Genome Biology and Evolution Advance Access published July 21, 2010 doi:10.1093/gbe/evq045

The effect of chaperonin buffering on protein evolution

Tom A. Williams[1] and Mario A. Fares[1,2]*

[1]Department of Genetics, University of Dublin, Trinity College, Dublin, Ireland

[2]Integrative Systems Biology Group, Institute of Molecular and Cellular Biology (CSIC-Universidad Politécnica de Valencia (UPV)), Valencia, Spain

*Corresponding author: faresm@tcd.ie

Address: Department of Genetics, University of Dublin, Trinity College, Dublin, Ireland

Phone number: +353 1 8963521

Abstract

Molecular chaperones are highly conserved and ubiquitous proteins that help other proteins in the cell to fold. Pioneering work by Rutherford and Lindquist suggested that the chaperone Hsp90 could buffer (that is, suppress) phenotypic variation in its client proteins, and that alternate periods of buffering and expression of these variants might be important in adaptive evolution. More recently, Tokuriki and Tawfik presented an explicit mechanism for chaperone-dependent evolution, in which the *E. coli* chaperonin GroEL facilitated the folding of clients that had accumulated structurally-destabilizing but neofunctionalizing mutations in the protein core. But how important an evolutionary force is chaperonin-mediated buffering in nature? Here, we address this question by modeling the per-residue evolutionary rate of the crystallized *E. coli* proteome, evaluating the relative contributions of chaperonin buffering, functional importance, and structural features such as residue contact density. Previous findings suggest an interaction between codon bias and GroEL in limiting the effects of misfolding errors. Our results suggest that the buffering of deleterious mutations by GroEL increases the evolutionary rate of client proteins. We then examine the evolutionary fate of GroEL clients in the *Mycoplasmas*, a group of bacteria containing the only known organisms that lack chaperonins. We show that GroEL was lost once in the common ancestor of a monophyletic subgroup of *Mycoplasmas*, and we evaluate the effect of this loss on the subsequent evolution of client proteins, providing evidence that client homologs in 11 *Mycoplasma* species have lost their obligate dependency on GroEL for folding. Our analyses indicate that individual molecules such as chaperonins can have significant effects on proteome evolution through their modulation of protein folding.

Introduction

Although many newly-synthesized proteins fold spontaneously into the correct, functional three-dimensional shape (Anfinsen 1973), some require the assistance of accessory proteins called molecular chaperones. Chaperones interact non-covalently with their client proteins, preventing the aggregation of unfolded polypeptides and promoting proper folding through a variety of mechanisms (Hartl, Hayer-Hartl 2009).

Through their modulation of the relationship between a protein's primary sequence and final structure – that is, between genotype and phenotype – chaperones have been proposed to facilitate the adaptive evolution of their client proteins (Rutherford, Lindquist 1998; Fares et al. 2002; Queitsch et al. 2002; Tokuriki, Tawfik 2009b; Tokuriki, Tawfik 2009a; Lindquist 2010). The pioneering work in this area was that of Rutherford and Lindquist (1998), who demonstrated that the chaperone Hsp90 suppresses (or buffers) the phenotypic effect of deleterious mutations in its clients, which are mainly signalling proteins. They found that the reduction of Hsp90 activity resulted in the expression of underlying developmental abnormalities in *Drosophila*. When subject to selection, these variants could be enriched in the population to the point where, combined in a single genome, they could no longer be suppressed by restored Hsp90 function. The fixation of a set of mutations in this way might cause an "adaptive leap" from one developmental pathway to another, explaining the phenomenon of "genetic assimilation" that had previously been observed by Waddington (1953). Since this initial discovery, Hsp90-buffered variation has been

documented in other eukaryotes including *Saccharomyces cerevisiae* (Cowen, Lindquist 2005) and *Arabidopsis thaliana* (Sangster et al. 2007; Sangster et al. 2008).

Work on the chaperonin GroEL/GroES of *Escherichia coli*, an unrelated molecular chaperone, has provided evidence for another mechanism by which chaperone buffering affects client protein evolution. Moran (1996) suggested that overexpression of GroEL/GroES in endosymbiotic bacteria was an evolutionary response to the high levels of genetic drift – and therefore high mutational load – experienced by these intracellular organisms, the idea being that higher levels of GroEL would enable the cell to continue functioning as deleterious mutations accumulated in the proteome. This hypothesis was supported by Fares et al. (2002), who showed that overexpression of GroEL recovered the fitness of *E. coli* strains exposed to strong genetic drift, while a recent bioinformatic analysis suggested that GroEL clients experience weaker selection for translationally-optimal codon usage in comparison to nonclients, perhaps due to a reduced need to prevent mistranslation (Warnecke, Hurst 2010). Far from being contradictory mechanisms, the authors suggested that GroEL buffering and codon usage may represent two complimentary ways by which organisms can limit protein misfolding errors (Warnecke, Hurst 2010).

The first concrete evidence that chaperonin buffering might act as more than a coping mechanism was provided by Tokuriki and Tawfik (2009a), who performed experimental evolution on four enzymes in *E. coli* with and without GroEL/GroES overexpression. Their results showed that GroEL/GroES could maintain the function of enzymes that had accumulated highly destabilizing mutations in their core. Even

more interesting was their attempt to enhance the inefficient esterase activity of one of the enzymes, *Pseudomonas* phosphotriesterase, by artificial selection in the presence and absence of GroEL/GroES. The esterase activity that evolved in the presence of GroEL was far more efficient than that which could be obtained without GroEL, because it depended upon a destabilizing mutation that reduced the rate of folding and greatly reduced enzyme activity in the absence of chaperonin buffering. Along with some existing evidence that functionally important mutations are often destabilizing (Wang et al. 2002; Tokuriki et al. 2008), this result provides a straightforward explanation for how chaperone buffering of deleterious mutations could be involved in the evolution of new functions in client proteins.

Despite this experimental evidence, the extent to which chaperones facilitate the evolution of their client proteins in nature remains unclear. In particular, chaperones may not only buffer deleterious variants, but also expose them to proteolysis (Kandror et al. 1994; Tomala, Korona 2008). Tokuriki and Tawfik (2009) performed their experimental evolution combining GroEL/GroES overexpression with strong purifying selection during each round of evolution: if chaperones really do buffer phenotypic variation in their clients, then the strength of selection acting on clients should be weaker than that acting on nonclients. Here, we evaluate the effect of chaperonin buffering on client protein evolutionary rate, using data from 85 gamma-proteobacterial genomes. This question can be approached bioinformatically due to two recent, systematic classifications of the *E. coli* proteome into client and nonclient portions (Kerner et al. 2005; Fujiwara et al. 2010). Kerner et al. (2005) identified 252 proteins that were repeatedly isolated from GroEL/GroES complexes, of which 85 were found so frequently as to suggest all copies of that protein required assistance

from the chaperonin complex in order to fold (obligate clients). Fujiwara et al. (2010) examined the solubility of these clients in GroEL/GroES-depleted cells, and found that 49/85 of the obligate clients of Kerner et al., along with another 8 proteins, were absolutely dependent on the chaperonin complex for folding. After controlling for several factors known to influence evolutionary rate, we compare the evolution of clients and nonclients under all these classifications.

We then examine the evolutionary fate of GroEL client proteins in the *Mycoplasmas*, a group of highly-derived bacteria with small genomes that contains the only organisms lacking GroEL/GroES yet described (Woese 1987; Lund 2009). We examine whether the loss of GroEL has lead to a loss of obligate client proteins, or whether *Mycoplasma* client homologs have adapted to life without GroEL, as has been reported for *Ureaplasma* (Fujiwara et al. 2010).

Materials and methods

Gamma-proteobacterial structures and alignments

All available crystallized protein structures for the gamma-proteobacteria (mostly from *Escherichia coli*) were downloaded from the Protein Data Bank (PDB, http://www.rcsb.org/pdb). The resulting dataset contained 1000 PDB entries (and 1075 protein chains – see Supplementary Material), representing 20-25% of the *E. coli* proteome and half (126/252) of known GroEL clients (Kerner et al. 2005), although it was not over-enriched for any of the functional categories in the Clusters of Orthologous Groups ontology system (Tatusov et al. 2003). Protein sequences

homologous to the structure-associated sequences were retrieved by reciprocal BLAST searching of 85 complete gamma-proteobacterial proteomes (see Supplementary Material), only considering reciprocal hits with E-values < 10^{-4} where the length of the whole protein was within the range of +/- 25% of the structure sequence. We limited the set of sequences to this rang of lengths in order to ensure that only proteins with the same structure and function would be included. Sets of homologs were aligned with ClustalW using the default parameters (Thompson et al. 1994), and the quality of the alignments was inspected mannually. Only those alignment columns that could be aligned to the structure sequence were used in our subsequent analyses.

Analysis of protein evolutionary rate

Classification of the *E. coli* proteome into clients and nonclients was carried out on the basis of the system of Kerner et al. (2005), who performed a proteome-wide screen for GroEL clients by trapping and then characterizing proteins encapsulated within GroEL/GroES complexes. GroEL interactors were further subdivided into facultative (class I and II) or obligate (class III) clients depending on the proportion associated with GroEL/GroES complexes versus the total amount of that protein in the cell. Recently, another study screened for obligate GroEL clients by identifying proteins that aggregate or are degraded in GroEL/GroES-depleted cells (Fujiwara et al. 2010). Their results overlap with, but do not exactly match, those of Kerner et al. (2005), because about 40% of Class III clients remain soluble during GroEL/GroES underexpression. In our analysis, we use both classifications when assessing the effect of chaperonin buffering. It is possible that these screens failed to identify all GroEL

clients in the E. coli proteome; however, we do not think that a (presumably small) proportion of unclassified clients among our set of nonclient proteins will have a serious effect on the analyses reported below – if anything, they ought to make the results more conservative.

Data on gene essentiality were downloaded from the SHIGEN Profiling of *E. coli* Chromosome database (Hashimoto et al. 2005; Kato, Hashimoto 2007). A gene is defined as essential if strains carrying a null mutation cannot grow under any conditions. Protein-protein interactions were quantified using the combined interaction dataset from Bacteriome.org, which contains 7613 experimentally-determined interactions between 2283 *E. coli* proteins (Peregrín-Alvarez et al. 2009). In order to avoid trivially biasing our results towards a greater number of client interactions, we removed all interactions involving GroEL/GroES from the dataset. We used gene expression data from the genome-wide study of Covert et. al (2004), using the dChip-normalized mean mRNA expression value across three replicates for wild-type *E. coli* cells growing in aerobic conditions. In the analyses reported below, we only used expression data when all three replicates were called as present on the array (resulting in data for 226/252 clients and 2889/3892 nonclients in the *E. coli* genome). Repeating the analyses using all expression data (regardless of quality) gave results which were qualitatively the same.

Per-residue estimates of evolutionary rate were calculated as follows: for each column in a protein sequence alignment, we counted the number of pairwise differences between residues x and the total number of comparisons n. To account for multiple

substitutions, we applied the Poisson correction to the proportion of differences $\frac{x}{n}$ to obtain a distance *d* for that column:

$$d = -\frac{19}{20}\log(1 - \frac{20}{19}.\frac{x}{n})$$

Per-residue amino acid contact density was defined as the number of other residues within 4 Angstroms of the site of interest (Toft, Fares 2010). For each atom in an amino acid, we calculated the Euclidean distance between it and all atoms in the other amino acids in the crystal structure. The distance between two amino acids was taken to be the minimum of the atomic distances between the two residues:

$$\min(\sqrt{(x_{1i} - x_{2j})^2 + (y_{1i} - y_{2j})^2 + (z_{1i} - z_{2j})^2})$$

where *i* and *j* represent all atoms in amino acids 1 and 2, respectively.

To evaluate the effect of chaperonin buffering on evolutionary rate after accounting for essentiality, amino acid contact density, gene expression level, and protein-protein interactions, we performed an analysis of covariance (ANCOVA) using the statistical software R (R Development Core Team, 2010). The ANCOVA was fit using the *lm* function. We used this approximation because this function represents a conservative relationship between the different factors and because modeling the relationships between more than two factors is both computationally expensive and combinatorially prohibitive. We are, however, aware of the fact that this linear modeling might represent a simplistic view of the interaction between factor effects, although a

systematic bias towards the covariance of two particular factors due to the model is unlikely. We compared the fit of models including (i) all main effects and interactions and (ii) just main effects with an ANOVA. The model without interactions fit the data significantly worse ($p < 10^{-15}$), prompting the retention of the more complex model. To circumvent the problem of model over-fitting, we assessed the significance of individual terms in the ANCOVA using the *step* function implemented in R, which uses Akaike's information criterion (AIC) to remove terms that do not significantly improve model fit – that is, models that increase the AIC value - resulting in the set of minimal adequate models discussed below.

In order to evaluate whether our results were due to bias introduced by phylogenetic non-independence of the 85 gamma-proteobacterial genomes used, we re-calculated Poisson distances using a reduced subset of our data comprising one species per genus and re-analyzed the data as described above. The representative sequence from each genus was chosen at random because none of the within-genus sequences presented distinctive characteristrics regarding genome size, codon composition, etc. The results were qualitatively very similar (see Supplementary Tables 8-12), suggesting that the effects discussed below are not an artifact of biased phylogenetic coverage. The numbers reported below are from the original analysis, which uses all of the available data. We would also like to stress that biases in our results due to the phylogenetic non-independence of sequences should affect clients and nonclients equally, and should not, therefore, bias tendencies systematically one way or the other. In other words, sequences are phylogenetically dependent but proteins are not.

Mycoplasma sequences and analysis

Four *Mycoplasma* genomes that contain a GroEL homolog (*Mycoplasma penetrans HF-2, Mycoplasma genitalium G37, Mycoplasma gallisepticum R*, and *Mycoplasma pneumoniae M129*) and seven that do not (*Mycoplasma pulmonis UAB CTIP, Mycoplasma capricolum subsp. capricolum ATCC 27343, Mycoplasma mobile 163K, Mycoplasma arthritidis 158L3-1, Mycoplasma mycoides subsp. mycoides SC str. PG1, Mycoplasma hyopneumoniae 232,* and *Mycoplasma synoviae*) were downloaded from NCBI (accession numbers provided in Supplementary Table 3), with the presence or absence of GroEL being assessed manually using NCBI Web-BLAST, using the *M. penetrans* HF-2 protein sequence (NP_757486.1) as the initial query. The *E. coli* proteome was divided into clients and nonclients as described above, and each set of genes was BLASTed against these 11 genomes. Only 29 of the 252 *E. coli* GroEL clients had significant hits against all 11 *Mycoplasma* genomes (defined as an E-value < 10^{-7}, which we found by manual experimentation to be a good trade-off between false positive and false negative presence/absence calls. In order to increase the size of our dataset, we also included genes which were present in at least 3/4 *Mycoplasma* genomes with GroEL and 6/7 genomes without. This resulted in a set of 57 *Mycoplasma* homologs of *E. coli* GroEL clients and 282 homologs of non-clients, with 9-11 *Mycoplasma* sequences per gene.

To evaluate whether GroEL client proteins in *E. coli* have been preferentially lost from *Mycoplasmas* that lack GroEL, we used an analysis of covariance (fitted with the *glm* function in R, with binomial errors) in which a binary response variable

reports the presence or absence of a homolog to each protein in the *E. coli* proteome in a given *Mycoplasma* species (where presence is defined as a BLASTP hit at $E < 10^{-7}$) and with client/nonclient status, essentiality, number of protein-protein interactions and mRNA expression level as the explanatory variables.

To investigate the evolution of GroEL client proteins within the *Mycoplasmas*, we built protein sequence alignments from the 57 genes homologous to *E. coli* GroEL clients using MUSCLE 3.7 (Edgar 2004) under the default parameters. These alignments were used to build 100-bootstrap maximum likelihood phylogenetic trees with RaxML 7.04 (Stamatakis 2006), using a substitution model chosen by ProtTest (Abascal et al. 2005) in each case. 49/57 consensus trees suggested the topology shown in Figure 2, in which a single loss of GroEL occured within the *Mycoplasmas*. This consensus topology was then used for comparison of selective constraint between *Mycoplasmas* with and without GroEL.

For each client and nonclient alignment, we calculated the nonsynonymous-to-synonymous substitution ratio (dN/dS) under maximum likelihood using the program codeml, from the PAML package version 4.0 (Yang 2007). In each case, we compared two models: one in which a single dN/dS ratio applies across the tree, and one in which the genomes with- and without GroEL evolve under different ratios. These models were compared with a likelihood ratio test for which the null distribution is a chi-squared distribution with one degree of freedom. The numbers of client and nonclient homologs which were evolving significantly faster in GroEL-lacking *Mycoplasma* were then compared with a chi-squared test.

Mycoplasma genomes lacking GroEL were not impoverished for GroEL clients when compared to *Mycoplasmas* with GroEL, raising the possibility that intrinsic changes in these proteins occurred in non-GroEL *Mycoplasmas* that made them independent from GroEL. To evaluate this possibility, we tested whether the amino acid compositions or molecular weights of the proteins from *Mycoplasmas* with GroEL differed significantly from those in *Mycoplasmas* without GroEL. The molecular weights of client homologs in *Mycoplasmas* with- and without-GroEL were calculated by summing the weights of their constituent amino acids and, for each protein, calculating a mean protein molecular weight for *Mycoplasmas* with- and those without-GroEL. Weights were compared with Wilcoxon two-sample paired signed rank test. Amino acid compositions were compared in a similar way, with mean proportions for each amino acid in each protein in Mycoplasmas with- and without-GroEL being compared with Wilcoxon two-sampled paired rank tests, using the Bonferroni correction to account for multiple testing.

Results and Discussion

The functional importance of GroEL client proteins

As outlined in the introduction, the idea that molecular chaperones buffer the phenotypic effects of mutations in their clients is critical to the hypothesis that chaperones facilitate adaptive evolution (Rutherford, Lindquist 1998; Tokuriki, Tawfik 2009a). Assuming that most mutations affecting phenotype are – at least individually - neutral or deleterious (Kimura 1983), if selection against such

mutations is weaker in GroEL clients than nonclients due to a buffering effect (Tokuriki, Tawfik 2009a), then clients ought to evolve faster than nonclients. However, precisely the opposite trend has been reported (Hirtreiter et al. 2009), with GroEL preferentially chaperoning slow-evolving proteins. The same trend was apparent in our dataset of 1,075 gamma-proteobacterial proteins, with clients evolving significantly more slowly than nonclients (mean Poisson distance in clients = 0.147, nonclients = 0.178, $p < 10^{-15}$, Mann-Whitney U test). Does this result falsify the chaperone buffering hypothesis? No, because it does not take into account the many factors that influence evolutionary rate. For instance, if clients are enriched for characteristics that constrain evolution, these might mask a buffering effect. We compared the functional importance of clients and nonclients in terms of essentiality, number of protein-protein interactions, and mRNA expression levels. All these factors have previously been observed to influence evolutionary rate (Krylov et al. 2003; Drummond, Wilke 2008; Wolf et al. 2010), although their relative importance is a matter of some debate (Bloom, Adami 2003; Jordan et al. 2003; Pal et al. 2003). We found striking differences between clients and nonclients in terms of essentiality and protein-protein interactions, with clients significantly more likely to prove essential upon single-gene knockout (43/248 essential clients, 242/3900 essential nonclients, $P < 10^{-3}$, chi-squared test), and participating in significantly more protein-protein interactions than nonclients (mean 13.7 for clients, 5.9 for nonclients, $P = 1.6 \times 10^{-14}$, Mann-Whitney U test), even after interactions with GroEL/GroES are removed from the dataset. At the level of mRNA expression, clients are expressed at a significantly higher level than nonclients in wild-type *E. coli* cells growing aerobically (mean client probe intensity 2186, nonclient 1290, $P < 10^{-15}$, Mann-Whitney U test),

although obligate clients were expressed at a lower level than facultative clients (1683 vs. 2425, P = 0.0008328, Mann-Whitney U test).

Taken together, these results suggest that client proteins are, on average, of greater functional importance than nonclients. Since a higher proportion of essential genes, a higher number of protein-protein interactions, and higher mRNA expression levels are all either weakly or strongly associated with a decrease in evolutionary rate (Krylov et al. 2003; Drummond, Wilke 2008; Wolf et al. 2010), their influence must be accounted for when evaluating the effect of chaperonin buffering on client protein evolution.

GroEL buffers the evolution of its obligate clients

To evaluate the relative contributions of chaperone buffering, essentiality, network connectivity (in terms of protein-protein interactions) and expression level on evolutionary rate, we performed an analysis of covariance with one response variable, per-residue Poisson distance, and five explanatory variables: two categorial (client/nonclient, essential/nonessential), and three continuous: number of protein-protein interactions, mRNA expression level (in mean probe intensity across three replicates), and amino acid contact density. This final covariate, which quantifies the number of other residues within a 4 Angstrom radius of a particular amino acid site, has previously been shown to correlate negatively with evolutionary rate: that is, amino acids surrounded by large numbers of other residues (such as in the protein core) evolve relatively slowly (Thorne et al. 1996; Goldman et al. 1998; Bustamante

et al. 2000; Mintseris, Weng 2005; Bloom et al. 2006; Conant, Stadler 2009; Toft, Fares 2010).

We performed four different analyses, in which GroEL clients were defined in four different ways: (i) all 252 GroEL/GroES interactors identified by Kerner et al. (2005) – that is, both facultative and obligate clients; (ii) 85 obligate clients only (as defined by Kerner et al. (2005)); (iii) 57 obligate clients as defined by Fujuwara et al. (2010); and (iv) the 34 obligate clients classified by Kerner et al. (2005) that do not depend on GroEL/GroES for solubility. An important difference exists between categories (ii) and (iii). Kerner et al. (2005) classified clients according to their enrichment in GroEL/GroES complexes. If more than 4% of the total cellular content of a particular protein was associated with GroEL/GroES, they inferred that all copies of that protein needed to interact with the chaperonin complex in order to reach their native conformation, making it an obligate "Class III" client. Proteins which were reliably isolated from GroEL/GroES complexes at lower levels of enrichment were assigned to two classes of facultative clients. Fujiwara et al. (2010) took a more direct approach, measuring the solubility of Class III clients in GroEL/GroES-depleted cells. They found that 34/85 of the Class III clients did not depend on GroEL/GroES for solubility ("Class III-" clients), suggesting that the enrichment of a protein in GroEL/GroES complexes is correlated with, but does not exactly predict, obligate dependency. Combining the 49/85 (60%) of Class III clients that are dependent on GroEL/GroES for solubility with another 8 proteins not previously included in Class III, these authors proposed a new class of obligate GroEL/GroES clients (Class IV). The results of our analyses are summarized in Table 1, which shows the effect of

chaperonin buffering on the evolution of each of these four groups of clients (all clients, Class III, Class IV, and Class III-).

Regardless of the way in which GroEL clients and nonclients are defined, our analysis recovers the well-documented negative correlations between expression levels, numbers of protein-protein interactions, residue contact density, and evolutionary rate (see Table 1), with two exceptions. Firstly, when the Class III- proteins are compared to the rest of the proteome, the main effect of expression level changes sign, with higher expression levels associated with a moderate increase in evolutionary rate. Deletion of the client/nonclient term recovers the negative correlation between expression level and evolutionary rate observed with all other client/nonclient classifications, suggesting that this effect is due to the interaction between these two terms. Class III- proteins are highly enriched in GroEL/GroES complexes, but do not depend on the chaperonin for solubility. Fujiwara et al. (2010) noted that half of the Class III- proteins bind RNA or DNA, and overall the class is enriched for positively-charged amino acids. On the basis of this evidence, they proposed that these proteins are frequently recovered from GroEL/GroES complexes because they can bind the negatively-charged interior surface of the GroEL/GroES cavity, not because they required GroEL/GroES for folding – an hypothesis that is supported by the lack of any significant chaperonin buffering effect in this class (see below). The division of our dataset into this group of proteins on the one hand, and a mix of the "genuine" clients and nonclients on the other, may have produced the interaction giving rise to the change in sign of the expression level main effect.

Secondly, we find that in two of our four analyses essential genes are evolving faster than nonessential ones when these other factors are taken into account. To explore the reason for this unexpected result, we compared essential and nonessential genes in several ways. A simple comparison of mean evolutionary rate recovers a moderate but statistically-significant reduction in rate in essential genes, as has previously been reported (mean Poisson distance in essential genes = 0.160, nonessential = 0.174, $P < 10^{-15}$, Mann-Whitney U test); (Koonin 2005; Wolf et al. 2010). Essential genes participate in more protein-protein interactions (16.1 vs 5.4, $P < 10^{-15}$, Mann-Whitney U test) and have higher expression levels (2500 vs. 1239, $P < 10^{-15}$, Mann-Whitney U test) than those that are nonessential, which may go some way to explaining why they are essential in the first place. To identify the factor(s) underlying the effect of essentiality in our analyses of covariance, we re-analyzed the data while dropping each one of the other factors in turn. Failing to account for residue contact density or client/nonclient status resulted in no change in sign or significance of the essentiality term, but its significance was abolished when either of the terms modelling the number of protein-protein interactions or expression level were droppped. In comparison, dropping any one of protein-protein interactions, expression levels, or essentiality from the ANCOVA neither changed the sign nor abolished the significance of the client/nonclient term (see Supplementary Table 13).

The method used to classify GroEL clients had a striking effect on the analysis: considering both facultative and obligate clients together, there was a marginally significant effect of chaperonin buffering on evolutionary rate, with an increase in rate associated with clients - although, unlike the effects discussed below, the significance of this term was abolished when we re-analyzed a nonredundant subset of our data to

test for biases arising from phylogenetic non-independence (see Methods). We note, however, that this could also have resulted from the reduction in statistical power when decreasing the number of sequences in our analyses. When only Class III clients were considered, the significance and effect size of this rate shift was greatly increased ($P = 0.00301$), and became even more striking among Class IV clients which absolutely depend on GroEL/GroES for solubility ($P < 10^{-15}$). The analysis suggested that once other factors are accounted for, these obligate clients show an increase in mean per-residue Poisson distance of 0.4263 relative to the rest of the gamma-proteobacterial proteome. As discussed above, there is no significant effect of chaperonin buffering when only Class III- proteins are considered – that is, proteins enriched in GroEL/GroES complexes but that remain soluble in GroEL/GroES-depleted cells. These results lead to two conclusions: (i) at least among the gamma-proteobacteria, GroEL/GroES facilitates the accumulation of amino acid substitutions in its obligate clients, but not in all proteins with which it regularly interacts; and (ii) this buffering effect is most pronounced in client proteins that depend on the GroEL/GroES system for solubility (Class IV clients), as opposed to all proteins which are highly enriched in GroEL/GroES complexes. This relationship is masked in simple comparisons of client and nonclient evolutionary rate due to the increased functional importance of clients. To identify the factors that most directly interfere with the buffering effect, we deleted individual factors from our Class IV ANCOVA and evaluated the effect upon the remaining terms (see Figure 1). This approach suggested that gene essentiality and the number of protein-protein interactions were the most important confounding factors.

The increase in evolutionary rate that we observed among GroEL clients might be taken as evidence in favour of the "chaperonin-facilitated adaptation" model of Tokuriki and Tawfik (2009), but we note that this will only hold if mutations which confer new functions are disproportionately likely to interfere with protein folding – that is, to make folding intermediates more difficult to reach. If the effect of chaperonin-mediated buffering is simply to broaden the spectrum of neutral mutations in clients, then the ability of positive selection to promote the fixation of adaptive mutations will be weakened – that is, buffering will mainly act to increase the strength of genetic drift operating on clients. If, however, neofunctionalizing mutations tend to be destabilizing – a proposition for which there is some evidence (Wang et al. 2002; Tokuriki et al. 2008) – then buffering could maintain such variants in the population, making them accessible to positive selection if they confer an advantageous phenotype.

Neutral evolution of GroEL clients in *Mycoplasmas*?

Certain species of *Mycoplasma* and *Ureaplasma* are unique among sequenced genomes in lacking a chaperonin homolog of any kind (Lund 2009). Although these bacteria have experienced extensive genome reduction (Woese 1987), the loss of GroEL is surprising. GroEL is an essential gene in *E. coli* at least in part because several other essential proteins depend on it for proper folding (Lund 2009). Presumably, the loss of GroEL in *Mycoplasmas* must have been accompanied by either the loss of client homologs or the loss of their dependency on GroEL for folding. One possibility is that *Mycoplasmas* invest more in protein degradation in order to prevent aggregation (Wong, Houry 2004). However, there is now

experimental evidence (Fujiwara et al. 2010) that at least some homologs of *E. coli* GroEL clients have lost their obligate chaperonin dependency in these bacteria, folding independently when expressed in *E. coli*. In the present study, our aim was to assess the effect of GroEL loss on the evolution of chaperonin clients in those *Mycoplasmas* that have lost GroEL. First, we used BLASTP to identify homologs of *E. coli* clients in 11 complete *Mycoplasma* genomes and *Ureaplasma*, comprising 4 genomes which retain a copy of GroEL and 8 which have lost it. Perhaps surprisingly, there was no significant difference in the rentention of obligate (Class III/IV) clients and nonclients in 9/12 of these genomes, and in the 3 genomes where the difference was significant (*M. capricolum, M. mycoides,* and *M. synoviae*), it reflected preferential retention of client proteins, even though these species have all lost GroEL (see Table 2). How can the loss of GroEL have no effect, or even a positive effect, on the retention of obligate clients? A simple comparison of the numbers of retained clients and nonclients does not take into account other factors that might influence the loss of genes in *Mycoplasmas*. To account for these, we performed an analysis of covariance that indicated that the heightened functional importance of client proteins (discussed above) plays some role in their retention in *Mycoplasmas* (Table 2). In particular, proteins involved in higher numbers of interactions and essential proteins are significantly more likely to be retained in *Mycoplasma* genomes. Controlling for these covariates, client/nonclient status did not in itself have a significant effect on retention in any of the 12 genomes we analyzed, suggesting that the uncoupling of obligate client folding from GroEL reported in *Ureaplasma* may also apply in the related *Mycoplasmas*.

Did the loss of GroEL dependency in client proteins occur before or after the loss of GroEL in *Mycoplasmas*? Although the two events might be expected to be coupled, there is evidence that GroEL is not essential in *M. genitalium* and *M. pneumoniae* (Hutchison et al. 1999; Wong, Houry 2004), two of the four species which still possess the chaperonin. We addressed this question from an evolutionary perspective, asking whether the loss of GroEL had an effect on the nonsynonymous-to-synonymous substitution ratio (dN/dS) in GroEL-lacking *Mycoplasmas*. Our set of *Mycoplasma* homologs of *E. coli* GroEL clients contained 57 proteins, which we used to build 100-bootstrapped, maximum likelihood phylogenetic trees. Interestingly, 49/57 of these unrooted trees had a topology in which the *Mycoplasmas* with GroEL were separated from those without GroEL, the most parsimonious interpretation of such an arrangement being a single loss of GroEL within the *Mycoplasmas* (see Figure 2).

If former GroEL clients have accumulated mutations that enable them to fold independently, then this process might be detectable as an elevated dN/dS ratio among client proteins in the *Mycoplasmas* that lack GroEL. We tested this hypothesis using maximum likelihood estimates of dN/dS calculated using codeml (Yang 2007) on the consensus tree obtained from our client phylogenies. In order to increase the size of our dataset, we considered any *E. coli* homolog, client or nonclient, if it was present in at least 3/4 of the *Mycoplasma* genomes with GroEL and 6/7 without. 28/57 client homologs and 100/282 nonclient homologs experienced significantly relaxed selective constraints in the *Mycoplasma* without GroEL (that is, a two-dN/dS model, with a higher value on the branches without GroEL, was a significantly better fit to the data), but the difference in these proportions did not attain statistical significance

($P = 0.0522$, chi-square test). Although this P-value exceeds the standard alpha value, we suggest that the analysis provides weak support for the idea of increased dN/dS in the client proteins of *Mycoplasma* that have lost GroEL, which might represent an evolutionary signature of adaptation to a GroEL-independent folding pathway. A plausible alternative explanation, however, is simply that GroEL-lacking *Mycoplasmas* experience a higher rate of genetic drift, which is supported by the remarkable observation that of all nonclient genes for which the two-ratio model fit better than the one-ratio, 100 showed a higher dN/dS on the GroEL-lacking branches, versus only 4 in the GroEL-possessing *Mycoplasmas.* Additional support for this neutral explanation comes from comparisons of amino acid composition and molecular weight between client homologs in *Mycoplasmas* with- or without-GroEL. Fujiwara et al. (2010) reported an enrichment of alanine and glycine residues in Class IV (obligate) clients versus the rest of the *E. coli* proteome, suggesting that this property might distinguish independently-folding from chaperonin-buffered proteins. Such a biasa, if also present in the client homologs of GroEL-posessing but not GroEL-lacking Mycoplasmas, would provide additional evidence for the acquisition of independent folding exclusively in GroEL-lacking *Mycoplasmas*. A comparison of amino acid frequencies in the client homologs of these two sets of genomes, however, revealed no such pattern. Although we did detect significant differences in the frequencies of certain amino acids (see Supplementary Table 15), there was no systematic bias in the biochemical properties of those amino acids enriched in one group or the other: for instance, valine was enriched in the clients of GroEL-possesing *Mycoplasmas* while isoleucine was enriched in those of GroEL-lacking species. We also compared the molecular weights of client homologs between the two sets of *Mycoplasma* genomes, with the idea that the acquisition of independent folding might

lead to increases in the mass of proteins no longer constrained by the volume of the GroEL protein-folding cavity. This test also allowed us to determine whether small changes in the frequencies of multiple amino acids in Mycoplasmas client homologs might have added up to a significant change in mass, with potential implications for the interaction of the proteins with GroEL. However, we failed to detect a significant difference ($P = 0.7771$, Wilcoxon two-sample paired signed rank test).

Taken together, these results suggest that the folding of *E. coli* client homologs has become uncoupled from GroEL in the *Mycoplasmas*, perhaps even in the species that have retained GroEL. Our conclusions are in agreement with those of Clark and Tillier (2010), who recently reported no differences in the folding properties of *Mycoplasma* client and nonclient homologs as predicted by the FoldIndex program (Prilusky et al. 2005). We also note that the results presented here do not exclude the possibility that GroEL clients in *E. coli* acquired chaperonin dependency after the divergence of the *E. coli* and *Mycoplasma* lineages: in this case, the equal retention of client and nonclient homologs in *Mycoplasma* genomes would not reflect the gain of independent folding in former clients, but rather the retention of the ancestral state.

Conclusions

Although the models of chaperone-facilitated adaptive change proposed by Rutherford and Lindquist (1998) and Tokuriki and Tawfik (2009) suggest that chaperone clients should evolve faster than nonclients, the opposite is observed in the case of the *E. coli* chaperonin clients and their homologs. Here we have shown that

this pattern is due to the increased functional importance of clients, and that once this is accounted for, client proteins are evolving faster than nonclients. As discussed above, our results support the hypothesis that chaperones facilitate adaptive evolution under the condition that functionally innovative mutations tend to interfere with protein folding. But why do clients tend to be more functionally important? We propose two hypotheses, based on the observation of increased evolutionary rates in clients. Firstly, proteins that are buffered by chaperones might be able to more easily fix functionally innovative mutations despite their structurally destabilizing effects. The acquisition of new functions by these proteins would then lead them to take on a more important role in the cell. Alternatively, proteins that are already performing important functions are highly constrained and therefore might have more need of chaperone-assisted folding following the fixation of functionally innovative mutations. However, if there is no connection between functional innovation and structural stability, then the effect of chaperonin buffering observed here would act to increase the strength of genetic drift acting on clients.

Acknowledgements

This work was supported by a grant from Science Foundation Ireland to MAF and a grant from the Irish Research Council for Science, Engineering and Technology to TAW. We are grateful to Christina Toft for providing the gamma-proteobacterial dataset for analysis, and to Brian E. Caffrey for providing useful feedback on the manuscript.

References

Abascal, F, Zardoya, R, Posada, D. 2005. ProtTest: selection of best-fit models of protein evolution. Bioinformatics 21:2104-2105.

Anfinsen, CB. 1973. Principles that govern the folding of protein chains. Science (New York, N.Y.) 181:223-230.

Bloom, JD, Adami, C. 2003. Apparent dependence of protein evolutionary rate on number of interactions is linked to biases in protein-protein interactions data sets. BMC Evol Biol 3:21.

Bloom, JD, Labthavikul, ST, Otey, CR, Arnold, FH. 2006. Protein stability promotes evolvability. Proc Natl Acad Sci U S A 103:5869-5874.

Bustamante, CD, Townsend, JP, Hartl, DL. 2000. Solvent accessibility and purifying selection within proteins of Escherichia coli and Salmonella enterica. Mol Biol Evol 17:301-308.

Clark, GW, Tillier, ER. 2010. Loss and gain of GroEL in the Mollicutes. Biochem Cell Biol 88:185-194.

Conant, GC, Stadler, PF. 2009. Solvent exposure imparts similar selective pressures across a range of yeast proteins. Mol Biol Evol 26:1155-1161.

Cowen, LE, Lindquist, S. 2005. Hsp90 potentiates the rapid evolution of new traits: drug resistance in diverse fungi. Science 309:2185-2189.

Drummond, DA, Wilke, CO. 2008. Mistranslation-Induced Protein Misfolding as a Dominant Constraint on Coding-Sequence Evolution. Cell 134:341-352.

Edgar, RC. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics.

Fares, MA, et al. 2002. Endosymbiotic bacteria: groEL buffers against deleterious mutations. Nature 417:398-398.

Fujiwara, K, et al. 2010. A systematic survey of in vivo obligate chaperonin-dependent substrates. EMBO J 29:1552-1564.

Goldman, N, Thorne, JL, Jones, DT. 1998. Assessing the impact of secondary structure and solvent accessibility on protein evolution. Genetics 149:445-458.

Hartl, FU, Hayer-Hartl, M. 2009. Converging concepts of protein folding in vitro and in vivo. Nat Struct Mol Biol 16:574-581.

Hashimoto, M, et al. 2005. Cell size and nucleoid organization of engineered Escherichia coli cells with a reduced genome. Molecular Microbiology 55:137-149.

Hirtreiter, AM, et al. 2009. Differential substrate specificity of group I and group II chaperonins in the archaeon Methanosarcina mazei. Molecular Microbiology 74:1152-1168.

Hutchison, CA, et al. 1999. Global transposon mutagenesis and a minimal Mycoplasma genome. Science 286:2165-2169.

Jordan, IK, Wolf, YI, Koonin, EV. 2003. No simple dependence between protein evolution rate and the number of protein-protein interactions: only the most prolific interactors tend to evolve slowly. BMC Evol Biol 3:1.

Kandror, O, Busconi, L, Sherman, M, Goldberg, AL. 1994. Rapid degradation of an abnormal protein in Escherichia coli involves the chaperones GroEL and GroES. J Biol Chem 269:23575-23582.

Kato, J, Hashimoto, M. 2007. Construction of consecutive deletions of the Escherichia coli chromosome. Molecular Systems Biology 3:132-132.

Kerner, MJ, et al. 2005. Proteome-wide analysis of chaperonin-dependent protein folding in Escherichia coli. Cell 122:209-220.

Kimura, M. 1983. The neutral theory of molecular evolution. Cambridge: Cambridge University Press.

Koonin, EV. 2005. Systemic determinants of gene evolution and function. Mol Syst Biol 1:2005 0021.

Krylov, DM, Wolf, YI, Rogozin, IB, Koonin, EV. 2003. Gene loss, protein sequence divergence, gene dispensability, expression level, and interactivity are correlated in eukaryotic evolution. Genome Res 13:2229-2235.

Lindquist, S. 2010. Protein Folding Sculpting Evolutionary Change. Cold Spring Harb Symp Quant Biol.

Lund, PA. 2009. Multiple chaperonins in bacteria - why so many? FEMS Microbiology Reviews 33:785-800.

Mintseris, J, Weng, Z. 2005. Structure, function, and evolution of transient and obligate protein-protein interactions. Proc Natl Acad Sci U S A 102:10930-10935.

Moran, NA. 1996. Accelerated evolution and Muller's rachet in endosymbiotic bacteria. Proceedings of the National Academy of Sciences of the United States of America 93:2873-2878.

Pal, C, Papp, B, Hurst, LD. 2003. Genomic function: Rate of evolution and gene dispensability. Nature 421:496-497; discussion 497-498.

Peregrín-Alvarez, JM, Xiong, X, Su, C, Parkinson, J. 2009. The Modular Organization of Protein Interactions in Escherichia coli. PLoS Comput Biol 5:e1000523-e1000523.

Prilusky, J, et al. 2005. FoldIndex: a simple tool to predict whether a given protein sequence is intrinsically unfolded. Bioinformatics 21:3435-3438.

Queitsch, C, Sangster, TA, Lindquist, S. 2002. Hsp90 as a capacitor of phenotypic variation. Nature 417:618-624.

R Development Core Team. 2010. R: A Language and Environment for Statistical Computing. http://www.R-project.org.

Rutherford, SL, Lindquist, S. 1998. Hsp90 as a capacitor for morphological evolution. Nature 396:336-342.

Sangster, TA, et al. 2007. Phenotypic diversity and altered environmental plasticity in Arabidopsis thaliana with reduced Hsp90 levels. PLoS ONE 2:e648-e648.

Sangster, TA, et al. 2008. HSP90-buffered genetic variation is common in Arabidopsis thaliana. Proceedings of the National Academy of Sciences of the United States of America 105:2969-2974.

Stamatakis, A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22:2688-2690.

Tatusov, RL, et al. 2003. The COG database: an updated version includes eukaryotes. BMC Bioinformatics 4:41-41.

Thompson, JD, Higgins, DG, Gibson, TJ. 1994. Clustal-W - Improving the Sensitivity of Progressive Multiple Sequence Alignment through Sequence Weighting, Position-Specific Gap Penalties and Weight Matrix Choice. Nucleic Acids Research 22:4673-4680.

Thorne, JL, Goldman, N, Jones, DT. 1996. Combining protein evolution and secondary structure. Mol Biol Evol 13:666-673.

Toft, C, Fares, MA. 2010. Structural Calibration of the Rates of Amino Acid Evolution in a Search for Darwin in Drifting Biological Systems. Mol Biol Evol.

Tokuriki, N, Stricher, F, Serrano, L, Tawfik, DS. 2008. How protein stability and new functions trade off. PLoS Comput Biol 4:e1000002.

Tokuriki, N, Tawfik, DS. 2009a. Chaperonin overexpression promotes genetic variation and enzyme evolution. Nature 459:668-673.

Tokuriki, N, Tawfik, DS. 2009b. Protein Dynamism and Evolvability. Science 324:203-207.

Tomala, K, Korona, R. 2008. Molecular chaperones and selection against mutations. Biology Direct 3:5-5.

Waddington, CH. 1953. Genetic Assimilation of an Acquired Character. Evolution 7:118-126.

Wang, X, Minasov, G, Shoichet, BK. 2002. Evolution of an antibiotic resistance enzyme constrained by stability and activity trade-offs. Journal of Molecular Biology 320:85-95.

Warnecke, T, Hurst, LD. 2010. GroEL dependency affects codon usage--support for a critical role of misfolding in gene evolution. Mol Syst Biol 6:340.

Woese, CR. 1987. Bacterial evolution. Microbiol Rev 51:221-271.

Wolf, YI, Gopich, IV, Lipman, DJ, Koonin, EV. 2010. Relative Contributions of Intrinsic Structural-Functional Constraints and Translation Rate to the Evolution of Protein-Coding Genes. Genome Biol Evol 2010:190-199.

Wong, P, Houry, WA. 2004. Chaperone networks in bacteria: analysis of protein homeostasis in minimal cells. Journal of Structural Biology 146:79-89.

Yang, Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Molecular Biology and Evolution 24:1586-1591.

Tables and figures

Figure 1: The relationships between chaperonin buffering, gene essentiality, protein-protein interactions, residue contact density, and expression levels. The effect of deleting each main term and its interactions on the remaining terms: the arrows point away from the term being deleted, with the width of the arrow proportional to the change in significance. Colors denote the direction of the change: blue indicates a decrease in the *P*-value, while orange indicates an increase. The raw data used to generate this figure is provided in Supplementary Table 13. The effect of chaperonin buffering becomes less significant when numbers of protein-protein interactions and gene essentiality are taken into account, suggesting that these are the most important confounding factors in simple comparisons of client and nonclient evolutionary rate.

Figure 2: Phylogeny of *Mycoplasma* genomes with and without GroEL. 49/57 *Mycoplasma* homologs of *E. coli* GroEL clients support a topology in which the *Mycoplasma* species that have retained GroEL (red) cluster to the exclusion of those that have lost it (black), suggesting a single loss of GroEL within this group of organisms. Each client-protein maximum-likelihood tree was built using RaxML, using 100 bootstraps.

Term	Slope			
	All clients	**Class III clients (Kerner et al. 2005)**	**Class IV clients (Fujiwara et al. 2010)**	**Class III- clients (Fujiwara et al. 2010)**
Nonclient	-2.561×10^{-2} (*)	-1.512×10^{-1} (**)	-4.623×10^{-1} (***)	-2.183×10^{-2}
Nonessential	-5.872×10^{-2} (**)	-6.043×10^{-2}	-5.11×10^{-1} (***)	6.44×10^{-2}
Residue contact density	-1.024×10^{-2} (***)	-2.075×10^{-2} (***)	-3.68×10^{-2} (**)	-1.449×10^{-2} (**)
Protein-protein interactions	-3.604×10^{-3} (***)	-1.178×10^{-2} (**)	-4.865×10^{-2} (**)	-5.445×10^{-3} (**)

Expression level	-1.782 x 10^{-5} (***)	-3.981 x 10^{-5} (**)	-1.317 x 10^{-4} (***)	3.180 x 10^{-5} (**)

Table 1: Main effects in the ANCOVAs evaluating influences on evolutionary rate. Chaperonin buffering (Nonclient), gene essentiality (Nonessential), residue contact density, number of protein-protein interactions, and mRNA expression level. Clients are classified in three ways: All clients (all 252 GroEL/GroES interactors identified by Kerner et al. (2005)); the 84 obligate clients identified by the same authors on the basis that >4% of the cellular content of the protein was interacting with GroEL/GroES at a given time; and the 57 proteins which become insoluble upon GroEL/GroES depletion in the experiments of Fujiwara et al. (2010). Significance levels: * = $P < 0.05$; ** = $P < 0.01$; *** = $P < 0.0001$. Full summaries of the analyses, including precise *P*-values, are provided as Supplementary Material.

Species	**Class IV Clients**	**Nonclients**	***P-value*** **(Chi-squared test)**	**GLM main terms**			
				Client/nonclient	**Essentiality**	**Protein-protein interactions**	**Expression**
M. genitalium	8/57	424/4087	*0.3691*	-2.655×10^{-1}	1.289 (***)	3.737×10^{-2} (***)	1.017×10^{-4} (*)
M. penetrans	11	600	*0.3288*	-2.867×10^{-1}	1.201 (***)	3.127×10^{-2} (***)	6.959×10^{-5}
M. gallisepticum	10	465	*0.1467*	-1.812×10^{-1}	1.352 (***)	3.828×10^{-2} (***)	9.936×10^{-5} (*)
M. pneumoniae	8	463	*0.5226*	-3.519×10^{-1}	1.191 (***)	3.728×10^{-2}	7.745×10^{-5}

						(***)	
M. pulmonis	10	515	*0.2652*	7.775×10^{-1}	3.723×10^{-1} (***)	4.373×10^{-2} (***)	-4.452×10^{-5}
M. capricolum	17	572	*0.0006768* (**)	5.784×10^{-1}	1.074 (***)	3.907×10^{-2} (***)	9.791×10^{-5} (*)
M. mycoides	17	581	*0.0008672* (**)	5.535×10^{-1}	9.998×10^{-1} (***)	3.622×10^{-2} (***)	1.026×10^{-4} (**)
M. mobile	11	483	*0.07522*	-3.995×10^{-2}	1.226 (***)	3.745×10^{-2} (***)	1.148×10^{-4} (*0.00434*)
M. arthriditis	8	410	*0.3189*	-7.284×10^{-1}	7.322×10^{-1} (*)	5.05×10^{-2} (***)	5.498×10^{-5}
M.	10	476	*0.1694*	1.187	4.190×10^{-1}	4.258×10^{-2}	-5.908×10^{-5}

hyopneumoniae						(***)	
M. synoviae	13	452	*0.00526* (**)	2.004	9.391×10^{-1} (**)	4.145×10^{-2} (***)	1.802×10^{-5}
Ureaplasma urealyticum	10	415	*0.0678*	1.581×10^{-1}	1.321 (***)	4.315×10^{-2} (***)	6.825×10^{-5}

Table 2: Loss of GroEL clients and nonclients from *Mycoplasma* genomes. The Chi-square *P*-value reported is for a test of association between retention in *Mycoplasma* genomes and client/nonclient status in *E. coli*. Proteins that are essential or involved in a high number of interactions are preferentially retained in *Mycoplasma* genomes, with higher mRNA expression levels in *E. coli* also being associated with retention in some cases. Client/nonclient status has no significant effect on retention in any species. Significance levels: * = $P < 0.05$; ** = $P < 0.01$; *** = $P < 0.0001$. The numbers reported here are for the Class IV clients of Fujiwara et al. (2010), but the results are qualitatively similiar for Class III clients (see Supplementary Material).

750