CURRENT TOPICS IN DEVELOPMENTAL BIOLOGY

"A meeting-ground for critical review and discussion of developmental processes"
A.A. Moscona and Alberto Monroy (Volume 1, 1966)

Academic Press is an imprint of Elsevier
50 Hampshire Street, 5th Floor, Cambridge, MA 02139, USA
525 B Street, Suite 1800, San Diego, CA 92101-4495, USA
The Boulevard, Langford Lane, Kidlington, Oxford OX5 1GB, UK
125 London Wall, London, EC2Y 5AS, UK

First edition 2016

Notices
Knowledge and best practice in this field are constantly changing. As new research and experience broaden our understanding, changes in research methods, professional practices, or medical treatment may become necessary.

Practitioners and researchers must always rely on their own experience and knowledge in evaluating and using any information, methods, compounds, or experiments described herein. In using such information or methods they should be mindful of their own safety and the safety of others, including parties for whom they have a professional responsibility.

To the fullest extent of the law, neither the Publisher nor the authors, contributors, or editors, assume any liability for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions, or ideas contained in the material herein.

ISBN: 978-0-12-417194-7
ISSN: 0070-2153

Publisher: Zoe Kruze
Acquisition Editor: Zoe Kruze
Editorial Project Manager: Sarah Lay
Production Project Manager: Radhakrishnan Lakshmanan
Cover Designer: Miles Hitchen

Typeset by SPi Global, India

CONTRIBUTORS

P. Balaresque
Laboratoire d´Anthropologie Moléculaire et Imagerie de Synthèse (AMIS), UMR5288-CNRS & Université Paul Sabatier-Toulouse III, Toulouse, France

J. Bergelson
University of Chicago, Chicago, IL, United States

M.-A. Félix
Institut de Biologie Ecole Normale Supérieure, CNRS, Paris, France

N. Goudemand
Univ Lyon, ENS Lyon, CNRS, Université Claude Bernard Lyon 1, Institut de Génomique Fonctionnelle de Lyon, UMR 5242, Lyon Cedex 07, France

M.D. Gupta
Max Planck Institute for Plant Breeding Research, Köln, Germany

T.E. King
University of Leicester, Leicester, United Kingdom

L.-F. Li
Washington University in St. Louis, St. Louis, MO, United States; Key Laboratory of Molecular Epigenetics of the Ministry of Education (MOE), Northeast Normal University, Changchun, PR China

J.H. Massey
University of Michigan, Ann Arbor, MI, United States

A. Monteiro
National University of Singapore; Yale-NUS College, Singapore, Singapore

B. Morizot
Université Aix-Marseille, CNRS UMR 7304, Aix-en-Provence, France

K.M. Olsen
Washington University in St. Louis, St. Louis, MO, United States

V. Orgogozo
Institut Jacques Monod, UMR 7592, CNRS-Université Paris Diderot, Sorbonne Paris Cité, Paris, France

S. Pantalacci
Univ Lyon, ENS Lyon, CNRS, Université Claude Bernard Lyon 1, Laboratory of Biology and Modelling of the Cell, UMR 5239, INSERM U1210, Lyon Cedex 07, France

A.E. Peluffo
Institut Jacques Monod, UMR 7592, CNRS-Université Paris Diderot, Sorbonne Paris Cité, Paris, France

C.C. Rittschof
University of Kentucky, Lexington, KY, United States

G.E. Robinson
Carl R. Woese Institute for Genomic Biology, University of Illinois, Urbana-Champaign, Urbana, IL, United States

F. Roux
INRA, Laboratoire des Interactions Plantes-Microorganismes (LIPM), UMR441; CNRS, Laboratoire des Interactions Plantes-Microorganismes (LIPM), UMR2594, Castanet-Tolosan, France

S. Urdy
University of Zürich, Institute of Physics, Zürich, Switzerland

P.J. Wittkopp
University of Michigan, Ann Arbor, MI, United States

PREFACE

Wilhelm Johannsen introduced the conceptual distinction between genotype and phenotype in 1911, and since then biologists have tried to disentangle the fascinating connection between genotypes and phenotypes, that is how nucleotide changes produce observable changes in living organisms. The elucidation of the actual genes and mutations responsible for observable differences between individuals and species started in the 1980s and has revolutionized the study of evolution. A wide array of powerful methods have been developed in recent years to find the genes and the mutations responsible for evolutionary changes so that empirical data on the "loci of evolution" are accumulating at a fast pace. The field of evolutionary genetics has moved from pure theoretical computations of changes in allele frequencies in populations to assessments of the actual mutations that occurred in populations and that cause observable changes in phenotype, bringing up new ideas and new questions. This volume aims to provide an overview of current knowledge and ongoing research on the genes and the mutations responsible for phenotypic evolution.

Chapter 1 highlights the fact that the term "gene" can have two different meanings in biology research papers and tries to clarify the distinction between both, to avoid confusion and misconceptions. Both concepts of "gene" are still extremely used and useful in today's research and are part of two relevant frameworks for explaining the biological world. Chapters 2–4 examine the entire set of genes and mutations (about a hundred of mutations/genes for each chapter) that have been identified so far as responsible for particular phenotypes, *Drosophila* pigmentation patterns (Chapter 2), Crop seed and fruit retention (Chapter 3), and biotic interactions involving *Arabidopsis thaliana* (Chapter 4). These chapters are particularly representative of the current state of evolutionary genetics research. Drosophila pigmentation (Chapter 2) is a premier model that has led to important insights on the precise genetic modifications responsible for phenotypic evolution and on the predominance of *cis*-regulatory changes. Chapter 3 points out that domestication generally involves the repeated evolution of similar traits in various species, with causal mutations often affecting the same genes. Collecting information on the genes causing domestication traits can thus help to define candidate genes for genetic improvement and domestication of other plant species. Chapter 4 will make readers realize that *A. thaliana* is, besides humans maybe, the Eukaryotic species whose genetic basis of

phenotypic variation is best understood. Examination of the genetics underlying interactions between *A. thaliana* and other species reveals that integrating multiple genetic effects within a natural environment and at the level of ecosystems is far from simple and poses interesting challenges.

Very little is known about the genetic basis of behavioral evolution. Indeed, most evolutionary changes in phenotype whose causal genes have been identified so far are morphological and physiological. Chapter 5 examines whether the concept of genetic toolkit, which derives from morphological studies, may also apply to behavioral evolution. Chapter 6 manipulates two important concepts in contemporary biology, gene networks and novelties, and argues that the mutations underlying the appearance of a novel trait are likely to affect particular positions within gene networks. Chapters 7 and 8 illustrate the fact that genes are not the unique factors that should be taken into account when trying to understand how phenotypes evolve. Chapter 7 shows that the shape and diversification of epithelial tissues result from a complex interplay between genes, signaling pathways, and forces. Chapter 8 examines precisely, with clear-cut and well-chosen examples, the multiple interactions and feedbacks between genes, organisms, and environment at various levels and timescales, which together lead to phenotypic changes. The last chapter provides a comprehensive overview of the genes and mutations that are currently known to cause phenotypic diversity in actual human populations and in our recent past history.

This volume is being published at a time when many research laboratories are focusing on the evolution of diverse phenotypic traits and trying to dissect their genetic basis. While these ongoing studies might look, to the eyes of certain biologists, quite repetitive and unoriginal, this volume provides a good material to reflect on our current state of research and to try to detect new future explanatory frameworks and ideas to be discovered from our examination of the genes and the mutations responsible for evolution.

I thank all the authors for their hard work and dedication to the project, as well as all the reviewers who have helped me to reach a level of definitely high quality: Matt Rockman, Jean-Michel Gibert, Lin-Feng Li, Noah Whiteman, Kim Hoke, Alexis Matamoro-Vidal, Sophie Pantalacci, Armin Moczek, Michel Raymond, and one anonymous reviewer. As a scientist, I feel greatly honored and lucky to collaborate with those nice, smart, and committed scientists. My love for biology and my productivity in research would not be the same were I not surrounded by those people.

Virginie Orgogozo
Paris
April 2016

CHAPTER ONE

The "Mendelian Gene" and the "Molecular Gene": Two Relevant Concepts of Genetic Units

V. Orgogozo*,[1], A.E. Peluffo*, B. Morizot†
*Institut Jacques Monod, UMR 7592, CNRS-Université Paris Diderot, Sorbonne Paris Cité, Paris, France
†Université Aix-Marseille, CNRS UMR 7304, Aix-en-Provence, France
[1]Corresponding author: e-mail address: virginie.orgogozo@normalesup.org

Contents

Abstract

We focus here on two prevalent meanings of the word gene in research articles. On one hand, the gene, named here "molecular gene," is a stretch of DNA that is transcribed and codes for an RNA or a polypeptide with a known or presumed function (as in "gene network"), whose exact spatial delimitation on the chromosome remains a matter of debate, especially in cases with alternative splicing, antisense transcripts, etc. On the other hand, the gene, called here "Mendelian gene," is a segregating genetic unit which is detected through phenotypic differences associated with different alleles at the same locus (as in "gene flow"). We show that the "Mendelian gene" concept is still extensively used today in biology research and is sometimes confused with the "molecular gene." We try here to clarify the distinction between both concepts. Efforts to delineate the beginning and the end of the DNA sequence corresponding to the "Mendelian gene" and the "molecular gene" reveal that both entities do not always match. We argue that both concepts are part of two relevant frameworks for explaining the biological world.

Current Topics in Developmental Biology, Volume 119
ISSN 0070-2153
http://dx.doi.org/10.1016/bs.ctdb.2016.03.002

1. INTRODUCTION

Since the early days, biologists have tried to extract general concepts from their observation of the living forms in order to increase their understanding of the surrounding world. Familiar examples include the concepts of species, ecosystem, symbiosis, or sexual selection. A concept becomes especially relevant when it can account for observations that were so far unexplained. In the history of biology, new discoveries and new theories have often challenged the underlying ideas and definitions behind existing concepts, and the meaning of certain biological concepts has evolved through time.

The concept of "gene" has, since its inception, been a central organizing notion within biology. The word "gene" was introduced by Johannsen (1911) from Hugo de Vries' "pangenes" (de Vries, 1889), themselves derived from Darwin's original, and erroneous, model of blending heredity, "pangenesis" (Darwin, 1868, 1871). According to Johannsen, the gene is "nothing but a very applicable little word" that helps to explain the inheritance of visible characters, and the sum of all genes is called the "genotype" (Johannsen, 1911). Johannsen insisted that "we do not know a genotype but we are able to demonstrate genotypical differences" and therefore that the genotypes are only accessible to the experimenter by comparing phenotypic traits in different organisms. Johannsen thought that what lies in the zygote are "potentialities" to develop a given phenotype and that it is these potentialities which segregate in the form of genes which are inherited (Johannsen, 1911). Looking back at Johannsen's writings, it is not clear whether in his view genes were necessarily connected to a phenotype: it seems theoretically possible to imagine that certain genes were simply transmitted to the progeny without having any phenotypic effect. Today biologists still struggle to find a consensual and generally accepted definition of the "gene." In 2006, 25 scientists of the Sequence Ontology Consortium, which ultimately aims to describe the features of DNA sequences, spent 2 days of long heated discussions to come up with a consensual definition of the gene (see Table 1; Pearson, 2006). More recently, several articles and books dealing with the definition of the term "gene" have been published (for example, Falk, 2010; Gerstein et al., 2007; Griffiths & Stotz, 2013; Pradeu, 2015), showing that the question of "what is a gene?" remains important.

Table 1 Definition of the Terms "Gene", "Allele," and "Locus" According to Several Biological Databases Consortia and Textbooks

Human Genome Nomenclature Organization http://www.genenames.org/about/guidelines#criteria A *gene* is defined as a DNA segment that contributes to phenotype/function. In the absence of demonstrated function a *gene* may be characterized by sequence, transcription or homology.
Sequence Ontology Consortium (Pearson, 2006) A *gene* is a locatable region of genomic sequence, corresponding to a unit of inheritance, which is associated with regulatory regions, transcribed regions and/or other functional sequence regions.
Ensembl Consortium http://www.ensembl.org/info/genome/genebuild/genome_annotation.html An Ensembl *gene* includes any spliced transcripts with overlapping coding sequence, with the exception of manually annotated readthrough genes which are annotated as a separate locus.
Population Genetics Textbook (Hedrick, 2011) *Allele*: Different form of a gene. *Gene*: Unit of inheritance that is transmitted from parents to offspring. *Locus*: Place where a particular gene resides in the genome.
Molecular Biology of the Cell (Alberts et al., 2008) *Allele*: One of several alternative forms of a gene. In a diploid cell each gene will typically have two alleles, occupying the corresponding position (locus) on homologous chromosomes. *Gene*: Region of DNA that is transcribed as a single unit and carries information for a discrete hereditary characteristic, usually corresponding to a single protein or a single RNA.
Genetics and Analysis of Quantitative Traits (Lynch & Walsh, 1998, p. 51) DNA sequences that encode for particular products (proteins and RNAs) are referred to as *genes*, and their chromosomal locations are called *loci*. Most organisms have two copies of each of several chromosomes, in which case they are said to be diploid. Since DNA replication is an imperfect process, mutations arise, and as a consequence the two "copies" of each gene carried by diploid individuals need not be identical. The various forms of a gene are called *alleles*.
Quantitative Genetics (Falconer & Mackay, 1996, pp. 1–2) Suppose for simplicity that we were concerned with a certain autosomal *locus*, A, and that two different *alleles* at this locus, A1 and A2, were present among the individuals. […] Then there would be three possible genotypes, A1A1, A1A2, A2A2 (we are concerned here, as throughout the book, exclusively with diploid organisms.) […] Each A1A1 individual contains two A1 *genes* and each A1A2 contains one A1 *gene*.

Continued

Table 1 Definition of the Terms "Gene", "Allele," and "Locus" According to Several Biological Databases Consortia and Textbooks—cont'd

Genes IX (Lewin, 2006, p. 845 and 852, Glossary) A *gene* is the segment of DNA specifying a polypeptide chain; it includes regions preceding and following the coding region (leader and trailer), as well as intervening sequences (introns) between individual coding segments (exons). An *allele* is one of several alternative forms of a gene occupying a given locus on a chromosome.

When given, we quote the exact definition. When not available, we provide the most representative quote of the authors' definition of gene.

The history of the concept of gene, after Mendel (1866) and Johannsen (1911), has been recounted in several recent publications (see for example, Deutsch, 2012; Gerstein et al., 2007; Keller, 2009; Portin, 2002; Weber, 2005). In brief, classical genetics first considered the gene as an abstract unit of inheritance which explained phenotypic similarities between parents and children. Then, with the advent of molecular biology, genes became segments of DNA which are used as template to make RNA, which can then be used to build proteins, with particular biochemical activities. Soon after, the simple original idea that a gene should be associated with a single transcript was overturned by the discovery of multiple exceptions (alternative splicing, overlapping transcripts on opposite strands, protein-coding genes nested within the intron of another gene, transcription of most chromosomal DNA, etc.), stirring debates about which piece of DNA should be considered as a gene. In this chapter, rather than exploring the evolution of the concept of gene over the years, we focus on the meaning of "gene" at present. We show that many definitions are still employed today by professional biologists and that it is important to try to understand the meaning of the term "gene" in each context to try to avoid confusion and misunderstandings. We argue that all present concepts of genes can be classified into two main categories, the "Mendelian gene" and the "molecular gene." Most writings regarding the different meanings of the term "gene" over the history of biology have presented the "Mendelian gene" as the precursor, now dead, of the "molecular gene" (Deutsch, 2012; Falk, 1984; Griffiths & Stotz, 2013; Weber, 2005). We argue here that the "Mendelian gene" concept is still alive and has not been completely replaced by the "molecular gene" concept. We provide several concrete examples to illustrate that the "Mendelian gene" and the "molecular gene" do not overlap and that both concepts are currently useful to explain different aspects of our biological world.

2. THE "MENDELIAN GENE" AND THE "MOLECULAR GENE"

Following the insight of most authors (Falk, 1984; Gilbert, 2000; Moss, 2003; Pradeu, 2015; Stern, 2000; Weber, 2005), we distinguish two main embodiments for the concept of "gene." On one hand, *a gene is considered as a stretch of DNA that is transcribed and codes for an RNA or a polypeptide with a known or presumed function* (Gerstein et al., 2007; Pearson, 2006). This is what we name here a "molecular gene." To our knowledge, all genome databases consider the "gene" as the "molecular gene" (Table 1). The "molecular gene" leads to the production of RNAs and proteins, which is translated into a phenotype at the level of the organism. The impact of mutations (changes in the nucleotide sequence) in the "molecular gene" is revealed at the level of the gene expression, whether they induce a change in the amount of RNA/protein produced or in the actual sequence that is expressed. This change can then affect the phenotype of interest, but not necessarily. Experimentally, a "molecular gene" is usually revealed by its expression, that is production of an RNA of the corresponding sequence.

On the other hand, a gene is considered as a genetic unit which is transmitted from parents to offspring and which is detected through phenotypic differences associated with different alleles at the same locus. This is what we call here a "Mendelian gene." We note that the "Mendelian gene" is different from what Mendel called "factors" (Mendel, 1866; Olby, 1979). In Mendel's notation, what we call today homozygous diploid individuals were written ***a*** or ***A*** (rather than ***aa*** or ***AA***), whereas heterozygous were written ***Aa***, indicating that Mendel was indeed focused on the phenotypic state which is passed on (Morange, 2016; Olby, 1979). Mendel factors may be seen as elements that combine into specific arrangements, where the two original factors can sometimes fuse into a single one if they are identical.

The "Mendelian gene" can only be revealed and dealt with experimentally if a genotype difference exists and is associated with a phenotype difference. In a previous paper (Orgogozo, Morizot, & Martin, 2015), we distinguished an abstract entity that encompasses both genetic and phenotypic levels that we named "gephe." A "gephe" consists of a phenotypic change (two distinct phenotypic states), its associated variation at a genetic locus (two alleles), and their relationships. For example, resistance to imidazolinone herbicides that inhibit acetolactate synthase (ALS) is

associated with mutations in the ALS gene in *Arabidopsis thaliana* (Sathasivan, Haughn, & Murai, 1991). In 57 other plant species, substitutions in the ALS gene have also been either linked or conclusively shown through functional tests to be responsible for resistance to such herbicides (Baucom, 2016). Here the ALS-resistance gephe is composed of two alleles of the ALS gene, two phenotypic states (resistance and sensitivity to imidazolinone) and the relationship between the genetic change in ALS and the phenotypic difference under consideration. The ALS-resistance gephe is present in over 58 plant species. "Mendelian genes" that are detected through phenotypic differences are part of a "gephe."

A genetic locus can be conceptualized as a position on the genome. However, it is important to mention that it is not strictly speaking a spatial localization, since the number of loci is invariant with the level of ploidy. For instance, a diploid individual will not have its number of loci divided by two in his haploid gametes. Because it can carry alternative alleles, the locus is a genomic position at which segregates genetic variation. A genetic locus thus harbors distinct "Mendelian genes," each associated with various phenotypic states. Noticeably, certain biologists sometimes assume that the "Mendelian gene" concept is synonymous to the concept of locus (A. Martin & M. Rockman, personal communication). Such assimilation may arise when trying to find a spatial localization for the idea of genotype difference that is inherent to the concept of "Mendelian gene," and this is especially apparent in sentences such as "The latter approach was recently used in sunflowers, for example, to identify several flowering-time *genes* that colocalize with flowering-time QTLs" (Olsen & Wendel, 2013) or "we mapped the *gene* to a 45.1-kb region between two markers pcc17 and pcc14 on chromosome 11" (Pei et al., 2012). However, a progeny cannot be said to inherit one locus from his mother and one locus from his father, it is the "Mendelian genes" and not the genetic loci which are inherited. The concept of "Mendelian gene" is therefore closer to the concept of molecular allele than to the one of genetic locus.

Importantly, the physical embodiment of the "Mendelian gene" does not necessarily correspond to a "molecular gene." For example, in yeast the deletion of a telomere, a chromosome extremity which contains no "molecular genes," leads to cell cycle arrest (Sandell & Zakian, 1993) (see also later for other examples).

In summary, for it to be defined and tackled in an operational manner, the "Mendelian gene" requires a phenotype difference associated with a genotype difference, whereas the "molecular gene" requires transcription.

3. CURRENT LITERATURE OFTEN CONFUSES THE "MENDELIAN GENE" AND THE "MOLECULAR GENE" CONCEPTS

Table 2 provides a compilation of several quotes extracted from recent scientific publications which employ the term "gene," and Table 3 lists various usages of the word "gene" in fixed expressions. Both tables show that in certain instances the word "gene" corresponds to the concept of "molecular gene" explained earlier, in others to the concept of "Mendelian gene" and in yet other contexts to an intermingled combination of both concepts.

Because the word gene is often used without specifying whether it is the "molecular" or the "Mendelian" gene, confusion can arise, especially at the crossroads between different fields. One interesting example can be found

Table 2 A Few Examples of Current Usage of the Word "Gene" in Recent Research Papers

Science (Blomen et al., 2015) Many of the *genes* not targeted by our library encode olfactory receptors that are unlikely to be cell-essential.
Nature (Boettiger et al., 2016) These Polycomb-repressed domains harbour *genes* encoding key developmental transcription factors, whose misexpression can have detrimental consequences in differentiated cells.
PLoS Genetics (Raab, Resnick, & Magnuson, 2015) ARID1B and ARID2 participate in wide-spread cooperation to repress hundreds of *genes*.
Scientific Reports (Versluis et al., 2015) There has not yet been sufficient time for the corresponding resistance *genes* to spread into environmental reservoirs.
Nature Reviews Neurology (Hou, Friedrich, Gounot, & Schacherer, 2015) Parkinson Disease is generally considered a multifactorial disorder that arises owing to a combination of *genes* and environmental factors.
PLoS Genetics (Schumer, Cui, Rosenthal, & Andolfatto, 2015) Simulations reveal that hybrid populations rapidly and frequently become isolated from parental species by fixing combinations of *genes* that hinder successful reproduction with parental species.

In the first three lines the word "gene" refers to the "molecular gene" and in the last three to the "Mendelian gene."

Table 3 Various Usages of the Word "Gene" in Fixed Expressions

Where "Gene" Means "Mendelian Gene"	Where "Gene" Means "Molecular Gene"	Where "Gene" Can Mean Both
Defective gene	Foreign gene	Chimeric gene
Dominant gene	Gene cluster	Gene amplification
Gene conversion	Gene expression	Gene manipulation
Gene flow	Gene family	Gene mapping
Gene frequency	Gene network	Gene sequencing
Gene pool	Gene number	Lateral gene transfer
Mutant gene	Gene polymorphism	Pleiotropic gene
Recessive gene		Resistance gene
Selfish gene	Reporter gene	
Susceptibility gene		

Please note that in molecular biology, what biologists mean by a "resistance gene" is a transcriptional unit whose mutation can cause a gain in resistance, in which case the word "gene" corresponds here to the "molecular gene."

on the Cambridge University Science Forum "The Naked Scientist," which denotes a situation often encountered by some of us during scientific discussions between molecular biologists and population geneticists. On the forum, someone wondered: "if as a human I share 98% of my genes with a chimpanzee and 60% of my genes with a banana, how come I only share 50% of my genes with my own daughter?" (http://www.thenakedscientists.com/HTML/questions/question/919/). The paradox occurs here because the first two instances of the term "gene" are used in the molecular sense whereas the last one is the "Mendelian gene." Inconsistencies and flawed reasoning can also occur in more specialized writings. For example, science writer David Dobbs wrote that "For a century, the primary account of evolution has emphasized the gene's role as architect: a gene (or gene variant) creates a trait that either proves advantageous or not, and is thus selected for, changing a species for the better, or not. […] But a number of biologists argue that we need to replace this gene-centric view with one that more heavily emphasizes the role of gene expression—that we need to see the gene less as an architect and more as a member of a collaborative remodeling and maintenance crew." (https://aeon.co/essays/the-selfish-gene-is-a-great-meme-too-bad-it-s-so-wrong). Here the "molecular gene" concept (gene expression) is mistakenly used within the explanatory framework featuring the "Mendelian gene" (the gene is "selected for"), and the gene is inaccurately seen as an entity which can produce a phenotype alone (Keller, 2010). As Steven Pinker blatantly put it: "Part of the blame goes

to molecular biologists, who hijacked the term "gene" for protein-coding sequences, confusing everyone." (https://richarddawkins.net/2013/12/adversarial-journalism-and-the-selfish-gene/).

The confusion between the two concepts is easily noticed in scientific publications and database resources. For example, the population genetics concept of gene flow, that is, "movement of genes among populations due to dispersal processes" (Petit & Excoffier, 2009) implies that the gene here is the "Mendelian gene" since this is what is transmitted from parent to offspring and therefore from one population to another. The "molecular gene" does not flow between populations, but its various copies/alleles can. If gene flow between populations of mosquitoes was to be observed, it would be the dynamics of the presence/absence of the actual sequences (each of them being a specific allele) which would be characterized. If by "gene" one means the "molecular gene," then the term "gene flow" should be replaced by "allele flow."

In model organisms' databases confusion also exists. Consider the *Drosophila melanogaster* gene *white*. On the database Flybase (http://flybase.org/reports/FBgn0039044.html), we can find, among many other features, the sequence of *white*, its position, molecular functions, biological role, and homology with genes in other species. What is meant by the "sequence" of *white* is the sequence of the wild-type (or reference) allele of the gene *white*. On the other hand, the molecular function and the biological role correspond indeed to the "molecular gene" *white*: they were characterized from the analysis of multiple alleles (some of which resulting from mutagenesis of the reference allele) and biochemical activity of different White proteins, all encoded at the *white* locus. When referring to the *white* gene (or any other gene) within the molecular framework, one pictures the "wild-type" sequence (and now, the database entry regarding this gene). Much like species before population thinking (Mayr, 1975), in the strict taxonomical sense, the "molecular gene" appears under the image of a type, or wild-type, sequence deposited into a database with essential properties (or functions). The corresponding alternative versions (alleles) are thought as variations from that reference sequence which share the same essential properties (locus, function, homology). As pointed out by multiple authors regarding the species (Hull, 1965; Sober, 1980), this fits an essentialist, and very Aristotelian picture of natural kinds, which are first envisioned as ideal types narrowly defined. In contrast, the "Mendelian gene" is defined based on an observed variation in phenotype and genotype, thus through a nontypological approach (also called variation approach, or population approach, to

refer to Mayr's dichotomy). Vagueness of definition seems much more tolerated for the "Mendelian gene," which can correspond to any piece of chromosome transmitted from parents to offspring, generally associated with a phenotype.

Another famous example is Dawkins' (1976) "selfish genes." There is no competition in Dawkin's sense between different "molecular genes" within an organism. Indeed, the *white* gene does not compete against the *p53* gene for survival in populations of *D. melanogaster*. It is the different alleles of a "molecular gene" that may compete against each other. Multiple authors have therefore switched to use the selfish allele terminology (Sterelny & Kitcher, 1988). In Dawkins' own words, "when two genes, like the brown eye and the blue eye gene, are rivals for the same slot on a chromosome, they are called alleles of each other." If talking about the "Mendelian gene" then the "selfish gene" terminology is correct. Because each diploid individual has two Mendelian genes at a given locus, competition will occur between them if they are different (meaning there are in different allelic states) and competition will not occur if they are the same. When saying that "one human being inherits 50% of her genes from her father and 50% of her genes from her mother," one is implying that each parental copy should be considered as one "Mendelian gene," even though the maternal copy and the paternal copy might in some cases correspond to the same allele.

In general, evolutionary biologists mean "Mendelian genes" when they speak about "genes," whereas molecular, cell, and developmental biologists mean "molecular genes." The concepts of pleiotropy and epistasis are particularly revealing in this respect. In broad terms, both fields consider that epistasis occurs when the effect of one *gene* on a phenotype is dependent on the presence of another *gene* (Cordell, 2002; Phillips, 2008) and that pleiotropy occurs when one *gene* affects two or more seemingly unrelated phenotypic traits (Paaby & Rockman, 2013; Stern, 2000, 2010). However, in this definition of pleiotropy and epistasis, the term "gene" is used either as the "Mendelian gene" or as the "molecular gene," and this produces radically different concepts. For example, when biochemical geneticists say that the *cid1* gene is epistatic to the *snf1* gene in the yeast *Saccharomyces cerevisiae* (Avery & Wasserman, 1992), what they mean is that first, loss-of-function mutations in these two genes produce distinct phenotypes, and second, the phenotype of the *cid1 snf1* double mutant is similar to the phenotype of the *cid1* gene. In contrast, in population genetics alleles can display epistatic relationships even though they do not correspond to null alleles that fully remove gene activity (Cordell, 2002; Phillips, 2008). To avoid confusion,

one has to be aware that multiple definitions of epistasis and pleiotropy are currently used and that it is important to pay attention to the context to understand what is meant in each case.

Because biology research fields are relatively well-defined and separated, the problem of using the same word for two different meanings does not always arise. However, in certain research areas, the problem is present and acute. In genome-wide association studies, analyses are mostly performed on "Mendelian genes" (Table 2), but results are often interpreted in terms of "molecular genes," with transcriptional units forming an essential part of the concluding explanatory statement that relates the phenotype to the genotype. The problem also occurs in evolutionary biology, especially in evolutionary genetics and eco-evo-devo, which aims to uncover the rules that underlie the interactions between an organism's environment, genes, and development and to incorporate this knowledge into the theory of evolution (Abouheif et al., 2014; Carroll, 2005). Because these fields have a tradition of coupling population genetics, molecular genetics, and developmental biology into one experimental framework, the term "gene" is used to denote either the "Mendelian gene" or the "molecular gene" depending on the context. For example, BMP4 is a "molecular gene" involved in beak shape differences between Darwin's finches species, in the sense that differences in BMP4 expression levels during beak development have been associated with distinct bill shapes, but BMP4 has not been shown to be a "Mendelian gene" involved in beak shape evolution, in the sense that the causing genetic locus and the causing mutation(s) have not been identified (Abzhanov, Protas, Grant, Grant, & Tabin, 2004). It is entirely possible that the change in BMP4 expression levels that is thought to have occurred during beak shape evolution was actually caused by a mutation in another "molecular gene" acting upstream of BMP4. Confusion between both meanings of the term "gene" may also arise in other interdisciplinary fields of biology, such as human genetics. In this chapter, we try to clarify the distinction and the relationship between the "Mendelian gene" and the "molecular gene."

4. HOW MANY GENES, ALLELES, AND LOCI WITHIN A GENOME?

According to most recent estimates, humans are now thought to carry approximately 19,000 genes in their genome (Ezkurdia et al., 2014). In such a statement, genome refers to the nuclear genome and genes to "molecular

genes," or protein-coding sequences. Let us consider one human being. Although his father and his mother gave him 19,000 genes each, we would agree that he has 19,000 genes and not 38,000. If by "gene" we mean the "Mendelian gene," then it is difficult and probably impossible to estimate the number of genes within a human genome, as there is no correlation between the number of "molecular genes" and the number of "Mendelian genes." If by "Mendelian gene" one means any DNA sequence difference, then the number of "Mendelian genes" within a genome is huge and correlated to the level of nucleotide polymorphism within the population. If one means any change in a chromosome region which is associated with a phenotypic change, then the estimation of the number of "Mendelian genes" is extremely difficult, in particular because of the immensity of the phenotype space (Houle, 2010), of G × G interactions and of the various environmental conditions that can affect phenotypes through G × E interactions. If we take one of those "Mendelian genes" and identify it as the one inherited from the father, then there is an equivalent copy which is inherited by the mother. Now, under this view, a diploid organism has in general, at each locus, two "Mendelian genes" which can be identical (homozygous genotype) or different (heterozygous), corresponding to one "molecular gene." To avoid confusion, the total number of genes is often given for the nuclear *haploid* genome.

Compared to the notion of gene, the concept of allele may, at first thought, seem more clearly defined, but it is not certain. According to certain biologists, a diploid homozygous individual carries *one* allele (and thus two copies of the same allele) whereas others affirm that a diploid homozygous individual has *two* alleles (which are identical). A key question which highlights the confusion is "what makes us diploid: the number of genes or the number of alleles?" One possibility is to reply that there are two Mendelian genes and only one allele (considering that an allele represents one version of a gene), and this fits the Mendelian definition. An alternative, close to the molecular view, is to say that there is one "molecular gene" and two copies of the same sequence, that is, two alleles which are identical. At a given locus, the number of molecular alleles is thus equal or higher than the number of Mendelian alleles. In summary, the Mendelian allele refers to an allelic version whereas the molecular allele refers to one of the copies (which can be identical). These two discordant views are found in various biology textbooks (Table 1), showing that there is no consensus.

Similarly, the term "locus" is loosely defined (Table 1). The word "locus" refers to a genomic or genetic position. A locus can be part of a "molecular

gene" or can correspond to several. As stated by the *Rules and Guidelines from the International Committee on Standardized Genetic Nomenclature for Mice*: "A locus is a point in the genome, identified by a marker, which can be mapped by some means. It does not necessarily correspond to a gene; it could, for example, be an anonymous noncoding DNA segment or a cytogenetic feature. A single gene may have several loci within it (each defined by different markers) and these markers may be separated in genetic or physical mapping experiments. In such cases, it is useful to define these different loci, but normally the gene name should be used to designate the gene itself, as this usually will convey the most information." (http://www.informatics.jax.org/mgihome/nomen/gene.shtml). Examination of the concept of "quantitative trait locus" (QTL) also reveals that a locus can encompass several "Mendelian genes." A QTL is a section of chromosome (the locus) that correlates with variation in a quantitative phenotype (Falconer & Mackay, 1996). In cases where one large-effect QTL is later found to be made of several closely linked QTL with smaller effects (McGregor et al., 2007; Orgogozo, Broman, & Stern, 2006), the original locus is found to be made of several "Mendelian genes." In its smallest size, a locus represents one nucleotide position within a genome and in its largest it can be an entire chromosome.

5. "GENES" AS CAUSAL AGENTS OF PHENOTYPES

The "Mendelian gene" and the "molecular gene" concepts are each part of two distinct frameworks for explaining the causes of phenotypes. The "Mendelian gene" explains phenotypic differences between individuals that can interbreed (members of a given population, parents, offspring, etc.) whereas the "molecular gene" explains the existence of a particular phenotype in a given individual (if the gene were to be absent then the phenotype in question would not be as such). Both concepts are part of a causal–mechanistic explanation of the living world (Salmon, 1994, 1997), as opposed to other types of explanations such as the Hempel–Oppenheim deductive-nomological model (Hempel & Oppenheim, 1948). At least two types of causal–mechanistic explanations can be distinguished, the "constitutive" one, which describes the temporal series of successive mechanisms that generate the phenomenon, and the "etiological" one, which identifies factors whose changes modify the phenomenon that needs to be explained (Waters, 2007; Woodward, 2005). In both cases, causes represent pertinent elements that account for the building up of the phenomenon to

be explained. The "molecular gene" is rather involved in a constitutive explanation and the "Mendelian gene" in an etiological explanation. The "Mendelian gene" concept is often used in a framework which does not allow the reconstitution of the entire chain of causal operations linking the genetic level to the phenotypic level. In contrast, the "molecular gene" is part of a continuous series of explanatory processes: the gene is transcribed into mRNA molecules, which are then translated into proteins, and the accumulation of proteins leads to such-and-such effects at the level of the cell and consequently at the level of the organism. Even though certain authors pointed out that current explanations on how molecular genes play a role in elaborating phenotypes are still not as extensive and constitutive as they could effectively be (for example, the effects of cytoplasmic water, gravity, etc., are generally not taken into account) (Gilbert & Epel, 2009; Keller, 2010; Lewontin, 2001; Oyama, 2000), explanations of phenotypic traits involving "molecular genes" are generally more constitutive than those involving "Mendelian genes." Both concepts are important and bring significant insights in their respective fields of research. The "molecular gene" connects better to molecular and cellular processes than the "Mendelian gene," while the "Mendelian gene" connects more directly to the phenotype at the level of the organism than the "molecular gene."

6. SEARCHING FOR THE CONCRETE OBJECTS REPRESENTED BY "GENES"

For any type of concept, the human mind has a tendency to try to make it correspond to a concrete object, that is, an object which can be isolated in time and space by our sensory system. Yet a concept does not necessarily represent such a concrete entity (Cassirer, 1910). For example, the concept of natural selection (Darwin, 1859; Lewontin, 1970) is fully relevant for our understanding of the living world even though it does not represent a concrete object. The concept of "gene" is particularly interesting in this respect. Even though the notion of "gene" was primarily apprehended as an abstract entity that explains the origin of visible characteristics observed in living organisms and how such phenotypic traits are passed from parents to child, biologists have, since the presence of this word in the scientific literature, struggled to find the physical molecular object embodied by the concept of "gene." Today, both the "Mendelian gene" and the "molecular gene" concepts are extremely used and useful to understand the origin of phenotypic traits, in their respective explanatory frameworks, yet biologists