Introduction: Connections by Sea

Hello, and welcome to the January 31, 2009 issue of DNA Tribes® Digest. In this issue’s feature article, “Patterns of Gene Flow through the Coastal Silk Routes,” we continue to explore relationships among peoples of the world. This exploration will cover peoples of many languages, cultures, religions, and ways of life, reflecting each region’s local environmental and geographical conditions. Nevertheless, we will observe substantial continuity between each of these regions, reflecting ancient origins, historical events, and (perhaps most importantly) long-term patterns of human interaction in part shaped by the lay of the land itself.

Best regards and I hope to speak with you soon,
Lucas Martin
DNA Tribes
Patterns of Gene Flow through the Coastal Silk Routes

The Silk Routes comprise an ancient trade network that has connected populations of Europe, the Near East, Africa, India, and East Asia for thousands of years. In this study, we examined the relationships among genetic regions along the coastal portions of the Silk Routes1, beginning with the South Chinese region in the east and ending with the East African region in the west. For each region studied, genetic contributions from all other world regions were estimated.

Figure 1: Genetic regions and populations along the coastal Silk Routes explored in this study.

1 Genetic relationships among regions and populations of the inland Silk Routes were explored in our December 28, 2008 issue of DNA Tribes® Digest, available at: http://www.dnatribes.com/dnatribes-digest-2008-12-26.pdf.
Middle Kingdom: the South Chinese Genetic Region

**Background:** The South Chinese genetic region characterizes populations living in central provinces of the People’s Republic of China (known in Mandarin as Zhōngguó, meaning “Central Nation” or “Middle Kingdom”\(^2\)), including populations of the demographically predominant Han ethnic group\(^3\). The Han people trace their descent to the semi-mythical Huangdi or Yellow Emperor, who is said to have lived more than four thousand years ago and begun Chinese civilization with a conquest and unification of his own bear clan with a rival bull clan.

China is sometimes said to be the oldest continuous civilization still extant today. The earliest dynasty recorded by Chinese tradition is the Xia, which some archaeologists have speculatively linked to Bronze Age Erlitou artifacts uncovered in present day Henan Province and dated to 2100 BC. Over the course of several millennia, Chinese culture has come in contact with many neighboring peoples and remains influential to the present day.

**Genetic analysis:** Genetic contributions to the South Chinese region from the other 35 world regions\(^4\) presently identified by DNA Tribes® analysis were estimated. Results are illustrated in Figure 2 and summarized in Table 1 below.

**Discussion:** Results in Table 1 indicate genetic contributions from four neighboring regions, with the largest contributions identified from the southerly Southeast Asian region (46.2%) and westerly Tibetan region (30.3%). The 12.6% contribution from the North Chinese region might reflect gene flow from northerly Mongolic and Tungusic peoples who have periodically invaded and at times ruled China. The 10.8% Japanese contribution might reflect maritime contacts with Japan, perhaps also mediated by populations of the Korean peninsula.

It is notable that in contrast to regions and populations of the inland Silk Routes explored previously\(^5\), no substantial American Indian genetic contributions were identified for this or any other more southerly and coastal regions of Eurasia discussed in the present study.

---

\(^2\) The name “China” used in English derives from the Qin State established by the first emperor of a unified China, Qin Shi Huangdi, whose tomb (which remains largely unexcavated to the present day) is famous for its army of lifelike terracotta warriors.

\(^3\) The ethnonym “Han” refers to the subsequent Han Dynasty, which is remembered for its abundant prosperity and cultural and technological achievements.

\(^4\) A map illustrating the genetic world regions presently identified by DNA Tribes® analysis can be viewed at: [http://dnatribes.com/populations.html](http://dnatribes.com/populations.html).

Figure 2: Estimated genetic contributions to the South Chinese genetic region.

<table>
<thead>
<tr>
<th>Genetic Region</th>
<th>Estimated Contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Southeast Asian</td>
<td>46.2%</td>
</tr>
<tr>
<td>Tibetan</td>
<td>30.3%</td>
</tr>
<tr>
<td>North Chinese</td>
<td>12.6%</td>
</tr>
<tr>
<td>Japanese</td>
<td>10.8%</td>
</tr>
</tbody>
</table>

Table 1: Estimated genetic contributions to the South Chinese genetic region.

A Union of Mountain and Sea: the Southeast Asian Genetic Region

Background: The Southeast Asian genetic region characterizes populations of mainland Southeast Asia, including Hmong (known in China as Miao), Thai, and Vietnamese peoples, as well as Han populations of some Chinese provinces including Macau, Hong Kong, Guangdong, and (to some extent) Sichuan. A Vietnamese myth traces Vietnamese ancestry to the mythical union of a dragon and sea fairy. This legend is sometimes interpreted as representing the

6 Notably, the variety of Chinese spoken in several of these provinces affiliated with the Southeast Asian genetic region is Cantonese. Cantonese is written using the Chinese writing system standardized under Emperor Qin Shi Huangdi. However, spoken Cantonese is considered mutually unintelligible to speakers of other varieties of Chinese such as Mandarin, the official spoken language of the People’s Republic of China. One feature that distinguishes Cantonese from Mandarin is the more complex system of tones in Cantonese. The use of tone is also a feature of neighboring languages to the south such as Hmong-Mien, Tai-Kadai, and Vietnamese; in contrast, tone is not a feature of northerly neighboring languages such as Mongolian, Korean, or Japanese.
blending of coastal peoples (symbolized by the dragon) with mountain dwelling peoples (symbolized by the fairy) in the distant past. Languages spoken in the Southeast Asian genetic region include Sino-Tibetan languages such as Cantonese, Austro-Asiatic languages such as Vietnamese, and Tai-Kadai languages such as Thai.

Genetic analysis: Genetic contributions to the Southeast Asian region from the other 35 world regions presently identified by DNA Tribes® analysis were estimated. Results are illustrated in Figure 3 and summarized in Table 2 below.

![Figure 3: Estimated genetic contributions to the Southeast Asian genetic region.](image)

<table>
<thead>
<tr>
<th>Genetic Region</th>
<th>Estimated Contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>South Chinese</td>
<td>60.5%</td>
</tr>
<tr>
<td>Malay Archipelago</td>
<td>39.5%</td>
</tr>
</tbody>
</table>

Table 2: Estimated genetic contributions to the Southeast Asian genetic region.

Discussion: Results in Table 2 indicate genetic contributions from the South Chinese (60.5%) and Malay Archipelago (39.5%) regions. This might correspond to the blending of mountain dwelling peoples from the north with coastal or maritime peoples alluded to in the Vietnamese origin myth.
The Great Navigators: the Malay Archipelago Genetic Region

**Background:** The Malay Archipelago is a vast region of many islands that were once connected by land bridges to mainland Southeast Asia during the previous ice age. “Java man” fossils found here were at one time the oldest known human-like fossils, until the discovery of older remains in the Great Rift Valley of present day Kenya (in the East African genetic region). The Malay Archipelago has been described in several ancient texts: as Malayodvipa (“Malay Island”) in the Hindu Puranas, as Maleu-kolon by the Roman-Egyptian geographer Ptolemy, and as Ma-la-yu in Chinese historical records. Today, the terms Alam Melayu and Nusantara describe the cultural sphere that encompasses this genetic region.

The Austronesian family of languages spoken throughout the Malay Archipelago is thought to have spread from Taiwan between approximately 10,000 and 6,000 BC, although an alternate theory suggests the Austronesian languages spread from the Philippines. The Malay-speaking peoples are thought to be descended from seafaring cultures who have navigated Persian, Indian, and Chinese trade routes for many centuries, and their linguistic traces have been identified in places as distant as Madagascar.

Malay peoples have come in extensive contact with neighboring cultures and adopted at various times Buddhism, Hinduism, and Islam. For instance, the Buddhist Srivijaya kingdom based in Sumatra traded with both China and India. In the seventh century AD, the Persian-Hindu kingdom of Kadaram was founded by Maharaja Derbar Raja after he was defeated in battle in Persia and escaped to Sri Lanka, only to be blown off course to Kedah in present day Malaysia, where the local Malay people made him king. The Malay population of Kadaram practiced Hinduism, but later converted to Islam. More recently, the Dutch, British, Portuguese, and Spanish established trade links and colonial governments among Malay peoples, before the emergence of independent states that exist in the present day.

**Genetic analysis:** Genetic contributions to the Malay Archipelago region from the other 35 world regions presently identified by DNA Tribes® analysis were estimated. Results are illustrated in Figure 4 and summarized in Table 3 below.

---

7 To give some perspective, the distance between Manila, Philippines and Jakarta, Indonesia (both in the Malay Archipelago genetic region) is more than 1,700 miles, which is greater than the distance between London, England and Istanbul, Turkey.

Figure 4: Estimated genetic contributions to the Malay Archipelago genetic region.

<table>
<thead>
<tr>
<th>Genetic Region</th>
<th>Estimated Contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Southeast Asian</td>
<td>75.6%</td>
</tr>
<tr>
<td>Australian</td>
<td>12.0%</td>
</tr>
<tr>
<td>European</td>
<td>6.3%</td>
</tr>
<tr>
<td>Polynesian</td>
<td>6.1%</td>
</tr>
</tbody>
</table>

Table 3: Estimated genetic contributions to the Malay Archipelago genetic region.

**Discussion:** Results in Table 3 identified a predominant Southeast Asian contribution of 75.6%, suggesting substantial gene flow from mainland Asia. However, genetic contributions were also observed from the Australian (12.0%) and Polynesian (6.1%) regions, which might reflect maritime contacts with peoples to the south and east of the Malay Archipelago. The European contribution of 6.3% might reflect contacts during the more recent colonial period, but might also reflect older relationships with westerly populations such as the Persian-founded dynasty of Kadaram.

**At the Edge of East Asia: the Eastern India Genetic Region**

**Background:** The Eastern India genetic region characterizes populations of eastern parts of India, including populations near the Bay of Bengal, eastern Indo-Gangetic Plain, and the Chota Nagpur Plateau, as well as Bangladesh. The dominant language spoken here is Bengali, which is considered part of the Indo-Aryan branch of the Indo-European family of languages. However, also spoken here are Austro-Asiatic Munda languages, which are related to Southeast Asian

---

9 Genetic contributions from all European regions were listed as “European.”
languages such as Vietnamese and are thought to pre-date Indo-European languages in this region.

Historical empires in Eastern India include the Magadha Empire founded in 684 BC (mentioned in classical Sanskrit texts such as the Ramayana, Mahabharata, and Puranas), in which several events in the life of Siddhārtha Gautama (the founder of Buddhism) took place. Other states in this region include the Hindu Sunga Empire in the first and second centuries BC and the Pala Empire known for its support of the Buddhist Nālandā University, one of the world’s earliest large universities. These lands later became part of the Islamic Mughal Empire in the sixteenth and seventeenth centuries AD. Kolkata (Calcutta), the largest city in Eastern India, served as the capital of all of India under the British Empire between 1858 and 1912.

**Genetic analysis:** Genetic contributions to the Eastern India region from the other 35 world regions presently were estimated. Results are illustrated in Figure 5 and summarized in Table 4 below.

![Figure 5: Estimated genetic contributions to the Eastern India genetic region.](image-url)


<table>
<thead>
<tr>
<th>Genetic Region</th>
<th>Estimated Contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>South India</td>
<td>56.0%</td>
</tr>
<tr>
<td>North India</td>
<td>19.2%</td>
</tr>
<tr>
<td>Malay Archipelago</td>
<td>10.9%</td>
</tr>
<tr>
<td>Southeast Asian</td>
<td>4.4%</td>
</tr>
<tr>
<td>Australian</td>
<td>3.0%</td>
</tr>
<tr>
<td>Tibetan</td>
<td>3.0%</td>
</tr>
<tr>
<td>Other</td>
<td>3.5%</td>
</tr>
</tbody>
</table>

Table 4: Estimated genetic contributions to the Eastern India genetic region.

Discussion: Results in Table 4 indicate predominant genetic contributions from the South India (56.0%) and North India (19.2%) regions, for a combined contribution of 75.2% from within the Indian Subcontinent. However, substantial contributions were also identified from regions of East Asia, including the Malay Archipelago (10.9%), Southeast Asian (4.4%), and Tibetan (3.0%) regions. These East Asian contributions might reflect genetic traces of contacts also involved in the formation of the Munda languages in eastern parts of India.

As in the Malay Archipelago studied above, a substantial Australian contribution of 3.0% was also identified for the Eastern India region. Small but substantial Australian contributions will be seen in more westerly coastal Silk Route regions later explored in this study as well.

Home of the Dravidians: the South India Genetic Region

Background: The South India genetic region characterizes populations of southern states of India such as Andhra Pradesh, Karnataka, and Tamil Nadu, as well as the independent island country of Sri Lanka. These lands are home to the Tamils and other Dravidian peoples, whose languages are possibly related to those spoken in the Indus Valley Civilization (located in the North India genetic region) prior to the introduction of Indo-Aryan languages.


11 Indigenous populations of Australia are usually thought to be descended from early waves of west-to-east migrations out of Africa along the coasts of southern Asia. The observations in this study suggest the additional possibility of some return gene flow from Australia in the opposite direction via the Indian Ocean. Any such east-to-west gene flow might reflect processes or events similar to those that have brought the Austronesian Malagasy language to the island of Madagascar off the coast of Southern Africa. Genetic relationships among South Seas populations, some of which remain unexplained, are discussed further in the September 2008 issue of DNA Tribes® Digest, available at: [http://www.dnatribes.com/dnatribes-digest-2008-09-27.pdf](http://www.dnatribes.com/dnatribes-digest-2008-09-27.pdf).

12 It has sometimes been proposed that Dravidian languages are related to Uralic (such as Finnish) and Altaic (such as Turkish) languages spoken further north in Eurasia. However, substantial differences exist between each of these three language families, and any relationships between them remain speculative.
Dravidian peoples are thought to be descended from older indigenous cultures of India who left notable Mesolithic (Middle Stone Age) burial sites in the area. Traditional Tamil Sangam legends describes Tamil lands as once extending further southward before being submerged by the sea, which (if accurate) might have taken place as glacial ice melted and sea levels rose in the early Holocene Epoch that began 10,000 years before present.

Tamil literature is recognized as the oldest body of secular literature in the entire Indian Subcontinent. Related Dravidian languages are also spoken by some ethnic groups in other parts of India. Other languages spoken in the South India genetic region include Indo-Aryan languages such as Sinhalese (spoken in Sri Lanka), and Sanskritized Dravidian languages such as Telegu (spoken in Andhra Pradesh).

**Genetic analysis:** Genetic contributions to the South India region from the other 35 world regions were estimated. Results are illustrated in Figure 6 and summarized in Table 5 below.

![Figure 6: Estimated genetic contributions to the South India region.](image)

<table>
<thead>
<tr>
<th>Genetic Region</th>
<th>Estimated Contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eastern India</td>
<td>85.4%</td>
</tr>
<tr>
<td>North India</td>
<td>11.8%</td>
</tr>
<tr>
<td>Other</td>
<td>2.9%</td>
</tr>
</tbody>
</table>

**Table 5:** Estimated genetic contributions to the South India region.
Discussion: Results in Table 5 indicate genetic contributions from both Eastern India (85.4%) and North India (11.8%). These findings are consistent with local genetic characteristics that are primarily autochthonous (native) to India, with contacts with populations outside of the Indian Subcontinent substantially mediated by populations of Eastern and North India. The greater contribution from Eastern India might reflect older indigenous contacts that predate the formation of the Indo-Aryan cultures that are today most characteristic of the North India region.

Desert of the Bedouin: the Arabian Genetic Region

Background: The Arabian genetic region characterizes populations of the Arabian Peninsula, whose shores face the Persian Gulf, Indian Ocean, and Red Sea. According to some geographers, the Arabian Peninsula is more properly described as a subcontinent, because it rests on its own tectonic plate (the Arabian Plate). The southern Arabian interior houses the Rub’ al Khali or “Empty Quarter,” which is one of the largest and most formidable deserts in the world and remains largely unexplored to the present day. This desert is home to the Bedouin, nomadic herders whose ancient culture is sometimes thought to be emblematic of this region.

However, the Arabian Peninsula is thought to have been more fertile in the past, having undergone gradual desertification. The more fertile lands to the south (in present day Yemen) were known to the ancient Romans as Arabia Felix (meaning “Happy or Fertile Arabia”) and have been an important hub of the spice trade since ancient times, while the interior desert region was called Arabia Deserta (“Desert Arabia.”). Early states active here included the Sabean Kingdom (sometimes thought to be the home of the Queen of Sheba mentioned in the Bible), as well as the Axumite Empire (later to be called Ethiopia) of the first century AD that ruled some lands of Yemen and Saudi Arabia but was centered across the Red Sea in northeastern Africa.

The Arabian region has been a source of multiple culturally influential migrations over the course of history. For instance, approximately 3500 BC, the Semitic-speaking Akkadians expanded into Mesopotamia to the north and displaced earlier Sumerians (whose non-Semitic language is classified by linguists as an isolate not clearly related to other neighboring languages). More recently, the Rashidun Caliphate expanded from here beginning in 632 AD to establish an Islamic Empire including the Levantine, North African, and Mesopotamian regions. These territories are sometimes proposed as the cradle of the early Proto-Semitic language, the hypothetical ancestor of modern Arabic and other related languages of the Near East.

Genetic analysis: Genetic contributions to the Arabian region from the other 35 world regions were estimated. Results are illustrated in Figure 7 and summarized in Table 6 below.

13 For more information on genetic relationships in the Near East, see the December 2008 issue of DNA Tribes® Digest, available at: http://www.dnatribes.com/dnatribes-digest-2008-12-26.pdf. Notably, this previous study identified substantial Arabian genetic contributions to the Levantine and Mesopotamian regions, but not to the Aegean genetic region, which remained outside the Islamic world until the Seljuk Turks converted to Islam and conquered Anatolia in the eleventh century AD.
Discussion: Results in Table 6 indicate genetic contributions from several neighboring regions, including the Mesopotamian (37.6%) and Levantine (36.7%) regions to the north and the northwest, which might reflect contacts with early civilization centers in Egypt and the Fertile Crescent. An East African contribution of (19.4%) was also observed, which might reflect contacts across the Red Sea during periods such as the Axumite Empire. The North India contribution of 3.9% might reflect similar maritime contacts, in this case by way of the Indian Ocean and Persian Gulf. As in the Eastern India genetic region, a smaller Australian genetic contribution of 2.3% was also observed, which might reflect genetic traces of an early coastal route from Africa to Australia and perhaps suggesting the possibility of some subsequent contact via the Indian Ocean.

<table>
<thead>
<tr>
<th>Genetic Region</th>
<th>Estimated Contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mesopotamian</td>
<td>37.6%</td>
</tr>
<tr>
<td>Levantine</td>
<td>36.7%</td>
</tr>
<tr>
<td>East African</td>
<td>19.4%</td>
</tr>
<tr>
<td>North India</td>
<td>3.9%</td>
</tr>
<tr>
<td>Australian</td>
<td>2.3%</td>
</tr>
</tbody>
</table>

Table 6: Estimated genetic contributions to the Arabian genetic region.

Figure 7: Estimated genetic contributions to the Arabian genetic region.
Cradle of Humanity: the East African Genetic Region

Background: The East African genetic region characterizes populations near the African Great Lakes\(^{14}\) of the Great Rift Valley, and to some extent populations to the east and north along the coastal Horn of Africa. The East African region faces the Indian Ocean to the east and the Arabian Peninsula to the northeast, and abuts not far from the Congo River Basin to the west. The highest mountains in Africa are found in this region, including the highest peak in Africa, Mount Kilimanjaro in northeastern Tanzania. Because the oldest human-like fossils (including the australopithecine skeleton “Lucy”) have been discovered here, this region is thought to be the birthplace of the first humans and ancestors of all people living today.

Historical states in this region have included the Kingdom of Kush in present day Sudan at the meeting of the Blue Nile and White Nile rivers, which at times controlled parts of Egypt. Also in this region was the Empire of Axum (later to be called Ethiopia), which at one time controlled parts of the Arabian Peninsula across the Red Sea as well as lands of present day Ethiopia, Eritrea, Sudan, and Egypt.

Languages spoken in the East African region include Bantu languages such as Swahili (from the word *Kiswahili* or “coastal language”), a Bantu language with substantial Arabic (Semitic) influence used as a *lingua franca* in this part of Africa. Also spoken here are Semitic languages such as Amharic (spoken in Ethiopia), Cushitic languages such as Somali (spoken in Somalia), Omotic languages (spoken in southwestern Ethiopia), and Nilo-Saharan languages such as Luo (spoken in Kenya, Uganda, and Tanzania) and Maasai (spoken in Kenya and Tanzania).

Genetic analysis: Genetic contributions to the East African region from the other 35 world regions were estimated. Results are illustrated in Figure 8 and summarized in Table 7 below.

Discussion: Results in Table 7 indicate genetic contributions of 52.4% from West Africa and 24.9% from Southern Africa, for a combined 77.3% contribution from regions south of the Sahara Desert. This suggests substantial contact with neighboring African populations to the west and south\(^ {15}\). Also observed was an 18.4% Arabian contribution, which might reflect contacts with populations to the north and east in periods such as the Axumite Empire.

As in the Malay Archipelago, Eastern India, and Arabian regions, a small Australian contribution of 4.3% was identified here as well, which might reflect genetic traces of an early coastal migration from Africa to Australia and might also suggest some degree of subsequent contact, perhaps by way of the Indian Ocean.

\(^{14}\) The African Great Lakes are: Lake Victoria (the largest of the African Great Lakes and the world’s second largest freshwater lake by surface area), Lake Tanganyika (the world’s second largest freshwater lake by volume and by depth), Lake Malawi, Lake Turkana, Lake Albert, and Lake Kivu. It should also be noted that based on available data, both baHutu and baTutsi peoples of Rwanda are genetically representative of the East African region. This suggests deep local roots specific to this part of Africa for both peoples.

\(^{15}\) These results suggest that East Africa has been a geographical mediator of genetic contacts between Caucasian (European and Near Eastern) and Sub-Saharan African populations in the past, which would indicate that types of contact sometimes described as “admixture” by present day cultural standards in fact have substantial precedent dating to historical and perhaps ancient times.
Conclusion

Despite the great distances between regions surveyed in this article, results have indicated substantial genetic continuity between neighboring peoples in each region of the coastal Silk Routes, just as we have seen in our previous study of regions and populations along the inland Silk Routes\textsuperscript{16}. This suggests that human populations have been in substantial contact both by land and by sea, both with their immediate neighbors and in some cases with more distant peoples.

\textsuperscript{16} For more information, see “Patterns of Gene Flow through the Inland Silk Routes,” available here: http://www.dnatribes.com/dnatribes-digest-2008-12-26.pdf.
New Populations added January 18, 2009

We are pleased to announce the addition of several new populations to our database (numbers in parentheses indicate that number of individuals represented in each new population sample):

New East Asian Populations:
- Han (Jinan, China) (420)
- Shanghai, China (1,000)

New European Populations:
- Bashkir (Russian Federation) (100)
- Belarus (2,196)
- Belarus (96)
- Chuvash (Russian Federation) (96)
- Kharkov, Ukraine (114)
- Komi (Russian Federation) (103)
- Mari (Russian Federation) (44)
- Moldova (92)
- Mordva-Erza (Russian Federation) (52)
- Mordva-Moksha (Russian Federation) (47)
- Moscow, Russia (187)
- Northern Dobruja, Romania (569)
- Russian (100)
- Tatar-Mishary (Russian Federation) (94)
- Transylvania, Romania (1,977)
- Udmurt (Russian Federation) (93)
- Ukrainians (97)

New Latin American Populations:
- Santa Catarina, Brazil (3,000)

New Siberian Populations:
- Yakut (Russian Federation) (94)

New South Asian Populations:
- Indian Tamil (Sri Lanka) (40)
- Sinhalese (Sri Lanka) (292)
- Sri Lankan Moor (Sri Lanka) (74)
- Sri Lankan Tamil (Sri Lanka) (107)

Updates:

Previous DNA Tribes customers who would like to update their results to include these new populations and our most up to date algorithms can order using the "Update Your Analysis" option through our secure online checkout.