Distinguish between Duplication of Essential Genes and Duplication of Dispensable Genes

Xun Gu¹

¹ Department of Genetics, Development and Cell Biology, Iowa State University, Ames, IA, USA.

**Corresponding author:* Xun Gu (xqu@iastate.edu)

Keywords: genetic robustness, gene duplication, essentiality, genetic buffering

Abstract

When a dispensable gene is duplicated (ancestral dispensability), genetic buffering and duplicate compensation together maintain the gene dispensability, whereas duplicate compensation is the only mechanism when an essential gene is duplicated (ancestral essentiality). To explore the distinct pattern of genetic robustness between these evolutionary scenarios, we formulated a probabilistic model with some biologically reasonable assumptions for analyzing a set of duplicate pairs with three possible states: double-dispensable (DD), semi-dispensable (one dispensable one essential, DE) or double-essential (EE). A computational pipeline is then developed to predict the distribution of three states (DD, DE and EE) conditional of ancestral dispensability or essentiality, respectively. This model was applied to yeast duplicate pairs from a whole-genome duplication, revealing that the process of essentiality of those duplicated from essential genes could be significantly higher than that of those duplicated from dispensable genes. We thus proposed a hypothesis that the process of sub-functionalization may be faster than neo-functionalization. Our analysis may provide some new insights about the role of duplicate compensation on genetic robustness.

Introduction

The role of functional compensation by duplicate genes has been examined in diverse organisms by comparing the proportion (P_E) of essential genes in duplicates to P_E in singletons (Conant and Wagner 2004; Gu, et al. 2003; Hanada, et al. 2009; Liang and Li 2007; Liao and Zhang 2007; Wagner 2000). If duplicates play a significant role in functional compensation, the P_E for duplicates should be lower than that of singletons. While this is indeed the case in yeasts, worms, and plants (Conant and Wagner 2004; Gu, et al. 2003; Hanada, et al. 2009; Hanada, et al. 2011; Qian, et al. 2010), no significant difference in P_E was found between mouse single-copy and duplicate genes (Liang and Li 2007; Liao and Zhang 2007). A number of explanations were proposed. For instance, Su and Gu (2008) noticed that most mouse knockout duplicates tend to be ancient, subject to the loss of duplicate compensation by the mechanism of sub-functionalization. Meanwhile, some studies showed that functional and protein connectivity bias between essential and dispensable duplicate genes may be the cause (Makino, et al. 2009; Liang and Li 2009).

The pattern of duplicate compensation is complex (Diss, et al. 2017; Hahn 2009; Keane, et al. 2014; Szklarczyk, et al. 2008; Teufel, et al. 2018). When an essential gene is duplicated (termed *ancestral essentiality*), duplicate compensation is the only mechanism to keep two resulting copies dispensable. On the other hand, when a dispensable gene is duplicated (termed *ancestral dispensability*), the ancient genetic buffering and duplicate compensation together keep both duplicate copies dispensable. However, all P_E -related analyses in the literature did not distinguish between these two possibilities.

This paper will address this issue as follows. We first develop a statistical method to analyze a set of duplicate pairs with three possible states: double-dispensable (DD), semi-dispensable (one dispensable one essential, DE) or double-essential (EE). With the help of some biologically reasonable assumptions, a probabilistic model is formulated, which takes the duplication of an essential gene or a dispensable gene into account. A computational pipeline is then developed to

predict the distribution of three states (DD, DE and EE) conditional of ancestral dispensability or essentiality, respectively. Exemplified by the yeast duplicate pairs from the whole genome duplication (WGD), the new model compared the proportion of essential genes between those duplicated from essential genes and those dispensable genes, and discussed new insights for the evolutionary pattern of genetic robustness after gene duplication.

Results and Discussion

Genetic robustness between duplicate genes

A gene is called ‘*essential*’ (denoted by d^-) if the single-gene deletion phenotype is severe or lethal, or ‘*dispensable*’ (denoted by d^+) if its deletion phenotype is normal or nearly-normal (Hsiao and Vitkup 2008; Ihmels, et al. 2007; Su and Gu 2008). Consider two paralogous genes (A and B) duplicated from a common ancestor (O) t time units ago. There are four combined states, denoted by (d_A, d_B) , representing double-dispensable (d^+, d^+), semi-dispensable (d^+, d^-) or (d^-, d^+), or double-essential (d^-, d^-), respectively.

First consider $P_t(d_A, d_B)$, the probability of any joint states (d_A, d_B) at time t . To derive $P_t(d_A, d_B)$, one should distinguish between the duplication of an essential genes (ancestral essentiality, denoted by O^-) and the duplication of a dispensable gene (ancestral dispensability, denoted by O^+). This idea can be statistically formulated as follows. Let $Q_t(d_A, d_B|O^-)$ be the probability of being (d_A, d_B) after t time units since gene duplication, conditional of the ancestral essentiality (O^-), and $Q_t(d_A, d_B|O^+)$ be the probability conditional of the ancestral dispensability (O^+). Since the ancestral state (dispensable or essential) for a duplicate pair is usually unknown, a mixture model is implemented as follows. Let $R_0 = P(O^+)$ be the probability of a gene pair duplicated from a dispensable gene, and $1 - R_0 = P(O^-)$ be that from an essential gene. Together, one can write

$$P_t(d_A, d_B) = (1 - R_0)Q_t(d_A, d_B|O^-) + R_0Q_t(d_A, d_B|O^+) \quad (1)$$

where $(d_A, d_B) = (d^+, d^+), (d^+, d^-), (d^-, d^+)$ or (d^-, d^-) , respectively.

Suppose we have a set (N) of duplicate pairs; all $2N$ genes have single-gene deletion phenotypes (dispensable or essential). Three types of duplicate pairs are considered, that is, DD for (d^+, d^+) , DE for (d^+, d^-) or (d^-, d^+) , and EE for (d^-, d^-) , and their frequencies are denoted by f_{DD} , f_{DE} and f_{EE} , respectively. An intriguing question we wish to address is how we can estimate $Q_t(d_A, d_B|O^-)$ and $Q_t(d_A, d_B|O^+)$ from observed frequencies f_{DD} , f_{DE} and f_{EE} . As shown below, to this end we develop a three-step approach: first develop simple models for; second estimate or specify model parameters from the observations; and third calculate $Q_t(d_A, d_B|O^-)$ and $Q_t(d_A, d_B|O^+)$.

Sub-functionalization model of $Q_t(d_A, d_B|O^-)$

When an essential gene was duplicated, the process of *sub-functionalization* has been thought to be the major evolutionary mechanism for duplicate preservation (Innan and Kondrashov 2010; Prince and Pickett 2002; Stark, et al. 2017). As a result, both duplicate copies can be preserved without the need to invoke positive selection. Here we develop a simple model for $Q_t(d_A, d_B|O^-)$ that is calculated from the data. Suppose a duplicate pair has m independent components, each of which is either ‘active’ (denoted by ‘1’) or ‘inactive’ (denoted by ‘0’). Let U_{11} be the probability of a component being active in both genes; U_{01} (or U_{10}) is that of being inactive in gene A but active in gene B (or active in A but inactive in B); and U_{00} be the probability of a component being inactive in both genes. Without loss of generality, one can assume $U_{01}=U_{10}$. Further, we claim $U_{00}=0$ according to the *not-all-inactive constraint*, i.e., each component is functionally active at least in one duplicate copy. Together, U_{ij} s can be concisely written as $U_{11}=1-2U$, $U_{10}=U_{01}=U$ and $U_{00}=0$, respectively. In other words, with a probability of $2U$, a functional component is active in one duplicate but inactive in another one, and with a probability of $1-2U$, a component is active in both duplicates.

Under the assumption that functional components of a gene are statistically independent and identical, one can derive $Q_t(d_A, d_B/O^-)$ in term of the single component parameter (U) and the number (m) of functional components, that is,

$$\begin{aligned} Q_t(d^+, d^+|O^-) &= (1 - 2U)^m \\ Q_t(d^+, d^-|O^-) &= (1 - U)^m - (1 - 2U)^m \\ Q_t(d^-, d^+|O^-) &= (1 - U)^m - (1 - 2U)^m \\ Q_t(d^-, d^-|O^-) &= 1 - 2(1 - U)^m + (1 - 2U)^m \end{aligned} \quad (2)$$

The rationale of Eq.(2) is as follows. Under the m -component model ($m>1$), two duplicate copies remain both dispensable only when each component is active in both duplicates (with a probability of $1-2U$), which directly leads to the derivation of $Q_t(d^+, d^+/O^-)$. Next we consider the (marginal) probability of dispensability (d^+) conditional of the ancestral essentiality (O^-), denoted by $Q_t(d^+/O^-)$. It appears that $Q_t(d^+/O^-)=(1-U)^m$ because the probability of a component for being active in one duplicate is given by $(1-U)$. Since $Q_t(d^+/O^-)=Q_t(d^+, d^+/O^-)+Q_t(d^+, d^-/O^-)$, it is straightforward to obtain the second and third equations of Eq.(2). The last equation of Eq.(2) is derived by the sum of probabilities to be one.

Rare neo-functionalization after duplication of dispensable gene

When a dispensable gene was duplicated, gene dispensability can be maintained through ancient genetic buffering and duplicate compensation. As a result, sub-functionalization becomes ineffective because the process of complementation between functional components (Force, et al. 1999) is difficult to achieve. While the neo-functionalization has been suggested for duplicate preservation (Chen, et al. 2010; Chen, et al. 2012), it is unlikely that both copies acquire new functions simultaneously. In this sense one can assume that

$$Q_t(d^-, d^-|O^+) = 0 \quad (3)$$

Intuitively speaking, this assumption holds well except for very ancient duplicates by independent new function acquiring that may be unrelated to the duplication event.

Eq.(3) implies that any double-essential duplicate pairs (d^- , d^-) were virtually from the duplication of an essential gene (O^-).

Estimation of conditional probabilities

Based on Eqs.(2) and (3), we implement a computational procedure to calculate $Q_t(d_A, d_B|O^-)$ and $Q_t(d_A, d_B|O^+)$, respectively.

- (i) The proportion of dispensable single-copy genes is used as a proxy of R_0 .
- (ii) Applying $Q_t(d^-, d^-|O^+)=0$ to Eq.(1), leading to $P_t(d^-, d^-)=(1-R_0) Q_t(d^-, d^-|O^-)$.

Replacing $P_t(d^-, d^-)$ with its observed frequency f_{EE} , we obtain

$$Q_t(d^-, d^-|O^-)=f_{EE}/(1-R_0) \quad (4)$$

- (iii) The parameter U can be estimated by replacing the last equation of Eq.(2) of $Q_t(d^-, d^-|O^-)$ by $f_{EE}/(1-R_0)$, that is,

$$1 - 2(1 - \hat{U})^m + (1 - 2\hat{U})^m = \frac{f_{EE}}{1 - R_0} \quad (5)$$

where m , the number of functional components, is treated as a known integer, i.e., $m=2, 3, \dots$

- (iv) When U is estimated by Eq.(5), the estimates of joint probabilities conditional of the ancestral essentiality (O^-), $Q(d_A, d_B|O^-)$, can be obtained by replacing U with its estimate in Eq.(2). Eqs.(4) and (5) indicate that f_{EE} , the frequency of double-essential duplicate pairs, determines the conditional probabilities of (d_A , d_B) after duplication of an essential gene (O^-), provided that R_0 and m are treated as known constants.
- (v) It is straightforward to estimate $Q(d^+, d^+|O^+)$ by equating $P_t(d^+, d^+)$ with f_{DD} in Eq.(1) in the case of $d_A=d^+$ and $d_B=d^+$, and replacing $Q(d^+, d^+|O^-)$ by its estimate. They are, respectively, given by

$$\begin{aligned}\hat{Q}_t(d^+, d^+ | O^+) &= \frac{f_{DD} - (1 - R_0)(1 - 2\hat{U})^m}{R_0} \\ 2\hat{Q}_t(d^+, d^- | O^+) &= 1 - \frac{f_{DD} - (1 - R_0)(1 - 2\hat{U})^m}{R_0} \\ \hat{Q}_t(d^-, d^- | O^+) &= 0\end{aligned}\tag{6}$$

Estimation of the sampling variances

The statistical property of those estimated conditional probabilities can be evaluated by calculating their approximate sampling variances. One approach invokes the bootstrapping algorithm. The other approach is the delta-method under a multinomial distribution of f_{DD} , f_{DE} and f_{EE} , as briefly introduced below. Denote the function g by

$$g = 1 - 2(1 - \hat{U})^m + (1 - 2\hat{U})^m\tag{7}$$

One can calculate the partial derivative of g by

$$\frac{\partial g}{\partial U} = 2m(1 - U)^{m-1} - 2m(1 - 2U)^m\tag{8}$$

Hence, applying the delta-method in Eq.(5), we obtain

$$\frac{Var(f_{EE})}{(1 - R_0)^2} \approx \left(\frac{\partial g}{\partial U} \right)^2 Var(\hat{U})\tag{9}$$

resulting in

$$Var(\hat{U}) \approx (2m)^{-2} \left[(1 - U)^{m-1} - (1 - 2U)^{m-1} \right]^{-2} \frac{Var(f_{EE})}{(1 - R_0)^2}\tag{10}$$

After some basic yet tedious algebras, the analytical forms for the sampling variances of those estimated conditional probabilities can be approximately derived in terms of the observed proportions f_{DD} , f_{DE} and f_{EE} , as well as N (the number of duplicate pairs).

Case-study: yeast duplicate pairs from whole genome duplications

We applied the new method to analyze 325 yeast duplicate pairs from the yeast WGD (whole-genome duplication) about 100 million years ago. The fitness dataset of single gene-deletions for yeast genes were from Gu et al. (2003). The

fitness value f_i is measured by the growth rate of the strain with gene i deleted relative to the average growth rate of wild strains under five growth media: YPD, YPDGE, YPE, YPG, and YPL. Similar to Gu et al. (2003), we classified these duplicate genes under study into four groups: The normal (N)-group is defined by $f_{min} > 0.95$, where f_{min} is the smallest f -value for all five growth conditions, i.e., the deletion has a weak or no fitness effect in all conditions. The M -group is defined by $0.8 < f_{min} < 0.95$ for the moderate effect, and the S -group by $0.5 < f_{min} < 0.8$ for a strong effect. Finally, the L -group is defined by $f_{min} < 0.5$ that includes the lethal effect ($f=0$). Tentatively, a gene is classified as d^+ if it belongs to the N or M -group, or d^- otherwise. Under this cutoff, the observed frequencies for patterns DD, DE and EE are $f_{DD}=0.803 \pm 0.022$, $f_{DE}=0.179 \pm 0.021$ and $f_{EE}=0.018 \pm 0.0074$, respectively. In addition, the proportion of dispensable single-copy genes (0.605) is used as a proxy of R_0 .

We first examine how the number (m) of functional components may affect our analysis. Note that single functional component ($m=1$) is not allowed. Hughes and Liberles (2007) suggested that between $m=2$ and $m=12$ regulatory regions would be biologically realistic. Stark et al. (2017) suggested that it is unlikely that a gene would have in excess of $m=20$ functional components. Within the range of $m=2-10$, the effects of m on our analysis of yeast duplicate pairs are shown in **Fig.1** (panel A or B). Overall it reveals little difference among cases of $m=3$ or more; only marginal differences appear when $m=2$, and all estimates are virtually the same between $m=20$ and $m=\infty$. We therefore conclude that the effect of variable m is usually negligible.

The proportion of essential genes for the yeast duplicate pairs can be calculated by $P_E = (f_{EE} + f_{DE})/2 = 0.103 \pm 0.012$, which is significantly lower than the proportion of essential genes in single-copy genes ($P_E = 0.395$); $p\text{-value} < 10^{-6}$. Based on the newly-developed analysis, one can further estimate P_E for genes duplicated from essential genes, which is given by $P_E(O^-) = Q_i(d^-, d^-/O^-) + Q_i(d^-, d^-/O^+) = 0.187 \pm 0.035$ ($m=6$), whereas P_E for genes duplicated from dispensable genes, which is given by $P_E(O^+) = Q_i(d^-, d^-/O^+) + Q_i(d^-, d^-/O^-) = 0.043 \pm 0.018$ ($m=6$). Bootstrapping analysis shows that $P_E(O^-)$ is significantly larger than by $P_E(O^+)$; $p\text{-value} < 0.001$. Therefore, a rapid increase

of essentiality, or loss of dispensability may occur after the duplication of essential genes, approximately four-fold as fast as the process after the duplication of dispensable genes.

Model assumptions

We acknowledge that the new probabilistic model for the evolution of genetic robustness after gene duplication is inevitably oversimplified. For instance, after duplication of an essential gene, the model assumes that two duplicate copies evolve under sub-functionalization, neglecting the chance of neo-functionalization or other mechanisms. Moreover, each functional component is assumed to undergo sub-functionalization independently. On the other hand, after duplication of a dispensable gene, the interactions between ancestral genetic buffering, duplicate compensation and neo-functionalization remain largely unknown. It should be noticed that some attributes of genetic mechanisms have not been taken into accounts, such as the effect of dosage balance, later-stage functional divergence, as described in Innan and Kondrashov (2010). Our future study will focus on the development of a more realistic model for the evolution of gene duplicates.

Effects of the assumption $Q_i(d^-, d^+/O^+)=0$

A key assumption in our analyses is Eq.(3), that is, after duplication of a dispensable gene (O^+), the chance for both duplicate copies to be essential is negligible. While it is biologically intuitive, it may cause some bias, especially for some very ancient duplicate pairs. We thus conducted a computer simulation study to examine this effect by letting $Q(d^-, d^+/O^+)=q$, where q is a small positive value. Our preliminary simulation result shows that the estimation bias of those conditional probabilities is usually marginal, except for the case after an extremely long evolutionary span after gene duplication (not shown).

Concluding remarks

This paper described a probabilistic model to explore the pattern of genetic robustness after the gene duplication, which makes a distinction between two evolutionary scenarios: duplication of an essential gene and duplication of a dispensable gene. This model was applied to yeast duplicate pairs from a whole-genome duplication, revealing that the proportion of essential genes for those duplicated from essential genes is approximately four-fold as high as that for those duplicated from dispensable genes. While more genome-wide single-gene deletion (knockout) data for duplicate pairs are available in different species, it is anticipated that the new model can provide some new insights about the role of duplicate compensation on genetic robustness.

References

- Baryshnikova A, Costanzo M, Kim Y, Ding HM, Koh J, Toufighi K, Youn JY, Ou JW, San Luis BJ, Bandyopadhyay S, et al. 2010. Quantitative analysis of fitness and genetic interactions in yeast on a genome scale. *Nature Methods* 7: 1017-U1110.
- Chen S, Zhang YE, Long M 2010. New genes in *Drosophila* quickly become essential. *Science* 330: 1682-1685.
- Chen WH, Trachana K, Lercher MJ, Bork P 2012. Younger Genes Are Less Likely to Be Essential than Older Genes, and Duplicates Are Less Likely to Be Essential than Singletons of the Same Age. *Molecular Biology and Evolution* 29: 1703-1706.
- Conant GC, Wagner A 2004. Duplicate genes and robustness to transient gene knock-downs in *Caenorhabditis elegans*. *Proceedings of the Royal Society B-Biological Sciences* 271: 89-96.
- Costanzo M, Baryshnikova A, Bellay J, Kim Y, Spear ED, Sevier CS, Ding HM, Koh JLY, Toufighi K, Mostafavi S, et al. 2010. The Genetic Landscape of a Cell. *Science* 327: 425-431.
- de Visser JAGM, Hermisson J, Wagner GP, Meyers LA, Bagheri HC, Blanchard JL, Chao L, Cheverud JM, Elena SF, Fontana W, et al. 2003. Perspective: Evolution and detection of genetic robustness. *Evolution* 57: 1959-1972.
- Dean EJ, Davis JC, Davis RW, Petrov DA 2008. Pervasive and Persistent Redundancy among Duplicated Genes in Yeast. *Plos Genetics* 4.
- Diss G, Gagnon-Arsenault I, Dion-Cote AM, Vignaud H, Ascencio DI, Berger CM, Landry CR 2017. Gene duplication can impart fragility, not robustness, in the yeast protein interaction network. *Science* 355: 630-633.
- Dowell RD, Ryan O, Jansen A, Cheung D, Agarwala S, Danford T, Bernstein DA, Rolfe PA, Heisler LE, Chin B, et al. 2010. Genotype to Phenotype: A Complex Problem. *Science* 328: 469-469.
- Flatt T 2005. The evolutionary genetics of canalization. *Quarterly Review of Biology* 80: 287-316.
- Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J 1999. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* 151: 1531-1545.
- Gu ZL, Steinmetz LM, Gu X, Scharfe C, Davis RW, Li WH 2003. Role of duplicate genes in genetic robustness against null mutations. *Nature* 421: 63-66.
- Guan YF, Dunham MJ, Troyanskaya OG 2007. Functional analysis of gene duplications in *Saccharomyces cerevisiae*. *Genetics* 175: 933-943.
- Hahn MW 2009. Distinguishing Among Evolutionary Models for the Maintenance of Gene Duplicates. *Journal of Heredity* 100: 605-617.
- Hanada K, Kuromori T, Myouga F, Toyoda T, Li WH, Shinozaki K 2009. Evolutionary Persistence of Functional Compensation by Duplicate Genes in *Arabidopsis*. *Genome Biology and Evolution* 1: 409-414.
- Hanada K, Sawada Y, Kuromori T, Klausnitzer R, Saito K, Toyoda T, Shinozaki K, Li WH, Hirai MY 2011. Functional Compensation of Primary and Secondary Metabolites by Duplicate Genes in *Arabidopsis thaliana*. *Molecular Biology and Evolution* 28: 377-382.
- Harrison R, Papp B, Pal C, Oliver SG, Delneri D 2007. Plasticity of genetic interactions in metabolic

networks of yeast. *Proceedings of the National Academy of Sciences of the United States of America* 104: 2307-2312.

Hartman JL, Garvik B, Hartwell L 2001. Cell biology - Principles for the buffering of genetic variation. *Science* 291: 1001-1004.

Hillenmeyer ME, Fung E, Wildenhain J, Pierce SE, Hoon S, Lee W, Proctor M, St Onge RP, Tyers M, Koller D, et al. 2008. The chemical genomic portrait of yeast: Uncovering a phenotype for all genes. *Science* 320: 362-365.

Hsiao TL, Vitkup D 2008. Role of duplicate genes in robustness against deleterious human mutations. *Plos Genetics* 4: e1000014.

Hughes T, Liberles DA 2007. The pattern of evolution of smaller-scale gene duplicates in mammalian genomes is more consistent with neo-than subfunctionalisation. *Journal of Molecular Evolution* 65: 574-588.

Ihmels J, Collins SR, Schuldiner M, Krogan NJ, Weissman JS 2007. Backup without redundancy: genetic interactions reveal the cost of duplicate gene loss. *Molecular Systems Biology* 3.

Innan H, Kondrashov F 2010. The evolution of gene duplications: classifying and distinguishing between models. *Nature Reviews Genetics* 11: 97-108.

Kabir M, Barradas A, Tzotzos GT, Hentges KE, Doig AJ 2017. Properties of genes essential for mouse development. *Plos One* 12.

Kafri R, Bar-Even A, Pilpel Y 2005. Transcription control reprogramming in genetic backup circuits. *Nature Genetics* 37: 295-299.

Kafri R, Springer M, Pilpel Y 2009. Genetic Redundancy: New Tricks for Old Genes. *Cell* 136: 389-392.

Kamath RS, Fraser AG, Dong Y, Poulin G, Durbin R, Gotta M, Kanapin A, Le Bot N, Moreno S, Sohrmann M, et al. 2003. Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. *Nature* 421: 231-237.

Keane OM, Toft C, Carretero-Paulet L, Jones GW, Fares MA 2014. Preservation of genetic and regulatory robustness in ancient gene duplicates of *Saccharomyces cerevisiae*. *Genome Research* 24: 1830-1841.

Le Meur N, Gentleman R 2008. Modeling synthetic lethality. *Genome Biology* 9.

Lee I, Lehner B, Vavouri T, Shin J, Fraser AG, Marcotte EM 2010. Predicting genetic modifier loci using functional gene networks. *Genome Research* 20: 1143-1153.

Li JJ, Yuan ZE, Zhang ZL 2010. The Cellular Robustness by Genetic Redundancy in Budding Yeast. *Plos Genetics* 6.

Liang H, Li WH 2009. Functional compensation by duplicated genes in mouse. *Trends in Genetics* 25: 441-442.

Liang H, Li WH 2007. Gene essentiality, gene duplicability and protein connectivity in human and mouse. *Trends in Genetics* 23: 375-378.

Liao BY, Zhang JZ 2007. Mouse duplicate genes are as essential as singletons. *Trends in Genetics* 23:

378-381.

Liu GB, Zou YY, Cheng QQ, Zeng YW, Gu X, Su ZX 2014. Age distribution patterns of human gene families: divergent for Gene Ontology categories and concordant between different subcellular localizations. *Molecular Genetics and Genomics* 289: 137-147.

Makino T, Hokamp K, McLysaght A 2009. The complex relationship of gene duplication and essentiality. *Trends in Genetics* 25: 152-155.

Makino T, McLysaght A 2010. Ohnologs in the human genome are dosage balanced and frequently associated with disease. *Proceedings of the National Academy of Sciences of the United States of America* 107: 9270-9274.

Muller FL, Colla S, Aquilanti E, Manzo VE, Genovese G, Lee J, Eisenson D, Narurkar R, Deng PN, Nezi L, et al. 2012. Passenger deletions generate therapeutic vulnerabilities in cancer. *Nature* 488: 337-+.

Musso G, Costanzo M, Huangfu MQ, Smith AM, Paw J, Luis BJS, Boone C, Giaever G, Nislow C, Emili A, et al. 2008. The extensive and condition-dependent nature of epistasis among whole-genome duplicates in yeast. *Genome Research* 18: 1092-1099.

Nowak MA, Boerlijst MC, Cooke J, Smith JM 1997. Evolution of genetic redundancy. *Nature* 388: 167-171.

Papp B, Pal C, Hurst LD 2004. Metabolic network analysis of the causes and evolution of enzyme dispensability in yeast. *Nature* 429: 661-664.

Plata G, Vitkup D 2014. Genetic robustness and functional evolution of gene duplicates. *Nucleic Acids Research* 42: 2405-2414.

Prince VE, Pickett FB 2002. Splitting pairs: The diverging fates of duplicated genes. *Nature Reviews Genetics* 3: 827-837.

Qian WF, Liao BY, Chang AYW, Zhang JZ 2010. Maintenance of duplicate genes and their functional redundancy by reduced expression. *Trends in Genetics* 26: 425-430.

Ramani AK, Chuluunbaatar T, Verster AJ, Na H, Vu V, Pelte N, Wannissorn N, Jiao A, Fraser AG 2012. The Majority of Animal Genes Are Required for Wild-Type Fitness. *Cell* 148: 792-802.

Stark TL, Liberles DA, Holland BR, O'Reilly MM 2017. Analysis of a mechanistic Markov model for gene duplicates evolving under subfunctionalization. *Bmc Evolutionary Biology* 17: 38.

Su ZX, Gu X 2008. Predicting the Proportion of Essential Genes in Mouse Duplicates Based on Biased Mouse Knockout Genes. *Journal of Molecular Evolution* 67: 705-709.

Su ZX, Wang JQ, Gu X 2014. Effect of Duplicate Genes on Mouse Genetic Robustness: An Update. *Biomed Research International*.

Szklarczyk R, Huynen MA, Snel B 2008. Complex fate of paralogs. *Bmc Evolutionary Biology* 8.

Teufel AI, Johnson MM, Laurent JM, Kachroo AH, Marcotte EM, Wilke CO 2018. The many nuanced evolutionary consequences of duplicated genes (Retraction Article, vol 35, msy210, 2018) (Retraction of Vol 35, art no MSY210, 2018). *Molecular Biology and Evolution* 35: E1-E1.

Tischler J, Lehner B, Chen NS, Fraser AG 2006. Combinatorial RNA interference in *Caenorhabditis*

C. elegans reveals that redundancy between gene duplicates can be maintained for more than 80 million years of evolution. *Genome Biology* 7.

Tong AHY, Lesage G, Bader GD, Ding HM, Xu H, Xin XF, Young J, Berriz GF, Brost RL, Chang M, et al. 2004. Global mapping of the yeast genetic interaction network. *Science* 303: 808-813.

Ulitsky I, Shamir R 2007. Pathway redundancy and protein essentiality revealed in the *Saccharomyces cerevisiae* interaction networks. *Molecular Systems Biology* 3.

VanderSluis B, Bellay J, Musso G, Costanzo M, Papp B, Vizeacoumar FJ, Baryshnikova A, Andrews B, Boone C, Myers CL 2010. Genetic interactions reveal the evolutionary trajectories of duplicate genes. *Molecular Systems Biology* 6.

Vavouri T, Semple JI, Lehner B 2008. Widespread conservation of genetic redundancy during a billion years of eukaryotic evolution. *Trends in Genetics* 24: 485-488.

Wagner A 2005. Distributed robustness versus redundancy as causes of mutational robustness. *Bioessays* 27: 176-188.

Wagner A 2000. Robustness against mutations in genetic networks of yeast. *Nature Genetics* 24: 355-361.

Wong SL, Zhang LV, Tong AHY, Li ZJ, Goldberg DS, King OD, Lesage G, Vidal M, Andrews B, Bussey H, et al. 2004. Combining biological networks to predict genetic interactions. *Proceedings of the National Academy of Sciences of the United States of America* 101: 15682-15687.

Fig.1. Estimated conditional probabilities of yeast WGD duplicate pairs plotting against the number of functional components $m=2, \dots, 20$. (A) $Q(d_A, d_B/O^-)$, predicted probabilities conditional of ancestral essentiality (O^-). (B) $Q(d_A, d_B/O^+)$, predicted probabilities conditional of ancestral dispensability (O^+).