UNIVERSITY OF DUBLIN

Gavin C. Conant

DEPARTMENT OF GENETICS
SMURFIT INSTITUTE
TRINITY COLLEGE
DUBLIN 2
IRELAND

September 28, 2005

Yves Brun, Ph. D.
Systems Biology/Microbiology Faculty Search,
Department of Biology,
Indiana University,
Jordan Hall 142, 1001 E 3rd St,
Bloomington IN 47405-7005

Professor Brun and members of the search committee,

Enclosed please find my application for your faculty position in Systems Biology/Microbiology. As you will see from my attached statement of research interests, the field of biological networks, mentioned in your advertisement in *Nature* magazine (ref. number 73583), is of particular interest to me and has led me to respond to your advertisement.

By way of introduction, I would describe myself as a computational evolutionary biologist. My undergraduate and graduate degrees are in biology with additional emphasis on statistics and computer science. As a graduate student, I was supported by the Department of Energy's Computational Sciences Graduate Fellowship, a program which allowed me to both pursue my own research interests as well as to study topics such as parallel programming and statistics. As a post-doctoral fellow first at the University of Leipzig and more recently at Trinity College here in Ireland, I have continued to explore the role of natural selection and genetic drift in shaping the genomes of extant organisms. Of particular interest to me is mapping the boundaries of biological changes that can be attributed to genetic drift, and I discuss several examples of this topic in my statement of research interests.

I have always found interactions with researchers of diverse interests to be both fascinating and useful for my own research, so the opportunity to work in a broadly-based department would be an exciting one. Future progress in biology

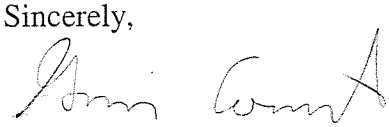
Tel: +353 1 6081288 **Fax: +353 1 6798558** **Email: conantg@tcd.ie**

will depend on applying innovative ideas to the massive existing frameworks of biological knowledge and data. It is only by constantly interacting with scientists from diverse fields that this sort of progress will be possible. Of course, the reputation of the Biology department and Biocomplexity Institute is not an unimportant attraction.

Enclosed please find a *curriculum vitae* as well as statements of research and teaching interests. In addition, I have requested four letters of recommendation to be sent to you from my undergraduate advisor, Prof. Paul Lewis, my doctoral advisor Prof. Andreas Wagner, and my two post-doctoral advisors, Profs. Peter Stadler and Kenneth Wolfe. Further information on publications and software tools I have coauthored is also available from my website (http://www/unm.edu/~gconant).

Finally, please accept my sincere thanks for your consideration of this application.

Sincerely,



Gavin Conant

Gavin Conant, Ph. D.
Department of Genetics
Trinity College
University of Dublin
Dublin 2,
Ireland

Statement of Teaching Interests:

While writing this application, I had the rather daunting thought that an undergraduate course is often a student's most significant exposure to fundamental biology. Thus, thinking back on best classes I had as a freshman and sophomore in college, I have tried to understand what made these classes outstanding. I suggest that four "themes" helped unify the material presented. I would further argue that these themes, to a greater or lesser extent, are applicable to almost all material in introductory biology. The first three, closely linked, themes are those of engineering, evolution and chemistry, while the fourth that of history. Introductory biology occasionally suffers from appearing to be only a mass of facts: providing context with these sorts of overarching perspectives can help students retain what they have learned beyond their freshman year.

The engineering principle is simply to think about parts of an organism as being made to do a job. Even something as simple as the base-pairing rules of DNA can be thought about in this way. This principle of course is intimately linked to evolution. I think one of the most interesting ways to point this out is to consider cases where organisms' design does not seem to match with how a human engineer would build a system. Sometimes the biological solution may be worse: the higher affinity of the Rubisco protein for molecular oxygen than for carbon dioxide appears to be an example of this. Such examples obviously lead to a discussion of the undirected nature of evolution. But there are also examples where I would argue that evolution has been much more clever than a naïve engineer would have been. The immune system is an excellent example.

The third related idea is that of chemistry. Because chemistry is challenging to many students, it is also a challenge to bring chemistry into biology in a way that enhances understanding. But it is important to remember than many of the factors that make chemistry difficult for students need not be introduced in biology classes. For a beginning biology student, a few basic chemical concepts, such as the ideas of solvents, activation energies and concentration gradients, are key.

The final perspective I found helpful was an historical one. One of the best ways to teach about how science works is to teach how it has worked in the past. Experiments by Avery and by Hershey and Chase, demonstrating that DNA is the hereditary material, teach both scientific facts and scientific methods simultaneously, which is efficient if nothing else.

One final general comment that I would add is that my experience as a computational biologist has led me to consider the role of computing in teaching. By this, I do *not* mean using Powerpoint to give classroom lectures, about which I have mixed feelings. Instead, I think that computers can help visually demonstrate ideas that may be difficult to relate in words. For instance, in one lecture, I used a simple animation of gene frequencies changing in a simulated population to help explain the

concept of genetic drift. I hope that this type of approach would especially help reach students who have difficulties with standard lecture formats.

I do not deny that my experience in the classroom is limited. This is only partly by choice. During most of my graduate studies I was supported, in lieu of a teaching assistantship, by a fellowship from the Department of Energy. I cannot regret the research opportunities this fellowship provided, but I do note that I arranged to teach one section of an Anatomy and Physiology laboratory, despite being somewhat discouraged from doing so by the fellowship terms. In addition, our department required the delivery of four classroom lectures as part of the Ph. D. program: I gave these lectures in both evolutionary genetics and genomics. These experiences have at least given me a sense of the amount of work required to prepare even a single lecture.

Finally, I think it is appropriate to comment on what variety of courses I might see myself teaching. At the introductory level, I would obviously expect to be asked to teach genetics, as well as classes in introductory biology, cellular biology or evolution. For more advanced students, a course in molecular evolution with a strong emphasis on computational methods and algorithms seems appropriate. My former advisor, Andreas Wagner, developed a class entitled Evolutionary Genetics, which gave students a fairly broad overview of topics such as phylogenetics and population genetics. Such a course would be appropriate for early graduate students or graduate students whose major interests are not in molecular evolution but who would like an acquaintance with the area. I would be pleased to undertake similar efforts, particularly in collaboration with other faculty members.

Gavin Conant, Ph. D.
Department of Genetics
Trinity College
University of Dublin
Dublin 2,
Ireland

Statement of Research Interests:

In the caricature of natural selection, organisms' phenotypes change slowly and uniformly over time as the population climbs a "hill" of increased fitness. While this picture is no worse than a simplified version of the truth, it is unhelpful and even a little misleading because it fails to make a connection between what we know about organismal complexity at the molecular level and the operation of natural selection. My interests as a biologist focus on trying to understand how genetic changes, many of which are well understood, lead to novel phenotypes, which are not. As overwhelming as this question seems, impressive new experimental techniques (such as gene expression microarrays and genome-scale transcription factor binding assays) have provided comprehensive data that allow us to measure both genetic and phenotypic differences, particularly among microbes. Not surprisingly, associating genetic changes with their phenotypic effects has proven difficult, and I continue to look for ways to study the appearance of novelty at differing levels of biological organization. Three levels of particular interest are gene sequences, the genomic gene content and genetic networks.

One very obvious question regarding evolutionary novelty, the origins of novel genes, has been studied for decades. The answer to that question, proposed by Haldane and Ohno [1,2], can be tritely stated as "new genes come from old genes." Research on the question of gene duplication has considered questions such has the frequency of duplication [3], the sizes of duplications [4] and the duplicates' evolutionary trajectory after duplication [5]. Together with Andreas Wagner, I developed a tool (GenomeHistory) which identifies all pairs of duplicate genes in a genome and calculates the pairwise divergence between them [6]. This tool allowed us to study questions such as what functional categories of genes were over or under-represented among duplicate genes. It also allowed us to demonstrate that large gene families in two species of yeasts were under less selective constraint than smaller families. Prof. Wagner and I also found that a considerable proportion of duplicate genes evolve asymmetrically [7] and that the presence of duplicate genes in the genome of the nematode *C. elegans* helped confer robustness to loss of gene function in that organism [8]. These last two results give a somewhat conflicting picture regarding the function of duplicate genes: asymmetric evolution suggests functional changes since duplication, while robustness argues for the maintenance of at least some functional overlap. How to reconcile these different roles for duplicate genes remains an open question.

More recently, Prof. Wagner and I have been studying another source of novel genes. We have estimated the rate at which recombination among genes creates novel chimeric genes over evolutionary timescales, a process we refer to as "gene-shuffling." Our estimates were admittedly rough, but our conclusion, that gene

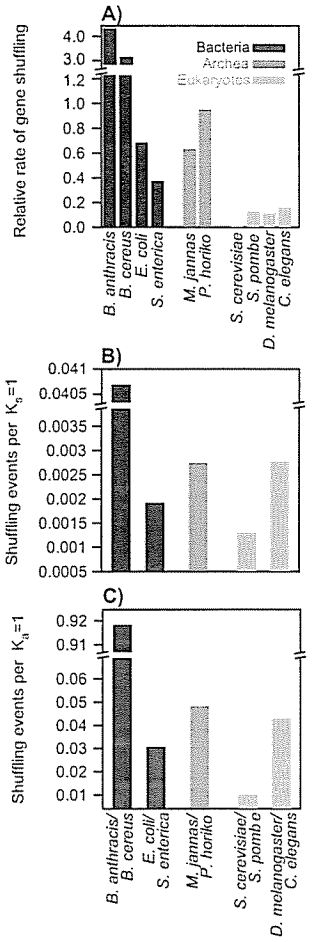


Figure 1: The incidence of gene shuffling across the three domains of life as compared to three other measures of evolutionary change: the rate of gene duplication, the rate of synonymous substitution and the rate of non-synonymous substitution.

shuffling is rather rare among conserved genes and especially rare in eukaryotes,[9] is a striking one (figure 1). As new genomic data becomes available, it will be interesting to test this conclusion by examining the shuffling rate in more rapidly evolving genes. Of particular interest are the yeasts, where we now have extraordinary phylogenetic depth of genomic sequences, which should allow us to study gene shuffling over multiple timescales.

The question of novelty is of course intimately linked to that of the action of natural selection. For gene duplication, we have a good sense of under what circumstances we would expect selection to allow a duplicate gene to survive. Generally, survival will require either the acquisition of a novel function in one duplicate, the subdivision of ancestral functions between the duplicates (subfunctionalization) [10], or the requirement for high dosages of the gene. Understanding all three fates is vital to understanding novelty because both subfunctionalization and changes in gene dosage can have knock-on effects elsewhere in the cell.

These general features of natural selection are relevant to other sources of genetic novelty as well. At the sequence level, properly detecting and measuring selection is vital for identifying sequences or regions of sequences where novel features have arisen. With the help of Günter Wagner and Peter Stadler, I have been developing models of protein-coding gene evolution which attempt to delimit the boundaries of neutral changes in protein sequence. While it might seem paradoxical to look for selection by modeling its opposite, it is only when we have a clear understanding of what neutral changes are permitted that we can identify selective changes. Our approach focuses on the differences between the amino acid residues, using physical and chemical properties to model substitution frequencies. We are currently studying how protein three-dimensional structure affects substitution rates, using a sample of approximately 50 yeast genes for which crystal structures are known. In future, we plan to study how substitution rates depend on protein secondary structure as well.

Finally, I am interested the origins of novelty in higher level structures such as genetic networks. By studying the distribution of small network motifs in the transcriptional regulatory networks of yeast and *E. coli*, Prof. A. Wagner and I have previously demonstrated that these abundant motifs evolved through convergent evolution rather than duplication [11]. This observation has interesting implications because new transcription factor binding sites may be relatively easy to create and yet can dramatically restructure the resulting regulatory network. Kenneth Wolfe and I

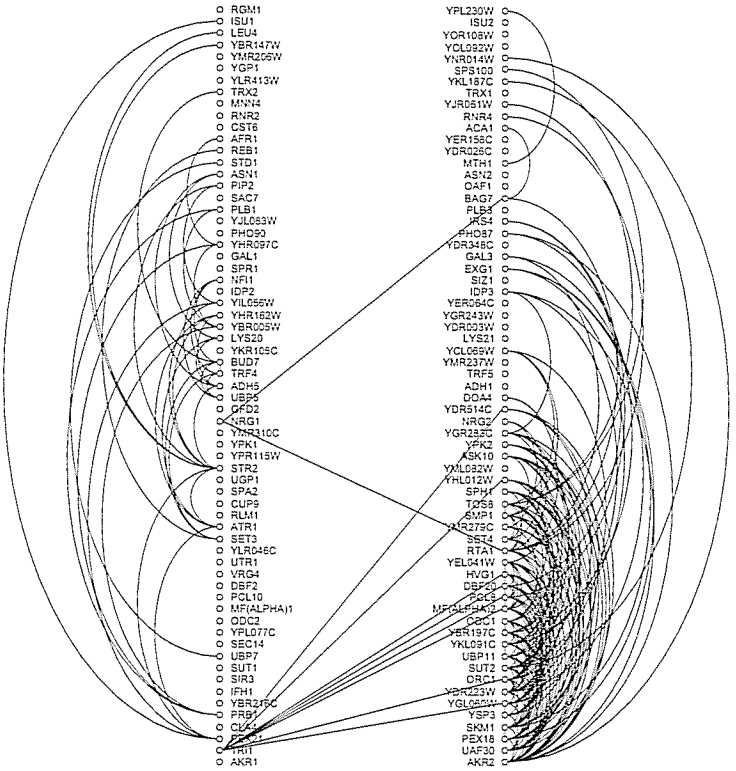


Figure 2: A network of coexpressed duplicate genes from *S. cerevisiae.* Duplicate gene pairs which originated at whole-genome duplicate are arranged opposite each other. Interactions are indicated for genes which share a Pearson's correlation of 0.75 or greater. This network has been arranged using an algorithm developed by Kenneth Wolfe and myself such that the minimum possible number of interactions cross between the two partitions. Analysis of simulated networks demonstrates that this network shows many fewer "crossing" interactions (more subfunctionalization) than we would expect by chance.

have followed up this work by studying how several cellular networks in yeast have evolved after a whole genome duplication which occurred roughly 100Mya [12,13]. In doing so, we have found evidence for "network subfunctionalization:" the division of sections of the ancestral gene expression network into discrete parts, each containing a single copy of a duplicate gene pair surviving from whole genome duplication (figure 2). This discovery is interesting because it potentially involves the creation of a complex novel structure (the subfunctionalized network) strictly through the actions of genetic drift and purifying selection.

In the future, I hope to study to what extent the potential for novelty is latent in organismal complexity. For instance, because genetic networks are evolutionary labile, relatively simple genetic alterations here may be able to produce large phenotypic changes. In addition, the fact that such networks can evolve under several different selective regimes implies that biological complexity may not be as difficult to achieve as has been thought. I am curious as to whether it is possible that a large set of potential networks might have similar or identical fitness in one environment and very different fitness in other environments. Under these circumstances, it would be possible to argue that some networks are more evolvable than others—whether selection could operate to preserve such networks would then be a very interesting question. Latent effects of complexity could also be found through an improved understanding of fields such as the dynamics of metabolism, where, for instance, subtle changes the expression of a single enzyme could have surprising effects on phenotype. As we better understand these questions of evolvablity, we will not only better understand how organisms have changed in the past but also how they function in the present.

References:

1. Haldane, J.B.S. The part played by recurrent mutation in evolution. *American Naturalist* **67**, 5-9 (1933).
2. Ohno, S. *Evolution by gene duplication*, 160pp. (Springer, New York, 1970).
3. Lynch, M. & Conery, J.S. The evolutionary fate and consequences of duplicate genes. *Science* **290**, 1151-1155 (2000).
4. Katju, V. & Lynch, M. The structure and early evolution of recently arisen gene duplicates in the *Caenorhabditis elegans* genome. *Genetics* **165**, 1793-1803 (2003).
5. Kondrashov, F.A., Rogozin, I.B., Wolf, Y.I. & Koonin, E.V. Selection in the evolution of gene duplicates. *Genome Biology* **3**, 0008.1-0008.9 (2002).
6. Conant, G.C. & Wagner, A. GenomeHistory: A software tool and its application to fully sequenced genomes. *Nucleic Acids Research* **30**, 3378-3386 (2002).
7. Conant, G.C. & Wagner, A. Asymmetric sequence divergence of duplicate genes. *Genome Research* **13**, 2052-2058 (2003).
8. Conant, G.C. & Wagner, A. Duplicate genes and robustness to transient gene knockouts in *Caenorhabditis elegans*. *Proceedings of the Royal Society, Biological Sciences* **271**, 89-96 (2004).
9. Conant, G.C. & Wagner, A. The rarity of gene shuffling in conserved genes. *Genome Biology* **6**, R60 (2005).
10. Force, A. et al. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* **151**, 1531-1545 (1999).
11. Conant, G.C. & Wagner, A. Convergent evolution of gene circuits. *Nature Genetics* **34**, 264-266 (2003).
12. Wolfe, K.H. & Shields, D.C. Molecular evidence for an ancient duplication of the entire yeast genome. *Nature* **387**, 708-713 (1997).
13. Kellis, M., Birren, B.W. & Lander, E.S. Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. *Nature* **428**, 617-624 (2004).