The genetic origin of Indians

The question of national and individual origins has a corporeal and concrete dimension, and a mythic and symbolic one. This is evident in the religious traditions which most of the world’s populations adhere to. Israel is both literally and figuratively a descent group. They issue from the tribes descended from the sons of Jacob. Those who convert into the Jewish religion customarily also convert into the Jewish nation, and so figuratively share the same descent. Similarly, among Muslims there is a particular prestige given over to the descendants of Muhammad, the Sayyids. Within Hinduism the importance of descent groups manifests generally in terms of the endogamy prevalent among South Asians, and also in specific cases, such as with gotras. The fundamental atomic basis of Confucian religious morality is arguably filial piety. Confucius’ descendants still play a prominent role in modern China promoting his ideas.

But descent also has a scientific and concrete aspect. Sometimes the mythic and scientific align. It does seem that the notional male line descendants of Genghis Khan are actually descended from one individual who flourished ~1,000 years ago. In other instances the connection is complex. Jews do seem to share common descent, but it is also evident that they have mixed greatly amongst the nations. And sometimes the inferences generated by science may warrant a reconsideration of treasured myths.
Most reasonable people will probably accede to the clear overwhelming descent of South Asian Muslims from the native people of the Indian subcontinent, but the genetics clinches that. True, there is quite often a clear trace of Middle Eastern and African ancestry among the Muslims of South Asia above and beyond what may be found amongst non-Muslims, but often this component is dwarfed by a minor East Asian element which seems to warrant no cultural memory! In this post I will not address specific cases as much as a general framework. I have been talking about genetics, and to a lesser extent South Asian genetics, since 2004 on this weblog. But we know so much more now than we did then. I thought it was time for me to sit down and actually condense the current state of knowledge as best as I can. I will not address the biomedical dimension of human population genetics in this post, only the historical ones.

First, a few notes. I understand that this is a controversial and fraught topic. One major issue I have when I bring up this area of knowledge in a South Asian forum is that people accuse me of promoting models which I barely understand. What I mean is that often I have to go and look things up to figure out what people are actually accusing me of implying. I didn’t grow up in South Asia, so I don’t know the political-cultural battles too well. Please be explicit and clear in your comments, and don’t assume I can connect the dots! Â Also, I’m going to apologize to some of you ahead of time for deleting your comments. I am going to track this thread and actually answer questions from interested parties, which means that I will need to shave off the noise. I won’t apologize to the people whose comments I delete because they address my comment moderation policy. Finally, I am going to use the word “Indian” from this point onward where in other cases I’d use “South Asian.” On the historical time scales that I’ll be addressing our ancestors were considered Indians (“Hindus”) by the rest of the world, and this seems a time where this clarity of terminology should trump contemporary geopolitical valences.

Why does any of this history matter? I have a hard time addressing this insofar as I have weak conditional effects based on my ancestry. By this, I mean that the details of my ancestry don’t matter much to me, except as a source of amusement or interest. I hope you don’t view me any differently if you find out that I seem to have a close genetic relationship to South Indian Dalits! (I do, probably far closer than you) You can also download a raw text file of my 23andMe v3 genotype if you want to poke around (I’ve made it public domain). But this sort of information matters for other people a great deal. I am, for example, kind of tired of listening to brown people talk about their non-Indian ancestry, whether it be Syrian Christians who claim Jewish antecedents, Jatts who claim Scythian antecedents, or Muslims who claim Arab, Turk, or Persian origins. From what I can tell reviewing the genetic data there is a grain of truth to many of these claims, but most brown people have ancestry that is overwhelmingly...brown. That’s pretty evident on our faces.

Second, I do know that finding ancestry from various groups can change how people view themselves. To give a personal example I have a friend who is a white American whose maternal grandparents were very racist against black people. After a detailed inspection of his genome it’s pretty clear that he’s ~5% African in ancestry. Some of his paternal relatives have been genotyped. This black ancestry doesn’t
show up on that branch of his family tree, so by elimination it seems likely that it was his anti-black side which had black ancestry (my friend told me that as a child he thought his maternal grandfather did look a touch black, an observation triggered by their vocal racism). A story is here, which he is only beginning to explore. There is something similar in my own family. My maternal grandmother comes from a family with some distant Middle Eastern ancestry. This obviously a point of pride. But a closer look at my mother’s genome makes two things clear: first, she does have a very small proportion of Middle Eastern ancestry. This could be noise, but it seems associated with a smaller African component, which is not uncommon among people of Muslim origin in the Indian subcontinent from what I have seen. But, a much larger fraction of my mother’s genome exhibits clear derivation from Southeast Asia, perhaps from an Austro-Asiatic or Tibeto-Burman group. But there is no mention of this in my family’s oral history.

But enough! Brass tacks, who are we as brown folk? The map at the top of this post gets at a big part of the answer. It was generated by the blogger behind The Jatt Gene using results from the Harappa Ancestry Project. It shows the rough distribution of a genetic element associated with the peoples of the Andaman Islands, and found from Pakistan to Vietnam to Indonesia. What does it mean? The Harappa Ancestry Project has thousands of individuals from hundreds of populations, and hundreds of thousands of genetic markers per person. This data set was then run through the program ADMIXTURE, which breaks apart the ancestry of individuals contingent upon the variation you throw into the program and the number of ancestral populations you want it to generate, the latter defined by the parameter “K.” This is just software, a dumb algorithm, so it needs to be used with care. But to give a concrete example, consider that you have three populations in your data set:

- White Americans
- Black Americans
- Nigerians

You tell ADMIXTURE to break apart the genomes of the individuals in your data set into at most two components. Two clusters if you will. The result in this case is going to be straightforward:

- The White Americans will be in one cluster
- The Nigerians in the other
- The black Americans will be a mix, with an average admixture fraction of 80% and 20%

The program is easy to interpret in this case, as we have a history, as well as other lines of evidence, to interpret these results. One component is clearly African ancestry, and the other is European. African Americans are on average 80% African and 20% European. So ADMIXTURE nicely popped out with that result.

What does ADMIXTURE tells us about South Asians? First, it depends on what reference populations
you use and how many clusters you want it to generate. I've addressed this detail before. But the Harappa Ancestry Project has lots of Indian populations. What you immediately see is that at higher K values a "South Asian" cluster breaks out. This cluster has the highest frequencies in southern and eastern India. It drops off as one moves west to Iran and east to Southeast Asia. Case closed?

Not quite. ADMIXTURE is a computer program. It can give strange results. It does not tell us reality, it tells us the the result of an algorithm. The "South Asian" cluster exhibits some peculiarities in terms of how it relates to other groups which can not be easily explained by history. I won't get into the details of that, but move to the main issue: deeper analytic techniques as well as moving up K's allows the "South Asian" cluster to fractionate into two dominant components. The major insight was unveiled nearly two years ago in a paper published in Nature, Reconstructing Indian population history:

India has been underrepresented in genome-wide surveys of human variation. We analyse 25 diverse groups in India to provide strong evidence for two ancient populations, genetically divergent, that are ancestral to most Indians today. One, the ‘Ancestral North Indians’ (ANI), is genetically close to Middle Easterners, Central Asians, and Europeans, whereas the other, the ‘Ancestral South Indians’ (ASI), is as distinct from ANI and East Asians as they are from each other. By introducing methods that can estimate ancestry without accurate ancestral populations, we show that ANI ancestry ranges from 39-71% in most Indian groups, and is higher in traditionally upper caste and Indo-European speakers. Groups with only ASI ancestry may no longer exist in mainland India. However, the indigenous Andaman Islanders are unique in being ASI-related groups without ANI ancestry. Allele frequency differences between groups in India are larger than in Europe, reflecting strong founder effects whose signatures have been maintained for thousands of years owing to endogamy. We therefore predict that there will be an excess of recessive diseases in India, which should be possible to screen and map genetically.

(ungated copy of the paper)

Using public data sets multiple bloggers have replicated the general shape of these results. The Harappa Ancestry Project has several populations from the Andaman Islands, and at K = 11 a component which is fixed in the Onge tribe correlates almost perfectly with the ANI/ASI ratios from the above paper.

Here's the short of it: Indians are hybrids between two ancient and very distinctive groups. If you want to know more details, I posted about it on my science blog. The top line is that the ANI is very much like Middle Eastern and European populations. In fact, ANI seems no closer to the ASI than these two other groups. Who were the ASI? The Andaman Islanders are their distant cousins, separated for tens of thousands of years. But the most current genomics shows a clear submerged substrate from the Indian subcontinent into Southeast Asia. Coincidentally Southeast Asia has been
strongly influenced by Indian culture. The ASI were closer to the populations of East Asia than to those of West Eurasia. Probably in part because East Asian populations are daughter groups from the modern humans who entered the Indian subcontinent from Africa tens of thousands of years ago. But the ASI are also quite distinct from East Asians. In some ways they represent a southern Eurasian population which seems to have been submerged within the last 10,000 years.

You can see shadows of their influence in this three dimensional visualization of genetic variation. Each point below is an individual projected onto a three dimensional space which is generated by the three largest components of variance within the data. The geographical clustering is pretty straightforward, but notice the “kink” in the South and Southeast Asians. That’s ASI’s shadow:
I just threw a lot out there for you to process. These results are pretty robust though. They’re based on hundreds of thousands of markers and there’s good population coverage. But their interpretation is more problematic. That’s because we don’t have records from prehistory. We are literally grappling with shadows. So let me address a few possibilities, and give my own take. All of these assertions are far less robust than what has come before because they are synthetic. They go beyond genomics, though they operate within the constraints that the new genomics imposes upon us.

- Who were the ANI? I think they derive from a set of farming populations from between the Black
Sea and the Caspian. The reason I think this is that there are suggestive associations with populations around the Caucasus with Indian groups, even more than with Iranians! This sort of “geographic leapfrog” requires a macrohistorical explanation.

- Were the ANI Aryans? I don’t think so. The admixture event with ASI is very old. Likely within the last 10,000 years, but probably older than 4,000 years (I know this from personal communication with one of the researchers who attempted linkage disequilibrium decay based time-from-admixture tests). Some of the Caucasian groups which have an affinity with Indians are not Indo-European speaking.

- So why did ANI arrive in India? I think it has to do with farming. Recent evidence is now pointing to massive reconfigurations of genetic variation across the world in the past 10,000 years. We have semi-historical evidence for nearly total replacement in Japan and Africa. But there is now a great deal of circumstantial evidence that the same occurred in Europe, at least once, and probably more than once. The ANI were one of the great farming Diasporas to pulse out of the Near East.

- But why didn't they replace ASI? I am not an archaeologist, so I am on weak ground here insofar as I’m relying heavily on others who know this stuff. But I suspect that the indigenous populations of the Indian subcontinent themselves had started an independent transition to farming. The ANI-ASI synthesis, both genetic and cultural, was that of two incipient farming toolkits. In contrast the relatives of the ASI in Southeast Asia did not enter into an independent phase of farming, and were marginalized to a far greater extent by populations from southern China (the exceptions being the Papuans). The Andaman Islanders then are exceptions, and not representative in their hunter-gatherer lifestyle.

- What about the Aryans? The data from Europe is far thicker than from the Indian subcontinent, and there is evidence for multiple movements and cultural influences. I believe that the Indo-Aryans arrived later, and are a minor overlay upon the ANI-ASI synthesis (South Indian tribals have 30-40% ANI, indicating how old and thoroughgoing the synthesis was). Some speculative suggestions can be made from the genetic data in regards to a post-ANI West Eurasian influence which does not seem Middle Eastern. I will leave that for now because we just don’t have much to go on, though I do suggest that one keep track of The Jatt Gene. I think the answers we’ve long been waiting for will be coming soon, especially with the imminent release of Indian populations from the 1000 Genomes.

- The northwest-southeast axis is the dominant genetic story of India, but not the only one. There is a northeast-southwest axis. It seems probable that the Munda are relative newcomers as well. Though mostly Indian, there is an element of ancestry in these populations which suggests relatively recent affinities with East Asians. This is probably at least part of my personal story, so I take an interest in this “third wheel” component of our heritage.
South Indian Brahmins claim northern Indo-Aryan origins. The genetics certainly bear this out, albeit with some probable admixture with the local substrate. There are many specific questions which can be asked and answered. The Cochin and Bene Israel Jews of the west coast of India clearly do have highly elevated Middle Eastern components of ancestry, though they are highly admixed with the native populations. My own question: do the Nasrani Christians truly descend from Jews? I would have dismissed this outright a few months ago, but I am not sure sure now. The western coast of India seems to have long-standing connections to southern Arabia, so we need to flesh out these patterns in more detail.

What's the biggest surprise from these results? For me I think it is the deep and incredibly thorough biological synthesis which characterizes the Indian subcontinent. We all know that there is a big difference between a Kashmiri Pandit and an Adivasi from South India. But about one third of the Pandit's ancestry is “Ancestral South Indian,” which is almost absent outside of the subcontinent. And about one third of the Adivasi's ancestry is “Ancestral North Indian,” which connects this individual with the populations which span the Atlantic, to the Urals, to the Sahara. The past is a strange and mysterious land. But the veil of ignorance is slowly lifting....

Note: Some might wonder why I didn’t address uniparental lineages. The post is long, that's why. The short of it is that ASI seems to have a much stronger impact on maternal lineages, while ANI is more dominant in paternal ones. Additionally, among the Munda the East Asian element is far more frequent on the paternal lineages than the maternal ones. This indicates a consistent trend of deep time events of sex-biased migration.

This entry was posted in Science by Razib Khan. Bookmark the permalink [http://sepiamutiny.com/blog/2011/04/22/the_genetic_ori/].

88 THOUGHTS ON “THE GENETIC ORIGIN OF INDIANS”

TTCUSM on April 23, 2011 at 6:56 pm said:

boston_mahesh, why did you italicize the term “niggard” when you were quoting the Rig Veda?

Razib Khan on April 23, 2011 at 7:06 pm said:
ok, watch it guys. the exchanges are starting to get less value-add for third parties, and more about your own inter-personal debate. may close the thread soon.

Surya  
on April 23, 2011 at 7:28 pm said:

"Who were the ANI? I think they derive from a set of farming populations from between the Black Sea and the Caspian. The reason I think this is that there are suggestive associations with populations around the Caucasus with Indian groups, even more than with Iranians! This sort of â€œgeographic leapfrogâ€ requires a macrohistorical explanation."

Aren't farmers tied to the land they farm? Why would farmers from the Black Sea area cross high mountain ranges like the Hindu Kush and vast arid expanses of land to come to India to farm?

Eurasian Sensation  
on April 23, 2011 at 7:54 pm said:

Razib, how does the movement of the Munda from the East correspond to the movement of ANI from the West?

You seem to indicate that ANI came earlier. So did the Munda leave their genetic imprint on a population that was already thoroughly mixed ANI/ASI?

Or was the ANI takeover of south and east India gradual? When the Munda arrived, would they have encountered people who were mostly ASI?

And are the Munda relicts of a population that was once much larger, or have they always been a small group?
Surya on April 23, 2011 at 8:26 pm said:

"We all know that there is a big difference between a Kashmiri Pandit and an Adivasi from South India. But about one third of the Pandit’s ancestry is Ancestral South Indian, which is almost absent outside of the subcontinent. And about one third of the Adivasi’s ancestry is Ancestral North Indian."

What does caste have to do with ANI and ASI? Don't we also know that:

A. There is a big difference between punjabi peasants (low caste) and south indian adivasis (tribals). We also know that south indian Brahmins are more indigenous than these punjabi peasants.

B. Kashmiri pundits are a very, very tiny fraction of brahmins. The majority of brahmins are found in Uttar Pradesh and Bihar, and these brahmins are also more indigenous than the punjabi peasants.

If even the supposedly aboriginal population of India from deep in the southeast corner of the subcontinent is of 1/3 ANI ancestry that necessarily implies that these few fair-skinned caucasian farmers from thousands of miles outside India somehow managed to father almost all of the current humongous mixed race population of India. That is quite an achievement, assuming this theory is correct.

Razib Khan on April 23, 2011 at 8:35 pm said:

Aren’t farmers tied to the land they farm? Why would farmers from the Black Sea area cross high mountain ranges like the Hindu Kush and vast arid expanses of land to come to India to farm?

malthusian limit. carrying capacity. partible inheritance. frontiers. these explain transient movements of populations. see peter bellwood’s book first farmers for an outline of the model. and they wouldn’t move as the crow flies. the expansion of farmers follows fertile valleys. e.g., helmund.
eurasian sensation, i will post on this soon. the munda seem to have mixed with a preexistent ANI/ASI group, not just ASI. the munda genetic imprint doesn't extend beyond northeast india (orissa, bengal, assam, etc.). but i think a lot of the rice farming lowlands were once munda. i think they were both indo-europeanized culturally and demographically assimilated with a second group from the northwest which was AN/ASI + more ANIlke stuff from the northwest. probably indo-europeans. perhaps dravidians.

1. There is a big difference between punjabi peasants (low caste) and south indian adivasis (tribals). We also know that south indian Brahmins are more indigenous than these punjabi peasants.

2. Kashmiri pundits are a very, very tiny fraction of brahmins. The majority of brahmins are found in Uttar Pradesh and Bihar, and these brahmins are also more indigenous than the punjabi peasants.

both true.

Surya
on April 23, 2011 at 8:57 pm said:

"are there not references to the "son of the black woman," referring to those warriors and priests who had indigenous mothers?"

Well Razib, that is the first time I have read that the Vedas says something like that. Like you my knowledge of the Vedas is second hand. Can you provide us the source that references the above?

What I have read is that the aryans considered people living to the west, who obviously were light skinned compared to themselves, to be barbarians or mlecchas. Even Punjab at one time was considered mleccha-land.
The next obvious question that enters my head is, "What did the ASI look like?" Maybe we can never know, but is there an extant population that most closely resembles them? Is it the Onge, or have the Onge also differentiated in phenotype considerably?

I’d also be interested to know how much ASI and ANI genes penetrate into Burma and beyond. There is obviously a “darker” element in SE Asian populations; does it correspond with ASI or is it something else? And does ANI stop at the east Indian border? Sorry, so many questions!

Surya on April 23, 2011 at 9:04 pm said:

“see peter bellwood’s book first farmers for an outline of the model. and they wouldn’t move as the crow flies. the expansion of farmers follows fertile valleys.”

I just don’t see a continuity of fertile valleys between the Black Sea region and the Indus River.

Ajita on April 23, 2011 at 9:31 pm said:

“I agree with Mithra’s argument above that it was most likely agricultural limitations that prevented the ANI from wiping out the ASI. South India’s diet today is primarily a rice-based one, as opposed to the dominance of wheat and corn (dry-land crops) in parts of Northern India.”

The idea is feasible, but corn would not have been grown by the invading ANIs. It's a New World species.

Eurasian Sensation on April 23, 2011 at 10:20 pm said:

The idea is feasible, but corn would not have been grown by the invading ANIs. It's a
Maybe I didn’t make my point clear, Ajita. Of course the ANI would not have grown corn/maize. My point was that today, South and North India are reliant on different crops due to the relative dryness of the north. Likewise, the early ANI were presumably reliant on their dry-land crops derived from the Fertile Crescent, but their spread through wetter southern India would have been slower as their agricultural toolkit was not as suitable there.

I just don’t see a continuity of fertile valleys between the Black Sea region and the Indus River.

It's leapfrog. That's what happened in Europe. If you don't accept this model, that's fine. But you should read the book if you are really interested.

The idea is feasible, but corn would not have been grown by the invading ANIs. It's a New World species.

FYI, corn is a generic term in the old world. Not always maize.

interesting points...there's obviously a substratum in Southeast Asia at low proportions which is related to ASI. Lower than in India. Not much Western Eurasian. Though obviously some because of Indian and middle eastern mercenaries (Aung San Suu Kyi's family is descended in part from West Asian or Indian Muslims who came into the service of the Burman kings). But why is the ASI-like element so
much more residual in southeast asia? you just offered one explanation: **the rice farmers of south china could push straight to indonesia without any ecological barrier.** in contrast, the ANI probably came into south asia and encountered the limits of their easily ecological push pretty earlier on, and had to pause.

---

**Razib Khan**

on **April 23, 2011 at 11:12 pm** said:

> Maybe we can never know, but is there an extant population that most closely resembles them? Is it the Onge, or have the Onge also differentiated in phenotype considerably?

this point doesn’t get made enough. human populations are evolving and change. blue eyes for example is almost certainly a novel feature of the last 10,000 years (you can trace when genes rise in frequency, it’s been done). the easiest way to do this is extract ancient subfossils and match the variants to modern populations. it will be feasible to easily do such reconstructions in a few years. the major issue will be finding DNA. it gets harder and harder the warmer and moister the climate. there’s a reason that ancient DNA has mostly been retrieved from temperate or arctic climes.

> I’d also be interested to know how much ASI and ANI genes penetrate into Burma and beyond. There is obviously a “darker” element in SE Asian populations; does it correspond with ASI or is it something else? And does ANI stop at the east Indian border?

ANI did not. ASI-ish elements seem to have spanned the area from the south china sea to the arabian sea.

---

**Wedplan**

on **April 23, 2011 at 11:22 pm** said:

I had posted this link on the News tab sometime back. Another data point in the cultural issues involved with the AIT theory.
http://www.huffingtonpost.com/rajiv-malhotra/how-europeans-misappropri_b_837376.html

Razib Khan  
on April 23, 2011 at 11:53 pm said:

AIT was dumb, but the indians who respond with nationalistic models are just as dumb. that's just an assertion. not an opening for content-free conversation which i'll have to delete....

Razib Khan  
on April 23, 2011 at 11:55 pm said:

Speculations based on genetic studies are prone to myth-making from all sides as well. The studies provide partial content, which then are filled with all kinds of content, by all kinds of people based on where in the ideological spectrum they come from, what kind of histories they were taught and so forth. As a materialist, gene-mapping is a good scientific endeavor but speculations based on that to tracing roots in pre-historic antiquity is ultimately a futile exercise. My 2 cents:)
i'm moving on to other things now.

thanks to commenters who made this interesting. to the rest, i'll know who to delete preemptively in the future.