

SUBSCRIBE

# GlobalCoffeeReport

The intelligent industry resource

Home > Features

## The history of Arabica

by Aisling Geraghty — July 29, 2024 in Features, Research & Development Reading Time: 8 mins read

AA



Image: NTU Singapore.

Share on Facebook

Share on Twitter

Pin and Share icons

Dr Jarkko Salojärvi of Nanyang Technological University details the family history of Arabica, highlighting the knowledge to establish breeding, the development of cultivars, and disease resistance.

Cast your mind back half a million years ago. It was a time when Neanderthals traversed Asia and Europe, megafauna roamed the Earth during the Pleistocene era, and according to new research, when Coffea arabica was first sighted.

In the study, published in April 2024 in *Nature Genetics* and entitled 'The genome and population genomics of allopolyploid Coffea arabica reveal the diversification history of modern coffee cultivars', researchers suggest that Arabica developed 350,000 to 610,000 years ago in the forests of Ethiopia via natural mating between two other coffee species: Coffea eugenioides (Eugenioides) and Coffea canephora (Robusta).

"We've used genomic information in plants alive today to go back in time and paint the most accurate picture possible of Arabica's long history, as well as determine how modern cultivated varieties are related to each other," says Jarkko Salojärvi, Assistant Professor at Nanyang Technological University and Principal Investigator at University of Helsinki.

The new reference genome was accomplished by using DNA sequencing technology and advanced data science. The research team sequenced 41 wild and cultivated accessions to facilitate in-depth analysis of Arabica history and dissemination routes, and the identification of candidate genomic regions associated with pathogen resistance. This included an 18th century type specimen, provided by the Linnaean Society of London, considered the world's oldest active society devoted to natural history; 12 cultivars with different breeding histories; the Timor hybrid and five of its backcrosses to Arabica; and 17 wild and three wild/cultivated accessions collected from the Eastern and Western sides of the Great Rift Valley, stretching from Southeast Africa to Asia.

"We did several things to come up with the best possible genome. Firstly, the accession we chose as reference is an aberrant individual with only one copy of each chromosome. Usually in diploid species, like us humans, our chromosomes come in pairs, one from the mother and the second from the father, but in this individual, there was only one copy of each chromosome. We call it a di-haploid ('di' meaning it copies genomes from Robusta and Eugenioides). With one copy per chromosome, we didn't have to worry about the differences between the paternal and maternal genomes," Salojärvi says.

"Second, we read the genome using the latest genome sequencing technology PacBio HiFi, which allows virtually error-free reading of genome fragments of over 20,000 to 30,000 bases. Like in the old technology, we still need to fragment the DNA into millions of small pieces that are then read in parallel, but because there are no errors finding overlaps between these fragments, it is much easier than using the old technology."

Salojärvi says the assembly step produces long continuous stretches of genome sequence, but they are not yet organised into chromosomes.

"To do this, we used one more technology, chromosome conformation capture, which finds the 3D-structure of the chromosomes and is able to tell which pieces of DNA belong to which chromosome, and also their relative position."

To sequence the 41 wild and cultivated accessions, Salojärvi says the team sequenced the samples using shotgun sequencing (short DNA fragments of only a few hundred bases) and aligned these reads against the reference genome.

"This made it possible to identify small differences and mutations between the different accessions. These mutations accumulate at a constant rate in the genome, which makes it possible to calculate the timings of different events," he says.



Image: NTU Singapore.

### Independent evolution

According to the study, the initial crossbreeding that created Arabica was done without any intervention from humans.

Scientists have had a hard time pinpointing exactly when — and where — the natural hybridisation between Robusta and Eugenioides took place, with estimates ranging everywhere from 10,000 to one million years ago.

"Usually the historical place of origin is identified with fossil evidence. However, the problem is that dead plants decompose easily and very rarely leave fossils that can be analysed. This means we are left with methods that use genome data, but we didn't have a genome we could use for this," says Salojärvi.



these areas.”

To assess the initial crossbreeding that created Arabica, Salojärvi says the team looked at the data in several different ways.

“One was to look at these mutations in the different accessions and estimate the time when they are all reduced to one common ancestor with no mutations. Secondly, was to compare the genes between the diploid progenitors and the Arabica subgenomes and calculate the number of mutations between them. Knowing how fast they accumulate in the genome tells us about the time between the common parent for Arabica and the diploid parents,” he says.

“Finally, we looked at the rates at which genes are lost in the genome. Gene loss is a standard process of how genomes evolve and how species become different from each other through time. This loss rate gave us one more estimate. Combined, these estimates gave us a time window of 350,000 to 610,000 years ago.

“In other words, the crossbreeding that created Arabica wasn’t something that humans did. It’s pretty clear that this polyploidy event predated modern humans and the cultivation of coffee.”

### Building the family tree

Coffee plants have long been thought to have developed in Ethiopia, but varieties that the team collected around the Great Rift Valley displayed a clear geographic split. Most of the wild varieties studied originated from the western side with three from the eastern side. All the cultivated varieties, however, originated from the eastern side closest to the Bab al-Mandab strait which separates Africa and Yemen.

This would align with evidence that coffee cultivation may have started principally in Yemen, around the 15th century. Indian monk Baba Budan is believed to have smuggled the fabled ‘seven seeds’ out of Yemen around 1600, establishing Indian Arabica cultivars and setting the stage for coffee’s global reach today.

“It looks like Yemeni coffee diversity may be the founder of all of the current major varieties. We tracked all the Bourbon, Typica, and Indian cultivars to a common origin in Yemen,” Salojärvi says. “Coffee is not a crop that has been heavily crossbred, such as maize or wheat, to create new varieties. People mainly chose a variety they liked and then grew it. So the varieties we have today have probably been around for a long time.”

According to the study, modern genomic tools and a detailed understanding of the origin and breeding history of contemporary varieties are vital to developing new Arabica cultivars, better adapted to climate change and agricultural practices.

“The genome allows us to use modern breeding techniques to develop improved cultivars much faster. There is still some work required, but the genome will help us get there,” Salojärvi says. “In future, we won’t have to wait for the individuals to bear berries but instead can predict already from the genome information which ones will be good for the next generation of breeding. We also have one candidate region to study for coffee leaf rust resistance.”

The reference genome was also able to shed more light on how one line of Arabica varieties obtained strong resistance to the coffee leaf rust disease.

The Timor variety formed in Southeast Asia as a spontaneous hybrid between Arabica and one of its parents, Robusta, a species more resistant to disease than Arabica.

“This means when Robusta hybridised itself back into Arabica on Timor, it brought some of its pathogen defence genes along with it,” says Salojärvi.

Breeders have tried replicating this crossbreeding to boost pathogen defence. The new Arabica reference genome allows present researchers to pinpoint a new region harbouring members of the RPP8 resistance gene family, as well as a general regulator of resistance genes, CPR1.

“These results suggest a novel target locus for potentially improving pathogen resistance in Arabica,” Salojärvi says.

The genome provided other new findings as well, including which wild varieties are closest to modern cultivated Arabica coffee. Researchers also found that the Typica variety, an early Dutch cultivar originating from either India or Sri Lanka, is likely the parent of the Bourbon variety, principally cultivated by the French.

“Our work has not been unlike reconstructing the family tree of a very important family,” Salojärvi says.

*This article was first published in the July/August 2024 edition of Global Coffee Report. Read more [HERE](#).*

