sion was not so much between the species (as would be expected if expression follows overall genetic divergence), but rather between the northern population and the two southern populations. These are then good candidate genes for investigation of a role in physiological adaptation to the marine thermal gradient.

All of these results are to a broad approximation in agreement with those from several other organisms. Two recent comparisons of haploid yeast strains (Cavalieri *et al*, 2000; Brem *et al*, 2002) suggest that up to 25% of the transcriptome differs among isolates, as did a study of *Drosophila melanogaster* adults (Jin *et al*, 2001). None of these studies compared individuals, but they both showed that there is a genetic component to transcription. By contrast, the *Fundulus* study shows that there is among-individual variation, but not formally that this is genetic. Perhaps, the next study of this sort will allow a true estimate of transcriptional heritability by including both among-individual and among-genotype comparisons. It is also worth noting that expression profiling of primate brains and livers, as well as of these tissues in three murine species, also indicates that perhaps between 5 and 20% of the transcriptome varies among individuals (Enard *et al*, 2002), though comparable statistical analyses of these data have not been published.

The value of these studies from the point of view of evolutionary genetics is not difficult to see, but should genome biologists in general take note? The simple answer is yes, because it is now clear that individual variation may be confounded with experimental contrasts.

Few would argue that when contrasting gene expression profiles of patients who have a rare liver disease with those of healthy people, it is essential to ensure that the individuals are matched for age and sex. For example, if the diseased sample was from a 50-year-old woman and the control was a 21-year-old man, there may be other factors that actually cause differences in the liver expression profiles. But how careful should we be to match the individuals for ethnicity or other types of population structure?

Suppose in a clinical comparison that the population from which the controls are drawn is subtly different in some unknown manner from the population from which the cases are drawn (Figure 1). This difference could be dietary or environmental or even genetic, and the populations could be admixed. Suppose further that the difference affects the frequency of the number of individuals who share a particular pattern of altered gene expression that is actually unrelated to the disease being studied. It turns out that even if the frequency of cases is the same in the two populations, a spurious association between transcript abundance and disease status could appear, much as population structure confounds genotype association mapping (Nielsen and Zaykin, 2001). There are no easy solutions to this problem, but it at least behooves us to respect differences in individual expression. ■

G Gibson is at the Department of Genetics, Gardner Hall, North Carolina State University, Raleigh, NC 27695-7614, USA.

e-mail: ggibson@unity.ncsu.edu

Brem RB, Yvert G, Clinton R, Kruglyak L (2002). *Science* **296**: 752–755.
Cavalieri D, Townsend JP, Hartl DL (2000). *Proc Natl Acad Sci (USA)* **97**: 12 369–12 374.
Crawford DL, Powers DA (1989). *Proc Natl Acad Sci (USA)* **86**: 9365–9369.
Enard W, Khaitovich P, Klose J, Zöllner S, Heissig F, Giaralsson P (2002). *Science* **296**: 340–343.
Jin W, Riley R, Wolfinger R, White KP, Passador-Gurgel G, Gibson G (2001). *Nat Genet* **29**: 389–395.
Kerr MK, Churchill GA (2001). *Genet Res* **77**: 123–128.
Nielsen DM, Zaykin D (2001). *Expert Rev Mol Diagn* **1**: 334–342.
Oleksiak MF, Churchill GA, Crawford DL (2002). *Nat Genet* **32**: 261–266.
Podrabsky JB, Javillonar C, Hand S, Crawford DL (2000). *Am J Physiol* **279**: R2344–R2348.

Plant genomes: cyanobacterial genes revealed

JM Archibald and PJ Keeling

Heredity (2003) **90,** 2–3. doi:10.1038/sj.hdy.6800204

More than 20 years after it was proposed that plant genomes must contain genes derived from the cyanobacterial ancestor of the plastid (Weeden, 1981), the full impact of endosymbiotic gene transfer is just being revealed. In a recent study published in the *Proceedings of the National Academy of Sciences*, Martin *et al* (2002) show that the contribution of cyanobacterial genes to the nuclear genome of the flowering plant *Arabidopsis* extends far beyond those associated with photosynthesis or the plastid. Cyanobacterial-derived genes appear to make up a large fraction of the plant genome and they not only encode proteins that service the plastid, but proteins for all other cellular functions. These results underscore the importance of endosymbiosis in shaping the biochemistry and cell biology of eukaryotic cells.

In 1981 Weeden crystallized a major component of the theory of endosymbiosis when he proposed that plant nuclear genomes contained genes originating from the cyanobacterium that gave rise to the plastid. Weeden knew that plastids contained far more proteins than their reduced genomes could possibly encode. He suggested that these proteins were originally encoded in the endosymbiont genome, but were transferred to the host nucleus during the early stages of endosymbiosis. He also proposed that the protein products of these genes were somehow targeted to the plastid after they were translated, so that the proteins ended up where they had always been. At the time this idea was galvanizing, now it is taken for granted: hundreds of genes for plastid-targeted proteins have now been described in plant nuclei and the process by which the proteins are targeted to the plastid is now largely understood (McFadden, 1999).

While Martin *et al* (2002) now elucidate the full impact of endosymbiosis in shaping the plant nuclear genome, the idea that the cyanobacterial endosymbiont contributed more genes to the nucleus than strictly necessary for plastid function is not new. For example, higher plants contain two nuclear-encoded isoforms of the enzyme phosphoglycerate kinase (PGK), one functioning in the plastid, the other in the cytosol. Despite the different evolutionary histories of the cellular compartments in which they function, *both* PGK isoforms are cyanobacterial in origin. Apparently the cyanobacterial PGK gene was duplicated during plant evolution, with one

copy servicing the plastid and the other taking over the role of the pre-existing (noncyanobacterial) cytosolic protein (Brinkmann and Martin, 1996). This phenomenon is known as *endosymbiotic gene replacement*, and while a few cases have been well documented, its overall contribution to the nuclear genome of plants has not been clear. Armed with the complete set of proteins encoded in the nuclear genome of *Arabidopsis* (The *Arabidopsis* Genome Initiative, 2000), Martin and co-workers were able to tackle this question on a large scale.

The researchers compared 24 990 *Arabidopsis* proteins to those encoded in a set of completely sequenced archaeal, bacterial, and cyanobacterial genomes, as well as those of yeast. From a set of 9368 proteins that produced a significant match in at least one reference genome, about 1700, or 18%, of the genes were most similar to a cyanobacterial homologue. Extrapolating to the genome as a whole, they estimated that about 4500 *Arabidopsis* nuclear genes are of cyanobacterial origin. Regardless of whether this is an overestimate or underestimate (there are arguments for both), this is an unexpectedly large number. Indeed, the estimated 4500 cyanobacterial genes in the *Arabidopsis* nucleus is over 1000 more genes than the total gene complement of the cyanobacterium *Synechocystis* (Kaneko *et al*, 1996) and over 60% of the number of genes encoded in the largest sequenced cyanobacterial genome, that of *Nostoc* (Meeks *et al*, 2001). While subsequent analysis and new data are certain to revise this estimate somewhat (eg, see Rujan and Martin, 2001), it is clear that the cyanobacterial endosymbiont gave vastly more of its genome to the host than previously appreciated.

However, the significance of this observation lies not so much in the sheer number of genes involved, but rather in the diversity of cellular functions predicted for the proteins they encode. Metabolism, cell growth and division, intracellular transport, cell organization, and transcription are all implicated. Even more remarkable, fewer than half of the cyanobacterial-like proteins in *Arabidopsis* are predicted to be targeted to the plastid, leading the authors to conclude that the impact of plastid endosymbiosis on the host was far greater than just acquiring an organelle (Martin *et al*, 2002). One of the steps in the textbook explanation of endosymbiotic organelle origins is the severe reduction of the endosymbiont and its genome. This may still be true in a fashion, but at least in plastids it appears that much of the endosymbiont genome has survived this reduction by relocating and finding a new role in the cell. Apparently endosymbiosis creates an influx of raw genetic material, and the mixing and matching of this material with existing host genes fosters a period of invention for the host.

Decades after the general acceptance of an endosymbiotic origin for plastids, various aspects of the process and its implications remain to be fully understood. One aspect of plastid evolution that may be interesting to consider in the light of these new findings is secondary endosymbiosis. While all plastids are ultimately derived from the original endosymbiosis between a eukaryote and a cyanobacterium, plastids have also spread laterally among eukaryotes. Secondary endosymbiosis occurs when a eukaryotic alga is swallowed by a second, heterotrophic, eukaryote and the two integrate to form a new algal lineage (Archibald and Keeling, 2002). This phenomenon accounts for much of algal diversity, and the genetic contribution of these endosymbionts to their hosts is particularly interesting since the endosymbiont brings with it a large, eukaryotic genome. The integration of endosymbiont nuclear genes into the secondary host nucleus should be easier, because the eukaryotic genes and the proteins they encode may be more easily incorporated into their new eukaryotic background than prokaryotic genes. However, such replacements will be far more difficult to detect since the host and the endosymbiotic alga are both eukaryotes, and therefore much more closely related to each other than to cyanobacteria. To date, little sequence information exists from the nuclear genomes of most of these organisms, but in time these genomes should provide another new glimpse into the effects of endosymbiotic mergers at the molecular level. ■

JM Archibald and PJ Keeling are in the Department of Botany, Canadian Institute for Advanced Research, University of British Columbia, #3529-6270 University Blvd, Vancouver, British Columbia, Canada V6T 1Z4.

e-mail: jarch@interchange.ubc.ca

The *Arabidopsis* Genome Initiative (2000). *Nature* **408**: 796–815.
Archibald JM, Keeling PJ (2002). *Trends Genet*. **18**: 577–584.
Brinkmann H, Martin W (1996). *Plant Mol Biol* **30**: 65–75.
Kaneko T *et al* (1996). *DNA Res* **3**: 109–136.
Martin W *et al* (2002). *Proc Natl Acad Sci USA* **99**: 12246–12251.
McFadden GI (1999). *J Eukaryot Microbiol* **46**: 339–346.
Meeks JC *et al* (2001). *Photosynth Res* **70**: 85–106.
Rujan T, Martin W (2001). *Trends Genet*. **17**: 113–120.
Weeden NF (1981). *J Mol Evol* **17**: 133–139.

Evolution and development

Making jaws

T Schilling

Heredity (2003) **90,** 3–5. doi:10.1038/sj.hdy.6800205

Until now, it was unclear as to which genes control differences between the upper and lower jaw of the vertebrate head. Depew and his colleagues take a bite out of this problem by creating mutant mice in which the lower jaw is transformed into a copy of the upper jaw (right down to the whiskers).

Jawed vertebrates evolved from jawless ancestors over 400 million years ago, and the evolution of a biting lower jaw was a critical step in vertebrate evolution. Comparative studies of vertebrate embryos suggest that lower jaws arose during evolution through changes in patterning along the proximodistal (PD) axis of the jaw as it forms.

The jaw is one of a series of segmentally reiterated structures called pharyngeal arches, and its PD axis, like appendages such as limbs, extends from the base of the arch to its tip (Figure 1). How this axis is established is still debated, but like limbs it clearly involves *Dlx* homeobox transcription factors related to *distalless*, a key regulator of appendage development in Drosophila.

Depew *et al* (2002) and his colleagues in a new article published in *Science* take our understanding a step further. They demonstrate that two closely related *Dlx*