

REVIEW

Distribution of HLA haplotypes across Japanese Archipelago: similarity, difference and admixture

Hirofumi Nakaoka and Ituro Inoue

The human leukocyte antigen (HLA) region is the most polymorphic region in the human genome. The polymorphic nature of the HLA region is thought to have been shaped from balancing selection. The complex migration events during the Out-of-Africa expansion have influenced geographic patterns of HLA allele frequencies and diversities across present-day human populations. Differences in the HLA allele frequency may contribute geographic differences in the susceptibility to many diseases, such as infectious, autoimmune and metabolic diseases. Here we briefly reviewed characteristics of frequency distribution of HLA alleles and haplotypes in Japanese population. A large part of HLA alleles and haplotypes that are common in Japanese are shared with neighboring Asian populations. The differentiations in HLA alleles and haplotypes across Japanese regional populations may provide clues to model for peopling of Japanese Archipelago and for design of genetic association studies. Finally, we introduce recent topics that new HLA alleles derived from ancient admixtures with Neanderthals and Denisovans are thought to have played an important role in the adaptation of modern humans to local pathogens during Out-of-Africa expansion.

Journal of Human Genetics (2015) 60, 683–690; doi:10.1038/jhg.2015.90; published online 23 July 2015

INTRODUCTION

Genetic and fossil evidence supports the ‘Out-of-Africa’ model, in which modern humans evolved in Africa and subsequently migrated through Eurasia to reach the Pacific and Americas.¹ The expansion is thought to have begun approximately 45 000–60 000 years ago.² The signature of a series of colonization events during this expansion persists as a continuous loss of genetic diversity with the increasing distance from the ancestral population known as ‘serial founder effect’.^{3–5} When humans encountered new environments, including distinct pathogens, the prevalence of advantageous alleles under the corresponding circumstances could have increased through natural selection. There are several lines of evidence that admixture events occurred between modern humans and archaic hominins.^{6,7} Such events along with random genetic drift have influenced geographic patterns of genetic variation in present-day human populations, which may contribute to geographical differences in the susceptibility to infectious diseases and common diseases, such as autoimmune and metabolic diseases.

The human leukocyte antigen (HLA) region is the human equivalent of the major histocompatibility complex, which spans approximately 3.6 mega bases on the short arm of chromosome 6. The HLA region contains many genes involved in immune function. Genetic variation in the HLA region is associated with many diseases, including autoimmune and infectious diseases.⁸ Recently, several HLA alleles have been reported to be associated with severe and fatal drug hypersensitivity reactions.^{9–11} Additionally, the HLA region is characterized as the most polymorphic region in the human genome.⁸ More than 13 000 HLA alleles have been deposited in the IMGT/HLA

Database.¹² The frequency distributions of HLA alleles and haplotypes have been used to track human evolutionary processes, such as migration, admixture and selection.¹³

Here we review characteristics of frequency distribution of HLA alleles and haplotypes in Japanese population. First, we address the property in HLA frequencies of Japanese population compared with other populations, which shows that Japanese population shares a large proportion of HLA alleles and haplotypes with East and South East Asian populations but has some unique properties. Second, we consider genetic differentiation in HLA frequencies across Japanese regional populations, which may provide clues to model for peopling of Japanese Archipelago and to design genetic association studies. Finally, we would like to introduce recent topics that new HLA variants derived from ancient admixtures with Neanderthals and Denisovans played an important role in the adaptation of humans to local pathogens during the Out-of-Africa expansion.

CHARACTERISTICS OF HLA ALLELES AND HAPLOTYPES IN JAPANESE POPULATION COMPARED WITH OTHER POPULATIONS

The polymorphic nature of the HLA region is thought to have been shaped from balancing selection and maintained by overdominance effect, in which heterozygotes are assumed to have greater fitness than homozygotes because of their capability of recognizing a wide range of pathogens.^{14,15} This pathogen-driven balancing selection model has been supported by the fact that the heterozygosity of HLA loci in a population is associated with the pathogen richness at the region where the corresponding population resides.^{16,17} At the same time, the

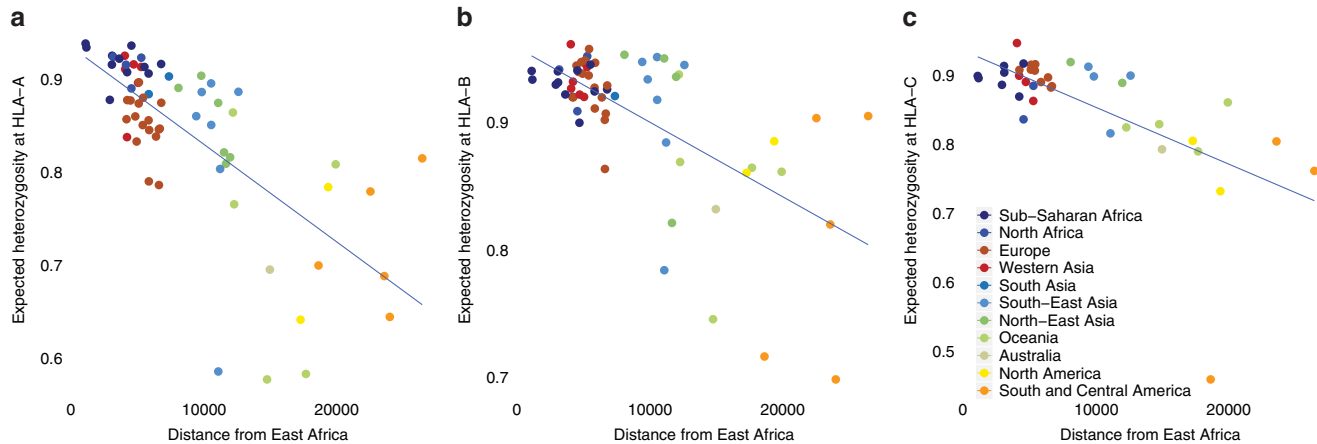


Figure 1 A continuous loss of HLA diversity with the increasing distance from East Africa. (a) *HLA-A* data from 66 countries, (b) *HLA-B* data from 64 countries and (c) *HLA-C* data from 38 countries. We retrieved one population with largest sample size and allele frequency data at four-digit level when there were multiple populations in a country in the AFND.²⁵ *y* axis corresponds to the expected heterozygosity of each population at the corresponding HLA locus calculated by the following equation: $\hat{H} = \frac{2n}{2n-1} (1 - \sum_i p_i^2)$, where *n* is the sample size and *p_i* is the allele frequency for *i*th allele in the population. *x* axis corresponds to the geographic distance (km) between the resident of a population and Addis Abeba. The geographic distances were calculated according to Ramachandran *et al.*³ by using five obligatory waypoints (Cairo, Egypt; Istanbul, Turkey; Anadyr, Russia; Phnom Pehn, Cambodia; and Prince Rupert, Canada).

heterozygosity at each HLA locus was demonstrated to be strongly associated with the distance from East Africa, reinforcing the presence of serial founder effect on HLA diversity (Figure 1).^{16,17} The serial founder model assumes that a small part of the population founded colonies nearby before the original population reached demographic equilibrium, which could have created allele frequency difference between the original and migrated populations. As the HLA region is characterized by strong linkage disequilibrium (LD),⁸ long haplotypes encompassing several HLA loci have been preserved in present-day populations. Therefore, the sharing of HLA haplotypes among populations can be used for inferring human migration routes.¹³

We will discuss characteristics of frequency distribution of HLA alleles and haplotypes in Japanese population. The frequencies of HLA alleles and haplotypes in Japanese population have been intensively analyzed^{18–21} to fulfill the demands of HLA matching between donors and recipients in hematopoietic stem cell transplantation^{22,23} and pharmacogenomics studies.²⁴ The Allele Frequency Net Database (AFND) provides the frequencies of HLA alleles and haplotypes from diverse populations.²⁵ In this review, we retrieve the frequency information mainly from the AFND database for the inference.

HLA alleles

Compared with the other HLA loci with many alleles, *HLA-A* has a unique property in which only a few alleles show predominantly high frequencies in Asian, Native American and European populations.²⁶ In Japanese population, the allele frequency of *HLA-A*24:02* reaches about 35%, indicating that approximately 60% of the Japanese people carry this allele.^{18–21} *A*24:02* distributes all over the world. This allele dominates especially in Taiwan's aborigines (~86.3%), Oceanian (~74.4%) and Native American populations (~61.4%). The allele frequencies of *HLA-A*02:01*, *A*02:06*, *A*11:01*, *A*26:01*, *A*31:01* and *A*33:03* have been reported to be >5% in Japanese populations.^{18–21} All of these *HLA-A* alleles are common in East Asian populations, including Chinese and Korean populations. *A*02:06* is common in Asian, Oceanian and Native American populations but of very low frequent in African and European populations. *A*26:01* is characterized by the high frequency in Japanese Ainu people who are descendants of indigenous Japanese.²⁷

The frequency distributions of *HLA-B*, *-C*, *-DRB1* and *-DQB1* alleles are relatively even, in which the frequencies of several representative alleles are fairly balanced. *B*07:02*, *B*15:01*, *B*35:01*, *B*40:01*, *B*40:02*, *B*44:03*, *B*46:01*, *B*51:01*, *B*52:01* and *B*54:01* are distributed evenly, with their frequencies ranging from 5% to 10% in Japanese population.^{18–21} At *HLA-C* locus, nine alleles represent frequencies >5% (*C*01:02*, *C*03:03*, *C*03:04*, *C*04:01*, *C*07:02*, *C*08:01*, *C*12:02*, *C*14:02* and *C*14:03*).^{18–21} *HLA-DRB1*01:01*, *DRB1*04:05*, *DRB1*08:03*, *DRB1*09:01*, *DRB1*13:02*, *DRB1*15:01* and *DRB1*15:02* are common in Japanese (5–15%).^{18–21,28} The common *DQB1* alleles whose frequencies are >5% are *DQB1*03:01*, *DQB1*03:02*, *DQB1*03:03*, *DQB1*04:01*, *DQB1*05:01*, *DQB1*06:01*, *DQB1*06:02* and *DQB1*06:04*.^{18,19,29} On the other hand, two alleles show predominantly high frequencies at *DPB1* locus (*DPB1*05:01*, 35%; *DPB1*02:01*, 25%) followed by several common alleles (>5%; *DPB1*04:01*, *DPB1*04:02* and *DPB1*09:01*).^{18–20}

*B*46:01* and *B*54:01* are prevalent only in Asian populations. The frequency of *DQB1*04:01* in Japanese is the highest in all the populations in the AFND database, which distributes from Siberia to Oceania.²⁵ *C*14:03* and *DPB1*09:01* are observed in the highest frequency in Japanese population. *C*14:03* is common in Korean population but of low frequent in Chinese and Taiwanese populations. Several ethnic groups in Cameroon, Kurdish people in Georgia and Khandesh people in India commonly harbor *C*14:03*.

There are many HLA alleles that are common in other populations but rare in Japanese.²⁰ For example, the following alleles are common in European populations but rare in Japanese: *A*01:01*, *A*03:01*, *B*08:01*, *B*18:01*, *B*44:02*, *C*02:02*, *C*05:01*, *C*12:03*, *DRB1*03:01*, *DRB1*07:01*, *DRB1*08:01*, *DRB1*11:04*, *DRB1*13:01*, and *DRB1*16:01*.

HLA haplotypes

The extent of LD differs according to the pair of HLA loci. Figure 2 shows pairwise LD between *HLA-A*, *-C*, *-B*, *-DRB1* and *DPB1* in Japanese. The extent of LD was measured by means of normalized global disequilibrium *W_n*, which was a measure of the total disequilibrium between two loci calculated by summing the deviations contributed by each of the individual haplotypes in a multiallelic

system.^{30,31} As expected from their close proximity, the alleles of *HLA-C* and *HLA-B* are in strong LD. As the recombination hot spots exist within the major histocompatibility complex especially between *DRB1* and *DPB1*,^{32,33} the alleles of *DPB1* are in weak LD with the alleles of the other loci. Though our data did not include *DQB1*, the previous report has demonstrated that the alleles of *DRB1* and *DQB1* exhibited strong LDs.¹⁸ According to HLA typing data for many families, the estimated recombination probabilities for *A-C*, *B-DRB1* and *A-DRB1* are 0.54, 0.54 and 1.08%, respectively.²¹

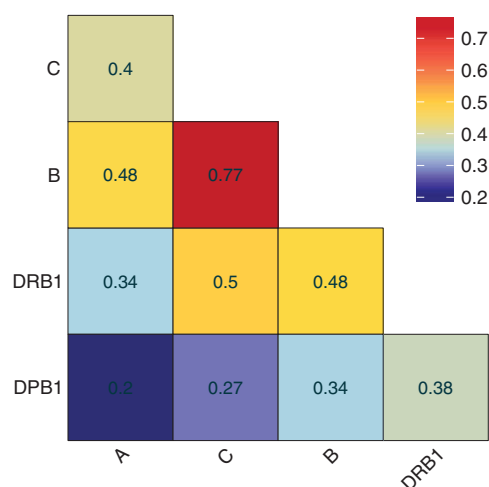


Figure 2 Extent of LD for each pair of HLA locus in Japanese population. LD was measured by means of normalized global disequilibrium *W_n*, which was a measure of the total disequilibrium between two loci calculated by summing the deviations contributed by each of the individual haplotypes in a multiallelic system. HLA genotyping data from mainland Japanese individuals analyzed in our previous study was used.²⁰

The four-locus HLA haplotypes (*HLA-A*, *-C*, *-B* and *-DRB1*) with the frequency >0.5% in Japanese population are represented in Table 1.²⁰ These haplotypes explained only 32.7% of chromosomes in mainland groups, indicating that most of Japanese individuals harbor short, segmented haplotypes. By scrutinizing the haplotype frequency data from the National Marrow Donor Program (NMDP) for about 3 000 000 subjects, including many Asian groups in the United States,³⁴ a substantial number of the HLA haplotypes that are common in Japanese are shared with Korean, Chinese Vietnamese and Filipino populations (Table 1). This is consistent with the previous report comparing Japanese and Korean populations.³⁵ Among them, only two haplotypes *A*26:01-C*03:04-B*40:02-DRB1*09:01* and *A*33:03-C*14:03-B*44:03-DRB1*08:03* are virtually absent in the other Asian populations from the NMDP.³⁴ As similar haplotypes containing parts of these two haplotypes are observed in Asian populations; therefore, it seems that these two haplotypes were generated by recombination events and preserved in Japanese population. We will discuss about the spatial distributions of some of these HLA haplotypes by linking with the consideration of differences among Japanese regional populations in the next section.

DIFFERENCES AMONG JAPANESE REGIONAL POPULATIONS ON THE BASIS OF HLA ALLELES

Genetic differentiation across Japanese Archipelago has been analyzed in several gene systems, such as mitochondrial DNA,^{36,37} Y chromosome³⁷⁻⁴⁰ and genome-wide single nucleotide polymorphisms.⁴¹⁻⁴³ Most of these researches were conducted to propose the model for peopling of Japanese Archipelago.

The origin of present-day Japanese has long been debated. It is thought that there were at least two waves of migrations to the Japanese Archipelago. The ancestors of the Jomon people migrated to Japanese Archipelago in the Upper Paleolithic age (approximately 30 000 years ago). The new migrants, the Yayoi people, came through

Table 1 The most common four-locus HLA haplotypes in mainland Japanese

Locus	Japanese ²⁰				National Marrow Donor Program ³⁴					
	A	C	B	DRB1	Mainland	Okinawa	Korean	Chinese	Vietnamese	Filipino
24:02	12:02	52:01	15:02	8.04	1.83	1.87				
33:03	14:03	44:03	13:02	5.01	0.46	3.97	0.25			
24:02	07:02	07:02	01:01	4.03	0.46	2.26				
24:02	01:02	54:01	04:05	3.01	7.34	1.19	0.48	0.37		
02:07	01:02	46:01	08:03	2.19	0.92	1.33	0.89	0.25		0.15
11:01	04:01	15:01	04:06	1.32		1.64	0.62			
11:01	01:02	54:01	04:05	0.90		0.44	0.37	0.55		
31:01	14:02	51:01	08:02	0.82		0.48	0.28			
24:02	01:02	59:01	04:05	0.79	2.75	0.75				
26:02	08:01	40:06	09:01	0.76		0.30				
24:02	08:01	40:06	09:01	0.76		0.47	0.21	0.14		
24:02	01:02	46:01	08:03	0.69		0.45	0.15			
02:06	08:01	40:06	09:01	0.66		0.25				
24:02	03:04	40:02	09:01	0.58	0.46	0.18				
26:01	03:04	40:02	09:01	0.53	0.46					
24:02	03:04	40:01	04:05	0.53			0.16			0.17
02:01	03:03	15:11	09:01	0.53	0.46	0.26	0.28	0.16		
01:01	06:02	37:01	10:01	0.53		0.92	0.31	0.37		
33:03	14:03	44:03	08:03	0.50						
26:03	03:03	15:01	15:01	0.50	0.92	0.33				

Abbreviation: HLA, human leukocyte antigen. HLA haplotype frequencies from Japanese individuals from the mainland and Okinawa groups analyzed in our previous study are represented.²⁰

the Korean Peninsula in the Aeneolithic period (1000 BC to 300 AD). The most prevailing model for peopling of Japan is the admixture model or 'dual structure model', in which modern Japanese was formed by admixture between the Jomon and Yayoi people.⁴⁴ The degree of admixture is thought to vary across the archipelago, which may influence genetic structure of the present-day Japanese population.⁴⁰ Here we introduce the results from researches on genetic differentiation across Japanese regional populations based on HLA alleles and haplotypes.

The phylogenetic studies using the frequencies of HLA alleles demonstrate that Okinawa and Ainu people who are thought to be descendants of indigenous Japanese have close affinity each other, and mainland Japanese is located in the middle of Okinawa and Ainu people and East Asian populations, including Korean population.^{27,45} The finding that the haplotype *A*24-B*54-DRB1*04:05* that are common in South East Asian populations are prevalent in Okinawa but virtually absent in Ainu people suggests that there was gene flow from South East Asia to Okinawa after differentiation of the ancestors of these two groups.⁴⁵ Additionally, it has been reported that Nivkh people in Sakhalin contributed to genetic constitution of Ainu people.^{37,43} These factors may influence the genetic distance between Okinawa and Ainu people.⁴⁵ Tokunaga *et al.*⁴⁶ pointed out genetic links between Ainu people and Native Americans.

We examined the differences in HLA frequencies across 10 regional populations (Hokkaido, Tohoku, Kanto-Koshin, Hokuriku, Tokai, Kinki, Chugoku, Shikoku, Kyushu and Okinawa).²⁰ Although genetic differentiation between Okinawa and mainland groups was notable, differences across mainland groups were also detected. We identified HLA alleles that were associated with the population structure of Japanese by means of the principal component scores (Figure 3). The identified HLA alleles were classified into four clusters according to the coordinates in the principal component score plot. The alleles of each

cluster exhibited similar frequency distributions across regions (Figure 3). Interestingly, the alleles in these clusters reside on the same haplotypes. By examining the extent of LD between pairs of HLA alleles on the haplotypes, we found that the alleles in the two clusters characterized by low frequency in Okinawa exhibited stronger LDs and were preserved as long haplotypes (*A*24:02-C*12:02-B*52:01-DRB1*15:02-DPBI*09:01* and *A*33:03-C*14:03-B*44:03-DRB1*13:02-DPBI*04:01*) (Figure 4). The alleles whose frequencies were high in Okinawa, Shikoku and Kyushu showed intermediate levels of LD and formed haplotype *C*01:02-B*54:01-DRB1*04:05-DPBI*05:01* (Figure 4). The LDs between the alleles whose frequencies were highest in Okinawa were much weaker (Figure 4). Interestingly, some of the alleles were frequent (*A*02:06*, 20.0%; *B*35:01*, 11.0%) in Ainu people.²⁷ The haplotype *A*02:06-B*35:01* was frequent in the Yupik people in Alaska (2.9%).⁴⁷

The long four-locus HLA haplotypes with >0.5% frequencies make up only 32.7% of chromosomes in mainland Japanese (Table 1), implying that the decay of LD generated short, segmented haplotypes during a long period of isolation of the Japanese population. The haplotypes whose constituent alleles were in strong LD (*A*24:02-C*12:02-B*52:01-DRB1*15:02* and *A*33:03-C*14:03-B*44:03-DRB1*13:02*) were commonly found in Japanese and Korean. *A*24-B*52-DR15* that was the serological equivalent encoded by *A*24:02-B*52:01-DRB1*15:02* was shared with Mongolians.¹³ It is plausible that if a haplotype is derived and goes through rapid expansion, its constituent alleles will show strong LD.⁴⁸⁻⁵⁰ Therefore, it is suggested that these haplotypes had been generated in North East Asia and the Korean Peninsula and then moved into Japan's mainland followed by the rapid expansion probably at the Yayoi period. The haplotype *C*01:02-B*54:01-DRB1*04:05* was in the highest frequency in Okinawa and shared by South East Asian populations. The LDs for pairs of the constituent alleles were in the intermediate level (that is,

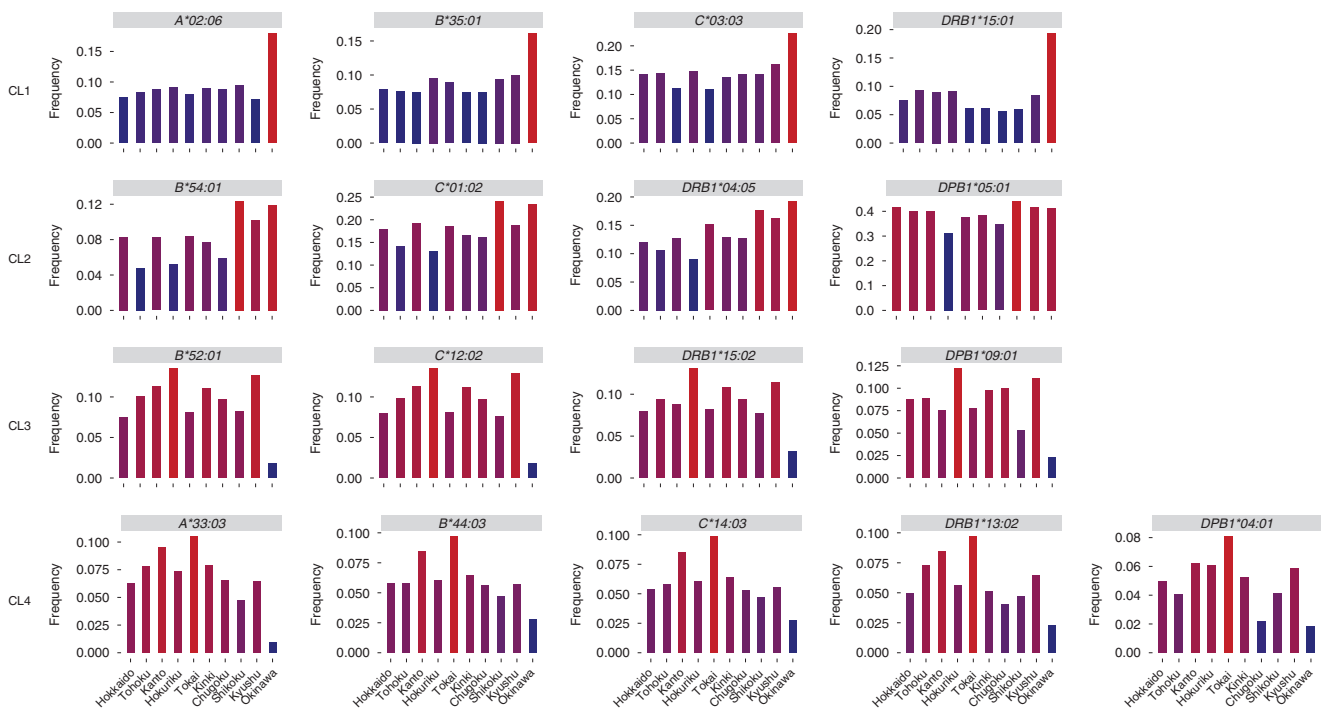


Figure 3 Frequency distribution of HLA alleles associated with population structure of Japanese population. HLA alleles in the same row are classified into clusters CL1, CL2, CL3 and CL4. The figure is reconstructed from our previous study.²⁰

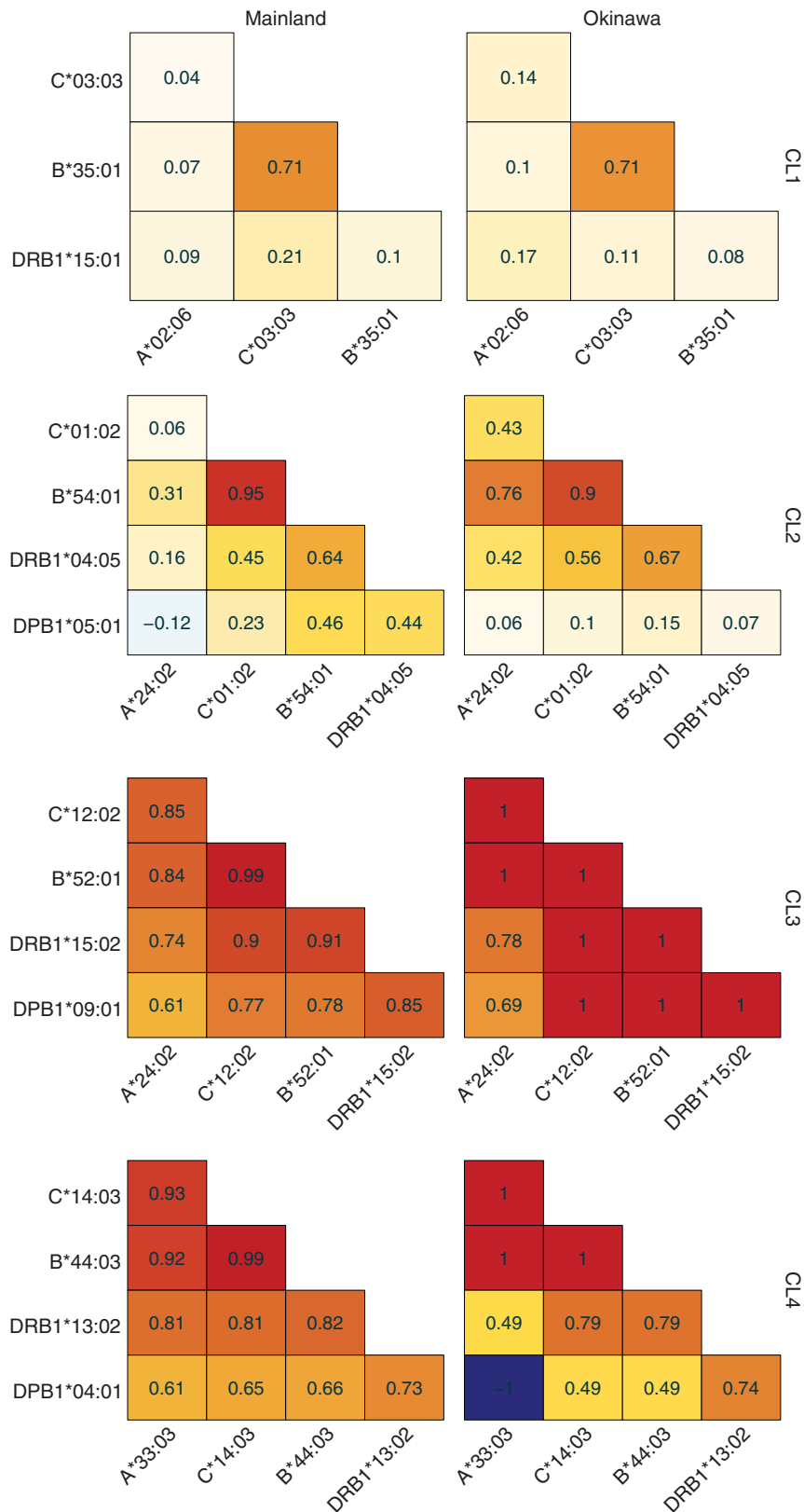


Figure 4 Extent of linkage disequilibrium (*D*) between pairs of HLA alleles in the same cluster represented in Figure 3.

decayed compared with the abovementioned two haplotypes). The property of *C*01:02-B*54:01-DRB1*04:05* can be interpreted as the signature of the gene flow from South East Asia in advance of that

from the Korean Peninsula. These results may support the admixed model for the peopling of Japan by at least three waves of migration events.

NEW HLA ALLELES ACQUIRED BY ADMIXTURES WITH ANCIENT HOMININS

The next-generation sequencing technology enabled to determine whole-genome sequences for degraded DNAs from ancient specimens, which settled a long controversy about interbreeding between modern humans and ancient hominins, such as Neanderthals and Denisovans.^{6,7,51,52} Neanderthals contributed 1.5–2.1% of Eurasian genomes.⁵² Denisovans-derived DNA was estimated to be 4–6% of New Geneans and Aboriginal Australians but was of no significance in mainland Asia.⁵¹ At the same time, strong signatures of gene flows from Neanderthals to Denisovans was detected in the HLA region.⁵²

Abi-Rached *et al.*⁵³ explored for signatures of archaic admixture in HLA class I genes. The authors reconstructed HLA allotypes for Neanderthals and Denisovan by reanalyzing whole-genome sequence data.^{6,7} The Denisova's allotypes were predicted to be identical or very close to *A*02:01/03/07/48*, *A*11:01/53*, *B*15:58*, *B*35:63*, *C*12:02* and *C*15:02/05/17* in modern humans. As the HLA-B allotypes were virtually absent in modern humans, they focused attention on the distributions of all the plausible HLA-A-C haplotypes. These HLA-A-C haplotypes were present in modern Asia and Oceania, rare in Europe but virtually absent in Africa. Among the constituent alleles, *A*11:01*, *C*12:02* and *C*15* are prevalent in Asia and Oceania. Therefore, the authors concluded that modern humans acquired these alleles through admixture with Denisovans. Based on the fact that an exceptionally divergent HLA allele, *B*73:01*, was in strong LD with *C*15:05* in present-day humans, the authors speculated that *B*73:01* was also derived from Denisovans.

Similarly, the haplotypes containing *B*07*, *B*51*, *C*07:02* and *C*16:02* were predicted to be derived from the admixture with Neanderthals. Additionally, some HLA-A alleles exhibiting notable LD decays (*A*11*, *A*26*, *A*02:01*, *A*02:06*, *A*24:02* and *A*31:01*) were predicted to be archaic alleles. As a consequence, the estimated contributions of the putative archaic HLA-A alleles reached >50% in Europeans, >70% in Asians and >95% in Papuans. According to the fact that these putative archaic alleles encode strong ligands for natural killer receptors, the authors concluded that the increasing frequencies of these alleles in Eurasia could be interpreted as adaptive introgressions.

Motivated by the fact that the amino-acid sequence motif in HLA-DPβ was shared between *DPB1*04:01* and Neanderthal, Temme *et al.*⁵⁴ proposed that *DPB1*04:01* was the signature of an adaptive introgression from Neanderthal. However, the phylogenetic analysis showed that *DPB1*04:01* coalesced with *DPB1*04:02* and the other modern human haplotypes before coalescing with Neanderthal haplotype.⁵⁵ Furthermore, *DPB1*04:01* was common in sub-Saharan Africans though its frequency was much less compared with Europeans.⁵⁵ These suggest that the sharing between *DPB1*04:01* and Neanderthal sequence is probably a consequence of long-standing balancing selection or interbreeding before the divergence of modern humans and Neanderthals. Several studies demonstrated that exceptionally divergent haplotypes in immunity-related genes *STAT2* and *OAS1* were derived from Neanderthals and Denisovans as a consequence of adaptive introgressions.^{56–58}

The findings suggesting that modern humans acquired immune systems from archaic hominins that had been adapted to local environments are very intriguing. The abovementioned HLA-A alleles are very common in Japanese population. The contributions of all the predicted archaic HLA alleles in Japanese are shown in Table 2, indicating very high archaic contributions to present-day Japanese on HLA variation. The proportions are much higher than those to Eurasians averaged over whole genome (1.5–2.1%).^{51,52} Japanese

Table 2 Contributions of the predicted archaic HLA alleles in Japanese population

Ancestry	Allele	Mainland	Okinawa
Denisovan	<i>A*11:01</i>	8.78	5.96
	<i>C*12:02</i>	10.44	1.83
	<i>C*15</i>	3.16	4.59
Neanderthal	<i>B*07</i>	5.30	2.29
	<i>B*51</i>	8.44	7.34
	<i>C*07:02</i>	12.37	13.30
Putative ancient	<i>C*16:02</i>	0.00	0.00
	<i>A*02:01</i>	11.18	17.43
	<i>A*02:06</i>	8.33	17.89
	<i>A*24:02</i>	36.45	34.86
	<i>A*26:01</i>	7.36	10.09
	<i>A*31:01</i>	9.23	4.13

Abbreviation: HLA, human leukocyte antigen. HLA allele frequencies from Japanese individuals from mainland and Okinawa groups analyzed in our previous study are represented.²⁰

Archipelago are thought to be far from the putative places where the archaic admixture events occurred; therefore, serial founder effect rather than positive selection could have increased the frequencies of these alleles. Considerations about the predicted ancient adaptive introgression is that the frequency of putative Denisovan-derived *C*12:02* is about 10% in present-day mainland Japanese; however, HLA-B alleles carried by Denisovan, which are likely to be in LD with HLA-C locus, are virtually absent. Additionally, genome-wide contribution of Denisovan was estimated to be of no significance in Japanese.⁵¹ It is possible that some of these alleles had been preserved by long-standing balancing selections.⁵⁹ More studies on selective advantages of the archaic HLA alleles are needed.

CONCLUSION

The complex migration events after the Out-of-Africa expansion have shaped spatial gradients of HLA allele frequencies and diversities. The sharing of HLA haplotypes among populations can be used a powerful tool to reconstruct past migration routes. However, it is notable that only a limited proportion of the overall variability in a population can be explained by common long haplotypes. In Japanese, the four-locus haplotypes whose frequencies are <0.5% account for about 70% of chromosomes. The short, segmented haplotypes due to LD decay may provide useful information for tracing back to much older ancestral relationships between populations. Indeed, a segmented haplotype *C*03:04-B*40:02* that is common in Japanese (mainland, 6.30%; Okinawa, 8.72%) seems to be one of the genetic footprints of the migration route of prehistoric ancient population from Asia to the New World.^{13,20}

Knowledge about population structure of Japanese is essential for the identification of genotype–phenotype correlations. A substantial level of population stratification in Japanese population, especially between Okinawa and other mainland groups, suggests a need for careful consideration on design of HLA association studies. The study design, where case–control samples are stratified into two groups (mainland and Okinawa), association studies are independently conducted and then meta-analysis is performed to integrate the results from the two groups, may be useful to avoid false positive findings owing to population stratification.^{20,41,60}

The improvement of sequencing technologies^{61–63} and the large-scale genotyping projects³⁴ have accelerated the accumulation of HLA frequency data. However, several aspects of HLA frequency data differ

according to populations in terms of genotyped genes, resolution and sample size. Current sequencing technologies allow to determine complete sequences of each HLA gene.⁶³ The high-resolution HLA sequencing will deliver a lot more value to several scientific fields, including medical genomics and population genetics.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

H.N. is supported by grant from Suzuken Memorial Foundation.

- 1 Ingman, M., Kaessmann, H., Pääbo, S. & Gyllensten, U. Mitochondrial genome variation and the origin of modern humans. *Nature* **408**, 708–713 (2000).
- 2 Henn, B. M., Cavalli-Sforza, L. L. & Feldman, M. W. The great human expansion. *Proc. Natl Acad. Sci. USA* **109**, 17758–17764 (2012).
- 3 Ramachandran, S., Deshpande, O., Roseman, C. C., Rosenberg, N. A., Feldman, M. W. & Cavalli-Sforza, L. L. Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. *Proc. Natl Acad. Sci. USA* **102**, 15942–15947 (2005).
- 4 Li, J. Z., Absher, D. M., Tang, H., Southwick, A. M., Casto, A. M., Ramachandran, S. *et al.* Worldwide human relationships inferred from genome-wide patterns of variation. *Science* **319**, 1100–1104 (2008).
- 5 DeGiorgio, M., Jakobsson, M. & Rosenberg, N. A. Out of Africa: modern human origins special feature: explaining worldwide patterns of human genetic variation using a coalescent-based serial founder model of migration outward from Africa. *Proc. Natl Acad. Sci. USA* **106**, 16057–16062 (2009).
- 6 Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M. *et al.* A draft sequence of the Neandertal genome. *Science* **328**, 710–722 (2010).
- 7 Reich, D., Green, R. E., Kircher, M., Krause, J., Patterson, N., Durand, E. Y. *et al.* Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* **468**, 1053–1060 (2010).
- 8 Shiina, T., Hosomichi, K., Inoko, H. & Kulski, J. K. The HLA genomic loci map: expression, interaction, diversity and disease. *J. Hum. Genet.* **54**, 15–39 (2009).
- 9 Mallal, S., Nolan, D., Witt, C., Masel, G., Martin, A. M., Moore, C. *et al.* Association between presence of HLA-B*5701, HLA-DR7, and HLA-DQ3 and hypersensitivity to HIV-1 reverse-transcriptase inhibitor abacavir. *Lancet* **359**, 727–732 (2002).
- 10 Chung, W. H., Hung, S. I., Hong, H. S., Hsieh, M. S., Yang, L. C., Ho, H. C. *et al.* Medical genetics: a marker for Stevens-Johnson syndrome. *Nature* **428**, 486 (2004).
- 11 Hung, S. I., Chung, W. H., Liou, L. B., Chu, C. C., Lin, M., Huang, H. P. *et al.* HLA-B*5801 allele as a genetic marker for severe cutaneous adverse reactions caused by allopurinol. *Proc. Natl Acad. Sci. USA* **102**, 4134–4139 (2005).
- 12 Robinson, J., Halliwell, J. A., Hayhurst, J. D., Flicek, P., Parham, P. & Marsh, S. G. The IPD and IMGT/HLA database: allele variant databases. *Nucleic Acids Res.* **43**, D423–D431 (2015).
- 13 Tokunaga, K., Imanishi, T., Takahashi, K. & Juji, T. In: *Prehistoric Mongoloid Dispersals* (eds Akazawa, T. & Szathmari, E. J. E.) 187–197 (Oxford University Press, Oxford, UK, 1996).
- 14 Hughes, A. L. & Nei, M. Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature* **335**, 167–170 (1988).
- 15 Solberg, O. D., Mack, S. J., Lancaster, A. K., Single, R. M., Tsai, Y., Sanchez-Mazas, A. *et al.* Balancing selection and heterogeneity across the classical human leukocyte antigen loci: a meta-analytic review of 497 population studies. *Hum. Immunol.* **69**, 443–464 (2008).
- 16 Prugnolle, F., Manica, A., Charpentier, M., Guégan, J. F., Guernier, V. & Balloux, F. Pathogen-driven selection and worldwide HLA class I diversity. *Curr. Biol.* **15**, 1022–1027 (2005).
- 17 Sanchez-Mazas, A., Lemaître, J. F. & Currat, M. Distinct evolutionary strategies of human leukocyte antigen loci in pathogen-rich environments. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **367**, 830–839 (2012).
- 18 Tokunaga, K., Ishikawa, Y., Ogawa, A., Wang, H., Mitsunaga, S., Moriyama, S. *et al.* Sequence-based association analysis of HLA class I and II alleles in Japanese supports conservation of common haplotypes. *Immunogenetics* **46**, 199–205 (1997).
- 19 Saito, S., Ota, S., Yamada, E., Inoko, H. & Ota, M. Allele frequencies and haplotypic associations defined by allelic DNA typing at HLA class I and class II loci in the Japanese population. *Tissue Antigens* **56**, 522–529 (2000).
- 20 Nakaoka, H., Mitsunaga, S., Hosomichi, K., Shyh-Yuh, L., Sawamoto, T., Fujiwara, T. *et al.* Detection of ancestry informative HLA alleles confirms the admixed origins of Japanese population. *PLoS ONE* **8**, e60793 (2013).
- 21 Ikeda, N., Kojima, H., Nishikawa, M., Hayashi, K., Futagami, T., Tsujino, T. *et al.* Determination of HLA-A, -C, -B, -DRB1 allele and haplotype frequency in Japanese population based on family study. *Tissue Antigens* **85**, 252–259 (2015).
- 22 Sasazuki, T., Juji, T., Morishima, Y., Kinukawa, N., Kashiwabara, H., Inoko, H. *et al.* Effect of matching of class I HLA alleles on clinical outcome after transplantation of hematopoietic stem cells from an unrelated donor. Japan Marrow Donor Program. *N. Engl. J. Med.* **339**, 1177–1185 (1998).
- 23 Morishima, S., Ogawa, S., Matsubara, A., Kawase, T., Nannya, Y., Kashiwase, K. *et al.* Impact of highly conserved HLA haplotype on acute graft-versus-host disease. *Blood* **115**, 4664–4670 (2010).
- 24 Kamitsujii, S., Matsuda, T., Nishimura, K., Endo, S., Wada, C., Watanabe, K. *et al.* Japan PGx Data Science Consortium Database: SNPs and HLA genotype data from 2994 Japanese healthy individuals for pharmacogenomics studies. *J. Hum. Genet.* **60**, 319–326 (2015).
- 25 González-Galarza, F. F., Takeshita, L. Y., Santos, E. J., Kempson, F., Maia, M. H., da Silva, A. L. *et al.* Allele frequency net 2015 update: new features for HLA epitopes, KIR and disease and HLA adverse drug reaction associations. *Nucleic Acids Res.* **43**, D784–D788 (2015).
- 26 Fernandez Vina, M. A., Hollenbach, J. A., Lyke, K. E., Sztein, M. B., Maier, M., Klitz, W. *et al.* Tracking human migrations by the analysis of the distribution of HLA alleles, lineages and haplotypes in closed and open populations. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **367**, 820–829 (2012).
- 27 Bannai, M., Ohashi, J., Harihara, S., Takahashi, Y., Juji, T., Omoto, K. *et al.* Analysis of HLA genes and haplotypes in Ainu (from Hokkaido, northern Japan) supports the premise that they descent from Upper Paleolithic populations of East Asia. *Tissue Antigens* **55**, 128–139 (2000).
- 28 Mitsunaga, S., Oguchi, T., Moriyama, S., Tokunaga, K., Akaza, T., Tadokoro, K. *et al.* Multiplex ARMS-PCR-RFLP method for high-resolution typing of HLA-DRB1. *Eur. J. Immunogenet.* **22**, 371–392 (1995).
- 29 Mitsunaga, S., Oguchi, T., Tokunaga, K., Akaza, T., Tadokoro, K. & Juji, T. High-resolution HLA-DQB1 typing by combination of group-specific amplification and restriction fragment length polymorphism. **42**. *Hum. Immunol.* 307–314 (1995).
- 30 Cohen, J. *Statistical Power Analysis for the Behavioral Sciences* 2nd edn (Lawrence Erlbaum, Hillsdale, NJ, USA, 1988).
- 31 Trachtenberg, E. A., Erlich, H. A., Rickards, O., DeStefano, G. F. & Klitz, W. HLA class II linkage disequilibrium and haplotype evolution in the Cayapa Indians of Ecuador. *Am. J. Hum. Genet.* **57**, 415–424 (1995).
- 32 Jeffreys, A. J., Kauppi, L. & Neumann, R. Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex. *Nat. Genet.* **29**, 217–222 (2001).
- 33 Miretti, M. M., Walsh, E. C., Ke, X., Delgado, M., Griffiths, M., Hunt, S. *et al.* A high-resolution linkage-disequilibrium map of the human major histocompatibility complex and first generation of tag single-nucleotide polymorphisms. *Am. J. Hum. Genet.* **76**, 634–646 (2005).
- 34 Gragert, L., Madbouly, A., Freeman, J. & Mair, M. Six-locus high resolution HLA haplotype frequencies derived from mixed-resolution DNA typing for the entire US donor registry. *Hum. Immunol.* **74**, 1313–1320 (2013).
- 35 Lee, K. W., Oh, D. H., Lee, C. & Yang, S. Y. Allelic and haplotypic diversity of HLA-A, -B, -C, -DRB1, and -DQB1 genes in the Korean population. *Tissue Antigens* **65**, 437–447 (2005).
- 36 Horai, S., Murayama, K., Hayasaka, K., Matsubayashi, S., Hattori, Y., Fuchuroen, G. *et al.* mtDNA polymorphism in East Asian Populations, with special reference to the peopling of Japan. *Am. J. Hum. Genet.* **59**, 579–590 (1996).
- 37 Tajima, A., Hayami, M., Tokunaga, K., Juji, T., Matsuo, M., Marzuki, S. *et al.* Genetic origins of the Ainu inferred from combined DNA analyses of maternal and paternal lineages. *J. Hum. Genet.* **49**, 187–193 (2004).
- 38 Hammer, M. F. & Horai, S. Y chromosomal DNA variation and the peopling of Japan. *Am. J. Hum. Genet.* **56**, 951–962 (1995).
- 39 Tajima, A., Pan, I. H., Fuchuroen, G., Fuchuroen, S., Matsuo, M., Tokunaga, K. *et al.* Three major lineages of Asian Y chromosomes: implications for the peopling of east and southeast Asia. *Hum. Genet.* **110**, 80–88 (2002).
- 40 Hammer, M. F., Karafet, T. M., Park, H., Omoto, K., Harihara, S., Stoneking, M. *et al.* Dual origins of the Japanese: common ground for hunter-gatherer and farmer Y chromosomes. *J. Hum. Genet.* **51**, 47–58 (2006).
- 41 Yamaguchi-Kabata, Y., Nakazono, K., Takahashi, A., Saito, S., Hosono, N., Kubo, M. *et al.* Japanese population structure, based on SNP genotypes from 7003 individuals compared to other ethnic groups: effects on population-based association studies. *Am. J. Hum. Genet.* **83**, 445–456 (2008).
- 42 Yamaguchi-Kabata, Y., Tsunoda, T., Kumasaka, N., Takahashi, A., Hosono, N., Kubo, M. *et al.* Genetic differences in the two main groups of the Japanese population based on autosomal SNPs and haplotypes. *J. Hum. Genet.* **57**, 326–334 (2012).
- 43 Japanese Archipelago Human Population Genetics Consortium, Jinam, T., Nishida, N., Hirai, M., Kawamura, S., Oota, H. *et al.* The history of human populations in the Japanese Archipelago inferred from genome-wide SNP data with a special reference to the Ainu and the Ryukyuan populations. *J. Hum. Genet.* **57**, 787–795 (2012).
- 44 Hanihara, K. Dual structure model for the population history of the Japanese. *Jpn. Rev.* 1–33 (1991).
- 45 Hata, Y., Ohashi, J., Imanishi, T., Kamiyama, H., Iha, M., Simabukuro, T. *et al.* HLA genes and haplotypes in Ryukyuan suggest recent gene flow to the Okinawa Islands. *Hum. Biol.* **71**, 353–365 (1999).
- 46 Tokunaga, K., Ohashi, J., Bannai, M. & Juji, T. Genetic link between Asians and native Americans: evidence from HLA genes and haplotypes. *Hum. Immunol.* **62**, 1001–1008 (2001).
- 47 Leffell, M. S., Fallin, M. D., Erlich, H. A., Fernandez-Vina, M., Hildebrand, W. H., Mack, S. J. *et al.* HLA antigens, alleles and haplotypes among the Yup'ik Alaska natives: report of the ASHI Minority Workshops, Part II. *Hum. Immunol.* **63**, 614–625 (2002).
- 48 Hayes, B. J., Visscher, P. M., McPartlan, H. C. & Goddard, M. E. Novel multilocus measure of linkage disequilibrium to estimate past effective population size. *Genome Res.* **13**, 635–643 (2003).
- 49 Fry, A. E., Trafford, C. J., Kimber, M. A., Chan, M. S., Rockett, K. A. & Kwiatkowski, D. P. Haplotype homozygosity and derived alleles in the human genome. *Am. J. Hum. Genet.* **78**, 1053–1059 (2006).

- 50 Nordborg, M. & Tavaré, S. Linkage disequilibrium: what history has to tell us. *Trends Genet.* **18**, 83–90 (2002).
- 51 Reich, D., Patterson, N., Kircher, M., Delfin, F., Nandineni, M. R., Pugach, I. *et al.* Denisova admixture and the first modern human dispersals into Southeast Asia and Oceania. *Am. J. Hum. Genet.* **89**, 516–528 (2011).
- 52 Prüfer, K., Racimo, F., Patterson, N., Jay, F., Sankararaman, S., Sawyer, S. *et al.* The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505**, 43–49 (2014).
- 53 Abi-Rached, L., Jobin, M. J., Kulkarni, S., McWhinnie, A., Dalva, K., Gragert, L. *et al.* The shaping of modern human immune systems by multiregional admixture with archaic humans. *Science* **334**, 89–94 (2011).
- 54 Temme, S., Zacharias, M., Neumann, J., Wohlfromm, S., König, A., Temme, N. *et al.* A novel family of human leukocyte antigen class II receptors may have its origin in archaic human species. *J. Biol. Chem.* **289**, 639–653 (2014).
- 55 Ding, Q., Hu, Y. & Jin, L. Non-Neanderthal origin of the HLA-DPB1*0401. *J. Biol. Chem.* **289**, 10252 (2014).
- 56 Mendez, F. L., Watkins, J. C. & Hammer, M. F. A haplotype at STAT2 Introgressed from neanderthals and serves as a candidate of positive selection in Papua New Guinea. *Am. J. Hum. Genet.* **91**, 265–274 (2012).
- 57 Mendez, F. L., Watkins, J. C. & Hammer, M. F. Global genetic variation at OAS1 provides evidence of archaic admixture in Melanesian populations. *Mol. Biol. Evol.* **29**, 1513–1520 (2012).
- 58 Mendez, F. L., Watkins, J. C. & Hammer, M. F. Neandertal origin of genetic variation at the cluster of OAS immunity genes. *Mol. Biol. Evol.* **30**, 798–801 (2013).
- 59 Ségurel, L. & Quintana-Murci, L. Preserving immune diversity through ancient inheritance and admixture. *Curr. Opin. Immunol.* **30**, 79–84 (2014).
- 60 Nakaoka, H. & Inoue, I. Meta-analysis of genetic association studies: methodologies, between-study heterogeneity and winner's curse. *J. Hum. Genet.* **54**, 615–623 (2009).
- 61 Wang, C., Krishnakumar, S., Wilhelmy, J., Babrzadeh, F., Stepanyan, L., Su, L. F. *et al.* High-throughput, high-fidelity HLA genotyping with deep sequencing. *Proc. Natl Acad. Sci. USA* **109**, 8676–8681 (2012).
- 62 Shiina, T., Suzuki, S., Ozaki, Y., Taira, H., Kikkawa, E., Shigenari, A. *et al.* Super high resolution for single molecule-sequence-based typing of classical HLA loci at the 8-digit level using next generation sequencers. *Tissue Antigens* **80**, 305–316 (2012).
- 63 Hosomichi, K., Jinam, T. A., Mitsunaga, S., Nakaoka, H. & Inoue, I. Phase-defined complete sequencing of the HLA genes by next-generation sequencing. *BMC Genomics* **14**, 355 (2013).