
Introduction to genetics and genomics*

Jonathan Hodgkin[§], Genetics Unit, Department of Biochemistry,
University of Oxford, Oxford OX1 3QU, UK

Caenorhabditis elegans owed much of its initial appeal as an experimental organism to the powerful genetic methods that could be developed for it. A further and major increase in popularity came in 1998, when it became the first multicellular organism (and only the second eukaryote, after *Saccharomyces cerevisiae*) for which a complete genome sequence was obtained. Annotation and exploitation of the worm's 100 Mb genome sequence is correspondingly more advanced than for most other eukaryotic genomes.

The interaction between genetic and genomic approaches continues at ever more sophisticated levels, and both approaches underlie most of current *C. elegans* research, so the associated techniques, principles and thinking can be encountered in almost every chapter of WormBook. Moreover, in both Genetics and Genomics, methodology and interpretation are intertwined, and as a result some of the central subjects of genetic analysis, such as gene mapping, are dealt with from a more practical standpoint in the methods section (WormMethods). This section is more concerned with general results and conclusions. Several subjects, such as mutagenesis, epistasis and pleiotropy (for Genetics) and chromosome organization, repeated sequences and promoter analysis (for Genomics) are only touched on briefly, although they deserve whole chapters.

In terms of genetic analysis, a fundamentally attractive feature of *C. elegans* is the ease of generating mutations. Several chemical mutagens, in particular EMS (ethyl methane sulfonate), work very efficiently on the worm, and mutations can also be generated with ionizing radiation and by transposon hopping and other means. *C. elegans* is diploid, so very deleterious mutations can be induced and propagated without killing the animal. However, the major mode of reproduction, by hermaphrodite self-fertilization, means that the effect of making any mutation homozygous can be automatically examined, as a result of Mendelian segregation. The efficiency of mutagenesis and of recessive phenotype screening, and also the ease of preserving mutant lines as archived frozen stocks, means that multiple mutations exist for many genes, with the number available growing all the time. The availability of multiple mutations is important in providing not just knockout mutations, which completely remove the activity of a gene, but also partial or conditional mutations, which allow the dissection of complex gene functions. The vast majority of these mutations are recessive, associated with partial or complete loss of gene function, but many rare gain-of-function mutations have also been generated, which have often provided essential tools for further investigation. Most recently, RNAi has provided a new way to manipulate gene activity.

The section on Genetics starts with a brief chapter, “Karyotype, ploidy, and gene dosage” (Jonathan Hodgkin), which summarizes some of the global features of the *C. elegans* genome from a genetic perspective, and reviews some of the gross alterations in karyotype that have been generated and analysed to date. The chapter also reviews the related subject of gene dosage, and the degree to which the animal can tolerate alterations in the dosage of single and multiple genes.

*Edited by Philip Anderson. Last revised April 25, 2005. Published September 6, 2005. This chapter should be cited as: Hodgkin J. Introduction to genetics and genomics (September 6, 2005), *WormBook*, ed. The *C. elegans* Research Community, WormBook, doi/10.1895/wormbook.1.17.1, <http://www.wormbook.org>.

Copyright: © 2005 Jonathan Hodgkin. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

[§]To whom correspondence should be addressed. E-mail: jah@bioch.ox.ac.uk

The second chapter, "Essential genes" (Kenneth Kemphues), discusses what might be regarded as the most important set of genes in the animal, without which it is either dead or sterile. Much work, using both classical genetic methods and more recently using RNAi, has gone into identifying, counting and analyzing such genes. Present estimates suggest that less than 25% of the genes have an essential function on their own. However, estimates of essential gene number are compromised by the knowledge that there is substantial gene redundancy in *C. elegans*, probably more so than in most other model organisms. This subject is discussed in the following chapter, "Gene duplications and genetic redundancy in *C. elegans*" (Alison Woollard). Levels of gene duplication appear to be higher in *C. elegans* than in *Drosophila* or yeast, which may sometimes complicate analysis but may also assist understanding, if the multiple functions of a single complex gene in an ancestral organism have been partitioned between two or more genes in *C. elegans*.

The four following chapters in this section address some of the major analytical methods for investigating the properties of mutations, and exploiting them in order to understand the underlying biology. The chapter on "Complementation" (Karen Yook) discusses an important first step in the mutational analysis of any phenomenon: testing sets of mutations to see if they affect the same or different genes, and thereby deriving information about the genetic complexity underlying any particular trait. Also, anomalous complementation patterns have sometimes been encountered in *C. elegans*, as in most well-studied genetic organisms, and the phenomenology of some of the more instructive cases is reviewed.

Further analysis of any mutation is greatly aided by the ability to search for interacting genes by means of modifier screens. These are discussed in the next two chapters, one on "Genetic suppression" (Jonathan Hodgkin) and one "Genetic enhancers" (Robert Herman and John Yochem). Suppression has been an extremely powerful tool in the arsenal of genetic methods for *C. elegans*, partly because the properties of self-fertilization and rapid growth make it so easy to search for suppressors of any given mutation in the worm. Different mechanisms of suppression are described in the chapter on "Genetic suppression," along with examples of how suppression analysis has been used in the dissection of molecular assemblies and signal transduction pathways.

The chapter on "Genetic enhancers" deals with the other side of the coin. Enhancement effects between mutations in different genes are commonly encountered, especially in the context of redundancy, and the possible kinds of enhancement effect and their implications are reviewed. Increasingly, genetic enhancers will be discovered as the result of specific screens for modifiers of particular mutant phenotypes in sensitized genetic backgrounds, an approach that has proved very powerful in both *Drosophila* and *C. elegans*.

Yochem and Herman also contribute a chapter on "Genetic mosaics," which provide a different kind of method for exploring the properties of genes, generating information about sites of expression and autonomy of action. Mosaic analysis in *C. elegans* has some unique aspects, because the invariance of the lineage and its complete description can be used to infer the genotype of every cell in a mosaic animal, allowing more rigorous interpretation of when and where genes have acted.

The chapters in the Genomics section are intended to provide overviews rather than exhaustive catalogues. Also, as noted, many related topics are discussed from a genomic perspective, in parts of chapters in other sections.

The worm's genome is currently unique amongst those of multicellular organisms in being truly finished, down to the last nucleotide. All other animal and plant 'complete' genomes still contain many megabases of undetermined sequence, mostly centric heterochromatin. *C. elegans* is in a different league, because its chromosomes are holocentric and each could therefore be fully sequenced, from telomere to telomere. Of course, a few mistakes must still persist even in the most recent version of the *C. elegans* sequence, but these errors are being steadily whittled away by improved annotation, comparative genomics, piecemeal re-sequencing and SNP detection. In addition, there are uncertainties such as the exact number of repeats in loci such as *rrn-1* and *rrs-1*, which contain tandem arrays of ribosomal RNA genes. There may well be individual variation in these repeat arrays, and there is certainly variability from cell to cell in the exact number of telomere repeats on each chromosome. Nevertheless, Chromosome V of *C. elegans* (20,921,718 basepairs in length, excluding telomeres) is the largest DNA molecule ever to be exactly and completely sequenced, and it is likely to retain that status for some time to come. The perfection of the whole genome sequence allows correspondingly more detailed annotation than is possible with other multicellular organisms.

The genome of *C. elegans* is very information-rich: about 25.5% appears to be protein-coding, and genes (including introns, 5' and 3' untranslated regions and regulatory regions) must take up altogether at least 40%. The chapter by John Spieth and Daniel Lawson, "[Overview of gene structure](#)," addresses general features of gene size and gene organization, including alternative splicing. A significant number of genes are arranged in operons, which are reviewed at greater length elsewhere in WormBook. ("[Trans-splicing and operons](#)," by Thomas Blumenthal). Spieth and Lawson also discuss the question of pseudogene number, for which a variety of estimates have been given. Certainly the worm contains some pseudogenes, particularly in large gene families that have undergone recent expansion, but it remains unclear what the real number is. Processed pseudogenes, which are very common in vertebrate genomes, are infrequent.

Most of the genes so far identified encode protein, and many of these fall into recognizable large families. The chapter by Erich Schwarz, "[Genomic classification of protein-coding gene families](#)," reviews the tools used for assigning proteins to different families, and provides lists and charts of the larger gene families. Some of these families are reviewed in specific chapters in other sections ("[Sperm motility and MSP](#)," by Harold Smith; "[G protein coupled receptors](#)" by Hugh Robertson and James Thomas).

The chapter by Sean Eddy et al., "[C. elegans noncoding RNA genes](#)," summarizes ribosomal RNA and transfer RNA from a genomic perspective, and moves on to review splice-leader RNAs, sno-RNAs, microRNAs, spliceosomal RNAs and other functional RNA species. Although microRNAs were first discovered in *C. elegans*, and have been extensively studied in this organism, it seems likely that the full complement and impact of these important regulators and other non-coding RNA molecules have yet to be defined.

A substantial part of the *C. elegans* genome is made up of transposons, according to the usual pattern in animal and plant genomes. The chapter on "[Transposons](#)" by Jean-Louis Bessereau reviews what is known about the various *C. elegans* transposon families. Transposons have provided important tools both for the analysis of *C. elegans* biology, and for studying transposon properties in general.

Lastly, Bernard Lemire's chapter on "[Mitochondrial genetics](#)" reviews what is known about the structure and function of the small mitochondrial genome of *C. elegans*. Studies of mitochondrial copy number, mitochondrial deletions and interactions with the nuclear genome are also discussed.

In conclusion, one can predict that both genetic and genomic investigations will continue to be pursued vigorously in *C. elegans* research, along well-established lines, but new discoveries and new techniques are already changing and expanding the possible kinds of research. In particular, the use of RNAi is providing a whole new dimension to the kinds of experiments and screens that can be done. Other new techniques will also become more important, such as the ability to carry out efficient gene replacement, and the use of more sophisticated transgene tools, which should allow both more precise control of gene expression, and better *in vivo* assays of gene expression and function. From a genomic point of view, the near-complete genome sequences that are being generated for an increasing number of close relatives of *C. elegans* (*Caenorhabditis briggsae*, *Caenorhabditis remanei*), other *Caenorhabditis* species and more distant cousins such as *Pristionchus pacificus* are supplying invaluable confirmation, new information and new insights, and the already extraordinary power of comparative genomics can only increase in the future.



All WormBook content, except where otherwise noted, is licensed under a [Creative Commons Attribution License](#).