

## An Introduction to Haplogroups: An Interactive Activity

Activity developed by Meredith T. Knight at Tufts University as part of David R. Walt's HHMI Professor's Award.

**Overview:** This activity will introduce participants to the concept of a haplogroup. It is an interactive non lab based activity requiring transparencies and a way to print on them. There is also a paper option.

**Learning Objectives:** Participants will learn

- The concept of a single nucleotide polymorphism (SNP)
- How to compare actual mitochondrial sequences
- How to sort those sequences into groups
- How to understand the concept of a haplogroup
- How to trace their haplogroup based on a pattern of SNPs using a haplogroup tree

### **Background Information**

DNA contains **single nucleotide polymorphisms**, or SNPs. A SNP is a mutation in the DNA sequence that is only one base pair long. At a certain place in his or her genome, a person may have a "T" (thymine) instead of a "G" (guanine). This can occur in both nuclear DNA and mitochondrial DNA. The sequences we are using today are actual sequences from mitochondrial DNA.

Mitochondrial DNA is passed down maternally and is not recombined. So, every individual inherits a pattern of SNPs in their mitochondrial DNA from their mother only. This is a useful clue for tracing "deep ancestry", in which we can look back for many generations. Scientists have used information from the mitochondrial DNA in people living today and the rate of mutation in mitochondria to help them understand the migration of *homo sapiens sapiens* out of Africa and across the Earth.

A **haplogroup** is a group of people who all share the same pattern of SNPs in their mitochondrial DNA. In this activity, students will compare actual sequences from mitochondrial DNA to see which SNPs they have on their sheet. They will then be able to determine which haplogroup they have, and learn more about haplogroups.

### **Naming conventions**

L3 G769A

Haplogroup – the name of the haplogroup is the first letter ("L3")

The "G" is the base pair for people in haplogroup L3

The "A" is the base pair for people not in haplogroup L3

The "769" is the location of the SNP site on the mitochondrial DNA loop.

### **Preparation**

To prepare for this activity, print out copies of the four sequences below. Print enough copies so each student can have one transparency. These can be used again.

You may want to have the haplogroup map and haplogroup "tree" available either on a transparency or projected on a screen for when you debrief this activity.

\*\*I have also created a "haplogroup card" in which students can check a "yes or no" for each sequence, depending on what nucleotide they have at the SNP site. At the top of the column it says "yes" or "no" based on which SNPs the students have on their cards.

### **Additional resources How to use "23andme.com" for learning about maternally inherited haplogroups.**

1. Log in, create an account with your name and email.
2. Go to "ancestry" on the home page.
3. On the left hand side, click on "maternal"
4. Click on "haplogroup tree" in the middle on the top.
5. You'll see Mitochondrial eve on the left, scroll right.
6. Click on your haplogroup of interest and then select "history" for descriptions.

L3 G769A

721 tcaccctcta aatcaccacg atcaaaagg**g** acaagcatca agcacgcagc aatgcagctc

M C10400T,

10381 aaaaggatt agac**ct**gaacc gaattggtat atagtttaaa caaaacgaat gatttcgact

N-T10873C

10861 tagcatcatc cc**ct**tactat tttttaacca aatcaacaac aacctattta gctgttcccc

N9- G5917A

5881 gccattttac ctcacccccca ctgatgttcg ccg**a**ccgttg actattctct acaaaccaca

A-A1736G

1681 gccccaaacc cactccacct tactaccaga caaccttagc caaacattt acccag**g**ataa

L3 G769A

721 tcaccctcta aatcaccacg atcaaaagg**g** acaagcatca agcacgcagc aatgcagctc

M C10400T

10381 aaaaggatt agat**tt**gaacc gaattggtat atagtttaaa caaaacgaat gatttcgact

N-T10873C

10861 tagcatcatc cc**tt**tactat tttttaacca aatcaacaac aacctattta gctgttcccc

N9- G5917A,

5881 gccattttac ctcacccccca ctgatgttcg ccg**g**ccgttg actattctct acaaaccaca

A-A1736G

1681 gccccaaacc cactccacct tactaccaga caaccttagc caaacattt acccag**g**ataa

L3 G769A

721 tcaccctcta aatcaccacg atcaaaagg**g** acaagcatca agcacgcagc aatgcagctc

M C10400T

10381 aaaaggatt agat**tt**gaacc gaattggtat atagtttaaa caaaacgaat gatttcgact

N-T10873C

10861 tagcatcatc cc**tt**tactat tttttaacca aatcaacaac aacctattta gctgttcccc

N9- G5917A

5881 gccattttac ctcacccccca ctgatgttcg ccg**a**ccgttg actattctct acaaaccaca

A-A1736G,

1681 gccccaaacc cactccacct tactaccaga caaccttagc caaacattt accca**a**ataa

L3 G769A,

721 tcaccctcta aatcaccacg atcaaaagg**a** acaagcatca agcacgcagc aatgcagctc

M C10400T

10381 aaaaggatt agat**tt**gaacc gaattggtat atagtttaaa caaaacgaat gatttcgact

N-T10873C

10861 tagcatcatc cc**ct**tactat tttttaacca aatcaacaac aacctattta gctgttcccc

N9- G5917A

5881 gccattttac ctcacccccca ctgatgttcg ccg**a**ccgttg actattctct acaaaccaca

A-A1736G

1681 gccccaaacc cactccacct tactaccaga caaccttagc caaacattt acccag**g**ataa

### Haplogroup Activity

Name: \_\_\_\_\_

Circle which SNP you have in your sample.

	In	Out
1. L3	G	A
2. M	C	T
3. N	T	C
4. N9	G	A
5. A	A	G

Which Haplogroup are you in? \_\_\_\_\_

Describe this Haplogroup in more detail.

6. When did they migrate out of Africa?

7. Which region of the Earth did they live?

8. What else did you learn about this haplogroup?

9. What more would you like to learn about this group?

### Haplogroup Activity

Name: \_\_\_\_\_

Circle which SNP you have in your sample.

	In	Out
1. L3	G	A
2. M	C	T
3. N	T	C
4. N9	G	A
5. A	A	G

Which Haplogroup are you in? \_\_\_\_\_

Describe this Haplogroup in more detail.

6. When did they migrate out of Africa?

7. Which region of the Earth did they live?

8. What else did you learn about this haplogroup?

9. What more would you like to learn about this group?

### Haplogroup Activity

Name: \_\_\_\_\_

Circle which SNP you have in your sample.

	In	Out
1. L3	G	A
2. M	C	T
3. N	T	C
4. N9	G	A
5. A	A	G

Which Haplogroup are you in? \_\_\_\_\_

Describe this Haplogroup in more detail.

6. When did they migrate out of Africa?

7. Which region of the Earth did they live?

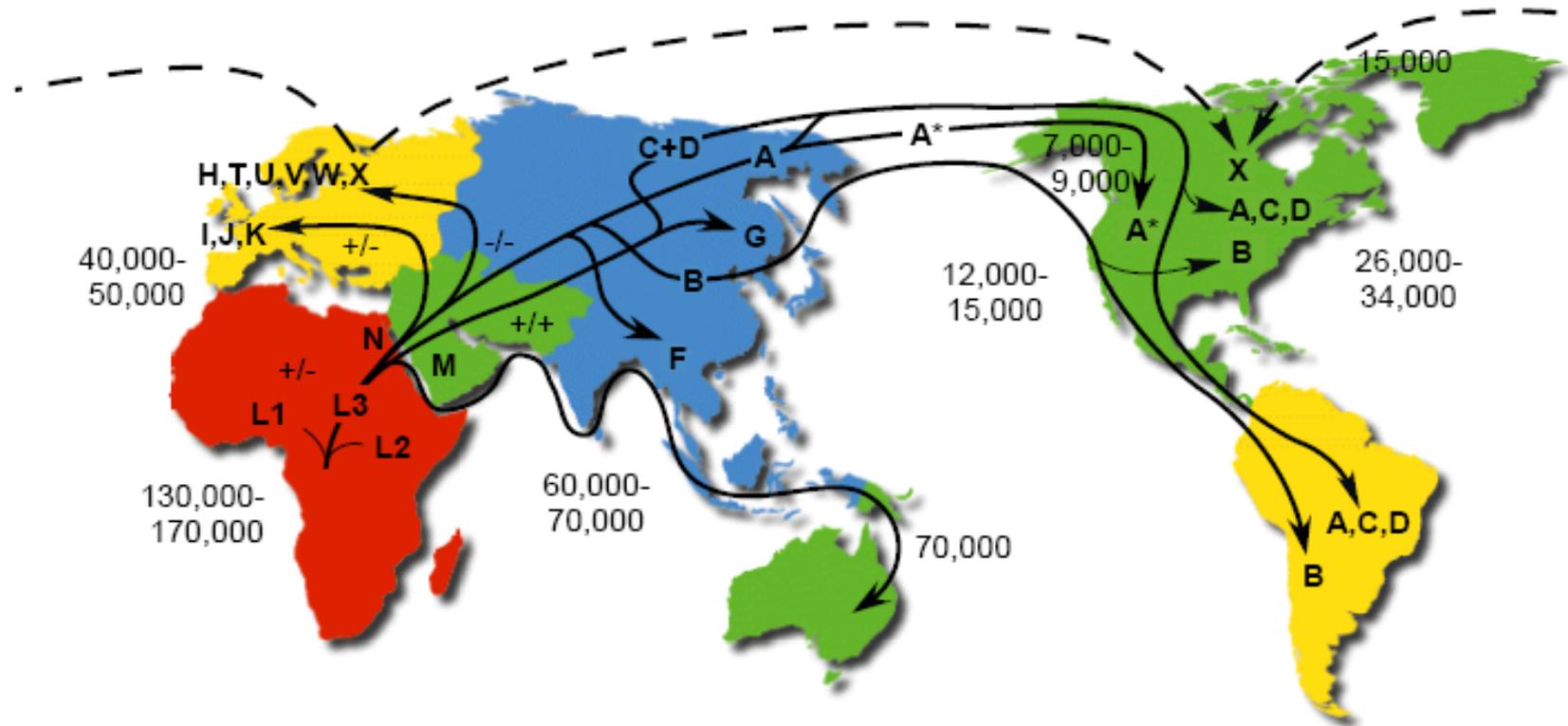
8. What else did you learn about this haplogroup?

9. What more would you like to learn about this group?

# Human mtDNA Migrations

<http://www.mitomap.org/mitomap/WorldMigrations.pdf>

Copyright 2002 © Mitomap.org



+/-, +/+, or -/- = Dde I 10394 / Alu I 10397  
 \* = Rsa I 16329

Mutation rate = 2.2 - 2.9 % / MYR  
 Time estimates are YBP



## **HAPLOGROUP L3 from [www.23andme.com](http://www.23andme.com)**

[Haplogroup](#) L3 has played a pivotal role in the history of the human species. Soon after the haplogroup arose in eastern Africa about 60,000 years ago a relatively small number of migrants carried it across the Red Sea to Arabia, inaugurating an intercontinental migration that eventually settled every major land mass on Earth except Antarctica. That small group also gave rise to every non-African haplogroup.

L3 has also been mobile within Africa, spreading south with migrations of Bantu-speaking populations over the past few thousand years. The Atlantic slave trade carried it to the Americas as well; various branches of L3 are found today among more than 25% of the African American population.

### **Haplogroup L3e**

[Haplogroup](#) L3e originated about 45,000 years ago in central Africa or present-day Sudan. It then spread southward with the eastern arm of an expansion by Bantu-speaking people that began about 4,000 years ago in the vicinity of present-day Cameroon and eventually reached most of sub-Saharan Africa.

Today L3e is most common in central and southeast Africa. It is also found among African Americans and about one-third of Brazilians whose mitochondrial lineages trace back to Africa in the slave-trading era, an indication that their ancestors may have originated in Angola rather than the more common slave-trading centers of the African west coast.

### **L3e1's Journey to the Americas**

The L3e1 branch of L3e offers another clue to the regional ancestry of some African Americans. L3e1 originated about 32,000 years ago in central Africa, and spread eastward into present-day Kenya during the migration of Bantu-speaking farmers that began

about 4,000 years ago. The [haplogroup](#) then spread south into Mozambique, where it persists today.

Because L3e1 is not found in other areas associated with the slave trade – and historical records indicate about 1 million people were transported from Mozambique – African Americans with [mitochondrial DNA](#) belonging to the haplogroup are likely to descend from ancestors who lived there.

### **The Isolation of L3e5**

Unlike other branches of L3e that are spread out across western Africa and among African-Americans, L3e5 is considerably more isolated. It is normally found only in the Chad Basin, the area surrounding Lake Chad and encompassing the borders of Niger, Nigeria, Chad, and Cameroon in western central Africa. L3e5 originated about 11,500 years ago in this region, and for the most part it has stayed put ever since.

Nearly 40% of the Fali, an indigenous tribe of 23,000 living in northern Cameroon, belong to [haplogroup](#) L3e5. This haplogroup has never been found among African-Americans, however, which suggests that unlike other areas of western and central Africa the Chad Basin probably was not plundered by slave traders.

## **HAPLOGROUP M – from 23andme.com**

[Haplogroup M](#) is one of two branches on the [mitochondrial DNA](#) tree that arose about 60,000 years ago, soon after humans first expanded out of Africa. M appears to have spread rapidly along the coast of the Indian Ocean, probably reaching southeastern Asia within a few thousand years.

Many scholars have speculated that haplogroup M traveled eastward at least in part by sea, because the route it took would have required a number of substantial ocean passages. The exit from Africa itself may have been across the Bab el-Mandeb, a narrow strait at the southern end of the Red Sea.

The harsh Ice Age climate 60,000 years ago would have made life away from the coast difficult for the new migrants. But with climatic improvements over the millennia people gradually carried haplogroup M inland. In Asia's interior they may have encountered migrants carrying haplogroup N, a sister branch of M that originated about 60,000 years ago in the Near East, possibly following an inland route to the central part of the the continent.

Once it had established itself in coastal Asia and some inland locations as well, M spawned multiple local haplogroups throughout the continent. Major branches of the M haplogroup include:

M1, which is centered in Ethiopia and appears to indicate a migration back into Africa about 30,000 years ago.

M2, the most common among myriad haplogroup M branches in India.

D, a major presence in central and eastern Asia and one of several haplogroups to reach the Americas during the Ice Age.

E, a Southeast Asian haplogroup found in Taiwan, Malaysia and the Pacific Islands.

G, a haplogroup found at high levels in northeastern Siberia among indigenous populations such as the Koryak and the Itel'men.

M7, a widespread haplogroup found in China, Japan, Southeast Asia and the Pacific Islands.

M8, a widespread haplogroup in central and eastern Asia that eventually sent an offshoot to the Americas.

M9, which appears to have arisen in Tibet.

### **M in India**

While [Haplogroup M](#) is widespread throughout South and East Asia, it is more diverse on the Indian sub-continent than anywhere else in the world. The high degree of diversity of M in India is likely tied to its ancient arrival here nearly 50,000 years ago. In addition to M2, which is found throughout the sub-continent, there are dozens of haplogroups branching off of M that exist in India. These branches are often connected to specific regions, tribes, or ethnic groups. For example, haplogroup M18 is found among the Oraon peoples of eastern India and Bangladesh, while haplogroup M41 is common among the Pardhan speakers of eastern India, and haplogroup M31a can be found on the Andaman Islands, just off the southeast coast of India.

There is an exceptional amount of diversity along the northeast corner of India and into Bangladesh. Some experts believe that the sheer number of haplogroups in this region may be the genetic signature of an ancient genetic corridor linking India with Southeast Asia, where several other branches of M exist today.

### **American Cousin**

The recent discovery of [haplogroup M](#) in ancient remains from North America could significantly expand the domain of this already widespread branch of the human [mitochondrial DNA](#) tree. M has never been found among living Native Americans, but two 5,000-year-old skeletons from British Columbia appear to belong to the haplogroup, according to a paper published in the April 2007 Journal of Archaeological Science. The discovery suggests that the small group of people who entered North America during the Ice Age belonged to not five, but six, haplogroups – A, B, C, D, X and M.

## **MATERNAL HAPLOGROUP N – from 23andme.com**

### **Introduction**

N is one of two major haplogroups involved in the initial modern human migration out of Africa. It arose about 64,000 years ago, not long after humans entered the Near East either via the Sinai peninsula or by crossing the Red Sea near its mouth. The [haplogroup](#) then spread to central Asia by an inland route, passing north of the Himalayas into present-day China and then south into southeastern Asia and Australia.

Subsequent migrations have spread N from Portugal to Polynesia. It has also spawned multiple offshoots, many of which are major haplogroups in their own right. H, V, U, W and X dominate in Europe and western Eurasia, where they were involved in the expansion of agriculture into those regions. A and B are widespread among both Asians and Native Americans, whose ancestors crossed a land bridge linking Siberia to Alaska more than 12,000 years ago.

## **MATERNAL HAPLOGROUP N9 – from 23andme.com**

**See N – no additional information given about N9**

## MATERNAL HAPLOGROUP A- from 23andme.com

### Introduction

[Haplogroup](#) A arose in Asia almost 60,000 years ago. It is now at its highest levels among Native American groups. Along with several other [mitochondrial DNA](#) types found only in Native American and Asians, haplogroup A provides clear evidence that the first people in the New World were migrants from Siberia and eastern Asia.

### A in Asia

[Haplogroup](#) A is widespread in Asia today, generally occurring at levels below 10% – but it reaches higher concentrations in some parts of China, Korea and Japan. Some ethnic Chinese populations, such as the Dong and the Yi, carry haplogroup A at levels as high as 30%. One branch of the haplogroup, A4, reaches levels of more than 15% among [mitochondrial DNA](#) samples collected in the city of Wuhan in central China.

*In the Spittoon...*

*Ancient China's famous Terracotta Army was constructed by men bearing haplogroup A. Check the [Spittoon](#) to learn more about these ancient builders.*

### Ancient DNA in Siberia

[Haplogroup](#) A was widespread in Siberia as recently as 7,000 years ago. One study of skeletal remains discovered near Siberia's Lake Baikal estimated the haplogroup was present in 13-26% of the region's population at the time. But the haplogroup is rare in the region today; it is found almost exclusively among the Chukchi and the Yupik, two small indigenous groups from northeastern Siberia.

### A in the Americas

At the peak of the Ice Age, between about 20,000 and 15,000 years ago, massive glaciers covered much of North America and Eurasia. So much water was locked up in the ice sheets that global sea level dropped 300 feet, creating connections between land masses that are isolated by wide straits or passages today. One of those connections was the Bering land bridge, an ice-free but frigid corridor hundreds of miles wide that linked Siberia and Alaska. Mammoths, bison, caribou and other Ice Age mammals roamed back and forth between Siberia and Alaska during this period, as did a few hardy hunter-gatherers who could cope with the region's extreme climate.

As the Ice Age ended, people began moving south from the Arctic into the heart of North America. Within a few thousand years, possibly even faster, the new arrivals had populated the Western Hemisphere down to the tip of South America.

### Coastal Connection

[Haplogroup](#) A is especially common among members of the Haida, Nuu-Chah-Nulth, Nuxálk (Bella Coola) and Chumash tribes of the Pacific coast. Humans had certainly reached what is now California by 10,000 years ago, as evidenced by the discovery of skeletal remains on islands off the state's coast that have been radiocarbon dated to that age. Other skeletal remains from near present-day Monterey, California have yielded ancient DNA tracing to the A haplogroup.

Haplogroup A is found in Central America and northern South America, but not farther south. That suggests that however people carrying the haplogroup moved into the Americas, their advance was ultimately impeded by earlier arrivals to the southern continent.

Haplogroup A certainly would have been found among the subjects of the Inca empire, which ruled the northern Andes until the arrival of Spanish conquistadors in 1526. [Mitochondrial DNA](#) belonging to the haplogroup was extracted from the "Ice Maiden,"

the mummified remains of a teenage Inca girl who died in a ritual sacrifice about 500 years ago.

### **Interior of North America**

While the distribution of A is patchier in the interior of North America, it does appear at high frequencies in many populations, particularly in the American Southwest, northern plains, and the southeastern United States.

About 50-60% of individuals from the Navajo and Apache carry [haplogroup A](#), while their neighbors rarely carry this haplogroup. Interestingly, the Navajo and Apache are both southern Athapaskan speakers that appear to have migrated from a homeland further north to the American Southwest only about 500 years ago. Athapaskan-speakers still reside in Alaska and Canada. Although they have adapted to the desert climate and the Pueblo lifestyle (at least the Navajo), their mitochondrial diversity still records their northern heritage.

Haplogroup A is also common in Algonquian-speaking populations from the Plains region of the United States and Canada, ranging from 30%-60%. Ojibwa/Chippewa, Cheyenne and Arapaho were all historically Algonquian-speaking populations from the Plains region and Great Lakes of United States and Canada.

Iroquoian-speaking groups such as the Mohawk also carry haplogroup group A at high frequencies (60%). It is not clear if this indicates that the Mohawk historically married women from other northern tribes or if the high frequency of A represents a recent bottleneck in the population. Iroquois from further south (e.g. Cherokee) have much lower frequencies of haplogroup A.

Haplogroup A is also quite common in Muskogean-speaking populations from the southeastern United States, reaching almost 75% in the Choctaw historically from Mississippi and 60% in the Seminoles of Florida. Genetic types in these southeastern

populations are different from other haplogroup A individuals throughout North America. This indicates that haplogroup A in the southeastern populations is the result of ancient common ancestry with the other Native Americans, rather than recent gene flow. Southeastern populations also display relatively low genetic diversity, indicative that the population size may have crashed after contact with European explorers.

### **The Polar Route**

A second migration of [haplogroup A](#) members into North America appears to have occurred more recently, as Eskimo-Aleut populations moved eastward from Siberia into Arctic Canada and Greenland about 5,000 to 6,500 years ago. That migration is marked by the wide distribution of A2 from the Haida on the Pacific coast of Canada to the Inuit in Greenland, where the subgroup accounts for 100% of the population.