The History and Geography of Human Genes

ABRIDGED PAPERBACK EDITION

L. LUCA CAVALLI-SFORZA
PAOLO MENOZZI
ALBERTO PIAZZA

PRINCETON UNIVERSITY PRESS
PRINCETON, NEW JERSEY

Published by Princeton University Press, 41 William Street,
Princeton, New Jersey 08540
In the United Kingdom: Princeton University Press,
Chichester, West Sussex

Library of Congress Cataloging-in-Publication Data
Cavalli-Sforza, L. L. (Luigi Luca), 1922–
The history and geography of human genes /
Luigi Luca Cavalli-Sforza, Paolo Menozzi, Alberto Piazza.
p. cm.
Includes bibliographical references and index.
ISBN 0-691-08750-4 (unabridged cloth ed.)
ISBN 0-691-02905-9 (abridged pbk. ed.)
1. Human population genetics—History. 2. Human evolution.
3. Human geography. 4. Human population genetics—Research.
I. Menozzi, Paolo, 1946– . II. Piazza, Alberto, 1941–
III. Title.
QH431.C395 1993
573.2'15—dc20 93-19339

This book has been composed in Times Roman
Designed by Jan Lilly

The cover illustration is a map of the world showing four major ethnic regions. Africans are yellow, Australians red, and Caucasians green. Mongoloids show the greatest variation retaining some similarities with Europeans on one side (a light brown greenish tinge in middle Siberia) and with Australians on the other (a pinkish color in parts of America and on the way to it). The extensive gradients due to admixtures between Africans and Caucasoids in North Africa, and between Caucasoids and Mongoloids in Middle Asia, are clearly visible. (See chapter two.)

Princeton University Press books are printed on acid-free paper and meet the guidelines for permanence and durability of the Committee on Production Guidelines for Book Longevity of the Council on Library Resources

First printing of the abridged paperback edition, 1996

Printed in the United States of America

10 9 8 7 6 5 4 3 2

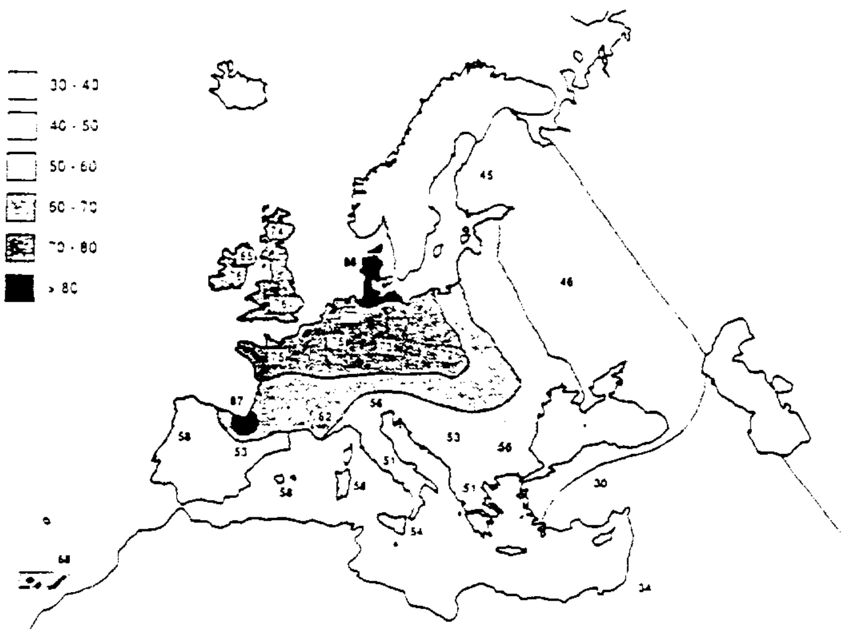


Fig. 2.14.10 Geographic distribution in Europe of the frequency (%) of Δ F-508 cystic fibrosis mutant, relative to other mutants determining the disease (data from European consortium on cystic fibrosis; modified from Devoto et al. 1990).

and therefore one amino acid) is very frequent, reaching approximately 70% of all known mutants determining the disease. Its distribution in Europe is shown in figure 2.14.10. The frequency at birth of the disease itself is less well known, but is found, on the average, in 1 of 2000 of all births in Europe. It varies somewhat from country to country, but the frequency of a genetic disease is not always easy to estimate accurately.

Very recent data indicate that the relative frequency of the common allele for cystic fibrosis (Δ *F*-508) is especially high in a relatively small sample of Basques (see fig 2.14.10). The geographic spread of the frequent allele is quite similar to the first synthetic map for Europe and therefore to that of the Mesolithic (pre-Neolithic) European gene pool. The conclusion would be that the most common cystic fibrosis allele in Europe is older than the spread of agriculture and must have been especially frequent among Mesolithics, while most other alleles come from the Middle East or originated in Europe after the Neolithic diffusion.

Many of the polymorphisms conferring resistance to malaria and other diseases may be relatively young, but the great majority of polymorphic alleles that we have studied in sections 2.10 and 2.13 are found in nearly all continents and therefore must have antedated the spread of a.m.h. Only a few are known at the DNA level, and it is impossible to say whether different mutations are hidden behind the same allele. The 100 RFLPs studied in section 2.4, however, are likely to have had a unique mutational origin, and the great majority of them are found in all continents (except the Americas which have not yet been studied for these polymorphisms). There are reasons to think that most human polymorphisms are much older than the date of spread from Africa. Theoretically, in the absence of selection, the average age of polymorphisms is comparable to that of the species. An estimate of the average time of appearance of the mutant allele of biallelic human RFLPs is 700,000 years (Mountain et al. unpubl.).

2.15. A BRIEF SUMMARY OF HUMAN EVOLUTION

The analysis of phylogenetic trees has used many different sets of data: nuclear polymorphisms tested by electrophoresis or immunological techniques, restriction fragment polymorphisms of nuclear genes, and mitochondrial DNA. All methods show a somewhat greater difference between Africans and non-Africans than between other human groups, and offer some information on dates supporting the interpretation that the origin of modern humans was in Africa, from which an expansion to the rest of the world started about 100 kya. The interpretation rests on the assumption that evolutionary rates are reasonably constant. Some further comfort is derived from the consideration that the rates involved in the evolution of polymorphisms of nuclear genes and those of mtDNA are different; the first are determined mostly by differences in gene frequencies separating two populations, and the second by numbers of mutations separating two individuals. The dates to which they lead for the bifurcation of Africans and non-Africans are different, being of the order of 100 kya and 200 kya

DOI 02175

respectively. Given the ways in which these dates were obtained, this difference should be expected, with the second being greater than the first by an amount that has not been determined, but perhaps might be estimated on the basis of theoretical considerations. All in all, there is basic agreement between the trees obtained with mtDNA and with nuclear markers, the latter giving much more detail. It is worth remembering that naming the mitochondrial ancestor "Eve" has generated the false belief that there was a time when there was only one woman alive.

On the question of place of origin, the archaeological field is divided. A number of paleoanthropologists believe that modern humans originated in Africa, from which they spread to the rest of the world beginning about 100 kya. This is in agreement with the genetic data. A fairly large number of anthropologists reserve their opinion. Another group believes that the evolution of *Homo sapiens*, and perhaps even its predecessor *H. erectus* proceeded in parallel all over the Old World, and there was no expansion from Africa. The mitochondrial data are, at this point, the most useful in helping to reject this hypothesis, given that the origin of extant types of Asian mtDNA is more recent than this hypothesis would imply.

It is not yet possible, however, to exclude completely a partial participation of archaic *H. sapiens* from the Old World. New data and methods of analysis may help in this direction. What is very difficult to conceive is a parallel evolution over such a vast expanse of land, given the limited genetic exchange that could have occurred in earlier times. The capacity of the human genus to expand rapidly over a large fraction of the Earth's surface is more in tune with the idea of specific expansions from a nuclear area of origin. Such expansions must have been determined by some important advantage, biological or cultural. It is not difficult to accept the idea that the expansion of modern humans must have been strongly influenced by the possession of greater skills in communication by language. This increased ability to communicate is likely to have been extremely useful in favoring exploration and travel to unknown lands. Other technical improvements may have favored a trend to expansion. Although modern humans have now been found to have lived outside of Africa (in the Middle East) by about 100 kya, humans of this time in both Africa and the Middle East were biologically very similar to modern humans but culturally much less developed than at the time the real expansion began, perhaps 50 or 60 ky later. Many things may have happened in the meantime, in terms of cultural maturation and, perhaps, forward and backward movements between Africa and West Asia. Neanderthals are found in the Middle East after the earliest local appearance of modern humans in the same areas, and it has been suggested that they may have gained, or regained, lost ground in that period. The time between 100 and 50 kya (or, perhaps more exactly, between 90 and 60 kya) is currently a blank from an archaeological point of view. We hope that new discoveries will illuminate it. At the moment, the indications are that at the end of the blank period modern humans emerged with a new stone technology and started a radiation that took them to Europe, Australia and New Guinea, and America. Whether they partially mixed with or totally supplanted earlier inhabitants—for example, Neanderthals in Europe and archaic *H. sapiens* in East Asia—is difficult to state precisely on the basis of present knowledge.

Linguistic and cultural diversity increased conspicuously after that time, and the major linguistic families probably began less than 50 kya. Most of them are between 25 and 5 ky old. Genetic dating of linguistic families can only be approximate, but it agrees with ideas expressed by a few linguists. Moreover, the archaeological record shows increasing diversification, probably parallel with that of language.

An unsolved problem is determining the route by which the East was reached. Differences between East Asia and Southeast Asia make it reasonable to hypothesize that there might have been two routes, one through Central Asia and one through South Asia. Very little, if any, evidence of them exists today (fig. 2.15.1). The occupation of Australia and New Guinea was the major success story of the southern route, but it eventually led to an evolutionary cul-de-sac, as the separation between Oceania and Southeast Asia increased with the rising of the sea levels in the times after the last glaciation. It was only with the development of new nautical skills, 5000–6000 years ago, by South-East Asian populations who were also good farmers, that the Pacific routes were increasingly used. In the last 3000–3500 years, the expansion that generated the colonization of Polynesia began, most probably originating in a nuclear area in Southeast Asia.

There are two weaknesses in the present analysis, which will certainly require future work. One of them is the very short branch linking Caucasoids and, in particular, Europeans to the phylogenetic tree. One hypothesis is that they might have originated from an admixture between their southwestern and northeastern neighbors, Africans and Mongoloids, between which Europeans are sandwiched. One cannot completely exclude other hypotheses. Particularly serious is the possible bias resulting from the fact that almost all known genetic polymorphisms have been detected in Europeans. It will be important to remove this bias, especially in future data collections. Another area of doubt is the relationship between New Guineans + Australians, Southeast Asians and Northeast Asians. Our results have not settled this question unequivocally. It seems likely that the uncertainty arises because Southeast Asia is poorly known and may be heterogeneous, with some populations having an important genetic component in common with northern Mongoloids and others with people from Oceania. The

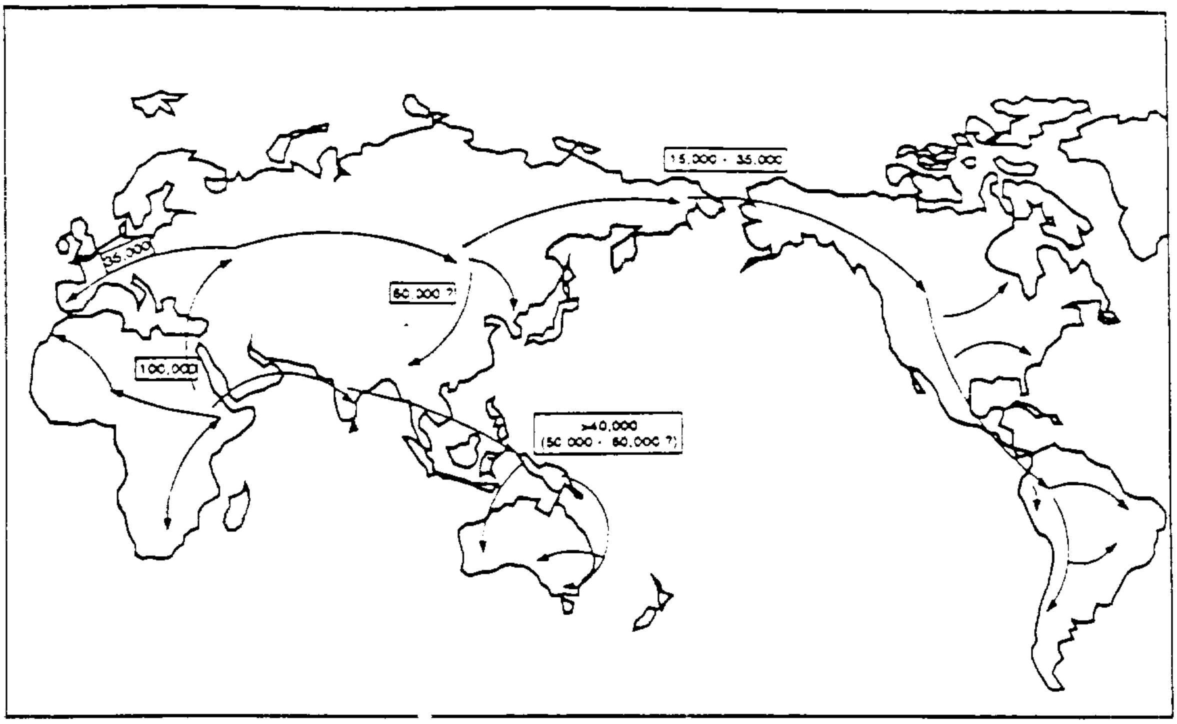


Fig. 2.15.1 Possible history and routes of expansion of modern humans in the last 100 ky.

heterogeneity may be in part due to ancient admixtures, and the arrow of northern Mongoloids pointing south in figure 2.15.1 express these considerations. There are also some undeniable physical similarities between northern and southern Mongoloids, leading one to wonder whether they have more in common than shown by the trees of sections 2.3 and 2.4. In other words, a fully dichotomous tree may be unsatisfactory in this part of the world, but more abundant and better evidence would be necessary for developing this explanation further.

The passage from Asia to America was later than that to Australia or Europe, perhaps because it first required a genetic and cultural adaptation to the more rigid climates of Northeast Asia. Genetic data, however, seem to agree with an early arrival, perhaps around 30 kya; possible uncertainties are discussed further in chapter 6.

Throughout the Paleolithic, population numbers remained small, leaving greater chance for random genetic drift to produce considerable diversification. Population size of a continental or subcontinental area at the beginning of expansion may have been on the order of 50,000–100,000 individuals. In the late Paleolithic, much of human action was in Asia, and the occupation of the rest of the world proceeded from this continent. Given the greater limitations on life in the north, Asia was like a relatively narrow, large landmass developed more in longitude than latitude. Because genetic divergence was subject more to random than selective forces, much of the gradient of the human gene pool goes from west to east. The first principal component therefore extends in this direction and explains 35% of the total human variation, showing only moderate, if any, influence of climatic factors at the level of the nuclear genes investigated, but a greater influence on genetic factors involved in the adaptation of body build and bodily surface characteristics, which notoriously respond to climate. A dichotomy is thus observed between genetic data and observations based on the physical constitution, which is detectable also on modern and fossil bones. This explains the discrepancy between the evolutionary histories reconstructed from data on genes and on skulls (or, in general, anthropometric data).

Only in the last 10 ky, perhaps under increasing population pressure and climatic changes, did humans develop new food technologies, culminating in several different agricultural developments. These innovations caused the beginning of more rapid population growth, and in some cases of local expansions, which extended to ecologically similar areas, allowing the exploitation of domesticated plants and animals developed in the three major nuclear areas of agriculture. The consequent increases of population densities began a progressive freezing of drift effects. Farmers' expansions, followed by those of nomadic pastoralists, contributed in an important way to changing the patterns of gene geography. In spite of this, opportunities still remained for the survival of much local diversity, especially in refugia, few of which have been well examined.

A major conclusion is that linguistic and genetic evolution are closely related. In this chapter we have seen this relationship at the global level, but several investigations on specific regions or people that we examine in the following chapters have given similar results. The main reason for the relationship is that the evolution of both depend on the same historic and geographic factors. We have seen that discrepancies are not impossible, given that genes can be partially or even almost completely replaced under certain conditions, and languages can also be replaced. Language replacement is more likely to happen, perhaps, in recent history, and there are well-known examples of it. One can also express the necessity of a relationship between genetic and linguistic evolution (and, more generally, certain types of cultural evolution of which the evolution of language is a key example), considering the similarity of the relevant mechanisms of transmission. Genes are clearly transmitted from parents to children; in traditional societies, especially in the absence of schools, cultural transmission (unfortunately a poorly investigated subject; see, however, Cavalli-Sforza and Feldman 1981; Hewlett and Cavalli-Sforza 1986) also takes place mostly from parents to children, as does, presumably, the transmission of language from generation to generation. Two phenomena transmitted in basically the same way are bound to be strongly correlated.

In our original paper (Cavalli-Sforza et al. 1988), we expressed the strong conviction that language must have been a great asset that considerably helped modern humans in their expansion, and that it also may have limited or prevented admixture with other forms of humans that were less developed linguistically. The linguistic inferiority of the Neanderthals (Lieberman 1975, 1989) is controversial (Falk 1975). Nevertheless, the extreme complexity shared by all existing human languages seems likely to be a product of a final step in linguistic evolution, which peaked in a.m.h. and was spread by them to the whole world. An interesting relationship has been observed by Foley (1991). Using the genetic tree (Cavalli-Sforza et al. 1988) and information on the numbers of languages per family given by Ruhlen (1987), he has shown that there is a very strong linear relation (r = 0.91) between genetic distance between two groups separated by a node of the genetic tree and the number of languages spoken by the two groups together (see fig. 2.15.2). Although this evidence is indirect, and the correlation coefficient is biased upwards because the nodes of the tree are not independent, it adds to the persuasion that linguistic evolution goes hand-in-hand with the spread of modern humans.

The analysis of the genetics of human populations requires an enormous mass of information. Unfortunately,

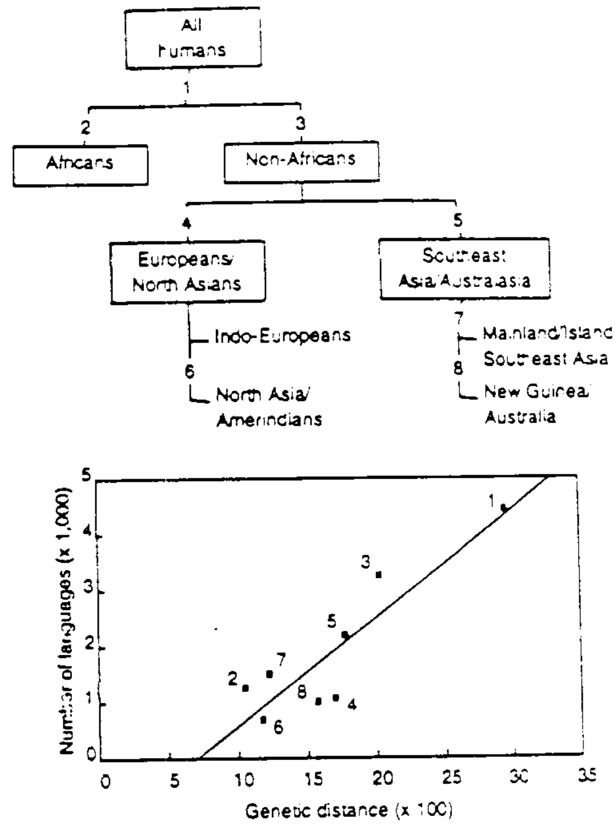


Fig. 2.15.2 According to Foley (1991), there is a strong relationship between human genetic variation and linguistic groups, showing that genes and languages may well diverge in similar ways. If genetic diversity found in each of the successive major groupings (top) of living human populations is plotted against the number of languages spoken by those groups, a strong linear relationship is obtained (bottom).

its retrieval has rarely been organized in an efficient way, and the data base available is the result of thousands of more or less haphazard collections and analyses of blood samples. An essential requirement of a sound analysis is that a large number of genes be thoroughly studied in parallel on all populations of interest. Today, there have been substantial advances in the techniques of analysis, unfortunately accompanied by nontrivial cost increases. The number of populations that can enlighten us on the past history of humanity is shrinking continuously. Only perhaps one or two decades remain in which we still have access to these populations. From the point of view of genetic history, we are an endangered species, and it is essential to avoid delay before taking the necessary steps to preserve this important knowledge about ourselves.

6 AMERICA

6.1. GEOGRAPHY AND ENVIRONMENT

The Americas, North and South, form 16% and 12% of the Earth's surface, respectively, and their cumulative area is slightly less than that of the largest continent, Asia, which comprises 29% of the Earth's surface. But today's total population of the Americas is only about 14% (including nonaborigines) of the inhabitants of the world, less than a quarter that of Asia (which is 60%). At the time of discovery, the population level was comparatively much lower, but is not precisely known. At that time, important population densities existed only in Mexico and in the northern and central Andes. Three major demographic changes took place after discovery (McEvedy and Jones 1978). The native population decreased practically everywhere and is now about 5% of the total population (much less in North America); it also underwent considerable admixture in many areas, and the mestizo population may be almost 20%. White immigrants and their descendants became the absolute majority of the population in North America (the United States and Canada) and in the southern part of South America. African slaves were imported for work on the plantations starting in 1650 and grew in numbers in most cases, especially in Brazil. Descendants of slaves now represent 15%–20% of the American population globally, an estimate made very imprecise by the extensive hybridization that took place. As usual, we confine our attention to the native populations living in the Americas before 1492, and begin by describing the environment.

North America. Two chains of mountains of very unequal altitude run along the eastern and western coasts of North America: the Appalachians in the east have been considerably flattened by erosion, whereas the Cordilleras in the west reach altitudes of 6194 m. They extend from Alaska to Mexico and in the region of their maximum width, near the fortieth parallel, they occupy about one-third of the surface of the continent.

The rest of North America is relatively flat: the central shield in the middle is 1400 feet (427 m) high, on the average, but it descends in altitude both in the north toward Hudson Bay and in the south and southeast, toward the lowlands and the Great Lakes region. The northern parts of the lowlands have been marked by moraines accumulated in four major glacial advances; the southern part remained ice free and was molded by rivers, of which the Mississippi is the most important.

Because the continent spans latitudes from 65° to a few degrees above the equator, climate and vegetation are very diverse. The Arctic is mostly a cold desert, with only two months in which temperatures exceed the freezing point. Below the Arctic, in southern Canada, the climate is temperate and cool with frosty winters, short springs, and moderately humid and warm summers. The continental United States has cold to mild winters, depending on latitude, and hot summers with ample rainfall. The western United States is very dry except on the coast, which enjoys, especially in its southern part, a Mediterranean climate. Central America has little variation in temperature with the seasons and has a mild climate with abundant precipitation, except in central areas, which can be very dry.

Two-thirds of North America was once forested, the type of trees depending on temperature and humidity. The rest of the continent is drier, with grassland or desert. In the Great Plains of North America, tallgrass prairies formed the habitat of the bison (often called buffalo) for

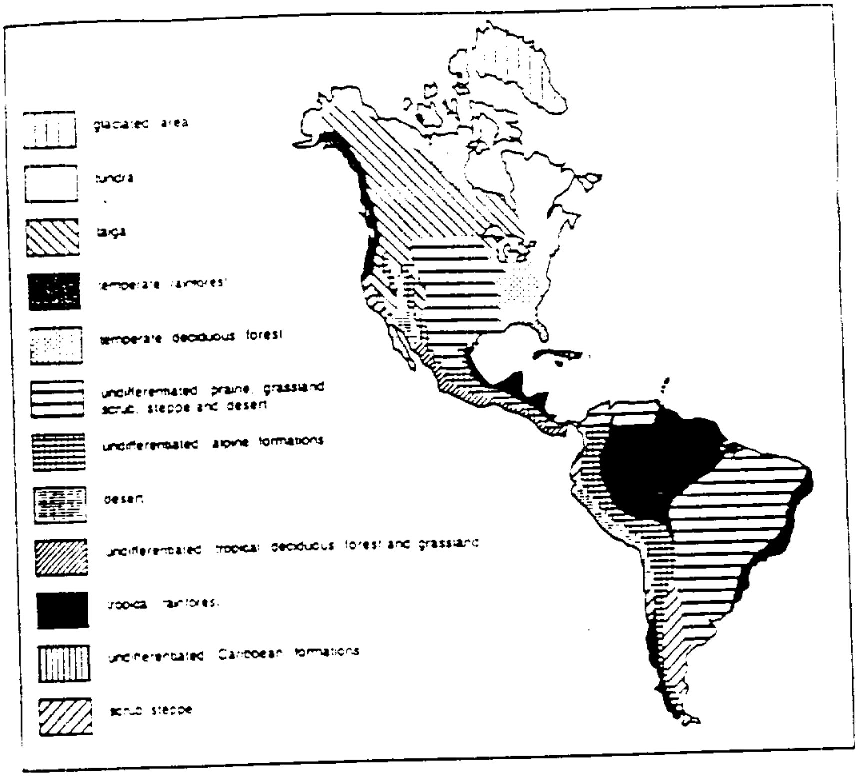


Fig. 6.1.1 Vegetation zones in America (Jennings 1983).

many millennia. Tropical savannas are found almost only in parts of Central America; the northern area, however, is mostly desert, whereas tropical forest is extensive in the southern lowlands. The map of vegetation illustrates the climate and ecological conditions (fig. 6.1.1).

South America. To some extent, South America is a mirror image of North America. Here too the western mountains border the Pacific and reach astounding heights; they go from the extreme north to the extreme south and are wider in the middle. Old, flattened highlands occur in the east in northern Guiana and in southern Brazil. Between these highlands is a very wide lowland, the Amazon basin. The Amazon basin occupies all the northeastern part of the continent and is covered by tropical rain forest, having very abundant precipitation and little change in rain or temperature throughout the year. A relatively small fraction, about 10% of the basin, is excellent for agriculture ("varzea") because it is flooded yearly when the rivers are high, but is not continuously submerged, so that it is naturally fertilized every year, but the rest ("terra firme") lends itself less to agriculture. Where the precipitation is not so heavy, the temperatures are higher and the seasons change, generating tropical savannas common to the Orinoco basin, just northwest of the Amazon in the Brazilian plateau. Farther south is dry forest; and still more to the south lies the basin of another great river, the Parana. Major grassland areas are the Pampas of northern Argentina; farther south lies the Patagonia desert.

The Andes vary in climate and flora, depending on altitude and local conditions, from tropical forest to grasses and plants of small and medium height ("paramos"), to steppe ("puna") that reach the snow line. The extreme south, at a latitude of 56°, has glaciers and mountains, and a frigid climate.

DOI 02180

6.2. Prehistory: Occupation of America

The prehistory of America is shorter than that of any other continent, and its beginnings are more obscure despite enormous interest among scientists who have contributed to the research. Thus, there is considerable uncertainty regarding the origins of native Americans and, as is often the case, uncertainty generates discussion to the point of passion.

There is essential agreement on the idea that the peopling of Americas took place with the passage of nomadic Siberian hunters from Northeast Asia to Alaska (Fagan 1987). Other hypotheses have posited extraordinary journeys—for instance, from Africa to America or from America to Polynesia—but they are not supported by hard evidence (Bellwood 1979). One problem, however, is that among the oldest sites those that are less in dispute (but certainly not entirely accepted) are in South America. Moreover, there are only a few Siberian sites that may have been inhabited by pioneers who later occupied North America. Well-established Siberian sites are more recent than the oldest American sites, which are few and difficult to date. The oldest American sites are not accepted by some archaeologists, whom others accuse of maintaining unreasonably high standards (Bray 1988). Briefly stated, there is strong disagreement between archaeologists who believe that the earliest entry into North America was 30–35 kya (there have even been claims of earlier sites), and those who are prepared to accept, on the basis of present evidence a first date of entry of 15 kya. We briefly review here some of the major finds that are generally accepted and indicate the major controversies.

There is substantial agreement on the lack of evidence of archaic *Homo sapiens* or earlier types in America. All widely accepted American site dates follow the disappearance of Neanderthals in Europe and in Northeast Asia, and there are no finds supporting the migration to America of human types preceding anatomically modern humans (a.m.h.).

The last glaciation occurred 30–13 kya, with a peak at 18 ky; the geography and environment of America and northern Asia when the migration from Siberia to America is believed to have taken place was different from today. In late glacial times (fig. 6.2.1), glaciers occupied almost all of Canada and part of the north-central United States. Temperate and tropical climates were found in North America at much lower latitudes than at present. The tropical forest had a somewhat smaller extension, especially in South America.

An ice-free corridor is believed to have existed between the eastern edge of the Rockies and the immense glaciers occupying the central and eastern parts of Canada, but the environmental conditions were undoubtedly fairly frigid in the corridor. Perhaps more importantly, at the presumed time of the crossing, the coast line was lower, due to water being retained in the polar ice. This exposed the continental shelf along the coast, causing the temporary disappearance of the Bering strait. A wide and flat land bridge, Beringia, replaced the strait connecting Asia and America, and is believed to have existed between 25 and 15 kya. It is not completely clear what the conditions for life were on Beringia; it was probably a largely treeless land with grasses, dwarf birch, and shrubs, a mosaic of steppe and tundra. It was cold and dry with strong winter winds. Nevertheless, there were mammoth, bison, horse, antelope, and smaller animals (Fagan 1987; Schweger 1990). Certainly the land bridge favored passage between the continents. Without it, the passage would have had to have been made by boat, but direct archaeological evidence of passage by water is difficult to find and, in this case, has not been discovered.

Conditions that permitted crossing from Asia to America by land existed for some time and may have favored the passage of different groups in different periods, some by land and some along the coast. The climate in Beringia was probably never too attractive, although perhaps not very different from that of the Siberian regions of origin, and it may have served as an incentive

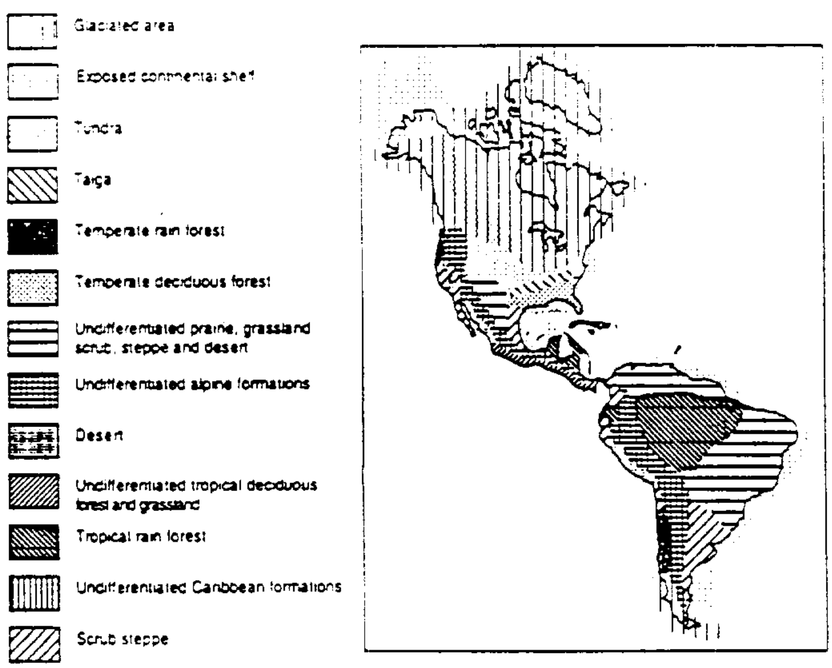
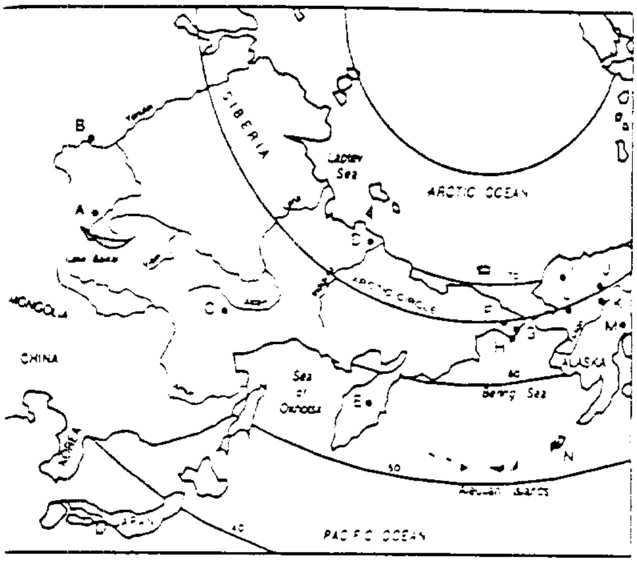


Fig. 6.2.1 Glacial environment in the Americas about 15 kya (Jennings 1983).

to continue migration in an eastern and, finally, southern direction.

Several Siberian sites could have been homes of the ancestors of the early Americans.

1. About 20 kya, in Mal'ta and Afontova, in southern Siberia (see fig. 6.2.2), there lived mammoth and reindeer hunters similar to the mammoth hunters of the western Russian steppes north of the Black Sea; among the latter, the best known lived at Mezirich on the Dnieper 18–14 kya (Fagan 1987). Some of their tools are similar to the "microblades" made in Northeast Asia at that time.



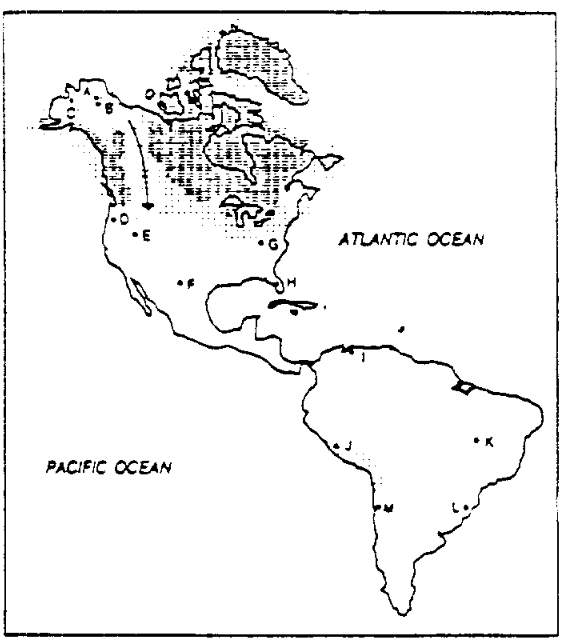
A. Mal'ta
B. Afontova
C. Dyukhtai
D. Berelekh
E. Ushki Lake
F. [illegible]
G. Putarak
H. [illegible]
I. Akmak
J. Old Crow
K. Bluefish Caves
L. Dry Creek
M. Denali Complex
N. Anangula

Fig. 6.2.2 Archaeological sites in Paleolithic Siberia and Alaska (Fagan 1987; Dikov 1988).

2. At the cave situated near Dyukhtai (also spelled Diuktai), near the Aldan River, an affluent of the Lena, a culture was found that was dated at 14–12 kya. By 14 kya, this culture had already spread even farther north, up to the Arctic Ocean where a mammoth burying ground was found at 71° latitude in Berelekh. The discoverer believes its beginning to be earlier and traces the origin of these people to northern China. The Diuktai people used microblades but, unlike Mal'ta people, also made bifacial tools (Fagan 1987). Microblades were used for inset tools and appeared in northern China 30–15 kya; they became common in Japan and perhaps Korea in the later part of this period.

3. A third site is Ushki Lake in Kamchatka where the oldest dates are around 14,000 B.C. The early Ushki cultures used stone-tipped spears, perhaps bows and arrows. The late Ushki culture (12,000–10,000 B.C.) is similar to the Diuktai culture, but more advanced, and has peculiarities of its own. A burial of a husky dated to 11 kya is the oldest northern find of a domesticated dog and may have been connected with the use of dog sleds. Many sites farther north on the Chuckchi peninsula (see fig. 6.2.2) seem to belong to the late Ushki culture and show some intermediacy with Alaskan sites (Dikov 1988).

The earliest archaeological scenario in North America includes sites in central Alaska (fig. 6.2.2) and others in the continental United States (in the parts that were not glaciated at the time) and Mexico (figs. 6.2.2, 6.2.3). Tool finds at Old Crow Flats in the northern



Ice Sheet
ca. 18 kya
ca. 12 kya

A. Old Crow
B. Bluefish Caves
C. Dry Creek
D. Fort Rock Cave
E. Wilson Butte Cave
F. Clovis
G. Meadowcroft Rockshelter
H. Little Salt Spring
I. El Jobo
J. Pikimachay
K. Pedra Furada
L. Alice Boer
M. Monte Verde

Fig. 6.2.3 Paleo-Indian sites in America (Fagan 1987; Guidon 1987).

Yukon (Canada) near the Alaskan border are undoubtedly human, but the date of 27 ky claimed for them is disputed because it comes from animal bones, and there is no consensus that they were "modified" by humans. A human artifact made of bone that had an older date has been redated to 13.9 kya. Another site close to Old Crow, Blue Fish Cave, has bones with dates of 15.5 kya and 12 kya, in addition to human artifacts, including microblades similar to those of the Diuktai caves. A lower layer at Blue Fish has broken bones dated 8,000–10,000 years earlier, but signs of human occupation are not as clear as for the later layers.

Many sites in Alaska have been dated to 12–10 kya: they contain bifaces and/or microblades reminiscent of the Siberian cultures (Denali complex, Dry Creek, Akmak). In summary, there is no evidence on which agreement has been reached that Alaska was occupied by humans before 15 kya.

In the central United States, there was a major explosion of archaeological finds marked by projectile points named after the *Clovis* site, which is dated to the period 11.5–11 kya. There are, however, several finds older than Clovis. A conservative analysis by Fagan (1987) lists a few places that are pre-Clovis and in his view more satisfactory (see location of sites in fig. 6.2.3): Fort Rock Cave, Oregon—13.25 kya; Wilson Butte Cave, Idaho—13–14.5 kya; Meadowcroft Rockshelter, Pennsylvania—more than 12 kya (up to 16.175 kya; Adovasio et al. 1982); Little Salt Spring, Florida—12 kya.

According to MacNeish (1978), Mexico has dates greater than 30 kya associated with chopping-chopper tools, followed by a phase 30–15 kya with bone tools and a unifacial industry (see criticism in Fagan 1987). Other archaeologists also believe dates earlier than 15 kya for Mexico and South America (see also Lynch 1990). Here we cite four major examples of early dates for South American sites.

The Pikimachay Caves in Peru have a more reliable later occupation at 14 kya and an older one at 20 kya considered less reliable.

Dates of 14,200±1150 at Alice Boer Site in south-central Brazil are more reliable than those of earlier tools from a lower layer at the same site, dated to 20–40 kya. Pedra Furada in the northeastern Brazilian plateau (Guidon 1987) has yielded various layers with signs of human occupation, the oldest of which was dated to 32 kya. Monteverde (south-central Chile) is an open settlement with excellent conservation. The people there were mammoth hunters living 12–14 kya.

It is difficult for nonarchaeologists to form a final opinion at this stage, but wide disagreement obviously exists among specialists. It is understandable that there is little tendency to rely on radiocarbon dates especially if they are unique, have high standard error, or come from samples that could have been contaminated with older material. Other often-cited objections are that the stratification is imperfect, or human occupation and use of implements uncertain. The lack of evidence for early, and totally satisfactory, sites in North America is clearly one of the motives for the resistance to accepting sites anterior to 14 kya or 15 kya years in Central and South America. The idea that there is too short an interval of time between occupation of Alaska and that of South America is not a major obstacle, since nomadic hunters could well have covered distances of many thousands of miles in a period of 1000 years. In fact, the whole journey from the extreme north to the extreme south might have taken about that long (Martin 1973). The problems that arise from accepting the hypothesis of this extremely rapid displacement are of two kinds: the hunters had little time to adjust to new environments if they moved so quickly from north to south across such a wide and diverse continent, and they must have reproduced at a high rate in order not to dilute themselves too much in the race toward the south. Approximate calculations indicate, however, that the hypothesis of rapid movement is not unacceptable (Cavalli-Sforza 1986b). Models of genetic consequences of this rapid advance are discussed in the last section of the chapter. The problem of adaptation to new environments must have been simplified by the availability of the same prey (mammoth, mastodon, and probably others) throughout the continent. The idea that South America was occupied before the north, either from the Pacific or the Atlantic Ocean, is more difficult to accept. Whatever trace of African genes are found among living people, it is much more likely to have originated from admixture with African slaves after the sixteenth century. The Pacific islands closest to South America are quite far away and were occupied only very late, in the last two thousand years.

There is no problem with the essentials of the Clovis culture, which developed around 11.5 kya on the Great Plains of North America and lasted for about 500 years. It is marked by mammoth and bison butchering places, where bones of other animals are also occasionally found. The mammoths were killed with spears headed with projectiles that had characteristically fluted stone points and were given additional thrust by using spear-throwers (known as *atlatl*). This culture takes its name from Clovis, one of the important sites; it was supported by a scarce and scattered population. Its origin is uncertain; its end coincided with the disappearance of mammoths from the plains. Shortly thereafter, these animals disappeared from all of America along with several other large mammals that became extinct between 12 and 10 kya, including the mastodon (another elephant), the saber-toothed cat, the horse, several camels, giant sloths, and others (Grayson 1987). One large mammal that survived and was still flourishing on the Great Plains until a few hundred years ago is the bison.

The disappearance of the big mammals has received different interpretations. Martin (1973) suggested that it was due to overkill that started in North America and was continued in South America by hunters that occupied the whole of America in pursuit of this prey. This hypothesis, however suggestive, is certainly simplistic. Pleistocene overkill has been advanced as an explanation for many similar extinctions that happened at about this time in many parts of the world. Although overhunting may have been a partial cause, it seems likely that the change of climate in the postglacial period also had a strong impact by causing profound ecological alterations. Evidence that it affected the fauna comes from the observation that large extinctions of birds also occurred at the same time, whereas small mammals survived and changed their range. Moving to other, more acceptable environments was certainly a mode of adaptation to climatic change (Grayson 1987) that was not equally open to large animals. The bison, however, could survive because it was not bound by its digestive system to eat only the tall grass of the archaic prairies but also the short grass that replaced it in postglacial times. After the disappearance of the mammoths, bison-hunting became the major source of food and other commodities (bones, hides, etc.). Weapons changed somewhat, and new projectile points were developed from the Clovis points. There was some slow evolution in the hunting techniques, but in the Plains the bison remained the major source of food for millennia. Only the introduction of the horse and the gun after the Spanish conquest in the early sixteenth century generated a dramatic change. The bison then came very close to extinction and was saved only by protection in government reserves at the beginning of this century.

Whatever the first date of entry, between 35 and 15 kya, it is clear that there was more than one migration. The linguistic and biological evidence is discussed in sections 6.8 and 6.9–6.13.

The oldest migration from Siberia was that of the Paleo-Indians, to which the above discussion refers, and led to the peopling of the entire continent. There may have been a series of migrational waves, not simply one, or there may even have been a continuous flow. The other two migrations were both later and led to the occupation of more limited and well-defined areas in the north.

Another migration, presumably a second one (15–10 kya) is named after the Na-Dene family of languages spoken by these people. They settled in southern Alaska and on the northwestern coast of North America, perhaps only a little later than the Paleo-Indians. Much more recently, at the beginning of the present millennium, some Na-Dene groups migrated farther south.

The third migration was that of the Eskimo-Aleut (ca. 10 kya), who kept to their Arctic and sub-Arctic habitats, with the Aleuts occupying the Aleutian islands and the Eskimos occupying Alaska and the northern coast of North America, spreading later as far as Greenland. There are still a few Eskimos in the extreme northeast of the USSR, but Siberian Eskimos are believed to have reentered Asia from the Americas and should not be considered, therefore, an aboriginal Asian group.

The original Asian locations of the Na-Dene and Eskimo-Aleut are not completely clear but are perhaps easier to fit into the general archaeological picture than that of Paleo-Indians, for whom the uncertainty of the time of origin (35–15 kya) is likely to be with us somewhat longer. It is possible that Na-Dene and Eskimo-Aleuts had common origins in Asia.

Dikov (1988) has suggested that the late Ushki culture, dated 10–12 kya and located on the eastern coast of the Kamchatka peninsula, shows similarities with cultures of Alaska and British Columbia and may have contributed to the Eskimo or the Na-Dene populations or both. Dikov also discovered a culture on the southeastern Chukchi peninsula at Puturak Pass, in close proximity to the Bering Sea, that has a technology different from other Asian cultures and similar to that of the Gallagher Flint station in the Brooks Range of northern Alaska. It is dated to 10,549±150 years ago and also has similarities with the culture of Anangula (fig. 6.2.2), a small island in the Aleutians near Umnak Island. The Anangula culture is the oldest known in the Aleutians (dated at 8.7 kya). Laughlin (1980) suggested that Eskimos and Aleuts both come from Anangula and that the occupation of the Aleutian Islands began from it, proceeding both westward and eastward from there. The earliest occupancy of the western and eastern ends of the chain of islands is currently dated to 3000 years ago, but the most interesting early sites of these fishermen and sea-mammal hunters may be submerged. The first known date of occupation of Anangula has also been suggested (Laughlin 1980) as the date of separation of Aleuts and Eskimos. Fagan (1987) indicated more conservatively a date before 4000 years ago. While Aleuts remained on the islands that carry their name and mostly maintained their primary skills in hunting sea mammals, Eskimos developed transportation skills across the Arctic and hunted not only sea but also land mammals (musk ox and caribou). The Dorset culture (Jennings 1983) ranged from the Northwestern Territory in Canada to the Hudson Bay, Labrador, Newfoundland, and Greenland by 1000 B.C., on the average, but there are signs of earlier occupancy of these regions by a pre-Dorset culture.

The difference in origin of Na-Dene and Eskimo-Aleut remains to be clarified. The coast of the Pacific Northwest was colonized by Na-Dene speakers, but the exact time sequence is not clear. Queen Charlotte Island, off the coast of British Columbia, was continuously inhabited between 7000 and 5000 years ago, but the area may have been occupied earlier. The populations of the northwest coast developed a special way of life, reaching high densities especially at the mouths of rivers where salmon was easy to catch. Their cultures at the time of European contact allow us to place them among the world's most successful foragers, and they were the subject of classical research in cultural anthropology.

The eastern coast of Greenland was settled by Vikings coming from Norway and Iceland in the ninth or tenth century A.D., but the Viking settlement lost contact with Europe and disappeared in the fifteenth century. Perhaps in that early time, and probably later after the Danes settled in Greenland (beginning in A.D. 1721), there was some degree of admixture with people of European origin.

In summary, there is little agreement about the first occupation of the Americas; possible dates vary from 35 to 15 kya. There is agreement that this first migration came from Siberia via Beringia and was followed by the rapid occupation of the whole continent by "Paleo-Indians." The next settlement, on the northwestern coast of North America, was between 15 and 10 kya and is attributed to Na-Dene-speaking people. The third, around 10 kya or later, led to the occupation of the Arctic coast by Eskimos. The three-migrations theory has been proposed by Greenberg et al. (1986; see also Greenberg and Ruhlen (1992). It is based on linguistic, dental, and genetic information, as we shall see in the rest of this chapter. A group of linguists (sec. 6.8) vigorously opposes the interpretation of linguistic data proposed by Greenberg (see Ruhlen 1987, 1991, also Rass 199_, Wright 1991).

Other useful references are Kirk and Szathmary (1985), Aikens (1990), and Ruhlen (1990), as well as chapters 54 and 55 of the Cambridge Encyclopedia of Archaeology.

6.3. BEGINNINGS OF AGRICULTURE

The development of human populations was very unequal in the various regions of Americas. The Paleo-Indian hunters occupied the continent with extraordinary rapidity; there later developed local hunting traditions that lasted for millennia in some areas, though inevitably with more or less continuous cultural changes and people displacements. The post-Paleo-Indian period is often called the *Archaic Period* or later hunting-and-foraging period.

The transition to food production from the foraging economy—that is, the hunting-gathering and, near the water, the fishing economy—is sometimes called the *Formative Period;* it occurred at very different times and in different ways in the various regions. In the periods preceding agriculture or in its early development, population density increased somewhat, a stimulus to technological advance in food production. The development of domesticated plants and animals and their adoption as staple food was always a relatively slow process, especially in the Americas, for reasons that depend in part on geography and in part on the nature of the domesticates themselves. Compared with Europe and East Asia, diffusion of agriculture to neighboring regions was slower and more limited. Therefore, at the time of European contact, plants had been cultivated for almost 10 millennia in areas like Mexico and the western part of South America, where important empires with large populations had developed. In many other regions, however, large numbers of American natives were still hunter-gatherers. This was true in particular of the Northwest coast North American Indians, the Na-Dene, and of Californians; but in both regions relatively high population densities had been reached at the time of contact and complex societies had developed, especially among the Na-Dene. The density and, according to some, social complexity of these hunter-gatherers were greater than in other parts of North America that offered only marginal resources and where agriculture, even if it had been adopted as a partial source of food, had only limited development.

The beginnings of agriculture in America are perhaps slightly later than those in the Middle East and in China. By the year 9000 B.P., Middle Eastern agriculture was already a complex economic system using both animal and plant domesticates that could be exported to nearby regions with a somewhat similar ecology. Initial developments in Mexico and the northern and central Andes took place in an environment and with domesticates not widely represented outside the original area. Few if any of the original crops had the potential of being easily exported to a wide area around that of origin before being more fully developed, unlike the Middle Eastern domesticates of wheat, barley, sheep, goats, and cattle. In addition, agriculture in America began in areas like central Mexico and the western part of South America (mostly Ecuador and Peru), which were to some extent unique or isolated. The Mexican plateau enjoyed a temperate climate not found in much drier northern Mexico nor in the tropical forest of the southern part of Central America.

The Andes were another unique environment in which extreme differences in altitude at a short distance provided a great variety of small niches, each suitable for very different types of economic activities. In time, this variety was cleverly used by what is called a "vertical pattern" of exploitation, namely by foraging, cultivating, or breeding very different plants and animals at different altitudes, often very close together, and exchanging these products by a complex network of trade and communications. Systems of seasonal migrations also developed, similar but not entirely comparable to "transhumance" in the Old World. It took time, however, before the social and political conditions of these populations were such that the extraordinary variety of available environments could be turned into a source of wealth.

DOI 02185

Native Americans developed a great number of domesticated plants for a variety of uses (Pickersgill and Heiser 1977). Many of them, like maize, potatoes, and tomatoes, were exported to Europe after their discovery in the New World and acquired primary importance as staple food in the Old World. Other American plants like manioc were exported to tropical Africa and radically altered the local food customs. The first plant domesticated in America may have been the bottle gourd (Lathrap 1977), at least 9500 but possibly 11,000 years ago, because of its usefulness as a water container. Maize was domesticated from local plants in Mexico at Tehuacán and Tamaulipas around 9500 years ago, but initially—and for many thousands of years—it remained a small component of the diet. Originally, maize cobs were one-tenth or less the size of modern cobs. Cob size grew with remarkable regularity over the millennia, presumably because of artificial selection exercised consciously or subconsciously by the breeders, who may have been systematically choosing the best cobs for reproduction. At the time of the Spanish conquest of Mexico, agriculture formed an important part of the food supply, which was augmented by the products of hunting and gathering. It is more or less arbitrarily assumed that agriculture became a major source of food supply at a "critical" time about 4000 years ago. At that time, the yield of maize was sufficient to support a sedentary population; pottery made its first appearance then, much later than in Europe and Asia, and almost certainly independently. Beans were also domesticated early in Mexico, with the first examples 9000–10,500 years old; they are a good complement to the maize diet because they supply essential amino acids deficient in maize. Squash was soon added to maize and beans, forming the American Indian triad of staple foods famous for being nutritionally well balanced. Potatoes probably came from Colombia (10 kya). Cotton was grown for use as a textile. Most of these crops could not grow in tropical environments, such as the lowlands of South America, where instead manioc was first domesticated. It later spread to other areas of tropical forest outside the continent.

Few animals were domesticated; however, the use of dog meat for food may be 6000 years old. The turkey is first found in Mexico from 300 B.C. In the central Andes considerable use was made of domestic camelids (llamas, alpacas), which became increasingly common in the last 8000 years for transportation and meat. Guinea pigs were domesticated in Colombia and Peru for meat probably in the last 4000 years. Figure 6.3.1 shows the sites of earliest domestication in America (Bray 1980).

At the time of European contact, American natives were still in the stone age; the only widely used metals, gold and silver, had almost entirely ornamental applications. Some native copper was used for weapons and ornaments. Even so, at the time of contact, two major empires with large populations had developed in Mexico

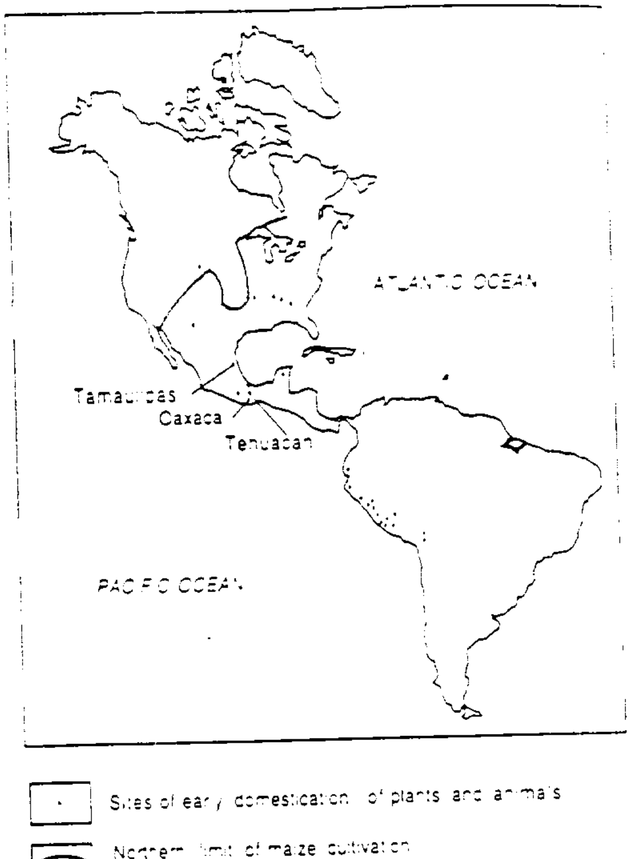


Fig. 6.3.1 Distribution of probable places of early domestication (Bray 1980).

and Peru. Elsewhere, population density was still low, although it had increased in the last millennia over the very low densities characteristic of the initial period. The high mobility of the Paleo-Indians allowed them to occupy the whole continent rapidly, but later population growth was slow until the last two or three millennia and increased almost exclusively in areas where previous important agricultural development had occurred. The number of American aboriginals at the time of contact is very imprecisely known and varies greatly with the authors. Early estimates by Kroeber (1939) and Mooney (1928) (whose estimates differ little from Kroeber's) give a total of 1.2 million for all of North America, of which the largest components (in thousands of individuals) come from California (260), Canada (190), the Gulf States (115), and the Plains (100). Later estimates are higher, up to 5 million for the United States (Russell 1987) and 300,000 or more in Canada (Charbonneau 1984). Meso-America was the most densely populated, with perhaps 6–25 million people (McEvedy and Jones 1978). For central Mexico, Cook and Borah (1971) suggested a population of almost 17 million in A.D. 1532, down to 6 million in A.D. 1548 and 1 million in 1608; but Zambardino (1980) corrected the 1548 estimate to

3.6 million. In Peru, the Spanish viceroy estimated 1.3 million in 1572, down to 600,000 in 1620 (Sanchez-Albornoz 1977).

The uncertainty of these estimates should not be surprising. Censuses are difficult even under optimal conditions; at the time of conquest, they were rare even in Europe and there was neither enough interest nor technical skills for carrying them out in the colonies, the occupation of which remained incomplete for a long time. In any case, qualitative evidence shows that population numbers declined rapidly after the conquest, with the spread of epidemics brought by the conquerors and the destruction of the preexisting civilizations. Later censuses are therefore of little use.

Aboriginal population densities largely reflected the degree of development of agriculture and social organization, being higher where the history of agricultural development was older. An exception was North America, where the nonagricultural societies of the west had relatively high densities because of exceptionally favorable environments and advanced sociocultural adaptations. By contrast, central and eastern North America had only a short agricultural history and had not reached high densities at the time of contact.

As noted elsewhere (sec. 2.7, 4.7, 5.2), the onset of agriculture and its successive development is of considerable importance from the point of view of population genetics in that the transition from food collection to food production usually increased population density and thus generally decreased the effect of random genetic drift. It also altered the pattern of migration in many ways, usually reducing individual migration by causing the population to become more sedentary. But migration was always higher in early agricultural times because initial agriculture was of the shifting type (moving to new fields as soil fertility was falling or for other reasons) and in many areas still remains at this stage. Population saturation following initial growth is expected to cause centrifugal migration toward new unexploited fields, when these are available, setting in a slow wave of advance of the agricultural population toward less dense areas. The wave of advance of farmers could begin only when and where cultivated crops had become the major source of food, and where strong physical barriers like mountains or deserts did not impede migration. These conditions occurred relatively late in America, after the Formative Period and therefore after 4000 years B.P. It is unclear whether rapid increases of population density in America caused major demic expansions as they did in Europe or Africa. Mexican agriculture was born in the highlands and expanded late to the north, but it is not clear if there was a demic component; the northern Mexican desert must have acted as a buffer that slowed northward expansions. Agriculture probably spread from Mexico to the south, but there may well have been retrograde flow. The development that took place in Meso-America has much in common with that in the northern Andes. Dates are probably not known in enough detail to allow study of the spread to the south. Lathrap (1977), however, has given tentative dates and directions of expansion for Central and South America. The Andean type of economy was suitable for the particular environment of the Andes, and much of it remained confined to it. However, manioc cultivation, which had an enormous impact on tropical agriculture, may have originated in the forest near the central Andes. The natural way of communication in the South American plains was along rivers, and it is not surprising if spread in this natural network was fast (Lathrop 1977).

As a direct consequence of the economic history we have outlined—mainly the late and limited expansion of agriculture—and its highly localized development, population density remained low in most areas, and the social structure stayed fragmentary, leading to high genetic drift and, with it, high local variation. In the following sections, the development of various regions before and after agriculture is briefly outlined. Surveys of the subject and references can be found in Jennings (1983) and in chapter 56 of the Cambridge Encyclopedia of Archaeology.

6.4. Development in North America

Agriculture arrived late in North America from Mexico, and never reached the western coast during the precontact period. For a general overview of the pre- and post-agricultural development, it is convenient to distinguish four large areas: the West, the Southwest, the central region (the Plains), and the East.

1. The *West* includes for our purposes California, the Great Basin (Nevada and Utah), and the Plateau (Idaho, eastern Washington, and northeastern Oregon). Here, as elsewhere, the more immediate descendants of the Paleo-Indian hunters had to cope with an environment that was becoming warmer and drier. Seven thousand years ago, the climate was already similar to the modern one. But even by 9000 years ago, there was some evidence of a beginning of local differentiation of cultures. A substantial development of the foraging population, accompanied by a trend toward population increase, began only about 3000 years ago but 500 years later in the interior (Aikens 1983). It was once believed that the social system was extremely simple, especially in California, but this view is being corrected. Without increasing sophistication, they would not have eventually reached relatively

high density and local wealth. The foraging peoples in the West were highly sedentary, and there was systematic exchange and trade between local populations.

2. Agriculture from Mexico moved first to its nearest neighbor, the *Southwest*. Defined geographically in various ways, it usually includes Arizona, New Mexico, Colorado, and southern Utah. It is a very dry and almost desert area, but in the Archaic Period, and sometimes even during the Paleo-Indian Period in the eastern moiety of the Southwest, there developed cultures of foragers that lasted for millennia, until the beginning of a sedentary-horticultural mode of living in the Formative Period. The introduction of some cultigens from Mexico, like maize, may be as old as 3000 years B.P. or more; a safer date is 2500 B.P. (Lipe 1983). The beginning of a radically new culture (see fig. 6.4.1) is seen with the *Hohokam* culture in southern Arizona, starting about 2000 years ago. According to some, the Hohokam were migrants from northern Mexico; to others, they were local inhabitants who were under cultural Meso-American influence (Lipe 1983). They grew maize, beans, squash and cotton, made ceramics, and with irrigation were able to colonize a vast area. The Hohokam are believed to be ancestral to the Papago and Pima, who still live in the same general region. The case for continuity of culture from the Hokoham to the Pima-Papago is reasonably strong.

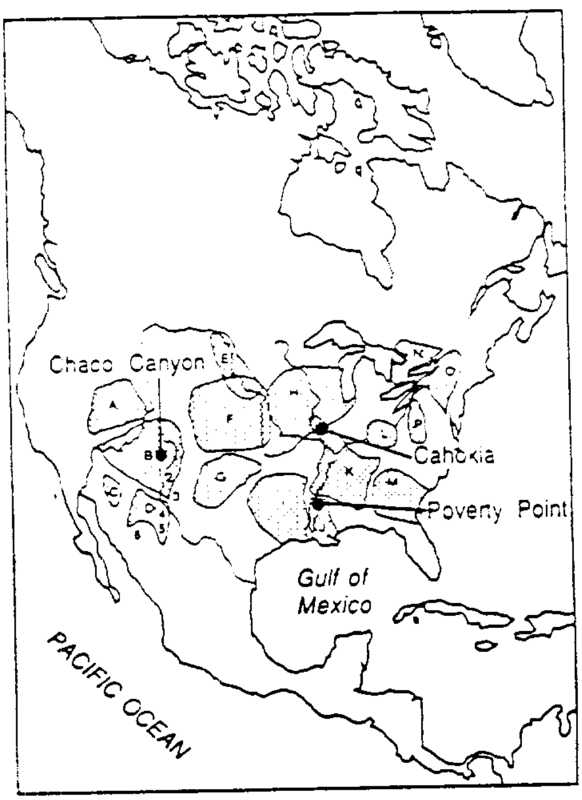


Fig. 6.4.1 Geographic location of the major North American agricultural groups (from Whitehouse and Whitehouse 1975; Griffin 1980). A, Fremont; B, Anasazi; C, Hohokam; D, Mogollon; E, Middle Missouri; F, Central plains; G, Southern plains; H, Onrota; I, Caddoan Mississippian; J, Plaquemine; K, Middle Mississippian; L, Fort Ancient; M, Appalachian Mississippian; N, Hurona; O, Iroquois; P, Monongahela. 1, Hopi; 2, Navaho; 3, Zuni, 4, Apache; 5, Pima; 6, Papago.

North of the Hohokam, the *Anasazi* culture may have developed directly from an earlier Archaic culture that lasted through the millennia (the Oshara), probably with the contribution of migrants. Maize, beans, and squash are well documented by A.D. 600, at which time the population, originally rather diffuse, began to collect in small separate settlements. Between A.D. 900 and 1100, large villages of Pueblo-type appear at Chaco Canyon in northwestern New Mexico. There were cycles in which large villages (Pueblos) were formed, then abandoned collectively when the population moved to other places, often to form larger pueblos. It is believed that the increase in village size made it possible to engage in irrigation works of greater magnitude. Conflicts with immigrants to the area like the Apache and Navajo were earlier believed to have been responsible for the movement of the pueblos, but it is now known that these Na-Dene speakers arrived in the area after A.D. 1200. Many new settlements were built and suddenly abandoned shortly thereafter, at dates that are accurately known thanks to the study of dendrochronology, the sequence of rings in trees. The reasons for movement are less clear. Among the current explanations for the abandonment of pueblos is the recent discovery of cooling and drying of the local climate around A.D. 1100, leading these people to search for areas more suitable for agriculture because more water was available. The descendants of the Anasazi are the modern Pueblo Indians (Hopi, Zuni, etc.; Lipe 1983).

Another culture, the *Mogollon*, started east of the Hohokam and at about the same time, reached its maximum extension around A.D. 900. It was eventually absorbed into the western Pueblo culture under the influence of the Anasazi. Other groups that developed a farming culture in the area, and that are not easily identified with modern descendants, include the Freemont in Utah, the most northern group in the Southwest.

3. Unlike the Southwest, which is dry, the *East* enjoys considerable rainfall, which favored the development of a rich flora and fauna. This area includes the valleys of two major rivers (the Mississippi and the Tennessee), the Appalachian region, and extends farther northeast. In the Paleo-Indian Period, the *Clovis* hunters were the dominant culture, followed by the *Dalton* culture, which clearly derives from the Clovis, but is adapted to a new target, deer. In Paleo-Indian and Archaic times the population was probably scarce and diffuse, made up of small mobile bands with no capacity for food storage.

The transition to a sedentary life was spread over a long period, and domestication of some native plants,

DOI 02188

like sunflower and amaranth, may have preceded the introduction of cultigens of Mexican origin. An innovation is the building of large mounds as at Poverty Point, Louisiana, with dates ranging from 1700 to 870 B.C. (Jennings 1983). The size of this mound (a diameter of 1200 m) indicates that a degree of social complexity had been reached that made it possible to build such monumental works. Smaller mounds, usually burials, are very common. Domesticated squash is known from the area, but could not have formed an important part of the food supply; maize came somewhat later. Pottery, rare in the beginning, was widespread by 700 B.C. The population clearly became more sedentary during this period, usually called the Hopewellian, but only later (A.D. 700–1000) did clear signs of shifting agriculture appear (the Mississippian period), still combined, as is usual in initial periods, with hunting and gathering. The principal site is Cahokia, near the Mississippi River, almost opposite St. Louis, Missouri. Production of maize and squash increased, and beans were added around A.D. 1000. Communities ranged in size from 100 to 1000, and the larger ones showed indications of social stratification, with chiefs or priests directing ceremonials, mound constructions, and agricultural operations. This culture, *Oneota*, spread north to north-western Illinois and southern Wisconsin after A.D. 1000 and had connections with other nearby cultures. Villages were often fortified (Jennings 1983).

4. Between the Southwest and the East are the *Plains*, which after the disappearance of the forest around 10,000 B.P. became a wide grassland occupied almost since the beginning by large herbivores, particularly bison. The numbers of bison fluctuated over the millennia; there are also fluctuations in the density of occupations and the archaeological record probably for the same reason. At Hell Gap, the archaeological complexes follow one another with few changes from 11,000 to 8000 years ago; after that time, a climate change may have set in. At Mummy Cave, Wyoming, there are 38 distinct fertile levels from 9300 years ago to A.D. 1580, indicating intermittent, perhaps seasonal, occupations for long periods of time (Jennings 1983). Agricultural activity with dependence on maize in the eastern Plains (the Plains Village Tradition) appeared between A.D. 600 and 1000 in South Dakota and nearby regions. The bison remained important, not only as a food source: bison scapulae were used as hoes. There were cultural contacts with the Pueblos and with the Caddoan Mississippian, and many villages were fortified.

In summary, quoting from the Cambridge Encyclopedia of Archaeology (chap. 57, which, along with Jennings [1983], is a good survey of the period), "all North American agricultural developments were related historically and were derived from prior appearances in Central Mexico and further south." The societies of the southeast reached the greatest degree of social complexity and development.

6.5. Development in Central America

The early development of agriculture at centers like Tehuacán, south of Mexico City, and Tamaulipas, northwest of the capital, has already been described. The slow emergence of an urban civilization reflects the long time necessary to develop an efficient agriculture in a challenging environment where techniques of irrigation were necessary in most of the area. The first indication of water control is in Tehuacán 6000 years ago. Places discussed later are shown in figure 6.5.1.

In the Formative Period (2500 B.C.–A.D. 300), the basis of the Meso-American civilization was laid through the development of intensive irrigation, astronomical observations, ceremonial centers and architecture, and hieroglyphic writing. The first great civilization was the *Olmec* (1200–600 B.C.) which developed its greatest monuments (the colossal stone heads of La Venta, San Lorenzo, and others) in an area of the Gulf coast. But the Olmecs established an exchange system that greatly extended and unified smaller-scale systems existing before in their area and in other areas of Meso-America, thus favoring the spread of cultural diffusion and trade throughout all Meso-America. After the decline of the Olmecs, important cultures and societies developed in the valley of Mexico (Cuicuilco first, then Teotihuacán) and in the valley of Oaxaca (Monte Albán), where major ceremonial centers were built. In Teotihuacán (200 B.C.–A.D. 800), the population in the later period may have been as high as 100,000 for the whole valley of Mexico, most of it in the capital.

The lowlands of Yucatán and Guatemala were occupied by *Mayas*, who extended also to the highlands in Guatemala. The conditions for agriculture in the Mayan regions were quite similar to those of the Gulf coast where the first urban civilization, that of the Olmecs, had earlier developed. These regions were excellent for sedentary, but not intensive, agriculture with two crops of maize a year. Soil fertility, however, is a serious problem: it is not clear how the Mayas solved it, but they may have employed several different solutions to make slash-and-burn farming more efficient (Jennings 1983). Ceremonial centers like Tikal in the Guatemalan lowlands and Kaminaljuyu in the Guatemalan highlands began developing in 30 and 500 B.C., respectively. The Mayan

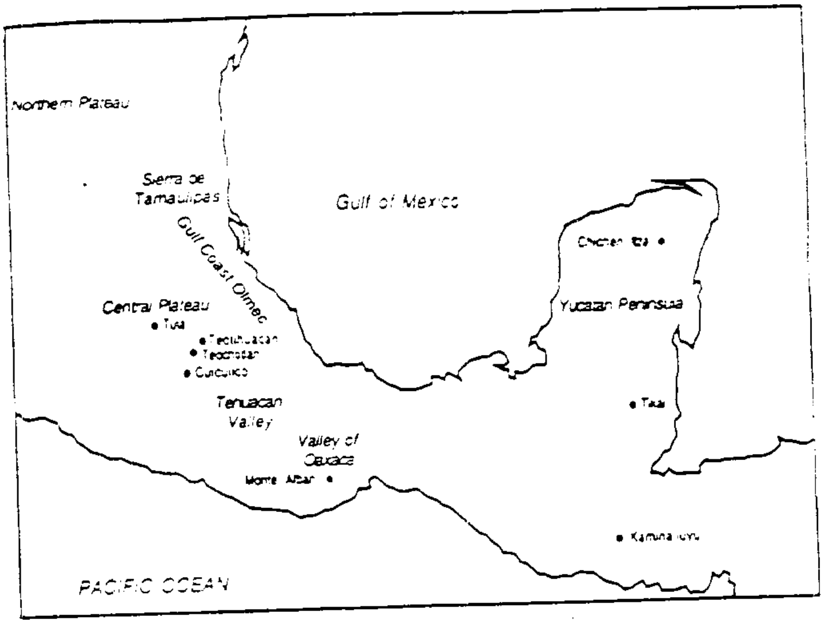


Fig. 6.5.1 Meso-America from Formative Period to European contact (from O'Shea 1980).

culture was strongly influenced by Teotihuacán. It was a multicentric, hierarchical society, with each center having majestic religious and ceremonial monuments. The major center in the Mayan classical period, Tikal, occupied an area of 60 km² (Jennings 1983) and had a population of tens of thousands of people. Outside the center, the population lived in small hamlets and was more diffuse. The classical Mayan period ended abruptly about A.D. 900 for unknown reasons.

The abandonment and destruction of Teotihuacán started a competition between Mexican regional centers, in which the *Toltecs*, from the city of Tollan near Tula in the central Plateau north of Teotihuacán, eventually gained control and became the first militaristic state of Meso-America. Their influence lasted from A.D. 900 to 1200 and extended as far as northern Yucatán, where Chichén Itzá (ended in A.D. 1224) became the most important center in the so called "Postclassic period" (A.D. 900–1520). Tula had been destroyed a little earlier. Power fell into the hands of the *Aztecs*, who came from the north to found a city at Tenochtitlán, where Mexico City is located. They were in power in 1519 when Hernán Cortés conquered Mexico. A survey of the period and region with additional references can be found in chapter 58 of the Cambridge Encyclopedia of Archaeology.

6.6: Development in South America

We have already discussed the difficulties associated with the very early dates of some South American sites. Clovis projectiles, indicating the Paleo-Indian Period, are found in most of South America as far south as Patagonia; in the north, at El Jobo, they may even antedate those found in North America. The beginnings of agriculture can be traced to a period between 9000 and 7000 B.P., mostly in the northern and central Andean region (fig. 6.6.1). There is no single, contained nuclear area, but a wide strip all along the northwestern coast that later radiated to other parts of the continent. By contrast, the tropical forest of the Amazon basin had a somewhat later, secondary, and less marked development, but hints of major novelties are already apparent.

There is broad consensus that maize came to the northern Andes from Mexico, along with perhaps squash and beans, but a number of plants were certainly domesticated locally in the early period in a variety of different environments. The rich marine fauna remained an important source of food on the coast, but it was later supplemented with agricultural products, and irrigation was developed in arid coastal regions. From the dry highlands came tubers like potatoes, while from higher-altitude forests or the eastern side of the Andes other products emerged, including apparently manioc, which later spread to the Amazon basin. As already mentioned, animal domesticates played a lesser role than in the Old World; however, in the south-central Andes the domestication of camelids provided an important contribution in terms of meat, wool, and animals of burden (for transportation). Around Lake Titicaca in southern Peru there were, at the time of conquest, some 500,000 camelids (llamas, alpacas). These animals had been food for highland hunters since very early times. Their natural range

Location of the Yanomame tribe

Fig. 6.6.1 Map of agricultural settlements in South America. The location of a modern tribe, the Yanomame, is also indicated (from Whitehouse and Whitehouse 1975; Barry 1980; Morris 1980).

is above 3000 m and their domestication may have begun very early (8000 B.P.; Bray 1980).

Agriculture played only a secondary role compared with foraging until about 5000 B.P., but after this date larger settlements supported by agriculture began to appear. Sites like Real Alto and San Pablo, on the Ecuadorian coast, are large stable preceramic farming villages; for example, El Paraiso had a population of 3000—4000 (Bray 1980). Pottery appeared around 5000 years ago at sites as diverse as Puerto Hormiga, near Cartagena, Colombia, and Valdivia, Ecuador.

Irrigation was practiced early and its sophistication increased to remarkable levels. Terracing of the steep Andean slopes was quite common and greatly improved water control and productivity. Cotton (possibly a local domesticate) and the manufacture of textiles soon acquired considerable importance. Improvement in trade networks made it possible to redistribute a variety of materials at long distances, and socioeconomic advances allowed people to make excellent use of the variety of microenvironments present in this region. Through ethnic and kin relations, in addition to trade, it became possible to develop the already mentioned pattern of a "vertical" economy whereby the same people had access to products made in very different environments, from the coast to the highest altiplanos. In the Andes with a day's walk, it is possible to go from one to another of a number of different ecological niches. By wise alliances or other social devices, an "archipelago" type of economy was created which gave people and small communities access to, or ownership of, pieces of land in a great variety of areas.

Population density must have risen steadily in this period, and it is not too surprising that the *Inca* empire, which at the time of conquest extended from southern Colombia to south-central Chile, may have been made up of 12 million people. Even if this estimate varies greatly according to sources, the area must have been very densely inhabited, perhaps as much as central Mexico. Complexity of society probably reached a new height about 3000 years ago, as shown by the high rate at which a new sophisticated art form, that of the *Chavin* culture (northern Peru), spread over a vast area, without any evidence of political or military occupations. The Moche site pictures of the north coast of Peru (200 B.C.–A.D. 600) show perhaps the first hints of organized military activity. A major influence was exercised by the *Tiahuanaco* culture, located on the southern rim of Lake Titicaca (1000 B.C. to A.D. 1000). There was progressive development of ceremonial centers and true imperial status was acquired in the last phases. At the peak, the urban population may have been 20,000—40,000 (Jennings 1983). This culture certainly had an important impact on the central Andes, probably initiating or advancing economic innovations later adopted by the Incas. After the collapse of the Moche, the *Huari* culture under Tiahuanaco influence established, probably through military conquest (Morris 1980), an empire that lasted until A.D. 800. Other states (e.g., Chimú, capital Chan Chan: perhaps 25,000 people) existed at the time in the central Andes. The only great South American empire started developing after 1438 when, near Cuzco, the Inca won a battle against a nearby state. They adopted an extremely effective military policy and by building an extensive network of excellent roads (15,000 km) across very difficult terrain, hundreds of road stations and state storehouses, and a well-trained army, they rapidly conquered an extensive territory. Called Tawantinsuyu ("Land of the Four Quarters") it was one of the greatest empires of the world. Inca was the name of the hereditary monarch. The nobility, the priests, and the bureaucrats formed 5%–10% of the population. The rest was a rural population on whom several types of taxes were levied, despite the lack of a currency. Of the agricultural products, roughly two parts of three went for the state and the nonproducing part of the population, and the rest was distributed by the village chief among villagers. Textile products were made by the women for the state. Time in the army and labor for the state were required of the men under the "mit'a" system, which was inherited and perfected from earlier states. It made possible very rapid military conquests and the monumental buildings dedicated to ceremonial and civil purposes for which the Incas are famous. Products taken by the state—food and textiles—

were redistributed to the population according to rank, and individual welfare was assured by an efficient state organization. The "khipu," a system of knotted strings of obscure origin, served the purposes of communication and accounting in lieu of writing.

The enormous Inca empire lasted about a century; at the time of conquest, the empire spanned 36° of latitude from near the present Equador-Colombia border to south central Chile, including much of the Andean region of Bolivia. It was destroyed by 250 *conquistadores* led to Peru by Pizarro in A.D. 1537. The Spaniards were greatly helped by epidemic diseases like smallpox and measles that they involuntarily imported to Peru, and decimated and disorganized the Indian population. They also ably exploited civil unrest.

The remarkable population density and degree of complexity and organization of the Andean states and empires were unmatched in the rest of South America, but a relatively dense population developed in the Amazon forests in spite of the difficulties met by farmers in much of this area. New crops were necessary for the wet soil and climate of the Amazon; the most successful of them was manioc. This plant exists in two varieties, sweet and bitter; the sweet variety was probably domesticated first. The bitter type requires a special fermentation treatment for destroying a poisonous substance that generates cyanide. Manioc cuttings can be easily planted, and propagation is extremely simple. It is especially suitable for tropical environments and provides roots rich in starch but poor in proteins, so it must be coupled with other food. Since manioc seeds are not used, it is difficult to trace it archaeologically; good clues are vats and special bowls employed to make chicha beer from it, or graters. Manioc may have been domesticated at an earlier date farther north, but the earliest well-dated find is from Yarinacocha on the upper Ucayali River in northern Peru (about 4000–3400 years ago). In the same area and time was also found the first pottery, probably derived from the Valdivia types. The Ucayali River is a tributary of the Amazon, and it has been suggested that there were close connections between Amazonia and the Andes during the Chavin culture. This would explain the Chavin paintings of tropical animals and plants that do not exist where this culture developed. The finding of pottery on the lower Amazon, and even at the mouth of the river (island of Marajo, Ananatuba culture, for location see fig. 6.6.1; date 980 B.C.) has suggested that cultural adaptations to the tropical environment, developed on the upper Ucayali River and other tributaries of the Amazon near the Andes, were spread downstream by colonists. There were also later migrations upstream, as in the case of the Omagua and Cocama tribes of the middle Amazon. At the time of European contact, the Omagua had villages of 300–3000 inhabitants, at short distances from each other, and the first visitors were impressed by the quality of pottery. High densities were possible only in areas very favorable for agriculture (varzea), from which natives were soon evicted after conquest, if they were not killed by disease or slave raids (Barry 1980).

Even today in the Orinoco and Amazon basin there exist tribes that have been relatively unchanged by European contact. Several of these—in particular, the *Yanomame* and the *Makiritare*—have been the subject of intensive biological investigations by Neel (1978, 1980; Neel et al. 1977) and his group, including among many others, population geneticists P. Smouse, R. Spielman, and R. Ward, linguist E. Migliazza, and cultural anthropologist N. Chagnon. The bibliography is too extensive for a complete listing, which can be found elsewhere (Chagnon et al. 1970; Smouse 1982; Chagnon 1983). The Yanomame are tropical gardeners who also rely on hunting-gathering activity. Like other hunter-gatherers they have a low number of births, because of long birth intervals. Despite their low fertility, they are at the moment in a period of demographic growth. Their present location is shown in figure 6.6.1. The history of Yanomame villages shows several fissions and fusions. Fissions reflect hostilities between groups and often take place along kinship lines. Although tendentially endogamous, there is migratory exchange between villages of the same tribe and, to a much lesser extent, with other Indian tribes of the region. There were only two documented instances of exchange (Neel, pers. comm.): one was due to the capture of two Makiritare women (Chagnon et al. 1970) and the other was due to the absorption of a few surviving members of a tribe that had come upon hard times (Weitkamp and Chagnon 1968). The genetic exchange between Yanomame villages, in spite of the fusion-fission history, is sufficiently limited that there is considerable genetic heterogeneity between villages, as described in detail in the original papers. In particular, the tendency to fissions following kinship relationships (lineal fission pattern) has the effect of reducing the effective population size of the village and therefore increases the effect of drift over that expected, assuming random fissions. Further strengthening of random genetic drift is due to the high polygamy of village chiefs. The Yanomame move frequently, often under pressure of hostile relationships within the tribe and with other tribes, and are currently drifting slowly southward. They occupied a part of the forest still sufficiently undeveloped at the time of the Neel study that they could keep to their traditional customs, a situation that is rapidly changing now.

The findings in other populations in southern Venezuela or in northern and central Brazil are similar to those of the Yanomame, but there are differences between tribes depending on their economy. Salzano and Callegari-Jacques (1988) have compared groups that they call stage-A tribes (hunters-gatherers with incipient agriculture, like the Yanomame, Trio, Cayapo, Xavante,

DOI 02192

and others), and stage-B tribes (technologically more advanced agriculturalists and fishermen like the Macushi, Wapishana, Ticuna, Makintare, Caingang, and many others). Fertility (number of children in completed families) is a little lower, intertribal marriages rarer, and variance of the number of children higher in stage A, but otherwise no major demographic differences were found. It is likely that the average size of villages is greater in stage B.

Contemporary but fragmentary information from other forest people of the Amazon-Orinoco basin shows that most are settling under pressure from governments, but the traditional way of life has been maintained in a few cases. Movements and admixtures are not uncommon, local economic development, especially mining, farming, and road building, are causes of serious encroachment. Temporary occupation in gold-mining operations and in oil fields is very destructive to traditional Amazon societies and bodes ill for the future of these populations. The extensive destruction of the forest following the opening of roads and modern agricultural and industrial plants create dangers that go well beyond the heavy damage to the local populations.

Surveys relevant to this period and additional references can be found in chapters 59 and 60 of the Cambridge Encyclopedia of Archaeology.

6.7. Physical Anthropology

Physical anthropologist C. S. Coon (1965) distinguished between Eskimos and Aleuts, on the one side, and American Indians on the other. The first two belong to the Siberian Mongoloids and came by a later migration; American Indians are stated to be Mongoloid in general and more uniform racially, "despite some of their peculiarities in blood groups" and are "more uniform racially than any other group of people occupying an equally vast area, but they are Mongoloids of a particular kind."

The origin of Mongoloids (see chap. 4) is believed to be either in northern China or north of it. According to Alexseev (1979), the maximum development of Mongoloid features is found in central and southern Siberia, especially among: (1) the Tungus-Manchu people of central Siberia, Kamchatka, and the lower part of the Amur Valley; (2) Turko-Mongolic people of southern Siberia and the Yakuts (middle Lena River); (3) the Nivkhs (= Gilyak), a small group in the northern part of Sakhalin and the mainland opposite it; (4) northern Asians like the Nganasan (Taymyr peninsula), Dolgans (a small group south of the Taymyr peninsula), Yukaghir (a small group east of the Lena River), and western Chuckchis. These people have somewhat variable pigmentation in skin and eyes, the lightest being the second group followed by the fourth and then the others. They all have extreme Mongoloid features, mostly reflected in the conformation of the skull and soft parts of the face, which include large cranial and facial dimensions, flattened face, nasal bones, and nasal bridge. It is difficult to give a "nuclear area," especially because the geographic distribution of Siberians has changed considerably in the last three centuries. Although Eskimos and Aleuts have peculiarities of their own, they tend to follow the same general pattern. Like most Mongoloids (with the exception of the Ainu), they have very little, if any, body and facial hair, but abundant and coarse dark hair with rare balding and late, if any, graying. Browridges are small, if any; the eyeballs are wide apart and smaller than in non-Mongoloids, placed forward in the orbits; the eye opening is narrowed to a slit by eyefolds, with the inner edge of the eye covered by the Mongolian or epicanthic eye fold in a percentage of individuals, which is especially high among Siberians. The lower margin of the orbit lies farther forward and the zygomatic bones protrude forward and laterally, generating the characteristic "high cheek-boned" appearance. The nasal bridge is usually low and flat, but there are also aquiline noses, with little, if any, intermediate forms.

American Indians have less flat faces than Siberians and often prominent, sometimes convex noses. This is perhaps the main difference, but, as just mentioned, the American Indian type of nose is also found in Asia: Coon (1965) cited the Tibetans and the Nagans of Assam. Pigmentation is usually darker among American Indians, but there is also variation among Siberians.

The mean stature of American Indians (Johnston and Schell 1979) varies considerably, being highest at high latitudes (Canada and Patagonia) and lowest in the tropical forests (Guatemala, Brazil). This follows the usual pattern of climate adaptation. In South America, mean stature was mapped for 43 tribes (Salzano and Callegari-Jacques 1988) and there is a slight difference between the northwest and the central-southeast (157 cm vs. 161.3 cm).

Of special interest are the studies of dental characteristics by Turner (1987,1989). Most northern Mongoloids have shovel-shaped incisors, which are also found in fossil skulls as far back as Chinese *Homo erectus*. This and other cranial peculiarities have been a major reason for claiming independent speciation of Mongoloids (Coon 1965; see also Wolpoff et al. 1984). The genetic exchange at various times and places between local human types, even archaic, and immigrant *H. sapiens sapiens* is a possibility worth considering, but the picture of migrations from Asia to America developed by Turner, and based essentially on dental clues, is unrelated to this question. It is important, however, that on the basis

DOI 02193

of this evidence, it was stated that a strong difference exists between East Asians from northern China and the Southeast Asian type. For instance, northern Mongoloids ("sinodonts" according to Turner) have 60%–92% shoveling, as against 13%–25% in southern Mongoloids ("sundadonts"). Different percentages refer to different populations sampled. Japanese of the Jomon period (chap. 4) show the lowest percentages, and, together with the Ainu, are classified by dental criteria among the southern Mongoloids, with Thailand, Malay-Java, and Polynesia. Two other traits showing major differences between northern and southern Mongoloids are the number of cusps and the number of roots on molars.

Turner's analysis is based on the premise that dental characteristics are highly inherited, stable in evolution, and not sensitive to evolutionary changes as a function of adaptation to different types of foods. These hypotheses require independent confirmation. Unquestionably, teeth have the advantage of being readable in fossil samples and perhaps also of offering greater detail than bones. Using dental microevolution, Turner calculated 14 kya as the date of the first crossing of the Bering land bridge by the Paleo-Indians. He also postulated that the Na-Dene migration was independent of that of Paleo-Indians and that it occurred 14–12 kya, just before the land bridge of Beringia was completely submerged. In addition, he hypothesized that the Na-Dene may have originated from the late Diuktai culture (fig. 6.2.2), passed along the southern edge of Beringia to Kodiak Island and then to the Northwest coast of the Pacific. He also stated that the third migration, that of the Eskimo-Aleuts, arrived just before the bridge was severed, but after the Na-Dene. These conclusions agree well with other independent sources of evidence (Greenberg et al. 1986; and, apart from dates, with our genetic analysis (sec. 6.9 et seq.).

As we have already briefly indicated in chapter 2, dental data on northern Asia, southeast Asia, and the Americas are generally in excellent agreement with those from single genes. How much further back this agreement will go remains to be seen. The question of how much further back dental data can take us in human evolution is also a matter of conjecture. Apart from the unknown role of natural selection and of dietetic customs—believed to be negligible by Turner—and the unknown level of heritability, an important consideration is the number of independent genes that can be detected by this approach. This is also unknown; only when this number is really large are conclusions insensitive to the addition of further information. Statements based on dental analysis are very interesting, but it would be unwise to rely on them alone until more is known about the problems just mentioned, especially if and when they disagree with other sources of evidence.

6.8. Linguistics

The nonlinguist who approaches the field of the classification of American Indian languages can only be shocked by the segregation of linguists into two groups that hold almost diametrically opposed beliefs: one, more numerous, refuses to recognize unity in these languages and chooses to list a large number of essentially unrelated small families or isolated languages, the interrelationships among which are considered beyond recognition; the other much smaller group proposes three families, corresponding to the three major migrations that are also recognized by other criteria, namely, in time sequence, Amerind, Na-Dene, and Eskimo-Aleut. One cannot fail to see this as the most dramatic example of the usual division between "splitters" and "lumpers," which has been observed repeatedly in almost every classification, be it of living organisms or inanimate objects. To increase the dismay, the group of splitters uses extremely strong language against the author of the unification of Amerind languages, Greenberg (1987) who has earned enormous respect from the whole linguistic community for all his other work. The diatribe has been the subject of articles of popular science (two rather extensive summaries by P. Ross in *Scientific American* and R. Wright in *Atlantic* appeared in April 1991). Another summary of the dispute is in a Postscript to the 1991 edition of Ruhlen (1987).

Ruhlen (1987 and references therein) summarizes the history of classification of Amerind languages, dividing it into three phases. The first was started by the famous anthropologist Alfred Kroeber (1876–1960), who, at the beginning of the century collaborated with R. Dixon to reduce the number of families of North American languages by combining some previously recognized taxonomic units. Edward Sapir carried this effort further, and in 1929 the number of North American families was six, two of which were Eskimo-Aleut and Na-Dene, the languages of the Pacific Northwest. This began a second phase, which can be called a "revolt," and the dismemberment of Sapir's families; after a 1976 conference, the number of independent units of North American languages was back to 63. The list of the results published in 1979 was stated to be "conservative and not very controversial" representing "current received opinion." The third phase was opened by the linguist J. Greenberg, who made the claim that there exist only three families: Eskimo-Aleut, Na-Dene, and Amerind (1987). The Amerind family includes most North American languages and all Central and South American languages,

DOI 02194

for which there had previously been only limited analysis. For South America, in particular, the information was a list of languages or language clusters rather than a true classification.

The exact meaning of the word "family" (for which some prefer "phylum" or "stock") need not concern us here; it usually refers to the highest "genetic" grouping recognized. Linguists use the word genetic to mean "common descent" similar to "phylogenetic" for geneticists. Today, some linguists have started forming "superfamilies" from the conventional families, hence some of the families are no longer the highest genetic unit.

Nonlinguists, like the authors of the present book, cannot make a contribution to a discussion based on linguistic arguments. From a general scientific point of view, the methodological analysis found in the recent book *Language in the Americas* by Greenberg (1987) is convincing. We accept Greenberg's work as a very serious attempt at a comprehensive classification, which has already achieved some important results by distinguishing the same three major groups found from totally independent sources. Even if this classification changes in the future, it supplies a starting point that is not provided by the extremely fragmentary classifications supported by other authors. As Greenberg's book convincingly shows, the difficulties encountered by the extreme splitters are methodological. They proceed by comparing two languages at a time, with an extremely detailed analysis that makes it impossible to test more than a small fraction of all possible pairs. Their conclusion is limited to the statement that the pair is either "related" or "not related," omitting an estimate of a degree of relationship, without which it is impossible to build a classification that goes beyond the recognition of scattered relationships. The decision on relatedness is based on extremely rigorous criteria, with which, according to Greenberg, it would be impossible to recognize even the unity of the Indo-European family, a step backward by universal consensus. One of these criteria is the belief that "sound correspondences" (rules of change of sounds established on the basis of historical examples) must be followed without exception. Greenberg uses a method of multilateral comparisons, in which many languages are compared for a number of words and other criteria selected for their evolutionary stability. We limit our treatment in the rest of this section to summarizing Greenberg's classification, as given by Ruhlen (1987).

We refer to the three families suggested by Greenberg, called phyla by Ruhlen, as families and to their subdivisions as subfamilies. The geographic distribution of the various subfamilies is shown in figures 6.8.1. A and B.

The ESKIMO-ALEUT family comprises 10 languages and 85,000 speakers; Aleut is presently spoken by 700 people in the Aleutian islands. Three Eskimo languages are spoken by 600 inhabitants of the USSR. The Asian Eskimo languages belong to the Yupik subgroup, found primarily in southwestern Alaska. The Eskimo living on the Arctic coast of North America and Greenland speak three languages: Alaskan Inuit, Canadian Inuit and Greenland Inuit. These are often considered three segments of a dialect chain stretching from northern Alaska to Greenland.

The NA-DENE family is spoken in northwestern North America and consists of two languages, Haida (300 speakers of a total 2000 Haida, living on Queen Charlotte and Vancouver islands) and Tlingit (2000 speakers of Tlingit, out of 10,000 living on the coast north of the Haida), as well as the Athabaskan subfamily made up of 30 languages. The Athabaskan languages are spoken by a northern group of some 70,000 speakers in eastern Alaska and all of western Canada, a few (mostly extinct) groups in California and Oregon, and a southern group of about 130,000 speakers, the Apache and Navajo.

The AMERIND family contains 583 languages, spoken by 18 million speakers. They are subdivided by Greenberg (1987) as follows (see also Ruhlen [1987] and fig. 6.8.1).

I. *Northern Amerind* includes as subfamilies Almosan-Keresiouan, Penutian, and Hokan.
   - A.1. *Almosan* consists of Kutenai (a single language), Algic (Algonquian and two isolated languages, Wiyot and Yurok) and Mosan (Wakashan, Salish, and Chimakuan); it covers most of Canada south of the zones occupied by Eskimos (the Arctic) and the Na-Dene (northwestern Canada and central Alaska). It also extends to the Midwest south of the Great Lakes and to New England.
   - A.2. *Keresiouan* includes Keres (essentially a single language) and the Siouan, Iroquoian, and Caddoan families; it covers the rest of the Midwest almost to the Atlantic coast.
   - B. *Penutian* is a northern group including much of Oregon and California, with outliers (Tsimshian) as far north as Canada; in southeastern North America, a Gulf group includes the Muskogean family and a few isolated languages; in New Mexico, Zuni; a southern group is found in Mexico (Huava, Mixe-Zoque, Totonacan, and the Maya in Yucatán and Guatemala).
   - C. *Hokan* is a northern group with small clusters in northern and southern California, Baja California, and parts of Arizona; a southern group in northeastern Mexico and Texas.

II. *Central Amerind* includes three distinct subfamilies: Tanoan, Uto-Aztecan, and Oto-Manguean
   - A. *Tanoan* includes Tewa (Arizona and New Mexico) and Kiowa (Oklahoma).
   - B. *Uto-Aztecan* is in most of the Southwest, including the Hopi and Pima groups.
   - C. *Oto-Manguean* is found in southern Mexico, especially the southwest; also includes the Zapotecan, Chinantecan, Mixtecan, and Mazatecan.

III. *Chibchan-Paezan* includes the Chibchan and Paezan families
   - A. *Chibchan* languages are found on the southwestern coast of Mexico and in almost all of Central America

DOI 02195

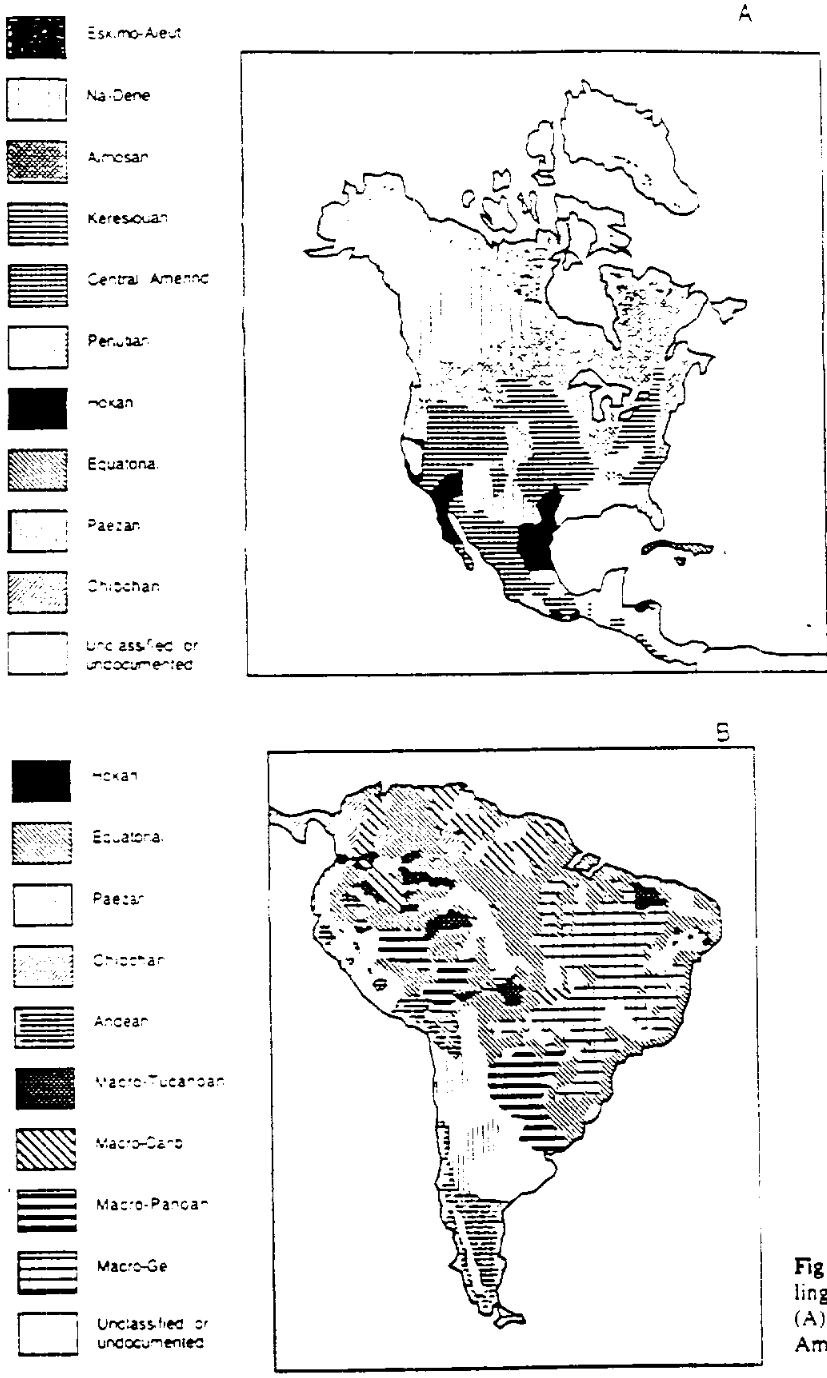


Fig. 6.8.1 Geographic distribution of linguistic families and subfamilies in (A) North America and (B) South America (Ruhlen 1987).

south of the Yucatán; other clusters in Venezuela and Brazil include the Yanomame.

B. The *Paezan* languages formerly found in northern Florida (one language, now extinct), now survive only in South America along the coast of Colombia and Ecuador and farther down in the Chilean Andes; there are also splinter groups in the Brazilian forest and on the northern coast of South America.

IV. The 20 *Andean* languages, of the 583 Amerind languages, account for half of the Amerind population because of the great diffusion of Quechua and Aymara in the central Andes. The Inca empire, and perhaps also the Spanish influence, were responsible for the spread, which is therefore recent. There are also a large number of speakers in the southern Andes, including the Mapuche (= Araucanians). Three small areas in the northern Andes also speak, or spoke, Andean languages.

V. *Equatorial-Tucanoan* includes the Equatorial and the Macro-Tucanoan subfamilies.

A. *Macro-Tucanoan* is found in nine geographic clusters, mostly in western Brazil, with a few in eastern Brazil.

B. *Equatorial* has the greatest number (25%) of all Amerind languages and is widespread from west to east and from the Caribbean islands to Uruguay, in Venezuela, Colombia, Ecuador, Peru, and central and eastern Brazil. The large number of languages is due to the inclusion of two important subfamilies, Arawakan and Tupi-Guarani.

VI. *Ge-Pano-Carib* includes Macro-Ge, Macro-Panoan, and Macro-Carib.

A. *Macro-Ge* was very widespread, but only a few languages survive, mostly in southern Brazil, in the highlands and farther south. The Kaingang language belongs to this group.

B. *Macro-Panoan* once extended from Peru to Uruguay; many languages are now extinct.

C. *Macro-Carib* languages were spoken in the northern regions of South America, mostly on the coast of Colombia, Venezuela, the Guianas, and northern Brazil, with outliers farther south.

Geographically, Almosan and Keresiouan are found only in North America; Penutian, Hokan, and Central Amerind are found in North and Central America, Paezan, Chibchan, and Equatorial in Central and South America; and Andean, Macro-Tucanoan, Macro-Carib, Macro-Panoan, and Macro-Ge only in South America. An important point is that the geographic distribution of Amerind languages is extremely fragmentary, especially in South America.

6.9. Phylogenetic analysis of America

Both anthropological and linguistic evidence points to three major groups that may have represented distinct migrations, all from Northeast Asia. The Paleo-Indians were the first, though their date of entry is uncertain, between 35 and 15 kya. There is greater consensus for later dates, but enough uncertainty that an earlier one must be entertained as a possibility. The northwestern American Indians, identified by the family of languages they speak as Na-Dene, were next, as indicated also by their remaining in a more northern area. The Eskimo-Aleut were the latest, and inhabit only the extreme northern region both in America and Asia. The presence of some Eskimos in Asia is believed to be a retrogression from the Americas to Asia, rather than an aboriginal Asian population. The date of entry of the last two groups is probably 15–10 kya.

The question of whether the three migrations can be distinguished on the basis of biological characteristics has recently received some tentative answers, all basically positive. In addition to Turner's (1987, 1989) dental analysis discussed in section 6.7, there is a study of Arctic populations by Szathmary (1981; see also 1985), who used data from 14 genetic loci and found the Athapascan (Na-Dene) are more similar to Eskimos and Chukchi than to northern Algonquians (non-Na-Dene North American Indians). Williams et al. (1985) collected *GM* and *KM* data from the Apache and Navajo (southern Na-Dene), and the Pima, Papago, Hopi, and Walapai (non-Na-Dene from the North American Southwest) and showed that these two groups differ genetically. The difference, however, is not striking and conclusions based on a single genetic system, even one as informative as *GM*, are unsatisfactory. In a more systematic analysis based on data from a larger number of genes and populations, Zegura (Greenberg et al. 1986) tentatively recognized the three migrations, but acknowledged the existence of difficulties for drawing final conclusions.

In our paper (Cavalli-Sforza et al. 1988), which summarizes some of the points made in chapter 2, all the Na-Dene were collected in one group and the rest of the American continent was divided into North, Central, and South America. The Central group was defined on a linguistic basis, taking the Central Amerind subfamily, which is actually partly in North America and does not include all people from Central America. In that analysis, Eskimos clustered with Chukchi and with Turkic-speaking populations of northern Asia, forming a small subcluster of the Northeast Asian cluster, while all American Natives including Na-Dene formed a separate, major subcluster of Northeast Asia. Na-Dene speakers, however, include two major groups, northern and southern. The southern Na-Dene are essentially the Apache and Navajo. Although the exact time of their migration from Canada is not known, it was probably late, and they are believed to have arrived in the Southwest around A.D. 1200. Until recently, there was a splinter Apache group in Kansas.

In the analysis of this section, the major criterion for grouping populations is linguistic. In view of the special linguistic interest, we also added Chukchi and Koriak in order to test possible similarities with Eskimos. Within a few linguistic groups, in particular the Na-Dene, we use a further subdivision on the basis of geography. Because subfamilies are dispersed in widely different areas, it would be especially interesting to distinguish subareas in other subfamilies; but unfortunately, even after the pooling of individual tribes into linguistic groups, there are not enough data to form as many geographic subgroups as would be desirable. Eliminating groups because they take unexpected positions would of course be unacceptable. The procedure adopted was to eliminate systematically groups or subgroups that had fewer markers. Here, as in other chapters, we have tried to limit gaps to not more than 50% in the data matrix. Populations for which there were clear signs of admixture with either Caucasoid or African people, according to the authors who collected the genetic data, were eliminated. We thought it useless to carry out a direct analysis of admixture considering that extreme drift in many American Native groups has generated exceptional gene-frequency variation. There is no assurance for any of the most informative markers, even some *RH* alleles, that they were truly absent in the original American Natives and can therefore be used for inferring admixture. We are reassured by the results of another study that the possible Caucasoid or African admixture of some data we used is not misleading: Salzano and Callegari-Jacques (1988) used 17 non-*RH* alleles potentially useful for evaluating the proportion of non-Indian genes and compared them with results using *RH* alleles, which might be better markers of admixture. There was a correlation, but it was doubtful whether the estimates of admixture could be considered quantitatively valid. Of 58 tribes, only 5 had estimated admixtures of over 25%; 11 between 10% and 25%. Trees from populations believed to have less than 10% admixture gave results very similar to those obtained using the general set. As to our own data, we find there is a clear effect of admixture only in North America, as shown by synthetic maps (sec. 6.13).

The groups for which the number of markers was considered adequate are listed below, together with the names of the major tribes that formed them. In almost every case, however, there were some other, less well-investigated tribes that are not named below but are listed in the tabulations; data from the tabulations were used to calculate the mean gene frequencies of each group. In this way it was possible to increase the representativeness of the data, at least for those genes for which data are more abundant. Such genes, because they are represented in more groups, inevitably have a more important influence on the final conclusions than genes more rarely investigated. Restricting the analysis exclusively to these genes, however, would have reduced its power.

In the list below, the tribes that are named are those that have supplied the most important part of the information, having been tested for more traits. We repeat here that, especially in the Americas, and not only in the southern part, there was enormous drift in many populations, generating great variation from one population to another. This is clearly visible, for instance, in the geographic maps of principal components (Suarez et al. 1985). The averaging over populations can help reduce the effects of drift of individual populations, as already explained in chapter 2.

Figure 6.8.1 shows the geographic distribution of the linguistic groups, and table 6.9.1 the F_{ST} genetic distances among groups. The 23 tribes or groups that contributed most to the genetic data used in the analysis are listed below, with the three-letter symbol used in the table.

I. Eskimo-Aleut
  A. Eskimos: U.S.A. other than Inuit (EUS); U.S.A. Inuit (EIN); Canadian Inuit (ECA); Greenland Inuit (EGR)
  B. Aleuts (only U.S.A.; USSR Aleuts had too few markers and tended to associate with Asian populations)

II. Na-Dene
  A. Northern Na-Dene (non-Athabascan): Haida, Tlingit (NDN)
  B. Canadian Na-Dene (Athabascan): Dogrib, Slave, Chipewyan (NDA)
  C. Southern Na-Dene (Athabascan): Apache, Navajo

III. Amerind (NDS)
  A. *Northern Amerind*
    1. Almosan (NAL): Blackfoot, Cree, Makah, Montagnais, Micmac + Penobscots, Naskapi, Nootka, Ojibwa, Salish + Mukleshoot + Flathead + Quinault + Okanagan
    2. Keresiouan (NKE): Caddoan (Caddo + Wichita + Pawnee), Cherokee
    3. North Penutian: Seminole (= Muskogee), Zuni
    4. South Penutian: Eastern Maya (Ixil, Kekchi, Cakchiquel, Kiche), Maya, Totonaca, Tzeltalan (Tzeltal + Toztil), Yucatecan

    Note that Penutian were tested jointly (PEN), and Hokan were eliminated because of strong admixture.
  B. *Central Amerind* (CAN): Chiapaneca, Choluteca, Nahua, Papago, Pima, Tarahumara, Zapoteca
  C. *Chibchan-Paezan*
    1. Chibchan (MCC): Guaymi, Ica, Misumalpan (Paya, Lenca, Miskito, Sumo), Rama, Talamanca (Cabecar, Bribri, Boruca, Teribe, San Blas), Tarascan, Tunebo, Yanomame
    2. Paez (MCS): Atacameno (= Kunza), Cayapa (Ecuador), Choco, Colorado, Noanama, Paez, Warao
  D. *Andean* (SAN): Alacaluf, Aymara, Mapuche, Ingano (Colombian Quechua), Quechua
  E. *Equatorial-Tucanoan*
    1. Equatorial (SEQ): Arawakan (Goajiro, Arawak, Paraujano), Baniwa, Bari, Campa (Maipuran), Chane, Chipaya, Emerillon, Guayaki, Jivaroan (Jivaro, Aguaruna, Yaruro, Cofan, Shuara), Maue, Oyampi, Pacas Novas (Chapacuna), Palikur, Parakana, Piaroa, Piro, Siriono, Wapishana, Zamucoan (Ayore, Imoro, Chamacoco)
    2. Macro-Tucanoan (SMT): Siona, Ticuna, others
  F. *Ge-Pano-Carib*
    1. Macro-Carib (SMC): Carib, Galibi, Macushi, Makiritare (= Yecuana), Panare, Pemon, Trio, Wayana, Yupa (= Northern Motilon)
    2. Macro-Panoan (SMP): Cashinahua, Choroti, Chulupi, Lengua, Mataco, Shipibo, Toba
    3. Macro-Ge (SMG): Caingang, Cayapo, Craho, Xavante

Table 6.9.1. Genetic Distances (in the lower left triangle) and Their Standard Errors (in the upper right triangle) of American Tribes, Grouped Mostly Linguistically (all values x 10,000)

	PEN	*CAN*	*CKC*	*CKO*	*CKR*	*ECA*	*EGR*	*EIN*	*EUS*	*ESR**	*MCC*	*MCS*	*NDA*	*NDN*	*NDS*	*NAL*	*NKE*	*SAN*	*SEQ*	*SMC*	*SMG*	*SMP*	*SMT*
PEN	0	76	181	280	209	234	300	329	224	178	92	70	110	171	50	104	139	40	37	38	98	128	115
CAN	199	0	259	386	246	139	320	265	245	207	85	138	133	124	48	79	39	60	71	88	126	115	126
CKC	896	968	0	170	537	229	198	359	244	84	265	250	313	366	178	165	257	228	180	247	184	281	247
CKO	1241	1367	335	0	678	294	264	627	548	210	413	299	495	480	269	275	333	364	259	362	241	405	326
CKR	1202	1051	980	1414	0	150	376	298	105	201	340	337	325	278	223	193	310	254	308	297	221	312	388
ECA	956	769	658	1070	621	0	86	56	69	66	276	231	191	128	231	75	193	267	197	226	245	257	195
EGR	1286	1152	682	1232	1067	302	0	84	69	156	296	405	314	189	345	247	333	357	298	214	411	307	258
EIN	1438	1211	1370	2015	1494	326	278	0	95	183	404	460	119	300	245	200	318	369	324	360	409	457	288
EUS	1033	977	562	1227	496	259	330	417	0	141	320	360	106	127	210	167	217	263	259	359	285	329	311
ESR	978	1002	246	680	583	407	677	799	351	0	364	385	247	190	238	151	289	289	291	304	237	370	347
MCC	436	359	1389	1871	1732	1267	1312	1548	1372	1669	0	105	211	362	117	208	92	103	55	79	134	81	161
MCS	334	451	1462	1857	1851	1454	1638	2129	1743	1708	488	0	311	531	123	116	68	125	148	96	189	98	288
NDA	744	734	1178	1587	1267	519	1054	492	681	930	1106	1742	0	63	142	111	116	126	129	191	172	213	206
NDN	744	589	1136	1516	1236	645	690	872	595	746	1265	1995	377	0	89	217	289	263	327	401	196	480	451
NDS	256	240	756	1019	892	836	1220	1051	719	787	560	655	426	425	0	52	102	73	68	81	144	145	149
NAL	335	419	618	963	950	367	831	814	669	704	737	634	483	736	217	0	100	167	101	100	111	92	141
NKE	257	146	1002	1136	1406	873	987	1263	703	899	401	237	703	719	319	295	0	41	61	97	153	94	116
SAN	168	280	1151	1586	1437	1144	1387	1374	1367	1335	394	588	965	1094	359	417	204	0	60	62	110	41	140
SEQ	195	230	1089	1522	1673	1072	1293	1450	1292	1437	264	352	906	1109	415	461	219	335	0	17	93	85	87
SMC	174	275	1208	1661	1471	1133	1109	1652	1390	1399	297	340	1067	1193	354	457	300	296	98	0	133	95	137
SMG	381	525	1118	1489	1449	1242	1722	1940	1268	1157	543	578	1118	1054	671	565	524	484	387	496	0	170	197
SMP	505	494	1430	1839	1878	1372	1362	1697	1460	1689	393	393	1037	1410	680	384	403	270	341	410	650	0	220
SMT	524	540	1429	1926	1910	1073	1179	1329	1676	1724	606	739	1075	1234	773	690	468	574	380	495	811	734	0

Note.– PEN, Penutian; CAN, North Central Amerind; CKC, Chukchi; CKO, Koryak; CKR, Reindeer; ECA, Canadian Eskimos; EGR, Greenland Eskimos; EIN, Inupik Eskimos; EUS, Yupik Eskimos; ESR, USSR Eskimos; MCC, Central Macro-Chibchan; MCS, South Macro-Chibchan; NDA, Canadian Na-Dene; NDN, North Na-Dene; NDS, South Na-Dene; NAL, Almosan; NKE, Keresiouan. SAN, Andean; SEQ, Equatorial, SMC, Macro-Carib; SMG, Macro-Ge; SMP, Macro-Panoan; SMT, Macro-Tucanoan. Triangles indicate more compact groups. Tribes included in the groups are listed in the text.

* USSR Eskimos had too few markers and were not used in the tree of Figure 6.9.1; they tend to associate with Chukchi.

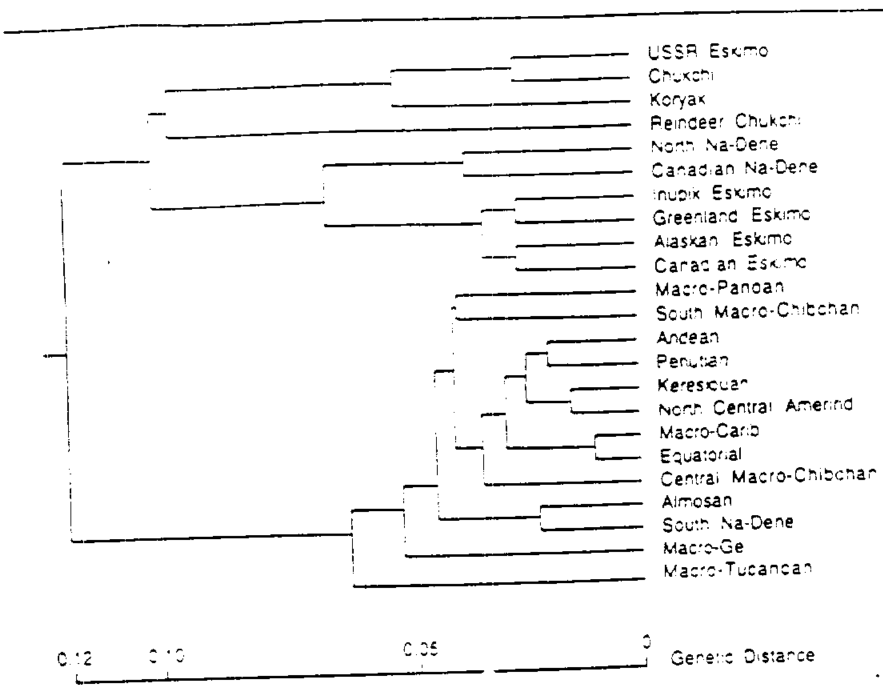


Fig. 6.9.1 Genetic tree of 23 American tribes grouped according to linguistic criteria

The tree obtained on the linguistic groups formed from subfamilies of North, Central and South America is given in figure 6.9.1 and the PC map in figure 6.9.2. Data include groups with an average number of 72.8 ± 6.8 genes. The PC map accounts for 52% of the original genetic variation.

The genetic tree shows a very clear separation between Eskimo-Aleut and Chukchi-Koryak (in northeastern Siberia, speaking non-American languages) on one side, and all American Indians other than Eskimos on the other. The Na-Dene separate into two groups, the most northern joining the Eskimo and Chukchi cluster and the southern ones the Amerind cluster. These conclusions are in agreement with those reached by studying the matrix of genetic distances. The average distance of southern Na-Dene to the two northern Na-Dene groups is 0.0426, and that between the two northern Na-Dene groups (Canadian and U.S.A.) is 0.0377 (difference not significant); but the northern and southern Na-Dene show average distances of 0.0693 and 0.0957 from the Eskimos. Table 6.9.2 shows the distances between the northern Na-Dene and southern Na-Dene on the one side, and the four most typical Northern Amerind groups on the other.

It is clear from the above distances that the Apache-Navajo, forming the southern Na-Dene, must have had

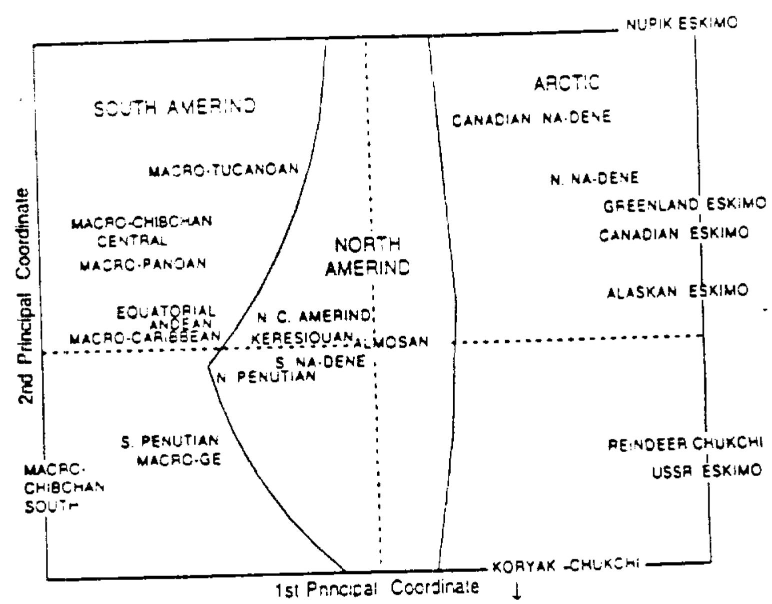


Fig. 6.9.2 Principal-component map of American tribes grouped by linguistic subfamilies.

DOI 02200

Table 6.9.2. Genetic distances (x 10,000) among Northern or Southern Na-Dene and Other American Natives from Northern and Central America

	N. Na-Dene	S. Na-Dene
Almosan	609	217
Keresiouan	712	319
Northern Penutian	739	266
North Central Amerindian	662	240

considerable admixture with northern Amerinds. It is also possible that the northern Na-Dene have had some admixture with northern Amerinds, but the data are insufficient to show it.

This observation can also explain why in our earlier world tree (chap. 2) a group made by averaging northern and southern Na-Dene tended to join the Amerinds, splitting from them, however, at an apparently very early time. We know that mixtures tend to attach to an average linkage tree at a higher level than the actual time at which the mixture occurred. The attachment of Na-Dene to other Amerinds indicates that the component in the mixture due to the latter is, on the average, strong enough that it outweighs an original, unknown, component responsible for the difference between Amerind and Na-Dene.

The five Eskimo groups are reasonably clustered in the tree, with Asiatic Eskimos showing greater similarity with their close geographic neighbors, the Chukchi, than with the American Eskimos. USSR Eskimos are a very small group and the separation sufficiently long that this result is not surprising. Furthermore, there are linguistic connections between the Chukchi and Eskimo languages, strengthening the case for a relatively recent common origin of the Eskimo and Chukchi.

Bootstrapping shows that the separation of the two major clusters is clear-cut.

Of 50 bootstraps, 19 show the identical first split of the tree of figure 6.9.1. This may seem a low proportion, but in the other 31 bootstraps, deviations from the tree of figure 6.9.1 are almost always minor.

In 14 bootstraps, the main change is the addition to the Arctic group of the southern Na-Dene; given the strong similarity between the southern Na-Dene and the northern Na-Dene this is not surprising. In 6 of these 14 bootstraps also, the Almosan follow the southern Na-Dene in joining the Arctic cluster. Because Almosan is the Amerind group geographically closest to the Eskimo and northern Na-Dene, the potential for admixture is not negligible.

In 17 bootstraps, one or two populations leave the Arctic cluster; they are, 11 times of 17, the pair of northern Na-Dene and Canadian Na-Dene, which almost always stay together and join the southern Na-Dene in the Amerind cluster. In the other 6 cases, Chukchi or the Reindeer Chukchi, or, more rarely, the USSR Eskimos join the Amerind cluster.

Even though the USSR Eskimos are today more similar genetically to the Chukchi than to the other Eskimos, the old relationship is still visible in bootstraps. The similarities with Almosan seem modest, and the admixture was probably not a major one, in harmony with the territorial and ecological segregation of Eskimos.

The similarity of northern and southern Na-Dene and their other associations are also clearly visible in the finer details of the bootstrap behavior. The group formed by Haida, Tlingit, and a few Athabascans on the coast is fairly similar genetically to the Canadian Athabascan, and they almost never part. Southern Na-Dene show their affinity with the northern Na-Dene, but they have an even greater affinity with Almosan, which manifests itself in pairing with Almosan in 25 of 50 bootstraps, while they pair with one or the other or both northern Na-Dene in 13 of 50 bootstraps; they show almost no tendency to pair with any other single population. This indicates that the admixture of Navajo-Apache with Amerinds probably happened mostly in earlier times in Canada before the move south.

The Amerind cluster has an internal subcluster of seven North, Central, and South American subfamilies. Two pairs of subfamilies, one central-southern and the other northern are next; Ge and Tucanoan are the outliers.

As mentioned more than once before, an outlier in a tree has several possible explanations. Assuming that evolutionary rates are constant, one can trust the tree structure to correspond to the order of separation of branches, and thus probably to the order of their migration away from the place or places of origin. When an outlier is a very small population that developed in a highly isolated area, the assumption of constant evolutionary rates is difficult to accept, as one would expect it to show a long branch because of high drift. In this case it seems more likely that outliers did not separate particularly early, but being of small size had a very high evolutionary rate because of extreme drift.

In order to avoid the consequences of extreme drift for individual small tribes, we have grouped them, in this case, according to linguistic subfamilies. If linguistic families are formed of groups with greater internal genetic similarity than randomly formed clusters, the pooling of tribes in linguistic groups can help reduce the effects of extreme drift. Although we did not know whether averaging by linguistic family would be truly useful, we attempted it nonetheless. We are currently not aware of better alternatives.

If drift of individual tribes is very high, one may need to average many tribes to obtain a substantial reduction of variation. This has no always been possible here because of a lack of adequate data. In fact, the two worst outliers: Macro-Ge, and Macro-Tucanoan are made up of only four and two populations, respectively. Moreover, the number of individuals in these tribes is small. The Ge are mostly represented by the Caingang (7000 in Brazil), Cayapo (3700 in Brazil), and Xavante (3000 in Brazil). The Tucano are represented by the Ticuna, who number 21,000 in Brazil, Peru, and Colombia combined. Each local population is likely to be a small fraction of the total for the tribe and to have little or no contact with other splinters of the tribe located in other, often distant, regions. These outliers are therefore likely to be cases of very high drift. The next South American outlier, Macro-Panoan, is represented by seven tribes, with numbers of individuals comparable to those above; Central Chibchan-Paezan is represented by 10 tribes. It seems that the greater the number of tribes, the less extreme is the position of the family in the tree. This supports the idea that drift is important in this case; further evidence that high drift is involved comes from geographic multivariate maps, and from other data to be given in later sections, which show extreme differences between geographic neighbors.

A third possible explanation for outliers is an agglomerative origin, with contributions from many groups belonging to very different sources. In urban civilizations, this is often observed in capitals that have received immigrants from widely different regions. They show, therefore, affinity with many other regions without forming close pairs with any particular one. This explanation can be excluded in the present case for forest populations like the Ge and the Tucano, who live (at least today) at a low economic level in isolated areas. The safest general conclusion from the tree, as we discuss later, is that, although the major fissions of the tree are in good agreement with information from other sources, it seems difficult to reconstruct a reasonable genetic history from it as far as the Amerinds of South America are concerned. We see in more detail in section 6.11 that this conclusion is correct. This does not necessarily mean that grouping by linguistic families leads to wrong conclusions, but simply that it was not adequate to improve on a difficult situation.

The PC map (fig. 6.9.2) is more useful, at least in showing the effect of geography: the first axis separates the Arctic populations at the right, puts all northern Amerinds in the center, and the central and southern Amerinds at the extreme left. It is thus in good agreement with basic geography. Arctic, northern Amerinds cluster neatly, whereas southern Amerinds show three major clusters: Tucanoan, Central Chibchan, Panoan; Carib, Equatorial, Andean; southern Penutian, southern Chibchan, and Ge. These results differ somewhat from those obtained with the tree, but they are based on two dimensions only.

At this point, we can ask the most important question: does the proposed three-migration theory agree with the results of genetic analysis? The answer is clearly positive. The two major clusters of the tree, Arctic and Amerind, could certainly be interpreted as separate migrations, and the Arctic cluster does contain a secondary split into Na-Dene and Eskimo, the other two postulated migrations. Thus, the tree is compatible with the three-migrations theory of Greenberg et al. (1987), as is the PC map. The analysis may also support the idea that the two later migrations, Na-Dene and Eskimo, had a related origin in Northeast Asia, in the sense of having come from a common ethnic group in that region. The separation of the Eskimo-Chukchi-northern Na-Dene cluster from the Amerind cluster is also visible in the first principal component of the PC map. The separation of northern Na-Dene from Eskimos is also seen in the second component, though not as clearly. The Na-Dene and Eskimo may have migrated independently to America, or they may have separated in Beringia, or even in Alaska; it is impossible to solve this problem with the present data.

The question of dating these major migrations may be reconsidered again here. In our 1988 paper (Cavalli-Sforza et al. 1988), the divergence between all Amerinds and all northern Mongoloids is in slightly better agreement with the first date of entry proposed, about 35 kya, than with the second. Using the constant calculated in table 2.5.1 we obtain here the date of 31 kya. However, northern Mongoloids are a very diverse population, which underwent considerable internal movement in the last three centuries (Alexeev, pers. comm.). With mixtures and other complications, the divergence between the average Siberian and the average Amerind is likely to be greater than the divergence of Amerinds and their direct Asian ancestors. It is also likely that some of the Siberian populations that remained in Siberia were exposed to more severe environmental conditions and decreased in size, undergoing even greater drift. In any case, our attempts at identifying one Siberian group closer to Amerinds have not been successful. On the basis of relatively few markers (6 loci), Spitsyn (1985) found that among all Siberian peoples, the Tungus, Even, and Yakut located in the northern part of central Siberia are genetically closest to the Athabaskan. The Asian ancestors of Amerinds may have come from a relatively small region, and their Asian descendants may now be diluted by admixture with other less closely related ones, to the point that they are no longer easily recognizable. It is also possible that the majority of the Asian ancestry of the American pioneers has effectively left Asia, as happened, for instance, for Eskimos. All these considerations, and the expectation of high drift in regions of very low density like Siberia, would tend to increase the distance between

Siberians and Americans and thus lead to an overestimate of the time of passage. One may also consider that these are dates of separation, presumably on the Asian mainland, and the date of passage may be later. There are several causes of uncertainty and a dating based on the divergence of Amerind from northern Mongoloids cannot yet be given complete confidence, but we are clearly within the range suggested by archaeology.

From the tree in figure 6.9.1, the genetic separation between Na-Dene and Eskimo is a little more than halfway between the separation of the Arctic group and the Amerind group. If the first is taken as representing the separation between Amerinds and Northeast Asians, for which we have a not completely convincing estimate (31 kya) discussed above, then the date of separation between Na-Dene and Eskimo, probably still in Asia, is about 18 kya. Note that this is not necessarily a date of entry to America of one, the other, or both: separation may precede entry, which may be later but perhaps not by a large amount.

The tribes were grouped in this section according to a linguistic criterion, modified to some extent by a geographic one. In the next section we consider the tribes that are better known genetically, independent of the linguistic grouping used in this section, for North and South America.

6.10. Phylogenetic analysis of individual tribes

The tribes tested for the greatest number of genes in our data files are here individually analyzed. Considering first North and Central America, we have the genetic tree in figure 6.10.1, based on a sample of 17 populations with an average number of 62.7 ± 5.8 markers. Table 6.10.1 shows the F_{ST} genetic distances. Cree and Naskapi, which are very similar linguistically (Voegelin and Voegelin 1977), were pooled; even after pooling, they remain the group with fewest genes.

The Arctic cluster has the same structure as before, with northern Na-Dene (Athabascan and Dogrib) connected to Eskimo, but separating from them in the first split; USSR Eskimos are the most peripheral of the Eskimo cluster.

In figure 6.10.2 the same genetic data are presented as a PC map which accounts for 59% of the original genetic variation. The clusters indicated are linguistic groups and are discussed further in the next section.

The analysis was repeated for 30 populations from South and Central America, including Central American linguistic groups because of the extensive linguistic similarities between some of them. The results of the analysis are shown in figure 6.10.3; distances are given in table 6.10.2. The average number of markers was 61.4 ± 5.7.

Difficult problems arise in the interpretation of this tree. The PC map from the same data (not given) does not bring any clarification. No simple geographic or linguistic correlation is found at first sight, a fact to be discussed further. The amount of genetic drift that has been going on for 10 ky has not abated even today, given that population densities in most of the area are still very small and may even have become smaller in some cases. There has clearly been an extensive geographic movement of tribes, as shown by, among other things, the fragmentation of the linguistic map, and also by modern ethnological observations. There also must have been in the past, and there certainly is at present, a complex network of genetic exchanges within and between tribes, which has been studied in detail only for

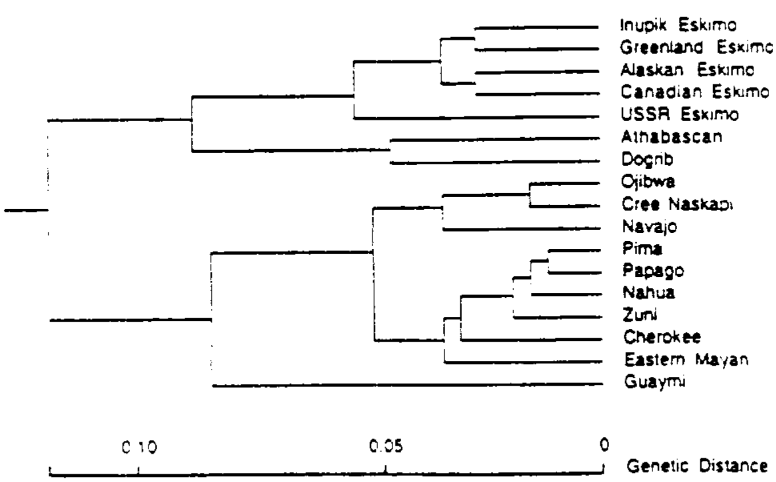


Fig. 6.10.1 Genetic tree of 17 single tribes or geographic groups of tribes from North and Central America.

DOI 02203

Table 6.10.1. Genetic distances (in the lower left triangle) and their standard errors (in the upper right triangle) among North American tribes or small groups of them (all values x 10,000). Triangles indicate relatively homogeneous groups.

	ECA	EGR	EIN	EUS	ESR	ANN	ANP	API	MSG	NCD	NAT	NNA	ACN	AOJ	KCH	PZU	PME
ECA	0	81	51	74	63	303	142	139	333	180	111	246	99	140	235	193	322
EGR	302	0	83	77	140	305	315	210	309	308	175	324	291	289	347	236	334
EIN	326	278	0	95	187	420	356	287	544	145	318	288	255	352	371	403	419
EUS	259	330	417	0	133	257	227	241	361	147	129	155	220	175	255	245	327
ESR	407	677	799	351	0	198	247	178	414	302	212	258	201	178	301	213	233
ANN	873	994	1285	675	875	0	40	38	266	334	200	79	148	146	62	47	142
ANP	951	1304	1509	1105	1289	174	0	32	216	211	232	66	80	142	118	55	94
API	814	953	1228	908	865	160	104	0	184	178	192	59	92	106	67	46	145
MSG	1882	1710	2615	2013	1952	714	793	705	0	578	591	300	381	375	317	111	179
NCD	775	1113	734	913	1298	1399	1445	1065	3153	0	66	174	202	308	396	307	131
NAT	754	746	952	676	889	866	810	791	2687	451	0	95	297	322	460	324	189
NNA	872	1253	1072	650	817	297	366	295	1123	629	369	0	55	115	100	123	82
ACN	535	1133	1117	939	930	406	331	340	1191	866	913	220	0	56	100	233	130
AOJ	619	991	1269	717	755	731	611	509	1277	1299	1236	433	157	0	122	221	162
KCH	1177	1168	1726	959	1277	170	266	314	565	2339	1489	540	374	496	0	101	180
PZU	1097	1088	1523	1251	1503	234	210	160	594	1682	1514	441	715	848	422	0	63
PME	1288	1342	1725	1037	848	278	347	317	611	1434	857	384	428	567	399	260	0

Note.– ECA, EGR, EIN, EUS, ESR, Eskimos (see table 6.9.1). ANN, Nahua, ANP, Papago, API, Pima; MSG, Guaymi; NCD, Dogrib; NAT, Athabascan; NNA, Navajo; ACN, Cree, Naskapi, Montaguais; AOJ, Ojibwa, KCH, Cherokee, PZU, Zuni, PME, Eastern Maya

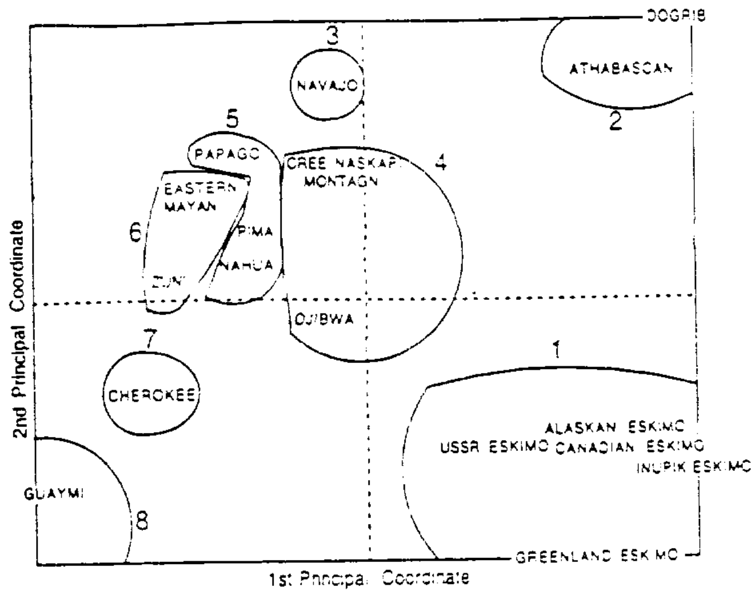


Fig. 6.10.2 Principal-component map of Northern and Central Amerind tribes or geographic groups. The clusters refer to linguistic groupings: 1. Eskimo; 2. Northern Na-Dene; 3. Southern Na-Dene; 4. Almosan; 5. Central Amerind; 6. Penutian; 7. Keresicuan; 8. Chibchan.

two tribes (Yanomame, Makiritare). These investigations are the only ones from which a model can be derived. One wonders how much one can generalize the conclusions reached for these examples, but it is encouraging to have excellent data even for only a few populations, which have not been seriously affected by contact with latecomers, or at least have shown little if any tendency to acculturation. The Yanomame may have originated at a considerable distance from their present location in the upper Orinoco (see fig. 6.6.1), probably in Panama (on the basis of linguistic considerations). They are still moving and expanding (Chagnon 1983). The story that emerges from the Yanomame or Makiritare is one of many scenarios which must exist in South America. It certainly should not be extended to regions with a long history of formation of towns or cities or even villages having a totally different demographic and mating structure. Rather, the Yanomame are a model for populations living as primitive horticulturists in the American tropical forest, which is a significant fraction of Central and South America.

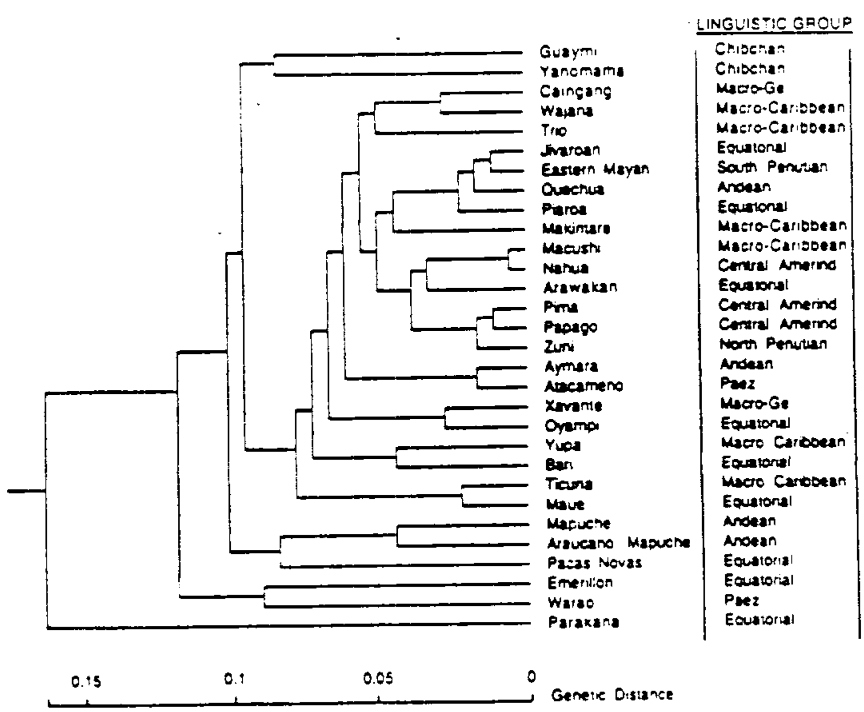


Fig. 6.10.3 Tree based on genetic distances of 30 South and Central or near-Central American Indian tribes.

Table 6.10.2. Genetic Distances (in the lower left triangle of the matrix) and Their Standard Errors (in the upper right triangle of the matrix) among Central and South American Tribes (all values x 10,000)

	ANN	ANP	API	MCA	MWR	MCY	MSG	PZU	PME	AAM	AAY	AMA	AQU	EAR	EBR	EEM	EJI	EMA	EOY	EPN	EPR	EPI	MMA	MMK	MTR	MWA	MYU	MGC	MGX	MTT
ANN	0	42	36	148	220	217	278	43	161	218	177	247	90	49	178	272	79	125	146	307	328	121	38	142	228	59	115	90	128	154
ANP	174	0	31	207	319	178	198	49	98	323	204	200	74	123	218	278	239	212	129	547	207	292	198	192	135	95	176	148	143	223
API	160	104	0	174	297	169	180	58	154	343	213	193	79	132	127	267	130	154	89	482	320	79	211	188	197	110	138	179	141	178
MCA	478	688	440	0	244	429	428	188	210	368	47	184	69	195	327	328	73	124	220	229	341	68	228	195	224	132	298	228	109	237
MWR	688	1168	1033	1102	0	378	857	250	318	438	146	436	211	228	236	245	333	539	263	538	638	415	258	241	397	239	318	195	458	408
MCY	889	921	701	1438	1368	0	238	121	204	728	267	252	227	343	274	254	194	181	198	463	498	380	175	325	279	292	260	348	732	239
MSG	714	793	705	1341	1529	788	0	110	189	669	192	185	201	333	391	372	180	199	157	267	773	344	142	335	287	302	648	340	484	309
PZU	234	210	160	558	1090	584	594	0	60	328	155	154	54	138	173	255	157	115	108	407	228	223	161	182	108	111	208	184	307	146
PME	278	347	317	585	1093	848	611	260	0	203	267	133	48	71	194	258	27	129	113	247	248	73	90	101	81	137	173	188	72	187
AAM	898	1398	1404	1212	1947	2807	2088	1180	822	0	343	263	390	480	491	447	195	403	653	188	471	285	334	673	192	188	321	380	289	373
AAY	414	598	499	190	578	910	844	469	716	1153	0	165	140	139	253	261	175	183	205	422	398	229	170	178	261	133	248	190	443	182
AMA	604	726	610	642	1422	1160	717	551	488	448	632	0	147	500	457	370	123	155	282	374	423	118	141	224	342	405	458	504	151	225
AQU	326	303	268	403	1045	941	778	208	150	940	462	466	0	112	195	280	71	128	104	358	255	63	135	148	102	97	158	150	141	219
EAR	168	480	548	895	531	854	821	598	375	1299	613	1399	585	0	152	227	149	228	115	267	507	228	92	218	111	109	224	89	168	353
EBR	545	639	477	1041	1170	714	809	584	697	1874	962	1381	855	392	0	358	162	155	235	304	671	251	157	222	373	318	148	205	221	260
EEM	834	1281	1185	1515	835	1361	1280	1146	1113	2049	1267	1713	1240	732	1292	0	434	378	308	447	540	638	258	204	282	230	333	202	271	337
EJI	244	650	329	285	1129	910	813	525	102	674	525	458	206	480	533	1240	0	137	110	103	275	71	82	58	172	160	198	141	490	149
EMA	329	852	593	430	1763	912	970	452	448	873	700	682	533	879	739	1401	429	0	207	244	290	148	128	165	136	217	170	343	318	80
EOY	588	895	578	1077	807	993	621	604	654	1953	1061	1227	552	418	963	854	468	895	0	299	398	227	108	183	140	114	225	68	113	408
EPN	618	1368	1284	1013	1722	1849	1408	1388	778	553	1405	1064	1059	814	1318	1735	643	1065	1257	0	625	164	238	198	288	193	161	154	658	427
EPR	1188	1148	1498	1554	2644	2094	2731	1316	980	1689	1783	2169	1359	1638	2227	2311	1550	1123	1848	2619	0	343	230	295	159	398	408	624	353	327
EPI	780	854	433	342	1887	1580	1195	710	351	898	682	618	239	1112	1082	1738	218	687	845	842	1881	0	118	138	198	185	347	178	287	178
MMA	82	835	394	852	995	490	597	537	147	1008	825	445	356	447	423	1047	294	417	497	888	1263	658	0	77	183	140	150	192	471	74
MMK	635	726	708	858	1110	1722	980	854	459	2072	922	1024	522	532	1328	1045	317	878	649	823	1665	489	383	0	98	105	238	102	536	304
MTR	455	882	887	760	1241	990	1010	500	249	782	802	1045	423	431	1004	1107	488	565	718	1028	853	872	470	602	0	120	275	230	419	228
MWA	203	458	505	650	768	1359	1028	548	432	415	547	1198	385	328	920	781	505	909	353	823	1329	753	712	497	298	0	198	111	443	337
MYU	489	767	897	1168	1040	850	988	713	558	707	933	1115	782	348	455	1111	528	778	925	731	1843	1009	898	1267	798	824	0	133	222	268
MGC	293	884	872	915	812	1328	1128	991	821	709	827	1534	749	328	842	887	488	1152	352	738	2105	880	788	655	727	271	437	0	359	478
MGX	294	501	657	682	1281	1825	928	880	267	888	1027	745	494	507	712	915	884	841	287	1498	1641	750	1092	1058	1144	813	855	705	0	552
MTT	674	824	718	953	2051	1080	1251	895	825	1321	907	925	820	1052	807	1882	397	240	1572	1267	1487	727	300	1287	835	1058	1043	1371	1311	0

Note.– ANN, Nahua; ANP, Papago; API, Pima; MCA, Atacamleno; MWR, Warao; MCY, Yanomama; MSG, Guaymi; PZU, Zuni; PME, Eastern Maya; AAM, Araucano; AAY, Aymara; AMA, Mapuche;AQU, Quechua; EAR Arawakan; EBR, Bari; EEM, Emerillon; EJI, Jivaroan; EMA, Mane; EOY, Oyampi; EPN, Pacos Novas; EPR, Parakana; EPI, Piaroa; MMA, Macushi; MMK, Makiritare; MTR, Tria; MWA, Wajana;MYU, Yupa; MGC, Cakiguap; MGX, Xavante; MTT, Ticuna.

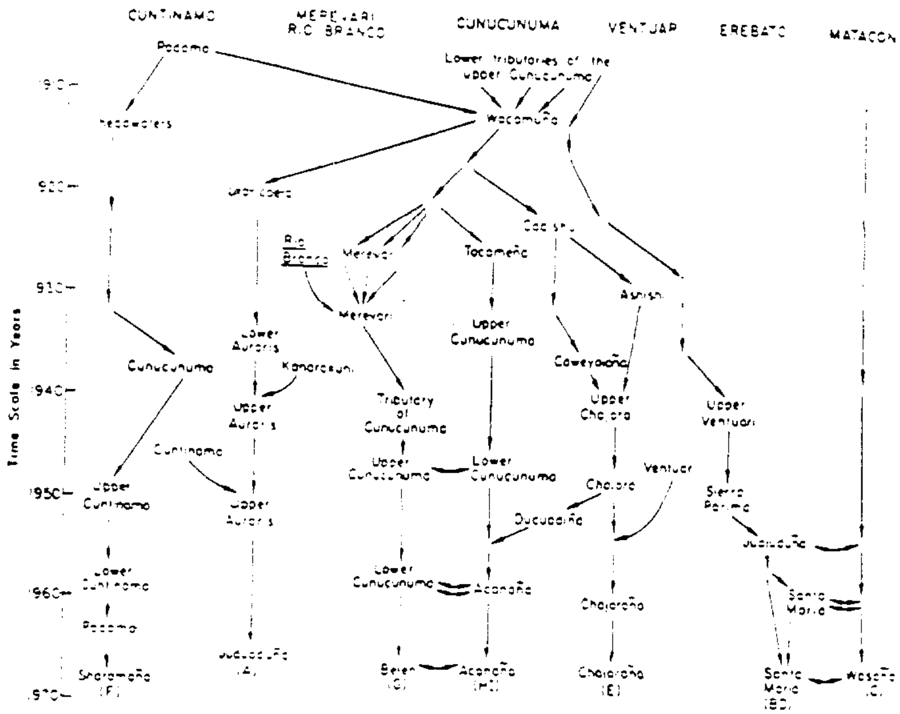


Fig. 6.10.4 History of fissions and fusions of seven Makiritare villages over 60 yr (from Ward and Neel 1970, fig. 1). The phylogenetic tree calculated from genetic similarities between villages can, to some extent, reconstruct the actual history of fissions (but not that of fusions which would demand other approaches. The changing temporal-spatial relationships are indicated by reference to the time scale on the left; headings at the top represent the six main geographical areas.

The history of fissions and fusions for the last few generations of the Makiritare (see fig. 6.10.4) shows a structure of very small groups, on the order of 100 individuals each, which split and reunite, to some extent according to kinship lines. Kinship groups, however, are not necessarily stable entities when viewed over several generations, and the whole picture is one of incomplete randomness of splits and fusions that is not easy to model quantitatively. The genetic variation between villages is about twice what would be expected (Wagener 1973) on the basis of the observed proportions of migration, assuming that migrants are a random sample of the population. Thus, drift is higher than expected from the observed migration and population size, probably because splits and perhaps reunions tend to follow kinship lines and are therefore not random (Smouse et al. 1981; Smouse 1982), as in regular models of population structure.

Another source of amplification of drift effects is strong differential fertility, especially of head men (Neel and Weiss 1975; Neel 1980). The Makiritare are largely endogamous within the village, and even more within the tribe, but give and receive nontrivial genetic contributions to and from neighboring tribes, usually of different linguistic groups since several tribes moved a long distance from their origin. One cannot exclude the possibility that immigrants from other tribes have closer kinship ties with the tribe, decreasing the outbreeding effect caused by mating with members of other tribes. Considering the frequency with which women are raided from other tribes, a certain amount of random or nearly random outbreeding with neighbors must also occur.

There is only limited information on other intertribal migration. According to a summary of information by Salzano and Callegari-Jacques (1988), genetic exchange is considerable, and it is higher for tribes at a more advanced economic level. Their tabulation does not distinguish between genetic exchange with neighbors and with different tribes. In unpublished data collected with H. Groot and A. Espinel in Colombia, genetic exchanges between different tribes on the upper Orinoco became very high at the end of a long period of intertribal hostilities; in a small area investigated near Puerto Inirida, it was difficult to find marriages where there had not been recent admixture between different tribes. The memory of genetic exchanges in older generations is frequently lost, and such findings make one suspicious about the real isolation of many South American tribes, at least today. Yet, there is enough genetic variation between South American tribes that some degree of isolation must have been maintained in many instances for a long time (Neel and Ward 1970). Our capacity to understand the genetic structure of southern Amerindian tribes can only benefit greatly by an extension of studies like those already cited by the Neel group, before they are made totally impossible by the disruption and disappearance of traditional customs.

6.11. Comparison of Genetics with Linguistics and Geography

In section 6.9 we have seen that genetic analysis fully confirms the division of American natives into three major clusters, Amerinds, Na-Dene, and Eskimos, which are also clearly distinct linguistically. The hypothesis that they correspond to three major migrations, all from Siberia via the Bering region, is in agreement with current archaeological knowledge, despite present uncertainties on dates. The general picture seems reasonably well established and further analysis given in section 6.10 has clarified possible doubts arising from the ambiguous position of the southern Na-Dene. Their geographic position and the peculiar genetic relationships with other Na-Dene and with Amerinds are best explained by admixtures with the latter that must have accompanied their southern migration. When we come to consider Amerinds, we find greater difficulties in fully reconciling genetic data with information provided by other approaches. In part, this is caused by the poverty of information. At this point however, we must summarize two previous investigations that show without doubt that Amerinds, too, provide good evidence of a strong correlation between genetics and language.

The first is an extensive analysis of the relations between the genetic, linguistic, and cultural similarities of 53 North American Indian tribes carried out by Spuhler (1979). The analysis used a subset of 13 gene frequencies from *ABO*, *MN*, *RH*; Diego blood group was tested only for a subset. Of seven linguistic groups: Arctic-Siberian, Na-Dene, Macro-Algonquian, Macro-Siouan, Hokan, Penutian, and Aztec-Tanoan, 34 (64.2%) of 53 tribes tested were classified correctly using gene frequencies. This indicates a substantial agreement between linguistic and genetic data, but also a number of discrepancies. Most misclassifications in the Spuhler sample are found among Na-Dene, Macro-Algonquian and Macro-Siouan, and in the Hokan group. In Spuhler's analysis by culture areas (Arctic, Subarctic, Northwest Coast, Plateau, California, Plains, Southwest, Northeast, Southeast) 31 of 53 tribes were correctly classified, or 58.5%. Considering that more groups were tested in the latter case, the two approaches gave approximately equivalent results. In conclusion, there is substantial, even if imperfect, agreement between genetic and linguistic or cultural classifications. Some of the discrepancies, especially that of northern and southern Na-Dene are of interest; note, however, that Apache and Navajo are not misclassified in Spuhler's analysis. The statistical approach used by Spuhler (stepwise discriminant analysis) is different from the usual one of calculating correlations between genetic and linguistic (sometimes also with geographic) distances. Moreover, we use more genes and fewer tribes. In figure 6.10.2, tribes belonging to the same linguistic group are circled. There clearly is a reasonable, though not perfect, agreement between a linguistic and a genetic classification. The small numbers do not permit a completely satisfactory assessment of the correlation. The incomplete agreement indicates that the estimates of genetic and linguistic similarities may need improvement. It may also result from frequent language or genetic replacements. In fact, these explanations are not mutually exclusive, and to some extent, all may have contributed to reduce the correlation without completely destroying it.

Using other more conventional approaches, Spuhler (1972) found no evidence of correlation between genetic and linguistic distance. This negative result may be more of an indictment of the method than of the general correlation between linguistics and genetics. A linear correlation can easily be destroyed by some outliers. The expectation of linearity may be naive when there is a complex fission and fusion pattern; simulations may be appropriate for a comparison of the different methodologies. However, Spuhler (1972) reanalyzed the same data by an analysis of variance, which escapes the strictures of linear-correlation analysis, and found that the variance of genetic distances among linguistic stocks is significantly higher than that within linguistic stocks. This is in line with his result by discriminant analysis. It is worth adding that Spuhler found a moderate but significant correlation between genetic and geographic distances and none between linguistic and geographic distances.

Apart from Spuhler's studies of the genetic-linguistic correlation on the North American continent, there have been many investigations of limited regions or groups of Central and South America. An early one by Spielman et al. (1974) compared the linguistic distances among seven Yanomami dialects and genetic distances among the people occupying the corresponding geographic areas. The matrices of genetic distance, distance calculated from lexical data, and from grammatical data showed in all three cases a significant congruence.

Chakraborty et al. (1976) found no linear correlation between genetic distances and linguistic distances in seven Chilean "highland" Andean populations. Linguistic distances were calculated on a scale based on an early classification by Greenberg. The scale of linguistic distance used may be responsible for the failure.

The same measurement of linguistic distance was used by Murillo et al. (1975) to compare linguistic and genetic distances of the Chipaya of Bolivia to nine South American Indian tribes. They found no correlation.

Salzano et al. (1977) investigated the intra- and intertribal genetic variation within the Ge-speaking Xavante, Kraho, and Cayapo of Brazil. They conclude that the average intertribal genetic distance within this linguistic group is about 63% as great as that between tribes speaking more differentiated languages. They found, however, a weak linear correlation ($r = 0.27$) between genetic distances and cognate percentages in a list of 100 words.

A very thorough and detailed study has been published recently by Barrantes et al. (1990) on the Chibchan-speaking groups of Costa Rica and Panama. Ten such populations were analyzed for 48 genetic loci. The genetic distances between pairs of populations were correlated to the linguistic distances based on cognate percentages. The observed correlation ($r = 0.74$) is high and highly significant, higher than that observed for genetic and geographic distances ($r = 0.49$, not significantly different from zero) and for geography and linguistics ($r = 0.52$, significant at $P = 0.05$).

When we look at figure 6.10.3, we are unable to find a simple interpretation linking genetics and linguistics in the whole of Central and South America. A similar failure is experienced in the related tree given for South America in Salzano and Callegari-Jacques (1988). It seems likely that, in these circumstance, a tree is highly inappropriate for detecting the correlation of interest, but it is also possible that the data are inadequate.

Even the usually strong relation between genetic and geographic distance is blurred in South America. The correlation calculated between the two distances is 0.191 ± 0.048 (standard error calculated by bootstrap). It is positive but low, and confirms the results obtained by plotting the genetic distance between population pairs against their geographic distance (sec. 2.9). Linguistic distance between families showed a negative correlation with genetic distance (-0.139 ± 0.051) and with geographic distance (-0.212 ± 0.051). These results (Minch and Cavalli-Sforza, unpubl.) will need further investigations.

There are many reasons why the correlation of linguistics with genetics and also with geography is especially difficult to study in South America. Part of the problem is tied to the major territorial, economic, and political changes that have taken and are taking place in South America, causing an epidemic of language extinctions that must have been especially dramatic in the last century and earlier. For instance, in Ruhlen's (1987) list, 71 languages of the 117 (61%) that form the Ge-Pano-Carib subfamily are extinct. Similar high percentages apply to many other subfamilies of South and Central America: Equatorial 67/145 (45%), Tucanoan 12/47 (26%), Andean 12/18 (67%), Chibchan 25/43 (58%).

Languages often become extinct when population numbers become too small, or when there is government pressure to expand those of another language, but this does not mean that the people also disappear. In fact, it seems reasonable to assume that in the modern situation, with the continuous shrinking of groups, an increasingly larger proportion of people stop speaking the traditional language and replace it, either with languages imported by the colonial powers or with more widely spoken, traditional languages from other groups. This would certainly contribute to the destruction of the correlation of languages and genes. There may be other important reasons that deserve more research.

One should remember that, as we have already discussed (sec. 2.9), American Natives show an extremely high geographic mobility, as measured by the relationship between genetic distance and geographic distance. Mobility is also detected by studying the distribution of language groups, which is extremely fragmented, with subfamilies forming very complex, interpenetrating patterns. This might be enough to destroy linear correlations between geographic and linguistic distances, and between genetic and linguistic distances. The ecological situation also contributes to this result: the Andean chain forms the backbone of the continent and is very different from the east. It runs from the extreme north to the extreme south and is relatively similar ecologically in spite of the great variation in latitude. It is occupied by people who are also relatively homogeneous genetically, as well as linguistically: only two major subfamilies of the nine spoken in the whole subcontinent occur in the Andean chain today. By contrast, the flatter, eastern part is more heterogeneous genetically and linguistically. Linear correlations are especially unsuitable for measuring the association among geographic, genetic, and linguistic distances in this case. Detailed studies of single linguistic groups that have not undergone too many disruptions and extinctions—for example, the Chibchan (Barrantes et al. 1990)—are best suited for showing the correlation between genetic and linguistic variation. Studies of other groups, that have not been excessively impoverished by extinctions may also be useful.

The studies of correlation between genetics and linguistics in America can give only a very partial answer to the general problem. Of the seven studies we have listed, only one that used linear correlation gave satisfactory results. One can see many reasons why this can happen even if there is a general congruence between the two phenomena. Other methods have given positive results when linear correlation failed. Moreover, even if this is generally overlooked, significant testing of linear correlations between distances calculated between pairs of populations is unsatisfactory because there is usually an internal correlation between the pairs. This does not apply to the sample by Murillo et al. (1977) in which the pairs of populations are independent. For further comments see Cavalli-Sforza et al. 1992.

In summary, three of seven studies favor the hypothesis of congruence between genetics and linguistics but for methodological, theoretical, and historical reasons, one may expect this type of analysis to fail in the Americas, especially using linear correlations. Further work on American data with more refined methods is clearly necessary.

6.12. Geographic Maps of Single Genes

The *ABO* system is remarkably different in America from other parts of the world: Amerinds are unique in having almost completely lost the *A* and *B* alleles. By contrast, *A* is conserved among Na-Dene and shows a remarkably high frequency among some Almosan, whereas among Eskimos, the *A* and *B* frequencies are much more similar to those of the rest of the world. Thus, the *ABO* locus is a fairly good, though not a perfect mirror, of the three major postulated migrations.

The reasons for the loss of one or two alleles of this system, which are present at relatively constant frequencies in all other world populations—and to some extent also in many Primates (Socha and Ruffié 1983)—are not entirely understood. The extent to which random variation in gene frequencies affected Amerind populations will be clear from several other examples in this section and suggests that genetic drift played a very important role in America. Did drift determine the irregularities of gene frequencies in America because of a very low number of initial migrants (an initial founder effect), or later bottlenecks, and perhaps persistence of low numbers for long periods? We may anticipate that the behavior of *HLA* loci indicates that the second or third hypothesis may be true, and that many tribes originated from a very small number of founders. Instead of the many alleles of an *HLA* locus commonly found elsewhere, even in small populations, a particular Amerind tribe has only a few alleles at a disproportionately high frequency, with other alleles rare or absent. In another tribe the same rarity of most alleles except a few is observed, but the frequent allele/s are different. This remarkable phenomenon is therefore unlikely to be due to natural selection, given its magnitude, or to the initial founder effect of a small number of first migrants from Asia. *ABO* has far fewer alleles than *HLA*, but in a way there is a somewhat similar phenomenon: an excess of *A* in a few groups, and an excess of *O* (up to 100%) in all the others. A high frequency of *B* is almost never found.

Even if there is a good chance that drift was responsible, at least in part, for the anomalous distribution of *ABO*, it is difficult, if not impossible, to exclude the effects of natural selection. As we have seen in section 2.10, *ABO* phenotypes (or genotypes) react differentially to many infectious diseases, and a popular explanation for the loss of *A* and *B* alleles among Amerinds is differential sensitivity to syphilis, because *O* individuals are more resistant. The origin of the hypothesis is the belief that syphilis was endemic in Central America in the fifteenth century and was spread to Europe by the crew of Christopher Columbus after their return to Spain. The evidence from direct studies of patients (Mourant et al. 1983) showed that *O* individuals heal more rapidly (as judged on the basis of immunological tests) after treatment with chemotherapeutics. The dates and geography of the European epidemic beginning shortly after the return of Columbus' crew correspond to the expectations of the hypothesis, but others have claimed that the disease originated in Africa from a closely related spirochete responsible for yaws, a nonvenereal disease (McNeill 1976). A search for a correlation between yaws and *ABO* was negative (Cavalli-Sforza 1986b).

The geographic distribution of the *ABO* alleles shown in the maps deserve some comments. Because of the rarity of *A* and *B*, and the omnipresence of *O*, all gene-frequency distributions are very skew. In North America *O* is lower, with allele *A* being high and reaching a peak above 45% (almost all *A1*) in western Canada. Elsewhere, *O* is almost never less than 50%. In the extreme south, there is a small patch with a maximum of *A* greater than 10%, and a corresponding trough in *O*. Greenland is also high in *A* (*A1*).

In Eskimos, *B* shows a peak in eastern and southern Canada, where *O* is low and there are also traces of *A*.

Apart from Eskimos, the simultaneous presence of *A* and *B* in proportions of 4:1 is a strong indication of admixture with Caucasoids. This is likely to be the case on the eastern coast of Canada, but the absence of *B* in the western part of Canada, despite the high frequency of *A*, is proof that this is not due to admixture. If Negroids were the donors of *ABO* genes, which is not the case in Canada, the proportion of *A* to *B* would be lower than for white admixture. We have tried to avoid using data from mixed populations but we will see that in the eastern part of the United States and Canada a fair number of mixed groups are present. More intensive contacts with Europeans occurred in this area and, therefore, it is not surprising that it is difficult to find "full-blood" (or even only 3/4 blood) Amerinds.

Variograms of *ABO* alleles have long initial linear segments, with rather small slopes.

Acid phosphatase (*ACP1*B*) shows an almost regular gradient from north to south. The distribution is almost bimodal, reflecting the major difference of Eskimos and Amerinds from the extreme north versus the rest of the continent. The variogram is approximately linear up to 4000 miles, with a fairly large slope.

Adenylate kinase *1* (*AK1*) is, like *ABO*, a marker of Caucasoid admixture. The less frequent allele, *AK1*2* has a frequency of about 5% among Europeans and is essentially absent in other populations. The band of low *AK1*1* (<97%) across the North American continent indicates Caucasoid admixture. It confirms and extends the observations with *ABO*. The variogram is uninformative and is not reported.

The Diego blood group (*DI*A*) is of special significance in America. It was first found in Amerinds.

DOI 02210

in which, as the map shows, the *A* allele varies from less than 5% to more than 35%; it is also found in some northern Mongoloids but at a lower frequency. It must therefore have originated in Northeast Asia. Its considerable variation in America is most probably due to drift. The maximum is in northern Brazil, but it is rare or absent in North America. The initial slope of the variogram is fairly high, and the linear portion is less than 1000 miles.

The Duffy blood group (*FY*) varies considerably with allele *A*, showing a maximum in the Arctic. The distribution spans almost the complete range, but is concentrated between 40% and 100%. Allele *B* has been studied much less extensively; it peaks with more than 40% frequency between northern Brazil and the Guianas. The variogram of allele *A* is fairly regular, whereas that of *B* has a strongly negative initial slope.

Allele *1* of esterase D (*ESD*1*) shows a maximum in Mato Grosso (southern Brazil) and the Paraguay basin, as well as in Central America; it also shows an absolute minimum in the extreme eastern part of Brazil. The variogram has a large slope and is linear until about 1500 miles.

Glyoxylase-1 allele *1* (*GLO1*1*) has a maximum in Central America and low values in South America; the regular decrease toward the north is artifactual and caused by the near absence of data in North America except in the extreme north. The variogram is approximately linear for almost 2000 miles with a largish slope.

The group-specific component or vitamin-D-binding protein allele *1* (*GC*1*) shows a minimum in central Brazil and a relative maximum farther west; the variogram is irregular, possibly because of the closeness of the minimum and maximum. The electrophoretically fast subtype of *GC*1*, *GC*1F* has two peaks on the western coast of South America, a relative minimum in the extreme south and one in the extreme north. The variogram shows a complex form.

Haptoglobin (*HP*1*) also has a very wide distribution, with gene frequencies ranging from 0% to 100%, with a mean of 55%. The peak is in the extreme south, but there are other secondary peaks in South America; the lowest values are in the extreme north. Basically, there is a north-south gradient, which, in the present case, cannot be attributed to climate. The variogram has a relatively short initial portion with a positive slope.

Antigens specified by *HLA* genes have revealed an unusually narrow range of alleles, especially in South America (Black et al. 1980). Only *HLAA*2*, *A*9*, *A*28*, *A*30*, *A*31*, *A*33*, *HLAB*5*, *B*15*, *B*16*, *B*17*, *B*27*, *B*35*, and *B*40* have average frequencies significantly different from zero. This restricted range of polymorphism is expected when the genetic diversity of an ancestral population has been reduced several times by passage through size bottlenecks.

A possible effect of selection should also be considered for *HLA*: in fact, evidence for heterosis in South American Indians had been advocated by Black and Salzano (1981), who found that, in a subpopulation of 122 people whose parents' *HLA* haplotypes were known, there were 56% fewer homozygotes than expected. If this phenomenon is due to differential mortality, it can be efficiently studied only in the few populations still subject to high prereproductive mortality.

The most frequent *HLAA* allele is *A*2*—37%, on the average—and it reaches maxima over 50% in southwestern North America and in Venezuela, with minima in the northern Andes and in eastern Greenland. The distribution is likely to have at least two modes. *HLAA*9* has an average frequency around 31% with a peak over 80% in eastern Greenland and the northwestern Arctic. A secondary peak (over 50%) is found in the northern Andes, whereas the rest of South America has frequencies below 20%. The distribution seems bimodal. Allele *A*19* has an average of 17.5% and a peak of more than 40% in northern Chile, with low frequencies north of Colombia. With an average frequency of 10%, *A*28* has a peak near 40% in the extreme south. Averaging only 1%, *A*30* has a peak of more than 4% in the southeastern United States. A subtype of *A19*, *A*31*, averages 15%, reaching more than 40% in northern Argentina. Again, the distribution seems bimodal. Although it has a maximum above 15% in the southeastern United States, *A*33* averages 1.8%.

With an average frequency of 12% and a peak over 50% in eastern Venezuela, *HLAB*5* has a secondary peak in eastern Greenland. Although its mean frequency is 1%, *B*7* reaches values above 10% in the western Arctic Ocean region. Allele *B*14*, with a 0.8% average, has a frequency greater than 10% in the southern Andes; and *B*15*, average 11.5%, has a peak in northern Chile. *B*16*, average 13%, has a peak in the north-central Andes greater than 50% and minor peaks elsewhere. *B*21*, averaging 1.5%, has a maximum above 10% in the extreme Southwest of the United States. *B*22*, with mean 0.7%, reaches more than 10% among central Eskimos. Well known for its strong association with ankylosing spondylitis, *B*27* has an average frequency of 3.8%, with a maximum above 20% in Alaska. It is interesting to note that the three tribes of the Southwest, the Pima, Papago and Zuni, have similar origins but significantly different frequencies for *B*27*. The most frequent *B* allele, *B*35*, has a 20% average and reaches about 70% in Brazil. With a mean frequency near 19%, *B*40* reaches over 50% among Eskimos of the western Canadian Arctic.

In sum, *HLA* shows great variation, most probably resulting from drift, like the other genetic systems, but as already noted, its multiallelic structure renders variation more evident. This genetic system is ordinarily represented by a great number of alleles in almost every population—even if very small—in the Old World, and all alleles tend to have relatively

DOI 02211

low frequencies. In the Americas, the situation is different. One or few alleles become definitely dominant in frequency, in one or a few tribes, sometimes reaching values above 50%, and the other alleles are correspondingly rare; but most populations are unique in that the dominant alleles differ from one to the other, sometimes even in neighboring populations. This is exactly what would be expected, at least qualitatively, under drift alone. In fact, in the total absence of cross-migration, drift would eventually lead to the survival of only one allele in each population. The surviving allele is chosen randomly from among those originally present, subject to the rule that the probability of an allele becoming the sole final survivor equals the initial frequency of that allele in the drifting population. Perhaps most alleles were represented at the beginning in Northeast Asia; many are still present in some, but not in all the other tribes.

Some alleles were probably lost, a few among the founders perhaps, but most in the process of evolution of individual tribes, as shown by the very different local patterns of each allele. It seems as if most local populations were started by such small numbers of individuals that they could only maintain two or three alleles at high frequency. Under these conditions, one does not need to postulate a very strong founder effect at the passage from Siberia to America (or even earlier). The remarkable variation among the Indian tribes of South America suggests the existence of a later bottleneck, perhaps more important than the first, if there was a first one. In other words, many alleles may have been present at the beginning and lost later. Only 17 alleles have sufficiently high average frequencies to generate maps of America; this is about half the number of European alleles, but one does not need to conclude that half of the alleles were lost. It is possible that there exist several undetected alleles, because the majority of reagents are of Caucasoid origin and do not necessarily detect all alleles present in other populations.

The variation with distance shows here, as in other *HLA* data, several negative or flat initial slopes: 5 of 17. The initial linear segments of those with positive slopes are in the usual range, and the initial linear portion may sometimes span 2000 miles.

GM (or *IGHG1G3*) also shows considerable local variation. The most common haplotype, *za;g*, varies from 40% to 100%, with several peaks and several minima. The next most important haplotype, *zax;g*, has a maximum in the center of South America and decreases almost regularly around it.

All the other *GM* haplotypes have lower average frequencies, but all show usually single, sometimes extreme peaks in different regions. Thus *za;b0b1b3b4b5*, a Negroid haplotype (very poorly represented in the maps for reasons of reagent availability), has an average frequency near 2%, but peaks at more than 6% in the Guianas where there is probable African admixture. An Oriental haplotype, *za;b0st;b3b5*, has an average frequency of 6% and peaks at more than 20% in Alaska and in Labrador. With an average of 1.6%, *fa;b0b1b3b4b5* has various peaks in the north and south, none too pronounced. A Caucasoid haplotype, *f;b0b1b3b4b5*, has an average frequency of 2.7% and peaks in Greenland and in the northern part of South America.

At first, one might be reluctant to believe that all these maxima and minima for *GM* haplotype frequencies are due to drift. One might hypothesize that this immunoglobulin marker reacts to local infectious diseases, and there is a little evidence for it as discussed earlier. However, drift is expected to operate with the same intensity for all markers. It is therefore likely that many *GM* gene-frequency peaks or troughs in America are due to drift.

The light immunoglobulin constant chain, *KM*(1&1,2)*, has a mean of 37%, with a wide distribution of 0% to 80%, minima in the north, but at least one in the south, and maxima around Panama.

The variograms of immunoglobulins tend to be irregular and uninformative. The Kell blood group (*KEL*K*) is a rare polymorphism almost homogeneously near zero. *KEL*Jsa* is also relatively rare (2% average), but shows a peak above 20% on the northern coast of South America. The Kidd group (*JK*A*) has a distribution of 0% to 80%, with minima in the extreme south and in the Panama region, and various maxima. Its complementary allele, *JK*B* is poorly studied directly; it shows a complementary maximum in Panama. The Lewis blood group *LE*Le* also varies greatly, from 10% to 100%, and has a maximum in Alaska. *LE*Le(a−)* has a maximum in a neighboring region, but has a much smaller range of variation. Almost all these blood groups have irregular variograms.

The *MNS* system shows somewhat less variation than other genes, judging by F_{ST} values, but the range of gene frequencies is not small. The *M* allele varies from 30% to 100%, and the *S* allele from 0% to more than 80%; both frequency distributions are probably unimodal, but both geographic maps are full of relative minima and maxima that span almost the whole range. Of the four haplotypes, only the rarest, *Ns* (6% average frequency), does not have a distribution extending from nearly 0% to nearly 100%; maxima and minima appear in regions already showing strong drift for other alleles, like the north-central Andes or the Arctic, or in new ones, like the coast of southern Brazil. All the variograms have positive initial increases with regular slopes, but with oscillations, except for *Ns* which is fairly flat.

The *P1* blood group, allele *1*, has a distribution varying from 5% to 100%, with a maximum in southern Chile and minima in many places, but mostly among Eskimos. The F_{ST} is elevated, and the variogram increases initially.

DOI 02212

Peptidase A (*PEPA*) is poorly studied and shows little variation: allele *2* has an average frequency of only 0.6%. The variogram is uninformative and is omitted. Taster (*PTC*T*) is poorly known in this part of the world; it varies between 30% and 100%, with maxima in southern Chile and the southwestern part of North America. Minima are among Eskimos. This geographic distribution is in some agreement with an advantage for tasters in an area where antithyroid substances containing plants may be common, at least to the extent that Eskimos, who eat essentially meat and fish, are less exposed to the danger. It is not clear whether the areas with highest frequencies of tasters have a particularly frequent occurrence of edible plants dangerous for thyroid function.

Phosphoglucomutase 1 (*PGM1*1*) varies from 55% to 100% with a mean of 83.5 for allele *1*; the maximum is in Venezuela, but a secondary peak is found in the Na-Dene region. There are various minima and an irregular variogram, as is almost usual. *PGM2* is less well known and, in any case, shows less variation, being confined to 80%–100% for allele *1*. A minimum is in the extreme south. The variogram of *PGM2* is uninformative.

6-phosphogluconate dehydrogenase (*PGD*) shows a low frequency of allele *B*, with some anomalies in northeastern North America and in northern Chile. Allele *C* is represented on the map, and *B* has the complementary pattern. The variogram has a moderate slope.

The *RH* system is also highly variable. Alleles *C* and *E* span essentially the whole range, while *D* is less variable, having, on the average, 96% frequency. *C* peaks in Panama and is lowest in the Arctic; *D* is universally high everywhere except for minima on the eastern coast of North America (possibly reflecting Caucasoid admixture, since Europeans have the highest world frequencies of the *d* allele [*RH-*]). *E* peaks in the Arctic and in the Andes; it is minimal in Panama.

The most frequent *RH* haplotypes are *CDe* (52%) and *cDE* (36%), and both span almost the entire 0%–100% range. The first peaks in Panama, and the second, in the Arctic. Next in frequency are *CDE* (4%), which also has several relative maxima in North and South America, up to about 30%; and *cDe* (4.6%), which peaks in the Southwest of North America. Ordinarily *cDe* is a good marker of Negroid admixture, which, however, seems very unlikely in the Southwest. The *cde* haplotype is, on the average, 2.5% and can be taken as a good indicator of Caucasoid admixture; not surprisingly, it shows a peak up to 20% on the eastern coast of North America, where we have seen other signs of admixture. It is uncertain if the relative maximum in the extreme northwest of Canada should also be interpreted as a result of Caucasoid admixture, because the other possible markers do not confirm it. Two rare haplotypes, *Cde* and *cdE*, show minor variations. Haplotype *cdE* surpasses 3% in a small area of Mexico and reaches 1%–2% in the extreme south of South America. *Cde* shows very low maxima in Mexico and in the Southwest of the United States. Their maps are omitted. The variograms of *RH* show less extreme oscillations around the curve than most other American alleles, probably because of the greater number of data, and slopes are fairly large on the average.

The secretor locus (*SE*) varies from less than 40% to 100% in frequency of the *Se* allele and has a maximum around the equator. Parts of the map are not supported by data and are unlikely to represent real variation: for example, the maximum in Florida, which is extrapolated from the high Mexican values, and the maximum in the extreme south. The minimum in Brazil seems well supported and is not surprising given the high drift observed throughout America.

Transferrin (*TF*) shows a few troughs of the common allele *C*, where the alternative allele *D* reaches relatively high frequencies, up to 30%: in Panama, northern Venezuela, and Labrador.

The major conclusion is that the Americas, especially South America, show extreme genetic variability. This is also shown by average F_{ST} values, which were calculated for the 491 populations selected for detailed analysis. Below we compare the American average with averages of world groups or regions of interest:

America	0.070 ± 0.006
Caucasoid (no exclusions)	0.043 ± 0.005
sub-Saharan Africa	0.035 ± 0.014
Australia	0.019 ± 0.004
New Guinea	0.039 ± 0.007
Polynesia	0.031 ± 0.004

In the various regions of Asia, F_{ST}s range from 0.021 (Southwest Asia) to 0.035 (Southeast Asia)

Of the various subdivisions of the Americas, South America has the greatest variation of gene frequencies: the average F_{ST} is 0.059 ± 0.006. The gene with the highest variation is *SE*Se* (0.30), followed by *KEL*Jsa* (0.19), *PGD*C* (0.18), and *TF*C* (0.16). After South America, the extreme North has the greatest variation: 0.051 ± 0.007 (including Eskimo, Aleut, all Na-Dene, and also the Chukchi, who cluster with Eskimos); the most variable genes are *FY*A* (0.26), *LE*Le* (0.21), *PCT*T* (0.13), and *KM*(1&1,2)* (0.10).

North and Central America combined, including Na-Dene but not Eskimos, has a comparatively low average F_{ST} (0.034 ± 0.004). The most variable gene is *ABO*A1* (0.17), followed by *A* (0.13), *HLAB*35* (0.12), and *O* (0.12). Of the various linguistic groups, Chibchan shows a variation comparable to that of South America as a whole: 0.059 ± 0.007, with *DI*A* being the most variable (0.17), *RH*cDE* and *CDe* next (0.13 and 0.11), and finally *TF*D* (0.11).

The impression from the geographic maps and distributions of gene frequencies is thus fully confirmed.

America, in particular South America, is genetically the most variable part of the world. As a consequence, there are extreme oscillations of mean F_{ST} values at various geographic distances around the interpolated variogram curves, that is, of the data points shown in variograms. These oscillations tend to be lower only for genes with high densities of observed frequencies, but even there the strong local geographic variation generates important fluctuations.

The F values indicated in the top right corner of the gene-frequency distributions given in each geographic map are F_{ST} values; but, unlike those given above, they are obtained from the original gene frequencies. They therefore include populations that have been excluded from the 491 selected as genetic references and, more importantly, they were pooled with neighbors. The data from the 491 populations are the basis for the F_{ST} values given above. Pooling neighbors decreases F_{ST} values (Cavalli-Sforza and Feldman 1990), and it is therefore not surprising that the F_{ST} values given in the maps are larger than the F_{ST}s calculated from the 491 populations.

An independent approach that leads to the same conclusions is the study of mitochondrial DNA. With a low-resolution technique, the restriction-fragment-length polymorphisms (RFLPs) of three tribes, Pima, Maya, Ticuna, were studied (Wallace et al. 1985; Schurr et al. 1990), and showed a variation of RFLPs similar to that of genes indicated above. Analyzing DNA markers makes it easier to identify specific mutants and may help us to follow specific migrations more closely. Inferences about the number of migrants to America that have been made in some mtDNA papers, even with techniques allowing higher resolution than those above, seem largely unwarranted at this stage of our knowledge.

6.13. Synthetic Maps of America

Table 6.13.1 shows the partition of the total variation among the first seven PCs, which cumulatively explain 74.3% of the total variation. The seventy-two genes used for the analysis correspond to the 69 genetic maps listed in the Table of Genetic Maps with the addition of *ABO*A2*, *AK1*1*, *GC*1F*. Table 6.13.2 shows correlations of the first six PCs with gene frequencies.

The analysis of single genes shows considerable local variation. Patterns found for different genes are rarely similar. By contrast, in other continents, several geographic patterns of single-gene frequencies were observed repeatedly with different genes. In those continents, one could easily anticipate, on the basis of the repeated patterns, and the number of repetitions of each, the general shape of synthetic maps obtained by PCs and their order of importance. In America we find this occurs clearly only for the first two synthetic maps, which correspond closely to the first two fissions in the genetic tree.

The first PC (fig. 6.13.1) shows a north-south gradient with the greatest slope in Canada, thus emphasizing the distinction between the Eskimos + Na-Dene group and Amerind populations closer to Eskimos on the one side, and the rest of America on the other side. In South America, there is a differentiation between east and west. According to some archaeological dates, not universally accepted (see sec. 6.2), the eastern area may also be the oldest part. There is a good correspondence with the first fission, which separates Eskimos and Na-Dene from all Amerinds. To note: the highest correlation of the first PC axis is with *IGHG1G3* za;b0stb3b5*, a typical marker of Asian origin.

Table 6.13.1. Percentage of Total Variance Explained by the First Seven Principal Components of American Gene Frequencies

Principal Component	*% of Total Variance*	*Principal Component*	*% of Total Variance*
1	32.6	5	5.7
2	12.7	6	4.8
3	8.6	7	3.9
4	6.0		

Most of the divergence found in the map of the second PC (fig. 6.13.2) is observed in North America. There is little variation in South America, though the east-west difference is always noticeable. In North America the major divergence is between Eskimos and non-Eskimos, with Na-Dene showing more similarities to the former than to the latter. The peak in the eastern part of North America most likely represents Caucasoid admixture; this is the area in which contact between Caucasoids and Amerinds has been longest. This area has *ABO*B*, relatively high *AK1*2*, *IGHG1G3*f; b0b1b3b4b5* and high *RH*cde*, strongly indicating Caucasoid admixture.

There is an inconsistency between the observations on the frequencies of the Caucasoid markers just indicated, which are drawn directly from the gene-frequency maps, and the correlations of this PC with the gene frequencies shown in table 6.13.2. The reason for this discrepancy is believed to be the existence of inordinate genetic variation in the Americas, which tends to cover other local regularities. The presence of important ethnic heterogeneity—that is, of Eskimos in the North—also tends to alter the meaning of the correlations of a

Table 6.13.2. Genes Showing the Highest Correlations with the First Six Principal Components of American Gene Frequencies

P.C.*	Range of Correlation Coefficient	Genes
1	1.00 – 0.90	(+) IGHG1G3•za;b0stb3b5, HLAB•27
		(−) —
	0.90 – 0.80	(+) ABC•A, ABC•A1, ACP1•A, AG•X, HLAA•9, LE•Le
		(−) ABO•O, AK1•1, DI•A, HLAA•1, HLAA•31, HP•1
	0.80 – 0.70	(+) HLAB•22, HLAB•40, HLAB•7
		(−) IGHG1G3•zax;g, KM•(1&1,2)
2	0.90 – 0.80	(+) HLAA•30
		(−) —
	0.80 – 0.70	(+) HLAA•33
		(−) PGD•A
	0.70 – 0.60	(+) JK•B, GC•1
		(−) —
	0.60 – 0.50	(+) HLAA•2, GLO1, HLAB•21
		(−) TF•C, MNS•Ms
	0.50 – 0.40	(+) ABO•A2, AG•X, GC•1F, HLAB•16, P1•1
		(−) FY•B, HLAB•15, LE•Le(a+), RH•D
3	0.80 – 0.70	(+) —
		(−) IGHG1G3•za;b0b1b3b4b5
	0.70 – 0.60	(+) RH•E
		(−) RH•CDe, RH•C, IGHG1G3•f;b0b1b3b4b5
	0.60 – 0.50	(+) RH•cdE, RH•cDE, IGHG1G3•fa;b0b1b3b4b5
		(−) HLAB•5
	0.50 – 0.40	(+) PTC•T, HLAB•14
		(−) PGM2•1, LE•Le(a+), JK•A
4	0.70 – 0.60	(+) HLAA•28
		(−) —
	0.60 – 0.50	(+) HLAA•2, HLAB•35, PGM1•1
		(−) HLAB•14
	0.50 – 0.40	(+) HLAB•21, PGM2•1
		(−) HLAB•15
5	0.60 – 0.50	(+) ESD•1, GLO1•1, JK•B
		(−) —
	0.50 – 0.40	(+) GC•1, FUT2(SE)•Se
		(−) IGHG1G3•za;g, MNS•S, MNS•MS
	0.40 – 0.30	(+) IGHG1G3•zax;g, HLAB•5, PEPA•1, RH•CDe
		(−) MNS•Ms, RH•cde
6	0.60 – 0.50	(+) —
		(−) FY•B, KEL•K
	0.50 – 0.40	(+) HLAB•22, RH•E, RH•cDE, FUT2(SE)•Se
		(−) ABO•B
	0.40 – 0.30	(+) CHE1•U, GC•1, GC•1F
		(−) —

Note.– Genes giving positive or negative correlation values are indicated by (+) or (−), respectively.
* P.C., Principal component.

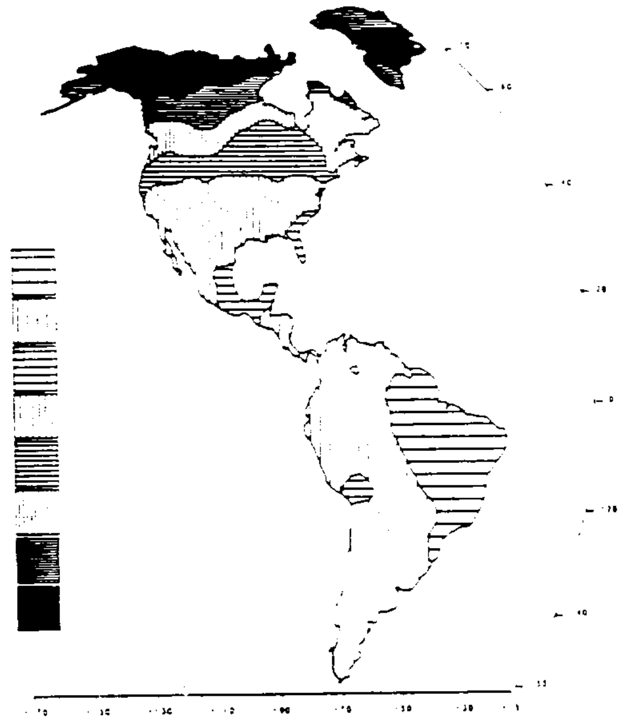


Fig. 6.13.1 Synthetic map of the Americas obtained by using the first principal component.

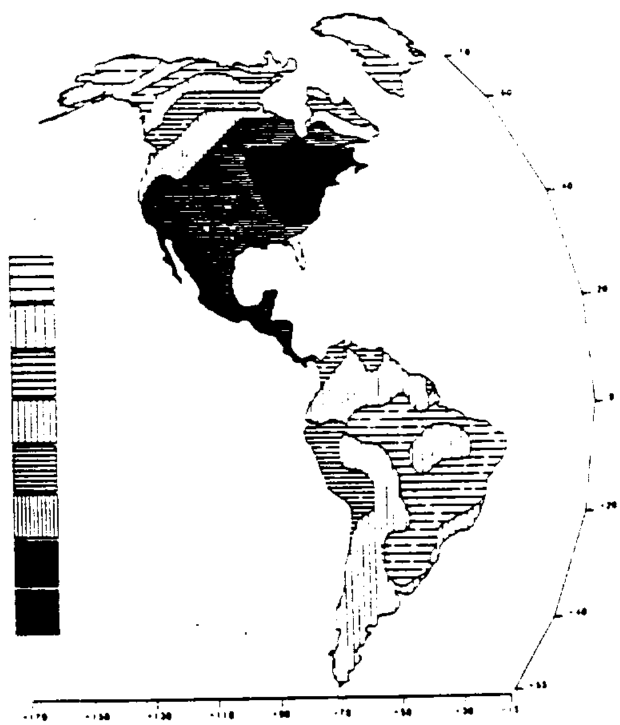


Fig. 6.13.2 Synthetic map of the Americas obtained by using the second principal component.

DOI 02215

PC with individual gene frequencies observed in a specific region.

Central America is more similar to North America than to South America. Thus, this map shows approximate correspondence with the fission between Na-Dene and Eskimo, but also with that between South America and the rest of the Americas. It also highlights Caucasoid admixture of the eastern part of North America.

Extreme values for the third PC (fig. 6.13.3) are found especially in South America, the contrast being remarkably strong between the northeastern and the southern Andes. North America also shows some variation between east and west, and in the same direction as in South America. It is possible that the east-west gradients observed in the north and in the south again express Caucasoid admixture which, as we have seen when discussing single genes, is especially prominent in the east-central area of North America, but is not missing in South America. Caucasoid admixture is also probably found among Greenland Eskimos, who were in contact with Vikings, especially on the eastern coast in the ninth to fourteenth centuries A.D.. Eventually, the Vikings died of starvation or were killed by the Eskimos (their fate was never clarified), but there may have been genetic exchange. If this is true, it is not surprising that one finds some similarity in the degree of shading of the three areas that may have had some Caucasoid contribution; some further clarification to this problem comes from the next PC. An admixture of another nature—that is, with Africans—is likely to have taken place in eastern Venezuela and the Guianas.

The fourth PC (fig. 6.13.4) also has a west-to-east gradient both in North America and in South America, but in contrast to the third PC, the direction of the gradient is inverted in the north and south. The similarity of the third and fourth PCs adds some evidence to the hypothesis that both eastern Greenland and the eastern coast of the United States have had some Caucasoid admixture, but the different behavior of the two components in Guiana may strengthen the hypothesis of admixture with Africans in this region.

The fifth component (fig. 6.13.5) stresses the difference between the Panama area and the rest of America. It is also indicative of migration to the south via Panama. The sixth map (not given) shows very little variation except in the extreme north, where it emphasizes the contrast between the Aleutian islanders and the Yupik Eskimos, occupying the southwestern part of Alaska, with the Eskimos of north-central Canada.

Other authors have used the synthetic map approach in America, O'Rourke et al. in both North (O'Rourke et al. 1986; Suarez et al. 1985) and South America (O'Rourke and Suarez 1986), and Salzano and Callegari-Jacques (1988) in South America. Both groups have found evidence of strong genetic drift in South America as we have, and their maps show less regular patterns than ours, being somewhat more similar to our single-gene maps. Our synthetic maps, however, seem less sensitive

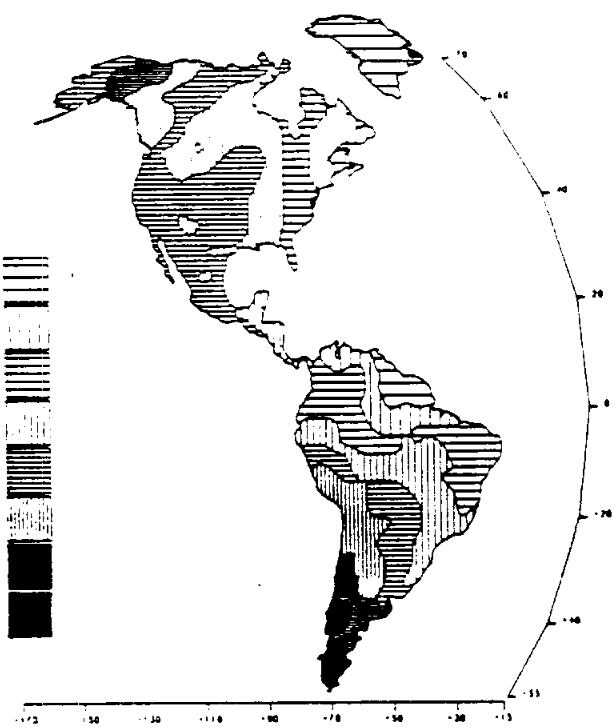


Fig. 6.13.3 Synthetic map of the Americas obtained by using the third principal component.

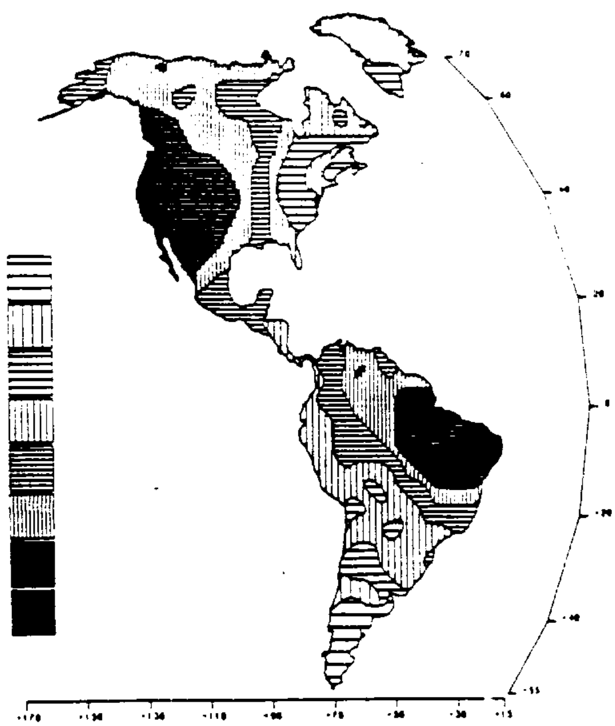


Fig. 6.13.4 Synthetic map of the Americas obtained by using the fourth principal component.

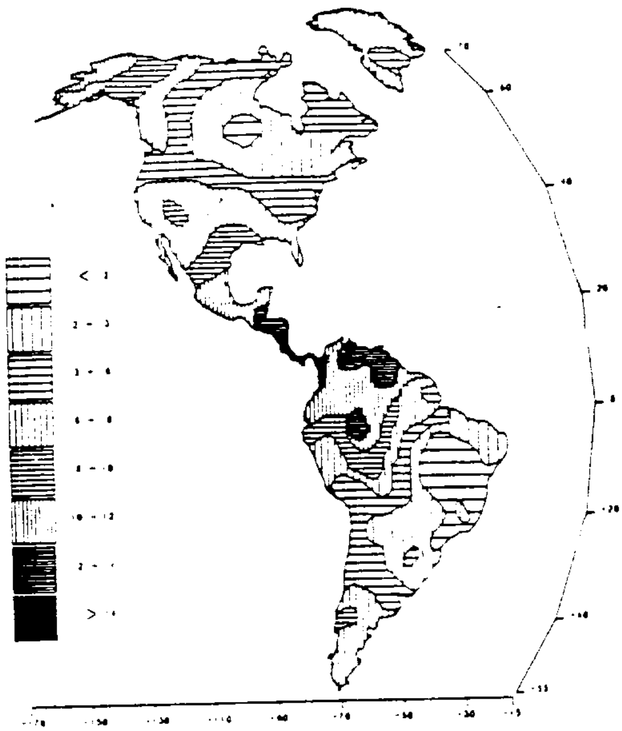


Fig. 6.13.5 Synthetic map of the Americas obtained by using the fifth principal component.

to drift than do individual genes. Our method obtains first maps of single genes and proceeds from them to obtain PCs and then their maps. This tends to smooth maps more than the direct calculation of PCs from original gene frequencies of selected groups or the slightly different mapping method used by O'Rourke and Suarez (1986). Differences in methods inevitably highlight one aspect or another: our synthetic maps are aimed at getting general similarities. Our single-gene maps are more useful than our synthetic maps for seeing highly localized effects of drift.

The conclusions from synthetic maps reinforce previous findings and help visualize major genetic regions. Eskimos, Na-Dene, and Almosan are well characterized and are even further differentiated into subgroups. The Caucasoid infiltration in the eastern United States, in eastern South America, and perhaps in Greenland are clear. The difference between the western and eastern coasts of North America is clear. In South America, several regions can be defined: the Andes show local homogeneity at the level of the higher PCs and always differ from the eastern part of South America. The lower PCs show differences between northern, and central and southern Andes, with the northern ones more similar to Central America. The fourth PC emphasizes the uniqueness of southern Chile. In the eastern part, one can distinguish a northern region formed by eastern Venezuela and the Guianas (see, e.g., the third PC), probably affected by African gene flow; a central one formed by northern Brazil, and a southern one corresponding to southern Brazil. There are important ecological differences among these areas, and there probably was greater exchange within, rather than between, different ecological regions.

The color map of the Americas conveys 63.9% of the regional variations.In North America there are green and yellow zones, the yellow being Na-Dene speakers and the green areas mostly northern Amerind. The color picture does not supply a clear distinction between these and Inuit (Eskimos), probably because the latter inhabit a very thin area on the coast. The southern part of North America is grayish, and the pink area at the boundary between southern Arizona, New Mexico, and northern Mexico is a sort of average from various local populations: southern Na-Dene (Apache and Navajo, who also have some genetic admixture with Amerinds) and neighboring speakers of Uto-Aztecan languages.

Central America shows a complicated mosaic of colors, as expected of a region that was probably crossed many times by many groups. The area occupied by Chibchan speakers is relatively homogeneous. The Caribbeans are passively stained: there are no aboriginals left.

South America is dominated by two colors, red and blue, neither of which is found in North America. Both colors appear, though not at the same intensity or with the same nuance, in Central America as well, indicating that there are some remnants of the passage across the funnel north of it. Blue extends to the north and northeast and must represent a dominant direction of migration, where languages of Tucanoan, Caribbean, and Ge stocks are spoken preferentially. The other dominant migration in bright red is found in the southern direction along the Andes, but it did expand from the Andes toward the east, mostly into the Amazon plains, as we have seen from archaeology. Is the white spot in the middle of the Andes near Bolivia and Peru, an indication of a possible inverse Thor Heyerdahl (1950) effect, the arrival of Polynesians to South America?

6.14. Summary of the Genetic History of America

The genetic patterns in the Americas fully confirm the three waves of migration suggested by dental and linguistic evidence: Amerinds, Na-Dene, and Eskimo. Their order in time is strongly suggested by their north-south geographical order. Further refinements may reveal that more than one entry contributed to the first wave.

but the archaeological information is contradictory and our understanding of the genetic pattern of Amerinds is incomplete, so that further investigations are required to settle this problem.

Eskimos, the last wave, fairly rapidly settled the Arctic coastal line and rarely occupied the interior. In the extreme east (Greenland) they may have mixed with Caucasoids, most probably because of contact with the Vikings who settled in Greenland and eventually vanished under fairly mysterious circumstances. It seems reasonable to assume that some of that population was reabsorbed by Greenland Eskimos.

The linguistic and geographic split between northern and southern Na-Dene is genetically clear-cut and probably reflects gene flow from other Amerinds, especially in the southern Na-Dene (Apache and Navajo), who had greater opportunities to receive it, because they were in more direct contact.

Amerinds show a much more complex picture. In North America, there is a band across the continent, which is wide in the east, of Caucasoid admixture. This admixture is also found elsewhere but it is less intense than in North America. In general, we have tried to avoid using populations in which admixture of some magnitude was suspected, but it was impossible to avoid mixed populations entirely without introducing an unwarranted bias.

In South America, one can use synthetic maps to distinguish three major genetic regions: the Andes, the Amazon basin, and the southern plateau. They are very different ecologically, and genetic exchange may have been less frequent among them than within. The genetic picture within the regions is so variable that an enormous amount of genetic drift clearly must have occurred. This variation is also found in North and Central America, but it is somewhat less extreme; besides, much of the genetic variation in North America is a direct consequence of major differences among ethnic groups like Eskimos, Na-Dene and Amerind, maintained over the millennia by ecological, behavioral, and social separation among the groups. No such obvious original ethnic differences exist in the rest of the continent. Clearly, fissions of tribes, and probably also fusions, have been numerous. Many tribes have probably originated from a small number of founders, justifying the enormous intertribal and interregional drift; they must also have moved around, as they still do, especially in areas like the Amazon and Orinoco basins. An important testimony to the extensive movements of Amerind tribes is the extreme fragmentation of the linguistic map, especially in South America.

It would be interesting to know whether some of the South American linguistic families existed before the passage through Panama and, if so, in which order they entered. The Andean family is found along the Andes, alternating with Paezan and, in some places, with Equatorial. It is not unreasonable to think that the Andeans entered before the Paezan, given that they extend farther south. The Paezan family is present in North America (Florida) and is most closely related to Chibchan, which is found mostly in Central, but also in South America. The relationship of Chibchan and Paezan may antedate their entry into South America.

It is very difficult to make inferences about the order of entry of the people who today speak Carib, Equatorial, Ge, and Panoan, on the basis of genetic data. On the basis of the geographic distribution of linguistic families, however, it seems natural to suggest that they entered in the order in which they are found in South America, those located farther south being first. Some subfamilies, however, have a very wide range: the Equatorial family, for instance, is spoken from Venezuela to Uruguay.

	West		*East*
North (latest)		Chibchan	
			Carib
		Tucanoan	
	Panoan	Equatorial	Ge
South (earliest)	Paezan		
	Andean		

These considerations could have more weight if there was a good correlation between linguistics and genetics in South America. Unfortunately, there is not, or it has not yet been found. Moreover, the considerable genetic noise caused by drift, and probably highly variable from place to place, makes an historical interpretation of the genetic tree less credible in South America than in other parts of the world. With very small populations, of variable size, evolutionary rate from drift is so variable that the length of the branches of the tree is hardly indicative of evolutionary time, using distances based on gene frequencies. It is difficult to say if other approaches—for example, using mtDNA—can be more useful.

At the moment, the simplest hypothesis is that fissions and movements of tribes, their complex gene flows and fusions, and the contrast that can be expected between the genetic and linguistic effects of fusions between tribes all contribute to dissociate genetic and linguistic evolution and to some extent even their relation with geography in this part of the world. Some regularities emerge from the genetic analysis of major geographic regions in South America but, at a microgeographic level, several poor or negative correlations among genetics, geography, and linguistics show the need for more detailed research, perhaps carried out with other methods. The research by Spuhler in North America and that on the Panama Chibchan (Barrantes et al. 1990) reassure us that we are on the right track in assuming a parallelism of genetic and linguistic differentiation in America, that this research model is productive, and some times even more informative than work at a macrogeographic level;

however, not every region will be equally favorable for microgeographic analysis.

In a model designed to test whether the settlement of the Americas could have produced the high genetic variation observed (Cavalli-Sforza 1986), five assumptions were made: (1) demes (tribes) were of census size 500; (2) they produced "buds" 25% of the size of the initial deme; (3) buds doubled in size every generation of 25 years (a rate of growth supported by many observations on populations in free growth; see sec. 2.7); they therefore reached the size of a full deme in 50 years; (4) in a budding cycle (two generations), a deme moved an average of 250 miles (5 miles per year). (5) It is likely that buds advancing in new territory had low mortality, living in environments either not contaminated or less contaminated by previous inhabitants; on the contrary, demes in regions behind the advancing frontier would soon slow down population increase. Perhaps increasing mortality was caused by rapid saturation of local population density. It is a necessary assumption of any expansion that population growth is rapid at the frontier and ceases or slows down considerably back of the frontier (Ammerman and Cavalli-Sforza 1984).

Under these conditions, the occupation of the Americas could be completed in few millennia, and, in the absence of admixture between demes, the final genetic variation between demes would even be too high with $N = 500$. Gene flow between demes would, of course, reduce genetic variation. Tribal fusions are bound to have played an important part because the genetic variation would be excessive if the models above are right.

A demic budding and expansion process in two dimensions would probably be random in direction, certainly unguided except by the search for game, safety, and comfort. The idea that a single band wandered across from Asia to America seems unrealistic. Along coasts and rivers, the process would be closer to unidimensional and unidirectional. The average rate of (random) movement of 5 miles per year is fast, because its randomness means that often, but not always, it would bring the group to new territory. It is, of course, possible that movement was by leaps and bounds, greater than 5 miles per move if people stayed in the same place for several years in a row. This pattern of repeated movement involves a specific behavior that is not typical of present-day hunter-gatherers (e.g., for African Pygmies: Cavalli-Sforza 1986), who move for long distances during the year but on established paths and repetitive, well-known circuits. In the past, Pygmies have certainly moved for long distances, in search of new abodes, but it is difficult to find comparable modern situations.

The model is very approximate, and only an accurate simulation could give more realistic values. Perhaps only at a later stage, closer to saturation of population density, fusion events would become more common. It is difficult to evaluate the saturation density in environments as diverse and poorly known as those in South America. Clearly, population density gradually rose in the Andes to levels much higher than in the rest of the subcontinent. Many urban developments, the skillful exploitation of the variety of ecological niches and astute social management in organized states must have gradually but greatly increased the carrying capacity of the Andean region in the last few millennia.

The most successful civilizations arose in the Andes and in many parts of Central America where the climate was more favorable. No such developments ever took place outside the Andes or other parts of Central America. But in the northern subcontinent, in times before European contact, Plains tribes were probably of relatively large size. More sedentary groups lived in communities that reached numbers in the thousands (sec. 6.4). Thus, wherever population numbers grew, the effects of drift were buffered and, especially where urban communities arose, they were eventually drastically reduced.