The Evolution Of The Folate Biosynthesis Gene Fusions In The Eukaryotes

Finlay Maguire, Guy Leonard and Thomas Richards

Folate Biosynthesis Pathway

- Folate is an essential metabolite in single carbon transfer reactions such as those involved in the biosynthesis of:
  - Adenine
  - Purine
  - Methione
  - Histidine
  - Formyl-tRNA (1)
- The loss of this pathway in the animals and its high degree of conservation makes it a potent anti-pathogen drug target (2)
- Biosynthetic enzymes in this pathway have been discovered in both monofunctional unfused forms and as a series of bi-functional and tri-functional fusion genes

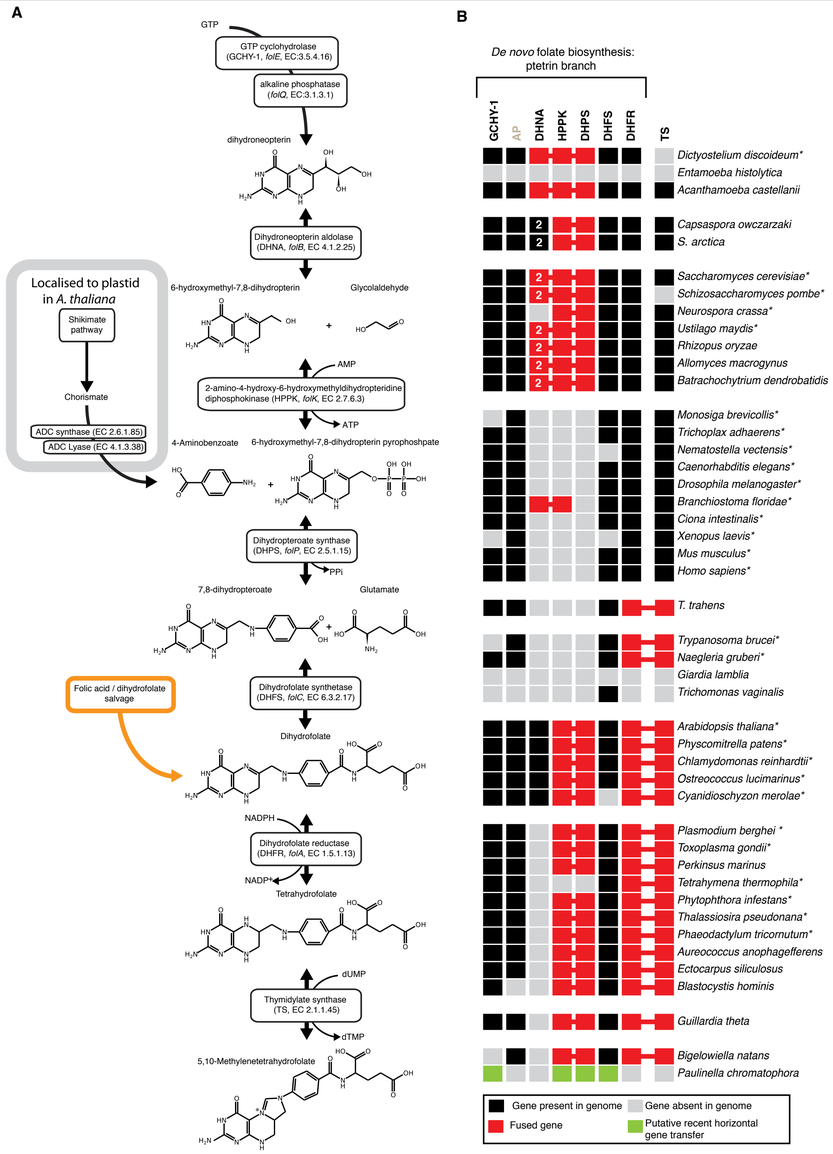

FIGURE (A): shows the pterin branch section of the folate biosynthetic pathway in which the DHNA-HPPK, HPPK-DHPS, DHFR-TS, and DHNA-HPPK-DHPS fusion forms occur. Data from the KEGG database.

FIGURE (B): shows the distribution of fused and unfused forms of the folate biosynthetic enzymes shown in figure A as determined by automated annotation of phylogenies using a modified version of the program featured in (3). Two on the figure indicates a duplication of that domain and the asterixes mark the taxa that are in the KEGG database.

Gene fusions

- Gene fusions occurs when two or more open reading frames (ORFs) become a single ORF (6)
- A gene fusion event occuring mainly by:
  - Chrosomal inversion
  - Interstitial deletion
  - Chrosomal translocation (5)
- Gene fusion/fissions, introns, insertion/deletions and horizontal gene transfers are rare genomic changes (RGCs) and can be used as a shared derived character (SDCs)

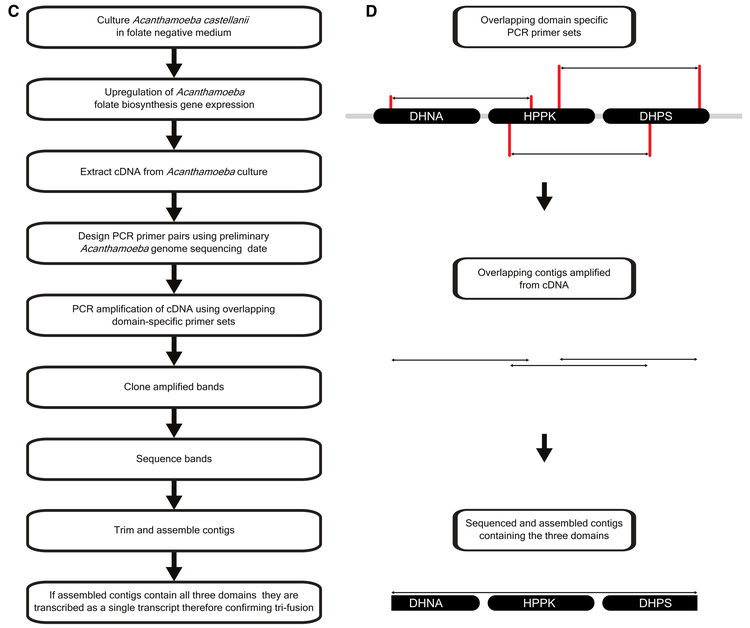

FIGURE (C): shows the proceedure used to determine whether the DHNA, HPPK and DHPS domains in *Acanthamoeba castellanii* are transcribed as a single gene fusion.

FIGURE (D): shows a simplified diagrammatic overview of the overlapping primer sets used to demonstrate that the folate biosynthesis genes are transcribed as a single transcript.

Shared derived characters

- SDCs are traits which have evolved in a common ancestor and are shared by the descendents of this ancestor
- The presence or absence of this structure can thus be used as a cladistic device (allowing trait polarisation)
- This can be useful in identifying deep-branching clades within the tree of life

Limitations

- The usefulness of a gene fusions as an SDCs is limited by four potential evolutionary phenomena which must be accounted for (6,7):
  - HGT
  - Hidden paralogy
  - Reversions (including both loss and fissions)
  - Convergent evolution

Acanthamoeba castellanii

- The DHNA-HPPK-DHPS tri-fusion has currently only been identified in the mycetozoan slime moulds
- *Acanthamoeba* represents a distant releative of the mycetozoa within the Amoebozoa
- Therefore, by identifying the presence of this gene fusion within *Acanthamoeba* it demonstrates the presence of the fusion in two highly divergent branches of the Amoebozoa and thus suggests the character across the Amoebozoa as a whole

Why is the folate tri-fusion a potentially useful SDC?

- Folate biosynthesis bi-fusions have been used to polarise ancient relationships within the eukaryotes (8)
- DHNA-HPPK-DHPS tri-fusion is most likely the product of 2 independent and rare fusion events therefore the probability of multiple incidences of the triple domain fusion occuring is relatively low compared to single RGC fusion characters such as DHFR-TS fusion and it is therefore less likely to independently arise in different taxa
- The folate biosynthesis pathway is present across the domains suggesting it is an ancient innovation
- By combining comparative genomics and use of concatenated phylogenies it was possible to identify and account for the phenomena limiting the use of gene fusions as SDCs in the case of this tri-fusion

Findings

- By sequencing the tri-fusion in *Acanthamoeba castellanii* it was possible to conclude that due to its presence in the mycetozoan slime moulds the tri-fusion is an SDC for unikonts with secondary losses
- Using phylogentics and comparative genomics it was possible to determine several key events in the evolution of the folate biosynthesis genes within the eukaryotes:
  - HPPK-DHPS bi-fusion as putative 'synapomorphy' for the eukaryotes (with secondary losses)
  - DHNA-HPPK-DHPS tri-fusion identified as possible 'synapomorphy' for the unikonts (Amoebozoa and Opisthokonta)
  - Duplication of the DHNA domain in the Dikarya (Fungi)
  - *Paulinella chromatophora* horizontal gene transfer

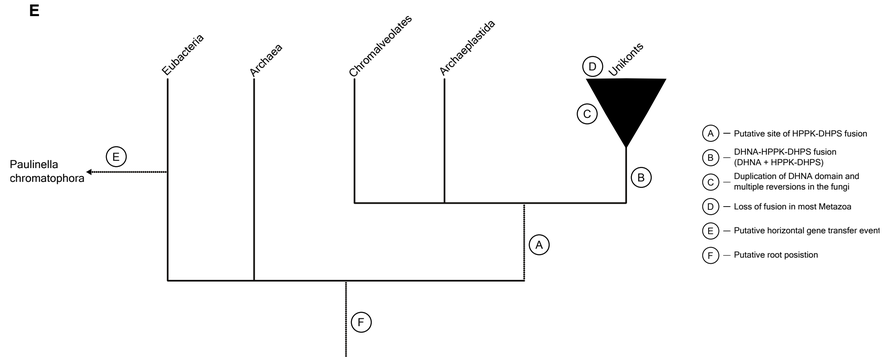

FIGURE (E): Evolutionary relationships predicted from the folate biosynthesis genes phylogenies and comparison of fusion states across taxa.

References

1. BROWN, G. M. 1971. The biosynthesis of pteridines. Advan. Enzymol. Relat. Areas Mol. Biol., 35, 35-77.
2. LAWRENCE, M., ILIADES, P., FERNLEY, R., BERGLEZ, J., PILLING, P. & MACREADIE, I. 2005. The three-dimensional structure of the bifunctional 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase/dihydropteroate synthase of Saccharomyces cerevisiae. Journal of Molecular Biology, 655-670.
3. WICKSTEAD, B., GULL, K. & RICHARDS, T. 2010. Patterns of kinesin evolution reveal a complex ancestral eukaryote with a multifunctional cytoskeleton. Bmc Evolutionary Biology
4. DOOLITTLE, R. F. 1995. The Origins and Evolution of Eukaryotic Proteins. Philosophical Transactions: Biological Sciences, 349, 235-240.
5. LEONARD, G. 2010. Development of fusion and duplication finder BLAST (fdfBLAST): systematic tool to detect differentially distributed gene fusions and resolve trifurcations in the tree of life. PhD Thesis, University of Exeter.
6. RICHARDS, T. 2005. Horizontal Gene Transfer and the Evolution of the Eukaryotes, D. Phil Thesis, University of Oxford.
7. SIMPSON, A. G. B. & ROGER, A. J. 2002. Eukaryotic evolution: Getting to the root of the problem. Current Biology,
12, R691-R693.
8. STECHMANN, A. & CAVALIER-SMITH, T. 2002. Rooting the eukaryote tree by using a derived gene fusion. Science,
297, 89-91.

Acknowledgements

Research was funded through an SGM summer studentship.
Acanthamoeba cDNA provided by Fiona Henriquez at the University of the West of Scotland