CSIRO Publishing
www.publish.csiro.au/journals/rs

The Royal Society of Victoria, 135, 34–37, 2023
10.1071/RS23005

GENOME BANKING OF ANCESTRAL HAPLOTYPES FOR FUTURE SURVIVAL

Erwin A. Paz[1,2,3], Lani A. Wade[1,2], Anthony J. Lloyd[1], Sally S. Lloyd[1,2], Roger L. Dawkins[1,2,4]

[1]C.Y. O'Connor ERADE Village Foundation, North Dandalup, Western Australia 6207 Australia
[2]The Marshall Centre for Rural Research, North Dandalup, Western Australia 6207 Australia
[3]Helicobacter Research Laboratory, The Marshall Centre for Infectious Diseases, Research and Training at UWA, University of Western Australia, Australia
[4]Melaleuka Stud, Piara Waters, Western Australia 6112 Australia

Correspondence: Roger L. Dawkins, rldawkins@tmgwa.com.au

ABSTRACT: The human genome contains Polymorphic Frozen Blocks (PFB) dedicated to the maintenance of haplotypes which are responsible for functional and survival differences between individuals. These Conserved Polymorphic Sequences (CPS) are protected from mutation and are retained as 'safe deposits' available in the future. Therefore, the genome has a reserve for use during cyclical challenges such as epidemics, famine and climate changes. In humans, this mechanism has survived since speciation and may help to explain success of humans in surviving diverse environments during this time frame. Less is known of other species. Domestic livestock have been bred to select for certain traits of commercial significance. Here we take the example of red Wagyu (Akaushi) cattle and show that their genome appears to have been restructured. As a consequence, some of their reserves may have been lost. There may be a need for back-crossing to maintain an optimal amount of diversity. We propose the establishment of genome banks to curate and preserve founder sequences containing a full complement of CPS.

Keywords: genome bank, conserved polymorphic sequences, ancestral haplotypes, species diversity, homologous recombination, founder populations

Ceppellini introduced the notion of a 'haplotype' to describe the co-inheritance of polymorphic alleles implying physical linkage of the loci without recombination (Ceppellini et al. 1967; Petersdorf 2017; Tait 2022; Alper et al. 2023). Haplotypes, rather than single genes, are now regarded as the principal unit for the inheritance of polygenic traits (Lloyd et al. 2016; Dawkins 2022).

Conserved Polymorphic Sequences (CPS) of kilobases and even megabases, also known as Extended or Ancestral Haplotypes, are inherited faithfully over many generations and therefore create differences between individuals of each generation (Figure 1).

The conservation of CPS was first shown in the Major Histocompatibility Complex (MHC) in humans. Identical haplotypes of hundreds of kilobases and even megabases were found in apparently unrelated individuals, implying inheritance from a founding population. Frequencies of common haplotypes differ between ethnicities, and there are many ethnic-specific haplotypes, implying formation by recombination within the founder population of that ethnic group (Gaudieri et al. 1997). By contrast, the 57.1 ancestral haplotype is present in multiple ethnic groups implying inheritance from earlier founders existing some 100,000 years ago, before the 'Out of Africa' migration (Kay et al. 1988; Dawkins & Lloyd 2019).

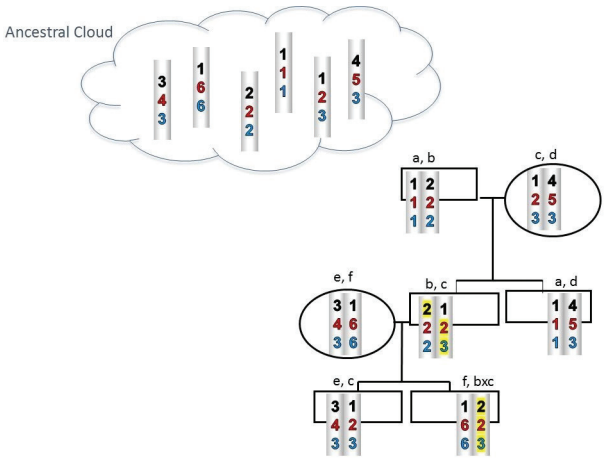

Figure 1: Conserved Polymorphic Sequences (CPS) are inherited unchanged from distant ancestors.
Blocks are conserved because sequence differences prevent recombination. Some reshuffling is possible when a homozygous block is present. An example is highlighted in yellow where the 123 haplotype has recombined with the 222 haplotype because it is homozygous at the second block thereby creating a 223 haplotype.

The partitioning of the genome into Polymorphic Frozen Blocks (PFB) is not restricted to the MHC (Gabriel et al. 2002). In fact, McLure et al. (2005, 2013) found evidence of blocks on many other chromosomes. Based on

multiple studies, we estimate that an average of 10% of the human genome consists of PFB (Gaudieri et al. 1997; Longman-Jacobsen et al. 2003; Kulski et al. 2021).

Miro Radman (2022) provides an explanation of how recombination can occur to form new haplotypes while generally conserving the critical cis-interactions within ancestral haplotypes. The sequence differences and the activity of the mismatch repair (MMR) maintain genome stability by allowing recombination only between homologous sequences (Figures 2,3). Unrestricted recombination would otherwise fragment CPS. Critically, polymorphism can be self-sustaining (Dawkins & Lloyd 2023).

The practical value of haplotyping has been demonstrated by comparing haplotype frequencies among several cattle breeds (Lloyd et al. 2013). For example, it is possible to distinguish Wagyu, and Angus masquerading as Wagyu. More importantly, the haplotypes control important traits such as marbling.

In principle, it should be possible to characterise these sequences which have contributed to survival and retained as safe deposits.

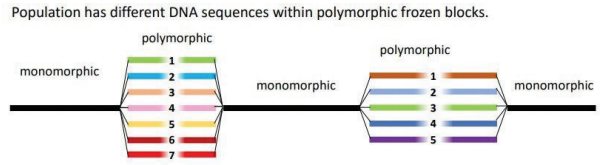

Figure 2: Polymorphic frozen blocks are linked to form extensive haplotypes.
Populations have different DNA sequences within polymorphic frozen blocks. Adapted from Dawkins and Lloyd 2023.

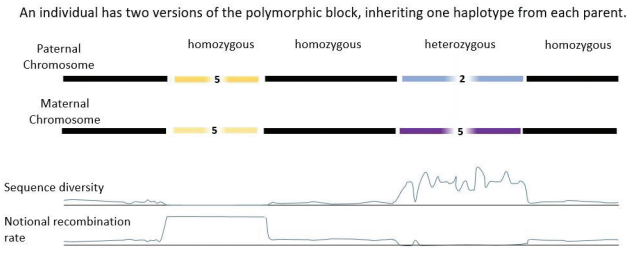

Figure 3: Recombination is suppressed by sequence diversity.
Sequence differences prevent recombination where an individual has inherited a different haplotype from each parent. Recombination can only occur in regions of low sequence diversity or if an individual is homozygous for a particular haplotype. Adapted from Dawkins and Lloyd 2023.

CONSERVED SEQUENCES IN BOS TAURUS RELATED TO MARBLING

Haplotyping and SNP testing of cattle herd

Chromosome 19 in cattle (Bota 19) has been analysed, extensively, particularly to understand the improvement of marbling in cattle (Williamson et al. 2011).

Samples of tissue, hair or blood were collected from a herd of Akaushi cattle for the purpose of DNA parentage verification and herd book registration with the breed association. Single Nucleotide Polymorphism (SNP) genotyping was carried out on 185 individual animals at a commercial laboratory in two batches — one using the Axiom_BovMD array and one using the Axiom_BovMDv3 array. These arrays determine in excess of 50,000 SNPs distributed across the genome, with 1128 SNPs reported on Bota 19. The minimum call rate for animals included in this study was 95%.

Haplotypes of Bota 19 were determined following the method of Williamson et al. 2011. DNA markers at SREBF, NT5M, MPRIP and TCAP were analysed to determine the segregation of alleles through the pedigree and identify haplotypes for each animal in this study (designated MRIP Haplotypes). The MRIP haplotypes were assigned before and independently of the SNP testing.

Results

Analysis of SNP results ordered by chromosome position (Figure 4) reveals some features of interest. There is a region between 27.5 Mb and 28 Mb, within which all the SNPs were homozygous in Akaushi cattle. These SNPs are not homozygous in other breeds. This pattern is a typical signature of selection where the population is bred for a trait determined by a gene or genes within this region. This region includes the gene SLC2A4 gene which encodes GLUT4, the insulin-regulated transporter responsible for glucose transport across the cell membrane.

Some adjacent animals show regions of identical SNP patterns (e.g. samples 94–98). In this case the animals were full siblings produced from an embryo transfer program which were submitted for testing at the same time. There are also some animals with substantial runs of homozygosity seen as vertical white stripes. This is often interpreted as a signature of recent inbreeding, but the same pattern can also occur with homozygous haplotypes conserved from more distant ancestors.

In Figure 5, the animals were grouped by their MRIP haplotypes and the SNP results reveal distinctive patterns of heterozygosity. The 27 SNP profiles on the left are from animals homozygous for the MRIP haplotype 60.7. All of these animals show substantial regions of homozygous SNPs around the location of the markers used to determine

Figure 4: Single Nucleotide Polymorphisms (SNPs) found in the Bovine Axiom array are arranged by chromosome position. Note the breed specific patterns. Each column represents an individual animal. Individuals are sorted by breed. Heterozygous SNP results are marked in green; homozygous SNPs (and any no call results) are white. The vertical axis shows chromosome position of the 1128 SNPs reported on Bota 19. The positions of some genes of importance to muscle and fat development are also shown. Vertical white stripes indicate individual animals with extended regions of homozygosity. Horizontal white stripes indicate localised groups of SNPs that are homozygous within the Akaushi breed. Reproduced with permission from Dawkins 2022.

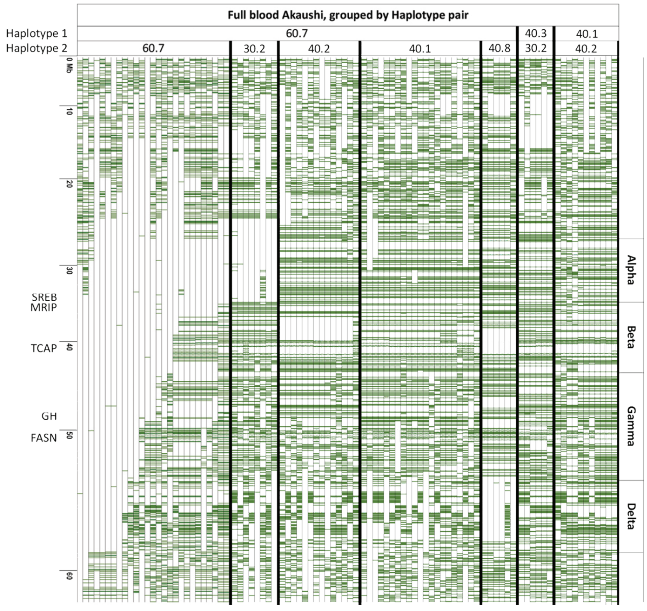

Figure 5: Haplotypes reveal uses for SNPs.
Selected SNP profiles from full blood Akaushi. Individual animals have been grouped by their MRIP Haplotypes displayed on the horizontal axis. Chromosome positions of some relevant genes and markers of the MRIP Haplotypes are shown on the left. Reproduced with permission from Dawkins 2022.

the MRIP haplotypes. Several of these animals have entirely homozygous SNPs from 25Mb to 55Mb i.e. 30 megabases.

The 60.7 haplotype is present at high frequency within the Akaushi breed. Paternal and maternal haplotypes are inherited from different ancestors over at least three generations. This haplotype is not found in either Black Wagyu or in Simmental, which, together with Hanwoo, have been proposed as the main foundation breeds in the development of Akaushi by selective mating. Those animals heterozygous for the MRIP haplotype have patterns of heterozygous SNPs where the SNPs of the paternal and maternal haplotypes differ. For example, there are 20 animals heterozygous for 60.7, with the second haplotype being 40.1. These animals show a distinctive pattern of SNP heterozygosity over the same region that the 60.7 homozygotes have homozygous SNPs. The heterozygous SNPs show the many SNPs at which the two haplotypes differ. It is possible to determine the SNP profile of the 40.1 haplotype by comparison with the 60.7 homozygotes. The present results reveal very extensive areas of conservation extending as far as 50Mb. On the right hand is the provisional classification into component polymorphic frozen blocks. 24 of the 27 MRIP 60.7 homozygotes are homozygous for a distinct block of SNPs, designated alpha.

CONCLUSION

We describe a strategy for the collection and application of founder sequences. By selecting for some traits and therefore CPS regions, necessarily with their adjacent regulatory sequences, breeders have been unintentionally reducing the genetic diversity of livestock. The Marshall Centre for Rural Research is collating a database of CPSs at http://cyo.edu.au/CPS_Database. The database is available for researchers wishing to submit CPSs of 400 nucleotides or longer shown to exist as at least three haplotypes within a species at a well-defined genomic location.

Acknowledgements

We are grateful to our colleagues at C.Y. O'Connor ERADE Village Foundation (CYOEVF) for the discussions which have led to this manuscript. This is publication number 2207 of the CYOEVF. The figures are reproduced with permission of the author and publisher of *Adapting Genetics*, 2nd edition.

Conflict of interest

The authors declare no conflicts of interest.

References

Alper, C.A., Dawkins, R.L., Kulski, J.K, Larsen, C.E., & Lloyd, S.S., 2023. Editorial: Population genomic architecture: conserved polymorphic sequences (CPSs), not linkage disequilibrium. *Frontiers in Genetics* 14:1140350. DOI: 10.3389/fgene.2023.1140350

Ceppellini, R., Curtoni, E.S., Mattuiz, P.L., Miggiano, V., and Scudeller, G. & Serra, A., 1967. Genetics of leukocyte antigens: a family study of segregation and linkage. In *Histocompatibility Testing* 1967, eds. E.S. Curtoni, P.L. Mattiuz & R.M. Tosi (Munksgaard, Copenhagen), 149–187

Dawkins, R.L., 2022. *Adapting Genetics. Educating the Genome: Safe Deposit Evolution*. Second Edition. Dallas, TX: Near Urban LLC (through Amazon). ISBN-10: 0994204043

Dawkins, R.L. & Lloyd, S.S., 2019. MHC genomics and disease: looking back to go forward. *Cells* 8(9):944. DOI: 10.3390/cells8090944.

Dawkins, R.L. & Lloyd, S.S., 2023. Commentary: conserved polymorphic sequences protect themselves for future challenges. *Frontiers in Genetics* 13:993944. DOI: 10.3389/fgene.2022.993944

Gabriel, S.B., Schaffner, S.F., Nguyen, H., Moore, J.M., Roy, J., Blumenstiel, B. & Altshuler, D., 2002. The structure of haplotype blocks in the human genome. *Science* 296(5576), 2225–2229. https://doi.org/10.1126/science.1069424

Gaudieri, S., Leelayuwat, C., Tay, G.K., Townend, D.C. & Dawkins, R.L., 1997. The major histocompatibility complex (MHC) contains conserved polymorphic genomic sequences that are shuffled by recombination to form ethnic-specific haplotypes. *Journal of Molecular Evolution* 45(1): 17–23. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9211730

Kay, P.H., Dawkins, R.L., Williamson, J., Tokunaga, K., Christiansen, F.T. & Charoenwong, P., 1988. Coexistence of an MHC chromosomal segment marked by HLA B17,BfS,C4A6,B1,DR7, and DQw9 in different ethnic groups. *Human Immunology* 27–35, ISSN 0198-8859, https://doi.org/10.1016/0198-8859(88)90015-8.

Kulski, J.K., Suzuki, S. & Shiina, T., 2021. SNP-Density Crossover Maps of Polymorphic Transposable Elements and HLA Genes Within MHC Class I Haplotype Blocks and Junction. *Frontiers in Genetics*, https://doi.org/10.3389/fgene.2020.594318

Lloyd, S.S, Bayard, D., Lester, S., Williamson, J. & Dawkins, R., 2013. The value of haplotyping. *Interbull Bulletin* 47, 252–255

Lloyd, S.S., Steele, E.J. & Dawkins, R.L., 2016. Analysis of haplotype sequences. In J.K. Kulski (ed.) *Next Generation Sequencing — Advances, Applications and Challenges*. InTech, 345–368

Longman-Jacobsen, N., Williamson, J.F., Dawkins, R.L. & Gaudieri, S., 2003. In polymorphic genomic regions indels cluster with nucleotide polymorphism: Quantum genomics. *Gene* 312(1–2), 257–261. https://doi.org/10.1016/S0378-1119(03)00621-8

McLure, C.A., Hinchliffe, P., Lester S., Williamson, J.F., Millman, J.A., Keating, P.J., Stewart, B.J. & Dawkins, R.L., 2013. Genomic evolution and polymorphism: segmental duplications and haplotypes at 108 regions on 21 chromosomes. *Genomics* 102(1):15–26. DOI: 10.1016/j.ygeno.2013.02.011. Epub 2013 Mar 13. PMID: 23501787

McLure, C.A., Williamson, J.F., Smyth, L.A., Agrawal, S., Lester, S., Millman, J.A., Keating, P.J., Stewart, B.J. & Dawkins, R.L., 2005. Extensive genomic and functional polymorphism of the Complement Control Proteins. *Immunogenetics* 57(11), pp. 805–815. Available at: https://doi.org/10.1007/s00251-005-0049-2

Petersdorf, E.W., 2017. In celebration of Ruggero Ceppellini: HLA in transplantation. *HLA* 89, 71–76. doi: 10.1111/tan.12955

Radman, M., 2022. Speciation of genes and genomes: conservation of DNA polymorphism by barriers to recombination raised by mismatch repair system. *Frontiers in Genetics* 13(February), 1–12. Available at: https://doi.org/10.3389/fgene.2022.803690.

Tait, B.D., 2022. The importance of establishing genetic phase in clinical medicine. *International Journal of Immunogenetics* 49, 1–7. DOI: 10.1111/iji.12567

Williamson, J.F., Steele, E.J., Lester, S., Kalai, O., Millman, J.A., Wolrige, L., Dayard, D., McLure, C. & Dawkins, R.L., 2011. Genomic evolution in domestic cattle: ancestral haplotypes and healthy beef. *Genomics* 2011/02/23, 97(5), 304–312. Available at: https://doi.org/S0888-7543(11)00037-1 [pii] 10.1016/j.ygeno.2011.02.006