Lecture 4:

Primary research vs. review articles:

A way to identify literature & conduct a search

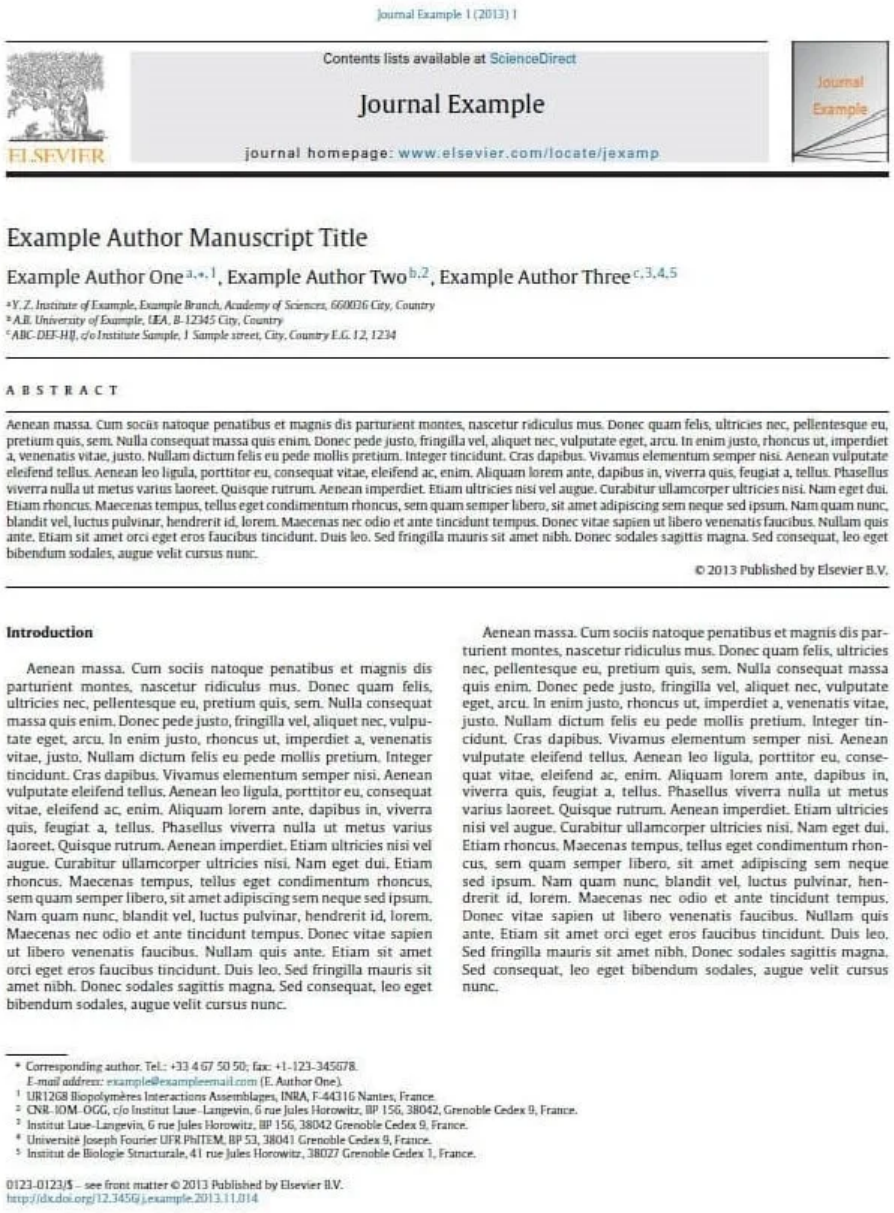
Course 501

Writing and Communication Skills

AIMS

- Understand the structure and components of review and research articles
- Learn how to evaluate journals, articles, and authors.
- Conduct literature research
- Learn about academic search engines and websites
- Identify recent published literature related to your research project or field of study.

Let’s revisit the primary and secondary articles



Journal Example 1 (2013) 1

Contents lists available at ScienceDirect

Journal Example

journal homepage: www.elsevier.com/locate/jexamp

Example Author Manuscript Title

Example Author One [a,*,1], Example Author Two [b,2], Example Author Three [c,3,4,5]

[a] Y. Z. Institute of Example, Example Branch, Academy of Sciences, 660026 City, Country
[b] A.B. University of Example, UEA, B-12345 City, Country
[c] ABC-DEF-HIJ, c/o Institute Sample, 1 Sample street, City, Country E.G. 12, 1234

ABSTRACT

Aenean massa. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Donec quam felis, ultricies nec, pellentesque eu, pretium quis, sem. Nulla consequat massa quis enim. Donec pede justo, fringilla vel, aliquet nec, vulputate eget, arcu. In enim justo, rhoncus ut, imperdiet a, venenatis vitae, justo. Nullam dictum felis eu pede mollis pretium. Integer tincidunt. Cras dapibus. Vivamus elementum semper nisi. Aenean vulputate eleifend tellus. Aenean leo ligula, porttitor eu, consequat vitae, eleifend ac, enim. Aliquam lorem ante, dapibus in, viverra quis, feugiat a, tellus. Phasellus viverra nulla ut metus varius laoreet. Quisque rutrum. Aenean imperdiet. Etiam ultricies nisi vel augue. Curabitur ullamcorper ultricies nisi. Nam eget dui. Etiam rhoncus. Maecenas tempus, tellus eget condimentum rhoncus, sem quam semper libero, sit amet adipiscing sem neque sed ipsum. Nam quam nunc, blandit vel, luctus pulvinar, hendrerit id, lorem. Maecenas nec odio et ante tincidunt tempus. Donec vitae sapien ut libero venenatis faucibus. Nullam quis ante. Etiam sit amet orci eget eros faucibus tincidunt. Duis leo. Sed fringilla mauris sit amet nibh. Donec sodales sagittis magna. Sed consequat, leo eget bibendum sodales, augue velit cursus nunc.

Introduction

Aenean massa. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Donec quam felis, ultricies nec, pellentesque eu, pretium quis, sem. Nulla consequat massa quis enim. Donec pede justo, fringilla vel, aliquet nec, vulputate eget, arcu. In enim justo, rhoncus ut, imperdiet a, venenatis vitae, justo. Nullam dictum felis eu pede mollis pretium. Integer tincidunt. Cras dapibus. Vivamus elementum semper nisi. Aenean vulputate eleifend tellus. Aenean leo ligula, porttitor eu, consequat vitae, eleifend ac, enim. Aliquam lorem ante, dapibus in, viverra quis, feugiat a, tellus. Phasellus viverra nulla ut metus varius laoreet. Quisque rutrum. Aenean imperdiet. Etiam ultricies nisi vel augue. Curabitur ullamcorper ultricies nisi. Nam eget dui. Etiam rhoncus. Maecenas tempus, tellus eget condimentum rhoncus, sem quam semper libero, sit amet adipiscing sem neque sed ipsum. Nam quam nunc, blandit vel, luctus pulvinar, hendrerit id, lorem. Maecenas nec odio et ante tincidunt tempus. Donec vitae sapien ut libero venenatis faucibus. Nullam quis ante. Etiam sit amet orci eget eros faucibus tincidunt. Duis leo. Sed fringilla mauris sit amet nibh. Donec sodales sagittis magna. Sed consequat, leo eget bibendum sodales, augue velit cursus nunc.

Aenean massa. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Donec quam felis, ultricies nec, pellentesque eu, pretium quis, sem. Nulla consequat massa quis enim. Donec pede justo, fringilla vel, aliquet nec, vulputate eget, arcu. In enim justo, rhoncus ut, imperdiet a, venenatis vitae, justo. Nullam dictum felis eu pede mollis pretium. Integer tincidunt. Cras dapibus. Vivamus elementum semper nisi. Aenean vulputate eleifend tellus. Aenean leo ligula, porttitor eu, consequat vitae, eleifend ac, enim. Aliquam lorem ante, dapibus in, viverra quis, feugiat a, tellus. Phasellus viverra nulla ut metus varius laoreet. Quisque rutrum. Aenean imperdiet. Etiam ultricies nisi vel augue. Curabitur ullamcorper ultricies nisi. Nam eget dui. Etiam rhoncus. Maecenas tempus, tellus eget condimentum rhoncus, sem quam semper libero, sit amet adipiscing sem neque sed ipsum. Nam quam nunc, blandit vel, luctus pulvinar, hendrerit id, lorem. Maecenas nec odio et ante tincidunt tempus. Donec vitae sapien ut libero venenatis faucibus. Nullam quis ante. Etiam sit amet orci eget eros faucibus tincidunt. Duis leo. Sed fringilla mauris sit amet nibh. Donec sodales sagittis magna. Sed consequat, leo eget bibendum sodales, augue velit cursus nunc.

* Corresponding author. Tel.: +33 4 67 50 50; fax: +1-123-345678.
E-mail address: example@exampleemail.com (E. Author One).
[1] UR1268 Biopolymères Interactions Assemblages, INRA, F-44316 Nantes, France.
[2] CNR-IOM-OGG, c/o Institut Laue-Langevin, 6 rue Jules Horowitz, BP 156, 38042, Grenoble Cedex 9, France.
[3] Institut Laue-Langevin, 6 rue Jules Horowitz, BP 156, 38042 Grenoble Cedex 9, France.
[4] Université Joseph Fourier UFR PhITEM, BP 53, 38041 Grenoble Cedex 9, France.
[5] Institut de Biologie Structurale, 41 rue Jules Horowitz, 38027 Grenoble Cedex 1, France.

0123-0123/$ - see front matter © 2013 Published by Elsevier B.V.
http://dx.doi.org/12.3456/j.example.2013.11.014

Primary literature

Peer-reviewed scientific articles

ideas

Primary literature

Peer-reviewed scientific articles

Citations and References

Chemistry Education Research and Practice

ROYAL SOCIETY OF CHEMISTRY

EDITORIAL

View Article Online
View Journal | View Issue

Cite this: *Chem. Educ. Res. Pract.*, 2021, **22**, 561

Writing a review article: what to do with my literature review

Nicole Graulich,[a] Scott E. Lewis,[b] Ajda Kahveci,[c] James M. Nyachwaya[d] and Gwendolyn A. Lawrie[*e]

DOI: 10.1039/d1rp90006d

rsc.li/cerp

Introduction

Science education and chemistry education articles have proliferated in the last two decades. For researchers new to the field, it can be hard to get an overview about a research area, for example what has been studied and general trends over time. This increases both the importance of, and the demand for, review articles. This type of article is typically based on already published work and is meant to summarize and collate available studies about a research topic. Reviews play an integrative role in synthesizing the body of literature under a thematic umbrella. A reviewing lens on published work can integrate and outline state-of-the-art research in a field, provide a discussion of controversies and inconsistencies in prior research, evaluate existing methodological approaches or possibly propose future research endeavors. Some review articles adopt a more quantitative effects estimation approach, whereas others are more narrative, seeking to synthesize qualitative findings. Not every individual review article can cover all of these potential objectives.

Review articles allow the readers to get a landscape view of a topic, but readers can also use the collection of references cited in a review article to dig deeper into a topic. Thus, they are valuable resources to consult. Well written review articles are often highly cited and could increase the visibility and reputation of the authors.

Decisions to make before starting to write a review article

Before starting to write a review article, it is helpful to clarify whether, especially for researchers early in their careers, review articles count towards their promotion or tenure benchmarks. Depending on the country or requirements, articles reporting original research work can have higher value than review articles for these processes. If tenure requirements are related to citation indices, a review article is probably worth the investment of time. Making this decision about writing a review article or not should be guided by this economic lens, as a well-written and well-researched review article can be very time-consuming. Time might be considered to be precious when just having started an academic position.

It might be tempting to consider adapting a literature review, that is part of an article, proposal or dissertation, into a published review article. Such a literature review can be used as a starting point to build a review article upon. However, a literature review often does not follow the quality criteria of a formal review article or specific types of reviews and therefore should be reworked based on the steps illustrated in this editorial.

Types of review articles suitable for chemistry education research and practice

The denomination of review types can vary depending on the field and on the resources used. This editorial does not encompass all types of review articles that are possible, but we endeavor to list and further explain the main types of review articles. *Chemistry Education Research and Practice* publishes three manuscript types: (1) original research articles, (2) perspectives and (3) review articles. The latter category includes narrative, integrative or systematic reviews and meta-analyses. Perspectives serve a different purpose than review articles, although like review articles perspectives should also be based on, or discuss, published research. Review articles need to align with the goals and scope of the journal. Thought experiments outlining a theoretical position or personal opinion without including a literature basis,

[a] *Institute of Chemistry Education, Justus-Liebig-University, Giessen, Germany*
[b] *Department of Chemistry, University of South Florida, USA*
[c] *Department of Chemistry, Fort Hays State University, USA*
[d] *Department of Chemistry and Biochemistry and School of Education, North Dakota State University, USA*
[e] *School of Chemistry & Molecular Biosciences, The University of Queensland, St Lucia, 4072, Australia. E-mail: g.lawrie@uq.edu.au*

Secondary literature

Review articles

ideas

PEER REVIEW

Secondary literature

Review articles

Citations and References

Can you identify the differences between review and research articles?

Can you tell if the article is worth your reading and citation?

You need to know what you are looking for and how to evaluate good sources

Review article

FUNDAMENTAL CONCEPTS IN GENETICS

Linkage disequilibrium — understanding the evolutionary past and mapping the medical future

Montgomery Slatkin

Abstract | Linkage disequilibrium — the nonrandom association of alleles at different loci — is a sensitive indicator of the population genetic forces that structure a genome. Because of the explosive growth of methods for assessing genetic variation at a fine scale, evolutionary biologists and human geneticists are increasingly exploiting linkage disequilibrium in order to understand past evolutionary and demographic events, to map genes that are associated with quantitative characters and inherited diseases, and to understand the joint evolution of linked sets of genes. This article introduces linkage disequilibrium, reviews the population genetic processes that affect it and describes some of its uses. At present, linkage disequilibrium is used much more extensively in the study of humans than in non-humans, but that is changing as technological advances make extensive genomic studies feasible in other species.

Linkage disequilibrium (LD) is one of those unfortunate terms that does not reveal its meaning. As every instructor of population genetics knows, the term is a barrier not an aid to understanding. LD means simply a nonrandom association of alleles at two or more loci, and detecting LD does not ensure either linkage or a lack of equilibrium. The term was first used in 1960 by Lewontin and Kojima[1] and it persists because LD was initially the concern of population geneticists who were not picky about terminology as long as the mathematical definition was clear. At first, there were few data with which to study LD, and its importance to evolutionary biology and human genetics was unrecognized outside of population genetics. However, interest in LD grew rapidly in the 1980s once the usefulness of LD for gene mapping became evident and large-scale surveys of closely linked loci became feasible. By then, the term was too well established to be replaced.

LD is of importance in evolutionary biology and human genetics because so many factors affect it and are affected by it. LD provides information about past events and it constrains the potential response to both natural and artificial selection. LD throughout the genome reflects the population history, the breeding system and the pattern of geographic subdivision, whereas LD in each genomic region reflects the history of natural selection, gene conversion, mutation and other forces that cause gene-frequency evolution. How these factors affect LD between a particular pair of loci or in a genomic region depends on local recombination rates. The population genetics theory of LD is well developed and is being widely used to provide insight into evolutionary history and as the basis for mapping genes in humans and in other species.

In this article, I will review the definitions of LD and the problems with assessing it, then outline the basic population genetics of LD that tells us how natural selection, genetic drift, recombination and mutation all affect levels of LD, and finally discuss some recent applications of LD to mapping genes, inferring the intensity of selection in the genome and estimating allele age.

Definitions

One pair of loci. LD between alleles at two loci has been defined in many ways (BOX 1), but all definitions depend on the quantity:

$$D_{AB} = p_{AB} - P_A P_B \quad (1)$$

which is the difference between the frequency of gametes carrying the pair of alleles A and B at two loci (p_{AB}) and the product of the frequencies of those alleles (p_A and p_B). Originally, the definition was in terms of gamete frequencies because that allows for the possibility that the

Department of Integrative Biology, University of California, Berkeley, California 94720-3140, USA. e-mail: slatkin@berkeley.edu
doi:10.1038/nrg2361
Published online
22 April 2008

Research article

Received: 12 July 2023 | Accepted: 19 August 2023
DOI: 10.1111/age.13352

RESEARCH ARTICLE

ANIMAL GENETICS WILEY

A *de novo* 2.3 kb structural variant in *MITF* explains a novel splashed white phenotype in a Thoroughbred family

R. R. Bellone[1,2] | J. Tanaka[1] | E. Esdaile[1] | R. B. Sutton[3] | F. Payette[4] | L. Leduc[4] | B. J. Till[1] | A. K. Abdel-Ghaffar[1] | M. Hammond[1,2] | K. G. Magdesian[5]

[1]Veterinary Genetics Laboratory, School of Veterinary Medicine, UC Davis, Davis, California, USA
[2]Department of Population Health and Reproduction, School of Veterinary Medicine, UC Davis, Davis, California, USA
[3]Cell Physiology and Molecular Biophysics, School of Medicine, Texas Tech University Health Sciences Center, Lubbock, Texas, USA
[4]Department of Clinical Studies, New Bolton Center, University of Pennsylvania School of Veterinary Medicine, University School of Veterinary Medicine, University of Pennsylvania, Kennett Square, Pennsylvania, USA
[5]Department of Medicine and Epidemiology, School of Veterinary Medicine, University of California-Davis, Davis, California, USA

Correspondence
R. R. Bellone, Veterinary Genetics Laboratory, School of Veterinary Medicine, UC Davis, Davis, CA, USA.
Email: rbellone@ucdavis.edu

Funding information
UC Davis Veterinary Genetics Laboratory

Abstract
Splashed white in horses is characterized by extensive white patterning on the legs, face and abdomen and may be accompanied by deafness. To date, seven variants in *microphthalmia-associated transcription factor* (*MITF*) and two variants in *Paired Box 3* (*PAX3*) have been identified to explain this phenotype. A splashed white Thoroughbred stallion, whose sire and dam were not patterned, was hypothesized to have a *de novo* variant leading to his white coat pattern. A whole-genome sequencing candidate gene approach identified two single nucleotide variants (SNVs) in *SOX10*, four SNVs in *MITF* and a 2.3 kb deletion in *MITF* with the alternative allele present in this stallion but absent in the other 18 horses analyzed. All six SNVs were annotated as modifiers and were not further considered. The deletion in *MITF* (NC_009159.3:g.21555811_21558139delinsAAAT) encompasses exon 9 encoding a part of the helix–loop–helix domain required for DNA binding. Sanger sequencing and parentage testing confirmed that this deletion was a *de novo* mutation of maternal origin. Consistent with the published nomenclature, we denote this likely causal variant as SW8. Genotyping three of this stallion's offspring identified SW8 only in the nearly all-white foal that was confirmed deaf by brainstem auditory evoked response testing. This foal was also a compound heterozygote for dominant white variants (W20/W22), but to date, W variants alone have not been connected to deafness. SW8 marks the fourth de novo *MITF* variant in horses reported to cause white patterning. The link between deafness and all *MITF* variants with and without other variants impacting melanocyte development and function needs to be further explored.

KEYWORDS
coat color, deafness, horse, pigmentation, white patterning

INTRODUCTION

Variation in white patterning has long fascinated animal enthusiasts and researchers alike. Moreover, some of these variations have a clinical connection and have similar counterparts in humans or other species. In horses, several different types of white patterning phenotypes have been described including frame overo, tobiano, roan, leopard complex spotting, dominant white, sabino and splashed white (Sponenberg & Bellone, 2017). The molecular bases of many of these have been determined, with the first to be unraveled being the frame overo pattern in which white patterning is prevalent on the ventral aspects of the horse with the pigment 'framing' the

Review articles:

Structure and components

FUNDAMENTAL CONCEPTS IN GENETICS

Linkage disequilibrium — understanding the evolutionary past and mapping the medical future

Montgomery Slatkin

Abstract | Linkage disequilibrium — the nonrandom association of alleles at different loci — is a sensitive indicator of the population genetic forces that structure a genome. Because of the explosive growth of methods for assessing genetic variation at a fine scale, evolutionary biologists and human geneticists are increasingly exploiting linkage disequilibrium in order to understand past evolutionary and demographic events, to map genes that are associated with quantitative characters and inherited diseases, and to understand the joint evolution of linked sets of genes. This article introduces linkage disequilibrium, reviews the population genetic processes that affect it and describes some of its uses. At present, linkage disequilibrium is used much more extensively in the study of humans than in non-humans, but that is changing as technological advances make extensive genomic studies feasible in other species.

Linkage disequilibrium (LD) is one of those unfortunate terms that does not reveal its meaning. As every instructor of population genetics knows, the term is a barrier not an aid to understanding. LD means simply a nonrandom association of alleles at two or more loci, and detecting LD does not ensure either linkage or a lack of equilibrium. The term was first used in 1960 by Lewontin and Kojima[1] and it persists because LD was initially the concern of population geneticists who were not picky about terminology as long as the mathematical definition was clear. At first, there were few data with which to study LD, and its importance to evolutionary biology and human genetics was unrecognized outside of population genetics. However, interest in LD grew rapidly in the 1980s once the usefulness of LD for gene mapping became evident and large-scale surveys of closely linked loci became feasible. By then, the term was too well established to be replaced.

LD is of importance in evolutionary biology and human genetics because so many factors affect it and are affected by it. LD provides information about past events and it constrains the potential response to both natural and artificial selection. LD throughout the genome reflects the population history, the breeding system and the pattern of geographic subdivision, whereas LD in each genomic region reflects the history of natural selection, gene conversion, mutation and other forces that cause gene-frequency evolution. How these factors affect LD between a particular pair of loci or in a genomic region depends on local recombination rates. The population genetics theory of LD is well developed and is being widely used to provide insight into evolutionary history and as the basis for mapping genes in humans and in other species.

In this article, I will review the definitions of LD and the problems with assessing it, then outline the basic population genetics of LD that tells us how natural selection, genetic drift, recombination and mutation all affect levels of LD, and finally discuss some recent applications of LD to mapping genes, inferring the intensity of selection in the genome and estimating allele age.

Definitions

One pair of loci. LD between alleles at two loci has been defined in many ways (BOX 1), but all definitions depend on the quantity:

$$D_{AB} = p_{AB} - p_A p_B \qquad (1)$$

which is the difference between the frequency of gametes carrying the pair of alleles A and B at two loci (p_{AB}) and the product of the frequencies of those alleles (p_A and p_B). Originally, the definition was in terms of gamete frequencies because that allows for the possibility that the

Department of Integrative Biology, University of California, Berkeley, California 94720-3140, USA.
e-mail: slatkin@berkeley.edu
doi:10.1038/nrg2361
Published online
22 April 2008

FUNDAMENTAL CONCEPTS IN GENETICS

Linkage disequilibrium — understanding the evolutionary past and mapping the medical future

Montgomery Slatkin

Abstract | Linkage disequilibrium — the nonrandom association of alleles at different loci — is a sensitive indicator of the population genetic forces that structure a genome. Because of the explosive growth of methods for assessing genetic variation at a fine scale, evolutionary biologists and human geneticists are increasingly exploiting linkage disequilibrium in order to understand past evolutionary and demographic events, to map genes that are associated with quantitative characters and inherited diseases, and to understand the joint evolution of linked sets of genes. This article introduces linkage disequilibrium, reviews the population genetic processes that affect it and describes some of its uses. At present, linkage disequilibrium is used much more extensively in the study of humans than in non-humans, but that is changing as technological advances make extensive genomic studies feasible in other species.

Linkage disequilibrium (LD) is one of those unfortunate terms that does not reveal its meaning. As every instructor of population genetics knows, the term is a barrier not an aid to understanding. LD means simply a nonrandom association of alleles at two or more loci, and detecting LD does not ensure either linkage or a lack of equilibrium. The term was first used in 1960 by Lewontin and Kojima[1] and it persists because LD was initially the concern of population geneticists who were not picky about terminology as long as the mathematical definition was clear. At first, there were few data with which to study LD, and its importance to evolutionary biology and human genetics was unrecognized outside of population genetics. However, interest in LD grew rapidly in the 1980s once the usefulness of LD for gene mapping became evident and large-scale surveys of closely linked loci became feasible. By then, the term was too well established to be replaced.

LD is of importance in evolutionary biology and human genetics because so many factors affect it and are affected by it. LD provides information about past events and it constrains the potential response to both natural and artificial selection. LD throughout the genome reflects the population history, the breeding system and the pattern of geographic subdivision, whereas LD in each genomic region reflects the history of natural selection, gene conversion, mutation and other forces that cause gene-frequency evolution. How these factors affect LD between a particular pair of loci or in a genomic region depends on local recombination rates. The population genetics theory of LD is well developed and is being widely used to provide insight into evolutionary history and as the basis for mapping genes in humans and in other species.

In this article, I will review the definitions of LD and the problems with assessing it, then outline the basic population genetics of LD that tells us how natural selection, genetic drift, recombination and mutation all affect levels of LD, and finally discuss some recent applications of LD to mapping genes, inferring the intensity of selection in the genome and estimating allele age.

Definitions

One pair of loci. LD between alleles at two loci has been defined in many ways (BOX 1), but all definitions depend on the quantity:

$$D_{AB} = p_{AB} - p_A p_B \qquad (1)$$

which is the difference between the frequency of gametes carrying the pair of alleles A and B at two loci (p_{AB}) and the product of the frequencies of those alleles (p_A and p_B). Originally, the definition was in terms of gamete frequencies because that allows for the possibility that the

Department of Integrative Biology, University of California, Berkeley, California 94720-3140, USA.
e-mail: slatkin@berkeley.edu
doi:10.1038/nrg2361
Published online
22 April 2008

Title

FUNDAMENTAL CONCEPTS IN GENETICS

Linkage disequilibrium — understanding the evolutionary past and mapping the medical future

Montgomery Slatkin

Abstract | Linkage disequilibrium — the nonrandom association of alleles at different loci — is a sensitive indicator of the population genetic forces that structure a genome. Because of the explosive growth of methods for assessing genetic variation at a fine scale, evolutionary biologists and human geneticists are increasingly exploiting linkage disequilibrium in order to understand past evolutionary and demographic events, to map genes that are associated with quantitative characters and inherited diseases, and to understand the joint evolution of linked sets of genes. This article introduces linkage disequilibrium, reviews the population genetic processes that affect it and describes some of its uses. At present, linkage disequilibrium is used much more extensively in the study of humans than in non-humans, but that is changing as technological advances make extensive genomic studies feasible in other species.

Linkage disequilibrium (LD) is one of those unfortunate terms that does not reveal its meaning. As every instructor of population genetics knows, the term is a barrier not an aid to understanding. LD means simply a nonrandom association of alleles at two or more loci, and detecting LD does not ensure either linkage or a lack of equilibrium. The term was first used in 1960 by Lewontin and Kojima[1] and it persists because LD was initially the concern of population geneticists who were not picky about terminology as long as the mathematical definition was clear. At first, there were few data with which to study LD, and its importance to evolutionary biology and human genetics was unrecognized outside of population genetics. However, interest in LD grew rapidly in the 1980s once the usefulness of LD for gene mapping became evident and large-scale surveys of closely linked loci became feasible. By then, the term was too well established to be replaced.

LD is of importance in evolutionary biology and human genetics because so many factors affect it and are affected by it. LD provides information about past events and it constrains the potential response to both natural and artificial selection. LD throughout the genome reflects the population history, the breeding system and the pattern of geographic subdivision, whereas LD in each genomic region reflects the history of natural selection, gene conversion, mutation and other forces that cause gene-frequency evolution. How these factors affect LD between a particular pair of loci or in a genomic region depends on local recombination rates. The population genetics theory of LD is well developed and is being widely used to provide insight into evolutionary history and as the basis for mapping genes in humans and in other species.

In this article, I will review the definitions of LD and the problems with assessing it, then outline the basic population genetics of LD that tells us how natural selection, genetic drift, recombination and mutation all affect levels of LD, and finally discuss some recent applications of LD to mapping genes, inferring the intensity of selection in the genome and estimating allele age.

Definitions

One pair of loci. LD between alleles at two loci has been defined in many ways (BOX 1), but all definitions depend on the quantity:

$$D_{AB} = p_{AB} - p_A p_B \qquad (1)$$

which is the difference between the frequency of gametes carrying the pair of alleles A and B at two loci (p_{AB}) and the product of the frequencies of those alleles (p_A and p_B). Originally, the definition was in terms of gamete frequencies because that allows for the possibility that the

Department of Integrative Biology, University of California, Berkeley, California 94720-3140, USA.
e-mail: slatkin@berkeley.edu
doi:10.1038/nrg2361
Published online
22 April 2008

What is unique about the titles of good review articles?

Title

FUNDAMENTAL CONCEPTS IN GENETICS

Linkage disequilibrium — understanding the evolutionary past and mapping the medical future

Montgomery Slatkin

Abstract | Linkage disequilibrium — the nonrandom association of alleles at different loci — is a sensitive indicator of the population genetic forces that structure a genome. Because of the explosive growth of methods for assessing genetic variation at a fine scale, evolutionary biologists and human geneticists are increasingly exploiting linkage disequilibrium in order to understand past evolutionary and demographic events, to map genes that are associated with quantitative characters and inherited diseases, and to understand the joint evolution of linked sets of genes. This article introduces linkage disequilibrium, reviews the population genetic processes that affect it and describes some of its uses. At present, linkage disequilibrium is used much more extensively in the study of humans than in non-humans, but that is changing as technological advances make extensive genomic studies feasible in other species.

Linkage disequilibrium (LD) is one of those unfortunate terms that does not reveal its meaning. As every instructor of population genetics knows, the term is a barrier not an aid to understanding. LD means simply a nonrandom association of alleles at two or more loci, and detecting LD does not ensure either linkage or a lack of equilibrium. The term was first used in 1960 by Lewontin and Kojima[1] and it persists because LD was initially the concern of population geneticists who were not picky about terminology as long as the mathematical definition was clear. At first, there were few data with which to study LD, and its importance to evolutionary biology and human genetics was unrecognized outside of population genetics. However, interest in LD grew rapidly in the 1980s once the usefulness of LD for gene mapping became evident and large-scale surveys of closely linked loci became feasible. By then, the term was too well established to be replaced.

LD is of importance in evolutionary biology and human genetics because so many factors affect it and are affected by it. LD provides information about past events and it constrains the potential response to both natural and artificial selection. LD throughout the genome reflects the population history, the breeding system and the pattern of geographic subdivision, whereas LD in each genomic region reflects the history of natural selection, gene conversion, mutation and other forces that cause gene-frequency evolution. How these factors affect LD between a particular pair of loci or in a genomic region depends on local recombination rates. The population genetics theory of LD is well developed and is being widely used to provide insight into evolutionary history and as the basis for mapping genes in humans and in other species.

In this article, I will review the definitions of LD and the problems with assessing it, then outline the basic population genetics of LD that tells us how natural selection, genetic drift, recombination and mutation all affect levels of LD, and finally discuss some recent applications of LD to mapping genes, inferring the intensity of selection in the genome and estimating allele age.

Definitions

One pair of loci. LD between alleles at two loci has been defined in many ways (BOX 1), but all definitions depend on the quantity:

$$D_{AB} = p_{AB} - p_A p_B \qquad (1)$$

which is the difference between the frequency of gametes carrying the pair of alleles A and B at two loci (p_{AB}) and the product of the frequencies of those alleles (p_A and p_B). Originally, the definition was in terms of gamete frequencies because that allows for the possibility that the

Department of Integrative Biology, University of California, Berkeley, California 94720-3140, USA.
e-mail: slatkin@berkeley.edu
doi:10.1038/nrg2361
Published online
22 April 2008

Short
Broad
Catchy
Contains keywords
No specific terminology

Author

FUNDAMENTAL CONCEPTS IN GENETICS

Linkage disequilibrium — understanding the evolutionary past and mapping the medical future

Montgomery Slatkin

Abstract | Linkage disequilibrium — the nonrandom association of alleles at different loci — is a sensitive indicator of the population genetic forces that structure a genome. Because of the explosive growth of methods for assessing genetic variation at a fine scale, evolutionary biologists and human geneticists are increasingly exploiting linkage disequilibrium in order to understand past evolutionary and demographic events, to map genes that are associated with quantitative characters and inherited diseases, and to understand the joint evolution of linked sets of genes. This article introduces linkage disequilibrium, reviews the population genetic processes that affect it and describes some of its uses. At present, linkage disequilibrium is used much more extensively in the study of humans than in non-humans, but that is changing as technological advances make extensive genomic studies feasible in other species.

Linkage disequilibrium (LD) is one of those unfortunate terms that does not reveal its meaning. As every instructor of population genetics knows, the term is a barrier not an aid to understanding. LD means simply a nonrandom association of alleles at two or more loci, and detecting LD does not ensure either linkage or a lack of equilibrium. The term was first used in 1960 by Lewontin and Kojima[1] and it persists because LD was initially the concern of population geneticists who were not picky about terminology as long as the mathematical definition was clear. At first, there were few data with which to study LD, and its importance to evolutionary biology and human genetics was unrecognized outside of population genetics. However, interest in LD grew rapidly in the 1980s once the usefulness of LD for gene mapping became evident and large-scale surveys of closely linked loci became feasible. By then, the term was too well established to be replaced.

LD is of importance in evolutionary biology and human genetics because so many factors affect it and are affected by it. LD provides information about past events and it constrains the potential response to both natural and artificial selection. LD throughout the genome reflects the population history, the breeding system and the pattern of geographic subdivision, whereas LD in each genomic region reflects the history of natural selection, gene conversion, mutation and other forces that cause gene-frequency evolution. How these factors affect LD between a particular pair of loci or in a genomic region depends on local recombination rates. The population genetics theory of LD is well developed and is being widely used to provide insight into evolutionary history and as the basis for mapping genes in humans and in other species.

In this article, I will review the definitions of LD and the problems with assessing it, then outline the basic population genetics of LD that tells us how natural selection, genetic drift, recombination and mutation all affect levels of LD, and finally discuss some recent applications of LD to mapping genes, inferring the intensity of selection in the genome and estimating allele age.

Definitions

One pair of loci. LD between alleles at two loci has been defined in many ways (BOX 1), but all definitions depend on the quantity:

$$D_{AB} = p_{AB} - p_A p_B \qquad (1)$$

which is the difference between the frequency of gametes carrying the pair of alleles A and B at two loci (p_{AB}) and the product of the frequencies of those alleles (p_A and p_B). Originally, the definition was in terms of gamete frequencies because that allows for the possibility that the

Department of Integrative Biology, University of California, Berkeley, California 94720-3140, USA.
e-mail: slatkin@berkeley.edu
doi:10.1038/nrg2361
Published online
22 April 2008

What is unique about the authors of good review articles?

Author

FUNDAMENTAL CONCEPTS IN GENETICS

Linkage disequilibrium — understanding the evolutionary past and mapping the medical future

Montgomery Slatkin

Abstract | Linkage disequilibrium — the nonrandom association of alleles at different loci — is a sensitive indicator of the population genetic forces that structure a genome. Because of the explosive growth of methods for assessing genetic variation at a fine scale, evolutionary biologists and human geneticists are increasingly exploiting linkage disequilibrium in order to understand past evolutionary and demographic events, to map genes that are associated with quantitative characters and inherited diseases, and to understand the joint evolution of linked sets of genes. This article introduces linkage disequilibrium, reviews the population genetic processes that affect it and describes some of its uses. At present, linkage disequilibrium is used much more extensively in the study of humans than in non-humans, but that is changing as technological advances make extensive genomic studies feasible in other species.

Linkage disequilibrium (LD) is one of those unfortunate terms that does not reveal its meaning. As every instructor of population genetics knows, the term is a barrier not an aid to understanding. LD means simply a nonrandom association of alleles at two or more loci, and detecting LD does not ensure either linkage or a lack of equilibrium. The term was first used in 1960 by Lewontin and Kojima[1] and it persists because LD was initially the concern of population geneticists who were not picky about terminology as long as the mathematical definition was clear. At first, there were few data with which to study LD, and its importance to evolutionary biology and human genetics was unrecognized outside of population genetics. However, interest in LD grew rapidly in the 1980s once the usefulness of LD for gene mapping became evident and large-scale surveys of closely linked loci became feasible. By then, the term was too well established to be replaced.

LD is of importance in evolutionary biology and human genetics because so many factors affect it and are affected by it. LD provides information about past events and it constrains the potential response to both natural and artificial selection. LD throughout the genome reflects the population history, the breeding system and the pattern of geographic subdivision, whereas LD in each genomic region reflects the history of natural selection, gene conversion, mutation and other forces that cause gene-frequency evolution. How these factors affect LD between a particular pair of loci or in a genomic region depends on local recombination rates. The population genetics theory of LD is well developed and is being widely used to provide insight into evolutionary history and as the basis for mapping genes in humans and in other species.

In this article, I will review the definitions of LD and the problems with assessing it, then outline the basic population genetics of LD that tells us how natural selection, genetic drift, recombination and mutation all affect levels of LD, and finally discuss some recent applications of LD to mapping genes, inferring the intensity of selection in the genome and estimating allele age.

Definitions

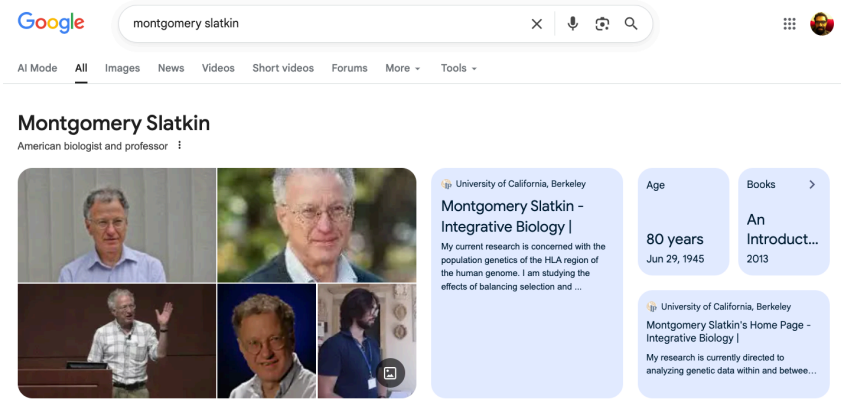
One pair of loci. LD between alleles at two loci has been defined in many ways (BOX 1), but all definitions depend on the quantity:

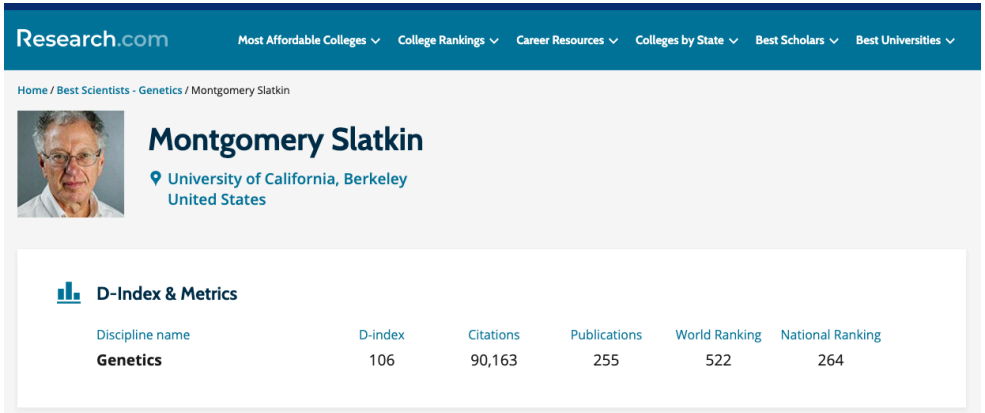
$$D_{AB} = p_{AB} - p_A p_B \qquad (1)$$

which is the difference between the frequency of gametes carrying the pair of alleles A and B at two loci (p_{AB}) and the product of the frequencies of those alleles (p_A and p_B). Originally, the definition was in terms of gamete frequencies because that allows for the possibility that the

Department of Integrative Biology, University of California, Berkeley, California 94720-3140, USA.
e-mail: slatkin@berkeley.edu
doi:10.1038/nrg2361
Published online
22 April 2008

One or a few
Well-known in the field
Senior


Google
montgomery slatkin
AI Mode
All
Images
News
Videos
Short videos
Forums
More
Tools
Montgomery Slatkin
American biologist and professor
University of California, Berkeley
Montgomery Slatkin - Integrative Biology |
My current research is concerned with the population genetics of the HLA region of the human genome. I am studying the effects of balancing selection and ...
Age
80 years
Jun 29, 1945
Books
An Introduct...
2013
University of California, Berkeley
Montgomery Slatkin's Home Page - Integrative Biology |
My research is currently directed to analyzing genetic data within and betwee...


Research.com
Most Affordable Colleges
College Rankings
Career Resources
Colleges by State
Best Scholars
Best Universities
Home / Best Scientists - Genetics / Montgomery Slatkin
Montgomery Slatkin
University of California, Berkeley
United States
D-Index & Metrics
Discipline name
D-index
Citations
Publications
World Ranking
National Ranking
Genetics
106
90,163
255
522
264

FUNDAMENTAL CONCEPTS IN GENETICS

Linkage disequilibrium — understanding the evolutionary past and mapping the medical future

Montgomery Slatkin

Abstract | Linkage disequilibrium — the nonrandom association of alleles at different loci — is a sensitive indicator of the population genetic forces that structure a genome. Because of the explosive growth of methods for assessing genetic variation at a fine scale, evolutionary biologists and human geneticists are increasingly exploiting linkage disequilibrium in order to understand past evolutionary and demographic events, to map genes that are associated with quantitative characters and inherited diseases, and to understand the joint evolution of linked sets of genes. This article introduces linkage disequilibrium, reviews the population genetic processes that affect it and describes some of its uses. At present, linkage disequilibrium is used much more extensively in the study of humans than in non-humans, but that is changing as technological advances make extensive genomic studies feasible in other species.

Linkage disequilibrium (LD) is one of those unfortunate terms that does not reveal its meaning. As every instructor of population genetics knows, the term is a barrier not an aid to understanding. LD means simply a nonrandom association of alleles at two or more loci, and detecting LD does not ensure either linkage or a lack of equilibrium. The term was first used in 1960 by Lewontin and Kojima[1] and it persists because LD was initially the concern of population geneticists who were not picky about terminology as long as the mathematical definition was clear. At first, there were few data with which to study LD, and its importance to evolutionary biology and human genetics was unrecognized outside of population genetics. However, interest in LD grew rapidly in the 1980s once the usefulness of LD for gene mapping became evident and large-scale surveys of closely linked loci became feasible. By then, the term was too well established to be replaced.

LD is of importance in evolutionary biology and selection, gene conversion, mutation and other forces that cause gene-frequency evolution. How these factors affect LD between a particular pair of loci or in a genomic region depends on local recombination rates. The population genetics theory of LD is well developed and is being widely used to provide insight into evolutionary history and as the basis for mapping genes in humans and in other species.

In this article, I will review the definitions of LD and the problems with assessing it, then outline the basic population genetics of LD that tells us how natural selection, genetic drift, recombination and mutation all affect levels of LD, and finally discuss some recent applications of LD to mapping genes, inferring the intensity of selection in the genome and estimating allele age.

Definitions

One pair of loci. LD between alleles at two loci has been defined in many ways (BOX 1), but all definitions depend on the quantity:

Department of Integrative Biology, University of California, Berkeley, California 94720-3140, USA.
e-mail: slatkin@berkeley.edu

doi:10.1038/nrg2361
Published online
22 April 2008

Author's affiliation, contact info, and corresponding author email

FUNDAMENTAL CONCEPTS IN GENETICS

Linkage disequilibrium — understanding the evolutionary past and mapping the medical future

Montgomery Slatkin

Abstract | Linkage disequilibrium — the nonrandom association of alleles at different loci — is a sensitive indicator of the population genetic forces that structure a genome. Because of the explosive growth of methods for assessing genetic variation at a fine scale, evolutionary biologists and human geneticists are increasingly exploiting linkage disequilibrium in order to understand past evolutionary and demographic events, to map genes that are associated with quantitative characters and inherited diseases, and to understand the joint evolution of linked sets of genes. This article introduces linkage disequilibrium, reviews the population genetic processes that affect it and describes some of its uses. At present, linkage disequilibrium is used much more extensively in the study of humans than in non-humans, but that is changing as technological advances make extensive genomic studies feasible in other species.

Linkage disequilibrium (LD) is one of those unfortunate terms that does not reveal its meaning. As every instructor of population genetics knows, the term is a barrier not an aid to understanding. LD means simply a nonrandom association of alleles at two or more loci, and detecting LD does not ensure either linkage or a lack of equilibrium. The term was first used in 1960 by Lewontin and Kojima[1] and it persists because LD was initially the concern of population geneticists who were not picky about terminology as long as the mathematical definition was clear. At first, there were few data with which to study LD, and its importance to evolutionary biology and human genetics was unrecognized outside of population genetics. However, interest in LD grew rapidly in the 1980s once the usefulness of LD for gene mapping became evident and large-scale surveys of closely linked loci became feasible. By then, the term was too well established to be replaced.

LD is of importance in evolutionary biology and human genetics because so many factors affect it and are affected by it. LD provides information about past events and it constrains the potential response to both natural and artificial selection. LD throughout the genome that cause gene-frequency evolution. How these factors affect LD between a particular pair of loci or in a genomic region depends on local recombination rates. The population genetics theory of LD is well developed and is being widely used to provide insight into evolutionary history and as the basis for mapping genes in humans and in other species.

In this article, I will review the definitions of LD and the problems with assessing it, then outline the basic population genetics of LD that tells us how natural selection, genetic drift, recombination and mutation all affect levels of LD, and finally discuss some recent applications of LD to mapping genes, inferring the intensity of selection in the genome and estimating allele age.

Definitions

One pair of loci. LD between alleles at two loci has been defined in many ways (BOX 1), but all definitions depend on the quantity:

$$D_{AB} = p_{AB} - p_A p_B \qquad (1)$$

which is the difference between the frequency of gametes

Department of Integrative Biology, University of California, Berkeley, California 94720-3140, USA.
e-mail: slatkin@berkeley.edu
doi:10.1038/nrg2361
Published online
22 April 2008

Digital Object Identifier

FUNDAMENTAL CONCEPTS IN GENETICS

Linkage disequilibrium — understanding the evolutionary past and mapping the medical future

Montgomery Slatkin

Abstract | Linkage disequilibrium — the nonrandom association of alleles at different loci — is a sensitive indicator of the population genetic forces that structure a genome. Because of the explosive growth of methods for assessing genetic variation at a fine scale, evolutionary biologists and human geneticists are increasingly exploiting linkage disequilibrium in order to understand past evolutionary and demographic events, to map genes that are associated with quantitative characters and inherited diseases, and to understand the joint evolution of linked sets of genes. This article introduces linkage disequilibrium, reviews the population genetic processes that affect it and describes some of its uses. At present, linkage disequilibrium is used much more extensively in the study of humans than in non-humans, but that is changing as technological advances make extensive genomic studies feasible in other species.

Linkage disequilibrium (LD) is one of those unfortunate terms that does not reveal its meaning. As every instructor of population genetics knows, the term is a barrier not an aid to understanding. LD means simply a nonrandom association of alleles at two or more loci, and detecting LD does not ensure either linkage or a lack of equilibrium. The term was first used in 1960 by Lewontin and Kojima[1] and it persists because LD was initially the concern of population geneticists who were not picky about terminology as long as the mathematical definition was clear. At first, there were few data with which to study LD, and its importance to evolutionary biology and human genetics was unrecognized outside of population genetics. However, interest in LD grew rapidly in the 1980s once the usefulness of LD for gene mapping became evident and large-scale surveys of closely linked loci became feasible. By then, the term was too well established to be replaced.

LD is of importance in evolutionary biology and human genetics because so many factors affect it and are affected by it. LD provides information about past events and it constrains the potential response to both natural and artificial selection. LD throughout the genome reflects the population history, the breeding system that cause gene-frequency evolution. How these factors affect LD between a particular pair of loci or in a genomic region depends on local recombination rates. The population genetics theory of LD is well developed and is being widely used to provide insight into evolutionary history and as the basis for mapping genes in humans and in other species.

In this article, I will review the definitions of LD and the problems with assessing it, then outline the basic population genetics of LD that tells us how natural selection, genetic drift, recombination and mutation all affect levels of LD, and finally discuss some recent applications of LD to mapping genes, inferring the intensity of selection in the genome and estimating allele age.

Definitions

One pair of loci. LD between alleles at two loci has been defined in many ways (BOX 1), but all definitions depend on the quantity:

$$D_{AB} = p_{AB} - p_A p_B \qquad (1)$$

which is the difference between the frequency of gametes carrying the pair of alleles A and B at two loci (p_{AB}) and

Department of Integrative Biology, University of California, Berkeley, California 94720-3140, USA.
e-mail: slatkin@berkeley.edu
doi:10.1038/nrg2361
Published online
22 April 2008

Date received and published

Was the review article peer-reviewed?

FUNDAMENTAL CONCEPTS IN GENETICS

Linkage disequilibrium — understanding the evolutionary past and mapping the medical future

Montgomery Slatkin

Abstract | Linkage disequilibrium — the nonrandom association of alleles at different loci — is a sensitive indicator of the population genetic forces that structure a genome. Because of the explosive growth of methods for assessing genetic variation at a fine scale, evolutionary biologists and human geneticists are increasingly exploiting linkage disequilibrium in order to understand past evolutionary and demographic events, to map genes that are associated with quantitative characters and inherited diseases, and to understand the joint evolution of linked sets of genes. This article introduces linkage disequilibrium, reviews the population genetic processes that affect it and describes some of its uses. At present, linkage disequilibrium is used much more extensively in the study of humans than in non-humans, but that is changing as technological advances make extensive genomic studies feasible in other species.

Linkage disequilibrium (LD) is one of those unfortunate terms that does not reveal its meaning. As every instructor of population genetics knows, the term is a barrier not an aid to understanding. LD means simply a nonrandom association of alleles at two or more loci, and detecting LD does not ensure either linkage or a lack of equilibrium. The term was first used in 1960 by Lewontin and Kojima[1] and it persists because LD was initially the concern of population geneticists who were not picky about terminology as long as the mathematical definition was clear. At first, there were few data with which to study LD, and its importance to evolutionary biology and human genetics was unrecognized outside of population genetics. However, interest in LD grew rapidly in the 1980s once the usefulness of LD for gene mapping became evident and large-scale surveys of closely linked loci became feasible. By then, the term was too well established to be replaced.

LD is of importance in evolutionary biology and human genetics because so many factors affect it and are affected by it. LD provides information about past events and it constrains the potential response to both natural and artificial selection. LD throughout the genome reflects the population history, the breeding system [illegible] the pattern of geographic subdivision, whereas LD [illegible]each genomic region reflects the history of natural [illegible]ection, gene conversion, mutation and other forces that cause gene-frequency evolution. How these factors affect LD between a particular pair of loci or in a genomic region depends on local recombination rates. The population genetics theory of LD is well developed and is being widely used to provide insight into evolutionary history and as the basis for mapping genes in humans and in other species.

In this article, I will review the definitions of LD and the problems with assessing it, then outline the basic population genetics of LD that tells us how natural selection, genetic drift, recombination and mutation all affect levels of LD, and finally discuss some recent applications of LD to mapping genes, inferring the intensity of selection in the genome and estimating allele age.

Definitions

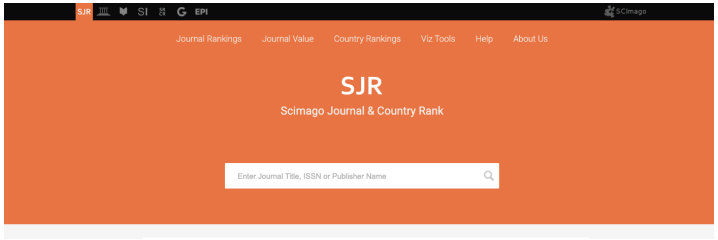
One pair of loci. LD between alleles at two loci has been defined in many ways (BOX 1), but all definitions depend on the quantity:

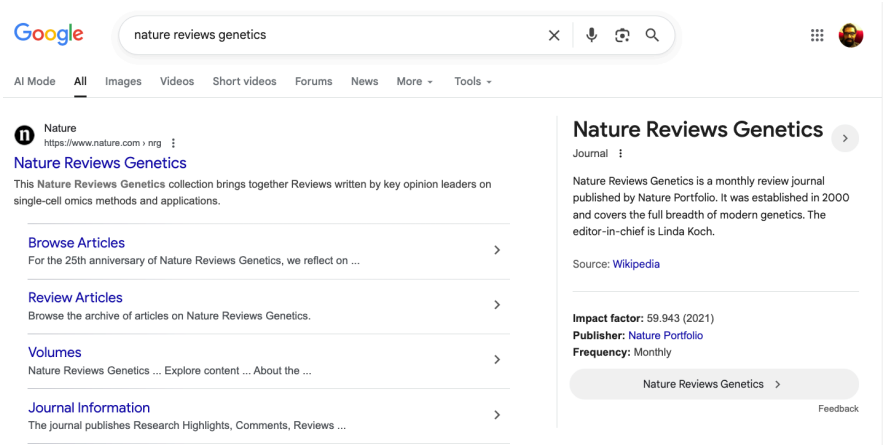
$$D_{AB} = p_{AB} - p_A p_B \qquad (1)$$

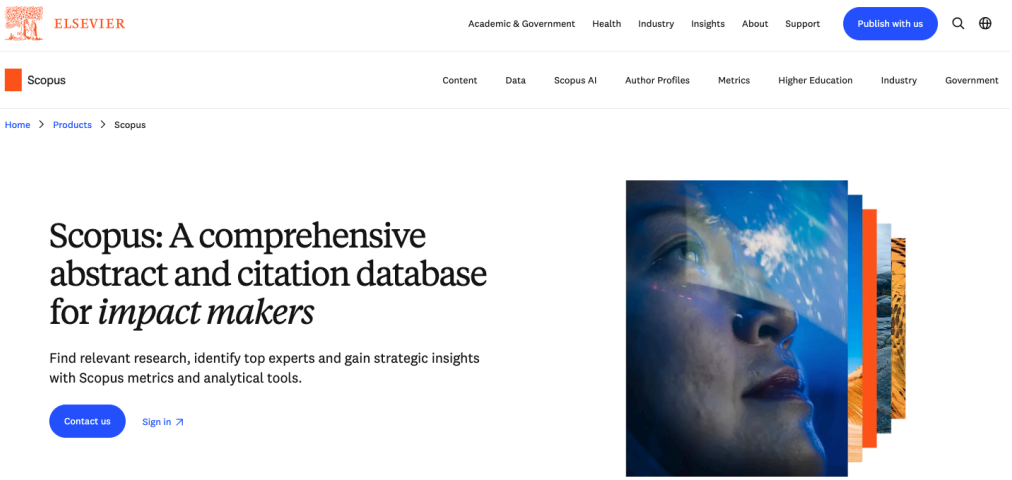
which is the difference between the frequency of gametes carrying the pair of alleles A and B at two loci (p_{AB}) and the product of the frequencies of those alleles (p_A and p_B). Originally, the definition was in terms of gamete frequencies because that allows for the possibility that the

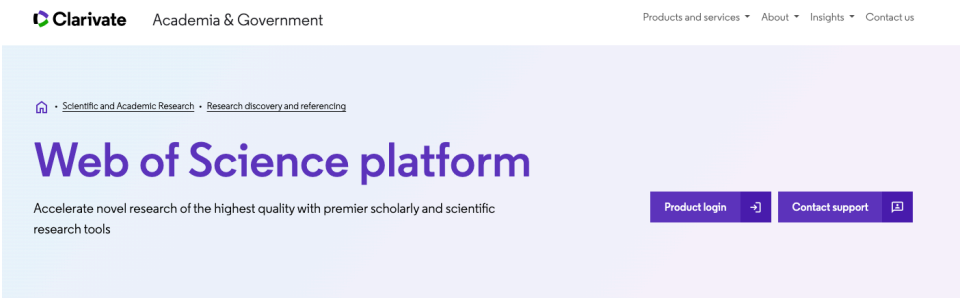
Department of Integrative Biology, University of California, Berkeley, California 94720-3140, USA.
e-mail: slatkin@berkeley.edu

How do I check if the journal is good?


Journal Rankings
Journal Value
Country Rankings
Viz Tools
Help
About Us
SJR
Scimago Journal & Country Rank
Enter Journal Title, ISSN or Publisher Name


Google
nature reviews genetics
AI Mode
All
Images
Videos
Short videos
Forums
News
More
Tools
Nature
https://www.nature.com › nrg
Nature Reviews Genetics
This Nature Reviews Genetics collection brings together Reviews written by key opinion leaders on single-cell omics methods and applications.
Browse Articles
For the 25th anniversary of Nature Reviews Genetics, we reflect on ...
Review Articles
Browse the archive of articles on Nature Reviews Genetics.
Volumes
Nature Reviews Genetics ... Explore content ... About the ...
Journal Information
The journal publishes Research Highlights, Comments, Reviews ...
Nature Reviews Genetics
Journal
Nature Reviews Genetics is a monthly review journal published by Nature Portfolio. It was established in 2000 and covers the full breadth of modern genetics. The editor-in-chief is Linda Koch.
Source: Wikipedia
Impact factor: 59.943 (2021)
Publisher: Nature Portfolio
Frequency: Monthly
Nature Reviews Genetics
Feedback


ELSEVIER
Academic & Government
Health
Industry
Insights
About
Support
Publish with us
Scopus
Content
Data
Scopus AI
Author Profiles
Metrics
Higher Education
Industry
Government
Home
Products
Scopus
Scopus: A comprehensive abstract and citation database for impact makers
Find relevant research, identify top experts and gain strategic insights with Scopus metrics and analytical tools.
Contact us
Sign in


Clarivate
Academia & Government
Products and services
About
Insights
Contact us
Scientific and Academic Research
Research discovery and referencing
Web of Science platform
Accelerate novel research of the highest quality with premier scholarly and scientific research tools
Product login
Contact support

FUNDAMENTAL CONCEPTS IN GENETICS

Linkage disequilibrium — understanding the evolutionary past and mapping the medical future

Montgomery Slatkin

Abstract | Linkage disequilibrium — the nonrandom association of alleles at different loci — is a sensitive indicator of the population genetic forces that structure a genome. Because of the explosive growth of methods for assessing genetic variation at a fine scale, evolutionary biologists and human geneticists are increasingly exploiting linkage disequilibrium in order to understand past evolutionary and demographic events, to map genes that are associated with quantitative characters and inherited diseases, and to understand the joint evolution of linked sets of genes. This article introduces linkage disequilibrium, reviews the population genetic processes that affect it and describes some of its uses. At present, linkage disequilibrium is used much more extensively in the study of humans than in non-humans, but that is changing as technological advances make extensive genomic studies feasible in other species.

Linkage disequilibrium (LD) is one of those unfortunate terms that does not reveal its meaning. As every instructor of population genetics knows, the term is a barrier not an aid to understanding. LD means simply a nonrandom association of alleles at two or more loci, and detecting LD does not ensure either linkage or a lack of equilibrium. The term was first used in 1960 by Lewontin and Kojima[1] and it persists because LD was initially the concern of population geneticists who were not picky about terminology as long as the mathematical definition was clear. At first, there were few data with which to study LD, and its importance to evolutionary biology and human genetics was unrecognized outside of population genetics. However, interest in LD grew rapidly in the 1980s once the usefulness of LD for gene mapping became evident and large-scale surveys of closely linked loci became feasible. By then, the term was too well established to be replaced.

LD is of importance in evolutionary biology and human genetics because so many factors affect it and are affected by it. LD provides information about past events and it constrains the potential response to both natural and artificial selection. LD throughout the genome reflects the population history, the breeding system and the pattern of g… in each genomic r… selection, gene cor… that cause gene-frequency evolution. How these factors affect LD between a particular pair of loci or in a genomic region depends on local recombination rates. The population genetics theory of LD is well developed and is being widely used to provide insight into evolutionary history and as the basis for mapping genes in humans and in other species.

In this article, I will review the definitions of LD and the problems with assessing it, then outline the basic population genetics of LD that tells us how natural selection, genetic drift, recombination and mutation all affect levels of LD, and finally discuss some recent applications of LD to mapping genes, inferring the intensity of selection in the genome and estimating allele age.

Definitions

One pair of loci. LD between alleles at two loci has been defined in many ways (BOX 1), but all definitions depend on the quantity:

$$D_{AB} = p_{AB} - p_A p_B \qquad (1)$$

which is the difference between the frequency of gametes carrying the pair of alleles A and B at two loci (p_{AB}) and the product of the frequencies of those alleles (p_A and p_B). Originally, the definition was in terms of gamete frequencies because that allows for the possibility that the

Department of Integrative Biology, University of California, Berkeley, California 94720-3140, USA.
e-mail: slatkin@berkeley.edu
doi:10.1038/nrg2361
Published online
22 April 2008

Is the publisher a credible one or predatory?

FUNDAMENTAL CONCEPTS IN GENETICS

Linkage disequilibrium — understanding the evolutionary past and mapping the medical future

Montgomery Slatkin

Abstract | Linkage disequilibrium — the nonrandom association of alleles at different loci — is a sensitive indicator of the population genetic forces that structure a genome. Because of the explosive growth of methods for assessing genetic variation at a fine scale, evolutionary biologists and human geneticists are increasingly exploiting linkage disequilibrium in order to understand past evolutionary and demographic events, to map genes that are associated with quantitative characters and inherited diseases, and to understand the joint evolution of linked sets of genes. This article introduces linkage disequilibrium, reviews the population genetic processes that affect it and describes some of its uses. At present, linkage disequilibrium is used much more extensively in the study of humans than in non-humans, but that is changing as technological advances make extensive genomic studies feasible in other species.

Linkage disequilibrium (LD) is one of those unfortunate terms that does not reveal its meaning. As every instructor of population genetics knows, the term is a barrier not an aid to understanding. LD means simply a nonrandom association of alleles at two or more loci, and detecting LD does not ensure either linkage or a lack of equilibrium. The term was first used in 1960 by Lewontin and Kojima[1] and it persists because LD was initially the concern of population geneticists who were not picky about terminology as long as the mathematical definition was clear. At first, there were few data with which to study LD, and its importance to evolutionary biology and human genetics was unrecognized outside of population genetics. However, interest in LD grew rapidly in the 1980s once the usefulness of LD for gene mapping became evident and large-scale surveys of closely linked loci became feasible. By then, the term was too well established to be replaced.

LD is of importance in evolutionary biology and human genetics because so many factors affect it and are affected by it. LD provides information about past events and it constrains the potential response to both natural and artificial selection. LD throughout the genome reflects the population history, the breeding system and the pattern of geographic subdivision, whereas LD in each genomic region reflects the history of natural selection, gene conversion, mutation and other forces that cause gene-frequency evolution. How these factors affect LD between a particular pair of loci or in a genomic region depends on local recombination rates. The population genetics theory of LD is well developed and is being widely used to provide insight into evolutionary history and as the basis for mapping genes in humans and in other species.

In this article, I will review the definitions of LD and the problems with assessing it, then outline the basic population genetics of LD that tells us how natural selection, genetic drift, recombination and mutation all affect levels of LD, and finally discuss some recent applications of LD to mapping genes, inferring the intensity of selection in the genome and estimating allele age.

Definitions

One pair of loci. LD between alleles at two loci has been defined in many ways (BOX 1), but all definitions depend on the quantity:

$$D_{AB} = p_{AB} - p_A p_B \qquad (1)$$

which is the difference between the frequency of gametes carrying the pair of alleles A and B at two loci (p_{AB}) and th[illegible]

Department of Integrative Biology, University of California, Berkeley, California 94720-3140, USA.
e-mail: slatkin@berkeley.edu
doi:10.1038/nrg2361
Published online
22 April 2008

Volume and/or issue

FUNDAMENTAL CONCEPTS IN GENETICS

Linkage disequilibrium — understanding the evolutionary past and mapping the medical future

Montgomery Slatkin

Abstract | Linkage disequilibrium — the nonrandom association of alleles at different loci — is a sensitive indicator of the population genetic forces that structure a genome. Because of the explosive growth of methods for assessing genetic variation at a fine scale, evolutionary biologists and human geneticists are increasingly exploiting linkage disequilibrium in order to understand past evolutionary and demographic events, to map genes that are associated with quantitative characters and inherited diseases, and to understand the joint evolution of linked sets of genes. This article introduces linkage disequilibrium, reviews the population genetic processes that affect it and describes some of its uses. At present, linkage disequilibrium is used much more extensively in the study of humans than in non-humans, but that is changing as technological advances make extensive genomic studies feasible in other species.

Abstract

Linkage disequilibrium (LD) is one of those unfortunate terms that does not reveal its meaning. As every instructor of population genetics knows, the term is a barrier not an aid to understanding. LD means simply a nonrandom association of alleles at two or more loci, and detecting LD does not ensure either linkage or a lack of equilibrium. The term was first used in 1960 by Lewontin and Kojima[1] and it persists because LD was initially the concern of population geneticists who were not picky about terminology as long as the mathematical definition was clear. At first, there were few data with which to study LD, and its importance to evolutionary biology and human genetics was unrecognized outside of population genetics. However, interest in LD grew rapidly in the 1980s once the usefulness of LD for gene mapping became evident and large-scale surveys of closely linked loci became feasible. By then, the term was too well established to be replaced.

LD is of importance in evolutionary biology and human genetics because so many factors affect it and are affected by it. LD provides information about past events and it constrains the potential response to both natural and artificial selection. LD throughout the genome reflects the population history, the breeding system and the pattern of geographic subdivision, whereas LD in each genomic region reflects the history of natural selection, gene conversion, mutation and other forces that cause gene-frequency evolution. How these factors affect LD between a particular pair of loci or in a genomic region depends on local recombination rates. The population genetics theory of LD is well developed and is being widely used to provide insight into evolutionary history and as the basis for mapping genes in humans and in other species.

In this article, I will review the definitions of LD and the problems with assessing it, then outline the basic population genetics of LD that tells us how natural selection, genetic drift, recombination and mutation all affect levels of LD, and finally discuss some recent applications of LD to mapping genes, inferring the intensity of selection in the genome and estimating allele age.

Definitions

One pair of loci. LD between alleles at two loci has been defined in many ways (BOX 1), but all definitions depend on the quantity:

$$D_{AB} = p_{AB} - p_A p_B \qquad (1)$$

which is the difference between the frequency of gametes carrying the pair of alleles A and B at two loci (p_{AB}) and the product of the frequencies of those alleles (p_A and p_B). Originally, the definition was in terms of gamete frequencies because that allows for the possibility that the

Department of Integrative Biology, University of California, Berkeley, California 94720-3140, USA.
e-mail: slatkin@berkeley.edu
doi:10.1038/nrg2361
Published online
22 April 2008

FUNDAMENTAL CONCEPTS IN GENETICS

Linkage disequilibrium — understanding the evolutionary past and mapping the medical future

Montgomery Slatkin

Abstract | Linkage disequilibrium — the nonrandom association of alleles at different loci — is a sensitive indicator of the population genetic forces that structure a genome. Because of the explosive growth of methods for assessing genetic variation at a fine scale, evolutionary biologists and human geneticists are increasingly exploiting linkage disequilibrium in order to understand past evolutionary and demographic events, to map genes that are associated with quantitative characters and inherited diseases, and to understand the joint evolution of linked sets of genes. This article introduces linkage disequilibrium, reviews the population genetic processes that affect it and describes some of its uses. At present, linkage disequilibrium is used much more extensively in the study of humans than in non-humans, but that is changing as technological advances make extensive genomic studies feasible in other species.

Abstract

Linkage disequilibrium (LD) is one of those unfortunate terms that does not reveal its meaning. As every instructor of population genetics knows, the term is a barrier not an aid to understanding. LD means simply a nonrandom association of alleles at two or more loci, and detecting LD does not ensure either linkage or a lack of equilibrium. The term was first used in 1960 by Lewontin and Kojima[1] and it persists because LD was initially the concern of population geneticists who were not picky about terminology as long as the mathematical definition was clear. At first, there were few data with which to study LD, and its importance to evolutionary biology and human genetics was unrecognized outside of population genetics. However, interest in LD grew rapidly in the 1980s once the usefulness of LD for gene mapping became evident and large-scale surveys of closely linked loci became feasible. By then, the term was too well established to be replaced.

LD is of importance in evolutionary biology and human genetics because so many factors affect it and are affected by it. LD provides information about past events and it constrains the potential response to both natural and artificial selection. LD throughout the genome reflects the population history, the breeding system and the pattern of geographic subdivision, whereas LD in each genomic region reflects the history of natural selection, gene conversion, mutation and other forces that cause gene-frequency evolution. How these factors affect LD between a particular pair of loci or in a genomic region depends on local recombination rates. The population genetics theory of LD is well developed and is being widely used to provide insight into evolutionary history and as the basis for mapping genes in humans and in other species.

In this article, I will review the definitions of LD and the problems with assessing it, then outline the basic population genetics of LD that tells us how natural selection, genetic drift, recombination and mutation all affect levels of LD, and finally discuss some recent applications of LD to mapping genes, inferring the intensity of selection in the genome and estimating allele age.

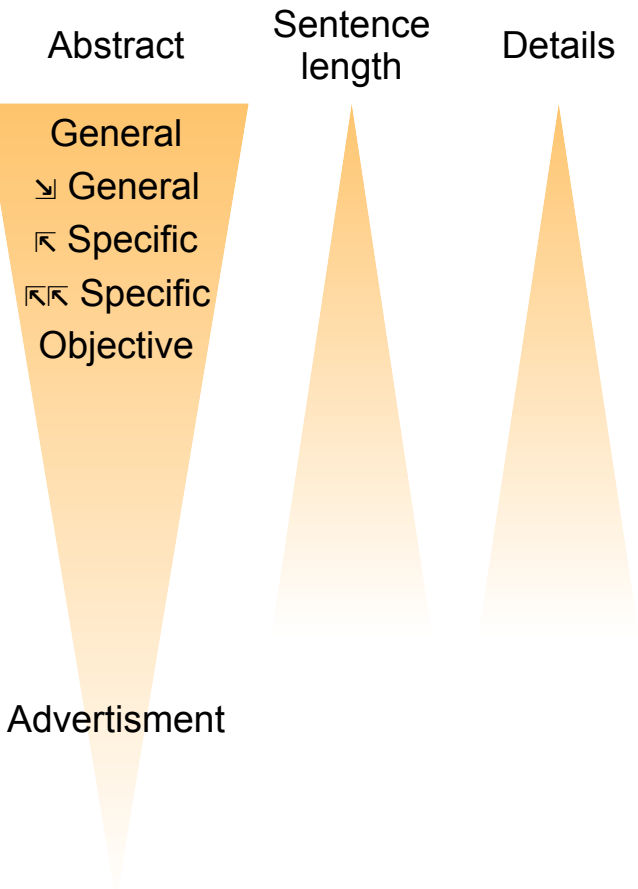
Definitions

One pair of loci. LD between alleles at two loci has been defined in many ways (BOX 1), but all definitions depend on the quantity:

$$D_{AB} = p_{AB} - p_A p_B \qquad (1)$$

which is the difference between the frequency of gametes carrying the pair of alleles A and B at two loci (p_{AB}) and the product of the frequencies of those alleles (p_A and p_B). Originally, the definition was in terms of gamete frequencies because that allows for the possibility that the

Department of Integrative Biology, University of California, Berkeley, California 94720-3140, USA.
e-mail: slatkin@berkeley.edu
doi:10.1038/nrg2361
Published online
22 April 2008



FUNDAMENTAL CONCEPTS IN GENETICS

Linkage disequilibrium — understanding the evolutionary past and mapping the medical future

Montgomery Slatkin

Abstract | Linkage disequilibrium — the nonrandom association of alleles at different loci — is a sensitive indicator of the population genetic forces that structure a genome. Because of the explosive growth of methods for assessing genetic variation at a fine scale, evolutionary biologists and human geneticists are increasingly exploiting linkage disequilibrium in order to understand past evolutionary and demographic events, to map genes that are associated with quantitative characters and inherited diseases, and to understand the joint evolution of linked sets of genes. This article introduces linkage disequilibrium, reviews the population genetic processes that affect it and describes some of its uses. At present, linkage disequilibrium is used much more extensively in the study of humans than in non-humans, but that is changing as technological advances make extensive genomic studies feasible in other species.

Introduction

Linkage disequilibrium (LD) is one of those unfortunate terms that does not reveal its meaning. As every instructor of population genetics knows, the term is a barrier not an aid to understanding. LD means simply a nonrandom association of alleles at two or more loci, and detecting LD does not ensure either linkage or a lack of equilibrium. The term was first used in 1960 by Lewontin and Kojima[1] and it persists because LD was initially the concern of population geneticists who were not picky about terminology as long as the mathematical definition was clear. At first, there were few data with which to study LD, and its importance to evolutionary biology and human genetics was unrecognized outside of population genetics. However, interest in LD grew rapidly in the 1980s once the usefulness of LD for gene mapping became evident and large-scale surveys of closely linked loci became feasible. By then, the term was too well established to be replaced.

LD is of importance in evolutionary biology and human genetics because so many factors affect it and are affected by it. LD provides information about past events and it constrains the potential response to both natural and artificial selection. LD throughout the genome reflects the population history, the breeding system and the pattern of geographic subdivision, whereas LD in each genomic region reflects the history of natural selection, gene conversion, mutation and other forces that cause gene-frequency evolution. How these factors affect LD between a particular pair of loci or in a genomic region depends on local recombination rates. The population genetics theory of LD is well developed and is being widely used to provide insight into evolutionary history and as the basis for mapping genes in humans and in other species.

In this article, I will review the definitions of LD and the problems with assessing it, then outline the basic population genetics of LD that tells us how natural selection, genetic drift, recombination and mutation all affect levels of LD, and finally discuss some recent applications of LD to mapping genes, inferring the intensity of selection in the genome and estimating allele age.

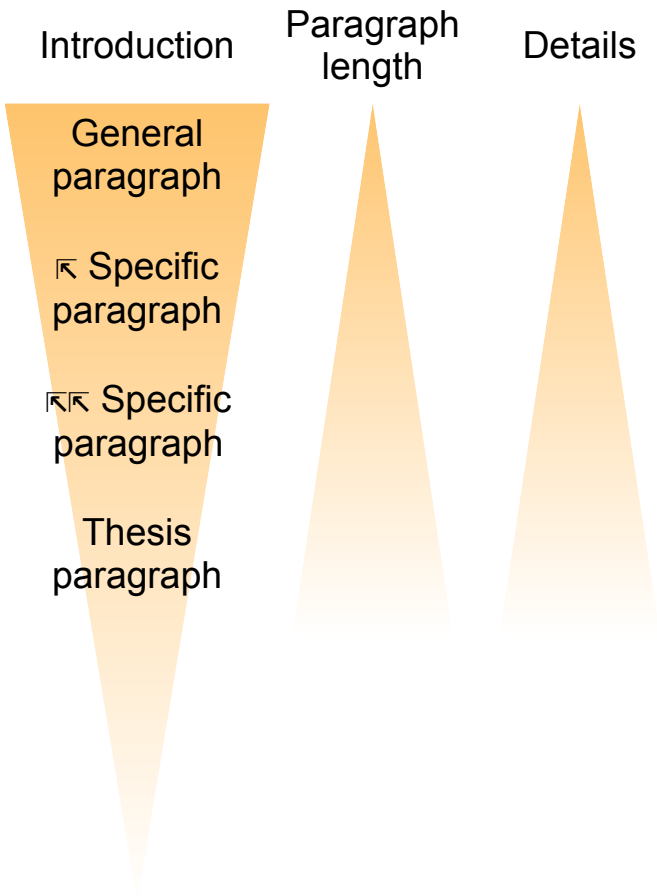
Definitions

One pair of loci. LD between alleles at two loci has been defined in many ways (BOX 1), but all definitions depend on the quantity:

$$D_{AB} = p_{AB} - p_A p_B \qquad (1)$$

which is the difference between the frequency of gametes carrying the pair of alleles A and B at two loci (p_{AB}) and the product of the frequencies of those alleles (p_A and p_B). Originally, the definition was in terms of gamete frequencies because that allows for the possibility that the

Department of Integrative Biology, University of California, Berkeley, California 94720-3140, USA.
e-mail: slatkin@berkeley.edu
doi:10.1038/nrg2361
Published online
22 April 2008



FUNDAMENTAL CONCEPTS IN GENETICS

Linkage disequilibrium — understanding the evolutionary past and mapping the medical future

Montgomery Slatkin

Abstract | Linkage disequilibrium — the nonrandom association of alleles at different loci — is a sensitive indicator of the population genetic forces that structure a genome. Because of the explosive growth of methods for assessing genetic variation at a fine scale, evolutionary biologists and human geneticists are increasingly exploiting linkage disequilibrium in order to understand past evolutionary and demographic events, to map genes that are associated with quantitative characters and inherited diseases, and to understand the joint evolution of linked sets of genes. This article introduces linkage disequilibrium, reviews the population genetic processes that affect it and describes some of its uses. At present, linkage disequilibrium is used much more extensively in the study of humans than in non-humans, but that is changing as technological advances make extensive genomic studies feasible in other species.

Linkage disequilibrium (LD) is one of those unfortunate terms that does not reveal its meaning. As every instructor of population genetics knows, the term is a barrier not an aid to understanding. LD means simply a nonrandom association of alleles at two or more loci, and detecting LD does not ensure either linkage or a lack of equilibrium. The term was first used in 1960 by Lewontin and Kojima[1] and it persists because LD was initially the concern of population geneticists who were not picky about terminology as long as the mathematical definition was clear. At first, there were few da… which to study LD, and its importance to evolut… biology and human genetics was unrecognized … of population genetics. However, interest in LI… rapidly in the 1980s once the usefulness of LD for gene mapping became evident and large-scale surveys of closely linked loci became feasible. By then, the term was too well established to be replaced.

LD is of importance in evolutionary … human genetics because so many factors aff… affected by it. LD provides information abou… and it constrains the potential response t… ral and artificial selection. LD throughout the genome reflects the population history, the breeding system and the pattern of geographic subdivision, whereas LD in each genomic region reflects the history of natural selection, gene conversion, mutation and other forces that cause gene-frequency evolution. How these factors affect LD between a particular pair of loci or in a genomic region depends on local recombination rates. The population genetics theory of LD is well developed and is being widely used to provide insight into evolutionary history and as the basis for mapping genes in humans and in other species.

In this article, I will review the definitions of LD and the problems with assessing it, then outline the basic population genetics of LD that tells us how natural … combination and mutation all … ally discuss some recent appli… genes, inferring the intensity of … nd estimating allele age.

Section

Definitions

Subsection

One pair of loci. LD between alleles at two loci has been defined in many ways (BOX 1), but all definitions depend …

$$D = p_{AB} - p_A p_B \quad (1)$$

which is the difference between the frequency of gametes carrying the pair of alleles A and B at two loci (p_{AB}) and the product of the frequencies of those alleles (p_A and p_B). Originally, the definition was in terms of gamete frequencies because that allows for the possibility that the

Department of Integrative Biology, University of California, Berkeley, California 94720-3140, USA.
e-mail: slatkin@berkeley.edu
doi:10.1038/nrg2361
Published online
22 April 2008

Box 1 | **Definitions of LD**

Different definitions of linkage disequilibrium (LD) have been proposed because they capture different features of nonrandom association. All of them are related to *D*, which is defined in equation 1 in the text. Although *D* completely characterizes the extent to which two alleles, A and B, are nonrandomly associated, it is often not the best statistic to use when comparing LD at different pairs of loci because the range of possible values of *D* for each pair is constrained by the allele frequencies. The

Box for additional technical material or examples

the presence of A and B. In general, r^2 is similar to D' in that it can be nearly one even if one or both alleles are in low frequency.

Still another measure is $\delta_{A|}$, defined to be $p_A + D/p_B$. It is the conditional probability that a chromosome carries an A allele, given that it carries a B allele. It is useful for characterizing the extent to which a particular allele is associated with a genetic diease[120].

loci are on different chromosomes. The usual application now is to loci on the same chromosome, in which case the allele pair AB is called a haplotype and p_{AB} is the haplotype frequency. As defined, D_{AB} characterizes a population; in practice, D_{AB} is estimated from allele and haplotype frequencies in a sample. Standard sampling theory has to be applied to find the confidence intervals of estimated values[2].

The quantity D_{AB} is the coefficient of linkage disequilibrium. It is defined for a specific pair of alleles, A and B, and does not depend on how many other alleles are at the two loci — each pair of alleles has its own *D*. The values for different pairs of alleles are constrained by the fact that the allele frequencies at both loci and the haplotype frequencies have to add up to 1. If both loci are diallelic, as is the case with virtually all SNPs, the constraint is strong enough that only one value of *D* is needed to charac LD between those loci. In fact, $D_{AB} = -D_{Ab} = -D_{aB}$ where a and b are the other alleles. In this case, th used without a subscript. The sign of *D* is arbitrar depends on which pair of alleles one starts with.

If either locus has more than two alleles, no single statistic quantifies the overall LD between them. Although several have been suggested[3,4], none has gained wide acceptance. Such a statistic is needed when both loci have numerous alleles, as is the case for many loci in the major histocompatibility complex in vertebrates, which have dozens or even hundreds of alleles, or for microsatellite loci, which often have 10 to 20 alleles. If there is no one pair of alleles of particular interest, the question D between one pair of D between a pair of loci

Subsection

Linkage equilibrium. If $D = 0$ there is linkage equilibrium (LE), which has similarities to the Hardy–Weinberg equilibrium (HWE) in implying statistical independence. When genotypes at a single locus are at HWE, whether an allele is present on one chromosome is independent of whether it is present on the homologue. Consequently, the frequency of the AA homozygote is the square of the frequency of A ($p_{AA} = p_A^2$) and the frequency of the Aa heterozygote is twice the product of p_A and p_a, the two being necessary to allow for both Aa and aA. The essential feature of HWE is that, regardless of the initial genotype frequencies, HWE is established in one generation of random mating. Any initial deviation from HWE disappears immediately. Significant departures from HWE indicate something interesting is going on, for example, extensive inbreeding, strong selection or genotyping error.

LE is similar to HWE because it implies that alleles at different loci are randomly associated. The frequency of the AB haplotype is the product of the allele frequencies ($p_A p_B$). LE differs from HWE, however, because it is not established in one generation of random mating. Instead, *D* decreases at a rate that depends on the recombination frequency, *c*, between the two loci:

$$D_{AB}(t+1) = (1-c)\,D_{AB}(t) \qquad (2)$$

where *t* is time in generations. Even for unlinked loci ($c = 0.5$), *D* decreases only by a factor of a half each generation, something proved by Weinberg[7] in 1909. The general formula was obtained first by Jennings[8].

Although LE will eventually be reached, it will occur slowly for closely linked loci. That is the basis for the uses of LD discussed in later sections. Other population genetic forces, including selection, gene flow, genetic drift and mutation, all affect *D*, so substantial LD will persist under many conditions. Now that very large numbers of polymorphic loci can be surveyed, the extent of LD in a genome can be quantified with great precision, allowing a fine-scale analysis of forces governing genomic variation.

The coefficient of LD and related quantities are descriptive statistics. Their magnitude does not indicate tistically significant associa- lotypes. Standard statistical ared and Fisher's exact test, for significance[2].

Section

Haplotype phase

D and related statistics implicitly assume that haploid individuals or gametes can be typed. But often, only diploid genotypes and not haplotypes can be determined. That is the case with all SNP surveys, other than those of the X chromosome in males (assuming males are the heterogametic sex) or when haploids can be typed. The problem is sketched in BOX 2. The extent of LD in genotypic data[2,9] can be quantified, but the lack of information about the haplotype phase weakens the signal of nonrandom association sufficiently that this approach is not often taken. It is more common to use a statistical method based on population genetics theory (BOX 2) to infer haplotype phase from genotypic data and then to treat the inferred haplotypes as if they were data.

Box 2 | **Genotype data and haplotype phase**

When the genotype of a dipoid individual is determined, the result is a list of genotypes for each locus surveyed. If three diallelic loci are surveyed, the genotypes of four individuals might be AA bb CC, Aa BB cc, aa Bb Cc and Aa Bb Cc. The haplotypes of the first two individuals are immediately apparent. Individual 1 has two copies of AbC and individual 2 has ABc and aBc. There is no uncertainty if no more than one locus is heterozygous. Otherwise, haplotypes cannot be determined without further

Box for additional technical material or examples

used in large genomic surveys.

It is much more common to use a statistical method based on the assumption that haplotypes are randomly joined into genotypes. The basic idea is that individuals that are homozygous at all loci or all but one locus provide some information about haplotype frequencies that can then be used to infer the haplotype phase of the other individuals. Various methods — including those based on maximum likelihood[122], parsimony[123], combinatorial theory[124] and *a priori* distribution derived from coalescent theory[125] — have been developed. The last method is the basis for the program PHASE, which has performed the best in extensive simulation studies[126]. The emerging view of this problem is that inferring haplotype phase is similar to other cases in which missing data (in this case the haplotype phase of a diploid genotype) has to be imputed[127].

Although this procedure is intuitively appealing and usually leads to reasonable results, especially for common haplotypes, it ignores the uncertainty that is inherent in the inference step and that might be important in some cases. Often, genotypes can be resolved into several possible haplotypes, and inferred frequencies of rare haplotypes can be quite wrong[10]. It is preferable, although sometimes difficult, to use methods that account for the

Section

otype frequencies, as is

Metropolis algorithm

ARC)[11] and some other

LD at more than two loci

Subsection

When more than two loci are considered toget

common practice is to distinguish graphically those pairs that have high levels of LD from those that do not[3]. The result is a graph of the type introduced by Miyashita and Langley[12] to describe patterns of LD in *Drosophila melanogaster*. A more recent example is shown in FIG. 1. This figure indicates that a 216 kb segment in the class II region of the major histocompatibility complex in humans is made up of non-overlapping sets of loci in strong LD with each other. Each group is called a 'haplotype block' and boundaries were shown to be associated with hot spots of recombination. Similar patterns were found in other genomic regions in humans[13,14], leading to the hypothesis that most of the human genome had a block-like pattern of LD. Haplotype blocks in humans vary in size from a few kb to more than 100 kb[15].

Haplotype blocks were a surprising discovery that was of great practical importance for the mapping of inherited diseases. Before their discovery, the prevailing view of LD in humans was represented by results from the simulation study of Kruglyak[16], which showed that, under assumptions that were intended to approximate the history of modern humans, little LD would be expected beyond 3 kb. The discovery of haplotype blocks showed that LD usually extended over much longer chromosomal distances and suggested that testing one SNP within each block for significant association with a disease might be sufficient to indicate association with every SNP in that block, thus reducing the number of SNPs that need to be tested in case–control studies of disease association[17]. The situation turned out to be more complex both because some genomic regions were found to not have a block-like structure[18] and because different ways of defining haplotype blocks resulted in different block boundaries[19]. Nevertheless, the observation that LD in humans extended over relatively large chromosomal distances provided a major part of the impetus for the International HapMap Project, which in its first generation identified over 1 million SNPs in humans and characterized the LD in 269 individuals in four ethnically different populations (European, Han Chinese, Japanese and Yoruban)[20]. The second generation HapMap published recently characterized 3.1 million SNPs in the same group of individuals[21].

Haplotype blocks vary somewhat among human populations — they tend to be somewhat shorter in African populations[15,20,21]. Haplotype blocks have been studied in other species as well, both model organisms, including the mouse and rat[22], and domesticated species, including cows[23] and dogs[24]. The isolation of strains and breeds in these species results in much longer block lengths than are found in humans.

Variance in heterozygosity. A simple and often useful statistic describing the overall extent of LD in a genomic region is the variance in heterozygosity across loci, which increases as a linear function of

when the density

goal is to obtain

recombination.

tatistic to assess

various pathogenic bacteria.

Higher-order disequilibria. When considering more than two loci, equation 1 can be generalized to define higher-order coefficients of LD. For alleles at three loci (A, B, and C) the third-order coefficient is:

$$D_{ABC} = p_{ABC} - p_A D_{BC} - p_B D_{AC} - p_C D_{AB} - p_A p_B p_C \qquad (3)$$

where D_{AB}, D_{BC} and D_{AC} are the pairwise disequilibrium coefficients. D_{ABC} is analogous to the three-way interaction term in an analysis of variance and can be interpreted as the non-independence among these alleles that is not accounted for by the pairwise

coefficients. The decay of these higher-order coefficients under random mating was studied by Geiringer[27] and has been worked out in some detail by later authors. Little practical use of these higher-order coefficients has been made, other than in the analysis of variation of human leukocyte antigen loci in humans, which suggested that two loci that are closely linked to a selected locus would display unusual patterns of LD[28]. It is worth considering whether higher-order disequilibrium coefficients can help to understand the patterns found in the HapMap and other large data sets.

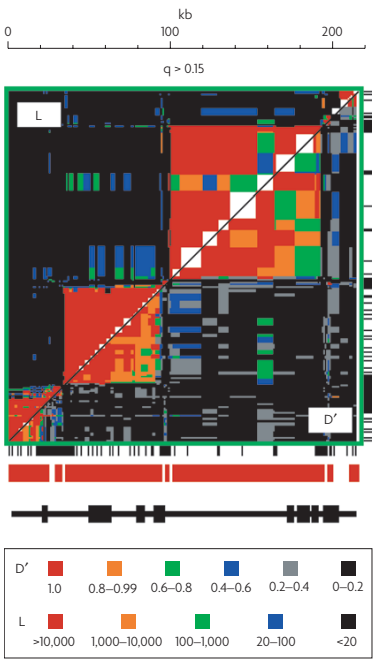


Figure 1 | **Haplotype blocks.** This graph provides some of the first evidence of haplotype blocks and their association with recombination hot spots. The figure shows the pattern of pairwise linkage disequilib… f the class II region of t… complex in humans fo… n for which the frequency… ommon) allele exceeded… agonal shows levels of L… s the null hypothesi… the test of linkage eq… (a measure of LD) to its maximum possible absolute value, given the allele frequencies. This figure is reproduced, with permission, from *Nature Genetics* REF. 130 © (2001) Macmillan Publishers Ltd.

LD within and between populations

When data for more than one population are available, LD between a pair of loci can be partitioned into contributions within and between populations. This partitioning, first suggested by Ohta[29,30], is similar to Wright's[31] partitioning of deviations from HWE frequencies into F_{IS}, the average deviation within populations, and F_{ST}, the average deviation that is attributable to differences in allele frequency among populations[31]. Ohta[30] partitioned D_T, the total disequilibrium in a subdivided population, into D_{IS}, the average disequilibrium within subpopulations, and D_{ST}, the contribution to the overall disequilibrium caused by differences in allele frequencies among subpopulations. Computer programs such as Genepop[32] are available to calculate D_{IS} and D_{ST}.

These statistics are used widely in the analysis of data from non-human populations but only rarely for human populations, probably because the focus in humans is on each population whereas the focus in other species is often on the overall pattern of LD. Natural selection favouring adaptations to local conditions will increase D_{ST} whenever alleles at different loci are favoured. Partitioning overall LD is an appropriate first step when trying to determine whether differences in LD result only from differences in allele frequency or from other factors that vary among populati…

Bypassing LD

Both D and measur… statistics that quantify deviations from random association of alleles. The statistics themselves provide no information about why alleles at different loci are nonrandomly associated. There is no agreement about which is the best or most useful statistic[2,33,34], in part because different statistics are sensitive to different population genetic processes that can cause nonrandom association. An alternative to using one or even several statistics is to ignore the coefficient of LD altogether and estimate parameters of the population genetic models as discussed in the following sections. Several methods to estimate recombination rates directly from haplotypes have been proposed[35–39] and they have been used successfully to estimate local rates of recombination in the human genome and to identify DNA sequence motifs associated with hot spots of recombination[40].

Bypassing descriptive statistics has the advantage of not having to decide which statistic best captures the underlying signal that is sought, but it has disadvantage of not providing a summary of the data independently of the model used.

Population genetics of LD

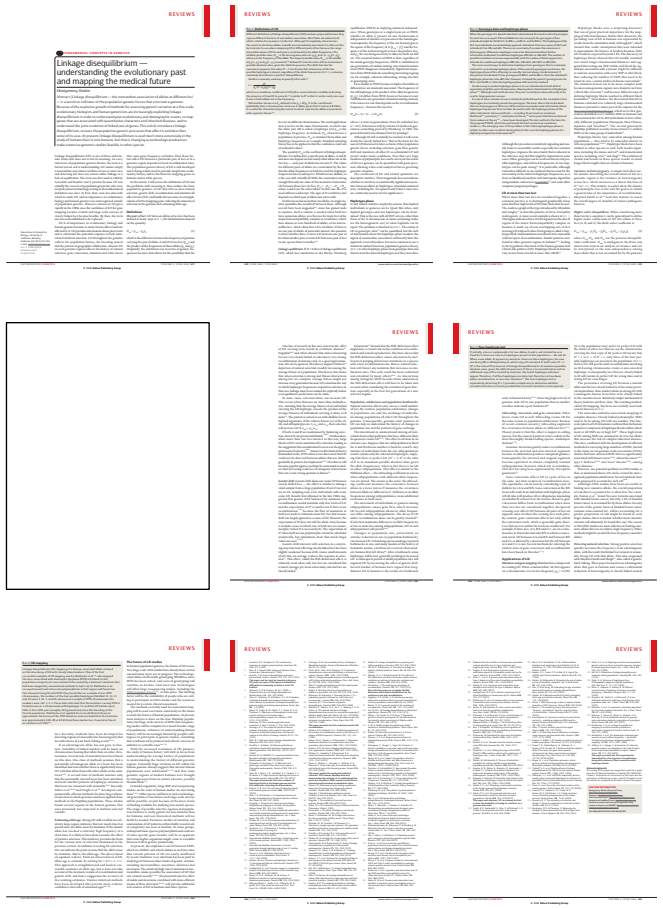
Natural selection. Initial interest in LD arose from questions about the operation of natural selection. If alleles at two loci are in LD and they both affect reproductive fitness, the response to selection on one locus might be accelerated or impeded by selection affecting the other.

Figure

Figure title

Figure details

Section



One line of research in this area concerns the effect of LD on long-term trends in evolution. Kimura[41], Nagylaki[42,43] and others showed that unless interacting loci are very closely linked or selection is very strong, recombination dominates and, to a good approximation, LD can be ignored. This theory supports Fisher's[44] depiction of natural selection steadily increasing the average fitness of a population. This theory also shows that when selection is strong and fitness interactions among loci are complex, average fitness might not increase every generation because LD constrains the way in which haplotype frequencies respond to selection. In that case, linkage must be accounted for explicitly before even qualitative predictions can be made.

In some cases, selection alone can increase LD. This occurs when fitnesses are more than multiplicative, meaning that the average fitness of an individual carrying the AB haplotype exceeds the product of the average fitnesses of individuals carrying A alone or B alone[45]. The pattern is easiest to see with diallelic loci in haploid organisms. If the relative fitness (w) of the ab, Ab, and aB haplotypes are 1, w_{Ab} and w_{aB}, then selection will increase LD if $w_{AB} > w_{Ab}w_{aB}$.

If both A and B are maintained by balancing selection, then LD can persist indefinitely[1,46,47]. Furthermore, when more than two loci interact in this way, large blocks of LD can be maintained b[illegible]se that one subpopulation is fixed the suggestion that an individual [illegible] another is fixed for a and b. Any priate unit of selection[48,49]. Interes[illegible]als from the two subpopulations diminished in the 1970s when it v[illegible]he AB and ab haplotypes, implying that there is perfect LD ($D' = 1$; D' is the ratio of D to its maximum possible absolute value, given the allele frequencies), when in fact there is no LD in either subpopulation. This effect is similar to the Wahlund effect — the inbreeding coefficient at a locus when subpopulations with different allele frequencies are mixed. The reason is the same: the inbreeding coefficient measures the covariance between alleles at a locus just as D measures the covariance between alleles at different loci[2]. Differences in allele frequencies among subpopulations create additional covariance in both cases.

Subsection

Felsenstein[59] showed that the Hill–Robertson effect might have a crucial role in the evolution of recombination and sexual reproduction. The basic idea is that the Hill–Robertson effect causes selection to be inefficient in purging deleterious mutations in a species with a low recombination rate. Hence, natural selection will favour any mutation that increases recombination rates. This early result has been confirmed and extended by many others[60,61]. As interactions among intragenic SNPs become better understood, the Hill–Robertson effect will have to be taken into account when considering the evolution of gene function, especially in the first few generations of a new selective regime.

Population subdivision and population bottlenecks. Natural selection affects only one or a small number of loci. By contrast, population subdivision, changes in population size and the exchange of individuals among populations all affect LD throughout the genome. Consequently, genome-wide patterns of LD can help us understand the history of changes in population size and the patterns of gene exchange.

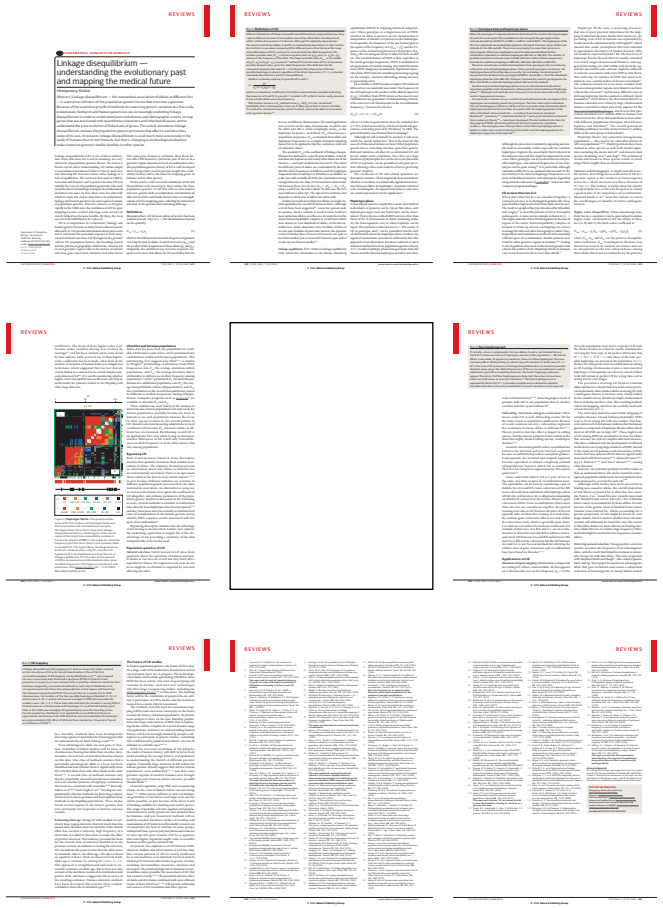
The intentional or unintentional mixing of individuals from subpopulations that have different allele frequencies creates LD[62,63]. The effect is obvious in an

The movement of individuals or gametes among subpopulations causes gene flow, which increases LD in each subpopulation whenever allele frequencies differ among subpopulations. The decay of LD under recombination alone can be greatly retarded[62]. If selection maintains differences in allele frequencies at two or more loci among subpopulations, LD in each subpopulation will persist[64,65].

Changes in population size, particularly an extreme reduction in size (a population bottleneck), can increase LD. Colonizing species undergo repeated bottlenecks in size, and many models of the history of hominids assume a bottleneck occurred when modern humans first left Africa[66]. After a bottleneck, some haplotypes will be lost, generally resulting in increased LD. A subsequent period of small population size will augment LD by increasing the effect of genetic drift. Several studies of humans have argued that long-distance LD in humans is the results of a bottleneck

Genetic drift. Genetic drift alone can create LD between closely linked loci — the effect is similar to taking a small sample from a large population. Even if two loci are in LE, sampling only a few individuals will create some LD. Results first obtained in the late 1960s suggested that genetic drift balanced by mutation and recombination would maintain only low levels of LD, and the expectation of D^2 is small even if there is no recombination[53,54] because the flux of mutations at both loci tends to eliminate most LD. For that reason, drift was largely ignored as a cause of LD. However, the expectation of D^2 does not tell the whole story because it includes cases in which one or both loci are monomorphic (when D is necessarily 0). The expectation of D^2 when both loci are polymorphic cannot be calculated analytically, but simulations show that much larger values are seen[35,55,56].

Genetic drift interacts with selection in a surprising way. Selection affecting closely linked loci becomes slightly weakened because drift creates small amounts of LD that, on average, reduces the response to selection[57]. This effect, called the Hill–Robertson effect, is relatively weak when only two loci are considered but is much stronger per locus when many selected loci are closely linked[58].



Box 3 | **Four-haplotype test**

Box for additional technical material or examples

early in human history[67–69]. Detecting higher levels of genome-wide LD in one population than in another can then indicate a past bottleneck[68].

Inbreeding, inversions and gene conversion. Other forces create LD as well. Inbreeding creates LD for [illegible] Because [illegible] segments [illegible] loci[70,71]. [illegible] selfing species, but the expected pattern is not evident in the most thoroughly studied selfing species, *Arabidopsis thaliana*[72,73].

Genomic inversions greatly reduce recombination between the inverted and non-inverted segments because recombination produces aneuploid gametes. Consequently, the inverted and original segments become equivalent to almost completely isolated subpopulations between which LD accumulates. This fact has long been appreciated by *Drosophila* geneticists[50].

Gene conversion affects LD at a pair of loci in the same way that reciprocal recombination does. The equivalence can be seen by considering a pair of diallelic loci A/a and B/b. Gene conversion at the B/b locus will result in an individual with haplotype phase AB/ab who will produce Ab or aB gametes depending on whether B converts b or the reverse. However, gene conversion differs from recombination when more than two loci are considered together. Reciprocal crossing over affects LD between all pairs of loci on opposite sides of where the crossing over took place. By contrast, gene conversion affects loci only within the conversion track, which is generally quite short. Loci that are not within the track are unaffected. For example, if three loci, A/a, B/b and C/c, are on a chromosome in that order and only B/b is within a conversion track, LD between A/a and B/b and between B/b [illegible]sion but the LD between [illegible]ethods for inferring the [illegible]sion and recombination [74–78].

Subsection

Section

Applications of LD

Mutation and gene mapping. Mutation has a unique role in creating LD. When a mutant allele, M, first appears on a chromosome, it is in low frequency, $p_M = 1/(2N)$ (N is the population size) and is in perfect LD with the alleles at other loci that are on the chromosome carrying the first copy of M; perfect LD means that $D' = 1$ (BOX 1). If $D' = 1$, only three of the four possible haplotypes are present in the population (BOX 3). Perfect LD will persist until recombination involving an M-bearing chromosome creates a non-ancestral haplotype. Consequently, loci that are closely linked to M will remain in perfect LD for a long time and in strong LD for even longer.

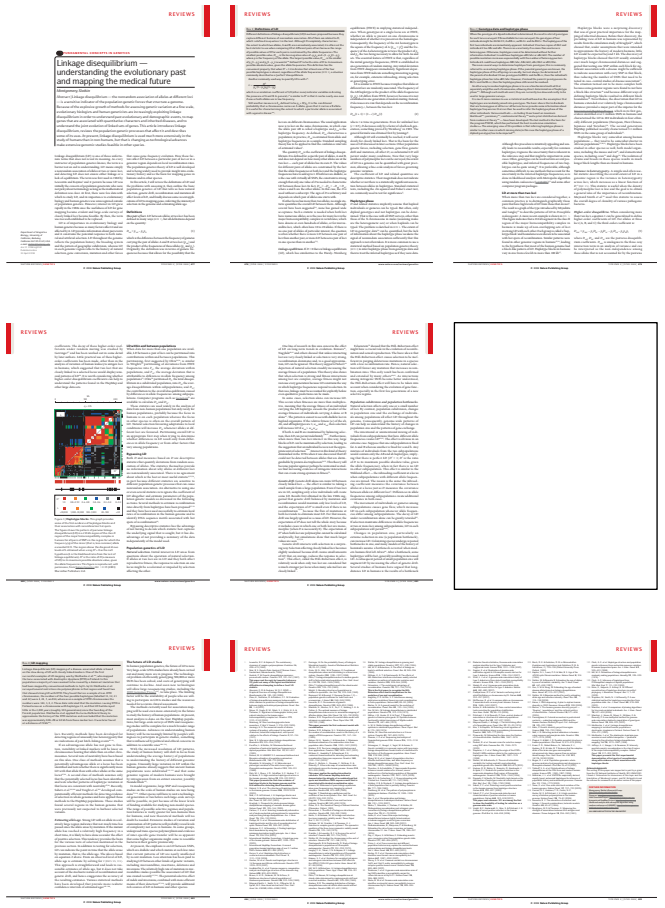
The persistence of strong LD between a mutant allele and the loci closely linked to it has many practical implications. Rare marker alleles in strong LD with a monogenic disease locus have to be closely linked to the causative locus. Relatively simple mathematical theory indicates just how close. The resulting method, called LD mapping, has been successfully used with several diseases (BOX 4).

The same idea underlies association mapping of complex diseases. Closely linked polymorphic SNPs tend to be in strong LD with one another. The fine-scale pattern of LD in humans confirms that the human genome is comprised of haplotype blocks within which most or all SNPs are in high LD[21]. These high levels of LD among SNPs are assumed to be true for alleles that increase the risk of complex inherited diseases. This idea, combined with the development of efficient methods for surveying large numbers of SNPs, has led to the many recent genome-wide association (GWA) studies that have detected SNPs that are significantly associated with breast cancer[79,80], colorectal cancer[81,82], type 2 diabetes[83–86] and heart disease[87,88], among other diseases.

However, one potential problem in GWA studies is that, as mentioned above, LD can be created by unrecognized population subdivision. Several methods have been proposed to account for such LD[89,90].

Although GWA studies have been successful in finding new causative alleles, the overall proportion of risk that is accounted for is often low. For example, Easton *et al.*[79] found five new variants associated with familial breast cancer, but only 3.6% of familial breast cancer is accounted for by those alleles. Seventy percent of the genetic basis of familial breast cancer remains unaccounted for. Alleles accounting for a greater proportion of risk might be found in even larger studies, but it is unclear whether most causative variants will ultimately be found this way. The reason is that GWA studies are more effective in finding causative alleles that are in relative high frequency. Other methods might be needed for low-frequency causative alleles.

Detecting natural selection. Strong positive selection quickly increases the frequency of an advantageous allele, with the result that linked loci remain in unusually strong LD with that allele. This idea originated with Maynard Smith and Haigh[91], who called it genetic hitch-hiking. Their paper focused on an advantageous allele that goes to fixation and causes a substantial reduction of heterozygosity at closely linked neutral



Box 4 | **LD mapping**

Box for additional technical material or examples

Finland arose on a chromosome with haplotype 11, and that LD had decayed little in the 2,000 years (roughly 100 generations) since the founding of the Finnish population. Hästbacka *et al.* applied the Luria–Delbrück theory to approximate the history of the DTD mutation and concluded that the mutation was approximately 0.06 cM or 60 kb from these marker loci. It was later found at a distance of 70 kb[129].

loci. Recently, methods have been developed for detecting regions of unusually low heterozygosity that are indications of past hitch-hiking events[92–94].

If an advantageous allele has not gone to fixation, variability at linked markers will be lower on chromosomes bearing that allele than on other chromosomes. Several tests of neutrality have been based on this idea. One class of methods assumes that a potentially advantageous allele at a locus has been identified and tests whether there is significantly more LD with that allele than with other alleles at the same locus[95,96]. A second class of methods assumes only that the potentially selected locus has been identified and tests whether patterns of haplotype variation at that locus are consistent with neutrality[97,98]. Recently, Sabeti *et al.*[99,100] and Voight *et al.*[101] developed computationally efficient methods for detecting evidence of selection in whole genomes and have applied those methods to the HapMap populations. These studies found several regions in the human genome that were previously not suspected to harbour selected variants.

Subsection

Estimating allele age. Strong LD with an allele in a relatively large region indicates that not much time has ...itation. If the mutant ...igh frequency in a ...e so under the effect ...cy provides the basis for the various tests of selection mentioned in the previous section. In addition to testing for selection, LD can indicate the point in time that the allele arose by mutation, that is, the allele age. The idea is based on equation 2 above. From an observed level of LD, allele age is estimate by solving for t (REFS 95,102). This approach is straightforward and leads to reasonable estimates of allele age, but it does not take account of the stochastic nature of recombination and genetic drift, and hence exaggerates the accuracy of the resulting estimates. Various statistical methods have been developed that provide more realistic confidence intervals of estimated ages[103–106].

The future of LD studies

In human population genetics, the future of LD is now. Very large-scale GWA studies have already been carried out and many more are in progress. The technological problem of efficiently genotyping 500,000 or more SNPs has been solved, and costs of genotyping will continue to decline. And soon new technologies will allow large resequencing studies, including the 1000 Genomes Project[107], to take place. The limiting factor will be the availability of people who are willing to participate in GWA studies and the resources needed for accurate clinical assessment.

The methods currently used for association mapping will be used even more extensively in the future to study the history of human populations. At present, most analysis is done on the four HapMap populations, but large-scale surveys of SNPs and resequencing studies will be complete for a much broader range of populations. Advances in understanding human history will be increasingly limited by people's willingness to participate in genetic studies, something that is influenced by political and ethical concerns in addition to scientific ones[108–110].

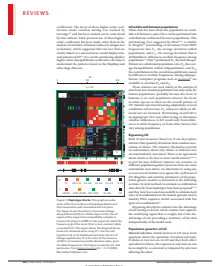
With the increased resolution of LD patterns, the study of human history will shift in focus from understanding the average history of populations to understanding the history of different genomic regions. Unusually large variation in LD within the human genome already suggests that ancient human populations were subdivided[111,112] and that some genomic regions of modern humans were brought by introgression from an extinct ancestor, possibly Neanderthals[113].

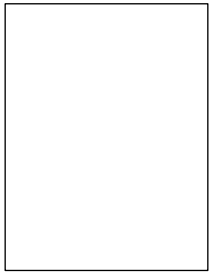
In model organisms, large SNP and resequencing studies on the scale of human studies are now being done[73,114]. Other species will have to wait a technological generation or two before such large-scale surveys will be possible, in part because of the lower levels of funding available for studying non-model species. The range of possible selective regimes and population histories is vastly greater for non-humans than for humans, and new theoretical methods will no doubt be needed. Extensive studies of variation and examination of LD patterns will probably reveal levels of complexity not seen in humans. In some groups, widespread trans-species polymorphism and evidence of inter-specific gene transfer will be so apparent that some higher organisms might come to resemble bacteria in their genetic promiscuity.

At present, the emphasis is on LD between SNPs, which are diallelic and which mutate at such low rates that current patterns of LD are nearly unaffected by recent mutation. Less attention has been paid to studying LD between other kinds of genetic variants, including microsatellites, insertions, deletions and inversions. The relatively high rate of mutation in microsatellites makes possible the assessment of LD that was created recently[115,116]. The potential selective effect of indels and inversions, combined with more efficient means of their detection[117,118], will provide additional rich sources of LD in humans and other species.

References

1. Lewontin, R. C. & Kojima, K. The evolutionary dynamics of complex polymorphisms. *Evolution* **14**, 458–472 (1960).
2. Weir, B. S. *Genetic Data Analysis II* (Sinauer Assoc., Sunderland, Massachusetts, 1996).
3. Hedrick, P. W. Genetic disequilibrium measures: proceed with caution. *Genetics* **117**, 331–341 (1987). **This paper and the reply by Lewontin (reference 33) point out many of the logical and statistical difficulties in attempting to define a 'best' LD statistic.**
4. Abecasis, G. R. & Cookson, W. O. C. GOLD — Graphical Overview of Linkage Disequilibrium. *Bioinformatics* **16**, 182–183 (2000).
5. Zhao, H., Nettleton, D. & Dekkers, J. C. M. Evaluation of linkage disequilibrium measures between multi-allelic markers as predictors of linkage disequilibrium between single nucleotide polymorphisms. *Genet. Res.* **89**, 1–6 (2007).
6. Meyer, D., Single, R. M., Mack, S. J., Erlich, H. A. & Thomson, G. Signatures of demographic history and natural selection in the human major histocompatibility complex loci. *Genetics* **173**, 2121–2142 (2006).
7. Weinberg, W. Uber vererbungsgesetze beim menschen. *Z. Abst V. Vererb.* **1**, 276–330 (1909).
8. Jennings, H. S. The numerical results of diverse systems of breeding, with respect to two pairs of characters, linked and independent, with special relation to the effects of linkage. *Genetics* **2**, 97–154 (1917).
9. Weir, B. S. Inferences about linkage disequilibrium. *Biometrics* **35**, 235–254 (1979).
10. Excoffier, L. & Slatkin, M. Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. *Mol. Biol. Evol.* **12**, 921–927 (1995).
11. Kuhner, M. K. LAMARC 2.0: maximum likelihood and Bayesian estimation of population parameters. *Bioinformatics* **22**, 768–770 (2006).
12. Miyashita, N. & Langley, C. H. Molecular and phenotypic variation of the white locus region in *Drosophila melanogaster*. *Genetics* **120**, 199–212 (1988).
13. Daly, M. J., Rioux, J. D., Schaffner, S. F., Hudson, T. J. & Lander, E. S. High-resolution haplotype structure in the human genome. *Nature* **29**, 229–232 (2001). **This paper presents the first clear evidence of haplotype blocks in the human genome and the first method for detecting block boundaries.**
14. Gabriel, S. B. *et al.* The structure of haplotype blocks in the human genome. *Science* **296**, 2225–2229 (2002).
15. Wall, J. D. & Pritchard, J. K. Haplotype blocks and linkage disequilibrium in the human genome. *Nature Rev. Genet.* **4**, 587–597 (2003).
16. Kruglyak, L. Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nature Genet.* **22**, 139–144 (1999).
17. Carlson, C. S. *et al.* Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. *Am. J. Hum. Genet.* **74**, 106–120 (2004).
18. Phillips, M. S. *et al.* Chromosome-wide distribution of haplotype blocks and the role of recombination hot spots. *Nature Genet.* **33**, 382–387 (2003).
19. Anderson, E. C. & Novembre, J. Finding haplotype block boundaries by using the minimum-description-length principle. *Am. J. Hum. Genet.* **73**, 336–354 (2003).
20. International HapMap Consortium. A haplotype map of the human genome. *Nature* **437**, 1299–1320 (2005).
21. International HapMap Consortium. A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449**, 851–861 (2007).
22. Guryev, V. *et al.* Haplotype block structure is conserved across mammals. *PLoS Genet.* **2**, 1111–1118 (2006).
23. Gautier, M. *et al.* Genetic and haplotypic structure in 14 European and African cattle breeds. *Genetics* **177**, 1059–1070 (2007).
24. Lindblad-Toh, K. *et al.* Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* **438**, 803–819 (2005).
25. Brown, A. H. D., Feldman, M. W. & Nevo, E. Multilocus structure of natural populations of *Hordeum spontaneum*. *Genetics* **96**, 523–536 (1980).
26. Maynard Smith, J., Smith, N. H., O'Rourke, M. & Spratt, B. G. How clonal are bacteria? *Proc. Natl Acad. Sci. USA* **90**, 4384–4388 (1993).
27. Geiringer, H. On the probability theory of linkage in Mendelian heredity. *Annals of Mathematical Statistics* **15**, 25–57 (1944).
28. Grote, M. N., Klitz, W. & Thomson, G. Constrained disequilibrium values and hitchhiking in a three-locus system. *Genetics* **150**, 1295–1307 (1998).
29. Ohta, T. Linkage disequilibrium due to random genetic
32. Raymond, M. & Rousset, F. Genepop (Version 1.2) — population-genetics software for exact tests and ecumenicism. *J. Hered.* **86**, 248–249 (1995).
33. Lewontin, R. C. On measures of gametic disequilibrium. *Genetics* **120**, 849–852 (1988).
34. Maniatis, N., Morton, N. E., Xu, C. F., Hosking, L. K. & Collins, A. The optimal measure of linkage disequilibrium reduces error in association mapping of affection status. *Hum. Mol. Genet.* **14**, 145–153 (2005).
35. Hudson, R. R. Properties of a neutral allele model with intragenic recombination. *Theor. Popul. Biol.* **23**, 183–201 (1983). **This paper presents the first coalescent model with recombination.**
36. Hudson, R. R. & Kaplan, N. L. Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics* **111**, 147–164 (1985).
37. Kuhner, M. K., Yamato, J. & Felsenstein, J. Maximum likelihood estimation of recombination rates from population data. *Genetics* **156**, 1393–1401 (2000).
38. Hudson, R. R. Two-locus sampling distributions and their applications. *Genetics* **159**, 1805–1817 (2001).
39. McVean, G., Awadalla, P. & Fearnhead, P. A coalescent-based method for detecting and estimating recombination from gene sequences. *Genetics* **160**, 1231–1241 (2002).
40. Myers, S., Bottolo, L., Freeman, C., McVean, G. & Donnelly, P. A fine-scale map of recombination rates and hotspots across the human genome. *Science* **310**, 321–324 (2005). **This paper applies the method described in reference 39 to human HapMap data and demonstrates the ubiquity of recombinational hot spots and identifies a DNA sequence motif that is associated with elevated recombination rates.**
41. Kimura, M. Attainment of quasi linkage equilibrium when gene frequencies are changing by natural selection. *Genetics* **52**, 875–890 (1965).
42. Nagylaki, T. Quasilinkage equilibrium and the evolution of two-locus systems. *Proc. Natl. Acad. Sci. USA* **71**, 526–530 (1974).
43. Nagylaki, T. The evolution of one and two-locus systems. *Genetics* **83**, 583–600 (1976).
44. Fisher, R. A. *The Genetical Theory of Natural Selection* (Clarendon, Oxford, 1930).
45. Felsenstein, J. The effect of linkage on directional selection. *Genetics* **52**, 349–363 (1965).
46. Karlin, S. & Feldman, M. W. Linkage and selection: two locus symmetric viability model. *Theor. Popul. Biol.* **1**, 39–71 (1970).
47. Feldman, M. W., Franklin, I. & Thomson, G. J. Selection in complex genetic systems I. The symmetric equilibria of the three-locus symmetric viability model. *Genetics* **76**, 135–162 (1974).
48. Franklin, I. & Lewontin, R. C. Is the gene the unit of selection? *Genetics* **65**, 707–734 (1970).
49. Slatkin, M. On treating the chromosome as the unit of selection. *Genetics* **72**, 157–168 (1972).
50. Charlesworth, B. & Charlesworth, D. Study of linkage disequilibrium in populations of *Drosophila melanogaster*. *Genetics* **73**, 351–359 (1973).
51. Langley, C. H., Tobari, Y. N. & Kojima, K. I. Linkage disequilibrium in natural populations of *Drosophila melanogaster*. *Genetics* **78**, 921–936 (1974).
52. Hamon, S. C. *et al.* Evidence for consistent intragenic and intergenic interactions between SNP effects in the *APOA1/C3/A4/A5* gene cluster. *Hum. Hered.* **61**, 87–96 (2006).
53. Hill, W. G. & Robertson, A. Linkage disequilibrium in finite populations. *Theor. Appl. Genet.* **38**, 226–231 (1968).
54. Ohta, T. & Kimura, M. Linkage disequilibrium at steady state determined by random genetic drift and recurrent mutation. *Genetics* **63**, 229–238 (1969).
55. Hudson, R. R. The sampling distribution of linkage disequilibrium under an infinite allele model without selection. *Genetics* **109**, 611–631 (1985).
56. Slatkin, M. Linkage disequilibrium in growing and stable populations. *Genetics* **137**, 331–336 (1994).
57. Hill, W. G. & Robertson, A. The effect of linkage on limits to artificial selection. *Genet. Res.* **8**, 269–294 (1966).
58. McVean, G. A. T. & Charlesworth, B. The effects of Hill–Robertson interference between weakly selected mutations on patterns of molecular evolution and variation. *Genetics* **155**, 929–944 (2000).
59. Felsenstein, J. The evolutionary advantage of recombination. *Genetics* **78**, 737–756 (1974). **This is the first paper to recognize the Hill–Robertson effect and its implications for the evolution of sex and recombination.**
60. Keightley, P. D. & Otto, S. P. Interference among deleterious mutations favours sex and recombination in finite populations. *Nature* **443**, 89–92 (2006).
61. Barton, N. H. A general model for the evolution of recombination. *Genet. Res.* **65**, 123–144 (1995).
62. Nei, M. & Li, W. Linkage disequilibrium in subdivided populations. *Genetics* **75**, 213–219 (1973).
63. Mitton, J. B., Koehn, R. K. & Prout, T. Population genetics of marine pelecypods. III. Epistasis between functionally related isoenzymes of *Mytilus edulis*. *Genetics* **73**, 487–496 (1973).
64. Li, W. H. Stable linkage disequilibrium without epistasis in subdivided populations. *Theor. Popul. Biol.* **6**, 173–183 (1974).
65. Slatkin, M. Gene flow and selection in a 2-locus system. *Genetics* **81**, 787–802 (1975).
66. Noonan, J. P. *et al.* Sequencing and analysis of Neanderthal genomic DNA. *Science* **314**, 1113–1118 (2006).
67. Schmegner, C., Hoegel, J., Vogel, W. & Assum, G. Genetic variability in a genomic region with long-range linkage disequilibrium reveals traces of a bottleneck in the history of the European population. *Hum. Genet.* **118**, 276–286 (2005).
68. Zhang, W. H. *et al.* Impact of population structure, effective bottleneck time, and allele frequency on linkage disequilibrium maps. *Proc. Natl. Acad. Sci. USA* **101**, 18075–18080 (2004).
69. Thornton, K. & Andolfatto, P. Approximate Bayesian inference reveals evidence for a recent, severe bottleneck in a Netherlands population of *Drosophila melanogaster*. *Genetics* **172**, 1607–1619 (2006).
70. Weir, B. S. & Cockerham, C. C. Group inbreeding with 2 linked loci. *Genetics* **63**, 711–742 (1969).
71. Golding, G. B. & Strobeck, C. Linkage disequilibrium in a finite population that is partially selfing. *Genetics* **94**, 777–789 (1980).
72. Nordborg, M. *et al.* The pattern of polymorphism in *Arabidopsis thaliana*. *PLoS Biol.* **3**, 1289–1299 (2005).
73. Kim, S. *et al.* Recombination and linkage disequilibrium in *Arabidopsis thaliana*. *Nature Genet.* **39**, 1151–1155 (2007).
74. Wiehe, T., Mountain, J., Parham, P. & Slatkin, M. Distinguishing recombination and intragenic gene conversion by linkage disequilibrium patterns. *Genet. Res.* **75**, 61–73 (2000).
75. Ardlie, K. *et al.* Lower-than-expected linkage disequilibrium between tightly linked markers in humans suggests a role for gene conversion. *Am. J. Hum. Genet.* **69**, 582–589 (2001).
76. Padhukasahasram, B., Marjoram, P. & Nordborg, M. Estimating the rate of gene conversion on human chromosome 21. *Am. J. Hum. Genet.* **75**, 386–397 (2004).
77. Gay, J., Myers, S. & McVean, G. Estimating meiotic gene conversion rates from population genetic data. *Genetics* **177**, 881–894 (2007).
78. Frisse, L. *et al.* Gene conversion and different population histories may explain the contrast between polymorphism and linkage disequilibrium levels. *Am. J. Hum. Genet.* **69**, 831–843 (2001).
79. Easton, D. F. *et al.* Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* **447**, 1087–1093 (2007).
80. Stacey, S. N. *et al.* Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer. *Nature Genet.* **39**, 865–869 (2007).
81. Tomlinson, I. *et al.* A genome-wide association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q24.21. *Nature Genet.* **39**, 984–988 (2007).
82. Zanke, B. W. *et al.* Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24. *Nature Genet.* **39**, 989–994 (2007).





83. Diabetes Genetics Initiative. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* **316**, 1331–1336 (2007).
84. Zeggini, E. *et al.* Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science* **316**, 1336–1341 (2007).
85. Scott, L. J. *et al.* A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* **316**, 1341–1345 (2007).
86. Sladek, R. *et al.* A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature* **445**, 881–885 (2007).
87. Gudbjartsson, D. F. *et al.* Variants conferring risk of atrial fibrillation on chromosome 4q25. *Nature* **448**, 353–357 (2007).
88. McPherson, R. *et al.* A common allele on chromosome 9 associated with coronary heart disease. *Science* **316**, 1488–1491 (2007).
89. Kohler, K. & Bickeboller, H. Case–control association tests correcting for population stratification. *Ann. Hum. Genet.* **70**, 98–115 (2006).
90. Pritchard, J. K. & Donnelly, P. Case–control studies of association in structured or admixed populations. *Theor. Popul. Biol.* **60**, 227–237 (2001).
91. Maynard Smith, J. & Haigh, J. The hitch-hiking effect of a favourable gene. *Genet. Res.* **23**, 23–35 (1974).
92. Kim, Y. & Stephan, W. Detecting a local signature of genetic hitchhiking along a recombining chromosome. *Genetics* **160**, 765–777 (2002).
93. Przeworski, M. Estimating the time since the fixation of a beneficial allele. *Genetics* **164**, 1667–1676 (2003).
94. Nielsen, R. *et al.* Genomic scans for selective sweeps using SNP data. *Genome Res.* **15**, 1566–1575 (2005).
95. Stephens, J. C. *et al.* Dating the origin of the CCR5–Delta32 AIDS-resistance allele by the coalescence of haplotypes. *Am. J. Hum. Genet.* **62**, 1507–1515 (1998).
96. Slatkin, M. & Bertorelle, G. The use of intra-allelic variability for testing neutrality and estimating population growth rate. *Genetics* **158**, 865–874 (2001).
97. Hudson, R. R., Bailey, K., Skarecky, D., Kwiatowski, J. & Ayala, F. J. Evidence for positive selection in the superoxide dismutase (Sod) region of *Drosophila melanogaster*. *Genetics* **136**, 1329–1340 (1994).
98. Depaulis, F. & Veuille, M. Neutrality tests based on the distribution of haplotypes under an infinite-site model. *Mol. Biol. Evol.* **15**, 1788–1790 (1998).
99. Sabeti, P. C. *et al.* Detecting recent positive selection in the human genome from haplotype structure. *Nature* **419**, 832–837 (2002).
100. Sabeti, P. C. *et al.* Genome-wide detection and characterization of positive selection in human populations. *Nature* **449**, 913–918 (2007).
This paper and reference 101 are among the first to show the feasibility of testing for selection on a genome-wide scale.
101. Voight, B. F., Kudaravalli, S., Wen, X. & Pritchard, J. K. A map of recent positive selection in the human genome. *PLoS Biol.* **4**, 446–458 (2006).
102. Reich, D. E. & Goldstein, D. B. in *Microsatellites: Evolution and Applications* (eds Goldstein, D. B. & Schlötterer, C.) 129–138 (Oxford University Press, Oxford, 1999).
103. Kaplan, N. L., Lewis, P. O. & Weir, B. S. Age of the ΔF508 cystic fibrosis mutation. *Nature Genet.* **8**, 216 (1994).
104. Slatkin, M. & Rannala, B. Estimating the age of alleles by use of intraallelic variability. *Am. J. Hum. Genet.* **60**, 447–458 (1997).
105. Guo, S. W. & Xiong, M. Estimating the age of mutant disease alleles based on linkage disequilibrium. *Hum. Hered.* **47**, 315–337 (1997).
106. Slatkin, M. A Bayesian method for jointly estimating allele age and selection intensity. *Genet. Res.* **90**, 119–128 (2008).
107. Kaiser, J. DNA sequencing: A plan to capture human diversity in 1000 Genomes. *Science* **319**, 395 (2008).
108. Barker, J. The human genome diversity project — 'Peoples', 'populations' and the cultural politics of identification. *Cultural Studies* **18**, 571–606 (2004).
109. Cunningham, H. Colonial encounters in postcolonial contexts — patenting indigenous DNA and the Human Genome Diversity Project. *Crit. Anthropol.* **18**, 205–233 (1998).
110. Kahn, P. Genetic diversity project tries again. *Science* **266**, 720–722 (1994).
111. Wall, J. D. Detecting ancient admixture in humans using sequence polymorphism data. *Genetics* **154**, 1271–1279 (2000).
112. Plagnol, V. & Wall, J. D. Possible ancestral structure in human populations. *Plos Genet.* **2**, 972–979 (2006).
113. Evans, P. D., Mekel-Bobrov, N., Vallender, E. J., Hudson, R. R. & Lahn, B. T. Evidence that the adaptive allele of the brain size gene microcephalin introgressed into *Homo sapiens* from an archaic *Homo* lineage. *Proc. Natl. Acad. Sci. USA* **103**, 18178–18183 (2006).
114. Begun, D. J. *et al.* Population genomics: whole-genome analysis of polymorphism and divergence in *Drosophila simulans*. *PLoS Biol.* **5**, e310 (2007).
115. Tishkoff, S. A. *et al.* Global patterns of linkage disequilibrium at the CD4 locus and modern human origins. *Science* **271**, 1380–1387 (1996).
116. Mountain, J. L. *et al.* SNPSTRs: empirically derived, rapidly typed, autosomal haplotypes for inference of population history and mutational processes. *Genome Res.* **12**, 1766–1772 (2002).
117. Tuzun, E. *et al.* Fine-scale structural variation of the human genome. *Nature Genet.* **37**, 727–732 (2005).
118. Conrad, D. F., Andrews, T. D., Carter, N. P., Hurles, M. E. & Pritchard, J. K. A high-resolution survey of deletion polymorphism in the human genome. *Nature Genet.* **38**, 75–81 (2006).
119. Lewontin, R. C. The interaction of selection and linkage. I. General considerations; heterotic models. *Genetics* **49**, 49–67 (1964).
120. Bengtsson, B. O. & Thomson, G. Measuring the strength of associations between HLA antigens and diseases. *Tissue Antigens* **18**, 356–363 (1981).
121. Clark, A. G. *et al.* Haplotype structure and population genetic inferences from nucleotide-sequence variation in human lipoprotein lipase. *Am. J. Hum. Genet.* **63**, 595–612 (1998).
122. Hill, W. G. Estimation of linkage disequilibrium in randomly mating populations. *Heredity* **33**, 229–239 (1974).
123. Clark, A. G. Inference of haplotypes from PCR-amplified samples of diploid populations. *Mol. Biol. Evol.* **7**, 111–122 (1990).
124. Eskin, E., Halperin, E. & Karp, R. M. Efficient reconstruction of haplotype structure via perfect phylogeny. *J. Bioinform. Comput. Biol.* **1**, 1–20 (2003).
125. Stephens, M., Smith, N. J. & Donnelly, P. A new statistical method for haplotype reconstruction from population data. *Am. J. Hum. Genet.* **68**, 978–989 (2001).
126. Marchini, J. *et al.* A comparison of phasing algorithms for trios and unrelated individuals. *Am. J. Hum. Genet.* **78**, 437–450 (2006).
127. Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nature Genet.* **39**, 906–913 (2007).
128. Hästbacka, J. *et al.* Linkage disequilibrium mapping in isolated founder populations: diastrophic dysplasia in Finland. *Nature Genet.* **2**, 204–211 (1992).
129. Hästbacka, J. *et al.* The diastrophic dysplasia gene encodes a novel sulfate transporter — positional cloning by fine-structure linkage disequilibrium mapping. *Cell* **78**, 1073–1087 (1994).
130. Jeffreys, A. J., Kauppi, L. & Neumann, R. Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex. *Nature Genet.* **29**, 217–222 (2001).
This paper presents the first experimental demonstration of hot spots of recombination along with evidence of their association with haplotype blocks.

Acknowledgements

The writing of this paper was supported in part by a grant from the US National Institutes of Health, R01-GM40282. I thank J. Felsenstein for discussions of this topic and the translation of Weinberg's paper, and M. Kirkpatrick and the referees for comments on an earlier version of this paper.

FURTHER INFORMATION

Montgomery Slatkin's Research Group: http://ib.berkeley.edu/labs/slatkin
Genepop: http://genepop.curtin.edu.au
1000 Genomes Project: http://www.1000genomes.org
International HapMap Project: http://www.hapmap.org
Likelihood analysis with metropolis algorithm using random coalescence (LAMARC): http://evolution.genetics.washington.edu/lamarc/index.html

ALL LINKS ARE ACTIVE IN THE ONLINE PDF

83. Diabetes Genetics Initiative. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* **316**, 1331–1336 (2007).
84. Zeggini, E. *et al.* Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science* **316**, 1336–1341 (2007).
85. Scott, L. J. *et al.* A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* **316**, 1341–1345 (2007).
86. Sladek, R. *et al.* A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature* **445**, 881–885 (2007).
87. Gudbjartsson, D. F. *et al.* Variants conferring risk of atrial fibrillation on chromosome 4q25. *Nature* **448**, 353–357 (2007).
88. McPherson, R. *et al.* A common allele on chromosome 9 associated with coronary heart disease. *Science* **316**, 1488–1491 (2007).
89. Kohler, K. & Bickeboller, H. Case–control association tests correcting for population stratification. *Ann. Hum. Genet.* **70**, 98–115 (2006).
90. Pritchard, J. K. & Donnelly, P. Case–control studies of association in structured or admixed populations. *Theor. Popul. Biol.* **60**, 227–237 (2001).
91. Maynard Smith, J. & Haigh, J. The hitch-hiking effect of a favourable gene. *Genet. Res.* **23**, 23–35 (1974).
92. Kim, Y. & Stephan, W. Detecting a local signature of genetic hitchhiking along a recombining chromosome. *Genetics* **160**, 765–777 (2002).
93. Przeworski, M. Estimating the time since the fixation of a beneficial allele. *Genetics* **164**, 1667–1676 (2003).
94. Nielsen, R. *et al.* Genomic scans for selective sweeps using SNP data. *Genome Res.* **15**, 1566–1575 (2005).
95. Stephens, J. C. *et al.* Dating the origin of the CCR5–Delta32 AIDS-resistance allele by the coalescence of haplotypes. *Am. J. Hum. Genet.* **62**, 1507–1515 (1998).
96. Slatkin, M. & Bertorelle, G. The use of intra-allelic variability for testing neutrality and estimating population growth rate. *Genetics* **158**, 865–874 (2001).
97. Hudson, R. R., Bailey, K., Skarecky, D., Kwiatowski, J. & Ayala, F. J. Evidence for positive selection in the superoxide dismutase (Sod) region of *Drosophila melanogaster*. *Genetics* **136**, 1329–1340 (1994).
98. Depaulis, F. & Veuille, M. Neutrality tests based on the distribution of haplotypes under an infinite-site model. *Mol. Biol. Evol.* **15**, 1788–1790 (1998).
99. Sabeti, P. C. *et al.* Detecting recent positive selection in the human genome from haplotype structure. *Nature* **419**, 832–837 (2002).
100. Sabeti, P. C. *et al.* Genome-wide detection and characterization of positive selection in human populations. *Nature* **449**, 913–918 (2007).
This paper and reference 101 are among the first to show the feasibility of testing for selection on a genome-wide scale.
101. Voight, B. F., Kudaravalli, S., Wen, X. & Pritchard, J. K. A map of recent positive selection in the human genome. *PLoS Biol.* **4**, 446–458 (2006).
102. Reich, D. E. & Goldstein, D. B. in *Microsatellites: Evolution and Applications* (eds Goldstein, D. B. & Schlötterer, C.) 129–138 (Oxford University Press, Oxford, 1999).
103. Kaplan, N. L., Lewis, P. O. & Weir, B. S. Age of the ΔF508 cystic fibrosis mutation. *Nature Genet.* **8**, 216 (1994).
104. Slatkin, M. & Rannala, B. Estimating the age of alleles by use of intraallelic variability. *Am. J. Hum. Genet.* **60**, 447–458 (1997).
105. Guo, S. W. & Xiong, M. Estimating the age of mutant disease alleles based on linkage disequilibrium. *Hum. Hered.* **47**, 315–337 (1997).
106. Slatkin, M. A Bayesian method for jointly estimating allele age and selection intensity. *Genet. Res.* **90**, 119–128 (2008).
107. Kaiser, J. DNA sequencing: A plan to capture human diversity in 1000 Genomes. *Science* **319**, 395 (2008).
108. Barker, J. The human genome diversity project — 'Peoples', 'populations' and the cultural politics of identification. *Cultural Studies* **18**, 571–606 (2004).
109. Cunningham, H. Colonial encounters in postcolonial contexts — patenting indigenous DNA and the Human Genome Diversity Project. *Crit. Anthropol.* **18**, 205–233 (1998).
110. Kahn, P. Genetic diversity project tries again. *Science* **266**, 720–722 (1994).
111. Wall, J. D. Detecting ancient admixture in humans using sequence polymorphism data. *Genetics* **154**, 1271–1279 (2000).
112. Plagnol, V. & Wall, J. D. Possible ancestral structure in human populations. *Plos Genet.* **2**, 972–979 (2006).
113. Evans, P. D., Mekel-Bobrov, N., Vallender, E. J., Hudson, R. R. & Lahn, B. T. Evidence that the adaptive allele of the brain size gene microcephalin introgressed into *Homo sapiens* from an archaic *Homo* lineage. *Proc. Natl. Acad. Sci. USA* **103**, 18178–18183 (2006).
114. Begun, D. J. *et al.* Population genomics: whole-genome analysis of polymorphism and divergence in *Drosophila simulans*. *PLoS Biol.* **5**, e310 (2007).
115. Tishkoff, S. A. *et al.* Global patterns of linkage disequilibrium at the CD4 locus and modern human origins. *Science* **271**, 1380–1387 (1996).
116. Mountain, J. L. *et al.* SNPSTRs: empirically derived, rapidly typed, autosomal haplotypes for inference of population history and mutational processes. *Genome Res.* **12**, 1766–1772 (2002).
117. Tuzun, E. *et al.* Fine-scale structural human genome. *Nature Genet.* **37**,
118. Conrad, D. F., Andrews, T. D., Carter Hurles, M. E. & Pritchard, J. K. A hi survey of deletion polymorphism in genome. *Nature Genet.* **38**, 75–81
119. Lewontin, R. C. The interaction of selection and linkage. I. General considerations; heterotic models. *Genetics* **49**, 49–67 (1964).
120. Bengtsson, B. O. & Thomson, G. Measuring the strength of associations between HLA antigens and diseases. *Tissue Antigens* **18**, 356–363 (1981).
121. Clark, A. G. *et al.* Haplotype structure and population genetic inferences from nucleotide-sequence variation in human lipoprotein lipase. *Am. J. Hum. Genet.* **63**, 595–612 (1998).
122. Hill, W. G. Estimation of linkage disequilibrium in randomly mating populations. *Heredity* **33**, 229–239 (1974).
123. Clark, A. G. Inference of haplotypes from PCR-amplified samples of diploid populations. *Mol. Biol. Evol.* **7**, 111–122 (1990).
124. Eskin, E., Halperin, E. & Karp, R. M. Efficient reconstruction of haplotype structure via perfect phylogeny. *J. Bioinform. Comput. Biol.* **1**, 1–20 (2003).
125. Stephens, M., Smith, N. J. & Donnelly, P. A new statistical method for haplotype reconstruction from population data. *Am. J. Hum. Genet.* **68**, 978–989 (2001).
126. Marchini, J. *et al.* A comparison of phasing algorithms for trios and unrelated individuals. *Am. J. Hum. Genet.* **78**, 437–450 (2006).
127. Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nature Genet.* **39**, 906–913 (2007).
128. Hästbacka, J. *et al.* Linkage disequilibrium mapping in isolated founder populations: diastrophic dysplasia in Finland. *Nature Genet.* **2**, 204–211 (1992).
129. Hästbacka, J. *et al.* The diastrophic dysplasia gene encodes a novel sulfate transporter — positional cloning by fine-structure linkage disequilibrium mapping. *Cell* **78**, 1073–1087 (1994).
130. Jeffreys, A. J., Kauppi, L. & Neumann, R. Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex. *Nature Genet.* **29**, 217–222 (2001).
This paper presents the first experimental demonstration of hot spots of recombination along with evidence of their association with haplotype blocks.

Acknowledgements
The writing of this paper was supported in part by a grant from the US National Institutes of Health, R01-GM40282. I thank J. Felsenstein for discussions of this topic and the translation of Weinberg's paper, and M. Kirkpatrick and the referees for comments on an earlier version of this paper.

Acknowledgements

International HapMap Project: http://www.hapmap.org
Likelihood analysis with metropolis algorithm using random coalescence (LAMARC): http://evolution.genetics.washington.edu/lamarc/index.html

ALL LINKS ARE ACTIVE IN THE ONLINE PDF

Research articles:

Structure and components

Received: 12 July 2023 | Accepted: 19 August 2023

DOI: 10.1111/age.13352

RESEARCH ARTICLE

ANIMAL GENETICS WILEY

A *de novo* 2.3 kb structural variant in *MITF* explains a novel splashed white phenotype in a Thoroughbred family

R. R. Bellone[1,2] | J. Tanaka[1] | E. Esdaile[1] | R. B. Sutton[3] | F. Payette[4] | L. Leduc[4] | B. J. Till[1] | A. K. Abdel-Ghaffar[1] | M. Hammond[1,2] | K. G. Magdesian[5]

[1]Veterinary Genetics Laboratory, School of Veterinary Medicine, UC Davis, Davis, California, USA

[2]Department of Population Health and Reproduction, School of Veterinary Medicine, UC Davis, Davis, California, USA

[3]Cell Physiology and Molecular Biophysics, School of Medicine, Texas Tech University Health Sciences Center, Lubbock, Texas, USA

[4]Department of Clinical Studies, New Bolton Center, University of Pennsylvania School of Veterinary Medicine, University School of Veterinary Medicine, University of Pennsylvania, Kennett Square, Pennsylvania, USA

[5]Department of Medicine and Epidemiology, School of Veterinary Medicine, University of California-Davis, Davis, California, USA

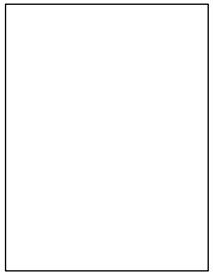
Correspondence
R. R. Bellone, Veterinary Genetics Laboratory, School of Veterinary Medicine, UC Davis, Davis, CA, USA.
Email: rbellone@ucdavis.edu

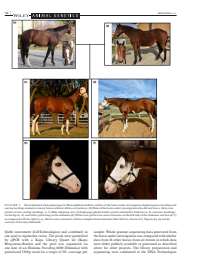
Funding information
UC Davis Veterinary Genetics Laboratory

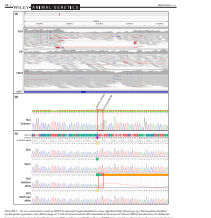
Abstract

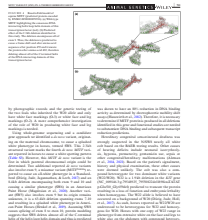
Splashed white in horses is characterized by extensive white patterning on the legs, face and abdomen and may be accompanied by deafness. To date, seven variants in *microphthalmia-associated transcription factor* (*MITF*) and two variants in *Paired Box 3* (*PAX3*) have been identified to explain this phenotype. A splashed white Thoroughbred stallion, whose sire and dam were not patterned, was hypothesized to have a *de novo* variant leading to his white coat pattern. A whole-genome sequencing candidate gene approach identified two single nucleotide variants (SNVs) in *SOX10*, four SNVs in *MITF* and a 2.3 kb deletion in *MITF* with the alternative allele present in this stallion but absent in the other 18 horses analyzed. All six SNVs were annotated as modifiers and were not further considered. The deletion in *MITF* (NC_009159.3:g.21555811_21558139delinsAAAT) encompasses exon 9 encoding a part of the helix–loop–helix domain required for DNA binding. Sanger sequencing and parentage testing confirmed that this deletion was a *de novo* mutation of maternal origin. Consistent with the published nomenclature, we denote this likely causal variant as SW8. Genotyping three of this stallion's offspring identified SW8 only in the nearly all-white foal that was confirmed deaf by brainstem auditory evoked response testing. This foal was also a compound heterozygote for dominant white variants (W20/W22), but to date, W variants alone have not been connected to deafness. SW8 marks the fourth de novo *MITF* variant in horses reported to cause white patterning. The link between deafness and all *MITF* variants with and without other variants impacting melanocyte development and function needs to be further explored.

KEYWORDS
coat color, deafness, horse, pigmentation, white patterning









INTRODUCTION

Variation in white patterning has long fascinated animal enthusiasts and researchers alike. Moreover, some of these variations have a clinical connection and have similar counterparts in humans or other species. In horses, several different types of white patterning phenotypes have been described including frame overo, tobiano, roan, leopard complex spotting, dominant white, sabino and splashed white (Sponenberg & Bellone, 2017). The molecular bases of many of these have been determined, with the first to be unraveled being the frame overo pattern in which white patterning is prevalent on the ventral aspects of the horse with the pigment 'framing' the

Received: 12 July 2023 | Accepted: 19 August 2023

DOI: 10.1111/age.13352

Date received and published

ANIMAL GENETICS WILEY

RESEARCH ARTICLE

A *de novo* 2.3 kb structural variant in *MITF* explains a novel splashed white phenotype in a Thoroughbred family

R. R. Bellone[1,2] | J. Tanaka[1] | E. Esdaile[1] | R. B. Sutton[3] | F. Payette[4] | L. Leduc[4] | B. J. Till[1] | A. K. Abdel-Ghaffar[1] | M. Hammond[1,2] | K. G. Magdesian[5]

[1]Veterinary Genetics Laboratory, School of Veterinary Medicine, UC Davis, Davis, California, USA

[2]Department of Population Health and Reproduction, School of Veterinary Medicine, UC Davis, Davis, California, USA

[3]Cell Physiology and Molecular Biophysics, School of Medicine, Texas Tech University Health Sciences Center, Lubbock, Texas, USA

[4]Department of Clinical Studies, New Bolton Center, University of Pennsylvania School of Veterinary Medicine, University School of Veterinary Medicine, University of Pennsylvania, Kennett Square, Pennsylvania, USA

[5]Department of Medicine and Epidemiology, School of Veterinary Medicine, University of California-Davis, Davis, California, USA

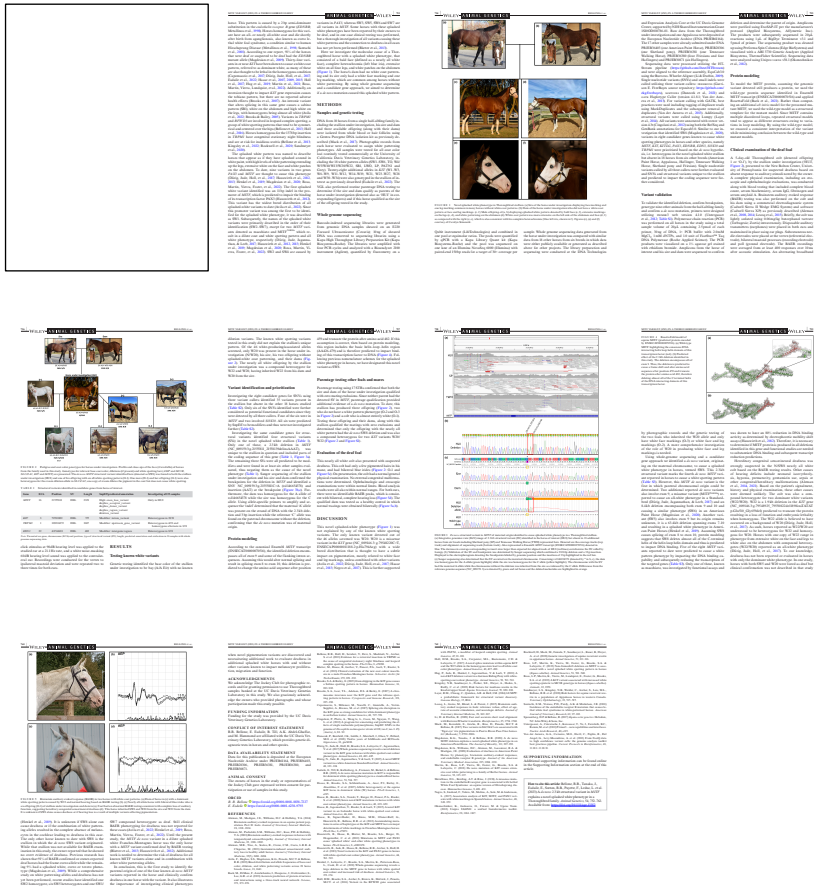
Correspondence
R. R. Bellone, Veterinary Genetics Laboratory, School of Veterinary Medicine, UC Davis, Davis, CA, USA.
Email: rbellone@ucdavis.edu

Funding information
UC Davis Veterinary Genetics Laboratory

Abstract

Splashed white in horses is characterized by extensive white patterning on the legs, face and abdomen and may be accompanied by deafness. To date, seven variants in *microphthalmia-associated transcription factor* (*MITF*) and two variants in *Paired Box 3* (*PAX3*) have been identified to explain this phenotype. A splashed white Thoroughbred stallion, whose sire and dam were not patterned, was hypothesized to have a *de novo* variant leading to his white coat pattern. A whole-genome sequencing candidate gene approach identified two single nucleotide variants (SNVs) in *SOX10*, four SNVs in *MITF* and a 2.3 kb deletion in *MITF* with the alternative allele present in this stallion but absent in the other 18 horses analyzed. All six SNVs were annotated as modifiers and were not further considered. The deletion in *MITF* (NC_009159.3:g.21555811_21558139delinsAAAT) encompasses exon 9 encoding a part of the helix–loop–helix domain required for DNA binding. Sanger sequencing and parentage testing confirmed that this deletion was a *de novo* mutation of maternal origin. Consistent with the published nomenclature, we denote this likely causal variant as SW8. Genotyping three of this stallion's offspring identified SW8 only in the nearly all-white foal that was confirmed deaf by brainstem auditory evoked response testing. This foal was also a compound heterozygote for dominant white variants (W20/W22), but to date, W variants alone have not been connected to deafness. SW8 marks the fourth de novo *MITF* variant in horses reported to cause white patterning. The link between deafness and all *MITF* variants with and without other variants impacting melanocyte development and function needs to be further explored.

KEYWORDS
coat color, deafness, horse, pigmentation, white patterning



INTRODUCTION

Variation in white patterning has long fascinated animal enthusiasts and researchers alike. Moreover, some of these variations have a clinical connection and have similar counterparts in humans or other species. In horses, several different types of white patterning phenotypes have been described including frame overo, tobiano, roan, leopard complex spotting, dominant white, sabino and splashed white (Sponenberg & Bellone, 2017). The molecular bases of many of these have been determined, with the first to be unraveled being the frame overo pattern in which white patterning is prevalent on the ventral aspects of the horse with the pigment 'framing' the

Received: 12 July 2023 |
DOI: 10.1111/age.13352

Digital Object Identifier

ENETICS WILEY

RESEARCH ARTICLE

A *de novo* 2.3 kb structural variant in *MITF* explains a novel splashed white phenotype in a Thoroughbred family

R. R. Bellone[1,2] | J. Tanaka[1] | E. Esdaile[1] | R. B. Sutton[3] | F. Payette[4] | L. Leduc[4] | B. J. Till[1] | A. K. Abdel-Ghaffar[1] | M. Hammond[1,2] | K. G. Magdesian[5]

[1]Veterinary Genetics Laboratory, School of Veterinary Medicine, UC Davis, Davis, California, USA

[2]Department of Population Health and Reproduction, School of Veterinary Medicine, UC Davis, Davis, California, USA

[3]Cell Physiology and Molecular Biophysics, School of Medicine, Texas Tech University Health Sciences Center, Lubbock, Texas, USA

[4]Department of Clinical Studies, New Bolton Center, University of Pennsylvania School of Veterinary Medicine, University School of Veterinary Medicine, University of Pennsylvania, Kennett Square, Pennsylvania, USA

[5]Department of Medicine and Epidemiology, School of Veterinary Medicine, University of California-Davis, Davis, California, USA

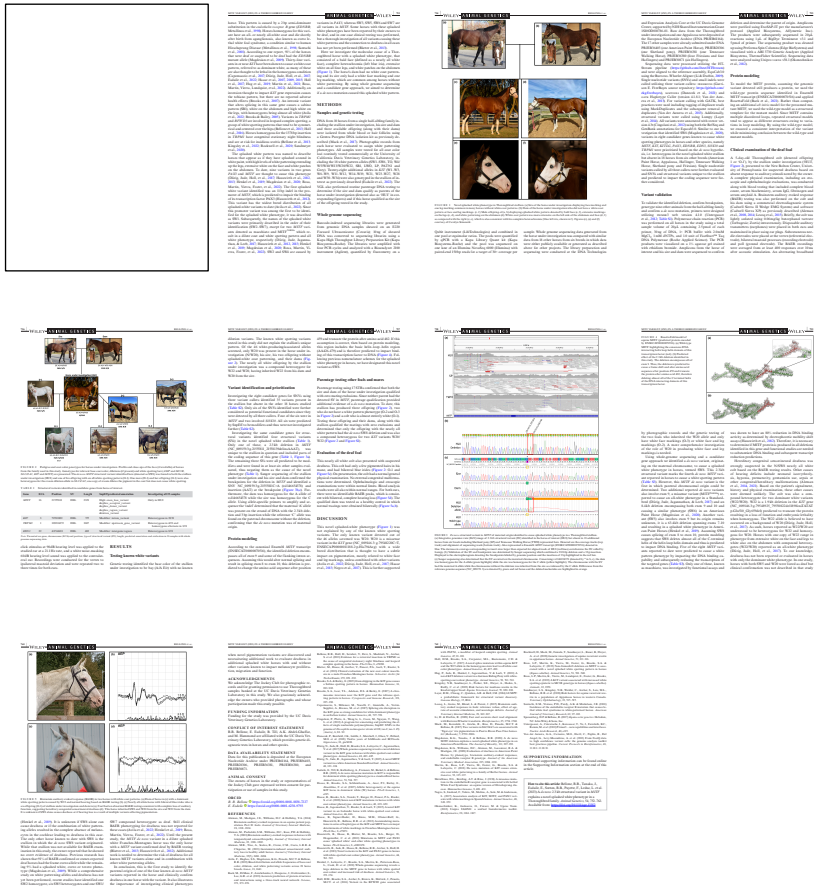
Correspondence
R. R. Bellone, Veterinary Genetics Laboratory, School of Veterinary Medicine, UC Davis, Davis, CA, USA.
Email: rbellone@ucdavis.edu

Funding information
UC Davis Veterinary Genetics Laboratory

Abstract

Splashed white in horses is characterized by extensive white patterning on the legs, face and abdomen and may be accompanied by deafness. To date, seven variants in *microphthalmia-associated transcription factor* (*MITF*) and two variants in *Paired Box 3* (*PAX3*) have been identified to explain this phenotype. A splashed white Thoroughbred stallion, whose sire and dam were not patterned, was hypothesized to have a *de novo* variant leading to his white coat pattern. A whole-genome sequencing candidate gene approach identified two single nucleotide variants (SNVs) in *SOX10*, four SNVs in *MITF* and a 2.3 kb deletion in *MITF* with the alternative allele present in this stallion but absent in the other 18 horses analyzed. All six SNVs were annotated as modifiers and were not further considered. The deletion in *MITF* (NC_009159.3:g.21555811_21558139delinsAAAT) encompasses exon 9 encoding a part of the helix–loop–helix domain required for DNA binding. Sanger sequencing and parentage testing confirmed that this deletion was a *de novo* mutation of maternal origin. Consistent with the published nomenclature, we denote this likely causal variant as SW8. Genotyping three of this stallion's offspring identified SW8 only in the nearly all-white foal that was confirmed deaf by brainstem auditory evoked response testing. This foal was also a compound heterozygote for dominant white variants (W20/W22), but to date, W variants alone have not been connected to deafness. SW8 marks the fourth de novo *MITF* variant in horses reported to cause white patterning. The link between deafness and all *MITF* variants with and without other variants impacting melanocyte development and function needs to be further explored.

KEYWORDS
coat color, deafness, horse, pigmentation, white patterning



INTRODUCTION

Variation in white patterning has long fascinated animal enthusiasts and researchers alike. Moreover, some of these variations have a clinical connection and have similar counterparts in humans or other species. In horses, several different types of white patterning phenotypes have been described including frame overo, tobiano, roan, leopard complex spotting, dominant white, sabino and splashed white (Sponenberg & Bellone, 2017). The molecular bases of many of these have been determined, with the first to be unraveled being the frame overo pattern in which white patterning is prevalent on the ventral aspects of the horse with the pigment 'framing' the

Received: 12 July 2023 | Accepted: 19 August 2023

DOI: 10.1111/age.13352

RESEARCH ARTICLE

Article type

ANIMAL GENETICS WILEY

A *de novo* 2.3 kb structural variant in *MITF* explains a novel splashed white phenotype in a Thoroughbred family

R. R. Bellone[1,2] | J. Tanaka[1] | E. Esdaile[1] | R. B. Sutton[3] | F. Payette[4] | L. Leduc[4] | B. J. Till[1] | A. K. Abdel-Ghaffar[1] | M. Hammond[1,2] | K. G. Magdesian[5]

[1]Veterinary Genetics Laboratory, School of Veterinary Medicine, UC Davis, Davis, California, USA

[2]Department of Population Health and Reproduction, School of Veterinary Medicine, UC Davis, Davis, California, USA

[3]Cell Physiology and Molecular Biophysics, School of Medicine, Texas Tech University Health Sciences Center, Lubbock, Texas, USA

[4]Department of Clinical Studies, New Bolton Center, University of Pennsylvania School of Veterinary Medicine, University School of Veterinary Medicine, University of Pennsylvania, Kennett Square, Pennsylvania, USA

[5]Department of Medicine and Epidemiology, School of Veterinary Medicine, University of California-Davis, Davis, California, USA

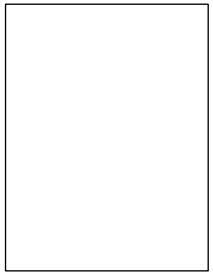
Correspondence
R. R. Bellone, Veterinary Genetics Laboratory, School of Veterinary Medicine, UC Davis, Davis, CA, USA.
Email: rbellone@ucdavis.edu

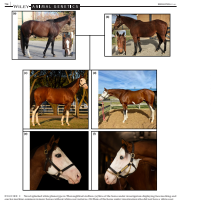
Funding information
UC Davis Veterinary Genetics Laboratory

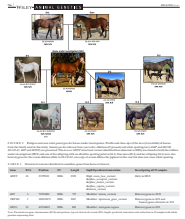
Abstract

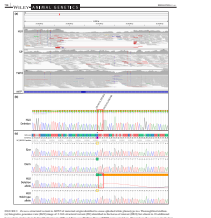
Splashed white in horses is characterized by extensive white patterning on the legs, face and abdomen and may be accompanied by deafness. To date, seven variants in *microphthalmia-associated transcription factor* (*MITF*) and two variants in *Paired Box 3* (*PAX3*) have been identified to explain this phenotype. A splashed white Thoroughbred stallion, whose sire and dam were not patterned, was hypothesized to have a *de novo* variant leading to his white coat pattern. A whole-genome sequencing candidate gene approach identified two single nucleotide variants (SNVs) in *SOX10*, four SNVs in *MITF* and a 2.3 kb deletion in *MITF* with the alternative allele present in this stallion but absent in the other 18 horses analyzed. All six SNVs were annotated as modifiers and were not further considered. The deletion in *MITF* (NC_009159.3:g.21555811_21558139delinsAAAT) encompasses exon 9 encoding a part of the helix–loop–helix domain required for DNA binding. Sanger sequencing and parentage testing confirmed that this deletion was a *de novo* mutation of maternal origin. Consistent with the published nomenclature, we denote this likely causal variant as SW8. Genotyping three of this stallion's offspring identified SW8 only in the nearly all-white foal that was confirmed deaf by brainstem auditory evoked response testing. This foal was also a compound heterozygote for dominant white variants (W20/W22), but to date, W variants alone have not been connected to deafness. SW8 marks the fourth de novo *MITF* variant in horses reported to cause white patterning. The link between deafness and all *MITF* variants with and without other variants impacting melanocyte development and function needs to be further explored.

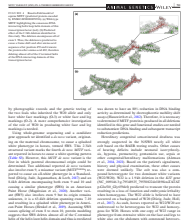
KEYWORDS
coat color, deafness, horse, pigmentation, white patterning











INTRODUCTION

Variation in white patterning has long fascinated animal enthusiasts and researchers alike. Moreover, some of these variations have a clinical connection and have similar counterparts in humans or other species. In horses, several different types of white patterning phenotypes have been described including frame overo, tobiano, roan, leopard complex spotting, dominant white, sabino and splashed white (Sponenberg & Bellone, 2017). The molecular bases of many of these have been determined, with the first to be unraveled being the frame overo pattern in which white patterning is prevalent on the ventral aspects of the horse with the pigment 'framing' the

Received: 12 July 2023 | Accepted: 19 August 2023

DOI: 10.1111/age.13352

RESEARCH ARTICLE

Journal ANIMAL GENETICS WILEY

A *de novo* 2.3 kb structural variant in *MITF* explains a novel splashed white phenotype in a Thoroughbred family

R. R. Bellone[1,2] | J. Tanaka[1] | E. Esdaile[1] | R. B. Sutton[3] | F. Payette[4] | L. Leduc[4] | B. J. Till[1] | A. K. Abdel-Ghaffar[1] | M. Hammond[1,2] | K. G. Magdesian[5]

[1]Veterinary Genetics Laboratory, School of Veterinary Medicine, UC Davis, Davis, California, USA

[2]Department of Population Health and Reproduction, School of Veterinary Medicine, UC Davis, Davis, California, USA

[3]Cell Physiology and Molecular Biophysics, School of Medicine, Texas Tech University Health Sciences Center, Lubbock, Texas, USA

[4]Department of Clinical Studies, New Bolton Center, University of Pennsylvania School of Veterinary Medicine, University School of Veterinary Medicine, University of Pennsylvania, Kennett Square, Pennsylvania, USA

[5]Department of Medicine and Epidemiology, School of Veterinary Medicine, University of California-Davis, Davis, California, USA

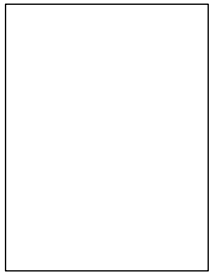
Correspondence
R. R. Bellone, Veterinary Genetics Laboratory, School of Veterinary Medicine, UC Davis, Davis, CA, USA.
Email: rbellone@ucdavis.edu

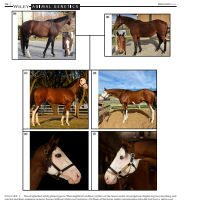
Funding information
UC Davis Veterinary Genetics Laboratory

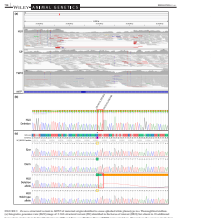
Abstract

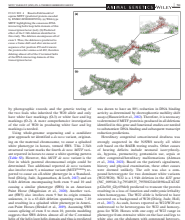
Splashed white in horses is characterized by extensive white patterning on the legs, face and abdomen and may be accompanied by deafness. To date, seven variants in *microphthalmia-associated transcription factor* (*MITF*) and two variants in *Paired Box 3* (*PAX3*) have been identified to explain this phenotype. A splashed white Thoroughbred stallion, whose sire and dam were not patterned, was hypothesized to have a *de novo* variant leading to his white coat pattern. A whole-genome sequencing candidate gene approach identified two single nucleotide variants (SNVs) in *SOX10*, four SNVs in *MITF* and a 2.3 kb deletion in *MITF* with the alternative allele present in this stallion but absent in the other 18 horses analyzed. All six SNVs were annotated as modifiers and were not further considered. The deletion in *MITF* (NC_009159.3:g.21555811_21558139delinsAAAT) encompasses exon 9 encoding a part of the helix–loop–helix domain required for DNA binding. Sanger sequencing and parentage testing confirmed that this deletion was a *de novo* mutation of maternal origin. Consistent with the published nomenclature, we denote this likely causal variant as SW8. Genotyping three of this stallion's offspring identified SW8 only in the nearly all-white foal that was confirmed deaf by brainstem auditory evoked response testing. This foal was also a compound heterozygote for dominant white variants (W20/W22), but to date, W variants alone have not been connected to deafness. SW8 marks the fourth de novo *MITF* variant in horses reported to cause white patterning. The link between deafness and all *MITF* variants with and without other variants impacting melanocyte development and function needs to be further explored.

KEYWORDS
coat color, deafness, horse, pigmentation, white patterning









INTRODUCTION

Variation in white patterning has long fascinated animal enthusiasts and researchers alike. Moreover, some of these variations have a clinical connection and have similar counterparts in humans or other species. In horses, several different types of white patterning phenotypes have been described including frame overo, tobiano, roan, leopard complex spotting, dominant white, sabino and splashed white (Sponenberg & Bellone, 2017). The molecular bases of many of these have been determined, with the first to be unraveled being the frame overo pattern in which white patterning is prevalent on the ventral aspects of the horse with the pigment 'framing' the

Received: 12 July 2023 | Accepted: 19 August 2023

DOI: 10.1111/age.13352

ANIMAL GENETICS WILEY

Publisher

RESEARCH ARTICLE

A *de novo* 2.3 kb structural variant in *MITF* explains a novel splashed white phenotype in a Thoroughbred family

R. R. Bellone[1,2] | J. Tanaka[1] | E. Esdaile[1] | R. B. Sutton[3] | F. Payette[4] | L. Leduc[4] | B. J. Till[1] | A. K. Abdel-Ghaffar[1] | M. Hammond[1,2] | K. G. Magdesian[5]

[1]Veterinary Genetics Laboratory, School of Veterinary Medicine, UC Davis, Davis, California, USA

[2]Department of Population Health and Reproduction, School of Veterinary Medicine, UC Davis, Davis, California, USA

[3]Cell Physiology and Molecular Biophysics, School of Medicine, Texas Tech University Health Sciences Center, Lubbock, Texas, USA

[4]Department of Clinical Studies, New Bolton Center, University of Pennsylvania School of Veterinary Medicine, University School of Veterinary Medicine, University of Pennsylvania, Kennett Square, Pennsylvania, USA

[5]Department of Medicine and Epidemiology, School of Veterinary Medicine, University of California-Davis, Davis, California, USA

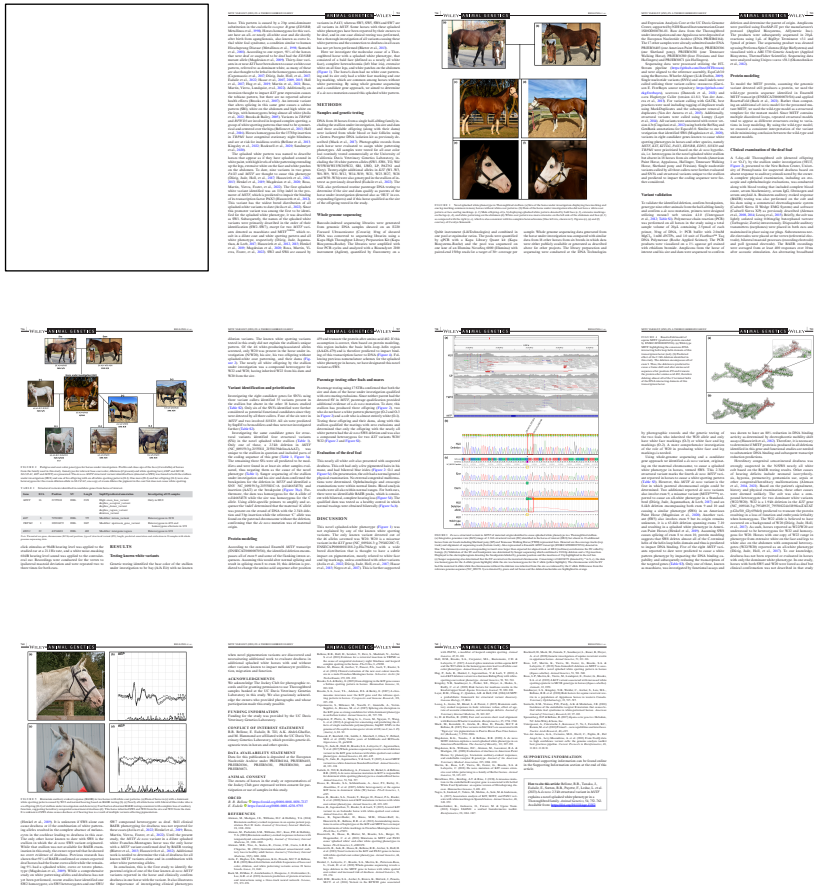
Correspondence
R. R. Bellone, Veterinary Genetics Laboratory, School of Veterinary Medicine, UC Davis, Davis, CA, USA.
Email: rbellone@ucdavis.edu

Funding information
UC Davis Veterinary Genetics Laboratory

Abstract

Splashed white in horses is characterized by extensive white patterning on the legs, face and abdomen and may be accompanied by deafness. To date, seven variants in *microphthalmia-associated transcription factor* (*MITF*) and two variants in *Paired Box 3* (*PAX3*) have been identified to explain this phenotype. A splashed white Thoroughbred stallion, whose sire and dam were not patterned, was hypothesized to have a *de novo* variant leading to his white coat pattern. A whole-genome sequencing candidate gene approach identified two single nucleotide variants (SNVs) in *SOX10*, four SNVs in *MITF* and a 2.3 kb deletion in *MITF* with the alternative allele present in this stallion but absent in the other 18 horses analyzed. All six SNVs were annotated as modifiers and were not further considered. The deletion in *MITF* (NC_009159.3:g.21555811_21558139delinsAAAT) encompasses exon 9 encoding a part of the helix–loop–helix domain required for DNA binding. Sanger sequencing and parentage testing confirmed that this deletion was a *de novo* mutation of maternal origin. Consistent with the published nomenclature, we denote this likely causal variant as SW8. Genotyping three of this stallion's offspring identified SW8 only in the nearly all-white foal that was confirmed deaf by brainstem auditory evoked response testing. This foal was also a compound heterozygote for dominant white variants (W20/W22), but to date, W variants alone have not been connected to deafness. SW8 marks the fourth de novo *MITF* variant in horses reported to cause white patterning. The link between deafness and all *MITF* variants with and without other variants impacting melanocyte development and function needs to be further explored.

KEYWORDS
coat color, deafness, horse, pigmentation, white patterning



INTRODUCTION

Variation in white patterning has long fascinated animal enthusiasts and researchers alike. Moreover, some of these variations have a clinical connection and have similar counterparts in humans or other species. In horses, several different types of white patterning phenotypes have been described including frame overo, tobiano, roan, leopard complex spotting, dominant white, sabino and splashed white (Sponenberg & Bellone, 2017). The molecular bases of many of these have been determined, with the first to be unraveled being the frame overo pattern in which white patterning is prevalent on the ventral aspects of the horse with the pigment 'framing' the

Received: 12 July 2023 | Accepted: 19 August 2023

DOI: 10.1111/age.13352

RESEARCH ARTICLE

Title

ANIMAL GENETICS WILEY

A *de novo* 2.3 kb structural variant in *MITF* explains a novel splashed white phenotype in a Thoroughbred family

R. R. Bellone[1,2] | J. Tanaka[1] | E. Esdaile[1] | R. B. Sutton[3] | F. Payette[4] | L. Leduc[4] | B. J. Till[1] | A. K. Abdel-Ghaffar[1] | M. Hammond[1,2] | K. G. Magdesian[5]

[1]Veterinary Genetics Laboratory, School of Veterinary Medicine, UC Davis, Davis, California, USA

[2]Department of Population Health and Reproduction, School of Veterinary Medicine, UC Davis, Davis, California, USA

[3]Cell Physiology and Molecular Biophysics, School of Medicine, Texas Tech University Health Sciences Center, Lubbock, Texas, USA

[4]Department of Clinical Studies, New Bolton Center, University of Pennsylvania School of Veterinary Medicine, University School of Veterinary Medicine, University of Pennsylvania, Kennett Square, Pennsylvania, USA

[5]Department of Medicine and Epidemiology, School of Veterinary Medicine, University of California-Davis, Davis, California, USA

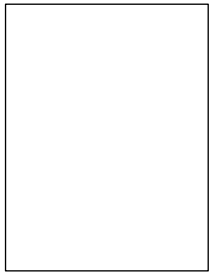
Correspondence
R. R. Bellone, Veterinary Genetics Laboratory, School of Veterinary Medicine, UC Davis, Davis, CA, USA.
Email: rbellone@ucdavis.edu

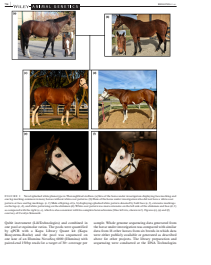
Funding information
UC Davis Veterinary Genetics Laboratory

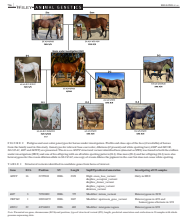
Abstract

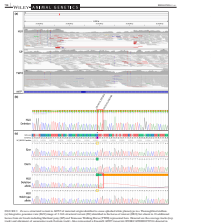
Splashed white in horses is characterized by extensive white patterning on the legs, face and abdomen and may be accompanied by deafness. To date, seven variants in *microphthalmia-associated transcription factor* (*MITF*) and two variants in *Paired Box 3* (*PAX3*) have been identified to explain this phenotype. A splashed white Thoroughbred stallion, whose sire and dam were not patterned, was hypothesized to have a *de novo* variant leading to his white coat pattern. A whole-genome sequencing candidate gene approach identified two single nucleotide variants (SNVs) in *SOX10*, four SNVs in *MITF* and a 2.3 kb deletion in *MITF* with the alternative allele present in this stallion but absent in the other 18 horses analyzed. All six SNVs were annotated as modifiers and were not further considered. The deletion in *MITF* (NC_009159.3:g.21555811_21558139delinsAAAT) encompasses exon 9 encoding a part of the helix–loop–helix domain required for DNA binding. Sanger sequencing and parentage testing confirmed that this deletion was a *de novo* mutation of maternal origin. Consistent with the published nomenclature, we denote this likely causal variant as SW8. Genotyping three of this stallion's offspring identified SW8 only in the nearly all-white foal that was confirmed deaf by brainstem auditory evoked response testing. This foal was also a compound heterozygote for dominant white variants (W20/W22), but to date, W variants alone have not been connected to deafness. SW8 marks the fourth de novo *MITF* variant in horses reported to cause white patterning. The link between deafness and all *MITF* variants with and without other variants impacting melanocyte development and function needs to be further explored.

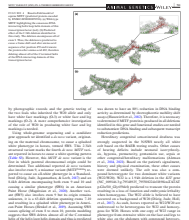
KEYWORDS
coat color, deafness, horse, pigmentation, white patterning











INTRODUCTION

Variation in white patterning has long fascinated animal enthusiasts and researchers alike. Moreover, some of these variations have a clinical connection and have similar counterparts in humans or other species. In horses, several different types of white patterning phenotypes have been described including frame overo, tobiano, roan, leopard complex spotting, dominant white, sabino and splashed white (Sponenberg & Bellone, 2017). The molecular bases of many of these have been determined, with the first to be unraveled being the frame overo pattern in which white patterning is prevalent on the ventral aspects of the horse with the pigment 'framing' the

What is unique about the titles of good research articles?

Received: 12 July 2023 | Accepted: 19 August 2023

DOI: 10.1111/age.13352

RESEARCH ARTICLE

ANIMAL GENETICS WILEY

A *de novo* 2.3 kb structural variant in *MITF* explains a novel splashed white phenotype in a Thoroughbred family

Title

R. R. Bellone[1,2] | J. Tanaka[1] | E. Esdaile[1] | R. B. Sutton[3] | F. Payette[4] | L. Leduc[4] | B. J. Till[1] | A. K. Abdel-Ghaffar[1] | M. Hammond[1,2] | K. G. Magdesian[5]

[1]Veterinary Genetics Laboratory, School of Veterinary Medicine, UC Davis, Davis, California, USA

[2]Department of Population Health and Reproduction, School of Veterinary Medicine, UC Davis, Davis, California, USA

[3]Cell Physiology and Molecular Biophysics, School of Medicine, Texas Tech University Health Sciences Center, Lubbock, Texas, USA

[4]Department of Clinical Studies, New Bolton Center, University of Pennsylvania School of Veterinary Medicine, University School of Veterinary Medicine, University of Pennsylvania, Kennett Square, Pennsylvania, USA

[5]Department of Medicine and Epidemiology, School of Veterinary Medicine, University of California-Davis, Davis, California, USA

Correspondence
R. R. Bellone, Veterinary Genetics Laboratory, School of Veterinary Medicine, UC Davis, Davis, CA, USA.
Email: rbellone@ucdavis.edu

Funding information
UC Davis Veterinary Genetics Laboratory

Abstract

Splashed white in horses is characterized by extensive white patterning on the legs, face and abdomen and may be accompanied by deafness. To date, seven variants in *microphthalmia-associated transcription factor* (*MITF*) and two variants in *Paired Box 3* (*PAX3*) have been identified to explain this phenotype. A splashed white Thoroughbred stallion, whose sire and dam were not patterned, was hypothesized to have a *de novo* variant leading to his white coat pattern. A whole-genome sequencing candidate gene approach identified two single nucleotide variants (SNVs) in *SOX10*, four SNVs in *MITF* and a 2.3 kb deletion in *MITF* with the alternative allele present in this stallion but absent in the other 18 horses analyzed. All six SNVs were annotated as modifiers and were not further considered. The deletion in *MITF* (NC_009159.3:g.21555811_21558139delinsAAAT) encompasses exon 9 encoding a part of the helix–loop–helix domain required for DNA binding. Sanger sequencing and parentage testing confirmed that this deletion was a *de novo* mutation of maternal origin. Consistent with the published nomenclature, we denote this likely causal variant as SW8. Genotyping three of this stallion's offspring identified SW8 only in the nearly all-white foal that was confirmed deaf by brainstem auditory evoked response testing. This foal was also a compound heterozygote for dominant white variants (W20/W22), but to date, W variants alone have not been connected to deafness. SW8 marks the fourth de novo *MITF* variant in horses reported to cause white patterning. The link between deafness and all *MITF* variants with and without other variants impacting melanocyte development and function needs to be further explored.

KEYWORDS
coat color, deafness, horse, pigmentation, white patterning

INTRODUCTION

Variation in white patterning has long fascinated animal enthusiasts and researchers alike. Moreover, some of these variations have a clinical connection and have similar counterparts in humans or other species. In horses, several different types of white patterning phenotypes have been described including frame overo, tobiano, roan, leopard complex spotting, dominant white, sabino and splashed white (Sponenberg & Bellone, 2017). The molecular bases of many of these have been determined, with the first to be unraveled being the frame overo pattern in which white patterning is prevalent on the ventral aspects of the horse with the pigment 'framing' the

Long
Specific
Technical
Contains keywords
Specific terminology

Received: 12 July 2023 | Accepted: 19 August 2023

DOI: 10.1111/age.13352

ANIMAL GENETICS WILEY

RESEARCH ARTICLE

A *de novo* 2.3 kb structu[illegible] explains a novel splashed white phenotyp[illegible] family

Authors

R. R. Bellone[1,2] | J. Tanaka[1] | E. Esdaile[1] | R. B. Sutton[3] | F. Payette[4] | L. Leduc[4] | B. J. Till[1] | A. K. Abdel-Ghaffar[1] | M. Hammond[1,2] | K. G. Magdesian[5]

[1]Veterinary Genetics Laboratory, School of Veterinary Medicine, UC Davis, Davis, California, USA

[2]Department of Population Health and Reproduction, School of Veterinary Medicine, UC Davis, Davis, California, USA

[3]Cell Physiology and Molecular Biophysics, School of Medicine, Texas Tech University Health Sciences Center, Lubbock, Texas, USA

[4]Department of Clinical Studies, New Bolton Center, University of Pennsylvania School of Veterinary Medicine, University School of Veterinary Medicine, University of Pennsylvania, Kennett Square, Pennsylvania, USA

[5]Department of Medicine and Epidemiology, School of Veterinary Medicine, University of California-Davis, Davis, California, USA

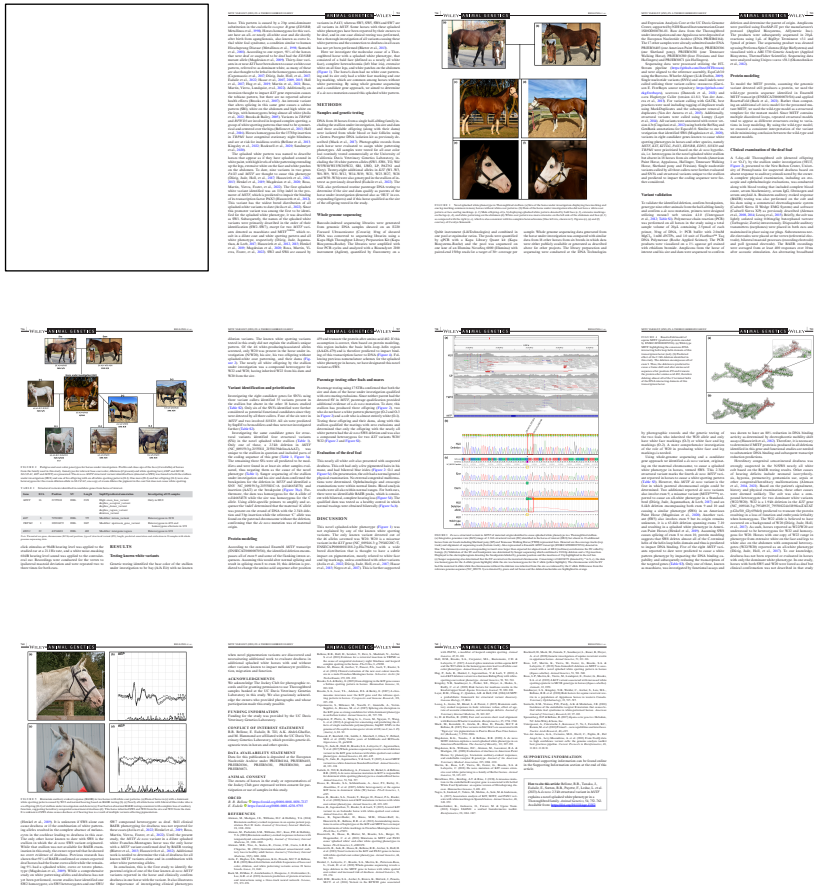
Correspondence
R. R. Bellone, Veterinary Genetics Laboratory, School of Veterinary Medicine, UC Davis, Davis, CA, USA.
Email: rbellone@ucdavis.edu

Funding information
UC Davis Veterinary Genetics Laboratory

Abstract

Splashed white in horses is characterized by extensive white patterning on the legs, face and abdomen and may be accompanied by deafness. To date, seven variants in *microphthalmia-associated transcription factor* (*MITF*) and two variants in *Paired Box 3* (*PAX3*) have been identified to explain this phenotype. A splashed white Thoroughbred stallion, whose sire and dam were not patterned, was hypothesized to have a *de novo* variant leading to his white coat pattern. A whole-genome sequencing candidate gene approach identified two single nucleotide variants (SNVs) in *SOX10*, four SNVs in *MITF* and a 2.3 kb deletion in *MITF* with the alternative allele present in this stallion but absent in the other 18 horses analyzed. All six SNVs were annotated as modifiers and were not further considered. The deletion in *MITF* (NC_009159.3:g.21555811_21558139delinsAAAT) encompasses exon 9 encoding a part of the helix–loop–helix domain required for DNA binding. Sanger sequencing and parentage testing confirmed that this deletion was a *de novo* mutation of maternal origin. Consistent with the published nomenclature, we denote this likely causal variant as SW8. Genotyping three of this stallion's offspring identified SW8 only in the nearly all-white foal that was confirmed deaf by brainstem auditory evoked response testing. This foal was also a compound heterozygote for dominant white variants (W20/W22), but to date, W variants alone have not been connected to deafness. SW8 marks the fourth de novo *MITF* variant in horses reported to cause white patterning. The link between deafness and all *MITF* variants with and without other variants impacting melanocyte development and function needs to be further explored.

KEYWORDS
coat color, deafness, horse, pigmentation, white patterning



INTRODUCTION

Variation in white patterning has long fascinated animal enthusiasts and researchers alike. Moreover, some of these variations have a clinical connection and have similar counterparts in humans or other species. In horses, several different types of white patterning phenotypes have been described including frame overo, tobiano, roan, leopard complex spotting, dominant white, sabino and splashed white (Sponenberg & Bellone, 2017). The molecular bases of many of these have been determined, with the first to be unraveled being the frame overo pattern in which white patterning is prevalent on the ventral aspects of the horse with the pigment 'framing' the

What is unique about the authors of good research articles?

Received: 12 July 2023 | Accepted: 19 August 2023

DOI: 10.1111/age.13352

RESEARCH ARTICLE

ANIMAL GENETICS WILEY

A *de novo* 2.3 kb structural variant in *MITF* explains a novel splashed white phenotyp[Authors]bred family

R. R. Bellone[1,2] | J. Tanaka[1] | E. Esdaile[1] | R. B. Sutton[3] | F. Payette[4] | L. Leduc[4] | B. J. Till[1] | A. K. Abdel-Ghaffar[1] | M. Hammond[1,2] | K. G. Magdesian[5]

[1]Veterinary Genetics Laboratory, School of Veterinary Medicine, UC Davis, Davis, California, USA

[2]Department of Population Health and Reproduction, School of Veterinary Medicine, UC Davis, Davis, California, USA

[3]Cell Physiology and Molecular Biophysics, School of Medicine, Texas Tech University Health Sciences Center, Lubbock, Texas, USA

[4]Department of Clinical Studies, New Bolton Center, University of Pennsylvania School of Veterinary Medicine, University School of Veterinary Medicine, University of Pennsylvania, Kennett Square, Pennsylvania, USA

[5]Department of Medicine and Epidemiology, School of Veterinary Medicine, University of California-Davis, Davis, California, USA

Correspondence
R. R. Bellone, Veterinary Genetics Laboratory, School of Veterinary Medicine, UC Davis, Davis, CA, USA.
Email: rbellone@ucdavis.edu

Funding information
UC Davis Veterinary Genetics Laboratory

Abstract

Splashed white in horses is characterized by extensive white patterning on the legs, face and abdomen and may be accompanied by deafness. To date, seven variants in *microphthalmia-associated transcription factor* (*MITF*) and two variants in *Paired Box 3* (*PAX3*) have been identified to explain this phenotype. A splashed white Thoroughbred stallion, whose sire and dam were not patterned, was hypothesized to have a *de novo* variant leading to his white coat pattern. A whole-genome sequencing candidate gene approach identified two single nucleotide variants (SNVs) in *SOX10*, four SNVs in *MITF* and a 2.3 kb deletion in *MITF* with the alternative allele present in this stallion but absent in the other 18 horses analyzed. All six SNVs were annotated as modifiers and were not further considered. The deletion in *MITF* (NC_009159.3:g.21555811_21558139delinsAAAT) encompasses exon 9 encoding a part of the helix–loop–helix domain required for DNA binding. Sanger sequencing and parentage testing confirmed that this deletion was a *de novo* mutation of maternal origin. Consistent with the published nomenclature, we denote this likely causal variant as SW8. Genotyping three of this stallion's offspring identified SW8 only in the nearly all-white foal that was confirmed deaf by brainstem auditory evoked response testing. This foal was also a compound heterozygote for dominant white variants (W20/W22), but to date, W variants alone have not been connected to deafness. SW8 marks the fourth de novo *MITF* variant in horses reported to cause white patterning. The link between deafness and all *MITF* variants with and without other variants impacting melanocyte development and function needs to be further explored.

KEYWORDS
coat color, deafness, horse, pigmentation, white patterning

INTRODUCTION

Variation in white patterning has long fascinated animal enthusiasts and researchers alike. Moreover, some of these variations have a clinical connection and have similar counterparts in humans or other species. In horses, several different types of white patterning phenotypes have been described including frame overo, tobiano, roan, leopard complex spotting, dominant white, sabino and splashed white (Sponenberg & Bellone, 2017). The molecular bases of many of these have been determined, with the first to be unraveled being the frame overo pattern in which white patterning is prevalent on the ventral aspects of the horse with the pigment 'framing' the

Several authors
Variable expertise
corresponding author
often senior

Received: 12 July 2023 | Accepted: 19 August 2023

DOI: 10.1111/age.13352

RESEARCH ARTICLE

ANIMAL GENETICS WILEY

A *de novo* 2.3 kb structural variant in *MITF* explains a novel splashed white phenotype in a Thoroughbred family

R. R. Bellone[1,2] | J. Tanaka[1] | E. Esdaile[1] | R. B. Sutton[3] | F. Payette[4] | L. Leduc[4] | B. J. Till[1] | A. K. Abdel-Ghaffar[1] | M. Hammond[1,2] | K. G. Magdesian[5]

[1]Veterinary Genetics Laboratory, School of Veterinary Medicine, UC Davis, Davis, California, USA

[2]Department of Population Health and Reproduction, School of Veterinary Medicine, UC Davis, Davis, California, USA

[3]Cell Physiology and Molecular Biophysics, School of Medicine, Texas Tech University Health Sciences Center, Lubbock, Texas, USA

[4]Department of Clinical Studies, New Bolton Center, University of Pennsylvania School of Veterinary Medicine, University School of Veterinary Medicine, University of Pennsylvania, Kennett Square, Pennsylvania, USA

[5]Department of Medicine and Epidemiology, School of Veterinary Medicine, University of California-Davis, Davis, California, USA

Correspondence
R. R. Bellone, Veterinary Genetics Laboratory, School of Veterinary Medicine, UC Davis, Davis, CA, USA.
Email: rbellone@ucdavis.edu

Funding information
UC Davis Veterinary Genetics Laboratory

Abstract

Splashed white in horses is characterized by extensive white patterning on the legs, face and abdomen and may be accompanied by deafness. To date, seven variants in *microphthalmia-associated transcription factor* (*MITF*) and two variants in *Paired Box 3* (*PAX3*) have been identified to explain this phenotype. ... am were not ... his white coat ... identified two ... *F* and a 2.3 kb ... on but absent ... modifiers and were not further considered. The deletion in *MITF* (NC_009159.3:g.21555811_21558139delinsAAAT) encompasses exon 9 encoding a part of the helix–loop–helix domain required for DNA binding. Sanger sequencing and parentage testing confirmed that this deletion was a *de novo* mutation of maternal origin. Consistent with the published nomenclature, we denote this likely causal variant as SW8. Genotyping three of this stallion's offspring identified SW8 only in the nearly all-white foal that was confirmed deaf by brainstem auditory evoked response testing. This foal was also a compound heterozygote for dominant white variants (W20/W22), but to date, W variants alone have not been connected to deafness. SW8 marks the fourth de novo *MITF* variant in horses reported to cause white patterning. The link between deafness and all *MITF* variants with and without other variants impacting melanocyte development and function needs to be further explored.

Author's affiliation & contact info

KEYWORDS
coat color, deafness, horse, pigmentation, white patterning

INTRODUCTION

Variation in white patterning has long fascinated animal enthusiasts and researchers alike. Moreover, some of these variations have a clinical connection and have similar counterparts in humans or other species. In horses, several different types of white patterning phenotypes have been described including frame overo, tobiano, roan, leopard complex spotting, dominant white, sabino and splashed white (Sponenberg & Bellone, 2017). The molecular bases of many of these have been determined, with the first to be unraveled being the frame overo pattern in which white patterning is prevalent on the ventral aspects of the horse with the pigment 'framing' the

Received: 12 July 2023 | Accepted: 19 August 2023

DOI: 10.1111/age.13352

RESEARCH ARTICLE

ANIMAL GENETICS WILEY

A *de novo* 2.3 kb structural variant in *MITF* explains a novel splashed white phenotype in a Thoroughbred family

R. R. Bellone[1,2] | J. Tanaka[1] | E. Esdaile[1] | R. B. Sutton[3] | F. Payette[4] | L. Leduc[4] | B. J. Till[1] | A. K. Abdel-Ghaffar[1] | M. Hammond[1,2] | K. G. Magdesian[5]

[1]Veterinary Genetics Laboratory, School of Veterinary Medicine, UC Davis, Davis, California, USA

[2]Department of Population Health and Reproduction, School of Veterinary Medicine, UC Davis, Davis, California, USA

[3]Cell Physiology and Molecular Biophysics, School of Medicine, Texas Tech University Health Sciences Center, Lubbock, Texas, USA

[4]Department of Clinical Studies, New Bolton Center, University of Pennsylvania School of Veterinary Medicine, University School of Veterinary Medicine, University of Pennsylvania, Kennett Square, Pennsylvania, USA

[5]Department of Medicine and Epidemiology, School of Veterinary Medicine, University of California-Davis, Davis, California, USA

Correspondence
R. R. Bellone, Veterinary Genetics Laboratory, School of Veterinary Medicine, UC Davis, Davis, CA, USA.
Email: rbellone@ucdavis.edu

Funding information
UC Davis Veterinary Genetics Laboratory

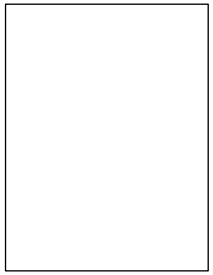
Abstract

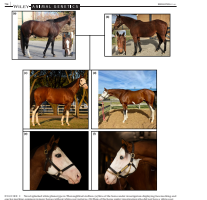
Splashed white in horses is characterized by extensive white patterning on the legs, face and abdomen and may be accompanied by deafness. To date, seven variants in *microphthalmia-associated transcription factor* (*MITF*) and two variants in *Paired Box 3* (*PAX3*) have been identified to explain this phenotype. A splashed white Thoroughbred stallion, whose sire and dam were not patterned, was hypothesized to have a *de novo* variant leading to his white coat pattern. A whole-genome sequencing candidate gene approach identified two single nucleotide variants (SNVs) in *SOX10*, four SNVs in *MITF* and a 2.3 kb deletion in *MITF* with the alternative allele present in this stallion but absent in the other 18 horses analyzed. All six SNVs were annotated as modifiers and were not further considered. The deletion in *MITF* (NC_009159.3:g.21555811_21558139delinsAAAT) encompasses exon 9 encoding a part of the helix–loop–helix domain required for DNA binding. Sanger sequencing and parentage testing confirmed that this deletion was a *de novo* mutation of maternal origin. Consistent with the published nomenclature, we denote this likely causal variant

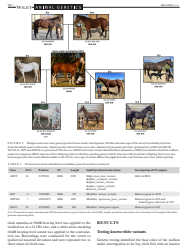
Corresponding author contacts and email

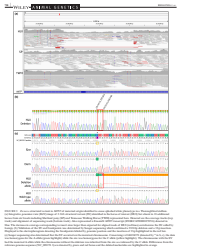
8 only in the tory evoked r dominant e not been nt in horses d all *MITF* variants with and without other variants impacting melanocyte development and function needs to be further explored.

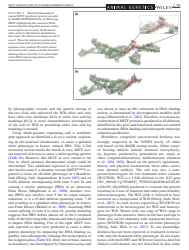
KEYWORDS
coat color, deafness, horse, pigmentation, white patterning











INTRODUCTION

Variation in white patterning has long fascinated animal enthusiasts and researchers alike. Moreover, some of these variations have a clinical connection and have similar counterparts in humans or other species. In horses, several different types of white patterning phenotypes have been described including frame overo, tobiano, roan, leopard complex spotting, dominant white, sabino and splashed white (Sponenberg & Bellone, 2017). The molecular bases of many of these have been determined, with the first to be unraveled being the frame overo pattern in which white patterning is prevalent on the ventral aspects of the horse with the pigment 'framing' the

Received: 12 July 2023 | Accepted: 19 August 2023

DOI: 10.1111/age.13352

RESEARCH ARTICLE

ANIMAL GENETICS WILEY

A *de novo* 2.3 kb structural variant in *MITF* explains a novel splashed white phenotype in a Thoroughbred family

R. R. Bellone[1,2] | J. Tanaka[1] | E. Esdaile[1] | R. B. Sutton[3] | F. Payette[4] | L. Leduc[4] | B. J. Till[1] | A. K. Abdel-Ghaffar[1] | M. Hammond[1,2] | K. G. Magdesian[5]

[1]Veterinary Genetics Laboratory, School of Veterinary Medicine, UC Davis, Davis, California, USA

[2]Department of Population Health and Reproduction, School of Veterinary Medicine, UC Davis, Davis, California, USA

[3]Cell Physiology and Molecular Biophysics, School of Medicine, Texas Tech University Health Sciences Center, Lubbock, Texas, USA

[4]Department of Clinical Studies, New Bolton Center, University of Pennsylvania School of Veterinary Medicine, University School of Veterinary Medicine, University of Pennsylvania, Kennett Square, Pennsylvania, USA

[5]Department of Medicine and Epidemiology, School of Veterinary Medicine, University of California-Davis, Davis, California, USA

Correspondence
R. R. Bellone, Veterinary Genetics Laboratory, School of Veterinary Medicine, UC Davis, Davis, CA, USA.
Email: rbellone@ucdavis.edu

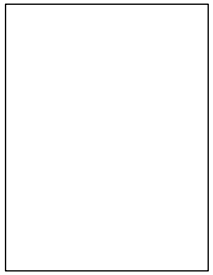
Funding information
UC Davis Veterinary Genetics Laboratory

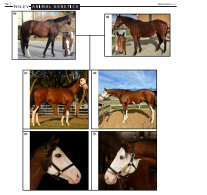
Funding

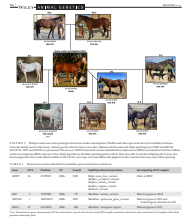
Abstract

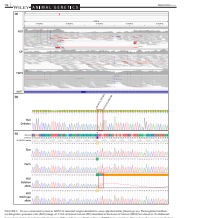
Splashed white in horses is characterized by extensive white patterning on the legs, face and abdomen and may be accompanied by deafness. To date, seven variants in *microphthalmia-associated transcription factor* (*MITF*) and two variants in *Paired Box 3* (*PAX3*) have been identified to explain this phenotype. A splashed white Thoroughbred stallion, whose sire and dam were not patterned, was hypothesized to have a *de novo* variant leading to his white coat pattern. A whole-genome sequencing candidate gene approach identified two single nucleotide variants (SNVs) in *SOX10*, four SNVs in *MITF* and a 2.3 kb deletion in *MITF* with the alternative allele present in this stallion but absent in the other 18 horses analyzed. All six SNVs were annotated as modifiers and were not further considered. The deletion in *MITF* (NC_009159.3:g.21555811_21558139delinsAAAT) encompasses exon 9 encoding a part of the helix–loop–helix domain required for DNA binding. Sanger sequencing and parentage testing confirmed that this deletion was a *de novo* mutation of maternal origin. Consistent with the published nomenclature, we denote this likely causal variant as SW8. Genotyping three of this stallion's offspring identified SW8 only in the nearly all-white foal that was confirmed deaf by brainstem auditory evoked response testing. This foal was also a compound heterozygote for dominant white variants (W20/W22), but to date, W variants alone have not been connected to deafness. SW8 marks the fourth de novo *MITF* variant in horses reported to cause white patterning. The link between deafness and all *MITF* variants with and without other variants impacting melanocyte development and function needs to be further explored.

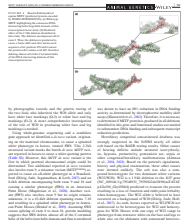
KEYWORDS
coat color, deafness, horse, pigmentation, white patterning











INTRODUCTION

Variation in white patterning has long fascinated animal enthusiasts and researchers alike. Moreover, some of these variations have a clinical connection and have similar counterparts in humans or other species. In horses, several different types of white patterning phenotypes have been described including frame overo, tobiano, roan, leopard complex spotting, dominant white, sabino and splashed white (Sponenberg & Bellone, 2017). The molecular bases of many of these have been determined, with the first to be unraveled being the frame overo pattern in which white patterning is prevalent on the ventral aspects of the horse with the pigment 'framing' the

Received: 12 July 2023 | Accepted: 19 August 2023

DOI: 10.1111/age.13352

RESEARCH ARTICLE

ANIMAL GENETICS WILEY

A *de novo* 2.3 kb structural variant in *MITF* explains a novel splashed white phenotype in a Thoroughbred family

R. R. Bellone[1,2] | J. Tanaka[1] | E. Esdaile[1] | R. B. Sutton[3] | F. Payette[4] | L. Leduc[4] | B. J. Till[1] | A. K. Abdel-Ghaffar[1] | M. Hammond[1,2] | K. G. Magdesian[5]

[1]Veterinary Genetics Laboratory, School of Veterinary Medicine, UC Davis, Davis, California, USA

[2]Department of Population Health and Reproduction, School of Veterinary Medicine, UC Davis, Davis, California, USA

[3]Cell Physiology and Molecular Biophysics, School of Medicine, Texas Tech University Health Sciences Center, Lubbock, Texas, USA

[4]Department of Clinical Studies, New Bolton Center, University of Pennsylvania School of Veterinary Medicine, University School of Veterinary Medicine, University of Pennsylvania, Kennett Square, Pennsylvania, USA

[5]Department of Medicine and Epidemiology, School of Veterinary Medicine, University of California-Davis, Davis, California, USA

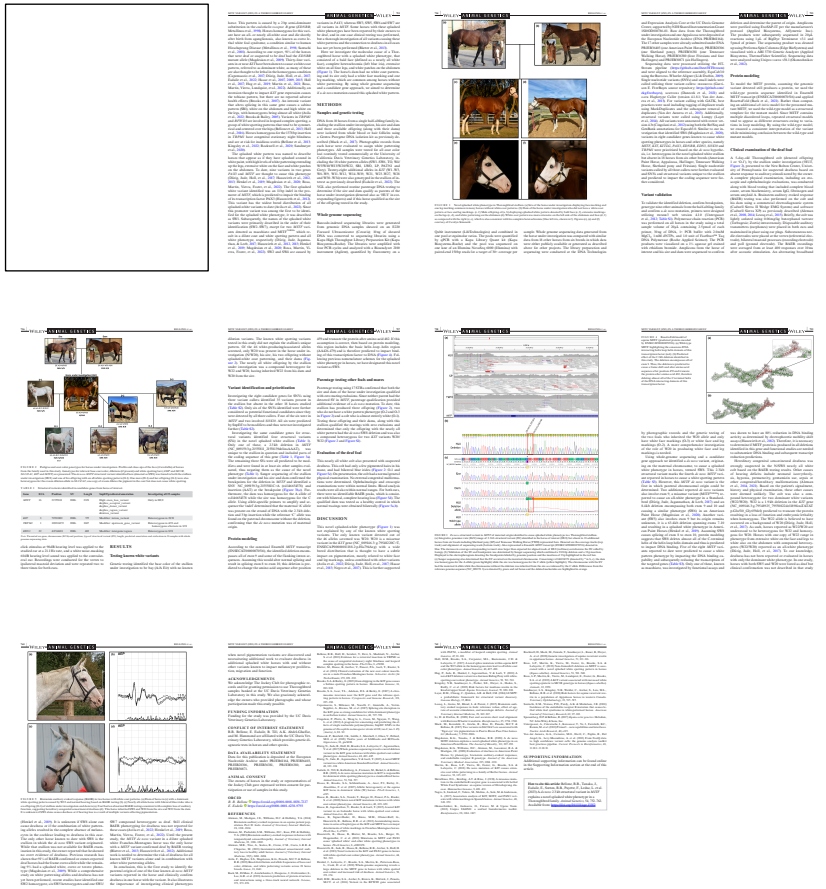
Correspondence
R. R. Bellone, Veterinary Genetics Laboratory, School of Veterinary Medicine, UC Davis, Davis, CA, USA.
Email: rbellone@ucdavis.edu

Funding information
UC Davis Veterinary Genetics Laboratory

Abstract

Splashed white in horses is characterized by extensive white patterning on the legs, face and abdomen and may be accompanied by deafness. To date, seven variants in *microphthalmia-associated transcription factor* (*MITF*) and two variants in *Paired Box 3* (*PAX3*) have been identified to explain this phenotype. A splashed white Thoroughbred stallion, whose sire and dam were not patterned, was hypothesized to have a *de novo* variant leading to his white coat pattern. A whole-genome sequencing candidate gene approach identified two single nucleotide variants (SNVs) in *SOX10*, four SNVs in *MITF* and a 2.3 kb deletion in *MITF* with the alternative allele present in this stallion but absent in the other 18 horses analyzed. All six SNVs were annotated as modifiers and were not further considered. The deletion in *MITF* (NC_009159.3:g.21555811_21558139delinsAAAT) encompasses exon 9 encoding a part of the helix–loop–helix domain required for DNA binding. Sanger sequencing and parentage testing confirmed that this deletion was a *de novo* mutation of maternal origin. Consistent with the published nomenclature, we denote this likely causal variant as SW8. Genotyping three of this stallion's offspring identified SW8 only in the nearly all-white foal that was confirmed deaf by brainstem auditory evoked response testing. This foal was also a compound heterozygote for dominant white variants (W20/W22), but to date, W variants alone have not been connected to deafness. SW8 marks the fourth de novo *MITF* variant in horses reported to cause white patterning. The link between deafness and all *MITF* variants with and without other variants impacting melanocyte development and function needs to be further explored.

KEYWORDS
coat color, deafness, horse, pigmentation, white patterning



INTRODUCTION

Variation in white patterning has long fascinated animal enthusiasts and researchers alike. Moreover, some of these variations have a clinical connection and have similar counterparts in humans or other species. In horses, several different types of white patterning phenotypes have been described including frame overo, tobiano, roan, leopard complex spotting, dominant white, sabino and splashed white (Sponenberg & Bellone, 2017). The molecular bases of many of these have been determined, with the first to be unraveled being the frame overo pattern in which white patterning is prevalent on the ventral aspects of the horse with the pigment 'framing' the

Volume, issue, page #s

Received: 12 July 2023 | Accepted: 19 August 2023

DOI: 10.1111/age.13352

RESEARCH ARTICLE

ANIMAL GENETICS WILEY

A *de novo* 2.3 kb structural variant in *MITF* explains a novel splashed white phenotype in a Thoroughbred family

R. R. Bellone[1,2] | J. Tanaka[1] | E. Esdaile[1] | R. B. Sutton[3] | F. Payette[4] | L. Leduc[4] | B. J. Till[1] | A. K. Abdel-Ghaffar[1] | M. Hammond[1,2] | K. G. Magdesian[5]

[1]Veterinary Genetics Laboratory, School of Veterinary Medicine, UC Davis, Davis, California, USA

[2]Department of Population Health and Reproduction, School of Veterinary Medicine, UC Davis, Davis, California, USA

[3]Cell Physiology and Molecular Biophysics, School of Medicine, Texas Tech University Health Sciences Center, Lubbock, Texas, USA

[4]Department of Clinical Studies, New Bolton Center, University of Pennsylvania School of Veterinary Medicine, University School of Veterinary Medicine, University of Pennsylvania, Kennett Square, Pennsylvania, USA

[5]Department of Medicine and Epidemiology, School of Veterinary Medicine, University of California-Davis, Davis, California, USA

Correspondence
R. R. Bellone, Veterinary Genetics Laboratory, School of Veterinary Medicine, UC Davis, Davis, CA, USA.
Email: rbellone@ucdavis.edu

Funding information
UC Davis Veterinary Genetics Laboratory

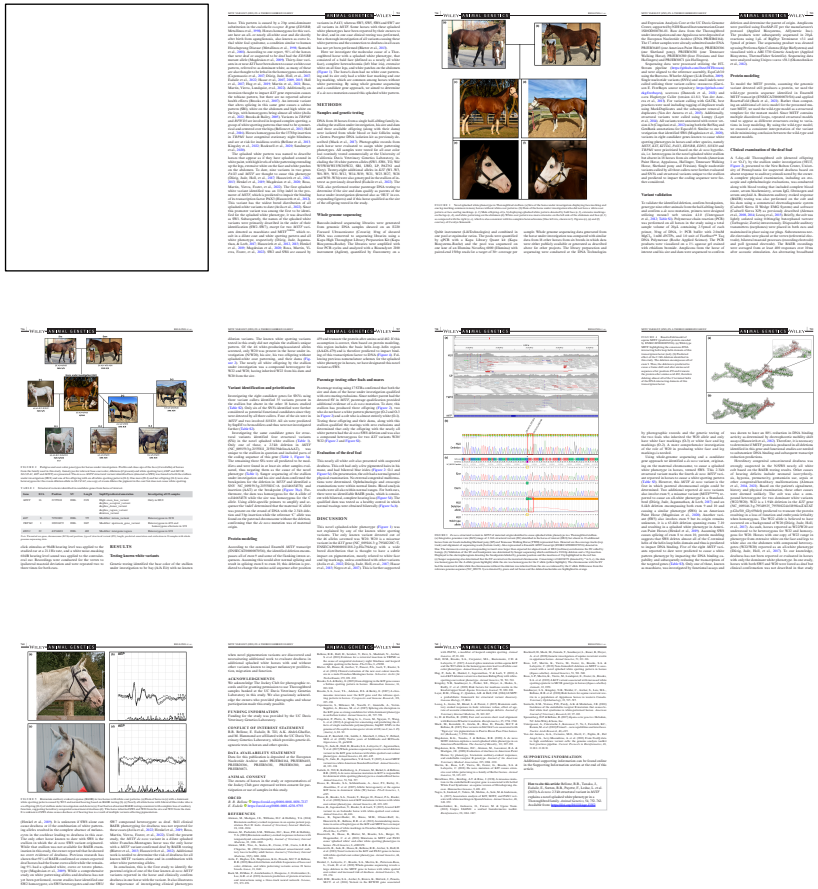
Abstract

Splashed white in horses is characterized by extensive white patterning on the legs, face and abdomen and may be accompanied by deafness. To date, seven variants in *microphthalmia-associated transcription factor* (*MITF*) and two variants in *Paired Box 3* (*PAX3*) have been identified to explain this phenotype. A splashed white Thoroughbred stallion, whose sire and dam were not patterned, was hypothesized to have a *de novo* variant leading to his white coat pattern. A whole-genome sequencing candidate gene approach identified two single nucleotide variants (SNVs) in *SOX10*, four SNVs in *MITF* and a 2.3 kb deletion in *MITF* with the alternative allele present in this stallion but absent in the other 18 horses [illegible] annotated as modifiers and were not further cons[illegible]7 (NC_009159.3:g.21555811_21558139delinsAAAT[illegible]g a part of the helix–loop–helix domain require[illegible] sequencing and parentage testing confirmed that this deletion was a *de novo* mutation of maternal origin. Consistent with the published nomenclature, we denote this likely causal variant as SW8. Genotyping three of this stallion's offspring identified SW8 only in the nearly all-white foal that was confirmed deaf by brainstem auditory evoked response testing. This foal was also a compound heterozygote for dominant white variants (W20/W22), but to date, W variants alone have not been connected to deafness. SW8 marks the fourth de novo *MITF* variant in horses reported to cause white patterning. The link between deafness and all *MITF* variants with and without other variants impacting melanocyte development and function needs to be further explored.

KEYWORDS
coat color, deafness, horse, pigmentation, white patterning

INTRODUCTION

Variation in white patterning has long fascinated animal enthusiasts and researchers alike. Moreover, some of these variations have a clinical connection and have similar counterparts in humans or other species. In horses, several different types of white patterning phenotypes have been described including frame overo, tobiano, roan, leopard complex spotting, dominant white, sabino and splashed white (Sponenberg & Bellone, 2017). The molecular bases of many of these have been determined, with the first to be unraveled being the frame overo pattern in which white patterning is prevalent on the ventral aspects of the horse with the pigment 'framing' the



Received: 12 July 2023 | Accepted: 19 August 2023

DOI: 10.1111/age.13352

RESEARCH ARTICLE

ANIMAL GENETICS WILEY

A *de novo* 2.3 kb structural variant in *MITF* explains a novel splashed white phenotype in a Thoroughbred family

R. R. Bellone[1,2] | J. Tanaka[1] | E. Esdaile[1] | R. B. Sutton[3] | F. Payette[4] | L. Leduc[4] | B. J. Till[1] | A. K. Abdel-Ghaffar[1] | M. Hammond[1,2] | K. G. Magdesian[5]

[1]Veterinary Genetics Laboratory, School of Veterinary Medicine, UC Davis, Davis, California, USA

[2]Department of Population Health and Reproduction, School of Veterinary Medicine, UC Davis, Davis, California, USA

[3]Cell Physiology and Molecular Biophysics, School of Medicine, Texas Tech University Health Sciences Center, Lubbock, Texas, USA

[4]Department of Clinical Studies, New Bolton Center, University of Pennsylvania School of Veterinary Medicine, University School of Veterinary Medicine, University of Pennsylvania, Kennett Square, Pennsylvania, USA

[5]Department of Medicine and Epidemiology, School of Veterinary Medicine, University of California-Davis, Davis, California, USA

Correspondence
R. R. Bellone, Veterinary Genetics Laboratory, School of Veterinary Medicine, UC Davis, Davis, CA, USA.
Email: rbellone@ucdavis.edu

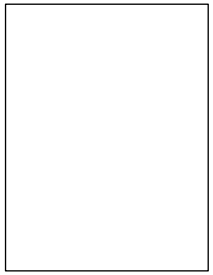
Funding information
UC Davis Veterinary Genetics Laboratory

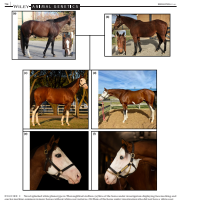
Abstract

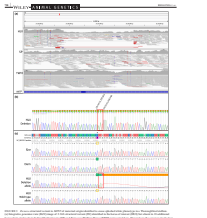
Splashed white in horses is characterized by extensive white patterning on the legs, face and abdomen and may be accompanied by deafness. To date, seven variants in *microphthalmia-associated transcription factor* (*MITF*) and two variants in *Paired Box 3* (*PAX3*) have been identified to explain this phenotype. A splashed white Thoroughbred stallion, whose sire and dam were not patterned, was hypothesized to have a *de novo* variant leading to his white coat pattern. A whole-genome sequencing candidate gene approach identified two single nucleotide variants (SNVs) in *SOX10*, four SNVs in *MITF* and a 2.3 kb deletion in *MITF* with the alternative allele present in this stallion but absent in the other 18 horses analyzed. All six SNVs were annotated as modifiers and were not further considered. The deletion in *MITF* (NC_009159.3:g.21555811_21558139delinsAAAT) encompasses exon 9 encoding a part of the helix–loop–helix domain required for DNA binding. Sanger sequencing and parentage testing confirmed that this deletion was a *de novo* mutation of maternal origin. Consistent with the published nomenclature, we denote this likely causal variant as SW8. Genotyping three of this stallion's offspring identified SW8 only in the nearly all-white foal that was confirmed deaf by brainstem auditory evoked response testing. This foal was also a compound heterozygote for dominant white variants (W20/W22), but to date, W variants alone have not been connected to deafness. SW8 marks the fourth de novo *MITF* variant in horses reported to cause white patterning. The link between deafness and all *MITF* variants with and without other variants impacting melanocyte development and function needs to be further explored.

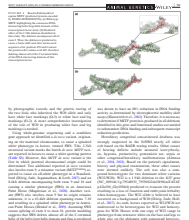
KEYWORDS
coat color, deafness, horse, pigmentation, white patterning

Keywords









INTRODUCTION

Variation in white patterning has long fascinated animal enthusiasts and researchers alike. Moreover, some of these variations have a clinical connection and have similar counterparts in humans or other species. In horses, several different types of white patterning phenotypes have been described including frame overo, tobiano, roan, leopard complex spotting, dominant white, sabino and splashed white (Sponenberg & Bellone, 2017). The molecular bases of many of these have been determined, with the first to be unraveled being the frame overo pattern in which white patterning is prevalent on the ventral aspects of the horse with the pigment 'framing' the

Received: 12 July 2023 | Accepted: 19 August 2023

DOI: 10.1111/age.13352

RESEARCH ARTICLE

ANIMAL GENETICS WILEY

A *de novo* 2.3 kb structural variant in *MITF* explains a novel splashed white phenotype in a Thoroughbred family

R. R. Bellone[1,2] | J. Tanaka[1] | E. Esdaile[1] | R. B. Sutton[3] | F. Payette[4] | L. Leduc[4] | B. J. Till[1] | A. K. Abdel-Ghaffar[1] | M. Hammond[1,2] | K. G. Magdesian[5]

[1]Veterinary Genetics Laboratory, School of Veterinary Medicine, UC Davis, Davis, California, USA

[2]Department of Population Health and Reproduction, School of Veterinary Medicine, UC Davis, Davis, California, USA

[3]Cell Physiology and Molecular Biophysics, School of Medicine, Texas Tech University Health Sciences Center, Lubbock, Texas, USA

[4]Department of Clinical Studies, New

[5]Department of Medicine and Epidemiology, School of Veterinary Medicine, University of California-Davis, Davis, California, USA

Correspondence
R. R. Bellone, Veterinary Genetics Laboratory, School of Veterinary Medicine, UC Davis, Davis, CA, USA.
Email: rbellone@ucdavis.edu

Funding information
UC Davis Veterinary Genetics Laboratory

Abstract

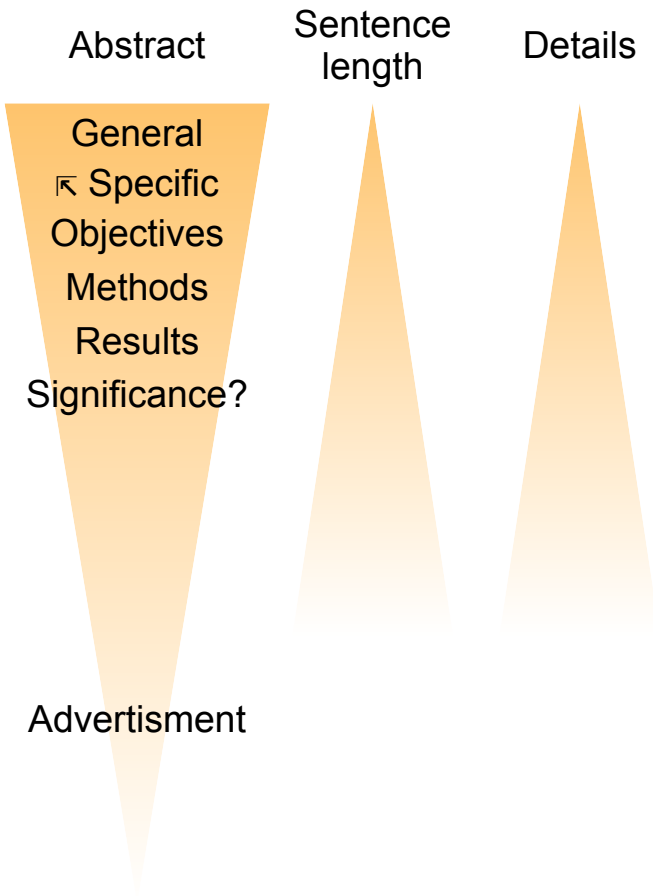
Abstract

Splashed white in horses is characterized by extensive white patterning on the legs, face and abdomen and may be accompanied by deafness. To date, seven variants in *microphthalmia-associated transcription factor* (*MITF*) and two variants in *Paired Box 3* (*PAX3*) have been identified to explain this phenotype. A splashed white Thoroughbred stallion, whose sire and dam were not patterned, was hypothesized to have a *de novo* variant leading to his white coat pattern. A whole-genome sequencing candidate gene approach identified two single nucleotide variants (SNVs) in *SOX10*, four SNVs in *MITF* and a 2.3 kb deletion in *MITF* with the alternative allele present in this stallion but absent in the other 18 horses analyzed. All six SNVs were annotated as modifiers and were not further considered. The deletion in *MITF* (NC_009159.3:g.21555811_21558139delinsAAAT) encompasses exon 9 encoding a part of the helix–loop–helix domain required for DNA binding. Sanger sequencing and parentage testing confirmed that this deletion was a *de novo* mutation of maternal origin. Consistent with the published nomenclature, we denote this likely causal variant as SW8. Genotyping three of this stallion's offspring identified SW8 only in the nearly all-white foal that was confirmed deaf by brainstem auditory evoked response testing. This foal was also a compound heterozygote for dominant white variants (W20/W22), but to date, W variants alone have not been connected to deafness. SW8 marks the fourth de novo *MITF* variant in horses reported to cause white patterning. The link between deafness and all *MITF* variants with and without other variants impacting melanocyte development and function needs to be further explored.

KEYWORDS
coat color, deafness, horse, pigmentation, white patterning

INTRODUCTION

Variation in white patterning has long fascinated animal enthusiasts and researchers alike. Moreover, some of these variations have a clinical connection and have similar counterparts in humans or other species. In horses, several different types of white patterning phenotypes have been described including frame overo, tobiano, roan, leopard complex spotting, dominant white, sabino and splashed white (Sponenberg & Bellone, 2017). The molecular bases of many of these have been determined, with the first to be unraveled being the frame overo pattern in which white patterning is prevalent on the ventral aspects of the horse with the pigment 'framing' the



Received: 12 July 2023 | Accepted: 19 August 2023

DOI: 10.1111/age.13352

RESEARCH ARTICLE

ANIMAL GENETICS WILEY

A *de novo* 2.3 kb structural variant in *MITF* explains a novel splashed white phenotype in a Thoroughbred family

R. R. Bellone[1,2] | J. Tanaka[1] | E. Esdaile[1] | R. B. Sutton[3] | F. Payette[4] | L. Leduc[4] | B. J. Till[1] | A. K. Abdel-Ghaffar[1] | M. Hammond[1,2] | K. G. Magdesian[5]

[1]Veterinary Genetics Laboratory, School of Veterinary Medicine, UC Davis, Davis, California, USA

[2]Department of Population Health and Reproduction, School of Veterinary Medicine, UC Davis, Davis, California, USA

[3]Cell Physiology and Molecular Biophysics, School of Medicine, Texas Tech University Health Sciences Center, Lubbock, Texas, USA

[4]Department of Clinical Studies, New Bolton Center, University of Pennsylvania School of Veterinary Medicine, University School of Veterinary Medicine, University of Pennsylvania, Kennett Square, Pennsylvania, USA

[5]Department of Medicine and Epidemiology, School of Veterinary Medicine, University of California-Davis, Davis, California, USA

Correspondence
R. R. Bellone, Veterinary Genetics Laboratory, School of Veterinary Medicine, UC Davis, Davis, CA, USA.
Email: rbellone@ucdavis.edu

Funding information
UC Davis Veterinary Genetics Laboratory

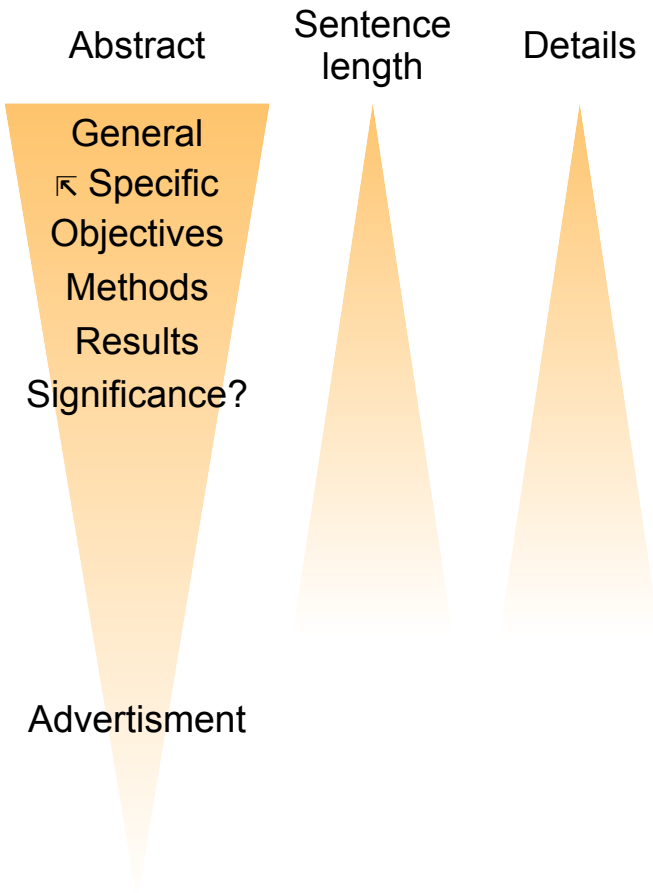
Abstract

Splashed white in horses is characterized by extensive white patterning on the legs, face and abdomen and may be accompanied by deafness.

KEYWORDS
coat color, deafness, horse, pigmentation, white patterning

INTRODUCTION

Variation in white patterning has long fascinated animal enthusiasts and researchers alike. Moreover, some of these variations have a clinical connection and have similar counterparts in humans or other species. In horses, several different types of white patterning phenotypes have been described including frame overo, tobiano, roan, leopard complex spotting, dominant white, sabino and splashed white (Sponenberg & Bellone, 2017). The molecular bases of many of these have been determined, with the first to be unraveled being the frame overo pattern in which white patterning is prevalent on the ventral aspects of the horse with the pigment 'framing' the



Received: 12 July 2023 | Accepted: 19 August 2023

DOI: 10.1111/age.13352

RESEARCH ARTICLE

ANIMAL GENETICS WILEY

A *de novo* 2.3 kb structural variant in *MITF* explains a novel splashed white phenotype in a Thoroughbred family

R. R. Bellone[1,2] | J. Tanaka[1] | E. Esdaile[1] | R. B. Sutton[3] | F. Payette[4] | L. Leduc[4] | B. J. Till[1] | A. K. Abdel-Ghaffar[1] | M. Hammond[1,2] | K. G. Magdesian[5]

[1]Veterinary Genetics Laboratory, School of Veterinary Medicine, UC Davis, Davis, California, USA

[2]Department of Population Health and Reproduction, School of Veterinary Medicine, UC Davis, Davis, California, USA

[3]Cell Physiology and Molecular Biophysics, School of Medicine, Texas Tech University Health Sciences Center, Lubbock, Texas, USA

[4]Department of Clinical Studies, New Bolton Center, University of Pennsylvania School of Veterinary Medicine, University School of Veterinary Medicine, University of Pennsylvania, Kennett Square, Pennsylvania, USA

[5]Department of Medicine and Epidemiology, School of Veterinary Medicine, University of California-Davis, Davis, California, USA

Correspondence
R. R. Bellone, Veterinary Genetics Laboratory, School of Veterinary Medicine, UC Davis, Davis, CA, USA.
Email: rbellone@ucdavis.edu

Funding information
UC Davis Veterinary Genetics Laboratory

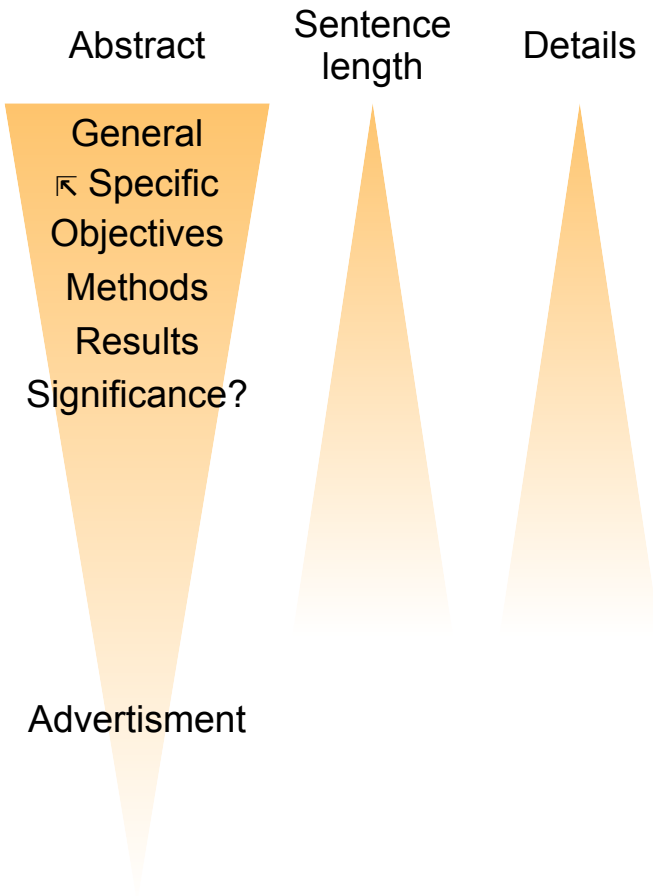
Abstract

To date, seven variants in *microphthalmia-associated transcription factor* (*MITF*) and two variants in *Paired Box 3* (*PAX3*) have been identified to explain this phenotype. A splashed white Thoroughbred stallion, whose sire and dam were not patterned, was hypothesized to have a *de novo* variant leading to his white coat pattern. A whole-genome sequencing candidate gene approach identified two single nucleotide variants (SNVs) in *SOX10*, four SNVs in *MITF* and a 2.3 kb deletion in *MITF* with the alternative allele present in this stallion but absent in the other 18 horses analyzed. All six SNVs were annotated as modifiers and were not further considered. The deletion in *MITF* (NC_009159.3:g.21555811_21558139delinsAAAT) encompasses exon 9 encoding a part of the helix–loop–helix domain required for DNA binding. Sanger sequencing and parentage testing confirmed that this deletion was a *de novo* mutation of maternal origin. Consistent with the published nomenclature, we denote this likely causal variant as SW8. Genotyping three of this stallion's offspring identified SW8 only in the nearly all-white foal that was confirmed deaf by brainstem auditory evoked response testing. This foal was also a compound heterozygote for dominant white variants (W20/W22), but to date, W variants alone have not been connected to deafness. SW8 marks the fourth de novo *MITF* variant in horses reported to cause white patterning. The link between deafness and all *MITF* variants with and without other variants impacting melanocyte development and function needs to be further explored.

KEYWORDS
coat color, deafness, horse, pigmentation, white patterning

INTRODUCTION

Variation in white patterning has long fascinated animal enthusiasts and researchers alike. Moreover, some of these variations have a clinical connection and have similar counterparts in humans or other species. In horses, several different types of white patterning phenotypes have been described including frame overo, tobiano, roan, leopard complex spotting, dominant white, sabino and splashed white (Sponenberg & Bellone, 2017). The molecular bases of many of these have been determined, with the first to be unraveled being the frame overo pattern in which white patterning is prevalent on the ventral aspects of the horse with the pigment 'framing' the



Received: 12 July 2023 | Accepted: 19 August 2023

DOI: 10.1111/age.13352

RESEARCH ARTICLE

ANIMAL GENETICS WILEY

A *de novo* 2.3 kb structural variant in *MITF* explains a novel splashed white phenotype in a Thoroughbred family

R. R. Bellone[1,2] | J. Tanaka[1] | E. Esdaile[1] | R. B. Sutton[3] | F. Payette[4] | L. Leduc[4] | B. J. Till[1] | A. K. Abdel-Ghaffar[1] | M. Hammond[1,2] | K. G. Magdesian[5]

[1]Veterinary Genetics Laboratory, School of Veterinary Medicine, UC Davis, Davis, California, USA

[2]Department of Population Health and Reproduction, School of Veterinary Medicine, UC Davis, Davis, California, USA

[3]Cell Physiology and Molecular Biophysics, School of Medicine, Texas Tech University Health Sciences Center, Lubbock, Texas, USA

[4]Department of Clinical Studies, New Bolton Center, University of Pennsylvania School of Veterinary Medicine, University School of Veterinary Medicine, University of Pennsylvania, Kennett Square, Pennsylvania, USA

[5]Department of Medicine and Epidemiology, School of Veterinary Medicine, University of California-Davis, Davis, California, USA

Correspondence
R. R. Bellone, Veterinary Genetics Laboratory, School of Veterinary Medicine, UC Davis, Davis, CA, USA.
Email: rbellone@ucdavis.edu

Funding information
UC Davis Veterinary Genetics Laboratory

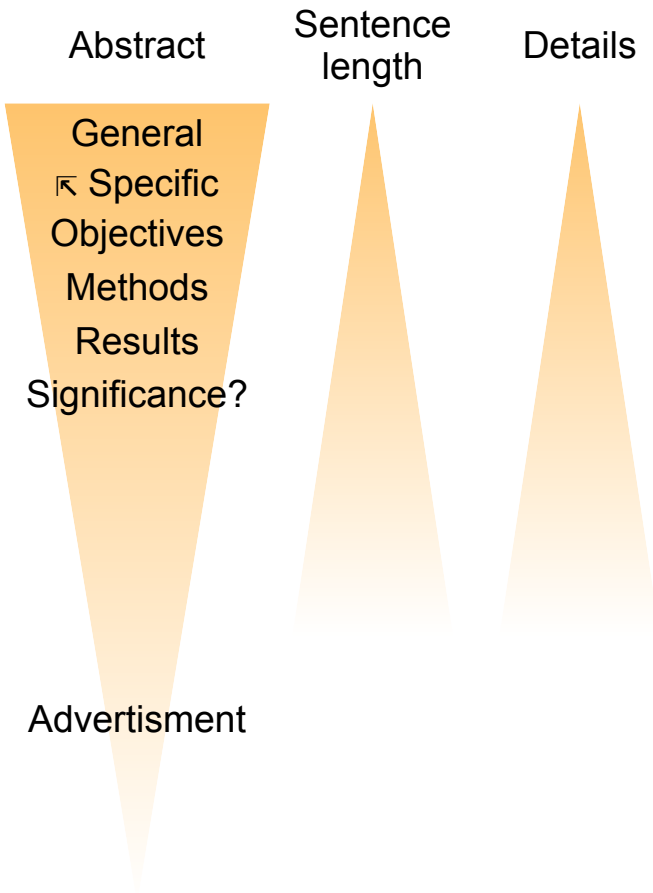
Abstract

variants in *microphthalmia-associated transcription factor* (*MITF*) and two variants in *Paired Box 3* (*PAX3*) have been identified to explain this phenotype. A splashed white Thoroughbred stallion, whose sire and dam were not patterned, was hypothesized to have a *de novo* variant leading to his white coat pattern. A whole-genome sequencing candidate gene approach identified two single nucleotide variants (SNVs) in *SOX10*, four SNVs in *MITF* and a 2.3 kb deletion in *MITF* with the alternative allele present in this stallion but absent in the other 18 horses analyzed. All six SNVs were annotated as modifiers and were not further considered. The deletion in *MITF* (NC_009159.3:g.21555811_21558139delinsAAAT) encompasses exon 9 encoding a part of the helix–loop–helix domain required for DNA binding. Sanger sequencing and parentage testing confirmed that this deletion was a *de novo* mutation of maternal origin. Consistent with the published nomenclature, we denote this likely causal variant as SW8. Genotyping three of this stallion's offspring identified SW8 only in the nearly all-white foal that was confirmed deaf by brainstem auditory evoked response testing. This foal was also a compound heterozygote for dominant white variants (W20/W22), but to date, W variants alone have not been connected to deafness. SW8 marks the fourth de novo *MITF* variant in horses reported to cause white patterning. The link between deafness and all *MITF* variants with and without other variants impacting melanocyte development and function needs to be further explored.

KEYWORDS
coat color, deafness, horse, pigmentation, white patterning

INTRODUCTION

Variation in white patterning has long fascinated animal enthusiasts and researchers alike. Moreover, some of these variations have a clinical connection and have similar counterparts in humans or other species. In horses, several different types of white patterning phenotypes have been described including frame overo, tobiano, roan, leopard complex spotting, dominant white, sabino and splashed white (Sponenberg & Bellone, 2017). The molecular bases of many of these have been determined, with the first to be unraveled being the frame overo pattern in which white patterning is prevalent on the ventral aspects of the horse with the pigment 'framing' the



Received: 12 July 2023 | Accepted: 19 August 2023

DOI: 10.1111/age.13352

RESEARCH ARTICLE

ANIMAL GENETICS WILEY

A *de novo* 2.3 kb structural variant in *MITF* explains a novel splashed white phenotype in a Thoroughbred family

R. R. Bellone[1,2] | J. Tanaka[1] | E. Esdaile[1] | R. B. Sutton[3] | F. Payette[4] | L. Leduc[4] | B. J. Till[1] | A. K. Abdel-Ghaffar[1] | M. Hammond[1,2] | K. G. Magdesian[5]

[1]Veterinary Genetics Laboratory, School of Veterinary Medicine, UC Davis, Davis, California, USA

[2]Department of Population Health and Reproduction, School of Veterinary Medicine, UC Davis, Davis, California, USA

[3]Cell Physiology and Molecular Biophysics, School of Medicine, Texas Tech University Health Sciences Center, Lubbock, Texas, USA

[4]Department of Clinical Studies, New Bolton Center, University of Pennsylvania School of Veterinary Medicine, University School of Veterinary Medicine, University of Pennsylvania, Kennett Square, Pennsylvania, USA

[5]Department of Medicine and Epidemiology, School of Veterinary Medicine, University of California-Davis, Davis, California, USA

Correspondence
R. R. Bellone, Veterinary Genetics Laboratory, School of Veterinary Medicine, UC Davis, Davis, CA, USA.
Email: rbellone@ucdavis.edu

Funding information
UC Davis Veterinary Genetics Laboratory

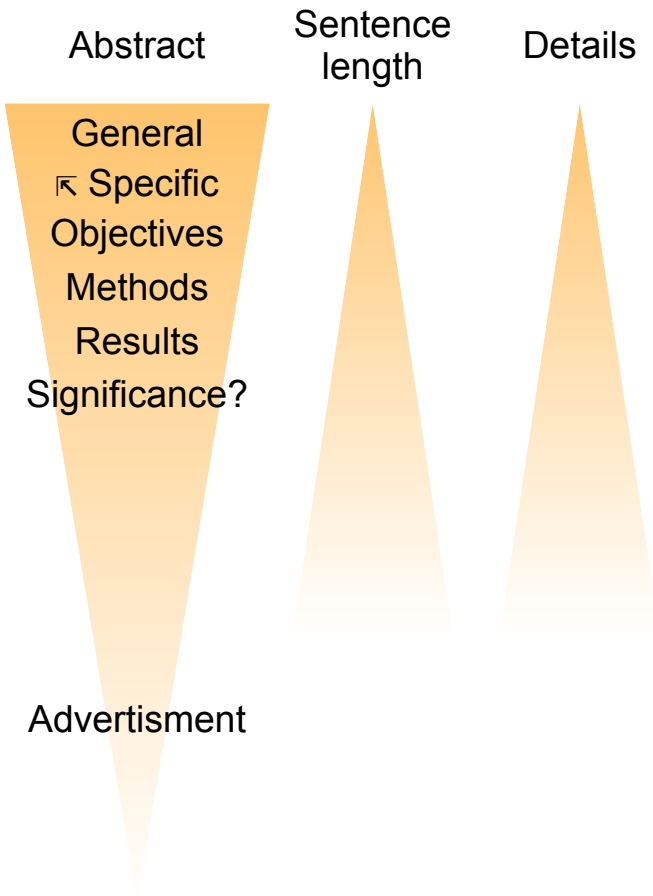
Abstract

variants in *microphthalmia-associated transcription factor* (*MITF*) and two variants in *Paired Box 3* (*PAX3*) have been identified to explain this phenotype. A splashed white Thoroughbred stallion, whose sire and dam were not patterned, was hypothesized to have a *de novo* variant leading to his white coat pattern. A whole-genome sequencing candidate gene approach identified two single nucleotide variants (SNVs) in *SOX10*, four SNVs in *MITF* and a 2.3 kb deletion in *MITF* with the alternative allele present in this stallion but absent in the other 18 horses analyzed. All six SNVs were annotated as modifiers and were not further considered. The deletion in *MITF* (NC_009159.3:g.21555811_21558139delinsAAAT) encompasses exon 9 encoding a part of the helix–loop–helix domain required for DNA binding. Sanger sequencing and parentage testing confirmed that this deletion was a *de novo* mutation of maternal origin. Consistent with the published nomenclature, we denote this likely causal variant as SW8. Genotyping three of this stallion's offspring identified SW8 only in the nearly all-white foal that was confirmed deaf by brainstem auditory evoked response testing. This foal was also a compound heterozygote for dominant white variants (W20/W22), but to date, W variants alone have not been connected to deafness. SW8 marks the fourth de novo *MITF* variant in horses reported to cause white patterning. The link between deafness and all *MITF* variants with and without other variants impacting melanocyte development and function needs to be further explored.

KEYWORDS
coat color, deafness, horse, pigmentation, white patterning

INTRODUCTION

Variation in white patterning has long fascinated animal enthusiasts and researchers alike. Moreover, some of these variations have a clinical connection and have similar counterparts in humans or other species. In horses, several different types of white patterning phenotypes have been described including frame overo, tobiano, roan, leopard complex spotting, dominant white, sabino and splashed white (Sponenberg & Bellone, 2017). The molecular bases of many of these have been determined, with the first to be unraveled being the frame overo pattern in which white patterning is prevalent on the ventral aspects of the horse with the pigment 'framing' the



Received: 12 July 2023 | Accepted: 19 August 2023

DOI: 10.1111/age.13352

RESEARCH ARTICLE

ANIMAL GENETICS WILEY

A *de novo* 2.3 kb structural variant in *MITF* explains a novel splashed white phenotype in a Thoroughbred family

R. R. Bellone[1,2] | J. Tanaka[1] | E. Esdaile[1] | R. B. Sutton[3] | F. Payette[4] | L. Leduc[4] | B. J. Till[1] | A. K. Abdel-Ghaffar[1] | M. Hammond[1,2] | K. G. Magdesian[5]

[1]Veterinary Genetics Laboratory, School of Veterinary Medicine, UC Davis, Davis, California, USA

[2]Department of Population Health and Reproduction, School of Veterinary Medicine, UC Davis, Davis, California, USA

[3]Cell Physiology and Molecular Biophysics, School of Medicine, Texas Tech University Health Sciences Center, Lubbock, Texas, USA

[4]Department of Clinical Studies, New Bolton Center, University of Pennsylvania School of Veterinary Medicine, University School of Veterinary Medicine, University of Pennsylvania, Kennett Square, Pennsylvania, USA

[5]Department of Medicine and Epidemiology, School of Veterinary Medicine, University of California-Davis, Davis, California, USA

Correspondence
R. R. Bellone, Veterinary Genetics Laboratory, School of Veterinary Medicine, UC Davis, Davis, CA, USA.
Email: rbellone@ucdavis.edu

Funding information
UC Davis Veterinary Genetics Laboratory

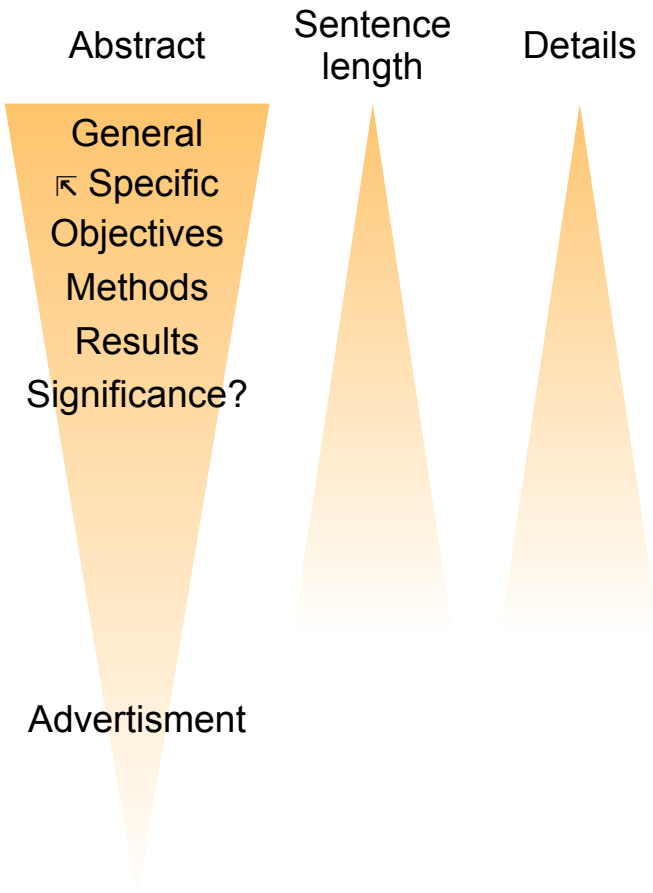
Abstract

variants in *microphthalmia-associated transcription factor* (*MITF*) and two variants in *Paired Box 3* (*PAX3*) have been identified to explain this phenotype. A splashed white Thoroughbred stallion, whose sire and dam were not patterned, was hypothesized to have a *de novo* variant leading to his white coat pattern. A whole-genome sequencing candidate gene approach identified two single nucleotide variants (SNVs) in *SOX10*, four SNVs in *MITF* and a 2.3 kb deletion in *MITF* with the alternative allele present in this stallion but absent in the other 18 horses analyzed. All six SNVs were annotated as modifiers and were not further considered. The deletion in *MITF* (NC_009159.3:g.21555811_21558139delinsAAAT) encompasses exon 9 encoding a part of the helix–loop–helix domain required for DNA binding. Sanger sequencing and parentage testing confirmed that this deletion was a *de novo* mutation of maternal origin. Consistent with the published nomenclature, we denote this likely causal variant as SW8. Genotyping three of this stallion's offspring identified SW8 only in the nearly all-white foal that was confirmed deaf by brainstem auditory evoked response testing. This foal was also a compound heterozygote for dominant white variants (W20/W22), but to date, W variants alone have not been connected to deafness. SW8 marks the fourth de novo *MITF* variant in horses reported to cause white patterning. The link between deafness and all *MITF* variants with and without other variants impacting melanocyte development and function needs to be further explored.

KEYWORDS
coat color, deafness, horse, pigmentation, white patterning

INTRODUCTION

Variation in white patterning has long fascinated animal enthusiasts and researchers alike. Moreover, some of these variations have a clinical connection and have similar counterparts in humans or other species. In horses, several different types of white patterning phenotypes have been described including frame overo, tobiano, roan, leopard complex spotting, dominant white, sabino and splashed white (Sponenberg & Bellone, 2017). The molecular bases of many of these have been determined, with the first to be unraveled being the frame overo pattern in which white patterning is prevalent on the ventral aspects of the horse with the pigment 'framing' the



Received: 12 July 2023 | Accepted: 19 August 2023

DOI: 10.1111/age.13352

RESEARCH ARTICLE

Title

ANIMAL GENETICS WILEY

A *de novo* 2.3 kb structural variant in *MITF* explains a novel splashed white phenotype in a Thoroughbred family

R. R. Bellone[1,2] | J. Tanaka[1] | E. Esdaile[1] | R. B. Sutton[3] | F. Payette[4] | L. Leduc[4] | B. J. Till[1] | A. K. Abdel-Ghaffar[1] | M. Hammond[1,2] | K. G. Magdesian[5]

[1]Veterinary Genetics Laboratory, School of Veterinary Medicine, UC Davis, Davis, California, USA

[2]Department of Population Health and Reproduction, School of Veterinary Medicine, UC Davis, Davis, California, USA

[3]Cell Physiology and Molecular Biophysics, School of Medicine, Texas Tech University Health Sciences Center, Lubbock, Texas, USA

[4]Department of Clinical Studies, New Bolton Center, University of Pennsylvania School of Veterinary Medicine, University School of Veterinary Medicine, University of Pennsylvania, Kennett Square, Pennsylvania, USA

[5]Department of Medicine and Epidemiology, School of Veterinary Medicine, University of California-Davis, Davis, California, USA

Correspondence
R. R. Bellone, Veterinary Genetics Laboratory, School of Veterinary Medicine, UC Davis, Davis, CA, USA.
Email: rbellone@ucdavis.edu

Funding information
UC Davis Veterinary Genetics Laboratory

Abstract

Splashed white in horses is characterized by extensive white patterning on the legs, face and abdomen and may be accompanied by deafness. To date, seven variants in *microphthalmia-associated transcription factor* (*MITF*) and two variants in *Paired Box 3* (*PAX3*) have been identified to explain this phenotype. A splashed white Thoroughbred stallion, whose sire and dam were not patterned, was hypothesized to have a *de novo* variant leading to his white coat pattern. A whole-genome sequencing candidate gene approach identified two single nucleotide variants (SNVs) in *SOX10*, four SNVs in *MITF* and a 2.3 kb deletion in *MITF* with the alternative allele present in this stallion but absent in the other 18 horses annotated as modifiers and were not further cons (NC_009159.3:g.21555811_21558139delinsAAAT g a part of the helix–loop–helix domain require sequencing and parentage testing confirmed that this deletion was a *de novo* mutation of maternal origin. Consistent with the published nomenclature, we denote this likely causal variant as SW8. Genotyping three of this stallion's offspring identified SW8 only in the nearly all-white foal that was confirmed deaf by brainstem auditory evoked response testing. This foal was also a compound heterozygote for dominant white variants (W20/W22), but to date, W variants alone have not been connected to deafness. SW8 marks the fourth de novo *MITF* variant in horses reported to cause white patterning. The link between deafness and all *MITF* variants with and without other variants impacting melanocyte development and function needs to be further explored.

Abstract

KEYWORDS
coat color, deafness, horse, pigmentation, white patterning

Keywords

INTRODUCTION

Variation in white patterning has long fascinated animal enthusiasts and researchers alike. Moreover, some of these variations have a clinical connection and have similar counterparts in humans or other species. In horses, several different types of white patterning phenotypes have been described including frame overo, tobiano, roan, leopard complex spotting, dominant white, sabino and splashed white (Sponenberg & Bellone, 2017). The molecular bases of many of these have been determined, with the first to be unraveled being the frame overo pattern in which white patterning is prevalent on the ventral aspects of the horse with the pigment 'framing' the

The title, abstract, and keywords are often the indexed items and are what you search for.

Received: 12 July 2023 | Accepted: 19 August 2023

DOI: 10.1111/age.13352

ANIMAL GENETICS WILEY

RESEARCH ARTICLE

A *de novo* 2.3 kb structural variant in *MITF* explains a novel splashed white phenotype in a Thoroughbred family

R. R. Bellone[1,2] | J. Tanaka[1] | E. Esdaile[1] | R. B. Sutton[3] | F. Payette[4] | L. Leduc[4] | B. J. Till[1] | A. K. Abdel-Ghaffar[1] | M. Hammond[1,2] | K. G. Magdesian[5]

[1]Veterinary Genetics Laboratory, School of Veterinary Medicine, UC Davis, Davis, California, USA

[2]Department of Population Health and Reproduction, School of Veterinary Medicine, UC Davis, Davis, California, USA

[3]Cell Physiology and Molecular Biophysics, School of Medicine, Texas Tech University Health Sciences Center, Lubbock, Texas, USA

[4]Department of Clinical Studies, New Bolton Center, University of Pennsylvania School of Veterinary Medicine, University School of Veterinary Medicine, University of Pennsylvania, Kennett Square, Pennsylvania, USA

[5]Department of Medicine and Epidemiology, School of Veterinary Medicine, University of California-Davis, Davis, California, USA

Correspondence
R. R. Bellone, Veterinary Genetics Laboratory, School of Veterinary Medicine, UC Davis, Davis, CA, USA.
Email: rbellone@ucdavis.edu

Funding information
UC Davis Veterinary Genetics Laboratory

Abstract

Splashed white in horses is characterized by extensive white patterning on the legs, face and abdomen and may be accompanied by deafness. To date, seven variants in *microphthalmia-associated transcription factor* (*MITF*) and two variants in *Paired Box 3* (*PAX3*) have been identified to explain this phenotype. A splashed white Thoroughbred stallion, whose sire and dam were not patterned, was hypothesized to have a *de novo* variant leading to his white coat pattern. A whole-genome sequencing candidate gene approach identified two single nucleotide variants (SNVs) in *SOX10*, four SNVs in *MITF* and a 2.3 kb deletion in *MITF* with the alternative allele present in this stallion but absent in the other 18 horses analyzed. All six SNVs were annotated as modifiers and were not further considered. The deletion in *MITF* (NC_009159.3:g.21555811_21558139delinsAAAT) encompasses exon 9 encoding a part of the helix–loop–helix domain required for DNA binding. Sanger sequencing and parentage testing confirmed that this deletion was a *de novo* mutation of maternal origin. Consistent with the published nomenclature, we denote this likely causal variant as SW8. Genotyping three of this stallion's offspring identified SW8 only in the nearly all-white foal that was confirmed deaf by brainstem auditory evoked response testing. This foal was also a compound heterozygote for dominant white variants (W20/W22), but to date, W variants alone have not been connected to deafness. SW8 marks the fourth de novo *MITF* variant in horses reported to cause white patterning. The link between deafness and all *MITF* variants with and without other variants impacting melanocyte development and function needs to be further explored.

KEYWORDS
coat color, deafness, horse, pigmentation, white patterning

INTRODUCTION

Section

Variation in white patterning has long fascinated animal enthusiasts and researchers alike. Moreover, some of these variations have a clinical connection and have similar counterparts in humans or other species. In horses, several different types of white patterning phenotypes have been described including frame overo, tobiano, roan, leopard complex spotting, dominant white, sabino and splashed white (Sponenberg & Bellone, 2017). The molecular bases of many of these have been determined, with the first to be unraveled being the frame overo pattern in which white patterning is prevalent on the ventral aspects of the horse with the pigment 'framing' the

Notice that the sections of a research article are often:

Introduction
Materials and methods
Results
Discussion

Received: 12 July 2023 | Accepted: 19 August 2023

DOI: 10.1111/age.13352

ANIMAL GENETICS WILEY

RESEARCH ARTICLE

A *de novo* 2.3 kb structural variant in *MITF* explains a novel splashed white phenotype in a Thoroughbred family

R. R. Bellone[1,2] | J. Tanaka[1] | E. Esdaile[1] | R. B. Sutton[3] | F. Payette[4] | L. Leduc[4] | B. J. Till[1] | A. K. Abdel-Ghaffar[1] | M. Hammond[1,2] | K. G. Magdesian[5]

[1]Veterinary Genetics Laboratory, School of Veterinary Medicine, UC Davis, Davis, California, USA

[2]Department of Population Health and Reproduction, School of Veterinary Medicine, UC Davis, Davis, California, USA

[3]Cell Physiology and Molecular Biophysics, School of Medicine, Texas Tech University Health Sciences Center, Lubbock, Texas, USA

[4]Department of Clinical Studies, New Bolton Center, University of Pennsylvania School of Veterinary Medicine, University School of Veterinary Medicine, University of Pennsylvania, Kennett Square, Pennsylvania, USA

[5]Department of Medicine and Epidemiology, School of Veterinary Medicine, University of California-Davis, Davis, California, USA

Correspondence
R. R. Bellone, Veterinary Genetics Laboratory, School of Veterinary Medicine, UC Davis, Davis, CA, USA.
Email: rbellone@ucdavis.edu

Funding information
UC Davis Veterinary Genetics Laboratory

Abstract

Splashed white in horses is characterized by extensive white patterning on the legs, face and abdomen and may be accompanied by deafness. To date, seven variants in *microphthalmia-associated transcription factor* (*MITF*) and two variants in *Paired Box 3* (*PAX3*) have been identified to explain this phenotype. A splashed white Thoroughbred stallion, whose sire and dam were not patterned, was hypothesized to have a *de novo* variant leading to his white coat pattern. A whole-genome sequencing candidate gene approach identified two single nucleotide variants (SNVs) in *SOX10*, four SNVs in *MITF* and a 2.3 kb deletion in *MITF* with the alternative allele present in this stallion but absent in the other 18 horses analyzed. All six SNVs were annotated as modifiers and were not further considered. The deletion in *MITF* (NC_009159.3:g.21555811_21558139delinsAAAT) encompasses exon 9 encoding a part of the helix–loop–helix domain required for DNA binding. Sanger sequencing and parentage testing confirmed that this deletion was a *de novo* mutation of maternal origin. Consistent with the published nomenclature, we denote this likely causal variant as SW8. Genotyping three of this stallion's offspring identified SW8 only in the nearly all-white foal that was confirmed deaf by brainstem auditory evoked response testing. This foal was also a compound heterozygote for dominant white variants (W20/W22), but to date, W variants alone have not been connected to deafness. SW8 marks the fourth de novo *MITF* variant in horses reported to cause white patterning. The link between deafness and all *MITF* variants with and without other variants impacting melanocyte development and function needs to be further explored.

KEYWORDS
coat color, deafness, horse, pigmentation, white patterning

INTRODUCTION

Section

Variation in white patterning has long fascinated animal enthusiasts and researchers alike. Moreover, some of these variations have a clinical connection and have similar counterparts in humans or other species. In horses, several different types of white patterning phenotypes have been described including frame overo, tobiano, roan, leopard complex spotting, dominant white, sabino and splashed white (Sponenberg & Bellone, 2017). The molecular bases of many of these have been determined, with the first to be unraveled being the frame overo pattern in which white patterning is prevalent on the ventral aspects of the horse with the pigment 'framing' the

Remember

We do not use the information presented in the introduction as the outcome and findings of the paper.

Received: 12 July 2023 | Accepted: 19 August 2023

DOI: 10.1111/age.13352

RESEARCH ARTICLE

ANIMAL GENETICS WILEY

A *de novo* 2.3 kb structural variant in *MITF* explains a novel splashed white phenotype in a Thoroughbred family

R. R. Bellone[1,2] | J. Tanaka[1] | E. Esdaile[1] | R. B. Sutton[3] | F. Payette[4] | L. Leduc[4] | B. J. Till[1] | A. K. Abdel-Ghaffar[1] | M. Hammond[1,2] | K. G. Magdesian[5]

[1]Veterinary Genetics Laboratory, School of Veterinary Medicine, UC Davis, Davis, California, USA

[2]Department of Population Health and Reproduction, School of Veterinary Medicine, UC Davis, Davis, California, USA

[3]Cell Physiology and Molecular Biophysics, School of Medicine, Texas Tech University Health Sciences Center, Lubbock, Texas, USA

[4]Department of Clinical Studies, New Bolton Center, University of Pennsylvania School of Veterinary Medicine, University School of Veterinary Medicine, University of Pennsylvania, Kennett Square, Pennsylvania, USA

[5]Department of Medicine and Epidemiology, School of Veterinary Medicine, University of California-Davis, Davis, California, USA

Correspondence
R. R. Bellone, Veterinary Genetics Laboratory, School of Veterinary Medicine, UC Davis, Davis, CA, USA.
Email: rbellone@ucdavis.edu

Funding information
UC Davis Veterinary Genetics Laboratory

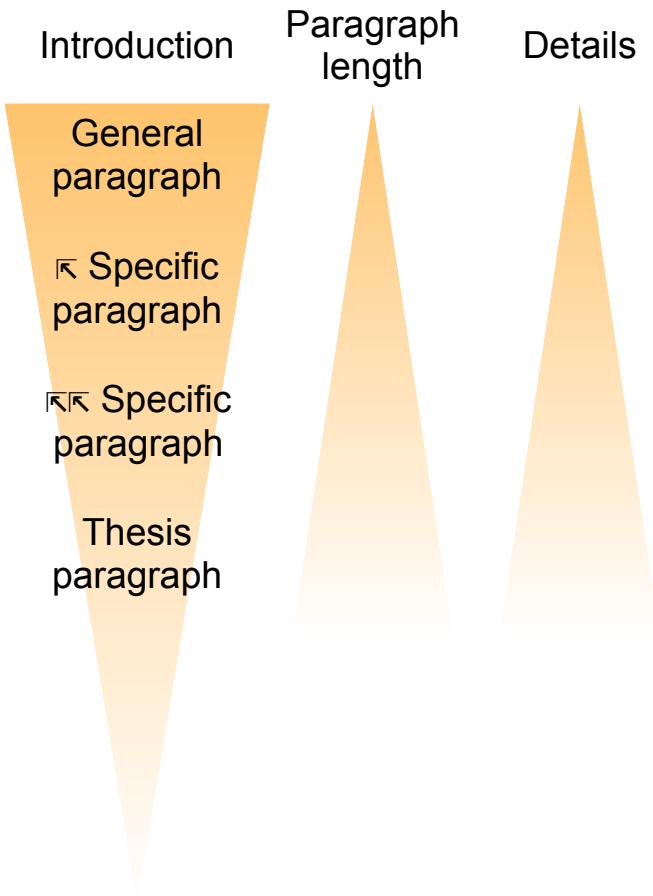
Abstract

Splashed white in horses is characterized by extensive white patterning on the legs, face and abdomen and may be accompanied by deafness. To date, seven variants in *microphthalmia-associated transcription factor* (*MITF*) and two variants in *Paired Box 3* (*PAX3*) have been identified to explain this phenotype. A splashed white Thoroughbred stallion, whose sire and dam were not patterned, was hypothesized to have a *de novo* variant leading to his white coat pattern. A whole-genome sequencing candidate gene approach identified two single nucleotide variants (SNVs) in *SOX10*, four SNVs in *MITF* and a 2.3 kb deletion in *MITF* with the alternative allele present in this stallion but absent in the other 18 horses analyzed. All six SNVs were annotated as modifiers and were not further considered. The deletion in *MITF* (NC_009159.3:g.21555811_21558139delinsAAAT) encompasses exon 9 encoding a part of the helix–loop–helix domain required for DNA binding. Sanger sequencing and parentage testing confirmed that this deletion was a *de novo* mutation of maternal origin. Consistent with the published nomenclature, we denote this likely causal variant as SW8. Genotyping three of this stallion's offspring identified SW8 only in the nearly all-white foal that was confirmed deaf by brainstem auditory evoked response testing. This foal was also a compound heterozygote for dominant white variants (W20/W22), but to date, W variants alone have not been connected to deafness. SW8 marks the fourth de novo *MITF* variant in horses reported to cause white patterning. The link between deafness and all *MITF* variants with and without other variants impacting melanocyte development and function needs to be further explored.

KEYWORDS
coat color, deafness, horse, pigmentation, white patterning

INTRODUCTION

Variation in white patterning has long fascinated animal enthusiasts and researchers alike. Moreover, some of these variations have a clinical connection and have similar counterparts in humans or other species. In horses, several different types of white patterning phenotypes have been described including frame overo, tobiano, roan, leopard complex spotting, dominant white, sabino and splashed white (Sponenberg & Bellone, 2017). The molecular bases of many of these have been determined, with the first to be unraveled being the frame overo pattern in which white patterning is prevalent on the ventral aspects of the horse with the pigment 'framing' the



horse. This pattern is caused by a 2 bp semi-dominant substitution in the *endothelin receptor B* gene (*EDNRB*; Metallinos et al., 1998). Horses homozygous for this variant have an all- or nearly all-white coat and die shortly after birth from aganglionosis, also known as overo lethal white foal syndrome, a condition similar to human Hirschsprung Disease (Metallinos et al., 1998; Santschi et al., 2001). According to one report, 91% of the horses that were deaf or suspected to be deaf had the *EDNRB* mutant allele (Magdesian et al., 2009). Thirty-four variants in or near *KIT* have been shown to cause a white coat pattern, referred to as dominant white, as many of these are also thought to be lethal in the homozygous condition (Capomaccio et al., 2017; Dürig, Jude, Holl, et al., 2017; Esdaile et al., 2022; Haase et al., 2007, 2009, 2015; Holl et al., 2017; Hug et al., 2019; Martin et al., 2021; Rosa, Martin, Vierra, Lundquist, et al., 2022). Additionally, an inversion thought to impact *KIT* gene expression causes the tobiano pattern, but there are no reported adverse health effects (Brooks et al., 2007). An intronic variant that alters splicing in this same gene causes a sabino pattern (SB1), white on the abdomen and high white on the legs, with homozygotes being almost all white (Avila et al., 2022; Brooks & Bailey, 2005). Variants in *TRPM1* and *RFWD3* are involved in leopard complex spotting, a group of white spotting patterns that tend to be symmetrical and centered over the hips (Bellone et al., 2013; Holl et al., 2016). Horses homozygous for the 1378 bp insertion in *TRPM1* have congenital stationary night blindness and are at risk for insidious uveitis (Bellone et al., 2013; Kingsley et al., 2022; Rockwell et al., 2020; Sandmeyer et al., 2020).

The splashed white pattern was named to describe horses that appear as if they have splashed around in white paint, with high levels of white patterning extending up the legs, extensive white on the face and white patches on the abdomen. To date, nine variants in two genes *PAX3* and *MITF* are thought to cause this phenotype (Dürig, Jude, Holl, et al., 2017; Hauswirth et al., 2012, 2013; Henkel et al., 2019; Magdesian et al., 2020; Rosa, Martin, Vierra, Foster, et al., 2022). The first splashed white variant identified was an 11 bp indel in the promoter of *MITF*, which is predicted to impair the binding of its transcription factor PAX3 (Hauswirth et al., 2012). This variant has the widest breed distribution of all splashed-white variants to date (Avila et al., 2022). Since this promoter variant was among the first to be identified for the splashed white phenotype, it was described as SW1. Subsequently, the names of the splashed white variants were primarily described in the order of their identification (SW2–SW7), except for two *MITF* variants denoted as macchiato and MITF^{244Glu} which result in a dilute coat and white spotting pattern and all white phenotype, respectively (Dürig, Jude, Jagannathan, & Leeb, 2017; Hauswirth et al., 2012, 2013; Henkel et al., 2019; Magdesian et al., 2020; Rosa, Martin, Vierra, Foster, et al., 2022). SW2 and SW4 are caused by variants in *PAX3*, whereas SW3, SW5, SW6 and SW7 are all variants in *MITF*. Some horses with these splashed white phenotypes have been reported by their owners to be deaf, and in one case clinical testing was performed, but a thorough investigation of all variants causing these white patterns and the connection to deafness in all cases has not yet been performed (Blatter et al., 2013).

Here we investigate the molecular cause of a Thoroughbred horse with a splashed white phenotype, that consisted of a bald face (defined as a nearly all white face), complete heterochromia (left blue iris), extensive white on all four legs, and white patches on the abdomen (Figure 1). The horse's dam had no white coat patterning and its sire only had a white face marking and one leg marking, which are common among horses without white patterning. By using whole genome sequencing and a candidate gene approach, we aimed to determine if a *de novo* mutation caused this splashed white pattern.

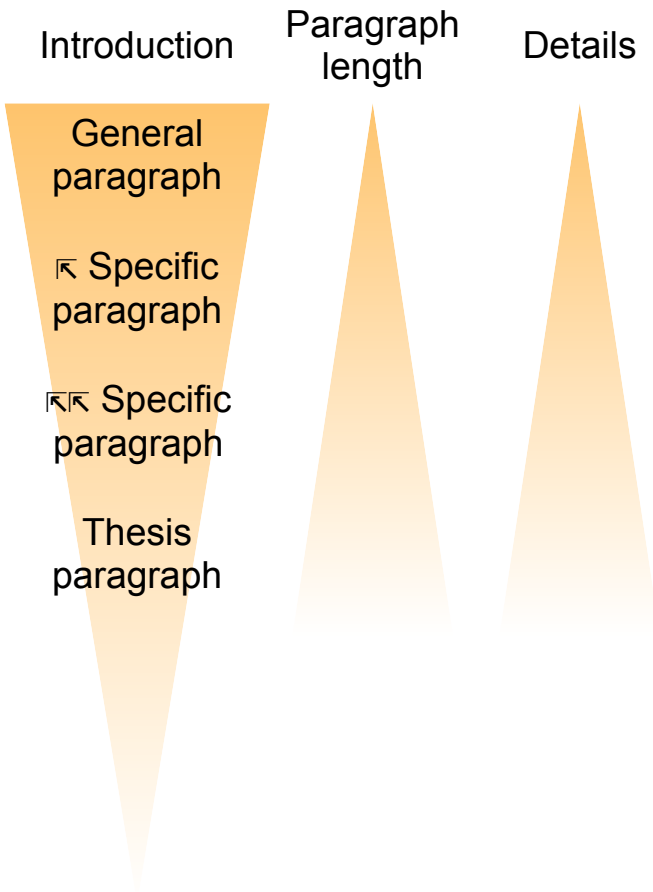
METHODS

Samples and genetic testing

DNA from 10 horses from a single half-sibling family including the stallion under investigation, his sire and dam and three available offspring (along with their dams) were isolated from whole blood or hair follicles using a Gentra Puregene DNA isolation kit as previously described (Mack et al., 2017). Photographic records from each horse were evaluated to assign white patterning phenotypes. All samples were tested for all coat color loci routinely tested commercially at the University of California Davis Veterinary Genetics Laboratory, including the 18 white pattern alleles (SW1–SW6, TO, W4/W5/W10/W13/W20/W22, SB1, LWO, LP, PATN1 and Grey). Twenty-eight additional alleles in *KIT* (W1–W3, W6–W9, W11–W12, W14–W19, W21, W23–W27, W28 and W30–W34) were also genotyped in the stallion of interest as previously described (Esdaile et al., 2022). The VGL also performed routine parentage DNA testing to determine if the sire and dam qualify as parents of the horse under investigation (denoted also as 'HUI' in corresponding figures) and if this horse qualified as the sire of the offspring tested in the study.

Whole genome sequencing

Barcode-indexed sequencing libraries were generated from genomic DNA samples sheared on an E220 Focused Ultrasonicator (Covaris); 10 ng of sheared DNA was converted to sequencing libraries using a Kapa High Throughput Library Preparation Kit (Kapa Biosystems-Roche). The libraries were amplified with four PCR cycles and analyzed with a Bioanalyzer 2100 instrument (Agilent), quantified by fluorometry on a



horse. This pattern is caused by a 2 bp semi-dominant substitution in the *endothelin receptor B* gene (*EDNRB*; Metallinos et al., 1998). Horses homozygous for this variant have an all- or nearly all-white coat and die shortly after birth from aganglionosis, also known as overo lethal white foal syndrome, a condition similar to human Hirschsprung Disease (Metallinos et al., 1998; Santschi et al., 2001). According to one report, 91% of the horses that were deaf or suspected to be deaf had the *EDNRB* mutant allele (Magdesian et al., 2009). Thirty-four variants in or near *KIT* have been shown to cause a white coat pattern, referred to as dominant white, as many of these are also thought to be lethal in the homozygous condition (Capomaccio et al., 2017; Dürig, Jude, Holl, et al., 2017; Esdaile et al., 2022; Haase et al., 2007, 2009, 2015; Holl et al., 2017; Hug et al., 2019; Martin et al., 2021; Ros Martin, Vierra, Lundquist, et al., 2022). Additionally, a inversion thought to impact *KIT* gene expression caus the tobiano pattern, but there are no reported adver health effects (Brooks et al., 2007). An intronic variant that alters splicing in this same gene causes a sabino pattern (SB1), white on the abdomen and high white on the legs, with homozygotes being almost all white (Avila et al., 2022; Brooks & Bailey, 2005). Variants in *TRPM1* and *RFWD3* are involved in leopard complex spotting, a group of white spotting patterns that tend to be symmetrical and centered over the hips (Bellone et al., 2013; Holl et al., 2016). Horses homozygous for the 1378 bp insertion in *TRPM1* have congenital stationary night blindness and are at risk for insidious uveitis (Bellone et al., 2013; Kingsley et al., 2022; Rockwell et al., 2020; Sandmeyer et al., 2020).

The splashed white pattern was named to describe horses that appear as if they have splashed around in white paint, with high levels of white patterning extending up the legs, extensive white on the face and white patches on the abdomen. To date, nine variants in two genes *PAX3* and *MITF* are thought to cause this phenotype (Dürig, Jude, Holl, et al., 2017; Hauswirth et al., 2012, 2013; Henkel et al., 2019; Magdesian et al., 2020; Rosa, Martin, Vierra, Foster, et al., 2022). The first splashed white variant identified was an 11 bp indel in the promoter of *MITF*, which is predicted to impair the binding of its transcription factor PAX3 (Hauswirth et al., 2012). This variant has the widest breed distribution of all splashed-white variants to date (Avila et al., 2022). Since this promoter variant was among the first to be identified for the splashed white phenotype, it was described as SW1. Subsequently, the names of the splashed white variants were primarily described in the order of their identification (SW2–SW7), except for two *MITF* variants denoted as macchiato and MITF^{244Glu} which result in a dilute coat and white spotting pattern and all white phenotype, respectively (Dürig, Jude, Jagannathan, & Leeb, 2017; Hauswirth et al., 2012, 2013; Henkel et al., 2019; Magdesian et al., 2020; Rosa, Martin, Vierra, Foster, et al., 2022). SW2 and SW4 are caused by variants in *PAX3*, whereas SW3, SW5, SW6 and SW7 are all variants in *MITF*. Some horses with these splashed white phenotypes have been reported by their owners to be deaf, and in one case clinical testing was performed, but a thorough investigation of all variants causing these white patterns and the connection to deafness in all cases has not yet been performed (Blatter et al., 2013).

Here we investigate the molecular cause of a Thoroughbred horse with a splashed white phenotype, that consisted of a bald face (defined as a nearly all white face), complete heterochromia (left blue iris), extensive white on all four legs, and white patches on the abdomen (Figure 1). The horse's dam had no white coat patterning and its sire only had a white face marking and one leg marking, which are common among horses without Using whole genome sequencing approach, we aimed to determine used this splashed white pattern.

Section

METHODS

Samples and genetic testing

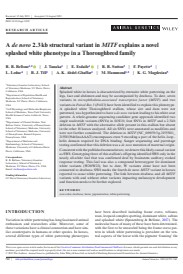
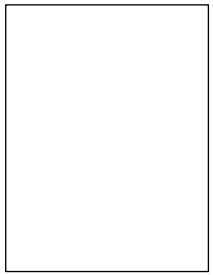
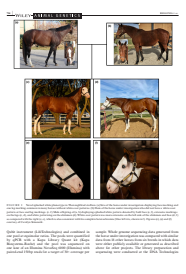
alf-sibling family in- ion, his sire and dam ng with their dams) r hair follicles using a Gentra Puregene DNA isolation kit as previously described (Mack et al., 2017). Photographic records from each horse were evaluated to assign white patterning phenotypes. All samples were tested for all coat color loci routinely tested commercially at the University of California Davis Veterinary Genetics Laboratory, including the 18 white pattern alleles (SW1–SW6, TO, W4/W5/W10/W13/W20/W22, SB1, LWO, LP, PATN1 and Grey). Twenty-eight additional alleles in *KIT* (W1–W3, W6–W9, W11–W12, W14–W19, W21, W23–W27, W28 and W30–W34) were also genotyped in the stallion of interest as previously described (Esdaile et al., 2022). The VGL also performed routine parentage DNA testing to determine if the sire and dam qualify as parents of the horse under investigation (denoted also as 'HUI' in corresponding figures) and if this horse qualified as the sire of the offspring tested in the study.

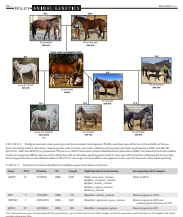
Subsection

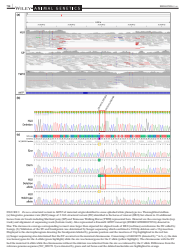
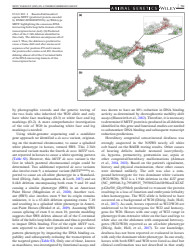
Whole genome sequencing

ries were generated eared on an E220); 10 ng of sheared g libraries using a Kapa High Throughput Library Preparation Kit (Kapa Biosystems-Roche). The libraries were amplified with four PCR cycles and analyzed with a Bioanalyzer 2100 instrument (Agilent), quantified by fluorometry on a

Subsection



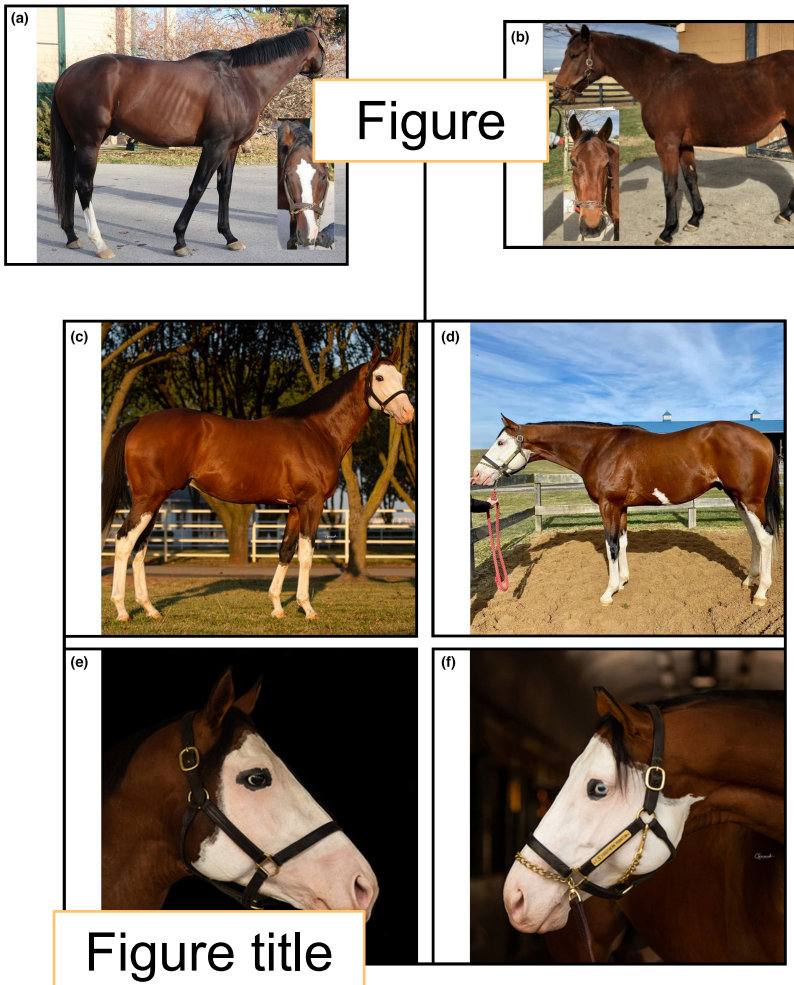
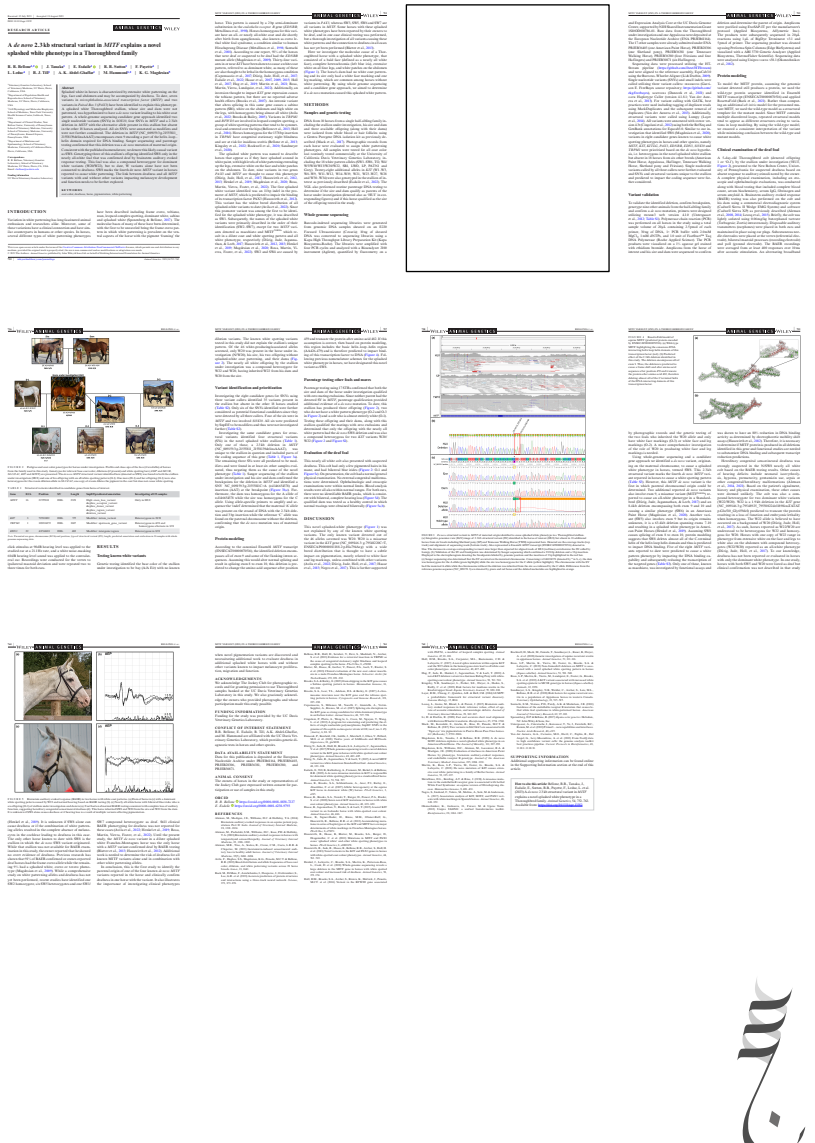



FIGURE 1 Novel splashed white phenotype in Thoroughbred stallion. (a) Sire of the horse under investigation displaying face marking and one leg marking common in many horses without white coat patterns. (b) Dam of the horse under investigation who did not have a white coat pattern or face and leg markings. (c–f) Male offspring of (a, b) displaying splashed white pattern denoted by bald face (e, f), extensive markings on the legs (c, d), and white patterning on t the left side of the abdomen and face (d, f) as compared with the right (c, e), which is a is, shown in f). Figures (c), (e) and (f) courtesy of Carolyn Simancik.

Figure details

Qubit instrument (LifeTechnologies) and combined in one pool at equimolar ratios. The pools were quantified by qPCR with a Kapa Library Quant kit (Kapa Biosystems-Roche) and the pool was sequenced on one lane of an Illumina NovaSeq 6000 (Illumina) with paired-end 150 bp reads for a target of 30× coverage per sample. Whole genome sequencing data generated from the horse under investigation was compared with similar data from 18 other horses from six breeds in which data were either publicly available or generated as described above for other projects. The library preparation and sequencing were conducted at the DNA Technologies



and Expression Analysis Core at the UC Davis Genome Center, supported by NIH Shared Instrumentation Grant 1S10OD010786-01. Raw data from the Thoroughbred under investigation and one Appaloosa were deposited at the European Nucleotide Archive (ENA PRJEB61164). The 17 other samples were already submitted under ENA PRJEB36403 (one American Paint Horse), PRJEB28306 (one Shetland pony), PRJEB36381 (one Tennessee Walking Horse), PRJEB36380 (four Friesians and four Haflingers) and PRJEB30871 (six Haflingers).

Sequencing data were processed utilizing the HTStream pipeline (https://github.com/ibest/HTStream) and were aligned to the reference assembly, EquCab3.0 using the Burrows–Wheeler Aligner (Li & Durbin, 2009). Single nucleotide variants (SNVs) and small indels were called utilizing three variant callers: FREEBAYES (G... son E. FreeBayes source repository; https://github.c... ekg/freebayes), SAMTOOLS (Danecek et al., 2021) ... GATK Haplotype Caller (version 4.1.8.1; Van der Auwera et al., 2013). For variant calling with GATK, best practices were used including tagging of duplicate reads using MarkDuplicates and the subsequent removal of duplicates (Van der Auwera et al., 2013). Additionally, structural variants were called using Lumpy (Layer et al., 2014). All variants were annotated with SNPEFF version 4.3t (Cingolani et al., 2012) using both the RefSeq and GenBank annotations for Equcab3.0. Similar to our investigation that identified SW6 (Magdesian et al., 2020), variants in eight candidate genes known to cause white spotting phenotypes in horses and other species, namely *MITF*, *KIT*, *KITLG*, *PAX3*, *EDNRB*, *EDN3*, *SOX10* and *TRPM1* were prioritized based on the *de novo* hypothesis, i.e. heterozygous in the novel splashed white stallion but absent in 18 horses from six other breeds (American Paint Horse, Appaloosa, Haflinger, Tennessee Walking Horse, Shetland pony and Friesian). Single nucleotide variants called by all three callers were further evaluated and SNVs and structural variants unique to the stallion and predicted to impact the coding sequence were further considered.

Variant validation

Subsection

...firm breakpoints, ...half-sibling family ...ers were designed utilizing PRIMER3 web version 4.1.0 (Untergasser et al., 2012; Table S1). Polymerase chain reaction (PCR) was performed on all horses in the study using a total sample volume of 20 μL containing 2.5 pmol of each primer, 30 ng of DNA, 1× PCR buffer with 2.0 mM $MgCl_2$, 1 mM dNTPs, and 1.0 unit of FastStart™ Taq DNA Polymerase (Roche Applied Science). The PCR products were visualized on a 1% agarose gel stained with ethidium bromide. Amplicons from the horse of interest and his sire and dam were sequenced to confirm deletion and determine the parent of origin. Amplicons were purified using ExoSAP-IT per the manufacturer's protocol (Applied Biosystems, Affymetric Inc.). The products were subsequently sequenced in 20 μL reactions using 1 μL of BigDye Terminator v3.1 and 5 pmol of primer. The sequencing product was cleaned up using Performa Spin Columns (Edge BioSystems) and visualized with a ABI 3730 Genetic Analyzer (Applied Biosystems, ThermoFisher Scientific). Sequencing data were analyzed using Unipro UGENE v38.1 (Okonechnikov et al., 2012).

Protein modeling

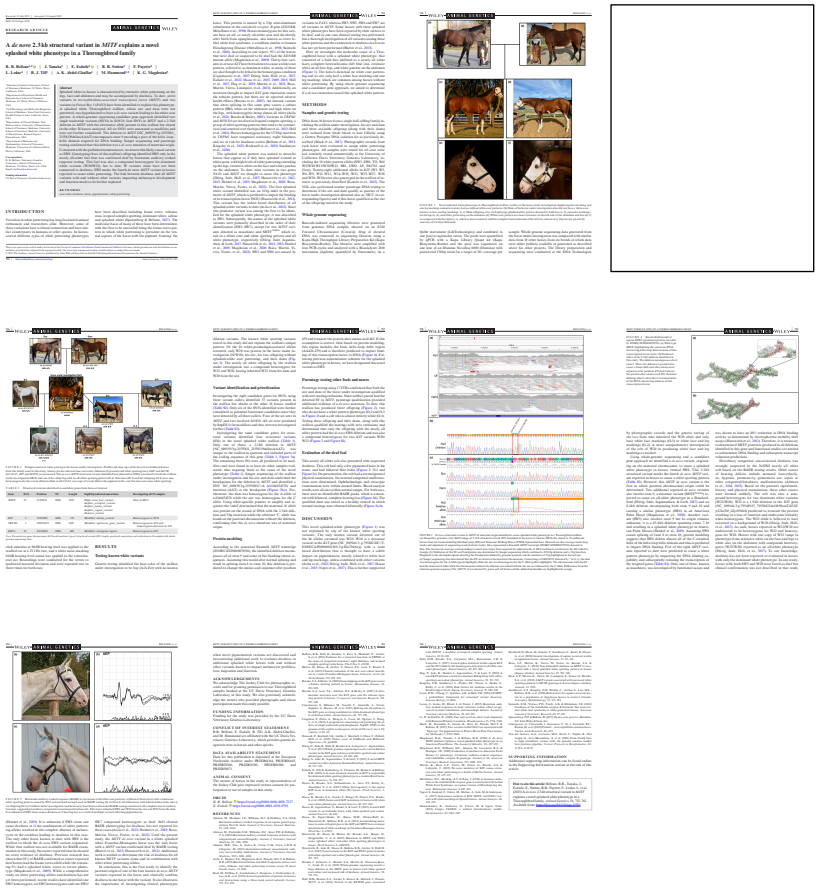
Subsection

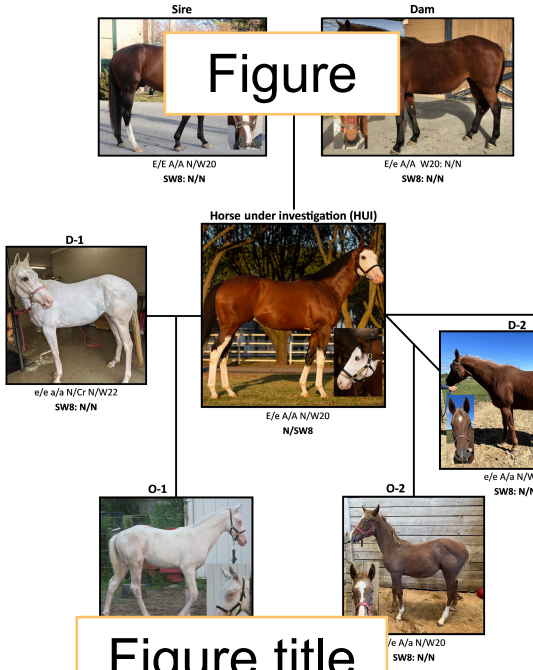
...l, assuming the genomic ...es a protein, we used the ... identified in Ensembl ...MITF transcript (ENSECAT00000070516) and applied RosettaFold (Baek et al., 2021). Rather than computing an additional *ab initio* model for the presumed mutant MITF, we used the wild-type model as a structural template for the mutant model. Since MITF contains multiple disordered loops, repeated structural models tend to appear as different structures owing to variations in loop modeling. By using the wild-type model, we ensured a consistent interpretation of the variant while minimizing confusion between the wild-type and mutant models.

Clinical examination of the deaf foal

Subsection

A ...oted offspring 1 ...tigation ('HUI', F... Center, University of Pennsylvania for suspected deafness based on absent response to auditory stimuli noted by the owner. A complete physical examination, including an otoscopic and ophthalmologic evaluations, was conducted along with blood testing that included complete blood count, serum biochemistry, serum IgG, fibrinogen and serum amyloid A. Brainstem auditory evoked response (BAER) testing was also performed on the colt and his dam using a commercial electrodiagnostic system (Cadwell Sierra II Wedge EMG System) and software (Cadwell Sierra XP) as previously described (Aleman et al., 2008, 2014; Lecoq et al., 2015). Briefly, the colt was lightly sedated using 0.06 mg/kg butorphanol tartrate (Torbugesic; Zoetis) intravenously. Disposable auditory transmitters (earphones) were placed in both ears and maintained in place using ear plugs. Subcutaneous needle electrodes were placed at the vertex (referential electrode), bilateral mastoid processes (recording electrode) and poll (ground electrode). The BAER recordings were averaged from at least 400 responses over 10 ms after acoustic stimulation. An alternating broadband





Figure

Figure title

FIGURE 2 Pedigree and coat color genotypes for horses under investigation. Pro
from the family used in this study. Genotypes for relevant base coat color, dilutions (i
SLC45A2, *KIT* and *MITF*) are presented. The *de novo MITF* structural variant identi
under investigation (HUI) and one of his offspring with an all-white spotting pattern (O-1). One mare (D-1) and her offspring (O-1) were also
heterozygous for the cream dilution allele in *SLC45A2*; one copy of cream dilutes the pigment in the coat but does not cause white spotting.

Figure details

TABLE 1 Structural variants identified in candidate genes from horse of interest.

Table title

Gene	ECA	Position	SV	Length	SnpEff predicated annotat	
MITF	16	21555810	DEL	2329	High: exon_loss_variant &splice_acceptor_variant &splice_donor_variant &splice_region_variant &intron_variant	Only in HUI
KIT	3	79591009	DEL	575	Modifier: intron_variant	Heterozygous in 19/19
TRPM1	1	109216072	DEL	3607	Modifier: upstream_gene_variant	Heterozygous in 4/19 and homozygous alternate in 3/19
EDN3	22	46714600	DEL	402	Modifier: intergenic region	Heterozygous in 8/19

Note: Presented are gene, chromosome (ECA) and position, type of structural variant (SV), length, predicted annotation and evaluation in 19 samples with whole genome sequencing data.

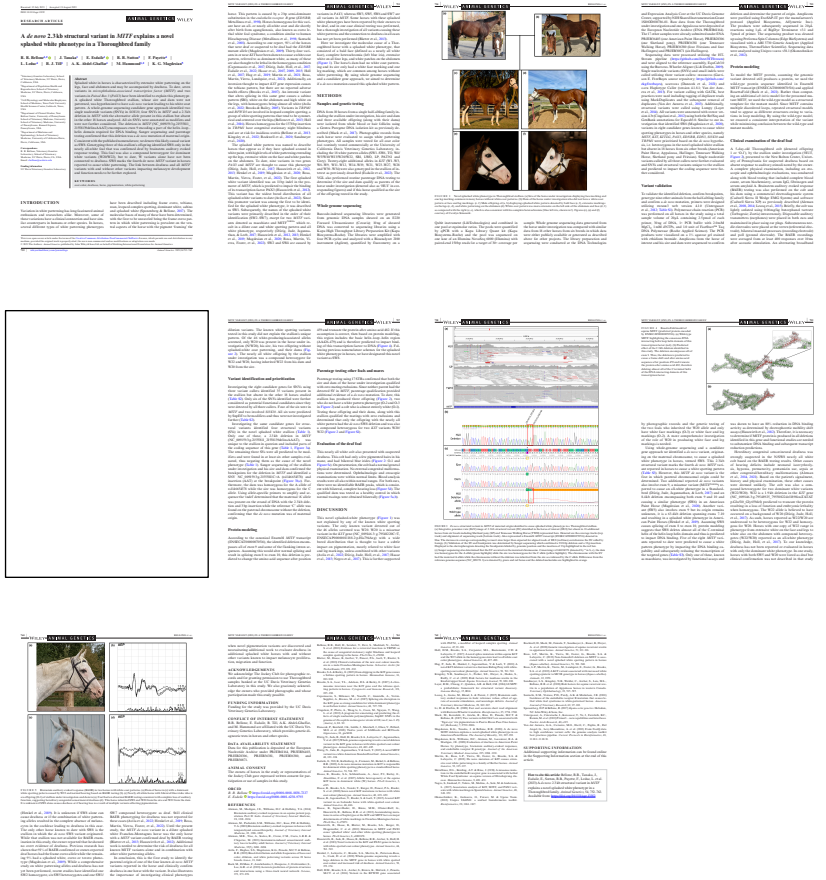
Table details

eral ear. Recordings were conducted for the vertex to ipsilateral mastoid deviation and were repeated two to three times for both ears.

RESULTS

Testing known white variants

Genetic testing identified the base color of the stallion under investigation to be bay (A/A E/e) with no known



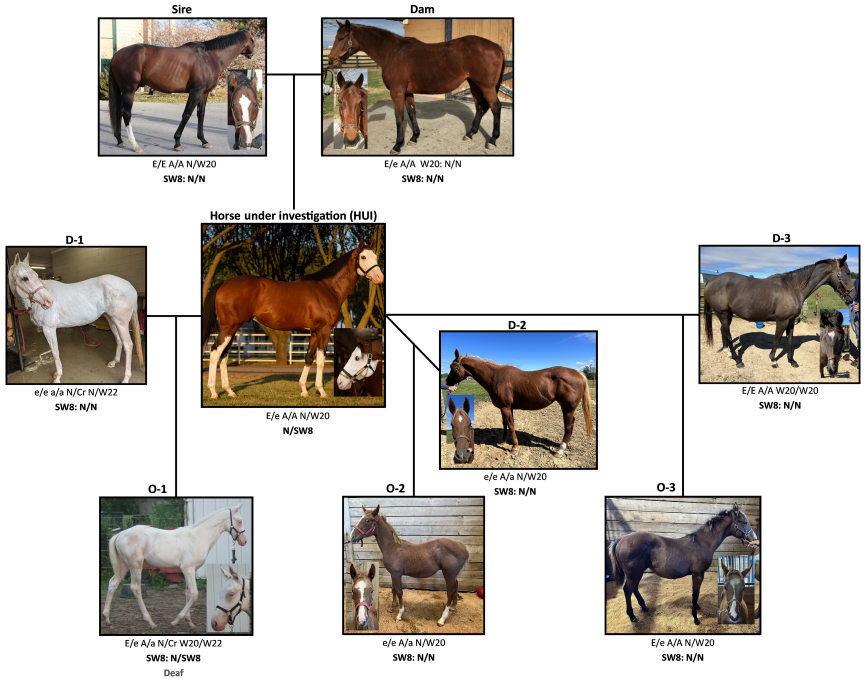


FIGURE 2 Pedigree and coat color genotypes for horses under investigation. Profile and close-ups of the face (if available) of horses from the family used in this study. Genotypes for relevant base coat color, dilutions (if present) and white spotting loci (*ASIP* and *MC1R*, *SLC45A2*, *KIT* and *MITF*) are presented. The *de novo MITF* structural variant identified here (denoted as SW8) was found in both the stallion under investigation (HUI) and one of his offspring with an all-white spotting pattern (O-1). One mare (D-1) and her offspring (O-1) were also heterozygous for the cream dilution allele in *SLC45A2*; one copy of cream dilutes the pigment in the coat but does not cause white spotting.

TABLE 1 Structural variants identified in candidate genes from horse of interest.

Gene	ECA	Position	SV	Length	SnpEff predicated annotation	Investigating all 19 samples
MITF	16	21555810	DEL	2329	High: exon_loss_variant &splice_acceptor_variant &splice_donor_variant &splice_region_variant &intron_variant	Only in HUI
KIT	3	79591009	DEL	575	Modifier: intron_variant	Heterozygous in 19/19
TRPM1	1	109216072	DEL	3607	Modifier: upstream_gene_variant	Heterozygous in 4/19 and homozygous alternate in 3/19
EDN3	22	46714600	DEL	402		Heterozygous in 8/19

Note: Presented are gene, chromosome (ECA) and position, type of structu ...tion and evaluation in 19 samples with whole genome sequencing data.

Section

RESULTS

Testing known white variants

Subsection

click stimulus at 90 dB hearing level was applied to the studied ear at a 21.1 Hz rate, and a white noise-masking 60 dB hearing level sound was applied to the contralateral ear. Recordings were conducted for the vertex to ipsilateral mastoid deviation and were repeated two to three times for both ears.

...color of the stallion E/e) with no known

Remember

We use the information presented in the results as the outcome and findings of the paper.

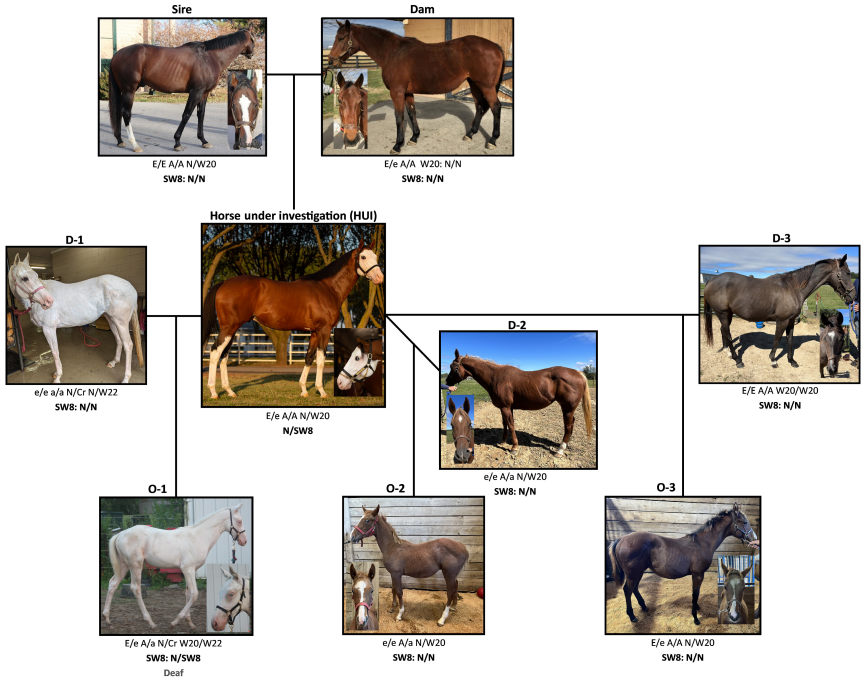


FIGURE 2 Pedigree and coat color genotypes for horses under investigation. Profile and close-ups of the face (if available) of horses from the family used in this study. Genotypes for relevant base coat color, dilutions (if present) and white spotting loci (*ASIP* and *MC1R*, *SLC45A2*, *KIT* and *MITF*) are presented. The *de novo MITF* structural variant identified here (denoted as SW8) was found in both the stallion under investigation (HUI) and one of his offspring with an all-white spotting pattern (O-1). One mare (D-1) and her offspring (O-1) were also heterozygous for the cream dilution allele in *SLC45A2*; one copy of cream dilutes the pigment in the coat but does not cause white spotting.

TABLE 1 Structural variants identified in candidate genes from horse of interest.

Gene	ECA	Position	SV	Length	SnpEff predicated annotation	Investigating all 19 samples
MITF	16	21555810	DEL	2329	High: exon_loss_variant &splice_acceptor_variant &splice_donor_variant &splice_region_variant &intron_variant	Only in HUI
KIT	3	79591009	DEL	575	Modifier: intron_variant	Heterozygous in 19/19
TRPM1	1	109216072	DEL	3607	Modifier: upstream_gene_variant	Heterozygous in 4/19 and homozygous alternate in 3/19
EDN3	22	46714600	DEL	402		Heterozygous in 8/19

Note: Presented are gene, chromosome (ECA) and position, type of structu ...tion and evaluation in 19 samples with whole genome sequencing data.

click stimulus at 90 dB hearing level was applied to the studied ear at a 21.1 Hz rate, and a white noise-masking 60 dB hearing level sound was applied to the contralateral ear. Recordings were conducted for the vertex to ipsilateral mastoid deviation and were repeated two to three times for both ears.

RESULTS

Testing known white variants

...color of the stallion ...E/e) with no known

Section

Subsection

Notice that the subsections of the results often mirror that displayed in the material and methods.

dilution variants. The known white spotting variants tested in this study did not explain the stallion's unique pattern. Of the 46 white-producing/associated alleles screened, only W20 was present in the horse under investigation (N/W20), his sire, his two offspring without splashed-white coat patterning, and their dams (Figure 2). The nearly all white offspring by the stallion under investigation was a compound heterozygote for W22 and W20, having inherited W22 from his dam and W20 from the sire.

Variant identification and prioritization

Subsection

Inv or SNVs using thr ats present in the horses studied (Table S2). Only six of the SNVs identified were further considered as potential functional candidates since they were detected by all three callers. Four of the six were in *MITF* and two involved *SOX10*. All six were predicted by SnpEff to be modifiers and thus were not investigated further (Table S2).

Investigating the same candidate genes for structural variants identified four structural variants (SVs) in the novel splashed white stallion (Table 1). Only one of these, a 2.3 kb deletion in *MITF* (NC_009159.3:g.21555811_21558139delinsAAAT), was unique to the stallion in question and included parts of the coding sequence of this gene (Table 1, Figure 3a). The remaining three SVs were all predicted to be modifiers and were found in at least six other samples evaluated, thus negating them as the cause of the novel phenotype (Table 1). Sanger sequencing of the stallion under investigation and his sire and dam confirmed the breakpoints for the deletion in *MITF* and identified a SNV NC_009159.3:g.21555811C>A (rs1146654378) and insertion (AAT) at the breakpoint (Figure 3b,c). Furthermore, the dam was homozygous for the A allele of rs1146654378 while the sire was homozygous for the C allele. Using allele-specific primers to amplify and sequence the 'indel' determined that the maternal 'A' allele was present on the strand of DNA with the 2.3 kb deletion and 3 bp insertion while the reference 'C' allele was found on the paternal chromosome without the deletion, confirming that the *de novo* mutation was of maternal origin.

Protein modeling

Subsection

MITF transcript d deletion encom- anking intron sequences. Assuming this would alter normal splicing and result in splicing exon 8 to exon 10, this deletion is predicted to change the amino acid sequence after position 459 and truncate the protein after amino acid 482. If this assumption is correct, then based on protein modeling, this region includes the basic helix–loop–helix region (AA426-479) and is therefore predicted to impact binding of this transcription factor to DNA (Figure 4). Following previous nomenclature schemes for the splashed white phenotype in horses, we have designated this novel variant as SW8.

Parentage testing other foals and mares

Par d that both the sire tion qualified wit parent had the detected SV in *MITF*, parentage qualification provided additional evidence of a *de novo* mutation. To date, this stallion has produced three offspring (Figure 2), two who do not have a white pattern phenotype (O-2 and O-3 in Figure 2) and a colt who is almost entirely white (O-1). Testing these offspring and their dams, along with this stallion qualified the matings with zero exclusions and determined that only the offspring with the nearly all white pattern had the *de novo* SW8 deletion and was also a compound heterozygous for two *KIT* variants W20/W22 (Figure 2 and Figure S1).

Evaluation of the deaf foal

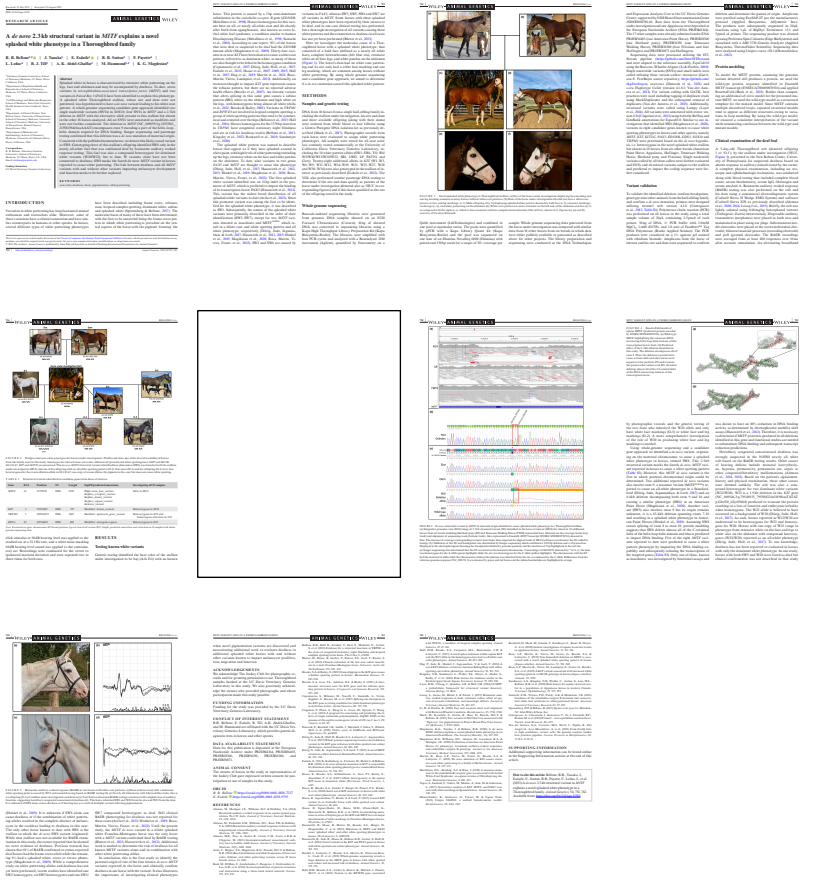
Subsection

sented with suspected pigmented hairs in his s (Figure 2: O-1 and Figure 3c). On presentation, the colt had a normal general physical examination. No external congenital malformations were determined. Ophthalmologic and otoscopic examinations were within normal limits. Blood analysis results were all also within normal ranges. For both ears, there were no identifiable BAER peaks, which is consist- te hearing loss (Figure 5d). The d as a healthy control in which ained bilaterally (Figure 5a,b).

Section

DISCUSSION

This novel splashed-white phenotype (Figure 1) was not explained by any of the known white spotting variants. The only known variant detected out of the 46 alleles screened was W20. W20 is a missense variant in the *KIT* gene (NC_009146.3: g.79548220C>T, ENSECAP00000011188.2:p.His794Arg) with a wide breed distribution that is thought to have a subtle impact on pigmentation, mostly related to white face and leg markings, unless combined with other variants (Avila et al., 2022; Dürig, Jude, Holl, et al., 2017; Haase et al., 2013; Negro et al., 2017). This is further supported



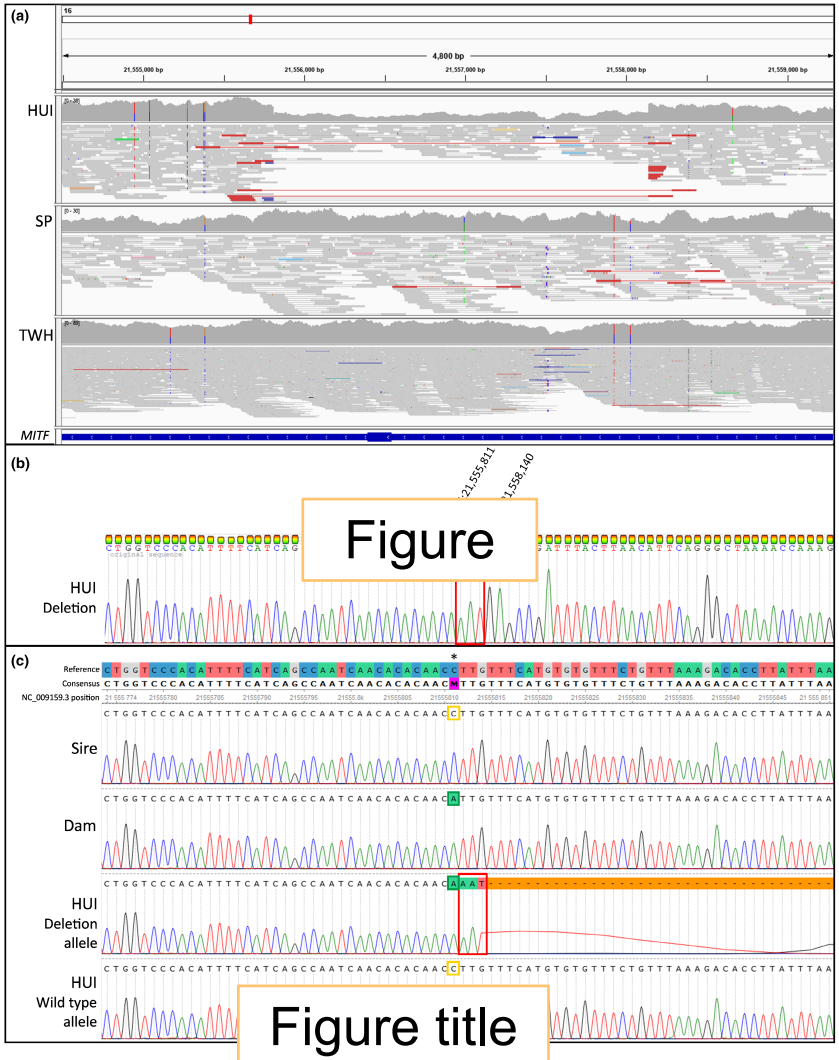
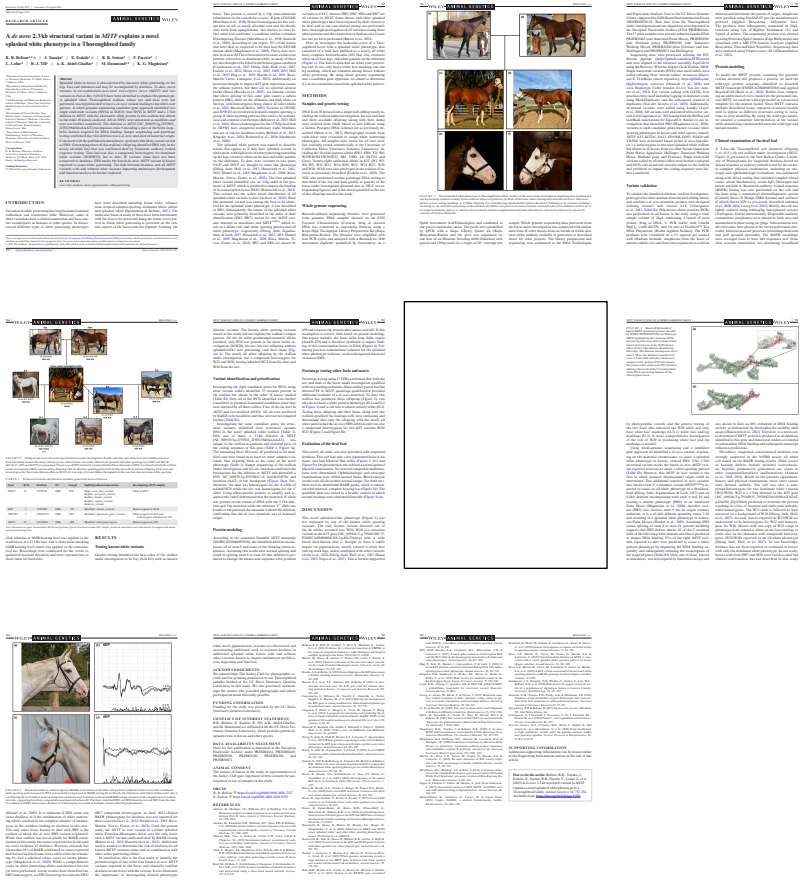


FIGURE 3 *De novo* structural variant in *MITF* of maternal origin identified to cause splashed white phenotype in a Thoroughbred stallion. (a) Integrative genomics view (IGV) image of 2.3 kb structural variant (SV) identified in the horse of interest (HUI) but absent in 18 additional horses from six breeds including Shetland pony (SP) and Tennessee Walking Horse (TWH) represented here. Denoted are the coverage tracks (top track) and alignment of sequencing reads (bottom track). Also represented is Ensembl *MITF* transcript (ENSECAT00000070516) denoted in blue. The decrease in coverage correspondin [illegible] (red lines) corroborates the SV called by Lumpy. (b) Validation of the SV and breakp [illegible] a 2326 bp deletion and a 3 bp insertion. Displayed is the electropherogram denoting [illegible] of 3 bp highlighted in the red box. (c) Sanger sequencing also determined that t [illegible] 146654378 (denoted by * in b, c), the dam was homozygous for the A allele (green high [illegible] highlight). The chromosome with the SV had the maternal A allele while the chromosome without the deletion was inherited from the sire as evidenced by the C allele. Differences from the reference genome sequence (NC_009159.3) are denoted by green and red boxes and the deleted nucleotides are highlighted in orange.



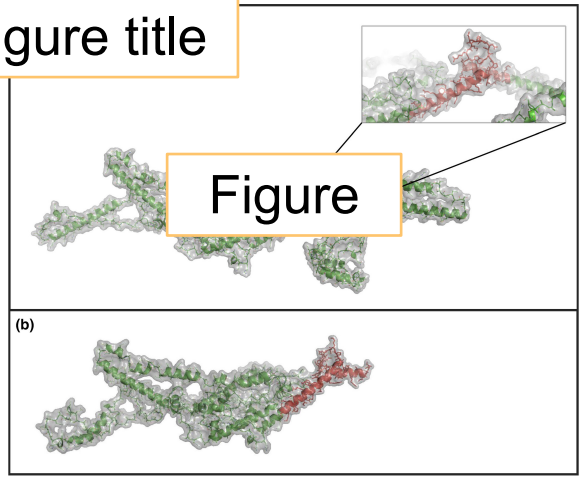


FIGURE 4 RosettaFold model of equine MITF (predicted protein encoded by ENSECAT00000070516). (a) Wild-type MITF highlighting the consensus DNA interacting helix loop helix domain of this transcription factor (red). (b) Predicted effect of the 2.3 kb deletion identified in this study. The deletion encompasses all of exon 9. Thus, the deletion is predicted to cause a frame shift and alter amino acid sequence after position 459 and truncate the protein after amino acid 482, therefore deleting almost all of the C terminal helix of the DNA interacting domain of this transcription factor.

Figure title

Figure

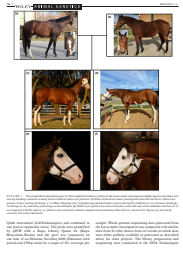
Figure details

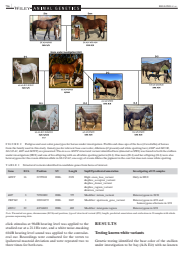
by photographic records and the genetic testing of the two foals who inherited the W20 allele and only have white face markings (O-3) or white face and leg markings (O-2). A more comprehensive investigation of the role of W20 in producing white face and leg markings is needed.

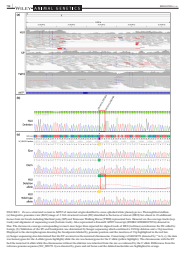
Using whole-genome sequencing and a candidate gene approach we identified a *de novo* variant, originating on the maternal chromosome, to cause a splashed white phenotype in horses, termed SW8. This 2.3 kb structural variant marks the fourth *de novo MITF* variant reported in horses to cause a white spotting pattern (Table S3). However, this *MITF de novo* variant is the first in which parental chromosomal origin could be determined. Two additional reported *de novo* variants also involve exon 9, a missense variant ($MITF^{244Glu}$) reported to cause an all-white phenotype in a Standardbred (Dürig, Jude, Jagannathan, & Leeb, 2017) and an 8.4 kb deletion encompassing both exon 9 and 10 and causing a similar phenotype (SW6) in an American Paint Horse (Magdesian et al., 2020). Another variant (SW5) also involves exon 9 but its origin remains unknown, it is a 63.4 kb deletion spanning exons 7–10 and resulting in a splashed white phenotype in American Paint Horses (Henkel et al., 2019). Assuming SW8 causes splicing of exon 8 to exon 10, protein modeling suggests that SW8 deletes almost all of the C-terminal helix of the helix loop helix domain and thus is predicted to impact DNA binding. Five of the eight *MITF* variants reported to date were predicted to cause a white pattern phenotype by impacting the DNA binding capability and subsequently reducing the transcription of the targeted genes (Table S3). Only one of these, known as macchiato, was investigated by functional assays and was shown to have an 80% reduction in DNA binding activity as determined by electrophoretic mobility shift assays (Hauswirth et al., 2012). Therefore, it is necessary to determine if MITF protein is produced in all deletions identified in this gene and functional studies are needed to substantiate DNA binding and subsequent transcript reduction predictions.

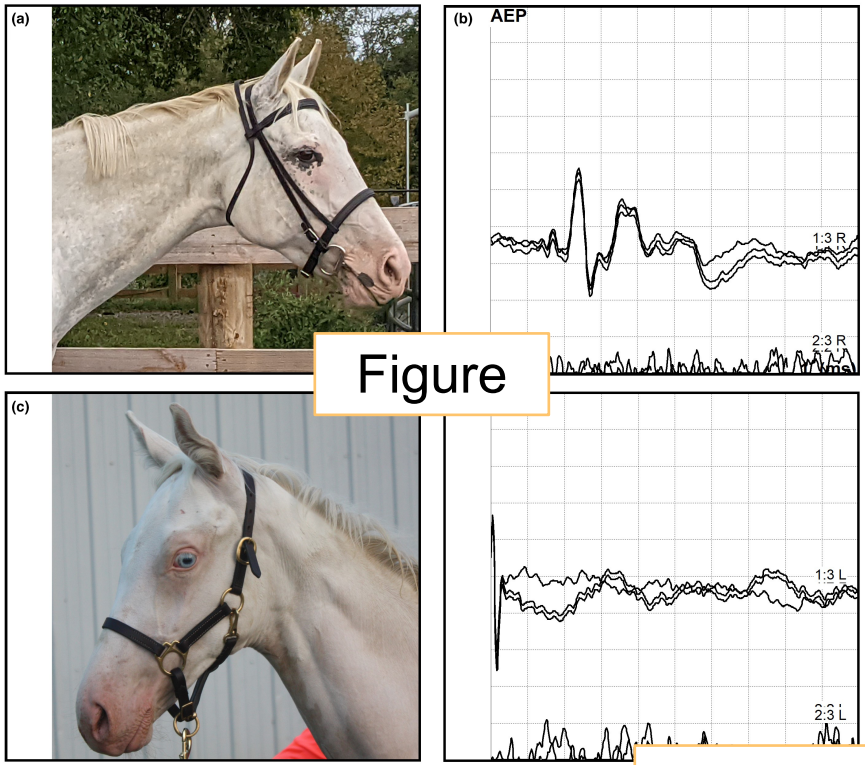
Hereditary congenital sensorineural deafness was strongly suspected in the N/SW8 nearly all white colt based on the BAER testing results. Other causes of hearing deficits include neonatal isoerythrolysis, hypoxia, prematurity, gentamicin use, sepsis or other congenital/hereditary malformations (Aleman et al., 2014, 2021). Based on the patient's signalment, history and physical examination, these other causes were deemed unlikely. The colt was also a compound heterozygote for two dominant white variants (W22/W20). W22 is a 1.9 kb deletion in the *KIT* gene (NC_009146.3:g.79548925_79550822del1898insTATAT p.Glu510_Gly659del) predicted to truncate the protein resulting in a loss of function and embryonic lethality when homozygous. The W22 allele is believed to have occurred on a background of W20 (Dürig, Jude, Holl, et al., 2017). As such, horses reported as W22/W20 are understood to be heterozygous for W22 and homozygous for W20. Horses with one copy of W22 range in phenotype from extensive white on the face and legs to white also on the abdomen with compound heterozygotes (W22/W20) reported as an all-white phenotype (Dürig, Jude, Holl, et al., 2017). To our knowledge, deafness has not been reported or evaluated in horses with only the dominant white phenotype. In one study, horses with both SW5 and W20 were listed as deaf but clinical confirmation was not described in that study









Figure

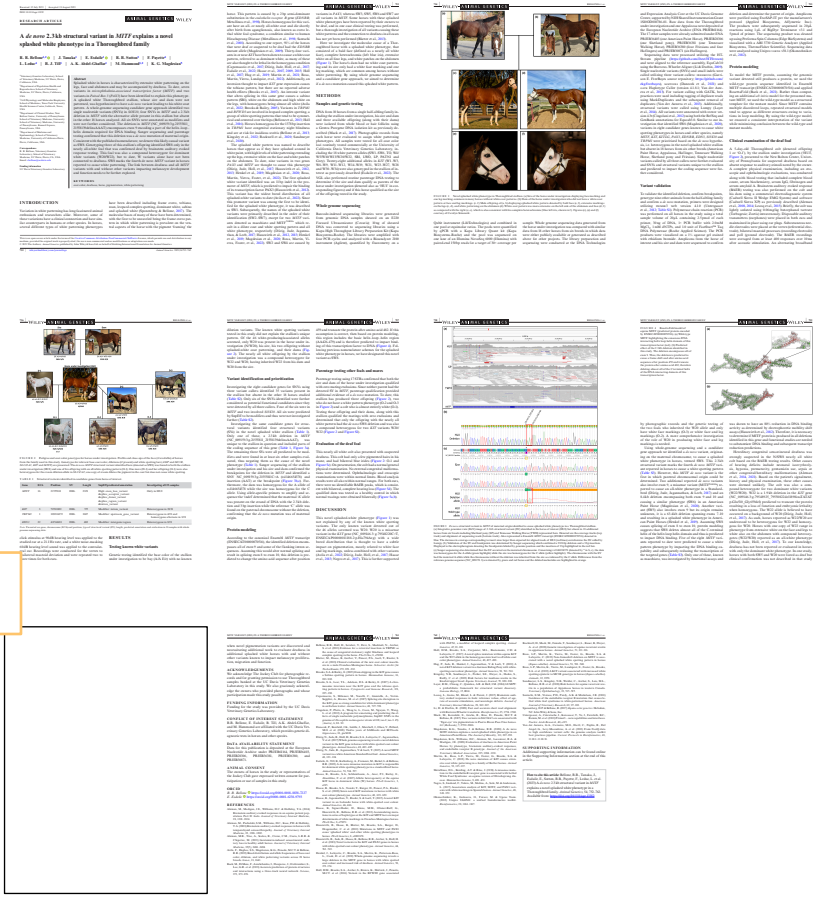
Figure title

FIGURE 5 Brainstem auditory evoked response (BAER) in two horses with white coat patterns. (a) white spotting pattern caused by W22 and normal hearing based on BAER testing (b). (c) Nearly all-wh an offspring (O-1) of stallion under investigation and dam in (a). Foal had an abnormal BAER testing consistent with complete loss of auditory function, suggesting hereditary congenital sensorineural deafness (d). This horse inherited SW8 and W20 from the sire and W22 from the dam. It is unknown if SW8 alone causes deafness or if hearing loss is a result of multiple variants affecting pigmentation.

Figure details

(Henkel et al., 2019). It is unknow eterozygote as deaf. Still clinical cause deafness or if the combinati for deafness was not reported for ing alleles resulted in the complete al., 2022; Henkel et al., 2019; Rosa, cytes in the cochleae leading to deafness in this case. Martin, Vierra, Foster, et al., 2022). Until the present The only other horse known to date with SW8 is the study, the *MITF de novo* variant in a dilute splashed stallion in which the *de novo* SW8 variant originated. white Franches-Montagnes horse was the only horse While that stallion was not available for BAER exam- with a *MITF* variant confirmed deaf by BAER testing ination in this study, the owner reported that he showed (Blatter et al., 2013; Hauswirth et al., 2012). Additional no overt evidence of deafness. Previous research has work is needed to determine the risk of deafness for all shown that 91% of BAER confirmed or owner-reported known MITF variants alone and in combination with deaf horses had the frame overo allele while the remain- other white patterning alleles.

ing 9% had a splashed white, overo or tovero pheno- In conclusion, this is the first study to identify the type (Magdesian et al., 2009). While a comprehensive parental origin of one of the four known *de novo MITF* study on white patterning alleles and deafness has not variants reported in the horse and clinically confirm yet been performed, recent studies have identified one deafness in one horse with the variant. It also illustrates SW2 homozygote, six SW5 heterozygotes and one SW1/ the importance of investigating clinical phenotypes



when novel pigmentation variants are discovered and necessitating additional work to evaluate deafness in additional splashed white horses with and without other variants known to impact melanocyte proliferation, migration and function.

ACKNOWLEDGEMENTS

We acknowledge The Jockey Club for photographic records and for granting permission to use Thoroughbred samples banked at the UC Davis Veterinary Genetics Laboratory in this study. We also graciously acknowledge the owners who provided photographs and whose participation made this study possible.

FUNDING INFORMATION

Funding for the study was provided by the UC Davis Veterinary Genetics Laboratory.

CONFLICT OF INTEREST STATEMENT

R.R. Bellone, E. Esdaile, B. Till, A.K. Abdel-Ghaffar, and M. Hammond are affiliated with the UC Davis Veterinary Genetics Laboratory, which provides genetic diagnostic tests in horses and other species.

DATA AVAILABILITY STATEMENT

Data for this publication is deposited at the European Nucleotide Archive under PRJEB61164, PRJEB36403, PRJEB28306, PRJEB36381, PRJEB36380, and PRJEB30871.

ANIMAL CONSENT

The owners of horses in the study or representatives of the Jockey Club gave expressed written consent for participation or use of samples in this study.

ORCID

R. R. Bellone https://orcid.org/0000-0001-8838-7227
E. Esdaile https://orcid.org/0000-0001-6238-9793

REFERENCES

Aleman, M., Madigan, J.E., Williams, D.C. & Holliday, T.A. (2014) Brainstem auditory evoked responses in an equine patient population. Part II: foals. *Journal of Veterinary Internal Medicine*, 28, 1318–1324.

Aleman, M., Puchalski, S.M., Williams, D.C., Kass, P.H. & Holliday, T.A. (2008) Brainstem auditory-evoked responses in horses with temporohyoid osteoarthropathy. *Journal of Veterinary Internal Medicine*, 22, 1196–1202.

Aleman, M.R., True, A., Scalco, R., Crowe, C.M., Costa, L.R.R. & Chigerwe, M. (2021) Gentamicin-induced sensorineural auditory loss in healthy adult horses. *Journal of Veterinary Internal Medicine*, 35(5), 2486–2494.

Avila, F., Hughes, S.S., Magdesian, K.G., Penedo, M.C.T. & Bellone, R.R. (2022) Breed distribution and allele frequencies of base coat color, dilution, and white patterning variants across 28 horse breeds. *Genes*, 13, 1641.

Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G.R. et al. (2021) Accurate prediction of protein structures and interactions using a three-track neural network. *Science*, 373, 871–876.

Bellone, R.R., Holl, H., Setaluri, V., Devi, S., Maddodi, N., Archer, S. et al. (2013) Evidence for a retroviral insertion in TRPM1 as the cause of congenital stationary night blindness and leopard complex spotting in the horse. *PLoS One*, 8, e78280.

Blatter, M., Haase, B., Gerber, V., Poncet, P.A., Leeb, T., Rieder, S. et al. (2013) Clinical evaluation of the new coat colour macchiato in a male Franches-Montagnes horse. *Schweizer Archiv für Tierheilkunde*, 155, 229–232.

Brooks, S.A. & Bailey, E. (2005) Exon skipping in the KIT gene causes

225–230.

Capomaccio, S., Milanesi, M., Nocelli, C., Giontella, A., Verini-

(2017) Splicing site disruption in

or white dominant phenotype

, 48, 727–728.

oon, M., Nguyen, T., Wang,

otating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the

Dürig, N., Jude, R., Holl, H., Brooks, S.A., Lafayette, C., Jagannathan, V. et al. (2017) Whole genome sequencing reveals a novel deletion variant in the KIT gene in horses with white spotted coat colour

Esdaile, E., Till, B., Kallenberg, A., Fremeux, M., Bickel, L. & Bellone, R.R. (2022) A de novo missense mutation in KIT is responsible

e195.

variant in an Icelandic horse with white-spotted coat colour. *Animal Genetics*, 46, 466.

Haase, B., Signer-Hasler, H., Binns, M.M., Obexer-Ruff, G., Hauswirth, R., Bellone, R.R. et al. (2013) Accumulating mutations in series of haplotypes at the KIT and MITF loci are major determinants of white markings in Franches-Montagnes horses. *PLoS One*, 8, e75071.

Hauswirth, R., Haase, B., Blatter, M., Brooks, S.A., Burger, D., Drogemuller, C. et al. (2012) Mutations in MITF and PAX3 cause 'splashed white' and other white spotting phenotypes in horses. *PLoS Genetics*, 8, e1002653.

Hauswirth, R., Jude, R., Haase, B., Bellone, R.R., Archer, S., Holl, H. et al. (2013) Novel variants in the KIT and PAX3 genes in horses with white-spotted coat colour phenotypes. *Animal Genetics*, 44, 763–765.

Henkel, J., Lafayette, C., Brooks, S.A., Martin, K., Patterson-Rosa, L., Cook, D. et al. (2019) Whole-genome sequencing reveals a large deletion in the MITF gene in horses with white spotted coat colour and increased risk of deafness. *Animal Genetics*, 50, 172–174.

Holl, H.M., Brooks, S.A., Archer, S., Brown, K., Malvick, J., Penedo, M.C.T. et al. (2016) Variant in the RFWD3 gene associated

Acknowledgements

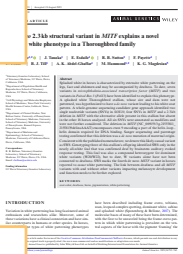
Funding

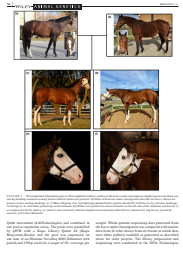
Conflict of interest

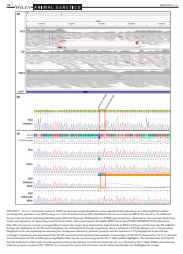
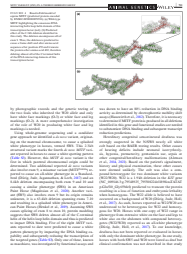
Data used

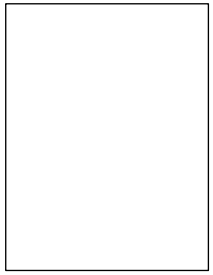
Ethics statement

Digital author identifier








when novel pigmentation variants are discovered and necessitating additional work to evaluate deafness in additional splashed white horses with and without other variants known to impact melanocyte proliferation, migration and function.

ACKNOWLEDGEMENTS

We acknowledge The Jockey Club for photographic records and for granting permission to use Thoroughbred samples banked at the UC Davis Veterinary Genetics Laboratory in this study. We also graciously acknowledge the owners who provided photographs and whose participation made this study possible.

FUNDING INFORMATION

Funding for the study was provided by the UC Davis Veterinary Genetics Laboratory.

CONFLICT OF INTEREST STATEMENT

R.R. Bellone, E. Esdaile, B. Till, A.K. Abdel-Ghaffar, and M. Hammond are affiliated with the UC Davis Veterinary Genetics Laboratory, which provides genetic diagnostic tests in horses and other species.

DATA AVAILABILITY STATEMENT

Data for this publication is deposited at the European Nucleotide Archive under PRJEB61164, PRJEB36403, PRJEB28306, PRJEB36381, PRJEB36380, and PRJEB30871.

ANIMAL CONSENT

The owners of horses in the study or representatives of the Jockey Club gave expressed written consent for participation or use of samples in this study.

ORCID

R. R. Bellone https://orcid.org/0000-0001-8838-7227
E. Esdaile https://orcid.org/0000-0001-6238-9702

REFERENCES

Section

Aleman, M., Madigan, J.E. [illegible] 2014) Brainstem auditory evoked responses in an equine patient population. Part II: foals. *Journal of Veterinary Internal Medicine*, 28, 1318–1324.

Aleman, M., Puchalski, S.M., Williams, D.C., Kass, P.H. & Holliday, T.A. (2008) Brainstem auditory-evoked responses in horses with temporohyoid osteoarthropathy. *Journal of Veterinary Internal Medicine*, 22, 1196–1202.

Aleman, M.R., True, A., Scalco, R., Crowe, C.M., Costa, L.R.R. & Chigerwe, M. (2021) Gentamicin-induced sensorineural auditory loss in healthy adult horses. *Journal of Veterinary Internal Medicine*, 35(5), 2486–2494.

Avila, F., Hughes, S.S., Magdesian, K.G., Penedo, M.C.T. & Bellone, R.R. (2022) Breed distribution and allele frequencies of base coat color, dilution, and white patterning variants across 28 horse breeds. *Genes*, 13, 1641.

Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G.R. et al. (2021) Accurate prediction of protein structures and interactions using a three-track neural network. *Science*, 373, 871–876.

Bellone, R.R., Holl, H., Setaluri, V., Devi, S., Maddodi, N., Archer, S. et al. (2013) Evidence for a retroviral insertion in TRPM1 as the cause of congenital stationary night blindness and leopard complex spotting in the horse. *PLoS One*, 8, e78280.

Blatter, M., Haase, B., Gerber, V., Poncet, P.A., Leeb, T., Rieder, S. et al. (2013) Clinical evaluation of the new coat colour macchiato in a male Franches-Montagnes horse. *Schweizer Archiv für Tierheilkunde*, 155, 229–232.

Brooks, S.A. & Bailey, E. (2005) Exon skipping in the KIT gene causes a Sabino spotting pattern in horses. *Mammalian Genome*, 16, 893–902.

Brooks, S.A., Lear, T.L., Adelson, D.L. & Bailey, E. (2007) A chromosome inversion near the KIT gene and the tobiano spotting pattern in horses. *Cytogenetic and Genome Research*, 119, 225–230.

Capomaccio, S., Milanesi, M., Nocelli, C., Giontella, A., Verini-Supplizi, A., Branca, M. et al. (2017) Splicing site disruption in the KIT gene as strong candidate for white dominant phenotype in an Italian trotter. *Animal Genetics*, 48, 727–728.

Cingolani, P., Platts, A., Wang le, L., Coon, M., Nguyen, T., Wang, L. et al. (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)*, 6, 80–92.

Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O. et al. (2021) Twelve years of SAMtools and BCFtools. *Gigascience*, 10, giab008.

Dürig, N., Jude, R., Holl, H., Brooks, S.A., Lafayette, C., Jagannathan, V. et al. (2017) Whole genome sequencing reveals a novel deletion variant in the KIT gene in horses with white spotted coat colour phenotypes. *Animal Genetics*, 48, 483–485.

Dürig, N., Jude, R., Jagannathan, V. & Leeb, T. (2017) A novel MITF variant in a white American Standardbred foal. *Animal Genetics*, 48, 123–124.

Esdaile, E., Till, B., Kallenberg, A., Fremeux, M., Bickel, L. & Bellone, R.R. (2022) A de novo missense mutation in KIT is responsible for dominant white spotting phenotype in a standardbred horse. *Animal Genetics*, 53, 534–537.

Haase, B., Brooks, S.A., Schlumbaum, A., Azor, P.J., Bailey, E., Alaeddine, F. et al. (2007) Allelic heterogeneity at the equine KIT locus in dominant white (W) horses. *PLoS Genetics*, 3, e195.

Haase, B., Brooks, S.A., Tozaki, T., Burger, D., Poncet, P.A., Rieder, S. et al. (2009) Seven novel KIT mutations in horses with white coat colour phenotypes. *Animal Genetics*, 40, 623–629.

Haase, B., Jagannathan, V., Rieder, S. & Leeb, T. (2015) A novel KIT variant in an Icelandic horse with white-spotted coat colour. *Animal Genetics*, 46, 466.

Haase, B., Signer-Hasler, H., Binns, M.M., Obexer-Ruff, G., Hauswirth, R., Bellone, R.R. et al. (2013) Accumulating mutations in series of haplotypes at the KIT and MITF loci are major determinants of white markings in Franches-Montagnes horses. *PLoS One*, 8, e75071.

Hauswirth, R., Haase, B., Blatter, M., Brooks, S.A., Burger, D., Drogemuller, C. et al. (2012) Mutations in MITF and PAX3 cause 'splashed white' and other white spotting phenotypes in horses. *PLoS Genetics*, 8, e1002653.

Hauswirth, R., Jude, R., Haase, B., Bellone, R.R., Archer, S., Holl, H. et al. (2013) Novel variants in the KIT and PAX3 genes in horses with white-spotted coat colour phenotypes. *Animal Genetics*, 44, 763–765.

Henkel, J., Lafayette, C., Brooks, S.A., Martin, K., Patterson-Rosa, L., Cook, D. et al. (2019) Whole-genome sequencing reveals a large deletion in the MITF gene in horses with white spotted coat colour and increased risk of deafness. *Animal Genetics*, 50, 172–174.

Holl, H.M., Brooks, S.A., Archer, S., Brown, K., Malvick, J., Penedo, M.C.T. et al. (2016) Variant in the RFWD3 gene associated

References are written in the style of the journal

References are very specific and often primary literature

with PATN1, a modifier of leopard complex spotting. *Animal Genetics*, 47, 91–101.
Holl, H.M., Brooks, S.A., Carpenter, M.L., Bustamante, C.D. & Lafayette, C. (2017) A novel splice mutation within equine KIT and the W15 allele in the homozygous state lead to all white coat color phenotypes. *Animal Genetics*, 48, 497–498.
Hug, P., Jude, R., Henkel, J., Jagannathan, V. & Leeb, T. (2019) A novel KIT deletion variant in a German Riding Pony with white-spotting coat colour phenotype. *Animal Genetics*, 50, 761–763.
Kingsley, N.B., Sandmeyer, L., Parker, S.E., Dwyer, A., Heden, S., Reilly, C. et al. (2022) Risk factors for insidious uveitis in the Knabstrupper breed. *Equine Veterinary Journal*, 55, 820–830.
Layer, R.M., Chiang, C., Quinlan, A.R. & Hall, I.M. (2014) LUMPY: a probabilistic framework for structural variant discovery. *Genome Biology*, 15, R84.
Lecoq, L., Gains, M., Blond, L. & Parent, J. (2015) Brainstem auditory evoked responses in foals: reference values, effect of age, rate of acoustic stimulation, and neurologic deficits. *Journal of Veterinary Internal Medicine*, 29, 362–367.
Li, H. & Durbin, R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25, 1754–1760.
Mack, M., Kowalski, E., Grahn, R., Bras, D., Penedo, M.C.T. & Bellone, R. (2017) Two variants in SLC24A5 are associated with 'Tiger-eye' iris pigmentation in Puerto Rican Paso Fino horses. *G3 (Bethesda)*, 7, 2799–2806.
Magdesian, K.G., Tanaka, J. & Bellone, R.R. (2020) A de novo MITF deletion explains a novel splashed white phenotype in an American Paint Horse. *The Journal of Heredity*, 111, 287–293.
Magdesian, K.G., Williams, D.C., Aleman, M., Lecouteur, R.A. & Madigan, J.E. (2009) Evaluation of deafness in American Paint Horses by phenotype, brainstem auditory-evoked responses, and endothelin receptor B genotype. *Journal of the American Veterinary Medical Association*, 235, 1204–1211.
Martin, K., Rosa, L.P., Vierra, M., Foster, G., Brooks, S.A. & Lafayette, C. (2021) De novo mutation of KIT causes extensive coat white patterning in a family of Berber horses. *Animal Genetics*, 52, 135–137.
Metallinos, D.L., Bowling, A.T. & Rine, J. (1998) A missense mutation in the endothelin-B receptor gene is associated with Lethal White Foal Syndrome: an equine version of Hirschsprung disease. *Mammalian Genome*, 9, 426–431.
Negro, S., Imsland, F., Valera, M., Molina, A., Sole, M. & Andersson, L. (2017) Association analysis of KIT, MITF, and PAX3 variants with white markings in Spanish horses. *Animal Genetics*, 48, 349–352.
Okonechnikov, K., Golosova, O., Fursov, M. & Ugene Team. (2012) Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics*, 28, 1166–1167.
Rockwell, H., Mack, M., Famula, T., Sandmeyer, L., Bauer, B., Dwyer, A. et al. (2020) Genetic investigation of equine recurrent uveitis in appaloosa horses. *Animal Genetics*, 51, 111–116.
Rosa, L.P., Martin, K., Vierra, M., Foster, G., Brooks, S.A. & Lafayette, C. (2022) Non-frameshift deletion on MITF is associated with a novel splashed white spotting pattern in horses (*Equus caballus*). *Animal Genetics*, 53, 538–540.
Rosa, L.P., Martin, K., Vierra, M., Lundquist, E., Foster, G., Brooks, S.A. et al. (2022) A KIT variant associated with increased white spotting epistatic to MC1R genotype in horses (*Equus caballus*). *Animals*, 12, 1958.
Sandmeyer, L.S., Kingsley, N.B., Walder, C., Archer, S., Leis, M.L., Bellone, R.R. et al. (2020) Risk factors for equine recurrent uveitis in a population of Appaloosa horses in western Canada. *Veterinary Ophthalmology*, 23, 515–525.
Santschi, E.M., Vrotsos, P.D., Purdy, A.K. & Mickelson, J.R. (2001) Incidence of the endothelin receptor B mutation that causes lethal white foal syndrome in white-patterned horses. *American Journal of Veterinary Research*, 62, 97–103.
Sponenberg, D.P. & Bellone, R. (2017) *Equine color genetics*. Hoboken, NJ: John Wiley & Sons, Inc.
Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M. et al. (2012) Primer3—new capabilities and interfaces. *Nucleic Acids Research*, 40, e115.
Van der Auwera, G.A., Carneiro, M.O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A. et al. (2013) From FastQ data to high confidence variant calls: the genome analysis toolkit best practices pipeline. *Current Protocols in Bioinformatics*, 43, 11.10.1–11.10.33.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Bellone, R.R., Tanaka, J., Esdaile, E., Sutton, R.B., Payette, F., Leduc, L. et al. (2023) A *de novo* 2.3 kb structural variant in *MITF* explains a novel splashed white phenotype in a Thoroughbred family. *Animal Genetics*, 54, 752–762. Available from: https://doi.org/10.1111/age.13352

44 How many references were used in the article compared to the review article?

with PATN1, a modifier of leopard complex spotting. *Animal Genetics*, 47, 91–101.
Holl, H.M., Brooks, S.A., Carpenter, M.L., Bustamante, C.D. & Lafayette, C. (2017) A novel splice mutation within equine KIT and the W15 allele in the homozygous state lead to all white coat color phenotypes. *Animal Genetics*, 48, 497–498.
Hug, P., Jude, R., Henkel, J., Jagannathan, V. & Leeb, T. (2019) A novel KIT deletion variant in a German Riding Pony with white-spotting coat colour phenotype. *Animal Genetics*, 50, 761–763.
Kingsley, N.B., Sandmeyer, L., Parker, S.E., Dwyer, A., Heden, S., Reilly, C. et al. (2022) Risk factors for insidious uveitis in the Knabstrupper breed. *Equine Veterinary Journal*, 55, 820–830.
Layer, R.M., Chiang, C., Quinlan, A.R. & Hall, I.M. (2014) LUMPY: a probabilistic framework for structural variant discovery. *Genome Biology*, 15, R84.
Lecoq, L., Gains, M., Blond, L. & Parent, J. (2015) Brainstem auditory evoked responses in foals: reference values, effect of age, rate of acoustic stimulation, and neurologic deficits. *Journal of Veterinary Internal Medicine*, 29, 362–367.
Li, H. & Durbin, R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25, 1754–1760.
Mack, M., Kowalski, E., Grahn, R., Bras, D., Penedo, M.C.T. & Bellone, R. (2017) Two variants in SLC24A5 are associated with 'Tiger-eye' iris pigmentation in Puerto Rican Paso Fino horses. *G3 (Bethesda)*, 7, 2799–2806.
Magdesian, K.G., Tanaka, J. & Bellone, R.R. (2020) A de novo MITF deletion explains a novel splashed white phenotype in an American Paint Horse. *The Journal of Heredity*, 111, 287–293.
Magdesian, K.G., Williams, D.C., Aleman, M., Lecouteur, R.A. &
Ma
Ho
and
Vet
Martin,
Laf
sive
Ge
Metallinos, D.L., Bowling, A.T. & Rine, J. (1998) A missense mutation in the endothelin-B receptor gene is associated with Lethal White Foal Syndrome: an equine version of Hirschsprung disease. *Mammalian Genome*, 9, 426–431.
Negro, S., Imsland, F., Valera, M., Molina, A., Sole, M. & Andersson, L. (2017) Association analysis of KIT, MITF, and PAX3 variants with white markings in Spanish horses. *Animal Genetics*, 48, 349–352.
Okonechnikov, K., Golosova, O., Fursov, M. & Ugene Team. (2012) Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics*, 28, 1166–1167.
Rockwell, H., Mack, M., Famula, T., Sandmeyer, L., Bauer, B., Dwyer, A. et al. (2020) Genetic investigation of equine recurrent uveitis in appaloosa horses. *Animal Genetics*, 51, 111–116.
Rosa, L.P., Martin, K., Vierra, M., Foster, G., Brooks, S.A. & Lafayette, C. (2022) Non-frameshift deletion on MITF is associated with a novel splashed white spotting pattern in horses (*Equus caballus*). *Animal Genetics*, 53, 538–540.
Rosa, L.P., Martin, K., Vierra, M., Lundquist, E., Foster, G., Brooks, S.A. et al. (2022) A KIT variant associated with increased white spotting epistatic to MC1R genotype in horses (*Equus caballus*). *Animals*, 12, 1958.
Sandmeyer, L.S., Kingsley, N.B., Walder, C., Archer, S., Leis, M.L., Bellone, R.R. et al. (2020) Risk factors for equine recurrent uveitis in a population of Appaloosa horses in western Canada. *Veterinary Ophthalmology*, 23, 515–525.
Santschi, E.M., Vrotsos, P.D., Purdy, A.K. & Mickelson, J.R. (2001) Incidence of the endothelin receptor B mutation that causes lethal white foal syndrome in white-patterned horses. *American Journal of Veterinary Research*, 62, 97–103.
Sponenberg, D.P. & Bellone, R. (2017) *Equine color genetics*. Hoboken, NJ: John Wiley & Sons, Inc.
Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M. et al. (2012) Primer3—new capabilities and interfaces. *Nucleic Acids Research*, 40, e115.
Van der Auwera, G.A., Carneiro, M.O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A. et al. (2013) From FastQ data to high confidence variant calls: the genome analysis toolkit best practices pipeline. *Current Protocols in Bioinformatics*, 43, 11.10.1–11.10.33.

Supplementary material

SUPPORTING INFORMATION
Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Bellone, R.R., Tanaka, J., Esdaile, E., Sutton, R.B., Payette, F., Leduc, L. et al. (2023) A *de novo* 2.3kb structural variant in *MITF* explains a novel splashed white phenotype in a Thoroughbred family. *Animal Genetics*, 54, 752–762. Available from: https://doi.org/10.1111/age.13352

Many fields require the inclusion of additional analyses, results, and raw data.

Such material is included as supplementary files.

Disclaimer

Figures, photos, and graphs in my lectures are collected using google searches. I do not claim to have personally produced the material (except for some). I do cite only articles or books used. I thank all owners of the visual aid that I use and apologize for not citing each individual item. If anybody finds the inclusion of their material in my lectures a violation of their copy rights, please contact me via email.

hhalhaddad@gmail.com