Guido Barbujani
A Genetics Primer for Beginners

SILLABARIO DI GENETICA PER PRINCIPIANTI

Ten lessons by one of the most important Italian geneticists to give us the bases of genetics and the tools to understand science involving all of us.

Our genome is a text: a book from the past and a handbook for the future. We have been able for some years to know what's written in the genome, or better we know the letters (the genes) but are quite unsure of their syntax (how each gene interacts with the others and reacts to what comes from the outside). That's why we know if a child will have cystic fibrosis, as it depends on just one gene, but we do not know if he or she will develop Parkinson Syndrome or cancer, as they are complex diseases and depend on thousands of genes. This is the most actual challenge geneticists have to face nowadays.

Guido Barbujani worked at the State University of New York and at the London, Padua and Bologna universities. Now he teaches Genetics at the Ferrara University. He is also a novelist and an essayist. Bompiani also published *Tutto il resto è provvisorio*, his latest novel, and *L'invenzione delle razze*, his bestselling essay on how human races have no scientific base.

Rights for *L'invenzione delle razze* sold:
World English (MIT Press).

BOMPIANI RIGHTS LIST | FICTION

Pages: 272
Publication: September 2019
Price: € 18,00

GUIDO BARBUJANI
A GENETICS PRIMER FOR BEGINNERS

translated from the Italian
by Sarah Victoria Barberis

SAMPLE COPY

BOMPIANI

MESSAGES FROM THE PAST

Each of us carries within us a message from the past. It's stored in the DNA of our every cell. Half comes from our mother and half from our father, and their DNA in turn is a mixture of their parent's DNA, i.e. our four grandparents. But those grandparents had grandparents and those their own grandparents, so on and so forth. The number of our forefathers doubles in each generation; you may be familiar with the old story of the chess inventor who asks, as a means of payment, one grain of rice for the first square, two for the second, four for the third and so on, and you may have already understood where this is going. Going back in time, our genealogies spread until they embrace an endless number of ancestors: ten generations ago, in Bach's time, those ancestors were a thousand, and each one of them had a thousand ancestors 300 years before; so we descend from one million ancestors who lived in a time during which the Brunelleschi discovered the rules of perspective and from a

billion ancestors from the time of the first crusade, and from a thousand billions at the time when Charles the Great was crowned and the Tang dynasty was ruling China (we will get back to these huge figures later). Each of those ancestors, one way or another, sent a message that, through time, has reached us safely kept in the DNA. That message from the past is an instruction booklet. Upon it, in the language of the DNA - a language that we have only partially deciphered - there are written instructions which have allowed the fertilised egg cell we come from to multiply in an orderly manner, until forming a complex organism, made up of 37 trillion of cells, that is us.

This very DNA allows us to live. We call a gene any trait of DNA that performs a specific function: therefore each gene is, in a certain sense, an instruction. It serves to make one or more proteins, or other molecules (of RNA, those too necessary to fabricate proteins). And other proteins, themselves codified, i.e. written in a gene, read those instructions.

Alphabet, lexicon, syntax of DNA

Discovering what's written in the genome, i.e. the overall of our DNA (or that of viruses, or of a platypus) has been a lengthy and tiring journey, but we're on our way; for a few years now, reading a genome is not only technically possible, but it can be done with little expense and on a large scale. "Writing" and "reading" are two metaphors, obviously, but of noble origin (Gregor Mendel, the father of genetics himself, started indicating the genes with the letters of the alphabet). Plus, they give quite a clear idea of it. The genome, which varies from individual to individual, but is identical in every cell of

an individual, is indeed a text: a lengthy one. We perfectly understand its alphabet, the four molecules (they are called adenine, cytosine, guanine, and thymine, indicated with the letters A, C, G, and T), placed in a row, one after the other, forming long chains, the chromosomes. We also understand its vocabulary quite well, i.e., the meaning of the elements that make it up, taken one at a time: the genes, and together with the genes the regions of DNA necessary to activate them or turn them off. We are, however, still far from understanding the syntax of this text, namely the way in which each of our 20,000 genes responds to the functioning of the other genes and to the messages from the environment. That's the reason why today we can predict (not in all cases, but in many) if an infant will have cystic fibrosis or muscular dystrophy. These are interesting diseases, and thanks to genetics, a lot has been done to prevent them. However one usually doesn't die from these diseases. The most serious and widespread diseases are diabetes, cancer, cardiovascular diseases and neurodegenerative disorders. The genes that make each of us more or less predisposed to develop them throughout life are numerous, and so are the environmental factors that could influence the process. In that instance our ability to understand if and when the disease will occur is still limited, not to mention the most interesting and complicated area of study among all: understanding cognitive functions, the brain, the intelligence. It suffices to control only one gene in order to ascertain whether a foetus or an infant will have cystic fibrosis, and we know all too well where to look for that gene; but in order to understand if we'll have diabetes, high blood pressure or Parkinson's, not to mention whether we'll possess any math

skills, we'd have to check out dozens or hundreds of genes, mostly unmapped. And even if we knew them all, we may not be able to predict much, because, as we'll see later, we only have a very vague idea on how a gene is influenced by others and by the thousand external factors that, in the absence of better terms, we call the environment.

Diabetes, cancer, Parkinson's, these are complex diseases. For medical researchers the challenge today is to navigate this formidable complexity. In the last twenty years the progress has been great in terms of general biological knowledge: we've figured out how cells are born and die; we have realised that their programmed death is a crucial biological process; we have understood a lot about the exchange of messages between two cells, and between molecules within the cell. Yet the progress in practical applications - namely treatments - has been, so far, limited.

However the outcome is not unsatisfactory. Reading the messages from the past contained in the DNA is already providing us with a basic understanding that is nevertheless indispensable in order to prevent and treat certain diseases, and that allows us to answer questions considered science-fiction up until a few years ago, putting back together pieces of our past that neither historical sources nor archaeological findings could ever reveal; it has allowed us to genetically improve crops and retrace the manipulations which, starting 10,000 years ago, brought humanity to create genetically modified organisms: practically all the plants and animals from which we derive food and textile fibres today.

Genetics and evolution

That understanding genetics is necessary in order to understand evolution, today may sound so obvious that there's no need to explain it. But it wasn't always like this: the two disciplines have developed independently of each other and, for some decades, even in controversy with each other. On October 2, 1836, when Charles Darwin lands in Cornwall after his trip around the world on the

Beagle, he is 27 years old. He still has 46 years to live on, and he'll let another 23 years pass before printing his basic text *On the Origin of Species*, or more precisely, *On the Origin of Species by means of Natural Selection*: a book that, a century and a half afterwards, remains the backbone of modern biology. During the five years of his voyage at sea around the world, and in the following years, barricaded in his house of Down, in Kent, Darwin understands very, very much about the relationships between different living forms and their origin: but not everything. The word *evolution* has not yet come up, but Darwin understands that, over time, new species evolve from the common ancestors, which acquire new bodily functions and new specialisations, adapting to the environment. The species are changeable: the species, he said, are transforming, and the environment drives their transformation.

But his brilliant reasoning lacked two elements, far from secondary. Even Lamarck, before him, had realised that different species descend from common ancestors, and Lamarck, too, had realised that it had something to do with the environment: the organs that perform well in a certain environment, are established and spread throughout. Lamarck, however, thought it was the environment itself generating differences among individuals. It is called inheritance of the acquired characters,

and the idea is that, by raising many weights, our powerful biceps would be passed on to our children, and their children, and so on. It doesn't work that way, as revealed by the famous example of the giraffes. The effort to reach the highest leaves of the trees, it was said, would stretch the vertebrae of the neck: an acquired character, which would then be passed down through the generations, producing giraffes with longer and longer necks (A less known fact, however, is that Lamarck took this example from Darwin: not Charles, but his grandfather Erasmus).

Grandfather liked the inheritance of the acquired characters, much less so his nephew. As a cultured and cautious scientist, Charles Darwin was willing to admit that, perhaps, some acquired characters could be passed on hereditarily: but it would still be an exception, not the rule.

The rule was, according to him and also according to us, that the individual differences, what we now call biodiversity, exist before and on their own, and they are not created by the environment; but in the mid-nineteenth no one could tell from where these differences sprung (today we know: from DNA mutations), nor how they were transmitted through generations (today we know, Mendel explained it well). These were the two missing elements in Darwin's reasoning. Of a fact Darwin was firmly convinced: the role of the environment is not that of creating biodiversity, but to select, within an existing biodiversity, the most suitable forms of life for survival and reproduction. It's called natural selection.

So, Darwin was in a difficult position: he refuted the inheritance of the acquired characters, while being unable to explain neither the origin of biological differences nor how they

were inherited. There was someone who could have given him a hand: in the same years, in a Moravian monastery,

by way of crossing many pea plants, Gregor Mendel was in fact discovering the fundamental laws of inheritance. But nobody knew Mendel and he also wasn't really keen on self-promotion. In 1865, i.e. six years after the publication of the *Origin of Species*, Mendel presents to the Natural Sciences Brno Society a speech with a very unappealing title, *Experiments on plant hybridisation*. He puts it in writing, prints 40 copies of it and sends it to as many colleagues; of 11 copies we know the recipients, of the other 29 we do not. According to a well-established legend, one of these copies lands on Darwin's table, and there it remains until his death, untouched.

Who knows if that's true. It's plausible: Darwin was one of the most famous scientists of his time, it would have made sense to ask for his feedback. And then Mendel owned a copy of the *On the Origin of Species*, diligently noted. Assuming that Mendel had tried to inform Darwin of his results, it mustn't be assumed that the latter would have found them interesting: it will take another forty years before the importance of Mendel's work finds its due recognition. At the time, it was not at all clear what Mendel had discovered: if a set of general inheritance laws, valid for all organisms, or just some bizarre characteristic of the pea plant. And in any event, his numerical processing of data would have possibly annoyed Darwin, who did not hold mathematics in good opinion.

In a story by Jorge Luis Borges, Averroes, tired and irritated, doesn't realise that a traveler reveals to him the meaning of two words, comedy and tragedy, whose meaning escapes him in an

Aristotle's text that he is trying to analyse and whose translation he will fail to grasp. Similarly, perhaps, the solution to one of the problems of Darwin's theory passed before his very nose, without him even realising it. That a copy of Mendel's article had landed on his desk may or may not be to be true; we know however that he owned a copy of Hermann Hoffmann's book on plant hybrids, and there are his handwritten notes on pages 50, 51, 53 and 54: however not on page 52, the one where Hoffmann summarised Mendel's work.

And so it took almost a century to merge Darwin's evolution theory and Mendel's inheritance, and more specifically to clear up the misunderstandings that prevented us from understanding how Mendel's legacy, with its clearly distinct characters, with its seeds that are either yellow or green, smooth or wrinkled, could explain how continuous variability could evolve: bird's beaks more or less pointed, deer's horns more or less large, those elements that are so dear to evolutionists.

Compare many texts

Nowadays we try to answer the question "How have we evolved?" (and by "we" I mean us living creatures) by reading the genomes and speculating about their differences. The genome is an immense text: six billion of characters and counting in humans, slightly more or slightly less in other mammals. To give you an idea, *The Bethroted* (Alessandro Manzoni novel) is about one million characters: it means that each of our cells contains an instruction manual equivalent to more than 6 thousand copies of *The Betrothed*. How the cell finds in a few moments the right

page is yet unknown. But in the meantime we understand that all living creatures descend from a common ancestor, who lived just under 4 billion years ago, because in all living beings the rules by which the information contained in the DNA passes to the RNA, and from there to proteins, are basically identical.

So, we share something in common with cyclamen and the flu virus; but, forgive the platitude, we are also separated by several differences. It is precisely by arguing about these differences, reading and interpreting the message from the past contained in our cells in the light of evolutionary theories, that we're recomposing the picture of life on earth, a vast and much surprising one in many instances. Current research uses very complicated laboratory technologies and sophisticated statistical analysis; however the general principles of genetics and evolution are simple, so we'll try to understand them, slowly and with gusto.

Nothing better for aficionados than to start by reading *On the Origin of Species*. I know, it's a thick book, over 450 pages in the unabridged edition. To keep it simple, you may just start by reading the index; and to make even less of an effort you can start by reading the titles of the first six chapters. One, "Variation under domestication"; two, "Variation under nature": Darwin argues on how breeders, by crossing horses, dogs and pigeons, have selected different varieties, and proposes that very similar selection phenomena may have occurred in nature to all creatures, including non-domestic creatures. Three, "Struggle for Existence"; four, "Natural Selection"; five, "Laws of variation": Darwin describes the processes that can cause beings to evolve. However he sees all too well that many

pieces of reasoning elude him, especially because he is unable to explain the mechanism that generates the variation (Once again, it's the process of mutation; today we know, but not at the time). And so here we are on chapter six, "Difficulties on theory": a severe examination of all the problematic or unresolved aspects of his theory, with a list of potential responses along with an analysis of their consequences. Good scientists are also the most critical of themselves, mainly because they know that if there are weak points in their reasoning, sooner or later someone will spot them. Thanks to his exceptional capacity for self-criticism, Darwin gives his theory the elasticity it needs to incorporate, for decades, new data, new knowledge, without the need to modify its basic structure. Modern biology, of which genetics is part, is not limited to Darwin's thought, but it still is absolutely Darwinian. As Theodosius Dobzhansky said in a famous aphorism, "Nothing in biology makes sense except in the light of evolution".

Neanderthal and Denisova

The genomes of Neanderthal and Denisova raise quite a lot of interesting questions. First and foremost: Neanderthals are closer to Europeans, Asians and Melanesians than they are to Africans. Just slightly closer - from 2% to 4% - but consistently. What could that mean? There are two possible explanations. The first is that when they came out of Africa our ancestors have mixed (technically: hybridised) with the Neanderthals. It resulted in an unbalanced genetic cocktail, with a 96-98% of the genome coming from Africa; but since outside Africa we all descend from that hybrid population, we all have in our cells a

tiny bit of Neanderthal legacy. The second explanation is best understood by an example. We look more like gorillas than kangaroos, not because we recently hybridised with gorillas, but because we have evolved together with them longer than we have with kangaroos. In the same way, it could be that Neanderthals (whose ancestors have moved out, perhaps one million years ago, of North Africa) have always had more DNA in common with those of us that 100,000 years ago came out of Africa (i.e. North Africa), than with the average Africans (the latter including also people from the West, East and South). Svante Pääbo and other excellent scientists rely on the first explanation, I have some doubts and Andrea Manica and Bill Amos, two geneticists working in Cambridge, even more so. It is demonstrated that, at least once, we have crossed paths with Neanderthal: an incomplete fossil, dated to 37,800 years ago, from the cave of Peştera cu Oase, in Romania, contains 6-9% of Neanderthal DNA. Do some math and it just means that this person had a Neanderthal great-great-grandfather. So the intersection has occurred: but that *all* Europeans and Asians and Papuans descend from similar crossbreeds is another matter. Isn't it a tad strange that in the European genomes the percentage of DNA similar to the Neanderthal is inferior to that of Asians and even of the Maoris of New Zealand, where the Neanderthal never even dreamed of going? People migrate, as we know, and carry with them the DNA of the ancestors, but we have dwelled in Europe for thousands of years in close contact with the Neanderthals, much less in Asia; how come we have stopped hybridising in Europe? We did not like them anymore, after the first try? There's another oddity, the mitochondrial DNA: on several thousand modern individuals

studied so far, no one displays a Neanderthal mitochondrial. The possible explanation was that the hybridisation could have been asymmetrical: females like us, anatomically modern, with Neanderthal males. However it seems at least unusual that the group invading a territory offers its women to the invaded populations, rather than the reverse. Eventually discovering that we don't even possess a single Neanderthal Y chromosome, the one fathers transmit to sons, increased the problem. In short, in my opinion, a certain amount of doubt regarding the hybridisation with Neanderthal is justified. There are few doubts, however, that Denisova may have left a contribution, modest but discernible, in the people of Southeast Asia. In one way or another, different human forms have met; they then have crossed, even if the effects of these crossings are yet to be discussed; and in the end only one remained, our own; some of us, or maybe many of us, carry in the cells some trace of the ancient hybridisations.

As for the theme mentioned earlier, i.e. if we and Neanderthal belong to the same species, given that we ran into each other, I'd say we'd do well to follow Charles Darwin's advice and don't give too much importance to names. Linnaeus has given names to species, and through these names we orient ourselves in the great world of the living. But Linnaeus did not know evolution; he thought that species existed, immutable, from the day of creation, so it was just a matter of putting the right label on each one. So, in elementary school they taught us that the horse and the donkey are in two different species because their hybrid, the mule, is sterile: right. But Lamarck already has clarified that different species derive, with modifications, from

common ancestors: at some point in the past donkey and horse were the same thing, later they became two different things that could make fertile hybrids, and (now) they can still cross but their hybrid is sterile. The formation of different species requires several thousands of years; the species is a historical entity, which in a certain moment is well defined (today no one can confuse, in good faith, horses and donkeys, or humans and chimpanzees), but a few millennia earlier may not be.

We and Neanderthal were obviously capable to cross, and the hybrids were not sterile, if in Peştera cu Oase we found one of their descendants. But fossil experts distinguish with certainty, unambiguously, our own skulls and those of Neanderthal. Therefore, it's wiser to continue calling us with different names, because we were different, and to admit that the concept of species still doesn't help us understanding what happened between them and us.

Environmental DNA

The DNA leaves even more surprising traces. While working in the Denisova cave, the members of Svante Pääbo's team decide to collect some sediment, i.e. the dust left by the crumbling of the rock, and take a look at it. They are surprised when they find that in that tiny debris they could still find some recognisable pieces of mitochondrial DNA. Where that DNA comes from is unclear: maybe from the stools, maybe from the rotting bodies. That DNA documents that both Neanderthals and Denisovans dwelled in that cave, as well as several animals: mammoths, rhinos, hyenas, bears, horses. The surprise is not so much in the fact that DNA can be found in absence of fossils. DNA

leaves traces wherever we go, and any police force is able to detect them, with very sophisticated techniques. We talk about environmental DNA, it allows you to demonstrate whether or not a certain person has been in the passenger seat of a car. The surprising aspect is that these traces were considered to be very labile: instead you can apply the same study technique to thousands years old samples and find out something unexpected. The limit of environmental DNA studies lies in the fact that the DNA doesn't carry a date with it. If it's Neanderthal's, no one may have left it there recently; but if it's a modern DNA, we can't know if it's been there for millennia or just for months.

The point is that today some layers of the excavations of the Denisova cave, corresponding to certain dates measured with some level of precision, reveal that the cave was occupied by Neanderthals, while other results say that it was in fact occupied by Denisovans.

Who are the Americans

One of the most beautiful results of ancient DNA studies is the analysis of the remains of a boy who died 24,000 years ago in Mal'ta, Siberia. It belonged to a population accustomed to harsh environments, where even today the temperatures never exceed zero from October to April (and back then it was worse); they lived in underground houses, reinforced with mammoth bones, presumably covered in animal skins. If we compare it with the genomes of contemporary individuals, the DNA of the boy from Mal'ta displays two special features: it very much resembles that of all contemporary Amerindian populations and little to that of the current Asian populations, including that from Siberia. It

means that the genomes of the indigenous Americans are all alike, and derived, if not from the boy of Mal'ta, from a population that is genetically close to his. We can't say if this means that the Americas have been colonised by a single major migratory wave or more, although derived from genetically undistinguishable populations. The second thing that the Mal'ta genome tells us is that either all the members of its population have moved to the Americas or that those who have remained are extinct. The current Siberian, and generally Asian populations, have little in common with the people who were in Siberia 24,000 years ago. In the next paragraph we will see that the succession in the same area of different populations, overlapping, mixing or not mixing, constitutes the rule, not the exception, for mankind. Our species has always been very mobile, and the studies of ancient DNA show that no one remained alone too long in the same place.

Who are the Europeans

Because the archaeological excavations are abundant and because the populations to be studied are accessible, we now know a lot about the DNA of Europeans. The ancient DNA has given depth to these studies: it has allowed us to understand not only how it is made up, but also in what moment the cocktail of genes that we find on our continent today was formed. Already half a century ago the studies of a great Italian geneticist, Luca Cavalli-Sforza, had shown that the European population has profoundly changed in the Neolithic Age. With the advent of agricolture, 10,000 years ago, the population of the Near East begins to have more food and grow more rapidly.

Gradually, this population expands more and more towards north and west and it reaches the Iberian Peninsula at the rate of one kilometre per year; it spreads new food production techniques and their own genes, throughout Europe. The presence of allelic frequency gradients, from the Near East to the Iberian peninsula, shows that agriculture spread through Europe not by means of cultural contacts and exchanges (which wouldn't have left traces in the genome of Europeans) but through the bodies of migrant populations that overlapped previous populations by occupying their territory, more or less as in Siberia.

More or less as it happened in Siberia but not quite the same. Over there, as far as we can see, there's no trace of the inhabitants from 24,000 years ago; instead, in Europe, these first inhabitants, the Paleolithic hunters from Africa, left a few traces in modern genomes. We can identify that genome and distinguish it from that of other migrants. In essence, by comparing the DNA of old Europeans with that of modern Europeans, we distinguish the contributions of three major waves of migration, though all ultimately from Africa. The first is Paleolithic, the second is Neolithic, and a third one in the Bronze Age comes from the steppes of present-day Ukraine. But beware: don't be tempted to think that some of us are Paleolithic, others are Neolithic and other are descended from those of the Bronze Age. In the cells of every European we find, mixed-up, all these contributions (and others, which are smaller and more difficult to identify), in variable proportions: in southern Europe, the Neolithic one prevails, in the North the Bronze Age, while the traces of the first Palaeolithic hunters are scarce everywhere, except in the

Iberian Peninsula; everyone has its own cocktail of genes, and therefore of ancestors. The study of ancient European genomes brilliantly confirms some of the things that we had apparently figured out, for example that the ability to digest lactose (we talked about it in chapter 3) appears in Europe when agriculture and dairy farming spread throughout.

Before then, if even some possessed the mutations that would allow them to use milk, there was no advantage in it, because there was no milk; since then onwards, being able to feed on milk and dairy products provided a richer diet, and the mutations that confer the ability to digest have spread, according to Darwin's mechanism as described in chapter 6. But the study of the old European genomes reserves more surprises. In 2018, a discovery was widely reported in the media according to which until 7 or 10 thousand years ago, Europeans (in England, Switzerland, Luxembourg and the Iberian Peninsula) still had dark skin, the one they had when they left Africa. Even though this pigmentation constituted a disadvantage in the European climate, the alleles necessary to enlighten it, simply hadn't arrived. Fair skin alleles appeared by mutation 12,000 years ago, however not in Europe, but in the Caucasus, as far as we know. They then came to Europe with the migration of Neolithic farmers, who in turn have inherited them from someone who came from the Caucasus. Just like what happened with the ability to digest lactose, a combination of migration and natural selection allowed the genes for fair skin to spread throughout Europe, although relatively late. The fact that Europeans are white-skinned has been true only for a few millennia.