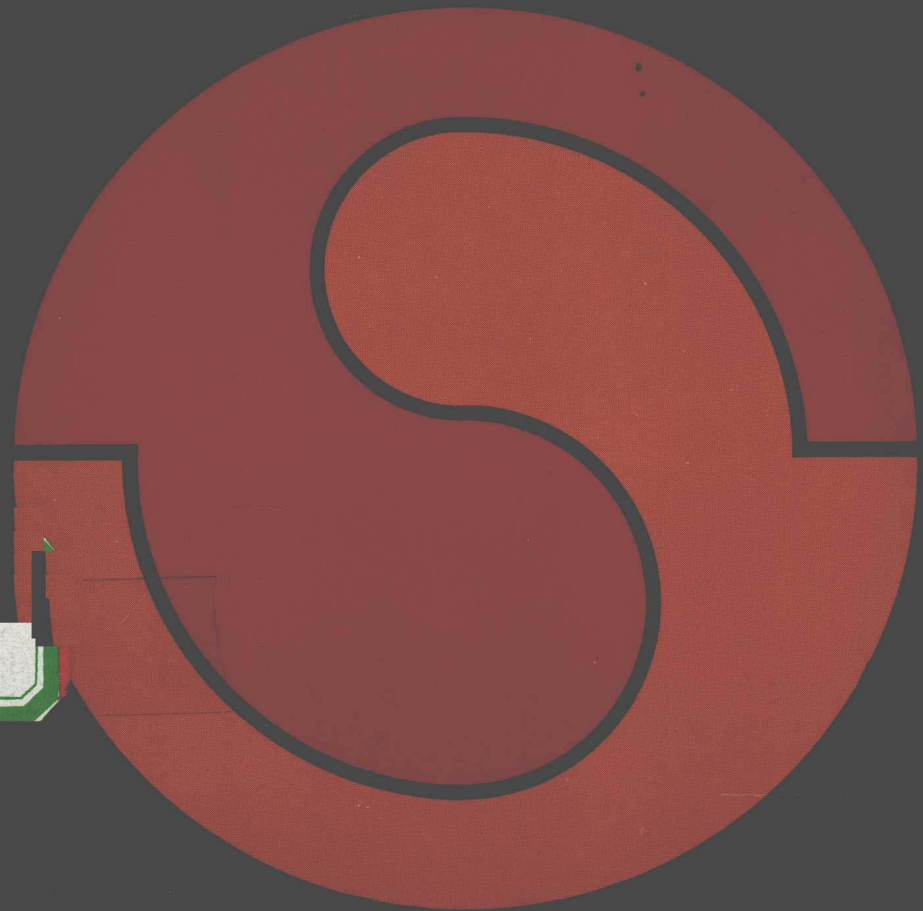
Outline Studies in Biology

Human Genetics

J.H. Edwards



List of titles

Already published

Cell Differentiation	J. M. Ashworth
Biochemical Genetics	R. A. Woods
Functions of Biological Membranes	M. Davies
Cellular Development	D. Garrod
Brain Biochemistry	H. S. Bachelard
Immunochemistry	M. W. Steward
The Selectivity of Drugs	A. Albert
Biomechanics	R. McN. Alexander
Molecular Virology	T. H. Pennington, D. A. Ritchie
Hormone Action	A. Malkinson
Cellular Recognition	M. F. Greaves
Cytogenetics of Man and other Animals	A. McDermott
RNA Biosynthesis	R. H. Burdon
Protein Biosynthesis	A. E. Smith
Biological Energy Conservation	C. Jones
Control of Enzyme Activity	P. Cohen
Metabolic Regulation	R. Denton, C. I. Pogson
Plant Cytogenetics	D. M. Moore
Population Genetics	L. M. Cook
Insect Biochemistry	H. H. Rees
A Biochemical Approach to Nutrition	R. A. Freedland, S. Briggs
Enzyme Kinetics	P. C. Engel
Polysaccharide Shapes	D. A. Rees
Transport Phenomena in Plants	D. A. Baker
Cellular Degradative Processes	R. J. Dean
Human Evolution	B. A. Wood
Human Genetics	J. H. Edwards

In preparation

The Cell Cycle	S. Shall
Microbial Metabolism	H. Dalton, R. R. Eady
Bacterial Taxonomy	D. Jones, M. Goodfellow
Metal Ions in Biology	P. M. Harrison, R. Hoare
Muscle	R. M. Simmons
Xenobiotics	D. V. Parke
Biochemical Systematics	J. B. Harborne
Biochemical Pharmacology	B. A. Callingham
Biological Oscillations	A. Robertson
Photobiology	K. Poff
Membrane Assembly	J. Haslam

OUTLINE STUDIES IN BIOLOGY

Editor's Foreword

The student of biological science in his final years as an undergraduate and his first years as a graduate is expected to gain some familiarity with current research at the frontiers of his discipline. New research work is published in a perplexing diversity of publications and is inevitably concerned with the minutiae of the subject. The sheer number of research journals and papers also causes confusion and difficulties of assimilation. Review articles usually presuppose a background knowledge of the field and are inevitably rather restricted in scope. There is thus a need for short but authoritative introductions to those areas of modern biological research which are either not dealt with in standard introductory textbooks or are not dealt with in sufficient detail to enable the student to go on from them to read scholarly reviews with profit. This series of books is designed to satisfy this need. The authors have been asked to produce a brief outline of their subject assuming that their readers will have read and remembered much of a standard introductory textbook of biology. This outline then sets out to provide by building on this basis, the conceptual framework within which modern research work is progressing and aims to give the reader an indication of the problems, both conceptual and practical, which must be overcome if progress is to be maintained. We hope that students will go on to read the more detailed reviews and articles to which reference is made with a greater insight and understanding of how they fit into the overall scheme of modern research effort and may thus be helped to choose where to make their own contribution to this effort. These books are guidebooks, not textbooks. Modern research pays scant regard for the academic divisions into which biological teaching and introductory textbooks must, to a certain extent, be divided. We have thus concentrated in this series on providing guides to those areas which fall between, or which involve, several different academic disciplines. It is here that the gap between the textbook and the research paper is widest and where the need for guidance is greatest. In so doing we hope to have extended or supplemented but not supplanted main texts, and to have given students assistance in seeing how modern biological research is progressing, while at the same time providing a foundation for self help in the achievement of successful examination results.

J. M. Ashworth, Professor of Biology, University of Essex.

Human Genetics

J. H. Edwards

Professor of Human Genetics,
Infant Development Unit,
Birmingham Maternity Hospital
Birmingham

LONDON
CHAPM

A Halstec
John Wiley & Sons, New York

First published in 1978
by Chapman and Hall Ltd
11 New Fetter Lane, London EC4P 4EE

Typeset by Preface Ltd., Salisbury, Wilts.
and printed in Great Britain at the
University Printing House, Cambridge

ISBN 0 412 13170 6

Distributed in the USA
by Halsted Press, a division
of John Wiley & Sons, Inc., New York

Library of Congress Cataloging in Publication Data

Edwards, John Hilton.
Human genetics.

(Outline studies in biology)
"A Halsted Press Book."
Includes bibliographical references and index.
1. Human genetics. I. Title.
QH431.E35 573.2'1 77-21582
ISBN 0-470-99096-1

Contents

Introduction

The awareness that children resemble their parents probably arose early: it was certainly obvious to the authors of Genesis and assumed to be self-evident to the audience of Homeric epics.

The mechanism by which these resemblances were manifested, and the extent to which inborn variation could be either amplified or suppressed by natural environments, or by such artifical environments as those imposed by education or medical interference, remained wholly unknown until the major components of these mechanisms had been observed or inferred.

In human genetics, as in astronomy, the natural curiosity of man about his past and future led to various extensive beliefs, varying from the flatness of the earth and a belief in horoscopes to the irrelevance of biological parentage to intelligence, behaviour or disease.

This book aims to provide a basic introduction to the mechanism of heredity, and to the ways in which these mechanisms may fail and lead to disease.

The emphasis is on man, but since the basic mechanisms of man differ little from those of other mammals, it is hoped that the book may be of some use to those whose primary interests are related to the mouse or the farmyard. It is based on a course of lectures given to the medical students in the University of Birmingham; this course included lectures on cytogenetics by Dr. Alan McDermott, who has written in this series.

In general the approach will be mechanistic and deductive, with the main emphasis on the necessary consequences of self-reproducing and self-copying genetic elements which, in 1978, we may regard as being 'known' to exist.

This is not the most usual way to develop this subject, since medical students traditionally start on the adult in the dissecting room and work from the obvious to the unseen. However, the time now seems ripe to follow geometry, astronomy, physics and chemistry, in defining a basic set of components, and building from these until, like a child with a set of Lego, recognisable structures with self-evident properties can be developed.

1 The units of inheritance

A simple model of inheritance is to suppose a large number of units, each determining some character, such as eye-colour, or some part of a character, such as the adding of a hundredth of a cubit to the stature, which merge at conception, and then split at reproduction. This model, which was implicit in much early work, adequately explained the similarity of parent and child.

This blending of the hereditary units did not explain the dissimilarity of brothers and sisters, or the persistence of hereditary variation. Mendel's model in which the units, or, factors as he called them, remained distinct between conception and reproduction, and were passed on intact and independently, each unit having a half-chance of entering each gamete, had the important distinction that only their effects were 'blended', and then only sometimes. This provided a model in which unchanging units could maintain a stable variation, and explained both the similarities and the dissimilarities of children, or, in Mendel's case, of the progeny of the sweet-pea.

Mendel also advanced the concept of unequal effect, so that where two different factors were inherited, the effects of one might so dominate the other that its partner would not be evident. He called these effects dominance and recessivity, and used the adjectives dominant and recessive for the factors.

He introduced the use of upper and lower case letters, the upper being dominant, and introduced the diagram (Fig. 1.1) by which 'factors' can produce 'effects'.

This diagram, in which two gametes lead to all possible pairs of alleles, can also be used backwards, so that all possible pairs of genes can lead to two types of gametes (a and A). The output gametes, unlike the input gametes, would of course both be male or both female.

Mendel summarized this in two 'laws'. The first stated that any gamete produced by an organism is equally likely to contain the contribution (factor) from either the male or female gamete that gave rise to that organism. That is, there is an equal splitting (spaltung) of the genetic material, a

		Female contribution		
		A	a	
Male contribution	A	AA	Aa	**Fig. 1.1 Mendel's diagram**
	a	aA	aa	

word which Bateson intentionally mistranslated as segregation, since he could not accept the premise of passivity. Mendel's second law stated that pairs of loci are also conveyed independently, a law which is only true of loci not mechanically associated by lying on the same chromosome, that is, of loci which are not linked. As all advanced organisms have many chromosomes, most pairs of loci are unlinked.

The two 'factors' A or a coexist after fertilization; when the effect of A is manifest independently of a, the organisms with AA, Aa and aA being indistinguishable when A is dominant to a.

This model was not applied widely until early this century, over forty years after publication in an obscure journal. The model was at first considered incapable of explaining graded or continuous variation, such as height; this inadequacy was removed by Fisher's paper of 1918 which showed that a model with many independent factors could explain with precision the inheritance of height in man.

The nature of the 'factors' was inferred to be that of producers of 'ferments' by Garrod, an English physician; the 'one gene one enzyme' model was later substantiated by direct experiments on *Neurospora*, a fungus. The development of microscopy, both optical and electronic, and of methods for the separation, detection, and characterization of both the genetic determinants, or factors, and their products, has now provided us with an elaborate hierarchy of genetic elements within which Mendel's factor, now termed a 'gene', may be regarded as the unit of protein synthesis. As each gene has a position or locus, and as variant forms (known as 'alleles') usually exist, it is convenient to avoid the word 'gene' wherever possible, since it is often used ambiguously for either locus or allele, and to speak of loci and alleles. Fig. 1.2 shows the hierarchy of genetic elements: the numbers should not be regarded as exact (in man there are 23 chromosomes per haploid set, 20 in the mouse, 7 in a muntjak and 80 in the rhinoceros – the least and the most in numerical terms). All mammals have, within the precision of the methods, the same amount of DNA, between 3 and 4×10^9 nucleotide pairs per gamete.

In addition there are genetic elements of unknown origin or function composed of distinct varieties of DNA, distributed irregularly among the majority type and representing some 3% of the total in man, and some 10% in the mouse. The finding of these fractions, which vary greatly between related species, is the major genetical discovery of the last decade.

These fractions are known as satellites, since they usually descend alongside the main type of DNA at ultra-centrifugation, due to having different proportions of the four nucleotides, which are all of different molecular weight. They will not be discussed further since they have not, as yet, been related to either normal or morbid variation; it is not even clear whether they are functional and necessary components, as opposed to ingenious parasites which have imposed themselves on the thread of life.

Self-reproducing elements of DNA are also found in such organelles

as mitochondria; again we will ignore these in this outline since no consequence of their variation is known. Mitochondria appear to breed true, although many, if not most, of their associated enzymes are coded by nuclear DNA. Most, if not all, mitochondria are of maternal origin: it is not known if the few mitochondria at the root of the sperm tail survive conception.

1.1 The nucleotides

The elements of DNA, like the units of a zip-fastener, have a unit, the base, which mates with its inverse, and a structural component which holds the base in line through the desoxyribose segment being coupled to its neighbour by a phosphate bond. A zip-fastener contains only two types of connection, a bump and a dent, which alternate, each unit having both a bump and a dent. If units were flat on one side, and had either a bump or a dent on the other, a simple structure would exist in which the strands could be separated and then each partner reconstructed from a set of spare units of each kind, leading to two identical 'zips'. The process would be repeated indefinitely and, for a zip of n elements, there would be as many as 2^n different variations.

The genetic code contains four distinct elements, adenine, thymine, guanine and cytosine, or A, T, G and C; since thymine and cytosine have pyrimidine bases and adenine and guanine purine bases, and the former are smaller, having one ring, while purines have two, we may write these as:–

A t G c

The unconventional use of small letters defines the purines. Compatible pairs of nucleotides are those in which one is a pyrimidine and one a purine. There are also differences in bonding, cytosine and guanine having three hydrogen bonds and adenine and thymine only two; we may therefore write the code as

A t G c where heavy type defines a triple bond.

A stable pairing is only possible if the units have the same total length (that is, since pyrimidines and purines are of different sizes, if there is one of each) and if the hydrogen bonds are equal in number between each pair so that, given one strand of:–

A t G c g A t c t,

the other strand must be:–

t A c G c t A g A,

and each strand can therefore replicate, and, if the replica replicates, the original is reproduced. Replication implies a distinct inverse copy, as in printing or photography, and is not a synonym for reproduction. This is the genetic code. The information content for n units can be as high as that implied by 4^n different possibilities, or, in 'bits', where a 'bit' is the

information implicit in a single choice between two possibilities, up to $2n$.

This structure, which may be visualized as a zip fastener, or as a ladder whose rungs are the bases and whose supports the backbone of phosphate linkages, is folded into a helix, the 'double helix'. However, at an elementary level, its properties are adequately covered by a linear zip or ladder model.

1.2 The gene

The nucleotide sequence is both informative, through its variation (the rungs of the ladder), and coherent, through its uniform phosphodiester linkages. In organisms with a naked chromosome, or chromoneme, the gene is defined syntactically by the information it contains, rather than being a distinct functional unit on a structural basis, as in the 'bead on a string' model of earlier days.

Even though there is no evidence of any structural discontinuities between genes, it is often convenient to retain the word locus for the site of this functional unit.

The discontinuity implicit in the gene is manifest at the functional level, since a continuous string of genic material, or DNA, will direct the synthesis of a series of discontinuous strings of amino-acids, the proteins.

We may define the units coding for proteins as genes. Proteins are usually 200–300 amino-acids long, or compounded of units of this length, and three nucleotides are needed to specify an amino-acid, some 600–900 nucleotides are needed for synthesis, in addition to others for starting, stopping and other poorly understood functions. The haemoglobin coding segment for rabbit haemoglobins is known to be at least 200 nucleotides longer than the 450 necessary for the synthesis of 150 amino-acids.

Since each gamete conveys over 3×10^9 nucleotide pairs, and most proteins require less than 10^3 for their coding, there is room for 3×10^6 or three million genes. Even if generous allowance is made for 'structural' elements allowing for chromosomal pairing, for inert spacers, for repeated units synthesizing ribosomal RNA, and for satellites, there is adequate room for a million genes – or for 20 000 on the smallest human chromosome.

This may seem a somewhat generous supply of proteins but is by no means impossible; most of these would be involved in subordinate or minor activities, and many, like words in a large dictionary, would be of purely historical interest. Furthermore, many genes may be duplicated, or coded at multiple sites, as is known to occur in human haemoglobin.

1.3 Structural units

In bacteria many examples are known in which the enzymes involved in serial metabolic actions are coded at a series of neighbouring loci, and the whole genic sequence of protein-forming genes, sometimes called structural genes, is under the integrated control of a connected series of 'regulator genes'. Only a minority of genic loci are integrated in this way.

Few examples of integration between the location of genes defining associated functions have been found in mammals, and many counter-examples are known. The chains of the main haemoglobin in both fetus and adult are coded from distant loci: in mice they are known to be on distinct chromosomes. The heavy and light immunoglobulin chains are similarly distinct.

On the main glucose pathway phosphogluco-dehydrogenase is coded on chromosome 1, glucose-6-dehydrogenase on the X-chromosome.

Doubtless some coordinated and serial systems have retained the relationship of their remote past, but few are known. In the absence of knowledge of 'regulator' genes the term 'structural gene' is unnecessary. Unless otherwise specified, gene, or allele, will imply a unit coding for a protein.

1.4 Repeated units

Substantial numbers of ribosomes, the standardized organelle on which proteins are coded, are needed in all cells, and are replicated directly from the DNA from numerous sites, at each of which many identical segments of DNA are present. Transfer RNAs are similarly coded from repeated units. There is also evidence of other repeated units, including ones too short to be able to code a protein. These short units are particularly mysterious, since they are both dispersed throughout the genetic apparatus, or genome, and highly specific by species.

1.5 Chromosome arms

Most human chromosomes have two arms, and are usually drawn with the shorter above the larger. In mice, cows, sheep and horses almost all the chromosomes have one arm far longer than the other, which may not even be visible. The reason for this is unclear. The reason for the chromosomes in any species to resemble each other, or, as in man, to form distinct groups, is also unclear.

1.6 The haploid set of chromosomes

This is the unit of sexual reproduction.

1.7 The diploid set

This, the nucleus in normal mammals, is the unit of cellular reproduction.

1.8 The genetic hierarchy

This set of mechanisms forms a hierarchy, each element being composed of several or many elements from lower in the hierarchy.

All these elements must reproduce and, in doing so, all are liable to error or mutation, some of which will persist. In the absence of an error, evolution would not occur. In the presence of too much error the species afflicted would not be competitive with its rivals. All these elements mutate at different rates, the larger elements in the hierarchy mutating faster. At no level of the hierarchy is the mutation rate likely to average more than once per unit per generation.

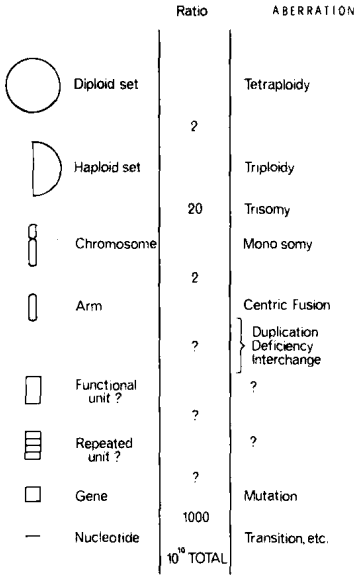


Fig. 1.2 The genetic hierarchy.

1.9 Protein synthesis

The primary functions of the genetic apparatus are, firstly, to duplicate itself and, secondly, to direct the synthesis of impermanent structures, the proteins, which will permit its survival. Protein synthesis is effected by the opening up of a segment of DNA to lay bare the bases; the synthesis of a matching chain of RNA from four components (adenine, guanine, cytosine and uridine: thymidine is not found in RNA); the release of this chain from the DNA; its transport from the nucleus to the cytoplasm; its capture by the two sub-units of a ribosome, which come together, trapping the chain by its beginning; and then the steady threading of this chain along the groove between the ribosomal sub-units, during which transport RNA (small molecules with a recognition site for RNA triplets at one end and a loosely attached amino-acid at the other), effects the translation of the genetic code to produce the linear segment of amino-acid, which is of course protein.

The processes are well understood in bacteria where transcription and translation take place alongside the chromosome. In nuclear organisms the transcription appears to involve a very long chain with regions which are added but not transcribed, this long chain then being segmented into transcription units which are then exported from the nucleus to the cytoplasm: other parts are apparently degraded in the nucleus.

In some unusual cells, such as immature mammalian red cells, messenger RNA may last many days, leading to the synthesis of millions of molecules at a rate of many nucleotides a second, or several protein molecules a minute per ribosome.

1.10 Disorders of transcription

Any defect in the enzymes involved might be expected to be incompatible with life. Since no cell transcribes more than a small minority of its DNA, differentiation must involve defining the segments to be transcribed; deficient transcription is a plausible cause of conditions in which the protein formed is normal in quality but deficient in quantity. Some beta thalassaemias, a class of conditions due to deficient synthesis of haemoglobin, are due to the degradation of the messenger before it has had time to be translated an adequate number of times.

1.11 Translation errors

Translation errors will arise when any mutant transfer RNA provides an unusual translation. In higher animals such a mutation is unlikely to be sufficiently viable to become established, and is likely to be present as a so-called lethal dominant. Certain nucleotide changes might lead to a messenger RNA which was so twisted upon itself that it could not be unravelled. This is a theoretical possibility which, while likely to arise, would be very difficult to detect.

2 The chromosomes

Man has 46 chromosomes, of which 44 are evenly paired in both sexes, and two sex chromosomes which are different in the male.

The chromosomes which are not sex chromosomes are called *autosomes*. The sex chromosomes are sometimes called *allosomes*, but this is not in common use in man, and the terms X-chromosome and sex-chromosomes are often used ambiguously and even erroneously. The chromosomes are usually observed by culturing lymphocytes in the presence of a stimulant and then allowing the dividing cells to accumulate in cell division, or *mitosis*, by adding a substance which blocks its completion. The chromosomes which are about to separate to form the nuclei of the daughter cells are then seen as double structures connected at their centromeres, the centromere being the constriction visible on all chromosomes and connected to the fibres by which the chromosomes are caused to separate after division. The metabolically active chromosome is a very attenuated single strand which forms almost the entire stained matter in the cell nucleus when observed on conventional microscopy.

On displaying the chromosomes by photography, followed by cutting out each chromosome and ranking them by size, shape, and staining characteristics, the characteristic karyotype is obtained (Fig. 2.1). Using modern staining techniques each chromosome can be identified, and, due

to small variations in many chromosomes, any individual can usually be distinguished from any other individual.

The chromosomes may deviate from the normal pattern either by being too few or too many, the *numerical abnormalities*, or by individual chromosomes being abnormal, the *structural abnormalities*: in some cases, as when two chromosomes are fused together, structural abnormalities may lead to a numerical abnormality.

It is convenient to discuss the abnormalities of the autosomes and sex chromosomes separately.

2.1 The autosomal aberrations

About 20% of conceptions are known to die *in utero* and to be aborted in consequence; there are usually some surviving cells and if these are established in tissue culture the chromosomes can be displayed. It is now clear that most of these abortions have numerical chromosomal errors, the commonest being an additional autosome, a missing sex chromosome, and the addition of a complete set of haploid chromosomes, so that 69 chromosomes are present, a condition called *triploidy*.

If the newborn are studied, either by survey or by clinical observation followed by the chromosomal study of unusual children, it is clear that only a minority of conceptions with abnormal numbers of chromosomes survive, the mortality varying from over 99.9% in triploidy and most varieties of extra autosome (trisomy) to about 98% in fetuses with an X-chromosome unaccompanied by another sex chromosome, to a mere slight majority in those with an extra chromosome 21, the survivors of which will have *mongolism*, often called *Down's* syndrome after the English physician who described it in 1866. The pattern of this mortality is shown in Fig. 2.2. There is a remarkable discrepancy between the mortality of conceptions and the morbidity of the survivors. Most children with a single X-chromosome slip through even expert neonatal scrutiny, and, although dwarfed and unable to mature sexually, are usually of average intellect. In mongolism almost all are identified at birth, and intellectual development is normally bounded by the limits of being able to dress but not to read.

Rather more than one baby in a thousand has trisomy 21 or mongolism. The only other trisomic conditions with a frequency exceeding once a paediatrician-lifetime are trisomy 18 and trisomy 13, with frequencies of about one in three thousand and one in five thousand respectively. Both are more severe than mongolism, as is to be expected as the offending chromosome is also bigger, and in the absence of enthusiastic resuscitation followed by tube feeding, natural death usually occurs in days or weeks.

Mongolism, or trisomy 21, provides the bulk of the disability due to chromosomal disorders, and, as will be discussed later, may arise in various ways. However, about 96% of cases are due to a simple trisomy in which the individual chromosomes are normal, and the incidence of this disorder in the newborn is greatly influenced by maternal age, rising from

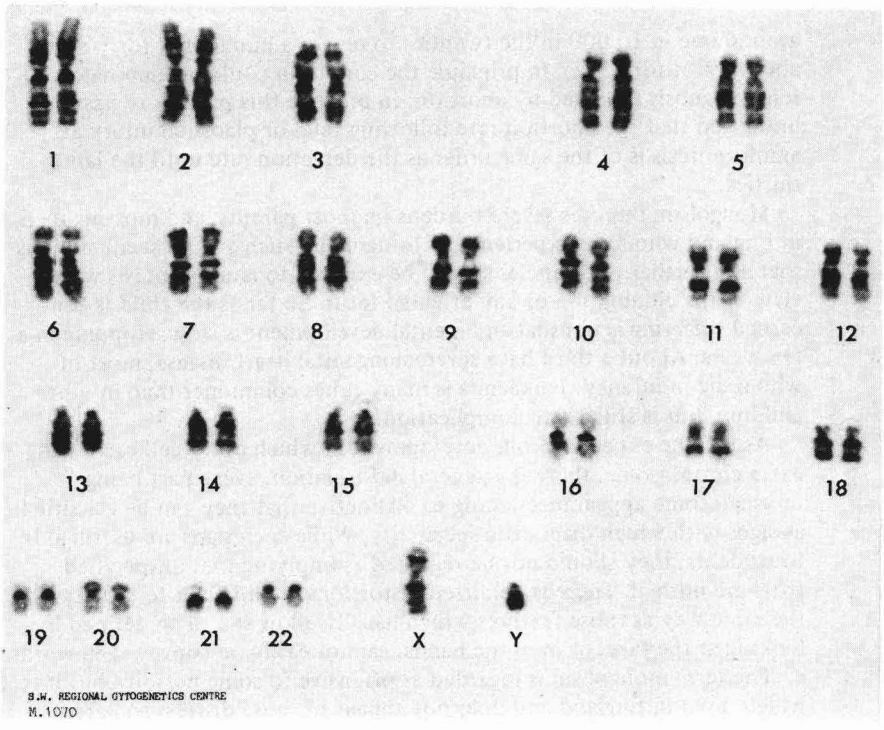


Fig. 2.1 Human male karyotype.

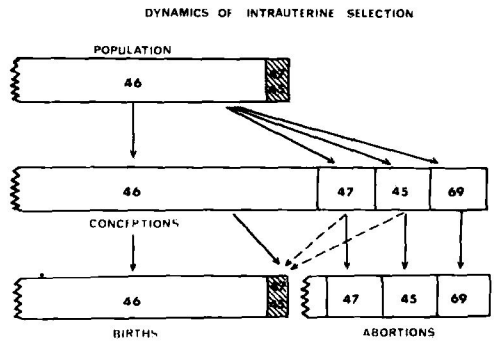


Fig. 2.2 Dynamics of intrauterine selection.

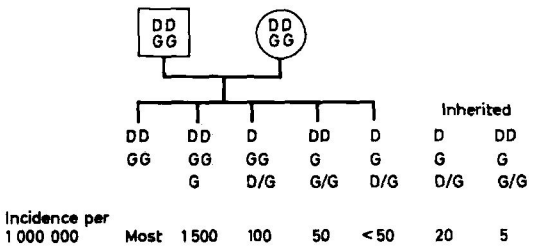


Fig. 2.3 Approximate incidence of types of mongolism (Down's syndrome) per million births.

around one in 10 000 in the twenties to one in a hundred at forty and about 3% at forty five. In principle the condition could be eliminated by fetal diagnosis followed by abortion: in practice this procedure has the limitation that the abortion rate following fetal or placental injury at amniocentesis is of the same order as the detection rate until the late thirties.

Mongolism imposes severe burdens on most parents, and most women in England who have experience of bringing up such a child seem anxious that any further pregnancies should be exposed to amniocentesis with a view to the elimination of any affected fetus. So far as the child is concerned suffering is unusual and mental development is slow, stopping in a few years. About a third have severe congenital heart disease, most of whom die in infancy: leukaemia is many times commoner than in other children, but is still a rare complication.

As is to be expected from development in which every cell carries an extra chromosome, there is a general deformation, every part being unusual, some appearances being so distinctive that they can be classified as signs with a high diagnostic specificity. While such signs are useful aids to students, they should not be regarded as implying that unspecified parts are normal. These generalized distortions are difficult to specify, in the same way as those features which usually allow sex to be defined by looking at the face, or even the hands, cannot easily be conveyed in words.

The term mongolism is regarded as offensive to some persons, but it is widely used in England and does not appear to cause distress to parents. In the United States, which has a substantial Chinese population, including several distinguished geneticists, the term has been almost completely replaced by Down's syndrome. Down considered the disorder to be a residual genetic trait following the Mongolian invasion of Europe.

2.2 Structural autosomal abnormalities

The karyotype, or chromosome set, may be composed of abnormal chromosomes due to a part being duplicated or deficient, or due to an interchange of parts between two chromosomes. Duplication without some associated deficiency is rare. Deficiencies occur, from loss of either the ends of chromosomes or of intervening segments, and it is likely that only a minority of deficiencies large enough to be disabling are large enough to be detectable on conventional microscopy. Human experience in man is consistent with the finding in Drosophila, the fruit fly, that a deficiency is likely to be more harmful than the equivalent bit in excess. About 4% of mongols have the additional chromosome 21 attached to another chromosome, so that, while they have as much chromosomal imbalance as if they had a freelying extra chromosome, they have only 46 chromosomes. In about half of these children the abnormal chromosome comes from a parent, usually the mother, and such parents, although not themselves handicapped by this abnormal attachment of normal chromosomes, are likely to afflict further children by passing on more than an exact half of their autosomes.