

Research article

Open Access

Recombination rate and protein evolution in yeast

Tim Connallon* and L Lacey Knowles

Address: Department of Ecology and Evolutionary Biology, Museum of Zoology, University of Michigan, Ann Arbor, MI 48109-1079, USA

Email: Tim Connallon* - tconnal@umich.edu; L Lacey Knowles - knowlesl@umich.edu

* Corresponding author

Published: 27 November 2007

Received: 22 February 2007

BMC Evolutionary Biology 2007, **7**:235 doi:10.1186/1471-2148-7-235

Accepted: 27 November 2007

This article is available from: <http://www.biomedcentral.com/1471-2148/7/235>

© 2007 Connallon and Knowles; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: Theory and artificial selection experiments show that recombination can promote adaptation by enhancing the efficacy of natural selection, but the extent to which recombination affects levels of adaptation across the genome is still an open question. Because patterns of molecular evolution reflect long-term processes of mutation and selection in nature, interactions between recombination rate and genetic differentiation between species can be used to test the benefits of recombination. However, this approach faces a major difficulty: different evolutionary processes (i.e. negative versus positive selection) produce opposing relationships between recombination rate and genetic divergence, and obscure patterns predicted by individual benefits of recombination.

Results: We use a combination of polymorphism and genomic data from the yeast *Saccharomyces cerevisiae* to infer the relative importance of nearly-neutral (i.e. slightly deleterious) evolution in different gene categories. For genes with high opportunities for slightly deleterious substitution, recombination substantially reduces the rate of molecular evolution, whereas divergence in genes with little opportunity for slightly deleterious substitution is not strongly affected by recombination.

Conclusion: These patterns indicate that adaptation throughout the genome can be strongly influenced by each gene's recombinational environment, and suggest substantial long-term fitness benefits of enhanced purifying selection associated with sexual recombination.

Background

Genetic drift is expected to overpower natural selection when selection is weak and effective population size (N_e) is small [1-3]. Recombination increases the effective population size in which genes evolve by reducing interference between linked loci under selection [4,5]. As a result, recombination is expected to facilitate the spread of beneficial mutations and the elimination of deleterious mutations [6,7]. Because recombination rates vary between different regions of a genome [e.g. yeast: [8]; *Drosophila*: [9]; Mammals: [10]; plants: [11]], adaptation at the molecular level might be strongly affected by each gene's

recombinational environment – genes evolving in low recombination regions are expected to be poorly adapted relative to those in high recombination regions [12,13].

Comparative genome analyses are potentially useful for assessing whether recombination promotes adaptation because long-term evolutionary processes are reflected in patterns of genetic divergence between species [13-19]. However, genomic approaches face a major challenge – multiple processes can contribute to evolutionary divergence between species and each predicts a different relationship between protein evolution and recombination

rate (Table 1). The rate of neutral substitution is unaffected by recombination and will tend to reduce correlations between recombination rate and total nucleotide divergence between species [20], mildly deleterious (i.e. nearly neutral) substitutions will generate a negative correlation between recombination and divergence [21], and adaptive substitutions generate a positive correlation between recombination and divergence [22].

The relationship between recombination and divergence will be shaped by the predominant process of molecular evolution (i.e. neutral; slightly deleterious; adaptive). To test whether recombination facilitates adaptation throughout the genome (by enhancing purifying and positive selection), genes evolving under purifying selection and those evolving via positive selection should be analyzed separately, as each predicts a different relationship between divergence and recombination rate. Unfortunately, inferring the processes causing molecular divergence has traditionally been problematic without detailed within- and between-species genetic data [23], which limits the extent of the genome that can be analyzed.

Here we take an alternative approach. By capitalizing on an extensive volume of yeast (*Saccharomyces* spp.) genomic and polymorphism data, individual genes can be partitioned by their 'opportunity' for slightly deleterious evolution, and benefits of recombination can be tested. Furthermore, direct estimates of local recombination rates are available for most genes in the *S. cerevisiae* genome, whole-genome sequencing projects provide data for estimating protein evolutionary rates, and previous studies have revealed that the average strength of selection predictably varies between genes with different functional attributes [reviewed in [22]; see below]. Evolutionary theory predicts that mildly deleterious substitutions will accumulate readily in genes subject to weak selection, but not in those subject to strong selection (see Fig. 1). To the extent that substitutions differentiating species are often deleterious (e.g., genes with weakly-selected mutations;

Table 1: Different processes of molecular evolution produce different correlations between recombination rate and nucleotide divergence between species¹.

Model of evolution	Fitness effect of substitutions ²	Correlation between recombination rate and divergence
Neutral	$s = 0$	None
Slightly deleterious	$-1/(2N_e) < s < 0$	Negative
Adaptive	$s > 0$	Positive

¹ See text for details

² The selection coefficient, s ($-1 \leq s < \infty$), reflects the strength and direction of selection acting upon new mutations during the process of fixation

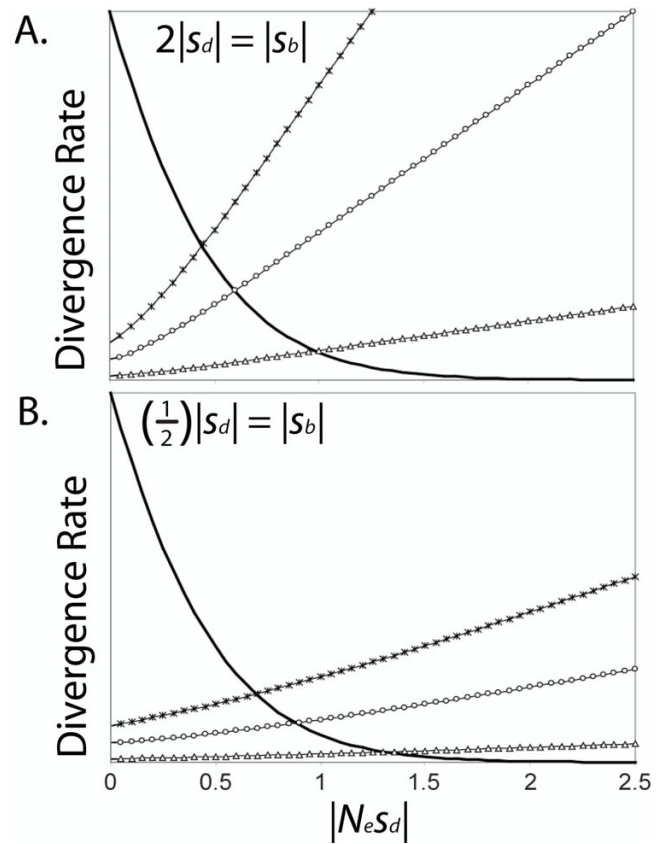


Figure 1

Substitution rate as a function of the effective strength of selection. The bold solid curve represents the slightly deleterious divergence rate between species. Remaining curves represent adaptive divergence under three scenarios: open triangles when the beneficial mutation rate (u_b) is 1% of the deleterious mutation rate (u_d); open circles when u_b is 5% of u_d ; 'X'-marks when u_b is 10% of u_d . The strength of selection against deleterious mutations ($|N_e s_d|$) is shown on the x-axis, where s_d is the average strength of purifying selection. A. The average strength of positive selection (s_b) is equal to the average strength of purifying selection (s_d); B. s_b is twice as strong as s_d ; C. s_d is twice as strong as s_b . When selection is weak, mildly-deleterious substitutions outnumber adaptive substitutions; divergence in weakly-selected genes is therefore predicted to be negatively correlated with local recombination rate. As the strength of selection increases, adaptive substitutions predominate; divergence in strongly-selected genes is therefore predicted to be positively correlated with recombination rate. The adaptive divergence rate is $u_b(1 - e^{-4N_e s_p})/(1 - e^{-4N_e s})$, where s is the average benefit conferred by each mutation, and p is the initial frequency of each mutation (results for $p = 0.0001$ are shown). The slightly deleterious rate is $u_d(1 - e^{-4N_e s_p})/(1 - e^{-4N_e s})$, where u_d is the deleterious mutation rate, and s is the average cost of each mutation (equations modified from [30]; $N_e = N$). The assumption that the beneficial mutation rate is much smaller than the deleterious mutation rate is supported by theory and mutation accumulation experiments [38,55-57; but see 58,59].

Fig. 1 – small $|N_e s_d|$, increased recombination is expected to decrease the rate of protein evolution by inhibiting the spread of deleterious mutations. However, in genes with little opportunity for slightly deleterious divergence (e.g.

genes subject to strong purifying selection; Fig. 1 – large $|N_e s_d|$), recombination will increase the rate of divergence by enhancing the spread of beneficial mutations (to the extent that such mutations frequently arise).

Results & Discussion

Inferring the fitness effects of deleterious mutations

Highly expressed genes appear to evolve under stronger purifying selection than low-expressed genes (although the mechanistic basis of this pattern is still debated; [24-27]). Consequently, gene expression level is a good predictor of the average fitness effect of deleterious mutations. Experimental gene knockouts have also identified suites of genes that are essential for survival, while many others are nonessential. To the extent that whole gene knockout phenotypes reflect the fitness effects of individual mutations, mutations in essential genes are predicted to have larger fitness effects than mutations in nonessential genes [25,28-30]. Lastly, proteins have variable numbers of interaction partners (protein-protein interactions per gene – PPI – range up to nearly 300 PPI in yeast; Connallon & Knowles unpub.), which indicate the level of constraint due to pleiotropy [25]. Because individual mutations are likely to disrupt more cellular processes in genes with many PPI compared to those with few PPI, purifying selection is expected to be stronger in genes with many PPI [31].

Previous inferences of the strength of purifying selection acting on different gene categories are based on an observed elevated rate of nonsynonymous substitution in low-expressed nonessential genes with few PPI [25], which assumes that elevated rates of substitution are caused by genetic drift. An alternative possibility is that rapidly evolving genes undergo frequent bouts of positive selection. To test this assumption, we analyzed available polymorphism data from *S. cerevisiae* genes (see Methods and Additional file 1). Under a neutral/nearly-neutral model, patterns of within species polymorphism are expected to mirror patterns of interspecific substitution (i.e. genes with high substitutions rates also exhibit high levels of polymorphism [32]). Positive selection decouples patterns of polymorphism and divergence and is expected to increase the number of substitutions relative to polymorphisms [33].

Low-expressed genes harbor more nonsynonymous polymorphisms (represented by the P_n/P_s ratio; P_n and P_s refer to nonsynonymous to synonymous polymorphisms, respectively) and substitutions (D_n/D_s), than highly expressed genes, consistent with the neutral/nearly-neutral model (Fig. 2; see Additional file 1). Low-expressed genes also harbor higher levels of moderate- to high-frequency polymorphism (i.e. "non-singleton" polymorphism). High frequency polymorphisms are not expected

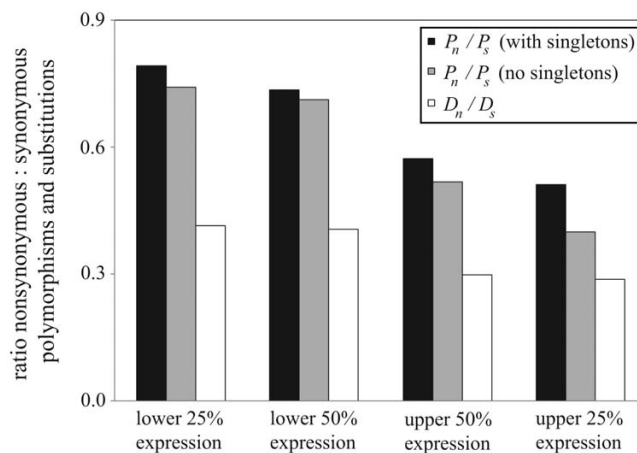


Figure 2

Ratios of replacement to silent polymorphism (P_n/P_s) in *S. cerevisiae*, and substitutions (D_n/D_s) between *S. cerevisiae* and *S. paradoxus*. Results were obtained by pooling polymorphism and divergence data for multiple genes within each expression category (see Fig. S1 for similar results using a different approach). P_n/P_s ratios are lower in highly expressed genes: upper vs. lower 50% with singletons, $P = 0.172$; upper vs. lower 25% with singletons, $P = 0.026$; upper vs. lower 50% without singletons, $P = 0.203$; upper vs. lower 25% without singletons, $P = 0.023$. D_n/D_s ratios are lower for highly expressed genes for all comparisons ($P < 0.0001$).

to be strongly deleterious (see [34]), but rather, will consist of neutral and slightly deleterious mutations – mutations that can potentially become fixed via genetic drift. D_n/D_s ratios are significantly lower than P_n/P_s ratios for all gene expression categories (G test; singletons included: $P < 0.0001$; singletons excluded: $P < 0.005$, except for upper 25% expressed genes: $P > 0.1$), indicating that the predominant pattern of selection on yeast genes is purifying. Similar results are reached by partitioning the data into gene essentiality categories (see Additional file 1; a meaningful statistical analysis based on PPI was not possible due to small sample size). The results strongly support previous inferences of selection intensity based on divergence data, and suggest that nonessential genes with low expression are relatively likely to evolve under a nearly-neutral process.

Recombination and protein divergence

Analysis of 4786 genes in the yeast species *Saccharomyces cerevisiae* and *S. paradoxus* shows that nonsynonymous divergence is weakly negatively correlated with recombination rate (partial $r = -0.052$, $P < 0.01$ after Bonferroni correction; as previously reported by Pal et al. [14]). This negative relationship is markedly stronger for low-expressed genes, particularly nonessential genes with few PPI (Fig. 3). In contrast, the divergence of highly expressed genes tends to correlate positively with recombination rate, though all such associations are weak. For both classes of nonessential genes as well as for the entire data-

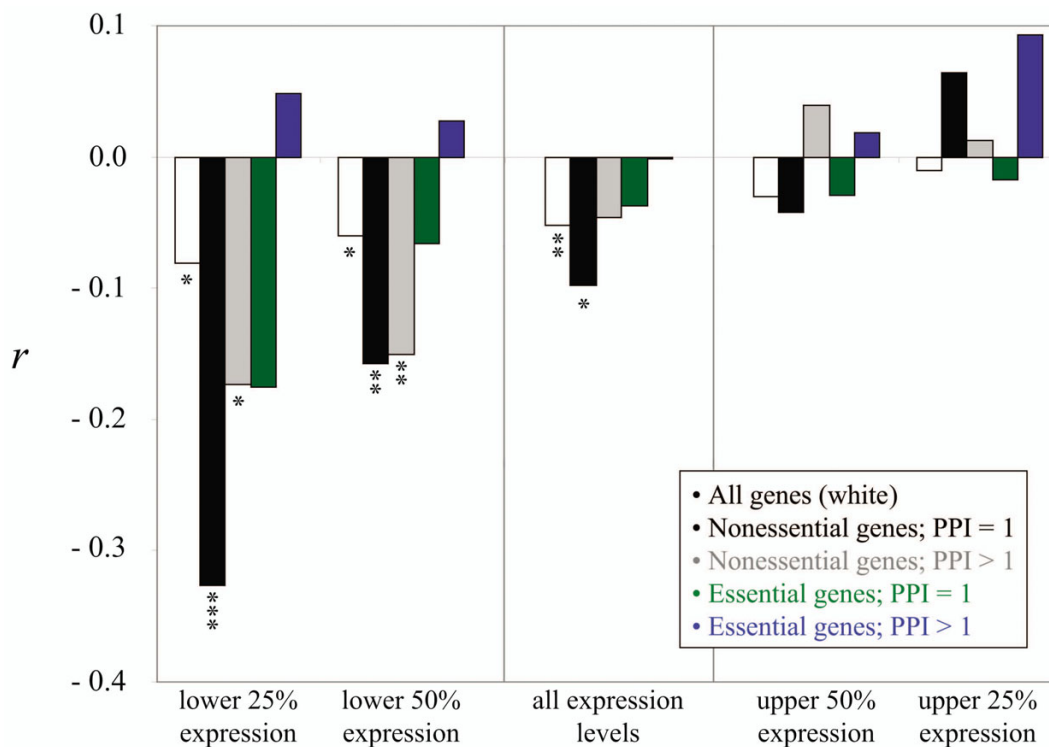


Figure 3

Recombination and protein evolutionary rate. The relationship (r = partial correlation coefficient; see materials and methods) between recombination rate and dN for five gene expression intervals: the lower gene expression quartile (2.25 to 3.15 log mRNA abundance), the lower 50% expression (3.31 to 4.54 log mRNA abundance), and the upper gene expression quartile (3.53 to 4.54 log mRNA abundance). * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$ (after Bonferroni correction for five comparisons).

set, the relationship between recombination and divergence in low-expressed genes is more negative than that of highly expressed genes (Upper vs. lower 25% expression quartiles: all genes, $n_{low} = 1197$, $n_{high} = 1197$, $P = 0.041$; nonessential PPI = 1, $n_{low} = 246$, $n_{high} = 218$, $P < 0.0001$; nonessential PPI > 1, $n_{low} = 222$, $n_{high} = 142$, $P = 0.041$. Upper vs. lower 50% expression quartiles: all genes, $n_{low} = 2393$, $n_{high} = 2393$, $P = 0.148$; nonessential PPI = 1, $n_{low} = 424$, $n_{high} = 411$, $P = 0.046$; nonessential PPI > 1, $n_{low} = 435$, $n_{high} = 327$, $P = 0.005$).

These results clearly show that recombination can influence the rate of protein evolution at a genome wide scale and that the impact of recombination rate variation is strongest for low-expressed, nonessential genes with few PPI. Associations between recombination and divergence rate cannot be explained by covariation between recombination rate and several variables that independently affect protein evolution (the effects were controlled for; see Methods). Estimates of the relative recombination rate between genes are coarse and limited by the quality of the *S. cerevisiae* recombination map, and there are potential evolutionary changes in recombination between *S. cerevi-*

siae and *S. paradoxus*. However, both of these factors will decrease the strength of associations between divergence and recombination, and will cause our test to be conservative.

Mutation bias is also unlikely to account for the effect of recombination on protein evolution. We present associations between recombination and divergence at nonsynonymous sites (dN) rather than between recombination and dN/dS ratios because synonymous sites are under selection in yeast [e.g. [35]]. Indeed, codon usage bias (F_{op}) is positively correlated with recombination (as previously reported [35]), most strongly for highly expressed genes, which presumably have stronger selection for optimal codons (highest 50% expression partial $r_{rec-Fop} = 0.214$ vs. lowest 50% $r_{rec-Fop} = 0.100$; highest 25% expression partial $r_{rec-Fop} = 0.249$ vs. lowest 25% $r_{rec-Fop} = 0.089$; $P < 0.0001$ for both comparisons). As a consequence, dS is negatively correlated with recombination (highest 50% expression partial $r_{rec-Fop} = -0.131$ vs. lowest 50% $r_{rec-Fop} = -0.077$, $P = 0.031$; highest 25% expression partial $r_{rec-Fop} = -0.122$ vs. lowest 25% $r_{rec-Fop} = -0.075$, $P = 0.123$). Furthermore, direct estimates indicate higher mutation rates in

regions of high recombination [36,37], which should make our tests conservative. Despite these caveats with respect to using dS to estimate underlying mutational dynamics across the genome, dN/dS produces nearly-identical patterns of covariation with recombination (see Additional file 1, Fig. S3).

The results are consistent with evolutionary theory suggesting that recombination enhances the efficacy of selection [e.g. [6,12]]. Mutations with weak fitness effects respond to selection when the effective population size (N_e) is large, but evolve via genetic drift when N_e is small. By increasing N_e , recombination enhances the power of selection and minimizes genetic drift. Furthermore, the adaptive consequences of recombination may be extreme in yeast since most genes in the yeast genome can be defined by weak purifying selection ($\sim 75\%$ of genes are nonessential; $\sim 50\%$ have one PPI; TC & LLK unpub.). Such genes also tend to reside in genomic regions with relatively low recombination frequencies (see Additional file 1, Fig. S4). The correlations revealed by the data are particularly striking when one considers the method by which the genes are partitioned. Functional genomic data permits classification of genes according to their relative opportunities for slightly deleterious evolution. However, multiple types of substitutions (i.e. slightly deleterious, neutral, and beneficial) are likely to contribute to each gene's total genetic divergence between species. This plurality should dampen patterns predicted by any single processes, and will cause the conclusions presented here to be conservative.

Why does highly-expressed gene divergence show no correlation with recombination rate? There are two major possibilities. If genes under stronger selection tend to experience fewer beneficial mutations, their overall divergence rate might be relatively unaffected by local recombination. This might occur because strongly-selected genes are closer to perfection than weakly-selected genes and therefore have less opportunity for improvement, or because tradeoffs via pleiotropy (as indicated by high PPI [31]) limit the opportunity for beneficial mutations [38-40]. Secondly, selection for beneficial mutations might be very strong. The adaptive impact of varying recombination rate (and thus N_e) is expected to decrease with the strength of selection (i.e. s ; [3,41]). If beneficial mutations tend to be strongly advantageous, they will tend to become fixed in low or in high recombinational environments.

Conclusion

This study shows that recombination reduces evolutionary divergence in genes under relatively weak purifying selection (e.g. low-expressed, nonessential, few PPI), and at best, marginally increases divergence in genes under

strong purifying selection (e.g. highly expressed, essential, many PPI). This pattern suggests that enhanced purifying selection is a primary long-term benefit of recombination in nature. The efficient removal of deleterious mutations might increase the competitive ability of sexual species and contribute to the observed ubiquity of sexual reproduction in eukaryotes [6,42].

While this interpretation is appealing and supported by both theory and data from other taxa [e.g. [12,13]], it should be noted that inferences about the processes driving nucleotide divergence between species are tentative and reflect a major limitation of molecular divergence data. Future studies using entire-genome polymorphism and divergence data can add resolution by estimating the proportion of adaptive substitutions per gene [see [23]]. Such estimates, combined with inferences about the slightly deleterious substitution rate (derived from expression, essentiality and PPI data), will permit a much improved analysis of the benefits of recombination.

Methods

Data

Publicly available polymorphism data was obtained via [43] and [44]. Genes with at least four samples from *S. cerevisiae* and at least one polymorphic site were included in the analysis, resulting in a dataset of 35 genes (lower 25% expression, $n = 11$; lower 50% expression, $n = 12$; upper 50% expression, $n = 23$; upper 25% expression, $n = 17$; essential genes, $n = 7$; nonessential genes, $n = 28$), comprising 34443 nonsynonymous, and 9975 synonymous nucleotide sites. The mean number of samples per gene was $\bar{n} = 22$. Orthologous sequences from *S. paradoxus* were obtained via BLAST search at [45].

Per gene recombination rates for *S. cerevisiae* are those reported by Gerton et al. [8] [data available at [46]]. These estimates refer to recombination rates per sexual generation. The total, per generation recombination rate during the evolutionary history of each gene is the product of the rate under sexual reproduction (R_{sex}) and the frequency of outcrossing (O_c); $R_{TOT} = R_{sex}O_c$ (modified from [47]). Because the exact value of O_c will be the same for all genes in the genome, R_{sex} per gene i , relative to R_{sex} for other genes, will be the same as R_{TOT} per gene i , relative to R_{TOT} for other genes

$$\left(i.e. \frac{R_{TOT}(i)}{R_{TOT}(\max)} = \frac{R_{sex}(i)O_c}{R_{sex}(\max)O_c} = \frac{R_{sex}(i)}{R_{sex}(\max)} \right), \quad \text{and} \quad R_{sex}$$

should accurately capture relative rates of total recombination for each gene. Furthermore, because $0 < O_c < 1$ [47], the critical transition between $R_{TOT} \approx 0$ to $R_{TOT} > 0$, pre-

dicted to most strongly impact the efficacy of selection [48], will be represented with yeast genes.

Protein divergence data for *S. cerevisiae* and *S. paradoxus* orthologous genes were kindly provided by D. Allan Drummond (see [24] for details). Gene expression values for *S. cerevisiae* were calculated from seven time period estimates during the diauxic shift by DeRisi et al. [49] [data available at [50]]. Average and maximum expression levels for the seven time periods were calculated using the method of Kliman et al. [35]. Results for average gene expression (E_{avg}) across the time periods are presented here; results do not differ when maximum expression values (E_{max}) are used. Essential genes were identified through the GeneMerge database [51]. Gene length, dispensability, protein-protein interactions and genome map positions were obtained from the *Saccharomyces* Genome Database. Genes with no known interaction partner were excluded from analyses involving PPI as a variable. Space between genes (SBG) was calculated by the method of Hey & Kliman [9]. Recombination, expression, length, SBG and divergence estimates were available for 4786 genes in total, including essential genes with two or more PPI ($n = 329$), essential genes with 1 PPI ($n = 195$), nonessential genes with two or more PPI ($n = 762$), and nonessential genes with 1 PPI ($n = 835$). Our dataset is available upon request.

Analysis

Population samples for each gene were aligned with ClustalW [52], available online, and manually adjusted. P_n , P_s , D_n , and D_s values were calculated with DnaSP, Version 4.10 [53]. Watterson's estimate of silent nucleotide diversity (θ) was calculated by hand (as described in [54]). The complete polymorphism dataset is provided in Additional file 1.

Genes were classified into low and high expression level categories, based on quantile partitions. These correspond with ranges of lower 25%: 2.25 to 3.15 ($n = 1197$); lower 50%: 2.25 to 3.31 ($n = 2393$); upper 50%: 3.31 to 3.53 ($n = 2393$); and upper 25%: 3.53 to 4.54 ($n = 1197$) log mRNA abundance.

All divergence estimates were \log_{10} transformed to facilitate linear comparisons, which are presented (values of $dN = 0$ were converted to $dN = 0.0001$ prior to log transformation); the results are robust and also obtained with nonparametric comparisons (TC & LLK unpub.). Partial correlation analysis was used to compare recombination rate with dN (the rate of nonsynonymous substitutions); the same results were obtained for the comparison between recombination rate and dN/dS . The partial r statistic reported here reflects the association between recombination rate and protein divergence after associa-

tions between gene expression, gene length, and SBG were removed. These factors are known to influence patterns of protein evolution [[25]; TC & LLK unpub.], are all correlated with one another (i.e. recombination is positively correlated with expression and gene density, but negatively correlated with length), and can therefore give rise to spurious correlations between the variables of interest. All statistical analyses were carried out with JMP (SAS Institute). Statistical comparisons between r for different gene categories were carried out with software available online <http://department.obg.cuhk.edu.hk/researchsupport/Correlation.asp>. Bonferroni corrections for multiple comparisons ($\alpha/5$; because of the 5 categories explored in Figure 2) were used to adjust P values of statistical significance.

Authors' contributions

TC and LLK participated in designing of the study, analyzing the data and writing the manuscript. Both authors read and approved the final manuscript.

Additional material

Additional file 1

(Connallon&Kowles Supplementary Data): includes the entire polymorphism dataset, four supplementary figures (Figs. S1, S2, S3, S4), and additional details of the polymorphism analysis.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2148-7-235-S1.pdf>]

Acknowledgements

We are grateful to D. Allan Drummond for providing genetic divergence estimates, and to Alex Kondrashov, Daven Presgraves, Akane Uesugi, Jianzhi Zhang, and two anonymous reviewers for valuable comments on an earlier version of the manuscript. This research was funded by a grant from the University of Michigan Rackham Graduate School and a training grant from the National Human Genome Research Institute (T32 HG00040) that supported T. C., and by a National Science Foundation grant (DEB-04-47224) to L. L. K.

References

1. Fisher RA: **The distribution of gene ratios for rare mutations.** *Proc Roy Soc Edinburgh* 1930, **50**:204-219.
2. Wright S: **Evolution in Mendelian populations.** *Genetics* 1931, **16**:97-159.
3. Kimura M: **On the probability of fixation of mutant genes in a population.** *Genetics* 1962, **47**:713-719.
4. Hill WG, Robertson A: **The effect of linkage on limits to artificial selection.** *Genet Res* 1966, **8**:269-294.
5. Felsenstein J: **The evolutionary advantage of recombination.** *Genetics* 1974, **78**:737-756.
6. Rice WR: **Experimental tests of the adaptive significance of sexual recombination.** *Nat Rev Genet* 2002, **3**:241-251.
7. Otto SP, Lenormand T: **Resolving the paradox of sex and recombination.** *Nat Rev Genet* 2002, **3**:252-261.
8. Gerton JL, DeRisi J, Schroff R, Lichten M, Brown PO, Petes TD: **Global mapping of meiotic recombination hotspots and cold-**

- pots in the yeast *Saccharomyces cerevisiae*.** *Proc Natl Acad Sci USA* 2000, **97**:11383-11390.
9. Hey J, Kliman RM: **Interactions between natural selection, recombination and gene density in the genes of *Drosophila*.** *Genetics* 2002, **160**:595-608.
 10. Jensen-Seaman MI, Furey TS, Payseur BA, Lu YT, Roskin KM, Chen CF, Thomas MA, Haussler D, Jacob HJ: **Comparative recombination rates in the rat, mouse, and human genomes.** *Genome Res* 2004, **14**:528-538.
 11. Mezard C: **Meiotic recombination hotspots in plants.** *Biochem Soc Trans* 2006, **34**:531-534.
 12. Barton NH: **Linkage and the limits to natural selection.** *Genetics* 1995, **140**:821-841.
 13. Presgraves DC: **Recombination enhances protein adaptation in *Drosophila melanogaster*.** *Curr Biol* 2005, **15**:1651-1656.
 14. Pal C, Papp B, Hurst LD: **Does the recombination rate affect the efficiency of purifying selection? The yeast genome provides a partial answer.** *Mol Biol Evol* 2001, **18**:2323-2326.
 15. Betancourt AJ, Presgraves DC: **Linkage limits the power of natural selection in *Drosophila*.** *Proc Nat Acad Sci USA* 2002, **99**:13616-13620.
 16. Bachtrog D: **Protein evolution and codon usage bias on the neo-sex chromosomes of *Drosophila miranda*.** *Genetics* 2003, **165**(3):1221-1232.
 17. Marais G, Domazet-Loso T, Tautz D, Charlesworth B: **Correlated evolution of synonymous and nonsynonymous sites in *Drosophila*.** *J Mol Evol* 2004, **59**:771-779.
 18. Zhang Z, Parsch J: **Positive correlation between evolutionary rate and recombination rate in *Drosophila* genes with male-biased expression.** *Mol Biol Evol* 2005, **22**:1945-1947.
 19. Paland S, Lynch M: **Transitions to asexuality result in excess amino acid substitutions.** *Science* 2006, **311**:990-992.
 20. Birkby CV, Walsh JB: **Effects of linkage on rates of molecular evolution.** *Proc Natl Acad Sci USA* 1988, **85**:6414-6418.
 21. Charlesworth B: **The effect of background selection against deleterious mutations on weakly selected, linked variants.** *Genet Res* 1994, **63**:213-227.
 22. Peck JR: **A ruby in the rubbish: beneficial mutations, deleterious mutations and the evolution of sex.** *Genetics* 1994, **137**:597-606.
 23. Eyre-Walker A: **The genomic rate of adaptive evolution.** *Trends Ecol Evol* 2006, **21**:569-575.
 24. Drummond DA, Bloom JD, Adami C, Wilke CO, Arnold FH: **Why highly expressed proteins evolve slowly.** *Proc Natl Acad Sci USA* 2005, **102**:14338-14343.
 25. Pal C, Papp B, Lercher MJ: **An integrated view of protein evolution.** *Nat Rev Genet* 2006, **7**:337-348.
 26. McInerney JO: **The causes of protein evolutionary rate variation.** *Trends Ecol Evol* 2006, **21**:230-232.
 27. Rocha EPC: **The quest for the universals of protein evolution.** *Trends Genet* 2006, **22**:412-416.
 28. Hirsh AE, Fraser HB: **Protein dispensability and rate of evolution.** *Nature* 2001, **411**:1046-1049.
 29. Wall DP, Hirsh AE, Fraser HB, Kumm J, Giaever G, Eisen MB, Feldman MW: **Functional genomic analysis of the rates of protein evolution.** *Proc Natl Acad Sci USA* 2005, **102**:5483-5488.
 30. Zhang J, He X: **Significant impact of protein dispensability on the instantaneous rate of protein evolution.** *Mol Biol Evol* 2005, **22**:1147-1155.
 31. He X, Zhang J: **Toward a molecular understanding of pleiotropy.** *Genetics* 2006, **173**:1885-1891.
 32. Kimura M: *The Neutral Theory of Molecular Evolution* Cambridge, Cambridge University Press; 1983.
 33. McDonald JH, Kreitman M: **Adaptive protein evolution at the *Adh* locus in *Drosophila*.** *Nature* 1991, **351**:652-654.
 34. Bierne N, Eyre-Walker A: **The genomic rate of adaptive amino acid substitution in *Drosophila*.** *Mol Biol Evol* 2004, **21**:1350-1360.
 35. Kliman RM, Irving N, Santiago M: **Selection conflicts, gene expression, and codon usage trends in yeast.** *J Mol Evol* 2003, **57**:98-109.
 36. Datta A, Jinksrobertson S: **Association of increased spontaneous mutation rates with high levels of transcription in yeast.** *Science* 1995, **268**:1616-1619.
 37. Holbeck SL, Strathern JN: **A role for REV3 in mutagenesis during double stranded break repair in *Saccharomyces cerevisiae*.** *Genetics* 1997, **147**:1017-1024.
 38. Fisher RA: *The genetical theory of natural selection* Oxford: Clarendon Press; 1930.
 39. Orr HA: **The population genetics of adaptation: the distribution of factors fixed during adaptive evolution.** *Evolution* 1998, **52**:935-949.
 40. Otto SP: **Two steps forward, one step back: the pleiotropic effects of favoured alleles.** *Proc Roy Soc Lond B* 2004, **271**:705-714.
 41. Hedrick PW: *Genetics of populations* Boston: Science Books International; 1983.
 42. Bell G: *The masterpiece of nature* Berkeley: University of California Press; 1982.
 43. Genbank [<http://www.ncbi.nlm.nih.gov/sites/entrez?db=Popset>]
 44. Polymorphix Database [<http://pbil.univ-lyon1.fr/polymorphix/>]
 45. The *Saccharomyces* Genome Database [<http://www.yeastgenome.org/>]
 46. Data from Gerton et al. 2000 [<http://derisilab14.ucsf.edu/hotspots/>]
 47. Ruderfer DM, Pratt SC, Seidel HS, Kruglyak L: **Population genomic analysis of outcrossing and recombination in yeast.** *Nat Genet* 2006, **38**:1077-1081.
 48. McVean GAT, Charlesworth B: **The effects of Hill-Robertson interference between weakly selected mutations on patterns of molecular evolution and variation.** *Genetics* 2000, **155**:929-944.
 49. DeRisi JL, Iyer VR, Brown PO: **Exploring the metabolic and genetic control of gene expression on a genomic scale.** *Science* 1997, **278**:680-686.
 50. Data from DeRisi et al. 1997 [<http://cmgm.stanford.edu/pbrown/explore/array.txt>]
 51. Castillo-Davis CI, Hartl DL: **GeneMerge – post-genomic analysis, data mining, and hypothesis testing.** *Bioinformatics* 2003, **19**:891-892 [<http://www.oeb.harvard.edu/hartl/lab/publications/GeneMerge/GeneMerge.html>].
 52. ClustalW online [<http://www.ch.embnet.org/software/ClustalW-XXL.html>]
 53. Rozas J, Sánchez-DelBarrio JC, Messegyer X, Rozas R: **DnaSP, DNA polymorphism analyses by the coalescent and other methods.** *Bioinformatics* 2003, **19**:2496-2497.
 54. Hein J, Schierup MH, Wiuf C: *Gene Genealogies, Variation and Evolution: A Primer in Coalescent Theory* New York: Oxford University Press; 2005.
 55. Wloch DM, Szafranec K, Borts RH, Korona R: **Direct estimate of the mutation rate and the distribution of fitness effects in the yeast *Saccharomyces cerevisiae*.** *Genetics* 2001, **159**:441-452.
 56. Keightley PD, Lynch M: **Toward a realistic model of mutations affecting fitness.** *Evolution* 2003, **57**:683-685.
 57. Joseph SB, Hall DW: **Spontaneous mutations in diploid *Saccharomyces cerevisiae*: More beneficial than expected.** *Genetics* 2004, **168**:1817-1825.
 58. Shaw FH, Geyer CJ, Shaw RG: **A comprehensive model of mutations affecting fitness and inferences for *Arabidopsis thaliana*.** *Evolution* 2002, **56**:453-463.
 59. Shaw RG, Shaw FH, Geyer C: **What fraction of mutations reduces fitness? A reply to Keightley and Lynch.** *Evolution* 2003, **57**:686-689.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

