Assessing selection hypotheses for the CCR5-Δ32 mutation in Europeans

Introduction

The human immunodeficiency virus (HIV), the causative agent leading to acquired immunodeficiency syndrome (AIDS), has become a pandemic that has devastated many regions of the world. Yet in some European populations, there is a mutation called CCR5-Δ32 that confers almost complete protection against HIV in homozygotes and partial protection in heterozygotes. Many scientists suggest some force other than HIV has exerted positive selection for this mutation allele. Several hypotheses and models have been proposed as to how this mutation arose as well as how its frequency may have increased over time. This paper seeks to investigate the cause(s) of positive selection proposed by these hypotheses and models. The hypotheses and models of interest include: Duncan et al's plague hypothesis, Galvani and Slatkin's smallpox hypothesis, Balanovsky et al's ecological factors model, and a Bronze Age hypothesis suggested by Sabeti et al and Hedrick and Verrelli.

Background

CCR5 is a transmembrane chemokine receptor that strains of HIV-1 use to enter immune cells such as macrophages and CD4 T-cells. A 32 base-pair deletion in CCR5 (called CCR5-Δ32) prevents the expression of the CCR5 receptor on the cell surface (Duncan et al, 2005). The CCR5-Δ32 mutation allele thus interferes with the ability of HIV-1 to infect immune cells. Other mutations in the CCR5 gene and genes of the chemokine receptor gene family also reduce infection rate, but their effects on HIV infection are significantly less compared to the effects of CCR5-Δ32 (Balanovsky et al, 2005).

Individuals homozygous for CCR5-Δ32 have almost complete resistance to HIV-1 infection (Galvani and Slatkin, 2003; Balanovsky et al, 2005; Duncan et al, 2005). Heterozygous individuals for CCR5-Δ32 have partial resistance conferring a 70% reduced risk of HIV infection compared to individuals without the mutation allele (Duncan et al, 2005). Heterozygous individuals who do become infected with HIV have a slow progression towards AIDS where onset is delayed for two to three additional years (Balanovsky et al, 2005).

The frequency of CCR5-Δ32 is estimated to be about 10%, on average, in European populations (Galvani and Slatkin, 2003; Balanovsky, 2005; Duncan et al, 2005). In addition, there is a north-south gradient in allele frequencies with the highest allele frequencies in northern European populations – about 16% in Finnish and Russian populations – and the lowest in southern European populations – about 4% in Sardinia (Balanovsky et al, 2005; Duncan et al, 2005). The CCR5-Δ32 allele is found in low frequencies in regions neighboring Europe such as North Africa, the Middle East, and Central Asia (Balanovsky et al, 2005), but is absent in East and South-east Asia, sub-Saharan Africa, and indigenous American populations (Balanovsky et al, 2005; Duncan et al, 2005).

According to Balanovsky et al, this pattern is unusual for genes in human populations. In 1998, Libert et al determined the CCR5-Δ32 mutation to be about 2000 years old using microsatellites linked with the CCR5 gene. Another study that same year by Stephens et al found strong linkage disequilibrium between two microsatellite markers and CCR5-Δ32, showing the CCR5-Δ32 mutation to be only about 700 years old (Balanovsky et al, 2005; Duncan et al, 2005; Sabeti et al, 2005). Despite this seemingly wide range in age estimates for CCR5-Δ32, 95% confidence intervals in both cases span a few thousand years (Balanovsky et al, 2005). Some of the studies assessed in this paper appear to use the age estimate of 700 years old.

Most studies on CCR5-Δ32 believe this mutation has a single and recent origin in Europe where the frequency of this mutation increased dramatically from 0% to 10% in a historically short period of time. Balanovsky et al also state that such a dramatic increase in allele frequency in a large population cannot be due to random genetic drift. As a result, most studies also believe positive selection should be considered (Galvani and Slatkin, 2003; Balanovsky et al, 2005; Duncan et al, 2005).

Plague Hypothesis

In 1998, Stephens et al hypothesized that a plague epidemic caused the rapid increase in the mutation frequency (Balanovsky et al; 2005). The Black Death between 1346 and 1352 killed up to 40% of Europe's population and has been suggested to be this single plague epidemic (Galvani and Slatkin, 2003; Duncan et al, 2005). The bubonic plague was assumed to be the cause of the Black Death pandemic and after the Black Death, plague epidemics occurred approximately every 10 years for about 400 years (Duncan et al, 2005). A second pandemic known as the "Great Plague" in 1665 and 1666 killed 15-20% of Europe's population, but after this pandemic the plague declined in Europe and virtually disappeared from Europe in 1750 (Galvani and Slatkin, 2003). Furthermore, the area where CCR5-Δ32 is found today corresponds with the range of the plagues (Duncan et al, 2005). Because the plagues resulted in high mortality, there would be significant selection for any "inherited plague-resistance-inducing factor" (Balanovsky et al, 2005).

While this hypothesis and model for the dramatic frequency increase of CCR5-Δ32 is largely accepted, it has also recently fallen under heavy criticisms. Balanovsky et al point out that this hypothesis relies heavily on statistical analyses and that it has not been proven clinically

or molecularly. According to Duncan et al, the Black Death would at most double the frequency of CCR5-Δ32 from an initial frequency of $5x10^{-5}$ to 10^{-4}. Galvani and Slatkin suggest that the indirect transmission from rodent reservoirs to humans via fleas explains the intermittent emergence of plague epidemics, and that this may further limit positive selection. Furthermore, the pathogen that causes the bubonic plague, the bacteria *Yersinia pestis*, does not use the CCR5 receptor to infect (Duncan et al, 2005).

Duncan et al have thus hypothesized a modified version of the plague model. They suggest that an unknown virus that causes hemorrhagic fever with a 100% case mortality was the true pathogen responsible for the Black Death. This virus enters immune cells via the CCR5 receptor and the regular plague epidemics of Europe's Middle Ages serve to dramatically increase the frequency of CCR5-Δ32.

The results of Duncan et al's modeling show how the frequency of CCR5-Δ32 could be dramatically increased from 10^{-5} to today's 10% through continuous selection pressure over 320 years due to this hemorrhagic fever virus. While the age estimate of 700 years coincides with the Black Death, Duncan et al acknowledge that the original CCR5-Δ32 mutation likely arose much earlier. They mention a study where the rates of crossing over between the IRI3.1 microsatellite and CCR5-Δ32 occurred about 3500 years ago, though they suggest the CCR5-Δ32 probably appeared over 2500 years ago. Duncan et al also indicate that sporadic outbreaks of plague before the Black Death could have partly accounted for the increase in the CCR5-Δ32 mutation allele.

Finally, the plague largely disappeared in Europe after 1670; without any selective pressure, the frequency of CCR5-Δ32 would decrease over the next 300 years as a result of genetic drift (Duncan et al, 2005). According to Duncan et al, smallpox and sustained plague in

northeastern Europe could have accounted for its continued high frequency. Since the *Myxoma poxvirus*, a relative of smallpox, uses the CCR5 receptor to enter the white blood cells of rabbits, CCR5-Δ32 would likely have provided at least partial resistance to smallpox in the 17th and 18th centuries, maintaining selective pressure almost until the emergence of HIV (Duncan et al, 2005). Furthermore, there is evidence that the hemorrhagic plague proposed by Duncan et al may have persisted in Scandinavia and Russia well after 1670, thus maintaining selective pressures in that region of Europe. This could also explain why the CCR5-Δ32 allele is found in higher frequencies around Scandinavia and Russia.

Although Duncan et al's hypothesis appears plausible, it is very difficult to prove. Due the rapid decline of plague in Europe around the 18th century, the unknown virus that caused the plagues of hemorrhagic fever presumably no longer afflicts humans. Furthermore, there is not sufficient proof to support that the plagues ravaging across Europe for centuries was indeed due to this virus and not the bubonic plague. The greatest weakness in Duncan et al's hypothesis, however, is that it is based on speculations about a virus that no longer exists and thus cannot be studied.

Smallpox Hypothesis

Galvani and Slatkin proposed an alternative to the plague hypothesis. They believed that the plague epidemics and outbreaks occurred too intermittently to have applied consistent selective pressure to dramatically increase the frequency of CCR5-Δ32. Galvani and Slatkin state that, while the plagues killed a significant percentage of Europe's population at a given time, the total mortality from the plagues only impacted a small percent of Europe's overall population

across time. Thus, they suggested that smallpox is the selective pressure that caused the dramatic increase of CCR5-Δ32.

Galvani and Slatkin first stated that the direct person-to-person transmission of smallpox provided a far more continuous selective pressure. Smallpox epidemics were frequent and mostly affected children under ten-years-old; it was thus considered a childhood disease (Galvani and Slatkin, 2003; Balanovsky et al, 2005). Furthermore, smallpox had approximately a 30% fatality rate (Galvani and Slatkin, 2003) and since it affected children, the allele frequency changed over generations more effectively (Galvani and Slatkin, 2003; Balanovsky et al, 2005). Therefore, whereas the plague would result in strong but interspersed selection, smallpox results in weaker yet continuous selection.

Galvani and Slatkin's model shows that the age range of an affected population can affect the selection of a resistance allele when the population is afflicted episodically by a disease with a high mortality. A crucial point was that the initial frequency of CCR5-Δ32, p_0, was the same for both the bubonic plague and smallpox. When they assumed the resistance allele was dominant, the 400 years of bubonic plague epidemics did not provide sufficient selective pressure to drive the frequency up to 1%. In contrast, the "more continuous smallpox mortality" on European children could provide the necessary selective pressure in that time. When incomplete dominance was assumed, Galvani and Slatkin calculated that 1135 years of smallpox epidemics would have been required; this is well within the 95% confidence interval of the age estimate of 700 years provided by Stephen et al. Galvani and Slatkin further state that heterozygous fitness was important to how quickly evolution occurs in the case of a resistance allele, as most of the selection is through the heterozygotes when the frequency of the resistance allele is initially very low.

Galvani and Slatkin thus believe the original CCR5-Δ32 mutation originated 1000-2000 years ago in northern Europe and was then dispersed across Europe down a continuous gradient to the south via Viking conquests as proposed by Lucotte in 2001 (Balanovsky et al, 2005). Interestingly, smallpox also affected Scandinavia more than the rest of Europe, which could lead to its higher frequencies of CCR5-Δ32 (Galvani and Slatkin, 2003). And again, smallpox is clinically more likely than the bubonic plague to cause the dramatic increase of CCR5-Δ32 because of the differing modes of infection. Poxviruses enter immune cells via chemokine receptors like the CCR5 receptor, whereas the bubonic plague pathogen does not (Galvani and Slatkin, 2003; Duncan et al, 2005).

But the smallpox hypothesis is not without criticisms. Balanovsky et al argue that the smallpox hypothesis also relies heavily on statistical analyses from historical accounts, and is not based on real clinical or molecular genetic data. Duncan et al mentioned that a lethal variant of smallpox did not arrive until around 1628 and before that date, smallpox was "not reputed to be a serious malady." Thus the continuous selection provided by smallpox could only have been effective between about 1700 and 1830, and according to Duncan et al's modeling, over 600 years of smallpox epidemics would be required to increase the frequency of CCR5-Δ32 to 10%.

Ecological Factors

A third hypothesis looked at ecological factors and what role they may have had in the distribution of the CCR5-Δ32 allele. Balanovsky et al reasoned that some ecological factors, such as climate, could provide selective pressures in differing parts of the world. Balanovsky et al thus compiled a global database on CCR5-Δ32 using data from 51 published sources, totaling 36,436 samples from 264 populations. They also included 3,784 samples from 35 populations in

Russia, Ukraine, and Moldova for the first time. With such a large sample size, Balanovsky et al were able to construct an updated and more detailed map of CCR5-Δ32 frequency distributions. In addition, Balanovsky et al also gathered data on the spatial distribution of atmospheric pressure, temperature, and precipitation for the first half of the 20th century before global warming.

Based on this more detailed map of CCR5-Δ32 distributions, Balanovsky et al found the highest frequency of CCR5-Δ32 to be in the areas surrounding the Baltic Sea and subsequently decreasing gradually in all directions away from that region. Additional modeling modified the map distribution of CCR5-Δ32 according to the "clinal" hypothesis and the "diffusion from one center" hypothesis. Balanovsky et al found strong correlations between the initial map and the maps modified for these two hypotheses. Additional correlation analysis found the correlation between temperature and CCR5-Δ32 to be higher than all other climatic parameters. Balanovsky et al also further tested the second hypothesis by obtaining correlations between frequencies and the studied populations and their geographic distances from the CCR5-Δ32 origin.

Balanovsky et al admit that their study might be better used to reject, rather than prove, some hypotheses. As they acquired new data on CCR5-Δ32 distribution, Balanovsky et al found that the correlation with climate actually decreased, though it still remained significant. They reason that, while the relationship between CCR5-Δ32 and climate is not straightforward, it may play a part in selection, for example having an effect on plague in colder climates. Furthermore, Balanovsky et al asserted that gene flow, genetic drift, and migration cannot explain the continued high frequencies of CCR5-Δ32 across Europe. While the Viking dispersal across Europe might have initially resulted in the gradient spread of CCR5-Δ32, selection would have been required to maintain the frequencies in the European populations (Balanovsky et al, 2005).

It is difficult to determine how much of an effect, if any, ecological factors had in the selection of CCR5-Δ32. Balanovsky et al provided no mechanism or convincing reason why certain environmental factors would cause the frequency of CCR5-Δ32 to increase over time, or what effect climate had on plague or smallpox. They also did not explain how the frequency of CCR5-Δ32 could have been maintained after the Viking dispersal across Europe. Certainly, the hypothesis proposed by Balanovsky et al is weaker compared to either the plague or smallpox models.

Bronze Age Hypothesis

In a radical break from previous hypotheses asserting a single and recent origin, a study by Sabeti et al found CCR5-Δ32 to be less remarkable as a mutation, whereas a study by Hedrick and Verrelli found CCR5-Δ32 to be far older than previous thought.

Sabeti et al carried out SNP (single nucleotide polymorphism) genotyping around CCR5 in multiple populations. When the data was compared against the human genome and genetic maps, Sabeti et al found CCR5-Δ32 was not unique in terms of genetic diversity relative to other variants at CCR5 or even throughout the human genome. They then assessed the high CCR5-Δ32 frequencies by looking at the heterozygosity about CCR5 to further determine its genetic diversity; if a selection sweep has occurred, diversity would decrease, but if balancing selection occurred, then diversity would increase (Sabeti et al, 2005; Hedrick and Verrelli, 2006). If CCR5-Δ32 was subjected to a selection sweep by positive selection, it would remove linked variants as its frequency increases in a population (Sabeti et al, 2005; Hedrick and Verrelli, 2006). Thus heterozygosity in SNPs near CCR5 should decrease, but when analyzed for recent directional positive selection and multiple selection sweeps at a locus, Sabeti et al did not find

this to be the case. They assert that while these tests found no evidence of positive selection, it cannot be wholly ruled out.

Furthermore, if the dramatic increase in CCR5-Δ32 frequency was a relatively recent event, there should be high linkage disequilibrium near CCR5-Δ32 (Sabeti et al, 2005; Hedrick and Verrelli, 2006). The linkage disequilibrium in regions distal (further away) to CCR5 showed similar frequencies to other CCR5 haplotypes, but regions proximal (near) to CCR5 showed higher disequilibrium (Sabeti et al, 2005; Hedrick and Verrelli, 2006). This may suggest that although the CCR5-Δ32 allele might have increased recently, it might also be much older than previous studies.

By analyzing DNA samples, a study by Hummel et al found similar frequencies of CCR5-Δ32 in Bronze Age skeletons 2900 years ago and in modern Germans (Hedrick and Verrelli, 2006). In addition, Hummel et al did not find reduced CCR5-Δ32 frequencies in a sample of 14th century plague victims, and thus might not have provided any protection against the plague. Using a new genetic map with 32 SNP markers closer to CCR5 increased the estimate age of CCR5-Δ32 to about 5075 years, consistent with the Bronze Age data (Hedrick and Verrelli, 2006). Hedrick and Verrelli assert that the CCR5-Δ32 allele was present long before either plague or smallpox could have exerted positive selection.

The studies by Sabeti et al and Hedrick and Verrelli suggest that CCR5-Δ32 is not as unique a variant as popularly believed and may in fact be older than previous estimates. While these two studies conclude that the original selection for CCR5-Δ32 is older than previously thought, and with a lack of evidence for positive selection since, nothing can be ruled out. Furthermore, the findings by Hummel et al in 14th century plague victims who also possessed the CCR5-Δ32 mutation allele does, however, support Duncan et al's modified plague hypothesis or

Galvani and Slatkin's smallpox hypothesis in that CCR5-Δ32 probably conferred no protection against the bubonic plague. More than anything, the data from these two studies present questions rather than answers.

Discussion and Conclusion

Each of the five studies addresses different hypotheses for the origin of CCR5-Δ32 and its continued high frequencies in European populations. All the studies agree that the bubonic plague is probably not the source of selective pressure for CCR5-Δ32. In refutation of this widely accepted hypothesis, each study attempts to create a new hypothesis and model that better explains the high frequency of CCR5-Δ32 as well as its age estimate.

All of the five studies have some compelling evidence to support their hypothesis, but can a synthesis between all five unite them into a single hypothesis? It certainly may be possible that each of the hypotheses represents only a piece of the whole. Both Sabeti et al and Hedrick and Verrelli state that nothing can be conclusively ruled out given CCR5-Δ32's unique nature, despite its lack of uniqueness in terms of genetic diversity. Also, as CCR5-Δ32 is a null allele that causes a loss in gene function, it would normally be subjected to purifying selection and removed from the population (Hedrick and Verrelli, 2006). CCR5-Δ32 is, however, a unique null allele due to its selective advantage against HIV-1 although Sabeti et al found no evidence of positive selection before the emergence of HIV.

Assuming CCR5-Δ32 is about 5075 years old as reported by Hedrick and Verrelli, the original selection for CCR5-Δ32 predates both the plague and smallpox. Duncan et al mentioned examples of several episodes of plague epidemics prior to the Black Death. Assuming the pathogen of these plague epidemics up through the 1700s is a virus that causes hemorrhagic

fever with 100% fatality, as suggested by Duncan et al, it is plausible that this plague may have provided the original selective pressure. Selection could have been further affected by ecological factors such as a cold climate that may have altered the selection from plague epidemics, as suggested by Balanovsky et al. In addition, if CCR5-Δ32 originated in the Baltic Sea area, it could very well have been dispersed by Vikings, providing the initial north-sound gradient.

The intersection of the emergence of smallpox with the disappearance of the plague in Europe might have provided the required selective pressure to maintain CCR5-Δ32 at high frequencies until the 20th century. The evidence cited by Duncan et al and Galvani and Slatkin, in which plague and smallpox persisted longer in northern European regions, could provide a means by which the north-sound gradient in CCR5-Δ32 frequencies could have been sustained well after the Viking conquests.

References

Balanovsky O, Pocheshkhova E, Pshenichnov A, Solovieva D, Kuznetsova M, Voronko O, Churnosov M, Tegako O, Atramentova L, Lavryashina M, Eseeva I, Borinska S, Boldyreva M, Dubova N, Balanovska E. 2005. Is Spatial Distribution of the HIV-1-resistant CCR5Δ32 Allele Formed by Ecological Factors? *Journal of Physiological Anthropology and Applied Human Science* 24(4):375-82.

Duncan SR, Scott S, Duncan J. 2005. Reappraisal of the historical selective pressures for the CCR5-Δ32 mutation. *Journal of Medical Genetics* 42(3):205-8.

Galvani AP, Slatkin M. 2003. Evaluating plague and smallpox as historical selective pressures for the CCR5-Δ32 HIV-resistance allele. *Proc National Academy of Science* 100(25):15276-9.

Hedrick PW, Verrelli VC. 2006. "Ground truth" for selection on CCR5-Δ32. *TRENDS in Genetics* 22(6):293-6.

Sabeti PC, Walsh E, Schaffner SF, Varilly P, Fry B, Hutchenson HB, Cullen M, Mikkelsen TS, Roy J, Patterson N, Cooper R, Reich D, Altshuler D, O'Brien S, Lander ES. 2005. The Case for Selection at CCR5-Δ32. *PLoS Biology* 3(11):e378.