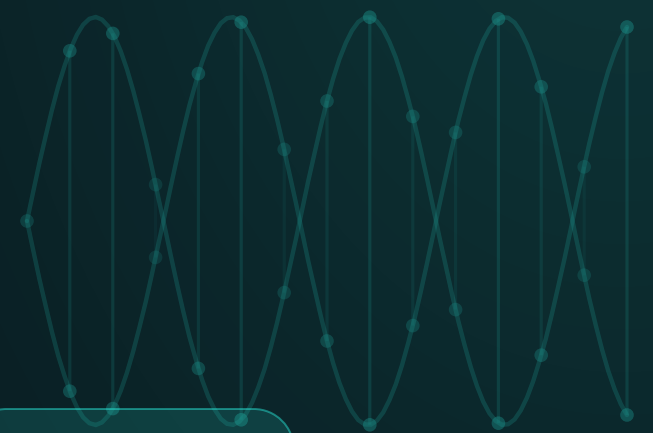


Lesson: lactose tolerance and natural selection



Biology

Genetics

Statistics

Social Studies

Age 13+

Intermediate

A free genetics lesson on natural selection, with no equipment needed. The class computes a prediction from real allele frequencies, gets a clean answer, and then finds out it is wrong — because the same solution evolved four separate times in the last few thousand years.

Free to read, use and adapt. No account, nothing to buy, nothing collected from anyone.

Full lesson, figures and updates: <https://www.mygenelog.com/updates/lesson-the-milk-mutation>

What this is. A genetics lesson about natural selection, free to use and adapt in any school, college or university. It needs no equipment and no genetic testing — a calculator and a board is the whole kit.

- Level: about age 14 through first-year undergraduate.
- Time: 45 minutes, or 30 if you set the reading beforehand.
- You need: nothing to buy. Calculators, or phones.
- It covers: Hardy–Weinberg, recessive inheritance, gene regulation versus coding change, convergent evolution, how selection is dated from DNA, and why a correct calculation can still give the wrong answer.

Start here: most adults in the world cannot drink milk

Every mammal drinks milk as an infant and stops being able to digest it afterwards. The enzyme that breaks lactose down, lactase, is switched off after weaning. That is the normal mammalian programme, humans included.

Some people carry on making it for life. That state is called lactase persistence, and it is the unusual one — the mutation, the exception. If you grew up somewhere where adults drink milk without thinking about it, the surprising thing is you.

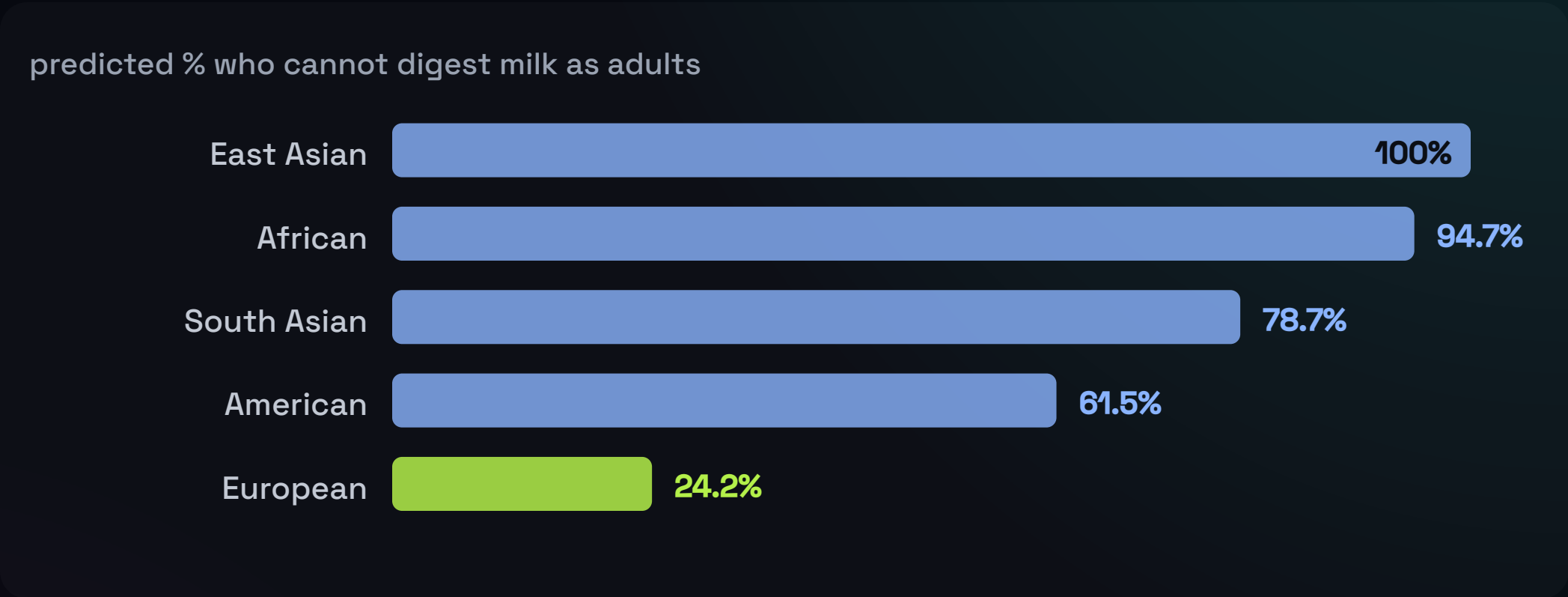
This lesson is about how that happened, how recently, and how we can tell.



Do this — the calculation

Here are real allele frequencies at a single position, rs4988235. The G allele is the ancestral one: two copies and lactase switches off after childhood. The other allele keeps it on, and one copy is enough.

- 1. Ask for a guess first. "Worldwide, what share of adults do you think cannot digest milk?" Write the class's guess on the board. Come back to it at the end.
- 2. Fill in the last column. Two copies of G are needed, so the predicted share is the frequency multiplied by itself. Five multiplications; give it three minutes.
- 3. Put the answers up and look at them. Ask which one surprises anybody.
- 4. Then ask the question this lesson turns on: is any of these numbers wrong? Not "surprising" — wrong. How would you find out?



The answers. One of these bars is badly wrong, and finding out which one is the lesson.



The number that is wrong

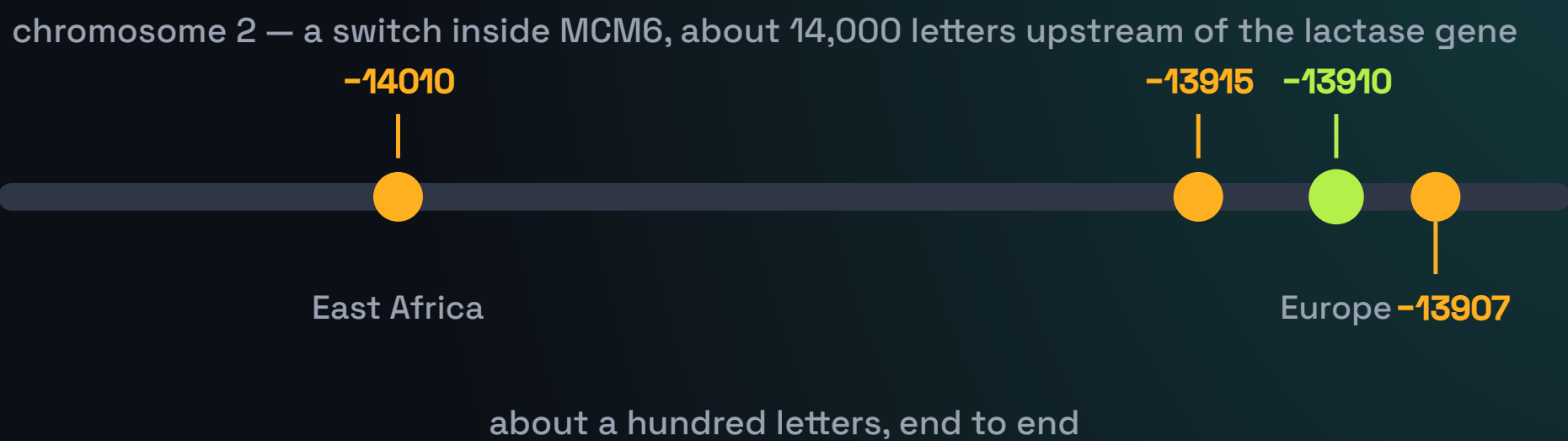
The African bar says almost 95% of adults should be unable to digest milk. Across the continent as a whole that is not far off. But it is badly wrong for particular peoples, and the ones it is wrong about are exactly the ones you would least expect: the cattle-herding communities of East Africa, whose diet has depended on milk for thousands of years.

Many of them digest milk perfectly well as adults. And they do not carry the European variant.



Four separate answers to the same problem

In 2007 a study genotyped 470 people in Tanzania, Kenya and Sudan and found three more variants associated with lactase persistence — none of them the European one.



The four positions, to scale with each other. Three of them were found in East Africa and arose on different chromosomes from the European one and from each other — so they are not one mutation that spread, but four that happened separately and do the same job.

All four sit within about a hundred letters of each other, in the same regulatory switch. All four turn the lactase gene back up. And each arose on a different chromosome background from the others, which is how we know they are separate events rather than one mutation that travelled.

This is convergent evolution: the same solution, found more than once, independently. Students usually meet it as wings in birds and bats — separated by hundreds of millions of years and unmistakably different underneath. Here it is the same species, the same gene, the same hundred base pairs, and the last few thousand years.

Where the switch actually is

Worth pausing on, because it is the part that surprises people who already know some genetics.

When the European variant was found in 2002, the researchers had already sequenced the lactase gene itself — the coding sequence and its promoter — and found nothing that tracked with who could digest milk. The gene is not the thing that differs.

What differs is a switch about fourteen thousand letters away, sitting inside an intron of a completely different gene, MCM6, which has nothing to do with digestion. The lactase gene is fine in everybody. What changed is the instruction telling it when to stop.

Most of what genome-wide studies find looks like this. Not broken proteins — altered instructions about when and how loudly a normal protein gets made.

How we know it is recent

Here is the part that lets you put a date on evolution, and it is a genuinely clever argument that students can follow.

When a new mutation appears, it appears on one particular chromosome, carrying its neighbours with it — a long stretch of DNA that travels as a block. Every generation, recombination cuts those blocks shorter. So a common variant is normally surrounded by short shared blocks: it has been around long enough to be chopped up.

A variant that is both common and still surrounded by a long unbroken block has not had time to be chopped up. It got common fast. And getting common fast is what selection looks like.

In northern Europe the persistence variant sits on a common haplotype running largely undisrupted for more than a million letters. In East Africa, one of the African variants shows haplotype homozygosity extending beyond two million. Both are far longer than their frequencies should allow.

The dates that come out: within the last 5,000–10,000 years in Europe, and about 7,000 for the East African one. Both are after animals were domesticated. The researchers describing the European signal called it among the strongest yet seen for any gene in the human genome.

Culture came first

The order matters and students usually get it backwards.

People did not start keeping cattle because they could drink milk. They kept cattle, and in populations that then depended on milk, anyone who could keep digesting it had an advantage — so the variant spread. The invention is cultural; the genetic change is the response to it.

Which is why it happened four times over. Four populations, four independent mutations, and the same cultural practice behind all of them.

Questions

To warm up

1. Why is lactase persistence described as the unusual state, when many people in the room may have it?
2. The predicted share needs the frequency multiplied by itself. Why squared, and not just the frequency?
3. Look back at the class's opening guess. How close was it, and which way was it wrong?

Core

1. Our calculation for Africa was arithmetically correct and gave the wrong answer. Say precisely what was wrong with it — in one sentence, about the assumption rather than the sum.
2. The four variants sit within a hundred letters of each other. Give one reason they might all be in the same tiny stretch of DNA, when the genome is three billion letters long.
3. Researchers say these arose separately rather than spreading from one origin. What evidence supports that? Describe the reasoning, not just the conclusion.
4. The lactase gene itself is identical in people who can and cannot digest milk. What does that tell you about where useful genetic differences tend to live?
5. Explain, to someone who has not read this, why a long unbroken stretch of shared DNA around a common variant means it became common recently.

To stretch

1. A company offers a test for "lactose intolerance" that reads only rs4988235. For which customers will it be wrong, and will it be wrong in a way they could detect?
2. Hardy–Weinberg assumes no selection. We just used it on a variant under some of the strongest selection known in the human genome. Does that invalidate the calculation? Argue both sides, then decide.
3. The 1000 Genomes labels — "African", "European" — cover hundreds of peoples with different histories. Say how that grouping produced our wrong answer, and how you would fix the study design rather than the arithmetic.
4. Design a study that would find a fifth persistence variant in a population nobody has looked at. Who would you recruit, what would you measure, and what result would convince you?

Notes for whoever is teaching — where these questions are going

2. Non-persistence is recessive: one working copy is enough to keep lactase on, so only people with two copies of G lose it. That is the frequency squared. Students who answer "because there are two chromosomes" have the mechanism and should be pushed to the arithmetic.

4. The assumption was that this one position is the only way to be lactase persistent. The sum was fine; the model behind it was too small. This is the sentence worth getting onto the board in the students' own words, because it generalises to every genetic test anyone will ever be offered.

5. Because that stretch is the switch. A mutation anywhere else in the genome does not turn lactase back on, so those mutations happened too and did nothing. We only see the ones that landed in the one place where landing does something. Survivorship, in evolution.

6. Different haplotype backgrounds. If they had one origin, they would sit on the same neighbouring stretch of DNA; they sit on different ones, so each began on a different chromosome in a different person.
8. Recombination shuffles a bit more each generation, so shared blocks get shorter with time. Long block plus high frequency is the contradiction: normally getting common takes long enough for the block to be cut down. Something made it common in a hurry.
9. Wrong for anyone whose persistence comes from one of the African variants — told they are lactose intolerant while they drink milk daily with no trouble. And no, they could not detect it from the result alone, which is the point: a test that reports one position cannot tell you it looked in the wrong place.
10. Both sides are real. Hardy–Weinberg is violated here, so the prediction is approximate. But the failure we found was not caused by selection — it was caused by leaving three variants out. Distinguishing “the model’s assumptions are imperfect” from “we measured the wrong thing” is a genuinely hard idea and worth the argument.
11. A continental label averages over peoples whose milk-drinking histories differ completely. Averaging East African pastoralists together with populations that never kept cattle produces a number describing nobody. The fix is sampling that matches the question, not a better formula.

Take it further

- rs4988235 — the position itself, with its sources.
- Lactose intolerance and lactase persistence — the full page, including what a diagnosis actually involves.
- Lesson: the bitter paper — the other classroom lesson here. It needs paper strips; this one needs nothing.
- What is a SNP — if the class needs the vocabulary first.

Use this freely. Print it, copy it, cut it up, translate it, change the questions, put the table on your own worksheet. No permission needed and nothing to pay. If you credit it, mygenelog.com is enough.

No student needs to be tested for anything, and nobody should be asked whether they or their family can drink milk. The whole lesson runs on published numbers.

Sources: the European variant — Enattah et al., Nature Genetics 2002, PMID 11788828.

Dating the European selection signal — Bersaglieri et al., American Journal of Human Genetics 2004, PMID 15114531. The three African variants and convergent evolution — Tishkoff et al., Nature Genetics 2007, PMID 17159977. Allele frequencies — 1000 Genomes Project via Ensembl. This post is educational and is not medical advice.