (and many more others) was driven by the hope that we can understand and/or cure all the diseases and anomalies in humans. Now with the sequence of the human genome in hand [40] scientists hope to do more and better to approach this goal. Since 2003, when Human Genome Project released a composite human genome sequence containing DNA from many people, DNA-sequencing technology has exploded, offering the possibility for some companies to plan sequencing 1000 human genomes this year, and 20.000 in 2010, at a cost of around $1000 each [41]. This is a tremendous drop in costs since June 2007 when James Watson was the first person to have his own complete genome sequenced at a cost of $ 1-1.5 milion.

We are approaching, much faster that we thought, an era when each of us can have the genome sequenced fast and cheap. Comparing whole genomes might reveal the genes underlying an illness making the development of finely tuned and specific drugs more cost efficient. Such studies will gain power as more genomes are sequenced, and each patient will become its own model in the process of developing specific drugs, thus reducing the risk of adverse reactions, with the help of the information revealed by its own genome sequence. The potential offered together by the DNA-sequence related technologies and the stem cell based therapies will definitely change the face of medicine.

Concluding remarks

Model organisms achieved their status only after they have been discovered or chosen for use in a particular research programme – they all began as model systems. Research that was initiated in many organisms has finally converged to a few models. Of course, many organisms continued to be used in basic science departments of physiology, neurobiology, plant sciences, ecology and evolution, in an integrative approach. Diverse research organisms are at present used in the institutions of applied research. In university departments of plant pathology, many bacterial, fungal, insect and nematode species continue to be investigated due to the economic importance of agricultural crops. Industry is also supporting microbial research with applications in food processing and preservation, animal and plant pathogens, pharmacology, exotic biochemicals and fermentation products [1].

Whereas in the past development of a new model system required enormous investment, with the recent advances including the decreasing cost of DNA sequencing and the power of reverse genetics to study gene function, together with improved bioinformatics tools, the process is greatly facilitated. However, the development of a new model system, necessitates a significant investment both in terms of money and time; thus, it is crucial to make careful choices of the most appropriate organism to study. Although we expect major model systems to possess the full complement of molecular tools, model systems that are especially developed to address evolutionary and ecological questions might not possess this entire spectrum of resources. Depending on the questions being asked, some tools are more relevant than others [31].

In the recent years, the development of genetic engineering techniques has brought new challenges in functional biology by promoting interest in sequencing entire genomes. On the way to the human genome [40], or shortly after, genomes of almost all the model organisms discussed above were sequenced: *S. cerevisiae*, 1996 [7]; *E. coli*, 1997 [29]; *C. elegans*, 1998 [42]; *A. thaliana*, 2000 [28]; *D. melanogaster*, 2000 [43]; mouse, 2002 [44]; *O. sativa L. ssp. indica*, 2002 [45]; *O. sativa L. ssp. japonica*, 2002 [46]; *Neurospora crassa*, 2003 [47], *Populus trichocarpa*, 2006 [30]. Furthermore, since the coordinated *International Rice*

Genome Sequencing Project (IRGSP) began in 2000, several other crop plant genome projects have been started. These include *Lotus japonicus* and *Medicago truncatula*, representatives of the legume group that harbour symbiotic nitrogen-fixing bacteria in specialized root nodules; tomato, representing the solanaceous group of important vegetable species; and canola, an oil-producing member of the *Brassica* family [48]. A analyse of the first draft covering 91.3% of the gene space of *L. japonicus* genome has recently being published [49] and the *Medicago truncatula* genome is expected to be soon completed.

Now, with the help of comparative genomics we can test how representative these models organisms are by comparing both the sequence and the function of the genes of one organism to those of any other. Finally, in the twenty-first century, by employing comparative genomics, we can return to the study of complexity and diversity of life forms that was so prevalent at the end of the nineteenth century. It is difficult to imagine how we would reach this point without the intense study of 'genetically domesticated' model organisms in the twentieth century [1]. A focus on model organisms will continue but with a new approach of interdisciplinary exchange, fusions of fields and cross-cutting initiatives [5]. Still, in the post-genomic era, one of the most successful ways of making consistent progress in particular fields of biology is by choosing a good model system on which to focus experimental efforts.

Acknowledgements

I would like to thank Monica L. Păcurar for critical reading of the manuscript.

Bibliography

1. R.H. DAVIS, *Nature Rev. Genet.* 5, 69-77 (2004).
2. EMBO *reports*, 9 (8), 717-720 (2008).
3. D.E. KOSHLAND, *Science* 240, 1385 (1988).
4. M.G. KIDWELL & M.B. EVGENIEV, *Genetica* 107, 103-111 (1999).
5. H.S. BLAIR, *Nature* 3, 838-849 (2002).
6. P.M. KIRK, P.F. CANNON, J.K. DAVID, J.A. STALPERS, Dictionary of Fungi, *CAB International*, Surrey, UK, (2001).
7. A. GOFFEAU, B.G. BARRELL, H. BUSSEY, R.W. DAVIS, B. DUJON, H. FELDMANN, F. GALIBERT, J.D. HOHEISEL, C. JACQ, M. JOHNSTON, E.J. LOUIS, H.W. MEWES, Y. MURAKAMI, P. PHILIPPSEN, H. TETTELIN, S.G. OLIVER, *Science* 274, 546-567 (1996).
8. V. WOOD, R. GWILLIAM, M.A. RAJANDREAM, M. LYNE, R. LYNE, A. STEWART, J. SGOUROS, N. PEAT, J. HAYLES, S. BAKER, D. BASHAM, S. BOWMAN, K. BROOKS, D. BROWN, S. BROWN, T. CHILLINGWORTH, C. CHURCHER, M. COLLINS, R. CONNOR, A. CRONIN, P. DAVIS, T. FELTWELL, A. FRASER, S. GENTLES, A. GOBLE, N. HAMLIN, D. HARRIS, J. HIDALGO, G. HODGSON, S. HOLROYD, HORNSBY T, HOWARTH S, HUCKLE EJ, HUNT S, JAGELS K, JAMES K, JONES L, JONES M, LEATHER S, S. MCDONALD, J. MCLEAN, P. MOONEY, S. MOULE, K. MUNGALL, L. MURPHY, D. NIBLETT, C. ODELL, K. OLIVER, S. O'NEIL, D. PEARSON, M.A. QUAIL, E. RABBINOWITSCH, K. RUTHERFORD, S. RUTTER, D. SAUNDERS, K. SEEGER, S. SHARP, J. SKELTON, M. SIMMONDS, R. SQUARES, S. SQUARES, K. STEVENS, K. TAYLOR, R.G. TAYLOR, A. TIVEY, S. WALSH, T. WARREN, S. WHITEHEAD, J. WOODWARD, G. VOLCKAERT, R. AERT, J. ROBBEN, B. GRYMONPREZ, I. WELTJENS, E. VANSTREELS, M. RIEGER, M. SCHÄFER, S. MÜLLER-AUER, C. GABEL, M. FUCHS, A. DÜSTERHÖFT, C. FRITZC, E. HOLZER, D. MOESTL, H. HILBERT, K. BORZYM, I. LANGER, A. BECK, H. LEHRACH, R. REINHARDT, T.M. POHL, P. EGER, W. ZIMMERMANN, H. WEDLER, R. WAMBUTT, B. PURNELLE, A. GOFFEAU, E. CADIEU, S. DRÉANO, S. GLOUX, V. LELAURE, S. MOTTIER, F. GALIBERT, S.J. AVES, Z. XIANG, C. HUNT, K. MOORE, S.M. HURST, M. LUCAS, M. ROCHET, C. GAILLARDIN, V.A. TALLADA, A. GARZON, G. THODE, R.R. DAGA, L. CRUZADO, J. JIMENEZ, M. SÁNCHEZ, F. DEL REY, J. BENITO, A. DOMÍNGUEZ, J.L. REVUELTA, S. MORENO, J. ARMSTRONG, S.L. FORSBURG, L. CERUTTI,

T. LOWE, W.R. MCCOMBIE, I. PAULSEN, J. POTASHKIN, G.V. SHPAKOVSKI, D. USSERY, B.G. BARRELL, P. NURSE, *Nature* 415, 871-880 (2002).
9. I.J. VAN DER KLEI, M. VEENHUIS, *Biochimica et Biophysica Acta* 1763, 1364-1373 (2006).
10. D.J. LEW, T. WEINERT & J.R. PRINGLE, The Molecular Biology of the Yeast *Saccharomyces*, Vol. 3 *Cell Cycle and Cell Biology*, *Cold Spring Harbor Laboratory Press*, Cold Spring Harbor, New York, (1997).
11. D.I. PĂCURAR, H. THORDAL-CHRISTENSEN, K.K. NIELSEN, I. LENK, *Transgenic Res.*, 17(5), 965-975 (2008).
12. E.S. COEN, R. CARPENTER, C. MARTIN, *Cell* 47, 285-296 (1986).
13. R. CARPENTER & E.S. COEN, *Genes Dev.* 4, 1483-1493 (1990).
14. C. MARTIN, A. PRESCOTT, S. MACKAY, J. BARTLETT, E. VRIJLANDT, *Plant J.* 1, 37-49 (1991).
15. R. MESSNER, Y. JACOBSON, S. MELEMED, S. LEVYATUV, G. SHALEV, A. ASHRY, Y. ELKIND, A. LEVY, *The Plant Journal* 12 (6), 1465-1472 (1997).
16. M.L. CUI & H. EZURA, *Plant Science* 165, 863-870 (2003).
17. T. KRÜGEL, M. LIM, K. GASE, R. HALITSCHKE AND I. T. BALDWIN, *Chemoecology* 12, 177-183 (2002).
18. O. TOLDI, S. TÓTH, T. PÓNYI, P. SCOTT, *Plant Cell Rep* 21, 63-69 (2002).
19. M.J. HARRISON, *Trends in plants science* 5 (10), 414-415 (2000).
20. D.D. SCHMIDT, A. KESSLER, D. KESSLER, S. SCHMIDT, M. LIM, K. GASE AND I.T. BALDWIN, *Molecular Ecology* 13, 981-995 (2004).
21. D.D. SCHAEFER, *Annu. Rev. Plant Biol.* 53, 477-501 (2002).
22. T.H. MORGAN, A.H. STURTEVANT, H.J. MULLER & C.B. BRIDGES, *The Mechanism of Mendelian Heredity* (1915). Reprinted with an introduction by Garland Allen, *Johnson Reprint Corporation*, New York (1972).
23. R.E. KOHLER, *Drosophila Genetics and the Experimental Life* 43, *Univ. Chicago Press*, Chicago (1994).
24. R. ANKENY, *Nature Rev.Genet.* 2, 474-479 (2001).
25. G.P. REDEI, Ann. Rev. Genet. 9, 111-127 (1975).
26. L.S., LEUTWILER, B.R., HOUGHEVANS & E.M. MEYEROWITZ, *Mol. Gen. Genet.* 194, 15-23 (1984).
27. C. SOMMERVILLE & M. KOORNNEEF, *Nature Rev. Genet.* 3, 883-889 (2002).
28. K.A. Feldmann & M.D. Marks, *Mol. Gen. Genet.* 208, 1-9 (1989).
29. THE ARABIDOPSIS GENOME INITIATIVE, *Nature* 408, 796-815 (2000).
30. G.A. TUSKAN, S.DIFAZIO, S. JANSSON, J. BOHLMANN, I. GRIGORIEV, U. HELLSTEN, N. PUTNAM, S. RALPH, S. ROMBAUTS, A. SALAMOV, J. SCHEIN, L. STERCK, A. AERTS, R. R. BHALERAO, R. P. BHALERAO, D. BLAUDEZ, W. BOERJAN, A. BRUN, A. BRUNNER, V. BUSOV, M. CAMPBELL, J. CARLSON, M. CHALOT, J. CHAPMAN, G.L. CHEN, D. COOPER, P.M. COUTINHO, J. COUTURIER, S. COVERT, Q. CRONK, R. CUNNINGHAM, J. DAVIS, S. DEGROEVE, A. DÉJARDIN, C. DEPAMPHILIS, J. DETTER, B. DIRKS, I. DUBCHAK, S. DUPLESSIS, J. EHLTING, B. ELLIS, K. GENDLER, D. GOODSTEIN, M. GRIBSKOV, J. GRIMWOOD, A. GROOVER, L. GUNTER, B. HAMBERGER, B. HEINZE, Y. HELARIUTTA, B. HENRISSAT, D. HOLLIGAN, R. HOLT, W. HUANG, N. ISLAM-FARIDI, S. JONES, M. JONES-RHOADES, R. JORGENSEN, C. JOSHI, J. KANGASJÄRVI, J. KARLSSON, C. KELLEHER, R. KIRKPATRICK, M. KIRST, A. KOHLER, U. KALLURI, F. LARIMER, J. LEEBENS-MACK, J.C. LEPLÉ, P. LOCASCIO, Y. LOU, S. LUCAS, F. MARTIN, B. MONTANINI, C. NAPOLI, D.R. NELSON, C. NELSON, K. NIEMINEN, O. NILSSON, V. PEREDA, G. PETER, R. PHILIPPE, G. PILATE, A. POLIAKOV, J. RAZUMOVSKAYA, P. RICHARDSON, C. RINALDI, K. RITLAND, P. ROUZÉ, D. RYABOY, J. SCHMUTZ, J. SCHRADER, B. SEGERMAN, H. SHIN,11 A. SIDDIQUI, F. STERKY, A. TERRY, C.J. TSAI, E. UBERBACHER, P. UNNEBERG, J. VAHALA, K. WALL, S. WESSLER, G. YANG, T. YIN, C. DOUGLAS, M. MARRA, G. SANDBERG, Y. VAN DE PEER, D. ROKHSAR9, *Science* 313 (5793), 1596-1604 (2006).
31. A. ABZHANOV, C.G. EXTAVOUR, A. GROOVER, S.A. HODGES, H.E. HOEKSTRA, E.M. KRAMER AND A. MONTEIRO, *Trends in Genetics*, 24 (7), 353-360 (2008).
32. S. FIELDS, M. JOHNSTON M, *Science* 307, 1885-1886 (2005).
33. N. SHANTHLY, M.R. ARUVA, K. ZHANG, B.MATHEW, M.L. THAKUR, *QJ Nucl Med Mol Imaging* 50, 205-216 (2006).

34. J.A. THOMSON, J.I. ELDOR, S.S. SHAPIRO, M.A. WAKNITZ, J.J. SWIERGIEL, V.S. MARSHALL, J.M. JONES, *Science* 282, 1145-1147 (1998).
35. I.L. WEISSMANN, *Cell* 400, 157-168 (2000).
36. A. COLMAN & J.C. BURLEY, *EMBO Rep.* 2(1), 2-5 (2001).
37. K.S. TWEEDELL, *Curr Stem Cell Res Ther.* 3(3), 151-162 (2008).
38. S. NISHIKAWA, R.A. GOLDSTEIN, C.R. NIERRAS, *Nat Rev Mol Cell Biol*. 9(9), 725-729 (2008).
39. C.T. SCOTT, R.A. REIJO PERA, *Hum Mol Genet.* 17:3-9 (2008).
40. J.C. VENTER, *Science* 291, 1304-1351 (2001).
41. E.C. HAYDEN, *Nature* news.2008.1151 (2008).
42. *C. ELEGANS* SEQUENCING CONSORTIUM, *Science* 282, 2012-2018 (1998).
43. M.D. ADAMS, S.E. CELNIKER, R.A. HOLT, C.A. EVANS, J.D. GOCAYNE, P.G. AMANATIDES, S.E. SCHERER, P.W. LI, R.A. HOSKINS, R.F. GALLE, R.A. GEORGE, S.E. LEWIS, S. RICHARDS, M ASHBURNER, S.N. HENDERSON, G.G. SUTTON, J.R. WORTMAN , M.D. YANDELL, Q. ZHANG, L.X. CHEN, R.C. BRANDON, Y.H. ROGERS, R.G. BLAZEJ, M. CHAMPE, B.D. PFEIFFER, K.H. WAN, C. DOYLE, E.G. BAXTER, G. HELT, C.R. NELSON, G.L. GABOR, J.F. ABRIL, A. AGBAYANI, H.J. AN, C. ANDREWS-PFANNKOCH, D. BALDWIN, R.M. BALLEW, A. BASU, J. BAXENDALE, L. BAYRAKTAROGLU, E.M. BEASLEY, K.Y. BEESON, P.V. BENOS, B.P. BERMAN, D. BHANDARI, S. BOLSHAKOV, D. BORKOVA, M.R. BOTCHAN, J. BOUCK, P. BROKSTEIN, P. BROTTIER, K.C. BURTIS, D.A. BUSAM, H. BUTLER, E. CADIEU, A. CENTER, I. CHANDRA, J.M. CHERRY, S. CAWLEY, C. DAHLKE, L.B. DAVENPORT, P. DAVIES, B. DE PABLOS, A. DELCHER, Z. DENG, A.D. MAYS, I. DEW, S.M DIETZ, K. DODSON, L.E. DOUP, M. DOWNES, S. DUGAN-ROCHA, B.C. DUNKOV, P. DUNN, K.J. DURBIN, C.C. EVANGELISTA, C. FERRAZ, S. FERRIERA, W. FLEISCHMANN, C. FOSLER, A.E. GABRIELIAN, N.S. GARG, W.M. GELBART, K. GLASSER, A. GLODEK, F. GONG, J.H. GORRELL, Z. GU, P. GUAN, M. HARRIS, N.L. HARRIS, D. HARVEY, T.J. HEIMAN, J.R. HERNANDEZ, J. HOUCK, D. HOSTIN, K.A. HOUSTON, T.J. HOWLAND, M.H. WEI, C. IBEGWA, M. JALALI, F. KALUSH, G.H. KARPEN, Z. KE, J.A. KENNISON, K.A. KETCHUM, B.E. KIMMEL, C.D. KODIRA, C. KRAFT, S. KRAVITZ, D. KULP, Z. LAI, P. LASKO, Y. LEI, A.A. LEVITSKY, J. LI, Z. LI, Y. LIANG, X. LIN, X. LIU, B. MATTEI, T.C. MCINTOSH, M.P. MCLEOD, D. MCPHERSON, G. MERKULOV, N.V. MILSHINA, C. MOBARRY, J. MORRIS, A. MOSHREFI, S.M. MOUNT, M. MOY, B. MURPHY, L. MURPHY, D.M. MUZNY, D.L. NELSON, D.R. NELSON, K.A. NELSON, K. NIXON, D.R. NUSSKERN, J.M. PACLEB, M. PALAZZOLO, G.S. PITTMAN, S. PAN, J. POLLARD, V. PURI, M.G. REESE, K. REINERT, K. REMINGTON, R.D. SAUNDERS, F. SCHEELER, H. SHEN, B.C. SHUE, I. SIDÉN-KIAMOS, M. SIMPSON, M.P. SKUPSKI, T. SMITH, E. SPIER, A.C. SPRADLING, M. STAPLETON, R. STRONG, E. SUN, R. SVIRSKAS, C. TECTOR, R. TURNER, E. VENTER, A.H. WANG, X. WANG, Z.Y. WANG, D.A. WASSARMAN, G.M. WEINSTOCK, J. WEISSENBACH, S.M. WILLIAMS, T. WOODAGE, K.C. WORLEY, D. WU, S. YANG, Q.A. YAO, J. YE, R.F. YEH, J.S. ZAVERI, M. ZHAN, G. ZHANG, Q. ZHAO, L. ZHENG, X.H. ZHENG, F.N. ZHONG, W. ZHONG, X. ZHOU, S. ZHU, X. ZHU, H.O. SMITH, R.A. GIBBS, E.W. MYERS, G.M. RUBIN, J.C. VENTER, *Science* 287, 2185-2195 (2000).
44. MOUSE GENOME SEQUENCING CONSORTIUM, *Nature* 420, 520-562 (2002).
45. J.YU, S. HU, J. WANG, G. KA-SHU WONG, S. LI, B. LIU, Y. DENG, L. DAI, Y. ZHOU, X. ZHANG, M. CAO, J. LIU, J. SUN, J. TANG, Y. CHEN, X. HUANG, W. LIN, C. YE, W. TONG, L. CONG, J. GENG, Y. HAN, L. LI, W. LI, G. HU, X. HUANG, W. LI, J. LI, Z. LIU, L. LI, J. LIU, Q. QI, J. LIU, L. LI, T. LI, X. WANG, H. LU, T. WU, M. ZHU, P. NI, H. HAN, W. DONG, X. REN, X. FENG, P. CUI, X. LI, H. WANG, X. XU, W. ZHAI, Z. XU, J. ZHANG, S. HE, J. ZHANG, J. XU, K. ZHANG, X. ZHENG, J. DONG, W. ZENG, L. TAO, J. YE, J. TAN, X. REN, X. CHEN, J. HE, D. LIU, W. TIAN, C. TIAN, H. XIA, Q. BAO, G. LI, H. GAO, T. CAO, J. WANG, W. ZHAO, P. LI, W. CHEN, X. WANG, Y. ZHANG, J. HU, J. WANG, S. LIU, J. YANG, G. ZHANG, Y. XIONG, Z. LI, L. MAO, C. ZHOU, Z. ZHU, R. CHEN, B. HAO, W. ZHENG, S. CHEN, W. GUO, G. LI, S. LIU, M. TAO, J. WANG, L. ZHU, L. YUAN, H. YANG, *Science* 296 (5565), 79-92 (2002).
46. S.A. GOFF, D. RICKE, T.H. LAN, G. PRESTING, R. WANG, M. DUNN, J. GLAZEBROOK, A. SESSIONS, P. OELLER, H. VARMA, D. HADLEY, D. HUTCHISON, C. MARTIN, F. KATAGIRI, B.M. LANGE, T. MOUGHAMER, Y. XIA, P. BUDWORTH, J. ZHONG, T. MIGUEL, U. PASZKOWSKI, S. ZHANG, M. COLBERT, W.L. SUN, L. CHEN, B. COOPER, S. PARK, T.C. WOOD, L. MAO, P. QUAIL, R. WING, R. DEAN, Y. YU, A. ZHARKIKH, R. SHEN, S. SAHASRABUDHE, A. THOMAS, R. CANNINGS, A. GUTIN, D. PRUSS, J. REID, S. TAVTIGIAN, J. MITCHELL, G. ELDREDGE, T. SCHOLL, R.M. MILLER, S. BHATNAGAR, N. ADEY, T.

RUBANO, N. TUSNEEM, R. ROBINSON, J. FELDHAUS, T. MACALMA, A. OLIPHANT, S. BRIGGS, *Science* 296 (5565), 92-100 (2002).

47. J.E. GALAGAN, S.E. CALVO, K.A. BORKOVICH, E.U. SELKER, N.D. READ, D.JAFFE, W. FITZHUGH, L.J. MA, S.SMIRNOV, S.PURCELL, B.REHMAN, T.ELKINS, R. ENGELS, S. WANG, C.B. NIELSEN, J. BUTLER, M.ENDRIZZI, D. QUI, P. IANAKIEV, D.BELL-PEDERSEN, M.A. NELSON, M. WERNER-WASHBURNE, C.P. SELITRENNIKOFF, J.A. KINSEY, E.L. BRAUN, A.ZELTER, U. SCHULTE, G.O. KOTHE, G. JEDD, W. MEWES, C. STABEN, E MARCOTTE, D. GREENBERG, A. ROY, K. FOLEY, J. NAYLOR, N. STANGE-THOMANN, R. BARRETT, S. GNERRE, M. KAMAL, M. KAMVYSSELIS, E. MAUCELI, C. BIELKE, S. RUDD, D. FRISHMAN, S. KRYSTOFOVA, C. RASMUSSEN, R.L. METZENBERG, D.D. PERKINS, S. KROKEN, C. COGONI, G. MACINO, D. CATCHESIDE, W. LI, R.J. PRATT, S.A. OSMANI, C.P.C. DESOUZA, L. GLASS, M.J. ORBACH, J.A. BERGLUND, R.VOELKER, O. YARDEN, M. PLAMANN, S. SEILER, J. DUNLAP, A. RADFORD, R. ARAMAYO, D.O. NATVIG, L.A. ALEX, G. MANNHAUPT, D.J. EBBOLE, M. FREITAG, I. PAULSEN, M.S. SACHS, E.S. LANDER, C. NUSBAUM AND B. BIRREN, *Nature* 422, 859-868 (2003).
48. M. Bevan, *Science* 300, 1514-1515 (2003).
49. S. SATO, Y. NAKAMURA, T. KANEKO, E. ASAMIZU, T. KATO, M. NAKAO, S. SASAMOTO, A. WATANABE, A. ONO, K. KAWASHIMA, T. FUJISHIRO, M. KATOH, M. KOHARA, Y. KISHIDA, C. MINAMI, S. NAKAYAMA, N. NAKAZAKI, Y. SHIMIZU, S. SHINPO, C. TAKAHASHI, T. WADA, M. YAMADA, N. OHMIDO, M. HAYASHI, K. FUKUI, T. BABA, T. NAKAMICHI, H. MORI, S. TABATA, *DNA Res*. 15(4), 227-239 (2008).