HIGHLIGHTS

HUMAN GENETICS

Genetically speaking

According to the linguist Noam Chomsky, all languages share a universal grammar, and underlying this commonality are innate language skills. But practically nothing is understood of how directly these language skills derive from individual genes. Potentially informing this discussion, Anthony Monaco and colleagues now reveal the genetic basis of a severe speech and language disorder.

The KE family has been the focus of debate among psychologists and neurologists. It is a large, three-generation family beleaguered with a severe language and speech impairment that is inherited in an autosomal-dominant fashion. Affected members experience difficulty in identifying phonemes (the smallest unit of sound in a spoken language) and understanding sentences. The language skills that most of us exercise unconsciously — such as the use of plurals, verb tenses, and various word order and combination rules — must be learnt by heart by affected members of the KE family. But these individuals are also impaired in the mouth and facial movements needed to form and articulate words. So, whether their language deficiencies stem from motor-neural problems associated with speech and hearing, or from difficulty with grammatical rules, has been controversial.

Monaco's team, who previously mapped the gene to chromosome 7, has now tracked the so-called *SPEECH1* locus to a region containing a gene, *FOXP2*, that encodes a novel putative DNA-binding protein belonging to the family of forkhead/winged-helix transcription factors. All affected members of the KE family carry a point mutation within this gene, which alters an amino acid in the DNA-binding domain of FOXP2. Additional evidence to support the pathological significance of defects in *FOXP2* comes from the finding that the gene is disrupted by a chromosomal rearrangement in an unrelated individual who has a similar speech and language disorder to that of the KE family.

So how do defects in *FOXP2* contribute to language deficits? Brain imaging studies have previously shown that affected members of the KE family have basal ganglia pathology and so it is feasible that mutation of *FOXP2* leads to perturbation in basal ganglia formation during development. Even so, it would remain an open question whether such a developmental abnormality is a key link between language deficits and the networks in the brain that underlie grammar and linguistics.

The challenge will evidently be to determine the role of FOXP2 and how it contributes to language function. Although animal models of FOXP2 defects might shed light on the function of the transcription factor at the cellular level, it will be difficult to resolve its function in linguistics. Nevertheless, an exciting avenue for future research is the possibility of a better understanding of human language through comparative genomics. By comparing the *FOXP2* gene in humans and in our close primate relatives it might be possible to trace the thread of language evolution.

Carina Dennis
Senior Editor, Nature

References and links

ORIGINAL RESEARCH PAPER Lai, C. S. L. *et al.* A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature* **413**, 519–523 (2001)
FURTHER READING Pinker, S. Talk of genetics and vice versa. *Nature* **413**, 465–466 (2001)
WEB SITE
Anthony Monaco's lab:
http://www.well.ox.ac.uk/monaco/index.shtml

HUMAN GENETICS

A bridge to complex disease

The positional cloning of a human disease gene is a landmark event. Good examples are provided by Bardet–Biedl syndrome (BBS), for which several genes have recently been identified (see Highlights, July). But these landmarks have now assumed greater significance, because some cases of BBS seem to require at least three mutant alleles. As a clear-cut example of 'triallelic' inheritance, BBS might therefore provide some general lessons about the interaction between genes, and its influence on phenotype.

BBS is usually thought of as an autosomal-recessive disorder that is characterized by a broad and variable phenotype, involving retinal dystrophy, obesity, mental retardation, renal defects and polydactyly. BBS is also genetically heterogeneous — defects at several loci can cause the disease. So, when Katsanis *et al.* screened patients for mutations in two known BBS genes (*BBS2* and *BBS6*), it was no surprise to find mutations in these genes in only some of the patients. What was more surprising was that some affected individuals carried three mutant alleles — for example, two in *BBS2* and one in *BBS6*.

The authors propose that BBS is inherited, at least in some individuals, in a triallelic fashion, although further BBS genes will need to be identified to determine the generality of this observation. In an accompanying Perspective, Arthur Burghes and colleagues, describe BBS inheritance as autosomal recessive with a modifier of penetrance, and liken BBS to other disorders in which modifier loci have been identified. However these unexpected findings are described, BBS provides a clear example of a phenotype that is determined by interactions between a small number of genes. BBS inheritance therefore lies somewhere between Mendelian and polygenic. As researchers move towards a more molecular understanding of BBS pathology, this disorder will become a valuable source of information about potential genetic interactions in more complex genetic disease.

Mark Patterson

References and links

ORIGINAL RESEARCH PAPER Katsanis, N. *et al*. Triallelic inheritance in Bardet–Biedl syndrome, a Mendelian recessive disorder. *Science* **293**, 2256–2259 (2001)

FURTHER READING Burghes, A. H. M. *et al*. The land between Mendelian and multifactorial inheritance. *Science* **293**, 2213–2214 (2001) | Nadeau, J. H. Modifier genes in mice and humans. *Nature Rev. Genet*. **2**, 165–174 (2001)

WEB SITES

James Lupski's lab: http://imgen.bcm.tmc.edu/molgen/lupski/

Bardet–Biedl syndrome links: http://www.kumc.edu/gec/support/laurmoon.html

GENOME EVOLUTION

A primate example of positive selection

Segmental duplications are a common feature of the human genome and they can be very difficult to sequence. But they should not be ignored. Through careful analysis of one set of segmental duplications, Matthew Johnson and colleagues have discovered a new human gene family. Furthermore, this gene family seems to have been subjected to powerful positive selection.

The authors focused on a 20-kb repeated segment that is confined to a 15-Mb region of chromosome 16. There are 15 copies of this segment, which have very high levels of sequence similarity. To study the evolutionary history of the duplicated segments, the authors identified the orthologous sequences in a series of primates. Their analysis showed that the segment is only present in one or two copies in Old World monkeys, such as the baboon. By contrast, in great apes such as gorillas, which are more closely related to humans, there are 9–30 copies of the segment. Overall, the segment seems to have been duplicated recently and independently in several primate lineages, after the divergence of humans and great apes from Old World monkeys.

When the sequence of the human segment was used in a database search, the authors found that an expressed sequence lurks within the repeated segment, although no homologues could be found in other organisms.

Surprisingly, sequence comparisons of the human repeats showed that the putative protein-coding regions are five times more divergent compared with the non-coding regions of the repeat. This indicates that the gene might be under positive selection for adaptive mutations. In support of this, the ratio of nonsynomous to synonymous amino-acid changes was significantly greater than 1, and on the basis of comparisons with the primate sequences, evidence of positive selection could be found during the divergence of the great ape and human lineages. Despite the relatively recent duplication events, the gene has undergone major evolutionary change.

Positive selection for amino-acid substitutions indicates an important function, which, for this gene family, is as yet unknown. Nevertheless, these observations attest to the importance of duplication and divergence as key evolutionary mechanisms. So, although segmental duplications can be a positive nuisance for genome sequencers, they provide some fascinating clues about the history of our genome.

Mark Patterson

References and links

ORIGINAL RESEARCH PAPER Johnson, M. E. *et al*. Positive selection of a gene family during the emergence of humans and African apes. *Nature* **413** , 514–519 (2001)

WEB SITE

Evan Eichler's lab: http://genetics.gene.cwru.edu/eichler/

TECHNOLOGY

New delivery vehicles for Cre

As discussed in last month's special focus on mouse genomic technology (see below), new techniques are constantly emerging for switching on the site-specific recombinase Cre in spatially and temporally controlled ways. Now, two papers report new approaches to delivering the recombinase to undifferentiated and terminally differentiated cells — a self-deleting Cre-lentiviral vector that overcomes the cytotoxic effects of prolonged *cre* expression and a cell-permeable form of Cre. Because its uptake depends on protein trafficking, this cell-permeable Cre will probably both facilitate future gene-function studies and provide cell biologists with a useful new tool.

Traditionally, there have been two main ways to deliver Cre to target cells — by using *cre*-expressing transgenic mice or *cre*-carrying viral vectors. Although such vectors have their advantages, they predominantly infect dividing, undifferentiated cells. So to overcome this limitation, Alexander Pfeifer and colleagues created a new Cre-lentiviral delivery vector (LV-Cre) because lentiviruses can infect both dividing and non-dividing cells. When the authors injected LV-Cre into Rosa26–Cre (R26R) reporter mice, which ubiquitously express a Cre-activatable form of *lacZ*, β-gal staining was evident in both undifferentiated and terminally differentiated cells. However, when injected into the brains of R26R mice, this vector caused brain abnormalities to develop after several weeks. On closer inspection of Cre's effects, Pfeifer *et al.* found that its prolonged presence causes cell-cycle arrest and apoptosis in cells *in vitro* and *in vivo* — possibly because of Cre-mediated illegitimate recombination at 'pseudo' *loxP* sites in the mouse genome. So to prevent this cytotoxic activity, the authors developed a self-deleting form of LV-Cre (LV-Cre-SD), by inserting a single *loxP* site into the vector's 3′ UTR. When the vector is reverse-transcribed and inserted into the genome, this 3′ UTR region is duplicated and introduced into the 5′ UTR, creating a vector that is flanked with *loxP* sites. LV-Cre-SD mediates recombination both *in vitro* and *in vivo* at the same efficiency rates as LV-Cre but without its cytotoxic effects, proving that this self-regulating Cre vector is likely to be a useful new tool for activating or inactivating gene expression in dividing and differentiated cells.

Jo *et al.* took a different approach to delivering Cre to cells by developing a cell-permeable form of the protein that carries a membrane-trafficking sequence. When this protein is repeatedly injected intravenously or intraperitoneally into R26R mice over 3–5 days, it causes widespread *lacZ* expression throughout organs such as the brain, kidney and liver, without adversely affecting the mice. However, there are a couple of disadvantages to using this modified Cre — its widespread dissemination limits its use in tissue-specific studies and it mediates recombination in fewer cells than does Cre that is delivered by transgene expression. Nevertheless, this cell-permeable Cre is likely to become a useful tool for future gene-function studies, given its ease of use and systemic effects. Furthermore, because it provides a stable record of protein trafficking and uptake, it could provide a new approach to developing protein-based therapies for treating human disease and equip cell biologists with a handy new tool.

Jane Alfred

References and links

ORIGINAL RESEARCH PAPERS Pfeifer, A. *et al.* Delivery of the Cre recombinase by a self-deleting lentiviral vector: efficient gene targeting *in vivo*. *Proc. Natl Acad. Sci. USA* **98**, 11450–11455 (2001) | Jo, D. *et al.* Epigenetic regulation of gene structure and function with a cell-permeable Cre recombinase. *Nature Biotechnol.* **19**, 929–933 (2001)

WEB SITE

Special focus on mouse genome technology: http://www.nature.com/nrg/focus/mousegen/

IN BRIEF

VIRAL EVOLUTION

Recombination in the hemagglutinin gene of the 1918 "Spanish flu".

Gibbs, M. J. *et al. Science* **293**, 1842–1845 (2001)

A reanalysis of the gene sequence of the influenza virus that caused the 1918 pandemic — the 'Spanish flu' — has revealed that the major virulence determinant, encoded by the haemagglutinin (HA) locus, originated by recombination. Sequence alignment of three HA genes taken from the 1918 victims against modern-day isolates from humans, pigs and birds indicates that a recombination event occurred between swine- and human-lineage-derived HA sequences just before the 1918 pandemic, suggesting a causal link between the two events.

MOUSE GENOMICS

An SSLP marker-anchored BAC framework map of the mouse genome.

Cai, W.-W. *et al. Nature Genet.* **29**, 133–134 (2001)

A radiation hybrid transcript map of the mouse genome.

Avner, P. *et al. Nature Genet.* **29**, 194–200 (2001)

A radiation hybrid map of mouse genes.

Hudson, T. J. *et al. Nature Genet.* **29**, 201–205 (2001)

These papers report new resources for the sequencing and functional analysis of the mouse genome. The first reports an SSLP-marker-anchored BAC map of the mouse genome that was generated by using specific probes for library screenings, selected by improved oligo-designer software. This map covers 94% of the genome in 600 contigs. Avner *et al.* constructed a radiation-hybrid (RH) transcript map of 5,904 mapped EST and STS markers. The mapped ESTs were isolated from a mouse embryonic endoderm library to enrich for transcripts expressed in early development and unlikely to have been previously mapped in humans. Together with the RH map reported by Hudson *et al.* — which contains 11,109 genes positioned relative to a map of 2,280 markers — these maps provide resources for sequencing the mouse genome, for orthology mapping in humans and for rapidly identifying genes mutated in ENU mutagenesis screens.

FUNCTIONAL GENOMICS

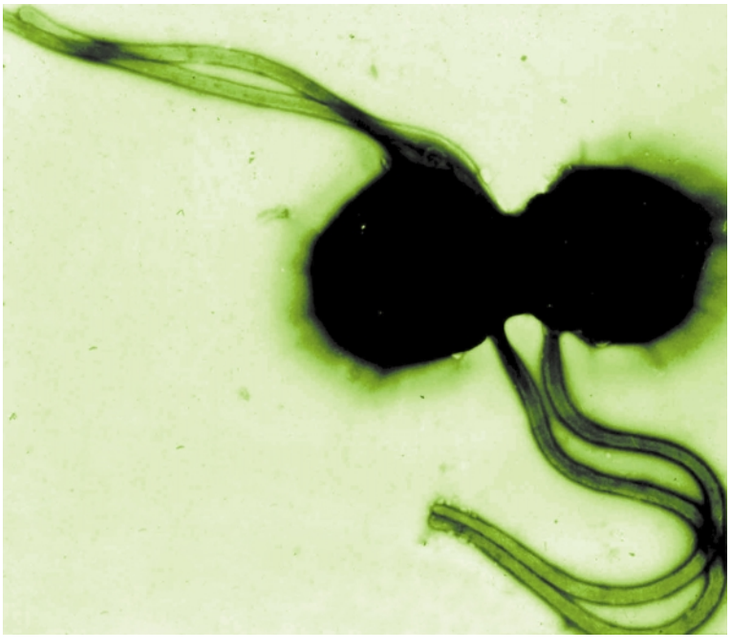
Large-scale identification of mammalian proteins localized to nuclear sub-compartments.

Sutherland, H. G. E. *et al. Hum. Mol. Genet.* **10**, 1995–2011 (2001)

Incorrect nuclear protein localization has been associated with human disease and cancer. Sutherland *et al.* tracked subnuclear protein localization in mouse embryonic stem cells using a previously developed gene-trap screen, in which a *lacZ* marker is inserted into introns and the chimeric proteins are detected by immunostaining. The authors found that proteins that share subnuclear localization contain similar domains, indicating that the localization of these proteins can be predicted from their sequence.

GENE REGULATION

Probing zygotic gene control



Courtesy of William Snell, University of Texas Southwestern, USA.

Although it is known that fertilization triggers events that ultimately lead to the transcriptional activation of the zygotic genome, we know surprisingly little about these events — a factor that probably contributes to the low success rates of current mammalian-cloning efforts. So, to investigate the mechanisms that control zygotic gene transcription, Zhao *et al.* turned to *Chlamydomonas reinhardtii* — a unicellular alga — in which early zygotic development is easier to study. Here they report that the gamete-specific, homeodomain protein GSP1 can activate the transcription of some zygotic genes in the absence of gamete fusion. Their findings show that both gametes contribute proteins that are required for zygote development.

Chlamydomonas's quirky life cycle makes it amenable to studies of fertilization. Upon nutrient starvation, its haploid vegetative cells of two mating types — mt^- and mt^+ — undergo gametogenesis. When gametes of opposite mating type meet, they fuse together (see picture), triggering rapid zygote-specific gene expression. This expression occurs independently of protein synthesis, indicating that pre-exisiting factors in one or both gametes regulate the expression of these genes. As zygote maturation continues, zygotes clump together to form large aggregates and mt^--derived choloroplasts are selectively degraded — zygotes germinate on return to a nutrient-rich medium.

The previous finding that GSP1 is present only in mt^+ gametes led the authors to consider it as a candidate regulator of zygotic gene transcription in *Chlamydomonas*. To test this, they expressed *gsp1* in mt^- cells, and found that although the vegetative mt^- transformants appeared normal, they formed zygote-like aggregates on undergoing gametogenesis — a behaviour not seen in transformed mt^+ cells. The transformed mt^- gametes expressed six out of seven zygote-specific genes — the unexpressed gene, *ezy1*, is thought to be required for mt^- chloroplast destruction. Its lack of expression in mt^- transformants indicates that it is under separate regulatory control or perhaps is imprinted in mt^- gametes.

DEVELOPMENTAL BIOLOGY

Hedgehog shows the way

Cells that originate in one place in the embryo often have to migrate some distance to reach their ultimate place of residence. A good example of this is the *Drosophila* germ cells. These cells are specified by maternal factors and lie outside the posterior of the embryo, segregated from the soma until gastrulation. At this stage, they undergo a complex pattern of migration that brings them inside the embryo and into contact with two lateral clusters of somatic gonadal precursor (SGP) cells — an association that creates a functional gonad. But how do germ cells know where to go? In their study of this process, Girish Deshpande and colleagues have found that, by secreting Hedgehog (Hh), SGP cells could provide the attractive cue that guides germ cells to their destination.

The study, published in *Cell*, was based on the premise that germ cells would follow attractive signals derived from the SGP cells, and that the identity of such molecules could best be found by looking at genes involved in specifying SGP cell fate. Of several such genes, Hh seemed a particularly promising candidate — it is expressed in the SGPs (as shown by the distribution of Hh–*lacZ*) and is known to be a short- and long-range signalling molecule. A crucial experiment reinforced this prediction: when Hh was ectopically expressed in the embryo, using four different Gal4 drivers, it redirected the migration of a subset of germ cells towards the new source of Hh. An alternative explanation for these results — that germ cells were drawn to ectopic SGP cells newly induced by Hh — is unlikely, as the Hh-expressing cells were negative for a SGP-cell-specific marker.

If germ cells respond directly to Hh, rather than to a secondary signal, then it follows that the germ line should require cell-autonomous components of the Hh pathway for proper migration. Indeed, when mutant for positively acting members of the pathway (such as *smoothened* or *fused*), germ cells scattered randomly and failed to

But if GSP1 regulates zygote-specific genes in mt$^-$ cells, why does it not affect zygotic gene expression in mt$^+$ gametes? Perhaps this is because, as the authors suggest, zygote-specific gene promoters are inaccessible to GSP1 in mt$^+$ gametes or because GSP1 associates with a pre-existing mt$^-$-specific partner molecule to form a transcription-regulatory complex that activates zygote-specific gene expression. The parallels between these events and those in budding yeast indicate that atypical, gamete-specific, homeodomain proteins, such as GSP1 and budding yeast's MATα2, might have evolved to act as regulators of zygotic gene expression before animals and plants diverged. If so, then such studies could inform our understanding of how zygote gene expression is controlled in mammals.

Jane Alfred

References and links

ORIGINAL RESEARCH PAPER Zhao, H. *et al.* Ectopic expression of a *Chlamydomonas* mt$^+$-specific homeodomain protein in mt$^-$ gametes initiates zygote development without gamete fusion. *Genes Dev.* **15**, 2767–2777 (2001)

WEB SITE

The ***Chlamydomonas*** **genome database**: http://www.biology.duke.edu/chlamy_genome/

associate with SGP cells, as if they had lost all sense of direction — consistent with loss of Hh. Conversely, *patched* or *protein kinase A* mutant germ cells, in which Hh signalling is constitutively active, clumped together in the middle of the embryo and failed to migrate at all.

Interest in germ-cell migration is not new and so several genes that affect this process, such as *wunen* and *Columbus*, have already been identified. How does Hh fit in with previous models of germ-cell migration? As there are countless sources of Hh in the embryo, how is specificity of migration achieved? This is probably only one leg of a longer journey to find out how germ cells reach their targets.

Tanita Casci

References and links

ORIGINAL RESEARCH PAPER Deshpande, G. *et al.* Hedgehog signaling in germ cell migration. *Cell* **106**, 759–769 (2001)

WEB SITE **Paul Schedl's homepage:** http://www.molbio.princeton.edu/labs/schedl/index.htm

HUMAN GENETICS

Closing in on palatal disorders

The formation of the lip and palate is a complex and delicate process in craniofacial development, requiring the careful joining of tissues from two opposite sides of the mouth. Cleft lip/palate (CL/P) — in which the lip and palate have failed to close — affects up to 0.2% of live births, with most instances occurring in families with no history of the disease. However, ~30% of cases occur as part of single-gene syndromes. Understanding the genetics behind this class of common birth disorder has not been easy, but the recent identification of two loci, one for CL/P and the other for isolated cleft palate (CP), has provided clues to the developmental defects that underlie these malformations. The importance of these studies is underscored by the finding that mutations at the same locus could be responsible for both the inherited and sporadic forms of CL/P, indicating a model that could lead to the identification of genes for other common, complex birth defects.

In the first of two studies, Claire Braybrook and co-workers went in search of the causative locus for a specific subclass of CP — cleft palate with ankyloglossia (CPX) — which is inherited as a semi-dominant X-linked disorder. The locus was delimited to a region of Xq21; of the three plausible transcripts within the candidate interval, only one, the conserved *TBX22* (T-box 22) gene, was mutated in affected males from an Icelandic family. Mutations in *TBX22* — missense, nonsense, splice site and frameshift — were also observed in individuals with CPX from five other families of different ethnic backgrounds, and are predicted to cause a complete loss of function of *TBX22*. This mutation distribution, the expression of *TBX22* in the palate and the involvement of T-box family genes in early development, make *TBX22* a likely determinant in palate morphogenesis.

The starting point for the second study, by Mehmet Sözen *et al.*, was their earlier finding of a gene responsible for the inherited CL/P-ectodermal dysplasia syndrome (CLPED1), an autosomal-recessive disorder attributable to mutations in the poliovirus receptor-related 1 gene, *PVRL1*. In the Venezuelan community on Margarita Island that they studied, CLPED1 is very frequent and is caused by homozygosity for the *PVRL1* nonsense mutation, W185X. Because the level of sporadic CL/P is also high on this island, the authors were curious to find out whether the same W185X variant was involved in both familial and sporadic forms of CL/P. Although there was no significant difference between the heterozygosity for W185X in sporadic CL/P patients and normal, unrelated islanders, a difference was observed in a population on the adjacent Venezuelan mainland. It seems likely that heterozygosity for W185X is a moderate genetic risk factor for sporadic CL/P, at least in this population, but is only one of many genetic and environmental contributors.

The story will not end here, as more susceptibility loci for CL/P defects will no doubt emerge. For all of them, the identification of the molecular lesion must be followed by a characterization of the resulting developmental pathology. In the case of *PVRL1*, which encodes nectin 1 — a cell–cell adhesion molecule important for cell fusion — this process has already begun. A key message to emerge from these two papers is that rare developmental syndromes can indicate candidate loci for more common disorders — a strategy that is especially welcome when standard mapping approaches are not an option.

Tanita Casci

References and links

ORIGINAL RESEARCH PAPERS Braybrook, C. *et al*. The T-box transcription factor gene *TBX22* is mutated in X-linked cleft palate and ankyloglossia. *Nature Genet.* **29**, 179–183 (2001) | Sözen, M. A. *et al.* Mutation of *PVRL1* is associated with sporadic non-syndromic cleft lip/palate in northern Venezuela. *Nature Genet.* **29**, 141–142 (2001)

FURTHER READING Murray, J. Time for T. *Nature Genet*. **29**, 107–109 (2001)

WEB WATCH

Interactive evolution

- http://www.evotutor.org/
- http://www.pbs.org/evolution/

Do those dry evolutionary genetics formulae fail to leap off the page? If so, then *EvoTutor* — a free online resource for interactive simulations of evolutionary processes — might be a site that you'll greet with open arms.

By clicking on 1 of 12 evolutionary processes, such as drift or adaptation, you can access a selection of subtopics, which provide background information on the process and explanations of its associated formulae. Alternatively, you can go straight from a subtopic to an interactive simulation of a process. To illustrate gene flow, for example, individuals in two starting populations are depicted as coloured dots; by altering the rate of migration between populations and the strength of selection against migrants and hybrids, you can watch the make-up of the populations change as the coloured dots intermingle.

Although you won't become an evolutionary genetics buff just by visiting this site (having a textbook to hand might help), the simulations do help bring to life what formulae and your imagination might not. Stay tuned for *PhyloTutor* and *GenTutor*, sister sites that are now under construction.

If you're after a broader picture, then make the *Evolution* web site — linked to a US PBS television miniseries — your first stop. In seven episodes, the series explored evolution from the life of Darwin to the struggle between science and religion. Going beyond the series, the site offers video previews of each programme, links to hundreds of online sites and movies, a glossary, and courses and manuals for teaching. The *Evolution* project's goal is to heighten public understanding of evolution, and how it works. In this respect, it certainly makes the grade.

Tanita Casci

MICROBIAL GENOMICS

Bugs inside out

The genomes of important human pathogens are under intense scrutiny — the goals being to understand the basis of their virulence and to design more effective antimicrobial drugs. Important steps towards unravelling the secrets of virulence evolution are described in recent articles that report the genome sequences of two microbial pathogens — *Yersinia pestis*, the agent of plague, and *Rickettsia conorii*, which causes Mediterranean spotted fever. In a third study, which takes a more functional approach, Yinduo Ji *et al.* have devised a method to find potential virulence genes in *Staphylococcus aureus*.

The 4.6-Mb genome of *Y. pestis* seems to be in flux. Parkhill *et al.* found signs of old genomic expansions — probably a result of horizontal transfer — that preceded the split of *Y. pestis* from its close cousin *Yersinia pseudotuberculosis*. This expansion has been followed by the first signs of genome downsizing, associated with a decay of previously active genes into pseudogenes. In part, this seems to reflect a change in the pathogen's lifestyle: the genes necessary for the enteropathogen *Y. pseudotuberculosis* in its environment within the host's gut, such as adhesion genes, have become pseudogenes in the bloodstream-dwelling *Y. pestis*. Adaptation to different environments is also reflected in the accumulation of mutations in genes responsible for the uptake and transport of different nutrients in the two species.

Similar studies were conducted by Ogata *et al.*, who sequenced the *R. conorii* genome (1.2 Mb) and compared it with that of its close relative, *Rickettsia prowazekii*. These intracellular parasites diverged 40–80 million years ago, so this comparison provided interesting insight into adaptations associated with different intracellular lifestyles. As expected, there seems to be a tendency to streamline the genome. For example, the authors discovered evidence of decaying orthologues — a group of 229 genes in *R. conorii* had homology to non-coding sequences in *R. prowazekii*. Nevertheless, some gene families do exist, and they point to processes, such as importing the host cell's ATP, which represent important adaptations of these pathogens.

In an independent approach to identifying essential genes in *S. aureus* (the sequence of which was published last year), Ji and colleagues subcloned random 200–800-bp fragments of *S. aureus* DNA into plasmids, downstream of a tetracycline-inducible promoter. Bacteria transformed with these constructs were replica plated and grown in the presence or absence of a tetracycline analogue. This allowed the authors to look for constructs that prevented bacterial growth as a result of induced antisense RNA expression. Of the 600 constructs that were identified, ~200 contained an open reading frame in an antisense orientation, and 30% of these had no known function. To test whether this approach could be extended to an *in vivo* context, Ji *et al.* infected mice with bacteria that contain a construct expressing antisense RNA to a known essential gene. Whereas control mice developed a heavy kidney infection within 72 h, mice induced to express the antisense construct suffered no infection.

Comparative and functional genomics of pathogenic species and their relatives will continue to provide invaluable insights into the evolution of pathogenicity and adaptation. The potential for substantial benefits to human health is tantalizing.

Magdalena Skipper

References and links

ORIGINAL RESEARCH PAPERS Parkhill, J. *et al.* Genome sequence of *Yersinia pestis*, the causative agent of plague. *Nature* **413**, 523–527 (2001) | Ogata, H. *et al.* Mechanisms of evolution of *Rickettsia conorii* and *R. prowazekii*. *Science* **293**, 2093–2098 (2001) | Ji, Y. *et al.* Identification of critical staphylococcal genes using conditional phenotypes generated by antisense RNA. *Science* **293**, 2266–2269 (2001)

FURTHER READING Cole, S. T. *et al*. A plague o' both your hosts. *Nature* **413**, 467–470 (2001)

WEB SITES

Yersinia pestis: http://www.sanger.ac.uk/Projects/Y_pestis/

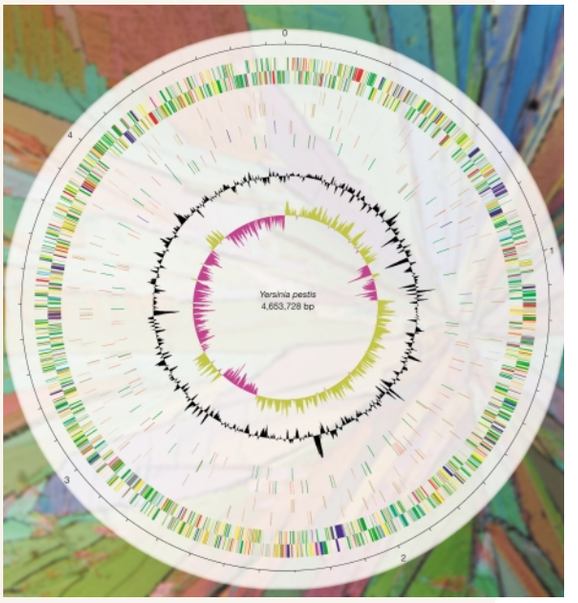
Rickettsia conorii: http://www.ncbi.nlm.nih.gov/cgi-bin/Entrez/framik?db=Genome&gi=199

Staphylococcus aureus: http://www.tigr.org/tigr-scripts/CMR2/GenomePage3.spl?database=ntsa01

Yersinia pseudotuberculosis: http://bbrp.llnl.gov/bbrp/html/y.pseudo.htm

Rickettsia prowazekii: http://www.tigr.org/tigr-scripts/CMR2/GenomePage3.spl?database=ntrp01

Comprehensive Microbial Resource at TIGR: http://www.tigr.org/tigr-scripts/CMR2/CMRHomePage.spl



Circular representation of the *Yersinia pestis* genome. (Modified with permission from Parkhill, J. *et al*. *Nature* **413**, 523–527 © (2001) Macmillan Magazines Ltd.)

GENE EXPRESSION

Genetic landscape in 3D

Last year, data from several yeast DNA microarray experiments were pooled to create a compendium of gene expression profiles in *Saccharomyces cerevisiae*. Data comparisons on such a massive scale can provide rich insights into gene function and biology. Kim *et al.* have now pooled data from 30 different laboratories to create a similar compendium that covers ~93% of the *Caenorhabditis elegans* genome. By visualizing data in three dimensions to create a gene expression landscape — in which genes that share expression patterns cluster together to form expression mountains — the authors were able to obtain a comprehensive view of how genes are co-regulated in the worm.

Motivated by the need to develop high-throughput functional genomics approaches to analyse gene function, Stuart Kim and colleagues combined data from 553 microarray experiments, in which gene expression was compared between wild-type and mutant worms, and between worms grown under different conditions. To identify genes that are co-regulated, they first assembled a gene expression matrix that contains the relative expression level for each gene in each microarray experiment programme. A two-dimensional scatter plot, in which genes with similar expression profiles lie close to each other, was converted into a three-dimensional map, in which the *z* axis corresponds to gene density in a given area. Each gene was then assigned to a cluster, or expression mountain, and each of these was numbered — the biggest mountain was zero and the smallest was 43.

The expression landscape is visually impressive, but is it biologically meaningful? Overlaps between certain expression mountains and groups of genes that are known to share a common function are significant. In addition, the authors confirmed the validity of this representation, for example by randomizing the data and adding noise.

Once satisfied with their method, Kim *et al.* went off to explore the genetic landscape they had generated. They found that some mountains group together genes that are expressed in similar tissues, so for example there is a muscle mountain and a germ-line mountain. Other mountains are made up of genes with similar cellular functions — there is a histones mountain and a ribosomal genes mountain. By showing that the mountains were enriched in particular sets of genes, Kim *et al.* were able to attribute physiological significance for 30 out of 44 of the mountains. Zooming in on particular mountains confirmed and augmented already known genetic links between genes. For example, 89% of previously identified sperm-enriched genes clustered on mountain 4 and genes that encode principal sperm proteins, protein kinases and phosphatases fell into three separate subclusters on mountain 4.

The exploratory trips of Kim and colleagues into this expression landscape have already yielded much information on gene co-regulation, and they have amply shown the superiority of this visual data representation. New gene interactions, unexpected gene co-regulation, assignment of function to new genes and much more await future explorers.

Magdalena Skipper

References and links

ORIGINAL RESEARCH PAPER Kim, S. K. *et al.* A gene expression map for *Caenorhabditis elegans*. *Science* **293**, 2087–2092 (2001)

FURTHER READING Gifford, D. K. Blazing pathways through genetic mountains. *Science* **293**, 2049–2051 (2001)

WEB SITE

Stuart Kim's lab: http://cmgm.stanford.edu/~kimlab/topomap/c._elegans_topomap.htm

IN BRIEF

GENE REGULATION

Identifying regulatory networks by combinatorial analysis of promoter elements.

Pilpel, Y. *et al. Nature Genet.* **29**, 153–159 (2001)

The use of microarray technology for transcriptional profiling and cluster analysis is a powerful approach for finding co-regulated genes and the promoter motifs that control transcription. Pilpel *et al.* have extended this approach by studying the combinatorial action of regulatory motifs in *Saccharomyces cerevisiae*. They identified pairs of motifs that act in a synergistic manner to provide tighter transcriptional regulation than either motif alone. The resulting network of interactions shows how a relatively small number of transcription factors might act in concert to regulate gene expression under a broad range of growth conditions.

DEVELOPMENT

A murine model of the Holt–Oram syndrome defines roles of the T-box transcription factor Tbx5 in cardiogenesis and disease.

Bruneau, B. G. *et al. Cell* **109**, 709–721 (2001)

Several T-box (Tbx) gene family members are involved in dominant congenital heart disorders — *TBX5* in Holt–Oram syndrome (HOS) and probably *TBX1* in DiGeorge syndrome (see Elizabeth Lindsay's review, p858). This *Tbx5*-knockout study provides new insight into how *Tbx5* haploinsufficiency causes HOS-like heart defects in mice. $Tbx^{+/-}$ mice show the reduced expression of many genes, most markedly that of atrial natriuretic factor (*Anf*) and connexin 40 (*Cx40*). The authors discovered that *Anf* and *Cx40* are activated when Tbx5 and Nkx2-5 interact at their promoters, indicating a mechanism for the haploinsufficient effects of *Tbx5*. Misregulation of *Cx40* could underlie the heart-conduction defects in $Tbx5^{+/-}$ mice and HOS patients. Their findings also shed light on the condition's variability and on cardiac malformations caused by other transcription factor mutations.

TRANSPOSABLE ELEMENTS

Mobilization of a *Drosophila* transposon in the *Caenorhabditis elegans* germ line.

Bessereau, J.-L. *et al. Nature* **413**, 70–74 (2001)

Although *C. elegans* has several native transposon types, they cannot be used to manipulate the worm genome because they are too numerous to serve as unique gene tags and because different transposon types can be activated at the same time, making them difficult to trace. This paper shows the successful mobilization of a foreign element — the *Drosophila* mariner element, *Mos1* — in the worm. *Mos1* has the essential qualities of a regulable transposon: it can be mobilized in the soma and germ line — where its insertion point can be identified by PCR — and it can be mutagenic by imperfect excision.