Inside JEB is a twice monthly feature, which highlights the key developments in the *Journal of Experimental Biology*. Written by science journalists, the short reports give the inside view of the science in JEB.

Inside JEB

THE GENOMIC REVOLUTION

The history of biology is peppered with groundbreaking discoveries: Charles Darwin's publication of 'The Origin of Species', the determination of the structure of DNA, and, in 2000, the sequence of the human genome was finally added to this illustrious roll call. Two international teams of scientists announced that they had decoded the blueprint of life, but what exactly did the human genome tell us about life itself? According to Andrew Cossins from the University of Liverpool, 'the basic information that the genome gives us is a parts list', but this 'parts list' came without an 'instruction manual'. However, scientists still know little about how the 20–25 000 protein coding genes on the human genome, which make up the genotype, interact with the environment to generate the observable phenotype that scientists study.

While unravelling this instruction manual seems like a daunting task, Cossins explains that it has the potential to revolutionise the study of animal function. With this in mind, Cossins and George Somero from the Hopkins Marine Station at Stanford University have compiled and edited the reviews published in this issue of *The Journal of Experimental Biology*, with the aim of demonstrating how genomic approaches could be added to the physiologists' toolkit. Written by leading scientists working at the forefront of the genomics revolution, each article discusses concepts, ideas and techniques for understanding how an organism's genotype contributes to the functioning of the whole organism under the complex suite of environmental factors it encounters in its habitat.

DEEP SEQUENCING AND NEW CONCEPTIONS

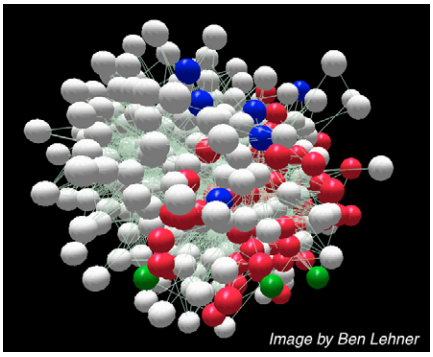
Scientists now have huge amounts of information at their fingertips, but what does it all mean? Piero Carninci (p. 1497) begins the review collection with a discussion on the techniques used by scientists to understand what the genome codes for. To do this, he writes, they need to categorise the transcriptome, the parts of the genome that are transcribed into mRNA; then, list which mRNAs are translated into proteins. One method is to create large libraries of cDNAs – complementary DNA sequences that are synthesised from RNAs – and use these libraries to identify those that correspond to proteins. This work has shown that the transcriptome is a lot more complex than originally thought. While many mRNA sequences are ultimately translated into proteins, some RNA transcripts do not code for proteins, and Carninci explains that these RNA transcripts are likely to play a role in the transcription of protein-coding genes.

While the Human Genome Project gave us a 'parts list', John Quackenbush explains that researchers currently lack a 'circuit diagram' linking all of the newly described genes into functional groups (p. 1507). He writes that powerful new data analysis techniques are needed to pick out which of the many thousands of genes in the genome are important for particular kinds of physiological responses. Computer technology developed as a result of the Human Genome Project, for example statistical techniques can be used to group genes that are controlled together and to model these groups of genes in complicated networks. Quackenbush hopes that these powerful new analytical techniques will ultimately allow scientists to develop predictions on how complex systems composed of many elements can function, for example how a mouse's genotype will affect its immunity.

The technology revolution is not just about advances in data analysis, explains Neil Hall (p. 1518). There is a huge lack of sequence data on non-model species, and advances in sequencing could be used to explore the 'vast microbial diversity in the natural environment'; for example, there could be as many as 10^7 bacterial species in 10 g soil. Hall adds that many prokaryotes are also human pathogens, such as *Plasmodium falciparum* and *Mycobacterium tuberculosis*, so understanding the sequence variation between strains has a clear clinical relevance. Sequencing these smaller, but still important, genomes will become easier with the development of desktop sequencing machines, bringing cheap sequencing technology into university campuses, potentially leading to a 'renaissance' in genome sequencing.

While many researchers are focussed on the coding parts of the genome, only a small fraction of the genome encodes functional proteins. Returning to the question of non-protein-coding DNA, John Mattick focuses on the 98% of the human genome that isn't translated into functional proteins (p. 1526). Organisms are incredibly complex, and the more complicated they are, the larger and more intricate the networks regulating their function have to be. The conventional view is that this complexity is controlled by interactions between proteins and signals

between tissues; however, as organism complexity increases, there is an even greater increase in the complexity of regulation. However, this increased complexity does not seem to have been matched by a significant increase in the number of protein genes encoded by genomes, suggesting that there is an upper limit to how complicated regulation can be. Mattick suggests that this is where the majority of the genome – so-called 'junk' DNA – comes in. Most of it is transcribed into RNA, apparently in a controlled and regulated way. But these RNA transcripts are not translated into proteins, suggesting that they may form another layer of genetic regulation, largely hidden up until now, controlling cell differentiation and development.



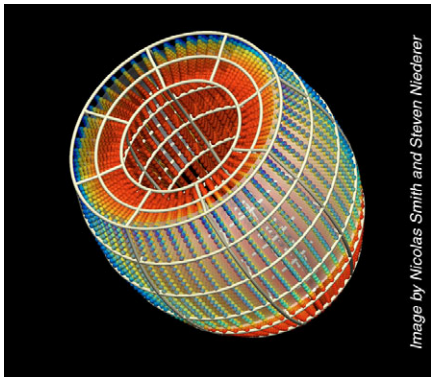
COMPLEX BIOLOGICAL SYSTEMS

Biological systems are remarkably complex, and making sense of the huge volumes of data available has meant that methods that can analyse these vast datasets are becoming increasingly important. Eivind Almaas (p. 1548) describes how network theory can be applied to whole biological systems, for example metabolism, and can help researchers understand the complex interactions between different parts of the system, in space and through time. Networks are represented by nodes linked to each other; in a biological system such as metabolism, the proteins in the system are the network nodes while protein–protein interactions form the links between the nodes. By representing and analysing a complex system such as metabolism as a network scientists can learn how its constituent parts interact to contribute to the function of the whole system. The next step is to combine information from different networks, from gene regulation to metabolism, to understand how whole cells function.

Ben Lehner develops the idea of gene and protein networks and how different types of data can be used together to describe networks and how networks in different species have common basic structures. For example, describing networks that regulate gene transcription in worms and yeast can inform researchers about the equivalent networks in humans (p. 1559). Because humans are so complicated, we know little about how genes interact to produce phenotypes, especially when it comes to hereditary diseases, which result from mutations in many genes. One approach to help understand disease is to use mutants in yeast or worms to systematically investigate how genes interact to produce a phenotype, often on a genome-wide scale. One feature of studying networks is the finding that a few so-called 'hub' genes can be key, and could influence many unrelated diseases.

While hub genes are an important feature of regulatory networks, Patricia Wittkopp expands on their basic structure, discussing how modulation of networks alters gene expression, which in turn underlies phenotypic plasticity and variation within and between species (p. 1567). By understanding the modulation of networks, researchers can provide 'insight into the molecular mechanisms of ecological responses and phenotypic evolution', says Wittkopp. However, to understand changes in gene expression, researchers need to understand how the changes in the regulatory networks alter this expression, such as the interactions between DNA, RNA and proteins, and how they affect transcription. There are common motifs – patterns in how the components in a network are arranged relative to each other – that emerge from networks. These involve interactions between molecules in groups or cascades, feedback loops or hub genes. Many regulatory networks have a hierarchical structure, where genes that control the earliest events occur at the top of the hierarchy and those controlling the final stages of differentiation are found at the bottom.

Nicolas Smith and colleagues use engineering principles to build predictive mathematical models of biological systems, with the aim of understanding how complex systems work based on knowledge of how the individual elements function (p. 1576). Again Smith highlights the need to integrate data from many different sources, from the cellular level to the whole organism, so that researchers can understand how the bits work together to contribute to the functioning of the whole, stating that mathematical models are an ideal way to do this. The model of the heart developed by Smith and his colleagues factors in ion pumps and leaks, muscle contraction, muscle energetics, tissue structure and properties. He cautions that when building such models, it's important that the data which underlies them are relevant and up-to-date. It is also essential that the researchers who build these models have a framework in place, such as Internet forums, to discuss, peer-review and compare their models.



'OMICS APPROACHES

One of the physiologist's goals is to understand how animals cope with fluctuations in their environment. Andrew Gracey discusses how scientists can relate physiological responses to environmental change to shifts in gene expression (p. 1584). Of the many techniques available to analyse changes in gene expression, cDNA microarrays remain the most powerful technique for screening non-model organisms, because of the large numbers of genes that can be analysed. Highlighting work carried out in the carp, Gracey and his colleagues have created a cDNA microarray to investigate the gene regulatory mechanisms underlying cold and hypoxia acclimation. The challenge is to integrate all the genetic and physiological mechanisms with the ultimate goal of predicting how an organism will respond to environmental and physiological perturbations.

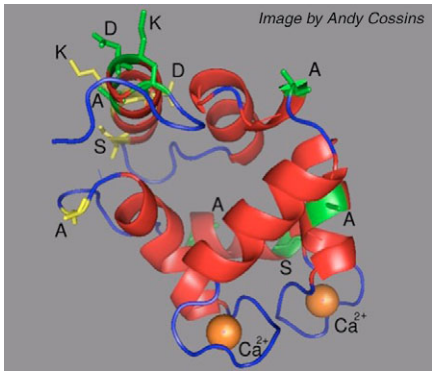
Again focussing on non-model species, Dietmar Kültz (p. 1593) writes about the challenges and techniques used to understand osmoregulation and coping with salinity in creatures such as tilapia, sharks and sponges. While researchers rely on gene databases for model organisms such as the stenohaline zebrafish to find out which genes are involved in physiological responses, there could be problems relying on these data to learn about the equivalent responses in the euryhaline tilapia. Not only do these fish operate over different salinity ranges, there will also be a high degree of plasticity in the physiological response and in the underlying gene and

protein networks in response to salinity. The solution, Kültz suggests, is not to be too ambitious too soon but to focus on one biological process at a time rather than trying to understand the whole non-model organism at once.

According to Bradley Buckley (p. 1602), one technique available to researchers to compare broad scale patterns of gene expression between species is heterologous hybridization, where a microarray from one species, such as the eurythermal goby, is used to probe for gene expression in another, such as species of cold-adapted Antarctic fish. While this approach saves researchers the effort of having to construct a new cDNA microarray for each new species they wish to study, there are caveats that have to be addressed. One issue is that the two species need to be closely related enough for the genes from one species to accurately identify the same genes in the other species. The length of the DNA probes on the array also affects success: longer probes reduce the chance of random mismatching.

It's not just comparisons between species that benefit from microarray technology; the technique can also be used to look at the well-studied problem of aging. Stuart Kim writes that aging is a complex process, which results in cumulative changes in the expression of many genes (p. 1607). By using microarrays to perform genome-wide scans, researchers can define the aging process, by comparing young and old organisms, such as worms and flies, and tissues in mice and humans. While many differences in gene expression between young and old are specific to a particular species, there are some common features. For example, the 95 genes that encode components of the electron transport pathway in mitochondria 'show common age regulation from worms to humans', says Kim; their expression decreases about twofold in older animals.

The final paper in this section, by Douglas Crawford and Marjorie Oleksiak, offers a word of caution (p. 1613). They stress the importance of measuring individual variation, because pooled samples can hide important physiological information. Levels of gene expression can differ greatly, even between closely related individuals. Using microarray analysis, they compared gene expression in isolated heart ventricles from killifish (*Fundulus heteroclitus*). They found that hearts from different individuals had differences in their metabolism: for example, which metabolic substrate was preferred by the tissues. 81% of this variation was explained by altered patterns of gene expression in various sets of genes coding for proteins in different parts of the metabolic pathway in separate groups of individuals. The implications of this are that results from inbred strains of animals should be interpreted with caution.



CLOSING THE GENOTYPE–PHENOTYPE GAP

Understanding how the genotype results in the phenotype that scientists observe is one of the biggest challenges facing comparative physiologists. Asking which animals are the best to test this relationship in, Kevin Strange (p. 1622) revisits the Krogh principle: 'for many problems there is an animal in which it can be most conveniently studied'. The ideal genetic model organism in which to answer the question of how the parts of a biological system work individually, and with each other, must be easy to manipulate genetically but still complex enough to be interesting. Strange highlights why the nematode *C. elegans* is an ideal animal to study: its well-developed muscular and nervous systems are of interest to physiologists, while it is also easy to manipulate genetically and its development is well categorised.

Julian Dow continues the discussion of how scientists can understand how phenotypes are created from genotypes, and how they can use model organisms to answer physiological questions (p. 1632). Dow argues that integrative physiology benefits from the investigation of gene function in the context of the intact animal. This implies that researchers need to use a genetically tractable model organism to answer physiological questions on the 'general principles of function,' he says, and can extend the Krogh principle a little further by choosing organisms on the basis of how easy they are to study experimentally. *Drosophila* is such a model organism, and studies on these flies have, for example, increased our understanding of circadian clocks through the scrutiny of emergence times and their genetic control.

How animals are adapted to their environment is a fascinating question, but Michael Berenbrink wants to answer the question 'how did it come to work as it does?' (p. 1641). By using a method of evolutionary reconstruction, Berenbrink discusses how molecular phylogenetic trees can be used to piece together the evolutionary steps of a system's development and thus offer another route for understanding physiological diversity. Focussing on the evolution of the swimbladder in fishes, Berenbrink relates how changes in the pH dependence of the oxygen-binding ability of haemoglobin and its specific buffer ability facilitated the evolutionary development in some fish of an inflatable swimbladder to achieve neutral buoyancy at great depths.

Continuing with the evolutionary theme, Martin Feder (p. 1653) concludes the issue by discussing how mutation influences gene function, which in turn influences how adaptations arise. While a lot of focus has been on single nucleotide mutations, which can have a large impact, they only affect existing genes. Other mechanisms at work include gene duplication, lateral gene transfer, or hybridisation, and other processes that can scramble and reassemble a nucleotide sequence. Understanding these mechanisms will allow researchers to detail the evolution of complex physiological and biochemical traits.

THE FUTURE

Cossins and Somero hope that the papers in this volume will inspire comparative and integrative physiologists to use genomic technologies to learn how the different parts of an organism work together and to integrate this information to understand the organism as a whole and how it responds to changes in its environment. The technological advances that have been made are making it possible to deal with thousands or genes simultaneously and to discover how they interact with each other. In addition, 'we have a much more discrete ability to use the knowledge from well-known species to learn more about unusual, yet related, species', says Cossins.

10.1242/jeb.02785

Laura Blackburn
laura@biologists.com