Insight

Traces of strong selective pressures in the genomes of C_4 grasses

Pascal-Antoine Christin

Animal and Plant Sciences, University of Sheffield, Western Bank, Sheffield S10 2TN, UK
p.christin@sheffield.ac.uk

C_4 photosynthesis is nature's response to CO_2 limitations, and evolved recurrently in several groups of plants. To identify genes related to C_4 photosynthesis, Huang *et al.* looked for evidence of past episodes of adaptive evolution in the genomes of C_4 grasses. They identified a large number of candidate genes that evolved under divergent selection, indicating that, besides alterations to expression patterns, the history of C_4 involved strong selection on protein-coding sequences.

The C_4 syndrome relies on a series of anatomical and biochemical adaptations that function together to concentrate CO_2 in some parts of the leaf (Hatch, 1987). This effectively boosts photosynthesis, and increases growth rates in subtropical and tropical conditions (Atkinson *et al.*, 2016). The prospect of improving non-C_4 crops, such as rice and wheat, by engineering an efficient C_4 cycle in them is therefore very appealing, and several projects have been set up in an attempt to realize this ambitious goal. Unfortunately, while the main enzymes of the C_4 pathway were identified long ago and have been characterized in detail, the genetic mechanisms underlying regulation of the pathway, transport of metabolites, and leaf anatomy remain poorly understood.

Engineering a complex biochemical pathway, which requires the action and coordination of numerous proteins, is virtually impossible when some of the underlying genes are yet to be identified. Evolution successfully engineered this intricate pathway, and did it a remarkably large number of times for such a complex trait (Sage *et al.*, 2011). While the details of how this happened are still to be elucidated, the traces of this accomplishment should still be present in the genomes of extant species. Any single genome consists of a 'long list of letters' that is difficult to decipher, and yet the comparison of multiple genomes has the power to reveal changes that happened during evolution. Obviously, the significance of these changes is another problem, but past evolutionary pressures left specific footprints on the small fraction of genomes that correspond to protein-coding genes.

Because each amino acid can be encoded by different nucleotide triplets, some nucleotide changes (substitutions) do not affect the protein. These are synonymous substitutions, while non-synonymous substitutions change the amino acid and so result in a slightly different protein. Under a purely stochastic model, the rates of fixation of these two types of substitutions should be similar and, as such, their ratio (dN/dS) should equal one (Yang, 1998). Most non-synonymous changes will, however, be detrimental and thus preferentially removed by selection, leading to an observed dN/dS much smaller than one in most cases. Exceptionally, when a change in the catalytic properties of the encoded enzyme benefits the organism, the rate of fixation of non-synonymous substitutions will increase, leading to an observed dN/dS that can exceed one, at least for some parts of a gene. Such instances of positive selection are classically associated with 'arms races' between hosts and pathogens, leading to sustained elevated dN/dS throughout the history of the gene (Endo *et al.*, 1996). However, episodic changes to the catalytic environment can also increase dN/dS for limited periods, corresponding to a few branches of a phylogenetic tree. Huang *et al.* looked for such traces of past episodes of adaptive evolution linked to C_4 photosynthesis by comparing the genomes of C_4 and non-C_4 grasses (Box 1).

Evidence of past positive selection reveals candidate genes for C_4 photosynthesis

Tracking evolutionary modifications to identify changes linked to C_4 photosynthesis is not a new idea. The many independent origins of the C_4 trait make it particularly amenable to comparative studies, enabling identification of the ecological and physiological changes linked to its evolution (e.g. Edwards and Smith 2010; Atkinson *et al.*, 2016). In recent years, attempts to identify all of the C_4-related genes similarly relied on evolution-based comparisons, but these mainly focused on gene expression (Brautigam *et al.*, 2014; Mallmann *et al.*, 2014). It is only more recently that attention has turned to genomic changes, such as duplication of genes (Emms *et al.*, 2016). While adaptive evolution of C_4 enzymes involving kinetic changes has been reported (Svensson *et al.*, 2003; Christin *et al.*, 2007), one might hypothesize that this would concern only a handful of enzymes – specifically, those

Box 1. Adaptive evolution in C$_4$ grasses

Phylogeny of the six species included in the analysis from Huang *et al.*: maize (*Zea mays*), *Sorghum bicolor*, *Setaria viridis* (wild progenitor of *Setaria italica*), *Dichanthelium oligosanthes*, *Brachypodium distachyon* and rice (*Oryza sativa*). Red branches indicate where the C$_3$ to C$_4$ transitions occurred. Photos courtesy of Pu Huang, James Schnable and Elizabeth Kellogg.
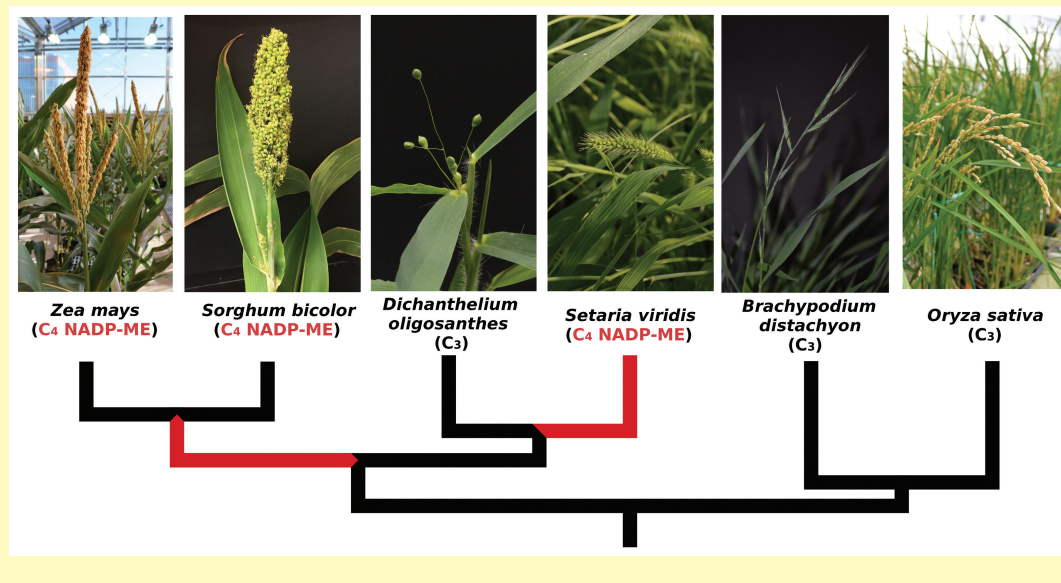
linked to core C$_4$ reactions and their very high catalytic rates. Huang and colleagues decided to challenge this assumption and adopt bioinformatic approaches to identify all the genes that evolved under elevated dN/dS, specifically in C$_4$ grasses.

After screening the genomes of six grasses, including three C$_4$ taxa belonging to two independent C$_4$ origins (Box 1), Huang *et al.* identified 88 genes that evolved under elevated dN/dS on branches belonging to one or several of the C$_4$ lineages. This type of genome scan is inherently subject to false positives. In addition, the methodology cannot strictly differentiate between adaptive evolution and relaxed selection. Finally, the genes might have been under divergent selection along these branches for reasons other than C$_4$ evolution. Fortunately, the putative link with the C$_4$ trait was confirmed for a number of candidates by independent evidence, including *a priori* knowledge for a few of them and high expression in C$_4$ tissues for many others. The list produced by Huang *et al.* therefore includes many promising candidates, some of which might be linked to C$_4$ anatomy. If confirmed, their identification would represent a major breakthrough for the engineering of C$_4$ photosynthesis into non-C$_4$ crops. In the short term the results already affect the way we should envision C$_4$ evolution.

Physiological innovation through adaptive evolution of numerous protein-coding genes

For the most part, previous studies have linked phenotypic variation to alterations in gene expression and regulation (King and Wilson, 1975; Brawand *et al.*, 2011). While these have certainly also played a key role in the evolution of C$_4$ photosynthesis, the new results show that the modification of promoter sequences and regulatory networks is only one part of the story, which also includes adaptive changes in the coding sequences of a large number of genes. This is impressive, providing even more evidence that the recurrent transition to C$_4$ photosynthesis represents a considerable evolutionary feat. The observations of Huang *et al.* also reveal a new set of questions; in particular, why did the coding sequences of so many proteins need to be adapted, both in terms of biochemical properties and evolutionary drivers? The biochemical component of this question will remain unanswered until extensive characterization is performed, and yet hints about the evolutionary pressures can already be proposed.

The precise timing of positive selection episodes is beyond the scope of this comparative work because of the limited number of species sampled, which corresponds to the few grasses for which a complete genome is currently available. Indeed, these episodes are inferred along phylogenetic branches that expand from the last divergence of C$_4$ and non-C$_4$ taxa to the first divergence of two C$_4$ taxa within the same lineage (Box 1). With only six species, this interval is initiated long before the transition to C$_4$ photosynthesis and continues for a long period after C$_4$ evolved, spanning up to 20 million years of changes. As more genomes become available, similar analyses will be able to pin down the timing of these episodes of positive selection with more precision. Until then, we can only speculate.

As with any complex trait, the numerous changes that define extant C$_4$ plants were probably spread over long

periods of evolutionary time, from the occurrence of capacitating mutations in non-C_4 ancestors to changes directly responsible for the emergence of a C_4 physiology, and continuous adaptive alterations after its origin (Christin and Osborne, 2014). Modelling efforts predict that changes in expression patterns can lead to the emergence of a C_4 cycle in plants with C_4-like anatomical traits (Heckmann *et al.*, 2013; Mallmann *et al.*, 2014). However, evolution did not stop after the initial transition to C_4 photosynthesis, and the presence of a working C_4 cycle, even if rudimentary, probably created a selective impetus for the fixation of substitutions that improved the C_4 syndrome. The genes detected by Huang *et al.* probably underwent adaptive mutations that improved the fit of the proteins to the new catalytic environment. Their impressive number suggests that the selective pressure for improving the C_4 syndrome was very strong or maintained over a long evolutionary period, possibly throughout the diversification of C_4 plants.

Until now, research on C_4 evolution has focused mainly on the events that led to a C_4 cycle, largely ignoring those that followed its emergence. The discovery of widespread C_4-related selection on coding genes should motivate new research into the changes that contributed to the improvement or diversification of the C_4 syndrome. A first step in this direction is to acknowledge the diversity of C_4-related traits within each C_4 lineage, and design comparative experiments that capture this diversity. With continuous advances in sequencing technology, this goal might soon become achievable for comparative genomics, contributing towards a full elucidation of the changes that were selected, both before and after the first C_4 plants emerged.

Acknowledgements

References

Atkinson RR, Mockford EJ, Bennett C, Christin PA, Spriggs EL, Freckleton RP, Thompson K, Rees M, Osborne CP. 2016. C_4 photosynthesis boosts growth by altering physiology, allocation and size. Nature Plants **2,** 16038.

Bräutigam A, Schliesky S, Külahoglu C, Osborne CP, Weber AP. 2014. Towards an integrative model of C_4 photosynthetic subtypes: insights from comparative transcriptome analysis of NAD-ME, NADP-ME, and PEP-CK C_4 species. Journal of Experimental Botany **65,** 3579–3593.

Brawand D, Soumillon M, Necsulea A, *et al*. 2011. The evolution of gene expression levels in mammalian organs. Nature **478,** 343–348.

Christin PA, Osborne CP. 2014. The evolutionary ecology of C_4 plants. New Phytologist **204,** 765–781.

Christin PA, Salamin N, Savolainen V, Duvall MR, Besnard G. 2007. C_4 photosynthesis evolved in grasses via parallel adaptive genetic changes. Current Biology **17,** 1241–1247.

Edwards EJ, Smith SA. 2010. Phylogenetic analyses reveal the shady history of C_4 grasses. Proceedings of the National Academy of Sciences, USA **107,** 2532–2537.

Emms DM, Covshoff S, Hibberd JM, Kelly S. 2016. Independent and parallel evolution of new genes by gene duplication in two origins of C_4 photosynthesis provides new insight into the mechanism of phloem loading in C_4 species. Molecular Biology and Evolution **33,** 1796–1806.

Endo T, Ikeo K, Gojobori T. 1996. Large-scale search for genes on which positive selection may operate. Molecular Biology and Evolution **13,** 685–690.

Hatch MD. 1987. C_4 photosynthesis: a unique blend of modified biochemistry, anatomy and ultrastructure. Biochimica et Biophysica Acta **895,** 81–106.

Heckmann D, Schulze S, Denton A, Gowik U, Westhoff P, Weber AP, Lercher MJ. 2013. Predicting C_4 photosynthesis evolution: modular, individually adaptive steps on a Mount Fuji fitness landscape. Cell **153,** 1579–1588.

Huang P, Studer AJ, Schnable JC, Kellogg EA, Brutnell TP. 2017. Cross species selection scans identify components of C_4 photosynthesis in the grasses. Journal of Experimental Botany **68,** 127–136.

King MC, Wilson AC. 1975. Evolution at two levels in humans and chimpanzees. Science **188,** 107–116.

Mallmann J, Heckmann D, Bräutigam A, Lercher MJ, Weber AP, Westhoff P, Gowik U. 2014. The role of photorespiration during the evolution of C_4 photosynthesis in the genus *Flaveria*. elife **3,** e02478.

Sage RF, Christin PA, Edwards EJ. 2011. The C_4 plant lineages of planet Earth. Journal of Experimental Botany **62,** 3155–3169.

Svensson P, Bläsing OE, Westhoff P. 2003. Evolution of C_4 phosphoenolpyruvate carboxylase. Archives of Biochemistry and Biophysics **414,** 180–188.

Yang Z. 1998. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. Molecular Biology and Evolution **15,** 568–573.