Chapter 1

Introduction

While biological research in the second half of the 20th century embraced 'molecule-centered' approaches to defining the properties of individual components of living systems, this reductionist view has changed with the availability of whole genome sequences for numerous model organisms. By employing integrated computational-experimental approaches, it had become a reachable goal to unravel the molecular functions of each individual gene. These advances led to the appreciation that for a better understanding of living systems, one needs to take a step beyond studying gene functions one at a time. Researchers realized that properties of systems — e.g. organisms, gene networks, protein complexes — are more than merely the sum of their parts. This view led to defining a new area of research, 'Systems Biology' (Ideker *et al.*, 2001; Kitano, 2002a; Kitano, 2002b; Pennisi, 2003). The central aim of Systems Biology is to gain insights into the 'emergent' properties of systems, features that cannot be attributed to any of their individual components. In order to get an understanding of such emergent properties, one needs firstly to identify all of a system's individual components, secondly to unravel the function(s) of each individual component and, finally, to decipher how these individual components interact to result in system-level dynamics (reviewed in Strange, 2006).

1.1. Outline of introduction

In this introduction, I will discuss the major steps that have been taken in the post-genomics era to address the fundamental biological question: 'How does an organism's genotype relate to its phenotype?' Since I used *C. elegans* as model organism for my studies, I will start by introducing 'the worm' and its versatility as model system. I will then describe the various genomics approaches that have been taken to generate genome-scale views of gene function in *C. elegans*, *D. melanogaster*, and *S. cerevisiae*. Moreover, I will discuss the insights gained from large-scale studies of gene functions in worm, fly, and yeast and consequently the hypotheses that have been brought forward. I will conclude by introducing the rationale, aim, and significance of my study.

1.2. *Caenorhabditis elegans* as a model system

The nematode *Caenorhabditis elegans* — also referred to as 'the worm' — has been extensively used as an experimental organism for genetics, development and neurobiology in the 1970s (Brenner, 1974). Numerous unique attributes make *C. elegans* a powerful model system (reviewed in Jorgensen and Mango, 2002; Strange, 2006): Worms have a short life–cycle, with a three-day generation time at room temperature. Animals pass through four larval stages (L1-L4), before they reach adulthood and become fertile. Worms are small in size (~1 mm), produce ~ 300 progeny per animal and can be grown easily and inexpensively in the laboratory on agar plates or in liquid cultures. *C. elegans* feed on bacteria, but can enter a specific developmental programme — called the 'dauer stage' — under limiting food conditions, in which they can survive for months. Furthermore, worms can be kept as frozen stocks.

C. elegans reproduces through self-fertilization in hermaphrodites or by mating with males. Hermaphrodites carry two X chromosomes, whereas males are of X0 karyotype, which arises through occasional meiotic non-disjunction. Self-fertilization allows worms to be propagated clonally and greatly facilitates the isolation of mutants through genetic screens, while males are used for inter-crossing mutant strains.

C. elegans is a highly differentiated animal but of simple anatomy — hermaphrodites comprise of 959 somatic cells only. Nonetheless, worms share many tissues with more complex animals, such as the nervous system, muscles, a reproductive and gastro-intestinal tract and an epidermis. Heroic efforts have been made to trace the fate of every somatic cell in *C. elegans*, starting from the first division (Sulston and Horvitz, 1977; Sulston *et al.*, 1983). This observation culminated in a documentation of the complete cell lineage of the worm, which was found to be relatively invariant. This study was followed by a comprehensive description of the structure of the nervous system, which resulted in a wiring diagram of all 302 *C. elegans* neurons (White *et al.*, 1986).

1.3. Uncovering gene functions in *C. elegans*

1.3.1. Forward genetic screens

Traditionally, ethyl methane sulphonate- (EMS-) or N-ethyl-N-nitrosourea- (ENU-) induced mutagenesis followed by genetic screening has been used to identify genes that function in a biological process or pathway of interest (reviewed in Jorgensen and Mango, 2002). Typically, about one null mutation in any single gene is recovered in ~ 2,000 genomes through such forward genetic approaches. By using mutagenesis screens, Sydney Brenner identified 619 mutants with visible phenotypes (Brenner, 1974). These were instrumental in establishing *C. elegans* as a key model organism, and many similar screens have been performed since then.

1.3.2. Reverse genetic approaches

As an alternative to forward genetic analysis, injection of anti-sense RNAs homologous to any gene of interest has been used as a means to reduce gene expression from endogenous loci to study loss-of-function phenotypes (Fire *et al.*, 1991; Guo and Kemphues, 1995). Surprisingly, Guo and Kemphues found that both anti-sense and sense RNAs were equally efficient for suppressing gene expression (Guo and Kemphues, 1995). This result led Andrew Fire and Craig Mello to their breakthrough. They discovered that introduction of double-stranded RNAs (dsRNAs) into *C. elegans* was substantially more effective in reducing gene expression than introducing single-stranded RNAs (Fire *et al.*, 1998). This potent RNA-mediated interference effect, however, was not observed when dsRNAs homologous to promoters or intronic sequences were used (Fire *et al.*, 1998). These observations, together with the finding that injection of dsRNA resulted in reduced or undetectable levels of corresponding mRNAs, suggested a posttranscriptional gene silencing mechanism. Studies in plant systems further supported this notion (de Carvalho *et al.*, 1992; Jones *et al.*, 2001; Ruiz *et al.*, 1998). Finally, biochemical approaches verified that the observed interference effect resulted from dsRNA-induced degradation of the endogeneous mRNA (Hammond *et al.*, 2000; Kennerdell and Carthew, 1998; Tuschl *et al.*, 1999). This mechanism was termed RNA interference (RNAi).

1.3.2.1. RNA interference phenomena

Strikingly, injection of dsRNA into the germline or extracellular body cavity of *C. elegans* was found to result in an interference effect in a broad region of the animal, demonstrating that dsRNA has the remarkable capacity to cross cell boundaries ('spreading'; Timmons and Fire, 1998). This observation led to the significant discovery that feeding worms on *Escherichia coli* engineered to express dsRNA could also confer specific interference effects; just as worms normally feed on bacteria, dsRNA-expressing bacteria are ingested, dsRNA absorbed through the gut and distributed to somatic tissues and the germ line ('RNAi by feeding'; Timmons and Fire, 1998). Subsequently, it was shown that soaking worms in a solution of dsRNA ('RNAi by soaking') could also induce specific interference with gene expression (Tabara *et al.*, 1998). Finally, it was found that introduction of dsRNA into hermaphrodite worms can also produce a specific and robust interference effect in the progeny (Tabara *et al.*, 1998). Thus, dsRNA-induced gene silencing can be used to study the loss-of-function phenotype of any gene of known sequence. However, RNAi by bacterial feeding or soaking is less effective than direct injection of dsRNA (Tabara *et al.*, 1998; Timmons and Fire, 1998).

1.3.2.2. Mechanism of double-stranded RNA-induced gene silencing

Genetic studies in *C. elegans* and plants together with biochemical approaches using *Drosophila* embryonic extracts or S2 cells have provided fundamental insights into the mechanisms underlying dsRNA-induced gene silencing (reviewed in Hannon, 2002; Joshua-Tor, 2006; Matzke and Birchler, 2005; Zamore and Haley, 2005). In the current model (Figure 1.1.), Dicer, an evolutionarily conserved member of the family of RNase III ribonucleases, recognizes and cleaves dsRNAs into ~ 22-nucleotide fragments, with two-nucleotide 3' overhangs and 5' phosphorylated termini. The protein structure of RNase III enzymes led to the suggestion that the ~22-nucleotide RNAs are generated by association of two Dicer homo-dimers in antiparallel orientation. In this view, one of the two catalytic domains in each homo-dimer is inactive, with the two active catalytic domains being spaced by ~ 22 nucleotides. This configuration results in cleavage of dsRNAs into small interfering RNAs (siRNAs) of ~ 22-nucleotide length. siRNAs are

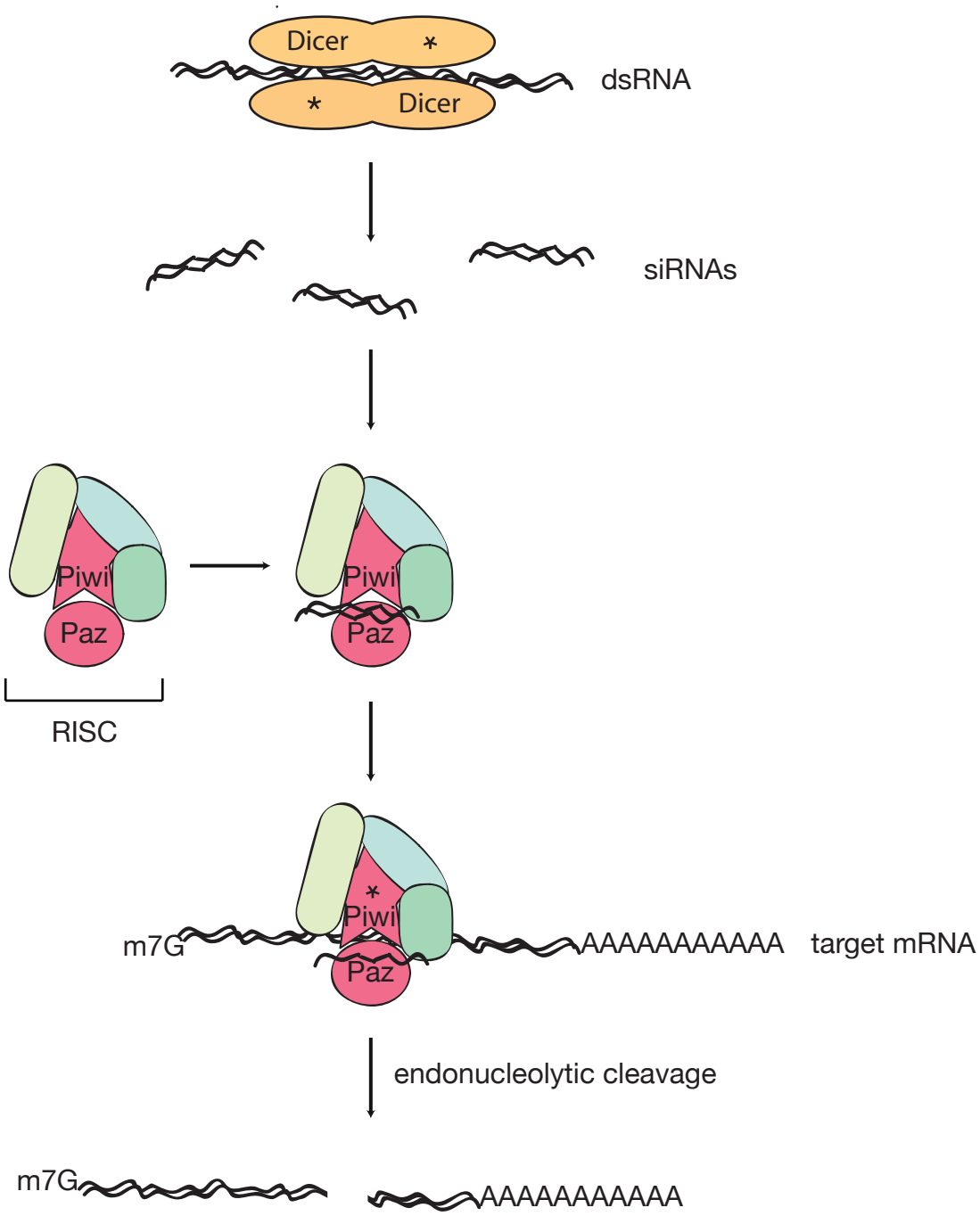

Figure 1.1. Mechanism of RNA interference
Two Dicer homo-dimers associate in anti-parallel orientation to cleave double-stranded RNA (dsRNA). Only one catalytic centre in each Dicer homo-dimer is active (*). Active catalytic domains are spaced by ~ 22 nucleotides (nt), resulting in cleavage products - small interfering RNAs (siRNAs) - of ~ 22 nt length. siRNAs are incorporated into the RNA-induced silencing complex (RISC), a ribonuclease-containing protein complex, which is activated (*) through unwinding of siRNAs. Watson-Crick base-pairing with siRNAs identifies homologous target mRNAs. The Piwi domain of the ribonuclease Slicer (shown in red, comprising of two RNA-binding domains, Piwi and Paz) mediates cleavage of target mRNA.

subsequently incorporated into a ribonuclease-containing protein complex, termed 'RNA-induced silencing complex' (RISC), targeting complementary mRNAs for degradation. Base complementarity with siRNAs identifies target mRNAs, which are subsequently endonucleolytically cleaved by Slicer, the catalytic core of the effector complex RISC. Slicer is a member of the Argonaute protein family, containing two RNA binding domains, the Piwi and PAZ domain. The Piwi domain, which is structurally homologous to ribonuclease H, comprises the endonuclease for cleavage of target mRNAs. *In vitro*, the inactive RISC precursor complex becomes activated upon addition of ATP. Remarkably, a correlation between RISC activation and unwinding of siRNAs was identified.

Intriguingly, in *C. elegans* and plants, RNAi can spread throughout the organism, whereas silencing in *Drosophila* and mammals appears to be cell-autonomous. The requirement of RNA-directed RNA polymerases (RdRPs) for dsRNA-induced gene silencing in *C. elegans* and plants led to the suggestion that siRNAs prime the synthesis for additional dsRNAs by RdRPs, thereby amplifying the silencing signal. While in plants, spreading of dsRNAs can occur by movement through plasmodesmata, in *C. elegans*, a transmembrane protein encoded by *sid-1* is suggested to mediate spreading of the silencing signal. Interestingly, orthologues of SID-1 are also encoded in mammalian genomes, whereas they were not identified in *Drosophila*. However, systemic gene silencing has thus far not been demonstrated in mammals.

1.4. The *C. elegans* genome

Being the first multicellular organism for which a complete genome sequence was available, *C. elegans* set a milestone in genomics in 1998 (The *C. elegans* Sequencing Consortium, 1998). To begin the annotation and analysis of the 97–megabase worm genome, GENEFINDER, an algorithm for the identification of putative coding regions, was used (The *C. elegans* Sequencing Consortium, 1998). By comparing computational gene predictions to experimental genome annotations based on extensive collections of expressed sequence tags (ESTs), over 90% of computationally predicted genes were found to overlap with experimentally verified introns. Additional manual annotation was

used to further refine computational gene structure predictions. With more genomic information becoming available and the development of better genome annotation tools, the analysis and annotation of the *C. elegans* genome is expected to further improve in the future.

The *C. elegans* genome is predicted to encode ~ 19,000 protein-coding genes, with an average density of 1 gene per 5kb, distributed on five autosomes and the X chromosome.

1.5. WormBase

A comprehensive web-accessible database for information on *C. elegans* and related nematodes has been generated for the *C. elegans* research community (www.wormbase.org). This repository is based on the *C. elegans* genome database architecture ('AceDB') that was originally generated for the storage of sequence information (The *C. elegans* Sequencing Consortium, 1998). WormBase is a very navigable database, providing extensive information on the sequence and structure on the genomes of *C. elegans* and its related nematode *C. briggsae*. WormBase also stores information on mutant strains and alleles — these are publicly available from two stock centres, the *C. elegans* Genetics Center, USA (http:// www.cbs.umn.edu/CGC/) and the National Bioresources Project, Japan (http:// shigen.lab.nig.ac.jp/c.elegans/index.jsp). Results from RNAi experiments, gene expression patterns, functional annotations, comparative data, such as orthologues and syntenic regions between species (reviewed in Strange, 2006) are also deposited in WormBase.

1.6. RNA interference in *C. elegans* by bacterial feeding

While the penetrance of the interference effects obtained through delivering dsRNAs by bacterial feeding or soaking are not as strong as those obtained by direct injection of dsRNAs (Tabara *et al.*, 1998; Timmons and Fire, 1998), RNAi by feeding has nonetheless numerous advantages for large-scale applications of RNAi. Firstly, RNAi by feeding is reasonably cost-effective, because it circumvents the need for expensive *in*

vitro synthesis of dsRNA that is necessary when using RNAi by injection or soaking. Secondly, bacterial strains expressing dsRNA can be kept as a reusable resource and thus can be distributed indefinitely. Finally, RNAi by bacterial feeding is less labour-intensive than injection, thereby allowing RNAi experiments to be scaled up to a reasonably high throughput.

This powerful technique led Julie Ahringer's lab to generate a whole-genome RNAi feeding library (also referred to as the 'Ahringer RNAi library'; Fraser *et al.*, 2000; Kamath and Ahringer, 2003; Kamath *et al.*, 2003). In this resource, gene-specific DNA fragments of roughly 1000 – 1500 bp were inserted between inverted repeats of the bacteriophage T7 promoter into a bacterial plasmid vector (L4440; Figure 1.2.). These plasmids were transformed into an RNaseIII-deficient *E. coli* strain (HT115(DE3), also referred to as 'bacterial feeding strain'), which was engineered to express T7 RNA polymerase under an isopropyl-β-D-thiogalactopyranoside- (IPTG-) inducible promoter. Thus, expression of dsRNA could be induced upon addition of IPTG (Timmons *et al.*, 2001; Timmons and Fire, 1998). RNaseIII-deficiency was found to improve the efficacy of RNAi by bacterial feeding, presumably because it results in increased stability of dsRNA that is produced in bacteria (Timmons *et al.*, 2001).

Using this approach, DNA fragments corresponding to ~86% of the predicted *C. elegans* genes were cloned, resulting in a collection of 16,757 dsRNA-expressing bacterial strains (Fraser *et al.*, 2000; Kamath *et al.*, 2003), each targeting one predicted gene. With the generation of this potent resource, RNAi by feeding has become a powerful reverse genetic tool for studying *C. elegans* gene functions on a large scale.

1.6.1. Experimental setup for RNA interference by bacterial feeding

The experimental procedure for RNAi by bacterial feeding, optimized by Kamath and Ahringer (2003), involves three major steps (Figure 1.3.). Firstly, selected bacterial strains of the *C. elegans* whole-genome library are grown to saturation and seeded onto IPTG-containing 12-well assay plates ('plate feeding'). Secondly, late larval-stage worms are placed onto pre-seeded assay plates and are clonally propagated after an appropriate

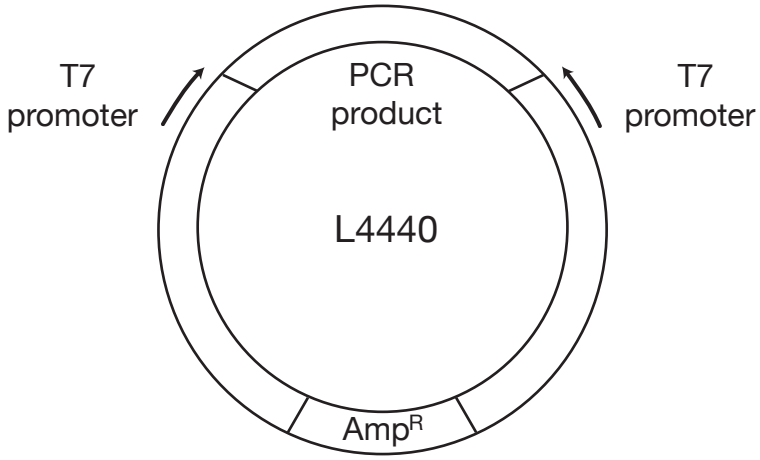

Figure 1.2. L4440 RNA interference feeding vector
A PCR product homologous to a target gene of interest is cloned between inverted T7 promoter sites. Induction of T7 RNA polymerase expression in the bacterial host genome (HT115(DE3)) results in transcription of anti-parallel single-stranded RNAs (ssRNAs). These ssRNAs anneal and form double-stranded RNAs, which trigger RNA interference.

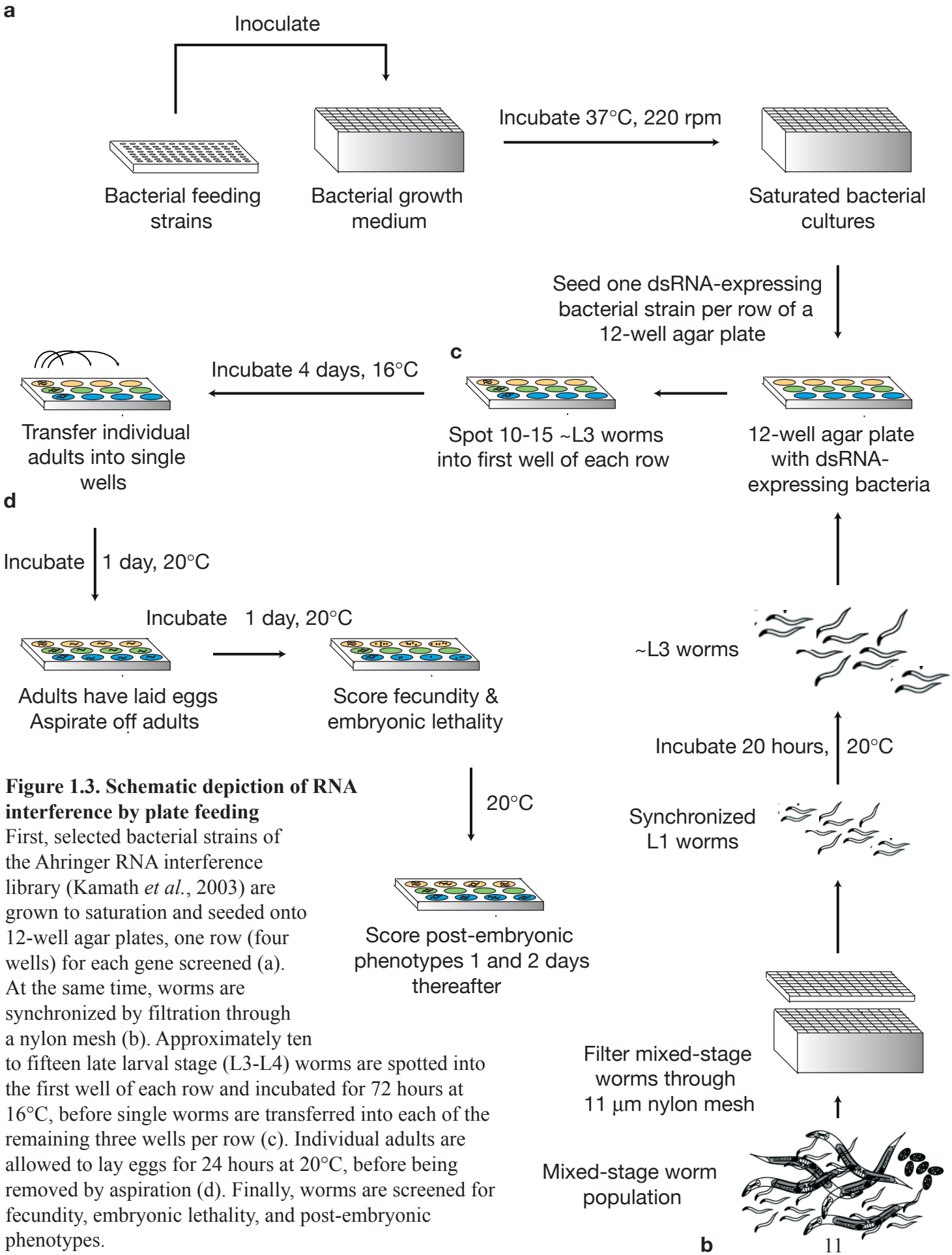

Figure 1.3. Schematic depiction of RNA interference by plate feeding
First, selected bacterial strains of the Ahringer RNA interference library (Kamath *et al.*, 2003) are grown to saturation and seeded onto 12-well agar plates, one row (four wells) for each gene screened (a). At the same time, worms are synchronized by filtration through a nylon mesh (b). Approximately ten to fifteen late larval stage (L3-L4) worms are spotted into the first well of each row and incubated for 72 hours at 16°C, before single worms are transferred into each of the remaining three wells per row (c). Individual adults are allowed to lay eggs for 24 hours at 20°C, before being removed by aspiration (d). Finally, worms are screened for fecundity, embryonic lethality, and post-embryonic phenotypes.

incubation time. Finally, individual adults and their progeny are screened for viability, fecundity, and post-embryonic growth and development. Using this methodology, phenotypes were assigned to 16,757 predicted *C. elegans* genes in wild-type worms (Fraser *et al.*, 2000; Kamath *et al.*, 2003).

1.6.2. Genome-wide RNA interference by bacterial feeding

When using the Ahringer library to target 16,757 predicted *C. elegans* genes in wild-type hermaphrodites, mutant phenotypes were detected for ~10% of targeted genes. While ~7% of genes were found to be essential for viability, knockdown of ~ 3% of genes generated worms with slowed post-embryonic growth or defects in post-embryonic development (Kamath *et al.*, 2003). By assessing the ability to correctly identify the known loss-of-function phenotypes of previously studied loci, the effectiveness of RNAi by bacterial feeding was determined to be, on average, ~64%. The detection rate was higher for loci with non-viable phenotypes (~78%) than for loci with post-embryonic phenotypes (~42%).

Notably, the results obtained by genome-wide RNAi by bacterial feeding correlated well with other previously reported large-scale studies, in which RNAi by injection and soaking, respectively, was used (Gonczy *et al.*, 2000; Maeda *et al.*, 2001). Thus, results obtained by RNAi are highly reproducible regardless of the method used. Most importantly, the false-positive rate of RNAi by bacterial feeding is below 1%, as assessed by targeting 225 genes that were known not to affect viability when deleted. RNAi against one gene only caused a mutant phenotype.

This initial screen was subsequently repeated in a genetic mutant deficient for the putative RNA-directed RNA polymerase RRF-3, which was found to be hypersensitive to RNAi in a forward genetic screen (Simmer *et al.*, 2002). Using the *rrf-3* background, mutant phenotypes could be assigned to an additional 400 genes, thereby increasing the percentage of *C. elegans* genes with a detectable RNAi phenotype to approximately 12% (Simmer *et al.*, 2003).

1.7. Using RNA interference in cell-based approaches for uncovering gene functions in *D. melanogaster* on a large scale

The completion of the *Drosophila* genome sequence (Adams *et al.*, 2000) together with the finding that addition of long dsRNAs to *Drosophila* cells was a potent method to reduce the expression of specific target genes (Clemens *et al.*, 2000; Hammond *et al.*, 2000) led to the assembly of a collection of dsRNAs targeting most of the roughly 13,600 predicted genes in the genome of *D. melanogaster* (Boutros *et al.*, 2004). When using this resource to target over 90% of predicted fly genes in cell-based assays, approximately 3% of genes were found to be indispensable for growth and viability in *D. melanogaster* (Boutros *et al.*, 2004). Recently, further progress has been made to refine the analysis of loss-of-function phenotypes generated by RNAi in cell-based approaches (Bjorklund *et al.*, 2006). Flow cytrometry has been used to study the role of ~70% of *D. melanogaster* genes in cell cycle progression. That way, numerous genes governing cell size, cytokinesis, apoptosis, and cell cycle progression were identified and ordered into known pathways. RNAi reagents have now also become available for genome-scale screens in mammalian cells (reviewed in Moffat and Sabatini, 2006). Thus, cell-based approaches similar to the ones taken for systematically deciphering gene function in *D. melanogaster* will help unravel gene functions on a genome-wide scale in mammals.

1.8. Uncovering gene functions in *S. cerevisiae* on a genome-wide scale

For the systematic unraveling of gene functions in the yeast *S. cerevisiae*, precise start-stop codon deletions were generated for each of the ~ 6,000 predicted genes (Giaever *et al.*, 2002; Winzeler *et al.*, 1999). These were constructed by targeted disruption of each predicted open reading frame (ORF) through homologous recombination (Baudin *et al.*, 1993; Lorenz *et al.*, 1995; Wach *et al.*, 1994). Targeting constructs for each gene were generated through polymerase chain reaction (PCR) amplification of a selectable marker gene, by using primers with a short sequence of homology to the predicted start and stop codons of each gene. In addition, unique 20-bp sequences (‘barcode’ sequences) — referred to as ‘up’ and ‘down’ tags (Hensel *et al.*,

1995; Shoemaker *et al.*, 1996) — flanking the marker gene and universal PCR priming sites just outside the barcodes were included in the targeting cassette. This design allows individual gene deletions to be identified in pools of mutant strains through PCR amplification using primers homologous to the universal sequence and subsequent hybridization of PCR products to a barcode microarray.

Mutant strains are maintained as a collection of heterozygous diploid strains, each of which is carrying a deletion in one copy of a specific gene, whereas the other copy is maintained as wild-type. The accomplishment of precisely deleting every single predicted ORF in the genome of *S. cerevisiae* paved the way for providing a genome-scale view of gene functions in yeast. When analyzing haploid meiotic progeny after induced sporulation, it was found that almost 5,000 of the predicted ~6,000 yeast genes can be eliminated entirely without any apparent deleterious consequences (Giaever *et al.*, 2002).

1.9. Identification of orthologous genes

To be able to transfer functional information between organisms, computational algorithms have been developed that enable the identification of orthologous genes in different species. Traditionally, phylogenetic trees have been constructed for the detection of orthologues (Yuan *et al.*, 1998). However, orthologue assignment by phylogenetic methods is difficult to automate and requires immense computing power. Thus, phylogenetic tree-based approaches are not suitable for assigning orthologues on a genome-wide scale.

The automatic clustering based on two-way best matches provides an alternative method (Chervitz *et al.*, 1998; Mushegian *et al.*, 1998; Rubin *et al.*, 2000; Wheelan *et al.*, 1999). Orthologue assignment, however, can be complicated through the existence of genes in multiple copies (‘co-orthologues’ or ‘paralogues’). While it might be desired to identify paralogues that arose after divergence of one species from another, it had initially not been possible to separate paralogues that predated species split from paralogues that arose after by using conventional clustering approaches based on two-way best pairwise matches.

To overcome this obstacle, an algorithm for the identification of orthologues and paralogues that arose after divergence of one species from another (so-called 'in-paralogues', to distinguish them from 'out-paralogues', paralogues that arose before species split) was devised (Remm *et al.*, 2001). This method (INPARANOID) is based on two-way best pairwise matches for detecting orthologues, with an algorithm added to identify in-paralogues between any two genomes. First, pairwise similarity scores are calculated using BLAST, with the bi-directional best hits being assigned the main orthologue pair. Adjustable cut-off values are applied to separate significant similarity scores from erroneous matches and thus to avoid inclusion of false positives. INPARANOID is based on the assumption that genes within one species that are more similar in sequence to the main orthologue than to any sequence from the other species represent in-paralogues (Figure 1.4.). Consequently, INPARANOID assigns confidence values (on a scale of 0% to 100%) as a measure of sequence similarity of a given in-paralogue to the main orthologue pair (which is assigned 100%).

By comparing orthologue assignments generated by INPARANOID with manual tree-based orthologue detection approaches, INPARANOID was demonstrated to generate data sets with a high degree of confidence.

Thus, while orthologues and in-paralogues have previously been detected through the construction of phylogenetic trees, a rather slow approach that is difficult to automate, alternative methods, based on two-way best pairwise matches, could not separate in-paralogues from out-paralogues. The INPARANOID algorithm, however, provides a powerful method for assigning orthologues and in-paralogues between any two species. Notably, INPARANOID is conservative and rather underpredicts orthologues and paralogues by excluding insignificant hits, though I recognize that this may exclude some true in-paralogues.

1.10. Gene dispensability and its potential underlying causes

The availability of whole genome-sequences for numerous model organisms together with the development of technological platforms has paved the way for the systematic investigation of gene functions on genome-wide scales (Bjorklund *et al.*,

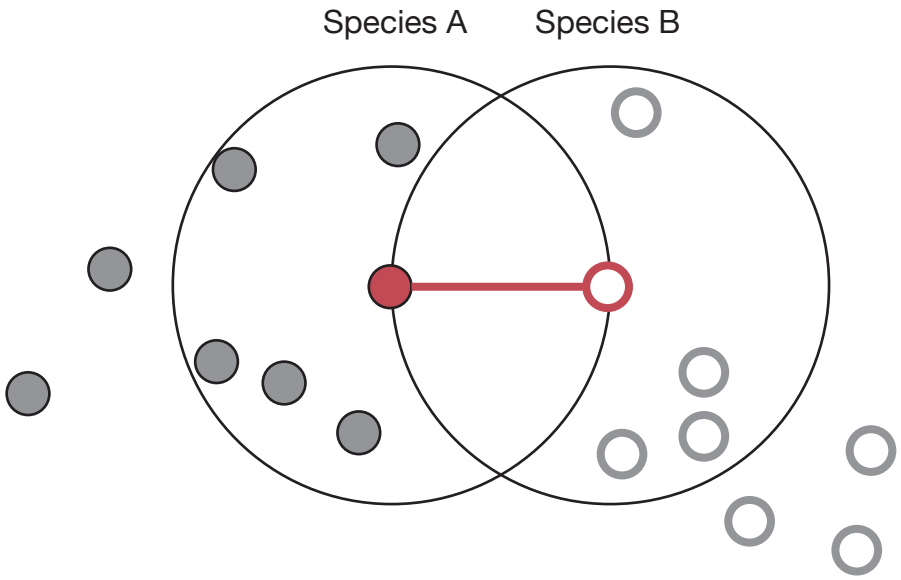

Figure 1.4. Graphical model for orthologue assignment using INPARANOID
Reciprocal best pair-wise matches of genes in two species (here species A and B) are assigned the main orthologue pair (shown as red filled and open circles, respectively). Additional orthologues within each species (co-orthologues) that are more similar in sequence to the main orthologue than to any other gene in the other species are assigned 'in-paralogues' (shown as grey filled and open circles within species A and B, respectively), whereas all other orthologues are assigned 'out-paralogues' (drawn outside the black circles). In-paralogues are considered co-orthologues that arose after divergence of one species from another, whereas out-paralogues represent co-orthologues that predate the species split. Re-drawn based on figure from Remm *et al.*, 2001.

2006; Boutros *et al.*, 2004; Giaever *et al.*, 2002; Kamath *et al.*, 2003). One striking insight gained from these systematic studies was the discovery that the great majority of genes encoded in eukaryotic genomes are dispensable for viability under laboratory conditions. Several theories have been put forward to explain this 'lack-of-phenotype' phenomenon; I will discuss these below.

Although intense theoretical work has been done on this subject, the causes of gene dispensability have remained controversial (reviewed in Wagner, 2005). One hypothesis suggests that apparently dispensable genes might be essential for survival only under specific environmental conditions that have not yet been probed in the laboratory. In this view, gene dispensability might reflect the system's ability to adjust to changing environments ('environmental adaptation'; Papp *et al.*, 2004).

Conversely, another theory postulates functional redundancy as genetic mechanism underlying gene dispensability (Wagner, 2000a; reviewed in Hartman *et al.*, 2001; Wagner, 2005). This view suggests the existence of 'back-up' mechanisms or compensatory pathways, such that loss of one gene has little effect because redundant or alternative pathways can counteract this loss ('mutational robustness').

Both theories suggest that the genetic networks that underlie viability are not constant, but can adjust their mode of operation under different external or internal conditions (reviewed in Kitano, 2004; Wagner, 2005). The two models for gene dispensability, however, might not be mutually exclusive. It has been increasingly noted that mutational robustness might have arisen as a 'by-product' of environmental adaptation (reviewed in Kitano, 2004; Wagner, 2005). I will discuss both theories in more detail below.

1.10.1. Environmental adaptation

While it has long been appreciated that seemingly dispensable genes can have important biological functions under specific environmental conditions, this notion has recently been investigated systematically through the computational modeling of metabolic networks in yeast (Papp *et al.*, 2004). When calculating the metabolic flux

under various nutrient conditions, roughly half of the apparently non-essential genes were predicted to be active under specific growth conditions. Recent experimental support for this computational approach came from large-scale phenotypic analyses, in which yeast deletion mutants were monitored for fitness defects under numerous environmental conditions. These studies led to the suggestion that at least 20% of the approximately 5,000 seemingly dispensable genes in *S. cerevisiae* might have essential functions under specific environmental conditions (Chang *et al.*, 2002; Davis-Kaplan *et al.*, 2004; Dudley *et al.*, 2005; Enyenihi and Saunders, 2003; Kuepfer *et al.*, 2005; Martinez *et al.*, 2004; Smith *et al.*, 2006). Most conditional essential genes were found to be indispensable under a limited number of growth conditions only, further demonstrating their environment-specific essentiality.

It had further been hypothesized that if seemingly dispensable genes were to have essential functions under specific environmental conditions, these genes might not be conserved in species that never encountered these conditions (Papp *et al.*, 2004). In agreement with this theory, enzymes that were predicted to be active under restricted nutrient conditions only by *in silico* modeling of the metabolic flux in yeast were found to have a limited phylogenetic distribution as compared to unconditionally active enzymes (Papp *et al.*, 2004). This computational study was supported by direct experimental evidence from studies on condition-specific genes in *E. coli* (Glasner *et al.*, 2003; discussed in Papp *et al.*, 2004). Furthermore, while the fraction of dispensable genes in yeast is approximated to be above 60%, the proportion of non-essential genes in the parasite *Mycoplasma genitalium* is estimated to be below 25%, suggesting a low number of condition-specific genes, which is consistent with the parasite's narrow host specificity (discussed in Papp *et al.*, 2004).

1.10.2. Functional redundancy and mutational robustness

One obvious source of genetic redundancy is through gene duplication. Duplicated genes that retain at least partially overlapping functions can confer robustness to mutation in the other copy (Force *et al.*, 1999; Lynch and Force, 2000). Gene duplicates can arise through homologous recombination and DNA repair processes, the

action of retrotransposons, or may be relics from whole-genome duplication events. Following a duplication event, both genes are initially performing the same function. However, for proper functions to be maintained, genes need to be under selective pressure. This notion led to the suggestion of various models for the evolutionary fate of duplicated genes.

The classical model for the evolution of gene duplicates predicts that the ancestral gene might be selected to maintain its original function, while its duplicate copy is free to diverge (Ohno, 1970). In this view, the most common fate of the duplicate copy was suggested to be extinction by the accumulation of deleterious mutations ('non-functionalization'; Kimura and King, 1979). Under this model, however, the duplicate copy, while it is not protected against mutations, might acquire new functions through rare beneficial mutations. This scenario has long been considered as the only mechanism for the preservation of gene duplicates and is regarded an important driving force for evolutionary innovation ('neo-functionalization'; Ohno, 1970).

The rationale that the probability for a gene to acquire a degenerative mutation might be higher than the likelihood to acquire a beneficial mutation (Lynch and Walsh, 1998), however, led to the proposal of an alternative model for the evolutionary fate of gene duplicates (Force *et al.*, 1999). This model ('duplication-degeneration-complementation', thereafter also referred to as 'sub-functionalization') suggests degenerative mutations as the central mechanism underlying duplicate gene preservation: Immediately after a duplication event, both copies might experience a period of relaxed selection, in which they might acquire deleterious mutations that impair different ancestral sub-functions. Such partially compromising mutations might result in a duplicate gene pair with complementary functions that can perform the original gene function in combination only. This sub-functionalization model is supported by findings from genome-wide studies, which reveal that, following a duplication event, both copies appear to be under selective pressure, with stronger purifying selection acting on older than on more recent gene duplicates (Kondrashov *et al.*, 2002; reviewed in Wagner, 2002). In addition, available experimental evidence suggests that sub-functionalization through complementary mutations frequently results in the partitioning of tissue-specific expression (Lynch and Force, 2000).

Various theoretical models have been proposed to explain how genes with redundant functions can be evolutionarily stable (Nowak *et al.*, 1997). One model predicts the maintenance of redundancy under the assumption that two genes perform their functions with equal efficiencies and under the same mutation rates. In this model, even if mutation rates differ marginally, thus rendering redundancy evolutionary unstable, it would still require a substantial amount of time until the overlap in function were eliminated, provided mutation rates are low. Conversely, two alternative theories suggest scenarios in which functional redundancy can be maintained indefinitely. While one model predicts the evolutionarily stability of redundancy based on the assumption that two genes perform the same function with different efficacies, with the more efficient gene experiencing higher mutation rates, the other theory relates pleiotropy to redundancy; two genes are maintained by selection because of their individual functions, while they are redundant with respect to another function.

Indirect experimental evidence for the redundant functions of duplicated genes comes from the systematic analysis of loss-of-function phenotypes of single genes: In both yeast and worms, it was found that inactivation of a duplicated gene is less likely to result in a non-viable phenotype than inactivation of a single-copy gene (Conant and Wagner, 2004; Gu *et al.*, 2003; Kamath and Ahringer, 2003). However, there are strong biases in the types of genes that are duplicated in genomes, which complicates the interpretation of these results (Castillo-Davis and Hartl, 2002) and at the time this study began, no attempt had been made to examine the extent of redundancy between duplicated genes *in vivo* directly and systematically.

The first systematic investigation into the mechanistic basis underlying gene dispensability was provided by an *in silico* analysis of the metabolic flux in *S. cerevisiae* (Papp *et al.*, 2004). Notably, while metabolic networks do not contain any redundant biochemical reactions — that is any metabolite is produced by one enzyme only —, approximately half of a system's unique enzymatic reactions can be perturbed without negatively affecting metabolic output (discussed in Wagner, 2005). Computational modeling of yeast metabolic fluxes was used to predict the effects of single-gene deletions under different environmental conditions and the concomitant changes in metabolic flux distribution (Papp *et al.*, 2004). Results from this *in silico* analysis

suggested that compensation for loss of individual gene function by a duplicate copy could account for roughly one quarter of dispensable genes in the yeast metabolic network. In addition, non-essentiality of fewer than 20% of metabolic genes could be explained by a re-routing of the metabolic flux through alternative, unaffected pathways. Remarkably, enzymes with unrelated activities can confer mutational robustness of metabolic networks by co-operation and flux re-organization.

Similar observations were made in developmental biology (reviewed in Wagner, 2005). Regulatory networks, such as the segment polarity network in *D. melanogaster*, do not comprise genes with equivalent functions. These networks, however, can maintain their functions despite perturbations (von Dassow *et al.*, 2000). Thus, genes that do not resemble one another at the sequence level and do not have related molecular roles can nonetheless compensate for loss of one another. This 'higher-order' functional redundancy is often referred to as 'distributed robustness', to distinguish it from genuine functional redundancy of gene duplicates ('redundancy of parts'; reviewed in Wagner, 2005).

1.10.2.1. Systematic experimental approaches for uncovering genetic interactions and functional genetic redundancy

While available evidence supports both conditional essentiality and functional redundancy as potential origins of gene dispensability, direct systematic experimental approaches are needed in order to examine the relative contribution of either source to the high proportion of seemingly non-essential genes. For the scope of this study, I will focus on experimental approaches that have been used to investigate genomes for functional genetic redundancy.

Traditionally, suppressor or enhancement genetics have been used to unravel functional relationships between genes (reviewed in Guarente, 1993; Hartman *et al.*, 2001). Genetic interactions were uncovered by screening with mutant alleles of known genes with defined phenotypes for mutations in other genes that can modulate this phenotype. That way, new genes with functions in the same or related molecular process were uncovered. Suppression, in which mutation in one gene alleviates the effects of

mutation in another gene, can unravel genes with roles in parallel biochemical or genetic pathways, if the second-site mutation increases pathway function ('bypass suppression'). Suppressor screens can also unravel genes functioning in the same pathway and have been instrumental in uncovering regulatory hierarchies. 'Synthetic enhancement', in which mutation in one gene worsens the effects of mutation in another gene, may occur between genes acting in the same biochemical pathways, or in distinct, but functionally redundant pathways. Synthetic lethality represents the most severe form of synthetic enhancement. This phenomenon was first observed by fly geneticists early last century who uncovered mutations in specific pairwise combinations of genes that resulted in lethality, whereas animals carrying mutations in each individual gene were viable (Dobzhansky, 1946; Sturtevant, 1956).

The identification of genetic interactions has provided major insights into key regulatory processes and pathway organization (Avery and Wasserman, 1992; Guarente, 1993; Hartman *et al.*, 2001; Lu and Horvitz, 1998; Thomas, 1993). However, before the availability of complete genome sequences and the feasibility of reverse genetic approaches, genetic interactions have been uncovered on a rather small scale to assist functional biological studies.

Considered a promising approach to uncover functional redundancy on a large scale, high-throughput technological platforms have been developed for the systematic mapping of synthetic lethal (SL) interactions in the yeast *S. cerevisiae*. I will discuss these in more detail below.

1.10.2.1.1. Identifying synthetic lethal interactions in the yeast *S. cerevisiae* on a large scale

Synthetic lethal (SL) interactions represent the most severe form of synthetic enhancement and thus are considerably straightforward to map. Therefore, thus far, most systematic studies focused on synthetic lethality as a framework for identifying non-additive genetic interactions.

Synthetic genetic array (SGA) analysis was the first methodology that had been developed for the genome-scale mapping of synthetic sick or synthetic lethal (SL) interactions in the genome of *S. cerevisiae* (Tong *et al.*, 2001). Using an automated approach, haploid strains carrying defined mutations are intercrossed and, following sporulation, fitness of double mutant meiotic progeny is assessed and compared to fitness of each single mutant. Therefore, a haploid strain of one mating type (α), carrying a mutation in a query gene of interest, is crossed into an array of haploid mutant strains of opposite mating type (a). Mutations in both strains are linked to different antibiotic resistance markers, allowing for selection of double mutant meiotic progeny. Importantly, to prevent mating of meiotic progeny — this would give rise to false-negative results — the query mutant strain is engineered to carry a selectable marker under a mating type a-specific promoter (*MFA1pr-HIS3*), permitting growth of meiotic progeny of one mating type (a) only. This approach has been pioneered by Amy Tong in Charlie Boone's lab and has since then been used extensively to systematically map SL interactions in *S. cerevisiae* (Tong *et al.*, 2004).

SGA analysis has further been extended to screen for genetic interactions between yeast essential genes (Davierwala *et al.*, 2005). Therefore, conditional expression alleles and temperature-sensitive (ts) conditional alleles of essential genes have been created. For the generation of conditional expression alleles, the endogenous promoter is replaced by a tetracycline (tet) – repressible promoter that can be shut off by the addition of doxycycline, a tetracycline analog (Mnaimneh *et al.*, 2004). Ts conditional alleles can be readily created by linking a heat-inducible destabilizing protein ('degron', Arg-DHFR(ts)) to the N-terminus of an essential gene, which results in protein degradation (Dohmen and Varshavsky, 2005). Using these approaches, extensive collections of tet-promoter mutants and degron alleles have been generated (Kanemaki *et al.*, 2003; Mnaimneh *et al.*, 2004). Furthermore, for the systematic construction of hypomorphic alleles, 3' untranslated regions of essential genes are replaced with selectable markers, which results in lower transcript levels through mRNA destabilization ('decreased abundance by mRNA perturbation, DAmP'; Schuldiner *et al.*, 2005). Recently, SGA technology has been coupled with quantitative approaches to allow genetic interactions to be identified more accurately (Schuldiner *et al.*, 2005).

Diploid synthetic lethal analysis by microarray (dSLAM) provides an alternative method to SGA analysis for uncovering SL interactions in the yeast *S. cerevisiae* (Pan *et al.*, 2004). In this approach, a deletion construct for a query gene of interest is transformed into a pool of barcode-tagged heterozygous deletion stains, which are carrying a selectable marker under a mating type-specific promoter (see SGA analysis). Following induced sporulation, DNA from pooled mutants is amplified by polymerase chain reaction (PCR) using a pair of primers binding common sequences outside the barcode tags (as discussed in 'Uncovering gene functions in *S. cerevisiae* on a genome-wide scale'). The abundance of each double mutant in the pool is determined by hybridization of PCR fragments to a barcode miniarray. This approach has been used to uncover SL interactions between genes with functions in the maintenance of DNA integrity (Pan *et al.*, 2006).

While the accuracy of identifying SL interactions by using SGA versus dSLAM approaches has not been assessed systematically, overlaps in data sets are found. Each method, however, also identifies SL interactions not uncovered by one or other approach. Beyond the differences in methodological approach, SGA technology uses haploid deletion strains, whereas dSLAM technology uses the yeast gene deletion collection in heterozygous diploid format. Notably, strains maintained as haploid deletion mutants might be subject to considerable selection pressure, which might lead to the accumulation of compensatory mutations. Results were found to differ when using homozygous haploid and diploid deletion mutants (Pan *et al.*, 2004). In contrast, maintaining mutant strains as heterozygous diploids reduces selection pressures through the existence of a wild-type copy for each deletion allele. Thus — with the exception of ~ 3% haplo-insufficient genes — the great majority of heterozygous yeast deletion mutants show normal growth on rich medium (Deutschbauer *et al.*, 2005).

1.10.2.1.2. Defining non-additive, 'synthetic' genetic interactions by a multiplicative model

For the unambiguous identification of non-additive, 'synthetic' genetic interactions, phenotypes are quantified and subject to statistical analysis. A

'multiplicative model' best describes the contributions of independent genetic loci to a phenotype; that is, the quantitative effects of mutations in individual genes should combine multiplicatively (Phillips *et al.*, 2000; Puniyani *et al.*, 2004). For additive genetic interactions, a double mutant's phenotype is expected to be the product of phenotypes for both individual genes. Divergence of the double mutant phenotype from the expected multiplicative values related to phenotypes for both individual genes is suggestive of a non-additive, synthetic genetic interaction. The multiplicative model also represents an application of Fisher's definition of 'epistacy', in which he describes epistasis as a phenomenon where the double mutant shows an unpredicted phenotype that is deviating from the expected product of phenotypes for both individual genes (Fisher, 1918). Deviations can have either negative or positive values, representing aggravating (fitness of the double mutant is lower than expected, with synthetic lethality as the most dramatic form) and alleviating (fitness of the double mutant is higher than anticipated) synthetic genetic effects. While aggravating interactions often occur between genes acting in distinct, but compensatory pathways ('between-pathway' interactions), alleviating interactions often reflect genes functioning in the same biochemical pathway ('within-pathway' interactions; Segre *et al.*, 2005; St Onge *et al.*, 2007). Notably, databases that store genetic interaction data for model organisms do often not discriminate between additive and non-additive (i.e. aggravating or alleviating) genetic interactions.

1.10.2.1.3. Inferring functional relationships from genetic interaction screens

When systematically mapping genetic interactions in the yeast *S. cerevisiae,* using synthetic genetic array (SGA) or diploid synthetic lethal analysis by microarray (dSLAM) technology, synthetic lethal (SL) interactions were found to be enriched between functionally related genes (Davierwala *et al.*, 2005; Pan *et al.*, 2006; Tong *et al.*, 2001; Tong *et al.*, 2004). Notably, the interaction density of genetic networks was found to exceed that of physical interactions (Tong *et al.*, 2004). Genetic interactions appear to be largely non-overlapping with physical interactions and tend not to occur between components of the same linear biochemical pathway, with the exception of essential

genes. This exception might be explained by the use of hypomorphic alleles when compiling the essential gene network — each mutant allele might partially reduce the flux through a linear pathway, but a combination of both mutant alleles might abolish pathway function (Bader *et al.*, 2003; Kelley and Ideker, 2005; Ye *et al.*, 2005). Most importantly, SL interactions can be used to infer pathway topology. Genes with similar SL interaction partners tend to encode components of the same biochemical pathway (Bader *et al.*, 2003; Kelley and Ideker, 2005; Ye *et al.*, 2005). Thus, genetic interaction and protein interaction networks provide complementary information. Accordingly, by using integrative approaches, such as combining SL data with protein-protein, mRNA co-expression, and phenotype data, the systematic mapping of SL interactions provides a powerful method for inferring functional relationships between genes (Wong *et al.*, 2004).

Together, these findings confirmed the classical interpretation of SL genetic interactions as the results of inactivating two functionally redundant pathways in the cell, either of which is individually dispensable (Figure 1.5.), and led to suggest the existence of abundant 'back-up' pathways conferring robustness to mutation in genetic networks (reviewed in Hartman *et al.*, 2001).

1.10.2.1.4. Features of genetic interaction networks

The systematic mapping of synthetic lethal (SL) interactions in *S. cerevisiae* shed light on the global properties of gene interaction networks. First, genetic interaction networks were found to follow a power-law distribution, with many genes interacting with few others, whereas few genes were found to have numerous interaction partners (Tong *et al.*, 2004). Second, genetic interaction networks appear to have a small-world topology, with short characteristic path length and densely connected local neighborhoods (Tong *et al.*, 2004). The average interaction density of genetic networks was found to be in the range of 1%, with a higher interaction frequency between essential than between non-essential genes (Davierwala *et al.*, 2005; Pan *et al.*, 2006; Tong *et al.*, 2004).

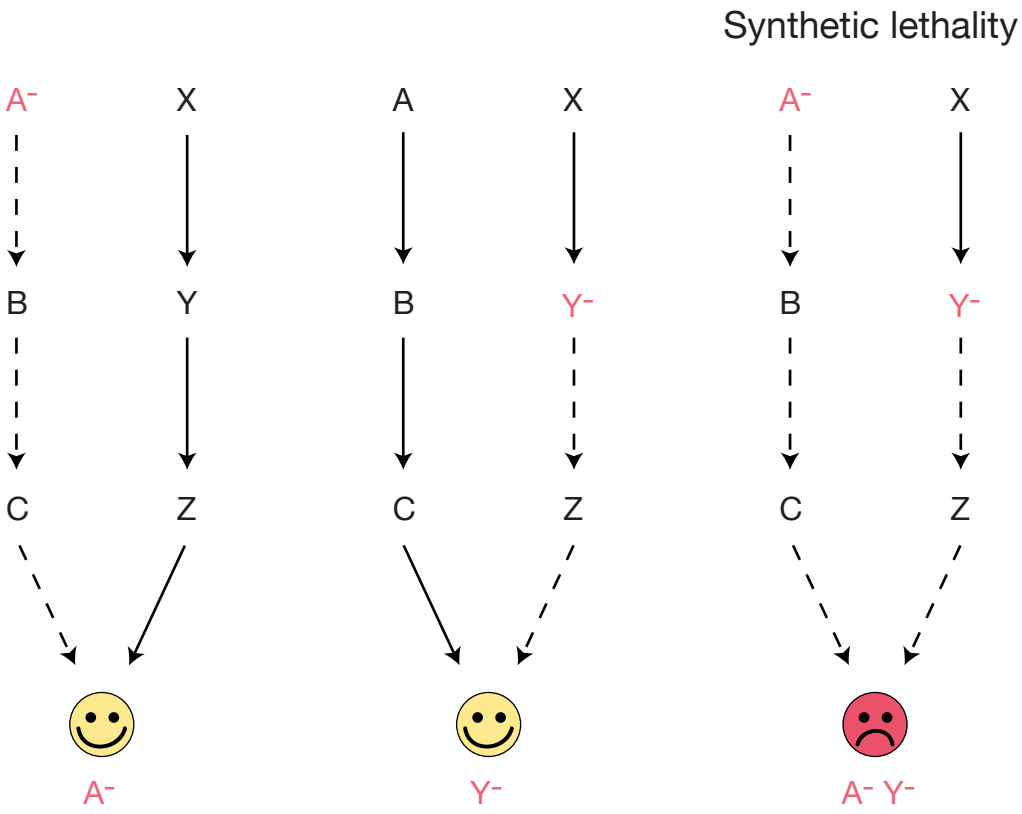

Figure 1.5. Model for the interpretation of synthetic lethal interactions
In the 'classical' model, synthetic lethal interactions are considered the result of inactivating genes with functions in two redundant essential pathways in the cell. In this view, loss of function of one pathway (here A-B-C) can be compensated for by a functionally redundant pathway (here X-Y-Z), and vice versa. Inactivation of both pathways through deleterious mutations (here depicted as A- and Y-, respectively, results in nonviability ('synthetic lethality') of the organism.

Remarkably, although SL interactions were enriched for genes encoding homologous proteins, their overall contribution was negligible (Tong *et al.*, 2004). Notably, however, the proportion of gene duplicates that had been sampled in large-scale SL screens had been fairly low. The majority of genetic interactions were uncovered between pairs of unrelated genes (Tong *et al.*, 2004).

Only recently has the contribution of gene duplicates to the robustness of yeast genetic networks been tested by systematic experimental approaches. Whereas the fraction of duplicate pairs that were found to genetically interact was significantly higher as compared to random pairs in the genome of *S. cerevisiae*, SL interactions were uncovered between ~25% of duplicate gene pairs only (Ihmels *et al.*, 2007). Notably, this fraction relates well to the proportion of duplicate gene pairs in the metabolic yeast network that were predicted to have compensatory capacities by *in silico* flux analysis (Papp *et al.*, 2004).

Taken together, results obtained both from systematic experimental approaches and computational studies led to the suggestion that although gene duplicates can — to some extent — provide robustness to mutation, their overall contribution to gene dispensability is limited. Conversely, most SL interactions take place between genes that do not share sequence similarity.

1.10.2.1.5. Condition-specificity of synthetic lethal interactions

Recently, a computational study has provided insights into the condition-dependence of synthetic genetic interactions. Flux balance analysis of the metabolic networks of *S. cerevisiae* was used to predict the effects of single- and double-gene deletions in 53 different nutritional environments (Harrison *et al.*, 2007). Notably, half of the predicted synthetic lethal (SL) relationships appear to be restricted to one or two nutrient conditions only, as compared to ~14% of SL interactions that appear to take place under all conditions investigated. Conversely, more than half of SL interactions remain undetected if examined in one nutritional environment only. These findings highlight the narrow condition-dependence of many SL interactions. Consequently, numerous SL interactions are likely to be missed by performing SL interactions screens

in one environmental condition only. A fraction of predicted SL interactions was tested experimentally, confirming the feasibility of the *in silico* approach. The study of condition-dependence of SL interactions in the yeast metabolic network was further extended to genetic interactions among non-metabolic genes. SL interaction data from literature were combined with available experimental data on the condition-dependence of single-gene deletion data. Remarkably, more than half of ~2,700 investigated SL interactions were found to take place between gene pairs of which either one or both appeared essential for viability under specific nutrient conditions. Furthermore, when assessing their phylogenetic distribution across species, members of SL pairs were not found to co-occur more frequently than random gene pairs. Thus, the condition-dependent restriction of SL interactions might be — at least for some interactions — explained by the individual essential functions of numerous genes under specific environmental conditions. Together, these findings support the notion that robustness to mutation might have arisen as a by-product during the evolution of adaptive responses to changing environmental conditions (reviewed in Kitano, 2004; Wagner, 2005).

1.10.2.1.6. Synthetic genetic interactions and disease

Enormous advances in genetics have uncovered numerous genes involved in human diseases. While many disorders are caused by mutations in single genes only, increasing numbers of human genetic diseases are identified to result from combinations of mutations in multiple genes (reviewed in Badano and Katsanis, 2002). Thus, many inherited mutations that alone have little effect can combine to result in severe defects. However, predicting and identifying such genetic interactions is a major obstacle in human genetics. Recently, synthetic lethal (SL) interactions have been systematically mapped in the yeast *S. cerevisiae* (Davierwala *et al.*, 2005; Pan *et al.*, 2006; Tong *et al.*, 2004). To date, however, approaches similar to the ones pioneered in yeast are not feasible in mammalian systems. Therefore, it is a major open question in genetics whether individual SL interactions are conserved between species and hence may be directly predicted in humans using interactions identified in simple model organisms. While functional studies in yeast have greatly contributed to our understanding of

individual gene functions in human (reviewed in Dolinski and Botstein, 2007), it remains to be determined whether SL interactions mapped in *S. cerevisiae* can be directly transferred into higher organisms and thus provide us with insights into complex human diseases.

In order to test the power of predicting candidate genetic interactions, comparative experimental approaches are needed. Ultimately, such a comparative study involves a tractable model system that is of higher complexity than yeast and that allows the analysis of systematic genetic perturbations. To date, *C. elegans* is the main animal model in which to carry out systematic functional studies *in vivo* in the context of a developing organism. The ability to inhibit gene function on a genome-wide scale by RNAi by bacterial feeding (Kamath *et al.*, 2003), together with its small size, its simple reproductive cycle and short generation time make the nematode *C. elegans* an ideal animal model system (Brenner, 1974) for the systematic study of synthetic genetic interactions. In particular, the comprehensive mapping of SL interactions in *C. elegans* would allow the comparison of *in vivo* genetic interaction networks between yeast, a single cellular organism and a multicellular animal. Such a study would have major practical implications for the use of yeast SL interaction data to predict SL interactions between genes in complex human diseases. Beyond the direct practical implications, this approach would also increase our understanding of the evolution and conservation of SL interactions.

1.11. Rationale of my study

Studies in *S. cerevisiae*, *C. elegans*, and *D. melanogaster* have shown that inactivation of most genes has little discernable effect on the organisms' fitness under laboratory conditions (Boutros *et al.*, 2004; Giaever *et al.*, 2002; Kamath *et al.*, 2003). These findings have led to various hypotheses explaining the observed high degree of gene dispensability. Whereas many of the apparently non-essential genes might have significant biological functions under specific environmental conditions ('environmental adaptation'; Papp *et al.*, 2004), an alternative theory suggests functional genetic redundancy as the principal source underlying the observed lack of phenotype upon loss

of individual gene function (‘mutational robustness’; Wagner, 2000a). This hypothesis proposes the existence of compensatory pathways as ‘fail-safe’ mechanisms to back-up essential biological pathways. While several lines of evidence support both the ability to adapt to new environmental conditions and the use of genetic compensatory pathways as potential sources underlying the apparent dispensability of a high proportion of genes in eukaryotic genomes, considerably more attention has been paid to the latter mechanism. This is not least because of the likely implications of non-additive, synthetic genetic interactions in complex human diseases (reviewed in Badano and Katsanis, 2002). Recently, enormous efforts have been made to systematically map synthetic lethal (SL) interactions in *S. cerevisiae* (Davierwala *et al.*, 2005; Pan *et al.*, 2006; Tong *et al.*, 2004). These studies appear to have uncovered thousands of gene pairs with redundant functions. However, at the time my study began, the systematic mapping of genetic interactions had not been extended to higher organisms.

The aim of my research was to begin to investigate functional redundancy in the genome of *C. elegans*. Therefore, I wished to systematically map SL interactions in the worm. To do so, it was crucial to establish protocols for simultaneously perturbing the functions of any pairwise combination of genes by using RNAi. First, I sought to investigate whether — and to which extent — *C. elegans* gene duplicates share redundant functions. Second, I wished to study functional redundancy in complex gene networks. In particular, I sought to investigate whether general modes of genetic redundancy are conserved between *S. cerevisiae* and *C. elegans* and whether I can find evidence for parallel pathways and back-up mechanisms in the worm. Therefore, I set out to investigate whether individual SL interactions are conserved between *S. cerevisiae* and *C. elegans*.