

TEACHER'S GUIDE

DNA *to Darwin Case Study*

Malaria and the human genome

**Steve Cross,
Bronwyn Terrill
and colleagues**

Wellcome Trust Sanger Institute
Hinxton

Version 1.0



Charles Darwin

Malaria and the human genome

Each year, the malaria parasite *Plasmodium falciparum* kills over a million African children and causes debilitating illness in over half a billion people worldwide. Malaria is the strongest known selective force in the recent history of the human genome. Many types of genetic variation have evolved in humans due to selection by the malarial parasite, causing variation in red blood cell regulation, structure and antigen expression.

In this *Case Study*, students investigate the origin and action of mutations that are thought to have arisen in human populations in response to selection pressure from malaria.

Outline of the activity

Malaria is a debilitating illness that affects more than 40% of the world's population that is caused by parasites of the genus *Plasmodium*. This disease is thought to be the strongest selective force on our species in recent history. Researchers believe that this is responsible for the diverse range of genetic adaptations that protect against malaria in different populations' genomes.

In this activity, students will use a common statistical test (chi-squared) to work out whether a genetic mutation is associated with incidence of the disease, or whether the two events are independent.

General reading

Reading the story in DNA: A beginner's guide to molecular evolution by Lindell Bromham (2008) Oxford University Press (Paperback) ISBN: 978 0199290918. *An engaging textbook on molecular evolution, which assumes no specialist mathematical knowledge and takes the reader from first principles. Although it's aimed at undergraduates, this superb book contains sufficient detail for PhD students, yet parts will appeal equally to 16-19 year-olds.*

The making of the fittest. DNA and the ultimate forensic record of evolution by Sean B. Carroll (2009) Quercus Books (Paperback) ISBN: 978 1847247247. *A popular lay account of some of the molecular evidence for evolution.*

The making of a tropical disease: A short history of malaria by Randall M. Packard (2011) The Johns Hopkins University Press (Paperback) ISBN: 978 1421403960.

Mosquito: The story of man's deadliest foe by Michael D'Antonio and Andrew Spielman (2002) Faber & Faber (Paperback) ISBN: 978 0571209859.



Scientific publications

Kwiatkowski, D.P. (2005) How malaria has affected the human genome and what human genetics can teach us about malaria. *American Journal of Human Genetics*, 77 (2) 171–92. The full text of this article is available at www.pubmedcentral.nih.gov/articlerender.fcgi?tool=pubmed&pubmedid=16001361

World Health Organisation (2008) *Malaria*
www.who.int/topics/malaria/en/

Kwiatkowski, D.P. (2007) *The Sanger Institute Malaria research programme*
www.sanger.ac.uk/pathogens/malaria/

Wellcome Trust (2008) *Malaria microsite*
<http://malaria.wellcome.ac.uk>

Pauling, L. (1977) *Sickle cell anaemia as a molecular disease*, from: ‘It’s in the blood’. Oregon State University. Video clip with interview and imagery available at <http://osulibrary.oregonstate.edu/specialcollections/coll/pauling/blood/videos/1977v.1-sicklecellanemia.html>

Requirements

Software

The data is presented for students to analyse as a *Microsoft Excel* spreadsheet. *Excel* can also be used to calculate the expected and chi-squared values, or students can be asked to work through the process manually using the tables in the spreadsheet provided.

JavaScript DNA Translator 1.1, is a simple tool provided for analysing DNA sequences within a web browser. It was written in 2002 by William Perry, who has allowed the code to be freely distributed as long as it remains unmodified; additional permission can be sought from the author at bperry@lilly.com.

Microsoft Excel

Students will need a computer equipped with *Microsoft Excel*.

Students’ worksheets

Students will require copies of Student’s Guide, pages 2–11.

Educational aims

Sickle cell trait, sickle cell anaemia and malaria are all rich areas of content to develop through and beyond the curriculum requirements. This activity provides you with the option of revising transcription and translation, while introducing aspects of How Science Works (hypotheses and data analysis) while processing DNA data.

Prerequisite knowledge

Students will need to understand the basic structure of DNA and how it codes for amino acids (the *Introductory Activities*, which are in a separate document, will be useful here).

It is essential that they understand the Chi-squared test and are competent at using spreadsheets in *Microsoft Excel*.

Using this activity in the classroom

In the classroom, each group could follow through the exercises as stated: first characterising the mutation and then considering hypotheses based on geographical spread. Alternatively, small groups could characterise the sickle cell mutation, research the symptoms and incidence of sickle cell trait and anaemia, research the *Plasmodium* life cycle, research the symptoms and spread of malaria, and analyse the SNP data. Each team would then have something to contribute to a comprehensive picture of the sickle cell mutation and its possible association with malarial incidence.

Answers to the questions on the worksheets

Exercise 1

Standard DNA sequence:

**ATG GTG CAT CTG ACT CCT GAG GAG AAG TCT GCC GTT
ACT GCC CTG TGG GGC AAG GTG AAC**

This translates to an amino acid sequence of:

**M V H L T P E E K S A V T A L W G K V N or
Met Val His Leu Thr Pro Glu Glu Lys Ser Ala Val Thr Ala
Leu Trp Gly Lys Val Asn**

Mutated (sickle cell) sequence:

**ATG GTG CAT CTG ACT CCT GTG GAG AAG TCT GCC GTT
ACT GCC CTG TGG GGC AAG GTG AAC**

This translates to an amino acid sequence of:

**M V H L T P V E K S A V T A L W G K V N or
Met Val His Leu Thr Pro Val Glu Lys Ser Ala Val Thr Ala
Leu Trp Gly Lys Val Asn**

Questions on Page 8, referring to the above sequences

- The sequence difference is an A to a T substitution.
- An E (Glu; Glutamic acid) is replaced by a V (Val; Valine).

Ideas for further research/investigation

How does this single amino acid alteration make cells sickled?

At low oxygen concentrations, Hb^S tends to polymerise, or form long chains of molecules, causing the blood cell to elongate and 'sickle'.

A 3-D animation of sickled cells and the DNA changes associated with them can be found at:

<http://www.dnai.org/text/mediashowcase/index2.html?id=609>

Why might that make a difference to incidence of malaria?

In people with sickle cell trait, the presence of the modified form of haemoglobin in red blood cells could mean that the cells no longer support the mass-reproduction of the malaria parasites during the merozoite stage of their lifecycle. Malarial protection in sickle cell trait could also come from an increased clearance of red blood cells (including those that contained proliferating parasites).

Exercise 2

These results were obtained from *MalariaGEN*, a consortium of researchers in 20 countries. Small changes have been made to the dataset to comply with ethical data release guidelines.

Technically the probability curves in this exercise are binomial, not normal. However, the large sample sizes involved mean that it's valid to approximate with normal distributions (and therefore use the chi-squared test). Any factor that researchers would use to compensate for this approximation would be incredibly small compared to the data set and the effect seen.

In the *Excel* document called **malaria_data_teacher**, there are four worksheets: **Cases**, **Controls**, **Student_table** and **Teacher_table**. The **malaria_data_student** file has only the **Cases** and **Controls** worksheets. On the **Cases** and **Controls** worksheets, you will find 500 people's genotypes at specific locations in the genome. The **Student_table** worksheet tallies the data to make things easier and shorten the exercise but doesn't complete the calculations. It has been included in the teacher's materials for you to copy and paste into the students' *Excel* files if you think this is appropriate.

In the **Case** and **Controls** worksheets, the column you're interested in is Column E: HbS. However, the exercise can be extended to the other locations to compare the strong effect of HbS against other weaker or non-significant effects.

Category	Observed (O)	Expected* (E) (see below)	O-E	(O-E) ²	$\frac{(O-E)^2}{E}$
Has malaria Has Hb ^S (T)	8 from AT + 3 from TT + 3 from TT = 14	43	-29	841	19.558
Has malaria No Hb ^S (A)	489 + 489 both from AA + 8 from AT = 986	957	29	841	0.879
No malaria Has Hb ^S (T)	64 from AT + 4 + 4 from TT = 72	43	29	841	19.558
No malaria No Hb ^S (A)	432 + 432 both from AA + 64 from AT = 928	957	29	841	0.879
TOTALS	2000				40.47 (χ^2)

* To calculate the Expected values, you will need to count the number of Hb^S alleles found across the entire population under test (1,000 people and therefore 2,000 chromosomes). This figure, divided by the total chromosome population size (2,000) will give you the prevalence of the Hb^S allele. Multiplying the prevalence of each form of the allele by the number of people's chromosomes in each test group, will give you the expected values.

Category	Size of population (ignoring alleles)	Prevalence of allele	Expected
Has Hb ^S	86 ÷ 2000	0.043 × 1000	43
No Hb ^S	1914 ÷ 2000	0.957 × 1000	957

The total of $\frac{(O-E)^2}{E}$ across all four groups is the chi-squared value (χ^2). There is only one degree of freedom in this experiment. Using this information, calculate the probability of the null hypothesis being true using this lookup table, where p is the probability.

p	0.25	0.2	0.15	0.1	0.05	0.025	0.01	0.005	0.001	0.0005
χ^2	1.32	1.64	2.07	2.71	3.84	5.02	6.63	7.88	10.83	12.12

An alternative approach is to use the **student_table** or **teacher_table** worksheet and to carry out all the calculations in *Excel*:

Observed

Category	Case	Control	Total
Has Hb ^S (T)	14	72	86
No Hb ^S (A)	986	928	1914
TOTALS	1000	1000	2000

Estimated

Category	Case	Control
Has Hb ^S (T)	43	43
No Hb ^S (A)	957	957

Questions on Page 11

- Less than 0.00005
- Yes; this means that the null hypothesis should be rejected and the alternative hypothesis accepted.
- This information reinforces the assumptions from geographical spread.

Important note

This data set is a subset of the 1,000 case and 1,500 control samples collected for the *MalariaGEN* study. Because the sample size has been decreased for this exercise, a statistically significant result is only obtained for the Hb^S data. A similar calculation with the ABO-related SNP subset (column D), produces a p value too high to be statistically significant. However, the full set of data, has given the researchers a result of $p = 0.0002$. This could be an additional discussion point.

Possible extension activities

There are several other genetic disorders that are found predominantly in the same regions as the malarial parasite. These include β -thalassaemia. Research β -thalassaemia to find out about the symptoms and spread. Why do you think it is found mainly in these regions?

There is another genetic variant of the β -globin gene found in some parts of West Africa called Hb^C. This is a change in the sequence immediately before the Hb^S mutation. Hb^C confers a resistance to malaria that is about 80% as effective as that conferred by Hb^S. In the homozygous form, Hb^C also causes a less severe form of sickle cell anaemia than Hb^S does. This allele seems to have arisen more recently than Hb^S. What do you predict will happen to the frequency of this allele in the population, given that Hb^C Hb^C homozygotes are much more likely to pass on their genes by having children than Hb^S Hb^S homozygotes?

Acknowledgements

This activity was developed by Steve Cross with members of the Wellcome Trust Sanger Institute Malaria Genetics team (Dominic Kwiatkowski and Bronwyn MacInnis) and the Communication and Public Engagement team (Bronwyn Terrill, Elizabeth Cooper-Williams, and Preeti Deshpande).

Thanks to the Malaria Genetics team for contributing and anonymising the sample data for this exercise.