

## Lab 05: The evolution of human resistance to malaria

### Introduction

In this case study, you will evaluate data that reveal the relationship between frequencies of a rare hemoglobin allele in human populations, and resistance to malaria, a disease that today is the third most deadly infectious disease in the world. Several mechanisms of evolution come into play in this story: mutation, migration, and natural selection. All of these mechanisms shed light on the intricate interrelationship between humans and this disease. You will be given several datasets and several analytical techniques, and provided with the opportunity to evaluate the nature of the interaction between humans and malaria. By answering a series of questions, you will discover whether there is evidence for ongoing evolution in human populations.

### Human variation in hemoglobin<sup>1</sup>

Hemoglobin, the molecule that carries oxygen in the red blood cells of human children and adults, is a protein made of two alpha subunits and two beta subunits. The genes that encode the alpha and beta subunits are both known to have several alleles. Many of these alleles derive from point-mutations in the DNA sequence that lead to single amino-acid substitutions in the protein. Most of the mutant alleles are very rare, because they reduce the efficiency of oxygen transport by red blood cells.

In sub-Saharan Africa, there is a surprisingly high frequency of an allele that produces an abnormal form of the  $\beta$ -subunit of the hemoglobin molecule (see figure 1).

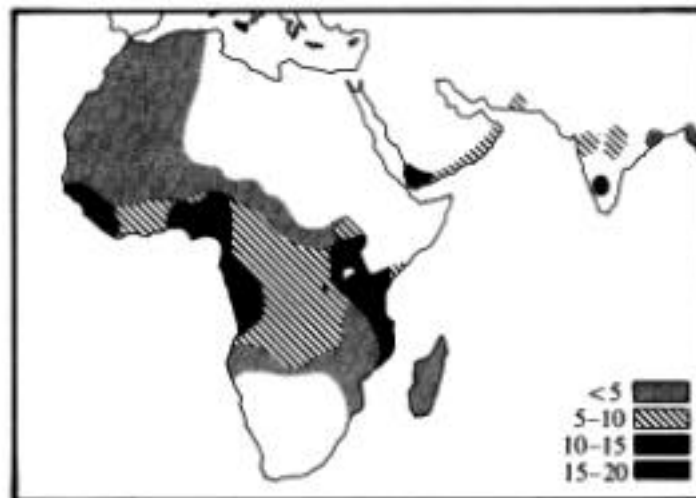


FIGURE 4.8  
Frequency of the sickle-cell gene  $Hb^S$  in various parts of the Old World. The key in the right-hand corner shows the percentages of the populations that bear the gene. (After Allison, 1961.)

**Figure 1.** The distribution of  $Hb^S$  in Africa.

From: Cavalli-Sforza, L. L. and W. F. Bodmer. 1971. *The Genetics of Human Populations*. W. H. Freeman and Company, San Francisco, CA. Figure 4.8 on p. 149.

---

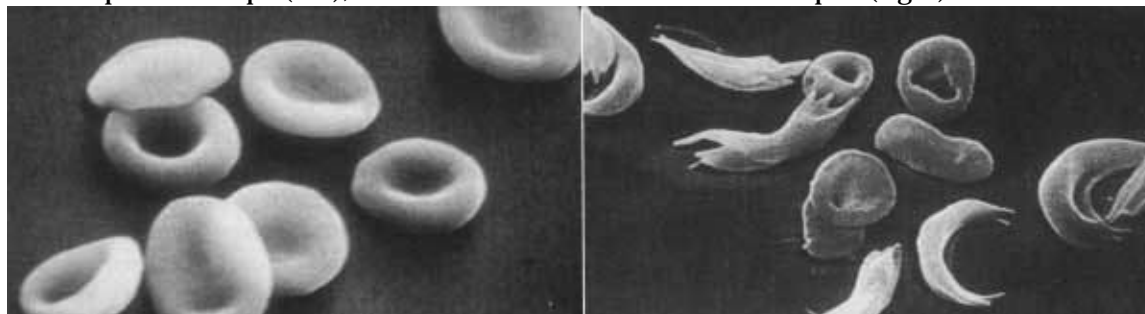
<sup>1</sup> Hemoglobin references:

Cavalli-Sforza, L. L. and W. F. Bodmer. 1971. *The Genetics of Human Populations*. W. H. Freeman and Company, San Francisco, CA.

Collins, F. H. and S. M. Paskewitz. 1995. Malaria: Current and future prospects for control. *Annu. Rev. Entomol.* 40: 195-219.

DNA Learning Center, Cold Spring Harbor Laboratory, One Bungtown Road, Cold Spring Harbor, New York 11724. <http://darwin.cshl.org/sickle.html>

The allele is the result of a single point mutation in the coding region of the  $\beta$ -subunit gene. This mutation results in a substitution of valine for glutamic acid in position six of the  $\beta$ -subunit. This substitution alters the shape of the hemoglobin molecule. Hemoglobin molecules containing altered  $\beta$ -subunits will crystallize under conditions of low oxygen tension. When this happens, the red blood cells which carry the hemoglobin molecules change shape drastically (see figure below). Instead of their normal biconcave shape (left), the red blood cells become sickle-shaped (right).



(from: Freeman, S. and J. Herron. 1998. Evolutionary Analysis. Prentice-Hall, Inc. NJ. p. 105).

Figure 2.

We will use the symbol  $A$  to represent the allele for the normal form of the  $\beta$ -subunit, and the symbol  $S$  to represent the mutant form responsible for cell sickling. The three possible genotypes in a population containing both alleles are  $AA$ ,  $AS$ , and  $SS$ . People with the genotype  $AA$  make normal hemoglobin and have a normal phenotype. People with genotype  $SS$  produce hemoglobin that is highly prone to sickling, and as a result suffer from anemia caused by poor oxygen delivery by the sickle-shaped cells in the blood. Furthermore, periodic and painful crises can lead to serious organ damage. This condition, known as sickle-cell anemia, is often fatal. Many  $SS$  homozygotes die before the age of reproduction.<sup>2</sup> People with the genotype  $AS$  produce a mixture of normal and abnormal hemoglobin. For the most part, the production of normal hemoglobin compensates for the production of mutant hemoglobin, and the heterozygotes do not suffer from sickle cell anemia (although they are, of course, carriers of the sickle-cell trait). In other words, allele  $S$  is recessive to  $A$ .

**Calculate the expected frequency of alleles  $A$  and  $S$  in a population containing both.**

As seen in figure 1, in sub-Saharan Africa the frequency of allele  $S$ , which is greater than 0.12 in some areas, is surprisingly high. The basis for this claim is a calculation of what the frequencies of  $S$  and  $A$  ought to be, given the information we have just reported about the fitnesses of phenotypes associated with each of the three possible genotypes. Let  $p$ =frequency of  $A$ , and  $q$ =frequency of  $S$ .

Fill out the following table:

genotype	frequency if in H-W	relative fitness (in terms of $s$ , not numbers)	effect of selection	new frequency
AA				
AS				
SS				

What is the new value frequency of  $S$ ?

<sup>2</sup> Medical researchers recently discovered that treatment with hydroxyurea dramatically decreases the frequency of crises. See Charache, S., et al. 1995. Effect of hydroxyurea on the frequency of painful crises in sickle cell anemia. New England Journal of Medicine 332: 1317-1322.

What is  $\Delta q$ , the change in the frequency of *S*?

Since we want to know what *p* and *q* should be if the information above is complete, we want the equilibrium frequencies of *p* and *q*, called  $\hat{p}$  and  $\hat{q}$ . Equilibrium is reached when the change in frequency is zero. What are  $\hat{p}$  and  $\hat{q}$ ?

*Because the frequency of the S allele in sub-Saharan Africa is much greater than zero, there must be some process competing with natural selection against the S allele that maintains the S allele in the population.*

**Is that process mutation?**

The rate of mutations converting allele *A* to allele *S* would have to be  $10^4$  or higher to suggest that a balance between selection and mutation is maintaining the *S* allele at a frequency of just 0.03 in a population (see Hartl and Clark 1997<sup>3</sup>). This mutation rate is extraordinarily high, and the frequency of the *S* allele is considerably higher than 3% in many parts of Africa (see map above). You can confirm this conclusion using the program *Simul8*. Thus mutation alone can't account for the high frequency of *S* in sub-Saharan Africa.

**Is it something else?**

What evolutionary processes could account for the high frequency of *S*?

What evolutionary processes could not?

*Clearly we have insufficient information to understand the frequency of the S allele. Here is more.*

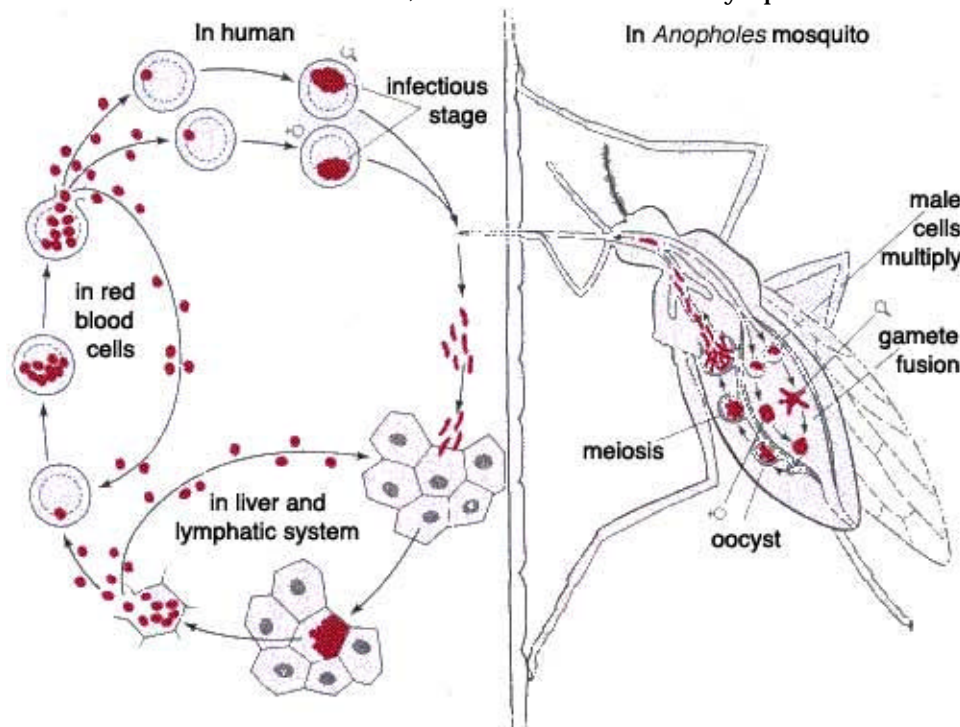
---

<sup>3</sup> Hartl, D. L., and A. G. Clark 1997. *Principles of Population Genetics*. 3rd edition. Sinauer Associates, Inc. Sunderland, MA.

### The biology of malaria<sup>4</sup>

Malaria, a disease which can cause debilitating anemia and potentially fatal brain blood clots in people, is caused by four species of protozoa (single-celled eukaryotes) in the genus *Plasmodium*. By far the most pernicious of the four is *Plasmodium falciparum* which is common in sub-Saharan Africa. The plasmodium is transmitted from one person to another by way of a vector, the mosquito.

Several tropical mosquito species are capable of transmitting the plasmodium, and as a result, the disease is endemic to many tropical areas. In Africa, where 90% of the world's malaria cases exist, the mosquito *Anopheles gambiae* is the main vector for the plasmodium. Mosquitoes transmit the disease by drawing blood from an infected person. With the blood, they also draw up plasmodia. These plasmodia live in the gut of the mosquito, where they reproduce sexually, and move to the mosquito's salivary gland. When the mosquito then draws blood from an uninfected person, the plasmodium is transferred to that person in the saliva of the mosquito. The plasmodia invade the human blood stream and quickly move to the liver where they multiply by asexual reproduction. The offspring of these invaders move back into the blood and inhabit red blood cells. Further asexual reproduction causes the red blood cells to burst, which causes the severe symptoms of the disease.



(modified from: Freeman, S. and J. Herron. 1998. Evolutionary Analysis. Prentice-Hall, Inc. NJ. p. 159).

Figure 3.

### The Relationship between S and malaria

In 1949, J. B. S. Haldane<sup>5</sup> suggested that the reason that the deleterious S allele occurs in high frequency in some human populations is because individuals who are heterozygous for the allele (genotype AS) do not suffer from the severe anemia due to cell sickling, and also enjoy some resistance to malaria. Thus the heterozygote has higher fitness than either homozygote. He based this proposition solely on the observation of higher than expected frequencies of allele S in regions where malaria is

<sup>4</sup> Malaria references:

Collins, F. H. and S. M. Paskewitz. 1995. Malaria: Current and future prospects for control. *Annu. Rev. Entomol.* 40: 195-219.

World Health Organization, Division of Control of Tropical Diseases.

<http://www.who.ch/programmes/ctd/diseases/mala/malamain.htm>

<sup>5</sup> Haldane, J. B. S. 1949. Disease and evolution. *Ricerca Sci.* 19:(Suppl. 1):3-10.

endemic. He undoubtedly generated a list of possible evolutionary processes, just like yours, and chose the most plausible.

If individuals with genotype *AA* are less resistant to malaria than *AS* individuals, their fitness will be lowered by some fraction, *t*, relative to individuals with the genotype *AS*. Of course, individuals with genotype *SS* still suffer from sickle-cell anemia, so their fitness is still lowered by some fraction, *s*. Thus we can represent the fitnesses of the three genotypes in a population as follows:

Hemoglobin genotype	<i>AA</i>	<i>AS</i>	<i>SS</i>
Fitness	1- <i>t</i>	1	1- <i>s</i>

**Calculate the expected frequency of alleles *A* and *S* in a population containing both.**

We need to recalculate our prediction of the equilibrium values  $\hat{p}$  and  $\hat{q}$  based on this new hypothesis.

Fill out the following table:

genotype	frequency if in H-W	relative fitness (in terms of <i>s</i> , not numbers)	effect of selection	new frequency
<i>AA</i>				
<i>AS</i>				
<i>SS</i>				

What is the new frequency of *S*?

What is  $q$ ?

What is  $\hat{q}$ ?

Does Haldane's hypothesis successfully explain the high frequency of *S* in sub-Saharan Africa?

*The equilibrium frequency of **S** in the population is a function of the competing strengths of the two opposing forces of natural selection: selection against **SS** individuals because of sickle-cell anemia, and selection against **AA** individuals because of susceptibility to malaria. Through this process, wherein the heterozygote has a fitness advantage over either homozygote, the frequency of **S** may remain at levels seen in sub-Saharan Africa, even if the fitness of **SS** homozygotes is extremely low.*

**What does the selective advantage of having genotype AS have to be in order to see the observed frequency of S (as high as 0.15) in sub-Saharan African populations? (In other words, how high does t have to be if  $s = 1$ , and the frequency of S is 0.15 in a population?)**

*Even though Haldane's hypothesis works, the question remains whether the S allele actually influences resistance to malaria.*

**Do the data support Haldane's prediction that individuals with the S allele have greater resistance to malaria than AA individuals?**

It wasn't until the 1960's that data that could address this prediction became available. Below you will find the original data<sup>6</sup> that were used to test Haldane's hypothesis. They are a compilation of 10 independent studies on the incidence of severe plasmodium infection among children with or without the sickle-cell allele.

Examine the data, and show whether the evidence supports or rejects the hypothesis that the S allele confers resistance to infection by the plasmodium that causes malaria. (Most of the individuals in the category 'with S allele' were heterozygotes, but a few were SS homozygotes.) Use the method for analysis described below.

Population	Region	With S allele		AA Homozygotes	
		Infected	Uninfected	Infected	Uninfected
1	Uganda	12	31	113	134
2	Kenya	131	110	154	87
3	Uganda	73	118	494	515
4	S. Ghana	42	131	270	572
5	N. Ghana	11	4	165	12
6	Nigeria	162	51	680	210
7	N. Ghana	13	6	109	18
8	Nigeria	51	40	245	97
9	Tanganyika	77	59	272	135
10	S. Ghana	34	89	176	417

Method for Analysis:

For each of the ten populations, calculate the frequency of infection among people with the S allele, and among people without it. If Haldane's hypothesis is supported, then there should be higher rates of infection among AA individuals than among AS or SS individuals. Does this appear to be true?

frequency of infection, with S:

frequency of infection, without S:

Is Haldane's hypothesis supported?

<sup>6</sup> from: Allison, A. C. 1965. Polymorphism and natural selection in human populations. Cold Spring Harbor Symposium in Quantitative Biology. 29:137-149.

Could we have gotten this result by chance, even if the S allele has no effect on resistance to malaria? Answering this question requires a statistical test.

- 1) Specify a null hypothesis. In this case, our null hypothesis is that there is no association between genotype and resistance to malaria.
- 2) Calculate a test statistic. We will take as our test statistic the number of populations in our set of ten for which the rate of infection is higher among AA individuals than among AS or SS individuals.

test statistic:

- 3) Determine the probability that chance alone could have made the test statistic as high as it is. Under the null hypothesis that genotype does not affect resistance to malaria, determining whether the frequency of infection is higher in individuals with or without the S allele is like tossing a coin. The true frequency of infection should be the same in both groups, with the group that appears to have the higher frequency in any given study determined by chance. If we toss a coin ten times and count the number of heads, we could get a number anywhere from 0 to 10. Figure 4 below shows the probability of each possible result. The table shows the total probability of getting 6 or more heads, 7 or more heads, and so on.

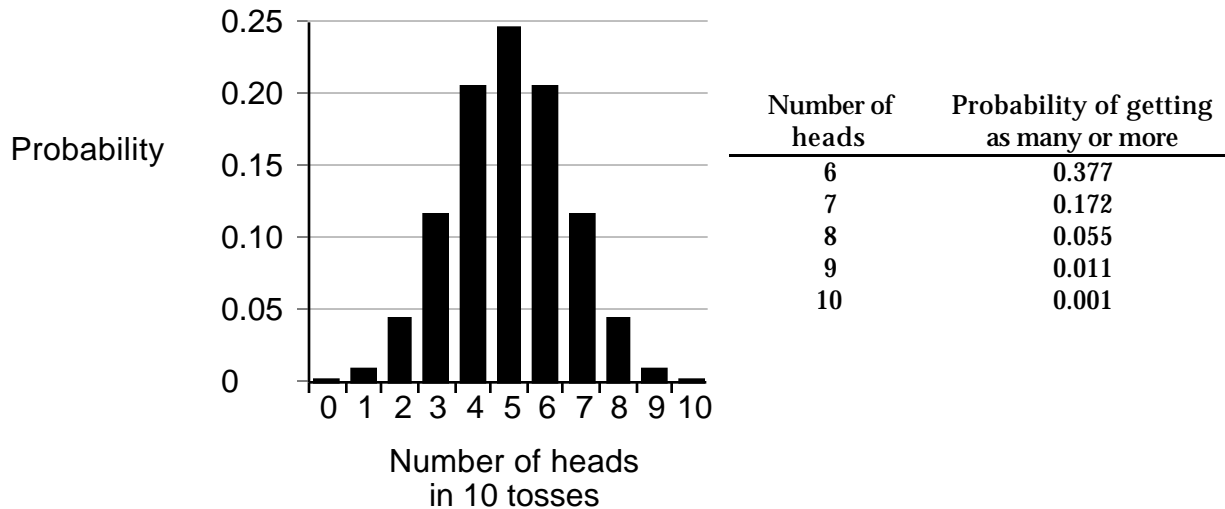


Figure 4.

- 4) Decide whether the outcome is statistically significant. If the value of the test statistic is 9 or 10, then we can consider our result to be significant. In other words, we can conclude that the data demonstrate that having at least one copy of the S allele makes a person more resistant to malaria.

probability of getting test statistic by chance:

What is your conclusion about the null hypothesis, based on this statistical test?

*Now that we have compelling evidence that the sickle cell allele protects individuals from infection by the plasmodium that causes malaria, we can test the hypothesis that it is the process of natural selection that explains high frequencies of this allele in human populations from regions where malaria is endemic. To do this, we need to evaluate the kinds of data that can show evidence that tolerance to malaria is the mechanism by which natural selection maintains high frequency of the S allele in human populations.*

### Approaches and Data:

Approach 1: Examine historical data showing an increase in the frequency of the allele with the appearance of endemic malaria.

We have no historical record of human hemoglobin allele frequencies prior to the emergence of malaria as an endemic tropical disease. In fact, data suggest that the plasmodium that causes malaria may have been around as long as the genus *Homo* (Cavalli-Sforza et al. 1994). However, current research is in progress that may approach this question. In our DNA are the footprints of natural selection. When balancing selection occurs, it leaves characteristic traces in the variability of the DNA sequence surrounding the locus under selection (See textbook chapter four, pages 228-230; and chapter seven, pages 376-377.) Michael Nachman at the University of Arizona and his colleagues are now analyzing the sequence variation surrounding hemoglobin genes to look for evidence of a history of balancing selection on the *S* allele.

Approach 2: Examine correlative data showing an association of high frequencies of the allele in regions of endemic malaria.

We have maps of the distribution of the *S* allele in Africa, as well as maps of the distribution of malaria prior to the 1950's, when global intervention against the disease began. These maps, shown below (figure 5), show that there is some correspondence between high frequency of *S* and endemic malaria. The maps, however, do not correspond perfectly, and we can only conclude that these patterns are not inconsistent with the hypothesis that malaria-resistance and the sickle cell trait are causally related.

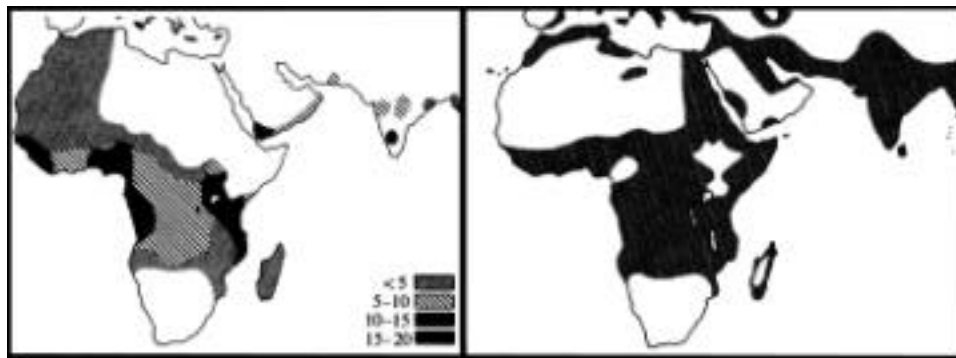


Figure 5. The distribution of  $Hb^S$  in Africa (left), and the distribution of malaria prior to 1950 (right). Modified from: Cavalli-Sforza and Bodmer 1971, figures 4.8 and 4.9 on p. 149.

Approach 3: Evaluate experimental data<sup>7</sup> showing that when malaria is removed, the frequency of the *S* allele declines.

Several hundred years ago, African populations were forcibly removed from Africa and brought to the United States and sold into slavery. Some of these people moved to the Caribbean, and their descendants have lived there ever since. Some Caribbean regions have endemic malaria and some do not. Thus, the setting of a “natural experiment” exists. We can evaluate the frequency of *S* in populations that originated in Africa where malaria was endemic and ask whether the frequency of *S* remains high when these people were removed from Africa but relocated to malarial regions. We can also ask whether the frequency of *S* has declined among these Caribbean peoples living in regions without malaria.

---

<sup>7</sup> data from: Allison, A. C. 1965. Polymorphism and natural selection in human populations. Cold Spring Harbor Symposium in Quantitative Biology. 29:137-149 (table 4) .

**Malaria present**

Population	SS or AS	AA
Honduras	164	541
Surinam 1	87	432
Surinam 2	35	137
Surinam 3	31	244
Surinam 4	51	292
Jamaica	117	901
St. Lucia	115	710

**Malaria absent**

Population	SS or AS	AA
Curaçao	108	1394
St. Vincent	65	683
Dominique	63	601
Barbados	64	848

**Analyze these data and draw your own conclusions.**

One method for analyzing the data is as follows:

First, calculate the frequency of the *S* allele in each of the 11 populations. For this calculation, assume that *SS* individuals are rare, and that to a first approximation all individuals in the category “*SS* or *AS*” are heterozygotes. Does it appear to be the case that the frequency of the *S* allele has fallen where malaria is absent? (Show your work.)

Now perform a statistical test as follows (this test is Wilcoxon’s rank sum test, described in many basic statistics books):

- 1) Take as the null hypothesis the claim that the frequency of the *S* allele is the same in regions where malaria is present as in regions where malaria is absent.
- 2) Calculate a test statistic as follows. Arrange all 11 populations in a list, in order of their frequency of the *S* allele. The population with the lowest frequency should be first; the population with the highest frequency should be last. The position of each population in the list is its rank. Take the sum of the ranks of the four populations from areas without malaria. For example (not the actual result), if Curaçao is first, St. Vincent seventh, Dominique fourth, and Barbados ninth, then the sum of their ranks is 21.

- 3) Under the null hypothesis, our procedure for calculating the test statistic is just like picking at random four numbers between 1 and 11 (such that no number is picked twice), then adding them up. The answer could be anything from  $1 + 2 + 3 + 4 = 10$  to  $8 + 9 + 10 + 11 = 38$ . Figure 6 below shows the

probability of each possible result. The table shows the total probability of getting a sum of 16 or smaller, 15 or smaller, and so on.

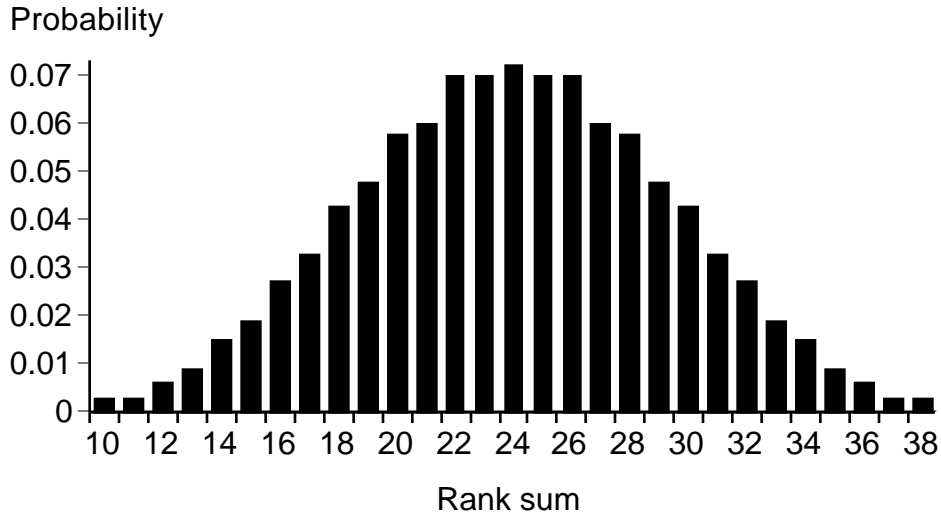


Figure 6.

Rank sum	Probability of getting a rank sum this small or smaller
16	0.082
15	0.055
14	0.036
13	0.021
12	0.012
11	0.006
10	0.003

4) If the value of the rank sum we calculate from the actual data is 14 or smaller, we can consider our result to be statistically significant. In other words, we can conclude that the frequency of the *S* allele is lower in African-Caribbean populations living in areas without malaria than it is in similar populations living in areas with malaria.

What is your result and your conclusion?

|