The Joy of Research
Mendel, Jones, and Human Origins[1]

William S-Y. Wang
City University of Hong Kong

Professor William S-Y. Wang grew up in Shanghai, and received his PhD from the University of Michigan in 1960. From 1965 to 1994, he was Professor of Linguistics at the University of California at Berkeley. He joined the City University of Hong Kong in 1995 as Chair Professor of Language Engineering. He is currently Adjunct Professor of several universities, including the Hong Kong University of Science and Technology.

He is the Founding Editor of the Journal of Chinese Linguistics [1973] and the Founding President of the International Association of Chinese Linguistics [1991]. The honors he has received include fellowships from the Fulbright Commission, Guggenheim Foundation, the Center for Advanced Studies at Stanford in California, and at Bellagio in Italy. He is an Academician of Academia Sinica.

Professor Wang's interest in language within an evolutionary perspective began in the late 1960s, when he explored parallels between biological change and linguistic change. He finds the recent convergence of research on human origins from several disciplines particularly exciting, and hopes the present essay will convey some sense of that excitement.

[1] Based on a lecture given at the symposium, Broadening Research Frontiers, sponsored by the School of Graduate Studies of the City University of Hong Kong, March 2001. I thank Professor Merritt Ruhlen of Stanford University for providing the language maps for Eurasiatic and Dene-Caucasian used here, and Dr. Lisa E. Husmann, Ms. Jinyun Ke, and Mr. Charles Chienjer Lin for many helpful discussions.

I am happy to have this opportunity to participate in these festivities of 'Broadening Research Frontiers'. Research lies at the heart of any university that takes its mission seriously; and broadening frontiers certainly enhances the quality of research. I hope to illustrate this point this afternoon. Furthermore, research and teaching are really two sides of the same intellectual coin — the *yin* and *yang* of knowledge. An inspired teacher is one who has directly investigated what he teaches, and a successful researcher is one who is eager to tell others what he has discovered.

In each, there is a joy for which there is no substitute. There is a pure pleasure to help others learn — to witness someone's eyes light up upon grasping a new concept or mastering a new method. And there is simply no substitute for the privilege and thrill that comes with discovery — to feel for a moment that you and only you have achieved this new insight!

The topic of research frontiers is an immense one, so I need to be very selective. Let me begin with the stories of two exceptional men. The first — a Czech monk in 19th century Brno; the second — an English judge in 18th century Calcutta. The modest streams of research they started in linguistics and in genetics have each grown into roaring intellectual rivers, and started to flow together toward the end of the 20th century. Their legacy is that we now have a much better idea of who we are, we the human race, and where we come from.

A hundred some years ago, an unassuming monk was growing peas in a monastery in Brno, in what is now the Czech Republic. He was curious about how these peas behaved from generation to generation. He noticed that among these plants, some stems were short while others were tall. The shape of the seeds was also different: some were smooth and round, whereas others were wrinkled. Furthermore, the color of the pods varied: some were green while others were yellow. So the monk grew many generations of these peas, sometimes keeping a strain pure and sometimes cross-breeding one strain with another. He grew many thousands of these peas, keeping count of how their traits were passed on from generation to generation.

The experiments gave interesting results, which the monk presented to a local society of scientists in 1865. He suggested that one trait can be dominant over another, such as the tall stem dominating over the short stem. The trait of stem length is inherited quite independently of other traits, such as the one for pod color. Such a relation he called the

'principle of independent assortment'. He offered several such principles to explain these facts, based on statistical analyses of the counts he made per generation of plants. The small audience applauded the monk politely, and shifted their attention to the 'hot' science of the day. Darwin's controversial *Origin of Species* was published just a few years earlier, and everyone was debating the theory of evolution.

Figure 1:
Double helix structure
of DNA.

Such were the humble beginnings of the science of genetics. The monk was, of course, Gregor Mendel. If there were some powerful, prescient intellect in that small audience, it might have recognized that Mendel was supplying precisely the mechanism that Darwin needed to make the theory of evolution tangible. In fact, Darwin himself was tripped up by genetics, and made several basic errors in his book. However, the time for realizing the importance of this work had not yet come, and the monk was essentially ignored by the local society. The report he published in 1866 lay dormant for thirty some years, gathering dust on library shelves. The time finally came in 1900, when, amazingly, Mendel's report was independently discovered by three scientists working in different parts of Europe — a Dutchman, a German, and an Austrian. Not long thereafter, genetic insights became a central part of Darwinian thinking, giving rise to the so-called 'synthetic theory of evolution.'

Over the past century, genetics has made great strides since growing peas in Brno, with some stunning discoveries, collecting a rich complement of Nobel prizes along the way. An early Nobel prize went to Thomas Morgan, who worked with *drosophila*, the common fruit fly. By working with the eye color of these flies, Morgan was able to discover genes linked to the determination of sex, the distinction between chromosomes which are XX for female and XY for male. Another Nobel prize went to the team of Francis Crick and James Watson, who unraveled the structure of the DNA molecule in the form of a double helix, giving us a deeper understanding of how the four letters of the genetic code, A and T, C and G, are arranged. As the artist's illustration in Figure 1 shows[2], this structure is not only

[2] Dr Michael Yang of the City University of Hong Kong kindly provided the illustration of the double helix used here.

biologically efficient but also aesthetically very pleasing. Recently, a Nobel prize went to Kary Mullis, for inventing PCR, a method of chain reaction that can produce millions of copies of DNA in a test tube. Such a technology is essential for the work for decoding the billions of base pairs in the human genome — the DNA instructions for manufacturing the human body.

By the beginning of the 21st century, 140 years after Mendel, the entire sequence of the three billion some letters of human genome has been worked out. Knowing the sequence of the four letters is a far cry from understanding what they mean, just as recognizing the letters of the Greek alphabet is a far cry from being able to read a book written in Greek. Knowing the alphabet is only the beginning of learning the language — the language of the genes. We are now able to recognize a few of the 30,000 some genes contained in the human body. Some of the genes are remarkably short: the gene on the Y chromosome that controls maleness, for instance, is only 240 letters long, and is shared by all mammals, from mice to men. On the other hand, the abnormal gene that is responsible for the sickness called cystic fibrosis contains 250,000 letters, over 1,000 times as long. This gene, in the middle of chromosome 7, was discovered in 1989.

An important reward of being able to read the book of genes is to understand the basis of many of the conditions which are inherited, whether or not we are able to provide any cure at present. These conditions include relatively mild dysfunctions such as color blindness, which is linked to the X chromosome, to serious diseases such as Huntington's Chorea, cystic fibrosis, susceptibility to cancer, and a host of other life threatening ailments. Given these implications for medicine, it is not surprising that research directed at reading these genes has spawned multi-billion technologies almost overnight.

With the discovery that the genetic sequence of each individual is unique, another industry has also sprung into being: genetic fingerprinting or genetic profiling. Many crime cases have been resolved since the 1980s when genetic fingerprinting first entered the courts. A recent example is that of a dentist in Florida who was accused of transmitting AIDS to his patients, presumably because he had a small cut while working on them.

A scientific by-product of this interest in identifying individuals or populations, whether for medicine or for law, is that genealogies gradually gained a solid footing. Written records are useful, of course, but they have limited time-depths, and may be falsified. The history written in one's DNA cannot lie, and it allows inferences further back in time, much earlier than the invention of writing. In fact, some of exciting breakthroughs in recent years have been the application of molecular genetics to human remains that have

lain buried for thousands of years. I will return to the questions of population genealogy later.

As genetics advances, numerous applications emerge which apply the new knowledge. With these applications come a whole host of new questions, involving not only health and the law, but ethics and morals as well. Is food produced by genetic alterations safe for consumption? How much of the genetic information of an individual should be disclosed, to his employer, to his government, or even to himself? What is the status of organs or organisms created by genetic engineering, by cloning? Are we to differentiate between life forms created by cloning from those created by nature, and if so, how?

This new world created by genetics was certainly not envisioned by the modest monk, counting peas in his garden in Brno. Yet his were the experiments that led to the avalanche of knowledge that continues to rush forward today. The lesson we can draw from this is that there is no good research that is useless. The knowledge we gain today may find immediate application tomorrow. Or, it may set the stage for decades of cumulative research ahead, which will completely change the world. The distinction we customarily use between basic research and applied research is overly simplistic. Any work that leads to a deeper understanding of ourselves or of the world around us is good research, regardless of whether or when it will impact the market.

Let us turn now to my other story. Eighty years before Mendel reported his experiments in Brno, an Englishman was taking his seat on the Bengal Supreme Court in India. Even though his profession was that of a jurist, William Jones had already established by that time an enviable reputation as a great scholar of classical languages. His book, *A Grammar of the Persian Language*, published in 1771, was to go through nine editions in English and two editions in French. In addition to Persian, he had also translated major works from Greek and Arabic. His sojourn into India offered him ample opportunities to study the languages there, particularly Sanskrit, the classical language of the sacred Vedas. Presumably Jones was aided in his efforts by the fact that India has a long tradition in linguistic description, extending back some five or six centuries before the Common Era.

Against such a background, it was expected that Jones would make a major scholarly contribution soon after he reached India. The world was not disappointed, for in 1786, in an address to the Bengal Asiatic Society in Calcutta, Jones uttered the famous words which laid the foundations for the science of modern linguistics:

> 'The Sanskrit language, whatever be its antiquity, is of a wonderful structure; more perfect than the Greek, more copious than the Latin, and more exquisitely refined than either, yet bearing to both of them a stronger affinity, both in the

> roots of verbs and in the forms of grammar, than could possibly have been produced by accident; so strong indeed, that no philologer could examine them all three, without believing them to have sprung from some common source, which, perhaps, no longer exists: there is a similar reason, though not quite so forcible, for supposing that both the Gothick and the Celtick, though blended with a very different idiom, had the same origin with the Sanskrit, and the old Persian might be added to the same family . . .'

There are things from this famous passage which a modern linguist would shy away from saying. However one might feel privately about such matters, it is surely politically incorrect these days to say that one language is 'more perfect' than another, or that it is 'more exquisitely refined', particularly when no criterion is given for such comparisons. Parenthetically, however, my own feeling is that the pendulum of political correctness has swung too far the other way — that there are indeed scales of communicative effectiveness against which languages can be objectively evaluated, and linguists would do well to study these.

Another flaw in this remarkable passage is the 'perhaps' he inserted after the phrase 'sprung from some common source'. If X is the source from which A, B, and C have descended, then X cannot be a contemporary of these languages since it has already evolved into them. 'Perhaps' is a hedge word there that reveals a blur in Jones's thinking, and detracts from the force of his insight.

But these are minor detractions from a major intellectual breakthrough, and there are positive sides even to Jones's failings. Spurred on by successes of the Industrial Revolution, Europe was developing a complacency that increasingly saw itself as the apex of civilization, surrounded by peoples who were either barbarians or savages. This arrogance was to be epitomized by authors like Kipling, for whom only 'lesser breeds' are born beyond the English Channel, and with whom we associate the concept of the 'White Man's Burden.' An influential thinker like Robert Chambers, for instance, was able to publish this kind of nonsense about the Chinese language in 1887:

> 'The Chinese have a language which resembles that of children, or deaf and dumb people. The sentence of short, simple, unconnected words, in which an infant amongst us attempts to express its wants and its ideas — the equally broken and difficult terms which the deaf and dumb express by signs . . . — these are like the discourse of the refined people of the so-called Celestial Empire.'

William Jones was not tainted by these kinds of prejudice. Nonetheless, it must have required intellectual courage for him to write that an unknown language in far away India

was ‘more perfect’ and ‘more exquisitely refined’ than the revered Latin and Greek, and to propose that all these languages are related. Such an announcement from an eminent English scholar must have shocked the mindset that put England in the center of the civilized world, much like Darwin’s message did later to directly relate our species to the apes. Whether or not he intended the shock, the effect was surely a useful antidote to the unhealthy Eurocentrism of the time.

With the foundation Jones laid, progress came steadily. Rather than speculating endlessly on which was the language spoken in the Garden of Eden, as had been the focus before, scholars began the immense task of comparing languages for resemblances among them, and of reasoning probabilistically on what were the causes for these resemblances. Jones himself was already aware that some resemblances among languages were due to inheritance from a common ancestor, and some resemblances were due to one language imitating another. In this respect, Jones was also aware of the importance of basic words for establishing inheritance, such as names ‘of material elements, parts of the body, natural objects and relations, affections of the mind, and other ideas common to the whole race of man.’ Furthermore, he realized that at least three languages should be compared in evaluating these resemblances. Indeed, these insights are basic in linguistics today.

Unlike Mendel contemplating alone in the garden, Jones had an enormous net of correspondents with whom he discussed his findings. Step by step the family of languages that has ‘*sprung from some common source*’ was assembled as the changes which differentiated them became identified. These changes in language are like mutations in the DNA. As mutations accumulate over time in organisms, they eventually can no longer interbreed and evolve into distinct species. Similarly, as changes accumulate over time in languages, they eventually become no longer mutually intelligible. The ‘source’ that Jones pointed to is now known as Indo-European, a family of several hundred languages, including some of the most widely spoken in the world today, such as English, Spanish, Russian and Hindi. The linguistic hypothesis is that this family of languages was once a single language. Archeologists tell us that Indo-European was spoken perhaps in the region of modern Turkey some 7,000 years ago.

We can illustrate the resemblances that interested Jones with the names of the integers, shown in the following table[3].

English	Gothic	Latin	Greek	Sanskrit	Chinese
one	ains	unus	heis	ekas	yi
two	twai	duo	duo	dva	er
three	threis	trs	treis	trayas	san
four	fidwor	quattuor	tettares	catvaras	si
five	fimf	quinque	pente	panca	wu
six	saihs	sex	heks	sat	liu
seven	sibun	septem	hepta	sapta	qi
eight	ahtau	octo	okto	asta	ba
nine	niun	novem	ennea	nava	jiu
ten	taihun	decem	deka	dasa	shi

Table 1: Numerals in six languages.

In spite of the great time depth which separates these languages, the principles of comparative linguistics can still be clearly seen in some of the forms. Thus we see that for the integer '2', English and Gothic, which are Germanic languages, have words which begin with the consonant 't', whereas Latin, Greek and Sanskrit all begin with 'd', i.e., a correspondence of t:t:d:d:d. Furthermore, the second element in these words for all five languages is a labial sound, 'u' or 'v' or 'w', which strengthens the resemblance among these words. We see that the t:t:d:d:d correspondence is repeated for the integer '10'. Taking other evidence into account, this tells us clearly that an earlier 'd' has changed to a 't'.

In fact, this correspondence in sounds is part of a major discovery made by Jakob Grimm. He was a great linguist as well as the co-author of the popular Grimm's fairy tales, and the 1822 edition of his *Deutsche Grammatik* gave a comprehensive account of the relations between Indo-European and the Germanic languages. We could couch his discovery in genetic terms, and say that a series of mutations took place in the Germanic languages which differentiated them from the other Indo-European languages. Some other parts of Grimm's discovery can also be seen in fragmentary form, such as the correspondence f:f:q:p:p for the integer '5'. We can easily think up other words which support this correspondence, such as English 'father' and Latin 'pater', English 'foot' and the Latin root in 'pedal', and so on. Yet another partial correspondence can be seen in

[3] The forms in Table 1 are adapted from those given by Colin Renfrew, reprinted on p. 47 of Wang 1991.

the middle of the words for ‘8’. Latin and Greek have ‘k’, whereas Germanic languages have ‘h’. The ‘h’ is no longer pronounced in English, though it is preserved in the spelling ‘eight’, whereas it is pronounced in German ‘acht’. Fragmentary supporting data can be seen in the words for ‘10’.

Integer	**English**	**Gothic**	**Latin**	**Greek**	**Sanskrit**
Two, ten	t-	t-	d-	d-	d-
Three	th-	Th-	t-	t-	t-
Eight, ten	-gh-	-h-	-k-	-k-	-s-
Six, seven	s-	s-	s-	h-	s-

Table 2: Some Indo-European correspondences exemplified in the integers

A large number of such correspondences were worked out by scholars in the 19th century that it was possible to add new languages to the family as these correspondences were discovered. The early family tree published by August Schleicher in the 1860s, based on Jones’s famous remarks, became increasingly incomplete, as more and more such correspondences were worked out by linguists which pointed to new languages to add to the family.

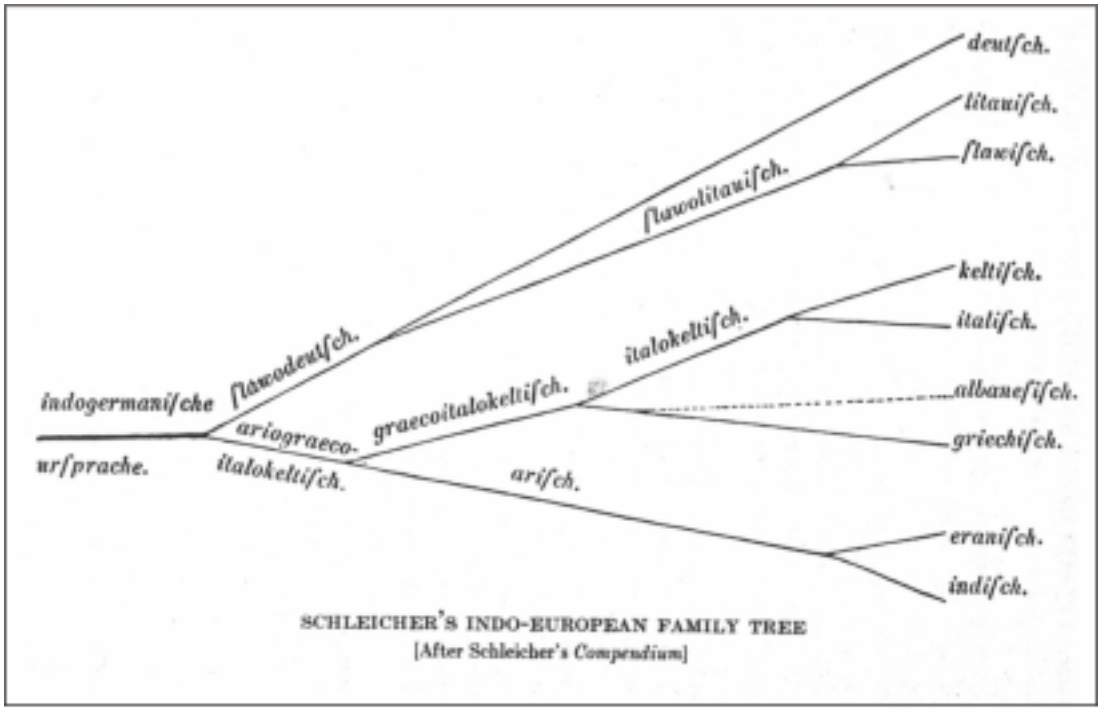


Figure 2: Schleicher’s tree.

The new members include several Hittite languages carved on stone tablets in cuneiform form, found in Turkey. Their relation to Indo-European was proposed in 1902. Closer to us here in Hong Kong are the finds in western Xinjiang, also around the beginning of the 20th century. These are of manuscripts and wooden tablets, written in Brahmi scripts which derive from India. The extremely dry desert climate and the shifting sands which bury and protect did a wonderful job at preserving these relics from the past. The language, Tocharian, is related to Indo-European. It is particularly intriguing because in many features Tocharian resembles the western Indo-European languages far away, rather than the Indian or Iranian languages much closer by. Recently, the discovery of several 4,000 year old mummies, of peoples who probably once spoke this language, gained a great deal of media attention as 'Caucasians' living in ancient China.

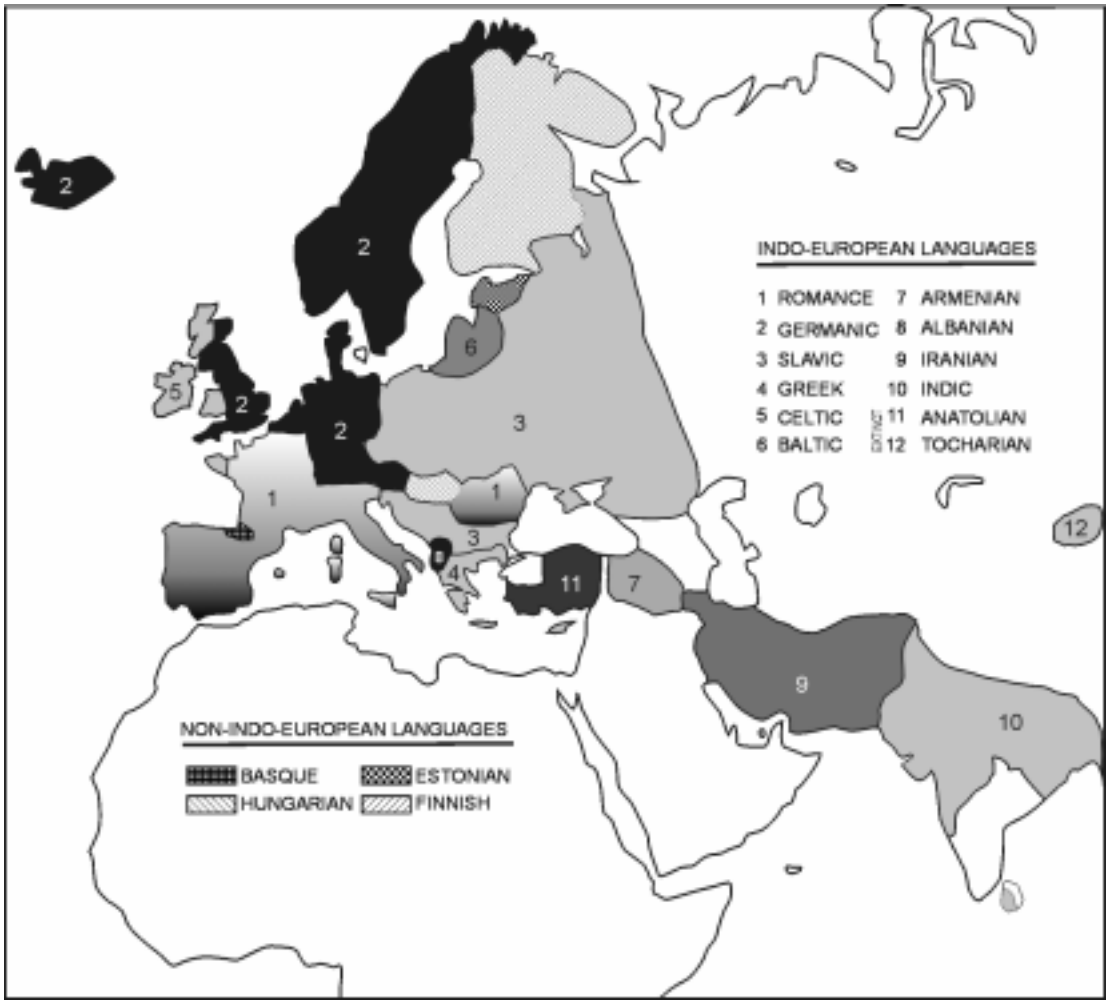


Figure 3: Renfrew's map.

In parallel with the advances in genetics, the study of language has also made important progress over the last century. While genetics now has mapped out the sequences in our DNA, linguists have achieved a rough knowledge of what the 6,000 some languages of the world are like — their grammatical structures and their geographical distributions. The question naturally arises as to how Indo-European is related to the other languages of the world. While numerous speculations have been made before, the first substantive hypothesis was recently proposed by Joseph Greenberg, a linguist at Stanford University, who has made enormous contributions to the classification of the world's languages.

Greenberg considers Indo-European to be but one branch of a larger super-family of languages, just as Germanic is a branch of Indo-European. This super-family he calls Eurasiatic. The evidence he has presented so far includes 70 some features of grammar. For example, we may note that in English, the first person singular pronoun has an 'm', as in 'me' ; and the second person singular pronoun used to be 'thou', where the consonant 'th' derived from a 't' via one of Grimm's correspondences. (The set of 'thou' pronouns was replaced by 'you' in English only in recent centuries.) This use of the m/t forms for first and second person pronouns respectively is found not only in Greek, Latin and Sanskrit, but spreads all the way across Asia to Korean and Japanese, and hops to the northern edge of the Americas in the languages of the Eskimos. All these languages, according to Greenberg, have sprung from a common Eurasiatic source.

Another super-family that has been recently proposed, primarily by a Moscow linguist, Sergei Starostin, is called Dene-Caucasian. Several members of this super-family are spoken by scattered pockets of small languages, tucked away in inaccessible mountainous regions: Basque in western Europe, Burushaski in northern India, and a group of languages in the Caucasus. Others are the Kets in northern Siberia, and the Na-Dene Native Americans in North America. The blotchy distribution of Dene-Caucasian languages suggests that its speakers were perhaps driven to these less fertile regions by the expanding Eurasiatics, who arrived on the scene later with a more advanced technology. Next to the expanding colonial languages, it is difficult to know how much longer these languages will survive.

The only branch of Dene-Caucasian that contains languages that have significant numbers of speakers with long literary traditions is the Sino-Tibetan family. The Chinese language, with numerous dialects spoken by over a billion speakers, is a member of this family. Burmese and Tibetan are the other two. Here too, these languages have expanded over the millennia at the expense of the hundreds of minority languages in China and in Southeast Asia, and many of these minority languages are on the brink of extinction.

In addition to Eurasiatic and Dene-Caucasion, other super-families are also being investigated in current research, to varying degrees of depth and precision. Greenberg

has propsed that the 6,000 some languages of the world can be classified into a dozen super-families, each one with a time depth of well over 10,000 years. This is a powerful working hypothesis that will influence research on human evolution for many decades to come.

Recall now that the ancestral language Indo-European has been dated to about 7,000 years ago. If we accept Greenberg's view that Indo-European was but one branch of Eurasiatic, then Eurasiatic must have started differentiating much earlier than 10,000 years ago. Similarly, Sino-Tibetan has been dated to about 6,000 years ago. Again, if Sino-Tibetan is but one branch of Dene-Caucasian, then Dene-Caucasian must have started differentiating much earlier than 10,000 years ago also. This is before the invention of agriculture, and considerably earlier than the invention of writing. Is there any way for us to know whether these hypotheses are true?

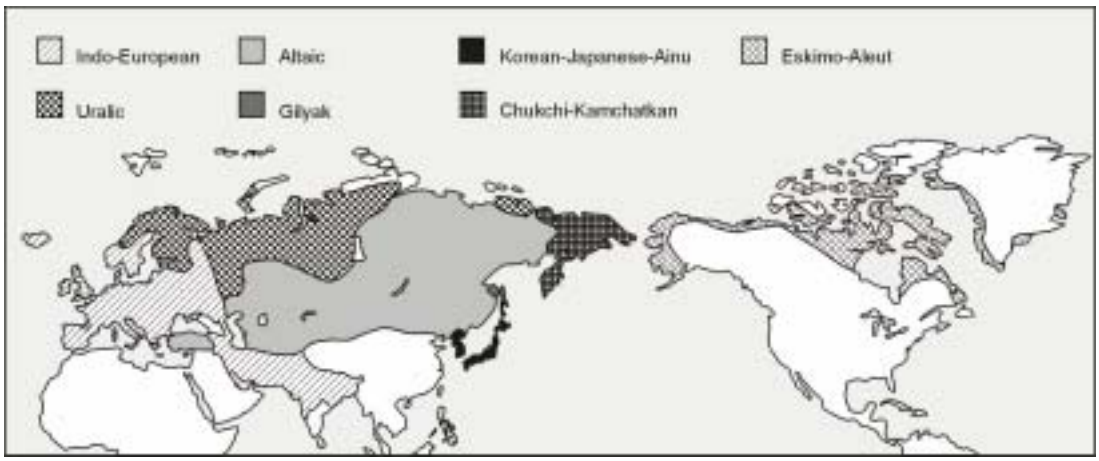


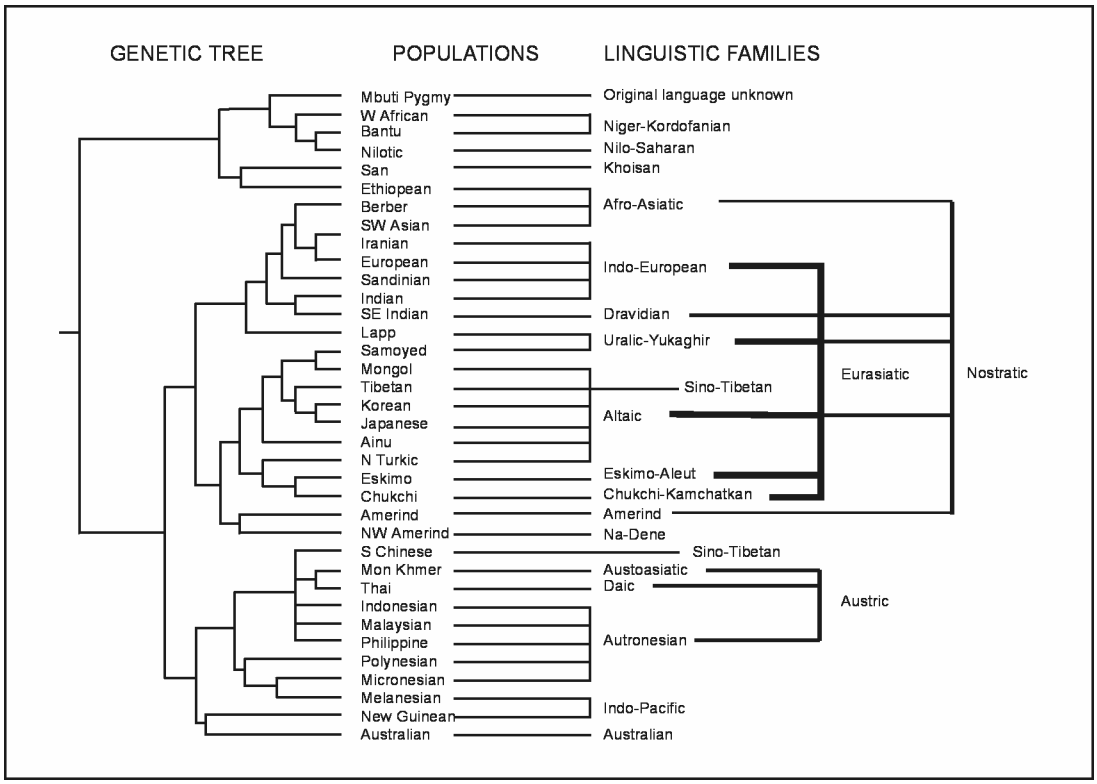
Figure 4: Eurasiatic

Figure 5: Dene-Caucasian

The first hint to the answer of this question came from Charles Darwin, who wrote these prescient words in his Origin of Species:

> 'If we possessed a perfect pedigree of mankind, a genealogical arrangement of the races of man would afford the best classification of the various languages now spoken throughout the world, and if all extinct languages, and all intermediate and slowly changing dialects, were to be included, such an arrangement would be the only possible one.'

Darwin's suggestion was that we should look both at words and at genes. As we saw earlier, when Darwin was writing in 1859, Mendel was busy counting peas in Brno. The research that Mendel started eventually led to our understanding of DNA, an understanding that has become so precise that every individual can be uniquely identified, and every population differentiated from one another. But all this happened after 1900, and Darwin had no idea how such a human genealogy can be constructed. The question remained dormant until a paper of 1988, by a team of researchers at Stanford University led by Luca Cavalli-Sforza. They published for the first time, a tree based on genetics on the left, side by side with a tree based on linguistics on the right.



Cavalli-Sforza, L.L., et al. 1988. Reconstruction of human evolution: bringing together genetic, archaeological and linguistic data. *Proc. Natl. Acad. Sci.* (USA) 85.6002–6006.

Figure 6: Double tree diagram.

As the title of their paper plainly states, archeology, genetics, and linguistics can each provide an independent and complementary window. The double-tree diagram published by these researchers, which has become justly famous, can serve as a sort of emblem of the flowing together of the many disciplines that bear on the big question of our origins. Many new results have been reported since their paper, using mtDNA and Y chromosome in genetics, which allows us to trace our genealogy either along just the maternal line, or just along the paternal line. It is a fascinating discovery that these two genealogies tell similar but distinct stories. This is perhaps not surprising, considering the different roles men and women played in prehistoric societies. New results have also been achieved connecting languages between Asia and America which strengthen the hypotheses for super-families in linguistics.

Only by bringing the disciplines together can we ever hope to achieve the knowledge of the ultimate origin of the human species, and arrive at, in Darwin's words, 'the only possible one'. The big picture that is unfolding with increasing clarity from the interdisciplinary research is that the most recent common ancestors of our species originated in Africa quite recently, around 100,000 years ago; and these were the peoples that invented the language from which all modern languages have sprung.

As part of this big picture, we are also getting a better understanding of where the peoples of China come from. I have referred to these recent findings from linguistics, genetics, and archeology as three windows on the past. My interpretation of these findings, to be verified by much future research, is that modern peoples entered China along two major routes. The northern route was taken by the Sino-Tibetans as they split off from the Dene-Caucasians. The southern route was taken by Austric peoples, who radiated out from the high plateaus of Xizang and Yunnan along major rivers to populate south China, Southeast Asia, and eventually Oceania and beyond.

The heart of leading edge research, as can be illustrated from many fields, is to make connections among seemingly disparate ideas that no one has seen before, and in so doing, deepen our understanding and broaden our research frontier. I have seen this again and again as I served on various committees which sponsor research: the most creative work is often associated with investigators who can forge relations across disciplines. The Chinese language expresses this idea well with the word *tong* 通, which means 'connect.' There must be no walls to impede the flow of knowledge.

The flowing together of different aspects of knowledge also has a broader meaning. The scientific name that describes our species is *Homo sapiens*, where the word 'sapiens' has the same root as 'sage', and refers to wisdom. Do we really deserve such a compliment? Are we really a wise species? As research has advanced across the centuries, some sectors of our knowledge have moved forward faster and outstripped the others. While science

and engineering have enabled us to launch mammoth vehicles deep into space, violence keeps on erupting all over the world, from wars in the Balkans to shootings in California high schools. How shall we be served by cloning artificial human beings when the natural ones, namely ourselves, are in constant conflict?

It is clear that a good deal of research is urgently required on the how's and why's of behaviors and values — of individuals as well as of diverse cultures — to keep pace with the awesome technological power that is still growing day to day. Information must flow at an accelerated rate among those who split atoms and decode molecules, and those who ponder the questions of morality, ethics, and aesthetics. Only then can we veer toward the vision recently expressed by the biologist, Stephen J. Gould:

> 'Science can then forge true partnerships with philosophy, religion, and the arts and humanities, for each must supply a patch in that ultimate coat of many colors, the garment called wisdom.'

The two men whose stories I told today, an English judge and a Czech monk, are but two of many similar stories we can find in the history of scientific research. The research they each started began as mere trickles, but gathered into major intellectual currents, joined by many tributaries along the way. These rivers flowed together toward the end of the last century, though they had very different beginnings. The combined insights of the research which they began are teaching us a lot about not only peas, or Sanskrit, but also about the ultimate origin of where we all come from. This is surely one of the most challenging questions that the human mind can grasp.

Time and again, we find that great discoveries come when we venture beyond the artificial confines of any one discipline, to combine and integrate insights across several disciplines. Perhaps the Chinese philosopher, Huainanzi, had something like this in mind, when he wrote 2000 years ago: 百川異源 而皆歸于海 'The myriad streams start from distant sources, they all return to the ocean.' To be able to add to this great ocean of human knowledge, that is indeed the joy of research.

Selected References:

Bodmer, Walter and Robin McKie. 1994. *The Book of Man — the Human Genome Project and the Quest to Discover our Genetic Heritage*. Oxford University Press.

Cavalli-Sforza, L.L., A.Piazza, P.Menozzi and J.Mountain. 1988. 'Reconstruction of human evolution: bringing together genetic, archeological and linguistic data.' *Proc. Natl .Acad. Sci*. 85.6002-6006.

Claverie, Jean-Michel. 2001. 'What if there are only 30,000 human genes?' *Science* 291.1255-7.

Gould, S.J. 1999. *Science* 284.2087, June 25.

Greenberg, Joseph H. 2000. *Indo-European and its Closest Relatives: the Eurasiatic Language Family*. Stanford University Press.

Jones, Steve. 1993. 'The Language of the Genes.' London.

Lamb. S.M. and E.D.Mitchell, eds. 1991. *Sprung from Some Common Source*. Stanford University Press.

Orel, Vitezslav. 1996. *Gregor Mendel, The First Geneticist*. Oxford University Press.

Pfeiffer, John. 1984. *The Cell*. Time-Life Books.

Ruhlen, Merritt. 1991. *A Guide to the World's Languages: Volume 1, Classification, with a postscript on recent developments*. Stanford University Press.

Wang, William S-Y., ed. 1991. *The Emergence of Language: Development and Evolution*. W.H.Freeman.

Wang, William S-Y. 1998. 'Three windows on the past.' 508-534 in *The Bronze Age and Early Iron Age Peoples of Eastern Central Asia*. V.H.Mair, ed. University of Pennsylvania Museum Publications.

研究之樂，樂在其中

孟德爾、瓊斯、人類起源

（譯文[1]）

香港城市大學 王士元

王士元教授（Professor William S-Y. Wang），1933 年生於上海，1960 年獲美國密芝根大學語言學博士學位。1965 至 1994 年任美國加州大學(柏克萊) 語言學教授，1995 年起受聘為香港城市大學電子工程學系語言工程學講座教授。目前兼任多所大學的特約教授，包括香港科技大學。

王教授曾任《中國語言學報》創刊人及主編，和國際中國語言學學會會長。他還享有多個院士榮銜，包括富布賴特委員會院士、古根海姆紀念基金會院士、加州史丹福及意大利莫爾貝尼奧的高等研究中心院士，及台灣中央研究院院士。

王教授於 60 年代後期開始探討生物進化與語言轉變的類同，數十年來，致力於推動中國語言學研究，他提出的詞匯擴散理論，在語言的演變和發展的研究中，具有全局性的指導作用。近年來，各個學術範疇對人類起源的研究日漸趨同，讓王教授十分興奮，希望藉著本文與大家分享箇中的樂趣。

---

1 本文乃根據 2001 年 3 月在香港城市大學「開拓研究前沿」會議中所作的一場演講整輯而成。我在此感謝美國史丹福大學的 Merritt Ruhlen 教授提供本文所使用的歐亞語系及提納—高加索語系的語圖。另外我要感謝胡雙虹博士、柯津雲女士及林千哲先生與我諸多的討論。

此份中文譯文由盧少光先生翻譯，並由林千哲先生校正，希望能對更廣大的讀者群有所幫助。

我很高興能有機會參加「開拓研究前沿」各項姿彩繽紛的節目。研究是任何充滿使命感的大學精神所在，而不斷開拓前沿一定會提高其研究質量。再者，研究與教學其實是智性工作的一體兩面，是知識的「陰」與「陽」。一位能春風化雨的老師，會親身思考研究授予學生的知識，而一位出色的研究者則會樂於將自己的發現公諸於世。上述兩種工作都有本身的樂趣，無可替代。看到學生因為了解一個新觀念或新方法而雀躍，或是本身在研究上有所突破，能入無人之境，這二者所帶來的喜悅，都是獨一無二的。

有關研究的前沿這個命題，範圍廣闊，因此，我必須有所選擇，就讓我從兩位偉人的故事說起。第一位是生活在十九世紀捷克布爾諾 (Brno) 的修道士，而第二位是在十八世紀印度加爾各答的英國法官。他們分別研究的遺傳學及語言學，起初時像小溪，但今天已各自發展成為知識的長河，並在二十世紀末開始匯合成流。兩位學問家留下的遺產，讓我們更了解自己、人類及人類的起源。

一百多年前，在今天捷克共和國布爾諾有一位毫不起眼的修道士在種植豌豆，他有興趣知道豌豆的特徵是如何由一代傳至下一代。他發現豌豆莖部長短參差不齊，種子的形狀也不盡相同：有圓滑的，但也有起皺摺的。另外，豆莢的顏色也有分別：有綠的，也有黃的。這位修道士於是種植了無數代次的豌豆，有時用純正的品種，有時用雜交配種，他前前後後種了好幾千棵豌豆，並且記錄下它們的特徵如何由一個世代傳到下一個世代。

這些實驗引出了有趣的結果，1865 年那位修道士向當地的科學家協會作了一個報告。他提出豌豆某個特徵可以比另一特徵更具顯性性狀的論點，例如長莖特徵可以壓倒短莖特徵；而莖的長短是獨立遺傳的，不受其他特徵——例如豆莢顏色—— 的影響，他把這種現象稱為「獨立分配定律」(Principle of Independent Assortment)。另外，基於他對每一代植物進行的統計分析，他也提出了若干類似的定律去解釋這些事實。會場的聽眾只是報以禮貌性的掌聲，隨即便把注意力轉移到當時「熱門」的科學話題上。幾年前，達爾文剛發表了富爭議性的《物種起源》一書，社會上下都對進化論議論紛紛。

這就是遺傳科學卑微的開端，而那位修道士就是孟德爾 (Gregor Mendel)。倘若當日在座的小量聽眾中有幾位是知識淵博及思考精深之輩，他們便會明白到孟德爾的

論點，剛好為達爾文進化論的具體化提供了方法。其實，達爾文本人在遺傳學的研究上並不順利，在其著作中犯了幾個基本的錯誤。但這距離真正明白孟德爾發現的重要性還有一段日子，基本上當時的科學家是把他遺忘了。他在 1866 年出版的研究報告，就一直擱在圖書館的書架上被塵埃封蓋了三十多年。最後，直至 1900 年，三位在歐洲不同地方工作的荷蘭、德國及奧地利科學家各自發現了孟德爾的報告，其後不久，遺傳學上的突破才成為達爾文進化論思想的核心部分，並因此造就了所謂「綜合進化論 (Synthetic Theory of Evolution)」的誕生。

在布爾諾的豌豆以來的一個世紀裡，遺傳學已經歷了多次驚人的突破及發現，並贏得多次諾貝爾獎所帶來的讚譽。最早的一個諾貝爾獎是授予以研究果蠅的摩根 (Thomas Morgan)，他從研究果蠅眼睛的顏色中，發現了決定性別的基因，亦即區分男女性別的 XX 和 XY 染色體。另一個諾貝爾獎則是由克里克和華生(Francis Crick 及 James Watson) 所領導的小組取得，他們成功破解了雙縲旋狀的 DNA 分子結構，讓我們能更深入了解基因密碼中四個字母 A, T, C 及 G 的排列組合。圖一 (英文版第 4 頁) 顯示 DNA 的結構不僅在生物學上很簡約，亦極見美感。最近，穆利斯 (Kary Mullis) 因發明了聚合酶連鎖反應（PCR）——種能在試管內產生千萬個 DNA 拷貝的連鎖反應方法——而獲得諾貝爾獎，這項技術對於幫助破解人類基因圖譜內數以億計的鹼基對（即製造人體的 DNA 指令）功不可沒。

至二十一世紀初，即孟德爾發表研究結果之後 140 年，科學家已能定出整個人類基因圖譜內全部三十億個字母的排列次序。但有了這四個字母排列次序的知識，還遠遠未能說得上是掌握了箇中意義，就如懂得希臘字母還離開能夠閱讀希臘文書籍很遠的道理一樣。認識字母，只是學習語言——基因的語言——的起步而已，我們現在只能辨別人體內 30,000 多個基因中的寥寥幾個。某些基因相當的短小，例如控制男性性徵的 Y 染色體，只有 240 個字母的長度，並且能在所有哺乳類動物——由老鼠到人——的身上找到。另一方面，一種造成囊腫性纖維化症的基因，則有 250,000 個字母，是前者的 1,000 多倍。後者於 1989 年在第七對染色體中找到。

人類能夠解讀基因譜圖的主要好處之一，是明白到人體多種疾病的起因是由遺傳而來，無論我們現在能否找到醫治方法。這些病症包括了由輕度的身體障礙，例如與 X 染色體有關的色盲，到嚴重的疾病，例如抗廷頓斯舞蹈病(Huntington's Chorea)、囊腫性纖維化症、癌病以及一大堆威脅生命的重病等。鑑於此等發現對醫學影響重大，難怪以解讀基因為目的的研究如雨後春筍般出現。

在發現每個人的基因排列都是獨一無二之後，另一門工業——建立個人基因指紋或基因檔案——亦應運而生。80 年代以來，若干法庭案件亦拜基因指紋技術之賜，才

得以審結。最近的一個案例，是美國佛羅里達州一名牙醫被控把愛滋病轉染給病人，據說是因為他在為病人行醫時身上負有小傷口。

無論是出於醫學或法律的需要，這項用於識別個人或整個民族的科學發展，令基因研究變得更穩健。歷史文字資料固然是十分有價值，但可惜年期效用有限，甚至易於偽造。由個人 DNA 所寫成的歷史則是無法篡改的，並可追溯到更早的時期，甚至是在發明文字記錄之前。近年遺傳學上有不少令人振奮的突破，將分子遺傳學應用到已埋藏了幾千年的人體骸骨研究上。讓我稍後再談談人口基因學的問題。

隨著遺傳學的發展，不少將新知識付諸實踐的應用方案相繼面世，但亦因此而衍生一大堆新問題，不單牽涉保健醫療及法律問題，更與倫理道德息息相關。經過基因改造的食物是否可以安全食用？究竟有那些個人基因資訊可以向僱主、政府或其本人透露？經由基因工程及複製技術而產生的人體器官或生物究竟享有甚麼地位？究竟我們應否區分由複製而來及由自然產生的生命？如應該的話，又應如何入手？

這個由遺傳學衍生的新世界，並不是那位終日在布爾諾園子裡數算豌豆的修道士所能預見。然而，他的實驗卻掀起了一場知識的大革命，至今仍不斷湧向前沿。我們可以從這個故事學到一個道理，那就是好的研究不愁沒有用處。我們今天取得的知識，明天或會找到實實在在的應用。或者，它將為未來幾十年的累積性研究打下根基，做好準備，最終徹底改變世界。我們習以為常會將基礎與應用研究區分，這往往是過於簡單，任何能令我們對自己或身處的世界加深認識的研究都是好的，無論它何時會在市場上發揮作用。

讓我現在轉到另外一個故事上去。在孟德爾發表在布爾諾的實驗結果之前 80 年，有一位英國人出任印度孟加拉高等法院法官一職。雖然當時的瓊斯(William Jones)以法律為職業，但他已是出色的古典語言學家，聲譽蜚然。他在 1771 年出版的《波斯語文法》一書，總共出了九個英文版及兩個法文版。除了波斯語外，他亦有將希臘及阿拉伯的重要著作翻譯成英語。瓊斯在印度的日子，讓他有更多的時間及機會研究語言，特別是神聖的吠陀文本（Vedas）中所用的古典語言——梵語。身處印度，相信對瓊斯的工作有所幫助，因為印度擁有悠長的語言描述傳統，可追溯至公元前五、六個世紀。

在此前提下，可以預見的，瓊斯在抵達印度後的短時間內，便會作出重大的學術貢獻。果然，他並沒有令人失望。1786 年瓊斯在加爾各答的亞洲學會上，發表了以下的演說，為現代語言科學奠下了基礎：

「姑勿論它的古老歷史背景，梵語的結構奇妙，比希臘文更完美，比拉丁文更豐富，比這兩種語言更為優雅，然而它的動詞根與文法形式與兩者的相似程度，絕非偶然。它們是那麼相似，任何研究這三門語言的學者都會相信三者是來自某個共同起源，這個源頭或許已不復存在；我們有類似的理由相信，縱使理由不那麼有力，我們可以推斷歌德語系及凱爾特語系兩者雖然已滲入一些不同的詞語，但與梵語應屬同一起源。另外，古老的波斯語也可以一併歸入此一語系內. . .」

這篇著名文章中有若干的觀點，現代語言學者會避而不談。無論他們私下的意見如何，但基於政治正確的考慮，肯定沒有人會在這個年代公然宣稱甲語言會比乙語言更完美或更優美細緻，特別是在沒有提出任何比較標準的情況下。讓我在這裡加入一些個人意見，我認為政治正確的鐘錘已擺過了頭，現在確實有客觀的標準去量度語言的傳意效率，而語言學者應多加研究，從中獲益。

上述那段著名的文章裡的另一道瑕疵，是作者在「來自某個共同起源」的短句後加上了「或許」，倘若 X 是 A, B 及 C 的來源，那麼 X 便再不能與這些語言屬於同一時代，因為 X 已演變成為它們一部分。這裡的「或許」是個投機用語，顯露了瓊斯思想裡的模糊點，也削弱了他的洞察力。

但這些只是重大知識突破裡的小瑕疵而已，就是瓊斯的缺失也帶來正面的意義。由於歐洲工業革命成功，當時社會上瀰漫著一股沾沾自喜的態度，歐洲人逐漸把自己看成是人類文明發展的頂峰，四週都是野蠻人。這種傲慢態度表現在作家吉卜林（J. Rudyard Kipling）筆下，他認為在英倫海峽以外出生的人都屬「次一等族類」，並提出了「白人的負累」觀點。舉例來說，像錢伯斯 (Robert Chambers) 這樣偉大的思想家，也在 1887 年發表了一篇對中文充滿謬論的文章：

「中國人的語言，好像是孩子、聾人或啞人的語言一樣。中文句子充塞了互不相連的簡單短字，好像我們的嬰孩在嘗試表達自己的需要及意念，或聾人透過手勢表達不連貫而難懂的詞語. . . . . .這就是所謂天朝裡有教養文化的子民的溝通方法。」

瓊斯卻沒有受這類偏見所影響，他實在需要有知識份子的勇氣，方能提出在遙遠印度一門未為人知的語言，比普世尊崇的拉丁文及希臘文「更優美細緻」這個論點，並且說這幾門語言都互相關連。由一位英國學術泰斗口中說出這樣的見解，一定給當時英國人自視為世界文明中心的觀念帶來衝擊，就好像其後不久達爾文把人

類與猴子的進化拉上關係的論點一樣，無論瓊斯是否存心掀起如此震盪，結果還是給當時那套病態的狹隘歐洲中心思想進行了一次消毒。

在瓊斯的根基上，日後的發展接踵而來。人們不再像以往一樣，把注意力集中在揣測伊甸園內說的究竟是那一種語言，而是開始著手研究和比較各種語言，從中尋找相似之處，並推論其背後的可能成因。瓊斯本人已意識到語言之間存在某些相似之處是由共同祖先遺傳而來，而其他相似之處則是語言之間互相模仿所致。在這方面，瓊斯亦明白到基本詞彙，例如那些描述「物件、身體各部位、自然界事物及人際關係、情緒思維以及全人類某些共同意念」的名稱，對於確立語言遺傳關係的重要性。另外，他也了解到最少需要有三種語言方可研究和比較語言的相似性。其實，這個見解也是今天語言學所通用的。

跟孟德爾在園子裡獨自思考有所不同，瓊斯得到一大群志同道合的人幫助，並與他們討論他的研究發現。隨著引致語言差別的變化得到識辨，由「共同起源」衍生的語言族系也慢慢建立起來。這些語言的變化就好像 DNA 的突變一樣：生物經過長時間的累積變化，不能再進行跨品種之間的雜交，最終要演進成獨特的物種。同樣地，語言的變化經過長時間的累積，最後變成不能互通的地步。瓊斯所指的「起源」，今天我們稱之為印歐語系，它涵蓋數百種語言，包括當今世界最通用的英語、西班牙語、俄語及印度語。語言學上假設這個語系曾經是單一語言，考古學家發現 7,000 年前印歐語可能在今日的土耳其地區內流行。

我們可以透過以下數字的名稱，說明令瓊斯感興趣的跨語言相似之處：[2]

英語	哥德語	拉丁文	希臘文	梵文	中文
one	ains	unus	Heis	Ekas	yi
two	twai	duo	Duo	Dva	er
three	threis	trs	Treis	Trayas	san
four	fidwor	quattuor	Tettares	Catvaras	si
five	fimf	quinque	Pente	Panca	wu
six	saihs	sex	Heks	sat	liu
seven	sibun	septem	Hepta	sapta	qi
eight	ahtau	octo	Okto	asta	ba
nine	niun	novem	Ennea	nava	jiu
ten	taihun	decem	Deka	dasa	shi

表 1：六種語言的數字

[2] 表 1 乃根據 Colin Renfrew 所提供的資料重製，曾刊印於 Wang 1991，第 47 頁。

雖然這些語言之間相距悠長的歲月，但某些比較語言學的規律仍然有跡可尋。所以，我們看見屬日耳曼語族的英語與哥德語的數字 2，是擁有以輔音 t 為首的字，而拉丁文、希臘文及梵文則全部以 d 為首，所以出現了是 t:t:d:d:d 的對應。另外，全部五種語言裡這批字的第二個元素都是唇音：u，v 或 w，加強了這些字之間的相似性。我們看見，在數字 10 中亦重覆出現 t:t:d:d:d 的對應，再加上其他證據，這明顯指出較早期的 d 音已演變成 t 音。

其實，語音對應性是格林 (Jakob Grimm) 重大發現的一部分，他既是偉大的語言學家，又是家傳戶曉的《格林童話》作者之一。在 1882 年版的《德文文法》(Deutsche Grammatik) 一書中，他詳細論述了印歐語系與日耳曼語系之間的關係。我們可以借用遺傳學的術語去表達他的發現，那就是說在日耳曼語中發生了連串的突變，令它從其他印歐語言中區別出來。格林在其他方面的發現，也可以透過片斷方式理解，例如整數 5 的 f:f:q:p:p 對應。我們可以容易想起足以證明這種特性的其他字詞，例如英語的 father（父親）與拉丁文的 pater（父親），英語的 foot（腳）與拉丁文字根 pedąl（足）等等。另一個局部相應現象可在 8 字中間找見，拉丁文及希臘文的是 k，而日耳曼語的則是 h。在英語中 h 已不發音，雖然在書寫的 eight 字中仍得以保存，但德語則讀成 acht。這些零星的證據也可以在多個 10 字裡找到。

整數	英語	歌德語	拉丁文	希臘文	梵文
Two, ten	t-	t-	d-	d-	d-
Three	th-	Th-	t-	t-	t-
Eight, ten	-gh-	-h-	-k-	-k-	-s-
Six, seven	s-	s-	s-	h-	s-

表 2：印歐語系中整數的某些相應現象

十九世紀的學者已找到大量這樣的相應現象，因此日後陸續發現的對應規則便能把新語言納入語言族系內。基於瓊斯的著名論點，施萊克爾 (August Schleicher) 在 1860 年代發表了早期的語族圖譜 （參看英文版第 10 頁、圖 2）。但隨著語言學家發現更多相應現象，以及不少新語言加入語系，圖譜已漸漸變得不夠全面。

新加入的成員，包括刻在土耳其出土的石塊上，用楔形文字組成的若干希泰語，於 1902 年有學者提出它們與印歐語系的關係。回到與香港較接近的新疆西部，在大約二十世紀初的考古中，發現了一批文獻及木板塊，上面載有由印度傳入的婆羅門文。當地沙漠氣候極為乾燥且經常刮風沙，令這批古老的文物得以好好的埋藏及保存。文物上所用的語言，是與印歐語系相關的吐火羅語 (Tocharian)。特別令人

感興趣的，是吐火羅語與在地域上距離較遠的西印歐語有多種相似的特徵，而與較接近的印度或伊朗語比較不同。近期考古發現幾個 4,000 年前的木乃伊及曾經使用這種語言的民族，引來傳媒對曾經在古中國生活的「高加索人」作了不少報導。（參看英文版第 11 頁、圖 3。）

在過往的世紀裡，語言學與遺傳學並肩前進，取得重大進展。至今，遺傳學已勾劃出人類 DNA 的排列次序，語言學者亦對全世界 6,000 多門語言的情況——包括它們的文法結構及地域分佈——取得概括的了解。下一個順理成章的問題是，究竟印歐語系與世界其他語言有甚麼關連？雖然以前在這方面已有不少的揣測，但首次具有實質性的假設是由史丹福大學語言學家格林堡 (Joseph Greenberg) 最近提出，他為全球語言分類作出了巨大的貢獻。

格林堡認為印歐語系只不過是更龐大的超語系裡一個分支而已，就好像日耳曼語是印歐語系的分支一樣，他把這個超語系稱為歐亞語系（Eurasiatic）。至今，他已提出了 70 多項文法特點作為佐證，例如，英語第一人稱單數代名詞 me 中有 m，而以往第二人稱單數代名詞是 thou，而輔音 th 則是透過如格林所提出相應性的 t 轉化而來。（英語中的代名詞 thou 由 you 取代，只是近幾個世紀的事）。而在第一及第二人稱代名詞分別採用 m/t 格式，不單在希臘文、拉丁文及梵文中可以找到，更遠伸至亞洲的朝鮮語及日語裡，甚至跨過美洲北部在愛斯基摩語中出現。按格林堡之見，所有這些語言都出自同一起源的歐亞語族。

最近，主要由莫斯科語言學家史塔羅斯丁(Sergei Starostin) 提出的另一個超語系，名為提納—高加索語系 (Dene-Caucasian) 其中若干成員的語言規模較小，零散地分佈於與外界隔絕的山區，例如：西歐的巴斯克語、印度北面的布魯沙斯基語(Burushaski)、以及在高加索區的一帶的語言，其他包括西伯利亞北部的凱特語（Kets）及北美那提納美洲原住民用的語言。由其零散的分佈看來，這些民族是被歐亞語族趕到這些較貧瘠的區域。相對於那些不斷擴散的殖民地語言，我們很難知道這些小規模語言可以生存多久。（參看英文版第 13 頁、圖 4 及 5。）

漢藏語系是提納一高加索語系中唯一有較多人口及較長文字傳統的語系。擁有超過十億使用者且方言眾多的漢語亦屬此一語族。緬甸語及藏語是另外兩例。這些語言數千年來擴張郤造成了在中國及南亞地區產生了幾百個少數民族語言，其中有許多均已瀕臨滅絕。漢藏語是提納語系中唯一有國家地位的分支，而有超過十億人使用的中文更是這個語系的成員。

現在對其他超語系也有各種不同深度和準確性的研究，以格林堡的見解，現今世上 6,000 多種語言可歸納為十數個超語系；每一個都超過一萬年。這個假設將對人類演化的研究有深遠的影響。

讓我們重溫一下，古老的印歐語系有 7,000 年的歷史。如果我們接受了格林堡的觀點，即印歐語系只是歐亞語系的分支，那麼後者應在 10,000 年前開始分化。同樣，漢藏語族也有 6,000 年的歷史，如果漢藏語是提納—高加索超語系的分支之一，那麼後者也必然是在 10,000 年前開始分化，這是在農耕出現之前，比發明文字還要早。我們是否有方法判斷這些假設是否真實？

為以上問題提供第一道線索的是達爾文，他在《物種起源》裡寫下了這樣的真知灼見：

> 「倘若我們擁有完美的人類族譜，那麼我們把各種族的人按其系譜排列，便能將當今世上各種各類的語言有效劃分。如果我們要將所有消失了的語言，以及在過渡期間慢慢變化的方言也包括在內，這樣的排列能讓我們找到唯一可能的答案。」

達爾文的建議，是我們應從語言文字及遺傳基因兩方面入手。正如我們上述所說，1859 年當達爾文在著書時，孟德爾則正在布爾諾忙於種植豌豆。孟德爾開展的研究，最後令我們了解 DNA。今天這門知識的精確程度，已到可以識別每一個人及每一個民族的地步。這些都是 1900 年以後發生的事，達爾文是沒法知道如何建立人類系譜的。這個問題一直蟄伏著，直至 1988 年由史丹福大學卡華里—蘇弗札 (Luigi Luca Cavalli-Sforza) 所領導的研究小組發表的一篇論文，才首先披露了分別以遺傳基因（在左）及語言學（在右）為基礎的系譜（參看英文版第 14 頁、圖 6）。

正如他們論文題目直截了當地指出，考古學、遺傳學和語言學各自提供了既相獨立又相互補足的觀點，這群研究學者所發表的雙系譜圖實至名歸，象徵了多個學科合力解答人類起源大問題的努力。自從這篇論文面世之後，不少新的研究成果陸續湧現。現在利用遺傳學的 mtDNA 及 Y 染色體的幫助，我們可以追尋父母任何一方的祖先，有趣的是這兩種方法卻導致相似但不相同的故事，這並不奇怪，因為史前的男性與女性在社會中必定扮演著相當不同的角色。新的發現試圖將亞洲、美洲的語言聯繫起來，從而加強了語言學有關超語系假設的可信性。

唯有透過將不同學科的整合，我們才可望在尋找人類最初起源的知識上取得成果，找到達爾文所說「唯一可能的答案」。跨學科研究所展示的輪廓愈來愈清晰，

最近發現人類的共同祖先在 100,000 年前左右生活在非洲，他們發明的語言，是所有現今語言的起源。同時，我們對於中國人從哪兒來也有了更深的認識。我曾經把語言學、遺傳學及考古學的發現作為研究過去的三個窗口，根據這些發現，我認為現代人循二條路線進入中國：漢藏語族在由提納一高加索語系獨立出來後由北方進入中國；沿著西藏、雲南一帶的高原順流而下的澳斯人(Austric Peoples)，則由南方進入，他們逐漸分佈於南中國、東南亞、甚至大洋洲及更遠的地方。來自多門科學反覆的例子證明，創新性研究的精髓是將看似無關的意念組織連繫起來，展示前人所未能見的道理，而在這過程中，加深和擴闊了我們對研究領域的認識。我曾多次參與研究獎助的審查工作，往往發現最有創意的研究是那些能夠整合不同領域人才的研究，中文裡將這個意念表達得最好的是「通」字，即接通的意思，知識之流通不應受到任何阻礙限制。

把各個知識領域互相接通、匯合成流，亦具有更廣闊的意義。人類的科學名稱是「智人」(Homo sapiens)，其中 sapiens 與 sage（智慧）享有共同字根。我們對此美譽是否當之無愧？我們是否真是智慧族類？多個世紀以來研究不斷向前，我們在某些領域的知識突飛猛進，比其他的為快，科學與工程讓我們可以將體積龐大的火箭發射上太空，但由巴爾幹半島的戰爭到美國加州中學的槍擊案等暴力事件都是依然不絕於世，當我們身為自然物種卻仍然陷於衝突和矛盾之中，又焉能從生物複製技術中得到益處？

顯而易見，我們急需對個人及不同文化的行為及價值觀進行大量的研究，探究其背後的成因，方能趕得上仍在不斷躍進的巨大科技力量。從研究原子分裂和解破分子的科學家和從事道德、倫理及美學思考的學者之間應有更多、更快的資訊交流。唯有這樣，我們才可以步近生物學家顧爾德 (Stephen J. Gould) 最近描述的視界：

> 「那個時候，科學可以與哲學、宗教、藝術和人文學科結合成為真正的伙伴，因為每門學科只是最終被編織成名叫智慧彩衣上的一小塊布而已。」

我今天講及兩位人物——一位英國法官及一位捷克修道士——是科學研究歷史上眾多類似故事中的兩個，他們的研究開始時好像小溪涓流，但後來經過多條支流慢慢匯合，變成知識的洪流。雖然它們有不同的起點，但在上世紀末已連接起來，他們的研究所匯聚貫通的觀點，已不單是對豌豆和梵文的認識，而是對人類最初起源的探索，這當然是人類思想中最富挑戰性的問題。

偉大的發現往往在我們敢於打破個別學科的人為規限和揉合多門學科的見解時出現。或許，二千年前中國哲學家准南子也有同感，他說：「百川異源，而皆歸於海」。能夠對人類浩瀚的知識之洋有所貢獻，就能真正領會研究之樂，樂在其中。

Published by:
University Publications Office
City University of Hong Kong
Tat Chee Avenue, Kowloon, Hong Kong
Fax: (852) 2628 2812
Email: puo@cityu.edu.hk

出版：
香港城市大學 出版事務處
香港 九龍 達之路
電傳：(852) 2628 2812
電郵：puo@cityu.edu.hk