Chapter 4

Functional redundancy between *C. elegans* gene duplicates

4.1. Introduction

Having validated combinatorial RNA interference (RNAi) as a robust method to simultaneously perturb the expression of any pairwise combination of genes, I sought to use this approach to uncover functional genetic redundancy in the *C. elegans* genome.

One obvious possible cause of genetic redundancy is through gene duplication (as discussed in the Introduction). Gene duplicates with at least partially overlapping functions can confer robustness to mutation in the other copy (Force *et al.*, 1999; Lynch and Force, 2000). While genome-wide loss-of-function screens provide indirect evidence that gene duplicates may often share redundant functions (Conant and Wagner, 2004; Gu *et al.*, 2003; Kamath *et al.*, 2003), this hypothesis has not been extensively tested with systematic experimental approaches at the time my study began.

I therefore set out to investigate whether *C. elegans* gene duplicates have redundant functions by using combinatorial RNAi. I reasoned that if gene duplicates were genuinely functionally redundant, targeting both genes of a duplicate pair would result in a more severe loss-of-function phenotype than observed when targeting each gene individually. In the most dramatic case, if gene duplicates together confer an essential redundant function, inactivation of both genes of such a pair would be expected to result in synthetic lethality.

4.2. Examining the redundancy of duplicated genes in the genome of *C. elegans*

To investigate the extent of functional redundancy between gene duplicates in the worm, I focused on *C. elegans* gene pairs that correspond to single orthologues in *S. cerevisiae* or *D. melanogaster* genomes. These genes have thus been duplicated in the genome of *C. elegans* since divergence from either species.

Using the INPARANOID algorithm to identify such gene pairs, the *C. elegans* genome was found to comprise a total of 293 gene pairs that have been duplicated since split from yeast or fly (Table 4.1.). To determine whether these gene duplicates share redundant functions, I set out to examine whether targeting both genes of a duplicate pair affected

***C. elegans* gene duplicates**	***S. cerevisiae***	***D. melanogaster***	***S. cerevisiae* & *D. melanogaster***	**Total**
Identified	79	160	54	293
RNAi clones available	53	105	37	195
Amenable to analysis	49	75	29	153

Table 4.1. *C. elegans* duplicate gene pairs that correspond to single orthologues in *S. cerevisiae* and *D. melanogaster* genomes

Gene pairs that have been duplicated in the genome of *C. elegans* since divergence from *S. cerevisiae* and *D. melanogaster*, respectively, were investigated for potential redundant functions. Shown are numbers for *C. elegans* gene duplicates that were identified by using the INPARANOID algorithm (Remm *et al.*, 2001) ('identified'), that could both be targeted by double-strand RNA- (dsRNA-) expressing clones using the Ahringer RNAi feeding library (Kamath *et al.,* 2003; 'RNAi clones available'), and that were amenable to analysis after excluding cross-reacting RNAi clones with inserts having more than 80% nucleotide identity over 200 base pairs with other genes ('amenable to analysis').

viability, fecundity, or growth in a non-additive, synergistic manner compared with the effects of targeting the individual genes.

For 195 out of 293 *C. elegans* gene duplicates that I had identified, RNAi clones were available from the *C. elegans* whole-genome RNAi library to target each gene of a pair. Of these, I excluded all genes that were targeted by bacterial clones with inserts having more than 80% nucleotide identity over 200 bp with the other copy — this is the threshold for cross-reaction used in Kamath *et al.* (2003) — to ensure that I am not targeting both genes of a duplicate pair with one RNAi clone. This left me with 153 duplicate gene pairs amenable to analysis for synthetic phenotypes using combinatorial RNAi (see Appendix Table 4.1.).

For each duplicate gene pair, I compared the RNAi phenotypes for each gene individually with combinatorial RNAi phenotypes side by side, using the HTP liquid-feeding assay, and the RNAi-hypersensitive *rrf-3* strain (as described in Chapter 3 and Materials and Methods; see Figure 3.6). At that stage, 10 duplicate gene pairs had to be excluded from the screen for functional redundancy, because one or other of the individual genes resulted in first-generation larval growth arrest — a phenotype that cannot be enhanced any further — following RNAi.

After two initial rounds of qualitative analysis, all duplicate gene pairs that appeared to show a stronger combinatorial RNAi phenotype as compared to the contributions of each single-gene RNAi phenotype were further verified by quantification (as described in Chapter 3 and Materials and Methods). Quantitative phenotype data were subsequently subject to statistical analysis under a multiplicative model (as described in Chapter 3 and Materials and Methods). Briefly, for each duplicate gene pair, brood size and embryonic survival, respectively, following combinatorial RNAi were compared to the measurements after RNAi against each gene individually, and the expected product associated with single-gene phenotypes using a Student's t-Test (two-tailed distribution, two-sample equal variance). I interpret a synthetic enhancement interaction under a multiplicative model — that is, where the combined phenotype is significantly stronger (as represented by a p-value below 5.0×10^{-2}) than the product of the individual phenotypes — as indicating functional genetic redundancy.

In total, of 143 duplicate gene pairs amenable to analysis by combinatorial RNAi, I identified 16 gene pairs as having synthetic lethal (SL) phenotypes by the criteria discussed above (Table 4.2. and Figure 4.1.). These data thus suggest that these duplicate pairs are — at least in part — functionally redundant. Of these gene pairs, only two have been previously identified as having redundant functions (Koh *et al.*, 2002; Lambie and Kimble, 1991). The pairs of genes that when co-targeted give SL phenotypes encode diverse molecular functions, ranging from structural constituents of the ribosome (e.g. *rpa-2* + C37A2.7, *rpl-25.1 + rpl-25.2*), signaling proteins (e.g. *lin-12 + glp-1*, C13G3.3 + W08G11.4), and transcription factors (e.g. *elt-6 + egl-18*) to polyadenylate-binding proteins (e.g. *pab-1 + pab-2*) (Table 4.3). Thus, the duplicate gene pairs that I have identified to share redundant functions do not appear to be enriched for specific biological function.

4.3. Transferring gene functions between *S. cerevisiae* and *C. elegans*

The duplicated genes that I focused on in the worm corresponded to single genes in either *S. cerevisiae* or *D. melanogaster* genomes. I wished to investigate whether the known function of a single gene in one organism was a good predictor of the synthetic RNAi phenotype identified by co-targeting the corresponding duplicated worm genes with redundant functions. If this were the case, then it is most likely that the redundancy that I observe is due to both duplicates retaining the ancestral molecular function.

As a preliminary to this study I sought to investigate whether the known function of an individual gene in one organism can predict the molecular function of its single orthologue in *C. elegans*. Testing the conservation of individual gene functions between species would allow me to assess the potential of predicting gene functions covered by pairs of redundant genes. I chose to focus on transferring individual gene functions between *S. cerevisiae* and *C. elegans*, because to date, yeast and worm are the main model organisms in which fully systematic functional studies can be performed *in vivo*. Moreover, I will be discussing the conservation of synthetic lethal interactions between yeast and worm in the next chapter.

Interaction Gene1 & Gene2	Gene1		Gene2		Gene1 & 2		p-value	p-value
	BS	ES	BS	ES	BS	ES	BS	ES
pab-1 + *pab-2*	15	100	88	100	0	n.s.	1.9E-04	n.s.
rpl-25.2 + *rpl-25.1*	6	50	17	63	0	n.s.	3.6E-04	n.s.
ptr-2 + *ptr-10*	*	53	*	98	*	n.s.	*	n.s.
unc-78 + *tag-216*	85	96	98	97	0	n.s.	6.4E-15	n.s.
rab-8 + *rab-10*	87	98	70	96	1	n.s.	7.3E-05	n.s
B0495.2 + ZC504.3	84	99	97	99	2	13	6.3E-09	1.4E-17
rpa-2 + C37A2.7	67	74	50	81	1	n.s.	1.9E-07	n.s.
C28H8.4 + *erd-2*	93	95	86	94	10	10	5.6E-08	2.2E-15
lin-12 + *glp-1*	90	95	99	83	16	75	1.2E-13	3.0E-01
C13G3.3 + W08G11.4	73	94	80	97	17	89	1.6E-06	3.5E-01
lin-53 + *rba-1*	74	63	51	5	16	75	1.1E-02	7.3E-17
Y53C12A.4 + R02E12.2	84	81	78	87	32	75	1.3E-03	6.9E-01
F37C12.7 + *acs-17*	95	100	77	98	44	73	9.4E-03	4.2E-06
C05G5.4 + F23H11.3	96	100	94	98	58	72	5.1E-06	1.5E-08
elt-6 + *egl-18*	100	97	86	88	63	73	4.0E-02	6.3E-03
dsh-1 + *dsh-2*	97	98	75	54	58	17	1.6E-02	1.1E-11

Table 4.2. *C. elegans* duplicate gene pairs with at least partially redundant functions
C. elegans duplicate gene pairs ('Interaction Gene1 & Gene2') displaying synthetic phenotypic effects upon combinatorial RNA interference (RNAi) in the RNAi-hypersensitive strain *rrf-3* are listed. Numbers shown are percentages of average wild-type brood size ('BS') and embryonic survival ('ES') rates after RNAi against each gene individually ('Gene1', 'Gene2') as well as after combinatorial RNAi against duplicate gene pairs ('Gene1 & 2'), and are the arithmetic mean of two independent biological repeats. Statistical significance of quantitative phenotype data (BS, ES) was evaluated under a multiplicative model (Phillips *et al.*, 2000; Puniyani *et al.*, 2004); p-values were assigned using a Student's t-test. n.s., given phenotype could not be quantified. * Note

that combinatorial RNAi against the duplicate gene pair *ptr-2 + ptr-10* resulted in an increased number of first generation larval growth arrested worms, rather than in reduced brood size; fraction of population which is wild-type, i.e. that does not arrest at an early larval stage: 70% (*ptr-2*), 100% (*ptr-10*), 0% (*ptr-2 + ptr-10*), *P* = 7.3E-09.

Figure 4.1. Quantitative analysis of synthetic lethal phenotypes following the simultaneous targeting of both genes of a duplicate pair

Phenotypes of duplicate gene pairs that yielded reproducible synthetic effects after combinatorial RNA interference (RNAi) were quantified. For each gene pair, brood size (BS) and embryonic survival (ES) after combinatorial RNAi against both duplicates (red bars), after RNAi against each gene individually (light- and dark-grey bars), and the calculated product of BS and ES measurements, respectively, of both individual genes (blue bars) are shown. Values plotted represent the percentage of average wild-type brood size and embryonic survival rates, respectively, and are the arithmetic mean of two independent RNAi experiments performed in the RNAi-hypersensitive *rrf-3* background. Duplicate gene pairs were considered to be synthetic lethal, if either BS or ES measurements were significantly reduced (P < 5.0E-02; Student's t-test) as compared to the multiplicative values of the single-gene BS and ES measurements, respectively. ***, P < 1.0E-03; **, P < 1.0E-02; *, P < 5.0E-02. Note that combinatorial RNAi against the gene pair *ptr-2 + ptr-10* resulted in a significantly increased number of first-generation larval growth arrested worms (P = 7.3E-09, Student's t-test), rather than a brood size defect, hence these data are not shown.

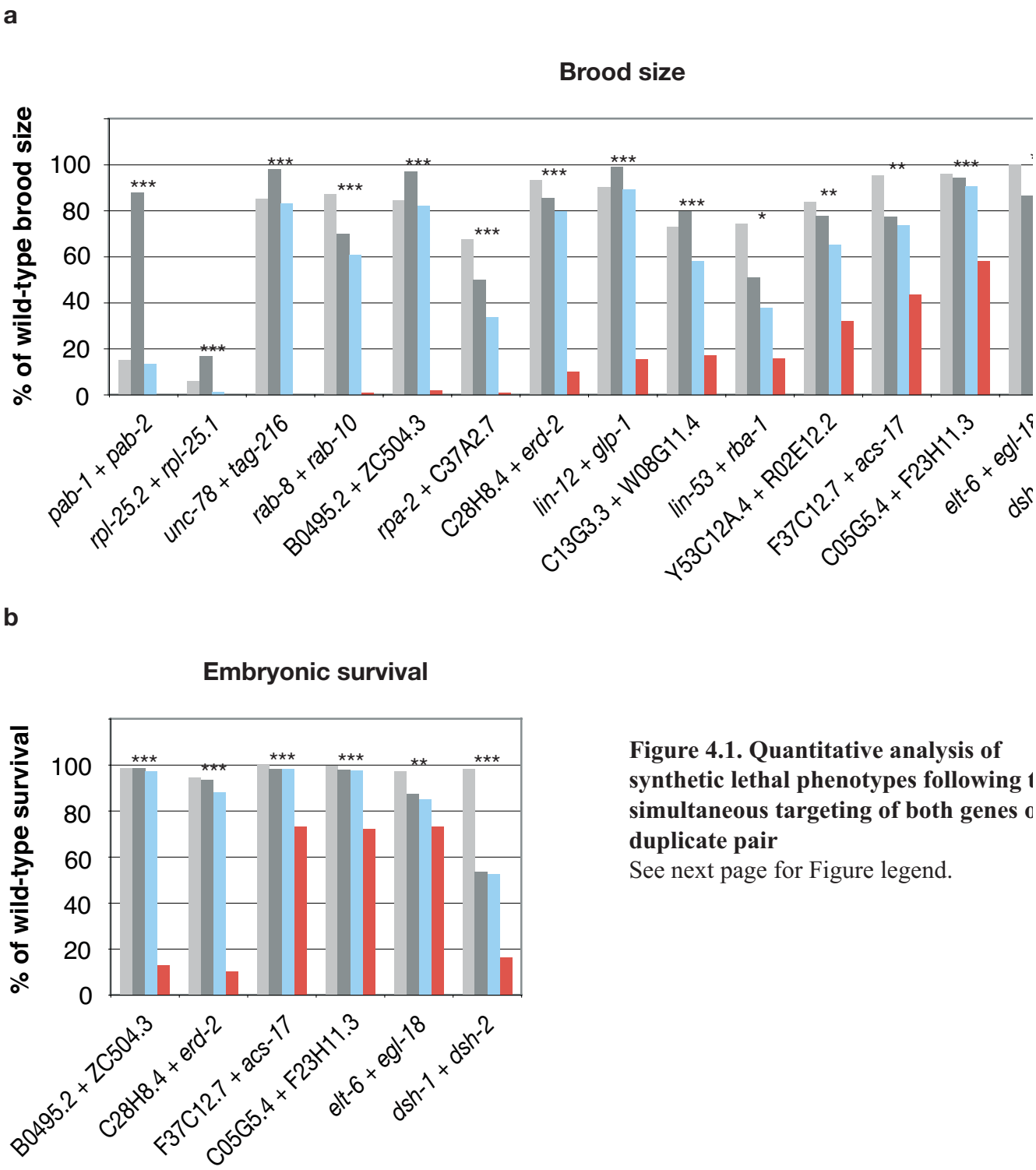

Figure 4.1. Quantitative analysis of synthetic lethal phenotypes following t simultaneous targeting of both genes o duplicate pair

See next page for Figure legend.

of the redundant function covered by a pair of duplicated genes in *C. elegans.* Based on the gene deletion phenotypes of the single-copy orthologues in yeast, I split the set of *C. elegans* duplicate gene pairs into those corresponding to essential or non-essential *S. cerevisiae* genes (see Appendix Table 4.2.). I found that five of eighteen worm duplicates (28%), that are orthologous to yeast essential genes, showed synthetic lethal phenotypes by combinatorial RNAi. In contrast, only five of fifty-five *C. elegans* duplicate gene pairs (9%) corresponding to *S. cerevisiae* non-essential genes were found to result in a synthetic viability defect when co-targeted. I thus conclude that duplicated genes in *C. elegans* that are related to an essential gene in yeast are about three times more likely to have an essential redundant function than those related to a non-essential yeast gene. Strikingly, this is the same enrichment for non-viable RNAi phenotypes as for non-duplicated genes: 61% of *C. elegans* single-copy orthologues of *S. cerevisiae* essential genes have non-viable RNAi phenotypes, compared to 20% of orthologues of yeast non-essential genes (Figure 4.2.) Thus, this finding is entirely consistent with a simple model of redundancy, suggesting that the function of a single gene identified in one organism is a good predictor of the redundant function covered by a pair of duplicated genes in a second organism.

4.4. Duplicated genes can maintain redundant functions for more than 80 million years of evolution

Having found that over 10% of genes (16 out of 143) that have been duplicated in the genome of *C. elegans* since the divergence from either *S. cerevisiae* or *D. melanogaster* share at least partially redundant functions, I next sought to address the underlying causes for this redundancy. Therefore, I wished to study the properties of gene duplicates with redundant functions, and whether these differ from duplicated gene pairs that were not identified as having redundant functions. For reasons of compactness, I will refer to these as ‘redundant’ and ‘non-redundant’ duplicate gene pairs, although of course I recognize that failure to detect a phenotype by RNAi does not preclude a genuine function.

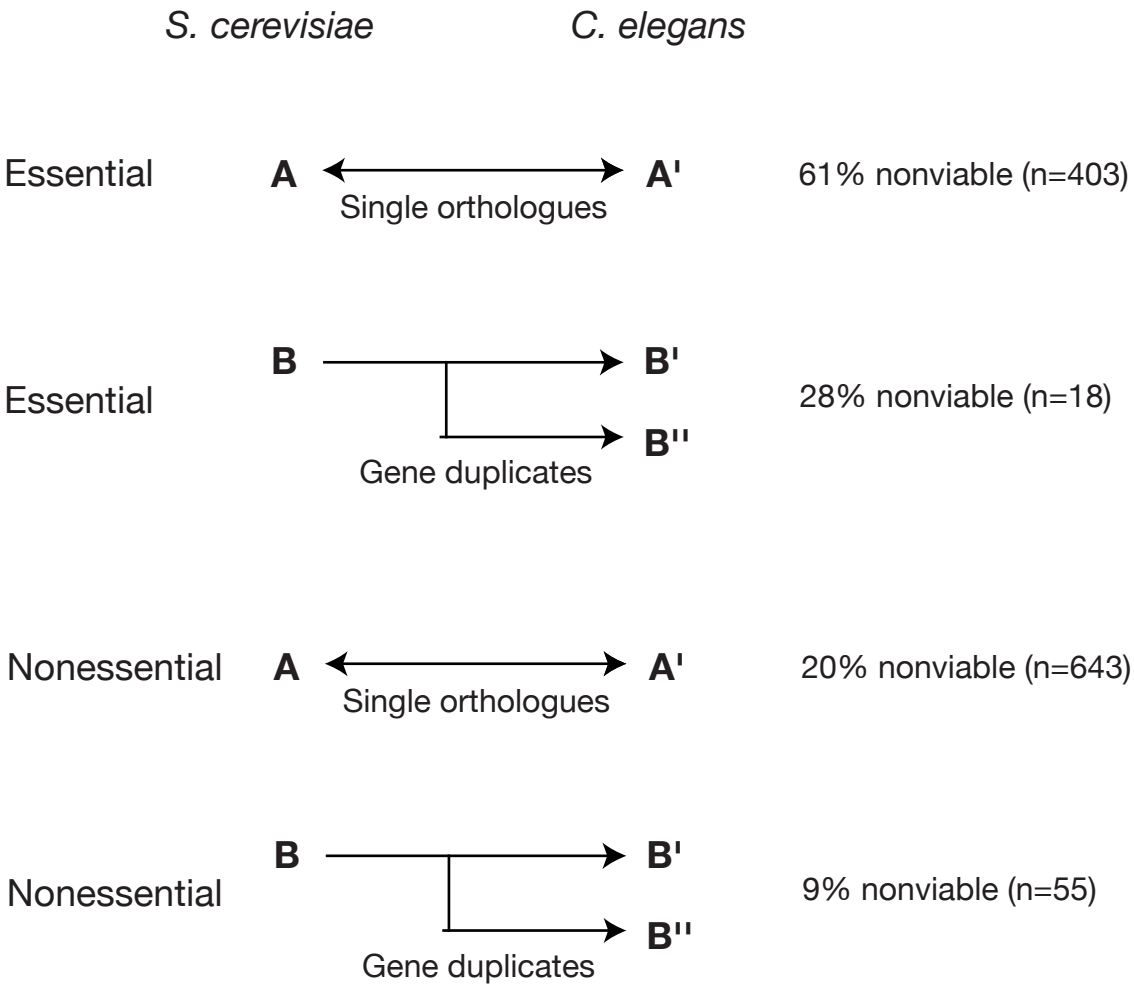

Figure 4.2. Transferring gene functions between *S. cerevisiae* and *C. elegans*
Orthologues of genes with essential functions in *S. cerevisiae* are very likely to have nonviable RNAi phenotypes in *C. elegans* ('Single orthologues'). Likewise, genes that are essential in yeast, but that have duplicated in *C. elegans* are likely to have a nonviable RNAi phenotype in *C. elegans* when both genes are targeted simultaneously by combinatorial RNAi ('Gene duplicates') The numbers indicate the percentage of tested genes or gene pairs with non-viable RNAi phenotypes in two independent experiments performed in the RNAi-hypersensitive strain *rrf-3*.

I considered two simple models that might explain why some duplicated genes appear to have redundant functions (as discussed in the Introduction). First, the redundancy may be a by-product resulting from a recent duplication event and thus represent a transient state; this initial functional overlap might get lost over time by functional divergence (Force *et al.*, 1999; Kimura and King, 1979; Lynch and Force, 2000; Ohno, 1970). In this case, the pairs of genes that I identified as having redundant essential functions would be expected to be more recent duplicates than those for which I found no functional overlap. Alternatively, several groups have established theoretical frameworks suggesting that redundant functions can be maintained by natural selection over substantial periods of evolutionary time (Nowak *et al.*, 1997; Wagner, 2000b). In this case, I would expect no clear difference in age between the sets of redundant and non-redundant duplicate gene pairs. Instead, I anticipated that there would be evidence that the redundant duplicated genes have been maintained relative to their ancestral sequence, thereby retaining their overlapping, redundant functions.

To investigate which of these two models can best explain the redundancy that I observed between some gene duplicates, I first examined whether there is evidence that the redundant gene pairs have duplicated more recently than non-redundant pairs. One would of course anticipate that more recently duplicated genes are more likely to have overlapping functions than more ancient duplicate gene pairs. Intriguingly, when investigating the number of synonymous nucleotide substitutions per synonymous site (Ks) as a measure of the evolutionary age of gene duplicates, I found the average rate of change to be 13.41 for redundant duplicates and 9.48 for non-redundant duplicates, indicating that both redundant and non-redundant duplicate gene pairs are ancient (see Appendix Table 4.3.), and their divergence time can no longer be reliably estimated. Having found no clear evidence that the redundant gene pairs represent more recent gene duplicates than the non-redundant gene pairs, I considered the possibility that the redundancy that I observe might simply be the consequence of a lack of evolutionary time for the duplicates to drift, as very unlikely.

Next, I set out to examine whether the duplicate gene pairs with essential redundant functions also do exist as gene duplicates in the related nematode *C. briggsae*. To do so, the INPARANOID algorithm was used to identify *C. briggsae* orthologues of

C. elegans genes (Remm *et al.*, 2001); I considered the gene duplication to predate the divergence of *C. elegans* from *C. briggsae*, if both *C. elegans* duplicates had a single identifiable orthologue in *C. briggsae*. Remarkably, 14 of the 16 pairs of duplicated genes that I identified as having essential redundant functions in *C. elegans* appear to have also been maintained as gene pairs in the related nematode *C. briggsae*. These findings suggested that these 14 duplicate gene pairs with redundant functions have arisen from a duplication event that predated the split from *C. briggsae*. In contrast, only 100 out of 127 non-redundant duplicate gene pairs also exist as gene pairs in *C. briggsae*. Thus, the frequency of conservation of redundant gene pairs between *C. elegans* and *C. briggsae* is significantly higher than the frequency observed for non-redundant duplicate gene pairs (χ^2 = 8.653, P = 0.0033, 1 degree of freedom; see Appendix Table 4.3.). *C. elegans* and *C. briggsae*, despite being morphologically very similar, last shared a common ancestor 80-110 million years ago (Stein *et al.*, 2003). Taking into account that *C. elegans* and *C. briggsae* only share ~60% of their genes as single orthologues, and a full 10% of genes encoded in either genome has no identifiable match in the other genome (Stein *et al.*, 2003), one would anticipate less than 40% of *C. elegans* duplicate gene pairs to be randomly conserved as pairs in *C. briggsae*. I thus consider the possibility that these 14 duplicate gene pairs with redundant essential functions in *C. elegans* have been retained as duplicate pairs in *C. briggsae* simply as a result of neutral evolution to be very unlikely. Instead, these data suggest that the redundancy between these duplicated genes might have been actively maintained for more than 80 million years of evolution.

Thus, I next sought to investigate whether there is evidence that the overlap in function has been actively retained by natural selection. If there has been selection for the maintenance of redundancy between duplicate gene pairs, then I would expect these duplicates to encode more similar proteins than non-redundant duplicates. To determine the percentage of identity between gene duplicates, protein sequences were aligned using the CLUSTAL W program (Thompson *et al.*, 1994). I found pairs of redundant duplicated genes to be more similar to each other at the amino-acid level ($P = 1.6 \times 10^{-2}$, Wilcoxon rank sum test, Figure 4.3.a) and to also have a greater similarity in alignable protein length ($P = 2.2 \times 10^{-2}$, Figure 4.3.b) than non-redundant duplicates and finally to also show a lower rate of non-synonymous nucleotide substitution per non-synonymous

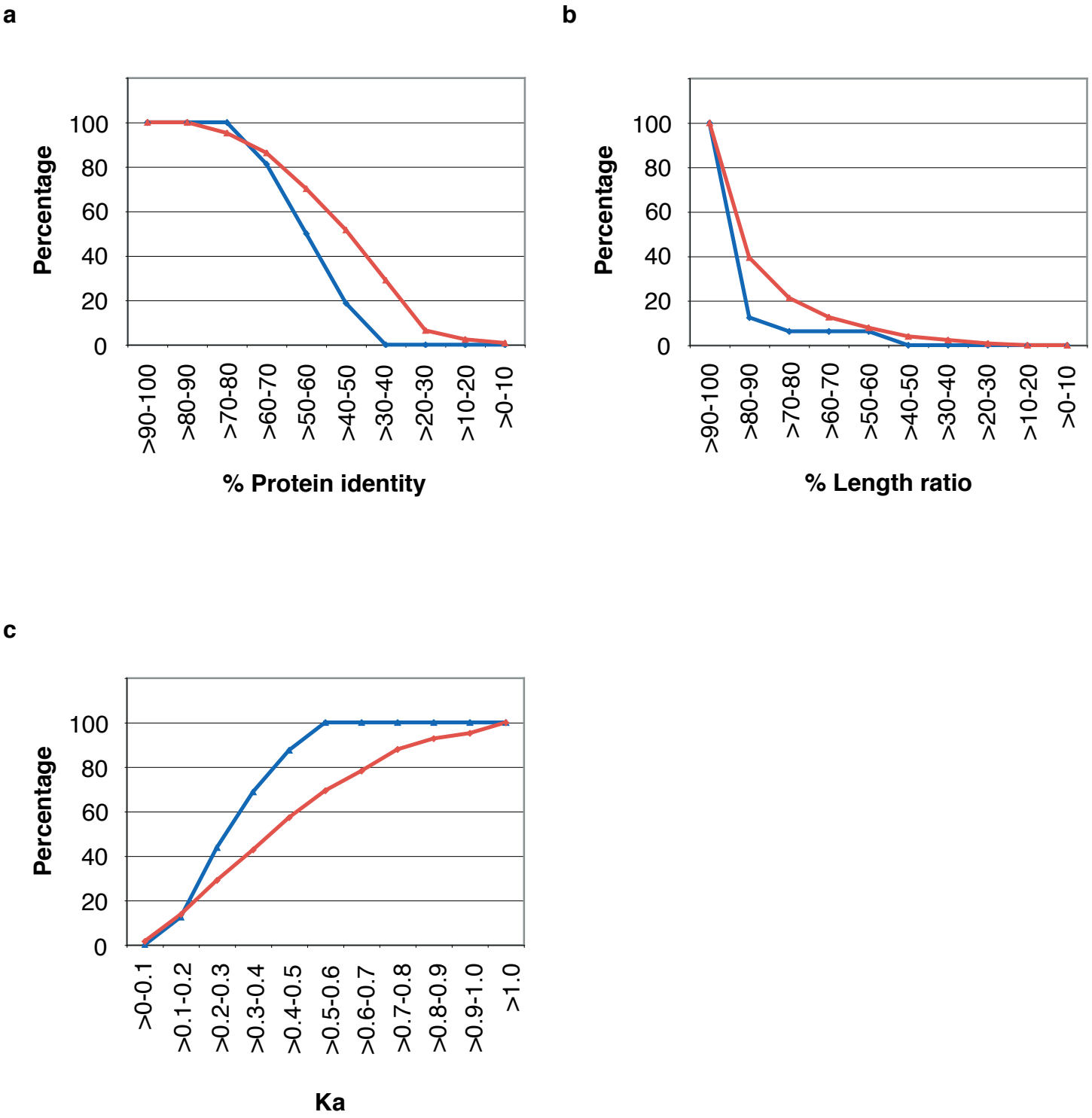

Figure 4.3. Higher sequence similarity between redundant versus non-redundant gene duplicates
Percentage similarity in protein sequence (a) and alignable protein length (b), and rate of non-synonymous nucleotide substitutions per non-synonymous site (Ka; c) are contrasted for gene duplicates with redundant functions (shown in blue) and for duplicate gene pairs that were not identified as having redundant functions (shown in red).

site (mean Ka for redundant duplicates = 0.34; mean Ka for non-redundant duplicates = 0.50; P = 3.8 x 10^{-2}; Figure 4.3.c) than non-redundant duplicates (see Appendix Table 4.3.). While I recognize that the redundant duplicate pairs appear only marginally more similar to each other than non-redundant gene duplicates, I found several independent lines of evidence that together suggest that redundant gene duplicates may have remained a higher degree of similarity as a consequence of stronger purifying selection than duplicate gene pairs that were not identified as having redundant functions.

Several theoretical models have been generated to explain how redundant functions can be maintained by natural selection (as discussed in the Introduction). One theory relates pleiotropy to redundancy. In this model, both copies are only redundant with respect to some sub-functions, while they also perform independent functions and thus are evolutionarily selected (sub-functionalization). While both experimental and theoretical studies support sub-functionalization as a likely evolutionary fate of gene duplicates (Kondrashov *et al.*, 2002; Lynch and Force, 2000) and a means to maintain gene duplicates, one would anticipate that the same mechanisms act on both redundant and non-redundant gene duplicates. Therefore, the sub-functionalization model cannot explain how redundant gene duplicates have maintained a higher degree of sequence similarity as compared to non-redundant gene duplicates. Further two models for the evolutionary stability of genetic redundancy are based on the assumption of very specific mutation rates and efficacies of protein function (Nowak *et al.*, 1997). These theories can — at least mathematically — explain how gene duplicates can maintain sequence similarity and perform the exact same function for very long or even infinite evolutionary timescales. I believe that my findings favour these latter models.

Taken together, I consider it as unlikely that the greater similarity between duplicate gene pairs with redundant functions that I observed is a trivial consequence of their having duplicated more recently. Rather, I suggest that the protein sequences of redundant gene pairs have been maintained relative to each other since duplication as the result of selective pressure to maintain their redundant functions.

4.5. Conclusion

In summary, in this chapter I have described how I have used combinatorial RNAi to systematically investigate whether there is functional redundancy between *C. elegans* gene duplicates. Focusing on genes that have been duplicated in the genome of *C. elegans* since divergence from either *S. cerevisiae* or *D. melanogaster*, I was able to analyse 143 duplicate gene pairs by combinatorial RNAi for their potentially redundant functions. Of these, 16 gene pairs showed unambiguous synthetic RNAi phenotypes, demonstrating that they are at least partially functionally redundant. I found that just as single-copy worm genes are more likely to have a non-viable RNAi phenotype if they are orthologous to an essential gene in *S. cerevisiae*, duplicated worm genes are more likely to have a redundant essential function if they are co-orthologous to an essential yeast gene. It therefore should be possible to predict the redundant functions of many duplicated genes in higher organisms based on the functions of single-copy orthologues in lower organisms.

Most intriguingly, the redundancy that I observed between duplicated genes cannot be explained simply because they are derived from a recent duplication event — 14 of the 16 redundant gene pairs were duplicated before the divergence of *C. elegans* and *C. briggsae* 80-110 million years ago (Stein *et al.*, 2003). The redundancy between these 14 gene pairs has thus been maintained for more than 80 million years of evolution. Therefore, I believe that it is extremely unlikely that the functional overlap between these 14 duplicated genes is present merely due to a lack of evolutionary time since duplication. Not only is the average half-life of a gene duplicate in eukaryotes typically about 4 million years (Lynch and Conery, 2000), but also, over this time period, the *C. elegans* and *C. briggsae* genomes have diverged enormously; they only share ~60% of their genes as single orthologues, and a further 10% of genes are present exclusively in one or other genome (Stein *et al.*, 2003). Rather, my findings are consistent with theoretical models, suggesting that under appropriate — but realistic — conditions it is possible to select, directly or indirectly, for redundancy between duplicates to be maintained (Nowak *et al.*, 1997).

Having provided the first large-scale analysis in any organism of the redundant functions of gene duplicates, I wished to further examine functional redundancy in complex genetic networks.