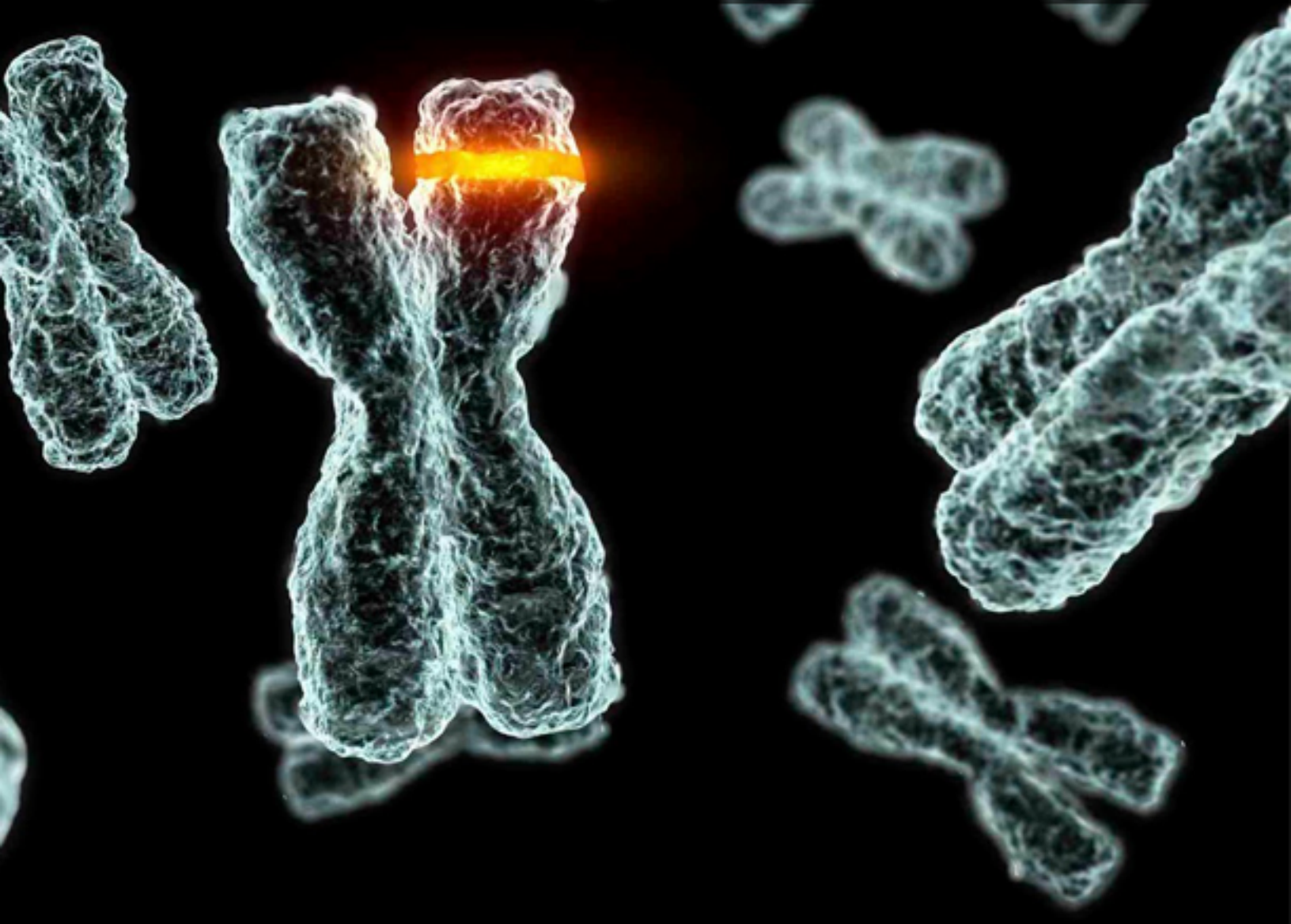
Branches of Genetics

Marie Linares



First Edition, 2012

ISBN 978-81-323-3252-7

© All rights reserved.

Published by:
Research World
4735/22 Prakashdeep Bldg,
Ansari Road, Darya Ganj,
Delhi - 110002
Email: info@wtbooks.com

Table of Contents

Chapter- 1

Behavioural Genetics, Classical Genetics and Ecological Genetics

Behavioural genetics

Behavioural genetics is the field of study that examines the role of genetics in animal (including human) behaviour. Often associated with the "nature versus nurture" debate, behavioural genetics is highly interdisciplinary, involving contributions from biology, genetics, ethology, psychology, and statistics. Behavioural geneticists study the inheritance of behavioural traits. In humans this often use the twin study or adoption study. In animal studies, breeding, transgenesis, and gene knockout techniques are common; psychiatric genetics is a closely related field.

History

Sir Francis Galton, a nineteenth-century intellectual, is recognized as one of the first behavioural geneticists. Galton, a cousin of Charles Darwin, studied the heritability of human ability, focusing on mental characteristics as well as eminence among close relatives in the English upper-class. In 1869, Galton published his results in *Hereditary Genius*. In his work, Galton "introduced multivariate analysis and paved the way towards modern Bayesian statistics" that are used throughout the sciences—launching what has been dubbed the "Statistical Enlightenment".

Behaviour genetics, *per-se*, gained recognition as a research discipline with the publication in 1960 of the textbook *Behavior Genetics* by J.L. Fuller and W.R. Thompson.

Underscoring the role of evolution in behavioural genetics, Theodosius Dobzhansky was elected the first president of the Behavior Genetics Association in 1972; the BGA bestows the Dobzhansky Award on researchers for their outstanding contributions to the field. In the early 1970s, Lee Ehrman, a doctoral student of Dobzhansky, wrote seminal papers describing the relationship between genotype frequency and mating success in

Drosophila, lending impetus to the pursuit of genetic studies of behaviour in other animals.

Notable behavioural geneticists

Notable behavioural geneticists include Dorret Boomsma, John DeFries, Lindon Eaves, David Fulker, John Hewitt, Kenneth Kendler, John Loehlin, Nick Martin, Gerald McClearn, Robert Plomin, Theodore Reich, who was a pioneer in psychiatric genetics, Hans van Abeelen, Avshalom Caspi, and Steven G. Vandenberg, the founding editor of the journal *Behavior Genetics*.

Journals

Behavioural geneticists are active in a variety of scientific disciplines including biology, medicine, pharmacology, psychiatry, and psychology; thus, behavioural-genetic research is published in a variety of scientific journals, including *Nature* and *Science*. Journals that specifically publish research in behavioural genetics include *Behavior Genetics*, *Molecular Psychiatry*, *Psychiatric Genetics*, *Twin Research and Human Genetics*, and *Genes, Brain and Behavior*.

Classical genetics

Classical genetics consists of the technique and methodologies of genetics that predate the advent of molecular biology. A key discovery of classical genetics in eukaryotes was genetic linkage. The observation that some genes do not segregate independently at meiosis broke the laws of Mendelian inheritance, and provided science with a way to map characteristics to a location on the chromosomes. Linkage maps are still used today, especially in breeding for plant improvement.

After the discovery of the genetic code and such tools of cloning as restriction enzymes, the avenues of investigation open to geneticists were greatly broadened. Some classical genetic ideas have been supplanted with the mechanistic understanding brought by molecular discoveries, but many remain intact and in use. Classical genetics is often contrasted with reverse genetics, and aspects of molecular biology are sometimes referred to as molecular genetics.

Ecological genetics

Ecological genetics is the study of genetics in the context of the interactions among organisms and between the organisms and their environment. While molecular genetics studies the structure and function of genes at a molecular level, ecological genetics (and the related field of population genetics) studies phenotypic evolution in natural populations of organisms. Research in this field is of traits of ecological significance — that is, traits related to fitness, which affect an organism's survival and reproduction (e.g., flowering time, drought tolerance, sex ratio).

Studies are often done on insects and other organisms that have short generation times, and thus evolve at high rates.

History

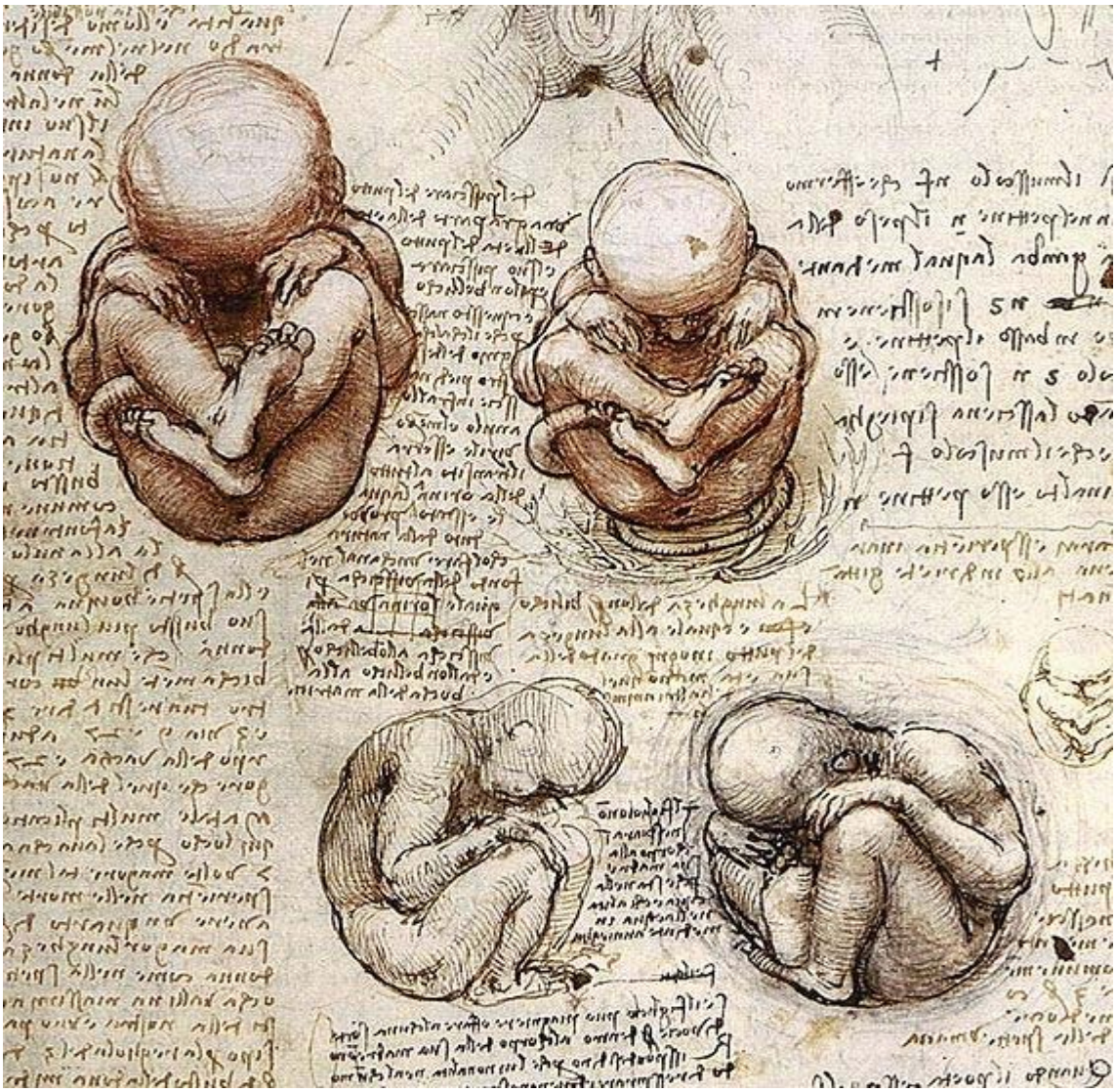
Although work on natural populations had been done previously, it is acknowledged that the field was founded by the English biologist E.B. Ford (1901-1988) in the early 20th century. Ford was taught genetics at Oxford University by Julian Huxley, and started research on the genetics of natural populations in 1924. *Ecological Genetics* is the title of his 1964 'magnum opus' on the subject (4th ed 1975). Other notable ecological geneticists would include Theodosius Dobzhansky who worked on chromosome polymorphism in fruit flies. As a young researcher in Russia, Dobzhansky had been influenced by Sergei Chetverikov, who also deserves to be remembered as a founder of genetics in the field, though his significance was not appreciated until much later.

Philip Sheppard, Cyril Clarke, Bernard Kettlewell and A.J. Cain were all strongly influenced by Ford; their careers date from the post WWII era. Collectively, their work on lepidopterans, and on human blood groups, established the field, and threw light on selection in natural populations where its role had been once doubted.

Work of this kind needs long-term funding, as well as grounding in both ecology and genetics. These are both difficult requirements. Research projects can last longer than a researcher's career; for instance, research into mimicry started 150 years ago, and is still going strongly. Funding of this type of research is still rather erratic, but at least the value of working with natural populations in the field cannot now be doubted.

Chapter- 2

Developmental Biology



"Views of a Fetus in the Womb", Leonardo da Vinci, ca. 1510-1512. The subject of prenatal development is a major subset of developmental biology.

Developmental biology is the study of the process by which organisms grow and develop. Modern developmental biology studies the genetic control of cell growth, differentiation and "morphogenesis", which is the process that gives rise to tissues, organs and anatomy.

Related fields of study

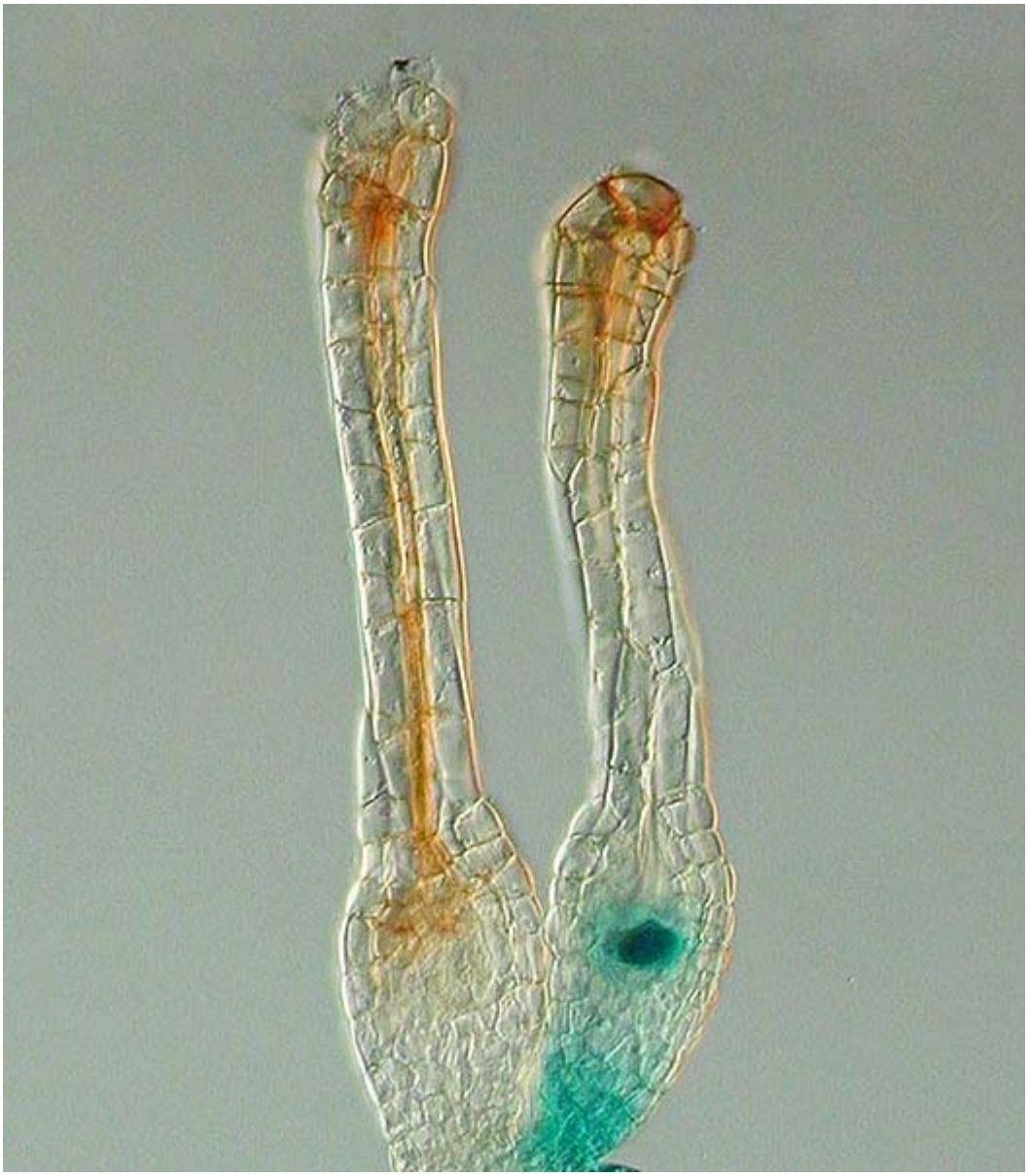
Embryology is a subfield, the study of organisms between the one-cell stage (generally, the zygote) and the end of the embryonic stage. Embryology was originally a more descriptive science until the 20th century. Embryology and developmental biology today deal with the various steps necessary for the correct and complete formation of the body of a living organism.

The related field of evolutionary developmental biology was formed largely in the 1990s and is a synthesis of findings from molecular developmental biology and evolutionary biology which considers the diversity of organismal form in an evolutionary context.

Perspectives

The development of a new life is a spectacular process and represents a masterpiece of temporal and spatial control of gene expression. Developmental genetics studies the effect that genes have in a phenotype, given normal or abnormal epigenetic parameters. The findings of developmental biology can help to understand developmental abnormalities such as chromosomal aberrations that cause Down syndrome. An understanding of the specialization of cells during embryogenesis has provided information on how stem cells specialize into specific tissues and organs. This information has led, for example, to the cloning of specific organs for medical purposes. Another biologically important process that occurs during development is apoptosis—programmed cell death or "suicide." Many developmental models are used to elucidate the physiology and molecular basis of this cellular process. Similarly, a deeper understanding of developmental biology can foster greater progress in the treatment of congenital disorders and diseases, e.g. studying human sex determination can lead to treatment for disorders such as congenital adrenal hyperplasia.

Developmental model organisms



Gene expression pattern determined by histochemical GUS assays in *Physcomitrella patens*. The Polycomb gene FIE is expressed (blue) in unfertilised egg cells of the moss Physcomitrella patens (right) and expression ceases after fertilisation in the developing diploid sporophyte (left). In situ GUS staining of two female sex organs (archegonia) of a transgenic plant expressing a translational fusion of FIE-uidA under control of the native FIE promoter

Often used model organisms in developmental biology include the following:

- Vertebrates

    - Zebrafish *Danio rerio*
    - Medakafish *Oryzias latipes*
    - Fugu (pufferfish) *Takifugu rubripes*
    - Frog *Xenopus laevis*, *Xenopus tropicalis*
    - Chicken *Gallus gallus*
    - Mouse *Mus musculus* (Mammalian embryogenesis)

- Invertebrates
    - Lancelet *Branchiostoma lanceolatum*
    - Ascidian *Ciona intestinalis*
    - Sea urchin *Strongylocentrotus purpuratus*
    - Roundworm *Caenorhabditis elegans*
    - Fruit fly *Drosophila melanogaster* (Drosophila embryogenesis)

- Plants (Plant embryogenesis)
    - *Physcomitrella patens*
    - *Arabidopsis thaliana*
    - Maize
    - Snapdragon *Antirrhinum majus*

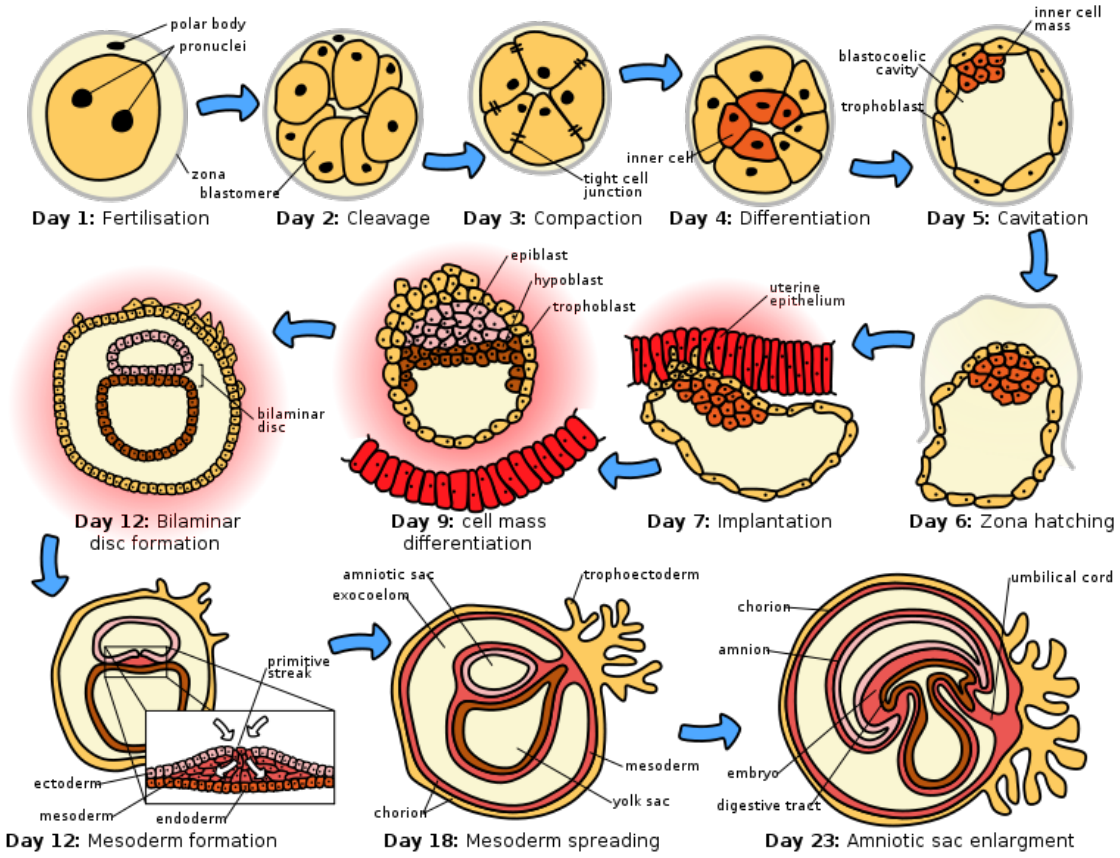
- Other
    - Slime mold *Dictyostelium discoideum*

Studied phenomena

Cell differentiation

Differentiation is the formation of cell types, from what is originally one cell – the zygote or spore. The formation of cell types like nerve cells occurs with a number of intermediary, less differentiated cell types. A cell stays a certain cell type by maintaining a particular pattern of gene expression. This depends on regulatory genes, e.g. for [[transcription factor] name=WolpertDifferentiationmodels>Wolpert L, Beddington R, Jessell T, Lawrence P, Meyerowitz E, Smith J (2002).

Embryonal development



The initial stages of human embryogenesis

Embryogenesis is the step in the life cycle after fertilisation – the development of the embryo, starting from the zygote (fertilised egg). Organisms can differ drastically in the how embryo develops, especially when they belong to different phyla. For example, embryonal development in placental mammals starts with cleavage of the zygote into eight uncommited cells, which then form a ball (morula). The outer cells become the trophectoderm or trophoblast, which will form in combination with maternal uterine endometrial tissue the placenta, needed for fetal nurturing via maternal blood, while inner cells become the inner cell mass that will form all fetal organs (the bridge between these two parts eventually forms the umbilical cord). In contrast, the fruit fly zygote first forms a sausage-shaped syncytium, which is still one cell but with many cell nuclei.

Patterning is important for determining which cells develop into which organs. This is mediated by signaling between adjacent cells by proteins on their surfaces, and by gradients of signaling secreted molecules. An example is retinoic acid, which forms a gradient in the head to tail direction in animals. Retinoic acid enters cells and activates Hox genes in a concentration-dependent manner – Hox genes differ in how much retinoic acid they require for activation and will thus show differential rostral expression boundaries, in a colinear fashion with their genomic order. As Hox genes code for

transcription factors, this causes different activated combinations of both Hox and other genes in discrete anteroposterior transverse segments of the neural tube (neuromeres) and related patterns in surrounding tissues, such as branchial arches, lateral mesoderm, neural crest, skin and endoderm, in the head to tail direction. This is important for e.g. the segmentation of the spine in vertebrates.

Embryonal development does not always proceed correctly, and errors can result in birth defects or miscarriage. Often the reason is genetic (mutation or chromosome abnormality), but there can be environmental influence (like teratogens) or stochastic events. Abnormal development caused by mutation is also of evolutionary interest as it provides a mechanism for changes in body plan.

Growth

Growth is the enlargement of a tissue or organism. Growth continues after the embryonal stage, and occurs through cell proliferation, enlargement of cells or accumulation of extracellular material. In plants, growth results in an adult organism that is strikingly different from the embryo. The proliferating cells tend to be distinct from differentiated cells. In some tissues proliferating cells are restricted to specialised areas, such as the growth plates of bones. But some stem cells migrate to where they are needed, such as mesenchymal stem cells which can migrate from the bone marrow to form e.g. muscle, bone or adipose tissue. The size of an organ frequently determines its growth, as in the case of the liver which grows back to its previous size if a part is removed. Growth factors, such as fibroblast growth factors in the animal embryo and growth hormone in juvenile mammals, also control the extent of growth.

Metamorphosis

Most animals have a larval stage, with a body plan different from that of the adult organism. The larva abrubtly develops into an adult in a process called metamorphosis. For example, caterpillars (butterfly larvae) are specialized for feeding whereas adult butterflies (imagos) are specialised for flight and reproduction. When the caterpillar has grown enough, it turns into an immobile pupa. Here, the imago develops from imaginal discs found inside the larva.

Regeneration

Regeneration is the reactivation of development so that a missing body part grows back. This phenomenon has been studied particularly in salamanders, where the adults can reconstruct a whole limb after it has been amputated. Researchers hope to one day be able to induce regeneration in humans. There is little spontaneous regeneration in adult humans, although the liver is a notable exception. Like for salamanders, the regeneration of the liver involves dedifferentiation of some cells to a more embryonal state.

Developmental systems biology

Computer simulation of multicellular development is a research methodology to understand the function of the very complex processes involved in the development of organisms. This includes simulation of cell signaling, multicell interactions and regulatory genomic networks in development of multicellular structures and processes. *Minimal genomes* for minimal multicellular organisms may pave the way to understand such complex processes *in vivo*.

Chapter- 3

Conservation Genetics

Conservation genetics is an interdisciplinary science that aims to apply genetic methods to the conservation and restoration of biodiversity. Researchers involved in conservation genetics come from a variety of fields including population genetics, molecular ecology, biology, evolutionary biology, and systematics. Genetic diversity is one of the three fundamental levels of biodiversity, so it is directly important in conservation of biodiversity, though genetic factors are also important in the conservation of species and ecosystem diversity. Conservation of genetic variability is important to the overall health of populations because decreased genetic variability leads to increased levels of inbreeding, and reduced fitness

Genetic Diversity

Genetic diversity is the variability of a genes in a species. It can be estimated by the mean levels of heterozygosity in a population, the mean number of alleles per locus, or the percentage of polymorphic loci.

The importance of genetic diversity

If genetic diversity becomes low at many genes of a species, that species becomes increasingly at risk. It has only one possible choice of information at all or nearly all of its genes—in other words, all the individuals are nearly identical. If new pressures (such as environmental disasters) occur, a population with high genetic diversity has a greater chance of having at least some individuals with a genetic makeup that allows them to survive. If genetic diversity is very low, none of the individuals in a population may have the characteristics needed to cope with the new environmental conditions. Such a population could be suddenly wiped out.

The genetic diversity of a species is always open to change. No matter how many variants of a gene are present in a population today, only the variants that survive in the next generation can contribute to species diversity in the future. Once gene variants are lost, they cannot be recovered.

Contributors to extinction

1. Inbreeding and inbreeding elevation which reduces the fitness of populations.
2. The accumulation of deleterious mutations
3. A decrease in frequency of heterozygotes in a population, or heterozygosity, which decreases a species' ability to evolve to deal with change in the environment.
4. Adapting to conditions in captivity
5. Outbreeding depression
6. Fragmented populations
7. Taxonomic uncertainties, which can lead to a reprioritization of conservation efforts
8. Genetic drift as the main evolutionary process, instead of natural selection
9. Management units within species
10. Use of molecular techniques, such as allozymes as molecular markers, to analyze species in depth

Techniques

Specific genetic techniques are used to assess the genetics of a species regarding specific conservation issues as well as general population structure. This analysis can be done in two ways, with current DNA of individuals or historic DNA.

Techniques for analysiing the differences between individuals and populations include

1. Alloenzymes
2. Random Fragment Length Polymorphisms
3. Amplified Fragment Length Polymorphisms
4. Random Amplification of Polymorphic DNA
5. Single strand conformation polymorphism
6. minisatellites
7. microsatellites
8. Single nucleotide polymorphisms
9. Sequence analysis
10. DNA fingerprinting

These different techniques focus on different variable areas of the genomes within animals and plants. The specific information that is required determines which techniques are used and which parts of the genome are analysed. For examplemitochondrial DNA in animals has a high substitution rate, which makes it useful for identifying differences between individuals. However, it is only inherited in the female line, and the mitochondrial genome is relatively small. In plants, the mitochondrial DNA has very high rates of structural mutations, so is rarely used for genetic markers, as the chloroplast genome can be used instead. Other sites in the genome that are subject to high mutation

rates such as the Major Histocompatibility Complex, and the microsatellites and minisatellites are also frequently used.

These techniques can provide information on long-term conservation of genetic diversity and expound demographic and ecological matters such as taxonomy.

Another technique is using historic DNA for genetic analysis. Historic DNA is important because it allows geneticists to understand how species reacted to changes to conditions in the past. This is a key to understanding the reactions of similar species in the future.

Techniques using historic DNA include looking at preserved remains found in museums and caves. Museums are used because there is a wide range of species that are available to scientists all over the world. The problem with museums is that, historical perspectives are important because understanding how species reacted to changes in conditions in the past is a key to understanding reactions of similar species in the future. Evidence found in caves provides a longer perspective and does not disturb the animals.

Another technique that relies on specific genetics of an individual is non invasive monitoring, which uses extracted DNA from organic material that an individual leaves behind, such as a feather. This too avoids disrupting the animals and can provide information about the sex, movement, kinship and diet of an individual.

Other more general techniques can be used to correct genetic factors that lead to extinction and risk of extinction. For example, when minimizing inbreeding and increasing genetic variation multiple steps can be taken. Increasing heterozygosity through immigration, increasing the generational interval through cryopreservation or breeding from older animals, and increasing the effective population size through equalization of family size all helps minimize inbreeding and its effects. Deleterious alleles arise through mutation, however certain recessive ones can become more prevalent due to inbreeding. Deleterious mutations that arise from inbreeding can be removed by purging, or natural selection. Populations raised in captivity with the intent of being reintroduced in the wild suffer from adaptations to captivity.

Inbreeding depression, loss of genetic diversity, and genetic adaptation to captivity are disadvantageous in the wild, and many of these issues can be dealt with through the aforementioned techniques aimed at increasing heterozygosity. In addition creating a captive environment that closely resembles the wild and fragmenting the populations so there is less response to selection also help reduce adaptation to captivity.

Solutions to minimize the factors that lead to extinction and risk of extinction often overlap because the factors themselves overlap. For example, deleterious mutations are added to populations through mutation, however the deleterious mutations conservation biologists are concerned with are ones that are brought about by inbreeding, because those are the ones that can be taken care of by reducing inbreeding. Here the techniques to reduce inbreeding also help decrease the accumulation of deleterious mutations.

Applications

These techniques have wide ranging applications. One application of these specific molecular techniques is in defining species and sub-species of salmonids. Hybridization is an especially important issue in salmonids and this has wide ranging conservation, political, social and economic implications. In Cutthroat Trout mtDNA and alloenzyme analysis, hybridization between native and non-native species was shown to be one of the major factors contributing to the decline in their populations. This led to efforts to remove some hybridized populations so native populations could breed more readily. Cases like these impact everything from the economy of local fishermen to larger companies, such as timber. Specific molecular techniques led to a closer analysis of taxonomic relationships, which is one factor that can lead to extinctions if unclear.

Implications

New technology in conservation genetics has many implications for the future of conservation biology. At the molecular level, new technologies are advancing. Some of these techniques include minisatellites and MHC. These molecular techniques have wider effects from clarifying taxonomic relationships, as in the previous example, to determining the best individuals to reintroduce to a population for recovery by determining kinship. These effects then have consequences that reach even further. Conservation of species has implications for humans in the economic, social, and political realms. In the biological realm increased genotypic diversity has been shown to help ecosystem recovery, as seen in a community of grasses which was able to resist disturbance to grazing geese through greater genotypic diversity. Because species diversity increases ecosystem function, increasing biodiversity through new conservation genetic techniques has wider reaching effects than before.

A short list of studies a conservation geneticist may research include:

1. Phylogenetic classification of species, subspecies, geographic races, and populations, and measures of phylogenetic diversity and uniqueness.
2. Identifying hybrid species, hybridization in natural populations, and assessing the history and extent of introgression between species.
3. Population genetic structure of natural and managed populations, including identification of Evolutionary Significant Units (ESUs) and management units for conservation.
4. Assessing genetic variation within a species or population, including small or endangered populations, and estimates such as effective population size (Ne).
5. Measuring the impact of inbreeding and outbreeding depression, and the relationship between heterozygosity and measures of fitness.
6. Evidence of disrupted mate choice and reproductive strategy in disturbed populations.
7. Forensic applications, especially for the control of trade in endangered species.

8. Practical methods for monitoring and maximizing genetic diversity during captive breeding programs and re-introduction schemes, including mathematical models and case studies.
9. Conservation issues related to the introduction of genetically modified organisms.
10. The interaction between environmental contaminants and the biology and health of an organism, including changes in mutation rates and adaptation to local changes in the environment (e.g. industrial melanism).
11. New techniques for noninvasive genotyping.

Chapter- 4

Genetic Engineering

Genetic engineering, also called **genetic modification**, is the direct human manipulation of an organism's genetic material in a way that does not occur under natural conditions. It involves the use of recombinant DNA techniques, but does not include traditional animal and plant breeding or mutagenesis. Any organism that is generated using these techniques is considered to be a genetically modified organism. The first organisms genetically engineered were bacteria in 1973 and then mice in 1974. Insulin producing bacteria were commercialized in 1982 and genetically modified food has been sold since 1994.

The most common form of genetic engineering involves the insertion of new genetic material at an unspecified location in the host genome. This is accomplished by isolating and copying the genetic material of interest, generating a construct containing all the genetic elements for correct expression, and then inserting this construct into the host organism. Other forms of genetic engineering include gene targeting and knocking out specific genes via engineered nucleases such as zinc finger nucleases or engineered homing endonucleases.

Genetic engineering techniques have been applied in numerous fields including research, biotechnology, and medicine. Medicines such as insulin and human growth hormone are now produced in bacteria, experimental mice such as the oncomouse and the knockout mouse are being used for research purposes and insect resistant and/or herbicide tolerant crops have been commercialized. Genetically engineered plants and animals capable of producing biotechnology drugs more cheaply than current methods (called pharming) are also being developed and in 2009 the FDA approved the sale of the pharmaceutical protein antithrombin produced in the milk of genetically engineered goats.

Definition

Genetic engineering alters the genetic makeup of an organism using techniques that introduce heritable material prepared outside the organism either directly into the host or into a cell that is then fused or hybridized with the host. This involves using recombinant nucleic acid (DNA or RNA) techniques to form new combinations of heritable genetic material followed by the incorporation of that material either indirectly through a vector

system or directly through micro-injection, macro-injection and micro-encapsulation techniques. Genetic engineering does not include traditional animal and plant breeding, in vitro fertilisation, induction of polyploidy, mutagenesis and cell fusion techniques that do not use recombinant nucleic acids or a genetically modified organism in the process. Cloning and stem cell research, although not considered genetic engineering, are closely related and genetic engineering can be used within them. Synthetic biology is an emerging discipline that takes genetic engineering a step further by introducing artificially synthesized genetic material from raw materials into an organism.

If genetic material from another species is added to the host, the resulting organism is called transgenic. If genetic material from the same species or a species that can naturally breed with the host is used the resulting organism is called cisgenic. Genetic engineering can also be used to remove genetic material from the target organism, creating a knock out organism. In Europe genetic modification is synonymous with genetic engineering while within the United States of America it can also refer to conventional breeding methods.

History

Humans have altered the genomes of species for thousands of years through artificial selection and more recently mutagenesis. Genetic engineering as the direct manipulation of DNA by humans outside breeding and mutations has only existed since the 1970s. The term "genetic engineering" was first coined by Jack Williamson in his science fiction novel *Dragon's Island*, published in 1951, one year before DNA's role in heredity was confirmed by Alfred Hershey and Martha Chase, and two years before James Watson and Francis Crick showed that the DNA molecule has a double-helix structure.

In 1972 Paul Berg created the first recombinant DNA molecules by combined DNA from the monkey virus SV40 with that of the lambda virus. In 1973 Herbert Boyer and Stanley Cohen created the first transgenic organism by inserting antibiotic resistance genes into the plasmid of an *E. coli* bacterium. A year later Rudolf Jaenisch created a transgenic mouse by introducing foreign DNA into its embryo, making it the world’s first transgenic animal. In 1976 Genentech, the first genetic engineering company was founded by Herbert Boyer and Robert Swanson and a year later and the company produced a human protein (somatostatin) in *E.coli*. Genentech announced the production of genetically engineered human insulin in 1978. In 1980, the U.S. Supreme Court in the Diamond v. Chakrabarty case ruled that genetically altered life could be patented. The insulin produced by bacteria, branded humulin, was approved for release by the Food and Drug Administration in 1982.

The first field trials of genetically engineered plants occurred in France and the USA in 1986, tobacco plants were engineered to be resistant to herbicides. The People’s Republic of China was the first country to commercialize transgenic plants, introducing a virus-resistant tobacco in 1992. In 1994 Calgene attained approval to commercially release the Flavr Savr tomato, a tomato engineered to have a longer shelf life. In 1994, the European Union approved tobacco engineered to be resistant to the herbicide bromoxynil, making it

the first genetically engineered crop commercialized in Europe. In 1995, Bt Potato was approved safe by the Environmental Protection Agency, making it the first pesticide producing crop to be approved in the USA. In 2009 11 transgenic crops were grown commercially in 25 countries, the largest of which by area grown were the USA, Brazil, Argentina, India, Canada, China, Paraguay and South Africa.

In 2010, scientists at the J. Craig Venter Institute, announced that they had created the first synthetic bacterial genome, and added it to a cell containing no DNA. The resulting bacterium, named Synthia, was the world's first synthetic life form.

Process

Isolating the Gene

Elements of genetic engineering

First, the gene to be inserted into the genetically modified organism must be chosen and isolated. Presently, most genes transferred into plants provide protection against insects or tolerance to herbicides. In animals the majority of genes used are growth hormone genes. Once chosen the genes must be isolated. This typically involves multiplying the gene using polymerase chain reaction (PCR). If the chosen gene or the donor organism's genome has been well studied it may be present in a genetic library. If the DNA sequence is known, but no copies of the gene are available, it can be artificially synthesized. Once isolated, the gene is inserted into a bacterial plasmid.

Constructs

The gene to be inserted into the genetically modified organism must be combined with other genetic elements in order for it to work properly. The gene can also be modified at this stage for better expression or effectiveness. As well as the gene to be inserted most constructs contain a promoter and terminator region as well as a selectable marker gene. The promoter region initiates transcription of the gene and can be used to control the location and level of gene expression, while the terminator region ends transcription. The selectable marker, which in most cases confers antibiotic resistance to the organism it is expressed in, is needed to determine which cells are transformed with the new gene. The constructs are made using recombinant DNA techniques, such as restriction digests, ligations and molecular cloning.

Gene Targeting

The most common form of genetic engineering involves inserting new genetic material randomly within the host genome. Other techniques allow new genetic material to be inserted at a specific location in the host genome or generate mutations at desired genomic loci capable of knocking out endogenous genes. The technique of gene targeting uses homologous recombination to target desired changes to a specific endogenous gene. This tends to occur at a relatively low frequency in plants and animals and generally requires the use of selectable markers. The frequency of gene targeting can be greatly enhanced with the use of engineered nucleases such as zinc finger nucleases, engineered homing endonucleases, or nucleases created from TAL effectors. In addition to enhancing gene targeting, engineered nucleases can also be used to introduce mutations at endogenous genes that generate a gene knockout.

Transformation

A. tumefaciens attaching itself to a carrot cell

About 1% of bacteria are naturally able to take up foreign DNA but it can also be induced in other bacteria. Stressing the bacteria for example, with a heat shock or an electric shock, can make the cell membrane permeable to DNA that may then incorporate into their genome or exist as extrachromosomal DNA. DNA is generally inserted into animal cells using microinjection, where it can be injected through the cells nuclear envelope directly into the nucleus or through the use of viral vectors. In plants the DNA is generally inserted using *Agrobacterium*-mediated recombination or biolistics.

In *Agrobacterium*-mediated recombination the plasmid construct must also contain T-DNA. *Agrobacterium* naturally inserts DNA from a tumor inducing plasmid into any susceptible plant's genome it infects, causing crown gall disease. The T-DNA region of this plasmid is responsible for insertion of the DNA. The genes to be inserted are cloned into a binary vector, which contains T-DNA and can be grown in both *E. Coli* and *Agrobacterium*. Once the binary vector is constructed the plasmid is transformed into *Agrobacterium* containing no plasmids and plant cells are infected. The *Agrobacterium* will then naturally insert the genetic material into the plant cells.

In biolistics particles of gold or tungsten are coated with DNA and then shot into young plant cells or plant embryos. Some genetic material will enter the cells and transform them. This method can be used on plants that are not susceptible to *Agrobacterium* infection and also allows transformation of plant plastids. Another transformation method for plant and animal cells is electroporation. Electroporation involves subjecting the plant or animal cell to an electric shock, which can make the cell membrane permeable to plasmid DNA. In some cases the electroporated cells will incorporate the DNA into their genome. Due to the damage caused to the cells and DNA the transformation efficiency of biolistics and electroporation is lower than agrobacterial mediated transformation and microinjection.

Selection

Not all the organism's cells will be transformed with the new genetic material; in most cases a selectable marker is used to differentiate transformed from untransformed cells. If a cell has been successfully transformed with the DNA it will also contain the marker gene. By growing the cells in the presence of an antibiotic or chemical that selects or marks the cells expressing that gene it is possible to separate the transgenic events from the non-transgenic. Another method of screening involves using a DNA probe that will only stick to the inserted gene. A number of strategies have been developed that can remove the selectable marker from the mature transgenic plant.

Regeneration

As often only a single cell is transformed with genetic material the organism must be regrown from that single cell. As bacteria consist of a single cell and reproduce clonally regeneration is not necessary. In plants this is accomplished through the use of tissue culture. Each plant species has different requirements for successful regeneration through tissue culture. If successful an adult plant is produced that contains the transgene in every cell. In animals it is necessary to ensure that the inserted DNA is present in the embryonic stem cells. When the offspring is produced they can be screened for the presence of the gene. All offspring from the first generation will be heterozygous for the inserted gene and must be mated together to produce a homozygous animal.

Confirmation

Further tests using PCR, Southern Blots and Bioassays are needed to confirm that the gene is expressed and functions correctly. The organism's offspring are also tested to ensure that the trait can be inherited and that it follows a Mendelian inheritance pattern.

Applications

Genetic engineering has applications in medicine, research, industry and agriculture and can be used on a wide range of plants, animals and micro organism.

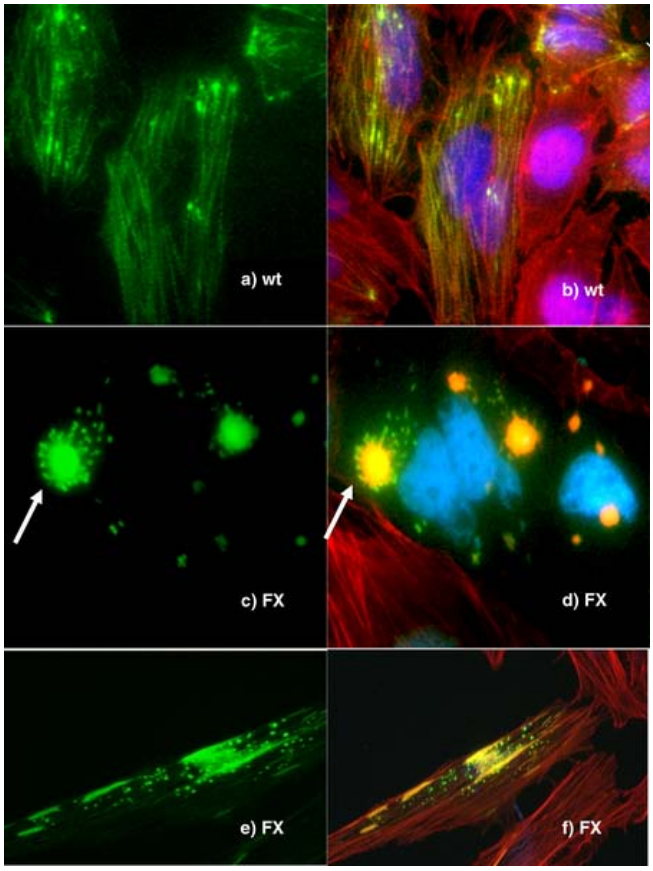
Medicine

In medicine genetic engineering has been used to mass-produce insulin, human growth hormones, follistim (for treating infertility), human albumin, monoclonal antibodies, antihemophilic factors, vaccines and many other drugs. Vaccination generally involves injecting weak live, killed or inactivated forms of viruses or their toxins into the person being immunized. Genetically engineered viruses are being developed that can still confer immunity, but lack the infectious sequences. Mouse hybridomas, cells fused together to create monoclonal antibodies, have been humanised through genetic engineering to create human monoclonal antibodies.

Genetic engineering is used to create animal models of human diseases. Genetically modified mice are the most common genetically engineered animal model. They have been used to study and model cancer (the oncomouse), obesity, heart disease, diabetes, arthritis, substance abuse, anxiety, aging and Parkinson disease. Potential cures can be tested against these mouse models. Also genetically modified pigs have been bred with the aim of increasing the success of pig to human organ transplantation.

Gene therapy is the genetic engineering of humans by replacing defective human genes with functional copies. This can occur in somatic tissue or germline tissue. If the gene is inserted into the germline tissue it can be passed down to that person's descendants. Gene therapy has been used to treat patients suffering from immune deficiencies (notably Severe combined immunodeficiency) and trials have been carried out on other genetic disorders. The success of gene therapy so far has been limited and a patient (Jesse Gelsinger) has died during a clinical trial testing a new treatment. There are also ethical concerns should the technology be used not just for treatment, but for enhancement, modification or alteration of a human beings' appearance, adaptability, intelligence, character or behavior. The distinction between cure and enhancement can also be difficult to establish. Transhumanists consider the enhancement of humans desirable.

Research



Human cells in which some proteins are fused with green fluorescent protein to allow them to be visualized

Knockout mice

Genetic engineering is an important tool for natural scientists. Genes and other genetic information from a wide range of organisms are transformed into bacteria for storage and modification, creating genetically modified bacteria in the process. Bacteria are cheap, easy to grow, clonal, multiply quickly, relatively easy to transform and can be stored at -80°C almost indefinitely. Once a gene is isolated it can be stored inside the bacteria providing an unlimited supply for research.

Organisms are genetically engineered to discover the functions of certain genes. This could be the effect on the phenotype of the organism, where the gene is expressed or what other genes it interacts with. These experiments generally involve loss of function, gain of function, tracking and expression.

- **Loss of function experiments**, such as in a gene knockout experiment, in which an organism is engineered to lack the activity of one or more genes. A knockout experiment involves the creation and manipulation of a DNA construct *in vitro*, which, in a simple knockout, consists of a copy of the desired gene, which has been altered such that it is non-functional. Embryonic stem cells incorporate the altered gene, which replaces the already present functional copy. These stem cells are injected into blastocysts, which are implanted into surrogate mothers. This allows the experimenter to analyze the defects caused by this mutation and thereby determine the role of particular genes. It is used especially frequently in developmental biology. Another method, useful in organisms such as Drosophila (fruit fly), is to induce mutations in a large population and then screen the progeny for the desired mutation. A similar process can be used in both plants and prokaryotes.
- **Gain of function experiments**, the logical counterpart of knockouts. These are sometimes performed in conjunction with knockout experiments to more finely establish the function of the desired gene. The process is much the same as that in knockout engineering, except that the construct is designed to increase the function of the gene, usually by providing extra copies of the gene or inducing synthesis of the protein more frequently.
- **Tracking experiments**, which seek to gain information about the localization and interaction of the desired protein. One way to do this is to replace the wild-type gene with a 'fusion' gene, which is a juxtaposition of the wild-type gene with a reporting element such as green fluorescent protein (GFP) that will allow easy visualization of the products of the genetic modification. While this is a useful technique, the manipulation can destroy the function of the gene, creating secondary effects and possibly calling into question the results of the experiment.

More sophisticated techniques are now in development that can track protein products without mitigating their function, such as the addition of small sequences that will serve as binding motifs to monoclonal antibodies.

- **Expression studies** aim to discover where and when specific proteins are produced. In these experiments, the DNA sequence before the DNA that codes for a protein, known as a gene's promoter, is reintroduced into an organism with the protein coding region replaced by a reporter gene such as GFP or an enzyme that catalyzes the production of a dye. Thus the time and place where a particular protein is produced can be observed. Expression studies can be taken a step further by altering the promoter to find which pieces are crucial for the proper expression of the gene and are actually bound by transcription factor proteins; this process is known as promoter bashing.

Industrial

By engineering genes into bacterial plasmids it is possible to create a biological factory that can produce proteins and enzymes. Some genes do not work well in bacteria, so yeast, a eukaryote, can also be used. Bacteria and yeast factories have been used to produce medicines such as insulin, human growth hormone, and vaccines, supplements such as tryptophan, aid in the production of food (chymosin in cheese making) and fuels. Other applications involving genetically engineered bacteria being investigated involve making the bacteria perform tasks outside their natural cycle, such as cleaning up oil spills, carbon and other toxic waste.

Agriculture

Bt-toxins present in peanut leaves (bottom image) protect it from extensive damage caused by European corn borer larvae (top image).

One of the best-known and controversial applications of genetic engineering is the creation of genetically modified food. There are three generations of genetically modified crops. First generation crops have been commercialized and most provide protection from insects and/or resistance to herbicides. There are also fungal and virus resistant crops developed or in development. They have been developed to make the insect and weed management of crops easier and can indirectly increase crop yield.

The second generation of genetically modified crops being developed aim to directly improve yield by improving salt, cold or drought tolerance and to increase the nutritional

value of the crops. The third generation consists of pharmaceutical crops, crops that contain edible vaccines and other drugs. Some agriculturally important animals have been genetically modified with growth hormones to increase their size while others have been engineered to express drugs and other proteins in their milk.

The genetic engineering of agricultural crops can increase the growth rates and resistance to different diseases caused by pathogens and parasites. This is beneficial as it can greatly increase the production of food sources with the usage of fewer resources that would be required to host the world's growing populations. These modified crops would also reduce the usage of chemicals, such as fertilizers and pesticides, and therefore decrease the severity and frequency of the damages produced by these chemical pollution.

Ethical and safety concerns have been raised around the use of genetically modified food. A major safety concern relates to the human health implications of eating genetically modified food, in particular whether toxic or allergic reactions could occur. Gene flow into related non-transgenic crops, off target effects on beneficial organisms and the impact on biodiversity are important environmental issues. Ethical concerns involve religious issues, corporate control of the food supply, intellectual property rights and the level of labeling needed on genetically modified products.

Other uses

In materials science, a genetically modified virus has been used to construct a more environmentally friendly lithium-ion battery. Some bacteria have been genetically engineered to create black and white photographs while others have potential to be used as sensors by expressing a fluorescent protein under certain environmental conditions. Genetic engineering is also being used to create BioArt and novelty items such as blue roses, and glowing fish.

Opposition and criticism

A 2010 study of Canola found transgenes in 80% of wild (uncultivated or "feral") varieties in North Dakota, meaning 80% of the plants which had established themselves in the area were genetically engineered varieties. The researchers stated that "we found the highest densities of [such transgene-containing] plants near agricultural fields and along major freeways, but we were also finding plants in the middle of nowhere" adding that "over time,..the build-up of different types of herbicide resistance in feral [natural] canola and closely related weeds, like field mustard, could make it more difficult to manage these plants using herbicides."

Chapter- 5

Heritability of IQ

The study of the **heritability of IQ** is a field of research that includes biology, genomics, psychology, philosophy, sociology, and anthropology. Heritability is "an estimate of the genetic and environmental contributions to the *variance* of any phenotypic measure around the mean for a given population." "Heritability refers to the genetic contribution to variance within a population and in a specific environment . . . ; if the environment changes, the heritability measure changes."

Some contend that heritability does not set any limit on how malleable the trait is under changes of environment, because "even highly heritable traits can be strongly manipulated by the environment, so heritability has little if anything to do with controllability.". However, others argue that heritability constrains malleability. The debate about IQ heritability touches on the nature versus nurture divide, and there has been significant controversy in the academic community about it ever since research began in the 19th century.

IQ is a polygenic trait under normal circumstances according to recent research. However, destructive mutation of individual genes associated with development can severely affect intelligence, with Phenylketonuria as an example.

Estimates in the academic research of the heritability of IQ have varied from below 0.5 to a high of 0.9. A 1996 statement by the American Psychological Association gave about .45 for children and about .75 during and after adolescence. A 2004 meta-analysis of reports in *Current Directions in Psychological Science* gave an overall estimate of around .85 for 18-year-olds and older. *The New York Times Magazine* has listed about three quarters as a figure held by the majority of studies.

Methods and results

Heritability calculations

Background

Heritability is defined as the proportion of variance in a trait which is attributable to genotype within a defined population in a specific environment. Heritability takes a value ranging from 0 to 1; a heritability of 1 indicates that all variation in the trait in question is genetic in origin and a heritability of 0 indicates that *none* of the variation is genetic. The determination of many traits can be considered primarily genetic under similar environmental backgrounds. For example, Visscher *et al.* (2006) found that adult height has a heritability estimated at 0.80, when a similar environmental background is present, to control for environment the study only looked at the contribution of heritability to variation within families. The paper stated that "one can never be sure that the estimates are correct, because nature and nurture can be confounded without one knowing it. The authors got around this problem by comparing the similarity between relatives as a function of the exact proportion of genes that they have in common, looking only within families." Other traits have low heritabilities, which indicate a large relative environmental influence. For example, a twin study on the heritability of depression in men calculated it as 0.29, while it was 0.42 for women in the same study.

Heritability for a trait is calculated by measuring how strongly traits covary in people of a given genetic and environmental similarity. The most common method is to consider identical twins reared apart, with any similarities which exists between such twin pairs attributed to genotype. In terms of correlation statistics, this means that theoretically the correlation of tests scores between monozygotic twins would be 1.00 if genetics alone accounted for variation in IQ scores; likewise, siblings and dizygotic twins share on average half of their alleles and the correlation of their scores would be 0.50 if IQ were affected by genes alone. Practically, however, the upper bound of these correlations are given by the reliability of the test, which tends to be 0.90 to 0.95 for typical IQ tests Thus, the actual heritability of IQ will tend to be slightly higher than attained by estimates derived from studies of monozygotic twins, though this effect is small.

In the case of the inheritance of IQ or a certain degree of giftedness, the relatives of probands with a high IQ exhibit a comparably high IQ with a much higher probability than the general population. In 1982, Bouchard and McGue reviewed such correlations reported in 111 original studies in the United States. The mean correlation of IQ scores between monozygotic twins was 0.86, between siblings, 0.47, between half-siblings, 0.31, and between cousins, 0.15. From such data the heritability of IQ was estimated at anywhere between 0.40 and 0.80 in the United States. The reason for this wide margin appeared to be that the heritability of IQ rises through childhood and adolescence, peaking at 0.68 and 0.78 in adults, leaving the overwhelming majority of IQ differences between individuals to be explained genetically.

The finding of rising heritability with age is counter-intuitive; it is reasonable to expect that genetic influences on traits like IQ should become less important as one gains experiences with age. However, that the opposite occurs is well documented. According to work by Robert Plomin, heritability estimates calculated on infant samples are as low as 20%, rising to around 40% in middle childhood, and ultimately as high as 80% in adult samples in the United States. This suggests that the underlying genes actually express themselves by affecting a person's predisposition to build, learn, and develop mental abilities throughout the lifespan.

Estimates and caveats to them

In 2006, *The New York Times Magazine* listed about three quarters as a figure held by the majority of studies, while a 2004 meta-analysis of reports in *Current Directions in Psychological Science* gave an overall estimate of around .85 for 18-year-olds and older. As well, a 1996 statement by the American Psychological Association gave about .45 for children and about .75 during and after adolescence.

The 2006 edition of *Assessing adolescent and adult intelligence* by Alan S. Kaufman and Elizabeth O. Lichtenberger reports correlations of 0.86 for identical twins raised together compared to 0.76 for those raised apart and 0.47 for siblings. A 1994 review in *Behavior Genetics* based on identical/fraternal twin studies found that it is as high as 0.80 in general cognitive ability but it also varies based on the trait, with .60 for verbal tests, .50 for spatial and speed-of-processing tests, and only .40 for memory tests.

There are a number of points to consider when interpreting heritability:

- A high heritability does not mean that the environment has no effect on the development of a trait, or that learning is not involved. Vocabulary size, for example, is very substantially heritable (and highly correlated with general intelligence) although every word in an individual's vocabulary is learned. In a society in which plenty of words are available in everyone's environment, especially for individuals who are motivated to seek them out, the number of words that individuals actually learn depends to a considerable extent on their genetic predispositions.
- A common error is to assume that because something is heritable it is necessarily unchangeable. This is wrong. Heritability does not imply immutability. As previously noted, heritable traits can depend on learning, and they may be subject to other environmental effects as well. The value of heritability can change if the distribution of environments (or genes) in the population is substantially altered. For example, an impoverished or suppressive environment could fail to support the development of a trait, and hence restrict individual variation. This could affect estimates of heritability. Another example is Phenylketonuria which previously caused mental retardation for everyone who had this genetic disorder. Today, this can be prevented by following a modified diet.
- On the other hand, there can be effective environmental changes that do not change heritability at all. If the environment relevant to a given trait improves in a

way that affects all members of the population equally, the mean value of the trait will rise without any change in its heritability (because the differences among individuals in the population will stay the same). This has evidently happened for height: the heritability of stature is high, but average heights continue to increase.
- Even in developed nations, high heritability of a trait within a given group has no necessary implications for the source of a difference between groups.

- In addition to strong evidence for heritability increasing with age, some studies suggest that heritability increases with social class,.Differences among children with higher than average social status are almost entirely due to inherited differences, while among very low social class groups, most of the differences in IQ-scores (at least in children and young adolescents) are attributable to differences between families (shared environment).
- Even among close groups such as families, different individuals—such as siblings—will still experience different environments, which matter in determining intelligence scores.

Test score differences

Intelligence tests measure many important abilities, such as verbal and quantitative reasoning, and can predict socially-relevant outcomes such as academic performance and occupational outcomes. However, intelligence test scores do not reflect all of the intricacies of the everyday meaning of intelligence, so researchers take care to distinguish between IQ test results and intelligence.

Some studies of intelligence tests use statistical methods to extract so-called latent variables from the IQ test scores. One such variable is the general intelligence factor, or *g*, which accounts for most of the differences in IQ test scores between individuals. There are other latent variables in addition to *g*, and IQ tests vary in their ability to measure these latent variables, if they measure them at all. IQ tests scores, while often summarized as a single overall number, are actually multidimensional in nature. Transforming IQ test scores into latent variables is an attempt to find one or dimensions on which to compare IQ test scores.

IQ scores can vary substantially for the same person, even on tests taken at the same age. (IQ score table data and pupil pseudonyms adapted from description of KABC-II norming study cited in Kaufman 2009.)

Pupil	KABC-II	WISC-III	WJ-III
Asher	90	95	111
Brianna	125	110	105
Colin	100	93	101
Danica	116	127	118
Elpha	93	105	93
Fritz	106	105	105

Georgi	95	100	90
Hector	112	113	103
Imelda	104	96	97
Jose	101	99	86
Keoku	81	78	75
Leo	116	124	102

Latent variables are also sometimes called factors or constructs. The construct validity of an IQ test score is a key criteria for judging whether the IQ test score differences are meaningful. Tests which do not measure difference in latent variables for some group are said to have measurement bias. The construct validity of most commonly used IQ tests has been fairly well established within multiple racial-ethnic groups in developing countries such as the United States. That is, test score difference within each racial-ethnic group are valid indicators of differences in latent variables such as *g*. A related question is whether test-score differences between groups are valid. There is a consensus that test score differences between Black and White people in the United States have predictive validity (also called predictive invariance), meaning that test scores predict the same socially-relevant outcomes regardless of the race of the person being tested. To further address this question, three studies using sophisticated statistical techniques have shown that Black-White differences in IQ test scores are not a result of measurement bias (a criterion called measurement invariance).

These studies imply that Black-White IQ differences reflect very general differences in some underlying latent variables, but they are unable to differentiate precisely which latent variables differ under a variety of models. These studies were performed in response to previous investigations which suggested that Black-White IQ differences are primarily differences in *g* in particular.

United States

There are observed differences in average test score achievement between racial-ethnic groups, which vary depending on the populations studied and the type of tests used. Self-defined black and white United States citizens have been the subjects of the greatest number of studies. Black-white average IQ differences appear to increase with age, reaching an average of nearly 17 points by age 24, which is slightly more than one standard deviation. According to James Flynn and others, the overall average black-white gap has reduced by one-third over the course of the 20th century.

For example, the black men inducted into the U.S. armed forces during World War II averaged about 1.5 standard deviations below their white counterparts. This improvement is also reflected in black-white differences on school achievement tests, which have shrunk from about 1.2 to about 0.8 standard deviations. However, these improvements may have stalled for people born after the early 1970s.

The average black-white IQ difference also varies depending on test content. For example, two subsections of the WISC IQ test, known as forward and reverse digit-span, ask children to repeat a long series of numbers either forwards or backwards. The black-white difference on forward digit span is relatively small, while the difference on reverse digit span is relatively large. Across a battery of tests, the size of the black-white gap is correlated with the extent to which the tests measure the psychometric factor *g*, which also accounts for most of the variation in interindividual differences in IQ test performance. Gaps are seen in other tests of cognitive ability or aptitude, including university admission exams such as the SAT and GRE as well as employment tests for corporate settings and the military.

The IQ distributions of other racial and ethnic groups in the United States are less well studied. Hispanic and Native American populations, including Arctic Natives, tend to score worse on average than white populations but better on average than black populations. East Asian populations may score higher on average than white populations in the United States as they do elsewhere. A 1960 study of 1,236 American teenagers calculated six IQ measures for Jews relative to white gentiles. The results found that the relative IQ of American Jews varied from a low of 91.3 (visual reasoning) to a high of 109.7 (mathematics). A recent review by Lynn (2004) used a 10 word vocabulary test to estimate the IQ of American Jews. The population of 150 Jews scored half a standard deviation above the 5,300 white gentiles in verbal IQ.

For each of these populations, there is some evidence that the mixture of ability factors that distinguish individuals are differentially distributed between groups. For example, East Asian populations tend to outscore white populations in performance IQ, whereas the test score differences skew towards higher verbal IQ for Ashkenazi Jew-white differences. However, the mixture of abilities within groups appears to be nearly identical across many ethnic groups. The stability of these differences is also less well studied than black-white differences.

Worldwide

According to Richard Lynn, J. Philippe Rushton, and others, IQ test score differences are observed cross-culturally and around the world. Lynn has published three books summarizing IQ test scores from around the world. The inaccuracy of the cross cultural IQ scores is well documented, but many scholars use the results as an estimate of worldwide IQ scores. Lynn's meta-analysis lists East Asians (105), Europeans (99), Inuit (91), Southeast Asians and Amerindians (87 each), Pacific Islanders (85), South Asians/North Africans (84), Non-Bushmen sub-Saharan Africans (67), Australian Aborigines (62) and Bushmen (54). However, critics point out that different researchers using different data get different results. For example, James Flynn found out that Lynn's comparisons between nations were skewed and that the IQ values should be recalculated.

International achievement test scores, including TIMSS and PISA, have also been used to estimate average IQ worldwide with similar results where data is available.

The very low IQ scores reported for sub-Saharan African populations are especially controversial. For example, Wicherts argues that the average IQ of sub-Saharan Africans is poorly measured and is more likely 78. According to anthropologist Mark Cohen, the frequently reported African mean IQ of 70 is "preposterous". Using Western standards, this would mean that African countries evidencing such a low IQ would be largely dysfunctional. Given that individuals in these countries lead "vibrant artistic, symbolic and spiritual lives", this is, according to Cohen, clearly not the case. Thus, he concludes, the IQ test results from Africa do not reflect actual intelligence levels.

Differences in education, prolonged malnutrition, exposure to toxin, exposure to stress, and exposure to disease are all generally expected to contribute to the lower scores observed in developing countries. However, direct experimental evidence to confirm the role of individual factors is difficult to acquire in most cases because each of these factors tends to also be associated with one another and with unfavorable socioeconomic conditions. In the case of some toxins, such as lead, a negative effect on IQ scores has been established. Two other factors that have as well established negative association with IQ are severely premature birth and severe low birth weight.

Developing nations

Almost all studies on heritability have been in the developed world, mostly in the United States. In developing nations there are many environmental factors affecting IQ which are much less important in developed nations. Examples include nutrition, diseases, environmental toxins, and health care. For example, iodine deficiency causes a fall, in average, of 12 IQ points in China.

Issues in the calculations

Family environment

In the developed world, nearly all personality traits show that, contrary to some expectations, environmental effects actually cause non-related children raised in the same family ("adoptive siblings") to be as different as children raised in different families (Harris, 1998; Plomin & Daniels, 1987). There are some family effects on the IQ of children, accounting for up to a quarter of the variance. However, by adulthood, this correlation disappears, such that adoptive siblings are not more similar in IQ than strangers, while adult full siblings show an IQ correlation of 0.6. Twin studies reinforce this pattern: monozygotic (identical) twins raised separately are highly similar in IQ (0.86), more so than dizygotic (fraternal) twins raised together (0.6) and much more than adoptive siblings (~0.0).

The American Psychological Association's report *Intelligence: Knowns and Unknowns* (1995) states that there is no doubt that normal child development requires a certain minimum level of responsible care. Severely deprived, neglectful, or abusive environments must have negative effects on a great many aspects of development,

including intellectual aspects. Beyond that minimum, however, the role of family experience is in serious dispute.

There is no doubt that such variables as resources of the home and parents' use of language are correlated with children's IQ scores, but such correlations may be mediated by genetic as well as (or instead of) environmental factors. But how much of that variance in IQ results from differences between families, as contrasted with the varying experiences of different children in the same family? Recent twin and adoption studies suggest that while the effect of the family environment is substantial in early childhood, it becomes quite small by late adolescence. These findings suggest that differences in the life styles of families whatever their importance may be for many aspects of children's lives make little long-term difference for the skills measured by intelligence tests. It also stated

"We should note, however, that low-income and non-white families are poorly represented in existing adoption studies as well as in most twin samples. Thus it is not yet clear whether these studies apply to the population as a whole. It remains possible that, across the full range of income and ethnicity, between-family differences have more lasting consequences for psychometric intelligence."

A study of French children adopted between the ages of four and six shows the continuing interplay of nature and nurture. The children came from poor backgrounds with IQs that initially averaged 77, putting them near retardation. Nine years later after adoption, they retook the IQ tests, and all of them did better. The amount they improved was directly related to the adopting family's socioeconomic status. "Children adopted by farmers and laborers had average IQ scores of 85.5; those placed with middle-class families had average scores of 92. The average IQ scores of youngsters placed in well-to-do homes climbed more than 20 points, to 98."

Biased older studies?

Stoolmiller (1999) found that the range restriction of family environments that goes with adoption, that adopting families tend to be more similar on for example SES than the general population, means that role of the shared family environment have been underestimated in previous studies. Corrections for range applied to adoption studies indicate that SE could account for as much as 50% of the variance in IQ. However, the effect of restriction of range on IQ for adoption studies was examined by Matt McGue and colleagues, who write that "restriction in range in parent disinhibitory psychopathology and family SES had no effect on adoptive-sibling correlations [in] IQ".

Eric Turkheimer and colleagues (2003), not using an adoption study, included impoverished US families. Results demonstrated that the proportions of IQ variance attributable to genes and environment vary nonlinearly with SES. They suggest that the role of shared environmental factors may have been underestimated in older studies which often only studied affluent middle class families.

When comparing late 1970s to pre-1963 recorded data, researches DeFries and Plomin found that IQ correlation between parent and child living together fell significantly, from 0.50 to 0.35.

Maternal (fetal) environment

A meta-analysis, by Devlin and colleagues in *Nature* (1997), of 212 previous studies evaluated an alternative model for environmental influence and found that it fits the data better than the 'family-environments' model commonly used. The shared maternal (foetal) environment effects, often assumed to be negligible, account for 20% of covariance between twins and 5% between siblings, and the effects of genes are correspondingly reduced, with two measures of heritability being less than 50%. They argue that the shared maternal environment may explain the striking correlation between the IQs of twins, especially those of adult twins that were reared apart.

Bouchard and McGue reviewed the literature in 2003, arguing that Devlin's conclusions about the magnitude of heritability is not substantially different than previous reports and that their conclusions regarding prenatal effects stands in contradiction to many previous reports. They write that:

Chipuer et al. and Loehlin conclude that the postnatal rather than the prenatal environment is most important. The Devlin et al. (1997a) conclusion that the prenatal environment contributes to twin IQ similarity is especially remarkable given the existence of an extensive empirical literature on prenatal effects. Price (1950), in a comprehensive review published over 50 years ago, argued that almost all MZ twin prenatal effects produced differences rather than similarities. As of 1950 the literature on the topic was so large that the entire bibliography was not published. It was finally published in 1978 with an additional 260 references. At that time Price reiterated his earlier conclusion (Price, 1978).

Dickens and Flynn model

Dickens and Flynn (2001) argue that the arguments regarding the apparent absence of influence from shared family environment should apply equally well to groups separated in time. This is contradicted by the Flynn effect. Changes between generations have happened too quickly to be explained by genetics. This paradox can be explained by observing that heritability estimates include both a direct effect of the genotype on IQ and also indirect effects through which the genotype changes the environment, in turn effecting IQ. That is, those with a higher IQ tend to seek out stimulating environments that further increase IQ. The direct effect can initially have been very small but feedback loops can create large differences in IQ. In their model an environmental stimulus can have a very large effect on IQ, even in adults, but this effect also decays over time unless the stimulus continues. (The model is also adaptable to possible factors, such as nutrition in early childhood, that may cause permanent effects.) The Flynn effect can be explained by a generally more stimulating environment for all people. The authors suggest that programs aiming to increase IQ would be most likely to produce long-term IQ gains if

they taught children how to replicate outside the program the kinds of cognitively demanding experiences that produce IQ gains while they are in the program and motivate them to persist in that replication long after they have left the program.

Regression toward the mean

The heritability of IQ measures the extent to which the IQ of a child is measurably influenced by the IQ of its parents. As IQ is a quantifiable phenotype, one can estimate the expected IQ of child using the equation $\hat{y} = \bar{x} + h^2 \left(\frac{m+f}{2} - \bar{x} \right)$, where

- $\hat{y}$ is the expected IQ of the child,
- $\bar{x}$ is the mean IQ of the population to which the parents belong,
- h^2 is the heritability of IQ,
- m and f are the IQs of the mother and father, respectively.

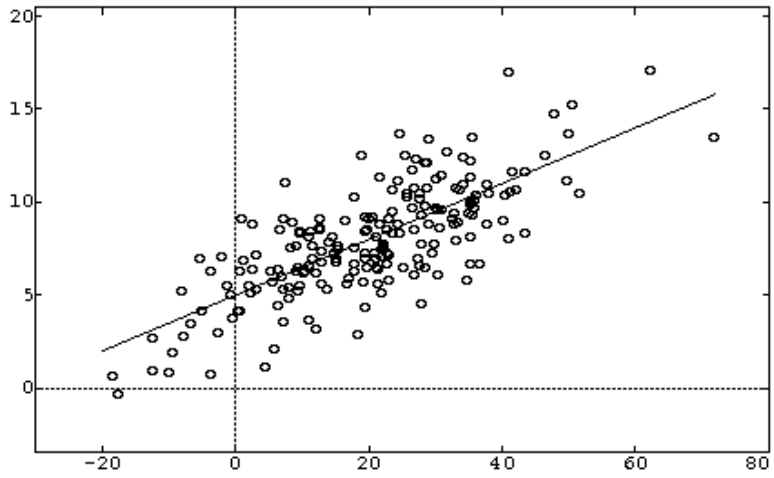


Illustration of linear regression on a data set

The equation asserts that, on average, the IQ of a child tends to the mean IQ of the population. For instance, if the heritability of IQ is 50% and the mean IQ of a population is 100, then a couple with an average IQ of 120 will, on average, have a child with an IQ of 110. Similarly, a couple with an average IQ of 80 will, on average, have a child with an IQ of 90.

It is noted that the above equation relates only statistical averages and is not deterministic. Furthermore, the equation is a general equation based in the inheritance of genetically-based characteristics (in this case, phenotypes), and so it is implicitly assumed that environmental factors are, for the sake of correctly assessing the genetic contribution to IQ, the same across the population.

Operating under the assumption that child and parent are raised in exactly the same environment (unlikely, but usually closer to the truth than in the completely dissimilar environment that the previous equation assumes), h^2 can be replaced by h, which is simply the correlation between parent and offspring IQ. In this case, regression towards the mean is no longer partially caused by environmental differences and therefore only by random genetic variation.

Finally, it is important to note that the expected IQ of the offspring is normally distributed around the mean calculated using the above equation, so in many cases regression towards the mean does not actually occur; as the values are normally distributed, there is a chance that offspring IQ will be more deviant from the mean than that of the parental average.

Historical research

As early as 1869, Francis Galton replaced mere speculations by statistical data through his book, *Hereditary Genius*:

Highly Gifted Men and the Percentage of their Highly Gifted Male Relatives
(classified by occupation and achievement)

	Galton	**Terman**	**Brimhall**		**Weiss**
	%	**%**	**%**	**%**	***n***
Probands	100	84[+]	100	97[+]	1972: *1329* 1994: *357*
Fathers	26	41	29	40	*346*
Brothers	47	–	49	49	*220*
Sons	60	64[*]	–	55	*77*
Grandfathers	14	–	9	9	*681*
Uncles	16	–	13	14	*615*
Nephews	23	–	–	22	*76*
Grandsons	14	–	–	–	–
Greatgrandfathers	0	–	–	4	*1290*
Uncles of the parents	5	–	–	5	*1996*
Cousins	16	–	9[#]	18	*570*
Greatgrandsons	7	–	–	–	–
Cousins of parents	–	–	–	11	*2250*

"+": classified by occupation; 100%, if classified by test
"*": classified only by IQ; classification by occupation gives about 55%; $n = 820$.
"#": some cousins were still too young and did not have full opportunity to become distinguished
"–": no data

Despite the differences in methods and societies, there is a notable parallelism in the published statistics. The ITO-method by Li and Sacks (1954) allows from this set of data the estimation of the underlying number of genes and their allele frequencies.

The inheritance of cognitive deficits

There are many genetic variants known to cause lower IQ. The number of such mutations already known is in the hundreds. For example, an allele of the gene GDI1 is associated with an IQ below 70.

Copy number variation has also been associated with idiopathic learning disability.

There are number of known cases where the homozygotes have severe cognitive deficits and the heterozygotes show a small decrease of IQ. In such cases further alleles are investigated to estimate their influence on IQ. For example, one minor allele of the gene ALDH5A1 is associated with an IQ difference of around 1.5 points.

Interindividual (between individuals) differences in learning ability are also known in mice, dogs and other animals, and the achievements of pure strains can be improved by selective breeding. In such a way also behavior genetics is contributing to our knowledge of the inheritance of mental traits.

The search for specific genes

Unfortunately, most of the research done about the heritability of intelligence have focused on children and young adults. Thus, the role of genetic factors to intelligence is mostly unknown. Many studies attempting to find loci in the genome relating to IQ have had little success. For example, a study by Robert Plomin using groups of around 100 people investigated 1,842 DNA markers in a high-IQ group and in an average-IQ control group. The study used a five-step replication process to eliminate false positives, and no gene met this rigid criterion for replicability.

The failure to find a specific gene associated with IQ indicates that cognitive abilities are very complex and are likely to involve several genes (polygenic). Some estimate that as much as 40% of all genes may contribute to IQ. The more genes that contribute to a trait the more the trait will be continuous instead of discrete. A 2008 study of 500,000 single nucleotide polymorphisms (SNPs) from 7,089 children did not substantially improve on earlier studies. The study did not find any SNPs that accounted for more than 0.5% of the variance in general intelligence.

A 2007 study did find that a gene called FADS2 along with breastfeeding adds about 7 IQ points to those with the "C" version of the gene. Those with the "G" version see no advantage.

There is "a highly significant association" between the CHRM2 gene and intelligence according to a 2006 Dutch family study. The study concluded that there was an

association between the CHRM2 gene on chromosome 7 and Performance IQ, as measured by the Wechsler Adult Intelligence Scale-Revised. The Dutch family study used a sample of 667 individuals from 304 families. A similar association was found independently in the *Minnesota Twin and Family Study* (Comings et al. 2003) and by the Department of Psychiatry at the Washington University. Microcephalin and ASPM are two genes that are associated with brain development. Mutations in these genes are associated with microcephaly, and hence they were initially associated with general intelligence. However recent studies have found no association with general cognitive abilities.

STX1A correlates significantly with intelligence in Williams syndrome patients.

Between-group heritability

Although IQ differences between individuals are shown to have a large genetic component, it does not automatically follow that mean group-level disparities (between-group differences) in IQ can be assumed to have a genetic basis. An analogy, attributed to Richard Lewontin, illustrates this point:

Suppose two handfuls are taken from a sack containing a genetically diverse variety of corn, and each grown under carefully controlled and standardized conditions, except that one batch is lacking in certain nutrients that are supplied to the other. After several weeks, the plants are measured. There is variability of growth within each batch, due to the genetic variability of the corn. Given that the growing conditions are closely controlled, nearly all the variation in the height of the plants within a batch will be due to differences in their genes. Thus, within populations, heritabilities will be very high. Nevertheless, the difference between the two groups is due entirely to an environmental factor - differential nutrition. Lewontin didn't go so far as to have the one set of pots painted white and the other set black, but you get the idea. The point of the example, in any case, is that the causes of between-group differences may in principle be quite different from the causes of within-group variation.

This nurture argument holds an intuitive appeal but is misleading since modern statistical comparisons control for such factors as differences in socio-economic and parental educational backgrounds specifically to eliminate such environmental effects and thereby isolate the genetic influence. In fact, differences in IQ between reportedly discrete genetic groups has been observed, and researchers, such as Arthur Jensen, maintain that environmental differences are too small to account for these differences. They propose, therefore, that genetic differences must provide the primary explanation.

This view is challenged by Peter Schönemann who claims that Arthur Jensen and others routinely confuse the first principal component (PC1) with *g* as Charles Spearman defined it. Schönemann argues that the high IQ heritability estimates reported in the literature derive from restrictive formal models whose underlying assumptions are rarely tested and usually violated by the data. Jensen's view is also rebutted by James Flynn in his book *What Is Intelligence*.

The issue of observed between-group IQ differences is controversial *vis-a-vis* considerations on both the nature of race and the meaning and measurement of intelligence. Important related questions include whether intelligence can be accurately described by a single number, and whether the nature of intelligence is the same across cultures.

Gene-by-Environment Interaction

Researchers have recently begun to empiricaly examine the hypothesis that genetic influences on IQ may depend on environmental inputs.

Eric Turkheimer and colleagues (2003) studied the heritability of IQ in a sample that included a substantial proportion of impoverished US families. Results demonstrated that, in seven-year-old twins, the proportions of IQ variance attributable to genes and environment vary nonlinearly with socioeconomic status. In impoverished families, 60% of the variance in early childhood IQ was accounted for by the shared family environment, and the contribution of genes is close to zero; in affluent families, the result is almost exactly the reverse. They suggest that the role of shared environmental factors may have been underestimated in older studies which often only studied affluent middle class families.

Harden and colleagues (2007) found a similar gene-environment interaction for adolescents. They found that, among higher income families, genetic influences accounted for approximately 55% of the variance in cognitive aptitude and shared environmental influences about 35%. Among lower income families, the proportions were in the reverse direction, 39% genetic and 45% shared environment."

Using a nationally representative sample of young twins, Tucker-Drob and colleagues (2011) found that genetic influences on mental ability emerged over the course of infancy, with larger increases in genetic variance for children living in higher socioeconomic status homes. At 10 months of age genetic influences on mental ability were negligible for children from all socioeconomic backgrounds. By 2 years of age, genes accounted for nearly 50% of the variation in mental ability of children raised in high-SES homes, but genes continued to account for negligible variation in mental ability of children raised in low-SES homes. These results suggest that family resources help to potentiate children's genetic capacities for intellectual development.

A 2007 study by Caspi and colleagues found that a gene called FADS2 along with breastfeeding adds about 7 IQ points to those with the "C" version of the gene. Those with the "G" version see no advantage.

Chapter- 6

Genomics

Genomics is a discipline in genetics concerning the study of the genomes of organisms. The field includes intensive efforts to determine the entire DNA sequence of organisms and fine-scale genetic mapping efforts. The field also includes studies of intragenomic phenomena such as heterosis, epistasis, pleiotropy and other interactions between loci and alleles within the genome. In contrast, the investigation of the roles and functions of single genes is a primary focus of molecular biology or genetics and is a common topic of modern medical and biological research. Research of single genes does not fall into the definition of genomics unless the aim of this genetic, pathway, and functional information analysis is to elucidate its effect on, place in, and response to the entire genome's networks.

For the United States Environmental Protection Agency, "the term "genomics" encompasses a broader scope of scientific inquiry associated technologies than when genomics was initially considered. A genome is the sum total of all an individual organism's genes. Thus, genomics is the study of all the genes of a cell, or tissue, at the DNA (genotype), mRNA (transcriptome), or protein (proteome) levels."

History

The first genomes to be sequenced were those of a virus and a mitochondrion, and were done by Fred Sanger. His group established techniques of sequencing, genome mapping, data storage, and bioinformatic analyses in the 1970-1980s. A major branch of genomics is still concerned with sequencing the genomes of various organisms, but the knowledge of full genomes has created the possibility for the field of functional genomics, mainly concerned with patterns of gene expression during various conditions. The most important tools here are microarrays and bioinformatics. Study of the full set of proteins in a cell type or tissue, and the changes during various conditions, is called proteomics. A related concept is materiomics, which is defined as the study of the material properties of biological materials (e.g. hierarchical protein structures and materials, mineralized biological tissues, etc.) and their effect on the macroscopic function and failure in their biological context, linking processes, structure and properties at multiple scales through a materials science approach. The actual term 'genomics' is thought to have been coined by

Dr. Tom Roderick, a geneticist at the Jackson Laboratory (Bar Harbor, ME) over beer at a meeting held in Maryland on the mapping of the human genome in 1986.

In 1972, Walter Fiers and his team at the Laboratory of Molecular Biology of the University of Ghent (Ghent, Belgium) were the first to determine the sequence of a gene: the gene for Bacteriophage MS2 coat protein. In 1976, the team determined the complete nucleotide-sequence of bacteriophage MS2-RNA. The first DNA-based genome to be sequenced in its entirety was that of bacteriophage Φ-X174; (5,368 bp), sequenced by Frederick Sanger in 1977.

The first free-living organism to be sequenced was that of *Haemophilus influenzae* (1.8 Mb) in 1995, and since then genomes are being sequenced at a rapid pace.

As of September 2007, the complete sequence was known of about 1879 viruses, 577 bacterial species and roughly 23 eukaryote organisms, of which about half are fungi. Most of the bacteria whose genomes have been completely sequenced are problematic disease-causing agents, such as *Haemophilus influenzae*. Of the other sequenced species, most were chosen because they were well-studied model organisms or promised to become good models. Yeast (*Saccharomyces cerevisiae*) has long been an important model organism for the eukaryotic cell, while the fruit fly *Drosophila melanogaster* has been a very important tool (notably in early pre-molecular genetics). The worm *Caenorhabditis elegans* is an often used simple model for multicellular organisms. The zebrafish *Brachydanio rerio* is used for many developmental studies on the molecular level and the flower *Arabidopsis thaliana* is a model organism for flowering plants. The Japanese pufferfish (*Takifugu rubripes*) and the spotted green pufferfish (*Tetraodon nigroviridis*) are interesting because of their small and compact genomes, containing very little non-coding DNA compared to most species. The mammals dog (*Canis familiaris*), brown rat (*Rattus norvegicus*), mouse (*Mus musculus*), and chimpanzee (*Pan troglodytes*) are all important model animals in medical research.

Human genomics

A rough draft of the human genome was completed by the Human Genome Project in early 2001, creating much fanfare. By 2007 the human sequence was declared "finished" (less than one error in 20,000 bases and all chromosomes assembled). Display of the results of the project required significant bioinformatics resources. The sequence of the human reference assembly can be explored using the UCSC Genome Browser.

Bacteriophage genomics

Bacteriophages have played and continue to play a key role in bacterial genetics and molecular biology. Historically, they were used to define gene structure and gene regulation. Also the first genome to be sequenced was a bacteriophage. However, bacteriophage research did not lead the genomics revolution, which is clearly dominated by bacterial genomics. Only very recently has the study of bacteriophage genomes

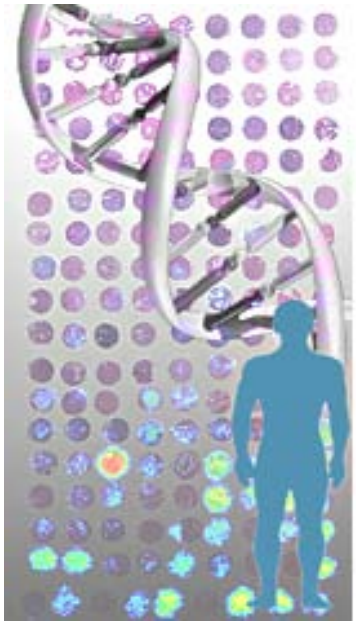
become prominent, thereby enabling researchers to understand the mechanisms underlying phage evolution. Bacteriophage genome sequences can be obtained through direct sequencing of isolated bacteriophages, but can also be derived as part of microbial genomes. Analysis of bacterial genomes has shown that a substantial amount of microbial DNA consists of prophage sequences and prophage-like elements. A detailed database mining of these sequences offers insights into the role of prophages in shaping the bacterial genome.

Cyanobacteria genomics

At present there are 24 cyanobacteria for which a total genome sequence is available. 15 of these cyanobacteria come from the marine environment. These are six *Prochlorococcus* strains, seven marine *Synechococcus* strains, *Trichodesmium erythraeum* IMS101 and *Crocosphaera watsonii* WH8501. Several studies have demonstrated how these sequences could be used very successfully to infer important ecological and physiological characteristics of marine cyanobacteria. However, there are many more genome projects currently in progress, amongst those there are further *Prochlorococcus* and marine *Synechococcus* isolates, *Acaryochloris* and *Prochloron*, the N_2-fixing filamentous cyanobacteria *Nodularia spumigena*, *Lyngbya aestuarii* and *Lyngbya majuscula*, as well as bacteriophages infecting marine cyanobaceria. Thus, the growing body of genome information can also be tapped in a more general way to address global problems by applying a comparative approach. Some new and exciting examples of progress in this field are the identification of genes for regulatory RNAs, insights into the evolutionary origin of photosynthesis, or estimation of the contribution of horizontal gene transfer to the genomes that have been analyzed.

Chapter- 7

Human Genetics



A small piece of human DNA

Human genetics describes the study of inheritance as it occurs in human beings. Human genetics encompasses a variety of overlapping fields including: classical genetics, cytogenetics, molecular genetics, biochemical genetics, genomics, population genetics, developmental genetics, clinical genetics, and genetic counseling. Genes can be the common factor of the qualities of most human-inherited traits. Study of human genetics can be useful as it can answer questions about human nature, understand the diseases and development of effective disease treatment, and understand genetics of human life.

Genetic differences and inheritance patterns

Inheritance of traits for humans are based upon Gregor Mendel's model of inheritance. Mendel deduced that inheritance depends upon discrete units of inheritance, called factors or genes.

Autosomal dominant inheritance

Autosomal traits are associated with a single gene on an autosome (non-sex chromosome)—they are called "dominant" because a single copy—inherited from either parent—is enough to cause this trait to appear. This often means that one of the parents must also have the same trait, unless it has arisen due to a new mutation. Examples of autosomal dominant traits and disorders are Huntington's disease, and achondroplasia.

Autosomal recessive inheritance

Autosomal recessive traits is one pattern of inheritance for a trait, disease, or disorder to be passed on through families. For a recessive trait or disease to be displayed two copies of the trait or disorder needs to be presented. The trait or gene will be located on a non-sex chromosome. Because it takes two copies of a trait to display a trait, many people can unknowingly be carriers of a disease. From an evolutionary perspective, a recessive disease or trait can remain hidden for several generations before displaying the phenotype. Examples of autosomal recessive disorders are albinism, Cystic Fibrosis, Tay-Sachs disease.

X-linked and Y-linked inheritance

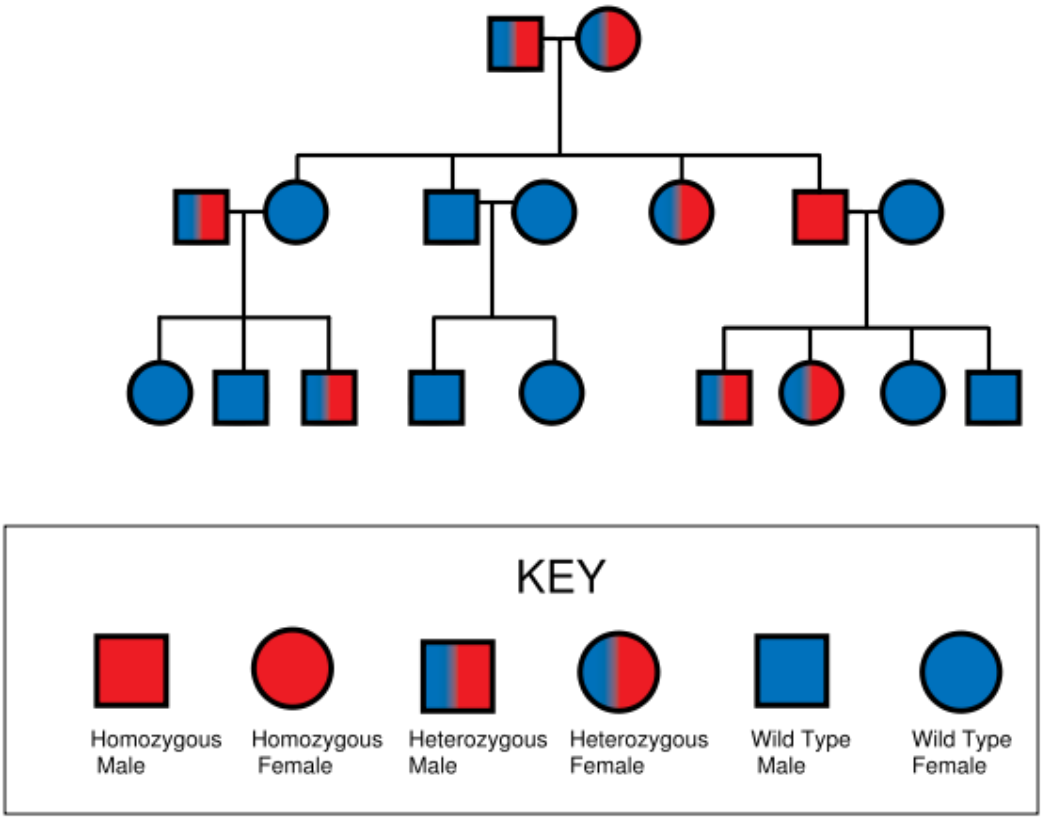
X-linked genes are found on the sex X chromosome. X-linked genes just like autosomal genes have both dominant and recessive types. Recessive X-linked disorders are rarely seen in females and usually only affect males. This is because males inherit their X chromosome and all X-linked genes will be inherited from the maternal side. Fathers only pass on their Y chromosome to their sons, so no X-linked traits will be inherited from father to son. Females express X-linked disorders when they are homozygous for the disorder and become carriers when they are heterozygous. X-linked dominant inheritance will show the same phenotype as a heterozygote and homozygote. Just like X-linked inheritance, there will be a lack of male-to-male inheritance, which makes it distinguishable from autosomal traits. One example of a X-linked trait is Coffin-Lowry syndrome, which is caused by a mutation in ribosomal protein gene. This mutation results in skeletal, craniofacial abnormalities, mental retardation, and short stature.

X chromosomes in females undergo a process known as X inactivation. X inactivation is when one of the two X chromosomes in females is almost completely inactivated. It is important that this process occurs otherwise a woman would produce twice the amount of normal X chromosome proteins. The mechanism for X inactivation will occur during the embryonic stage. For people with disorders like trisomy X, where the genotype has three X chromosomes, X-inactivation will inactivate all X chromosomes until there is only one X chromosome active. X inactivation is not only limited to females, males with Klinefelter syndrome, who have an extra X chromosome, will also undergo X inactivation to have only one completely active X chromosome.

Y-linked inheritance occurs when a gene, trait, or disorder is transferred through the Y chromosome. Since Y chromosomes can only be found in males, Y linked traits are only

passed on from father to son. The testis determining factor, which is located on the Y chromosome, determines the maleness of individuals. Besides the maleness inherited in the Y-chromosome there are no other found Y-linked characteristics.

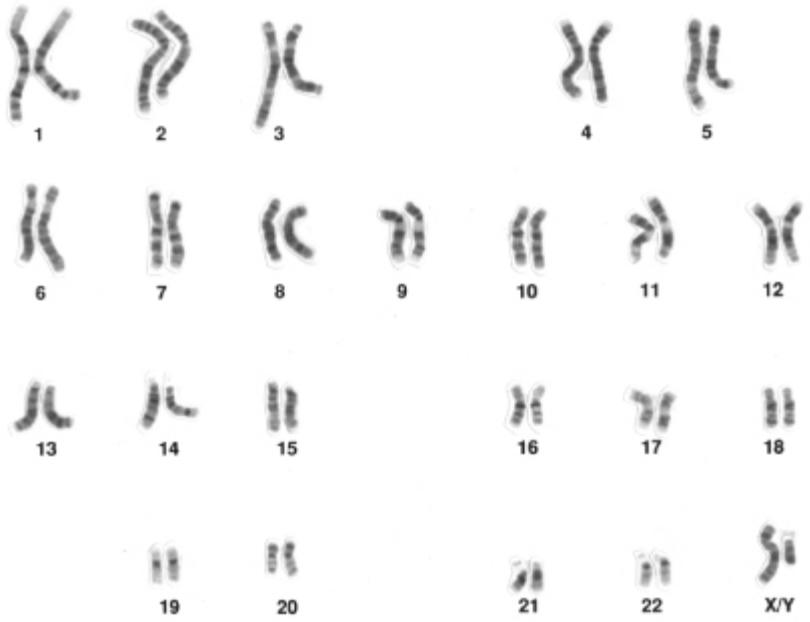
Pedigrees



An example of a family pedigree displaying an autosomal recessive trait

A pedigree is a diagram showing the ancestral relationships and transmission of genetic traits over several generations in a family. Pedigrees are used to help detect many different genetic diseases. A pedigree can also be used to help determine the chances for a parent to produce an offspring with a specific trait. Four different traits can be identified by pedigree chart analysis: autosomal dominant, autosomal recessive, x-linked, or y-linked. Partial penetrance can be shown and calculated form pedigrees. Penetrance is the percentage expressed frequency with which individuals of a given genotype manifest at

least some degree of a specific mutant phenotype associated with a trait. Inbreeding, the mating between closely related organisms of traits can clearly be seen on pedigree charts. Pedigree charts of royal families have a high degree of inbreeding, because it was customary and preferable for royalty to marry another member of royalty. Genetic councilors commonly use pedigrees to help couple determine if the parents will be able to produce healthy children.



A karyotype of a human male, showing 46 chromosomes including XY sex chromosomes

Karyotype

A karyotype is a very useful tool in cytogenetics. A karyotype is picture of all the chromosomes in the metaphase stage arranged according to length and centromere position. A karyotype can also be useful in clinical genetics, due to its ability to diagnose genetic disorders. On a normal karyotype, aneuploidy can be detected by clearly being able to observe any missing or extra chromosomes. Giemsa banding, g-banding, of the karyotype can be used to detect deletions, insertions, duplications, inversions, and translocations. G-banding will stain the chromosomes with light and dark bands unique to each chromosome. A FISH, fluorescent in situ hybridization, can be used to observe deletions, insertions, and translocations. FISH uses fluorescent probes to bind to specific sequences of the chromosomes that will cause the chromosomes to fluoresce a unique color.

Genomics

Genomics refers to the field of genetics concerned with structural and functional studies of the genome. A genome is all the DNA contained within an organism or a cell

including nuclear and mitochondrial DNA. The human genome is the total collection of genes in a human being contained in the human chromosome, composed of over three billion nucleotides. In April 2003, the Human Genome Project was able to sequence all the DNA in the human genome, to discover the human genome was composed around 20,000 protein coding genes.

Population genetics

Population genetics is the branch of evolutionary biology responsible for investigating processes that cause changes in allele and genotype frequencies in populations based upon Mendelian inheritance. Four different forces can influence the frequencies: natural selection, mutation, gene flow (migration), and genetic drift. A population can be defined as a group of interbreeding individuals and their offspring. For human genetics the populations will consist only of the human species. The Hardy-Weinberg principle is a widely used principle to determine allelic and genotype frequencies.

Hardy-Weinberg principle

The Hardy-Weinberg principle states that when no evolution occurs in a population the allele and genotype frequencies do not change from one generation to the next. No evolution refers to no mutation, no gene flow, no natural selection, and no genetic drift. To be in equilibrium two more assumptions need to be made that random mating occurs and there are discrete, non-overlapping generations.

Mitochondrial DNA

In addition to nuclear DNA, humans (like almost all eukaryotes) have mitochondrial DNA. Mitochondria, the "power houses" of a cell, have their own DNA because they are descended from a proteobacterium that merged with eukaryotic cells over 2 billion years ago—an assertion known as the endosymbiotic hypothesis. Mitochondria are inherited from one's mother, and its DNA is frequently used to trace maternal lines of descent. Mitochondrial DNA is only 16kb in length and encodes for 62 genes.

Genes and human characteristics

Genes are a fundamental unit of inheritance. Genes can be defined as a sequence of DNA in the genome that is required for production of a functional product. Genes have both minor and major effects on human characteristics. Human genes have become prominent in the nature versus nurture debate.

Genes and behavior

Genes have a strong influence on human behavior. IQ is largely heritable. However, this has been questioned. The stance that humans inherit substantial behavioral characteristics

is called psychological nativism, compared to the stance that human behavior and culture are virtually entirely constructed (tabula rasa).

In the early 20th century, eugenics was policy in parts of the United States and Europe. The goal was to reduce or eliminate traits that were considered undesirable. One form of eugenics was compulsory sterilization of people deemed mentally unfit. Hitler's eugenics programs turned social consciousness against the practice, and psychological nativism became associated with racism and sexism.

Evolutionary psychology

Evolutionary psychology explains many human behaviors as more or less moderated by genes that evolved in the hunter-gatherer stage of human cultural development.

Genetic disorders

Humans have several genetic diseases, often blamed on rare recessive genes. A few examples of human genetic diseases are: Turner Syndrome, Huntington's disease, Downs Syndrome (in some cases), and sickle cell anemia.

- Cri du Chat syndrome – A disorder caused from a deletion on the short arm of chromosome 5. This deletion results in a phenotype of mental retardation, behavioral problems, and a cat like call. About one in every 50,000 births will have the syndrome.
- Huntington's disease – A neurological disorder caused by a trinucleotide repeat sequence. Huntingtons is an autosomal dominant trait. Most individuals with the disease will first display the phenotype around 40 years of age. The symptoms are jerky uncontrollable movements, mental retardation, and behavioral problems.
- Turner syndrome – A condition that affects females caused by a 45, XO genotype instead of the normal XX genotype. These individuals have only one X chromosome. These individuals are phenotypically female, but will be sterile due to undeveloped ovaries.
- Klinefelter syndrome – A disorder in males caused by the presence of an extra X chromosome. These individuals have a genotype of 47, XXY instead of the normal XY genotype. The symptoms for this syndrome are enlarged breasts, small testes, and sterility.

Human traits with simple inheritance patterns

Dominant	Recessive
Widow's peak	No Widow's peak
Facial Dimples	No Facial Dimples
Able to taste PTC	Unable to taste PTC
Unattached earlobe	Attached earlobe
Cleft chin	No Cleft chin

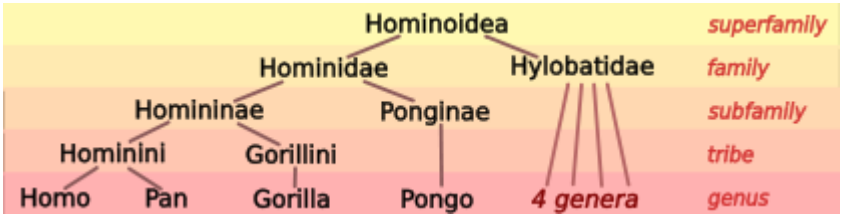
Brunette iris (anatomy)	Blue Iris (anatomy)
Color Vision	Color blindness
Brown Hair	Blonde Hair
normal	turned up nose
Ability to roll tongue (Able to hold tongue in a U shape)	No ability to roll tongue
Normal Pinkies	Crooked Pinkies
Normal Thumb	Hitchhiker's Thumb
Freckles	No Freckles
Wet-type earwax	Dry-type earwax
Curly Hair	Straight Hair

Chapter- 8

Human Evolutionary Genetics

Human evolutionary genetics studies how one human genome differs from the other, the evolutionary past that gave rise to it, and its current effects. Differences between genomes have anthropological, medical and forensic implications and applications. Genetic data can provide important insight into human evolution.

Origin of apes

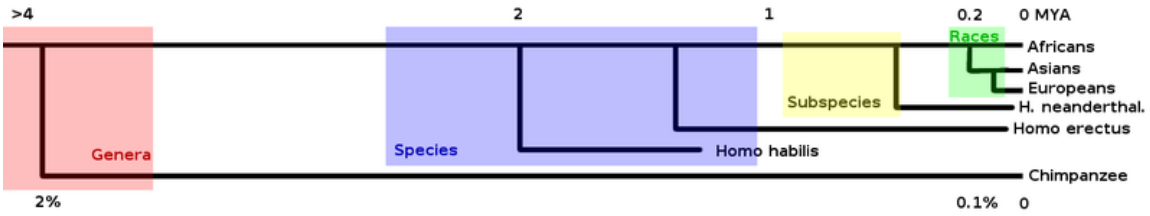


Taxonomic relationships of hominoids

Biologists classify humans, along with only a few other species, as great apes (species in the family Hominidae). The Hominidae include two distinct species of chimpanzee (the bonobo, *Pan paniscus*, and the common chimpanzee, *Pan troglodytes*), two species of gorilla (the western gorilla, *Gorilla gorilla*, and the eastern gorilla, *Gorilla graueri*), and two species of orangutan (the Bornean orangutan, *Pongo pygmaeus*, and the Sumatran orangutan, *Pongo abelii*).

Apes, in turn, belong to the primates order (>400 species). Data from both mitochondrial DNA (mtDNA) and nuclear DNA (nDNA) indicates that primates belong to the group of Euarchontoglires, together with Rodentia, Lagomorpha, Dermoptera, and Scandentia. This is further supported by Alu-like short interspersed nuclear elements (SINEs) which have been found only in members of the Euarchontoglires.

Cladistics



A phylogenetic tree like the one shown above is usually derived from DNA or protein sequences from populations. Often mitochondrial DNA or Y chromosome sequences are used to study ancient human demographics. These single-locus sources of DNA do not recombine and are almost always inherited from a single parent, with only one known exception in mtDNA (Schwartz and Vissing 2002). Individuals from the various continental groups tend to be more similar to one another than to people from other continents. The tree is rooted in the common ancestor of chimpanzees and humans, which is believed to have originated in Africa. Horizontal distance in the diagram corresponds to two things:

1. **Genetic distance**. Given below the diagram, the genetic difference between humans and chimps is less than 2%, or 20 times larger than the variation among modern humans.
2. **Temporal remoteness** of the most recent common ancestor. Rough estimates are given above the diagram, in millions of years. The mitochondrial most recent common ancestor of modern humans lived roughly 200,000 years ago, latest common ancestors of humans and chimps between four and seven million years ago.

Chimpanzees and humans belong to different genera, indicated in red. Formation of species and subspecies is also indicated, and the formation of "races" is indicated in the green rectangle to the right (note that only a very rough representation of human phylogeny is given). Note that vertical distances are not meaningful in this representation.

Speciation of humans and the African apes

The separation of humans from their closest relatives, the African apes (chimpanzees and gorillas), has been studied extensively for more than a century. Five major questions have been addressed:

- Which apes are our closest ancestors?
- When did the separations occur?
- What was the effective population size of the common ancestor before the split?
- Are there traces of population structure (subpopulations) preceding the speciation or partial admixture succeeding it?

- What were the specific events (including fusion of chromosomes 2a and 2b) prior to and subsequent to the separation?

General observations

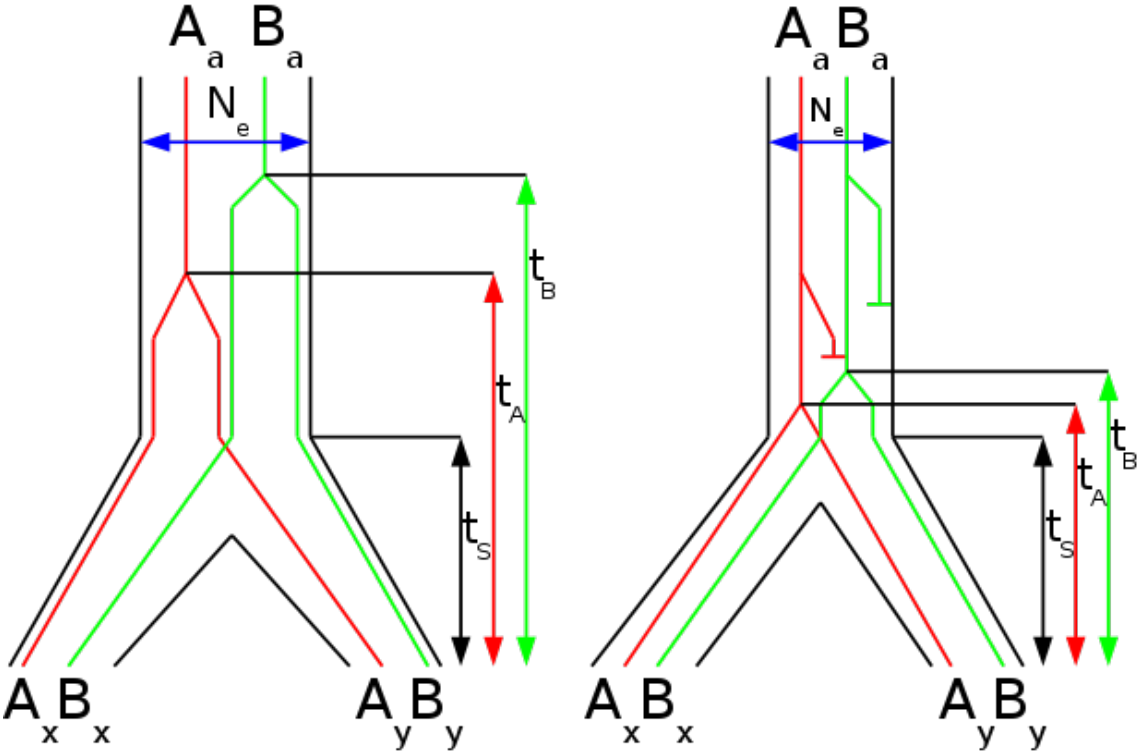
As discussed before, different parts of the genome show different sequence divergence between different hominoids. It has also been shown that the sequence divergence between DNA from humans and chimpanzees varies greatly. For example the sequence divergence varies between 0% to 2.66% between non-coding, non-repetitive genomic regions of humans and chimpanzees. Additionally gene trees, generated by comparative analysis of DNA segments, do not always fit the species tree. Summing up:

- The sequence divergence varies significantly between humans, chimpanzees and gorillas.
- For most DNA sequences, humans and chimpanzees appear to be most closely related, but some point to a human-gorilla or chimpanzee-gorilla clade.
- The human genome has been sequenced, as well as the chimpanzee genome. Humans have 23 pairs of chromosomes, while chimpanzees, gorillas, and orangutans have 24. Human chromosome 2 is a fusion between two chromosomes that remained separate in the other primates.

Divergence times

The divergence time of humans from other apes is of great interest. One of the first molecular studies, published in 1967 measured immunological distances (IDs) between different primates. Basically the study measured the strength of immunological response that an antigen from one species (human albumin) induces in the immune system of another species (human, chimpanzee, gorilla and Old World monkeys). Closely related species should have similar antigens and therefore weaker immunological response to each other's antigens. The immunological response of a species to its own antigens (e.g. human to human) was set to be 1. The ID between humans and gorillas was determined to be 1.09, that between humans and chimpanzees was determined as 1.14. However the distance to six different Old World monkeys was on average 2.46 indicating that the African apes are far closer related to humans than to monkeys. The authors consider the divergence time between Old World monkeys and hominoids to be 30 million years ago (MYA), based on fossil data, and the immunological distance was considered to grow at a constant rate. They concluded that divergence time of humans and the African apes to be roughly ~5 MYA. That was a surprising result. Most scientists at that time thought that humans and great apes diverged much earlier (>15 MYA). The gorilla was, in ID terms, closer to human than to chimpanzees, however the difference was so slight that the trichotomy could not be resolved with certainty. Later studies based on molecular genetics were able to resolve the trichotomy: chimpanzees are phylogenetically closer to humans than to gorillas. However, the divergence times estimated later (using much more sophisticated methods in molecular genetics) do not substantially differ from the very first estimate in 1967.

Divergence times and ancestral effective population size

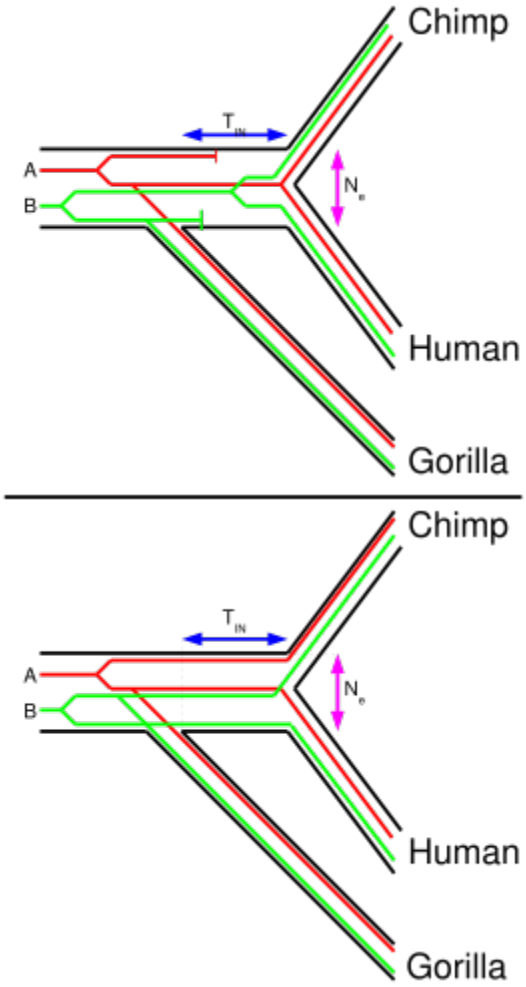


The sequences of the DNA segments diverge earlier than the species. A large effective population size in the ancestral population (left) preserves different variants of the DNA segments (=alleles) for a longer period of time. Therefore, on average, the gene divergence times (t_A for DNA segment A; t_B for DNA segment B) will deviate more from the time the species diverge (t_S) compared to a small ancestral effective population size (right).

Current methods to determine divergence times use DNA sequence alignments and molecular clocks. Usually the molecular clock is calibrated assuming that the orangutan split from the African apes (including humans) 12-16 MYA. Some studies also include some old world monkeys and set the divergence time of them from hominoids to 25-30 MYA. Both calibration points are based on very little fossil data and have been criticized. If these dates are revised, the divergence times estimated from molecular data will change as well. However, the relative divergence times are unlikely to change. Even if we can't tell absolute divergence times exactly, we can be pretty sure that the divergence time between chimpanzees and humans is about sixfold shorter than between chimpanzees (or humans) and monkeys.

One study (Takahata *et al.*, 1995) used 15 DNA sequence from different regions of the genome from human and chimpanzee and 7 DNA sequences from human, chimpanzee and gorilla. They determined that chimpanzees are more closely related to humans than gorillas. Using various statistical methods, they estimated the divergence time human-chimp to be 4.7 MYA and the divergence time between gorillas and humans (and

chimps) to be 7.2 MYA. Additionally they estimated the effective population size of the common ancestor of humans and chimpanzees to be ~100,000. This was somewhat surprising since the present day effective population size of humans is estimate to be only ~10,000. If true that means that the human lineage would have experienced an immense decrease of its effective population size (and thus genetic diversity) in its evolution.



A and B are two different loci. In the upper figure they fit to the species tree. The DNA that is present in today's gorillas diverged earlier from the DNA that is present in today's humans and chimps. Thus both loci should be more similar between human and chimp than between gorilla and chimp or gorilla and human. In the lower graph, locus A has a more recent common ancestor in human and gorilla compared to the chimp sequence. Whereas chimp and gorilla have a more recent common ancestor for locus B. Here the gene trees are incongruent to the species tree.

Another study (Chen & Li, 2001) sequenced 53 non-repetitive, intergenic DNA segments from a human, a chimpanzee, a gorilla, and orangutan. When the DNA sequences were concatenated to a single long sequence, the generated neighbor-joining tree supported the *Homo-Pan* clade with 100% bootstrap (that is that humans and chimpanzees are the

closest related species of the four). When three species are fairly closely related to each other (like human, chimpanzee and gorilla), the trees obtained from DNA sequence data may not be congruent with the tree that represents the speciation (species tree). The shorter internodal time span (T_{IN}) the more common are incongruent gene trees. The effective population size (N_e) of the internodal population determines how long genetic lineages are preserved in the population. A higher effective population size causes more incongruent gene trees. Therefore, if the internodal time span is known, the ancestral effective population size of the common ancestor of humans and chimpanzees can be calculated.

When each segment was analyzed individually, 31 supported the *Homo-Pan* clade, 10 supported the *Homo-Gorilla* clade, and 12 supported the *Pan-Gorilla* clade. Using the molecular clock the authors estimated that gorillas split up first 6.2-8.4 MYA and chimpanzees and humans split up 1.6-2.2 million years later (internodal time span) 4.6-6.2 MYA. The internodal time span is useful to estimate the ancestral effective population size of the common ancestor of humans and chimpanzees.

A parsimonious analysis revealed that 24 loci supported the *Homo-Pan* clade, 7 supported the *Homo-Gorilla* clade, 2 supported the *Pan-Gorilla* clade and 20 gave no resolution. Additionally they took 35 protein coding loci from databases. Of these 12 supported the *Homo-Pan* clade, 3 the *Homo-Gorilla* clade, 4 the *Pan-Gorilla* clade and 16 gave no resolution. Therefore only ~70% of the 52 loci that gave a resolution (33 intergenic, 19 protein coding) support the 'correct' species tree. From the fraction of loci which did not support the species tree and the internodal time span they estimated previously, the effective population of the common ancestor of humans and chimpanzees was estimated to be ~52 000 to 96 000. This value is not as high as that from the first study (Takahata), but still much higher than present day effective population size of humans.

A third study (Yang, 2002) used the same dataset that Chen and Li used but estimated the ancestral effective population of 'only' ~12,000 to 21,000, using a different statistical method.

Genetic differences between humans and other great apes

The alignable sequences within genomes of humans and chimpanzees differ by about 35 million single nucleotide substitutions. Additionally about 3% of the complete genomes differ by deletions, insertions and duplications.

Since mutation rate is relatively constant, roughly one half of these changes occurred in the human lineage. Only a very tiny fraction of those fixed differences gave rise to the different phenotypes of humans and chimpanzees and finding those is a great challenge. The vast majority of the differences are neutral and do not affect the phenotype.

Molecular evolution may act in different ways, through protein evolution, gene loss, differential gene regulation and RNA evolution. All are thought to have played some part in human evolution.

Gene loss

Many different mutations can inactivate a gene, but few will change its function in a specific way. Inactivation mutations will therefore be readily available for selection to act on. Gene loss could thus be a common mechanism of evolutionary adaptation (the "less-is-more" hypothesis).

80 genes were lost in the human lineage after separation from the last common ancestor with the chimpanzee. 36 of those were for olfactory receptors. Genes involved in chemoreception and immune response are overrepresented. Another study estimated that 86 genes had been lost.

Hair keratin gene KRTHAP1

A gene for type I hair keratin was lost in the human lineage. Keratins are a major component of hairs. Humans still have nine functional type I hair keratin genes but the loss of that particular gene may have caused the thinning of human body hair. The gene loss occurred relatively recently in human evolution—less than 240,000 years ago.

Myosin gene MYH16

Stedman *et al.* (2004) stated that the loss of the sarcomeric myosin gene MYH16 in the human lineage led to smaller masticatory muscles. They estimated that the mutation that led to the inactivation (a two base pair deletion) occurred 2.4 million years ago, predating the appearance of *Homo ergaster/erectus* in Africa. The period that followed was marked by a strong increase in cranial capacity, promoting speculation that the loss of the gene may have removed an evolutionary constraint on brain size in the genus *Homo*.

Another estimate for the loss of the MYH16 gene is 5.3 million years ago, long before *Homo* appeared.

Other

- CASPASE12, a cysteinyl aspartate proteinase

Gene addition

Segmental duplications (SDs or LCRs) have had roles in creating new primate genes and shaping human genetic variation.

Selection pressures

Human accelerated regions are areas of the genome that differ between humans and chimpanzees to a greater extent than can be explained by genetic drift over the time since the two species shared a common ancestor. These regions show signs of being subject to natural selection, leading to the evolution of distinctly human traits. Two examples are HAR1F, which is believed to be related to brain development and HAR2 (a.k.a HACNS1) that may have played a role in the development of the opposable thumb.

Genetic differences between humans and Neanderthals

An international group of scientists completed a draft sequence of the Neanderthal genome in May 2010. The results indicate some breeding between humans and Neanderthals as the genomes of non-African humans have 1-4% more in common with Neanderthals than do the genomes of subsaharan Africans. Neanderthals and most humans share a lactose-intolerant variant of the lactase gene that encodes an enzyme that is unable to break down lactose in milk after weaning. Humans and Neanderthals also share the FOXP2 gene variant associated with brain development and with speech in humans, indicating that Neanderthals may have been able to speak. Chimps have two amino acid differences in FOXP2 compared with human and Neanderthal FOXP2.

Sequence divergence between humans and apes

The draft sequence of the common chimpanzee genome published in the summer 2005 showed the regions that are similar enough to be aligned with one another account for 2400 million of the human genome's 3164.7 million bases – that is, 75.8% of the genome. This 75.8% of the human genome is 1.23% different from the chimpanzee genome in single nucleotide polymorphisms (changes of single DNA "letters" in the genome). Another type of difference, called indels (insertions/deletions) account for another ~3 % difference between the alignable sequences. In addition, variation in copy number of large segments (> 20 kb) of similar DNA sequence provides a further 2.7% difference between the two species. Hence the total similarity of the genomes could be as low as about 70%.

The figures above do not take into account differences in the organization of the alignable sequences within the genomes of humans and chimps. Short stretches of alignable sequence may be in very different orders and locations within the two genomes. At present we cannot fully assess the difference in structure of the two genomes, because the human genome was used as a scaffold when the chimpanzee draft genome was assembled. When genomes are sequenced, relatively short sequences of DNA are produced, and these sequences have to be fitted together like a jigsaw puzzle. This requires multiple overlapping reads to accurately assemble the overall sequence. The human genome sequence is relatively accurate, with 8 to 9-fold coverage, but the chimpanzee draft genome only has 3.6-fold coverage. The human genome was sequenced using a hierarchical shotgun method which can deal with duplications and difficult-to-

assemble sequences better than the whole genome shotgun method that was used for the chimpanzee draft genome. The human genome was used as a template for the assembly of the draft chimpanzee genome, on the assumption that the two genomes would be similar.

Almost half of that 1.23% SNP change belongs to the human at 0.53%, whose genetic variance is lower than a chimp, and just over half to the chimp at 0.7%. If we also take into account that random "genetic drift" takes up the bulk of the 0.54% difference, then that percentage difference where SNPs have a potential positive impact on human abilities, is between 0.01% and 0.02%. The bonobo is a sibling species of common chimpanzee and is genetically about as different from humans as are common chimps.

Percentage sequence divergence between humans and other hominids

Locus	Human-Chimp	Human-Gorilla	Human-Orangutan
Alu elements	2	-	-
Non-coding (Chr. Y)	1.68 ± 0.19	2.33 ± 0.2	5.63 ± 0.35
Pseudogenes (autosomal)	1.64 ± 0.10	1.87 ± 0.11	-
Pseudogenes (Chr. X)	1.47 ± 0.17	-	-
Noncoding (autosomal)	1.24 ± 0.07	1.62 ± 0.08	3.08 ± 0.11
Genes (K_s)	1.11	1.48	2.98
Introns	0.93 ± 0.08	1.23 ± 0.09	-
Xq13.3	0.92 ± 0.10	1.42 ± 0.12	3.00 ± 0.18
Subtotal for X chromosome	1.16 ± 0.07	1.47 ± 0.08	-
Genes (K_a)	0.8	0.93	1.96

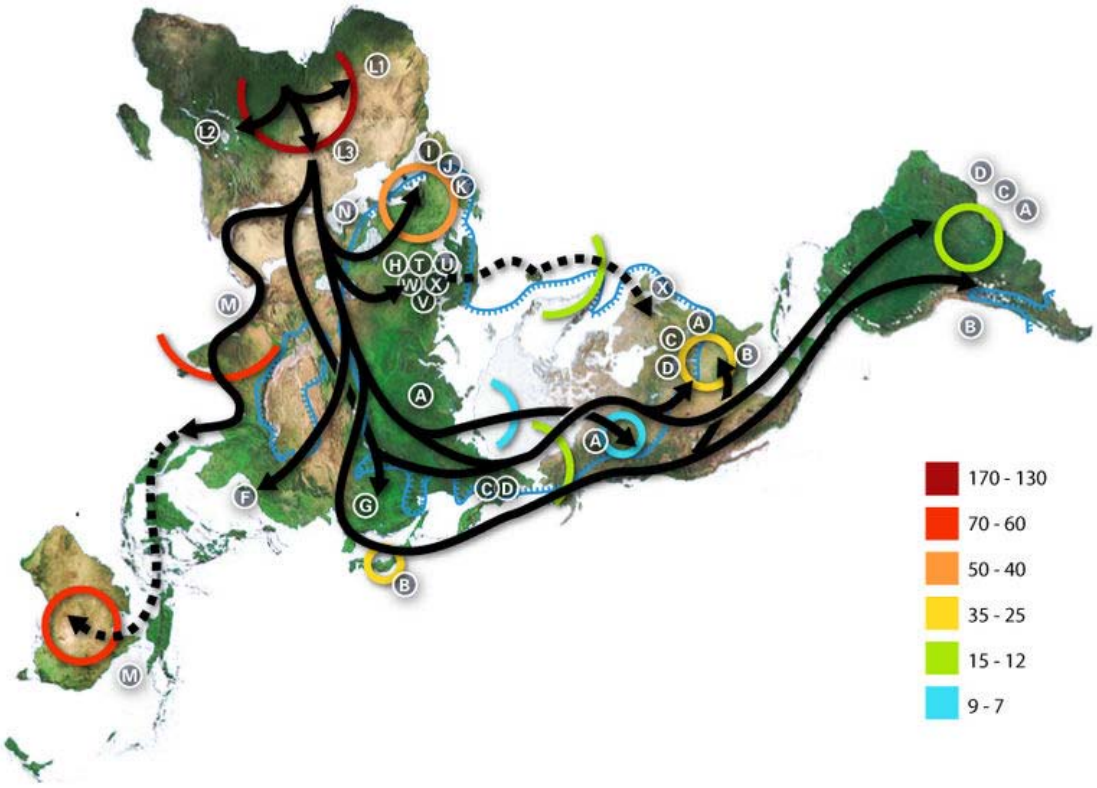
The sequence divergence has generally the following pattern: Human-Chimp < Human-Gorilla << Human-Orangutan, highlighting the close kinship between humans and the African apes. Alu elements diverge quickly due to their high frequency of CpG dinucleotides which mutate roughly 10 times more often than the average nucleotide in the genome. The mutation rate is higher in the male germ line, therefore the divergence in the Y chromosome—which is inherited solely from the father—is higher than in autosomes. The X chromosome is inherited twice as often through the female germ line as through the male germ line and therefore shows slightly lower sequence divergence. The sequence divergence of the Xq13.3 region is surprisingly low between humans and chimpanzees.

Mutations altering the amino acid sequence of proteins (K_a) are the least common. In fact ~29% of all orthologous proteins are identical between human and chimpanzee. The typical protein differs by only two amino acids.

The measures of sequence divergence shown in the table only take the substitutional differences, for example from an A (adenine) to a G (guanine), into account. DNA sequences may however also differ by insertions and deletions (indels) of bases. These are usually stripped from the alignments before the calculation of sequence divergence is

performed. The overall sequence divergence between humans and chimpanzees for example is close to 5% if indels would be included.

Modern humans



Map of the migration of modern humans out of Africa, based on mitochondrial DNA. Coloured rings indicate years before present, in thousands.

Molecular biologists starting with Wesley Brown on mtDNA and Allan Wilson on mtDNA have produced observations relevant to human evolution.

Age of the common ancestor

By estimating the rate at which mutations occur in mtDNA, the age of the common ancestral mtDNA type can be estimated: "the common ancestral mtDNA (type a) links mtDNA types that have diverged by an average of nearly 0.57%. Assuming a rate of 2%-4% per million years, this implies that the common ancestor of all surviving mtDNA types existed 140,000-290,000 years ago." This observation is robust, and this common direct female line ancestor (or mitochondrial most recent common ancestor (mtMRCA)) of all extant humans has become known as Mitochondrial Eve. The observation that the mtMRCA is the direct matrilineal ancestor of all living humans does not mean either that she was the first anatomically modern human, nor that no other female humans lived

concurrently with her. Other women would have lived at the same time and passed nuclear genes down to living humans, but their mitochondrial lineages were lost over time. This could be due to random events such as producing only male children.

African origin for modern humans

There is evidence that modern human mtDNA has an African origin: "We infer from the tree of minimum length... that Africa is a likely source of the human mitochondrial gene pool. This inference comes from the observation that one of the two primary branches leads exclusively to African mtDNAs... while the second primary branch also leads to African mtDNAs... By postulating that the common ancestral mtDNA... was African, we minimize the number of intercontinental migrations needed to account for the geographic distribution of mtDNA types."

The broad study of African genetic diversity headed by Sarah Tishkoff found the San people to express the greatest genetic diversity among the 113 distinct populations sampled, making them one of 14 "ancestral population clusters". The research also located the origin of modern human migration in south-western Africa, near the coastal border of Namibia and Angola.

Y chromosome findings

The Y chromosome is much larger than mtDNA, and is relatively homogeneous; therefore it has taken much longer to find distinct lineages and to analyse them. Conversely, because the Y chromosome is so large by comparison, it holds more genetic information. Y chromosome studies show similar findings to those made with mtDNA. The estimate for the age of the ancestral Y chromosome for all extant Y chromosomes is given at about 70,000 years ago and is also placed in Africa; the individual who contributed this Y chromosomal heritage is sometimes referred to as Y chromosome Adam. The difference in dates between Y chromosome Adam and mitochondrial Eve is usually attributed to a higher extinction rate for Y chromosomes due to greater differential reproductive success between individual men, which means that a small number of very successful men may produce many children, while a larger number of less successful men will produce far fewer children.

Chapter- 9

Human Mitochondrial Genetics

Human mitochondrial genetics is the study of the genetics of the DNA contained in human mitochondria. Mitochondria are small structures in cells that generate energy for the cell to use, and are hence referred to as the "powerhouses" of the cell.

Mitochondrial DNA (mtDNA) is not transmitted through nuclear DNA (nDNA). In humans, as in most multicellular organisms, mitochondrial DNA is inherited only from the mother's ovum.

Mitochondrial inheritance is therefore non-Mendelian, as Mendelian inheritance presumes that half the genetic material of a fertilized egg (zygote) derives from each parent.

Eighty percent of mitochondrial DNA codes for functional mitochondrial proteins, and therefore most mitochondrial DNA mutations lead to functional problems, which may be manifested as muscle disorders (myopathies).

Understanding the genetic mutations that affect mitochondria can help us to understand the inner workings of cells and organisms, as well as helping to suggest methods for successful therapeutic tissue and organ cloning, and to treatments or possibly cures for many devastating muscular disorders.

Mitochondrial function and genome

Because they provide 36 molecules of ATP per glucose molecule in contrast to the 2 ATP molecules produced by glycolysis, mitochondria are essential to all higher organisms for sustaining life. The mitochondrial diseases are genetic disorders carried specifically in mitochondrial DNA; slight problems with any one of the numerous enzymes used by the mitochondria can be devastating to the cell, and in turn, to the organism.

Membrane complexes

The processes carried out by the electron transport chain are mediated by protein complexes (named Complexes I-V, DHO-QO, ETF-QO, and ANT). Complex I, or NADH : coenzyme Q oxidoreductase, uses the energy in NADH to pump protons into the intermembrane space of the mitochondrion, pumping 2 protons per electron and passing 2 electrons via coenzyme Q to complex III or coenzyme Q : cytochrome c oxidoreductase. Complex II or succinate: coenzyme Q oxidoreductase accepts energy from succinate produced in the citric acid cycle and passes it via coenzyme Q to complex III. Complex III pumps 1 protons per electron and passes 1 electron via cytochrome c to complex IV or Cytochrome C : O_2 Oxidoreductase. Complex IV pumps 1 protons into the space between the mitochondrion's two membranes before passing the electron to O_2 which reacts to form water. Complex V (ATP synthase) is a rotary complex which allows approximately (determining the actual number is very difficult) 10 protons to enter the mitochondrial matrix along their concentration gradients. It uses the energy from the gradient to form the bond between ADP and the phosphate group to create ATP. The electron transfer flavoprotein : coenzyme Q oxidoreductase is also an electron-transporting molecule and is involved in the breakdown of fatty acids and amino acids. ANT (adenine nucleotide translocator) is also involved in oxidative phosphorylation as an energy carrying molecule. Each of these eight complexes plays a vital role in the health of the cell and any slight mutation in any one of the proteins that make up these complexes can lead to cell death or stress, which can both in turn lead to a number of diseases.

Genome

Mitochondrial DNA (mtDNA) is present in mitochondria as a circular molecule and in most species codes for 13 or 14 proteins involved in the electron transfer chain, 2 rRNA subunits and 22 tRNA molecules (all necessary for protein synthesis). The number of proteins involved in the electron transfer chain is much larger than 13 or 14, but the others are coded by the nuclear DNA.

In total, the mitochondrion hosts about 3000 proteins, but only about 13 of them are coded on the mitochondrial DNA. Most of the 3000 proteins are involved in a variety of processes other than ATP production, such as porphyrin synthesis. Only about 3% of them code for ATP production proteins. This means most of the genetic information coding for the protein makeup of mitochondria is in chromosomal DNA and is involved in processes other than ATP synthesis. This increases the chances that a mutation that will affect a mitochondrion will occur in chromosomal DNA, which is inherited in a Mendelian pattern. Another result is that a chromosomal mutation will affect a specific tissue due to its specific needs, whether those may be high energy requirements or a need for the catabolism or anabolism of a specific neurotransmitter or nucleic acid. Because several copies of the mitochondrial genome are carried by each mitochondrion (2-10 in humans), mitochondrial mutations can be inherited maternally by mtDNA mutations which are present in mitochondria inside the oocyte before fertilization, or (as stated above) through mutations in the chromosomes.

In humans, the heavy strand of mtDNA carries 28 genes and the light strand of mtDNA carries only 9 genes. Eight of the 9 genes on the light strand code for mitochondrial tRNA molecules. Human mtDNA consists of 16,569 nucleotide pairs. The entire molecule is regulated by only one regulatory region which contains the origins of replication of both heavy and light strands. The entire human mitochondrial DNA molecule has been mapped. The rate of mutation in mtDNA is calculated to be about ten times greater than that of nuclear DNA, possibly due to a paucity of DNA repair mechanisms. This high mutation rate leads to a high variation between mitochondria, not only among different species but even within the same species. It is calculated that if two humans are chosen randomly and their mtDNA is tested, they will have an average of between fifty and seventy different nucleotides. This may not seem like much, but when compared to the total number of nucleotides of a human mitochondrial DNA molecule (16,569), as much as 0.42% of the mtDNA varies between two people.

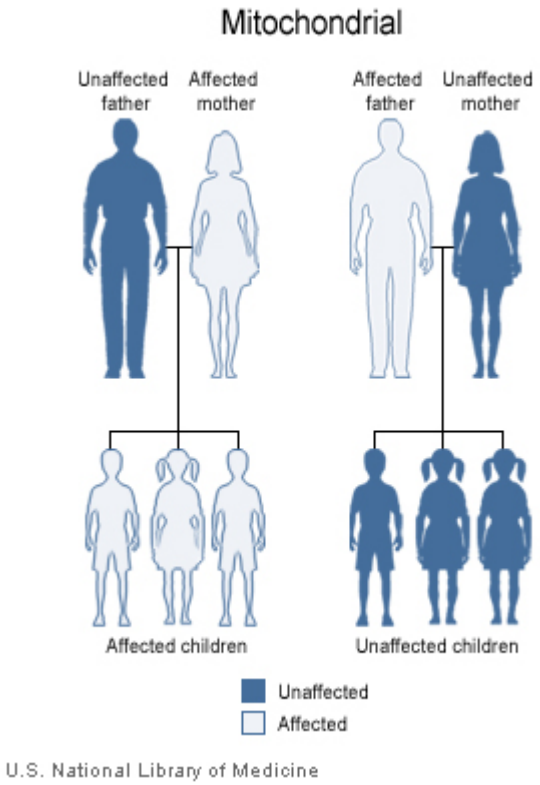
Genetic code variants

The genetic code is, for the most part, universal, with few exceptions: mitochondrial genetics includes some of these. For most organisms the "stop codons" are “UAA”, “UAG”, and “UGA”. In vertebrate mitochondria “AGA” and “AGG” are also stop codons, but not “UGA”, which codes for tryptophan instead. “AUA” codes for isoleucine in most organisms but for methionine in vertebrate mitochondrial mRNA.

There are many other variations among the codes used by other mitochondrial m/tRNA, which happened not to be harmful to their organisms, and which can be used as a tool (along with other mutations among the mtDNA/RNA of different species) to determine relative proximity of common ancestry of related species. (The more related two species are, the more mtDNA/RNA mutations will be the same in their mitochondrial genome).

Using these techniques, it is estimated that the first mitochondria arose around 1.5 billion years ago. A generally accepted hypothesis is that mitochondria originated as an aerobic prokaryote in a symbiotic relationship within an anaerobic eukaryote.

Inheritance patterns



Mitochondrial Inheritance Patterns

Because mitochondrial diseases (diseases due to malfunction of mitochondria) can be inherited both maternally and through chromosomal inheritance, the way in which they are passed on from generation to generation can vary greatly depending on the disease. Mitochondrial genetic mutations that occur in the nuclear DNA can occur in any of the chromosomes (depending on the species). Mutations inherited through the chromosomes can be autosomal dominant or recessive and can also be sex-linked dominant or recessive. Chromosomal inheritance follows normal Mendelian laws, despite the fact that the phenotype of the disease may be masked.

Because of the complex ways in which mitochondrial and nuclear DNA "communicate" and interact, even seemingly simple inheritance is hard to diagnose. A mutation in chromosomal DNA may change a protein that regulates (increases or decreases) the production of another certain protein in the mitochondria or the cytoplasm; this may lead to slight, if any, noticeable symptoms. On the other hand, some devastating mtDNA mutations are easy to diagnose because of their widespread damage to muscular, neural, and/or hepatic tissues (among other high-energy and metabolism-dependent tissues) and because they are present in the mother and all the offspring.

Mitochondrial genome mutations are passed on 100% of the time from mother to all her offspring. The number of affected mtDNA molecules inherited by a specific offspring can vary greatly because

- the mitochondria within the fertilized oocyte is what the new life will have to begin with (in terms of mtDNA),
- the number of affected mitochondria varies from cell (in this case, the fertilized oocyte) to cell depending both on the number it inherited from its mother cell and environmental factors which may favor mutant or wildtype mitochondrial DNA,
- the number of mtDNA molecules in the mitochondria varies from around two to ten.

It is possible, even in twin births, for one baby to receive more than half mutant mtDNA molecules while the other twin may receive only a tiny fraction of mutant mtDNA molecules with respect to wildtype (depending on how the twins divide from each other and how many mutant mitochondria happen to be on each side of the division). In a few cases, some mitochondria or a mitochondrion from the sperm cell enters the oocyte but paternal mitochondria are actively decomposed.

Replication, repair, transcription, and translation

Mitochondrial replication is controlled by nuclear genes and is specifically suited to make as many mitochondria as that particular cell needs at the time. Human mitochondrial DNA (mtDNA) has three promoters, H1, H2, and L (heavy strand 1, heavy strand 2, and light strand promoters). The H2 promoter transcribes almost the entire heavy strand and the L promoter transcribes the entire light strand. The H1 promoter causes the transcription of the two mitochondrial rRNA molecules. When transcription takes place on the heavy strand a polycistronic transcript is created. The light strand produces either small transcripts, which can be used as primers, or one long transcript. The production of primers occurs by processing of light strand transcripts with the Mitochondrial RNase MRP (Mitochondrial RNA Processing). The requirement of transcription to produce primers links the process of transcription to mtDNA replication. Full length transcripts are cut into functional tRNA, rRNA, and mRNA molecules. The process of transcription initiation in mitochondria involves three types of proteins: the mitochondrial RNA polymerase (POLRMT), mitochondrial transcription factor A (TFAM), and mitochondrial transcription factors B1 and B2 (TFB1M, TFB2M). POLRMT, TFAM, and TFB1M or TFB2M assemble at the mitochondrial promoters and begin transcription. The actual molecular events that are involved in initiation are unknown, but these factors make up the basal transcription machinery and have been shown to function in vitro. Mitochondrial translation is still not very well understood. In vitro translations have still not been successful, probably due to the difficulty of isolating sufficient mt mRNA, functional mt rRNA, and possibly because of the complicated changes that the mRNA undergoes before it is translated.

Mitochondrial DNA polymerase

The Mitochondrial DNA Polymerase (Pol gamma, encoded by the POLG gene) is used in the copying of mtDNA during replication. Because the two (heavy and light) strands on the circular mtDNA molecule have different origins of replication, it replicates in a D-loop mode. One strand begins to replicate first, displacing the other strand. This continues

until replication reaches the origin of replication on the other strand, at which point the other strand begins replicating in the opposite direction. This results in two new mtDNA molecules. Each mitochondrion has several copies of the mtDNA molecule and the number of mtDNA molecules is a limiting factor in mitochondrial fission. After the mitochondrion has enough mtDNA, membrane area, and membrane proteins, it can undergo fission (very similar to that which bacteria use) to become two mitochondria. Evidence suggests that mitochondria can also undergo fusion and exchange (in a form of crossover) genetic material among each other. Mitochondria sometimes form large matrices in which fusion, fission, and protein exchanges are constantly occurring. mtDNA shared among mitochondria (despite the fact that they can undergo fusion).

Damage and transcription error

Mitochondrial DNA is susceptible to damage from free oxygen radicals from mistakes that occur during the production of ATP through the electron transport chain. These mistakes can be caused by genetic disorders, cancer, and temperature variations. These radicals can damage mtDNA molecules or change them, making it hard for mitochondrial polymerase to replicate them. Both cases can lead to deletions, rearrangements, and other mutations. Recent evidence has suggested that mitochondria have enzymes that proofread mtDNA and fix mutations that may occur due to free radicals. It is believed that a DNA recombinase found in mammalian cells is also involved in a repairing recombination process. Deletions and mutations due to free radicals have been associated with the aging process. It is believed that radicals cause mutations which lead to mutant proteins, which in turn lead to more radicals. This process takes many years and is associated with some aging processes involved in oxygen-dependent tissues such as brain, heart, muscle, and kidney. Auto-enhancing processes such as these are possible causes of degenerative diseases including Parkinson’s, Alzheimer’s, and coronary artery disease.

Chromosomally mediated mtDNA replication errors

Because mitochondrial growth and fission are mediated by the nuclear DNA, mutations in nuclear DNA can have a wide array of effects on mtDNA replication. Despite the fact that the loci for some of these mutations have been found on human chromosomes, specific genes and proteins involved have not yet been isolated. Mitochondria need a certain protein to undergo fission. If this protein (made by the nucleus) is not present, the mitochondria grow but they do not divide. This leads to giant, inefficient mitochondria. Mistakes in chromosomal genes or their products can also affect mitochondrial replication more directly by inhibiting mitochondrial polymerase and can even cause mutations in the mtDNA directly and indirectly. Indirect mutations are most often caused by radicals created by defective proteins made from nuclear DNA.

Mitochondrial diseases

Mitochondrial diseases range in severity from asymptomatic to fatal, and are most commonly due to inherited rather than acquired mutations of mitochondrial DNA. A given mitochondrial mutation can cause various diseases depending on the severity of the

problem in the mitochondria and the tissue the affected mitochondria are in. Conversely, several different mutations may present themselves as the same disease. This almost patient-specific characterization of mitochondrial diseases makes them very hard to accurately recognize, diagnose and trace. Some diseases are observable at or even before birth (many causing death) while others do not show themselves until late adulthood (late-onset disorders). This is because the number of mutant versus wildtype mitochondria varies between cells and tissues, and is continuously changing. Because cells have multiple mitochondria, different mitochondria in the same cell can have different variations of the mtDNA. This condition is referred to as heteroplasmy. When a certain tissue reaches a certain ratio of mutant versus wildtype mitochondria, a disease will present itself. The ratio varies from person to person and tissue to tissue (depending on its specific energy, oxygen, and metabolism requirements, and the effects of the specific mutation). Mitochondrial diseases are very numerous and different. Apart from diseases caused by abnormalities in mitochondrial DNA, many diseases are suspected to be associated in part by mitochondrial dysfunctions, such as diabetes mellitus, forms of cancer and cardiovascular disease, lactic acidosis, specific forms of myopathy, osteoporosis, Alzheimer's disease, Parkinsons's disease, stroke, Male infertility and which are also believed to play a role in the aging process.

Chapter- 10

Molecular Genetics

Molecular genetics is the field of biology and genetics that studies the structure and function of genes at a molecular level. The field studies how the genes are transferred from generation to generation. Molecular genetics employs the methods of genetics and molecular biology. It is so-called to differentiate it from other sub fields of genetics such as ecological genetics and population genetics. An important area within molecular genetics is the use of molecular information to determine the patterns of descent, and therefore the correct scientific classification of organisms: this is called molecular systematics.

Along with determining the pattern of descendants, molecular genetics helps in understanding genetic mutations that can cause certain types of diseases. Through utilizing the methods of genetics and molecular biology, molecular genetics discovers the reasons why traits are carried on and how and why some may mutate.

Forward genetics

One of the first tools available to molecular geneticists is the forward genetic screen. The aim of this technique is to identify mutations that produce a certain phenotype. A mutagen is very often used to accelerate this process. Once mutants have been isolated, the mutated gene can be molecularly identified.

Reverse genetics

While forward genetic screens are productive, a more straightforward approach would be to determine the phenotype that results from mutating a given gene. This is called reverse genetics. In some organisms, such as yeast and mice, it is possible to induce the deletion of a particular gene, creating a gene knockout. Alternatives include the random induction of DNA deletions and subsequent selection for deletions in a gene of interest, the application of RNA interference and the creation of transgenic organisms that do not express a gene of interest.

Gene therapy

A mutation in a gene can result in a severe medical condition. A protein encoded by a mutated gene may malfunction and cells that rely on the protein might therefore fail to function properly. This can cause problems for specific tissues or organs, or for the entire body. This might manifest through the course of development (like a cleft palate) or as an abnormal response to stimuli (like a peanut allergy). Conditions related to gene mutations are called genetic disorders. One way to fix such a physiological problem is gene therapy. By adding a corrected copy of the gene, a functional form of the protein can be produced, and affected cells, tissues, and organs may work properly. As opposed to drug-based approaches, gene therapy repairs the underlying genetic defect.

Gene therapy is the process of treating or alleviating diseases by genetically modifying the cells of the affected person, causing the gene to function properly. When a human disease gene has been recognized, molecular genetics tools can be used to explore the process of the gene in both the normal and mutant states. From there, the gene is transferred either in vivo or ex vivo and the body begins to make proteins according to the instructions in the new gene. Gene therapy has to be repeated several times for the infected patient to continually be relieved, however, as repeated cell division and death slowly randomizes the body's ratio of functional-to-mutant genes.

Currently, gene therapy is still being experimented with and products are not approved by the U.S. Food and Drug Administration. There have been several setbacks in the last 15 years that have restricted further developments in gene therapy. As there are unsuccessful attempts, there continue to be a growing number of successful gene therapy transfers which have furthered the research.

Major diseases that can be treated with gene therapy include viral infections, cancers, and inherited disorders, including immune system disorders.

Classical gene therapy

Classical gene therapy is the approach which delivers genes, via a modified virus or "vector" to the appropriate target cells with a goal of attaining optimal expression of the new, introduced gene. Once inside the patient, the expressed genes are intended to produce a product that the patient lacks, kill diseased cells directly by producing a toxin, or activate the immune system to help the killing of diseased cells.

Nonclassical gene therapy

Nonclassical gene therapy inhibits the expression of genes related to pathogenesis, or corrects a genetic defect and restores normal gene expression.

In vivo gene transfer

During In vivo gene transfer, the genes are transferred directly into the tissue of the patient and this can be the only possible option in patients with tissues where individual cells cannot be cultured in vitro in sufficient numbers (e.g. brain cells). Also, in vivo gene transfer is necessary when cultured cells cannot be re-implanted in patients effectively.

Ex vivo gene transfer

During ex vivo gene transfer the cells are cultured outside the body and then the genes are transferred into the cells grown in culture. The cells that have been transformed successfully are expanded by cell culture and then introduced into the patient.

Principles for gene transfer

Classical gene therapies usually require efficient transfer of cloned genes into the disease cells so that the introduced genes are expressed at sufficiently high levels to change the patient's physiology. There are several different physicochemical and biological methods that can be used to transfer genes into human cells. The size of the DNA fragments that can be transferred is very limited, and often the transferred gene is not a conventional gene. Horizontal gene transfer is the transfer of genetic material from one cell to another that is not its offspring. Artificial horizontal gene transfer is a form of genetic engineering.

Techniques in molecular genetics

There are three general techniques used for molecular genetics: amplification, separation and detection, and expression. Specifically used for amplification is polymerase chain reaction, which is an “indispensable tool in a great variety of applications”. In the separation and detection technique DNA and mRNA are isolated from their cells. Gene expression in cells or organisms is done in a place or time that is not normal for that specific gene.

Amplification

There are other methods for amplification besides polymerase chain reaction. Cloning DNA in bacteria is also a way to amplify DNA in genes.

Polymerase chain reaction

The main materials used in polymerase chain reaction are DNA nucleotides, template DNA, primers and Taq polymerase. DNA nucleotides are the base for the new DNA, the template DNA is the specific sequence being amplified, primers are complementary nucleotides that can go on either side of the template DNA, and Taq polymerase is a heat stable enzyme that jump-starts the production of new DNA at the high temperatures

needed for reaction. This technique does not need to use living bacteria or cells; all that is needed is the base sequence of the DNA and materials listed above.

Cloning DNA in bacteria

The word cloning for this type of amplification entails making multiple identical copies of a sequence of DNA. The target DNA sequence is then inserted into a cloning vector. Because this vector originates from a self-replicating virus, plasmid, or higher organism cell when the appropriate size DNA is inserted the "target and vector DNA fragments are then ligated" and create a recombinant DNA molecule. The recombinant DNA molecules are then put into a bacteria strain (usually *E. coli*) which produces several identical copies by transformation. Transformation is the DNA uptake mechanism possessed by bacteria. However, only one recombinant DNA molecule can be cloned within a single bacteria cell, so each clone is of just one DNA insert.

Separation and detection

In separation and detection DNA and mRNA are isolated from cells (the separation) and then detected simply by the isolation. Cell cultures are also grown to provide a constant supply of cells ready for isolation.

Cell cultures

A cell culture for molecular genetics is a culture that is grown in artificial conditions. Some cell types grow well in cultures such a skin cells, but other cells are not as productive in cultures. There are different techniques for each type of cell, some only recently being found to foster growth in stem and nerve cells. Cultures for molecular genetics are frozen in order to preserve all copies of the gene specimen and thawed only when needed. This allows for a steady supply of cells.

DNA isolation

DNA isolation extracts DNA from a cell in a pure form. First, the DNA is separated from cellular components such as proteins, RNA, and lipids. This is done by placing the chosen cells in a tube with a solution that mechanically, chemically, breaks the cells open. This solution contains enzymes, chemicals, and salts that breaks down the cells except for the DNA. It contains enzymes to dissolve proteins, chemicals to destroy all RNA present, and salts to help pull DNA out of the solution.

Next, the DNA is separated from the solution by being spun in a centrifuge, which allows the DNA to collect in the bottom of the tube. After this cycle in the centrifuge the solution is poured off and the DNA is resuspended in a second solution that makes the DNA easy to work with in the future.

This results in a concentrated DNA sample that contains thousands of copies of each gene. For large scale projects such as sequencing the human genome, all this work is done by robots.

mRNA isolation

Expressed DNA that codes for the synthesis of a protein is the final goal for scientists and this expressed DNA is obtained by isolation mRNA (Messenger RNA). First, laboratories use a normal cellular modification of mRNA that adds up to 200 adenine nucleotides to the end of the molecule (poly(A) tail). Once this has been added, the cell is ruptured and its cell contents are exposed to synthetic beads that are coated with thymine string nucleotides. Because Adenine and Thymine pair together in DNA, the poly(A) tail and synthetic beads are attracted to one another, and once they bind in this process the cell components can be washed away without removing the mRNA. Once the mRNA has been isolated, reverse transcriptase is employed to convert it to single-stranded DNA, from which a stable double-stranded DNA is produced using DNA polymerase. Complementary DNA (cDNA) is much more stable than mRNA and so, once the double-stranded DNA has been produced it represents the expressed DNA sequence scientists look for.

The Human Genome Project

The Human Genome Project is a molecular genetics project that began in the 1990s and was projected to take fifteen years to complete. However, because of technological advances the progress of the project was advanced and the project finished in 2003, taking only thirteen years. The project was started by the U.S. Department of Energy and the National Institutes of Health in an effort to reach six set goals. These goals included:

1. identifying 20,000 to 25,000 genes in human DNA (although initial estimate were approximately 100,000 genes),
2. determining sequences of chemical based pairs in human DNA,
3. storing all found information into databases,
4. improving the tools used for data analysis,
5. transferring technologies to private sectors, and
6. addressing the ethical, legal, and social issues (ELSI) that may arise from the projects.

The project was worked on by eighteen different countries including the United States, Japan, France, Germany, and the United Kingdom. The collaborative effort resulted in the discovery of the many benefits of molecular genetics. Discoveries such as molecular medicine, new energy sources and environmental applications, DNA forensics, and livestock breeding, are only a few of the benefits that molecular genetics can provide.

Chapter- 11

Population Genetics

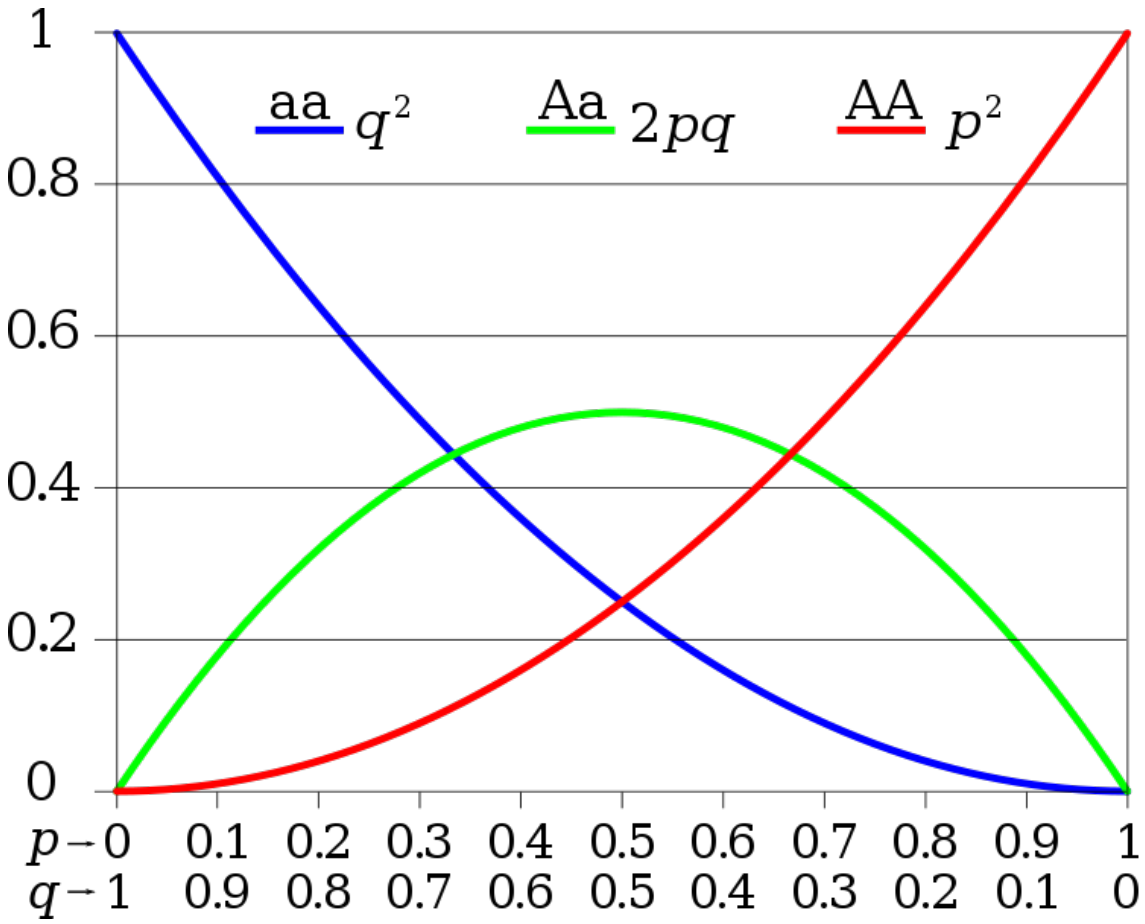
Population genetics is the study of allele frequency distribution and change under the influence of the four main evolutionary processes: natural selection, genetic drift, mutation and gene flow. It also takes into account the factors of recombination, population subdivision and population structure. It attempts to explain such phenomena as adaptation and speciation.

Population genetics was a vital ingredient in the emergence of the modern evolutionary synthesis. Its primary founders were Sewall Wright, J. B. S. Haldane and R. A. Fisher, who also laid the foundations for the related discipline of quantitative genetics.

Fundamentals

Population genetics concerns the genetic constitution of populations and how this constitution changes with time. A population is a set of organisms in which any pair of members can breed together. This implies that all members belong to the same species and live near each other.

For example, all of the moths of the same species living in an isolated forest are a population. A gene in this population may have several alternate forms, which account for variations between the phenotypes of the organisms. An example might be a gene for coloration in moths that has two alleles: black and white. A gene pool is the complete set of alleles for a gene in a single population; the allele frequency for an allele is the fraction of the genes in the pool that is composed of that allele (for example, what fraction of moth coloration genes are the black allele). Evolution occurs when there are changes in the frequencies of alleles within a population; for example, the allele for black color in a population of moths becoming more common.



Hardy–Weinberg principle for two alleles: the horizontal axis shows the two allele frequencies *p* and *q* and the vertical axis shows the genotype frequencies. Each graph shows one of the three possible genotypes.

Hardy–Weinberg principle

To understand the mechanisms that cause a population to evolve, it is useful to consider what conditions are required for a population not to evolve. The *Hardy-Weinberg principle* states that the frequencies of alleles (variations in a gene) in a sufficiently large population will remain constant if the only forces acting on that population are the random reshuffling of alleles during the formation of the sperm or egg, and random combination of the alleles in these sex cells during fertilization. Such a population is said to be in *Hardy-Weinberg equilibrium* as it is not evolving. Hardy Weinberg equilibrium is impossible in nature. Genetic equilibrium is an ideal state that provides a baseline to measure genetic change against.

Allele frequencies in a population remain static across generations, provided the following conditions are at hand: random mating, no mutation (the alleles don't change), no migration or emigration (no exchange of alleles between populations), infinitely large population size, and no selective pressure for or against any traits.

In the simplest case of a single locus with two alleles: the dominant allele is denoted **A** and the recessive **a** and their frequencies are denoted by p and q; freq(**A**) = p; freq(**a**) = q; $p + q = 1$. If the population is in equilibrium, then we will have freq(**AA**) = p^2 for the **AA** homozygotes in the population, freq(**aa**) = q^2 for the **aa** homozygotes, and freq(**Aa**) = $2pq$ for the heterozygotes.

Based on these equations, useful but difficult-to-measure facts about a population can be determined. For example, a patient's child is a carrier of a recessive mutation that causes cystic fibrosis in homozygous recessive children. The parent wants to know the probability of her grandchildren inheriting the disease. In order to answer this question, the genetic counselor must know the chance that the child will reproduce with a carrier of the recessive mutation. This fact may not be known, but disease frequency is known. We know that the disease is caused by the homozygous recessive genotype; we can use the Hardy–Weinberg principle to work backward from disease occurrence to the frequency of heterozygous recessive individuals.

Scope and theoretical considerations

The mathematics of population genetics were originally developed as part of the modern evolutionary synthesis. According to Beatty (1986), it defines the core of the modern synthesis.

According to Lewontin (1974), the theoretical task for population genetics is a process in two spaces: a "genotypic space" and a "phenotypic space". The challenge of a *complete* theory of population genetics is to provide a set of laws that predictably map a population of genotypes (G_1) to a phenotype space (P_1), where selection takes place, and another set of laws that map the resulting population (P_2) back to genotype space (G_2) where Mendelian genetics can predict the next generation of genotypes, thus completing the cycle. Even leaving aside for the moment the non-Mendelian aspects of molecular genetics, this is clearly a gargantuan task. Visualizing this transformation schematically:

$$G_1 \xrightarrow{T_1} P_1 \xrightarrow{T_2} P_2 \xrightarrow{T_3} G_2 \xrightarrow{T_4} G_1' \rightarrow \cdots$$

(adapted from Lewontin 1974, p. 12). XD

T_1 represents the genetic and epigenetic laws, the aspects of functional biology, or development, that transform a genotype into phenotype. We will refer to this as the "genotype-phenotype map". T_2 is the transformation due to natural selection, T_3 are epigenetic relations that predict genotypes based on the selected phenotypes and finally T_4 the rules of Mendelian genetics.

In practice, there are two bodies of evolutionary theory that exist in parallel, traditional population genetics operating in the genotype space and the biometric theory used in plant and animal breeding, operating in phenotype space. The missing part is the mapping between the genotype and phenotype space. This leads to a "sleight of hand" (as

Lewontin terms it) whereby variables in the equations of one domain, are considered parameters or *constants*, where, in a full-treatment they would be transformed themselves by the evolutionary process and are in reality *functions* of the state variables in the other domain. The "sleight of hand" is assuming that we know this mapping. Proceeding as if we do understand it is enough to analyze many cases of interest. For example, if the phenotype is almost one-to-one with genotype (sickle-cell disease) or the time-scale is sufficiently short, the "constants" can be treated as such; however, there are many situations where it is inaccurate.

The four processes

Natural selection

Natural selection is the process by which heritable traits that make it more likely for an organism to survive and successfully reproduce become more common in a population over successive generations.

The natural genetic variation within a population of organisms means that some individuals will survive more successfully than others in their current environment. Factors which affect reproductive success are also important, an issue which Charles Darwin developed in his ideas on sexual selection.

Natural selection acts on the phenotype, or the observable characteristics of an organism, but the genetic (heritable) basis of any phenotype which gives a reproductive advantage will become more common in a population. Over time, this process can result in adaptations that specialize organisms for particular ecological niches and may eventually result in the emergence of new species.

Natural selection is one of the cornerstones of modern biology. The term was introduced by Darwin in his groundbreaking 1859 book *On the Origin of Species*, in which natural selection was described by analogy to artificial selection, a process by which animals and plants with traits considered desirable by human breeders are systematically favored for reproduction. The concept of natural selection was originally developed in the absence of a valid theory of heredity; at the time of Darwin's writing, nothing was known of modern genetics. The union of traditional Darwinian evolution with subsequent discoveries in classical and molecular genetics is termed the *modern evolutionary synthesis*. Natural selection remains the primary explanation for adaptive evolution.

Genetic drift

Genetic drift is the change in the relative frequency in which a gene variant (allele) occurs in a population due to random sampling and chance. That is, the alleles in the offspring in the population are a random sample of those in the parents. And chance has a role in determining whether a given individual survives and reproduces. A population's allele frequency is the fraction or percentage of its gene copies compared to the total number of gene alleles that share a particular form.

Genetic drift is an important evolutionary process which leads to changes in allele frequencies over time. It may cause gene variants to disappear completely, and thereby reduce genetic variability. In contrast to natural selection, which makes gene variants more common or less common depending on their reproductive success, the changes due to genetic drift are not driven by environmental or adaptive pressures, and may be beneficial, neutral, or detrimental to reproductive success.

The effect of genetic drift is larger in small populations, and smaller in large populations. Vigorous debates wage among scientists over the relative importance of genetic drift compared with natural selection. Ronald Fisher held the view that genetic drift plays at the most a minor role in evolution, and this remained the dominant view for several decades. In 1968 Motoo Kimura rekindled the debate with his neutral theory of molecular evolution which claims that most of the changes in the genetic material are caused by genetic drift.

Mutation

Mutations are changes in the DNA sequence of a cell's genome and are caused by radiation, viruses, transposons and mutagenic chemicals, as well as errors that occur during meiosis or DNA replication. Errors are introduced particularly often in the process of DNA replication, in the polymerization of the second strand. These errors can also be induced by the organism itself, by cellular processes such as hypermutation.

Mutations can have an impact on the phenotype of an organism, especially if they occur within the protein coding sequence of a gene. Error rates are usually very low (1 error in every 10 million–100 million bases) due to the "proofreading" ability of DNA polymerases. Without proofreading, error rates are a thousandfold higher. Chemical damage to DNA occurs naturally as well, and cells use DNA repair mechanisms to repair mismatches and breaks in DNA. Nevertheless, the repair sometimes fails to return the DNA to its original sequence.

In organisms that use chromosomal crossover to exchange DNA and recombine genes, errors in alignment during meiosis can also cause mutations. Errors in crossover are especially likely when similar sequences cause partner chromosomes to adopt a mistaken alignment; this makes some regions in genomes more prone to mutating in this way. These errors create large structural changes in DNA sequence—duplications, inversions or deletions of entire regions, or the accidental exchanging of whole parts between different chromosomes (called translocation).

Mutation can result in several different types of change in DNA sequences; these can either have no effect, alter the product of a gene, or prevent the gene from functioning. Studies in the fly *Drosophila melanogaster* suggest that if a mutation changes a protein produced by a gene, this will probably be harmful, with about 70 percent of these mutations having damaging effects, and the remainder being either neutral or weakly beneficial. Due to the damaging effects that mutations can have on cells, organisms have evolved mechanisms such as DNA repair to remove mutations. Therefore, the optimal

mutation rate for a species is a trade-off between costs of a high mutation rate, such as deleterious mutations, and the metabolic costs of maintaining systems to reduce the mutation rate, such as DNA repair enzymes. Viruses that use RNA as their genetic material have rapid mutation rates, which can be an advantage since these viruses will evolve constantly and rapidly, and thus evade the defensive responses of e.g. the human immune system.

Mutations can involve large sections of DNA becoming duplicated, usually through genetic recombination. These duplications are a major source of raw material for evolving new genes, with tens to hundreds of genes duplicated in animal genomes every million years. Most genes belong to larger families of genes of shared ancestry. Novel genes are produced by several methods, commonly through the duplication and mutation of an ancestral gene, or by recombining parts of different genes to form new combinations with new functions.

Here, domains act as modules, each with a particular and independent function, that can be mixed together to produce genes encoding new proteins with novel properties. For example, the human eye uses four genes to make structures that sense light: three for color vision and one for night vision; all four arose from a single ancestral gene. Another advantage of duplicating a gene (or even an entire genome) is that this increases redundancy; this allows one gene in the pair to acquire a new function while the other copy performs the original function. Other types of mutation occasionally create new genes from previously noncoding DNA.

Gene flow

Gene flow is the exchange of genes between populations, which are usually of the same species. Examples of gene flow within a species include the migration and then breeding of organisms, or the exchange of pollen. Gene transfer between species includes the formation of hybrid organisms and horizontal gene transfer.

Migration into or out of a population can change allele frequencies, as well as introducing genetic variation into a population. Immigration may add new genetic material to the established gene pool of a population. Conversely, emigration may remove genetic material. As barriers to reproduction between two diverging populations are required for the populations to become new species, gene flow may slow this process by spreading genetic differences between the populations. Gene flow is hindered by mountain ranges, oceans and deserts or even man-made structures such as the Great Wall of China, which has hindered the flow of plant genes.

Depending on how far two species have diverged since their most recent common ancestor, it may still be possible for them to produce offspring, as with horses and donkeys mating to produce mules. Such hybrids are generally infertile, due to the two different sets of chromosomes being unable to pair up during meiosis. In this case, closely related species may regularly interbreed, but hybrids will be selected against and the species will remain distinct. However, viable hybrids are occasionally formed and

these new species can either have properties intermediate between their parent species, or possess a totally new phenotype. The importance of hybridization in creating new species of animals is unclear, although cases have been seen in many types of animals, with the gray tree frog being a particularly well-studied example.

Hybridization is, however, an important means of speciation in plants, since polyploidy (having more than two copies of each chromosome) is tolerated in plants more readily than in animals. Polyploidy is important in hybrids as it allows reproduction, with the two different sets of chromosomes each being able to pair with an identical partner during meiosis. Polyploids also have more genetic diversity, which allows them to avoid inbreeding depression in small populations.

Horizontal gene transfer is the transfer of genetic material from one organism to another organism that is not its offspring; this is most common among bacteria. In medicine, this contributes to the spread of antibiotic resistance, as when one bacteria acquires resistance genes it can rapidly transfer them to other species. Horizontal transfer of genes from bacteria to eukaryotes such as the yeast *Saccharomyces cerevisiae* and the adzuki bean beetle *Callosobruchus chinensis* may also have occurred. An example of larger-scale transfers are the eukaryotic bdelloid rotifers, which appear to have received a range of genes from bacteria, fungi, and plants. Viruses can also carry DNA between organisms, allowing transfer of genes even across biological domains. Large-scale gene transfer has also occurred between the ancestors of eukaryotic cells and prokaryotes, during the acquisition of chloroplasts and mitochondria.

Gene flow is the transfer of alleles from one population to another.

Migration into or out of a population may be responsible for a marked change in allele frequencies. Immigration may also result in the addition of new genetic variants to the established gene pool of a particular species or population.

There are a number of factors that affect the rate of gene flow between different populations. One of the most significant factors is mobility, as greater mobility of an individual tends to give it greater migratory potential. Animals tend to be more mobile than plants, although pollen and seeds may be carried great distances by animals or wind.

Maintained gene flow between two populations can also lead to a combination of the two gene pools, reducing the genetic variation between the two groups. It is for this reason that gene flow strongly acts against speciation, by recombining the gene pools of the groups, and thus, repairing the developing differences in genetic variation that would have led to full speciation and creation of daughter species.

For example, if a species of grass grows on both sides of a highway, pollen is likely to be transported from one side to the other and vice versa. If this pollen is able to fertilise the plant where it ends up and produce viable offspring, then the alleles in the pollen have effectively been able to move from the population on one side of the highway to the other.

Genetic structure

Because of physical barriers to migration, along with limited vagility, and natal philopatry, natural populations are rarely panmictic (Buston *et al.*, 2007). There is usually a geographic range within which individuals are more closely related to one another than those randomly selected from the general population. This is described as the extent to which a population is genetically structured (Repaci *et al.*, 2007). Genetic structuring can be caused by migration due to historical climate change, species range expansion or current availability of habitat.

Microbial population genetics

Microbial population genetics is a rapidly advancing field of investigation with relevance to many other theoretical and applied areas of scientific investigations. The population genetics of microorganisms lays the foundations for tracking the origin and evolution of antibiotic resistance and deadly infectious pathogens. Population genetics of microorganisms is also an essential factor for devising strategies for the conservation and better utilization of beneficial microbes (Xu, 2010).

History

Biston betularia f. typica is the white-bodied form of the peppered moth.

Biston betularia f. carbonaria is the black-bodied form of the peppered moth.

Population genetics

Population genetics was developed as a reconciliation of the Mendelian and biometrician models. A key step was the work of the British biologist and statistician R.A. Fisher. In a series of papers starting in 1918 and culminating in his 1930 book *The Genetical Theory of Natural Selection*, Fisher showed that the continuous variation measured by the biometricians could be produced by the combined action of many discrete genes, and that natural selection could change gene frequencies in a population, resulting in evolution (though lacking the knowledge of what an actual gene was at this time, it should be said in this sense he understood phenotypic trait frequency, rather than specifically identifiable gene frequency). In a series of papers beginning in 1924, another British geneticist, J.B.S. Haldane, applied statistical analysis to real-world examples of natural selection, such as the evolution of industrial melanism in peppered moths, and showed that natural selection worked at an even faster rate than Fisher assumed.

The American biologist Sewall Wright, who had a background in animal breeding experiments, focused on combinations of interacting genes, and the effects of inbreeding on small, relatively isolated populations that exhibited genetic drift. In 1932, Wright introduced the concept of an adaptive landscape and argued that genetic drift and inbreeding could drive a small, isolated sub-population away from an adaptive peak, allowing natural selection to drive it towards different adaptive peaks. Fisher and Wright had some fundamental disagreements and a controversy about the relative roles of selection and drift continued for much of the century between the Americans and the

British. The Frenchman Gustave Malécot was also important early in the development of the discipline.

The work of Fisher, Haldane and Wright founded the discipline of *population genetics*. This integrated natural selection with Mendelian genetics, which was the critical first step in developing a unified theory of how evolution worked.

John Maynard Smith was Haldane's pupil, whilst W.D. Hamilton was heavily influenced by the writings of Fisher. The American George R. Price worked with both Hamilton and Maynard Smith. American Richard Lewontin and Japanese Motoo Kimura were heavily influenced by Wright.

Modern evolutionary synthesis

In the first few decades of the 20th century, most field naturalists continued to believe that Lamarckian and orthogenic mechanisms of evolution provided the best explanation for the complexity they observed in the living world. However, as the field of genetics continued to develop, those views became less tenable. Theodosius Dobzhansky, a postdoctoral worker in T. H. Morgan's lab, had been influenced by the work on genetic diversity by Russian geneticists such as Sergei Chetverikov. He helped to bridge the divide between the foundations of microevolution developed by the population geneticists and the patterns of macroevolution observed by field biologists, with his 1937 book *Genetics and the Origin of Species*.

Dobzhansky examined the genetic diversity of wild populations and showed that, contrary to the assumptions of the population geneticists, these populations had large amounts of genetic diversity, with marked differences between sub-populations. The book also took the highly mathematical work of the population geneticists and put it into a more accessible form. In Great Britain E.B. Ford, the pioneer of ecological genetics, continued throughout the 1930s and 1940s to demonstrate the power of selection due to ecological factors including the ability to maintain genetic diversity through genetic polymorphisms such as human blood types. Ford's work would contribute to a shift in emphasis during the course of the modern synthesis towards natural selection over genetic drift.

Chapter- 12

Quantitative Genetics

Quantitative genetics is the study of continuous traits (such as height or weight) and their underlying mechanisms. It is effectively an extension of simple Mendelian inheritance in that the combined effect of the many underlying genes results in a continuous distribution of phenotypic values.

History

The field was founded, in evolutionary terms, by the originators of the modern synthesis, R.A. Fisher, Sewall Wright and J. B. S. Haldane, and aimed to predict the response to selection given data on the phenotype and relationships of individuals.

Analysis of Quantitative Trait Loci, or QTL, is a more recent addition to the study of quantitative genetics. A QTL is a region in the genome that affects the trait or traits of interest. Quantitative trait loci approaches require accurate phenotypic, pedigree and genotypic data from a large number of individuals.

Traits

Quantitative genetics is not limited to continuous traits, but to all traits that are determined by many genes. This includes:

- Continuous traits are quantitative traits with a continuous phenotypic range. They are often polygenic, and may also be influenced significantly by environmental effects.
- Meristic traits or other ordinal numbers are expressed in whole numbers, such as number of offspring, or number of bristles on a fruit fly. These traits can be either treated as approximately discontinuous traits or as threshold traits.
- Some qualitative traits can be treated as if they have an underlying quantitative basis, expressed as a threshold trait (or multiple thresholds). Some human diseases (such as, schizophrenia) have been studied in this manner.

Basic principles

The phenotypic value (P) of an individual is the combined effect of the genotypic value (G) and the environmental deviation (E):

$$P = G + E$$

The genotypic value is the combined effect of all the genetic effects, including nuclear genes, mitochondrial genes and interactions between the genes. It is worthwhile to note that the mathematics *is* related to the genetics: for which the brief following revision may be useful. In disomic (diploid) organisms, a nucleus gene is represented twice in the gene-set ("genotype"), one contribution being provided by each parent during sexual reproduction. Each "gene" is located at a particular place (a "locus" - the Latin word for place; plural "loci") on corresponding chromosomes (homologues), one from each parent. Any gene may have several functional forms in the species as a whole, and each of these may lead to outwardly different "effects" (= an average result in the phenotype considered over a large sample of gene-backgrounds and environments). These functional forms are "alleles" (or "allelomorphs", the original term). If both alleles at a gene have the same phenotypic effect (are the "same"), the gene is said to be "homozygous": if each allele at a gene is different in effect, the gene is "heterozygous". The average phenotypic outcome may also depend upon how alleles interact with their own homologous partner in the disomic genotype ("dominance"), and on how these alleles interact with those of other genes at other loci which also affect this phenotypic trait ("epistasis"). Notice that we have combined classical genetics ideas with those of statistics in this exposition. Terms such as "gene", "homologue", "allele", "homozygous" and "heterozygous" are genetical: but "effect" is statistical, and refers to the average observed over an infinity of backgrounds, both genetical and environmental. Thus, we have very sneakily defined the "genetic value" (or "genotypic" value) as the infinity mean of all the *phenotypes* it can ever produce in time and space! Before molecular genetics, there simply was no other way to do it! And after molecular genetics, this is still the most utilitarian way to tackle the idea of genetic value! Also notice that we have openly used fundamentals of reproductive biology behind the genetics. The founder of Quantitative Genetics - Sir Ronald Fisher - perceived all of this when he proposed the first mathematics of this branch of Genetics [Fisher R.A. (1930). The Genetical theory of Natural Selection. Oxford Clarendon Press. He sought to define a single statistical summary of all the variance arising from phenotypic change during the course of genetical assortment and segregation, which he called the "genetic" variance. His residual genotypic variance (which he called simply the "residual") represented that part of assortment which did not lead to phenotypic change, although the genes themselves had in fact been subject to meiosis and syngamy, of course. These partitions subsequently became the familiar subdivisions of the "additive" (A) and "dominance" (D) variances, respectively. These later names are utterly misleading and very unfortunate, and have led to much confusion as to what they mean genetically. [It would have been better for posterity had they been named "Assortative" (A), and "Stable" (S).] A more gene-focused partitioning was invented by Mather and Jinks in 1971, but they also were statisticians rather than geneticists, had their own rather opqaque symbolism, and became somewhat

overwhelmed by the Fisherian approach. Added to all of this was the problem that Fisher's underlying reproduction model (mating system) was unrealistically simplistic: whilst it facilitated solving the equations, it didn't describe many real-life scenarios. Fortunately, Wright (in the 1950s) did provide the means to overcome complex mating systems, but its own complexity minimised its popular adoption. Early in this millennium [Heredity (2003) 91: 85-89], a comprehensive linking together of all of these approaches reconciled their various meanings and relationships (as well as correcting an error), and suggested that new partitions reflecting real homozygosis [a] and heterozygosis [d] should replace the present "additive" [A] and "dominance" [D] subdivisions. This paper revealed, by the way, that the Additive Genetical variance consists of all the homozygote variance, plus part of the dominance variance, and a frequency-weighted covariance between homozygote and heterozygote gene effects. The so-called Dominance variance contains only the remainder of the overall dominance variance of the gene in question, being therefore very misleading indeed! It should be understood, however, that either method of partitioning still accounts for all of the genotypic variance in the model being used: it's the way it has been divided up which is being debated. At least the Environmental variance is much more straight-forward. This can be subdivided into a pure environmental component (E) and an interaction component (I) describing the gene-environment interaction. The overall "single gene" model can be written as:

$$P = a + d + E + I.$$

Expansion of the model to multiple genes (loci) is still not resolved satisfactorily, and until that is solved it is not possible to account for epistasis. The problem is being tackled currently. The contribution of those components cannot be determined in a single individual, but they can be estimated for whole populations by estimating the variances for those components, denoted as:

$$V_P = V_a + V_d + V_E + V_I$$

The heritability of a trait is the proportion of the total (i.e. phenotypic) variance (V_P) that is explained by the total genotypic variance (V_G). This is known also as the "broad sense" heritability (H^2). If only Additive genetic variance (V_A) is used in the numerator, the heritability is "narrow sense" (h^2). Unfortunately, this is often simply called "heritability", with little reflection about its true meaning. The broadsense heritability indicates the genotypic determination of the phenotype: while the latter estimates the degree of assortative disequlibrium in the trait. Fisher proposed that this narrow-sense heritability might be appropriate in considering the results of natural selection, focusing as it does on disequilibrium: and it has been used also for predicting the results of artificial selection. This latter usage seems to be inappropriate, however, as breeders are interested in steering attributes towards new phenotypes (that is in utilising all the gene effects), rather than simply exploiting disequilibrium. But old dogmas die hard!

Resemblance between relatives

Central in estimating the variances for the various components is the principle of relatedness. A child has a father and a mother. Consequently, the child and father share 50% of their alleles, as do the child and the mother. However, the mother and father normally do not share alleles as a result of shared ancestors. Similarly, two full siblings share also on average 50% of the alleles with each other, while half siblings share only 25% of their alleles. This variation in relatedness can be used to estimate which proportion of the total phenotypic variance (V_P) is explained by the above-mentioned components.

Correlated traits

Although some genes have only an effect on a single trait, many genes have an effect on various traits. Because of this, a change in a single gene will have an effect on all those traits. This is calculated using covariances, and the phenotypic covariance (Cov_P) between two traits can be partitioned in the same way as the variances described above. The genetic correlation is calculated by dividing the covariance between the additive genetic effects of two traits by the square root of the product of the variances for the additive genetic effects of the two traits:

$$\text{Genetic correlation} = \frac{\text{Cov}(A_1, A_2)}{\sqrt{V_{A1} * V_{A2}}}$$

Chapter- 13

Medical Genetics

Medical genetics is the specialty of medicine that involves the diagnosis and management of hereditary disorders. Medical genetics differs from Human genetics in that human genetics is a field of scientific research that may or may not apply to medicine, but medical genetics refers to the application of genetics to medical care. For example, research on the causes and inheritance of genetic disorders would be considered within both human genetics and medical genetics, while the diagnosis, management, and counseling of individuals with genetic disorders would be considered part of medical genetics.

In contrast, the study of typically non-medical phenotypes such as the genetics of eye color would be considered part of human genetics, but not necessarily relevant to medical genetics (except in situations such as albinism). *Genetic medicine* is a newer term for medical genetics and incorporates areas such as gene therapy, personalized medicine, and the rapidly emerging new medical specialty, predictive medicine.

Scope

Medical genetics encompasses many different areas, including clinical practice of physicians, genetic counselors, and nutritionists, clinical diagnostic laboratory activities, and research into the causes and inheritance of genetic disorders. Examples of conditions that fall within the scope of medical genetics include birth defects and dysmorphology, mental retardation, autism, metabolic and mitochondrial disorders, skeletal dysplasia, connective tissue disorders, cancer genetics, teratogens, and prenatal diagnosis. Medical genetics is increasingly becoming relevant to many common diseases. Overlaps with other medical specialties are beginning to emerge, as recent advances in genetics are revealing etiologies for neurologic, endocrine, cardiovascular, pulmonary, ophthalmologic, renal, psychiatric, and dermatologic conditions.

Subspecialties

In some ways, many of the individual fields within medical genetics are hybrids between clinical care and research. This is due in part to recent advances in science and technology that have enabled an unprecedented understanding of genetic disorders.

Clinical genetics

Clinical genetics is the practice of clinical medicine with particular attention to hereditary disorders. Referrals are made to genetics clinics for a variety of reasons, including birth defects, developmental delay, autism, epilepsy, short stature, and many others. Examples of genetic syndromes that are commonly seen in the genetics clinic include chromosomal rearrangements, Down syndrome, DiGeorge syndrome (22q11.2 Deletion Syndrome), Fragile X syndrome, Marfan syndrome, Neurofibromatosis, Turner syndrome, and Williams syndrome.

Metabolic/biochemical genetics

Metabolic (or biochemical) genetics involves the diagnosis and management of inborn errors of metabolism in which patients have enzymatic deficiencies that perturb biochemical pathways involved in metabolism of carbohydrates, amino acids, and lipids. Examples of metabolic disorders include galactosemia, glycogen storage disease, lysosomal storage disorders, metabolic acidosis, peroxisomal disorders, phenylketonuria, and urea cycle disorders.

Cytogenetics

Cytogenetics is the study of chromosomes and chromosome abnormalities. While cytogenetics historically relied on microscopy to analyze chromosomes, new molecular technologies such as array comparative genomic hybridization are now becoming widely used. Examples of chromosome abnormalities include aneuploidy, chromosomal rearrangements, and genomic deletion/duplication disorders.

Molecular genetics

Molecular genetics involves the discovery of and laboratory testing for DNA mutations that underlie many single gene disorders. Examples of single gene disorders include achondroplasia, cystic fibrosis, Duchenne muscular dystrophy, hereditary breast cancer (BRCA1/2), Huntington disease, Marfan syndrome, Noonan syndrome, and Rett syndrome. Molecular tests are also used in the diagnosis of syndromes involving epigenetic abnormalities, such as Angelman syndrome, Beckwith-Wiedemann syndrome, Prader-willi syndrome, and uniparental disomy.

Mitochondrial genetics

Mitochondrial genetics concerns the diagnosis and management of mitochondrial disorders, which have a molecular basis but often result in biochemical abnormalities due to deficient energy production.

There exists some overlap between medical genetic diagnostic laboratories and molecular pathology.

Genetic Counseling

Genetic counseling is the process of providing information about genetic conditions, diagnostic testing, and risks in other family members, within the framework of nondirective counseling. Genetic counselors are non-physician members of the medical genetics team who specialize in family risk assessment and counseling of patients regarding genetic disorders. The precise role of the genetic counselor varies somewhat depending on the disorder.

History

Although genetics has its roots back in the 19th century with the work of the Bohemian monk Gregor Mendel and other pioneering scientists, human genetics emerged later. It started to develop, albeit slowly, during the first half of the 20th century. Mendelian (single-gene) inheritance was studied in a number of important disorders such as albinism, brachydactyly (short fingers and toes), and hemophilia. Mathematical approaches were also devised and applied to human genetics. Population genetics was created.

Medical genetics was a late developer, emerging largely after the close of World War II (1945) when the eugenics movement had fallen into disrepute. The Nazi misuse of eugenics sounded its death knell. Shorn of eugenics, a scientific approach could be used and was applied to human and medical genetics. Medical genetics saw an increasingly rapid rise in the second half of the 20th century and continues in the 21st century.

Current practice

The clinical setting in which patients are evaluated determines the scope of practice, diagnostic, and therapeutic interventions. For the purposes of general discussion, the typical encounters between patients and genetic practitioners may involve:

- Referral to an out-patient genetics clinic (pediatric, adult, or combined) or an in-hospital consultation, most often for diagnostic evaluation.
- Specialty genetics clinics focusing on management of inborn errors of metabolism, skeletal dysplasia, or lysosomal storage diseases.

- Referral for counseling in a prenatal genetics clinic to discuss risks to the pregnancy (advanced maternal age, teratogen exposure, family history of a genetic disease), test results (abnormal maternal serum screen, abnormal ultrasound), and/or options for prenatal diagnosis (typically amniocentesis or chorionic villus sampling).
- Multidisciplinary specialty clinics that include a clinical geneticist or genetic counselor (cancer genetics, cardiovascular genetics, craniofacial or cleft lip/palate, hearing loss clinics, muscular dystrophy/neurodegenerative disorder clinics).

Diagnostic evaluation

Each patient will undergo a diagnostic evaluation tailored to their own particular presenting signs and symptoms. The geneticist will establish a differential diagnosis and recommend appropriate testing. Increasingly, clinicians use SimulConsult, paired with the National Library of Medicine Gene Review articles, to narrow the list of hypotheses (known as the differential diagnosis) and identify the tests that are relevant for a particular patient. These tests might evaluate for chromosomal disorders, inborn errors of metabolism, or single gene disorders.

Chromosome studies

Chromosome studies are used in the general genetics clinic to determine a cause for developmental delay/mental retardation, birth defects, dysmorphic features, and/or autism. Chromosome analysis is also performed in the prenatal setting to determine whether a fetus is affected with aneuploidy or other chromosome rearrangements. Finally, chromosome abnormalities are often detected in cancer samples. A large number of different methods have been developed for chromosome analysis:

- Chromosome analysis using a karyotype involves special stains that generate light and dark bands, allowing identification of each chromosome under a microscope.
- Fluorescence in situ hybridization (FISH) involves fluorescent labeling of probes that bind to specific DNA sequences, used for identifying aneuploidy, genomic deletions or duplications, characterizing chromosomal translocations and determining the origin of ring chromosomes.
- Chromosome painting is a technique that uses fluorescent probes specific for each chromosome to differentially label each chromosome. This technique is more often used in cancer cytogenetics, where complex chromosome rearrangements can occur.
- Array comparative genomic hybridization is a new molecular technique that involves hybridization of an individual DNA sample to a glass slide or microarray chip containing molecular probes (ranging from large ~200kb bacterial artificial chromosomes to small oligonucleotides) that represent unique regions of the genome. This method is particularly sensitive for detection of genomic gains or losses across the genome but does not detect balanced translocations or

distinguish the location of duplicated genetic material (for example, a tandem duplication versus an insertional duplication).

Basic metabolic studies

Biochemical studies are performed to screen for imbalances of metabolites in the bodily fluid, usually the blood (plasma/serum) or urine, but also in cerebrospinal fluid (CSF). Specific tests of enzyme function (either in leukocytes, skin fibroblasts, liver, or muscle) are also employed under certain circumstances. In the US, the newborn screen incorporates biochemical tests to screen for treatable conditions such as galactosemia and phenylketonuria (PKU). Patients suspected to have a metabolic condition might undergo the following tests:

- Quantitative amino acid analysis is typically performed using the ninhydrin reaction, followed by liquid chromatography to measure the amount of amino acid in the sample (either urine, plasma/serum, or CSF). Measurement of amino acids in plasma or serum is used in the evaluation of disorders of amino acid metabolism such as urea cycle disorders, maple syrup urine disease, and PKU. Measurement of amino acids in urine can be useful in the diagnosis of cystinuria or renal Fanconi syndrome as can be seen in cystinosis.
- Urine organic acid analysis can be either performed using quantitative or qualitative methods, but in either case the test is used to detect the excretion of abnormal organic acids. These compounds are normally produced during bodily metabolism of amino acids and odd-chain fatty acids, but accumulate in patients with certain metabolic conditions.
- The acylcarnitine combination profile detects compounds such as organic acids and fatty acids conjugated to carnitine. The test is used for detection of disorders involving fatty acid metabolism, including MCAD.
- Pyruvate and lactate are byproducts of normal metabolism, particularly during anaerobic metabolism. These compounds normally accumulate during exercise or ischemia, but are also elevated in patients with disorders of pyruvate metabolism or mitochondrial disorders.
- Ammonia is an end product of amino acid metabolism and is converted in the liver to urea through a series of enzymatic reactions termed the urea cycle. Elevated ammonia can therefore be detected in patients with urea cycle disorders, as well as other conditions involving liver failure.
- Enzyme testing is performed for a wide range of metabolic disorders to confirm a diagnosis suspected based on screening tests.

Molecular studies

- DNA sequencing is used to directly analyze the genomic DNA sequence of a particular gene. In general, only the parts of the gene that code for the expressed protein (exons) and small amounts of the flanking untranslated regions and introns are analyzed. Therefore, although these tests are highly specific and

sensitive, they do not routinely identify all of the mutations that could cause disease.
- DNA methylation analysis is used to diagnose certain genetic disorders that are caused by disruptions of epigenetic mechanisms such as genomic imprinting and uniparental disomy.
- Southern blotting is an early technique basic on detection of fragments of DNA separated by size through gel electrophoresis and detected using radiolabeled probes. This test was routinely used to detect deletions or duplications in conditions such as Duchenne muscular dystrophy but is being replaced by high-resolution array comparative genomic hybridization techniques. Southern blotting is still useful in the diagnosis of disorders caused by trinucleotide repeats.
- Short tandem repeats are unique markers that can be used to determine haplotypes and are used in identity testing for maternal cell contamination.

Treatments

Each cell of the body contains the hereditary information (DNA) wrapped up in structures called chromosomes. Since genetic syndromes are typically the result of alterations of the chromosomes or genes, there is no treatment currently available that can correct the genetic alterations in every cell of the body. Therefore, there is currently no "cure" for genetic disorders. However, for many genetic syndromes there is treatment available to manage the symptoms. In some cases, particularly inborn errors of metabolism, the mechanism of disease is well understood and offers the potential for dietary and medical management to prevent or reduce the long-term complications. In other cases, infusion therapy is used to replace the missing enzyme. Current research is actively seeking to use gene therapy or other new medications to treat specific genetic disorders.

Management of Metabolic disorders

In general, metabolic disorders arise from enzyme deficiencies that disrupt normal metabolic pathways. For instance, in the hypothetical example:

```
    A ---> B ---> C ---> D          AAAA ---> BBBBBB ---> CCCCCCCCCC ---
> (no D)
       X       Y       Z                    X            Y              (no
Z)
```

Compound "A" is metabolized to "B" by enzyme "X", compound "B" is metabolized to "C" by enzyme "Y", and compound "C" is metabolized to "D" by enzyme "Z". If enzyme "Z" is missing, compound "D" will be missing, while compounds "A", "B", and "C" will build up. The pathogenesis of this particular condition could result from lack of compound "D", if it is critical for some cellular function, or from toxicity due to excess "A", "B", and/or "C". Treatment of the metabolic disorder could be achieved through dietary supplementation of compound "D" and dietary restriction of compounds "A", "B", and/or "C" or by treatment with a medication that promoted disposal of excess "A", "B",

or "C". Another approach that can be taken is enzyme replacement therapy, in which a patient is given an infusion of the missing enzyme.

- Diet

Dietary restriction and supplementation are key measures taken in several well-known metabolic disorders, including galactosemia, phenylketonuria (PKU), maple syrup urine disease, organic acidurias and urea cycle disorders. Such restrictive diets can be difficult for the patient and family to maintain, and require close consultation with a nutritionist who has special experience in metabolic disorders. The composition of the diet will change depending on the caloric needs of the growing child and special attention is needed during a pregnancy if a woman is affected with one of these disorders.

- Medication

Medical approaches include enhancement of residual enzyme activity (in cases where the enzyme is made but is not functioning properly), inhibition of other enzymes in the biochemical pathway to prevent buildup of a toxic compound, or diversion of a toxic compound to another form that can be excreted. Examples include the use of high doses of pyridoxine (vitamin B6) in some patients with homocystinuria to boost the activity of the residual cystathione synthase enzyme, administration of biotin to restore activity of several enzymes affected by deficiency of biotinidase, treatment with NTBC in Tyrosinemia to inhibit the production of succinylacetone which causes liver toxicity, and the use of sodium benzoate to decrease ammonia build-up in urea cycle disorders.

- Enzyme replacement therapy

Certain lysosomal storage diseases are treated with infusions of a recombinant enzyme (produced in a laboratory), which can reduce the accumulation of the compounds in various tissues. Examples include Gaucher disease, Fabry disease, Mucopolysaccharidoses and Glycogen storage disease type II. Such treatments are limited by the ability of the enzyme to reach the affected areas (the blood brain barrier prevents enzyme from reaching the brain, for example), and can sometimes be associated with allergic reactions. The long-term clinical effectiveness of enzyme replacement therapies vary widely among different disorders.

Other examples

- Angiotensin receptor blockers in Marfan syndrome & Loeys-Dietz
- Bone marrow transplantation
- Gene therapy

Career paths and training

There are a variety of career paths within the field of medical genetics, and naturally the training required for each area differs considerably. It should be noted that the

information included here applies to the typical pathways in the United States and there may be differences in other countries. US Practitioners in clinical, counseling, or diagnostic subspecialties generally obtain board certification through the American Board of Medical Genetics.

Career	Degree	Description	Training
Clinical Geneticist	MD or MD/PhD	A **Clinical geneticist** is typically a physician who evaluates patients in the office or as a hospital consultation. This process includes a medical history, family history (pedigree), a detailed physical examination, reviewing objective data such as imaging and test results, establishing a differential diagnosis, and recommending appropriate diagnostic tests.	College (4 yrs) → Medical school (4 yrs) → Primary residency (2-3 yrs) → Residency in Clinical genetics (2 yrs). Some Clinical geneticists also obtain a PhD degree (4-7 yrs). A new residency track offers a 4 yr primary residency in Clinical genetics immediately after finishing Medical school.
Genetic Counselor	MS	A **Genetic counselor** specializes in communication of genetic information to patients and families. Genetic counselors often work closely with Clinical geneticists or other physicians (such as Obstetricians or Oncologists) and often convey the results of the recommended tests.	College (4 yrs) → Graduate program in Genetic counseling (2 yrs).
Metabolic nurse and/or nutritionist	BA/BS, MS, RN	One of the critical aspects of the management of patients with metabolic disorders is the appropriate nutritional intervention (either restricting the compound that cannot be metabolized, or supplementing compounds that are deficient as the result of an enzyme deficiency). The metabolic nurse and nutritionist play important roles in coordinating the dietary management.	College (4 yrs) → Nursing school or graduate training in nutrition.
Biochemical Diagnostics	PhD, MD, or MD/PhD	Individuals who specialize in **Biochemical genetics** typically work in the diagnostic laboratory, analyzing and interpreting specialized biochemical tests that	College (4 yrs) → Graduate school (PhD, usually 4–7 years) and/or Medical school (MD, 4 years)

		measure amino acids, organic acids, and enzyme activity. Some Clinical Geneticists are also board certified in Biochemical Genetics.	
Cytogenetic Diagnostics	PhD, MD, or MD/PhD	Individuals who specialize in **Cytogenetics** typically work in the diagnostic laboratory, analyzing and interpreting karyotypes, FISH, and comparative genomic hybridization tests. Some Clinical Geneticists are also board certified in Cytogenetics.	College (4 yrs) → Graduate school (PhD, usually 4–7 years) and/or Medical school (MD, 4 years)
Molecular Diagnostics	PhD, MD, or MD/PhD	Individuals who specialize in **Molecular genetics** typically work in the diagnostic laboratory, analyzing and interpreting specialized genetic tests that look for disease-causing changes (mutations) in the DNA. Some examples of molecular diagnostic tests include DNA sequencing and Southern blotting.	College (4 yrs) → Graduate school (PhD, usually 4–7 years) and/or Medical school (MD, 4 years)
Research Geneticist	PhD, MD, or MD/PhD	Any researcher who studies the genetic basis of human disease or uses model organisms to study disease mechanisms could be considered a Research Geneticist. Many of the clinical career paths also include basic or translational research, and thus individuals in the field of medical genetics often participate in some form of research.	College (4 yrs) → Graduate school (PhD, usually 4–7 years) and/or Medical school (MD, 4 years) → Post-doctoral research training (usually 3+ years)
Laboratory Technician	BS or MS	Technicians in the diagnostic or research labs handle samples and run the assays at the bench. Often these individuals are promoted to supervisory positions.	College (4 yrs), may have higher degree (MS, 2+ years)

Ethical, legal and social implications

Genetic information provides a unique type of knowledge about an individual and his/her family, fundamentally different than a typically laboratory test that provides a "snapshot"

of an individual's health status. The unique status of genetic information and inherited disease has a number of ramifications with regard to ethical, legal, and societal concerns.

Societies

The more empirical approach to human and medical genetics was formalized by the founding in 1948 of the American Society of Human Genetics. The Society first began annual meetings that year (1948) and its international counterpart, the International Congress of Human Genetics, has met every 5 years since its inception in 1956. The Society publishes the American Journal of Human Genetics on a monthly basis.

Medical genetics is now recognized as a distinct medical specialty in the U.S. with its own approved board (the American Board of Medical Genetics) and clinical specialty college (the American College of Medical Genetics). The College holds an annual scientific meeting, publishes a monthly journal, Genetics in Medicine, and issues position papers and clinical practice guidelines on a variety of topics relevant to human genetics.

Research

The broad range of research in medical genetics reflects the overall scope of this field, including basic research on genetic inheritance and the human genome, mechanisms of genetic and metabolic disorders, translational research on new treatment modalities, and the impact of genetic testing

Basic genetics research

Basic research geneticists usually undertake research in universities, biotechnology firms and research institutes.

Allelic architecture of disease

Sometimes the link between a disease and an unusual gene variant is more subtle. The genetic architecture of common diseases is an important factor in determining the extent to which patterns of genetic variation influence group differences in health outcomes. According to the common disease/common variant hypothesis, common variants present in the ancestral population before the dispersal of modern humans from Africa play an important role in human diseases. Genetic variants associated with Alzheimer disease, deep venous thrombosis, Crohn disease, and type 2 diabetes appear to adhere to this model. However, the generality of the model has not yet been established and, in some cases, is in doubt. Some diseases, such as many common cancers, appear not to be well described by the common disease/common variant model.

Another possibility is that common diseases arise in part through the action of combinations of variants that are individually rare. Most of the disease-associated alleles discovered to date have been rare, and rare variants are more likely than common variants

to be differentially distributed among groups distinguished by ancestry. However, groups could harbor different, though perhaps overlapping, sets of rare variants, which would reduce contrasts between groups in the incidence of the disease.

The number of variants contributing to a disease and the interactions among those variants also could influence the distribution of diseases among groups. The difficulty that has been encountered in finding contributory alleles for complex diseases and in replicating positive associations suggests that many complex diseases involve numerous variants rather than a moderate number of alleles, and the influence of any given variant may depend in critical ways on the genetic and environmental background. If many alleles are required to increase susceptibility to a disease, the odds are low that the necessary combination of alleles would become concentrated in a particular group purely through drift.

Population substructure in genetics research

One area in which population categories can be important considerations in genetics research is in controlling for confounding between population substructure, environmental exposures, and health outcomes. Association studies can produce spurious results if cases and controls have differing allele frequencies for genes that are not related to the disease being studied, although the magnitude of this problem in genetic association studies is subject to debate. Various methods have been developed to detect and account for population substructure, but these methods can be difficult to apply in practice.

Population substructure also can be used to advantage in genetic association studies. For example, populations that represent recent mixtures of geographically separated ancestral groups can exhibit longer-range linkage disequilibrium between susceptibility alleles and genetic markers than is the case for other populations. Genetic studies can use this admixture linkage disequilibrium to search for disease alleles with fewer markers than would be needed otherwise. Association studies also can take advantage of the contrasting experiences of racial or ethnic groups, including migrant groups, to search for interactions between particular alleles and environmental factors that might influence health.