14. Jones, G. & Rayner, J. M. V. Flight performance, foraging tactics and echolocation in free-living Daubenton's bats *Myotis daubentoni* (Chiroptera: Vespertilionidae). *J. Zool.* **215,** 113–132 (1988).
15. Krull, D., Schumm, A., Metzner, W. & Neuweiler, G. Foraging areas and foraging behavior in the notch-eared bat, *Myotis emarginatus* (Vespertilionidae). *Behav. Ecol. Sociobiol.* **28,** 247–253 (1991).
16. Schumm, A., Krull, D. & Neuweiler, G. Echolocation in the notch-eared bat, *Myotis emarginatus*. *Behav. Ecol. Sociobiol.* **28,** 255–261 (1991).
17. Britton, A. R. C., Jones, G., Rayner, J. M. V., Boonman, A. M. & Verboom, B. Flight performance, echolocation and foraging behaviour in pond bats, *Myotis dasycneme* (Chiroptera: Vespertilionidae). *J. Zool.* **241,** 503–522 (1997).
18. Siemers, B. M. & Schnitzler, H.-U. Natterer's bat (*Myotis nattereri* Kuhl, 1818) hawks for prey close to vegetation using echolocation signals of very broad bandwidth. *Behav. Ecol. Sociobiol.* **47,** 400–412 (2000).
19. Swift, S. M. & Racey, P. A. Gleaning as a foraging strategy in Natterer's bat *Myotis nattereri*. *Behav. Ecol. Sociobiol.* **52,** 408–416 (2002).
20. Fenton, M. B. & Bogdanowicz, W. Relationships between external morphology and foraging behaviour: bats in the genus *Myotis*. *Can. J. Zool.* **80,** 1004–1013 (2002).
21. Arlettaz, R., Jones, G. & Racey, P. A. Effect of acoustic clutter on prey detection by bats. *Nature* **414,** 742–745 (2001).
22. Siemers, B. M., Stilz, P. & Schnitzler, H.-U. The acoustic advantage of hunting at low heights above water: behavioural experiments on the European 'trawling' bats *Myotis capaccinii*, *M. dasycneme* and *M. daubentonii*. *J. Exp. Biol.* **204,** 3843–3854 (2001).
23. Harvey, P. H. & Pagel, M. D. *The Comparative Method in Evolutionary Biology* 138–162 (Oxford Univ. Press, Oxford, 1991).
24. Ruedi, M. & Mayer, F. Molecular systematics of bats of the genus *Myotis* (Vespertilionidae) suggests deterministic ecomorphological convergences. *Mol. Phyl. Evol.* **21,** 436–448 (2001).
25. Simmons, J. A. & Stein, R. A. Acoustic imaging in bat sonar: Echolocation signals and the evolution of echolocation. *J. Comp. Physiol. A* **135,** 61–84 (1980).
26. von Helversen, O. & von Helversen, D. Acoustic guide in bat pollinated flower. *Nature* **398,** 759–760 (1999).
27. Schmidt, S., Hanke, S. & Pillat, J. The role of echolocation in the hunting of terrestrial prey—new evidence for an underestimated strategy in the gleaning bat, *Megaderma lyra*. *J. Comp. Physiol. A* **186,** 975–988 (2000).
28. Morse, P. M. & Ingard, K. U. *Theoretical Acoustics* (Princeton Univ. Press, Princeton, 1986).
29. Kingston, T., Jones, G., Akbar, Z. & Kunz, T. H. Echolocation signal design in Kerivoulinae and Murinae (Chiroptera: Vespertilionidae) from Malaysia. *J. Zool.* **249,** 359–374 (1999).
30. Müller, R. & Kuc, R. Foliage echoes: A probe into the ecological acoustics of bat echolocation. *J. Acoust. Soc. Am.* **108,** 836–845 (2000).

Supplementary Information accompanies this paper on **www.nature.com/nature**.

Acknowledgements We thank all those who assisted with fieldwork. We also thank A. Boonman, A. Denzinger, J. Ostwald, D. Menne, E. Müller, P. Pilz, M. Sánchez-Villagra and P. Stilz for discussions, H. Harty for language assistance and B. Fenton for comments. Our research was funded by the Deutsche Forschungsgemeinschaft (DFG) and a PhD scholarship by Studienstiftung des deutschen Volkes to B.M.S.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to B.M.S. (bjoern.siemers@uni-tuebingen.de).

Metabolic network analysis of the causes and evolution of enzyme dispensability in yeast

Balázs Papp[1,2], Csaba Pál[1,3] & Laurence D. Hurst[1]

[1]*Department of Biology and Biochemistry, University of Bath, BA2 7AY Bath, Somerset, UK*
[2]*Collegium Budapest (Institute for Advanced Study) Szentháromság utca 2, Budapest H-1014, Hungary*
[3]*MTA, Ecology and Theoretical Biology Research Group, Eötvös Loránd University, Pázmány Péter Sétány 1/C, Budapest H-1117, Hungary*

Under laboratory conditions 80% of yeast genes seem not to be essential for viability[1]. This raises the question of what the mechanistic basis for dispensability is, and whether it is the result of selection for buffering or an incidental side product. Here we analyse these issues using an *in silico* flux model[2–5] of the yeast metabolic network. The model correctly predicts the knockout fitness effects in 88% of the genes studied[4] and *in vivo* fluxes. Dispensable genes might be important, but under conditions not yet examined in the laboratory. Our model indicates that this is the dominant explanation for apparent dispensability, accounting for 37–68% of dispensable genes, whereas 15–28% of them are compensated by a duplicate, and only 4–17% are buffered by metabolic network flux reorganization. For over one-half of those not important under nutrient-rich conditions, we can predict conditions when they will be important. As expected, such condition-specific genes have a more restricted phylogenetic distribution. Gene duplicates catalysing the same reaction are not more common for indispensable reactions, suggesting that the reason for their retention is not to provide compensation. Instead their presence is better explained by selection for high enzymatic flux.

Although many single-gene deletions have negligible effects on growth rates under laboratory conditions[1,6], the causes and evolution of gene dispensability has remained a controversial issue[7–9]. The capacity of organisms to compensate mutations partly stems from gene duplicates[8], whereas alternative metabolic pathways might also have a role[7,10–12]. The one previous systematic analysis on a eukaryotic organism[13] used a gene's rate of evolution as a proxy for dispensability, a supposition now considered questionable[14]. A third possibility, and one that has received relatively little attention, is that genes only seem to be non-essential, and that they have important roles under environmental conditions yet to be replicated in the laboratory[8,15].

To investigate the causes of gene dispensability, the metabolic capabilities of the *Saccharomyces cerevisiae* network were calculated using flux balance analysis (FBA)[16]. The previously reconstructed network[2,4] consists of 809 metabolites as nodes (including external metabolites), connected by 851 different biochemical reactions (including transport processes). The method first defines a solution space of fluxes of all metabolic reactions in the network that satisfy the governing constraints (that is, steady state of metabolites, flux capacity, direction of reactions, nutrients available in the environments; see Methods). Next, the optimal use of the metabolic network to produce major biosynthetic components for growth can be found among all possible solutions using various optimization protocols[3,4]. The FBA and MOMA[5] (minimization of metabolic adjustment) protocols enable us to predict the phenotypic behaviour of nutritional changes and gene deletions, along with the concomitant changes in flux distributions.

We start by asking how well the mathematical model predicts experimentally measured fluxes, and the effects of gene deletions. We then use it to address the relative importance of the suggested mechanisms for gene dispensability. Finally, we ask whether dispensability is a directly selected feature or a side consequence.

Owing to the availability of systematic knockout studies[1] and some experimentally measured fluxes under four different growth conditions[17], we can directly test the predictive power of the mathematical protocol. We initiated the model to mimic the growth conditions used in these experimental studies. The model correctly predicts: (1) relative differences in flux values; (2) presence or absence of fluxes in 91–95% of the cases; (3) the fitness effects of 88% of single-gene deletions under nutrient-rich growth conditions[4] (see Supplementary Tables S1–S3). Although the model ignores details of gene regulation, the predicted variations in the activity of metabolic pathways across environments are consistent with observations (Supplementary Tables S1 and S2; see also ref. 3). The method, although robust, is not perfect. Although the frequency of experimentally verified essential genes in the group of genes with zero predicted flux is low on rich medium, it is not zero (8.8% for genes with zero flux compared with 28.8% for the rest; $\chi^2 = 18.54$, $P < 10^{-4}$, 1 degree of freedom (d.f.)). The few essential genes in this group probably represent incomplete biochemical knowledge, missing components from the biomass equation, or pleiotropic gene functions[4].

To investigate the possible causes of empirically observed gene dispensability, we compared the predicted and experimentally measured effects of enzyme deletions under nutrient-rich conditions (see Methods). Enzymes were classified into five mutually exclusive groupings on the basis of the presence of isoenzymes, predicted dispensability and flux distribution: (A) enzymes that are inactive under nutrient-rich conditions but active under some other environments; (B) single-copy enzymes that encode essential reactions; (C) duplicated isoenzymes that encode essential reactions; (D) single-copy enzymes that encode dispensable reactions with non-zero enzymatic flux; (E) duplicated isoenzymes that encode dispensable reactions with non-zero enzymatic flux (Fig. 1; see also Supplementary Table S4).

One possible reason that a gene might be non-essential is that its function is not required under a given circumstance. Indeed, the model predicts that a large fraction of experimentally 'verified' non-essential genes should have zero enzymatic flux under nutrient-rich conditions (68.3%). This result indicates that many enzymes make no contribution to the production of biomass components under this condition.

Can we find conditions under which the apparently non-essential genes with zero predicted flux have an important fitness contribution? As with a previous study[3], we set up nine different growth conditions that might have been representative during the evolution of this species (Fig. 2), and performed enzyme-based deletion studies. Under nutrient-rich medium, the fraction of essential reactions and reactions with non-zero flux are especially low (Fig. 2). Importantly, more than half of the experimentally verified non-essential genes that are predicted to have zero flux under nutrient-rich condition appear to have non-zero flux under some other conditions (54.5%, 79 out of 145 cases). These we define as 'conditionally active' genes. This contrasts with the unconditionally active (non-zero flux under all conditions examined) genes and those for which we cannot find conditions under which they are active. These results suggest that 37–68% of the seemingly dispensable genes are environmentally specific (Supplementary Table S4).

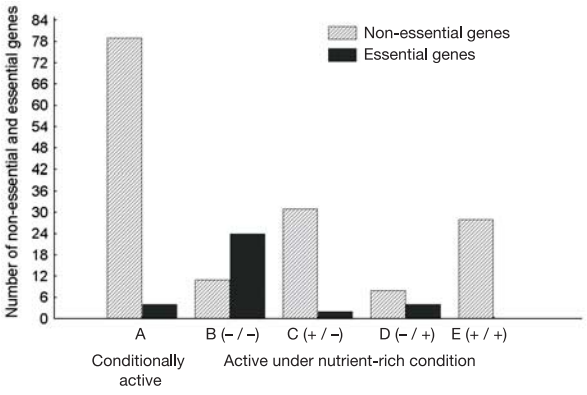

Figure 1 Number of experimentally verified essential and non-essential genes in different categories. The classes are: (A) predicted to have zero flux under nutrient-rich conditions, but non-zero flux in at least one other environment; (B) single-copy genes predicted to catalyse essential reactions; (C) duplicate genes predicted to catalyse essential reactions; (D) single-copy genes predicted to catalyse dispensable reactions; and (E) duplicate genes predicted to catalyse dispensable reactions. When comparing groups B and C, of the 68 metabolic genes that are predicted to catalyse essential reactions, 33 are known to have a duplicated isoenzyme. Only about 6% of those that have an isoenzyme are observed to be essential *in vivo*, whereas the proportion of essential genes is roughly 69% among those without an isoenzyme ($\chi^2 = 28.1$, d.f. $= 1$, $P < 10^{-6}$). When comparing groups B and D, of the 47 single-copy genes 35 are predicted to catalyse essential reactions whereas 12 are predicted to be dispensable. The fraction of essential genes is indeed higher in the former class (about 69% versus about 33%, $\chi^2 = 4.6$, $P < 0.05$). A plus sign indicates the presence and a minus sign the absence of isoenzymes/flux compensation.

Many of the conditionally active genes (76%) are predicted to catalyse reactions that are essential under specific conditions. It remains to be seen whether experiments will actually confirm the detailed predictions.

If the above classifications are correct, one should expect differences in the phylogenetic distribution of enzymes with unconditional and conditional activity, as the latter group can more easily be lost during evolution if the appropriate environment becomes rare. We investigated this issue using a database[18] of enzymatic reactions across 133 sequenced genomes. We found that enzymes having non-zero fluxes under only a few environmental conditions tend to have a more limited phylogenetic distribution than enzymes with unconditional activity (Fig. 3).

Why is it that there are dispensable genes associated with non-zero predicted fluxes under nutrient-rich conditions? To shed light on the relative importance of the compensation mechanisms (duplication versus flux reorganization in the network), we first compared the fraction of experimentally verified essential genes between single-copy enzymes and duplicated isoenzymes. For the comparison, only genes that are predicted to encode essential reactions were considered (Fig. 1, class B and C). The low fraction of essential enzymes with isoenzymes strongly supports previous claims that dispensability partially results from redundant gene duplicates[8]. The two exceptions (failure of compensation) might be due to lack of duplicate enzyme activity in the same subcellular compartment (Supplementary Table S5). Assuming that all or none of the non-essential genes of class E are due to gene duplication rather than flux reorganization, we obtained a lower (14.6%) and upper estimate (27.8%) for the contribution of gene duplication to dispensability (Supplementary Table S4).

The ability of duplicates to buffer each other's loss may be considered a special case of a more general mode of compensation, in which the metabolic network adjusts the metabolic flux, and, in so doing, mitigates the loss of individual genes. Compensation occurs only if the original enzyme has a contribution to biomass production (non-zero flux), but the underlying reaction is dispensable for growth[19] (class D and E, Fig. 1). To see the effect of flux reorganization on *in vivo* gene dispensability independent of duplicate gene copies, we compared the fraction of experimentally verified essential genes between class B and D under the assumption that essential and dispensable reactions should differ in the network's ability to compensate for their loss. Indeed, this is what we observed (Fig. 1). However, this mode of compensation can only explain 3.8–17% of gene dispensability (Fig. 1; see also Supplementary Table S4).

What factors might limit the compensatory capability of the metabolic network? Our model demonstrates that the extent of flux

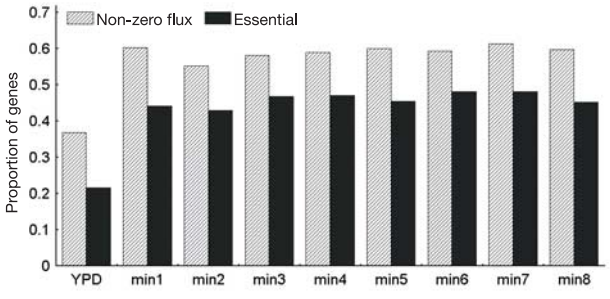

Figure 2 The proportion of genes predicted to have non-zero flux and to be essential under different growth conditions. Single-enzyme knockouts were simulated under nine different growth conditions listed below. The total number of investigated enzymes was 310 for all conditions (isoenzymes were counted only once). Environments were: YPD, rich glucose, low O_2; min1, minimal glucose, low O_2; min2, minimal glucose, anaerobic; min3, minimal ethanol, low O_2; min4, minimal acetate, low O_2; min5, minimal glucose, carbon limited; min6, minimal glucose, nitrogen limited; min7, minimal glucose, phosphate limited; min8, minimal glucose, sulphur limited.

reorganization positively correlates with the predicted fitness effect of the compensated knockout (Supplementary Fig. S1), suggesting that the yeast metabolic network has difficulties in tolerating large flux reorganization.

Although it is clear that the presence of isoenzymes has a large effect on gene dispensability, is it likely that dispensability has evolved to enable such compensation or might it instead be a side product? In the former case, one would expect gene duplicates to be preferentially maintained if they specify a crucial function to provide a shield against intracellular noise[20]. If it were so, then one would expect the most important reactions of the network to be under the control of isoenzymes. In contrast to expectations, essential reactions are not more likely to be catalysed by isoenzymes compared to non-essential reactions (Supplementary Table S6). One possible alternative explanation for the maintenance of isoenzymes is that selection favours enhanced dosage of the same product to provide high enzymatic flux[21]. We found strong support for this theory: the average predicted flux of reactions catalysed by isoenzymes is higher under all conditions than that of reactions catalysed by single-copy enzymes (Supplementary Table S7).

Although many suggest that the high number of dispensable genes is evidence for the selection of robustness to perturbation, our results support a different conclusion. For the most part, knockout studies performed under nutrient-rich conditions provide a substantial underestimate of the number of genes that are essential under some environmental conditions (Fig. 2). Moreover, non-essential genes may make small but significant contributions to fitness even under routine growth conditions, but the effects are not large enough to be detected[8,15]. Of those that seem to be truly dispensable (non-zero flux and viable knockout), at least in the case of gene duplicates, the dispensability is better explained as a side consequence, rather than the result of selection to favour resilience. These results, along with previous studies[21,22], indicate that the dosage requirements have an important influence on the evolutionary maintenance of gene duplicates in yeast.

Is it likely that environmental specificity explains much of the apparent dispensability seen in other organisms? Recent systematic deletion studies[1,23–25] suggest that despite the apparent differences in metabolic complexity and the extent of gene duplication across free-living bacterial and eukaryotic species, the fraction of essential genes under a given laboratory condition is generally low, in the range of 7–19%. In contrast to these low figures in free-living species, the fraction of essential genes is 55–73% in the *Mycoplasma genitalium* genome[26]. This is not simply due to a rarity of gene duplicates. We suggest that, being a parasite with strict host and tissue specificity, *M. genitalium* should have relatively few condition-specific genes. We can test this hypothesis by comparing the proportion of single-copy genes that are non-essential in yeast and in *Mycoplasma*. In agreement with the theory, in yeast this is at least 62%, whereas this drops to 24% in *Mycoplasma*. More direct evidence comes from a data set on the growth phenotypes of mutant strains in *Escherichia coli*[27]: most genes show severe fitness defects only under a small fraction (10%) of the 282 conditions investigated (Supplementary Fig. S2). Moreover, in agreement with the results on yeast metabolic genes, condition-specific genes of *E. coli* show limited phylogenetic distribution (Supplementary Fig. S3).

These issues are important, not least because they affect our ability to test reliably hypotheses concerning the evolution of genes and genomes. For example, the abundance of environmentally specific genes in yeast might explain why dispensability under nutrient-rich conditions only very weakly correlates with the rate of protein evolution[14]. □

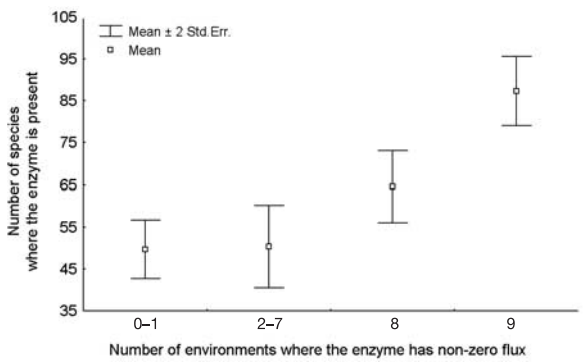

Figure 3 Relationship between phylogenetic distribution and condition specificity. Enzymes having non-zero fluxes in most of the simulated environments have wide phylogenetic distributions (analysis of variance: $F = 17.72$; d.f. $= 3, 281$; $P < 10^{-9}$). Data are means (square symbols) ± 2 standard errors.

Methods

Filtered data sets used in this study

To investigate the metabolic network we used a previously compiled list of enzymatic reactions in yeast[2]. The metabolic reconstruction gives accurate information on the stoichiometry and direction of enzymatic reactions and on the presence of isoenzymes. Cytosolic, mitochondrial and extracellular metabolites are treated separately, and the data set also includes a list of transport reactions between compartments. Reactions catalysed by isoenzymes were considered as a single flux, eliminating duplicate reactions. For data analyses we restricted our attention to unambiguously classified enzymes; that is, those with complete EC numbers. Sequence similarity between isoenzyme pairs was computed by a pair-wise BLASTP[28] search (we used an *E*-value of <0.01 as a cutoff to recognize even distant duplicated isoenzymes). We checked whether the duplicated isoenzymes act in protein complexes, using both the compiled list on yeast metabolism and the MIPS CYGD[29] catalogue on known protein complexes, and these pairs ($N = 21$) were excluded from further analysis. Classification of the dispensability of genes on glucose-rich medium (essential versus non-essential) was as provided by the *Saccharomyces* Genome Deletion Project (http://www-sequence.stanford.edu/group/yeast_deletion_project/), which contains information on large-scale knockout studies[1]. To minimize confounding factors in designation of dispensability, multienzyme polypeptides, genes participating in protein complexes (according to the MIPS CYGD catalogue of annotated complexes) and genes with overlapping reading frames were excluded from all of the analyses. The KEGG database[18] was used to identify the enzymatic reactions of 133 bacterial and eukaryotic species with complete genome sequences (a filtered set of genomes consisting of only one genome per genus gives similar results).

Basic metabolic network model

Flux distribution and metabolic network capabilities were investigated by modification of a previously elaborated genome-scale metabolic flux balance model of *S. cerevisiae*[2–4]. The model starts by specifying the mass balance constraints around intracellular metabolites. These constraints specify a series of linear equations of individual reaction fluxes that must be fulfilled to enable steady state of metabolites. Mathematically, this is represented by $S\mathbf{v} = 0$, where S is the $m \times n$ stoichiometric matrix, with m as the number of metabolites, and n as the number of reactions. An S_{ij} element of the stoichiometric matrix represents the contribution of a jth reaction to metabolite i. The vector $\mathbf{v}$ represents the individual fluxes of the network. Besides mass balance equations, reversibility/irreversibility constraints are also imposed on individual internal fluxes ($\mathbf{v}_i > 0$ for irreversible reactions). Import flux of external metabolites was constrained to be zero when not available under the studied environment. The system also includes a biomass reaction (with rate $\mathbf{v}_{growth}$) that represents the relative contribution of metabolites to the cellular biomass of yeast (see Supplementary equation S1). Linear programming was used to find a particular flux distribution that maximizes $\mathbf{v}_{growth}$ under the described constraints and defined nutrient uptake rates. We used this optimal flux configuration as the wild type under the given growth conditions. We have investigated nine different environments (see Fig. 2).

Calculating the fitness effect of gene knockouts

Enzyme deletions were simulated by constraining the flux of the corresponding reactions to zero and calculating the knockout flux configuration under the assumption that knockout metabolic fluxes undergo a minimal flux redistribution with respect to the flux configuration of the wild type (minimization of metabolic adjustment, MOMA protocol[5]). Using a different optimization protocol[3,4] gives almost exactly the same results (data not shown). Thus, calculation of knockout $\mathbf{v}_{growth}$ requires quadratic programming to find a point in flux space, which is closest to wild type. The software tool Cplex 7.5 was used to solve these linear and quadratic optimization problems. We scaled fitness relative to the wild type. Essential enzymes are defined as knockout strains having a growth rate of at most one-half of the wild type. We observed a clear bimodal distribution of knockout fitnesses: enzymes predicted to be non-essential have minimal or no effect on growth (Supplementary Fig. S4). If the optimization problem for a given knockout was infeasible we treated the enzyme as essential. Flux and knockout phenotype predictions were not attempted for enzymes located on dead-end pathways or for enzymes with functions not represented in the biomass equation[4] (for example, glycoprotein, haem and chitin metabolism, transfer RNA synthetases). In the case of reactions catalysed by isoenzymes, the

duplicates were deleted to obtain predictions on the dispensability of the underlying enzymatic reaction.

Comparison of *Mycoplasma* and *Saccharomyces* genomes

We calculated the frequency of non-essential genes in the *M. genitalium* and the *S. cerevisiae* genomes (only single-copy genes were considered). Gene duplicates were identified using a BLAST protein search, with at least 25% amino acid similarity (using different thresholds do not affect our results). The list of putative essential *Mycoplasma* genes is from ref. 26. We found 1,881 out of 3,003 single-copy yeast genes that are non-essential. This figure is 83 out of 356 genes for *Mycoplasma*.

Received 14 January; accepted 30 April 2004; doi:10.1038/nature02636.

1. Giaever, G. *et al.* Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* **418,** 387–391 (2002).
2. Forster, J., Famili, I., Fu, P., Palsson, B. O. & Nielsen, J. Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. *Genome Res.* **13,** 244–253 (2003).
3. Famili, I., Forster, J., Nielsen, J. & Palsson, B. O. *Saccharomyces cerevisiae* phenotypes can be predicted by using constraint-based analysis of a genome-scale reconstructed metabolic network. *Proc. Natl Acad. Sci. USA* **100,** 13134–13139 (2003).
4. Forster, J., Famili, I., Palsson, B. O. & Nielsen, J. Large-scale evaluation of *in silico* gene deletions in *Saccharomyces cerevisiae*. *OMICS* **7,** 193–202 (2003).
5. Segre, D., Vitkup, D. & Church, G. M. Analysis of optimality in natural and perturbed metabolic networks. *Proc. Natl Acad. Sci. USA* **99,** 15112–15117 (2002).
6. Steinmetz, L. M. *et al.* Systematic screen for human disease genes in yeast. *Nature Genet.* **31,** 400–404 (2002).
7. Wagner, A. Robustness against mutations in genetic networks of yeast. *Nature Genet.* **24,** 355–361 (2000).
8. Gu, Z. *et al.* Role of duplicate genes in genetic robustness against null mutations. *Nature* **421,** 63–66 (2003).
9. Gu, X. Evolution of duplicate genes versus genetic robustness against null mutations. *Trends Genet.* **19,** 354–356 (2003).
10. Stelling, J., Klamt, S., Bettenbrock, K., Schuster, S. & Gilles, E. D. Metabolic network structure determines key aspects of functionality and regulation. *Nature* **420,** 190–193 (2002).
11. Papin, J. A., Price, N. D., Wiback, S. J., Fell, D. A. & Palsson, B. O. Metabolic pathways in the post-genome era. *Trends Biochem. Sci.* **28,** 250–258 (2003).
12. Emmerling, M. *et al.* Metabolic flux responses to pyruvate kinase knockout in Escherichia coli. *J. Bacteriol.* **184,** 152–164 (2002).
13. Kitami, T. & Nadeau, J. H. Biochemical networking contributes more to genetic buffering in human and mouse metabolic pathways than does gene duplication. *Nature Genet.* **32,** 191–194 (2002).
14. Pál, C., Papp, B. & Hurst, L. D. Rate of evolution and gene dispensability. *Nature* **421,** 496–497 (2003).
15. Thatcher, J. W., Shaw, J. M. & Dickinson, W. J. Marginal fitness contributions of nonessential genes in yeast. *Proc. Natl Acad. Sci. USA* **95,** 253–257 (1998).
16. Edwards, J. S. & Palsson, B. O. The *Escherichia coli* MG1655 *in silico* metabolic genotype: its definition, characteristics, and capabilities. *Proc. Natl Acad. Sci. USA* **97,** 5528–5533 (2000).
17. Daran-Lapujade, P. *et al.* Role of transcriptional regulation in controlling fluxes in central carbon metabolism of *Saccharomyces cerevisiae*: a chemostat culture study. *J. Biol. Chem.* **279,** 9125–9138 (2004).
18. Nakao, M. *et al.* Genome-scale gene expression analysis and pathway reconstruction in KEGG. *Genome Inform. Ser. Workshop Genome Inform.* **10,** 94–103 (1999).
19. Schilling, C. H., Edwards, J. S., Letscher, D. & Palsson, B. O. Combining pathway analysis with flux balance analysis for the comprehensive study of metabolic systems. *Biotechnol. Bioeng.* **71,** 286–306 (2000).
20. Wagner, G. P., Booth, G. & Bagheri-Chaichian, H. A population genetic theory of canalization. *Evolution* **51,** 329–347 (2000).
21. Seoighe, C. & Wolfe, K. H. Yeast genome evolution in the post-genome era. *Curr. Opin. Microbiol.* **2,** 548–554 (1999).
22. Papp, B., Pal, C. & Hurst, L. D. Dosage sensitivity and the evolution of gene families in yeast. *Nature* **424,** 194–197 (2003).
23. Gerdes, S. Y. *et al.* Experimental determination and system level analysis of essential genes in *Escherichia coli* MG1655. *J. Bacteriol.* **185,** 5673–5684 (2003).
24. Kobayashi, K. *et al.* Essential *Bacillus subtilis* genes. *Proc. Natl Acad. Sci. USA* **100,** 4678–4683 (2003).
25. Kamath, R. S. *et al.* Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. *Nature* **421,** 231–237 (2003).
26. Hutchison, C. A. *et al.* Global transposon mutagenesis and a minimal *Mycoplasma* genome. *Science* **286,** 2165–2169 (1999).
27. Glasner, J. D. *et al.* ASAP, a systematic annotation package for community analysis of genomes. *Nucleic Acids Res.* **31,** 147–151 (2003).
28. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25,** 3389–3402 (1997).
29. Mewes, H. W. *et al.* MIPS: a database for genomes and protein sequences. *Nucleic Acids Res.* **30,** 31–34 (2002).

Supplementary Information accompanies the paper on **www.nature.com/nature**.

Acknowledgements We thank J. Förster and B. Palsson for discussions and providing information on the yeast flux balance model. We also thank J. Glasner for providing growth data on the *E. coli* wild-type strain. B.P. and C.P. were supported by the Hungarian National Research Grant Foundation (OTKA).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to L.D.H. (l.d.hurst@bath.ac.uk).

Temporal difference models describe higher-order learning in humans

Ben Seymour[1], John P. O'Doherty[1], Peter Dayan[2], Martin Koltzenburg[3], Anthony K. Jones[4], Raymond J. Dolan[1], Karl J. Friston[1] & Richard S. Frackowiak[1,5]

[1]*Wellcome Department of Imaging Neuroscience, 12 Queen Square, London WC1N 3BG, UK*
[2]*Gatsby Computational Neuroscience Unit, Alexandra House, 17 Queen Square, London WC1N 3AR, UK*
[3]*Institute of Child Health, University College London, 30 Guilford St, London WC1N 1EH, UK*
[4]*University of Manchester Rheumatic Diseases Centre, Hope Hospital, Manchester M6 8HD, UK*
[5]*Fondazione Santa Lucia, 00179 Rome, Italy*

The ability to use environmental stimuli to predict impending harm is critical for survival. Such predictions should be available as early as they are reliable. In pavlovian conditioning, chains of successively earlier predictors are studied in terms of higher-order relationships, and have inspired computational theories such as temporal difference learning[1]. However, there is at present no adequate neurobiological account of how this learning occurs. Here, in a functional magnetic resonance imaging (fMRI) study of higher-order aversive conditioning, we describe a key computational strategy that humans use to learn predictions about pain. We show that neural activity in the ventral striatum and the anterior insula displays a marked correspondence to the signals for sequential learning predicted by temporal difference models. This result reveals a flexible aversive learning process ideally suited to the changing and uncertain nature of real-world environments. Taken with existing data on reward learning[2], our results suggest a critical role for the ventral striatum in integrating complex appetitive and aversive predictions to coordinate behaviour.

Substantial evidence in humans and other animals has outlined a network of brain regions involved in the prediction of painful and aversive events[3–6]. Most of this work has concentrated on its simplest realization, namely first-order pavlovian fear conditioning; however, the predictions in this paradigm are rudimentary, showing little of the complexities associated with sequences of predictors that are critical in psychological investigations of prognostication[7]. These latter studies led to a computational account called temporal difference learning[1,8], which has close links with methods for prediction, and optimal action selection, in engineering[9]. When applied to first-order appetitive conditioning, temporal difference learning provides a compelling account of neurophysiological data, both with respect to the phasic activity of dopamine neurons in animal studies, and with blood-oxygenation-level-dependent (BOLD) activity in human functional neuroimaging studies[10–15]. However, beyond this simple paradigm, the utility of temporal difference models to describe learning remains largely unexplored. Here we provide a neurobiological investigation based on aversive and, importantly, sequential conditioning.

We used fMRI to investigate the pattern of brain responses in humans during a second-order pain learning task. Fourteen healthy subjects were shown two visual cues in succession, followed by a high- or low-intensity pain stimulus delivered to the left hand (Fig. 1a) (see Methods). Subjects were told that they were performing a study of reaction times and were asked to judge whether the cues appeared on the left or on the right side of a display monitor. The second cue in each sequence was fully predictive of the strength of the subsequently experienced pain; however, the first cue only allowed a probabilistic prediction. Thus, in a small percentage of