The Tic20 gene family: phylogenetic analysis and evolutionary considerations

Mats Töpel¹ and Paul Jarvis²*

¹Department of Plant and Environmental Sciences, Göteborg University, Box 461, SE-405 30 Göteborg, Sweden.

²Department of Biology, University of Leicester, University Road, Leicester LE1 7RH, United Kingdom.

Addendum to:

Submitted: 25 March 2011
Accepted: 

*Correspondence to:
Paul Jarvis; Email: rpj3@le.ac.uk

Abstract

Tic20 is a polytopic protein of the inner envelope membrane of chloroplasts, and it is proposed to act as a translocation channel during chloroplast protein import. By analysing 29 sequences from diverse organisms, it was evident that Tic20-related proteins form two distinct clades, termed Group 1 and Group 2. The former group includes canonical Tic20 proteins that are essential for chloroplast development, while members of the latter are of unknown function. An increased evolutionary rate, in connection with adaptation to terrestrial life, was detected in Group 1. Interestingly, the sub-cellular (genomic) localization of genes coding for Group 1 proteins differs between evolutionary lineages.
Chloroplast Protein Import and Tic20

While chloroplasts retain a functional endogenous genetic system as a vestige of their pre-endosymbiotic existence as a free-living photosynthetic prokaryote, the vast majority of the ~3000 proteins in these organelles are nucleus-encoded and must be imported from the cytosol. Most imported proteins possess an N-terminal extension, or transit peptide, and this is recognized by a protein import apparatus comprising the TOC and TIC (translocon at the outer / inner envelope membrane of chloroplasts) complexes at the organelle’s periphery.\(^1,2\) Several components of these translocon complexes have been identified through biochemical analyses of isolated pea (\textit{Pisum sativum}) chloroplasts. One such component is Tic20 of the inner envelope membrane, which has topological characteristics that led to its proposed role as a translocation channel.\(^3,4\)

The Evolution of the Tic20 Gene Family

Four Tic20-homologous proteins are found in \textit{Arabidopsis thaliana}.\(^5,6\) The roles of two of them (termed atTic20-II and atTic20-V) are still unknown, as no mutant phenotypes have been detected in the mutant genotypes analysed.\(^7\) Nonetheless, these genes are expressed at high levels and have been conserved over ~1.2 billion years (Ga) in many evolutionary lineages (see below). Hence, it seems likely that they do have a function, yet to be discovered, although this function may turn out to be unrelated to that of the other two proteins (atTic20-I and atTic20-IV), which are orthologous to pea Tic20 and essential for viability.\(^7,8\)

In a recent study\(^7\), we conducted extensive BLAST searches of the NCBI (www.ncbi.nlm.nih.gov) and JGI (genome.jgi-psf.org) repositories, using the Arabidopsis and pea proteins as query sequences, to reveal that Tic20 homologues are present in cyanobacteria, red and green alga, chromalveolates and land plants. Phylogenetic analyses of these sequences helped to explain \textit{in vivo} results from genetic
analyses, and provided a new understanding of the evolution of the Tic20 gene family.

For the analysis, we used a Bayesian method implemented in the software MrBayes that takes into account the uncertainty in the analysis when estimating the tree topology and branch lengths.\textsuperscript{9} Instead of searching for the optimal tree during the analysis (like in maximum likelihood and maximum parsimony methods), trees are sampled according to their posterior probability (the posterior probability of a tree can be interpreted as the probability that the tree is correct). Features that are common among the sampled trees, like for example clades or branch lengths, can then be distinguished and summarized in a consensus tree (see ref. 10 for an accessible review). A further advantage of the method is that instead of selecting an amino acid substitution model (e.g. Dayhoff, Blosum62 or Wag; refs. 11-13) \textit{a priori}, all of them can be used. Each model will then be allowed to contribute to the result in proportion to its posterior probability. This is a great advantage compared to the often used maximum likelihood method that has been shown to be inconsistent (i.e. it will give the wrong tree even if the amount of data is increased) when the evolutionary model used is inappropriate (e.g.\textsuperscript{14}). Differences between the methods used for phylogenetic reconstruction of the Tic20 gene family are possible explanations for why our results differ from previous analyses.\textsuperscript{6}

With this method, we found that the oldest gene duplication in the tree happened before the split between red algae (represented in the tree by \textit{Cyanidioschyzon merolae}) and viridiplantae (represented by various taxa of green algae, bryophytes and angiosperms), resulting in two clades termed Group 1 and Group 2 (Figure 1); the oldest fossils of a red alga have been found in the Huntington formation (Somerset Island, Canada), and have been dated to 1.2 Ga ago.\textsuperscript{15} On the other hand, cyanobacteria are not represented in either of Groups 1 and 2, but form the outgroup of the eukaryote proteins. Thus, the first gene duplication must have happened at least 1.2 Ga ago, but after the emergence of chloroplasts in the eukaryotic lineage (~1.5 Ga ago).
The Group 1 protein of the red alga *C. merolae* is plastid-encoded, while its parologue in Group 2 is nucleus-encoded. Another potentially significant observation we have made is that *Ectocarpus siliculosus* and *Fucus vesiculosus* (two brown macro algae) each have just one Tic20 gene, and both of these genes are plastid-encoded; no Group 2 proteins from these species were found in our analysis. Hence, the subcellular localization of Tic20 genes differs between evolutionary lineages, which indicates that nuclear localization of this gene is not essential for its function. *Ectocarpus* and *Fucus* belong to the Chromalveolata kingdom, and are believed to have acquired their plastids by secondary or tertiary endosymbiosis with red algae (see ref17 and references therein). Therefore, a plausible explanation for the observed pattern in these two species is that a nucleus-encoded Tic20 copy from the red alga (presumably a Group 2 protein) has been lost, and that only the Group 1 protein is required for a functional chloroplast in these species.

**Peculiar Phylogenetic Patterns**

During the course of our analysis, we also found peculiar phylogenetic patterns where branch lengths differed significantly between clades of paralogous proteins. We interpret this as an indication that an ancestral protein of one of the clades (branch 5 in Group 1, Figure 1) has experienced an increased mutation rate. Branch length in a phylogeny is a measure of the amount of evolution that has taken place between nodes on the tree. Hence, variation in branch length following a gene duplication means that the two copies have evolved at different rates. This may also indicate that a shift in selection pressure on one of the two copies has occurred. Whether this change in mutation rate involved directed positive selection (where natural selection favours mutations leading to changes in the amino acid sequence of the protein) or a relaxation of purifying selection (where natural selection loosens its constraints on the gene, resulting in an accumulation of mutations) remains to be investigated.

Interestingly, the two Arabidopsis homologues, atTic20-I and atTic20-IV, as well as the *Pisum sativum* homologue, psTic20, are derived from the rapidly evolving ancestral protein of Group 1.
Evidently, the mutations that accumulated along branch 5 have not been disadvantageous (they were perhaps even favourable) for the function of proteins derived from this branch. Furthermore, the Tic20 homologue in *Toxoplasma gondii* (a chromalveolate) is essential for apicoplast protein import, and is also part of Group 1 (Figure 1). This implies that proteins of Group 1 have been functional in the TIC complex for ~1.2 Ga, and have been able to retain their function during the period of inferred rapid evolution (less parsimonious explanations are that Tic20 has been incorporated into the TIC complex at different times in different lineages, or that it was temporarily lost and later regained its role as an essential component in chloroplast protein import).

As seen in Figure 1, the evolutionary burst in the Tic20 ancestor of Group 1 happened prior to, or during, adaptation to land life in plants. A variety of physiological and biochemical adaptations to the stresses associated with land life (e.g. desiccation, temperature, UV exposure) have been suggested, including the synthesis of phenolics, development of heat shock proteins, isoprene emission, and oxidative stress response (summarized in). It is reasonable to speculate that the TOC/TIC apparatus was also affected by the transition to land life, either directly or indirectly, and that it too has undergone adaptation to conform to terrestrial existence.

**References**


Figure Legend

Figure 1

Phylogenetic analysis of a subset of the amino acid sequences analysed in ref7. Posterior probability support (pp.) is indicated above the branches, and only branches with a pp. value >0.95 are shown. The subset of 29 protein sequences was prepared and analysed in the same way as the 59 sequences from the original analysis. The outgroup is here represented by the cyanobacterium, *Nodularia spumigena*. The dataset was also analysed with a more extensive selection of cyanobacterial sequences in order to determine if the gene duplication (1) happened in a prokaryotic organism before being engulfed by a eukaryote, or in the host organism before the split yielding red algae and viridiplantae (2 and 3); our results supported the latter possibility (data not shown). After the split (4) between green algae and land plants, the ancestral Group 1 Tic20 protein evolved rapidly which is evident from the inferred long branch (5); branch length (0.85 in this case) is equal to the expected number of amino acid changes per site. Whether the rate of mutations was elevated along the whole length of the branch, or only for short periods, cannot be determined without samples from more green algal groups. In Group 2, a second gene duplication (6) happened, presumably before land plants emerged; only one Group 2 sequence from each of *Volvox* and *Chlamydomonas*
was found in our analysis, suggesting that this duplication probably happened after the split between green algae and viridiplantae. The two branches (7 and 8) leading to land plant proteins are much shorter (0.37 and 0.18, respectively) than branch 5. Hence, approximately 2.3-4.7 times more amino acid substitutions per site are inferred to have happened in the ancestral Group 1 protein prior to, or during, adaptation to land life. Names on a black background are proteins that *in vivo* or biochemical experiments have shown to be important for plastid biogenesis. The *Pisum sativum* protein is designated psTic20, and *Arabidopsis thaliana* proteins are termed atTic20[-I, -IV, -II, -V] with a succeeding roman figure indicating the chromosome it is part of.