Evolution of bacterial diversity and the origins of modularity

Paul B. Rainey a,b,∗, Tim F. Cooper a

a School of Biological Sciences, University of Auckland, Private Bag 92019, Auckland, New Zealand
b Department of Plant Sciences, University of Oxford, South Parks Road, Oxford, OX1 3RB, UK

Received 1 October 2003; accepted 23 January 2004
Available online 19 March 2004

Abstract

A characteristic feature of all organisms is modular organisation: the tendency for groups of genes to interact in such a way as to limit the extent of pleiotropic effects among characters belonging to different functional complexes. While the implications of modularity for the evolution of variability have been much discussed the evolutionary origins remain obscure. Here we develop a model, with special reference to signal transduction cascades of bacteria, which predicts that in the face of ecological opportunity and lateral gene transfer, selection will favour modular genome architectures because such architectures minimise the pleiotropic effects associated with accommodation of potentially beneficial foreign DNA.

© 2004 Elsevier SAS. All rights reserved.

Keywords: Modularity; Bacterial diversity; Signal transduction

1. Introduction

The two domains of prokaryotic microbes, the Bacteria and Archaea, encompass the majority of genetic and phenotypic diversity within the biosphere. Extreme physiological diversity means that microbes are found in almost every conceivable environment, from deep sea thermal vents to the upper atmosphere; at extremes of acidity and alkalinity—even in environments where energy is derived solely from oxidation of obscure organic and inorganic compounds. The evolutionary causes of this diversity are of interest.

In a recent article Spiers et al. [46] considered the ecological and genetic causes of diversity within the genus Pseudomonas—one of the most phenotypically, genetically and ecologically diverse groups of Bacteria known. The genus Pseudomonas was singled out, not only because it is particularly species-rich (while at the same time forming a well circumscribed and coherent group), but also because even at the level of individual species, for example, P. aeruginosa, strains can be isolated from ecologically disparate environments, including soil, water, and the tissues and surfaces of plants and animals [47]. Thus, much of the diversity that manifests at the level of the genus can also be found within a single species. Such diversity and apparent adaptability led to questions concerning the causes: in particular, whether there might be some feature of the Pseudomonas genotype–phenotype map, that is, the mechanism by which genetic changes are translated to phenotypic variability that makes it especially evolvable.

Of the various possibilities, particular attention was paid to genetic mechanisms likely to increase the production of heritable phenotypic variability. The likelihood that Pseudomonas has evolved mechanisms for elevating mutation rate under stress was explored; while an attractive idea, it is one that has little support [29]. The possibility that Pseudomonas might be rich in hyper-mutable loci, such as the “contingency loci” found in certain pathogenic bacteria was also considered [25]. These loci, characterised by short, slippage-prone, repeated DNA sequences enable bacterial populations to generate variability within a short period of time. Analysis of the published Pseudomonas genomes reveals scant evidence of contingency loci. Finally the possibility that Pseudomonas might be more recombinogenic than other bacteria was explored, with such a capacity perhaps manifesting through special mechanisms for the uptake and/or successful incorporation of foreign DNA via lateral gene transfer. Not only is natural competency rare in Pseudomonas, but mosaicism (a hallmark of recombination)
is not especially pronounced in the genus [1,7,19,42,45]. Moreover, population genetic data indicate that overall Pseudomonas is in linkage disequilibrium [15,19].

These findings led Spiers et al. [46] to reject the notion that mechanisms causing high levels of genetic variability were responsible for the exceptional diversity of Pseudomonas. They subsequently concluded that diversification of the genus reflected nothing more (or less) than the outcome of a long evolutionary history (bringing with it a high cumulative opportunity for lateral gene transfer) in spatially and temporally complex environments. However, in a final note, these authors drew attention to the large genome of Pseudomonas (on average 6.5 Mbp) and the fact that it contains a relatively high proportion of genes encoding components of signal transduction pathways (Fig. 1). They then drew attention to the special modular arrangement of signal transduction pathways [17,20,50] and suggested that this modularity might confer a form of evolvability, more subtle in mechanism than had initially been envisaged. Indeed they even went so far as to suggest that the modular arrangement of signalling cascades might itself have been selected for by pressure to properly regulate and accommodate incoming DNA. Here we wish to develop this idea and to do so in general terms with no particular emphasis on Pseudomonas.

2. The evolutionary causes of modularity: a hypothesis

Lateral gene transfer has played, and continues to play, a central role in the evolution of bacterial diversity. Newly acquired sequences encoding genes (and even entire operons) provide opportunity for evolution to proceed in leaps [21,28]. At the same time, the number of evolutionarily successful recombinants arising from lateral gene transfer is likely to be but a small fraction of the total number of transfer events. Crucial for the long-term persistence of newly acquired DNA is that it confers a selectable function. But even in the face of a selectable phenotype, integration of foreign DNA is likely to carry significant costs. To our knowledge such costs have not been directly measured, but costs due to antagonistic pleiotropy are a common feature of adaptive evolution [51], for example, spontaneous antibiotic resistant mutants, are typically less fit than the ancestral sensitive type in the ancestral environment [43], as are phage-resistant types [2]; plasmid acquisition also carries a measurable cost [4,48]. Integration of novel DNA via lateral gene transfer is also likely to incur costs, for example, due to the disruption of genes at the site of insertion. Even if foreign DNA arrives complete with its own regulatory elements, the new regulatory (and structural) genes must still be accommodated within existing regulatory networks; expression is likely to need adjustment via alterations in regulatory sequences and regulatory cascades, critical levels of key signalling molecules (for example, ppGpp, cyclic-di-GMP) must be rebalanced, and so forth [46]. Nevertheless, despite the costs, the potential benefits arising from rapid and efficient integration of novel operons within bacterial genomes are considerable. As a consequence, selection might be expected to favour the evolution of mechanisms that minimise the negative impact of potentially beneficial foreign DNA on the host cell and thus maximise the likelihood that phenotypically useful DNA will be successfully integrated into recipient genomes. We suggest that modularity, especially of the type represented by signal transduction pathways, may have been one result of this selection.

3. Evolvability and modularity

Much has been written on the subjects of evolvability and modularity. Our intention here is to provide the briefest of overviews (for additional information see [3,20,22,31,50]).

Evolvability is the capacity of a genome to generate heritable, phenotypically useful variation and depends critically on the way that genotype and phenotype are connected [41,50]. A characteristic feature of all organisms and particularly notable in bacterial genomes is a modular organisation: the tendency for groups of genes to interact in such a way as to limit the extent of pleiotropic effects among characters belonging to different functional complexes (Fig. 2). This means that changes that occur within one complex have less chance of impacting negatively on others. Striking evidence of modularity stems from analysis of the phenotypic consequences of individually deleting all two-component regulators from the Escherichia coli genome [30,52]. In that study, nearly half of the deletions had no discernable phenotypic effect even when measured across hundreds of
different environmental conditions. Modular structures possess additional features that are amenable to the generation of variability. In addition to pleiotropic robustness, the organisation of organisms into parts means that the parts function as building blocks that can be reused in various combinations to increase the probability of generating phenotypically viable (and potentially useful) variation [20].

Bacterial signal transduction cascades provide a good example of modular structures [17,31,37]. On the one hand regulatory components usually show high fidelity and regulate a small number of structural genes which means that changes in one regulatory module are less likely to impact negatively on other regulatory modules [30,52] (Fig. 2). In addition, the regulatory components exhibit 'genetic' modularity which confers properties relevant to evolutionary change [20]. These properties include versatile protein elements and weak linkage among components. Versatility among protein elements is particularly common in protein kinases where minor mutational modification can alter the specific target of activity or affect the timing of activation [17]. Weak linkage is also a feature of signal transduction pathways; components have switch-like properties and signals act to release the activity, but do not act instructively. Such regulatory organisation facilitates a component's accommodation to novelty and reduces the cost of generating variation (for more information see [12,20]). Despite much work on the evolutionary consequences of modularity the origins remain obscure [11,20,22,50]. In the following section we outline a hypothesis that posits modularity to be a derived property of organisms that arose in the course of evolution, in part because of selection for genome architectures that minimise the pleiotropic effects associated with accommodation of newly acquired DNA sequences by lateral gene transfer.

4. A simple thought experiment for the evolution of modularity

Imagine a hypothetical bacterial lineage: the one in mind is ancient—on the order of 3.5 billion years old—it is a minimal chemotroph, with the pathways of central metabolism (glycolysis and the TCA cycle) intact (see [9]). It has a small genome that encodes a single extracellular enzyme, which enables utilisation of a single complex carbon source. The genotype–phenotype map is simple and based on a one (regulator) to many (genes) arrangement as depicted in Fig. 3a.

Assuming that the environment in which this population grows is homogeneous, with a single limiting carbon source supplied at a constant rate, then diversification will be minimal (although not necessarily non-existent: see [34, 39,40]). To a large extent evolution will proceed by clonal replacement with any diversity that arises by recurrent mutation being purged each time a new beneficial mutant arises and sweeps through the population. This model of evolution is likely to remain unaltered even in the face of substantial lateral gene transfer. Imagine the sudden introduction of a source of novel DNA from an organism that inhabits a different ecological niche. Incoming DNA sequences may well be successfully incorporated into the genomes of some of the evolving individuals, but opportunities for diversification will be limited given that the environment affords little ecological opportunity. Moreover the hybrid genomes are likely to be less fit [35,36]. Ecological opportunity, as empirical studies have shown, is crucial for diversification [5,33].

Imagine now that this evolving population finds itself in an environment with multiple carbon sources and thus subject to strong diversifying selection. Concomitant with enhanced ecological opportunity is diversification: mutation provides the variation upon which selection acts—competition among variant types fuels the process of diversification [34]. Mutants that gain additional metabolic benefit from one of the new resources enjoy a strong frequency-dependent advantage. The particular mutational events and scenarios are not too important. On one hand selection may favour increasingly specialised types if tradeoffs exist such that increases in the ability to grow on one carbon source reduce the capacity to metabolise another [6]. On the other hand, and dependent upon various ecological factors [18], types may emerge with the capacity to efficiently utilise more than one carbon source. Such generalist genotypes might arise as a consequence of duplication and subsequent divergence of the gene encoding the structural enzyme [16,24,44].

Let us now consider one of the generalist lineages: in one of these a mutation of some significance occurs, involving a duplication of the sole regulatory gene. Such a mutation is likely to occur spontaneously at an appreciable rate relative to its loss [38,44] and therefore such a duplication will be present at least some of the time in some members of the population (Fig. 3b). While duplication of the regulatory system may carry some selective benefit (for example, by supplying more of a limiting positive regulator, or by allowing more efficient repression of some genes [44]), there is no reason at this stage to suppose that such mutants are more fit than any other in the radiating population. Nevertheless, this
divergence has, according to the developing scenario, important implications for subsequent evolution of modularity and diversification.

It just so happens that the environment contains an unexploited ecological niche in the form of an abundance of unutilised organic matter that cannot be metabolised without a major evolutionary innovation, out of the reach of a simple process of duplication and divergence of structural genes. However, just such a capacity has evolved independently in a separate lineage evolving elsewhere in the hypothetical ecosystem. Cells from this foreign lineage are suddenly transported into the environment upon which our attention is focussed, but cannot survive: they lyse and release their DNA. As a consequence all radiating lineages have the opportunity to acquire this new DNA and via it the capacity to metabolise the novel carbon source. We predict that lineages which have undergone duplications to regulatory genes are more likely to do so successfully.

Let us imagine that different organisms, some with, and some without, duplicated regulatory genes acquire the same fragment of DNA at the same time. The one that most readily accommodates the novel DNA stands to reap most benefit. The regulatory genes in the organisms differ in number but not in identity; therefore both types have an equal chance of being able to regulate the incoming DNA (Fig. 3c). However, the ultimate fate of the incoming DNA is predicted to be very different in the two organisms. In organisms with a duplicated regulatory gene (whether this duplication occurs before introduction of the foreign DNA, as in the above scenario, or shortly thereafter, is not important), divergence of one to increase its affinity for regulatory regions of the incoming DNA and for the appropriate environmental cues can allow expression of the newly acquired trait to be more efficiently matched to the cells needs (e.g., by reducing wasteful expression in the absence of the cognate metabolisable compound). The other regulatory gene can remain specific to the original target genes (Fig. 3d). In organisms with only one regulatory gene, mutations allowing more efficient regulation of the incoming DNA can only occur to the extent that they do not compromise proper regulation of the original target genes. The higher the number of genes under the control of the regulator the less likely this is to occur [8,10]. The effect of this constraint is a bias in the number of successful integrations of incoming DNA in favour of cells with ‘redundant’ regulatory genes. Clearly the potential exists for this process to repeat itself as further ecological opportunity, and the DNA to exploit this, is presented, recreating selection for regulatory duplication and iteratively increasing modularisation. Furthermore, it is not difficult to envisage similar processes generating increasing modularity and sophistication within regulatory components to the point where simple regulators become increasingly compartmentalised within modules (by duplication and divergence) as exists today in multi-domain two-component regulatory networks [37].

5. Predictions, limitations and tests of the model

Our model encompasses a simple key expectation, namely, that in the face of ecological opportunity and lateral gene transfer, selection will favour modular genome architectures better able to accommodate potentially beneficial foreign DNA. We suggest that one way in which this can be achieved is the duplication, specialisation and subsequent divergence, of signal transduction pathways. Below we discuss a number of specific predictions made by the model regarding the phylogenetic relationships between acquired traits and the genes that control them.

If laterally acquired traits are incorporated into the regulatory system through duplication and subsequent divergence of a pre-existing host regulatory system, then we expect comparisons between the genomes of various bacteria to reveal positive correlations between the following: (i) the number of signal transduction pathways and the number of laterally acquired traits; (ii) the number of signal transduction pathways and the number of ecological niches a species can occupy; (iii) the amount of time elapsed since a trait was acquired and the divergence time of its corresponding regulator; (iv) the time since divergence of a signal transduction pathway and the number of genes under its control. These predictions all stem from the temporal pattern of events outlined in Fig. 3.
Of course, considered individually, these predictions are not necessarily unique to our model. For example, a positive correlation between the number of signal transduction pathways and the number of ecological niches an organism can occupy might be the cause rather than the consequence of ecological opportunity. This caveat notwithstanding, in our view, consistent positive correlations between those factors listed above would provide strong support for the claim that the mechanism proposed by our model has contributed to bacterial diversity.

Additionally, we can envision a simple experimental test to determine whether redundancy in components of signal transduction pathways facilitates accommodation of novel DNA. This would involve constructing isogenic strains that differ in the number of copies of one (or several) signal transduction pathway(s). Potentially beneficial DNA (for example, conferring the ability to utilise a novel carbon source) would then be introduced into these strains, allowed to integrate randomly into the chromosome, and the recombinant populations allowed to compete in an environment that provides the appropriate unexploited niche. If our model is correct then the strain with the duplicate regulatory pathway would be expected to more effectively incorporate the foreign DNA and therefore be more competitive in the new niche. Subsequent analysis could uncover the exact genetic basis by which the introduced DNA became successfully integrated.

Although to our knowledge such an experiment has not been performed, several recent studies provide some insight into the efficacy of the mechanism proposed in our model. For example, in one study Ferenci and co-workers observed changes to regulatory systems usually involved in the uptake of maltose and galactose during evolution of *E. coli* under conditions of glucose limitation [26,27]. Neither maltose or galactose were present in the environment, therefore in a sense these regulatory systems can be considered as ‘duplicates’ in that they were free to vary in the absence of selection for their original function. Improvements in glucose metabolism were caused by changes in the maltose and galactose regulatory genes, demonstrating the potential for a ‘duplicated’ regulatory system to mutate to provide more efficient expression of a structural gene.

Other illuminating examples come from a variety of work on the fate of duplicated genes in eukaryotes [14,23,49]. The conclusion from such studies is that duplicate genes do offer a degree of robustness, suggesting that, for some time at least, they retain a degree of functional redundancy [14]. Nevertheless, it is clear that this redundancy is far from complete [13,14,49] and that duplicated genes often diverge from one another to gain new functions, for example, to allow more precise temporally or spatially specialized regulation (reviewed in [32]). In models explaining these results regulatory genes are usually assumed to be under some constraint limiting their ability to optimally regulate disparate target genes [23,32]. In a manner analogous to that proposed in our model, duplication and subsequent divergence removes this constraint and allows regulation to be optimised.

6. Concluding comments

Here we have developed a model describing a simple selection pressure for the evolution of modularity, with special reference to bacteria. Central to the model is the hypothesis that selection will favour modular connections between genotype and phenotype because such connections minimise the effects of antagonistic pleiotropy that arise from integration of selectively beneficial foreign DNA. We propose that this selection might underlie the relative abundance of signalling networks found in physiologically and ecologically diverse bacterial genera such as *Pseudomonas*, *Vibrio* and *Myxococcus*.

Our model differs from previous attempts to explain the evolution of modularity and evolvability in that it does not predict modularity to be a result of indirect selection on individuals, lineages or clades, to increase effective variability. We agree with previous authors that modularity may have special properties for evolvability (see [20]), but argue that such properties are likely to be a consequence (not a cause) of modularity. In this regard, our model seems unique in positing that modularity can be caused by direct selection on individuals. As signal transduction pathways are a universal feature of life, it is possible that our simple model has implications for understanding the evolution of modularity in higher organisms.

Acknowledgements

We thank John Roth and an anonymous reviewer for critical comment.

References


