Tame affairs: domesticated transposase and domestic pigs

Jean-Nicolas Volff

Our opinion on transposable elements has evolved considerably during the past decade. First considered as genome parasites and selfish elements, mobile sequences have been rehabilitated by the demonstration of their crucial role in gene and genome evolution. Notably, transposable elements have repeatedly contributed regulatory and coding sequences to host genes during evolution. A fascinating facet of the evolutionary impact of transposable elements is their ability to act as a source of new genes with novel functions. In eukaryotes, genes from both retrotransposable elements—which transpose through the reverse transcription of an RNA intermediate—and DNA transposons have been domesticated or exapted to fulfil functions that are useful to the host (Volff, 2006). Proteins encoded by transposons are of obvious interest for the host cell, as they can bind to, replicate or degrade nucleic acids, cut and rearrange DNA, and interact with or process proteins.

Genome sequencing has uncovered many protein-coding genes that seem to be derived from domesticated transposable elements in phyla as diverse as fungi, plants and animals—including humans. These sequences have lost their ability to transpose and are thus single-copy elements located at the same position in the genome of divergent species. The proteins they encode are also conserved, indicating that they are useful to the host, although their functions remain largely unknown. However, two recent papers—one in this issue of *EMBO reports*—show that a domesticated hAT transposase functions as an important transcription factor involved in development, cell proliferation and growth in placental mammals (Butter, 2010; Markjlung, 2009).

The domesticated transposase was identified through a search for genomic sequences that have been involved in domestication

and selection in the pig. An evolutionarily conserved, paternally expressed quantitative trait locus that controls muscle mass and back-fat thickness in European pigs had previously been found in intron 3 of the insulin-like growth factor 2 gene (*IGF2*; van Laere *et al*, 2003). In domestic pigs selected for lean growth, a G to A nucleotide substitution was shown to be associated with threefold increase in *IGF2* mRNA expression in post-natal muscle, as well as with higher muscle content, increased heart size and reduced subcutaneous fat deposition. This mutation was thought to abrogate the binding of a putative transcriptional repressor of *IGF2*, which both studies identify by using DNA–protein interaction screens based on quantitative mass spectrometry. The repressor—a 110 kDa protein known as MGR or ZBED6—is related to transposases encoded by DNA transposons from the hAT superfamily. It has a typical hAT dimerization domain and BED zinc fingers necessary for DNA binding, which suggest it is a domesticated transposase. ZBED6 is encoded by an intronless gene located in the first intron of an unrelated gene and is a member of a family that includes at least five other genes in placental mammals.

Several lines of evidence are presented in both studies to demonstrate that ZBED6 is the genuine repressor of *IGF2*: the protein can bind to the wild-type but not to the mutated regulatory sequence *in vitro* and binds to the *IGF2* locus *in vivo*; its knockdown in a mouse myoblast cell line increases *IGF2* expression and promotes cell proliferation, wound healing and myotube formation; and its ectopic expression represses the activity of a reporter gene under the control of the wild-type *IGF2* regulatory sequence. Together, the results show that the transcriptional repressor ZBED6 is a domesticated transposase with an important role in muscle growth regulation (Fig 1). The mutation of its binding site in *IGF2*

abrogates repression and leads to increased muscle mass and reduced fat deposition in many commercially bred pigs.

In addition, the identification of more than 1,000 ZBED6 binding sites in the mouse genome suggests that many other target genes—and possibly biological processes—are regulated by this transcription factor, which is consistent with the fact that ZBED6 is expressed broadly in mammals. ZBED6 putative target genes are involved in development, transcriptional regulation, cell differentiation, morphogenesis, neurogenesis, cell–cell signalling and muscle development. Notably, the human orthologues of ZBED6 target genes seem to be particularly associated with developmental disorders, cancer, cardiovascular diseases or neurological diseases (Markjlung *et al*, 2009).

There is increasing evidence that transposases from the hAT superfamily function as novel transcription factors in various phyla. For example, an *Arabidopsis* hAT-like transposase encodes a global gene regulator that is essential for normal plant growth (Bundock & Hooykaas, 2005) and other domesticated hAT transposases have been identified in vertebrates, insects, nematodes and plants (Feschotte & Pritham, 2007). In addition, other superfamilies of DNA transposons have evolved into new genes in various organisms, such as the recombination-activating gene *Rag1*, which is derived from a Transib transposon (Kapitonov & Jurka, 2005) and catalyses the somatic recombination responsible for immunoglobulin and T-cell receptor diversity in humans and other jawed vertebrates. Retrotransposable elements are also a source of new genes, such as those derived from retrotransposon *gag* and from endogenous retrovirus *envelope* that are necessary for placenta formation in mammals (Volff, 2006).

The precise evolutionary origin of *ZBED6* remains to be elucidated. At least five other

Fig 1 | Possible evolutionary scenario for *MGR/ZBED6*. Before the divergence between mammals and birds about 300 million years ago, a transposase gene from a hAT transposon was domesticated into a new transcription factor, the molecular progenitor of mammalian ZBED genes. This gene was duplicated, leading to the formation of *MGR/ZBED6* and other paralogues. The binding sites that allow gene regulation by ZBED proteins might be derived from the extremities of the ancestral transposon originally bound by the transposase and their dispersion might have occurred through transposition. During the selection of domestic pigs, the disruption of ZBED6-mediated regulation of *IGF2* led to increased *IGF2* expression, higher muscle mass and reduced fat deposition.

ZBED genes have been identified in mammalian genomes, which might suggest an initial event of molecular domestication followed by sequential events of gene duplication (Fig 1). As at least one ZBED sequence is present in chicken, this domestication event might be at least 300 million years old. Alternatively, several independent events of molecular domestication might have been involved in the formation of the ZBED gene family. *ZBED6* is present and highly conserved in mammals but is a pseudogene in marsupials and monotremes. Its absence from non-mammalian tetrapod species suggests that transposase domestication took place about 180 million years ago, implying that it occurred long before pig domestication—which happened around 9,000 years ago. The origin of the thousands of ZBED6 binding sites dispersed in mammalian genomes remains an open question. One possibility—also suggested in the case of RAG1 (Kapitonov & Jurka, 2005)—is that the extremities of the ancestral hAT transposon bound by the transposase have been co-domesticated to provide binding sites for the new transcription factor. Interestingly, the A/G polymorphism in the regulatory region of *IGF2* that is bound by ZBED6 is directly adjacent to a CAGCG sequence in many

placental mammals (van Laere *et al*, 2003), which matches the consensus sequence (T/C)A(A/G)NG that has been established for terminal inverted repeats of hAT transposons (Rubin *et al*, 2001). To clarify these evolutionary considerations, classical hAT transposons related to *ZBED6* would need to be identified; a quest that should be performed in amphibians, fish or an even more divergent group, as the ancestral transposon is not present in placental mammalian genomes. Domesticated genes and their transposable element progenitor do not generally coexist in genomes, possibly to avoid mobilization of the neogene in *trans* and allow better control of its copy number and expression pattern after domestication.

The studies by the Mann and Andersson groups not only confirm that transposable elements can constitute coding material for the formation of new genes with useful functions for the host, but also demonstrate how transposon-derived DNA proteins can be exapted as master gene regulators for the formation of new regulatory networks (Feschotte, 2008). Future work in the fields of transcriptomics and functional genomics will probably provide insight into the numerous target genes and biological processes regulated by ZBED6 and

other hAT-related genes—at least one other ZBED gene, *ZBED1/hDREF*, also encodes a transcription factor that regulates cell proliferation (Yamashita *et al*, 2007). The approaches taken by these two groups have paved the way for such analyses, which will contribute strongly to our understanding of the evolutionary impact of transposable elements on the evolution of gene networks and biological processes.

REFERENCES

- Bundock P, Hooykaas P (2005) *Nature* **436**: 282–284
- Butter F *et al* (2010) *EMBO Rep* **11**: 305–311
- Feschotte C (2008) *Nat Rev Genet* **9**: 397–405
- Feschotte C, Pritham EJ (2007) *Annu Rev Genet* **41**: 331–368
- Kapitonov VV, Jurka J (2005) *PLoS Biol* **3**: e181
- Markjlung E *et al* (2009) *PLoS Biol* **7**: e1000256
- Rubin E *et al* (2001) *Genetics* **158**: 949–957
- van Laere AS *et al* (2003) *Nature* **425**: 832–836
- Volf JN (2006) *Bioessays* **28**: 913–922
- Yamashita D *et al* (2007) *Mol Cell Biol* **27**: 2003–2013

Jean-Nicolas Volf is at the Institut de Génomique Fonctionnelle de Lyon, Université de Lyon, Université Lyon 1, CNRS, INRA, Ecole Normale Supérieure de Lyon, Lyon, France. E-mail: jean-nicolas.volf@ens-lyon.fr

Published online 26 February 2010

EMBO reports (2010) **11**, 241–242.
doi:10.1038/embor.2010.31