

## GENOMICS

## A matched set of frog sequences

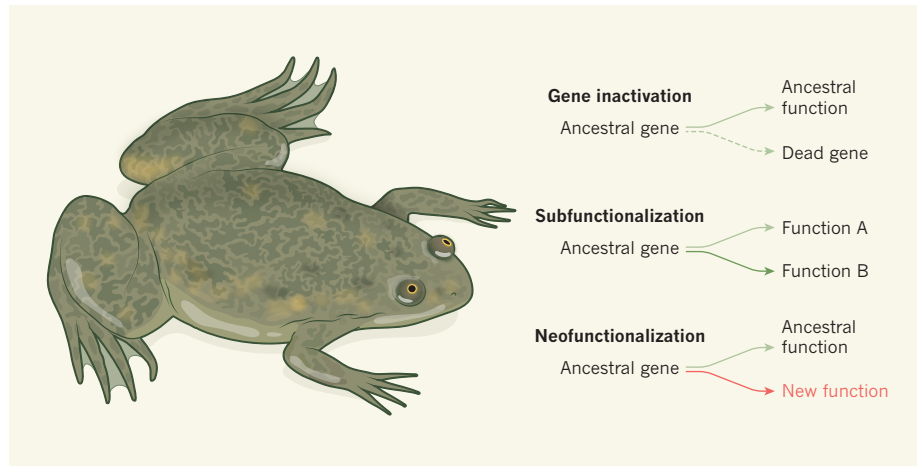
A whole-genome duplication that occurred around 34 million years ago in the frog *Xenopus laevis* made generating a genome sequence for this valuable model organism a challenge. This obstacle has finally been overcome. [SEE ARTICLE P.336](#)

SHAWN BURGESS

Ask a developmental biologist to name the most valuable animal models for their field and they will probably put the African clawed frog, *Xenopus laevis*, at or near the top of their list. Ask any geneticist the same question and this species is unlikely to even make the top ten. One reason for this disparity is that *X. laevis* has undergone a whole-genome duplication, which makes genome assembly — an essential tool of modern genetics — extremely difficult. But on page 336, Session *et al.*<sup>1</sup> report the successful sequencing of the *X. laevis* genome. The authors took advantage of ever-improving technologies and the hard work of a large, international consortium to complete this challenging project.

During the genome-assembly process for a diploid organism (one, like humans, that has two sets of chromosomes), a single reference chromosome sequence is generated to correspond to each chromosome pair. *X. laevis*, by contrast, is tetraploid — it has four sets of chromosomes, and so a reference sequence will contain two copies of most genes, instead of one. This leads to problems when using the typical shotgun approach to genome assembly, in which hundreds of millions of random short sequence reads are taken and assembled by computer into logical, continuous sequences. With a duplicated genome, it can be difficult to tell which of the two gene copies a short sequence comes from. If the sequences of the copies are too similar, the computer's assembly algorithms 'collapse' the duplicated sequence into a single copy, confounding attempts to make correct, end-to-end assemblies across all chromosomes.

Two laborious approaches that enable distinctions between duplicated chromosomes made Session and colleagues' effort successful. In the first, the authors isolated DNA from a frog and inserted long stretches into DNA constructs called bacterial artificial chromosomes (BACs). They then systematically identified 798 BACs that contained large fragments (100 kilobases or more) of DNA encoding one copy of a duplicated gene, and that could be paired with another BAC containing the other copy.



**Figure 1 | Pathways to rediploidization.** After an organism undergoes a whole-genome duplication (WGD), as has occurred in the frog *Xenopus laevis*, its genome will slowly return to the normal diploid state (in which it inherits just one set of chromosomes from each parent) through three mechanisms. First, the most common outcome is for one copy of an ancestral gene to pick up enough mutations that it becomes inactivated and 'dies'. Second, if a gene has multiple functions, these roles can be split between the two copies (subfunctionalization), and both copies will be maintained. Third, a less common outcome is neofunctionalization, in which one gene duplicate evolves a new function. By sequencing the genome of *X. laevis*, which has undergone a WGD, Session *et al.*<sup>1</sup> investigate these processes. (Graphic adapted from an original by Darryl Leja.)

The researchers used these paired BACs to make pairs of DNA 'probes', which bind to the two DNA sequences and are each labelled with a different fluorescent molecule, and then simultaneously hybridized the two probes to intact chromosomes. This process allowed them to assign the two duplicated genes to the correct chromosomes on the basis of which colour probe bound specifically to which chromosome, effectively re-separating collapsed sequences. The technique improved genome assembly by enabling stretches of assembled sequence to be strung together into larger, chromosome-assigned chunks.

The second technique was tethered conformation capture, in which regions of tightly packaged DNA are cross-linked together and the joined DNA pieces are subsequently sequenced as a pair. Most cross-linked sequence pairs come from the same chromosome, but can be up to hundreds of kilobases apart. As such, sequences can be linked to others from the same chromosome, creating larger continuous sequences. Together, these two techniques enabled the separation

of duplicated sequences into distinct chromosomes, resulting in a high-quality genome sequence.

We could stop here and Session and colleagues' work would already be of major interest. But species that have undergone a whole-genome duplication (WGD) also provide an opportunity to watch evolution happen within a species, instead of piecing together evolutionary paths by comparing distinct species. After a WGD, duplicated genes that share the same function can undergo several types of change over time: inactivating mutations can arise in one copy (the mutated copy 'dies'); the original functions can be split between the two copies; or one copy can develop a new function while the other gene retains its ancestral role (Fig. 1). Given enough time, the organism will return to a diploid state, in which all remaining genes have unique and evolutionarily important functions. This process of rediploidization has happened several times during vertebrate evolution<sup>2</sup>.

*Xenopus* is the third vertebrate with a duplicated genome to be sequenced in the past three

years, joining the common carp (*Cyprinus carpio*)<sup>3</sup> and the Atlantic salmon (*Salmo salar*)<sup>4</sup>. Of the three, the WGD in carp occurred most recently, only 8 million years ago. The Atlantic salmon genome was duplicated 80 million years ago, and the *X. laevis* genome provides us with an intermediate, at approximately 34 million years.

Session and colleagues made some interesting observations when looking at within-species evolution in *X. laevis*. They found that protein-coding genes were retained at a higher rate than expected, suggesting that maintaining balanced expression levels is necessary for more genes than previously thought. Conversely, conserved non-coding elements (CNEs) — the regions of the genome most likely to be sequences such as enhancers or promoters that regulate gene expression — were retained at a significantly lower rate. This fits with the idea that regulatory elements have more freedom to change in a duplicated genome, accelerating evolution.

Another interesting phenomenon was that one paired set of chromosomes (dubbed S) was almost four times more likely to have a gene die or be deleted than the other (X). It is unclear why this would occur — perhaps certain aspects of the physiology of the new frog species that emerged from the WGD were more compatible with the ancestor that contributed the X chromosomes, thus favouring retention of this set. Session *et al.* also noticed that certain categories of gene were more likely to be specifically retained in two copies — in particular, those encoding DNA-binding proteins and proteins of developmentally regulated signalling pathways. One reason given by the authors for this is that transcription factors and signalling molecules that often rely on gradients of expression for their effect on development might be more sensitive to alterations in copy number than most other proteins.

Finally, Session and colleagues showed that many pairs of duplicated *X. laevis* genes have divergent spatio-temporal expression. These alterations in gene expression are a good opportunity for scientists to connect gene expression to the molecular evolution of duplicated CNEs. In other words, alterations in enhancer sequences can be correlated with alterations in gene expression, and the causative proof of these changes is particularly compelling because the changes can be measured against another copy of the same gene within the same species.

The genome sequence for the African clawed frog gives us much to celebrate. Developmental biologists now have at their disposal the detailed genomic information so essential to modern biology. Genome biologists have proof that even large, complex genome duplications can ultimately be resolved into high-quality assemblies. And evolutionary

biologists have another powerful tool with which to examine the birth and death of genes and their regulatory elements during evolution. *Xenopus* has made a huge leap forward as a model organism — scientists will surely follow. ■

Shawn Burgess is in the Translational and Functional Genomics Branch, National

Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland 20892, USA.

e-mail: [burgess@mail.nih.gov](mailto:burgess@mail.nih.gov)

1. Session, A. M. *et al.* *Nature* **538**, 336–343 (2016).
2. Ohno, S. *Evolution by Gene Duplication* (Springer, 1970).
3. Xu, P. *et al.* *Nature Genet.* **46**, 1212–1219 (2014).
4. Lien, S. *et al.* *Nature* **533**, 200–205 (2016).

#### ASTROPHYSICS

## Unexpected X-ray flares

Two sources of highly energetic flares have been discovered in archival X-ray data of 70 nearby galaxies. These flares have an undetermined origin and might represent previously unknown astrophysical phenomena. [SEE LETTER P.356](#)

SERGIO CAMPANA

Since our beginning, humanity has observed and chronicled stars, comets and supernova explosions. Nowadays, we have instruments that can cover a wide range of the electromagnetic spectrum, from low-energy radio waves to high-energy X-rays and  $\gamma$ -rays. The high-energy sky is the theatre for many time-varying and energetic phenomena, often involving compact objects such as neutron stars or black holes. On page 356, Irwin *et al.*<sup>1</sup> report what looks like a previously

undescribed kind of bright X-ray flaring activity in neighbouring galaxies.

In 2005, an unexplained source of X-ray flares was discovered<sup>2</sup> in proximity to the galaxy NGC 4697. Motivated by this discovery, Irwin and colleagues analysed archival X-ray data from the Chandra and XMM-Newton space observatories and found two further X-ray sources that exhibit particularly bright flares. The authors' sources are located in the outskirts of nearby galaxies: one near the Virgo galaxy NGC 4636 and the other near the Centaurus A galaxy (NGC 5128; Fig. 1). The



**Figure 1 | The environment of Centaurus A.** Irwin *et al.*<sup>1</sup> observe a source of ultraluminous X-ray flares (indicated by the white circle) near the elliptical galaxy Centaurus A (NGC 5128; centre).

NASA/CXC/UNIV. BIRMINGHAM/M. BURKE ET AL.