THE UNIVERSITY of ADELAIDE

AUSTRALIAN CENTRE FOR ANCIENT DNA

Ancient plant DNA to the rescue: unlocking crop genetic diversity from the past

Oscar Andres Estrada Santamaria

Australian Centre for Ancient DNA (ACAD)

School of Biological Sciences

Faculty of Sciences

The University of Adelaide

AUSTRALIA

Thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

March 2019

Ancient plant DNA to the rescue: unlocking crop genetic diversity from the past

A Thesis Submitted By

Oscar Andres Estrada Santamaria

B.Sc. (Hons), M.Sc.

For the Degree Of

Doctor of Philosophy (Ph.D.)

March 2019

Australian Centre for Ancient DNA (ACAD)

School of Biological Sciences

Faculty of Sciences

The University of Adelaide

Supervisors:

Prof. Alan Cooper

Dr James Breen

Dr Stephen M Richards

Table of Contents

The University of Adelaide

Abstract

Australian Centre for Ancient DNA (ACAD)
School of Biological Sciences
Faculty of Sciences

Doctor of Philosophy

Ancient plant DNA to the rescue: unlocking crop genetic diversity from the past

By Oscar Andres Estrada Santamaria

The domestication of animals and plants around 12,000 years ago triggered one of the most transcendental revolutions in human society; from a hunter-gatherer lifestyle to a sedentary, agricultural-based society. Since that time, it is estimated that almost 2,500 species of plants have been subject to domestication. However, the domestication history of most of these species remains unclear, with fundamental questions about the identity of wild ancestors, selected phenotypes (domestication traits), and timing and origin of domestication still to be answered. The impact of the extensive selective pressures exerted by humans on the genomes of staple crops can potentially be understood by reconstructing the genomes of ancient plant specimens. This doctorate thesis aims to develop and apply new methods to access the genetic information of historical and ancient plant specimens of staple crops such as wheat and quinoa. Using modern genomic techniques such as hybridisation capture, high-throughput sequencing (HTS) and bioinformatic data mining and analysis, I reconstructed the phylogeny and examined the genetic diversity of modern and historical specimens of wheat (*Triticum aestivum*, *T. timopheevii*, and *T. turgidum*) and ancient specimens of quinoa (*Chenopodium quinoa*). Two newly designed hybridisation capture arrays unearthed a large amount of genetic variability occurring 100 years ago in historical wheat species from Georgia, a region with a significant number of native varieties. Additionally, deep sequencing of ancient quinoa specimens (~1,400 years-old) from the highlands of north Argentina led to the reconstruction of the first known ancient genomes reported from a crop domesticated in South America and identified a bottleneck in the recent history of quinoa. The new knowledge gained provides a potential resource for further research on plant ancient DNA and plant domestication, as well as the investigation of genetic changes that occurred in loci of breeding value.

Declaration of Authorship

I, Oscar Andres Estrada Santamaria, certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name, in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the test. In addition, I certify that no part of this work will, in the future be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint-award of this degree.

I give consent to this copy of my thesis, when deposited in the University Library, being made available for loan and photocopying, subject to the provisions of the Copyright Act 1968.

I acknowledge that copyright of published works contained within this thesis resides with the copyright holder(s) of those works.

I also give permission for the digital version of my thesis to be made available on the web via the University's digital research repository, the Library Search and also through the web search engines, unless permission has been granted by the University to restrict access for a period of time.

…….. ……..

Oscar Andres Estrada Santamaria Date: March 8, 2019

Acknowledgements

Firstly, I would like to express my sincere gratitude to my supervisor Dr Jimmy Breen for his unconditional support, continuous advice, and providing valuable feedback on my work. His guidance on both research as well as on my career has been invaluable.

My sincere thanks also go to my principal supervisor Prof Alan Cooper for giving me the opportunity to join the Australian Centre for Ancient DNA (ACAD) and for his great advice, mentorship and tuition during my candidature. I would also like to thank my co-supervisor Dr Steve Richards for sharing his molecular biology expertise so willingly and provide feedback on my work.

I would like to extend my thanks to the staff and students at ACAD (both past and present), especially to Maria Lekis for her administrative support and my fellow labmates in for the stimulating discussions and collaborative work environment. Thanks also to all my collaborators, particularly to Pilar Babot (The University of Tucuman), Salomon Hocsman (The University of Tucuman), Sabrina Costa (CONICET), Nana Rusishvili (Georgian National Museum), Nelli Hovhannisyan (Yerevan State University) Boris Gasparyan (Armenian Academy of Sciences), Mark Tester (King Abdullah University of Science and Technology) and David Jarvis (Brigham Young University).

I would like to thank the financial support towards my PhD from the Administrative Department of Science, Technology and Innovation of Colombia (COLCIENCIAS) and the Graduate Centre of the University of Adelaide.

Finally, I am most grateful to my family and my partner Andrea for keeping me together and for their invaluable support throughout my doctorate.

Chapter 1

Introduction

Introduction

Earth's first plants began to colonise the land over 470 million years ago, triggering the evolution of climate and subsequent diversification of life (Figure 1) (Heckman *et al.*, 2001; Beerling *et al.*, 2010; Strother *et al.*, 2011). Photosynthesis transformed the atmosphere by simultaneously increasing the percentage of oxygen on Earth and decreasing the percentage of carbon dioxide (Heckman *et al.*, 2001; Knauth & Kennedy, 2009). The development of roots in early plants promoted fixation of carbon dioxide and the formation and acidification of soils (Mora *et al.*, 1996; Knauth & Kennedy, 2009). The evolution of stomata and vascular tissues facilitated gas exchange and distribution of water and nutrients through the plant (Kenrick, 2001; Lucas *et al.*, 2013). Subsequent innovations in nutrient storage, life cycles, and reproduction paved the way for the colonisation of most terrestrial ecosystems, which in-turn enabled animal and human evolution (Cronk, 2001). The ability of plants to evolve in such a way has been driven by genetic changes and epigenetic mechanisms, which are now valuable markers of large-scale climate change and evolution.

Plants have also been essential to the evolution of the modern human society. The rise of agriculture 12,000 years ago enabled humans to transition from a hunter-gathering lifestyle to a sedentary and agriculture-based society. Domestication of plant and animal species allowed humans to have control over their food resources, store food surpluses and protect them from climatic and environmental changes (Salamini *et al.*, 2002; Brown *et al.*, 2009). As a result of a stable food supply, agriculture allowed the development of a dense human population and a sedentary lifestyle, which led to the stratification of society, subsequent specialisation and division of labour, development of art, trading economies, and political structures (Salamini *et al.*, 2002; Brown *et al.*, 2009; Zohary *et al.*, 2013).

Evolution by domestication is recognised as an exceptional model for the study of evolutionary processes (Meyer & Purugganan, 2013). In fact, Charles Darwin in *The Origin of Species by Means of Natural Selection*, dedicated his first chapter to *Variation under Domestication* (Darwin, 1859), followed by a complete publication on *Variation of Plant and Animals under Domestication* (Darwin, 1868). Darwin stressed that intensive selection during domestication has produced domestic varieties that in most cases are not able to live in a wild state. Darwin also pointed out the high morphological variability observed in domestic varieties when

compared with their wild relatives and discusses the difficulty in identifying wild progenitor species and the centres of origin of crops (Darwin, 1859).

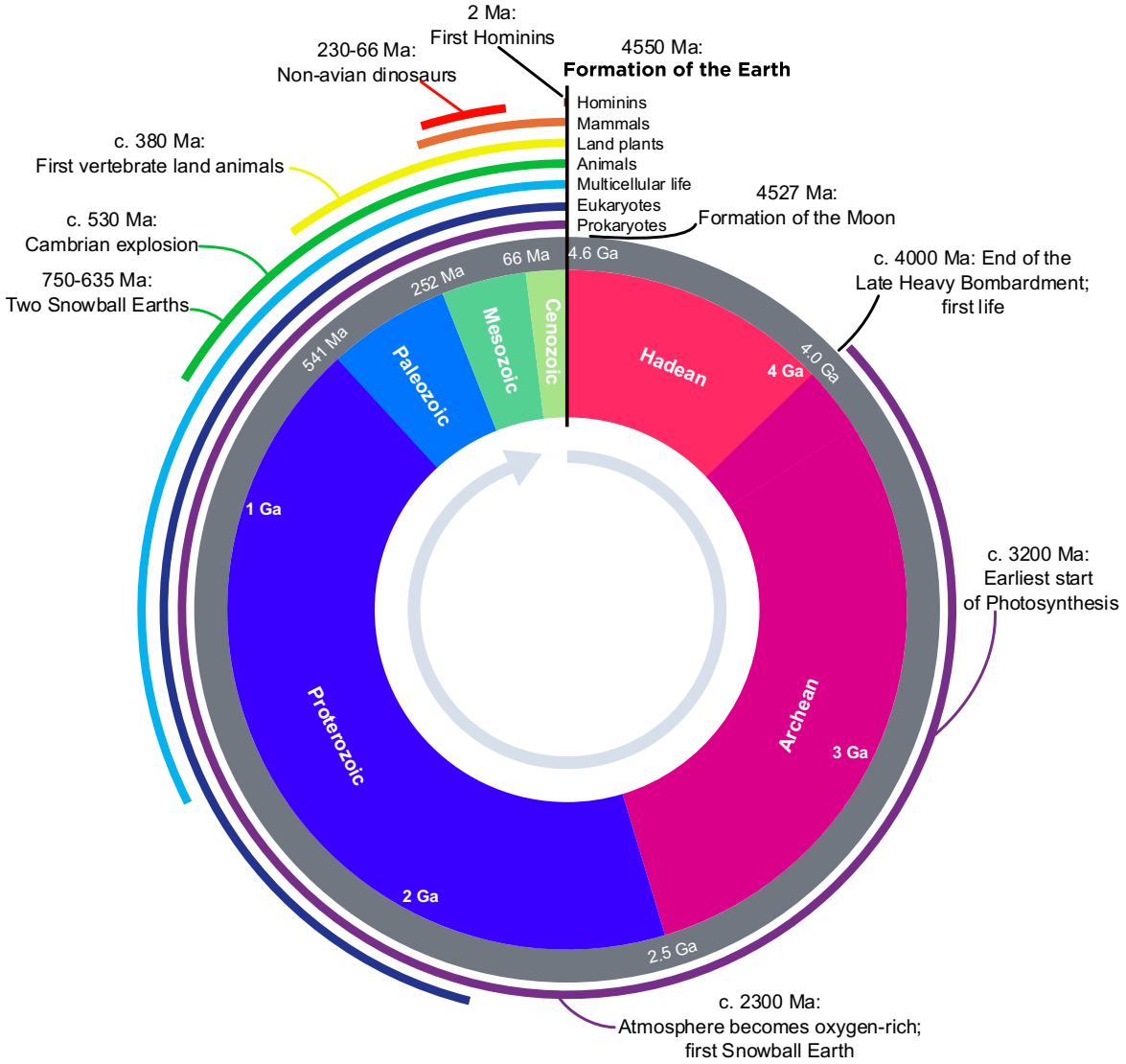


Figure 1. Geological clock representation showing the major units of geological time and definitive events of Earth history (Source: Wikimedia Commons). Notice the colonisation of land plants in the Palaeozoic era, 470 million years ago (Ma).

Domestication could be defined as a mutualistic relationship in which one species has a significant degree of control over the growth and reproduction of another species in order to obtain a resource of interest (Zeder, 2015). The domestication process involves altering key traits that lead to the establishment of a new domesticated species and often increasing the fitness of both species under the mutualistic relationship (Meyer & Purugganan, 2013; Zeder, 2015). Evolution by domestication has been documented in several non-human species including the fungus-growing ants, the fungus-growing termites and the ambrosia beetles

(Mueller *et al.*, 2005). The domestication of fungi by ant and termite species has been a gradual process spanning millions of years (Mueller *et al.*, 2005; Zeder, 2015). However, the most successful domesticator species, humans, has domesticated hundreds of plant and animal species in a short time (Figure 2) (Mueller *et al.*, 2005; Meyer *et al.*, 2012).

Evidence from the archaeological record suggests that the first animal (the dog) was domesticated between 20,000 and 40,000 BP (Skoglund *et al.*, 2015; Botigué *et al.*, 2017) and that the first plants (*e.g.* wheat and barley) were domesticated around 12,000 years before present (BP) (Figure 2) (Salamini *et al.*, 2002; Brown *et al.*, 2009). Archaeological data also shows two major chronological periods of plant domestication: early Holocene from about 12,000 to 9,000 BP, and the middle Holocene between 7,000 and 4,000 BP (Figure 2). The combination of archaeological, molecular, diversity, production, trade, and consumption data have led to the identification of about 20 main geographical regions of domestication (Figure 2) (Larson et al., 2014; Khoury et al., 2016).

The evolution from wild to domesticated species involves genetic changes associated with the phenotypic traits selected in the cultivated form from the wild progenitor; this set of traits is recognised as the "domestication syndrome" (Gepts, 2004; Kilian *et al.*, 2010). The characteristics selected during domestication can vary depending on the species but often include the loss of seed dormancy, increase in seed size and changes in the reproductive structures (Meyer & Purugganan, 2013). In annual plants such as grasses, the most critical traits modified during domestication were the free-threshing state (whereby seeds are released from the rachis at threshing) and the brittle rachis, a trait that allowed stable capture of seeds without the head shattering on the ground (Salamini *et al.*, 2002; Brown *et al.*, 2009; Kilian *et al.*, 2010). Further changes during domestication and breeding involve changes in plant height, seed hardness, photoperiod, vernalisation, synchronised flowering, nutritional content, and enhanced culinary chemistry (Brown *et al.*, 2009; Kilian *et al.*, 2010; Meyer & Purugganan, 2013).

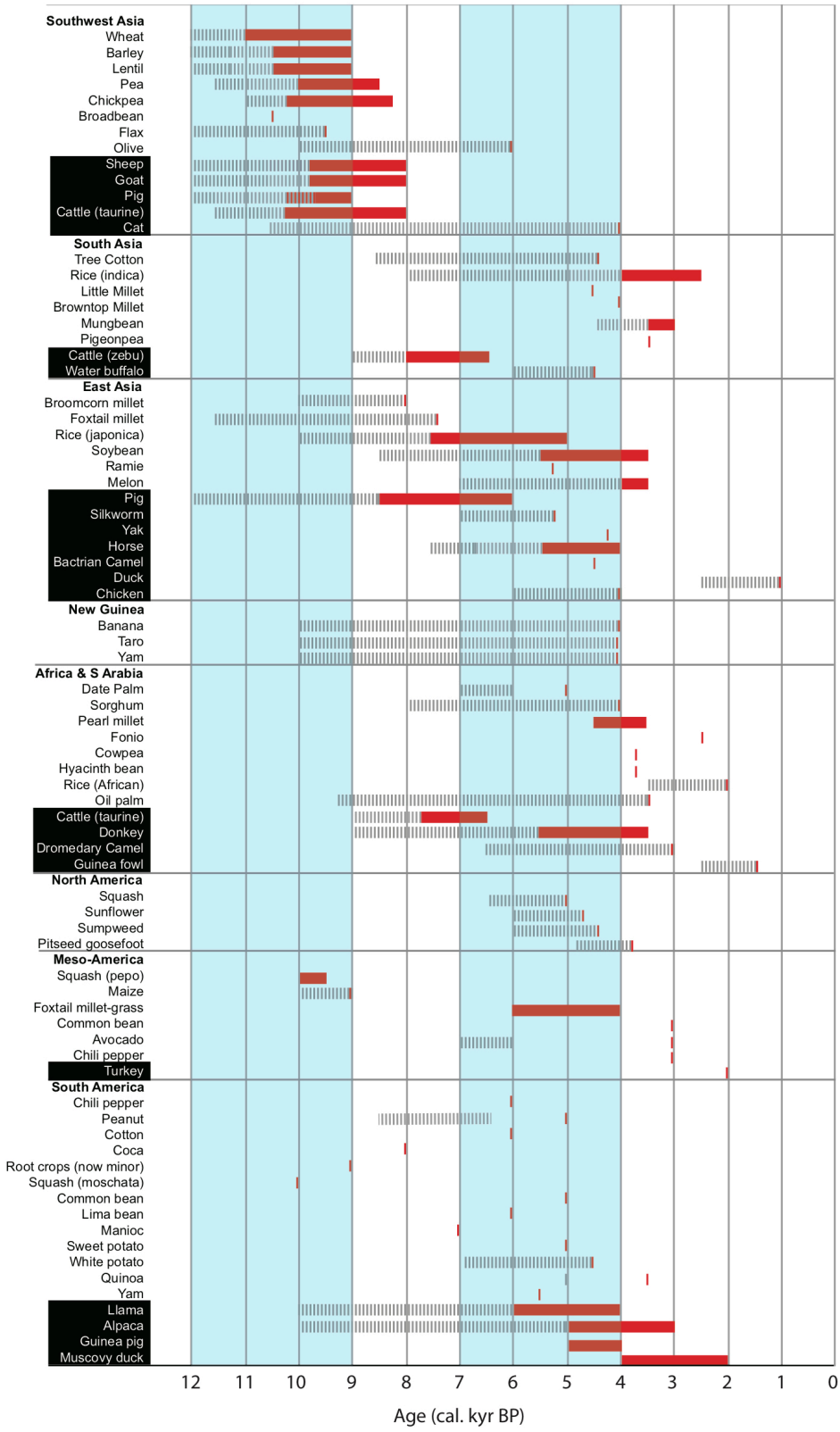


Figure 2. A chronological chart listing the regions and the time frames over which key plants and animals were domesticated. Grey dashed lines represent documented exploitation or management before domestication. Red bars show the documented time of domestication. Blue shaped areas denote two major chronological periods of plant domestication: early Holocene from about 12,000 to 9,000 BP, and the middle Holocene between 7,000 and 4,000 BP. Plot modified from (Larson *et al.*, 2014).

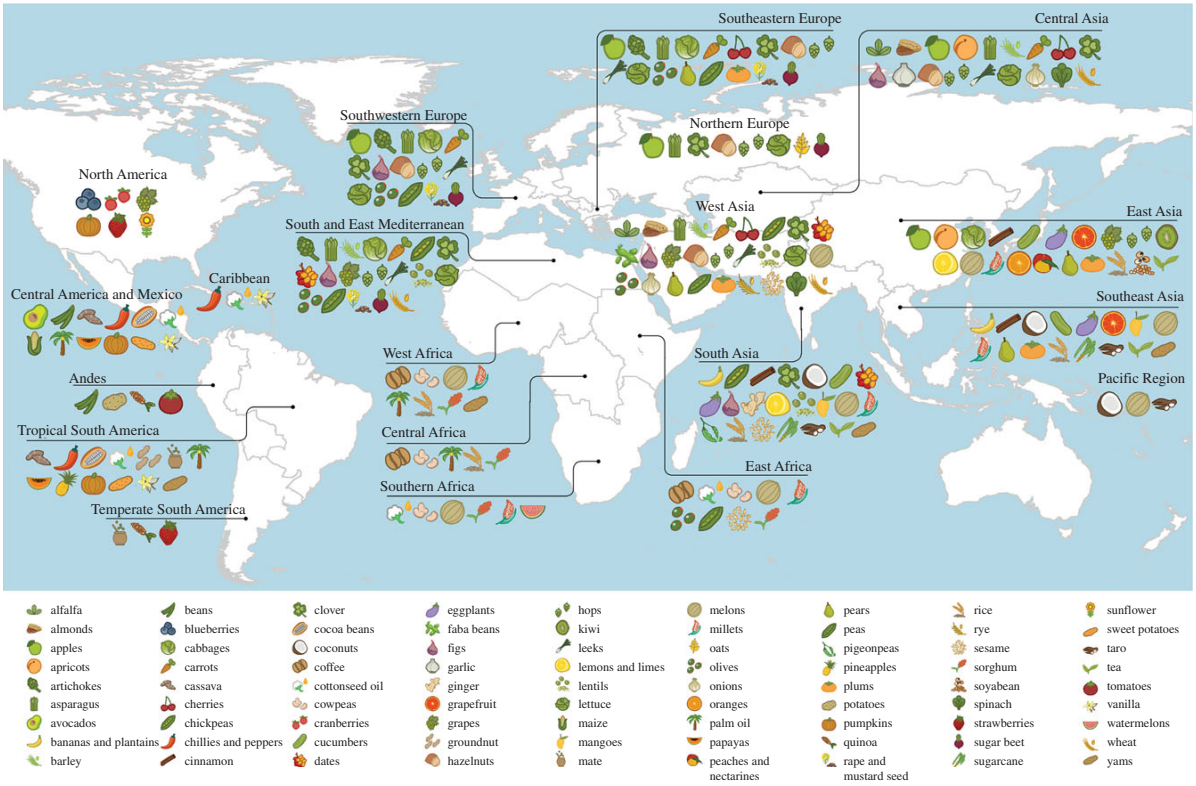


Figure 3. Primary geographical regions of domestication and diversity of major agricultural crops worldwide (Khoury *et al.*, 2016).

Molecular studies have greatly enhanced our ability to understand the genetic architecture of crop domestication and the nature of selection in domesticated species (Doebley *et al.*, 2006). Domestication-related genes have mostly been identified by quantitative trait locus (QTL) mapping and cloning, genome-wide association studies (GWAS) and whole-genome sequencing (Meyer & Purugganan, 2013; Bevan *et al.*, 2017). *Teosinte branched1* (*tb1*) was the first domestication-related gene identified in maize via QTL mapping, which is a locus controlling branching differences between maize and its wild progenitor, teosinte (Doebley *et al.*, 1997). After the identification of *tb1* in maize, genetic studies have identified around 50 domestication loci in important crops such as wheat, barley, rice, and maize (Meyer & Purugganan, 2013; Bevan *et al.*, 2017). However, despite the progress in identifying genes related to domestication traits in the main crops, several loci remain unidentified. For instance, a genomic study in maize estimates that more than 1,700 genes show evidence for selection during domestication (Hufford *et al.*, 2012), yet only 13 loci have been identified through QTL mapping or GWAS studies (Meyer & Purugganan, 2013; Stitzer & Ross-Ibarra, 2018).

To date, knowledge of crop domestication has largely been inferred from the morphological analysis of archaeobotanical remains, analysis of microfossils, and genetic analysis of present-day samples (Salamini *et al.*, 2002; Brown *et al.*, 2009; Zohary *et al.*, 2013). Most of this knowledge is derived from the study of model crops such as maize, rice, wheat, and barley. However, it is estimated that more than two thousand plant species have been subject to domestication (Meyer *et al.*, 2012). Although the archaeological record has shed light on the timing of domestication of some species, the fundamental questions proposed by Darwin regarding the origin of domestication, wild ancestors, and domestication traits are yet to be answered. Additionally, domestication and intensive breeding have modified the genomes of domesticated species to increase desirable traits but have also reduced undesirable and non-target traits. This process of selection has led to a major reduction in the diversity of domesticated species, with such genetic bottlenecks presenting limitations for present-day crop improvement (Doebley *et al.*, 2006; Bevan *et al.*, 2017).

Ancient DNA (aDNA) research opens a new path to go back in time and understand the role of artificial selection and the evolutionary process that led to the generation of domesticated species (Orlando *et al.*, 2015; Der Sarkissian *et al.*, 2015). Access to the genetic makeup of dated archaeobotanical remains provides a powerful opportunity to examine the genomic regions, alleles, genes and genome combinations that have changed at various stages of domestication. However, due to the degraded nature of the archaeobotanical specimens, studies on plant ancient DNA pose several limitations. First, most archaeobotanical remains are found in regions of temperate climates, and therefore exhibit low preservation due to extreme variations in temperature and humidity (Palmer *et al.*, 2009; Allaby *et al.*, 2015). Second, standard methods used in aDNA studies are not optimised for botanical remains, and thus, do not consider intrinsic characteristics of plants such as the high content of secondary compounds (Wales *et al.*, 2014). Lastly, plant genomes are extremely large (0.5-22 Gigabases), include a high proportion of repetitive elements (30-90%) and can display several levels of ploidy (2x-12x), which challenges sequencing data recovery and analysis (Jiao & Schneeberger, 2017).

This thesis aims to develop and apply new methods to access the genetic information of degraded plant materials of important crops such as wheat and quinoa. Using modern techniques such as hybridisation capture (*chapter 4*), high-throughput sequencing (HTS) and bioinformatics data mining and analysis (*chapter 3, 4* and *5*), I was able to reconstruct the phylogeny and examine the genetic diversity of modern and historical specimens of wheat (*Triticum* spp. L.) and ancient specimens of quinoa (*Chenopodium quinoa* Willd.). The knowledge gained provides new resources for further research on plant ancient DNA and crop domestication as well as to investigate in detail the genetic changes that occurred in loci of breeding value. The thesis is divided into seven chapters outlining the main research questions, objectives and results.

Firstly, I discuss (*Chapter 2*) and develop (*Chapter 3*) methods that enable the efficient analysis of ancient plant DNA. In *Chapter 2* I summarise the recent achievements in the field of ancient plant DNA, discuss the limitations involved in the isolation and analysis of plant aDNA, and suggest potential improvements that will facilitate the contributions that plant aDNA could have on future research. Subsequently, in *Chapter 3* I evaluate the overall efficiency of four methods to extract DNA from modern and historical seeds (60-100-years-old) of three wheat species with different levels of ploidy and examine the potential use in downstream analysis using quantitative PCR (qPCR) and high-throughput sequencing. Within this chapter, I also assess the utility of the four DNA extraction protocols for the retrieval of DNA in ancient seeds of bread wheat (*Triticum aestivum* L.) from over 1,200 years of age.

HTS and hybridisation capture are important tools for recovering aDNA from poorly preserved materials. Thus, in *Chapter 4* I describe the development of two hybridisation capture arrays to study the plastid and nuclear genetic diversity of a panel of historical samples (60-100-years-old) of polyploid wheat species (*T. aestivum*, *T. timopheevii*, and *T. turgidum*) from Georgia, a region containing a significant array of native varieties. In *Chapter 5* I report on the aDNA sequencing and analysis of four ~1,400 years old archaeological samples of *C. quinoa* from the highlands of northern Argentina, which enabled the recovery of full chloroplast genomes for phylogenetic analysis. The dataset also provides evidence of a bottleneck in the chloroplast diversity of modern quinoa, with two of the identified ancient haplotypes absent in the modern accessions. Notably, these are the first known ancient plant genomes sequences reported from a crop domesticated in South America.

To summarise the work, in *chapter 6* I discuss the thesis findings along with the concluding remarks of the doctoral dissertation. The relevance of the knowledge achieved in the thesis is discussed in the context of the plant domestication research, as well as the potential implications for plant breeding through the application of novel approaches such as mutant screening or genome editing. Finally, in *chapter 7*, I outline the significance and perspectives of future ancient plant DNA research, which was recently published in *Nature Plants* (Estrada *et al.*, 2018).

References

Allaby RG, Gutaker R, Clarke AC, Pearson N, Ware R, Palmer SA, Kitchen JL, Smith O. **2015**. Using archaeogenomic and computational approaches to unravel the history of local adaptation in crops. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **370**: 20130377.

Beerling DJ, Bidartondo MI, Franks PJ, Humphreys CP, Rees M, Leake JR. **2010**. Mutualistic mycorrhiza-like symbiosis in the most ancient group of land plants. *Nature Communications* **1**: 103.

Bevan MW, Uauy C, Wulff BBH, Zhou J, Krasileva K, Clark MD. **2017**. Genomic innovation for crop improvement. *Nature* **543**: 346–354.

Botigué LR, Song S, Scheu A, Gopalan S, Pendleton AL, Oetjens M, Taravella AM, Seregély T, Zeeb-Lanz A, Arbogast R-M, *et al.* 2017. Ancient European dog genomes reveal continuity since the Early Neolithic. *Nature Communications* **8**: 16082.

Brown TA, Jones MK, Powell W, Allaby RG. **2009**. The complex origins of domesticated crops in the Fertile Crescent. *Trends in Ecology and Evolution* **24**: 103–109.

Cronk QCB. **2001**. Plant evolution and development in a post-genomic context. *Nature Reviews Genetics* **2**: 607–619.

Darwin C. **1859**. *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. London: John Murray.

Darwin C. **1868**. *Variation of plants and animals under domestication*. John Murray.

Doebley JF, Gaut BS, Smith BD. **2006**. The molecular genetics of crop domestication. *Cell* **127**: 1309–21.

Doebley J, Stec A, Hubbard L. **1997**. The evolution of apical dominance in maize. *Nature* **386**: 485–488.

Estrada O, Breen J, Richards SM, Cooper A. **2018**. Ancient plant DNA in the genomic era . *Nature Plants*.

Gepts P. **2004**. Crop domestication as a long-term Selection Experiment. *Plant Breeding Reviews* **24**: 1–44.

Heckman DS, Geiser DM, Eidell BR, Stauffer, R.L., Kardos, N.L., Hedges SB. **2001**. Molecular evidence for the early colonization of land plants by fungi and plants. *Science* **293**: 1129–1133.

Hufford MB, Xu X, van Heerwaarden J, Pyhäjärvi T, Chia J-M, Cartwright RA, Elshire RJ, Glaubitz JC, Guill KE, Kaeppler SM, *et al.* 2012. Comparative population genomics of maize domestication and improvement. *Nature Genetics* **44**: 808–811.

Jiao WB, Schneeberger K. **2017**. The impact of third generation genomic technologies on plant genome assembly. *Current Opinion in Plant Biology* **36**: 64–70.

Kenrick P. **2001**. Turning over a new leaf. *Nature* **410**: 309–310.

Khoury CK, Achicanoy HA, Bjorkman AD, Navarro-Racines C, Guarino L, Flores-Palacios X, Engels JMM, Wiersema JH, Dempewolf H, Sotelo S, *et al.* 2016. Origins of food crops connect countries worldwide. *Proceedings of the Royal Society B: Biological Sciences* **283**: 20160792.

Kilian B, Martin W, Salamini F. **2010**. *Genetic Diversity, Evolution and Domestication of Wheat and Barley in the Fertile Crescent* (M Glaubrecht, Ed.).

Knauth LP, Kennedy MJ. **2009**. The late Precambrian greening of the Earth. *Nature* **460**: 728–732.

Larson G, Piperno DR, Allaby RG, Purugganan MD, Andersson L, Arroyo-Kalin M, Barton L, Climer Vigueira C, Denham T, Dobney K, *et al.* 2014. Current perspectives and the future of domestication studies. *Proceedings of the National Academy of Sciences* **111**: 6139–6146.

Lucas WJ, Groover A, Lichtenberger R, Furuta K, Yadav S-R, Helariutta Y, He X-Q, Fukuda H, Kang J, Brady SM, *et al.* 2013. The Plant Vascular System: Evolution, Development and Functions F. *Journal of Integrative Plant Biology* **55**: 294–388.

Meyer RS, DuVal AE, Jensen HR. **2012**. Patterns and processes in crop domestication: an historical review and quantitative analysis of 203 global food crops. *The New Phytologist* **196**: 29–48.

Meyer RS, Purugganan MD. **2013**. Evolution of crop species: Genetics of domestication and diversification. *Nature Reviews Genetics* **14**: 840–852.

Mora C, Driese SG, Colarusso LA. **1996**. Middle to Late Paleozoic Atmospheric CO $ _ 2 $ Levels from Soil Carbonate and Organic Matter Author (s): Claudia I . Mora , Steven G . Driese and Lee Ann Colarusso. **271**: 1105–1107.

Mueller UG, Gerardo NM, Aanen DK, Six DL, Schultz TR. **2005**. The Evolution of Agriculture in Insects. *Annual Review of Ecology, Evolution, and Systematics* **36**: 563–595.

Orlando L, Gilbert MTP, Willerslev E. **2015**. Reconstructing ancient genomes and epigenomes. *Nature Reviews Genetics* **16**: 395–408.

Palmer SA, Moore JD, Clapham AJ, Rose P, Allaby RG. **2009**. Archaeogenetic evidence of ancient nubian barley evolution from six to two-row indicates local adaptation. *PLoS ONE* **4**: 2–8.

Salamini F, Ozkan H, Brandolini A, Schäfer-Pregl R, Martin W. **2002**. Genetics and geography of wild cereal domestication in the near east. *Nature reviews. Genetics* **3**: 429–441.

Der Sarkissian C, Allentoft ME, Avila-Arcos MC, Barnett R, Campos PF, Cappellini E, Ermini L, Fernandez R, da Fonseca R, Ginolhac A, *et al.* 2015. Ancient genomics. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **370**: 20130387–20130387.

Skoglund P, Ersmark E, Palkopoulou E, Dalén L. **2015**. Ancient wolf genome reveals an early divergence of domestic dog ancestors and admixture into high-latitude breeds. *Current Biology* **25**: 1515–1519.

Stitzer MC, Ross-Ibarra J. **2018**. Maize domestication and gene interaction. *New Phytologist* **220**: 395–408.

Strother PK, Battison L, Brasier MD, Wellman CH. **2011**. Earth's earliest non-marine eukaryotes. *Nature* **473**: 505–509.

Wales N, Andersen K, Cappellini E, ??vila-Arcos MC, Gilbert MTP. **2014**. Optimization of DNA recovery and amplification from non-carbonized archaeobotanical remains. *PLoS ONE* **9**.

Zeder MA. **2015**. Core questions in domestication research. *Proceedings of the National Academy of Sciences* **112**: 3191–3198.

Zohary D, Hopf M, Weiss E. **2013**. *Domestication of plants in the Old World : the origin and spread of domesticated plants in Southwest Asia, Europe, and the Mediterranean Basin*. Oxford : Oxford University Press.

Chapter 2

Limitations, progress and prospects of ancient plant DNA

Manuscript prepared for publication

Statement of authorship

Title of Paper	Limitations, progress and prospects of ancient plant DNA
Publication Status	☐ Published ☐ Accepted for Publication ☐ Submitted for Publication ☑ Unpublished and Unsubmitted w ork w ritten in manuscript style
Publication Details	Manuscript prepared for submission to the journal *Frontiers in Plant Science*

Principal Author

Name of Principal Author (Candidate)	Oscar Estrada		
Contribution to the Paper	Contributed to the conception and design of the manuscript, reviewed literature, wrote and edited the manuscript.		
Overall percentage (%)	70		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	08/03/2019

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	James Breen		
Contribution to the Paper	Contributed to the conception and design of the manuscript, reviewed and edited the manuscript.		
Signature		Date	08/03/219

Name of Co-Author	Stephen M Richards		
Contribution to the Paper	Contributed to the conception and experimental design, advised laboratory analyses, reviewed and edited the manuscript.		
Signature		Date	09/03/2019

Please cut and paste additional co-author panels here as required.

Name of Co-Author	Alan Cooper		
Contribution to the Paper	Contributed to the conception and experimental design, reviewed and edited the manuscript		
Signature		Date	8/3/19

Limitations, progress and prospects of ancient plant DNA

Oscar Estrada[1], Jimmy Breen[2], Stephen M Richards[1], Alan Cooper[1].

[1]Australian Centre for Ancient DNA, University of Adelaide, North Terrace, South Australia 5005, Australia.
[2]Robinson Research Institute, University of Adelaide, SA, 5005, Australia.

Abstract

Ancient DNA (aDNA) research has experienced substantial progress in recent years through the application of high-throughput sequencing (HTS) technologies together with improved methods for the recovery and authentication of endogenous molecules. However, despite the value in identifying how plants have evolved in the face of selective breeding and domestication, studies in ancient plant DNA have lagged behind those in other organisms like humans and vertebrate taxa. Nonetheless, plant aDNA research has progressed from the analysis of a few loci to the retrieval of complete genomes dating back to the early and middle Holocene. In this review, we summarise the recent achievements in the field of ancient plant DNA, discuss the limitations involved in the isolation and analysis of aDNA from botanical remains, and suggest potential improvements that will facilitate the significant contributions that ancient plant DNA could have on future research.

Introduction

Since the recovery and sequencing of a 229 nucleotide-long DNA amplicon from an extinct subspecies of zebra more than three decades ago (Higuchi *et al.*, 1984), the analysis of nucleic acids recovered from degraded and subfossil specimens has become an essential source of information for studies in evolutionary biology, palaeontology, archaeology, and forensics (Hofreiter *et al.*, 2001; Willerslev & Cooper, 2005; Der Sarkissian *et al.*, 2015). Early ancient DNA (aDNA) studies made used of conventional PCR amplification, cloning and chain-termination sequencing (Sanger *et al.*, 1977) to reconstruct sequences of small DNA fragments from ancient organisms (Hofreiter *et al.*, 2001; Paabo *et al.*, 2004). Twenty years later, the introduction of high-throughput sequencing (HTS) technologies revolutionised the field, allowing the sequencing of thousands of short DNA molecules (as short as 25 bp) at once (Margulies et al. 2005) while a range of methods for authentication and target-sequence enrichment have also been developed (Knapp & Hofreiter, 2010; Hofreiter *et al.*, 2014; Der Sarkissian *et al.*, 2015).

By applying modern genomic technologies such as HTS and capture enrichment, studies in aDNA have provided pivotal insights into human movement and evolution, animal phylogenetics, plant and animal domestication, palaeoenvironments, and most recently, changes in human microbiota (Orlando & Cooper, 2014; Weyrich *et al.*, 2015; MacHugh *et al.*, 2016; Estrada *et al.*, 2018). For instance, studies of aDNA from archaeobotanical remains have provided pivotal information to address evolutionary and anthropological questions, such as the origin and domestication of crops, adaptive evolution of cereals, and how agriculture spread (Jaenicke-Després *et al.*, 2003; Li *et al.*, 2011; Palmer *et al.*, 2012a; Fordyce *et al.*, 2013; da Fonseca *et al.*, 2015; Li *et al.*, 2016; Mascher *et al.*, 2016; Ramos-Madrigal *et al.*, 2016a).

Despite advances in genomic technologies, there have been fewer studies in ancient plant DNA compared to those on other organisms (Gugerli *et al.*, 2005; Schlumbaum *et al.*, 2008; Palmer *et al.*, 2012b; Parducci *et al.*, 2017). Studies from bones of large animals spanning hundreds of thousands of years have potentially been more attractive to researchers than studies from plant materials. However, several ancient plant DNA studies have failed to recover authentic endogenous DNA molecules from archaeobotanical remains (Brown *et al.*, 1998; Blatter *et al.*, 2002; Oliveira *et al.*, 2012; Nistelberger *et al.*, 2016). Although this is mainly due to the lack of well-preserved samples suitable for aDNA research, plant samples are also rich in secondary

compounds that hinder DNA retrieval and may act as inhibitors in molecular biology reactions applied in aDNA research (Bessetti, 2007; Samarakoon *et al.*, 2013). Likewise, the sequencing and assembly of plant genomes such as cereals can be challenging due to the presence of large genomes sizes (4-22 Gigabases), high levels of repetitive elements (often more than 80%), extensive variation in organelle genome size, and diverse levels of ploidy observed in the genomes of various species (Egan *et al.*, 2012; Cronn *et al.*, 2012; Jiao & Schneeberger, 2017).

In this review, we examine the principal methodological challenges of ancient plant DNA research from archaeobotanical remains of domesticated species. We discuss the availability of samples suitable for aDNA studies, the limitations of isolating quality aDNA, and the technical challenges facing ancient plant DNA analysis using high-throughput sequencing technologies. Finally, we provide recommendations on technical issues and suggest directions for future research on ancient plant DNA.

Ancient DNA

Ancient DNA is generally described as the DNA recovered from remains of organisms that have undergone taphonomic processes over time, and may also be present as extracellular DNA in the absence of any physical remains (Hofreiter *et al.*, 2001; Paabo *et al.*, 2004). The ability to repair DNA molecules ceases when an organism becomes deceased, and as a result, endogenous nucleases cleave the phosphodiester backbone of nucleic acids degrading and fragmenting the DNA molecules (Lindahl, 1993; Hofreiter *et al.*, 2001). Preservation under frozen, dried, or anoxic environments slows and inhibits enzymes activity, reducing DNA damage after death. However, other factors such as hydrolysis and oxidation continue to cause DNA damage.

Hydrolytic damage results in single-strand DNA breaks by direct cleavage or following depurination (Lindahl, 1993; Dabney *et al.*, 2013). Hydrolysis of DNA molecules also gives rise to the deamination of cytosine (C) to uracil (U), and its homologue 5-methylcytosine to thymine (T) (Lindahl, 1993; Dabney *et al.*, 2013). Importantly, deaminated cytosines act as miscoding lesions in many sequencing technologies, resulting in the misincorporation of adenine (A) instead of guanine (G) on the complementary strand, and subsequently T instead of C at the original site in downstream copies (Dabney *et al.*, 2013). Secondly, oxidation modifies the nitrous bases in DNA resulting in the fragmentation of the sugar ring, and this

also limits polymerase activity by either stopping amplification or leading to the production of artefacts or chimeric sequences (Lindahl, 1993; Dabney *et al.*, 2013). Another form of damage that also limits polymerase activity is the formation of crosslinks either between DNA strands by alkylation, or between DNA and other molecules by Maillard reactions (Pääbo, 1989; Poinar *et al.*, 1998; Hansen *et al.*, 2006; Heyn *et al.*, 2010). The cumulative impacts of all these destructive factors means the length of aDNA fragments retrieved with current ancient DNA methods is limited to very short fragments, typically <100bp (Figure 1a) (Slon *et al.*, 2017; Gutaker *et al.*, 2017).

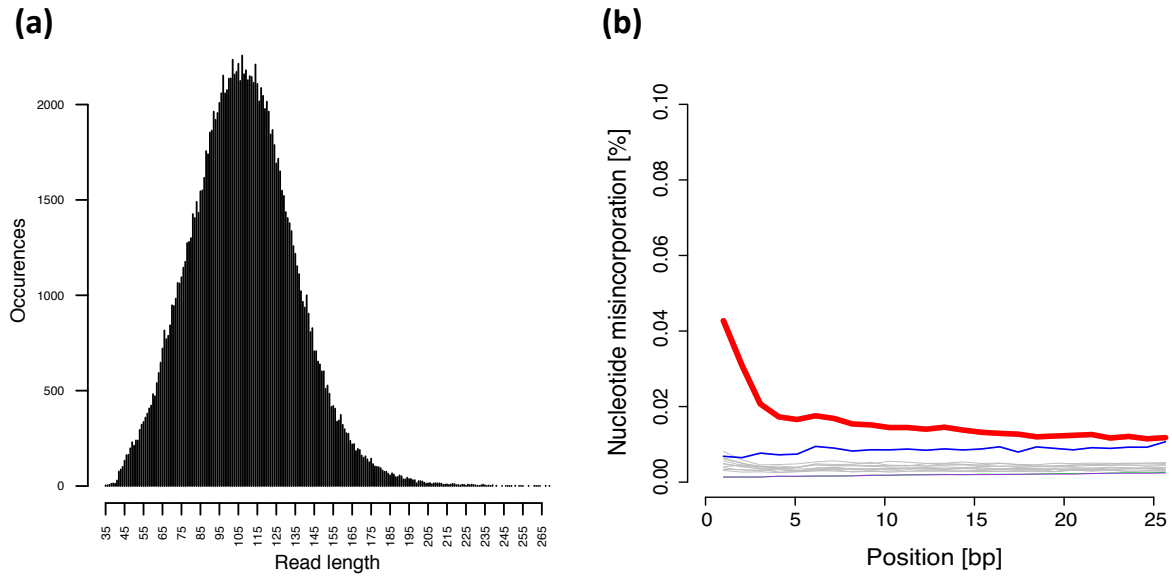


Figure 1. Biochemical characteristics of ancient DNA based on whole-genome shotgun data from a 1,400-year-old specimen of *Chenopodium quinoa*. Data presented in chapter 5. **(a)** Distribution of fragment lengths of sequenced reads. **(b)** Nucleotide misincorporation profile at the 5'-end of sequenced DNA fragments. The red line shows an excess of cytosine to thymine substitutions that increase exponentially towards the end of the molecules. Guanine to adenine substitutions are shown in blue, whereas all other substitutions in grey.

Due to the degraded nature of aDNA, the application of polymerase chain reaction (PCR) and chain-termination sequencing (Sanger *et al.*, 1977) restricted the scope of earlier aDNA studies to the analysis of a small portion of the genetic information present in the best preserved specimens (Paabo *et al.*, 2004). PCR and sanger sequencing approaches proved to have limitations when scaling to multiple loci, in that these methods need large DNA input, require fragments longer than the average length of aDNA (more than 100bp), and PCR primers must be designed based on known sequences (Knapp & Hofreiter, 2010).

The introduction of high-throughput sequencing methods greatly expanded the amount of genetic information that can be recovered from ancient specimens (Der Sarkissian *et al.*, 2015). HTS is defined as a group of technologies that can sequence thousands of short DNA molecules in parallel without the need for sample-specific primers (Margulies et al. 2005). Importantly, HTS technologies require the generation of DNA sequencing libraries, a crucial innovation for the recovery and authentication of aDNA molecules (Briggs & Heyn, 2012; Hofreiter *et al.*, 2014). In shotgun DNA library construction methods, universal oligonucleotide adapters are ligated to DNA fragments and provide binding sites for primers that amplify the entire sequencing library simultaneously (Meyer & Kircher, 2010; Gansauge & Meyer, 2013). DNA fragments as small as 25 nucleotides can be recovered through sequencing libraries, thus increasing the total amount of molecules available for amplification and sequencing (Meyer & Kircher, 2010; Gansauge & Meyer, 2013; Gansauge *et al.*, 2017).

Also, in HTS data an excess of C to T and G to A misincorporations are observed at the end of aDNA fragments, as cytosine deamination occurs in higher proportion in single-stranded overhanging ends (Figure 1b) (Briggs *et al.*, 2007; Brotherton *et al.*, 2007; Dabney *et al.*, 2013). This pattern of miscoding lesions together with the fragment length distribution has been proposed as a means to validate the presence of ancient DNA in a given sample (Ginolhac *et al.*, 2011; Jónsson *et al.*, 2013). In order to avoid large amounts of incorrect nucleotide insertions during sequencing, uracils derived from cytosine deamination can also be removed or repaired during library construction by enzymatic treatment with *E. coli* uracil-DNA-glycosylase and endonuclease VIII (Briggs *et al.*, 2010; Gansauge & Meyer, 2014; Rohland *et al.*, 2015).

Another key aspect of aDNA research is the susceptibility of the samples to contamination with exogenous DNA from organisms that were in contact before and after death (Paabo *et al.*, 2004; Willerslev & Cooper, 2005; Llamas *et al.*, 2017). The non-specific nature of adapter ligation in HTS and the high sensitivity of PCR to any source of DNA make the amplification and sequencing of abundant, or less damaged, contaminating DNA more likely than the low content of damaged endogenous aDNA (Willerslev & Cooper, 2005; Llamas *et al.*, 2017). Contamination complicates and undermines the authenticity of many aDNA sequences from ancient specimens as modern DNA from the environment may easily lead to false positive results (Gilbert *et al.*, 2005; Malmström *et al.*, 2005; Skoglund *et al.*, 2014). In fact, early studies on plant material claiming DNA sequences surviving for millions of years (Golenberg

et al., 1990; Soltis *et al.*, 1992; Poinar *et al.*, 1993; Kim *et al.*, 2004) are all now believed to be PCR artefacts or the result of contamination with modern DNA (Paabo *et al.*, 2004).

Protocols to avoid contamination and validate ancient DNA studies have been adopted in the aDNA field (Cooper & Poinar, 2000; Paabo *et al.*, 2004; Llamas *et al.*, 2017). These protocols include conducting DNA extractions under sterile conditions in facilities isolated from laboratories where PCR is performed, using non-template controls and extraction blanks, applying aDNA-specific bioinformatics pipelines, and cross-facility validation to ensure sample authenticity (Cooper & Poinar, 2000; Paabo *et al.*, 2004; Llamas *et al.*, 2017). The adoption of these minimal standards together with the inclusion of technological innovations have facilitated the authentication and publication of remarkable aDNA studies such as the genomes of the extinct mammoth (Miller *et al.*, 2008), the Neanderthal (Green *et al.*, 2010), and the Denisovan (Meyer *et al.*, 2012).

Ancient plant DNA studies

While animal and human aDNA studies have seen rapid progress in the last three decades, plant aDNA research continues to grow at a much lower rate (Gugerli *et al.*, 2005; Schlumbaum *et al.*, 2008; Palmer *et al.*, 2012b; Parducci *et al.*, 2017). In a recent review, Parducci *et al.* (2017) reported that from 2,104 aDNA-papers published until mid-2016 only 5% are focused on plants. Agricultural species have received most of the attention except for a small number of wild plants (Gugerli *et al.*, 2005; Schlumbaum *et al.*, 2008; Palmer *et al.*, 2012b; Brown *et al.*, 2015). One key area of research has been the evolution and domestication of plants, in which technologies like PCR have been used to identify the taxonomic status of archaeobotanical samples and to address questions such as the origin and spread of agricultural species such as wheat (Allaby *et al.*, 1999; Li *et al.*, 2011; Oliveira *et al.*, 2012; Fernández *et al.*, 2013; Bilgic *et al.*, 2016), maize (Freitas *et al.*, 2003; Jaenicke-Després *et al.*, 2003; Lia *et al.*, 2007), barley (Palmer *et al.*, 2009; Fernández *et al.*, 2013), broomcorn millet (Li *et al.*, 2016), grape (Manen *et al.*, 2003; Cappellini *et al.*, 2010), bottle gourd (Erickson *et al.*, 2005), chenopod (Kistler & Shapiro, 2011), and olives (Elbaum *et al.*, 2006).

The use of conventional PCR and chain-termination sequencing (Sanger *et al.*, 1977) limited the focus of plant aDNA research to a range of loci previously annotated in modern species. Regions of chloroplast and nuclear-ribosomal DNA loci have been widely used since both occur in multiple copies in a plant cell, and therefore, are likely to be recovered from an ancient specimen (Schlumbaum *et al.*, 2008; Palmer *et al.*, 2012b). In contrast to fungal and animal studies, plant mitochondrial DNA, which also often exists in high copy number, is rarely employed because mitochondrial genes are very conserved across plant species and plant mitochondrial genomes are massive and vary significantly in size between closely related species (0.2-11 Mb) (Palmer *et al.*, 2000; Gualberto & Newton, 2017). Additionally, plant mitochondrial genomes are largely constituted of intergenic regions, which contain a high proportion of repetitive sequences and frequently incorporate insertions from chloroplastic or nuclear DNA (Palmer *et al.*, 2000; Gualberto & Newton, 2017). Nuclear DNA, with only a few copies per plant cell, contains most of the evolutionary information and thus has been used to examine variation in functional genes (Allaby *et al.*, 1999; Freitas *et al.*, 2003; Jaenicke-Després *et al.*, 2003; Lia *et al.*, 2007).

The adoption of sequencing by high throughput methods and hybridisation capture significantly expanded the studies possible with ancient DNA, enabling the analysis of whole ancient genomes (Knapp & Hofreiter, 2010; Hofreiter *et al.*, 2014; Brown *et al.*, 2015; Der Sarkissian *et al.*, 2015). Hybridisation capture is a method that allows simultaneous enrichment of multiple targets from complex DNA extracts by removing non-target molecules (Hodges *et al.*, 2007; Gnirke *et al.*, 2009). To date, few published aDNA studies have utilised HTS and hybridisation capture in ancient botanical remains (Table 1). However, significant advances in recent plant aDNA research have been made; for instance, chloroplast (Kistler *et al.*, 2015; Wales *et al.*, 2016, 2018; Pérez-Zamorano *et al.*, 2017), mitochondrial (Pérez-Zamorano *et al.*, 2017; Wales *et al.*, 2018), and nuclear genomes (Mascher *et al.*, 2016; Vallebueno-Estrada *et al.*, 2016; Ramos-Madrigal *et al.*, 2016b; Swarts *et al.*, 2017) have been sequenced from archaeological plant samples dating back to the early-mid Holocene epoch (Table 1).

Table 1. List of published ancient plant DNA studies of domesticated species using high-throughput sequencing until March 2019. Age is provided in calibrated years before present (BP). N/A indicates no information provided in the publication.

Publication	Species	Tissue	Preservation	Age (BP)	Genomic target
Ávila-Arcos *et al.,* (2011)	*Zea mays* L.	Kernels Cobs	Desiccated	700 – 1,400	Chloroplast Domestication genes
Bunnings *et al.,* (2012)	*Triticum monococcum* L. *Triticum turgidum* L. *Hordeum vulgare* L. *Panicum miliaceum* L.	Seeds	Charred	3,300	Nuclear
Palmer *et al.,* (2012)	*Gossypium barbadense* L. *Gossypium herbaceum* L.	Seeds	Desiccated	750 – 1,600.	Transposable elements.
Wales *et al.,* (2012)	*Vitis vinifera* L. *Zea mays* L.	Seeds Branch Kernels Cobs	Desiccated	700 – 1,400	Chloroplast Nuclear
Fordyce *et al.,* (2013)	*Zea mays* L.	Kernels	Desiccated	700	RNA
Wales *et al.,* (2014)	*Vitis vinifera* L. *Zea mays* L. *Cornus mas* L. *Gossypium* spp. *Olea europaea* L.	Seed Branch Kernels	Desiccated Waterlogged	400 – 2,400	*rbcL* gene
Smith *et al.,* (2014)	*Hordeum vulgare* L.	Seeds	Desiccated	750	Viral RNA
Smith *et al.,* (2014)	*Hordeum vulgare* L.	Seeds	Desiccated	200 – 800	Methylated DNA RNA
Kistler *et al.,* (2014)	*Lagenaria siceraria* Standl.	Rind	Desiccated Waterlogged	100 – 9,000	Chloroplast
Wales *et al.,* (2015)	*Vitis vinifera* L. *Zea mays* L.	Seeds Kernels Cobs	Waterlogged Desiccated	1,700 – 4,500	Nuclear
Kistler *et al.,* (2015)	*Cucurbita* spp.	Rind Seed Peduncle	N/A	N/A	Chloroplast

Da Fonseca *et al.,* (2015)	*Zea mays* L.	Cobs	Desiccated	670 – 6,000	Exome
Mascher *et al.,* (2016)	*Hordeum vulgare* L.	Seeds	Desiccated	6,000	Exome Nuclear
Ramos-Madrigal *et al.,* (2016)	*Zea mays* L.	Cobs	Desiccated	5,310	Exome
Wales *et al.,* (2016)	*Vitis vinifera* L.	Seeds	Desiccated Waterlogged	500 – 4,000	Chloroplast
Vallebueno-Estrada *et al.,* (2016)	*Zea mays* L.	Basal stalk Cobs Leaf	Desiccated	1,530 – 5,300	Nuclear
Nistelberger *et al.,* (2016)	*Vitis Vinifera* L. *Hordeum vulgare* L. *Zea mays* L. *Oryza sativa* L	Seeds Kernels Cobs	Charred	550 – 4,450	Nuclear
Perez-Zambrano *et al.,* (2017)	*Zea mays* L.	Cobs	Desiccated	5,300	Chloroplast Mitochondria
Smith *et al.,* (2017)	*Hordeum vulgare* L.	Seeds	Desiccated	600 – 900	RNA
Swarts *et al.,* (2017)	*Zea mays* L.	Cobs	Desiccated	1,900	Nuclear
Wales *et al.,* (2018)	*Helianthus annuus* L.	Disk Pericarp Kernel	Desiccated	790 – 3,000	Chloroplast Mitochondria
Kistler *et al.,* (2018)	*Zea mays* L.	N/A	N/A	100 – 1,010	Nuclear
Wagner *et al.,* (2018)	*Querqus* spp.	Wood	Waterlogged	500 – 9,800	Chloroplast Nuclear

One of the most important findings of research into ancient plant remains has been the recovery of RNA sequences from ancient barley and maize seeds (Fordyce *et al.*, 2013; Smith *et al.*, 2017). In seeds, several RNA molecules are synthesised to protect the embryo, guarantee dormancy over time and promote germination when the environmental conditions are favourable (Nakabayashi *et al.*, 2005). Based on this feature, Fordyce and collaborators (2013) applied HTS to complementary DNA (cDNA) synthesised from RNA extracted from a 750-year-old ancient maize kernel, proving that not only aDNA but also ancient RNA molecules can be isolated from specimens with extraordinary preservation. Subsequent plant ancient DNA studies also retrieved interfering RNAs (siRNAs) and microRNAs (miRNAs) from desiccated barley grains (1,100-1,400-years-old) and sequenced the first complete ancient RNA genome from a plant pathogen; the Barley Stripe Mosaic Virus isolated from a 750-year-old barley grain from Africa (Smith *et al.*, 2014a,b, 2017). Also, the potato blight fungus *Phytophthora infestans*, the cause of the Irish potato famine the 19th century, has been sequenced from foliar tissues of herbarium specimens (Martin *et al.*, 2013; Yoshida *et al.*, 2013).

Technological developments in ancient DNA research have also been implemented to characterise patterns of methylation as a means to access ancient epigenetic signals. Epigenetics is defined as "an inherited change in a phenotype that is not solely due to a change in DNA sequences" (Springer & Schmitz, 2017), and has been a critical strategy for plant species to respond to environmental stimuli and adapt to biotic and abiotic stresses (He *et al.*, 2011; Diez *et al.*, 2014). Recent ancient DNA publications have provided evidence that pre-mortem epigenetic signals such as methylation of cytosines can be reconstructed from aDNA sequences by methyl-enrichment (Seguin-Orlando *et al.*, 2015), bisulphite sequencing (Llamas *et al.*, 2012) and from the distribution of DNA nucleotide misincorporations in aDNA sequencing data (Gokhman *et al.*, 2014). Bisulphite sequencing and methyl-enrichment methods have even been applied to a set of virus-infected archaeological barley seeds spanning ~2,600 years, which revealed epigenetic responses to the ancient viral infection (Smith *et al.*, 2014b).

One important aspect hindering the genetic analysis of plant domestication is the availability of archaeobotanical remains suitable for ancient DNA research. Most of the plant domestication processes were carried out in temperate and tropical latitudes where the fluctuation of environmental conditions such as temperature and humidity do not favour the long-term preservation of nucleic acids in plant specimens (Palmer *et al.*, 2009; Allaby *et al.*,

2015). Samples preserved under these conditions could have undergone hydrolytic and oxidative damage as well as Maillard reactions, which can produce cross-links between proteins and DNA, and thus, making endogenous DNA inaccessible to current methods of DNA extraction (Pääbo, 1989; Poinar *et al.*, 1998; Hansen *et al.*, 2006; Heyn *et al.*, 2010). Also, archaeological plant materials are frequently found in a charred state (Hather, 1991; Zohary *et al.*, 2013). Charred preservation is biased towards hard tissues such as seeds, chaff, branches, and wood that come in contact with fire through human activities such as cooking, crop management or fuel use (Zohary *et al.*, 2013; Pearsall, 2015). The exposure of plant specimens to high temperatures turns carbohydrates and proteins into high-molecular-weight melanoidins and other organic compounds into charcoal (Styring *et al.*, 2013; Nitsch *et al.*, 2015). Such compounds are not easily degraded by saprophytic organisms, which favours the preservation of charred samples under different environmental conditions (Styring *et al.*, 2013; Nitsch *et al.*, 2015). However, the extensive molecular damage caused by high temperatures makes the preservation and recovery of nucleic acids far less likely (Nistelberger *et al.*, 2016). Despite such damage, some plant aDNA studies have used charred material - reporting contrasting results (Palmer *et al.*, 2012b). For instance, in PCR-based studies, while some reported successful retrieval and amplification of DNA from charred samples (Fernández *et al.*, 2013; Smýkal *et al.*, 2014; Mikić, 2015; Bilgic *et al.*, 2016), others reported a failure to amplify and authenticate endogenous DNA molecules (Brown *et al.*, 1998; Blatter *et al.*, 2002; Oliveira *et al.*, 2012).

To assess the ability of HTS methods to recover endogenous DNA molecules from charred material, Bunning *et al.* (2012) applied this technology to a 3,300-year-old sample of charred cereal grains. The study reported the recovery of 0.008% of reads mapping to the target genomes from a panel of 21 million sequences. Even though the number of endogenous reads was low, the authors suggested that this was a promising result and that the enrichment by capture methods was likely to improve the endogenous content. Conversely, Nistelberger and collaborators (2016) reanalysed the sequences reported by Bunning *et al.* (2012) and concluded that the endogenous reads were likely to be contamination or mis-assigned due to non-stringent analyses. Nistelberger *et al.* (2016) also reported failure to recover informative ancient DNA sequences from 38 charred specimens of four plant species (barley, grape, maize, and rice), using enrichment by hybridisation capture and HTS. These contradictory results and the low replicability of experiments question the utility of charred material for aDNA research. These results emphasise that for successful ancient DNA studies in plants, it is imperative that the source of the material is of high quality.

Sources of ancient plant DNA

Ancient plant DNA molecules have been reported from different sources including cereal seeds, maize cobs, bottle gourd rinds, foliar tissues, wood, and even from sediments, ice cores, coprolites and dental calculus (Gugerli *et al.*, 2005; Palmer *et al.*, 2012b; Weyrich *et al.*, 2017; Parducci *et al.*, 2017). Archaeological collections of maize cobs and kernels have proven to be the most successful source of plant aDNA molecules from domesticated species, and form the basis of the pioneering studies of plant aDNA that used HTS and hybridisation capture as well as the recovery of ancient RNA molecules (Table 1) (Ávila-Arcos *et al.*, 2011; Fordyce *et al.*, 2013). Studies on desiccated maize samples from Mexico and North and South America spanning 6,000 years have yielded several ancient chloroplast, mitochondrial and nuclear genomes (Ávila-Arcos *et al.*, 2011; da Fonseca *et al.*, 2015; Vallebueno-Estrada *et al.*, 2016; Ramos-Madrigal *et al.*, 2016b; Swarts *et al.*, 2017; Pérez-Zamorano *et al.*, 2017; Kistler *et al.*, 2018). Similarly, barley seeds from the Judean desert in Israel have provided the oldest barley nuclear genome (6,000-years-old) yet sequenced (Mascher *et al.*, 2016), and archaeological barley seeds from Egypt have allowed the generation of ancient RNA sequences and epigenetic profiles (Table 1) (Smith *et al.*, 2014a,b, 2017). Other ancient plant that have been reported to yield aDNA using HTS include bottle gourd rinds, sunflower disk fragments, foliar tissues, pollen and wood (Table 1) (Kistler *et al.*, 2014, 2015; Vallebueno-Estrada *et al.*, 2016; Parducci *et al.*, 2017; Wales *et al.*, 2018; Wagner *et al.*, 2018). The potential yield of aDNA molecules using HTS from other botanical subfossils such as phytoliths are subject to investigation.

Biological collections such as those held by museums of natural history and herbariums are a well-recognised source of ancient plant DNA. Such collections are reservoirs of a large number of plant specimens dating back to the 17th century and represent species and populations that can provide valuable information about the recent history of plant evolution and domestication (Wandeler *et al.*, 2007; Palmer *et al.*, 2012b; Bakker, 2017). The use of HTS methods on herbarium specimens has allowed the estimation of plant DNA decay rates (Staats *et al.*, 2011; Weiß *et al.*, 2016) and the recovery of historical plant pathogen sequences (Martin *et al.*, 2013; Yoshida *et al.*, 2013; Smith *et al.*, 2014a). A detailed review of the application of HTS methods on herbarium and museum specimens is provided by Bakker and collaborators (2017).

The preservation of DNA in ancient plant specimens is influenced by a variety of factors such as environmental conditions, depositional contexts, and taphonomic processes. Most ancient plant DNA sequences obtained using high-throughput methods have been retrieved from plant

remains found under desiccated and waterlogged conditions (Table 1). Desiccated remains come from extremely dry and arid environments such as cave sites or desert areas, and often exhibit remarkable preservation. Most of the successful plant aDNA studies have made use of dried specimens, with investigations reporting the recovery of up to 96% endogenous molecules (Mascher *et al.*, 2016). Likewise, the anoxic conditions in peat bogs, deep sediments of lakes and seas, or even in wells and latrines allow good preservation of biological remains. Waterlogged plant samples collected from these environments have also been a good source of aDNA sequences (Kistler *et al.*, 2014; Wales *et al.*, 2014, 2015, 2016; Wagner *et al.*, 2018).

The quality and preservation of botanical remains also depend on a variety of cultural and methodological factors. Cultural practices associated with historical human activities such as charring plant material during handling, seed storage in sealed containers, plant processing for medicinal and food use, and plant offerings in burials directly impact the probability of preservation of plant remains (Zohary *et al.*, 2013; Pearsall, 2015). In the same way, methodological approaches used during sample collection at paleontological and archaeological sites can affect sample quality and preservation. For example, in collection methods such as water flotation or wet sieving, samples are separated from the soil by adding water and then dried in the shade or by heating (Pearsall, 2015). The immersion in water and fluctuation in temperature may induce hydrolytic and oxidative damage to the DNA molecules and also increase the risk of contamination and degradation by microorganisms (Lindahl, 1993; Dabney *et al.*, 2013). Therefore, recovery by dry sieving, despite being laborious and time-consuming, is the most amenable method to avoid additional damage (Pearsall, 2015). Equally important, practices of curation and storage of specimens in biological collections play a key role in the utility of a sample for further molecular analysis. Samples should be stored in dry and cold conditions, and avoid treatments of specimens with substances that can cause molecular damage such as formalin; a fixative often used to preserve herbarium and museum specimens (Wandeler *et al.*, 2007; Staats *et al.*, 2011).

The complex conditions in which plant aDNA is preserved, together with the bias of plant aDNA studies towards cultivated species, have limited studies to the time-frame and locations where agriculture took place. Nevertheless, the lack of availability of well-preserved samples is also influenced by the fast rate of DNA degradation in plant specimens. A recent study on herbarium specimens spanning 300 years, estimated a per nucleotide decay rate of 1.66×10^{-4} per year (Weiß *et al.*, 2016), which is six times faster than the rate estimated for ancient bones (Allentoft *et al.*, 2012). However, more research is necessary to understand the nature of

degradation of nucleic acids in plant remains across different times and preservation conditions. This knowledge is fundamental to define effective technical approaches for the recovery and molecular analysis of ancient plant DNA and the exploration of new sources.

Molecular biology of ancient plant DNA

Most of the current methods applied in aDNA studies have been established and optimised for animal tissues, especially for bones and teeth (Shapiro & Hofreiter, 2012). However, the molecular structure and biochemical makeup of plant tissues are significantly different to animal tissues, and therefore modifications to standard procedures are needed to optimise yield from a wide range of ancient plant remains preserved in different conditions. Optimisation of protocols for ancient plant specimens is particularly challenging not only because of the rapid degradation of plant material but also because of the presence of secondary compounds which can inhibit processes used in molecular biology reactions (Bessetti, 2007; Samarakoon *et al.*, 2013).

Plant tissues are characterised to be rich in polyphenols, proteins, and polysaccharides (Varma *et al.*, 2007). Ancient samples additionally contain humic acids and salts produced by the organic decomposition or leakage from associated sediments (Kistler, 2012; Wales *et al.*, 2014). Proteins and polyphenols can form crosslinks with nucleic acids during plant degradation and early steps of DNA extraction, affecting the quality and quantity of DNA isolated (Varma *et al.*, 2007; Schrader *et al.*, 2012). Most crosslinked macromolecules are washed away during purification steps – simply reducing nucleic acid yields; however, a fraction of crosslinked macromolecules, polyphenols, polysaccharides, and humic acids can be coextracted with nucleic acids and inhibit downstream enzymatic reactions. The presence of crosslinked macromolecules has been reported to inhibit polymerase activity, interfere with absorbance-based quantification methods, and cause fluorescence inhibition in real-time PCR (Kistler, 2012; Schrader *et al.*, 2012; Sidstedt *et al.*, 2015). Currently, it is unknown if there are potential effect of secondary compounds on other critical steps of HTS such as library preparation and enrichment methods. However, the presence of these molecules is likely to interfere with the activity of key enzymes such as ligases and endonucleases.

Several methods have been implemented in modern plant DNA extraction protocols to reduce contamination by inhibitor molecules, all of which are suitable for the extraction of ancient plant remains. DNA extraction protocols based on surfactants have proven to yield aDNA suitable with HTS methods. SDS (Sodium Dodecyl Sulphate), an anionic surfactant commonly included in ancient plant DNA extraction buffers, is known to disrupt cell walls and to denature proteins - helping to release the DNA from histones and other DNA binding proteins (Dellaporta *et al.*, 1983). Other protocols incorporate the cationic surfactant CTAB (Cetyl Trimethylammonium Bromide), which is effective at isolating DNA from tissues rich in polysaccharides (Doyle & Doyle, 1987). The combination of these surfactants with other reagents is often needed to remove polyphenols and humic acids. Flocculants, such as magnesium chloride, ferric chloride, calcium chloride, and aluminium ammonium sulphate are employed to cause aggregation and precipitation of inhibitor molecules (Braid *et al.*, 2003; Sharma *et al.*, 2014). Likewise, PVPP (Polyvinylpolypyrrolidone) and its soluble counterpart PVP (Polyvinylpyrrolidone), are used together with antioxidants such as β-mercaptoethanol and DTT (Dithiothreitol) to remove polyphenols and humic acids by agglomeration (John, 1992; Kim *et al.*, 1997).

Commercial kits, specifically designed for isolation and purification of plant DNA, have also been used in combination with additional reagents to purify plant ancient DNA (Kistler, 2012). PTB (N-phenacylthiazolium bromide), a compound known to disrupt covalent crosslinks between carbohydrates and proteins (Vasan *et al.*, 1996), has been incorporated into standard lysis buffers applied in plant aDNA studies (Kistler, 2012). However, while PTB has increased the yield of aDNA in coprolite extractions (Poinar *et al.*, 1998; Hofreiter *et al.*, 2000), no advantages or disadvantages have been observed in ancient bones and teeth (Rohland & Hofreiter, 2007). Further investigations to improve current methods of plant aDNA isolation, especially to overcome crosslink formations, seem crucial for ancient plant DNA research.

Sample preparation for HTS requires the construction of DNA sequencing libraries, a critical step in the recovery of aDNA where endogenous molecules can be lost or over-amplified. Novel and improved methods have been proposed in ancient DNA research to assemble DNA sequencing libraries from either double-stranded or single-stranded templates. In the double-stranded library preparation, two different adapters are ligated to enzymatically blunt end-repaired double-stranded DNA templates (Margulies *et al.*, 2005; Meyer & Kircher, 2010), or a single Y-shaped adapter with a T-overhang is ligated to both ends of DNA fragments previously modified with A-overhangs (Bentley *et al.*, 2008; Gansauge *et al.*, 2017). The

single-stranded library method first dephosphorylates and denatures DNA fragments, and then single-stranded DNA is ligated to a first adapter and immobilised on streptavidin-coated magnetic beads (Gansauge *et al.*, 2017). Subsequent enzymatic reactions to copy DNA templates, create blunt ends, and ligate a second adapter are executed with the DNA attached to the magnetic beads (Gansauge *et al.*, 2017).

The single-stranded library method reduces the amount of DNA molecules washed away during purification steps, while increasing the yield of aDNA molecules in sequencing libraries (Gansauge & Meyer, 2013; Gansauge *et al.*, 2017). Furthermore, DNA molecules with single-strand breaks that are lost in double-stranded library approaches can still be converted into single-stranded sequencing libraries, thus increasing, even more, the information retrieved from ancient specimens (Gansauge & Meyer, 2013; Gansauge *et al.*, 2017). Both single and double-stranded DNA library preparation methods have been successfully applied across a range of aDNA studies (Orlando *et al.*, 2015), with the latter being recommended for highly degraded specimens as it has shown to perform better on degraded bone and plant materials (Gansauge & Meyer, 2013; Wales *et al.*, 2015; Gansauge *et al.*, 2017).

Shotgun sequencing has been used in several plant aDNA studies, including the sequencing of complete genomes from well-preserved samples of maize and barley (Table 1) (Mascher *et al.*, 2016; Ramos-Madrigal *et al.*, 2016a). Shotgun sequencing also provides a means to assess the levels of contamination, the proportion of endogenous molecules, and the quality of a sequencing library. However, shotgun sequencing can be inefficient in samples with low endogenous DNA or when specific genomic loci are targets (Knapp & Hofreiter, 2010; Hofreiter *et al.*, 2014). In such cases, abundant non-target DNA fragments will be sequenced preferentially over less abundant target molecules (Knapp & Hofreiter, 2010).

An alternative to direct shotgun sequencing is to reduce the genomic complexity of the DNA library by targeting specific molecules of interest for selective enrichment (Cronn *et al.*, 2012). Conventional PCR is the most straightforward approach to enrich key genomic targets, but selective enrichment by hybridisation capture has been shown to be highly efficient at enriching thousands of targets in parallel and is now optimised for aDNA studies (Orlando *et al.*, 2015). In hybridisation capture, DNA targets are hybridised with specifically designed probes that are immobilized on microarrays or capture beads, unbound non-target molecules are washed away, and the remaining target library molecules are enriched with PCR prior to sequencing (Hodges *et al.*, 2007; Gnirke *et al.*, 2009; Carpenter *et al.*, 2013). In plant aDNA

research, the application of hybridisation capture has allowed the sequencing of short nuclear loci, genome-wide SNPs, exomes, and complete chloroplasts, mitochondrial, and nuclear genomes (Table 1).

Hybridisation capture is a flexible technology with many variables, such as probe design, hybridisation temperature, and buffer composition that can be adjusted to the desired outcome of a study. For example, the complexity of a sequencing library can be reduced by removing unwanted sequences with targeted probes instead of enriching specific loci (Cronn *et al.*, 2012). For example, repetitive elements, that may account for up to 90% of the genome size in plants (Mehrotra & Goyal, 2014), can be removed from a DNA library using a two-step or "subtractive" hybridisation capture protocol, where the first set of probes sequesters repetitive elements from the library and the second enrich for target loci (Fu *et al.*, 2010).

Plant genomes present additional obstacles such as large genome size, high levels of repetitive elements, extensive variation in the organelle genome size, and diverse levels of ploidy that can hinder DNA retrieval and the sequencing process (Egan *et al.*, 2012; Cronn *et al.*, 2012; Der Sarkissian *et al.*, 2015). Although there are other strategies available to reduce the complexity of plant genomic extracts such as C0t and methylation-filtration, none of them have been successfully applied to ancient botanical remains.

C0t-filtration uses the rapid re-association kinetics of repetitive elements to enrich single and low copy elements from the sequencing library (Peterson *et al.*, 2002; Lamoureux *et al.*, 2005). In principle, the rate at which a specific sequence anneals after heat denaturation is proportional to the number of times it occurs in the genome (Peterson *et al.*, 2002). Therefore, in C0t-filtration, genomic DNA is heated to denaturation and allowed to re-anneal to a given C0t value (C0t value = the product of nucleotide concentration in moles per litre (C0) and reassociation time in seconds (t)), at which point most of the low copy DNA remains single-stranded, repetitive elements are filtered using hydroxyapatite columns, and the single-stranded fraction is used to create target-enriched libraries (Peterson *et al.*, 2002; Lamoureux *et al.*, 2005; Guerrero *et al.*, 2010; Barrero *et al.*, 2016).

Methylation-filtration relies on the assumption that DNA methylation in plants occurs more often in repetitive elements than in genes (Rabinowicz *et al.*, 2003). A methylation-sensitive restriction endonuclease can be used to digest methylated regions of the genome while leaving lower copy regions (Whitelaw, 2003; Palmer *et al.*, 2003; Bedell *et al.*, 2005; Grativol *et al.*,

2014). Sections of the genomes of wheat (Lamoureux *et al.*, 2005), maize (Palmer *et al.*, 2003), sorghum (Peterson *et al.*, 2002; Bedell *et al.*, 2005), and sugarcane (Grativol *et al.*, 2014) have been sequenced using combined C0t, methylation filtration and sequencing techniques. The inclusion of such strategies in future methodological improvements is likely to have a positive impact on the reduction of genomic complexity of plant aDNA libraries, where desired.

Concluding remarks and future directions

High-throughput technologies have markedly improved our understanding of the characteristics and limitations of aDNA studies, providing insights into the best approaches for the retrieval and authentication of genetic information from ancient specimens. The introduction of HTS has also triggered a significant change in the scope of ancient plant DNA research, moving from the analysis of a few loci to the generation of ancient genomes, transcriptomes and epigenomes (Fordyce *et al.*, 2013; Smith *et al.*, 2014b, 2017). This expansion of the plant aDNA field has provided essential information to answer evolutionary questions such as the origin and adaptation of cultivated plants. Notably, the recent technological advances open the possibility to investigate the functional variability of the genome at key developmental stages and may provide insights into how plants have adapted to past environments (Orlando *et al.*, 2015; Der Sarkissian *et al.*, 2015).

However, a better understanding of how and why plant DNA decay occurs is crucial for the exploration of new aDNA sources and effective technical improvements in extraction procedures. The increased resolution that HTS data provides has permitted the critical evaluation of plant remains preserved under different conditions (Wales *et al.*, 2014; Nistelberger *et al.*, 2016). Targeted enrichment methods, such as hybridisation capture are helping to reduce the genomic complexity of plant aDNA libraries and allow the investigation of a wider variety of samples (Egan *et al.*, 2012; Cronn *et al.*, 2012; Der Sarkissian *et al.*, 2015).

Given the importance of plants to food security, investigating the genetic makeup of ancient crops promises to not only facilitate our understanding of the evolution of cultivated plants but also provide valuable information for crop breeding (Der Sarkissian *et al.*, 2015). Domestication, as an intensive selection process has changed the genomes of domesticated species to reduce undesirable traits while increasing desirable traits. However, this intensive selection process has also caused loss of non-target traits and reduction of the genetic diversity

of domesticated species, which could threaten the continued ability to improve cultivated plants (Doebley *et al.*, 2006; Bevan *et al.*, 2017). Ancient plant DNA studies may help to identify the genetic changes that have occurred at various stages of domestication, potentially enabling the identification of lost genetic diversity at key loci (Orlando *et al.*, 2015; Der Sarkissian *et al.*, 2015; Estrada *et al.*, 2018). Modern breeding technologies such as genetic engineering, mutant screening, or targeted gene editing can be applied to test the functionality of the ancient diversity and evaluate the potential reintroduction of extinct variants of breeding value.

The spread of crops from the centres of domestication required adaptation to new environments, supported by favourable alleles at critical genetic loci (Larson *et al.*, 2014). The recovery of RNA molecules from ancient plant specimens (Fordyce *et al.*, 2013; Smith *et al.*, 2014a, 2017) now makes it possible to reconstruct ancient transcriptomes and observe evolutionary changes in gene expression. Ancient plant transcriptomes would allow characterisation of the functional variability of the genome at key developmental stages and may provide insights into how plants have responded to past environments (Fordyce *et al.*, 2013; Der Sarkissian *et al.*, 2015). Equally important, recent aDNA research has detected epigenetic signals such as cytosine methylation levels in ancient plant specimens (Smith *et al.*, 2014b). There is growing evidence that epigenetic mechanisms significantly contribute to plant responses to abiotic and biotic stresses (Springer & Schmitz, 2017). Consequently, in addition to the genome, studying ancient plant transcriptomes and epigenomes promise to significantly increase the understanding of plant adaptive mechanisms over time (Fordyce *et al.*, 2013).

Humans have impacted the evolutionary history of domesticated plants, but the cultivation of plants has also had a major role in human history. Plant domestication triggered the human transition from a hunter-gathering lifestyle to a sedentary and agriculture-based society, starting around 12,000 years ago (Diamond, 2002). In this regard, the analysis of ancient plant domesticates can also shed light on the recent history of human behaviour and evolution (Marciniak & Perry, 2017) and past human diets, nutrition, and health (Schlumbaum *et al.*, 2008; Marciniak & Perry, 2017). Also, the evolutionary origins and spread of domesticated plants can be used as a proxy to reconstruct past agriculture practices, economies, migrations, and trade (Schlumbaum *et al.*, 2008; Marciniak & Perry, 2017).

Plant aDNA research also has a significant role to play in studying how non-domesticated species have evolved and adapted to environmental changes through time. The application of HTS on permafrost soil, ice-cores, and sediments samples has proven to be a powerful tool for the identification of ancient plant material in environmental samples, providing critical information for the reconstruction of palaeoenvironments (Rawlence *et al.*, 2014; Hofman *et al.*, 2015; Parducci *et al.*, 2017). By investigating the genetic variability of ancient and modern plants, the historical distribution of species, population dynamics, hybridisations, genetic bottlenecks, and extinctions can be examined in greater depth (Paabo *et al.*, 2004; Orlando & Cooper, 2014). Perhaps the most significant contribution of plant ancient DNA to plant evolution and conservation is the identification of ancient variability associated with plant responses to past climate changes and other environmental pressures (Orlando & Cooper, 2014; Rawlence *et al.*, 2014; Hofman *et al.*, 2015). The identification of functional variability in past natural population unlocks the possibility to employ ancient variants for the genetic rescue of modern endangered populations and to select plant populations with better adaptability to future environmental changes.

References

Allaby RG, Banerjee M, Brown T a. **1999**. Evolution of the high molecular weight glutenin loci of the A, B, D, and G genomes of wheat. *Genome* **42**: 296–307.

Allaby RG, Gutaker R, Clarke AC, Pearson N, Ware R, Palmer SA, Kitchen JL, Smith O. **2015**. Using archaeogenomic and computational approaches to unravel the history of local adaptation in crops. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **370**: 20130377.

Allentoft ME, Collins M, Harker D, Haile J, Oskam CL, Hale ML, Campos PF, Samaniego JA, Gilbert MTP, Willerslev E, *et al.* 2012. The half-life of DNA in bone: measuring decay kinetics in 158 dated fossils. *Proceedings. Biological sciences / The Royal Society* **279**: 4724–33.

Ávila-Arcos MC, Cappellini E, Romero-Navarro JA, Wales N, Moreno-Mayar JV, Rasmussen M, Fordyce SL, Montiel R, Vielle-Calzada J-P, Willerslev E, *et al.* 2011. Application and comparison of large-scale solution-based DNA capture-enrichment methods on ancient DNA. *Scientific Reports* **1**: 74.

Bakker FT. **2017**. Herbarium genomics: skimming and plastomics from archival specimens. *Webbia* **7792**: 1–11.

Barrero RA, Guerrero FD, Black M, McCooke J, Chapman B, Schilkey F, Pérez de León AA, Miller RJ, Bruns S, Dobry J, *et al.* 2016. Gene-enriched draft genome of the cattle tick Rhipicephalus microplus: Assembly by the hybrid Pacific Biosciences/Illumina approach enabled analysis of the highly repetitive genome. *International Journal for Parasitology* **47**: 569–583.

Bedell JA, Budiman MA, Nunberg A, Citek RW, Robbins D, Jones J, Flick E, Rohlfing T, Fries J, Bradford K, *et al.* 2005. Sorghum genome sequencing by methylation filtration. *PLoS Biology* **3**.

Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, Hall KP, Evers DJ, Barnes CL, Bignell HR, *et al.* 2008. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **456**: 53–59.

Bessetti J. **2007**. An Introduction to PCR Inhibitors. *Promega Notes*: 9–10.

Bevan MW, Uauy C, Wulff BBH, Zhou J, Krasileva K, Clark MD. **2017**. Genomic innovation for crop improvement. *Nature* **543**: 346–354.

Bilgic H, Hakki EE, Pandey A, Khan MK, Akkaya MS. **2016**. Ancient DNA from 8400 Year-Old Çatalhöyük Wheat: Implications for the Origin of Neolithic Agriculture. *Plos One* **11**: e0151974.

Blatter RHE, Jacomet S, Schlumbaum A. **2002**. Little Evidence for the Preservation of a Single--Copy Gene in Charred Archaeological Wheat. *Ancient Biomolecules* **4**: 65.

Braid MD, Daniels LM, Kitts CL. **2003**. Removal of PCR inhibitors from soil DNA by chemical flocculation. *Journal of Microbiological Methods* **52**: 389–393.

Briggs AW, Heyn P. **2012**. Preparation of Next-Generation Sequencing Libraries from Damaged DNA. In: Shapiro B, Hofreiter M, eds. Ancient DNA: Methods and Protocols. Totowa, NJ: Humana Press, 143–154.

Briggs AW, Stenzel U, Johnson PLF, Green RE, Kelso J, Prufer K, Meyer M, Krause J, Ronan MT, Lachmann M, *et al.* 2007. Patterns of damage in genomic DNA sequences from a Neandertal. *Proceedings of the National Academy of Sciences* **104**: 14616–14621.

Briggs AW, Stenzel U, Meyer M, Krause J, Kircher M, Pääbo S. **2010**. Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA. *Nucleic Acids Research* **38**: e87–e87.

Brotherton P, Endicott P, Sanchez JJ, Beaumont M, Barnett R, Austin J, Cooper A. **2007**. Novel high-resolution characterization of ancient DNA reveals C > U-type base modification events as the sole cause of post mortem miscoding lesions. *Nucleic Acids Research* **35**: 5717–5728.

Brown T a, Allaby RG, Sallares R, Jones G. **1998**. Ancient DNA in charred wheats:

Taxonomic identification of mixed and single grains. *Ancient Biomolecules* **2**: 185–193.

Brown TA, Cappellini E, Kistler L, Lister DL, Oliveira HR, Wales N, Schlumbaum A. **2015**. Recent advances in ancient DNA research and their implications for archaeobotany. *Vegetation History and Archaeobotany* **24**: 207–214.

Cappellini E, Gilbert MTP, Geuna F, Fiorentino G, Hall A, Thomas-Oates J, Ashton PD, Ashford DA, Arthur P, Campos PF, *et al.* 2010. A multidisciplinary study of archaeological grape seeds. *Naturwissenschaften* **97**: 205–217.

Carpenter ML, Buenrostro JD, Valdiosera C, Schroeder H, Allentoft ME, Sikora M, Rasmussen M, Gravel S, Guillén S, Nekhrizov G, *et al.* 2013. Pulling out the 1%: Whole-Genome Capture for the Targeted Enrichment of Ancient DNA Sequencing Libraries. *The American Journal of Human Genetics* **93**: 852–864.

Cooper A, Poinar H. **2000**. Ancient DNA: Do It Right or Not at All. *Science* **289**: 1139b–1139.

Cronn R, Knaus BJ, Liston A, Maughan PJ, Parks M, Syring J V., Udall J. **2012**. Targeted enrichment strategies for next-generation plant biology. *American Journal of Botany* **99**: 291–311.

Dabney J, Meyer M, Pa S. **2013**. Ancient DNA Damage. *Cold Spring Harb Perspect Biol* **5**: 1–7.

Dellaporta SL, Wood J, Hicks JB. **1983**. A plant DNA minipreparation: Version II. *Plant Molecular Biology Reporter* **1**: 19–21.

Diamond J. **2002**. Evolution, consequences and future of plant and animal domestication. *Nature* **418**: 700–707.

Diez CM, Roessler K, Gaut BS. **2014**. Epigenetics and plant genome evolution. *Current Opinion in Plant Biology* **18**: 1–8.

Doebley JF, Gaut BS, Smith BD. **2006**. The molecular genetics of crop domestication. *Cell* **127**: 1309–21.

Doyle JJ, Doyle JL. **1987**. A rapid procedure for DNA purification from small quantities of fresh leaf tissue. *Phytochemical Bulletin* **19**: 11–15.

Egan AN, Schlueter J, Spooner DM. **2012**. Applications of next-generation sequencing in plant biology. *American Journal of Botany* **99**: 175–185.

Elbaum R, Melamed-Bessudo C, Boaretto E, Galili E, Lev-Yadun S, Levy AA, Weiner S. **2006**. Ancient olive DNA in pits: Preservation, amplification and sequence analysis. *Journal of Archaeological Science* **33**: 77–88.

Erickson DL, Smith BD, Clarke AC, Sandweiss DH, Tuross N. **2005**. An Asian origin for a 10,000-year-old domesticated plant in the Americas. *Proceedings of the National Academy*

of Sciences of the United States of America **102**: 18315–20.

Estrada O, Breen J, Richards SM, Cooper A. **2018**. Ancient plant DNA in the genomic era . *Nature Plants*.

Fernández E, Thaw S, Brown T a., Arroyo-Pardo E, Buxó R, Serret MD, Araus JL. **2013**. DNA analysis in charred grains of naked wheat from several archaeological sites in Spain. *Journal of Archaeological Science* **40**: 659–670.

da Fonseca RR, Smith BD, Wales N, Cappellini E, Skoglund P, Fumagalli M, Samaniego JA, Carøe C, Ávila-Arcos MC, Hufnagel DE, *et al.* 2015. The origin and evolution of maize in the Southwestern United States. *Nature Plants* **1**: 14003.

Fordyce SL, Ávila-Arcos MC, Rasmussen M, Cappellini E, Romero-Navarro JA, Wales N, Alquezar-Planas DE, Penfield S, Brown T a., Vielle-Calzada JP, *et al.* 2013. Deep Sequencing of RNA from Ancient Maize Kernels. *PLoS ONE* **8**: 1–9.

Freitas FO, Bendel G, Allaby RG, Brown TA. **2003**. DNA from primitive maize landraces and archaeological remains: implications for the domestication of maize and its expansion into South America. *Journal of Archaeological Science* **30**: 901–908.

Fu Y, Springer NM, Gerhardt DJ, Ying K, Yeh CT, Wu W, Swanson-Wagner R, D'Ascenzo M, Millard T, Freeberg L, *et al.* 2010. Repeat subtraction-mediated sequence capture from a complex genome. *Plant Journal* **62**: 898–909.

Gansauge M-T, Gerber T, Glocke I, Korlević P, Lippik L, Nagel S, Riehl LM, Schmidt A, Meyer M. **2017**. Single-stranded DNA library preparation from highly degraded DNA using *T4* DNA ligase. *Nucleic Acids Research* **29**: gkx033.

Gansauge M-T, Meyer M. **2013**. Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. *Nature Protocols* **8**: 737–748.

Gansauge M-T, Meyer M. **2014**. Selective enrichment of damaged DNA molecules for ancient genome sequencing. *Genome research* **24**: 1543–9.

Gilbert MTP, Rudbeck L, Willerslev E, Hansen AJ, Smith C, Penkman KEH, Prangenberg K, Nielsen-Marsh CM, Jans ME, Arthur P, *et al.* 2005. Biochemical and physical correlates of DNA contamination in archaeological human bones and teeth excavated at Matera, Italy. *Journal of Archaeological Science* **32**: 785–793.

Ginolhac A, Rasmussen M, Gilbert MTP, Willerslev E, Orlando L. **2011**. mapDamage: Testing for damage patterns in ancient DNA sequences. *Bioinformatics* **27**: 2153–2155.

Gnirke A, Melnikov A, Maguire J, Rogov P, LeProust EM, Brockman W, Fennell T, Giannoukos G, Fisher S, Russ C, *et al.* 2009. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nature biotechnology* **27**: 182–9.

Gokhman D, Lavi E, Prufer K, Fraga MF, Riancho JA, Kelso J, Paabo S, Meshorer E,

Carmel L. 2014. Reconstructing the DNA Methylation Maps of the Neandertal and the Denisovan. *Science* **344**: 523–527.

Golenberg EM, Giannasi DE, Clegg MT, Smiley CJ, Durbin M, Henderson D, Zurawski G. 1990. Chloroplast DNA sequence from a Miocene Magnolia species. *Nature* **344**: 656–658.

Grativol C, Regulski M, Bertalan M, McCombie WR, Da Silva FR, Zerlotini Neto A, Vicentini R, Farinelli L, Hemerly AS, Martienssen RA, *et al.* 2014. Sugarcane genome sequencing by methylation filtration provides tools for genomic research in the genus Saccharum. *Plant Journal* **79**: 162–172.

Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, Patterson N, Li H, Zhai W, Fritz MHY, *et al.* 2010. A Draft Sequence of the Neandertal Genome. *Science* **328**: 710–722.

Gualberto JM, Newton KJ. 2017. Plant Mitochondrial Genomes: Dynamics and Mechanisms of Mutation. *Annual Review of Plant Biology* **68**: annurev-arplant-043015-112232.

Guerrero F, Moolhuijzen P, Peterson D, Bidwell S, Caler E, Bellgard M, Nene V, Djikeng A. 2010. Reassociation kinetics-based approach for partial genome sequencing of the cattle tick, Rhipicephalus (Boophilus) microplus. *BMC Genomics* **11**: 374.

Gugerli F, Parducci L, Petit RJ. 2005. Ancient plant DNA: review and prospects. *New Phytologist* **166**: 409–418.

Gutaker RM, Reiter E, Furtwängler A, Schuenemann VJ, Burbano HA. 2017. Extraction of ultrashort DNA molecules from herbarium specimens. *BioTechniques* **62**: 76–79.

Hansen AJ, Mitchell DL, Wiuf C, Paniker L, Brand TB, Binladen J, Gilichinsky DA, Rønn R, Willerslev E. 2006. Crosslinks rather than strand breaks determine access to ancient DNA sequences from frozen sediments. *Genetics* **173**: 1175–1179.

Hather JG. 1991. The identification of charred archaeological remains of vegetative parenchymous tissue. *Journal of Archaeological Science* **18**: 661–675.

He G, Elling AA, Deng XW. 2011. The Epigenome and Plant Development. *Annual Review of Plant Biology* **62**: 411–435.

Heyn P, Stenzel U, Briggs AW, Kircher M, Hofreiter M, Meyer M. 2010. Road blocks on paleogenomes--polymerase extension profiling reveals the frequency of blocking lesions in ancient DNA. *Nucleic acids research* **38**: e161.

Higuchi R, Bowman B, Freiberger M, Ryder OA, Wilson AC. 1984. DNA sequences from the quagga, an extinct member of the horse family. *Nature* **312**: 282–284.

Hodges E, Xuan Z, Balija V, Kramer M, Molla MN, Smith SW, Middle CM, Rodesch MJ, Albert TJ, Hannon GJ, *et al.* 2007. Genome-wide in situ exon capture for selective resequencing. *Nature genetics* **39**: 1522–7.

Hofman CA, Rick TC, Fleischer RC, Maldonado JE. **2015**. Conservation archaeogenomics: ancient DNA and biodiversity in the Anthropocene. *Trends in Ecology & Evolution* **30**: 540–549.

Hofreiter M, Paijmans JLA, Goodchild H, Speller CF, Barlow A, Fortes GG, Thomas JA, Ludwig A, Collins MJ. **2014**. The future of ancient DNA: Technical advances and conceptual shifts. *BioEssays* **37**: 284–293.

Hofreiter M, Poinar HN, Spaulding WG, Bauer K, Martin PS, Possnert G, Pääbo S. **2000**. A molecular analysis of ground sloth diet through the last glaciation. *Molecular Ecology* **9**: 1975–1984.

Hofreiter M, Serre D, Poinar HN, Kuch M, Pääbo S. **2001**. Ancient DNA. *Nature Reviews Genetics* **2**: 353–359.

Jaenicke-Després V, Buckler ES, Smith BD, Gilbert MTP, Cooper A, Doebley J, Pääbo S. **2003**. Early allelic selection in maize as revealed by ancient DNA. *Science (New York, N.Y.)* **302**: 1206–1208.

Jiao WB, Schneeberger K. **2017**. The impact of third generation genomic technologies on plant genome assembly. *Current Opinion in Plant Biology* **36**: 64–70.

John ME. **1992**. An efficient method for isolation of RNA and DNA from plants containing polyphenolics. *Nucleic acids research* **20**: 2381.

Jónsson H, Ginolhac A, Schubert M, Johnson PLF, Orlando L. **2013**. MapDamage2.0: Fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* **29**: 1682–1684.

Kim C, Lee C, Shin J, Chung Y, Hyung N. **1997**. A simple and rapid method for isolation of high quality genomic DNA from fruit trees and conifers using PVP. *Nucleic Acids Research* **25**: 1085–1086.

Kim S, Soltis DE, Soltis PS, Suh Y. **2004**. DNA sequences from Miocene fossils: An ndhF sequence of Magnolia latahensis (Magnoliaceae) and an rbcL sequence of Persea pseudocarolinensis (Lauraceae). *American Journal of Botany* **91**: 615–620.

Kistler L. **2012**. Ancient DNA Extraction from Plants. In: Shapiro B, Hofreiter M, eds. Ancient DNA: Methods and Protocols. Totowa, NJ: Humana Press, 71–79.

Kistler L, Maezumi SY, Gregorio de Souza J, Przelomska NAS, Malaquias Costa F, Smith O, Loiselle H, Ramos-Madrigal J, Wales N, Ribeiro ER, *et al.* 2018. Multiproxy evidence highlights a complex evolutionary legacy of maize in South America. *Science* **362**: 1309–1313.

Kistler L, Montenegro A, Smith BD, Gifford J a, Green RE, Newsom L a, Shapiro B. **2014**. Transoceanic drift and the domestication of African bottle gourds in the Americas.

Proceedings of the National Academy of Sciences of the United States of America **111**: 1–5.

Kistler L, Newsom LA, Ryan TM, Clarke AC, Smith BD, Perry GH. **2015**. Gourds and squashes (Cucurbita spp.) adapted to megafaunal extinction and ecological anachronism through domestication. *Proceedings of the National Academy of Sciences of the United States of America* **112**: 15107–12.

Kistler L, Shapiro B. **2011**. Ancient DNA confirms a local origin of domesticated chenopod in eastern North America. *Journal of Archaeological Science* **38**: 3549–3554.

Knapp M, Hofreiter M. **2010**. Next generation sequencing of ancient DNA: Requirements, strategies and perspectives. *Genes* **1**: 227–243.

Lamoureux D, Peterson DG, Li W, Fellers JP, Gill BS. **2005**. The efficacy of Cot-based gene enrichment in wheat (Triticum aestivum L.). *Genome* **48**: 1120–1126.

Larson G, Piperno DR, Allaby RG, Purugganan MD, Andersson L, Arroyo-Kalin M, Barton L, Climer Vigueira C, Denham T, Dobney K, *et al.* 2014. Current perspectives and the future of domestication studies. *Proceedings of the National Academy of Sciences* **111**: 6139–6146.

Li C, Dong Y, Liu M, Lu P, Li W, Wang Y, Cui X, Zhou H, Xu Y. **2016**. Ancient DNA analysis of Panicum miliaceum (broomcorn millet) from a Bronze Age cemetery in Xinjiang, China. *Vegetation History and Archaeobotany*.

Li C, Lister DL, Li H, Xu Y, Cui Y, Bower MA, Jones MK, Zhou H. **2011**. Ancient DNA analysis of desiccated wheat grains excavated from a Bronze Age cemetery in Xinjiang. *Journal of Archaeological Science* **38**: 115–119.

Lia V V, Confalonieri V a, Ratto N, Hernández J a C, Alzogaray AMM, Poggio L, Brown T a. **2007**. Microsatellite typing of ancient maize: insights into the history of agriculture in southern South America. *Proceedings. Biological sciences / The Royal Society* **274**: 545–554.

Lindahl T. **1993**. Instability and decay of the primary structure of DNA. *Nature* **362**: 709–715.

Llamas B, Holland ML, Chen K, Cropley JE, Cooper A, Suter CM. **2012**. High-Resolution Analysis of Cytosine Methylation in Ancient DNA (C Lalueza-Fox, Ed.). *PLoS ONE* **7**: e30226.

Llamas B, Valverde G, Fehren-Schmitz L, Weyrich LS, Cooper A, Haak W. **2017**. From the field to the laboratory: Controlling DNA contamination in ancient DNA research. *STAR-Science and Technology in Archaeological Research* **??** ??-??

MacHugh DE, Larson G, Orlando L. **2016**. Taming the Past: Ancient DNA and the Study of Animal Domestication. *Annual Review of Animal Biosciences* **5**: 329–351.

Malmström H, Storå J, Dalén L, Holmlund G, Götherström A. **2005**. Extensive human DNA contamination in extracts from ancient dog bones and teeth. *Molecular Biology and*

Evolution **22**: 2040–2047.

Manen J-F, Bouby L, Dalnoki O, Marinval P, Turgay M, Schlumbaum A. **2003**. Microsatellites from archaeological Vitis vinifera seeds allow a tentative assignment of the geographical origin of ancient cultivars. *Journal of Archaeological Science* **30**: 721–729.

Marciniak S, Perry GH. **2017**. Harnessing ancient genomes to study the history of human adaptation. *Nature Reviews Genetics* **18**: 659–674.

Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben L a, Berka J, Braverman MS, Chen Y-J, Chen Z, *et al.* 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**: 376–80.

Martin MD, Cappellini E, Samaniego J a, Zepeda ML, Campos PF, Seguin-Orlando A, Wales N, Orlando L, Ho SYW, Dietrich FS, *et al.* 2013. Reconstructing genome evolution in historic samples of the Irish potato famine pathogen. *Nature communications* **4**: 2172.

Mascher M, Schuenemann VJ, Davidovich U, Marom N, Himmelbach A, Hübner S, Korol A, David M, Reiter E, Riehl S, *et al.* 2016. Genomic analysis of 6,000-year-old cultivated grain illuminates the domestication history of barley. *Nature Genetics*.

Mehrotra S, Goyal V. **2014**. Repetitive Sequences in Plant Nuclear DNA: Types, Distribution, Evolution and Function. *Genomics, Proteomics and Bioinformatics* **12**: 164–171.

Meyer M, Kircher M. **2010**. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor Protocols* **5**: 1–11.

Meyer M, Kircher M, Gansauge M-T, Li H, Racimo F, Mallick S, Schraiber JG, Jay F, Prufer K, de Filippo C, *et al.* 2012. A High-Coverage Genome Sequence from an Archaic Denisovan Individual. *Science* **338**: 222–226.

Mikić AM. **2015**. The first attested extraction of ancient DNA in legumes (Fabaceae). *Frontiers in Plant Science*.

Miller W, Drautz DI, Ratan A, Pusey B, Qi J, Lesk AM, Tomsho LP, Packard MD, Zhao F, Sher A, *et al.* 2008. Sequencing the nuclear genome of the extinct woolly mammoth. *Nature* **456**: 387–390.

Nakabayashi K, Okamoto M, Koshiba T, Kamiya Y, Nambara E. **2005**. Genome-wide profiling of stored mRNA in Arabidopsis thaliana seed germination: Epigenetic and genetic regulation of transcription in seed. *Plant Journal* **41**: 697–709.

Nistelberger HM, Smith O, Wales N, Star B, Boessenkool S. **2016**. The efficacy of high-throughput sequencing and target enrichment on charred archaeobotanical remains. *Scientific Reports* **6**: 37347.

Nitsch EK, Charles M, Bogaard A. **2015**. Calculating a statistically robust δ13C and δ15N offset for charred cereal and pulse seeds. *Science and Technology of Archaeological Research*

1: 1–8.

Oliveira HR, Civáň P, Morales J, Rodríguez-Rodríguez A, Lister DL, Jones MK. **2012**. Ancient DNA in archaeological wheat grains: preservation conditions and the study of pre-Hispanic agriculture on the island of Gran Canaria (Spain). *Journal of Archaeological Science* **39**: 828–835.

Orlando L, Cooper A. **2014**. Using Ancient DNA to Understand Evolutionary and Ecological Processes. *Annual Review of Ecology, Evolution, and Systematics* **45**: 573–598.

Orlando L, Gilbert MTP, Willerslev E. **2015**. Reconstructing ancient genomes and epigenomes. *Nature Reviews Genetics* **16**: 395–408.

Pääbo S. **1989**. Ancient DNA: extraction, characterization, molecular cloning, and enzymatic amplification. *Proceedings of the National Academy of Sciences of the United States of America* **86**: 1939–1943.

Paabo S, Poinar H, Serre D, Jaenicke-Despres V, Hebler J, Rohland N, Kuch M, Krause J, Vigilant L, Hofreiter M. **2004**. Genetic analyses from ancient DNA. *Annual Review of Genetics* **38**: 645–679.

Palmer JD, Adams KL, Cho Y, Parkinson CL, Qiu YL, Song K. **2000**. Dynamic evolution of plant mitochondrial genomes: mobile genes and introns and highly variable mutation rates. *Proceedings of the National Academy of Sciences of the United States of America* **97**: 6960–6.

Palmer SA, Clapham AJ, Rose P, Freitas FO, Owen BD, Beresford-Jones D, Moore JD, Kitchen JL, Allaby RG. **2012a**. Archaeogenomic evidence of punctuated genome evolution in gossypium. *Molecular Biology and Evolution* **29**: 2031–2038.

Palmer SA, Moore JD, Clapham AJ, Rose P, Allaby RG. **2009**. Archaeogenetic evidence of ancient nubian barley evolution from six to two-row indicates local adaptation. *PLoS ONE* **4**: 2–8.

Palmer LE, Rabinowicz PD, O'Shaughnessy AL, Balija VS, Nascimento LU, Dike S, de la Bastide M, Martienssen RA, McCombie WR, O 'shaughnessy AL. **2003**. Maize Genome Sequencing by Methylation Filtration. *Science* **302**: 2115–2117.

Palmer SA, Smith O, Allaby RG. **2012b**. The blossoming of plant archaeogenetics. *Annals of Anatomy* **194**: 146–156.

Parducci L, Bennett KD, Ficetola GF, Alsos IG, Suyama Y, Wood JR, Pedersen MW. **2017**. Ancient plant DNA in lake sediments. *New Phytologist* **214**: 924–942.

Pearsall DM. **2015**. *Paleoethnobotany, Third Edition: A Handbook of Procedures*. CA: Left Coast Press.

Pérez-Zamorano B, Vallebueno-Estrada M, Martínez González J, García Cook A, Montiel R, Vielle-Calzada J-P, Delaye L. **2017**. Organellar genomes from a ~5,000 years

old archaeological maize sample are closely related to NB genotype. *Genome Biology and Evolution*.

Peterson DG, Schulze SR, Sciara EB, Lee SA, Bowers JE, Nagel A, Jiang N, Tibbitts DC, Wessler SR, Paterson AH. **2002**. Integration of Cot Analysis, DNA Cloning, and High - Throughput Sequencing Facilitates Genome Characterization and Gene Discovery. *Genome Research* **12**: 795–807.

Poinar HN, Cano RJ, Poinar GO. **1993**. DNA from an extinct plant. *Nature* **363**: 677.

Poinar HN, Hofreiter M, Spaulding WG, Martin PS, Stankiewicz BA, Bland H, Evershed RP, Possnert G, Pääbo S, ___., *et al.* 1998. Molecular coproscopy: dung and diet of the extinct ground sloth Nothrotheriops shastensis. *Science (New York, N.Y.)* **281**: 402–6.

Rabinowicz PD, Palmer LE, May BP, Hemann MT, Lowe SW, McCombie WR, Martienssen RA. **2003**. Genes and transposons are differentially methylated in plants, but not in mammals. *Genome Research* **13**: 2658–2664.

Ramos-Madrigal J, Smith BD, Moreno-Mayar JV, Gopalakrishnan S, Ross-Ibarra J, Gilbert MTP, Wales N. **2016a**. Genome Sequence of a 5,310-Year-Old Maize Cob Provides Insights into the Early Stages of Maize Domestication. *Current Biology* **26**: 3195–3201.

Ramos-Madrigal J, Smith BD, Moreno-Mayar JV, Gopalakrishnan S, Ross-Ibarra J, Gilbert MTP, Wales N. **2016b**. Genome Sequence of a 5,310-Year-Old Maize Cob Provides Insights into the Early Stages of Maize Domestication. *Current Biology* **26**: 3195–3201.

Rawlence NJ, Lowe DJ, Wood JR, Young JM, Churchman GJ, Huang Y-TT, Cooper A. **2014**. Using palaeoenvironmental DNA to reconstruct past environments: Progress and prospects. *Journal of Quaternary Science* **29**: 610–626.

Rohland N, Harney E, Mallick S, Nordenfelt S, Reich D. **2015**. Partial uracil-DNA-glycosylase treatment for screening of ancient DNA. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **370**: 20130624.

Rohland N, Hofreiter M. **2007**. Comparison and optimization of ancient DNA extraction. *BioTechniques* **42**: 343–352.

Samarakoon T, Wang SY, Alford MH. **2013**. Enhancing PCR Amplification of DNA from Recalcitrant Plant Specimens Using a Trehalose-Based Additive. *Applications in Plant Sciences* **1**: 1200236.

Sanger F, Nicklen S, Coulson a R. **1977**. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America* **74**: 5463–7.

Der Sarkissian C, Allentoft ME, Avila-Arcos MC, Barnett R, Campos PF, Cappellini E, Ermini L, Fernandez R, da Fonseca R, Ginolhac A, *et al.* 2015. Ancient genomics. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **370**:

20130387–20130387.

Schlumbaum A, Tensen M, Jaenicke-Després V. **2008**. Ancient plant DNA in archaeobotany. *Vegetation History and Archaeobotany* **17**: 233–244.

Schrader C, Schielke A, Ellerbroek L, Johne R. **2012**. PCR inhibitors - occurrence, properties and removal. *Journal of Applied Microbiology* **113**: 1014–1026.

Seguin-Orlando A, Gamba C, Sarkissian C Der, Ermini L, Louvel G, Boulygina E, Sokolov A, Nedoluzhko A, Lorenzen ED, Lopez P, *et al.* 2015. Pros and cons of methylation-based enrichment methods for ancient DNA. *Scientific Reports* **5**: 11826.

Shapiro B, Hofreiter M. **2012**. *Ancient DNA: Methods and Protocols* (B Shapiro and M Hofreiter, Eds.). Totowa, NJ: Humana Press.

Sharma S, Sharma KK, Kuhad RC. **2014**. An efficient and economical method for extraction of DNA amenable to biotechnological manipulations, from diverse soils and sediments. *Journal of Applied Microbiology* **116**: 923–933.

Sidstedt M, Jansson L, Nilsson E, Noppa L, Forsman M, Rådström P, Hedman J. **2015**. Humic substances cause fluorescence inhibition in real-time polymerase chain reaction. *Analytical Biochemistry* **487**: 30–37.

Skoglund P, Northoff BH, Shunkov M V, Derevianko AP, Pääbo S, Krause J, Jakobsson M. **2014**. Separating endogenous ancient DNA from modern day contamination in a Siberian Neandertal. *Proceedings of the National Academy of Sciences of the United States of America* **111**: 2229–34.

Slon V, Hopfe C, Weiß CL, Mafessoni F, de la Rasilla M, Lalueza-Fox C, Rosas A, Soressi M, Knul M V., Miller R, *et al.* 2017. Neandertal and Denisovan DNA from Pleistocene sediments. *Science* **9695**: eaam9695.

Smith O, Clapham A, Rose P, Liu Y, Wang J, Allaby RG. **2014a**. A complete ancient RNA genome: identification, reconstruction and evolutionary history of archaeological Barley Stripe Mosaic Virus. *Scientific reports* **4**: 4003.

Smith O, Clapham AJ, Rose P, Liu Y, Wang J, Allaby RG. **2014b**. Genomic methylation patterns in archaeological barley show de-methylation as a time-dependent diagenetic process. *Scientific Reports* **4**: 5559.

Smith O, Palmer SA, Clapham AJ, Rose P, Liu Y, Wang J, Allaby RG. **2017**. Small RNA activity in archaeological barley shows novel germination inhibition in response to environment. *Molecular Biology and Evolution*.

Smýkal P, Jovanović Ž, Stanisavljević N, Zlatković B, Ćupina B, Đorđević V, Mikić A, Medović A. **2014**. A comparative study of ancient DNA isolated from charred pea (Pisum sativum L.) seeds from an Early Iron Age settlement in southeast Serbia: inference for pea

domestication. *Genetic Resources and Crop Evolution* **61**: 1533–1544.

Soltis PS, Soltis DE, Smiley CJ. **1992**. An rbcL sequence from a Miocene Taxodium (bald cypress). *Proceedings of the National Academy of Sciences of the United States of America* **89**: 449–51.

Springer NM, Schmitz RJ. **2017**. Exploiting induced and natural epigenetic variation for crop improvement. *Nature Reviews Genetics* **18**: 563–575.

Staats M, Cuenca A, Richardson JE, Ginkel RV van, Petersen G, Seberg O, Bakker FT. **2011**. DNA damage in plant herbarium tissue. *PLoS ONE* **6**.

Styring AK, Manning H, Fraser RA, Wallace M, Jones G, Charles M, Heaton THE, Bogaard A, Evershed RP. **2013**. The effect of charring and burial on the biochemical composition of cereal grains: Investigating the integrity of archaeological plant material. *Journal of Archaeological Science* **40**: 4767–4779.

Swarts K, Gutaker RM, Benz B, Blake M, Bukowski R, Holland J, Kruse-Peeples M, Lepak N, Prim L, Cinta Romay M, *et al.* 2017. Genomic estimation of complex traits reveals ancient maize adaptation to temperate North America. *Science* **357**: 512–515.

Vallebueno-Estrada M, Rodríguez-Arévalo I, Rougon-Cardoso A, Martínez González J, García Cook A, Montiel R, Vielle-Calzada J-P. **2016**. The earliest maize from San Marcos Tehuacán is a partial domesticate with genomic evidence of inbreeding. *Proceedings of the National Academy of Sciences* **113**: 14151–14156.

Varma A, Padh H, Shrivastava N. **2007**. Plant genomic DNA isolation: An art or a science. *Biotechnology Journal* **2**: 386–392.

Vasan S, Zhang X, Zhang X, Kapurniotu A, Bernhagen J, Teichberg S, Basgen J, Wagle D, Shih D, Terlecky I, *et al.* 1996. An agent cleaving glucose-derived protein crosslinks in vitro and in vivo. *Nature* **382**: 275–278.

Wagner S, Lagane F, Seguin-Orlando A, Schubert M, Leroy T, Guichoux E, Chancerel E, Bech-Hebelstrup I, Bernard V, Billard C, *et al.* 2018. High-Throughput DNA sequencing of ancient wood. *Molecular Ecology* **27**: 1138–1154.

Wales N, Akman M, Watson RHB, Sánchez Barreiro F, Smith BD, Gremillion KJ, Gilbert MTP, Blackman BK. **2018**. Ancient DNA reveals the timing and persistence of organellar genetic bottlenecks over 3,000 years of sunflower domestication and improvement. *Evolutionary Applications* **12**: 38–53.

Wales N, Andersen K, Cappellini E, ??vila-Arcos MC, Gilbert MTP. **2014**. Optimization of DNA recovery and amplification from non-carbonized archaeobotanical remains. *PLoS ONE* **9**.

Wales N, Carøe C, Sandoval-Velasco M, Gamba C, Barnett R, Samaniego JA, Madrigal

JR, Orlando L, Gilbert MTP. **2015**. New insights on single-stranded versus double-stranded DNA library preparation for ancient DNA. *BioTechniques* **59**.

Wales N, Ramos J, Cappellini E, Carmona A, Pe F, Avila-arcos MC, Gasparyan B, Zardaryan D, Bagoyan T, Smith A, *et al.* 2016. The limits and potential of paleogenomic techniques for reconstructing grapevine domestication. **72**: 5–7.

Wandeler P, Hoeck PEA, Keller LF. **2007**. Back to the future: museum specimens in population genetics. *Trends in Ecology and Evolution* **22**: 634–642.

Weiß CL, Schuenemann VJ, Devos J, Shirsekar G, Reiter E, Gould BA, Stinchcombe JR, Krause J, Burbano HA. **2016**. Temporal patterns of damage and decay kinetics of DNA retrieved from plant herbarium specimens. *Royal Society Open Science* **3**: 160239.

Weyrich LS, Dobney K, Cooper A. **2015**. Ancient DNA analysis of dental calculus. *Journal of Human Evolution* **79**: 119–124.

Weyrich LS, Duchene S, Soubrier J, Arriola L, Llamas B, Breen J, Morris AG, Alt KW, Caramelli D, Dresely V, *et al.* 2017. Neanderthal behaviour, diet, and disease inferred from ancient DNA in dental calculus. *Nature* **544**: 357–361.

Whitelaw CA. **2003**. Enrichment of Gene-Coding Sequences in Maize by Genome Filtration. *Science* **302**: 2118–2120.

Willerslev E, Cooper A. **2005**. Review Paper. Ancient DNA. *Proceedings of the Royal Society B: Biological Sciences* **272**: 3–16.

Yoshida K, Schuenemann VJ, Cano LM, Pais M, Mishra B, Sharma R, Lanz C, Martin FN, Kamoun S, Krause J, *et al.* 2013. The rise and fall of the Phytophthora infestans lineage that triggered the Irish potato famine. *eLife* **2**: 1–25.

Zohary D, Hopf M, Weiss E. **2013**. *Domestication of plants in the Old World : the origin and spread of domesticated plants in Southwest Asia, Europe, and the Mediterranean Basin*. Oxford : Oxford University Press.

Chapter 3

DNA isolation from single seeds: applications in high-throughput sequencing and degraded plant DNA

Manuscript prepared for publication

Statement of authorship

Title of Paper	DNA isolation from single seeds: applications in high-throughput sequencing and degraded plant DNA
Publication Status	☐ Published ☐ Accepted for Publication ☐ Submitted for Publication ☑ Unpublished and Unsubmitted w ork w ritten in manuscript style
Publication Details	Manuscript prepared for submission to the journal *Plant Methods*

Principal Author

Name of Principal Author (Candidate)	Oscar Estrada		
Contribution to the Paper	Contributed to the conception and experimental design, performed DNA extractions, optimised and performed PCR and qPCR assays, constructed Illumina DNA sequencing libraries, performed bioinformatics data analyses, performed statistical analyses, interpreted results, wrote and edited the manuscript.		
Overall percentage (%)	85		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	08/03/2019

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	James Breen		
Contribution to the Paper	Contributed to the conception and experimental design, collected samples, advised the data analyses, reviewed and edited the manuscript.		
Signature		Date	08/03/2019

Name of Co-Author	Stephen M Richards		
Contribution to the Paper	Contributed to the conception and experimental design, advised laboratory analyses, reviewed and edited the manuscript.		
Signature		Date	09/03/2019

Please cut and paste additional co-author panels here as required.

Name of Co-Author	Alan Cooper		
Contribution to the Paper	Contributed to the conception and experimental design, reviewed and edited the manuscript .		
Signature		Date	8/3/19

DNA isolation from single seeds: applications in high-throughput sequencing and degraded plant DNA

Oscar Estrada[1], Stephen M Richards[1], James Breen[2], Alan Cooper[1]

[1]Australian Centre for Ancient DNA, University of Adelaide, North Terrace, South Australia 5005, Australia.
[2]Robinson Research Institute, University of Adelaide, SA, 5005, Australia.

Abstract

Seeds are an abundant plant material preserved in biological collections and also often found as macro-remains in archaeological sites. Accessing the genetic information contained in seeds is pivotal for several areas of plant research such as ancient plant DNA and plant breeding where seeds are one of the primary sources of genetic material. However, DNA retrieval from seeds is challenging owing to the presence of plant compounds that can limit DNA quality and quantity. We evaluated the overall efficiency of four methods to extract DNA from modern and historical seeds (60-100-years-old) of three wheat species (*Triticum* spp.) and examined the suitability of the DNA for downstream analysis using qPCR and high-throughput sequencing. Extraction methods were compared on the basis of DNA yield, DNA purity, and copy number of the *rbcL* gene as measured by qPCR. Further, the extracted DNA was converted into sequencing libraries and assayed for endogenous DNA content by mapping to chloroplast, mitochondrial and nuclear genome references. In both modern and historical seeds, all methods recovered similar proportions of endogenous nuclear sequences (~18.4% for *T. monococcum* L., ~46.9% for *T. turgidum* L., and ~59.4% in *T. aestivum* L.), demonstrating the efficiency of these protocols to recover highly-complex nuclear DNA. In historical samples, the DNeasy PowerPlant Pro protocol (QPP) retrieved more mitochondrial and chloroplast sequences than other methods. In addition, the QPP protocol was more efficient at recovering endogenous nuclear sequences from ancient samples of *T. aestivum*. These findings may help several areas of plant research such as ancient plant DNA, phylogenetics, and plant breeding, where the isolation of DNA directly from seeds could save time, effort, and provide access to genetic information that otherwise be unobtainable.

Introduction

Seeds have been an essential development in the evolution and success of the plants that now inhabit most environments on earth. An extensive collection of compounds synthesised during seed development guarantee the preservation of the embryo for long periods of time, which in turn hosts the genetic information required for the development and survival of a new plant (Sreenivasulu & Wobus, 2013). Accessing the genetic information contained in seeds is pivotal for several areas of plant research such as ancient plant DNA and plant breeding where seeds are one of the primary sources of genetic material. Methods of isolating plant DNA from modern seeds often include germination and subsequent DNA extraction from seedlings. However, such methodologies are not compatible with archaeological or historical seeds, which cannot be germinated owing to the biological degradation and the non-viability of the embryos (Burrieza *et al.*, 2016). Hence, there is a need for the development and improvement of methodologies for DNA isolation that are compatible with archaeological and historical samples.

Nucleic acid isolation from plant seeds is often affected by the presence of secondary compounds such as tannins, alkaloids, and polyphenols and the high starch and protein content of the endosperm (Varma *et al.*, 2007; Abdel-Latif & Osman, 2017). Plant secondary compounds are complex chemicals such as tannins, alkaloids, and polyphenols that are not essential to the growth of plants but are important for the response of plants to biotic and abiotic stress (Varma *et al.*, 2007; Abdel-Latif & Osman, 2017). To avoid the presence of secondary compounds in DNA extracts, protocols implemented in modern seeds often involve the extraction of nucleic acids from seedlings or isolated tissues such as the aleurone (Betts *et al.*, 2017) or embryos (Abdel-Latif & Osman, 2017). Such protocols involve incubation of seeds in water for long periods of time, removing tissues with scalpels, and physical powdering using liquid nitrogen or crushing tissues in mortars. These procedures are time-consuming and not compatible with historical or ancient samples as immersion in water is likely to induce hydrolytic and oxidative damage in surviving nucleic acids (Lindahl, 1993; Dabney *et al.*, 2013).

The recovery of high-quality DNA without contaminants or inhibitors is the goal of any extraction protocol as these compounds can hinder the use of technologies such as high throughput sequencing (HTS) technologies in genomic studies. However, recovery of such DNA from plant material is difficult due to the presence of a rigid cellulose cell wall, the high

content of polysaccharides and proteins, and the presence of secondary compounds such as tannins, alkaloids, and polyphenols (Varma *et al.*, 2007; Abdel-Latif & Osman, 2017). Seeds are particularly challenging owing to the vast array of storage compounds synthesised during development. In cereals, various carbohydrates (mostly starch), storage proteins, and lipids are accumulated in the endosperm, which undergoes programmed cell death without degeneration during seed development (Sreenivasulu & Wobus, 2013). Ancient plant remains and historical specimens may additionally contain humic substances resulting from biological decomposition, leaching from soils, or chemical treatments during storage (Kistler, 2012; Wales *et al.*, 2014), further complicating the recovery of DNA.

DNA isolation from plant tissues has generally involved physical and chemical methods to disrupt cell walls and release nucleic acids. During these processes, the components of the seed are homogenised, and DNA can become crosslinked through the formation of covalent bonds with other macromolecules such as polysaccharides, polyphenols, and proteins (Varma *et al.*, 2007; Schrader *et al.*, 2012). Most crosslinked molecules are washed away during purification steps but reduce nucleic acid yields. However, a fraction of secondary substances can be coextracted with nucleic acids, and both interfere with absorbance-based quantification methods and inhibit enzymatic reactions in downstream applications such as HTS (Kistler, 2012; Schrader *et al.*, 2012; Sidstedt *et al.*, 2015).

Numerous protocols have been developed to purify chloroplast, mitochondrial or nuclear DNA from modern plant tissues (Asif & Cannon, 2005; Varma *et al.*, 2007; Shi *et al.*, 2012; Ahmed & Fu, 2015), but there are currently no standard methods to isolate DNA from degraded plant samples. DNA extraction protocols based on sodium dodecyl sulphate (SDS) and cetyltrimethylammonium bromide (CTAB) lysis buffers have been reported to yield aDNA from ancient samples (Palmer *et al.*, 2009; Cappellini *et al.*, 2010; Wales *et al.*, 2014). Also, a custom SDS lysis buffer including N-phenacylthiazolium bromide (PTB), a reagent that breaks glucose-derived protein crosslinks (Vasan *et al.*, 1996), has been applied together with the DNeasy Plant mini Kit (Qiagen, USA) to isolate DNA from ancient plant specimens (Kistler, 2012). Likewise, commercial kits designed to isolate DNA from soil samples that are developed for dealing with high levels of humic acids have also been applied to extract DNA from herbarium specimens (Heenan *et al.*, 2018).

Although the protocols described above have been able to recover DNA from degraded materials, is still unclear which plant DNA extraction method is more efficient and best suited to current ancient DNA research methods, specifically the application of HTS. Therefore, in this study, we tested the ability of four protocols to extract DNA from modern and historical seeds (60-100-years-old) of three wheat species (*Triticum* spp.). For each combination of age, species, and protocols we measured DNA yield using fluorometric quantification, DNA quality using spectrophotometric quantification, and gene copy number of the chloroplast ribulose-bisphosphate carboxylase (*rbcL*) gene via quantitative PCR (qPCR). We constructed DNA sequencing libraries and generated shotgun HTS data to evaluate the endogenous DNA recovery and behaviour for each method. In addition, we assessed the utility of the four DNA extraction protocols for the retrieval of DNA in ancient seeds of wheat *T.* cf *aestivum* L.

Materials and methods

Plant material

Modern seeds of three wheat species were obtained from the Australian Grains Genebank (Victoria, Australia). The historical wheat seeds (age 60-100-years-old) were obtained from the Georgian National Museum (Tiblisi, Georgia; Dr Nana Rusishvili). Ancient desiccated wheat seeds were excavated from the Areni-1 cave located in southern Armenia (Wilkinson *et al.*, 2012; Smith *et al.*, 2014) and were provided by Dr Nelli Hovhannisyan (Yerevan State University) and Dr Boris Gasparyan (Armenian Academy of Sciences). The Areni-1 seeds were recovered from a wine juice collection tank from Late Chalcolithic II phase occupation and were taxonomically classified *Triticum* cf. *aestivum* L. based on morphological characters (Wilkinson *et al.*, 2012). The ancient wheat seeds have not been dated but wheat and barley seeds from the same unit of the cave have been carbon-dated to 4000 – 3800 Cal BC at 2 sigma calibration range (95.4%) (Boris Gasparyan, personal communication). No permits were required for the use of the historical or ancient wheat seeds. Details of origin and ploidy level for each accession are given in Table 1. The GRIN taxonomy system was used to denote wheat species (https://npgsweb.ars-grin.gov).

Table 1. List of wheat samples analysed in this study. Details of country of origin, age, ploidy, and accession numbers are provided. Ploidy levels have been determined based on the taxonomic identification of the samples and taking as reference the wheat genomic information presented by Feuillet & Muehlbauer (2009). The GRIN taxonomy system is used to denote wheat species (https://npgsweb.ars-grin.gov). Australian Centre for Ancient DNA (ACAD), Australian Grains Genebank (AGG).

Species	Ploidy	Country	Age	Number of seeds	Accession number
T. monococcum L.	2x (AA)	Georgia	Historical	24	ACAD19329
T. turgidum L.	4x (AABB)	Georgia	Historical	24	ACAD19330
T. aestivum L.	6x (AABBDD)	Georgia	Historical	24	ACAD16052
T. monococcum L.	2x (AA)	Russia	Modern	24	AGG33300
T. turgidum L.	4x (AABB)	Syria	Modern	24	AGG26449
T. aestivum L.	6x (AABBDD)	Armenia	Modern	24	AGG31137
T. cf. *aestivum L.*	6x (AABBDD)	Armenia	Ancient	4	ACAD11290

Sample preparation

Modern and historical samples

Modern and historical samples were processed in a standard DNA extraction laboratory at the Australian Centre for Ancient DNA, University of Adelaide. We analysed a set of 144 seeds which included 72 modern and 72 historical (60-100-year-old) wheat seeds with different levels of ploidy (*T. aestivum* 6x, *T. turgidum* 4x and *T. monococcum* 2x) (Table 1). The historical seeds were desiccated and exhibited remarkable preservation, with an appearance similar to the modern seeds except for shrinkage of the embryo (Figure 1).

Seed coats from modern and historical seeds were removed using a clean scalpel blade, and the seeds were then weighed on an AB204-S analytical balance (Mettler Toledo, Greifensee, Switzerland). Two different grinding treatments were applied to evaluate if the amount of starch released during grinding may influence DNA yield. One treatment consisted of partially grinding the seeds in 2 ml tubes using ball bearings with four 2.3 mm Zirconia/Silica beads (Biospec, USA) in an MM300 Mixer Mill (Retsch, Germany) for 5 min at 4000 rpm. The second treatment consisted of fully powdering the grains before DNA extraction. For this, grains were cut into small pieces with a clean scalpel blade and ground as above, for 10 min at 4000 rpm. Three biological replicates were processed for each combination of species, protocols, and treatments (Figure 2).

Also, we included one DNA extraction protocol (DNeasy PowerPlant Pro Kit (Qiagen, USA) which uses a technology that combines mechanical and chemical methods for cell lysis. In this case, wheat seeds were processed in a 2 ml bead tube containing the lysis buffer and four 2.38 mm metal beads provided by the manufacturer. Two grinding treatments were applied as described previously for 5 min or 10 min at 4000 rpm.

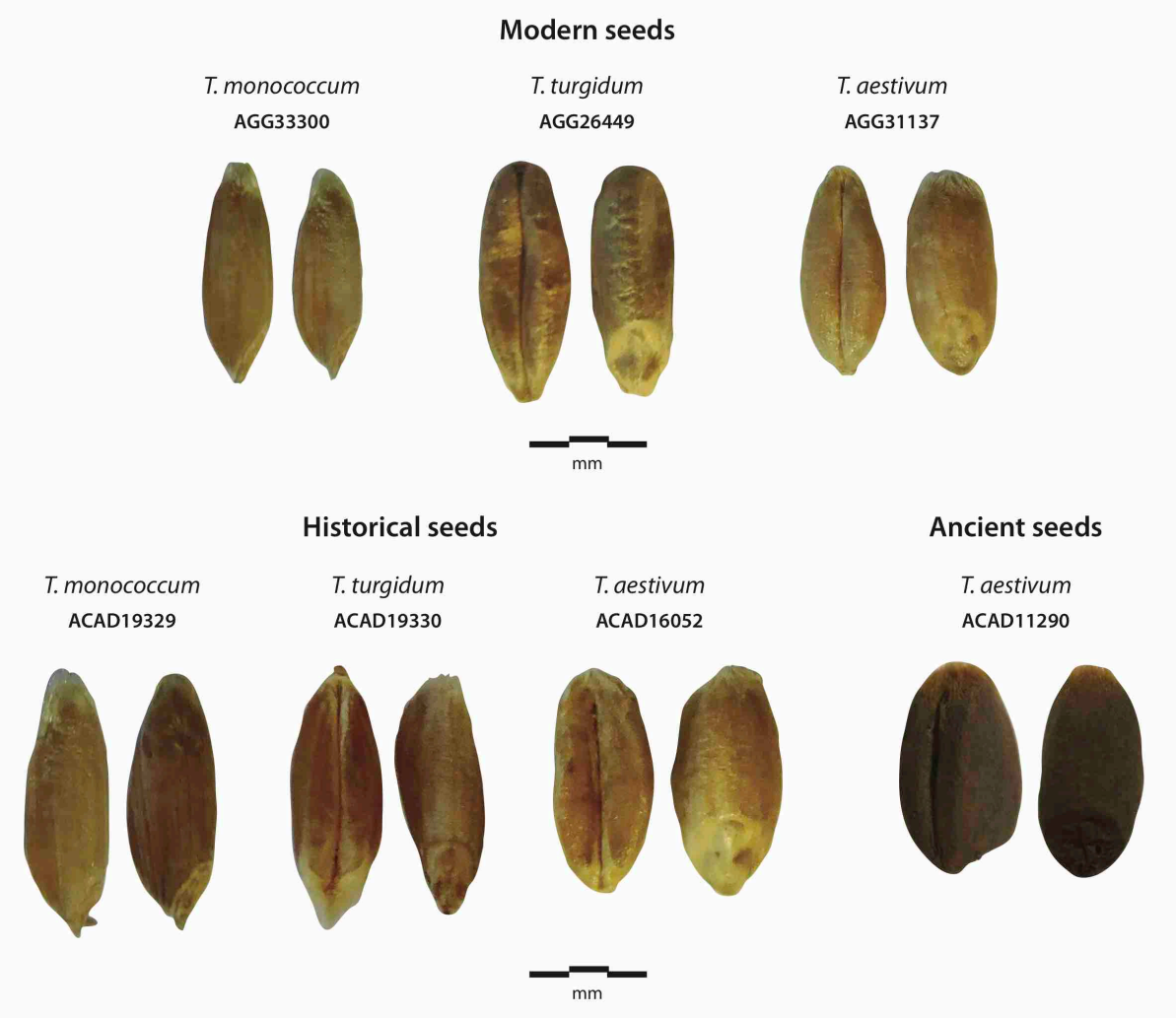


Figure 1. Photographs of the modern (top) and historical and ancient samples (bottom) used in this study. Details of the origin of the samples are given in table 1. The historical samples exhibit good preservation, with an appearance similar to the modern seeds except for shrinkage of the embryo. The dark colour of the ancient seeds is indicative of the taphonomic process of the samples (e.g. generation of Maillard products which lead to the browning colouration).

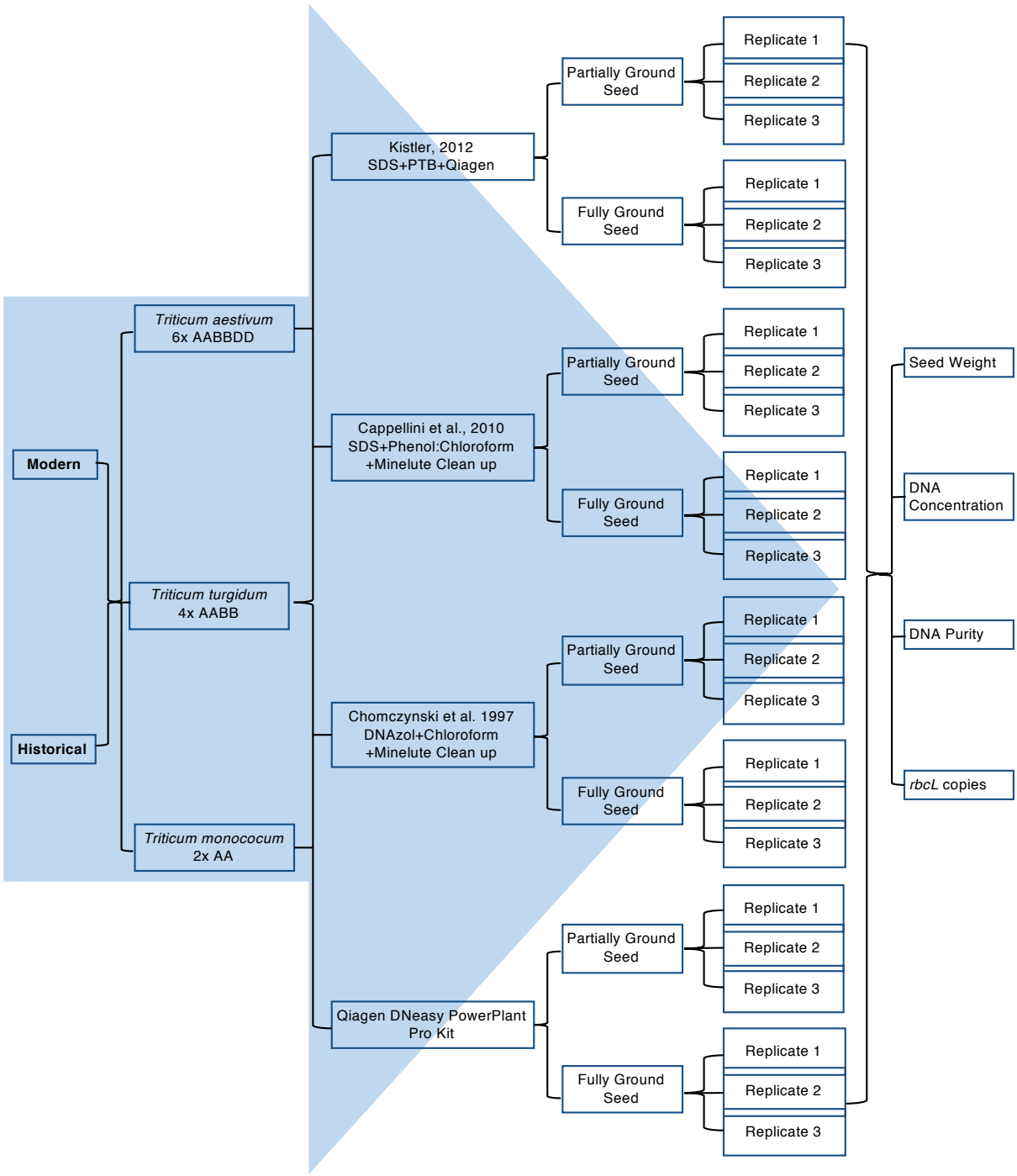


Figure 2. Summary of the experimental design to evaluate the performance of four protocols to isolate DNA from fresh and museum seeds of three species of wheat (*Triticum* spp.).

Ancient samples

Sample preparation, DNA extraction and library construction of ancient material were performed in the purpose-built, physically isolated, ancient DNA laboratory at the Australian Centre for Ancient DNA, University of Adelaide. Standard ancient DNA practices were followed which included negative controls for DNA extractions and amplifications (Cooper & Poinar, 2000; Paabo *et al.*, 2004; Llamas *et al.*, 2017). The archaeological wheat seeds were desiccated, lightweight, fragile, and dark coloured (Figure 1). The grains were washed in a RMS6 rotor (Ratek Instruments, Australia) for two minutes with 1 ml absolute ethanol and rinsed twice with 2 ml of ultrapure water. After rinsing, grains were transferred to a 2 ml screw cap microcentrifuge tubes and macerated using ball bearings with four 2.3 mm Zirconia/Silica beads (Biospec, USA) in a FastPrep 120 instrument (Thermo Savant, USA) at 1 min at 4000 rpm. Due to their rarity, only one ancient grain was tested for each DNA extraction protocol.

DNA extraction

We evaluated three DNA extraction protocols that have previously been used in ancient plant DNA studies. Two protocols are based on the SDS surfactant and contain either SDS or a mixture of SDS and PTB, and have been applied to ancient plant remains such as barley seeds, maize cobs, grape seeds and stems (da Fonseca *et al.*, 2015; Mascher *et al.*, 2016; Wales *et al.*, 2016). The third protocol is based on the Plant DNAzol® reagent (Invitrogen, USA) which was developed for the isolation of DNA from plants (Chomczynski *et al.*, 1997) and that has been recently used to isolate DNA from ancient millet seeds (Richards *et al.,* unpublished data). Lastly, we tested the DNeasy PowerPlant Pro Kit (Qiagen, USA) which includes an inhibitor removal technology to separate polyphenolics and polysaccharides by breaking the bonds between DNA and phenolics and polysaccharides. The DNeasy PowerPlant Pro Kit has been reported to yield high-quality DNA from plant tissues with a high content of secondary compounds such as cotton and coffee seeds.

DNA isolation from modern and historical plant materials was performed as below. Ancient plant samples were processed in a similar manner except that RNase A was not applied, incubation in lysis buffers was for 24 hours, and purification steps were completed using the MinElute spin-columns with the PCR protocol provided with the MinElute kit (Qiagen, USA).

SDS method

Macerated seeds were mixed with 750 µL of lysis buffer containing 2% w/v SDS, 10 mM tris(hydroxymethyl)aminomethane (Tris-HCl) pH 8.0, 10 mM sodium chloride (NaCl), 5 mM calcium chloride $CaCl_2$, 2.5 mM ethylenediaminetetraacetic acid (EDTA) pH 8.0, 40 mM DTT, and 10% proteinase K (20 mg/mL). Four microliters of RNase A (100 mg/ml) were added, and the mixture was incubated at 65 °C for 10 min in a thermo-shaker with constant agitation. The samples were centrifuged at 20000 x g for 5 min, and supernatants were extracted twice with an equal volume of phenol equilibrated with 10 mM Tris HCl, pH 8.0, 1 mM EDTA followed by one round of chloroform:isoamyl alcohol 24:1. The aqueous phase was collected and purified with the DNeasy Plant Mini Kit (Qiagen, USA) following manufacturer's directions.

PTB method

Powdered seeds were processed according to the protocol described by Kistler *et al.,* 2012 with some modifications. The lysis buffer was composed of 1% w/v SDS, 2.5 mM PTB, 10 mM Tris-HCl pH 8.0, 5 mM NaCl, 10 mM EDTA pH 8.0, 50 mM DTT, and 0.4 mg/mL proteinase K. Macerated grains were incubated in 500 µL of lysis buffer and 4 µL of RNase A (100 mg/ml) at 65 °C in a thermo-shaker with constant agitation. The mixture was centrifuged at 10000 x g for 5 min, the supernatant transferred to a new tube and the recovered volume estimated. The supernatant was mixed with 0.325 volumes of DNeasy Plant Mini Kit buffer P3 (Qiagen, USA), incubated for 5 min on ice, and purified with the DNeasy Plant Mini Kit (Qiagen, USA) according to the instructions of the manufacturer.

Plant DNAzol method

Ground seeds were digested in a buffer consisting of 300 µL of Plant DNAzol® reagent (Invitrogen, USA), 2% w/v PVP, and 4 µL of RNase A (100 mg/ml). Digestion was conducted at 25 °C for 10 min in a thermo-shaker with constant agitation. Samples were centrifuged for 5 min at 10000 x g, and the supernatant was extracted with chloroform:isoamyl alcohol 24:1. The aqueous phase was collected and cleaned up with the DNeasy Plant Mini Kit (Qiagen, USA) following manufacturer's directions.

DNeasy PowerPlant Pro Kit (QPP method)

Before grinding, seeds were mixed with the lysis buffer containing 410 µL of bead solution, 40 µl of the phenolic separation solution, 50 µL of solution SL, and 3 µl of RNase A. After powdering, DNA purification was conducted following the manufacturers protocol.

DNA quantification

Purified DNA was quantified with a Qubit 2.0 fluorometer (Life Technologies, Germany) using the double-stranded DNA broad range kit (ThermoFisher Scientific, Germany) with a quantification range from 2-1000 ng/µL. DNA purity was determined using a NanoDrop 1000 spectrophotometer (ThermoFisher Scientific, Germany) measuring sample absorbance ratios at both A260/A280 nm and A260/A230 nm.

rbcL copy number analysis by quantitative PCR (qPCR)

Primer3 online software (http://primer3.ut.ee) was used to design a set of qPCR primers to amplify a 55 bp fragment of the *rbcL* gene (Table 2) based on the *T. aestivum* reference sequence (NCBI accession number AY328025.1). All qPCR assays were carried on a LightCycler® 96 instrument (Roche, Switzerland) using Brilliant II SYBR Green QPCR Master Mix (Agilent Technologies, USA) and 2.5 ng of purified DNA in 10 µl reaction volume. The following thermocycling conditions were utilised: 94 ˚C for 10 min, 50 cycles of 94 ˚C for 30 s, 63 ˚C for 30 s, and 72 ˚C for 40 s. Primer amplification efficiency was tested and optimised on genomic DNA extracted from *T. aestivum* (AGG accession number AGG31137), and specificity was confirmed by shotgun sequencing of PCR amplified products. A qPCR standard curve was generated with a dilution series of amplified and purified *rbcL* amplicons containing 10^1 to 10^6 copies. Each biological sample was evaluated by the qPCR assay using three technical replicates, and the *rbcL* copy number was determined using the LightCycler® 96 Software based on the average cycle threshold (Ct) values.

Table 2. Primer sequences used to amplify a 55 bp fragment of the *rbcL* gene.

Amplicon	Sequence 5' to 3'	Fragment size (bp)	Annealing (ºC)
rbcL	(F) AGCAGCTTGCAAATGGAGTC (R) CTCGAATTTGATCGCCTTCC	55	63

DNA library construction and sequencing

A total of 48 DNA extracts from modern and historical materials representing each combination of species, protocols, and grinding were sheared with an ultrasonicator (Covaris S220, Covaris, USA) to a mean fragment size of 200 bp following the manufacturers settings. Double-stranded DNA Illumina sequencing libraries were constructed from 50 ng of fragmented DNA using a modified protocol (Meyer & Kircher, 2010). First, the ragged ends

of the sheared DNA were polished using T4 Polynucleotide Kinase PNK (NEB, USA) and T4 DNA polymerase (NEB, USA). Then, truncated Illumina adapters containing seven bp internal barcodes were ligated to the DNA using T4 ligase (ThermoFisher Scientific, Germany). The standard structure of the Illumina molecule was subsequently completed by amplifying with an indexed primer set. This indexing amplification reaction was split into eight PCR replicates to reduce clonality in the amplified products. Individual PCR reactions of 25 µl consisted of 5 µl of library DNA, 1X High Fidelity PCR buffer, 2 mM $MgSO_4$, 0.2 mM dNTPs, 0.2 mM of each primer, and 0.1 µl Platinum Taq DNA Polymerase High Fidelity (5 U/µL). The following thermocycling conditions were utilised: 94 ˚C for 2 min, 13 cycles of 94 ˚C for 10 s, 58 ˚C for 30 s, and 68 ˚C for 40 s.

Four DNA libraries from archaeological samples were prepared following the methods described above with some modifications. Ancient DNA extracts were treated with *E. coli* uracil-DNA-glycosylase (UDG) and endonuclease VIII to remove deaminated cytosines, a product of DNA degradation (Gansauge & Meyer, 2014; Rohland *et al.*, 2015). The number of PCR amplification cycles for the ancient libraries was determined via real-time PCR in a LightCycler® 96 instrument (Roche, Switzerland).

Indexed libraries were purified using Ampure XP magnetic beads (Beckman Coulter, USA) and quantified using the Agilent 2200 TapeStation system (Agilent Technologies, USA) for sequencing. Libraries were then pooled at equimolar concentrations and sequenced on the NextSeq500 sequencer at the Australian Cancer Research Foundation (ACRF) Cancer Genomics Facility (Adelaide, South Australia). The modern and historical libraries were sequenced with a NextSeq Mid-output run using 300 cycle chemistry (2 x150 bp), while the ancient libraries were sequenced with a NextSeq Mid-output run using 150 cycle chemistry (2 x 75 bp).

Data analysis

DNA yields and quality

DNA extraction yields, DNA quality, and qPCR quantification were evaluated for the modern and historical material using three biological replicates. There were no replicates for the ancient DNA extractions as there was only enough material for one biological sample per DNA extraction protocol. Calculations were performed in R 3.5.1 (R CoreTeam, 2018) and plots and figures were generated using R base graphics or the ggplot2 package (Wickham, 2016).

Shotgun sequencing analysis

DNA sequence reads were demultiplexed based on the unique P5 and P7 barcoded adapters using *Sabre* version 1.0 (https://github.com/najoshi/sabre) with default parameters and no mismatches allowed. Demultiplexed reads were collapsed and trimmed of adapters using *AdapterRemoval* version 2 (Schubert *et al.*, 2016). Bases with quality lower than four were trimmed and collapsed reads shorter than 25 bp were discarded. Read quality was analysed before and after trimming using *fastQC* version 0.11.5 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc).

To compare sequencing results from modern and historical samples across extraction protocols, 1.4 million reads were randomly subsampled using *seqtk* (https://github.com/lh3/seqtk) with a 100 reads seed. Subsampled reads were mapped to the chloroplast, mitochondrial and nuclear genome references using the *BWA* version 0.7.15 (Li & Durbin, 2009; Li, 2013). Sequenced libraries of *T. monococcum* were mapped to the plastid reference genome (NCBI Accession: NC_021760) and the nuclear reference genome of the closest relative, *T. urartu* Tumanian ex Gandilyan ASM34745v1. Libraries from *T. turgidum* were aligned to the plastid reference genome (NCBI Accession: NC_024814) and the nuclear reference genome of the wild relative *T. turgidum* L. subsp. *dicoccum* (Schrank ex Schübl.) Thell. WEWSeq_v.1.0 (Avni *et al.*, 2017). Samples of *T. aestivum* were mapped to the plastid reference genome (NCBI Accession: NC_002762) and the nuclear reference genome IWGSC RefSeq v1.0 (Appels *et al.*, 2018). As there are no mitochondrial genomes sequenced for *T. monococcum* and *T. turgidum*, all sequenced libraries were mapped to the *T. aestivum* mitochondrial reference genome (NCBI Accession: NC_036024). Alignment files were created using *SAMTools* version 1.3.1 (Li *et al.*, 2009) keeping reads with minimum mapping quality of 30. Alignments were coordinate sorted using *SortSam* and PCR duplicates were removed with *MarkDuplicates*, both available from the *Picard* package version 2.1.1 (https://broadinstitute.github.io/picard).

Ancient libraries were mapped in a similar setting, with the exception that sequenced reads were mapped to reference genomes using the *BWA* aln and samse algorithms version 0.7.15 (Li & Durbin, 2009; Li, 2013) with recommended mapping parameters for aDNA (-l 1024 -n 0.01 -o 2) (Schubert *et al.*, 2012). Unique mapped alignments were also analysed for patterns of DNA damage using *mapDamage* version 2.0.6 (Jónsson *et al.*, 2013).

Summary statistics for each sequenced library were estimated using *SAMTools* version 1.3.1 (Li *et al.*, 2009) and *Qualimap* version 2.2.1 (Okonechnikov *et al.*, 2016), such as number of mapped reads, number of unique reads, library complexity, clonality, percentage of unique endogenous reads, coverage, GC content, and fragment lengths.

Results

DNA yields and quality

To assess the performance of the DNA extraction methods, the yield from each protocol was measured by determining the amount of nanograms of DNA recovered from each milligram of seed. All extraction protocols successfully recovered DNA from the modern and historical samples investigated in this study. A summary of the results of seed weight, DNA yield, and purity obtained for all samples is given in Supplementary Table 1. DNA yield ranged from 45-422 ng/mg in modern seeds and between 12-364 ng/mg in historical seeds. As expected, the average amount of DNA recovered from modern materials (mean= 181 ng/mg) was higher than that obtained from historical samples (mean= 128 ng/mg).

Of the four protocols tested, the PTB and QPP methods demonstrated to recover the highest DNA yield in modern seeds pulverised with both grinding procedures, with an average yield of 259 ng/mg and 234 ng/mg, respectively (Figure 3). Also, in historical samples, the PTB method produced the highest yields of DNA for both grinding treatments (mean= 229 ng/mg) followed by the QPP (mean= 167 ng/mg) (Figure 3). A three-factor analysis of variance (ANOVA) ($P < 0.05$) determined that protocol is significantly influencing DNA yield across species (*T. monococcum* F= 234.0857, P=0.001; *T. turgidum* F= 17.5843, P=0.001; *T. aestivum* F= 244.8475, P=0.001). The application of the PTB reagent in the PTB protocol and the combination of mechanical and chemical lysis in the QPP method seem to be more effective in recovering DNA molecules.

The DNAzol method recovered the lowest yields of DNA across species and grinding treatments in both modern (mean= 70 ng/mg) and historical (mean= 25 ng/mg) seeds (Figure 3). The DNAzol protocol uses a guanidine-detergent lysing buffer and these type of reagents has been reported to promote solidification of seed extracts with high levels of starch (Li & Trick, 2005; Wang *et al.*, 2012). Wheat seeds contain high levels of starch, which could have

caused the solidification of the extracts used in this study and therefore lowered the DNA yields with the DNAzol protocol.

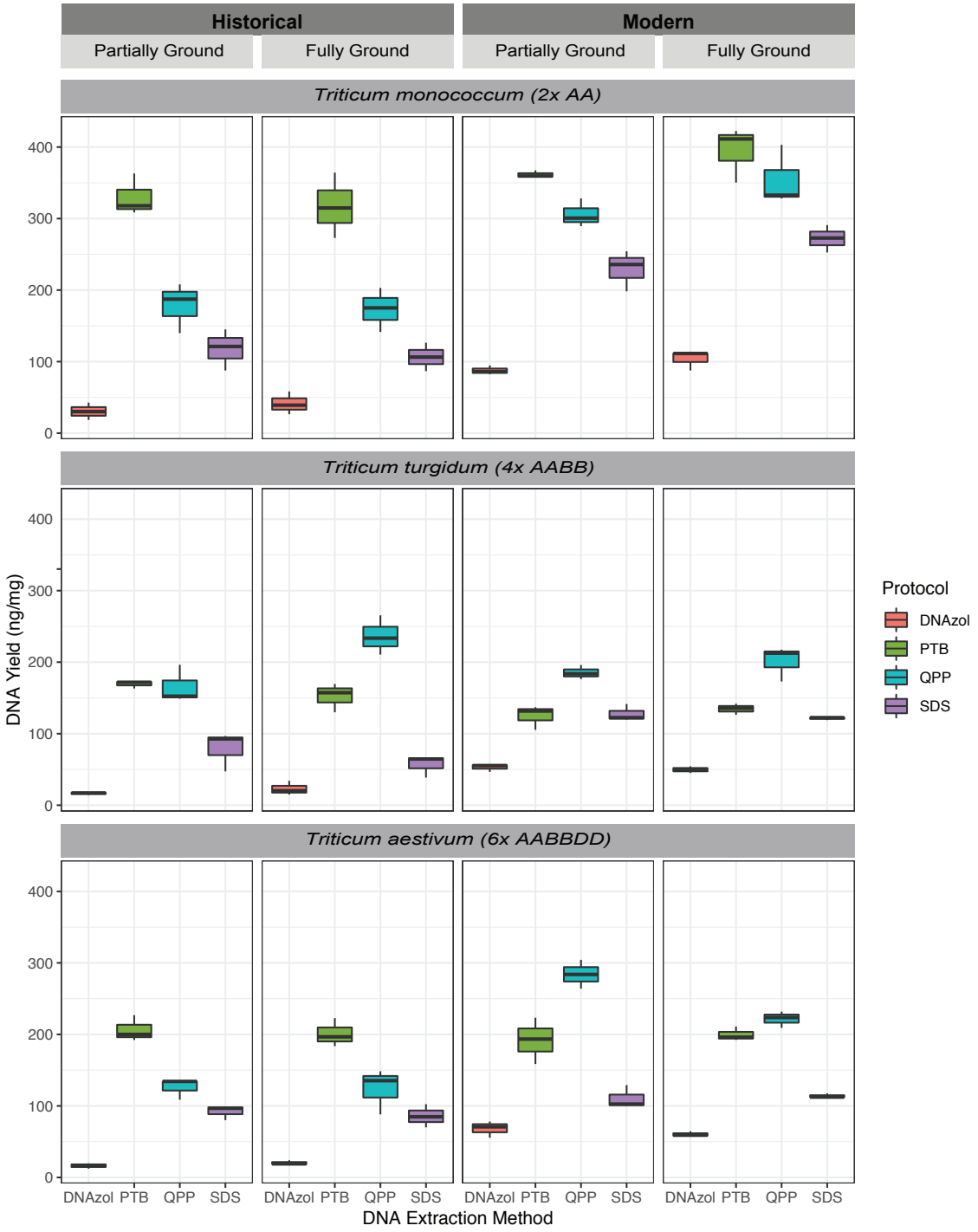


Figure 3. DNA yield (ng/mg) measured for each combination of age, species, grinding treatments, and protocols. The PTB and QPP protocols produced the highest yields across species, age and grinding treatments.

We applied two different grinding treatments as a strategy to avoid the use of pulverisation with liquid nitrogen or mortars and to evaluate if the amount of starch released during grinding may influence DNA yield and quality. Three-factor ANOVA ($P < 0.05$) determined that the grinding treatment did not have a significant effect on the yield of DNA (per milligram of seed) recovered in each species (*T. monococcum* F= 3.7629, P= 0.1; *T. turgidum* F= 2.0490, P= 1; *T. aestivum* F= 3.1200, P= 0.1). Partially powdering seeds produced DNA yields (mean= 152.405 ng/mg) similar to those of thoroughly pulverised seeds (mean= 156.888 ng/mg). The application of this type of grinding treatment is simple, avoids the need for highly laborious procedures such as grinding with liquid nitrogen or crushing tissues in mortars, and yields DNA molecules suitable for downstream procedures such as qPCR and DNA sequencing.

Plant nucleic acid extracts are often contaminated with residual secondary compounds that may result in an overestimation of the nucleic acid concentration and interfere in DNA downstream analysis. The ratios of absorbance at 260/280 nm and 260/230 nm are commonly used to assess DNA purity. A 260/280 ratio of approximately 1.8 and a 260/230 ratio between 2.0 and 2.2 are accepted as an indication of high pure DNA (Desjardins & Conklin, 2011). The spectrophotometric analysis of DNA extracts showed that all four DNA extractions protocols isolated relatively pure DNA with an average ratio of 1.7 at 260/280 nm and 1.95 at 260/230 nm (Figure 4). Although the DNAzol ratio at 260/230 was lower than the other protocols indicating carbohydrate carryover or residual guanidine, the DNA extracts proved suitable for qPCR amplification and DNA sequencing.

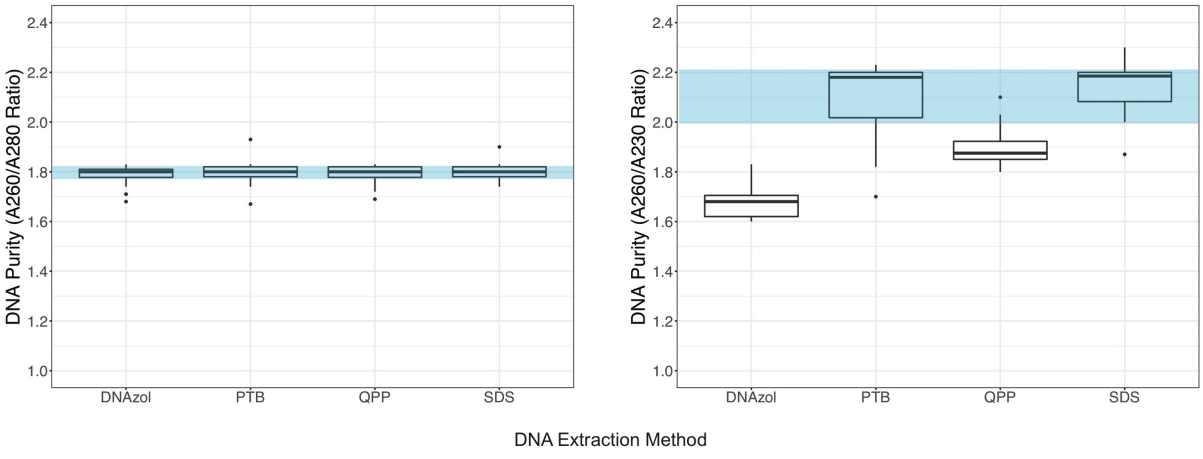


Figure 4. DNA purity evaluated as the ratios of absorbance at 260/280 nm (left) an 260/230 nm(right) for each DNA extraction protocol. The blue shaped area denotes the optimal absorbance ratios.

rbcL copy number analysis by quantitative PCR (qPCR)

To test the recovery of endogenous DNA, we designed a qPCR assay to determine the copy number of the chloroplast *rcbL* gene in all extracts. Amplification of a 55 bp fragment of the *rbcL* gene was successful in all DNA extracts. The number of *rbcL* copies did not reflect the amount of DNA (ng/mg) measured by fluorometry (Figures 3 and 5). This result may indicate that the protocols evaluated in this study recover chloroplast DNA differentially, in comparison to total DNA yield. This finding is statistically supported by a three-factor ANOVA ($P < 0.05$) which determined that protocol have a significant effect on the number of copies of *rbcL* per nanogram of DNA isolated from each species (*T. monococcum* F= 21.8839, P= 0.001; *T. turgidum* F= 19.8024, P=0.001; *T. aestivum* F= 8.3335, P=0.001). The QPP method yielded more *rbcL* copies in historical samples for all three species and grinding treatments (QPP mean= 959 copies/ng, PTB mean= 660 copies/ng, SDS mean= 628 copies/ng, DNAzol mean= 525 copies/ng) (Figure 5). However, in modern specimens all DNA extraction methods performed relatively similarly with DNAzol recovering a higher number of *rbcL* copies (DNAzol mean= 644 copies/ng, QPP mean= 561 copies/ng, SDS mean= 546 copies/ng, PTB mean= 542 copies/ng) (Figure 5).

Surprisingly, the average *rbcL* number of copies recovered from historical grains (mean= 693 copies/ng) was higher than those retrieved from the modern seeds (mean= 573 copies/ng) (Figure 5). A three-factor ANOVA ($P < 0.05$) determined that the interaction between age and protocol have a significant effect on the number of copies of *rbcL* per nanogram of DNA isolated from each species (*T. monococcum* F= 57.5915, P= 0.001; *T. turgidum* F= 22.4047, P=0.001; *T. aestivum* F= 5.2582, P=0.01). The development of seeds in flowering plants involves rapid reprogramming events in several maternal and embryogenic tissues from fertilisation until seed maturation. In early and intermediate stages of seed development, tissues such as the endosperm, the aleurone, the chlorenchyma (a chloroplast-rich tissue), and the pericarp (a maternal tissue) dramatically grow to nourish the embryo (Sreenivasulu & Wobus, 2013). In cereals, after the establishment of the embryo, the endosperm undergoes programmed cell death without degeneration, becoming a storage tissue that accumulates starch and proteins (Sreenivasulu & Wobus, 2013). It is possible that the historical seeds were collected in an intermediate or late stage of maturation and therefore have more chloroplast molecules available. It is also likely that the modern seeds, which were mature seeds ready for germination, have more starch content and thus limiting PCR performance. However, a further

estimation of the content of secondary compounds content in the different seed would be needed to confirm these hypotheses.

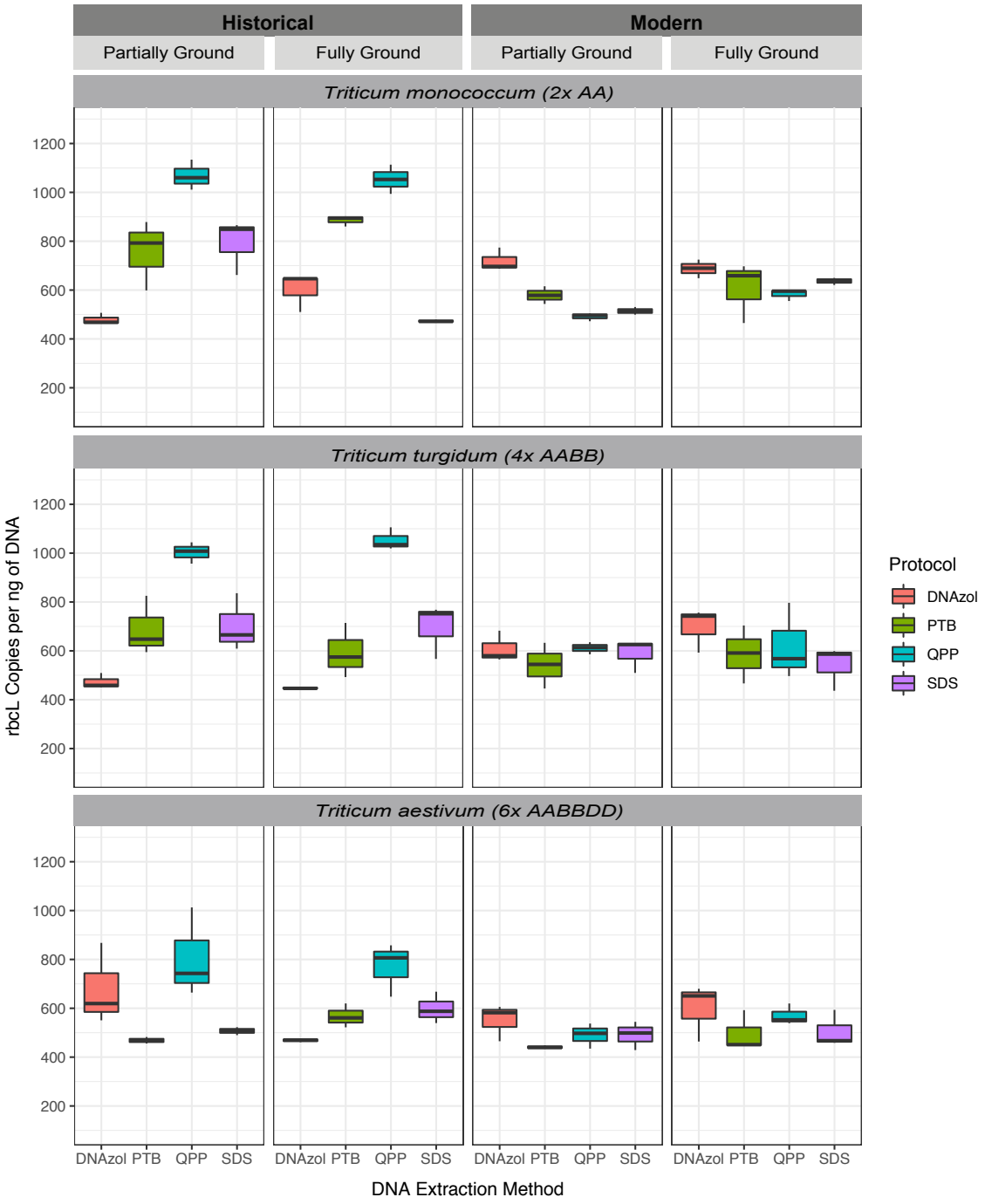


Figure 5. *rcbL* copy number per ng of DNA determined by qPCR for each combination of age, species, grinding treatments, and protocols.

Shotgun sequencing analysis

Modern and historical samples

A total of 48 DNA libraries were shotgun sequenced yielding between 1,549,502 and 6,140,147 raw reads. Summary statistics for each sequenced library were estimated for a subsample of 1.4 million reads per library (Supplementary Table 2-5).

Endogenous content was calculated as the proportion of unique reads mapping to the reference genomes over the total number of subsampled reads after quality filtering and removal of PCR duplicates. The percentage of unique reads mapping to the mitochondrial and chloroplast genomes was less than 1% for all DNA extraction protocols (Figure 6), which is roughly consistent with estimates of total plant DNA containing only 1-10% of chloroplast and mitochondrial DNA (Oldenburg & Bendich, 2004). In addition, seeds in dormancy contain a low proportion of plastid organelles (Sreenivasulu & Wobus, 2013), so our organellular DNA results are not unexpected.

The shotgun sequencing results were consistent with the number of *rbcL* copies measured by qPCR (Figure 5 and 6). DNA libraries constructed from partially powdered grains yielded a slightly higher number of unique reads mapping to the chloroplast than completely ground seeds (mean= 0.13% vs 0.11%), while libraries generated from historical grains produced more unique reads mapping to the chloroplast than modern seeds (mean= 0.15% vs 0.10%). The QPP method recovered more endogenous chloroplast reads in historical specimens across species (mean= 0.32%) while DNAzol retrieved more endogenous chloroplast reads in modern seeds (mean= 0.19%) (Figure 6). The mitochondrial endogenous content of the DNA libraries mirrored the chloroplast results, other than modern samples produced more endogenous mitochondrial DNA reads (mean= 0.49%) than historical samples (mean= 0.43%). (Figure 6)

In contrast to the chloroplast and mitochondrial results, all DNA extraction protocols recovered a similar proportion of unique reads mapping to the nuclear genomes of each species (Figure 7). For the historical specimens of *T. monococcum* this was around 19.1% nuclear endogenous content (QPP method), for *T. turgidum* around 47.8% (DNAzol), and *T. aestivum* around 58.9% (DNAzol). The low endogenous content of *T. monococcum* specimens may be due to the absence of a nuclear reference genome for this species, which was mapped to the closest relative with a sequenced genome *T. turgidum subsp. dicoccoides*. Modern seeds yielded slightly more endogenous nuclear DNA reads (mean= 42.1%), with partially ground seeds producing similar amounts (mean= 42%) to completely ground seeds (mean= 41%) (Figure 7).

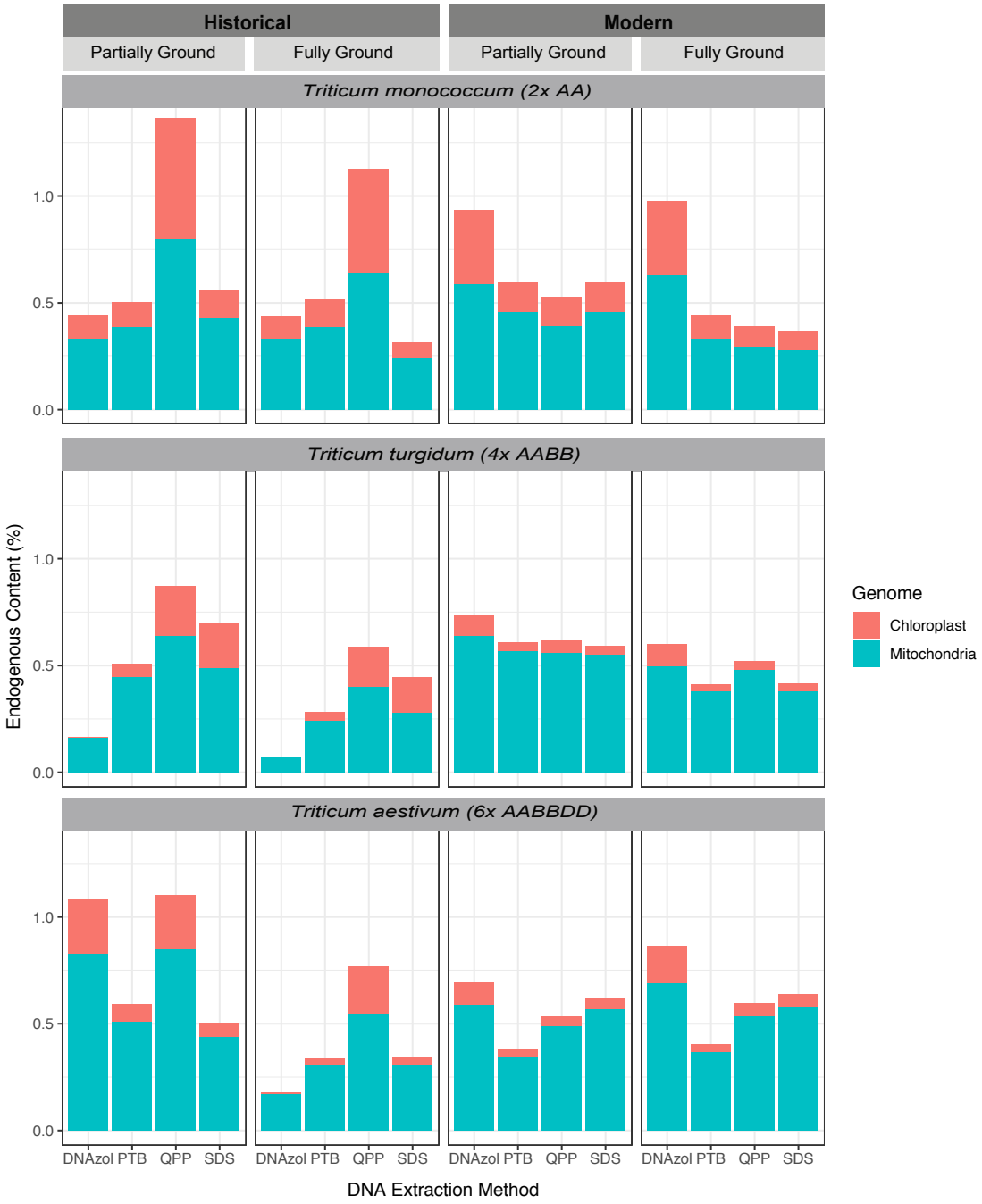


Figure 6. Percentage of endogenous unique reads for chloroplast and mitochondria recovered for each combination of age (historical or modern), species, grinding treatments, and protocols.

In this dataset, genomic library complexity remained high regardless of species, protocol, age or grinding treatment, and ranged from 97.6% to 99.5% in modern samples and from 96.8% to 99.7% in historical specimens (Supplementary Table 3-5). The QPP method had a slightly higher proportion of unique reads mapping to the nuclear genome for historical (99.3-98.5%) and modern specimens (99.1-98.87%) but differences between protocols were minimal.

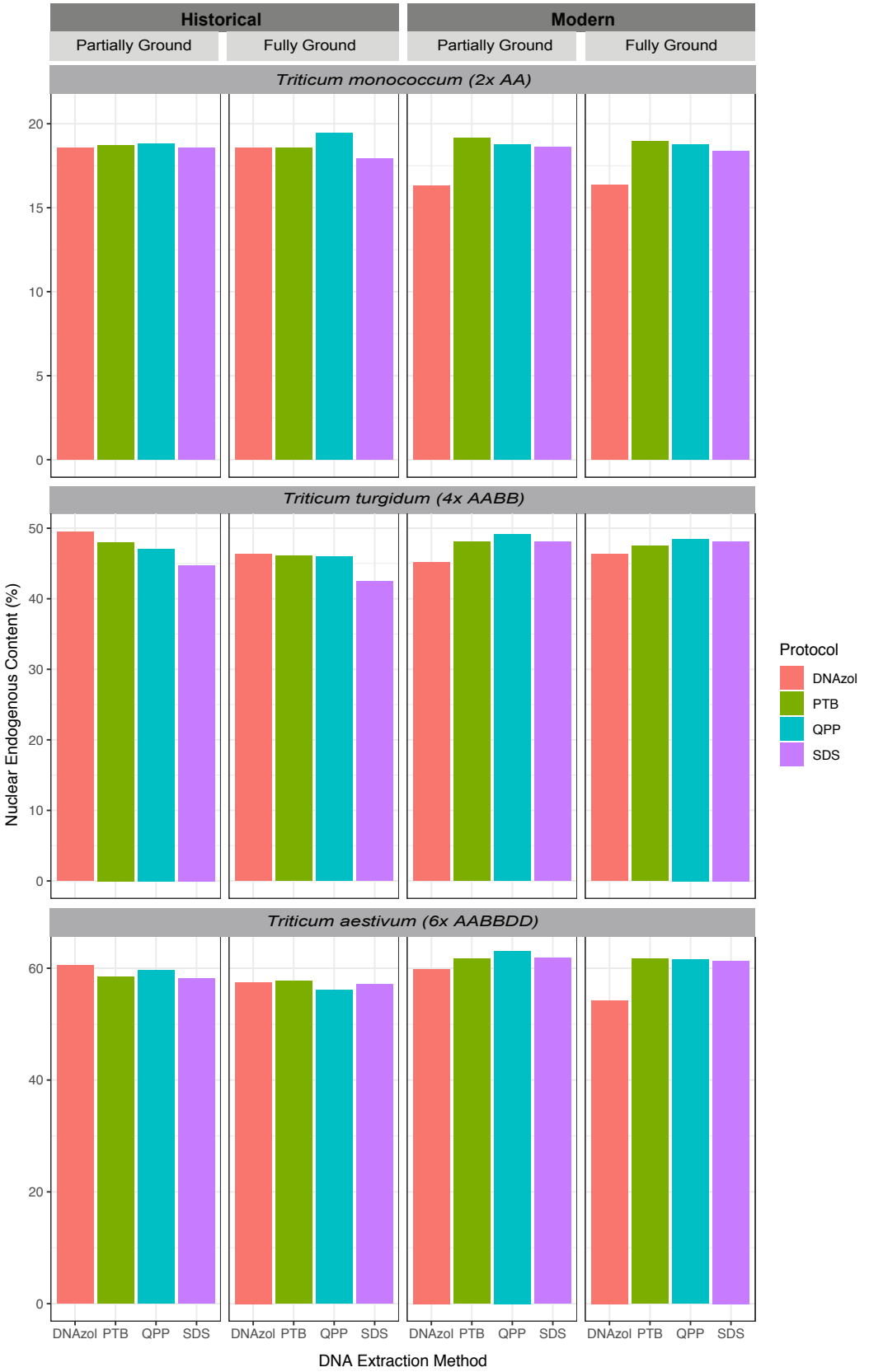


Figure 7. Percentage of nuclear endogenous unique reads recovered for each combination of age, species, grinding treatments, and protocols. The low endogenous content of *T. monococcum* specimens may be due to the absence of a nuclear reference genome for this species, which was mapped to the closest relative with a sequenced genome *T. turgidum subsp. dicoccoides*.

Ancient samples

Shotgun sequencing of the four DNA libraries of ancient wheat (*T.* cf. *aestivum*) yielded between 6.5 and 8.3 million raw reads but had an endogenous nuclear content lower than 0.2% (Table 3). Despite the good morphological preservation of the ancient samples, the low endogenous content indicates poor DNA preservation in the specimens, which were low mass, fragile, and dark coloured (Figure 1). DNA extraction supernatants also showed a dark colour which indicates the presence of secondary compounds such as melanoidins and Maillard products generated through the degradation of the samples. Unfortunately, either no or a very low non-informative number of reads mapped to mitochondria and chloroplast genomes (Supplementary Table 5) and therefore no further analysis was carried on these genomes.

Table 3. Summary of the shotgun sequencing results obtained from four ancient wheat DNA libraries isolated with four DNA extraction protocols. Additional statistics are listed in Supplementary Table 5.

Protocol	Total number of reads	Number of reads aligned to the nuclear genome	Endogenous content	Average read length (bp)	Average GC content (%)
DNAzol	8,893,019	2216	0.007%	36.2	64.45
PTB	8,310,657	1213	0.013%	33.25	72.15
QPP	6,370,977	7705	0.125%	57.33	58.23
SDS	6,500,510	1711	0.020%	34.51	71.81

Despite the low preservation of the samples, DNA extraction protocols recovered between 1,213 to 7,705 reads mapping to the wheat nuclear genome (Table 3). All samples, irrespective of extraction method, had average read fragment lengths lower than 60 base pairs and GC content higher than 58% (Table 3). The small fragment size and high GC content found in all samples supports the finding that these specimens are highly degraded (Briggs *et al.*, 2007; Dabney *et al.*, 2013), and is similar to reports in previous ancient plant DNA publications (Fordyce *et al.*, 2013; Mascher *et al.*, 2016; Ramos-Madrigal *et al.*, 2016).

We employed mapDamage version 2.0.6 (Jónsson *et al.*, 2013) to assess the patterns of DNA damage across all samples (Supplementary Figures 3-6). The specimen extracted with QPP method showed a high proportion of C to T and G to A misincorporations at the first and last position of each fragment (Supplementary Figure 3). This pattern together with the small fragment length distribution is consistent with the characteristics of degraded ancient DNA (Briggs *et al.*, 2007; Dabney *et al.*, 2013; Jónsson *et al.*, 2013). The specimens extracted with DNAzol, PTB, and SDS protocols exhibited a mixed pattern of substitutions across reads,

preventing a clear pattern of misincorporations at the end of the fragments (Supplementary Figures 1, 2, and 4). These mismatches may be the result of the relatively low number of sequencing reads aligning to the reference genome. Given the archaeological value of the ancient samples only one seed per DNA extraction method was tested. A deeper sequencing effort of the aDNA libraries or the application of the extraction protocols to better-preserved plant specimens might allow a better understanding of the performance of each method in ancient plant remains.

Discussion

Nucleic acid isolation from plant seeds is often affected by the presence of secondary compounds and the high starch and protein content of the endosperm (Varma *et al.*, 2007; Abdel-Latif & Osman, 2017). To avoid the presence of secondary compounds in DNA extracts, protocols implemented in modern seeds often extract nucleic acids from seedlings or isolated tissues such as the aleurone (Betts *et al.*, 2017) or embryos (Abdel-Latif & Osman, 2017). Such protocols involve incubation of seeds in water long periods of time, removing tissues with scalpel, and powdering using liquid nitrogen or crushing tissues in mortars. These procedures are time-consuming and not compatible with historical or ancient samples as the immersion in water may induce hydrolytic and oxidative damage to nucleic acids (Lindahl, 1993; Dabney *et al.*, 2013). We aimed to address these limitations by evaluating the overall efficiency of four methods to extract DNA from modern and historical seeds (60-100-years-old) of three wheat species (*T. monococcum, T. turgidum* and *T. aestivum,*) and examine their potential application for downstream analysis such as qPCR and HTS. Also, we assessed the utility of the four DNA extraction protocols for the retrieval of DNA in ancient seeds of wheat *T.* cf. *aestivum*.

The four extraction protocols successfully recovered high-quality DNA from modern and historical seeds that was suitable for qPCR amplification and DNA sequencing. All methods recovered a similar proportion of endogenous nuclear sequences that could be used to produce high complexity DNA sequencing libraries. We also evaluated two bead-based grinding treatments, in which seeds were partially pulverised or cut into small pieces with a scalpel and completely pulverised. We demonstrated that partially powdering seeds retrieved DNA concentrations and endogenous sequences similar to those of thoroughly pulverised seeds, and both produced suitable DNA for HTS analysis. The application of this grinding treatment is simple and avoids the immersion of seeds in water or the utilisation of highly laborious

procedures such as grinding with liquid nitrogen or crushing tissues in mortars which are likely to be difficult to keep free of contaminating modern DNA.

The DNeasy PowerPlant Pro (QPP) protocol proved to be the most effective at generating complex DNA libraries from modern and historical samples, producing a greater amount of reads that mapped to mitochondrial and chloroplast genomes, and the highest amount of endogenous nuclear sequences from the ancient samples of *T.* cf. *aestivum*. The QPP protocol combines mechanical and chemical cell lysis, together with PCR inhibitor removal technology, which appears to improve the recovery of endogenous molecules over the other DNA extraction protocols examined in this study.

The retrieval of high-quality DNA without secondary compounds and inhibitors is a crucial factor in plant genetic analysis. The results presented in this study demonstrate the recovery of high-quality DNA from modern and historical wheat seeds known to be rich in secondary compounds. These findings are important for studies of ancient plant DNA, phylogenetics, and plant breeding, where the isolation of DNA directly from seeds could save time, effort, and provide access to genetic information that otherwise would not be accessible.

References

Abdel-Latif A, Osman G. **2017**. Comparison of three genomic DNA extraction methods to obtain high DNA quality from maize. *Plant Methods* **13**: 1.

Ahmed Z, Fu Y-B. **2015**. An improved method with a wider applicability to isolate plant mitochondria for mtDNA extraction. *Plant methods* **11**: 56.

Appels R, Eversole K, Feuillet C, Keller B, Rogers J, Stein N, Pozniak CJ, Choulet F, Distelfeld A, Poland J, *et al.* 2018. Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* **361**.

Asif MJ, Cannon CH. **2005**. DNA extraction from processed wood: A case study for the identification of an endangered timber species (Gonystylus bancanus). *Plant Molecular Biology Reporter* **23**: 185–192.

Avni R, Nave M, Barad O, Baruch K, Twardziok SO, Gundlach H, Hale I, Mascher M, Spannagl M, Wiebe K, *et al.* 2017. Wild emmer genome architecture and diversity elucidate wheat evolution and domestication. *Science* **357**: 93–97.

Betts NS, Berkowitz O, Liu R, Collins HM, Skadhauge B, Dockter C, Burton RA, Whelan

J, Fincher GB. **2017**. Isolation of tissues and preservation of RNA from intact , germinated barley grain. : 754–765.

Briggs AW, Stenzel U, Johnson PLF, Green RE, Kelso J, Prufer K, Meyer M, Krause J, Ronan MT, Lachmann M, *et al.* 2007. Patterns of damage in genomic DNA sequences from a Neandertal. *Proceedings of the National Academy of Sciences* **104**: 14616–14621.

Burrieza HP, Sanguinetti A, Michieli CT, Bertero HD, Maldonado S. **2016**. Death of embryos from 2300-year-old quinoa seeds found in an archaeological site. *Plant Science* **253**: 107–117.

Cappellini E, Gilbert MTP, Geuna F, Fiorentino G, Hall A, Thomas-Oates J, Ashton PD, Ashford DA, Arthur P, Campos PF, *et al.* 2010. A multidisciplinary study of archaeological grape seeds. *Naturwissenschaften* **97**: 205–217.

Chomczynski P, Mackey K, Drews R, Wilfinger W. **1997**. DNAzol(R): A reagent for the rapid isolation of genomic DNA. *Biotechniques* **22**: 550–553.

Cooper A, Poinar H. **2000**. Ancient DNA: Do It Right or Not at All. *Science* **289**: 1139b–1139.

Dabney J, Meyer M, Pa S. **2013**. Ancient DNA Damage. *Cold Spring Harb Perspect Biol* **5**: 1–7.

Desjardins PR, Conklin DS. **2011**. Microvolume quantitation of nucleic acids. *Current Protocols in Molecular Biology*: 1–4.

Feuillet C, Muehlbauer GJ. **2009**. *Genetics and Genomics of the Triticeae* (GJ Muehlbauer and C Feuillet, Eds.). New York, NY: Springer US.

da Fonseca RR, Smith BD, Wales N, Cappellini E, Skoglund P, Fumagalli M, Samaniego JA, Carøe C, Ávila-Arcos MC, Hufnagel DE, *et al.* 2015. The origin and evolution of maize in the Southwestern United States. *Nature Plants* **1**: 14003.

Fordyce SL, Ávila-Arcos MC, Rasmussen M, Cappellini E, Romero-Navarro JA, Wales N, Alquezar-Planas DE, Penfield S, Brown T a., Vielle-Calzada JP, *et al.* 2013. Deep Sequencing of RNA from Ancient Maize Kernels. *PLoS ONE* **8**: 1–9.

Gansauge M-T, Meyer M. **2014**. Selective enrichment of damaged DNA molecules for ancient genome sequencing. *Genome research* **24**: 1543–9.

Heenan PB, Wood JR, Cole TL. **2018**. A partial cpDNA trnL sequence from the extinct legume *Streblorrhiza speciosa* confirms its placement in the tribe Coluteae (Fabaceae). *Phytotaxa* **374**: 87–91.

Jónsson H, Ginolhac A, Schubert M, Johnson PLF, Orlando L. **2013**. MapDamage2.0: Fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* **29**: 1682–1684.

Kistler L. **2012**. Ancient DNA Extraction from Plants. In: Shapiro B, Hofreiter M, eds. Ancient DNA: Methods and Protocols. Totowa, NJ: Humana Press, 71–79.

Li H. **2013**. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint arXiv:1303.3997*.

Li H, Durbin R. **2009**. Fast and accurate short read alignment with Burrows – Wheeler transform. **25**: 1754–1760.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. **2009**. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078–2079.

Li Z, Trick HN. **2005**. Rapid method for high-quality RNA isolation from seed endosperm containing high levels of starch. *BioTechniques* **38**: 872–876.

Lindahl T. **1993**. Instability and decay of the primary structure of DNA. *Nature* **362**: 709–715.

Llamas B, Valverde G, Fehren-Schmitz L, Weyrich LS, Cooper A, Haak W. **2017**. From the field to the laboratory: Controlling DNA contamination in ancient DNA research. *STAR-Science and Technology in Archaeological Research* **??** ??-??

Mascher M, Schuenemann VJ, Davidovich U, Marom N, Himmelbach A, Hübner S, Korol A, David M, Reiter E, Riehl S, *et al.* 2016. Genomic analysis of 6,000-year-old cultivated grain illuminates the domestication history of barley. *Nature Genetics*.

Meyer M, Kircher M. **2010**. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor Protocols* **5**: 1–11.

Okonechnikov K, Conesa A, Garcı F. **2016**. Genome analysis Qualimap 2 : advanced multi-sample quality control for high-throughput sequencing data. **32**: 292–294.

Oldenburg DJ, Bendich AJ. **2004**. Changes in the structure of DNA molecules and the amount of DNA per plastid during chloroplast development in maize. *Journal of Molecular Biology* **344**: 1311–1330.

Paabo S, Poinar H, Serre D, Jaenicke-Despres V, Hebler J, Rohland N, Kuch M, Krause J, Vigilant L, Hofreiter M. **2004**. Genetic analyses from ancient DNA. *Annual Review of Genetics* **38**: 645–679.

Palmer SA, Moore JD, Clapham AJ, Rose P, Allaby RG. **2009**. Archaeogenetic evidence of ancient nubian barley evolution from six to two-row indicates local adaptation. *PLoS ONE* **4**: 2–8.

R CoreTeam. **2018**. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2016.

Ramos-Madrigal J, Smith BD, Moreno-Mayar JV, Gopalakrishnan S, Ross-Ibarra J, Gilbert MTP, Wales N. **2016**. Genome Sequence of a 5,310-Year-Old Maize Cob Provides

Insights into the Early Stages of Maize Domestication. *Current Biology* **26**: 3195–3201.

Rohland N, Harney E, Mallick S, Nordenfelt S, Reich D. **2015**. Partial uracil-DNA-glycosylase treatment for screening of ancient DNA. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **370**: 20130624.

Schrader C, Schielke A, Ellerbroek L, Johne R. **2012**. PCR inhibitors - occurrence, properties and removal. *Journal of Applied Microbiology* **113**: 1014–1026.

Schubert M, Ginolhac A, Lindgreen S, Thompson JF, Al-rasheid KAS, Willerslev E, Krogh A, Orlando L. **2012**. Improving ancient DNA read mapping against modern reference genomes.

Schubert M, Lindgreen S, Orlando L. **2016**. AdapterRemoval v2: Rapid adapter trimming, identification, and read merging. *BMC Research Notes* **9**: 1–7.

Shi C, Hu N, Huang H, Gao J, Zhao Y-J, Gao L-Z. **2012**. An Improved Chloroplast DNA Extraction Procedure for Whole Plastid Genome Sequencing (Y Xu, Ed.). *PLoS ONE* **7**: e31468.

Sidstedt M, Jansson L, Nilsson E, Noppa L, Forsman M, Rådström P, Hedman J. **2015**. Humic substances cause fluorescence inhibition in real-time polymerase chain reaction. *Analytical Biochemistry* **487**: 30–37.

Smith A, Bagoyan T, Gabrielyan I, Pinhasi R, Gasparyan B. **2014**. Late Chalcolithic and Medieval archaeobotanical remains from Areni-I (Birds'Cave), Armenia. *In: Stone Age of Armenia. A Guidebook to the Stone Age Archaeology in the Republic of Armenia. Monograph of the JSPS-Bilateral Joint Research Project (Vol. 1).*

Sreenivasulu N, Wobus U. **2013**. Seed-Development Programs : A Systems Biology – Based Comparison Between Dicots and Monocots.

Varma A, Padh H, Shrivastava N. **2007**. Plant genomic DNA isolation: An art or a science. *Biotechnology Journal* **2**: 386–392.

Vasan S, Zhang X, Zhang X, Kapurniotu A, Bernhagen J, Teichberg S, Basgen J, Wagle D, Shih D, Terlecky I, *et al.* 1996. An agent cleaving glucose-derived protein crosslinks in vitro and in vivo. *Nature* **382**: 275–278.

Wales N, Andersen K, Cappellini E, ??vila-Arcos MC, Gilbert MTP. **2014**. Optimization of DNA recovery and amplification from non-carbonized archaeobotanical remains. *PLoS ONE* **9**.

Wales N, Ramos J, Cappellini E, Carmona A, Pe F, Avila-arcos MC, Gasparyan B, Zardaryan D, Bagoyan T, Smith A, *et al.* 2016. The limits and potential of paleogenomic techniques for reconstructing grapevine domestication. **72**: 5–7.

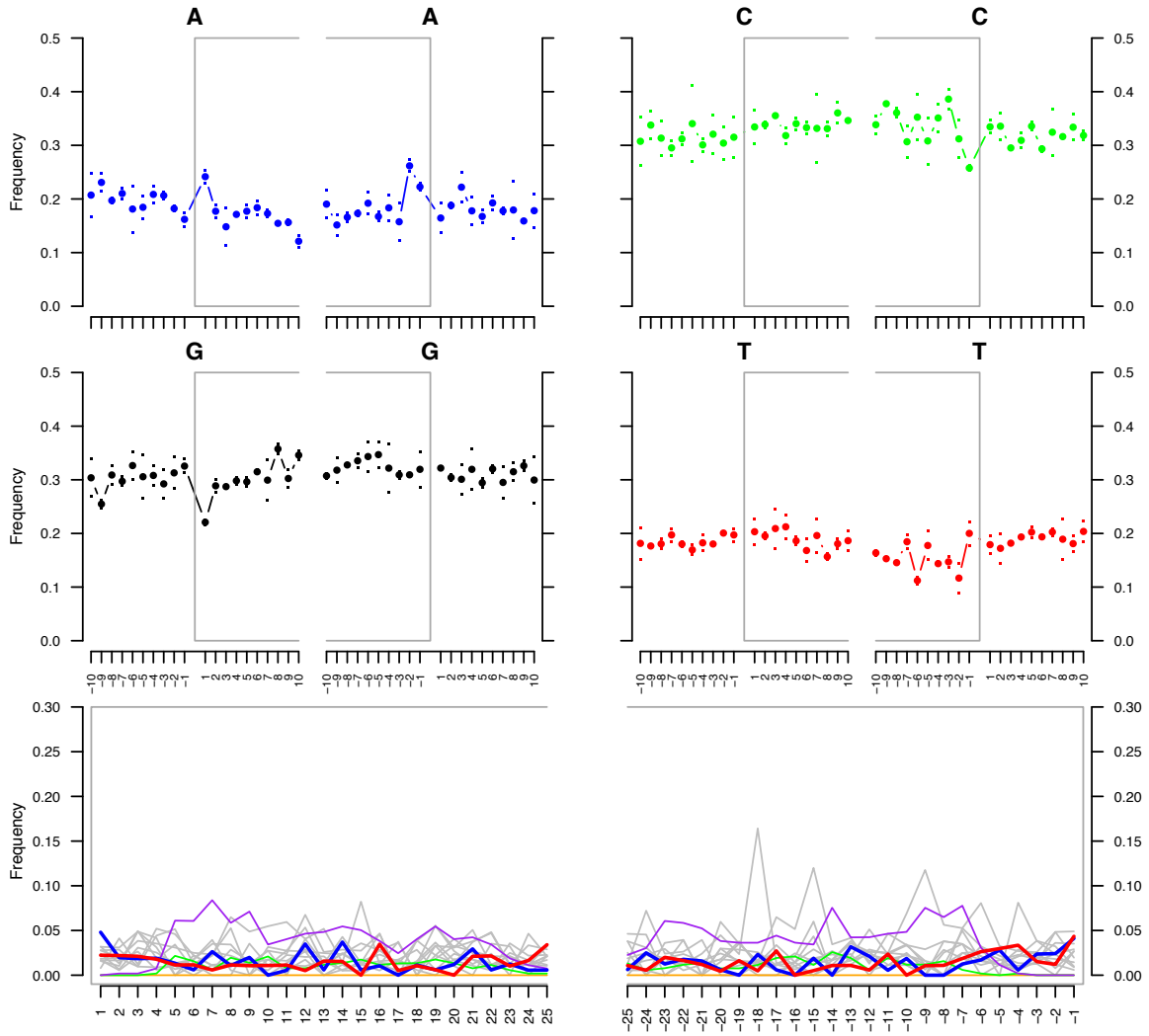
Wang G, Wang G, Zhang X, Wang F, Song R. **2012**. Isolation of high quality RNA from

cereal seeds containing high levels of starch. *Phytochemical Analysis* **23**: 159–163.

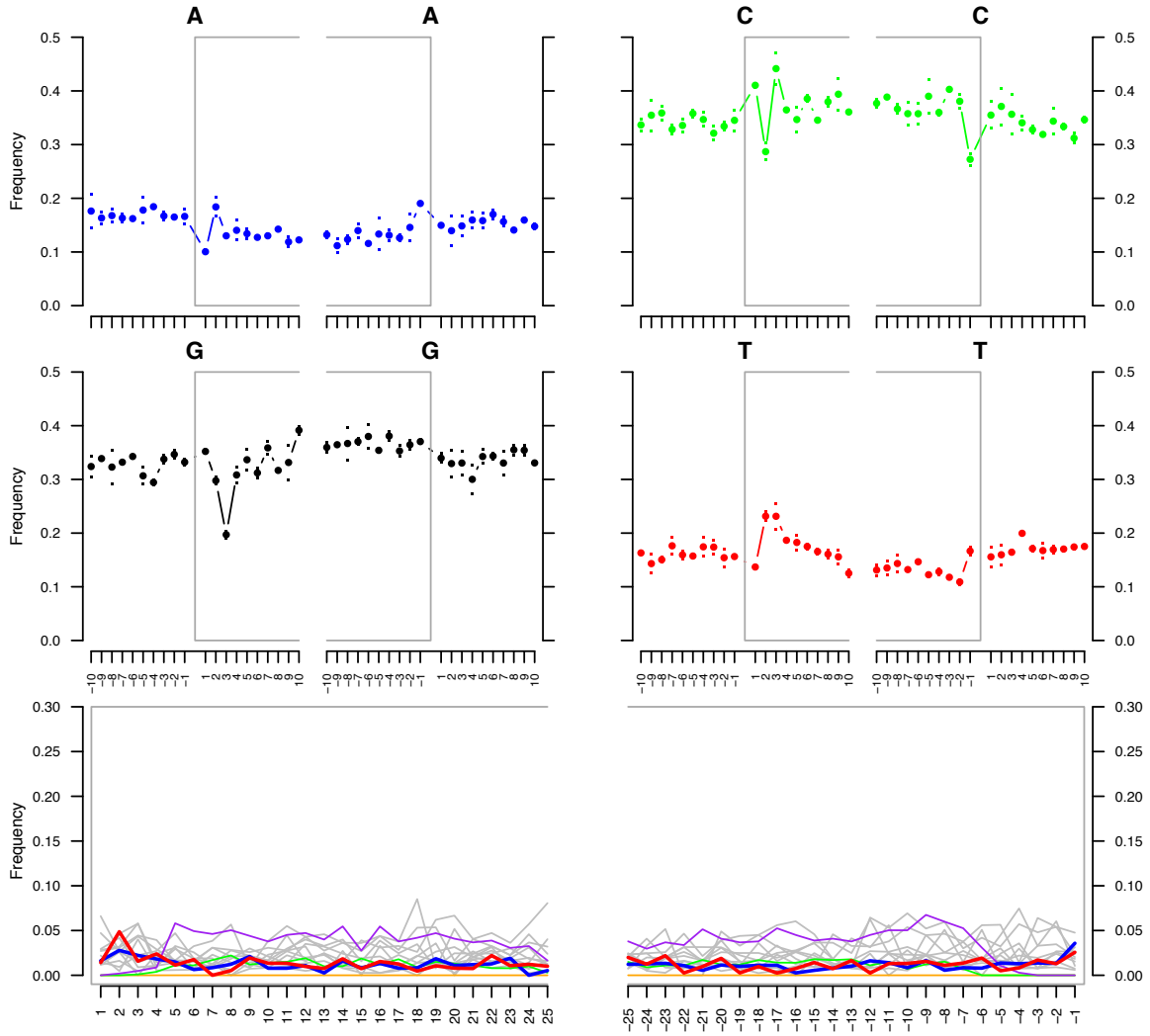
Wickham H. **2016**. *ggplot2: elegant graphics for data analysis*. Springer.

Wilkinson KN, Gasparian B, Pinhasi R, Avetisyan P, Hovsepyan R, Zardaryan D, Areshian GE, Bar-Oz G, Smith A. **2012**. Areni-1 Cave, Armenia: A Chalcolithic–Early Bronze Age settlement and ritual site in the southern Caucasus. *Journal of Field Archaeology* **37**: 20–33.

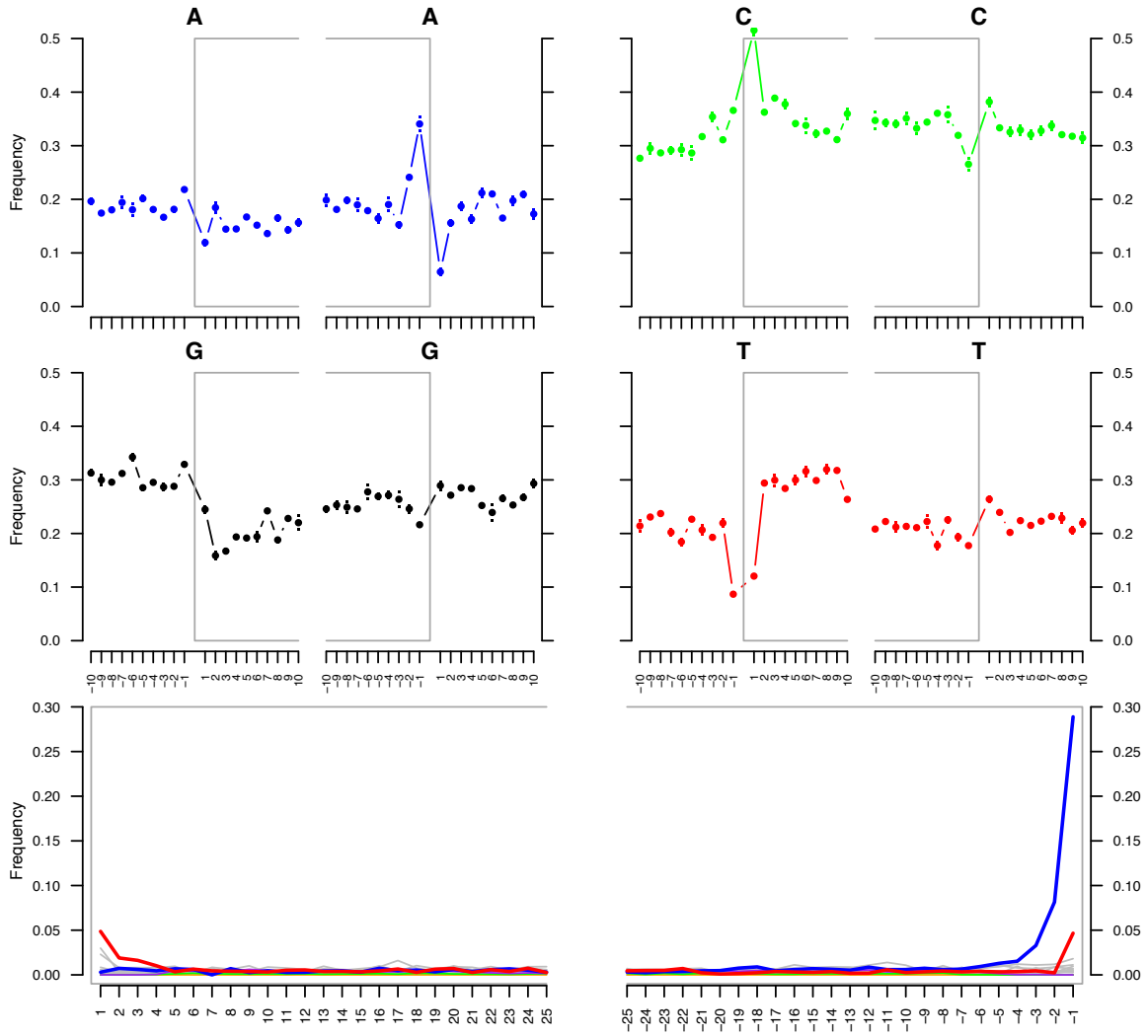
Supplementary information



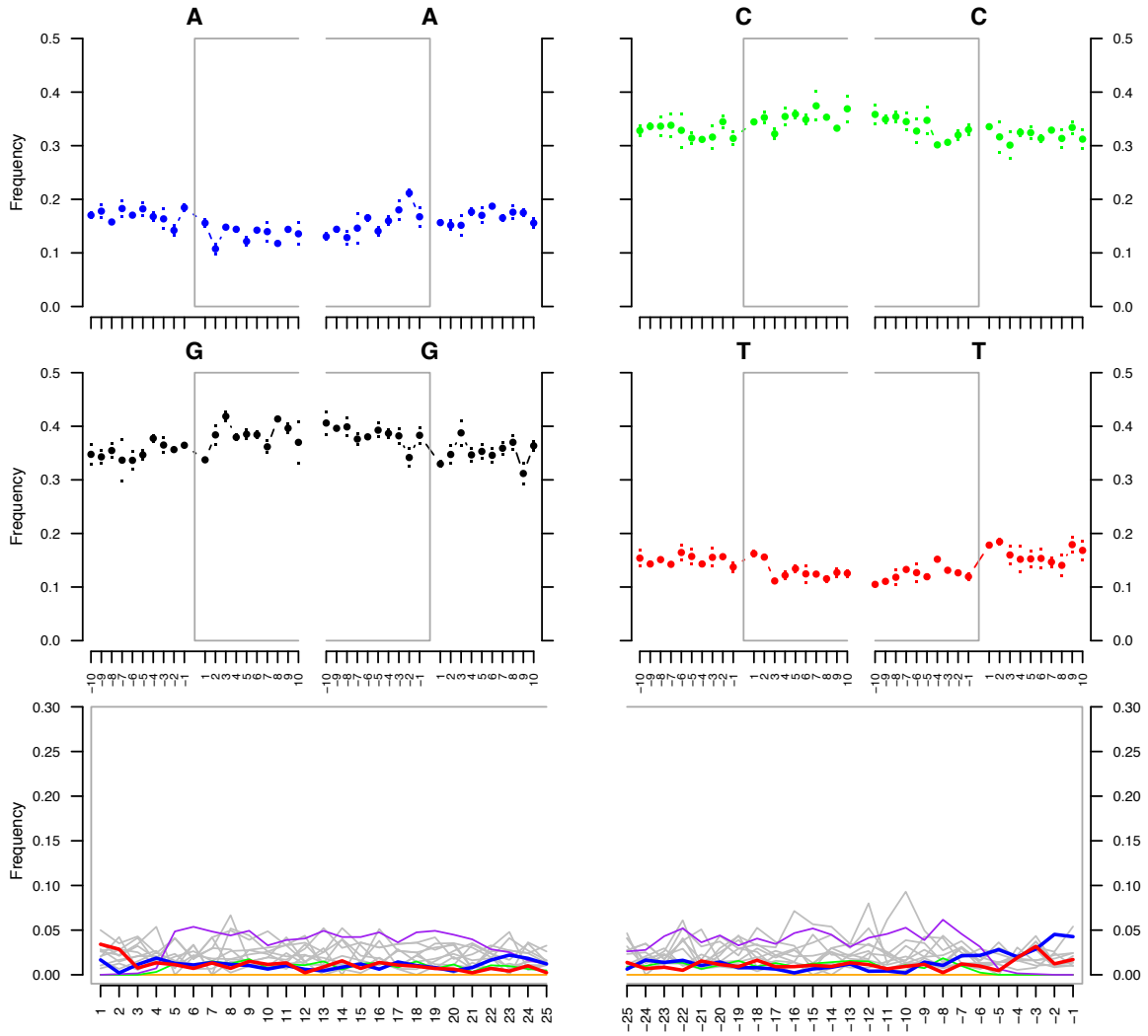
Supplementary Figure 1. MapDamage report for the shotgun sequenced library of the aDNA isolated with the DNAzol protocol. The two top panels show the frequency of bases outside and in the reads. The two lower panels show the specific positions in which substitutions occurred. C to T and G to A substitutions characteristic of ancient DNA are shown red and blue, respectively. All other substitutions are shown in grey. Deletions and insertions are shown in green and blue. Soft-clipped bases are shown in orange.



Supplementary Figure 2. MapDamage report for the shotgun sequenced library of the aDNA isolated with the PTB protocol. The two top panels show the frequency of bases outside and in the reads. The two lower panels show the specific positions in which substitutions occurred. C to T and G to A substitutions characteristic of ancient DNA are shown in red and blue, respectively. All other substitutions are shown in grey. Deletions and insertions are shown in green and blue. Soft-clipped bases are shown in orange.



Supplementary Figure 3. MapDamage report for the shotgun sequenced library of the aDNA isolated with the QPP protocol. The two top panels show the frequency of bases outside and in the reads. The two lower panels show the specific positions in which substitutions occurred. Note the high proportion of 5' C to T and 3' G to A substitutions accumulated at the start of the read characteristic of ancient DNA, shown in red and blue, respectively. All other substitutions are shown in grey. Deletions and insertions are shown in green and blue. Soft-clipped bases are shown in orange.



Supplementary Figure 4. MapDamage report for the shotgun sequenced library of the aDNA isolated with the PTB protocol. The two top panels show the frequency of bases outside and in the reads. The two lower panels show the specific positions in which substitutions occurred. C to T and G to A substitutions characteristic of ancient DNA are shown in red and blue, respectively. All other substitutions are shown in grey. Deletions and insertions are shown in green and blue. Soft-clipped bases are shown in orange.

Supplementary Table 1. Summary of the results of seed weight (mg), DNA yield (ng/mg), purity in 260/280 and 260/230 ratios, and rbcL copy number quantified by qPCR. Australian Centre for Ancient DNA (ACAD), Australian Grains Genebank (AGG).

Sample ID	Accession number	Species	Age	Grinding Treatment	Protocol	Rep	Weight (mg)	Total DNA (ng)	Yield (ng/mg)	Purity 260/280	Purity 260/230	*rbcL* copies/ng
DHAA_1	ACAD16052	*T. aestivum*	Historical	Partially ground	DNAzol	1	43.5	855	19.7	1.8	1.63	868
DHAA_2	ACAD16052	*T. aestivum*	Historical	Partially ground	DNAzol	2	45	554	12.3	1.82	1.62	619
DHAA_3	ACAD16052	*T. aestivum*	Historical	Partially ground	DNAzol	3	44.8	755	16.9	1.77	1.68	551
DHAW_1	ACAD16052	*T. aestivum*	Historical	Fully ground	DNAzol	1	36.8	665	18.1	1.78	1.63	470
DHAW_2	ACAD16052	*T. aestivum*	Historical	Fully ground	DNAzol	2	38.3	920	24.0	1.79	1.7	461
DHAW_3	ACAD16052	*T. aestivum*	Historical	Fully ground	DNAzol	3	39	740	19.0	1.81	1.6	476
DHMA_1	ACAD19329	*T. monococcum*	Historical	Partially ground	DNAzol	1	29.5	547	18.5	1.8	1.77	507
DHMA_2	ACAD19329	*T. monococcum*	Historical	Partially ground	DNAzol	2	32.5	977	30.1	1.78	1.62	468
DHMA_3	ACAD19329	*T. monococcum*	Historical	Partially ground	DNAzol	3	29.9	1276	42.7	1.82	1.8	463
DHMW_1	ACAD19329	*T. monococcum*	Historical	Fully ground	DNAzol	1	33.5	1310	39.1	1.82	1.8	510
DHMW_2	ACAD19329	*T. monococcum*	Historical	Fully ground	DNAzol	2	25.5	675	26.5	1.8	1.76	647
DHMW_3	ACAD19329	*T. monococcum*	Historical	Fully ground	DNAzol	3	21.7	1265	58.3	1.83	1.68	648
DHTA_1	ACAD19330	*T. turgidum*	Historical	Partially ground	DNAzol	1	34.6	592	17.1	1.81	1.76	459
DHTA_2	ACAD19330	*T. turgidum*	Historical	Partially ground	DNAzol	2	36.5	520	14.2	1.78	1.6	451
DHTA_3	ACAD19330	*T. turgidum*	Historical	Partially ground	DNAzol	3	33.6	601	17.9	1.8	1.6	509
DHTW_1	ACAD19330	*T. turgidum*	Historical	Fully ground	DNAzol	1	36.7	740	20.2	1.82	1.7	445
DHTW_2	ACAD19330	*T. turgidum*	Historical	Fully ground	DNAzol	2	38.4	1310	34.1	1.74	1.62	446
DHTW_3	ACAD19330	*T. turgidum*	Historical	Fully ground	DNAzol	3	35.9	534	14.9	1.83	1.7	452
DMAA_1	AGG31137	*T. aestivum*	Modern	Partially ground	DNAzol	1	31.2	2440	78.2	1.8	1.68	605
DMAA_2	AGG31137	*T. aestivum*	Modern	Partially ground	DNAzol	2	29.9	1660	55.5	1.71	1.62	582
DMAA_3	AGG31137	*T. aestivum*	Modern	Partially ground	DNAzol	3	35.9	2540	70.8	1.8	1.68	465
DMAW_1	AGG31137	*T. aestivum*	Modern	Fully ground	DNAzol	1	35.2	2270	64.5	1.71	1.65	680

Sample ID	Accession number	Species	Age	Grinding Treatment	Protocol	Rep	Weight (mg)	Total DNA (ng)	Yield (ng/mg)	Purity 260/280	Purity 260/230	*rbcL* copies/ng
DMAW_2	AGG31137	*T. aestivum*	Modern	Fully ground	DNAzol	2	34.6	2045	59.1	1.8	1.7	651
DMAW_3	AGG31137	*T. aestivum*	Modern	Fully ground	DNAzol	3	32.2	1860	57.8	1.68	1.6	464
DMMA_1	AGG33300	*T. monococcum*	Modern	Partially ground	DNAzol	1	13.1	1080	82.4	1.76	1.83	774
DMMA_2	AGG33300	*T. monococcum*	Modern	Partially ground	DNAzol	2	12.2	1155	94.7	1.8	1.68	697
DMMA_3	AGG33300	*T. monococcum*	Modern	Partially ground	DNAzol	3	12.8	1100	85.9	1.82	1.72	687
DMMW_1	AGG33300	*T. monococcum*	Modern	Fully ground	DNAzol	1	12.3	1370	111.4	1.8	1.68	725
DMMW_2	AGG33300	*T. monococcum*	Modern	Fully ground	DNAzol	2	12.5	1095	87.6	1.76	1.77	649
DMMW_3	AGG33300	*T. monococcum*	Modern	Fully ground	DNAzol	3	12.3	1370	111.4	1.79	1.8	690
DMTA_1	AGG26449	*T. turgidum*	Modern	Partially ground	DNAzol	1	48.9	2280	46.6	1.8	1.64	580
DMTA_2	AGG26449	*T. turgidum*	Modern	Partially ground	DNAzol	2	45.3	2510	55.4	1.77	1.6	683
DMTA_3	AGG26449	*T. turgidum*	Modern	Partially ground	DNAzol	3	45.1	2560	56.8	1.82	1.62	565
DMTW_1	AGG26449	*T. turgidum*	Modern	Fully ground	DNAzol	1	49.3	2220	45.0	1.81	1.65	593
DMTW_2	AGG26449	*T. turgidum*	Modern	Fully ground	DNAzol	2	47.6	2365	49.7	1.77	1.63	757
DMTW_3	AGG26449	*T. turgidum*	Modern	Fully ground	DNAzol	3	41.8	2270	54.3	1.8	1.7	742
PHAA_1	ACAD16052	*T. aestivum*	Historical	Partially ground	PTB	1	41	8200	200.0	1.78	2.2	482
PHAA_2	ACAD16052	*T. aestivum*	Historical	Partially ground	PTB	2	42.3	9600	227.0	1.82	2.23	456
PHAA_3	ACAD16052	*T. aestivum*	Historical	Partially ground	PTB	3	41.1	7900	192.2	1.83	2.18	468
PHAW_1	ACAD16052	*T. aestivum*	Historical	Fully ground	PTB	1	40.3	7400	183.6	1.67	1.7	561
PHAW_2	ACAD16052	*T. aestivum*	Historical	Fully ground	PTB	2	41.7	8200	196.6	1.8	1.82	522
PHAW_3	ACAD16052	*T. aestivum*	Historical	Fully ground	PTB	3	37.7	8400	222.8	1.8	2.01	620
PHMA_1	ACAD19329	*T. monococcum*	Historical	Partially ground	PTB	1	13.5	4900	363.0	1.77	2.2	879
PHMA_2	ACAD19329	*T. monococcum*	Historical	Partially ground	PTB	2	17.3	5500	317.9	1.82	2.21	599
PHMA_3	ACAD19329	*T. monococcum*	Historical	Partially ground	PTB	3	17.5	5400	308.6	1.8	2.23	793
PHMW_1	ACAD19329	*T. monococcum*	Historical	Fully ground	PTB	1	15.1	5500	364.2	1.79	1.91	895

Sample ID	Accession number	Species	Age	Grinding Treatment	Protocol	Rep	Weight (mg)	Total DNA (ng)	Yield (ng/mg)	Purity 260/280	Purity 260/230	*rbcL* copies/ng
PHMW_2	ACAD19329	*T. monococcum*	Historical	Fully ground	PTB	2	16.2	5100	314.8	1.82	2.2	902
PHMW_3	ACAD19329	*T. monococcum*	Historical	Fully ground	PTB	3	26.6	7260	272.9	1.78	2.1	860
PHTA_1	ACAD19330	*T. turgidum*	Historical	Partially ground	PTB	1	33.9	5860	172.9	1.82	2	595
PHTA_2	ACAD19330	*T. turgidum*	Historical	Partially ground	PTB	2	37	6360	171.9	1.8	2.15	648
PHTA_3	ACAD19330	*T. turgidum*	Historical	Partially ground	PTB	3	33.7	5500	163.2	1.93	2.2	825
PHTW_1	ACAD19330	*T. turgidum*	Historical	Fully ground	PTB	1	33.4	5660	169.5	1.8	1.89	714
PHTW_2	ACAD19330	*T. turgidum*	Historical	Fully ground	PTB	2	35.7	4640	130.0	1.78	2.02	493
PHTW_3	ACAD19330	*T. turgidum*	Historical	Fully ground	PTB	3	35.6	5600	157.3	1.78	1.86	575
PMAA_1	AGG31137	*T. aestivum*	Modern	Partially ground	PTB	1	31.4	4980	158.6	1.82	2.2	434
PMAA_2	AGG31137	*T. aestivum*	Modern	Partially ground	PTB	2	24.9	5560	223.3	1.78	2	448
PMAA_3	AGG31137	*T. aestivum*	Modern	Partially ground	PTB	3	29.1	5630	193.5	1.8	2.23	440
PMAW_1	AGG31137	*T. aestivum*	Modern	Fully ground	PTB	1	29	6120	211.0	1.82	2.2	592
PMAW_2	AGG31137	*T. aestivum*	Modern	Fully ground	PTB	2	33.9	6520	192.3	1.78	2.2	447
PMAW_3	AGG31137	*T. aestivum*	Modern	Fully ground	PTB	3	35.7	7000	196.1	1.82	2.18	451
PMMA_1	AGG33300	*T. monococcum*	Modern	Partially ground	PTB	1	14.1	5180	367.4	1.82	2.2	543
PMMA_2	AGG33300	*T. monococcum*	Modern	Partially ground	PTB	2	15.1	5400	357.6	1.82	2.17	578
PMMA_3	AGG33300	*T. monococcum*	Modern	Partially ground	PTB	3	13.1	4710	359.5	1.78	2.2	615
PMMW_1	AGG33300	*T. monococcum*	Modern	Fully ground	PTB	1	13.4	5660	422.4	1.8	2.1	659
PMMW_2	AGG33300	*T. monococcum*	Modern	Fully ground	PTB	2	14.1	5800	411.3	1.78	2.2	697
PMMW_3	AGG33300	*T. monococcum*	Modern	Fully ground	PTB	3	14.3	5010	350.3	1.81	2.18	465
PMTA_1	AGG26449	*T. turgidum*	Modern	Partially ground	PTB	1	47.8	5040	105.4	1.8	2.1	633
PMTA_2	AGG26449	*T. turgidum*	Modern	Partially ground	PTB	2	41.3	5440	131.7	1.78	2.18	446
PMTA_3	AGG26449	*T. turgidum*	Modern	Partially ground	PTB	3	51.1	7000	137.0	1.82	2.2	545
PMTW_1	AGG26449	*T. turgidum*	Modern	Fully ground	PTB	1	44.5	5620	126.3	1.74	2.15	591

Sample ID	Accession number	Species	Age	Grinding Treatment	Protocol	Rep	Weight (mg)	Total DNA (ng)	Yield (ng/mg)	Purity 260/280	Purity 260/230	*rbcL* copies/ng
PMTW_2	AGG26449	*T. turgidum*	Modern	Fully ground	PTB	2	43	5830	135.6	1.82	2.2	467
PMTW_3	AGG26449	*T. turgidum*	Modern	Fully ground	PTB	3	35.8	5080	141.9	1.75	2	704
QHAA_1	ACAD16052	*T. aestivum*	Historical	Partially ground	QPP	1	28.3	3800	134.3	1.74	2	1013
QHAA_2	ACAD16052	*T. aestivum*	Historical	Partially ground	QPP	2	28.3	3800	134.3	1.82	1.93	664
QHAA_3	ACAD16052	*T. aestivum*	Historical	Partially ground	QPP	3	36.8	4000	108.7	1.8	2.03	743
QHAW_1	ACAD16052	*T. aestivum*	Historical	Fully ground	QPP	1	39.2	5820	148.5	1.81	1.93	858
QHAW_2	ACAD16052	*T. aestivum*	Historical	Fully ground	QPP	2	39.7	3500	88.2	1.76	1.87	806
QHAW_3	ACAD16052	*T. aestivum*	Historical	Fully ground	QPP	3	44.5	6020	135.3	1.8	1.85	647
QHMA_1	ACAD19329	*T. monococcum*	Historical	Partially ground	QPP	1	11.5	2392	208.0	1.8	1.82	1134
QHMA_2	ACAD19329	*T. monococcum*	Historical	Partially ground	QPP	2	9	1686	187.3	1.69	1.85	1060
QHMA_3	ACAD19329	*T. monococcum*	Historical	Partially ground	QPP	3	28.9	4042	139.9	1.72	1.81	1011
QHMW_1	ACAD19329	*T. monococcum*	Historical	Fully ground	QPP	1	30.4	4300	141.4	1.82	2.1	1053
QHMW_2	ACAD19329	*T. monococcum*	Historical	Fully ground	QPP	2	10	2030	203.0	1.75	2	1113
QHMW_3	ACAD19329	*T. monococcum*	Historical	Fully ground	QPP	3	21.7	3800	175.1	1.79	1.98	994
QHTA_1	ACAD19330	*T. turgidum*	Historical	Partially ground	QPP	1	32.8	5000	152.4	1.78	1.85	1044
QHTA_2	ACAD19330	*T. turgidum*	Historical	Partially ground	QPP	2	32.7	6420	196.3	1.8	1.87	957
QHTA_3	ACAD19330	*T. turgidum*	Historical	Partially ground	QPP	3	32.9	4900	148.9	1.82	1.84	1008
QHTW_1	ACAD19330	*T. turgidum*	Historical	Fully ground	QPP	1	22.7	4780	210.6	1.8	1.88	1106
QHTW_2	ACAD19330	*T. turgidum*	Historical	Fully ground	QPP	2	28.6	6680	233.6	1.77	1.85	1035
QHTW_3	ACAD19330	*T. turgidum*	Historical	Fully ground	QPP	3	22.9	6080	265.5	1.82	1.8	1019
QMAA_1	AGG31137	*T. aestivum*	Modern	Partially ground	QPP	1	32.2	9140	283.9	1.8	1.83	537
QMAA_2	AGG31137	*T. aestivum*	Modern	Partially ground	QPP	2	35.9	9480	264.1	1.82	1.89	497
QMAA_3	AGG31137	*T. aestivum*	Modern	Partially ground	QPP	3	30.6	9310	304.2	1.79	1.93	435
QMAW_1	AGG31137	*T. aestivum*	Modern	Fully ground	QPP	1	30.6	6400	209.2	1.75	1.98	539

Sample ID	Accession number	Species	Age	Grinding Treatment	Protocol	Rep	Weight (mg)	Total DNA (ng)	Yield (ng/mg)	Purity 260/280	Purity 260/230	*rbcL* copies/ng
QMAW_2	AGG31137	*T. aestivum*	Modern	Fully ground	QPP	2	31.7	7350	231.9	1.82	1.88	619
QMAW_3	AGG31137	*T. aestivum*	Modern	Fully ground	QPP	3	32.2	7200	223.6	1.8	1.88	553
QMMA_1	AGG33300	*T. monococcum*	Modern	Partially ground	QPP	1	15.3	4600	300.7	1.82	1.82	496
QMMA_2	AGG33300	*T. monococcum*	Modern	Partially ground	QPP	2	11.7	3840	328.2	1.8	1.87	505
QMMA_3	AGG33300	*T. monococcum*	Modern	Partially ground	QPP	3	13.3	3850	289.5	1.76	1.88	472
QMMW_1	AGG33300	*T. monococcum*	Modern	Fully ground	QPP	1	6.4	2580	403.1	1.8	1.86	596
QMMW_2	AGG33300	*T. monococcum*	Modern	Fully ground	QPP	2	11.7	3840	328.2	1.77	1.87	555
QMMW_3	AGG33300	*T. monococcum*	Modern	Fully ground	QPP	3	11.3	3760	332.7	1.83	1.83	598
QMTA_1	AGG26449	*T. turgidum*	Modern	Partially ground	QPP	1	43.1	7600	176.3	1.80	1.8	613
QMTA_2	AGG26449	*T. turgidum*	Modern	Partially ground	QPP	2	39.8	7800	196.0	1.82	1.92	586
QMTA_3	AGG26449	*T. turgidum*	Modern	Partially ground	QPP	3	38.1	7000	183.7	1.79	1.88	635
QMTW_1	AGG26449	*T. turgidum*	Modern	Fully ground	QPP	1	43.6	9480	217.4	1.82	1.89	797
QMTW_2	AGG26449	*T. turgidum*	Modern	Fully ground	QPP	2	45.1	7800	172.9	1.78	1.87	568
QMTW_3	AGG26449	*T. turgidum*	Modern	Fully ground	QPP	3	44.6	9480	212.6	1.82	1.89	497
SHAA_1	ACAD16052	*T. aestivum*	Historical	Partially ground	SDS	1	41.5	4125	99.4	1.9	2.21	490
SHAA_2	ACAD16052	*T. aestivum*	Historical	Partially ground	SDS	2	44.3	3550	80.1	1.8	2	509
SHAA_3	ACAD16052	*T. aestivum*	Historical	Partially ground	SDS	3	42.9	4150	96.7	1.78	2.1	523
SHAW_1	ACAD16052	*T. aestivum*	Historical	Fully ground	SDS	1	39.3	2750	70.0	1.82	2.18	668
SHAW_2	ACAD16052	*T. aestivum*	Historical	Fully ground	SDS	2	37.6	3850	102.4	1.8	2.14	588
SHAW_3	ACAD16052	*T. aestivum*	Historical	Fully ground	SDS	3	44.5	3775	84.8	1.83	2.2	539
SHMA_1	ACAD19329	*T. monococcum*	Historical	Partially ground	SDS	1	30.5	4425	145.1	1.8	2.2	662
SHMA_2	ACAD19329	*T. monococcum*	Historical	Partially ground	SDS	2	15.7	1375	87.6	1.74	2.18	849
SHMA_3	ACAD19329	*T. monococcum*	Historical	Partially ground	SDS	3	29.5	3575	121.2	1.8	2.19	866
SHMW_1	ACAD19329	*T. monococcum*	Historical	Fully ground	SDS	1	12.7	1100	86.6	1.82	2.09	479

Sample ID	Accession number	Species	Age	Grinding Treatment	Protocol	Rep	Weight (mg)	Total DNA (ng)	Yield (ng/mg)	Purity 260/280	Purity 260/230	*rbcL* copies/ng
SHMW_2	ACAD19329	*T. monococcum*	Historical	Fully ground	SDS	2	30.8	3275	106.3	1.8	2.2	472
SHMW_3	ACAD19329	*T. monococcum*	Historical	Fully ground	SDS	3	17.8	2250	126.4	1.78	2	469
SHTA_1	ACAD19330	*T. turgidum*	Historical	Partially ground	SDS	1	35	1655	47.3	1.8	2.2	609
SHTA_2	ACAD19330	*T. turgidum*	Historical	Partially ground	SDS	2	37.5	3475	92.7	1.82	2.1	665
SHTA_3	ACAD19330	*T. turgidum*	Historical	Partially ground	SDS	3	33.6	3250	96.7	1.78	2.2	836
SHTW_1	ACAD19330	*T. turgidum*	Historical	Fully ground	SDS	1	34.6	2235	64.6	1.83	2.23	768
SHTW_2	ACAD19330	*T. turgidum*	Historical	Fully ground	SDS	2	37.7	1455	38.6	1.8	2.3	752
SHTW_3	ACAD19330	*T. turgidum*	Historical	Fully ground	SDS	3	34.6	2315	66.9	1.82	2	567
SMAA_1	AGG31137	*T. aestivum*	Modern	Partially ground	SDS	1	31.4	3220	102.5	1.78	2.16	499
SMAA_2	AGG31137	*T. aestivum*	Modern	Partially ground	SDS	2	27.8	3590	129.1	1.82	2.21	429
SMAA_3	AGG31137	*T. aestivum*	Modern	Partially ground	SDS	3	33.6	3350	99.7	1.8	2.06	544
SMAW_1	AGG31137	*T. aestivum*	Modern	Fully ground	SDS	1	27.4	3220	117.5	1.78	2	593
SMAW_2	AGG31137	*T. aestivum*	Modern	Fully ground	SDS	2	34.7	3900	112.4	1.8	2.3	459
SMAW_3	AGG31137	*T. aestivum*	Modern	Fully ground	SDS	3	32	3590	112.2	1.82	2.01	467
SMMA_1	AGG33300	*T. monococcum*	Modern	Partially ground	SDS	1	11.8	3000	254.2	1.78	2.2	513
SMMA_2	AGG33300	*T. monococcum*	Modern	Partially ground	SDS	2	12.3	2440	198.4	1.79	2.26	499
SMMA_3	AGG33300	*T. monococcum*	Modern	Partially ground	SDS	3	12.3	2900	235.8	1.8	2.19	531
SMMW_1	AGG33300	*T. monococcum*	Modern	Fully ground	SDS	1	11	3200	290.9	1.82	1.87	621
SMMW_2	AGG33300	*T. monococcum*	Modern	Fully ground	SDS	2	12.3	3355	272.8	1.75	2.26	639
SMMW_3	AGG33300	*T. monococcum*	Modern	Fully ground	SDS	3	11.3	2855	252.7	1.78	2.2	650
SMTA_1	AGG26449	*T. turgidum*	Modern	Partially ground	SDS	1	41	5800	141.5	1.81	2	632
SMTA_2	AGG26449	*T. turgidum*	Modern	Partially ground	SDS	2	40.2	4920	122.4	1.79	2.18	626
SMTA_3	AGG26449	*T. turgidum*	Modern	Partially ground	SDS	3	50.1	6000	119.8	1.78	2.2	510
SMTW_1	AGG26449	*T. turgidum*	Modern	Fully ground	SDS	1	35.3	4200	119.0	1.8	2.1	599

Sample ID	Accession number	Species	Age	Grinding Treatment	Protocol	Rep	Weight (mg)	Total DNA (ng)	Yield (ng/mg)	Purity 260/280	Purity 260/230	*rbcL* copies/ng
SMTW_2	AGG26449	*T. turgidum*	Modern	Fully ground	SDS	2	43.3	5300	122.4	1.82	2	587
SMTW_3	AGG26449	*T. turgidum*	Modern	Fully ground	SDS	3	34.2	4200	122.8	1.8	2.2	437

Supplementary Table 2. Shotgun sequencing statistics of 48 modern and historical wheat DNA libraries constructed from DNA extracts obtained with four extraction methods.

Sample ID	Species	Age	Protocol	Grinding Treatment	Total reads	Reads after trimming and collapsing	Percentage of reads collapsed	Sampled reads	Fraction of total reads sampled
DHAA_1	*T. aestivum*	Historical	DNAzol	Partially ground	6,140,147	5,374,823	87.54%	1,400,000	26.05%
DHAW_1	*T. aestivum*	Historical	DNAzol	Fully ground	2,028,281	1,821,983	89.83%	1,400,000	76.84%
DHMA_1	*T. monococcum*	Historical	DNAzol	Partially ground	2,772,823	2,521,960	90.95%	1,400,000	55.51%
DHMW_1	*T. monococcum*	Historical	DNAzol	Fully ground	2,188,258	1,994,787	91.16%	1,400,000	70.18%
DHTA_1	*T. turgidum*	Historical	DNAzol	Partially ground	3,626,061	3,009,255	82.99%	1,400,000	46.52%
DHTW_1	*T. turgidum*	Historical	DNAzol	Fully ground	1,549,502	1,417,111	91.46%	1,400,000	98.79%
DMAA_1	*T. aestivum*	Modern	DNAzol	Partially ground	2,028,280	1,703,443	83.98%	1,400,000	82.19%
DMAW_1	*T. aestivum*	Modern	DNAzol	Fully ground	3,815,700	3,454,521	90.53%	1,400,000	40.53%
DMMA_2	*T. monococcum*	Modern	DNAzol	Partially ground	5,066,341	4,549,689	89.80%	1,400,000	30.77%
DMMW_1	*T. monococcum*	Modern	DNAzol	Fully ground	3,359,670	3,083,019	91.77%	1,400,000	45.41%
DMTA_2	*T. turgidum*	Modern	DNAzol	Partially ground	2,186,928	1,923,902	87.97%	1,400,000	72.77%
DMTW_1	*T. turgidum*	Modern	DNAzol	Fully ground	3,499,890	2,953,854	84.40%	1,400,000	47.40%
PHAA_1	*T. aestivum*	Historical	PTB	Partially ground	2,101,857	1,868,172	88.88%	1,400,000	74.94%
PHAW_3	*T. aestivum*	Historical	PTB	Fully ground	3,059,759	2,802,527	91.59%	1,400,000	49.95%
PHMA_2	*T. monococcum*	Historical	PTB	Partially ground	1,748,342	1,552,789	88.81%	1,400,000	90.16%
PHMW_3	*T. monococcum*	Historical	PTB	Fully ground	1,940,427	1,752,322	90.31%	1,400,000	79.89%
PHTA_2	*T. turgidum*	Historical	PTB	Partially ground	3,258,108	2,830,176	86.87%	1,400,000	49.47%

Sample ID	Species	Age	Protocol	Grinding Treatment	Total reads	Reads after trimming and collapsing	Percentage of reads collapsed	Sampled reads	Fraction of total reads sampled
PHTW_2	*T. turgidum*	Historical	PTB	Fully ground	2,998,314	2,740,547	91.40%	1,400,000	51.08%
PMAA_2	*T. aestivum*	Modern	PTB	Partially ground	2,857,039	2,548,872	89.21%	1,400,000	54.93%
PMAW_1	*T. aestivum*	Modern	PTB	Fully ground	3,184,428	2,826,907	88.77%	1,400,000	49.52%
PMMA_1	*T. monococcum*	Modern	PTB	Partially ground	3,081,232	2,663,576	86.45%	1,400,000	52.56%
PMMW_1	*T. monococcum*	Modern	PTB	Fully ground	2,928,171	2,585,419	88.29%	1,400,000	54.15%
PMTA_1	*T. turgidum*	Modern	PTB	Partially ground	2,282,999	1,988,071	87.08%	1,400,000	70.42%
PMTW_1	*T. turgidum*	Modern	PTB	Fully ground	2,430,279	2,150,720	88.50%	1,400,000	65.09%
QHAA_1	*T. aestivum*	Historical	QPP	Partially ground	2,085,828	1,563,976	74.98%	1,400,000	89.52%
QHAW_1	*T. aestivum*	Historical	QPP	Fully ground	3,888,577	3,421,002	87.98%	1,400,000	40.92%
QHMA_2	*T. monococcum*	Historical	QPP	Partially ground	3,581,442	3,231,212	90.22%	1,400,000	43.33%
QHMW_1	*T. monococcum*	Historical	QPP	Fully ground	3,910,380	3,330,661	85.17%	1,400,000	42.03%
QHTA_2	*T. turgidum*	Historical	QPP	Partially ground	4,251,409	3,613,249	84.99%	1,400,000	38.75%
QHTW_2	*T. turgidum*	Historical	QPP	Fully ground	4,715,342	4,164,470	88.32%	1,400,000	33.62%
QMAA_1	*T. aestivum*	Modern	QPP	Partially ground	4,244,900	3,647,289	85.92%	1,400,000	38.38%
QMAW_2	*T. aestivum*	Modern	QPP	Fully ground	2,984,748	2,594,859	86.94%	1,400,000	53.95%
QMMA_1	*T. monococcum*	Modern	QPP	Partially ground	3,057,486	2,667,700	87.25%	1,400,000	52.48%
QMMW_2	*T. monococcum*	Modern	QPP	Fully ground	3,274,925	2,916,436	89.05%	1,400,000	48.00%
QMTA_2	*T. turgidum*	Modern	QPP	Partially ground	3,774,970	3,225,304	85.44%	1,400,000	43.41%
QMTW_1	*T. turgidum*	Modern	QPP	Fully ground	3,894,833	3,364,833	86.39%	1,400,000	41.61%
SHAA_2	*T. aestivum*	Historical	SDS	Partially ground	2,030,679	1,840,652	90.64%	1,400,000	76.06%
SHAW_3	*T. aestivum*	Historical	SDS	Fully ground	2,189,754	2,007,447	91.67%	1,400,000	69.74%
SHMA_1	*T. monococcum*	Historical	SDS	Partially ground	2,222,342	2,038,810	91.74%	1,400,000	68.67%
SHMW_2	*T. monococcum*	Historical	SDS	Fully ground	2,193,055	1,996,183	91.02%	1,400,000	70.13%

Sample ID	Species	Age	Protocol	Grinding Treatment	Total reads	Reads after trimming and collapsing	Percentage of reads collapsed	Sampled reads	Fraction of total reads sampled
SHTA_2	*T. turgidum*	Historical	SDS	Partially ground	2,642,666	2,365,262	89.50%	1,400,000	59.19%
SHTW_3	*T. turgidum*	Historical	SDS	Fully ground	2,515,211	2,279,736	90.64%	1,400,000	61.41%
SMAA_2	*T. aestivum*	Modern	SDS	Partially ground	2,130,405	1,845,844	86.64%	1,400,000	75.85%
SMAW_1	*T. aestivum*	Modern	SDS	Fully ground	2,487,134	2,188,156	87.98%	1,400,000	63.98%
SMMA_1	*T. monococcum*	Modern	SDS	Partially ground	2,752,141	2,439,309	88.63%	1,400,000	57.39%
SMMW_1	*T. monococcum*	Modern	SDS	Fully ground	3,119,800	2,813,222	90.17%	1,400,000	49.77%
SMTA_2	*T. turgidum*	Modern	SDS	Partially ground	2,758,983	2,398,497	86.93%	1,400,000	58.37%
SMTW_1	*T. turgidum*	Modern	SDS	Fully ground	2,954,232	2,568,779	86.95%	1,400,000	54.50%

Supplementary Table 3. Chloroplast assembly information of 1.4 million reads of 48 modern and historical wheat DNA libraries constructed from DNA extracts obtained with four extraction methods.

Sample ID	Species	Age	Protocol	Grinding Treatment	Reads mapped to chloroplast	Reads mapped to chloroplast after removing duplicates	Duplicated reads mapped to the chloroplast	Chloroplast clonality (Duplicate reads / Mapped reads)	Percentage of endogenous chloroplast reads (Mapped Unique/Sampled Reads)
DHAA_1	*T. aestivum*	Historical	DNAzol	Partially ground	3,596	3,536	60	1.7%	0.25%
DHAW_1	*T. aestivum*	Historical	DNAzol	Fully ground	83	82	1	1.2%	0.01%
DHMA_1	*T. monococcum*	Historical	DNAzol	Partially ground	1,756	1,525	231	13.2%	0.11%
DHMW_1	*T. monococcum*	Historical	DNAzol	Fully ground	1,711	1,485	226	13.2%	0.11%
DHTA_1	*T. turgidum*	Historical	DNAzol	Partially ground	59	58	1	1.7%	0.00%
DHTW_1	*T. turgidum*	Historical	DNAzol	Fully ground	43	42	1	2.3%	0.00%
DMAA_1	*T. aestivum*	Modern	DNAzol	Partially ground	1,472	1,444	28	1.9%	0.10%

Sample ID	Species	Age	Protocol	Grinding Treatment	Reads mapped to chloroplast	Reads mapped to chloroplast after removing duplicates	Duplicated reads mapped to the chloroplast	Chloroplast clonality (Duplicate reads / Mapped reads)	Percentage of endogenous chloroplast reads (Mapped Unique/Sampled Reads)
DMAW_1	*T. aestivum*	Modern	DNAzol	Fully ground	2,455	2,420	35	1.4%	0.17%
DMMA_2	*T. monococcum*	Modern	DNAzol	Partially ground	4,919	4,820	99	2.0%	0.34%
DMMW_1	*T. monococcum*	Modern	DNAzol	Fully ground	4,955	4,848	107	2.2%	0.35%
DMTA_2	*T. turgidum*	Modern	DNAzol	Partially ground	1,379	1,356	23	1.7%	0.10%
DMTW_1	*T. turgidum*	Modern	DNAzol	Fully ground	1,434	1,409	25	1.7%	0.10%
PHAA_1	*T. aestivum*	Historical	PTB	Partially ground	1,185	1,163	22	1.9%	0.08%
PHAW_3	*T. aestivum*	Historical	PTB	Fully ground	435	429	6	1.4%	0.03%
PHMA_2	*T. monococcum*	Historical	PTB	Partially ground	1,653	1,588	65	3.9%	0.11%
PHMW_3	*T. monococcum*	Historical	PTB	Fully ground	1,834	1,783	51	2.8%	0.13%
PHTA_2	*T. turgidum*	Historical	PTB	Partially ground	838	827	11	1.3%	0.06%
PHTW_2	*T. turgidum*	Historical	PTB	Fully ground	587	583	4	0.7%	0.04%
PMAA_2	*T. aestivum*	Modern	PTB	Partially ground	485	479	6	1.2%	0.03%
PMAW_1	*T. aestivum*	Modern	PTB	Fully ground	484	477	7	1.4%	0.03%
PMMA_1	*T. monococcum*	Modern	PTB	Partially ground	1,947	1,874	73	3.7%	0.13%
PMMW_1	*T. monococcum*	Modern	PTB	Fully ground	1,578	1,540	38	2.4%	0.11%
PMTA_1	*T. turgidum*	Modern	PTB	Partially ground	550	542	8	1.5%	0.04%
PMTW_1	*T. turgidum*	Modern	PTB	Fully ground	412	410	2	0.5%	0.03%
QHAA_1	*T. aestivum*	Historical	QPP	Partially ground	3,688	3,559	129	3.5%	0.25%
QHAW_1	*T. aestivum*	Historical	QPP	Fully ground	3,164	3,122	42	1.3%	0.22%
QHMA_2	*T. monococcum*	Historical	QPP	Partially ground	8,183	7,937	246	3.0%	0.57%
QHMW_1	*T. monococcum*	Historical	QPP	Fully ground	7,027	6,841	186	2.6%	0.49%
QHTA_2	*T. turgidum*	Historical	QPP	Partially ground	3,259	3,223	36	1.1%	0.23%

Sample ID	Species	Age	Protocol	Grinding Treatment	Reads mapped to chloroplast	Reads mapped to chloroplast after removing duplicates	Duplicated reads mapped to the chloroplast	Chloroplast clonality (Duplicate reads / Mapped reads)	Percentage of endogenous chloroplast reads (Mapped Unique/Sampled Reads)
QHTW_2	*T. turgidum*	Historical	QPP	Fully ground	2,615	2,587	28	1.1%	0.18%
QMAA_1	*T. aestivum*	Modern	QPP	Partially ground	689	682	7	1.0%	0.05%
QMAW_2	*T. aestivum*	Modern	QPP	Fully ground	830	816	14	1.7%	0.06%
QMMA_1	*T. monococcum*	Modern	QPP	Partially ground	1,938	1,865	73	3.8%	0.13%
QMMW_2	*T. monococcum*	Modern	QPP	Fully ground	1,435	1,401	34	2.4%	0.10%
QMTA_2	*T. turgidum*	Modern	QPP	Partially ground	859	848	11	1.3%	0.06%
QMTW_1	*T. turgidum*	Modern	QPP	Fully ground	548	534	14	2.6%	0.04%
SHAA_2	*T. aestivum*	Historical	SDS	Partially ground	938	922	16	1.7%	0.07%
SHAW_3	*T. aestivum*	Historical	SDS	Fully ground	512	507	5	1.0%	0.04%
SHMA_1	*T. monococcum*	Historical	SDS	Partially ground	1,859	1,808	51	2.7%	0.13%
SHMW_2	*T. monococcum*	Historical	SDS	Fully ground	1,050	1,030	20	1.9%	0.07%
SHTA_2	*T. turgidum*	Historical	SDS	Partially ground	3,024	2,981	43	1.4%	0.21%
SHTW_3	*T. turgidum*	Historical	SDS	Fully ground	2,344	2,317	27	1.2%	0.17%
SMAA_2	*T. aestivum*	Modern	SDS	Partially ground	768	756	12	1.6%	0.05%
SMAW_1	*T. aestivum*	Modern	SDS	Fully ground	798	788	10	1.3%	0.06%
SMMA_1	*T. monococcum*	Modern	SDS	Partially ground	1,926	1,877	49	2.5%	0.13%
SMMW_1	*T. monococcum*	Modern	SDS	Fully ground	1,259	1,217	42	3.3%	0.09%
SMTA_2	*T. turgidum*	Modern	SDS	Partially ground	585	577	8	1.4%	0.04%
SMTW_1	*T. turgidum*	Modern	SDS	Fully ground	517	504	13	2.5%	0.04%

Supplementary Table 4. Mitochondria assembly information of 1.4 million reads of 48 modern and historical wheat DNA libraries constructed from DNA extracts obtained with four extraction methods.

Sample ID	Species	Age	Protocol	Grinding Treatment	Reads mapped to mitochondria	Reads mapped to mitochondria after removing duplicates	Duplicated reads mapped to the mitochondria	Mitochondria clonality (Duplicate reads / Mapped reads)	Percentage of endogenous mitochondria reads (Mapped Unique/Sampled Reads)
DHAA_1	*T. aestivum*	Historical	DNAzol	Partially ground	11,764	11,559	205	1.7%	0.83%
DHAW_1	*T. aestivum*	Historical	DNAzol	Fully ground	2,522	2,422	100	4.0%	0.17%
DHMA_1	*T. monococcum*	Historical	DNAzol	Partially ground	5,153	4,660	493	9.6%	0.33%
DHMW_1	*T. monococcum*	Historical	DNAzol	Fully ground	5,145	4,669	476	9.3%	0.33%
DHTA_1	*T. turgidum*	Historical	DNAzol	Partially ground	2,263	2,200	63	2.8%	0.16%
DHTW_1	*T. turgidum*	Historical	DNAzol	Fully ground	980	954	26	2.7%	0.07%
DMAA_1	*T. aestivum*	Modern	DNAzol	Partially ground	8,490	8,251	239	2.8%	0.59%
DMAW_1	*T. aestivum*	Modern	DNAzol	Fully ground	9,883	9,720	163	1.6%	0.69%
DMMA_2	*T. monococcum*	Modern	DNAzol	Partially ground	8,337	8,209	128	1.5%	0.59%
DMMW_1	*T. monococcum*	Modern	DNAzol	Fully ground	9,039	8,851	188	2.1%	0.63%
DMTA_2	*T. turgidum*	Modern	DNAzol	Partially ground	9,173	8,940	233	2.5%	0.64%
DMTW_1	*T. turgidum*	Modern	DNAzol	Fully ground	7,173	7,049	124	1.7%	0.50%
PHAA_1	*T. aestivum*	Historical	PTB	Partially ground	7,269	7,126	143	2.0%	0.51%
PHAW_3	*T. aestivum*	Historical	PTB	Fully ground	4,437	4,365	72	1.6%	0.31%
PHMA_2	*T. monococcum*	Historical	PTB	Partially ground	5,533	5,397	136	2.5%	0.39%
PHMW_3	*T. monococcum*	Historical	PTB	Fully ground	5,574	5,443	131	2.4%	0.39%
PHTA_2	*T. turgidum*	Historical	PTB	Partially ground	6,372	6,271	101	1.6%	0.45%
PHTW_2	*T. turgidum*	Historical	PTB	Fully ground	3,417	3,384	33	1.0%	0.24%
PMAA_2	*T. aestivum*	Modern	PTB	Partially ground	5,042	4,954	88	1.7%	0.35%

PMAW_1	*T. aestivum*	Modern	PTB	Fully ground	5,248	5,162	86	1.6%	0.37%
PMMA_1	*T. monococcum*	Modern	PTB	Partially ground	6,612	6,443	169	2.6%	0.46%
PMMW_1	*T. monococcum*	Modern	PTB	Fully ground	4,715	4,636	79	1.7%	0.33%
PMTA_1	*T. turgidum*	Modern	PTB	Partially ground	8,115	7,920	195	2.4%	0.57%
PMTW_1	*T. turgidum*	Modern	PTB	Fully ground	5,408	5,313	95	1.8%	0.38%
QHAA_1	*T. aestivum*	Historical	QPP	Partially ground	12,218	11,856	362	3.0%	0.85%
QHAW_1	*T. aestivum*	Historical	QPP	Fully ground	7,815	7,717	98	1.3%	0.55%
QHMA_2	*T. monococcum*	Historical	QPP	Partially ground	11,558	11,260	298	2.6%	0.80%
QHMW_1	*T. monococcum*	Historical	QPP	Fully ground	9,157	8,968	189	2.1%	0.64%
QHTA_2	*T. turgidum*	Historical	QPP	Partially ground	9,067	8,909	158	1.7%	0.64%
QHTW_2	*T. turgidum*	Historical	QPP	Fully ground	5,685	5,610	75	1.3%	0.40%
QMAA_1	*T. aestivum*	Modern	QPP	Partially ground	7,050	6,923	127	1.8%	0.49%
QMAW_2	*T. aestivum*	Modern	QPP	Fully ground	7,684	7,528	156	2.0%	0.54%
QMMA_1	*T. monococcum*	Modern	QPP	Partially ground	5,561	5,458	103	1.9%	0.39%
QMMW_2	*T. monococcum*	Modern	QPP	Fully ground	4,120	4,048	72	1.7%	0.29%
QMTA_2	*T. turgidum*	Modern	QPP	Partially ground	7,959	7,821	138	1.7%	0.56%
QMTW_1	*T. turgidum*	Modern	QPP	Fully ground	6,905	6,701	204	3.0%	0.48%
SHAA_2	*T. aestivum*	Historical	SDS	Partially ground	6,222	6,090	132	2.1%	0.44%
SHAW_3	*T. aestivum*	Historical	SDS	Fully ground	4,414	4,340	74	1.7%	0.31%
SHMA_1	*T. monococcum*	Historical	SDS	Partially ground	6,166	6,033	133	2.2%	0.43%
SHMW_2	*T. monococcum*	Historical	SDS	Fully ground	3,466	3,406	60	1.7%	0.24%
SHTA_2	*T. turgidum*	Historical	SDS	Partially ground	6,908	6,813	95	1.4%	0.49%
SHTW_3	*T. turgidum*	Historical	SDS	Fully ground	3,977	3,937	40	1.0%	0.28%
SMAA_2	*T. aestivum*	Modern	SDS	Partially ground	8,133	7,922	211	2.6%	0.57%
SMAW_1	*T. aestivum*	Modern	SDS	Fully ground	8,386	8,177	209	2.5%	0.58%
SMMA_1	*T. monococcum*	Modern	SDS	Partially ground	6,579	6,432	147	2.2%	0.46%

SMMW_1	*T. monococcum*	Modern	SDS	Fully ground	4,049	3,966	83	2.0%	0.28%
SMTA_2	*T. turgidum*	Modern	SDS	Partially ground	7,917	7,714	203	2.6%	0.55%
SMTW_1	*T. turgidum*	Modern	SDS	Fully ground	5,420	5,320	100	1.8%	0.38%

Supplementary Table 5. Nuclear assembly information of 1.4 million reads of 48 modern and historical wheat DNA libraries constructed from DNA extracts obtained with four extraction methods.

Sample ID	Species	Age	Protocol	Grinding Treatment	Reads mapped to nuclear	Reads mapped to nuclear after removing duplicates	Duplicated reads mapped to the nuclear	Nuclear clonality (Duplicate reads / Mapped reads)	Percentage of endogenous nuclear reads (Mapped Unique/Sampled Reads)
DHAA_1	*T. aestivum*	Historical	DNAzol	Partially ground	849,332	846,839	2,493	0.3%	60.49%
DHAW_1	*T. aestivum*	Historical	DNAzol	Fully ground	811,808	804,718	7,090	0.9%	57.48%
DHMA_1	*T. monococcum*	Historical	DNAzol	Partially ground	268,485	259,906	8,579	3.2%	18.56%
DHMW_1	*T. monococcum*	Historical	DNAzol	Fully ground	268,856	260,266	8,590	3.2%	18.59%
DHTA_1	*T. turgidum*	Historical	DNAzol	Partially ground	695,298	692,028	3,270	0.5%	49.43%
DHTW_1	*T. turgidum*	Historical	DNAzol	Fully ground	654,847	647,984	6,863	1.0%	46.28%
DMAA_1	*T. aestivum*	Modern	DNAzol	Partially ground	848,092	837,764	10,328	1.2%	59.84%
DMAW_1	*T. aestivum*	Modern	DNAzol	Fully ground	762,720	759,608	3,112	0.4%	54.26%
DMMA_2	*T. monococcum*	Modern	DNAzol	Partially ground	230,973	228,552	2,421	1.0%	16.33%
DMMW_1	*T. monococcum*	Modern	DNAzol	Fully ground	232,054	229,147	2,907	1.3%	16.37%
DMTA_2	*T. turgidum*	Modern	DNAzol	Partially ground	637,811	631,515	6,296	1.0%	45.11%
DMTW_1	*T. turgidum*	Modern	DNAzol	Fully ground	653,703	648,908	4,795	0.7%	46.35%
PHAA_1	*T. aestivum*	Historical	PTB	Partially ground	825,405	818,958	6,447	0.8%	58.50%
PHAW_3	*T. aestivum*	Historical	PTB	Fully ground	812,110	807,930	4,180	0.5%	57.71%
PHMA_2	*T. monococcum*	Historical	PTB	Partially ground	267,869	262,039	5,830	2.2%	18.72%
PHMW_3	*T. monococcum*	Historical	PTB	Fully ground	265,036	259,630	5,406	2.0%	18.55%

PHTA_2	*T. turgidum*	Historical	PTB	Partially ground	675,466	672,224	3,242	0.5%	48.02%
PHTW_2	*T. turgidum*	Historical	PTB	Fully ground	647,986	644,550	3,436	0.5%	46.04%
PMAA_2	*T. aestivum*	Modern	PTB	Partially ground	868,860	864,145	4,715	0.5%	61.72%
PMAW_1	*T. aestivum*	Modern	PTB	Fully ground	868,216	864,011	4,205	0.5%	61.72%
PMMA_1	*T. monococcum*	Modern	PTB	Partially ground	274,002	268,415	5,587	2.0%	19.17%
PMMW_1	*T. monococcum*	Modern	PTB	Fully ground	270,000	265,155	4,845	1.8%	18.94%
PMTA_1	*T. turgidum*	Modern	PTB	Partially ground	677,670	672,822	4,848	0.7%	48.06%
PMTW_1	*T. turgidum*	Modern	PTB	Fully ground	668,589	664,312	4,277	0.6%	47.45%
QHAA_1	*T. aestivum*	Historical	QPP	Partially ground	845,643	835,155	10,488	1.2%	59.65%
QHAW_1	*T. aestivum*	Historical	QPP	Fully ground	789,103	785,968	3,135	0.4%	56.14%
QHMA_2	*T. monococcum*	Historical	QPP	Partially ground	266,098	263,556	2,542	1.0%	18.83%
QHMW_1	*T. monococcum*	Historical	QPP	Fully ground	274,903	272,070	2,833	1.0%	19.43%
QHTA_2	*T. turgidum*	Historical	QPP	Partially ground	661,527	659,102	2,425	0.4%	47.08%
QHTW_2	*T. turgidum*	Historical	QPP	Fully ground	645,477	643,374	2,103	0.3%	45.96%
QMAA_1	*T. aestivum*	Modern	QPP	Partially ground	886,857	883,277	3,580	0.4%	63.09%
QMAW_2	*T. aestivum*	Modern	QPP	Fully ground	866,128	861,147	4,981	0.6%	61.51%
QMMA_1	*T. monococcum*	Modern	QPP	Partially ground	268,074	262,842	5,232	2.0%	18.77%
QMMW_2	*T. monococcum*	Modern	QPP	Fully ground	267,915	262,812	5,103	1.9%	18.77%
QMTA_2	*T. turgidum*	Modern	QPP	Partially ground	690,672	687,650	3,022	0.4%	49.12%
QMTW_1	*T. turgidum*	Modern	QPP	Fully ground	681,667	678,902	2,765	0.4%	48.49%
SHAA_2	*T. aestivum*	Historical	SDS	Partially ground	821,204	813,904	7,300	0.9%	58.14%
SHAW_3	*T. aestivum*	Historical	SDS	Fully ground	805,699	799,669	6,030	0.7%	57.12%
SHMA_1	*T. monococcum*	Historical	SDS	Partially ground	264,923	259,793	5,130	1.9%	18.56%
SHMW_2	*T. monococcum*	Historical	SDS	Fully ground	255,974	250,918	5,056	2.0%	17.92%
SHTA_2	*T. turgidum*	Historical	SDS	Partially ground	630,360	626,474	3,886	0.6%	44.75%
SHTW_3	*T. turgidum*	Historical	SDS	Fully ground	598,925	595,082	3,843	0.6%	42.51%

SMAA_2	*T. aestivum*	Modern	SDS	Partially ground	871,998	865,075	6,923	0.8%	61.79%
SMAW_1	*T. aestivum*	Modern	SDS	Fully ground	864,565	858,821	5,744	0.7%	61.34%
SMMA_1	*T. monococcum*	Modern	SDS	Partially ground	265,663	260,452	5,211	2.0%	18.60%
SMMW_1	*T. monococcum*	Modern	SDS	Fully ground	262,710	257,368	5,342	2.0%	18.38%
SMTA_2	*T. turgidum*	Modern	SDS	Partially ground	677,009	672,800	4,209	0.6%	48.06%
SMTW_1	*T. turgidum*	Modern	SDS	Fully ground	677,556	673,840	3,716	0.5%	48.13%

Supplementary Table 6. Chloroplast and mitochondria assembly information of four ancient wheat DNA libraries constructed from DNA extracts obtained with four extraction methods.

Sample ID	Species	Protocol	Total reads	Reads after trimming and collapsing	Percentage of reads collapsed	Reads mapped to chloroplast	Reads mapped to mitochondria
ACAD11290	*T. aestivum*	DNAzol	8,893,019	8,324,624	93.61%	0	2
ACAD11290	*T. aestivum*	PTB	8,310,657	7,993,066	96.18%	0	7
ACAD11290	*T. aestivum*	QPP	6,370,977	5,541,512	86.98%	5	107
ACAD11290	*T. aestivum*	SDS	6,500,510	6,148,750	94.59%	0	2

Chapter 4

Target enrichment uncovers the genetic variability of historical polyploid wheat species from Georgia

Manuscript prepared for publication

Statement of authorship

Title of Paper	Target enrichment uncovers the genetic variability of historical polyploid wheat species from Georgia
Publication Status	☐ Published ☐ Accepted for Publication ☐ Submitted for Publication ☑ Unpublished and Unsubmitted w ork w ritten in manuscript style
Publication Details	Manuscript prepared for submission to the *Plant Biotechnology Journal*

Principal Author

Name of Principal Author (Candidate)	Oscar Estrada		
Contribution to the Paper	Contributed to the conception and experimental design, performed DNA extractions, designed hybridisation captured arrays, constructed Illumina DNA sequencing libraries, performed bioinformatics data analyses, performed phylogenetic analyses, interpreted results, wrote and edited the manuscript.		
Overall percentage (%)	85		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	08/03/2019

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);
ii. permission is granted for the candidate in include the publication in the thesis; and
iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	James Breen		
Contribution to the Paper	Contributed to the conception and experimental design, collected samples, designed hybridisation captured arrays, advised the data analyses, reviewed and edited the manuscript.		
Signature		Date	08/03/2019

Name of Co-Author	Stephen M Richards		
Contribution to the Paper	Contributed to the conception and experimental design, advised laboratory analyses, reviewed and edited the manuscript.		
Signature		Date	08/03/2015

Please cut and paste additional co-author panels here as required.

Name of Co-Author	Alan Cooper		
Contribution to the Paper	Contributed to the conception and experimental design, reviewed and edited the manuscript.		
Signature		Date	8/3/19

Target enrichment uncovers the genetic variability of historical polyploid wheat species from Georgia

Oscar Estrada[1], James Breen[2], Stephen M Richards[1], Alan Cooper[1]

[1]Australian Centre for Ancient DNA, University of Adelaide, North Terrace, South Australia 5005, Australia.
[2]Robinson Research Institute, University of Adelaide, SA, 5005, Australia.

Abstract

Wheat species have had a long history of domestication and selective breeding starting around 12,000 years ago. The decline in genetic diversity associated with domestication has been accelerated by the intensive breeding behind current commercial wheat varieties. The genetic diversity lost from modern commercial breeds might potentially still be preserved in historical specimens and wild relatives. However, the large size and complexity of the genomes of wheat species have limited the ability of high-throughput sequencing (HTS) assays to analyse them. In this study, we develop two hybridisation capture arrays to study plastid and nuclear genetic diversity within a panel of historical samples (60-100-years-old) of polyploid wheat species (*Triticum aestivum* L., *T. timopheevii* , and *T. turgidum*) from Georgia, a region with a wide array of native varieties. The plastid hybridisation capture array recovered the chloroplast genomes of all three species and enabled the phylogenetic reconstruction of their history. Additionally, by applying a Bayesian approach, we demonstrated the ability of the nuclear capture array to recover a large number of SNPs (20,244 SNPs) in a set of 81 genes (224 Kb) related to cell transport, domestication traits, and disease resistance. The newly generated information in this study will provide a resource for the analysis of genetic resources of wheat and future genomic and evolutionary studies.

Introduction

Wheat (*Triticum* spp.) have been an important staple food for the human population since the rise of modern agriculture (Salamini *et al.*, 2002; Brown *et al.*, 2009; Kilian *et al.*, 2010; Zohary *et al.*, 2013). Domestication of wheat started around 12,000 years ago during the "Neolithic Revolution" when humans began the transition from a hunter-gathering lifestyle to a sedentary and agriculture-based society (Zohary *et al.*, 2013). All wheat species were domesticated in the Fertile Crescent (Figure 1), a region of the Near East comprising the valleys of the Tigris, Euphrates and Jordan rivers (Salamini *et al.*, 2002; Feuillet *et al.*, 2008; Zohary *et al.*, 2013).

Figure 1. The Fertile Crescent (highlighted in yellow), a region of the Near East comprising the valleys of the Tigris, Euphrates and Jordan rivers, was the principal centre of wheat domestication and other cereals. Image modified from a Google Earth screenshot.

Early farmers made use of wild diploid wheat species, however as agriculture evolved, wild crops were gradually substituted with domesticated diploid and polyploid wheat varieties due to their increase in yield and ability to grow in warmer areas. Wild diploid and tetraploid species were altered substantially according to human-selected phenotypic traits - leading to domesticated forms such as Einkorn (*T. monococcum* L.; diploid genome AA), Emmer (*T. turgidum* L. subsp. *dicoccum* (Schrank ex Schübl.) Thell; tetraploid genome AABB) and Timopheevii wheat (*T. timopheevii* (Zhuk.) Zhuk; tetraploid genome AAGG) (Salamini *et al.*, 2002). It is hypothesised that the timopheevii wheat *T. timopheevii* (Zhuk.) Zhuk. was

domesticated from the wild from *T. timopheevii* (Zhuk.) Zhuk. subsp. armeniacum (Jakubz.) Slageren in the eastern part of the Fertile Crescent (Kilian *et al.*, 2010; Zohary *et al.*, 2013). Modern cultivated bread wheat (*T. aestivum* L.; hexaploid genome AABBDD) is believed to have arisen from a natural whole-genome hybridisation of cultivated tetraploid wheat (*T. turgidum*) and the diploid wild species *Aegilops tauschii* Coss. (diploid genome DD) approximately 10,000 years ago (Figure 2) (Salamini *et al.*, 2002; Marcussen *et al.*, 2014).

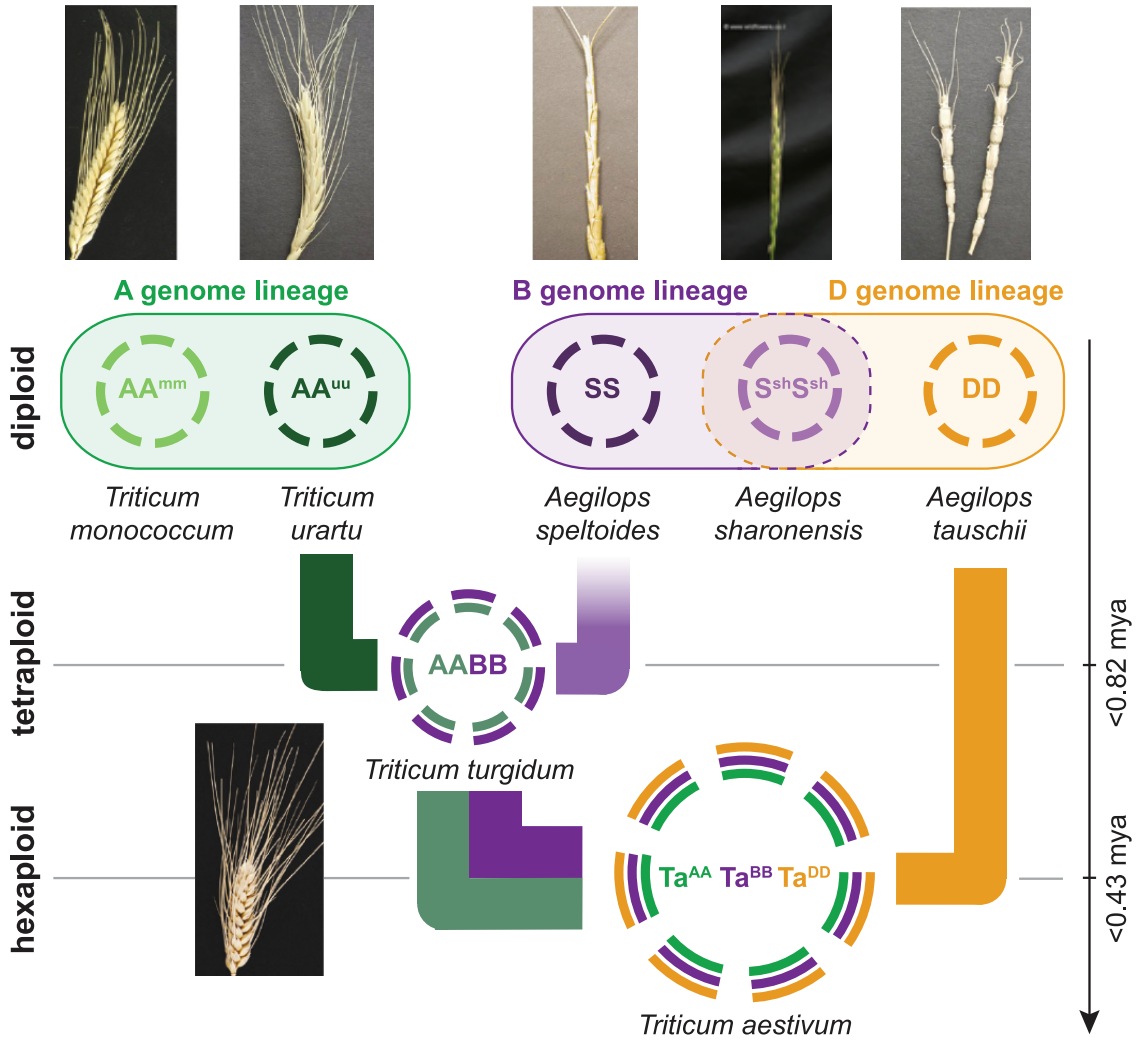


Figure 2. Diagrammatic representation of the relationships between wheat genomes, polyploidisation history, and genealogy. Modified from (Mayer *et al.*, 2014). AA^{mm} denotes the diploid genome of *T. monococcum* (einkorn wheat). AA^{uu} denotes the diploid genome of *T. urartu.* SS denotes the diploid genome of *A. speltoides.* $S^{sh}S^{sh}$ denotes the diploid genome of *A. sharonensis.* DD denotes the diploid genome of *A. tauschii.* AABB denotes the tetraploid genome of *T. turgidum* (emmer wheat). $Ta^{AA}Ta^{BB}Ta^{DD}$ denotes the hexaploid genome of *T. aestivum* (bread wheat).

The most important traits selected during wheat domestication were the free-threshing state (whereby seeds are released from the rachis at threshing) and the brittle rachis, a trait that allowed stable capture of seeds without the head shattering on the ground (Salamini *et al.*, 2002; Kilian *et al.*, 2010). Further changes during domestication and breeding of wheat involved an increase in seed size, kernel row type, plant height, grain hardness, tillering, seed dormancy, photoperiod, vernalisation (dependence of plants on exposure to cold temperatures in order to flower), heading date, nutritional content, and enhanced culinary chemistry (Kilian *et al.*, 2010; Cavanagh *et al.*, 2013).

In addition to agronomically-influenced traits, the spread of the domesticated wheat out of the Fertile Crescent required adaptation to new environments supported by favourable alleles at critical genetic loci (Kilian *et al.*, 2010), which in some cases represent local adaptation over thousands of years (Allaby, 2014). Transporter genes are thought to be particularly influential in this adaptation because they are involved in multiple physiological processes that include cellular homeostasis, metal detoxification and importantly - the absorption, transport and storage of water and micronutrients (Tripathi & Müller, 2015). When faced with abiotic stress such as salinity or drought conditions, these genes can become one of the main sources of phenotypic change (Wang *et al.*, 2003; Nevo & Chen, 2010). In wheat, for instance, high-affinity potassium transporter genes (HKT) have been associated with salinity and drought tolerance (Mondini *et al.*, 2012; Munns *et al.*, 2012; Zamani Babgohari *et al.*, 2013; Yang *et al.*, 2014; Byrt *et al.*, 2014). In the same way, plant aluminium-activated malate transporters (ALMT) have been related to wheat tolerance of acidic soils with high concentrations of aluminium (Sasaki *et al.*, 2004; Raman *et al.*, 2005, 2010; Ryan *et al.*, 2010; Liu *et al.*, 2015; Ramesh *et al.*, 2015).

Domestication and cultivation have also had a major role in the evolution of the wheat genome due to intense selection and breeding. One of the most intensive selection and breeding processes occurred 60 years ago during the “Green Revolution”. The “Green Revolution” was a technology package that doubled crop yields in developing nations in the late 1960s by introducing high-yielding varieties of cereals, especially dwarf photoperiod-insensitive wheat and rice varieties, capable of responding to chemical fertilisers and pesticides (Khush, 2001; Hedden, 2003). Since the Green Revolution, landraces and local varieties have been progressively replaced by modern high-yielding varieties leading to the reduction of crop diversity which could jeopardise the continued ability to improve crops (Reif *et al.*, 2005; Doebley *et al.*, 2006; Bevan *et al.*, 2017). Therefore, the genetic diversity of wild relatives and

historical landraces of wheat have the potential to contribute valuable information for wheat breeding programs and further understanding of the evolution and domestication of this cereal.

Georgia is one of the most important centres of origin and diversity for domesticated wheat. Out of about 20 early wheat species known in the world, 14 were cultivated in Georgia, and five of them were endemic to Georgia (Mosulishvili *et al.*, 2017). (Mosulishvili *et al.*, 2017). The four wheat genomes important in domesticated breeds (A, B, D and G) occur in Georgia, and wheat species containing the G genome *T. timopheevii* and *T. zhukovskyi* Menabde & Ericzjan were domesticated in western Georgia (Zohary *et al.*, 2013). Although *T. timopheevii* has been used to broaden the genetic diversity of cultivated wheat *T. aestivum* (Timonova *et al.*, 2013), most of the Georgian wheat genetic diversity is still yet to be assessed and remains a source of untapped potential.

The recent application of high-throughput sequencing (HTS) technologies has allowed the genomic characterisation of the diploid wheat *T. urartu* Tumanian ex Gandilyan, the tetraploid wheat *T. turgidum* L. subsp. *dicoccoides* (Korn. ex Asch. & Graebn.) Thell, and the hexaploid wheat *T. aestivum* (Mayer *et al.*, 2014; Avni *et al.*, 2017; Appels *et al.*, 2018; Ling *et al.*, 2018). However, the investigation of the genetic diversity of wheat species and historical landraces remains challenging owing to the different levels of ploidy (2x, 4x, 6x), large size of the genomes (4-16 Gbp), and high content of repetitive elements (80-90%) (Mayer *et al.*, 2014; Avni *et al.*, 2017; Appels *et al.*, 2018; Ling *et al.*, 2018). Hence, targeted gene methods such as hybridisation capture have been applied to both reduce the genomic complexity and sequence specific molecules of interest (Hodges *et al.*, 2007; Gnirke *et al.*, 2009; Cronn *et al.*, 2012). Selective enrichment by hybridisation capture has been shown to be highly efficient in tetraploid and hexaploid wheat species at enriching thousands of targets in parallel (Saintenac *et al.*, 2011; Winfield *et al.*, 2012; Wang *et al.*, 2014).

In this study, we used hybridisation capture and high-throughput sequencing to investigate the genetic variability of historical samples (60-100-years-old) of polyploid wheat species (*T.* cf. *aestivum* 6x, *T.* cf. *timopheevii* 4x, and *T.* cf. *turgidum* 4x) from Georgia. While historical samples may contain degraded DNA, their genetic diversity offers significant insight into early advantageous alleles. We designed a plastid RNA hybridisation capture array to recover full chloroplast genomes and reconstruct phylogenetic trees, and a plant domestication gene RNA capture array that contains probes for 185 nuclear genes, including transporter, domestication,

and disease resistance-related genes, to evaluate autosomal variation and relate it to plant domestication and adaptation to changing environments.

Materials and Methods

Plant material and DNA extraction

DNA was extracted from historical wheat seeds from Georgia the were approximately 60-100-years-old and morphologically identified as *T.* cf. *aestivum*, *T.* cf. *timopheevii*, and *T.* cf. *turgidum* (Table 1). The GRIN taxonomy system was used to denote wheat species (https://npgsweb.ars-grin.gov). All samples were provided by the Georgian National Museum and processed in a standard modern DNA laboratory at the Australian Centre for Ancient DNA, University of Adelaide.

One seed of each accession was selected for the isolation of genomic DNA. Seed coats were removed using a clean scalpel blade, and naked seeds were macerated in a 2 ml tubes using a ceramic ball bearing in an MM300 mechanical Mixer Mill (Retsch, Germany) for 5 min at 4000 rpm. Genomic DNA was isolated from single ground seeds using DNeasy Plant Mini Kit (Qiagen, USA) following manufacturer directions. Purified DNA was quantified with a fluorometer Qubit 2.0 (Life Technologies, Germany) using the double-stranded DNA broad range (dsDNA BR) kit (ThermoFisher Scientific, Germany) - quantification range 2-1000 ng/µL, and DNA quality was evaluated by electrophoresis in 2% agarose gels.

Table 1. List of historical (60-100-years-old) wheat samples from Georgia analysed in this study. The accession numbers of the Australian Centre for Ancient DNA (ACAD) biological collection are provided. The GRIN taxonomy system is used to denote wheat species (https://npgsweb.ars-grin.gov).

Species	Genome	ACAD accession number
Triticum cf. *aestivum*	AABBDD	ACAD16052
Triticum cf. *timopheevii*	AAGG	ACAD16053
Triticum cf. *turgidum*	AABB	ACAD16054
Triticum cf. *turgidum*	AABB	ACAD16055

Capture probe design

Customised sets of RNA probes were designed to enrich wheat plastid and nuclear sequences using MYcroarray MYbaits Custom Target Capture Kits (arbor biosciences, USA). An initial chloroplast set of 20,000 baits of 80 nucleotides with 4x tiling density was designed based on the chloroplast sequences of wheat *T. aestivum* (NCBI accession: NC_002762.1) and barley *Hordeum vulgare* (NCBI accession: NC_008590.1). A second nuclear set of 20,000 probes of 70 nucleotides with 4x tiling density was designed based on the sequences of 185 wheat and barley nuclear genes, including transporter, domestication, and disease resistance related genes (Table 2). *RepeatMasker* (http://www.repeatmasker.org) was used to mask interspersed repeats and low complexity DNA sequences against the TREP Triticeae repeats database (Wicker *et al.*, 2002) during the bait design.

DNA library preparation

DNA extracts were sheared with a Covaris ultrasonicator S220 (Covaris, USA) to a mean fragment size of 200 bp following the manufacturer settings. For each sample, 50 ng of fragmented DNA was used to construct double-stranded DNA Illumina sequencing libraries following a protocol modified from Meyer & Kircher, 2010. The fragmented DNA was end-repaired using T4 Polynucleotide Kinase PNK (NEB, USA) and T4 DNA polymerase (NEB, USA). After end-repair truncated Illumina adapters with seven bp internal barcodes were ligated to the DNA fragments using T4 ligase (ThermoFisher Scientific, Germany). Lastly, the Illumina molecules were amplified by PCR for eight cycles using Illumina short primer. Each library was split into eight PCR replicates and individual PCR reactions (25 µl) consisted of 5 µl of library DNA, 1X High Fidelity PCR buffer, 2 mM $MgSO_4$, 0.2 mM dNTPs, 0.2 mM of each primer, and 0.1 µl Platinum Taq DNA Polymerase High Fidelity (5 U/µL). The following thermocycling conditions were utilised: 94 ˚C for 2 min, 8 cycles of 94 ˚C for 10 s, 58 ˚C for 30 s, and 68 ˚C for 40 s. Amplified libraries were purified using the Ampure XP magnetic beads system (Beckman Coulter, USA), and quantified as before.

Table 2. List of 185 wheat and barley nuclear genes, including transporter, domestication, and disease resistance related genes selected for target enrichment. NCBI gene accession numbers are provided. Genes with accession numbers marked with (*) are taken from a private repository.

ID	Species	Accession number	Gene category	Gene family	Gene name	Length (bp)
001	*Hordeum vulgare*	AF250933	Disease resistance	Germin-like protein	HvGerA	1,243
002	*Hordeum vulgare*	AF250934	Disease resistance	Germin-like protein	HvGerB	2,689
003	*Hordeum vulgare*	AF250935	Disease resistance	Germin-like protein	HvGerF	2,798
004	*Hordeum vulgare*	AF250936	Disease resistance	Germin-like protein	HvGerD	1,272
005	*Hordeum vulgare*	AF493980	Disease resistance	Germin-like 8	HvGLP8	2,741
006	*Hordeum vulgare*	AJ310534	Disease resistance	Germin-like protein	HvGLP2	2,028
007	*Triticum aestivum*	AY939880	Disease resistance	Powdery mildew resistance	PM3A	4,448
008	*Triticum aestivum*	AY939881	Disease resistance	Powdery mildew resistance	PM3D	8,816
009	*Hordeum vulgare*	DQ324800	Disease resistance	Germin-like protein	GerF	1,174
010	*Hordeum vulgare*	DQ647625	Disease resistance	Germin-like protein	GER6a	1,581
011	*Aegilops tauschii*	DQ655791	Disease resistance	Multidrug resistance-associated protein	MRP1	3,024
012	*Triticum aestivum*	EU385606	Disease resistance	Homeobox-like resistance	HLRG	2,396
013	*Hordeum vulgare*	EU883790	Disease resistance	Steptoe stem rust resistance protein	Rpg5	5,280
014	*Hordeum vulgare*	FJ156744	Disease resistance	Blufensin	Bln1-3	1,211
015	*Hordeum vulgare*	FJ156749	Disease resistance	Blufensin	Bln2	605
016	*Triticum aestivum*	FJ815160	Disease resistance	Powdery mildew resistance-related	RGA2	1,008
017	*Triticum aestivum*	FJ815161	Disease resistance	Powdery mildew resistance-related	RGA2	1,008
018	*Triticum aestivum*	FJ831681	Disease resistance	Peroxidase precursor	TaPrx	1,374
019	*Hordeum vulgare*	HvPPT1	Disease resistance	Phosphopantetheinyl Transferase	HvPPT1	4,345

ID	Species	Accession number	Gene category	Gene family	Gene name	Length (bp)
020	*Triticum monococcum*	JF261156	Disease resistance	Powdery mildew resistance protein	Mla1	3,928
021	*Triticum aestivum*	JF439306	Disease resistance	Serine/threonine protein kinase	Stpk-A	2,381
022	*Triticum aestivum*	JF439307	Disease resistance	Serine/threonine protein kinase	Stpk-D	2,432
023	*Triticum aestivum*	JF439308	Disease resistance	Serine/threonine protein kinase	Stpk-B	2,584
024	*Triticum aestivum*	JF439309	Disease resistance	Serine/threonine protein kinase	Stpk-A(B)	999
025	*Triticum aestivum*	JF439310	Disease resistance	Serine/threonine protein kinase	Stpk-D	1,016
026	*Hordeum vulgare*	JN375539	Disease resistance	Stem rust resistance	HvRin4	3,890
027	*Triticum aestivum*	KF572030	Disease resistance	Powdery mildew resistance protein	Pm8	7,113
028	*Triticum aestivum*	KF572031	Disease resistance	Powdery mildew resistance protein	Pm3-1B	4,527
029	*Triticum aestivum*	M63223.1	Disease resistance	Wheat germin 9f-2.8 gene	Germin-9f-2.8	2,822
030	*Triticum aestivum*	M63224.1	Disease resistance	Wheat germin 9f-3.8 gene	Germin-9f-3.8	3,761
031	*Triticum aestivum*	AB181991.1	Domestication	mRNA for actin	ACT-1	1,163
032	*Triticum aestivum*	AB795034.1	Domestication	High-molecular-weight glutenin subunit	Glu-A1	2,505
033	*Triticum turgidum*	AY945221.1	Domestication	Grain softness protein	Gsp-1	467
034	*Triticum aestivum*	BK006460.1	Domestication	High-molecular-weight glutenin Dx2	Glu-D1-1a	2,508
035	*Triticum aestivum*	BK006773.1	Domestication	High-molecular-weight glutenin Bx7	Glu-B1	2,388
036	*Triticum urartu*	DQ269895.1	Domestication	Grain softness protein	Gsp-1	495
037	*Aegilops speltoides*	DQ269896.1	Domestication	Grain softness protein	Gsp-1	495
038	*Aegilops sharonensis*	DQ269908.1	Domestication	Grain softness protein	Gsp-1	492
039	*Triticum aestivum*	EF620907.1	Domestication	Puroindoline a	Pina-D1	447
040	*Hordeum vulgare*	HM222644	Domestication	Cellulose synthase	CesA4	5,700

ID	Species	Accession number	Gene category	Gene family	Gene name	Length (bp)
041	*Hordeum vulgare*	HvAP2*	Domestication	APETALA2-like transcription factor	HvAP2	1,015
042	*Hordeum vulgare*	HvCO1*	Domestication	CONSTANS-like protein	HvCO1	3,908
043	*Hordeum vulgare*	HvCO2*	Domestication	CONSTANS-like protein	HvCO2	2,097
044	*Hordeum vulgare*	HvFT1*	Domestication	Flowering like protein	HvFT1	2,804
045	*Hordeum vulgare*	HvFT2*	Domestication	Flowering like protein	HvFT2	4,008
046	*Hordeum vulgare*	HvFT3*	Domestication	Flowering like protein	HvFT3	1,853
047	*Hordeum vulgare*	HvFT4*	Domestication	Flowering like protein	HvFT4	2,419
048	*Hordeum vulgare*	HvFT5.1*	Domestication	Flowering like protein	HvFT5	1,629
049	*Hordeum vulgare*	HvFT5.2*	Domestication	Flowering like protein	HvFT5	1,407
050	*Hordeum vulgare*	HvGA20ox2*	Domestication	Gibberellin 20 oxidase 2-like	HvGA20ox2	783
051	*Hordeum vulgare*	HvGI*	Domestication	Gigantea-like protein	HvGI	7,393
052	*Hordeum vulgare*	HvPhyA*	Domestication	Phytochrome	HvPhyA	1,531
053	*Hordeum vulgare*	HvPhyB*	Domestication	Phytochrome	HvPhyB	1,336
054	*Hordeum vulgare*	HvPhyC*	Domestication	Phytochrome	HvPhyC	3,842
055	*Hordeum vulgare*	HvSOC1*	Domestication	Suppressor of overexpression of constans 1	HvSOC1	1,050
056	*Hordeum vulgare*	HvTFL1*	Domestication	Terminal flower-like protein	HvTFL1	1,410
057	*Hordeum vulgare*	HvVRN1*	Domestication	Vernalisation	HvVRN1	17,049
058	*Hordeum vulgare*	HvZCCTa*	Domestication	Vernalisation	HvZCCTa	2,407
059	*Hordeum vulgare*	HvZCCTb*	Domestication	Vernalisation	HvZCCTb	2,313
060	*Triticum aestivum*	JN626222.1	Domestication	Puroindoline b	Pinb-D1	447
061	*Aegilops tauschii*	JX173947.1	Domestication	High molecular weight glutenin subunit 1Dx	Glu-D1-1	2,568

ID	Species	Accession number	Gene category	Gene family	Gene name	Length (bp)
062	*Aegilops tauschii*	JX173953.1	Domestication	High molecular weight glutenin subunit 1Dy	Glu-D1-2	1,980
063	*Triticum aestivum*	AY319478.1	House-keeping	Cytosolic glyceraldehyde-3-phosphate dehydrogenase	GAPDH	1,042
064	*Hordeum vulgare*	AB447482	Transporter	HvLsi1 mRNA for silicon transporter	HvLsi1	1,368
065	*Hordeum vulgare*	AB447483	Transporter	HvLsi2 mRNA for silicon transporter	HvLsi2	1,969
066	*Hordeum vulgare*	AB447484	Transporter	HvLsi6 mRNA for silicon transporter	HvLsi6	1,588
067	*Triticum aestivum*	AB968529.1	Transporter	mRNA for boron transporter	TaBOR1.1	2,139
068	*Triticum aestivum*	AB968530.1	Transporter	mRNA for boron transporter	TaBOR1.2	2,139
069	*Triticum aestivum*	AB968531.1	Transporter	mRNA for boron transporter	TaBOR1.3	2,139
070	*Triticum aestivum*	AF015523	Transporter	Low-affinity cation transporter	TaLCT1	1,984
071	*Hordeum vulgare*	AF075270	Transporter	High affinity sulphate transporter	HvST1	993
072	*Triticum aestivum*	AF288688	Transporter	High affinity nitrate transporter	TaNRT2	1,796
073	*Triticum aestivum*	AF332214.1	Transporter	High affinity nitrate transporter mRNA	TaNRT2.1	1,750
074	*Hordeum vulgare*	AF543197	Transporter	Phosphate transporter	HvPht1-1	3,202
075	*Hordeum vulgare*	AF543198	Transporter	Phosphate transporter	HvPht1-6	2,876
076	*Aegilops tauschii*	AJ238245	Transporter	High affinity sulphate transporter	AtSt2	5,329
077	*Hordeum vulgare*	AJ297886	Transporter	High affinity potassium transporter	HvHak1	1,205
078	*Hordeum vulgare*	AJ300161	Transporter	High affinity potassium transporter	HvHak4	1,183
079	*Hordeum vulgare*	AB120306	Transporter	Tonoplast ABC transporter	ABC	966
080	*Triticum aestivum*	AF408845	Transporter	Sucrose transporter	TaSUT1D	6,582
081	*Hordeum vulgare*	AJ310644	Transporter	Plasma membrane proton ATPase	HvPPA1	808
082	*Triticum urartu*	AJ512492.1	Transporter	Sulphate transporter	st1.1b	3,856

ID	Species	Accession number	Gene category	Gene family	Gene name	Length (bp)
083	*Aegilops speltoides*	AJ512815.1	Transporter	Sulphate transporter	st1.1a	4,133
084	*Aegilops speltoides*	AJ512816.1	Transporter	Sulphate transporter	st1.1a	3,876
085	*Triticum aestivum*	AJ512818.1	Transporter	Sulphate transporter	stA1.1b	3,842
086	*Triticum aestivum*	AJ512820.1	Transporter	Sulphate transporter, genome D	stD1.1b	3,881
087	*Triticum aestivum*	AJ512821.1	Transporter	Sulphate transporter, genome B	stB1.1a	4,071
088	*Triticum aestivum*	AJ512822.1	Transporter	Sulphate transporter, genome D	stD1.1a	4,129
089	*Hordeum vulgare*	AM000057	Transporter	High-affinity sodium transporter	HvHKT1	2,155
090	*Hordeum vulgare*	AM182059	Transporter	Zinc transporter	ZIP7	1,165
091	*Hordeum vulgare*	AM286795	Transporter	Putative zinc transporter	mtp1	1,404
092	*Triticum aestivum*	AM747385.1	Transporter	Sulphate transporter	ST3.5	894
093	*Triticum aestivum*	AY053452.1	Transporter	High-affinity nitrate transporter	NRT2.3	1,744
094	*Hordeum vulgare*	AY187019	Transporter	Phosphate transporter	HvPT2	2,785
095	*Hordeum vulgare*	AY187021	Transporter	Phosphate transporter	HvPT5	3,187
096	*Hordeum vulgare*	AY187022	Transporter	Phosphate transporter	HvPT7	3,700
097	*Hordeum vulgare*	AY187023	Transporter	Phosphate transporter	HvPT8	2,276
098	*Hordeum vulgare*	AY187024	Transporter	Phosphate transporter	HvPT4	6,561
099	*Hordeum vulgare*	AY189896	Transporter	High-affinity sulphate transporter	HvST1	5,480
100	*Triticum aestivum*	AY293827.1	Transporter	Phosphate transporter	PT2-1	2,094
101	*Triticum aestivum*	AY390355.1	Transporter	Ammonium transporter	AMT1	1,852
102	*Triticum aestivum*	AY428038.1	Transporter	Ammonium transporter	Amt2.1	1,521
103	*Triticum aestivum*	AY525637.2	Transporter	Ammonium transporter	Amt1;1	2,036

ID	Species	Accession number	Gene category	Gene family	Gene name	Length (bp)
104	*Triticum aestivum*	AY525638.1	Transporter	Ammonium transporter	Amt1;2	821
105	*Triticum aestivum*	AY587264.1	Transporter	Low affinity nitrate transporter	NRT1.2	1,783
106	*Triticum aestivum*	AY587265.1	Transporter	Low affinity nitrate transporter	NRT1.1	1,811
107	*Triticum aestivum*	AY766367	Transporter	PDR-like ABC transporter	ABC	2,584
108	*Triticum aestivum*	DD412189	Transporter	Molybdate Transporter	MOT	778
109	*Triticum aestivum*	DQ009003	Transporter	High-affinity potassium uptake transporter	TaHKT1	1,981
110	*Triticum aestivum*	DQ072262	Transporter	Aluminium-activated malate transporter	ALMT1	3,951
111	*Aegilops tauschii*	DQ072271	Transporter	Aluminium-activated malate transporter	ALMT1	3,966
112	*Triticum aestivum*	DQ323065	Transporter	Cadmium tolerance factor mRNA	ABC	2,998
113	*Triticum aestivum*	DQ490131.1	Transporter	Putative zinc transporter	ZIP3	1,083
114	*Triticum aestivum*	DQ490132.1	Transporter	Putative zinc transporter	ZIP5	1,131
115	*Triticum aestivum*	DQ490133.1	Transporter	Putative zinc transporter	ZIP1	1,068
116	*Triticum aestivum*	DQ490134.1	Transporter	Putative zinc transporter	ZIP7	1,161
117	*Triticum monococcum*	DQ646339	Transporter	Cation transporter	TmHKT8	3,120
118	*Hordeum vulgare*	DQ912169	Transporter	Putative sodium transporter	HKT1;5	3,161
119	*Triticum monococcum*	EF062819.1	Transporter	Putative sodium transporter	HKT7-A2	1,665
120	*Triticum monococcum*	EF062820	Transporter	Putative sodium transporter	HKT7-A1	4,254
121	*Aegilops tauschii*	EF424085.1	Transporter	Aluminium-activated malate transporter	ALMT2	1,368
122	*Aegilops tauschii*	EF424086.1	Transporter	Aluminium-activated malate transporter	ALMT3	1,410
123	*Triticum aestivum*	EF599631	Transporter	Salt tolerance	TaSTRG	879
124	*Hordeum vulgare*	EF660436	Transporter	Boron transporter	Bot1.b/Bot1(Dp).b	6,396

ID	Species	Accession number	Gene category	Gene family	Gene name	Length (bp)
125	*Triticum urartu*	EU183487	Transporter	Calcium-dependent protein kinase	CPK1	4,118
126	*Aegilops tauschii*	EU183488	Transporter	Calcium-dependent protein kinase	CPK4	3,757
127	*Aegilops tauschii*	EU183489	Transporter	Calcium-dependent protein kinase	CPK5	3,552
128	*Aegilops tauschii*	EU183490	Transporter	Calcium-dependent protein kinase	CPK7	3,175
129	*Aegilops tauschii*	EU183491	Transporter	Calcium-dependent protein kinase	CPK8	3,492
130	*Triticum aestivum*	EU220225.1	Transporter	Boron transporter	TaBOR2	2,338
131	*Hordeum vulgare*	EU223365.1	Transporter	Putative boron transporter	HvBOR2	2,228
132	*Hordeum vulgare*	FJ208991	Transporter	Zinc transporter	ZIP3	1,089
133	*Hordeum vulgare*	FJ208992	Transporter	Zinc transporter	ZIP5	1,053
134	*Hordeum vulgare*	FJ208993	Transporter	Zinc transporter	ZIP8	1,080
135	*Triticum turgidum*	FN401377	Transporter	High affinity potassium transporter	htk1	525
136	*Aegilops tauschii*	FN401378	Transporter	High affinity potassium transporter	htk1	529
137	*Triticum aestivum*	FN432835.1	Transporter	Sulphate transporter	ST3.1	2,032
138	*Triticum aestivum*	FN599528.1	Transporter	Sulphate transporter	st3.2	959
139	*Triticum aestivum*	FN601348.1	Transporter	Putative sulphate/molybdate transporter	ST5.1	727
140	*Triticum aestivum*	FN601349	Transporter	Putative sulphate/molybdate transporter	ST5.2	999
141	*Hordeum vulgare*	FR873736	Transporter	Heavy metal ATPase 1	hma1	9,488
142	*Hordeum vulgare*	GL19623*	Transporter	Aluminium-activated malate transporter	HvALMT	1,767
143	*Hordeum vulgare*	GL7606*	Transporter	Aluminium-activated malate transporter	HvALMT	369
144	*Triticum aestivum*	GQ916634	Transporter	ADP-glucose brittle-1 transporter precursor	BT1	2,234
145	*Aegilops crassa*	GQ916635	Transporter	ADP-glucose brittle-1 transporter precursor	BT1	2,328

ID	Species	Accession number	Gene category	Gene family	Gene name	Length (bp)
146	*Triticum turgidum*	GU056165.1	Transporter	High affinity potassium transporter	hkt1	227
147	*Triticum turgidum*	GU825945.1	Transporter	Zinc transporter	ZIP1	1,083
148	*Triticum aestivum*	HF545004	Transporter	Low affinity nitrate transporter	NPF6.7	834
149	*Triticum aestivum*	HM803114.1	Transporter	Silicon transporter protein	si1	888
150	*Hordeum vulgare*	HQ696002	Transporter	High affinity potassium transporter	HKT4	2,473
151	*Hordeum vulgare*	HQ696004	Transporter	High affinity potassium transporter	HKT1	2,319
152	*Hordeum vulgare*	HV555321	Transporter	Mugineic Acid -Metal Complex Transporter	MAs	2,109
153	*Hordeum vulgare*	HV555322	Transporter	Mugineic Acid -Metal Complex Transporter	MAs	2,037
154	*Triticum turgidum*	JF301955	Transporter	Transmembrane protein inducible by TNF-α	TMPIT1	660
155	*Triticum aestivum*	JF489139.1	Transporter	Cadmium/zinc-transporter	HMA2	2,221
156	*Triticum aestivum*	JF489140.1	Transporter	Cadmium/zinc-transporter	HMA2	2,369
157	*Triticum aestivum*	JF489141.1	Transporter	Cadmium/zinc-transporter	HMA2	1,519
158	*Hordeum vulgare*	JF496205	Transporter	High affinity potassium transporter	HKT1	2,505
159	*Hordeum vulgare*	JX051321	Transporter	Manganese transporter	MTP8.1	1,203
160	*Hordeum vulgare*	JX051322	Transporter	Manganese transporter	MTP8.2	1,233
161	*Hordeum vulgare*	JX310334	Transporter	Calcium dependent protein kinase	CDPK12	3,454
162	*Triticum turgidum*	JX896648.1	Transporter	High-affinity sulphate transporter	Sultr1.1	2,132
163	*Triticum turgidum*	JX896649.1	Transporter	High-affinity sulphate transporter	Sultr1.3	2,114
164	*Triticum aestivum*	KC521451	Transporter	Copper transporter	CT1-5A	3,172
165	*Triticum aestivum*	KC521452	Transporter	Copper transporter	CT1-5B	3,249
166	*Triticum aestivum*	KC521453	Transporter	Copper transporter	CT1-5D	3,230

ID	Species	Accession number	Gene category	Gene family	Gene name	Length (bp)
167	*Triticum turgidum*	KF148632.1	Transporter	boron transporter (Bot) gene	Bot-B5a	9,828
168	*Triticum turgidum*	KF443078.1	Transporter	High affinity potassium transporter	HKT1;4-1	1,692
169	*Triticum turgidum*	KF443079.1	Transporter	High affinity potassium transporter	HKT1;4-2	1,686
170	*Triticum aestivum*	KJ170113.1	Transporter	Phosphate transporter	PHT1.2-B1	1,578
171	*Triticum aestivum*	KJ170127.1	Transporter	Phosphate transporter	PHT1.10-D1	1,578
172	*Aegilops cylindrica*	KR051076.1	Transporter	High affinity potassium transporter	HKT1_5	580
173	*Triticum aestivum*	KR422354.1	Transporter	High affinity potassium transporter	TaHKT2;2	1,623
174	*Triticum aestivum*	KR422355.1	Transporter	High affinity potassium transporter	TaHKT2;2	1,527
175	*Triticum aestivum*	KR422356.1	Transporter	High affinity potassium transporter	TaHKT2;2	1,659
176	*Triticum aestivum*	KR422358.1	Transporter	High affinity potassium transporter	TaHKT2;1	2,144
177	*Aegilops tauschii*	KU253615.1	Transporter	High affinity potassium transporter	HKT8D	1,009
178	*Hordeum vulgare*	SL1251*	Transporter	Aluminium-activated malate transporter	HvALMT	1,587
179	*Hordeum vulgare*	SL17626*	Transporter	Aluminium-activated malate transporter	HvALMT	1,458
180	*Hordeum vulgare*	SL17915*	Transporter	Aluminium-activated malate transporter	HvALMT	1,527
181	*Hordeum vulgare*	SL20601*	Transporter	Aluminium-activated malate transporter	HvALMT	1,962
182	*Hordeum vulgare*	SL5062*	Transporter	Aluminium-activated malate transporter	HvALMT	1,368
183	*Hordeum vulgare*	SL5945*	Transporter	Aluminium-activated malate transporter	HvALMT2	475
184	*Hordeum vulgare*	U34198	Transporter	High affinity nitrate transporter	pBCH1	1,837
185	*Hordeum vulgare*	U34290	Transporter	High affinity nitrate transporter	pBCH2	1,741

Hybridisation enrichment and sequencing

Two hundred nanograms of each short adapter-ligated library were enriched with the two sets of probes (chloroplast and nuclear arrays) following the manufacturer's protocol version three (https://arborbiosci.com/wp-content/uploads/2017/10/MYbaits-manual-v3.pdf). The hybridisation reactions were incubated for 40 hours with 5 hrs at 65°C, 5 hrs at 60°C, and 30 hrs at 55°C. One sample of each species (*T.* cf. *aestivum* ACAD16052, *T.* cf. *timopheevii* ACAD16053, and *T.* cf. *turgidum* ACAD16054) was used for plastid enrichment. Following chloroplast analysis (presented in results), an additional accession of *T.* cf. *turgidum* (ACAD16055) was included with the original three samples for enrichment with the nuclear array.

A quantitative PCR (qPCR) assay was used to evaluate the enrichment of the chloroplast and nuclear probe captured libraries. *Primer3* online software (http://primer3.ut.ee) was used to design a set of qPCR primers to amplify a 55 bp fragment of the chloroplast *rbcL* gene and a 90 bp fragment of the nuclear *TaHKT1-1* gene based on the *T. aestivum* reference sequence NCBI accession numbers AY328025.1 and DQ009003.1, respectively (Table 3). Quantitative PCR was performed in a LightCycler® 96 (Roche, Switzerland) using three replicates of each DNA library. Each qPCR reaction contained 1µl of library DNA and 9 µl of Brilliant II SYBR Green QPCR Master Mix (Agilent Technologies, USA). Thermocycling conditions were as follows: 94 ˚C for 10 min, 50 cycles of 94 ˚C for 30 s, annealing temperature for 30 s, and 72 ˚C for 40 s. Enrichment success was verified as the relative difference in qPCR cycle threshold between pre-capture and post-capture libraries for each locus, indicative of the starting number of template molecules.

Table 3. Details of the primer sequences utilised to amplify by qPCR a 55 bp fragment of the chloroplast *rbcL* gene and a 90 bp fragment of the nuclear *TaHKT1-1* gene.

Amplicon	Sequence	Fragment size (bp)	Annealing (ºC)
rbcL	(F) AGCAGCTTGCAAATGGAGTC (R) CTCGAATTTGATCGCCTTCC	55	63
TaHKT1-1	(F) ATGTCTTGGGCTTTCTGCTG (R) GCACGATGTTGATCCCTTTC	90	60

Enriched DNA libraries were amplified and indexed as before using Illumina GAII indexing primers with a variable number of cycles (20-24 cycles) previously estimated via real-time PCR in a LightCycler® 96 instrument (Roche, Switzerland). Indexed libraries were purified using Ampure magnetic beads and quantified using an Agilent 2200 TapeStation system (Agilent Technologies, USA). Plastid enriched DNA libraries were pooled at 1.2 nM and sequenced on the Illumina NextSeq platform (2 x 150, paired-end) (Illumina, USA) at the Australian Cancer Research Foundation (ACRF) Cancer Genomics Facility in South Australia. Libraries enriched with the nuclear array were pooled at 1.5 nM and run on an Illumina HiSeq X Ten using 300 cycles (2 x 150 bp, paired-end) sequencing chemistry at the Garvan Institute of Medical Research, Sydney, Australia.

Data analysis

The demultiplexing tool *sabre* (https://github.com/najoshi/sabre) was used with default parameters and no mismatches permitted to separate the DNA sequence reads according to the unique P5 and P7 barcoded adapters. *BBmerge* (*BBtools* version 36.64 https://sourceforge.net/projects/bbmap) was used to trim adapters and merge overlapping reads, with base quality >4 and length >25 bp. The qualities of the reads were checked using *fastQC* version 0.11.5 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc) before and after adapter clipping.

Chloroplast analysis

Collapsed reads from the three historical wheat samples enriched with the plastid array were mapped to the chloroplast reference genome of *T. aestivum* (NCBI accession: NC_002762.1) with one of the two inverted repeats, using *BWA* version 0.7.15 (Li & Durbin, 2009; Li, 2013). Binary sequence alignment files were created using *SAMTools* version 1.3.1 (Li *et al.*, 2009) keeping reads with minimum mapping quality of 30. Alignment files were coordinate-sorted, and PCR duplicates removed using *SortSam* and *MarkDuplicates* from the *Picard* package version 2.1.1 (https://broadinstitute.github.io/picard). Consensus sequences and variant profiles were generated using *Geneious* 10.2.6 (https://www.geneious.com) with *highest quality* chosen as the threshold for calling the consensus sequence and a minimum of 5X coverage and frequency of one to call a variant.

The consensus chloroplast sequences of each historical sample were annotated using *GeSeq* (Tillich *et al.*, 2017) through comparison with the chloroplast reference genome of *T. aestivum* (NCBI accession: NC_002762.1). Circular maps of each chloroplast genome were generated using *OGDraw* (Lohse *et al.*, 2013).

To reconstruct the phylogeny and confirm the taxonomic identity of each of the historical wheat specimens, consensus sequences from all three samples were aligned with eight previously published chloroplast genomes of species from the Triticeae tribe and one outgroup (Table 4) using the *MUSCLE* program (Edgar, 2004) in *Geneious* 10.2.6 (https://www.geneious.com). The chloroplast genome of barley *H. vulgare* (NCBI accession: NC_002762.1) was used as an outgroup. To avoid the use of one of the two inverted repeats, a region of 100,000 bp was selected from all species and samples to perform the multiple sequencing analysis. Following the multiple sequencing analysis, a maximum likelihood (ML) phylogenetic tree was built using *RAxML GTRGAMMA* model (Stamatakis, 2014) in *Geneious* 10.2.6 (https://www.geneious.com) with 1,000 bootstrapping replicates. The phylogenetic tree was visualised and edited using *Figtree* version 1.4.4 (http://tree.bio.ed.ac.uk/software/figtree).

Table 4. Details of the chloroplast genomes of nine species of the Triticeae tribe used in this study for the phylogenetic analyses. NCBI accession numbers are provided. The GRIN taxonomy system is used to denote wheat species (https://npgsweb.ars-grin.gov).

Species	Nuclear genome	NCBI accession number
Aegilops sharonensis Eig	$S^{sh}S^{sh}$	NC_024816.1
Aegilops speltoides Tausch	SS	NC_022135.1
Aegilops tauschii Coss	DD	NC_022133.1
Hordeum vulgare L.	HH	NC_008590.1
Triticum aestivum L.	AABBDD	NC_002762.1
Triticum monococcum L.	A^mA^m	NC_021760.1
Triticum timopheevii (Zhuk.) Zhuk.	AAGG	NC_024764.1
Triticum turgidum L.	AABB	AGG26449
Triticum urartu Tumanian ex Gandilyan	A^uA^u	NC_021762.1

Nuclear analysis

Merged reads from the four nuclear enriched libraries were mapped to the reference sequences of the 185 wheat and barley genes included in the nuclear array (Table 2) using *BWA* version 0.7.15 (Li & Durbin, 2009; Li, 2013). Alignments with reads of mapping quality >30 were selected using *SAMTools* version 1.3.1 (Li *et al.*, 2009) and duplicate reads were discarded using *Picard* package version 2.1.1 (https://broadinstitute.github.io/picard).

Since polyploids contain multiple distinct copies of each nuclear gene derived from the genomes of different ancestors, known as homoeologous copies, an increased depth of coverage and accurate approach is required to reliably detect nucleotide variants. We used the *Freebayes* program version 1.2.0 (Garrison & Marth, 2012) to generate the variant profiles of each sequencing library with minimum coverage of 15x to call a variant. *Freebayes* identifies variants based on the most likely combination of genotypes at each position in the reference sequence. With this Bayesian approach, *Freebayes* is capable of analysing sequences of organisms with ploidy level greater than 2N and detecting multi-allelic haplotypes (Garrison & Marth, 2012). Variant profiles for each sample were imported into *Geneious* 10.2.6 (https://www.geneious.com), checked by eye, and consensus sequences were obtained using the *highest quality* as threshold.

Mapping details for each DNA library were estimated using *SAMTools* version 1.3.1 (Li *et al.*, 2009) and *Qualimap* version 2.2.1 (Okonechnikov *et al.*, 2016), including the number of mapped reads, number of unique reads, library complexity, clonality, percentage of unique endogenous reads, coverage, GC content, and fragment lengths. Plots and figures were generated using R 3.5.1 base graphics (R CoreTeam, 2018) and the ggplot2 package (Wickham, 2016).

Results

Capture efficiency of the chloroplast sequences

Sequencing of chloroplast enriched libraries from three historical Georgian polyploid wheat species resulted in a total of 25.2 million high quality reads, with the number of reads sequenced per sample ranging from 19.5 million to 2.4 million reads (Table 5). The percentage of reads that mapped back to the chloroplast reference genome of *T. aestivum* (NCBI accession: NC_002762.1) was 42% with an average read length of 176.3 nucleotides. The plastid array efficiently recovered the full chloroplast genome from single seeds of each of the three historical wheat species with a mean coverage of 335.9X for *T.* cf. *aestivum*, 366.4X for *T.* cf. *timopheevii*, and 336.7X for *T.* cf. *turgidum* (Table 5). Gene maps of the chloroplast genomes of the three historical wheat samples are presented in Supplementary Figures 1-3.

Table 5. Summary of the chloroplast assembly information of three historical Georgian polyploid wheat samples. The accession numbers of the Australian Centre for Ancient DNA (ACAD) biological collection are provided.

ACAD accession number	Species	Total number of reads	Percentage of reads on target	Mean depth of coverage
ACAD16052	*Triticum* cf. *aestivum*	19,158,178	36%	335.9X
ACAD16053	*Triticum* cf. *timopheevii*	3,643,850	61%	366.4X
ACAD16054	*Triticum* cf. *turgidum*	2,415,172	52%	336.7X

Variations in the chloroplast genomes of the historical polyploid wheat species from Georgia were characterised using the chloroplast genome of *T. aestivum* (NCBI accession: NC_002762.1) as a reference, again minus one inverted repeat to assist with mapping. A total of 518 polymorphisms were identified in the three specimens with a mean coverage of 265X, including 481 single nucleotide variants (SNVs) and 37 indels (insertions/deletions) (Figure 3). Timopheevii wheat held 44% of the total chloroplast variations, with 211 SNVs and 16 indels and a mean coverage of 288.7X (Figure 3). *T.* cf. *turgidum* showed 169 chloroplast variants (156 SNVs and 13 indels) with a mean coverage of 242.4X, and the bread wheat *T.* cf. *aestivum* presented 122 variants (114 SNVs and 8 indels) with a mean coverage of 266.3X (Figure 3).

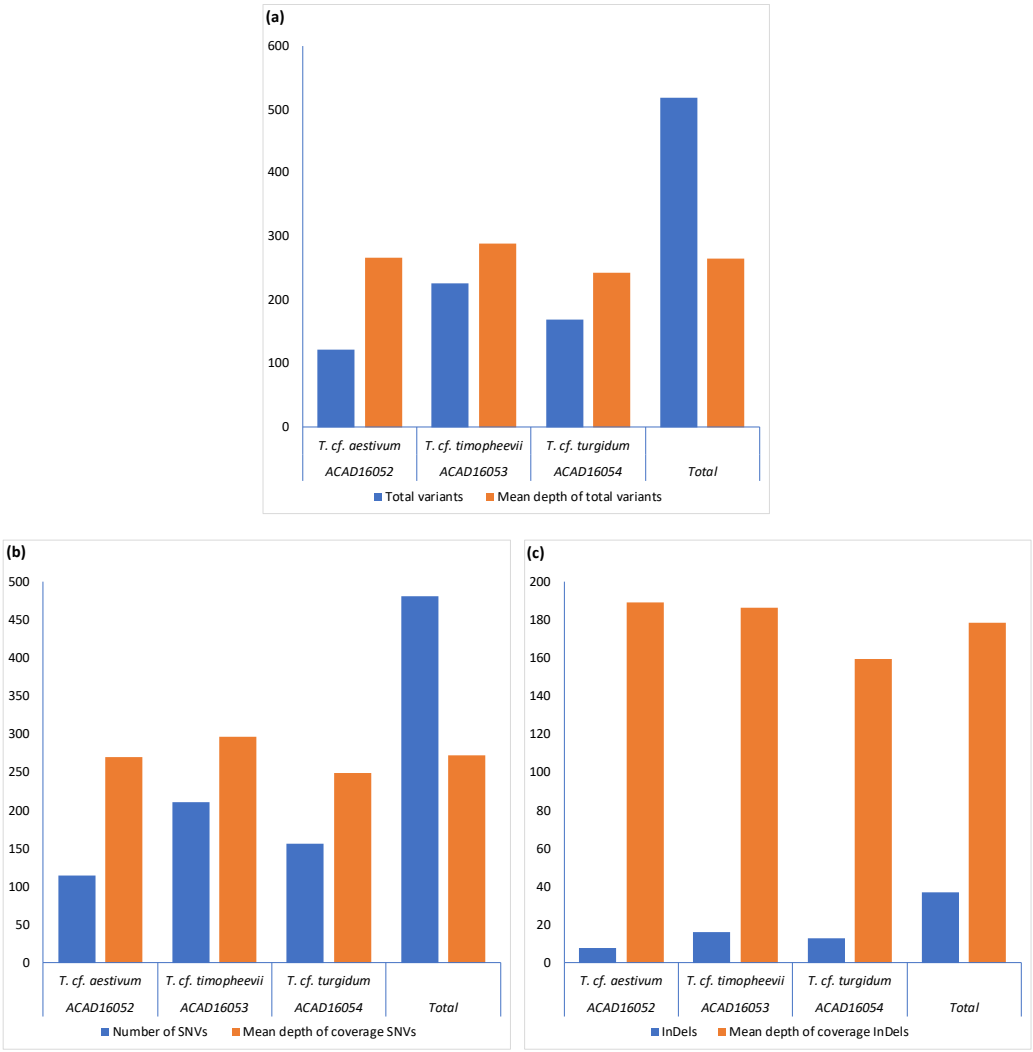


Figure 3. Number polymorphisms detected in the chloroplast genomes of three historical Georgian polyploid wheat samples. (a) Total number of variants (SNVs/indels) and mean depth of coverage. (b) Number of single nucleotide variants (SNVs) and mean depth of coverage. (c) Number of insertions and deletions (Indels) and mean depth of coverage.

While around 4.55 variants were observed per kilobase for the total SNVs and indels, the distribution across the chloroplast genome is uneven. Most of the variation in all samples is concentrated in four regions of the chloroplast genomes with 65% in the first 12,000 bp, including 364 polymorphisms, mostly in the following genes: *matK*, *trnK*, *trnQ*, *psbA*, *psbC*, *psbD*, and *psbK* (Figure 4). A second small region located between 60,000 to 62,000 bp accounts for 26 polymorphisms (~5%), including variants in the *petA*, *psbE*, *psbF*, and *psbL* genes. In the inverted repeat region, 92 (~18%) polymorphisms are found, involving variants in the *rpl2*, *rpl23*, *rps7*, and *ndhB* genes. The last block of variants is located between 100,000 to 106,000 bp and contains 29 variants (~5%), some of these variants positioned in the *ndhF* and *ccsA* genes (Figure 4).

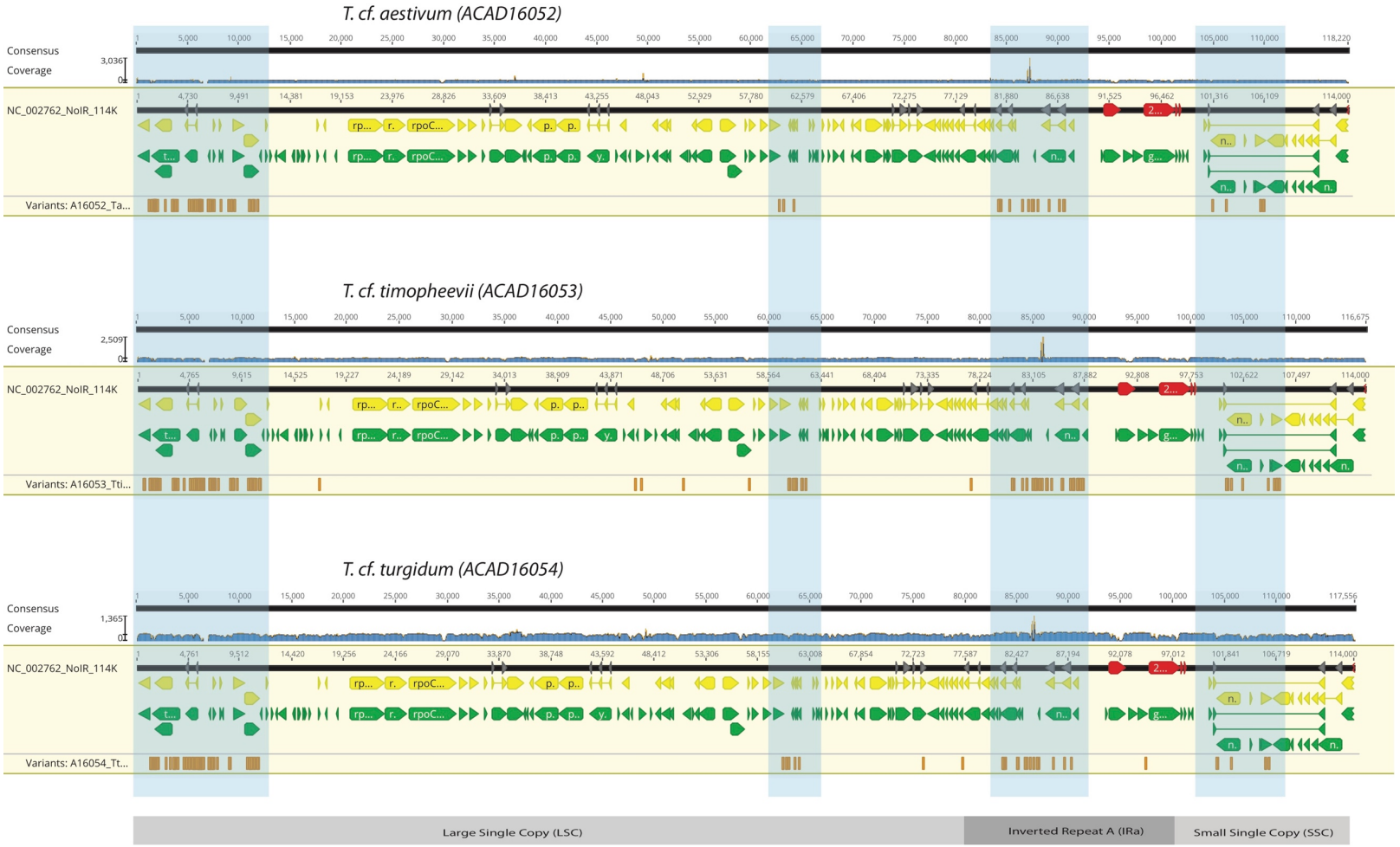


Figure 4. Overall distribution of variations (SNVs and indels) across the chloroplast genomes of three historical Georgian polyploid wheat samples. Variations of all three accessions were identified based on the chloroplast reference genome of *T. aestivum* (NCBI accession: NC_002762.1). Blue shaped area denotes regions of high density of nucleotide variants. The large single copy (LSC) of approximately 80 Kb, the inverted repeat A (IRa) of 20 Kb and the small single copy (SSC) regions of 12 Kb are indicated.

In order to reconstruct the phylogeny of the polyploid historical Georgian wheat samples, we combined the new recovered sequences with existing chloroplast genomes of different taxa from the Triticeae tribe (Table 4). The best phylogenetic tree from the maximum likelihood analysis showed two monophyletic groups separating most of the diploid species from the polyploid species (Figure 5). The diploid species *A. speltoides* (SS), together with *T. timopheevii* (AAGG), *T. turgidum* (AABB), and *T. aestivum* (AABBDD), form a monophyletic clade (100% bootstrap support), indicating that diploid species of the *Speltoides* lineage are the maternal donors of the chloroplast to the polyploid historical wheat samples from Georgia. All polyploid species carrying the A and B genomes, including the tetraploid wheat *T. turgidum* (AABB) and the hexaploidy wheat *T. aestivum* (AABBDD), form a monophyletic clade with 100% bootstrap support (Figure 5). This association of the A and B genomes demonstrates the origin of the chloroplast genome of *T. turgidum* and *T. aestivum* from the B subgenome donor (*A. speltoides*).

The phylogeny also revealed that the historical wheat *T.* cf. *timopheevii* is closely related to the reference genome of the *T. timopheevii* with 100% bootstrap support (Figure 5). In contrast, the historical samples *T.* cf. *aestivum* and *T.* cf. *turgidum* are more closely related to each other (100% bootstrap support) than to the reference chloroplast genomes of *T. aestivum* and *T. turgidum* (Figure 5). The chloroplast reference genome of *T. aestivum* is the sister-taxon of the historical samples of *T.* cf. *aestivum* and *T.* cf. *turgidum*. Although this clade does not have strong bootstrap support (63% bootstrap support), the phylogeny suggests that both accessions (*T.* cf. *aestivum* ACAD16052 and *T.* cf. *turgidum* ACAD16054) belong to the *T. aestivum* species (Figure 5).

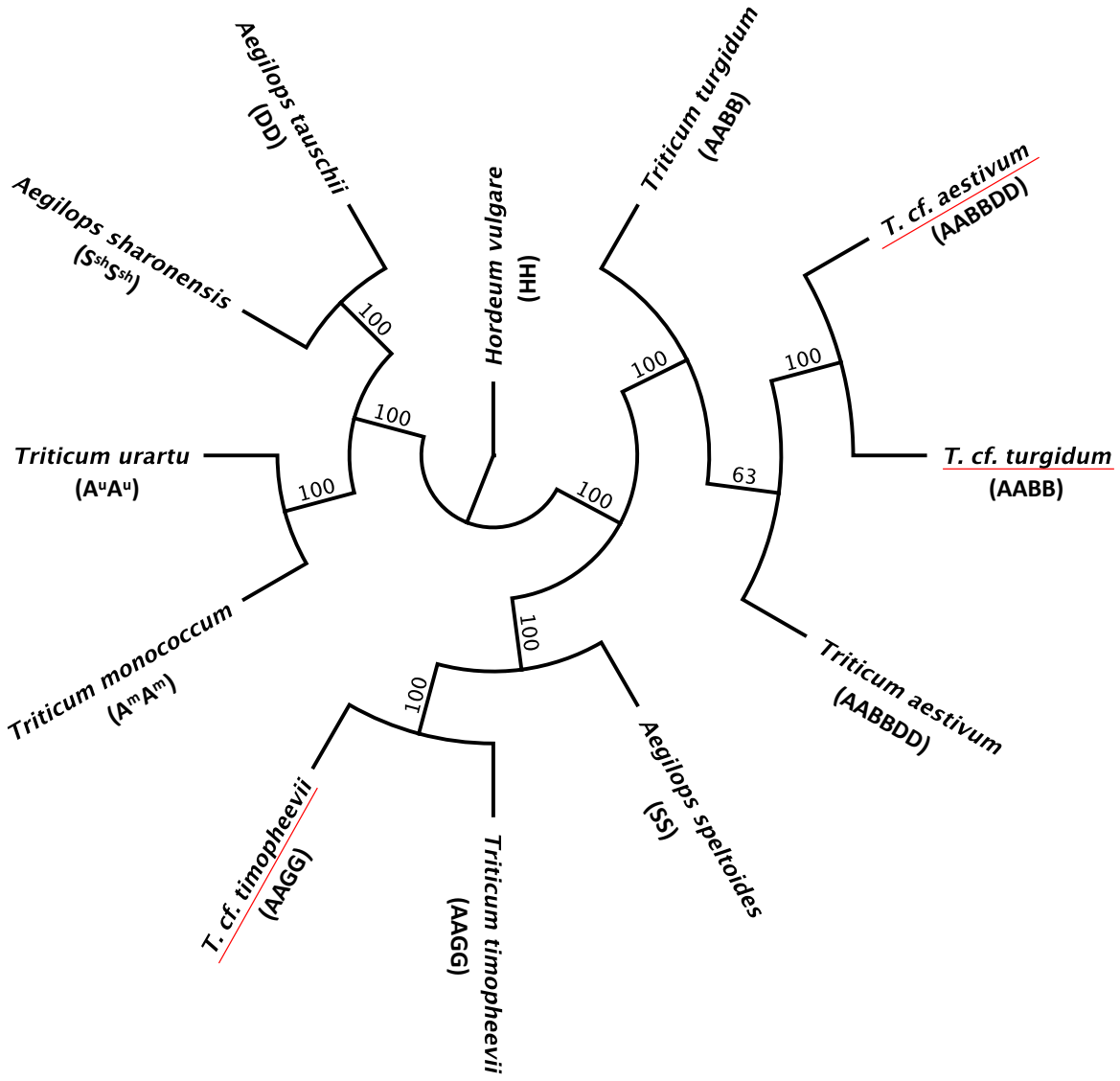


Figure 5. Chloroplast phylogeny of three historical Georgian polyploid wheat samples (underlined in red) estimated following a multiple sequence alignment with nine previously published chloroplast genomes of species from the Triticeae tribe. The best maximum likelihood tree was built using RAxML with 1,000 bootstrapping replicates and *H. vulgare* as outgroup. Numbers indicate bootstrap support values. Nuclear genomes for each accession are denoted in parenthesis. Details of accession numbers are listed in Table 4. The historical Georgian sample *T.* cf. *turgidum* forms a monophyletic group with the sample *T.* cf. *aestivum*, which are the sister-taxon of the reference chloroplast genome of the *T. aestivum* species, suggesting that both historical samples belong to the *T. aestivum* species.

Sequence capture efficiency of the nuclear variation

Given that the results of the chloroplast phylogeny positioned the historical sample of *T.* cf. *turgidum* ACAD16054 as more like *T. aestivum* (Figure 5), an additional sample *T.* cf. *turgidum* ACAD16055 was included in the nuclear enrichment analysis. Sequencing libraries were constructed from four samples of historical polyploid wheat species from Georgia (Table 1) and enriched using the customised nuclear array. The hybridisation capture array was designed to recover sequences from 185 wheat and barley nuclear genes with a total of 224 Mb of target sequences (Table 2).

Deep sequencing of the nuclear enriched libraries produced a total of 52.1 million high quality reads from which 2.6 million mapped to the reference sequences with an average coverage of 44X (Table 6). On average, five percent of reads from each accession mapped to the reference sequence of the genes in the array with a mean coverage of 38.8X for *T.* cf. *aestivum*, 20.8X for *T.* cf. *timopheevii*, and 57.5X for *T.* cf. *turgidum* (Table 6). The coverage per gene across samples varied from 0-250X with a mean coverage of 44X (Figure 6, 7, and Supplementary Table 1). Of the 185 genes in the array, 96% (177 genes) had an average coverage of more than 1X in at least one accession, eight genes had an average coverage of less than 1X, and there were no sequences retrieved for two genes in all samples (Figure 6, 7, and Supplementary Table 1). The nuclear enrichment recovered sequences from 171 genes in the accessions *T.* cf. *timopheevii* ACAD16053 and *T.* cf. *turgidum* ACAD16055, and 178 genes in the accessions *T.* cf. *aestivum* ACAD16052 and *T.* cf. *turgidum* ACAD16054 (Figure 6, 7, and Supplementary Table 1).

Table 6. The number of sequenced reads, percentage of reads on target, and coverage of four historical Georgian polyploid wheat samples following hybridisation capture of 185 nuclear genes. The accession numbers of the Australian Centre for Ancient DNA (ACAD) biological collection are provided. (*) Specimen identified as *T. aestivum* based on the chloroplast phylogeny.

Accession	Species	Total number of reads	Percentage of reads on target	Mean depth of coverage
ACAD16052	*Triticum* cf. *aestivum*	13,353,665	6%	38.8X
ACAD16053	*Triticum* cf. *timopheevii*	11,283,707	4%	20.8X
ACAD16054*	*Triticum* cf. *turgidum*	13,983,605	5%	56.7X
ACAD16055	*Triticum* cf. *turgidum*	13,569,585	5%	59.7X
Total		52,190,562	5%	44X

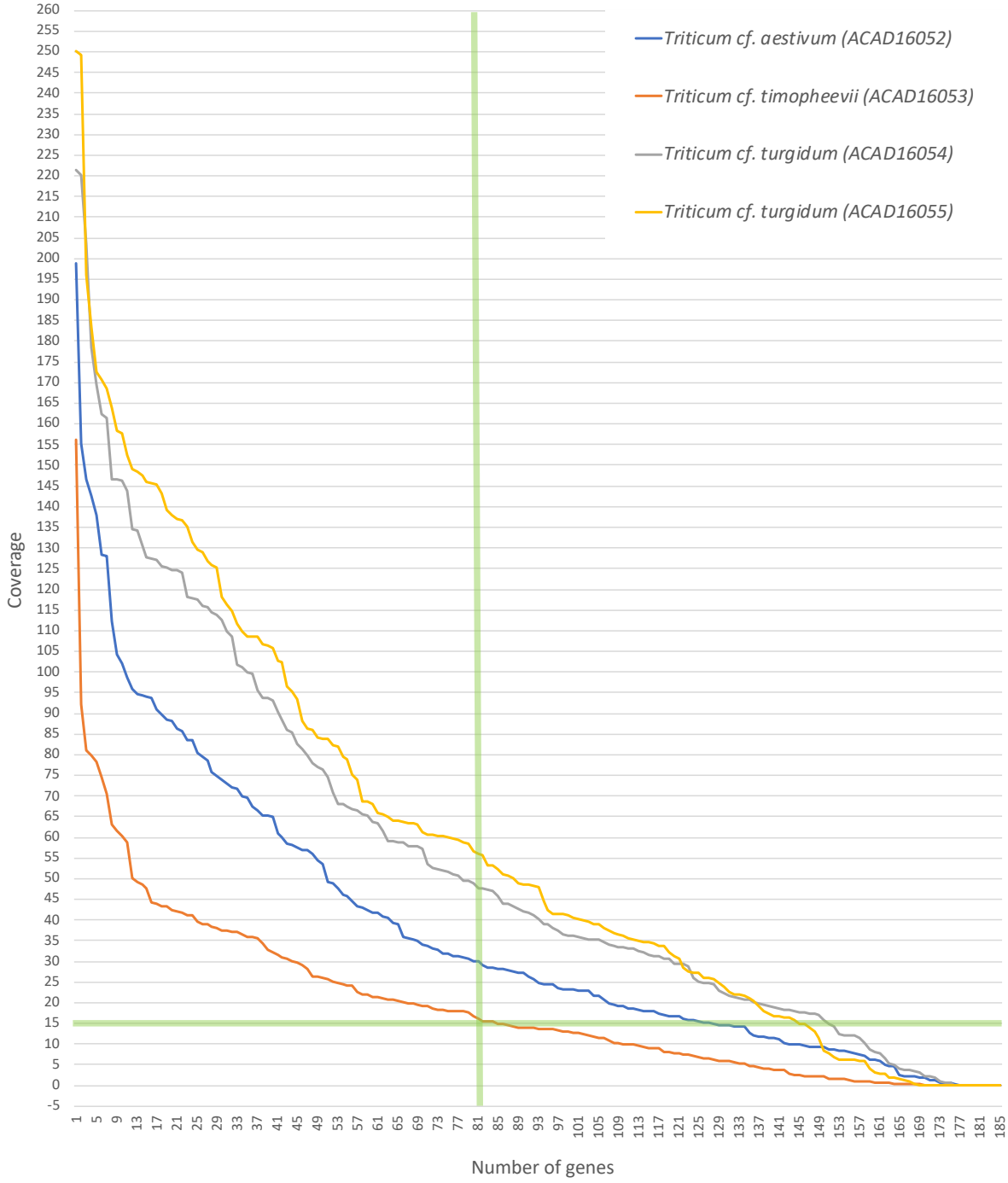


Figure 6. Mean depth of coverage plotted against the number of genes captured from four historical Georgian polyploid wheat samples following hybridisation capture of 185 nuclear genes. The digit in the X axis signifies the number of genes with a given coverage (Y axis). We selected a mean depth of coverage of 15X (green horizontal lane) as a threshold for variant calling, from which 81 genes are represented across all species (green vertical lane).

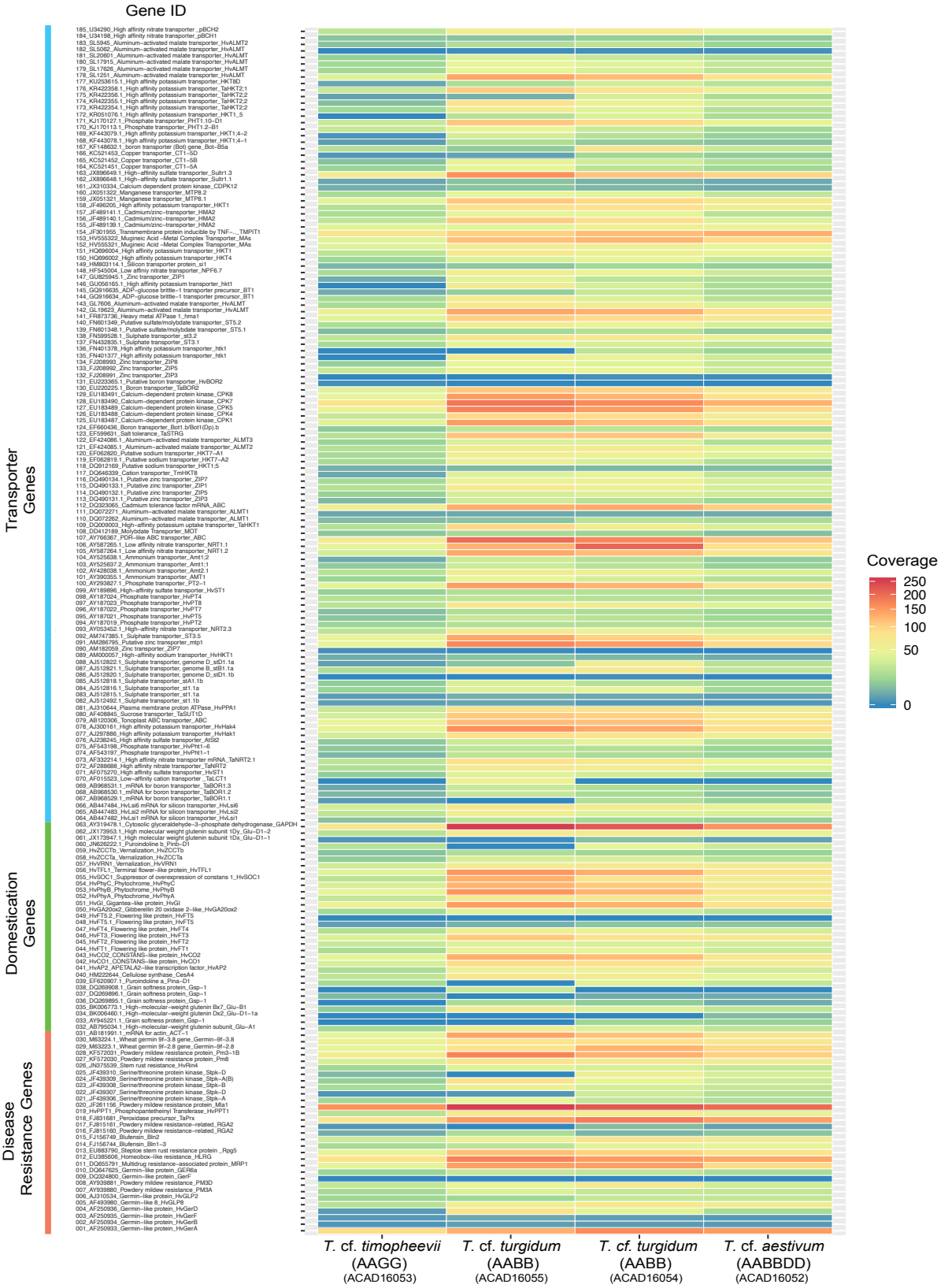


Figure 7. Heat map coverage for 185 nuclear genes sequenced from four historical Georgian polyploid wheat samples. Nuclear genomes for each accession are represented in parenthesis. The hybridisation capture array efficiently enriched most of the genes in the four samples with a coverage higher than 30X.

The nuclear array included a set of genes of the hardness locus (*Ha-D*) which can be used to diagnose the presence of the D subgenome of the hexaploid wheat *T. aestivum*. Grain texture in wheat is mainly associated with three genes of the *Ha-D* locus: *Puroindoline a* (*Pina-D1*), *Puroindoline b* (*Pinb-D1*), and *Grain softness protein-1* (*Gsp-D1*), which are located on chromosome 5DS of *T. aestivum* (Gautier *et al.*, 1994; Rahman *et al.*, 1994; Chantret, 2005) (Figure 8). In diploid species, soft grain texture is associated with alleles of both *Puroindoline* gene*s* (*Pina-D1a* and *Pinb-D1a*) (Gautier *et al.*, 2000) (Figure 8). In the tetraploid wheat *T. turgidum* (AABB), *Puroindoline* gene*s* were deleted during the hybridisation event that gave rise to the species, resulting in hard textured grain (Giroux & Morris, 1998; Chang *et al.*, 2006) (Figure 8). Grain softness was restored in the hexaploid bread wheat *T. aestivum* (AABBDD) following the allopolyploidisation between the diploid D genome donor *A. tauschii* and the tetraploid *T. turgidum* (Chantret, 2005) (Figure 8). Regarding the tetraploid *T. timopheevii* (AAGG), during the allopolyploidisation process *Pina* and *Pinb* were eliminated from the G genome but maintained in the A genome (Li *et al.*, 2008) (Figure 8).

The chloroplast phylogeny analysis highlighted that the historical samples *T.* cf. *aestivum* ACAD16052 and *T.* cf. *turgidum* ACAD16054 in reality likely both belong to the same species *T. aestivum* (Figure 5). The nuclear enrichment recovered sequences of both *Puroindoline* gene*s* (*Pina-D1* and *Pinb-D1*) from these specimens confirming they are indeed *T. aestivum* (Figure 8). Additionally, the nuclear enrichment did not recover sequences of the *Puroindoline* gene*s* from the accession *T.* cf. *turgidum* ACAD16055 suggesting that this accession is in fact likely to belong to *T. turgidum* species group (Figure 8).

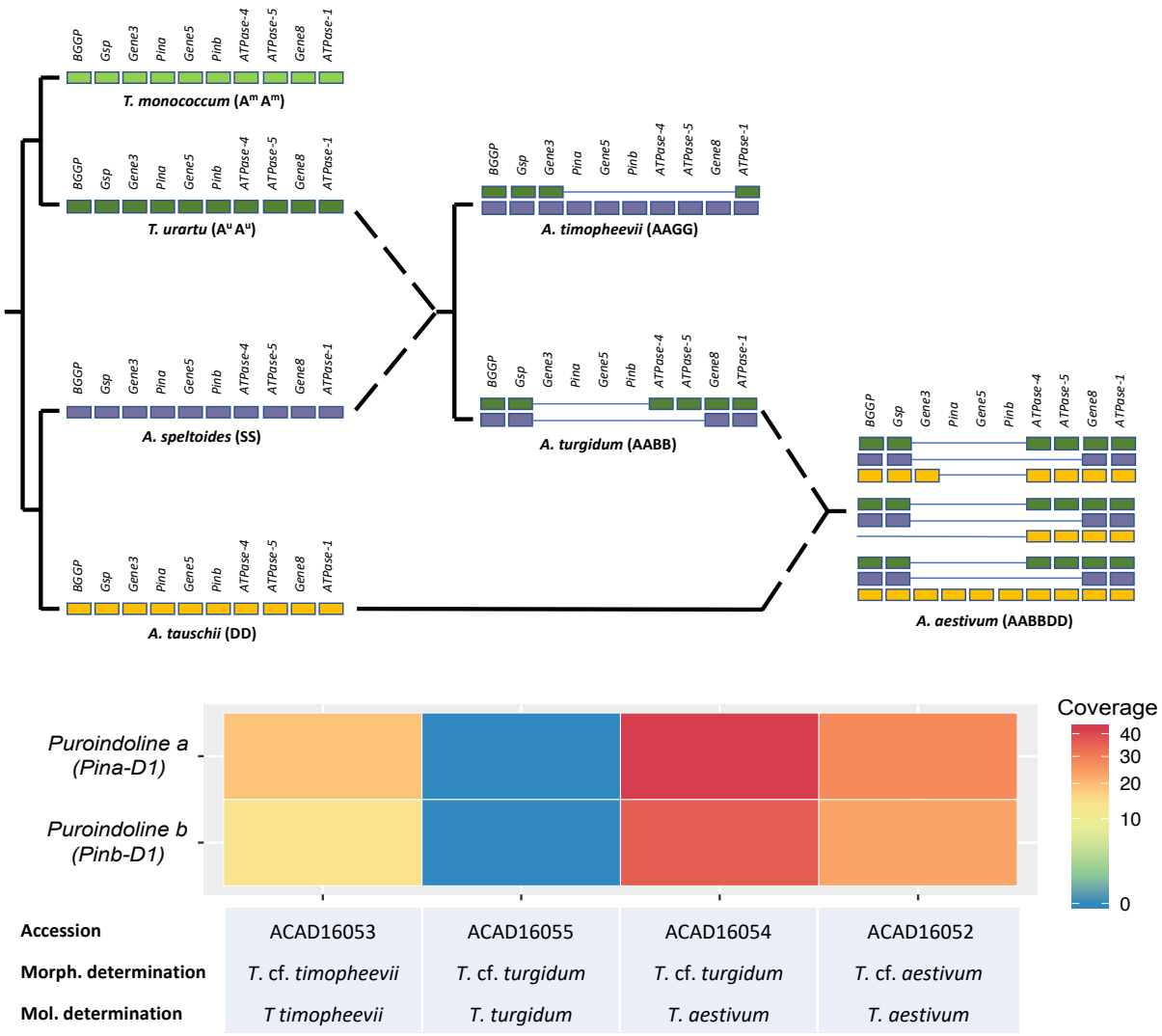


Figure 8. Top chart: Diagram representing the evolution of the hardness locus *Ha* in wheat species. The small coloured boxes represent the presence or absence of the ten genes of the hardness locus. Nuclear genomes for each species are represented in parenthesis. Bottom chart: heat map coverage of *Puroindoline a* (*Pina-D1*), *Puroindoline b* (*Pinb-D1*) genes sequenced from four historical Georgian polyploid wheat samples. The table underneath shows the accession number, morphological determination, and molecular determination for each sample. The *Pina-D1* and *Pinb-D1* are present in the sample ACAD16054 *T. cf. turgidum* confirming that it belongs to the *T. aestivum* species.

We used a minimum depth of coverage of 15X to detect nucleotide polymorphisms in the historical Georgian polyploid wheat species. From the 185 genes included in the nuclear capture array, 81 genes had a minimum depth of coverage of 15X across all species (Figure 6). Variant calling was performed in one accession of each species (*T. aestivum* ACAD16052, *T. timopheevii* ACAD16053, and *T. turgidum* ACAD16055) using the sequences of 81 wheat and barley genes as a reference.

A total of 20,244 single nucleotide variants were identified in 81 nuclear genes of the historical Georgian polyploid wheat species (Figure 9 and Supplementary Table 2). *T. turgidum* showed the highest number of variants (7,039 SNVs), followed by timopheevii wheat (6,859 SNVs) and bread wheat (6,346 SNVs) (Figure 9). The maximum number of SNVs was recorded in transporter genes (8,414 SNVs), followed by domestication related genes (6,767 SNVs) and disease resistance related genes (5,063 SNVs) (Figure 9). A similar frequency of single nucleotide variants per kilobase (SNVs/kb) was detected in all three historical polyploid wheat species. *T. turgidum* had 31.4 SNVs/kb, *T. timopheevii* 30.6 SNVs/kb and *T. aestivum* 28.3 SNVs/kb (Figure 9-10). Although transporter genes showed the highest number of SNVs in all species, domestication related genes showed the highest proportion of SNVs per kilobase across species with 117.2 SNVs/kb (Figure 9-10). Disease resistance and transporter genes exhibited 97.4 and 73.5 SNVs per kilobase, respectively (Figure 9-10).

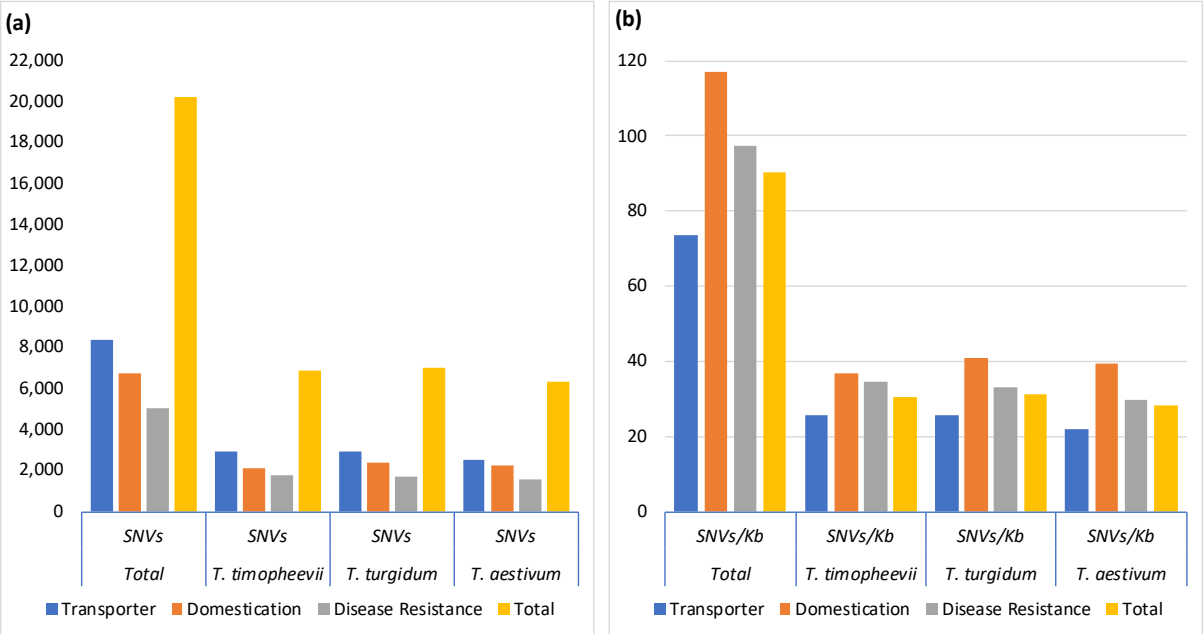


Figure 9. Number of single nucleotide variants detected in 81 nuclear genes sequenced from three historical Georgian polyploid wheat samples. (a) Number of single nucleotide variants (SNVs). (b) Number of single nucleotide variants per kilobase (SNVs/Kb).

Figure 10. Heat map of single nucleotide variants per kilobase (SNVs/Kb) detected in 81 nuclear genes sequenced from four historical Georgian polyploid wheat samples. Nuclear genomes for each accession are represented in parenthesis. Although all genes had a high number of nucleotide variants, domestication and disease resistance-related genes exhibited a higher number of nucleotide variants across all samples.

Discussion

The characterisation of the wheat genome has been revolutionised by the application of HTS methods, which have allowed the sequencing of the genomes of the diploid wheat *T. urartu*, the tetraploid wheat *T. turgidum*, and the hexaploid wheat *T. aestivum* (Mayer *et al.*, 2014; Avni *et al.*, 2017; Appels *et al.*, 2018; Ling *et al.*, 2018). We aimed to further address the history of wheat domestication by developing two hybridisation capture arrays to investigate the plastid and autosomal variability in historical samples (60-100-years-old) of polyploid wheat species (*T. aestivum*, *T. timopheevii*, and *T. turgidum*) from Georgia.

The application of the plastid array allowed the reconstruction of the chloroplast genomes of the three historical wheat samples with mean of coverage >335X. Sequence alignment to the chloroplast reference genome of *T. aestivum* (NCBI accession: NC_002762.1) identified 518 polymorphisms, including 481 SNVs and 37 indels in the chloroplast sequences of the three historical samples. Previous chloroplast studies in modern polyploid wheat species from Georgia have reported only 15 polymorphisms in emmer wheat (*T. turgidum* L. subsp. *paleocolchicum* Á. & D. Löve) and spelt wheat (*T. aestivum* L. subsp. *macha* (Dekapr. & A. M. Menabde) Mackey) when comparing with a de novo assembled reference chloroplast genome of spelt wheat (*T. aestivum* subsp. *macha*) (Gogniashvili *et al.*, 2018). Similarly, another study documented only 10 indels in the chloroplast genome of *T. timopheevii* when aligned to the *T. aestivum* chloroplast sequence (NCBI accession: KJ614396) (Gornicki *et al.*, 2014). As a result, the current study dramatically increases the number of known variants within these wheat accessions and highlights the potential information that could be accessed for evolutionary studies or breeding purposes.

A high number of polymorphisms identified in chloroplast sequences are often due to intra-individual polymorphisms. Although the chloroplast is a haploid organelle, sequence data have documented a large number of intra-individual polymorphisms in different species (Wolfe & Randle, 2004; Sabir *et al.*, 2014; Scarcelli *et al.*, 2016). Such polymorphisms can derive from sequencing errors, the presence of heteroplasmy, and the transfer of chloroplast sequences into the nuclear and mitochondrial genomes (Scarcelli *et al.*, 2016). In this study, we constructed DNA sequencing libraries from total DNA isolated from single seeds. Consequently, the DNA sequencing libraries are likely to contain chloroplast, mitochondrial and nuclear DNA. Therefore, we used a strict bioinformatics haploid approach to detect polymorphisms in the chloroplast sequence data. This approach includes keeping reads with mapping quality >30 and

calling variants with coverage >5X and a frequency of one (100% of the alignments include the variant). The application of this strict haploid methodology allowed us to discard any sequence errors, heteroplasmies, and exclude the interference of nuclear and mitochondrial sequences. In addition, the deep sequencing of the chloroplast enriched libraries allowed a high coverage alignment for each polymorphism (mean coverage of 265X). Therefore, the large number of nucleotide polymorphisms detected in the chloroplast enriched libraries represents the plastid diversity of the historical Georgian polyploid wheat species.

The historical wheat samples were previously identified as *T.* cf. *aestivum*, *T.* cf. *timopheevii*, and *T.* cf. *turgidum* based on morphological characteristics of the inflorescences, spikes, and seeds. However, morphological identification of wheat seeds to a species level is often difficult owing to the structural similarities between species (Goncharov, 2011). In historical and archaeological samples, morphological identification can be even more complex due to taphonomic and diagenetic processes (Nesbitt, 2016). The chloroplast-based phylogenetic analysis together with the hardness locus included in the nuclear array, clarified the phylogenetic position of one sample misidentified as *T.* cf. *turgidum* as actually being *T. aestivum*, while three samples were confirmed as *T. aestivum*, *T. timopheevii*, and *T. turgidum*. This result highlights the power of our two hybridisation capture arrays to reconstruct the phylogenetic history of museum, historical, and other ancient samples.

The application of HTS has allowed the massive identification of SNPs in several plant genomes large and complex ones of wheat species (Kumar *et al.*, 2012), However, SNP identification in polyploid genomes poses a significant challenge due to the need to distinguish homoeologous SNPs (polymorphic positions occurring across subgenomes) from allelic SNPs (polymorphic positions occurring within a single subgenome) (Wang *et al.*, 2014; Clevenger *et al.*, 2015). Large numbers of HTS-based SNPs have been reported in several studies applying hybridisation capture in tetraploid and hexaploid wheat species, varying from hundreds of thousands to millions of SNPs (Saintenac *et al.*, 2011; Winfield *et al.*, 2012; Cavanagh *et al.*, 2013; Wang *et al.*, 2014). Here, we developed a nuclear capture array to recover sequences from 185 genes related to cell transport, domestication traits, and disease resistance. We demonstrated the ability of the capture array to recover a high number of SNPs (20,244 SNPs) in a dataset of 81 genes (224 Kb) within just three historical wheat species from Georgia.

The large number of nucleotide polymorphisms recovered by both the plastid and nuclear arrays represents the genetic variability present 100 years ago. These historical samples were adapted to the environmental conditions and agricultural practices of the time and therefore can be used as a reference to understand the major genetic changes in the evolutionary history of wheat in the last 100 years. In particular, these historical samples existed before the Green Revolution and presumably contain genetic diversity that might has been lost during this major bottleneck. As a result, the genetic variability discovered in the historical wheat species from Georgia together with the hybridisation capture arrays developed in this study are a valuable source of information for the genetic mapping, gene discovery, germplasm characterisation, and population genomics in wheat and other cereals. As a result, the genetic variability discovered in the historical wheat species from Georgia together with the hybridisation capture arrays developed in this study are a valuable source of information for the genetic mapping, gene discovery, germplasm characterisation, and population genomics in wheat and other cereals.

Conclusion

We demonstrated the utility of two new plastid and nuclear arrays to reconstruct the phylogenetic history of historical wheat specimens and to recover a large number of single nucleotide polymorphisms. The arrays promise to provide a valuable resource for the analysis of wheat genetic diversity in modern, historical, and ancient samples. The new highly-detailed information about genetic diversity in historical Georgian wheat species will facilitate future detailed studies of the of the evolution and domestication of this cereal.

References

Allaby RG. **2014**. Genetics of Early Plant Domestication: DNA and aDNA. In: Smith C, ed. Encyclopedia of Global Archaeology. New York, NY: Springer New York, 3004–3007.

Appels R, Eversole K, Feuillet C, Keller B, Rogers J, Stein N, Pozniak CJ, Choulet F, Distelfeld A, Poland J, *et al.* 2018. Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* **361**.

Avni R, Nave M, Barad O, Baruch K, Twardziok SO, Gundlach H, Hale I, Mascher M, Spannagl M, Wiebe K, *et al.* 2017. Wild emmer genome architecture and diversity elucidate wheat evolution and domestication. *Science* **357**: 93–97.

Bevan MW, Uauy C, Wulff BBH, Zhou J, Krasileva K, Clark MD. **2017**. Genomic innovation for crop improvement. *Nature* **543**: 346–354.

Brown TA, Jones MK, Powell W, Allaby RG. **2009**. The complex origins of domesticated crops in the Fertile Crescent. *Trends in Ecology and Evolution* **24**: 103–109.

Byrt CS, Xu B, Krishnan M, Lightfoot DJ, Athman A, Jacobs AK, Watson-Haigh NS, Plett D, Munns R, Tester M, *et al.* 2014. The Na + transporter, TaHKT1;5-D, limits shoot Na + accumulation in bread wheat. *The Plant Journal* **80**: 516–526.

Cavanagh CR, Chao S, Wang S, Huang BE, Stephen S, Kiani S, Forrest K, Saintenac C, Brown-Guedira GL, Akhunova A, *et al.* 2013. Genome-wide comparative diversity uncovers multiple targets of selection for improvement in hexaploid wheat landraces and cultivars. *Proceedings of the National Academy of Sciences* **110**: 8057–8062.

Chang C, Zhang H, Xu J, Li W, Liu G, You M, Li B. **2006**. Identification of allelic variations of puroindoline genes controlling grain hardness in wheat using a modified denaturing PAGE. *Euphytica* **152**: 225–234.

Chantret N. **2005**. Molecular Basis of Evolutionary Events That Shaped the Hardness Locus in Diploid and Polyploid Wheat Species (Triticum and Aegilops). *THE PLANT CELL ONLINE* **17**: 1033–1045.

Clevenger J, Chavarro C, Pearl SA, Ozias-Akins P, Jackson SA. **2015**. Single Nucleotide Polymorphism Identification in Polyploids: A Review, Example, and Recommendations. *Molecular Plant* **8**: 831–846.

Cronn R, Knaus BJ, Liston A, Maughan PJ, Parks M, Syring J V., Udall J. **2012**. Targeted enrichment strategies for next-generation plant biology. *American Journal of Botany* **99**: 291–311.

Doebley JF, Gaut BS, Smith BD. **2006**. The molecular genetics of crop domestication. *Cell* **127**: 1309–21.

Edgar RC. **2004**. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**: 1792–1797.

Feuillet C, Langridge P, Waugh R. **2008**. Cereal breeding takes a walk on the wild side. *Trends in Genetics* **24**: 24–32.

Garrison E, Marth G. **2012**. Haplotype-based variant detection from short-read sequencing. : 1–9.

Gautier MF, Aleman ME, Guirao A, Marion D, Joudrier P. **1994**. Triticum aestivum puroindolines, two basic cystine-rich seed proteins: cDNA sequence analysis and developmental gene expression. *Plant Molecular Biology* **25**: 43–57.

Gautier MF, Cosson P, Guirao A, Alary R, Joudrier P. **2000**. Puroindoline genes are highly

conserved in diploid ancestor wheats and related species but absent in tetraploid Triticum species. *Plant Science* **153**: 81–91.

Giroux MJ, Morris CF. **1998**. Wheat grain hardness results from highly conserved mutations in the friabilin components puroindoline a and b. *Proceedings of the National Academy of Sciences* **95**: 6262–6266.

Gnirke A, Melnikov A, Maguire J, Rogov P, LeProust EM, Brockman W, Fennell T, Giannoukos G, Fisher S, Russ C, *et al.* 2009. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nature biotechnology* **27**: 182–9.

Gogniashvili M, Maisaia I, Kotorashvili A, Kotaria N, Beridze T. **2018**. Complete chloroplast DNA sequences of Georgian indigenous polyploid wheats (Triticum spp.) and B plasmon evolution. *Genetic Resources and Crop Evolution* **65**: 1995–2002.

Goncharov NP. **2011**. Genus Triticum L. taxonomy: the present and the future. *Plant Systematics and Evolution* **295**: 1–11.

Gornicki P, Zhu H, Wang J, Challa GS, Zhang Z, Gill BS, Li W. **2014**. The chloroplast view of the evolution of polyploid wheat. *New Phytologist* **204**: 704–714.

Hedden P. **2003**. The genes of the Green Revolution. *Trends in Genetics* **19**: 5–9.

Hodges E, Xuan Z, Balija V, Kramer M, Molla MN, Smith SW, Middle CM, Rodesch MJ, Albert TJ, Hannon GJ, *et al.* 2007. Genome-wide in situ exon capture for selective resequencing. *Nature genetics* **39**: 1522–7.

Khush GS. **2001**. Green revolution: the way forward. *Nature Reviews Genetics* **2**: 815–822.

Kilian B, Martin W, Salamini F. **2010**. *Genetic Diversity, Evolution and Domestication of Wheat and Barley in the Fertile Crescent* (M Glaubrecht, Ed.).

Kumar S, Banks TW, Cloutier S. **2012**. SNP Discovery through Next-Generation Sequencing and Its Applications. *International Journal of Plant Genomics* **2012**: 1–15.

Li H. **2013**. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint arXiv:1303.3997*.

Li H, Durbin R. **2009**. Fast and accurate short read alignment with Burrows – Wheeler transform. **25**: 1754–1760.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. **2009**. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078–2079.

Li W, Huang L, Gill BS. **2008**. Recurrent Deletions of Puroindoline Genes at the Grain Hardness Locus in Four Independent Lineages of Polyploid Wheat1,[W],[OA]. *PLANT PHYSIOLOGY* **146**: 200–212.

Ling H-Q, Ma B, Shi X, Liu H, Dong L, Sun H, Cao Y, Gao Q, Zheng S, Li Y, *et al.* 2018.

Genome sequence of the progenitor of wheat A subgenome Triticum urartu. *Nature* **557**: 424–428.

Liu M, Yu M, Li G, Carver BF, Yan L. **2015**. Genetic characterization of aluminum tolerance in winter wheat. *Molecular Breeding* **35**: 1–12.

Lohse M, Drechsel O, Kahlau S, Bock R. **2013**. OrganellarGenomeDRAW--a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. *Nucleic acids research* **41**: 575–581.

Marcussen T, Sandve SR, Heier L, Pfeifer M, Kugler KG, Zhan B, Spannagl M, Pfeifer M, Jakobsen KS, Wulff BBH, *et al.* 2014. Ancient hybridizations among the ancestral genomes of bread wheat. *Science (New York, N.Y.)* **345**: 1250092.

Mayer KFX, Rogers J, Dole el J, Pozniak C, Eversole K, Feuillet C, Gill B, Friebe B, Lukaszewski AJ, Sourdille P, *et al.* 2014. A chromosome-based draft sequence of the hexaploid bread wheat (Triticum aestivum) genome. *Science* **345**: 1251788–1251788.

Meyer M, Kircher M. **2010**. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor Protocols* **5**: 1–11.

Mondini L, Nachit M, Porceddu E, Pagnotta MA. **2012**. Identification of SNP mutations in DREB1, HKT1, and WRKY1 genes involved in drought and salt stress tolerance in durum wheat (Triticum turgidum L. var durum). *Omics : a journal of integrative biology* **16**: 178–87.

Mosulishvili M, Bedoshvili D, Maisaia I. **2017**. A consolidated list of Triticum species and varieties of Georgia to promote repatriation of local diversity from foreign genebanks. *Annals of Agrarian Science* **15**: 61–70.

Munns R, James R a, Xu B, Athman A, Conn SJ, Jordans C, Byrt CS, Hare R a, Tyerman SD, Tester M, *et al.* 2012. Wheat grain yield on saline soils is improved by an ancestral Na^+ transporter gene. *Nature biotechnology* **30**: 360–4.

Nesbitt M. **2016**. *Identification Guide for Near Eastern Grass Seeds*. Routledge.

Nevo E, Chen G. **2010**. Drought and salt tolerances in wild relatives for wheat and barley improvement. *Plant, Cell and Environment* **33**: 670–685.

Okonechnikov K, Conesa A, Garcı F. **2016**. Genome analysis Qualimap 2 : advanced multi-sample quality control for high-throughput sequencing data. **32**: 292–294.

R CoreTeam. **2018**. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2016.

Rahman S, Jolly CJ, Skerritt JH, Wallosheck A. **1994**. Cloning of a wheat 15-kDa grain softness protein (GSP). GSP is a mixture of puroindoline-like polypeptides. *European Journal of Biochemistry* **223**: 917–925.

Raman H, Stodart B, Ryan PR, Delhaize E, Emebiri L, Raman R, Coombes N, Milgate

A. **2010**. Genome-wide association analyses of common wheat (Triticum aestivum L.) germplasm identifies multiple loci for aluminium resistance. *Genome / National Research Council Canada = Génome / Conseil national de recherches Canada* **53**: 957–66.

Raman H, Zhang KR, Cakir M, Appels R, Garvin DF, Maron LG, Kochian L V, Moroni JS, Raman R, Imtiaz M, *et al.* 2005. Molecular characterization and mapping of ALMT1, the aluminium-tolerance gene of bread wheat (Triticum aestivum L.). *Genome* **48**: 781–791.

Ramesh S a, Tyerman SD, Xu B, Bose J, Kaur S, Conn V, Domingos P, Ullah S, Wege S, Shabala S, *et al.* 2015. GABA signalling modulates plant growth by directly regulating the activity of plant-specific anion transporters. *Nature Communications* **6**: 7879.

Reif JC, Zhang P, Dreisigacker S, Warburton ML, van Ginkel M, Hoisington D, Bohn M, Melchinger AE. **2005**. Wheat genetic diversity trends during domestication and breeding. *Theoretical and Applied Genetics* **110**: 859–864.

Ryan PR, Raman H, Gupta S, Sasaki T, Yamamoto Y, Delhaize E. **2010**. The multiple origins of aluminium resistance in hexaploid wheat include Aegilops tauschii and more recent cis mutations to TaALMT1. *The Plant journal : for cell and molecular biology* **64**: 446–55.

Sabir JSM, Arasappan D, Bahieldin A, Abo-Aba S, Bafeel S, Zari TA, Edris S, Shokry AM, Gadalla NO, Ramadan AM, *et al.* 2014. Whole mitochondrial and plastid genome SNP analysis of nine date palm cultivars reveals plastid heteroplasmy and close phylogenetic relationships among cultivars. *PLoS ONE* **9**.

Saintenac C, Jiang D, Akhunov ED. **2011**. Targeted analysis of nucleotide and copy number variation by exon capture in allotetraploid wheat genome. *Genome Biology* **12**: R88.

Salamini F, Ozkan H, Brandolini A, Schäfer-Pregl R, Martin W. **2002**. Genetics and geography of wild cereal domestication in the near east. *Nature reviews. Genetics* **3**: 429–441.

Sasaki T, Yamamoto Y, Ezaki B, Katsuhara M, Ahn SJ, Ryan PR, Delhaize E, Matsumoto H. **2004**. A wheat gene encoding an aluminum-activated malate transporter. *Plant Journal* **37**: 645–653.

Scarcelli N, Mariac C, Couvreur TLP, Faye A, Richard D, Sabot F, Berthouly-Salazar C, Vigouroux Y. **2016**. Intra-individual polymorphism in chloroplasts from NGS data: Where does it come from and how to handle it? *Molecular Ecology Resources* **16**: 434–445.

Stamatakis A. **2014**. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**: 1312–1313.

Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, Greiner S. **2017**. GeSeq - Versatile and accurate annotation of organelle genomes. *Nucleic Acids Research* **45**: W6–W11.

Timonova EM, Leonova IN, Röder MS, Salina EA. **2013**. Marker-assisted development and

characterization of a set of Triticum aestivum lines carrying different introgressions from the T. timopheevii genome. *Molecular Breeding* **31**: 123–136.

Tripathi BN, Müller M. **2015**. *Stress Responses in Plants* (BN Tripathi and M Müller, Eds.). Cham: Springer International Publishing.

Wang W, Vinocur B, Altman A. **2003**. Plant responses to drought, salinity and extreme temperatures: Towards genetic engineering for stress tolerance. *Planta* **218**: 1–14.

Wang S, Wong D, Forrest K, Allen A, Chao S, Huang BE, Maccaferri M, Salvi S, Milner SG, Cattivelli L, *et al.* 2014. Characterization of polyploid wheat genomic diversity using a high-density 90 000 single nucleotide polymorphism array. *Plant Biotechnology Journal* **12**: 787–796.

Wicker T, Matthews DE, Keller B. **2002**. TREP: A database for Triticeae repetitive elements. *Trends in Plant Science* **7**: 561–562.

Wickham H. **2016**. *ggplot2: elegant graphics for data analysis*. Springer.

Winfield MO, Wilkinson PA, Allen AM, Barker GLA, Coghill JA, Burridge A, Hall A, Brenchley RC, D'Amore R, Hall N, *et al.* 2012. Targeted re-sequencing of the allohexaploid wheat exome. *Plant Biotechnology Journal* **10**: 733–742.

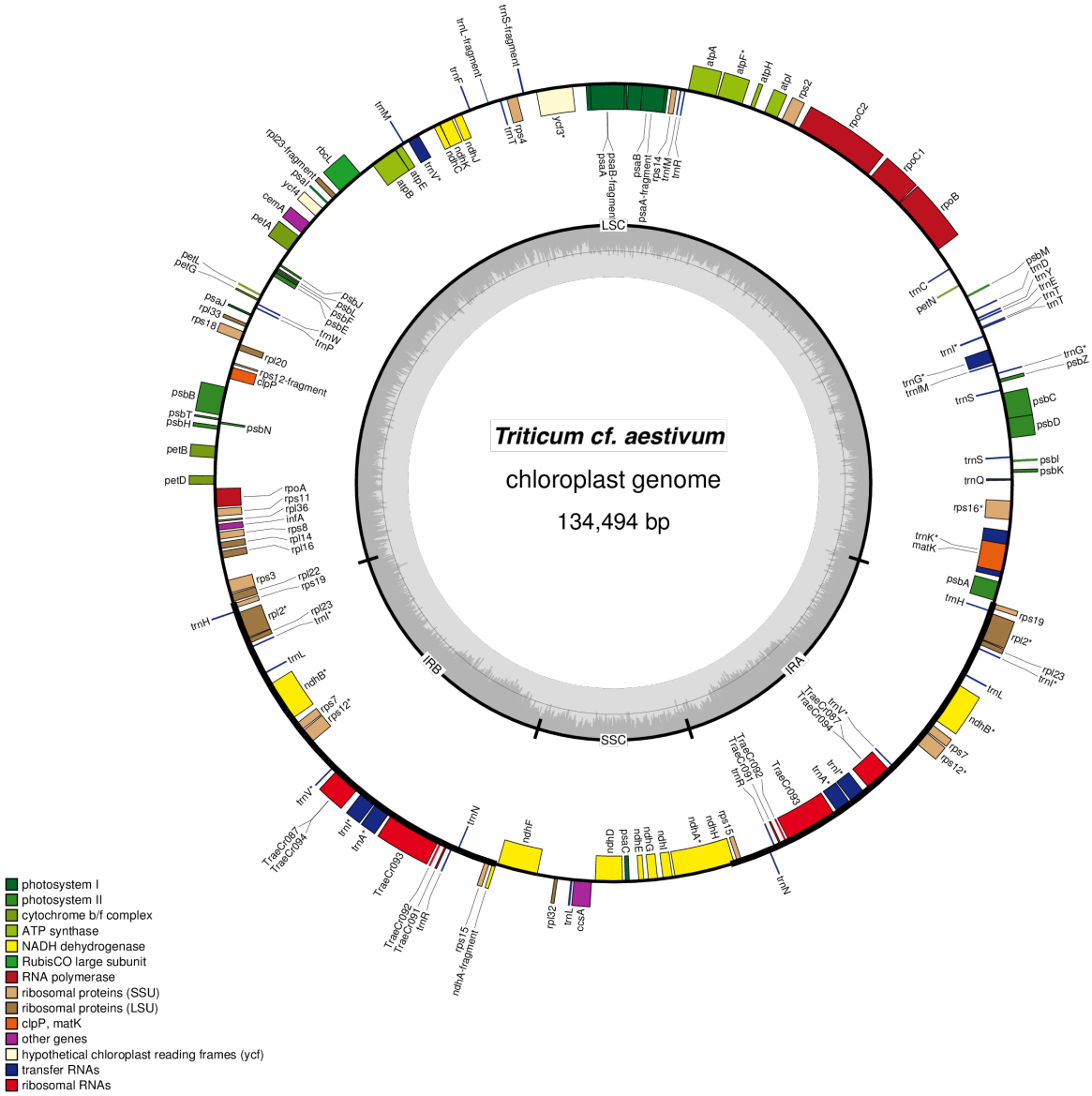
Wolfe AD, Randle CP. **2004**. Recombination, Heteroplasmy, Haplotype Polymorphism, and Paralogy in Plastid Genes: Implications for Plant Molecular Systematics. *Systematic Botany* **29**: 1011–1020.

Yang C, Zhao L, Zhang H, Yang Z, Wang H, Wen S, Zhang C, Rustgi S, von Wettstein D, Liu B. **2014**. Evolution of physiological responses to salt stress in hexaploid wheat. *Proceedings of the National Academy of Sciences of the United States of America* **111**: 11882–7.

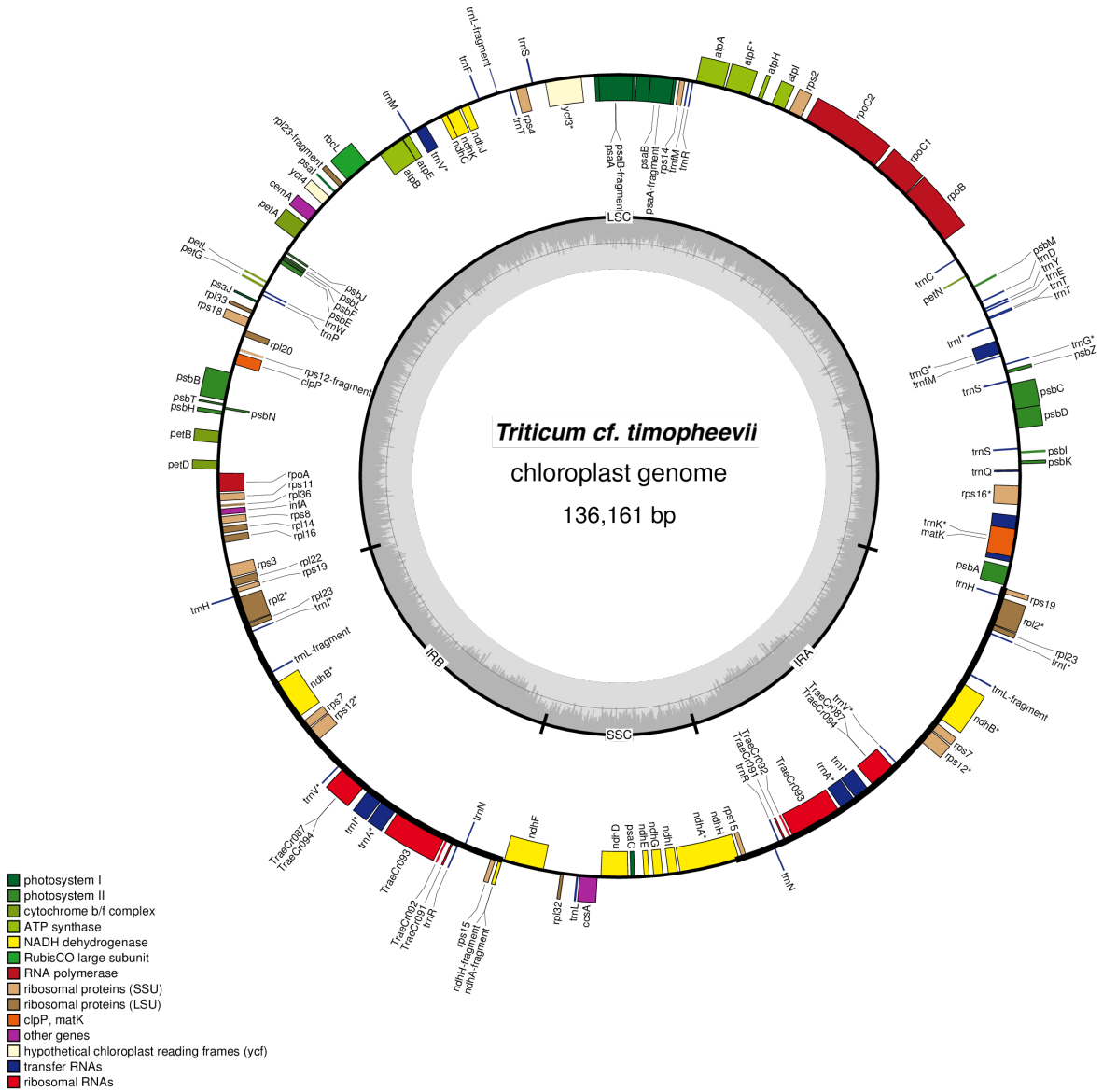
Zamani Babgohari M, Niazi A, Moghadam AA, Deihimi T, Ebrahimie E. **2013**. Genome-wide analysis of key salinity-tolerance transporter (HKT1;5) in wheat and wild wheat relatives (A and D genomes). *In Vitro Cellular & Developmental Biology - Plant* **49**: 97–106.

Zohary D, Hopf M, Weiss E. **2013**. *Domestication of plants in the Old World : the origin and spread of domesticated plants in Southwest Asia, Europe, and the Mediterranean Basin*. Oxford : Oxford University Press.

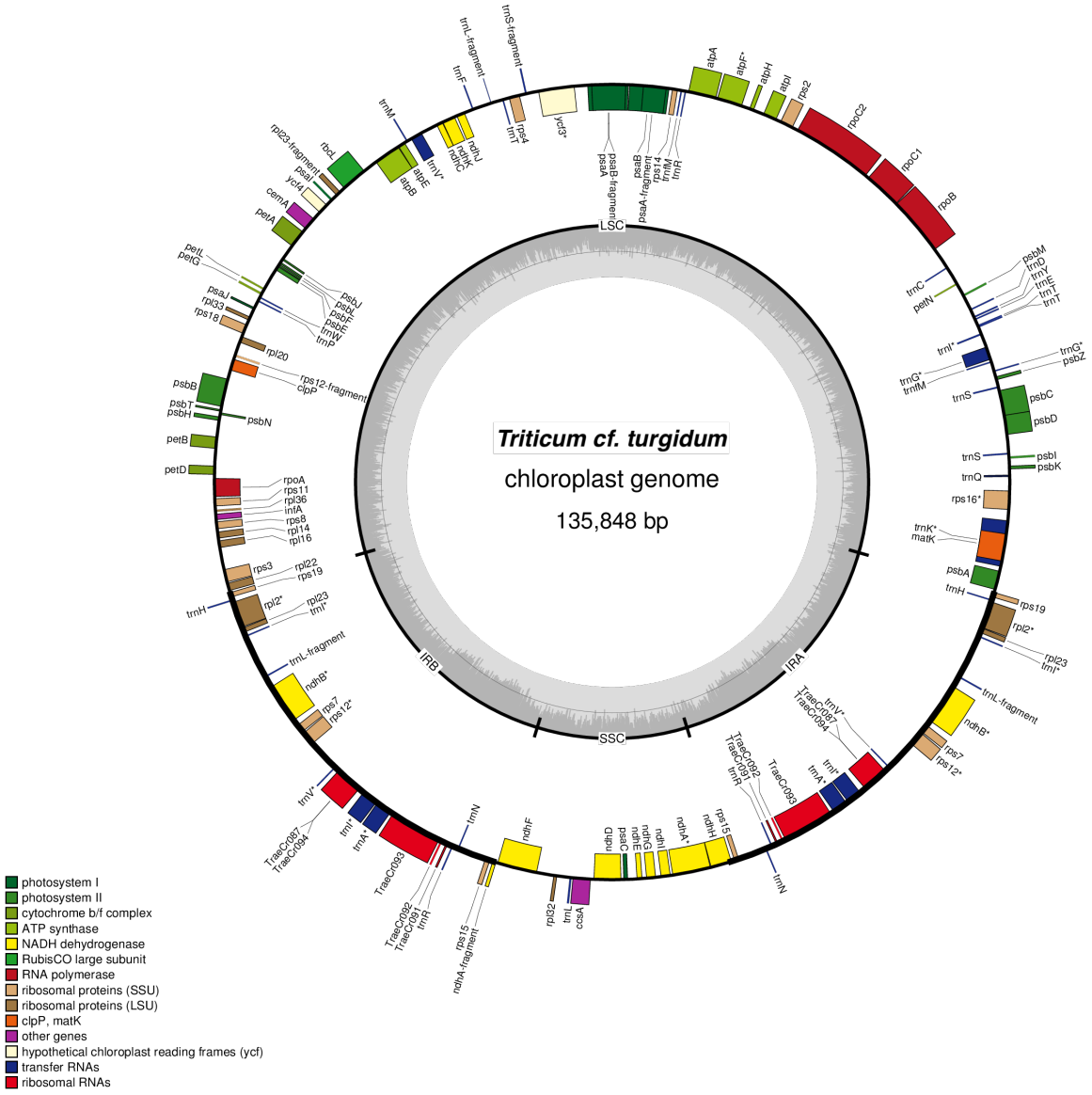
Supplementary information



Supplementary Figure 1. Gene map of the chloroplast genome of historical Georgian wheat *T.* cf. *aestivum*. Accession number ACAD16052. Small single copy (SSC), large single copy (LSC), and inverted repeats (IRa, IRb) are indicated. The bar plots in grey colour represent GC content.



Supplementary Figure 2. Gene map of the chloroplast genome of historical Georgian wheat *T.* cf. *timopheevii*. Accession number ACAD16053. Small single copy (SSC), large single copy (LSC), and inverted repeats (IRa, IRb) are indicated. The bar plots in grey colour represent GC content.



Supplementary Figure 3. Gene map of the chloroplast genome of the historical Georgian wheat *T.* cf. *turgidum*. Accession number ACAD16054. Small single copy (SSC), large single copy (LSC), and inverted repeats (IRa, IRb) are indicated. The bar plots in grey colour represent GC content.

Supplementary Table 1. Mean depth of coverage for 185 nuclear genes sequenced from four historical Georgian polyploid wheat samples. Additional information about gene accession numbers and is given in Table 2.

ID	Gene name	Gene category	*T.* cf. *aestivum* ACAD16052	*T.* cf. *timopheevii* ACAD16053	*T.* cf. *turgidum* ACAD16054	*T.* cf. *turgidum* ACAD16055
001	HvGerA	Disease resistance	146.7	92.3	161.4	128.9
002	HvGerB	Disease resistance	0.6	0.3	0.7	1.0
003	HvGerF	Disease resistance	0.5	0.6	0.9	1.9
004	HvGerD	Disease resistance	14.4	0.6	37.4	83.9
005	HvGLP8	Disease resistance	58.6	19.2	63.4	53.1
006	HvGLP2	Disease resistance	23.0	11.6	24.6	11.5
007	PM3A	Disease resistance	23.4	13.4	24.9	34.2
008	PM3D	Disease resistance	30.0	24.1	65.3	41.5
009	GerF	Disease resistance	0.0	0.0	0.0	0.0
010	GER6a	Disease resistance	15.1	10.2	22.3	24.7
011	MRP1	Disease resistance	94.1	39.5	143.7	143.2
012	HLRG	Disease resistance	137.9	80.9	146.7	170.8
013	Rpg5	Disease resistance	83.6	43.3	99.7	116.3
014	Bln1-3	Disease resistance	38.9	13.6	38.1	21.2
015	Bln2	Disease resistance	60.8	28.2	79.9	48.2
016	RGA2	Disease resistance	6.3	2.5	3.5	14.8
017	RGA2	Disease resistance	1.2	0.4	5.0	0.0
018	TaPrx	Disease resistance	142.7	74.7	178.4	148.5
019	HvPPT1	Disease resistance	39.3	19.6	58.0	64.9
020	Mla1	Disease resistance	198.8	156.2	221.5	249.3
021	Stpk-A	Disease resistance	16.7	12.5	22.9	40.3
022	Stpk-D	Disease resistance	17.2	0.3	28.7	0.4
023	Stpk-B	Disease resistance	24.5	14.2	31.6	50.1
024	Stpk-A(B)	Disease resistance	27.2	9.0	43.9	82.1
025	Stpk-D	Disease resistance	31.3	4.2	52.3	0.0
026	HvRin4	Disease resistance	24.5	11.9	32.6	36.2
027	Pm8	Disease resistance	56.1	32.2	65.6	79.6
028	Pm3-1B	Disease resistance	102.1	58.9	127.9	168.7
029	Germin-9f-2.8	Disease resistance	95.9	41.2	125.5	82.2
030	Germin-9f-3.8	Disease resistance	69.9	49.4	86.0	78.8
031	ACT-1	Domestication	74.9	41.7	113.9	139.1
032	Glu-A1	Domestication	21.8	13.8	34.7	58.9
033	Gsp-1	Domestication	9.2	0.0	11.4	0.0
034	Glu-D1-1a	Domestication	4.8	0.0	0.0	0.0
035	Glu-B1	Domestication	21.7	7.8	35.3	63.3
036	Gsp-1	Domestication	2.5	0.0	7.6	15.0

ID	Gene name	Gene category	*T.* cf. *aestivum* ACAD16052	*T.* cf. *timopheevii* ACAD16053	*T.* cf. *turgidum* ACAD16054	*T.* cf. *turgidum* ACAD16055
037	Gsp-1	Domestication	2.4	4.2	2.3	0.0
038	Gsp-1	Domestication	0.0	0.0	0.0	26.1
039	Pina-D1	Domestication	23.1	13.9	36.2	0.0
040	CesA4	Domestication	40.5	22.1	58.7	68.6
041	HvAP2	Domestication	28.6	15.4	47.5	48.9
042	HvCO1	Domestication	60.1	31.5	88.4	93.4
043	HvCO2	Domestication	72.2	38.9	117.6	129.6
044	HvFT1	Domestication	27.9	14.6	43.4	48.1
045	HvFT2	Domestication	30.0	25.6	33.3	34.8
046	HvFT3	Domestication	73.0	34.5	81.4	108.6
047	HvFT4	Domestication	31.0	17.5	49.0	52.2
048	HvFT5	Domestication	7.0	1.1	6.7	3.1
049	HvFT5	Domestication	0.0	0.0	0.0	0.0
050	HvGA20ox2	Domestication	18.3	12.8	29.5	36.4
051	HvGI	Domestication	88.5	47.8	127.3	145.9
052	HvPhyA	Domestication	49.2	24.1	68.1	61.2
053	HvPhyB	Domestication	79.6	41.1	124.8	152.4
054	HvPhyC	Domestication	66.5	32.7	108.6	126.0
055	HvSOC1	Domestication	57.0	20.3	74.6	126.8
056	HvTFL1	Domestication	85.8	37.3	130.6	149.0
057	HvVRN1	Domestication	64.9	29.2	68.2	60.7
058	HvZCCTa	Domestication	19.2	8.1	33.3	32.2
059	HvZCCTb	Domestication	10.0	5.9	12.2	16.5
060	Pinb-D1	Domestication	27.6	19.1	42.2	0.0
061	Glu-D1-1	Domestication	1.2	0.3	5.4	0.0
062	Glu-D1-2	Domestication	46.1	20.2	61.6	64.1
063	GAPDH	House-keeping	155.3	70.6	220.0	250.1
064	HvLsi1	Transporter	8.7	5.9	12.0	20.3
065	HvLsi2	Transporter	33.1	21.5	35.5	55.7
066	HvLsi6	Transporter	45.9	30.1	59.0	75.2
067	TaBOR1.1	Transporter	9.3	0.7	16.9	0.0
068	TaBOR1.2	Transporter	6.1	3.9	14.3	23.7
069	TaBOR1.3	Transporter	8.9	2.1	17.3	25.9
070	TaLCT1	Transporter	0.0	0.0	0.0	51.1
071	HvST1	Transporter	14.4	8.1	29.6	35.4
072	TaNRT2	Transporter	35.9	16.2	57.7	63.8
073	TaNRT2.1	Transporter	40.8	20.6	67.6	86.2
074	HvPht1-1	Transporter	9.6	3.8	14.9	21.7

ID	Gene name	Gene category	*T.* cf. *aestivum* ACAD16052	*T.* cf. *timopheevii* ACAD16053	*T.* cf. *turgidum* ACAD16054	*T.* cf. *turgidum* ACAD16055
075	HvPht1-6	Transporter	10.1	6.8	17.5	22.1
076	AtSt2	Transporter	16.7	7.0	41.8	27.5
077	HvHak1	Transporter	56.9	29.8	77.1	88.1
078	HvHak4	Transporter	89.7	44.0	134.6	158.3
079	ABC	Transporter	80.5	42.4	114.3	118.1
080	TaSUT1D	Transporter	65.3	30.9	101.2	108.5
081	HvPPA1	Transporter	54.5	26.2	78.0	83.9
082	st1.1b	Transporter	0.4	0.9	0.6	1.5
083	st1.1a	Transporter	0.8	0.8	1.9	2.7
084	st1.1a	Transporter	12.6	10.1	44.0	35.5
085	stA1.1b	Transporter	11.1	7.8	18.3	33.6
086	stD1.1b	Transporter	0.0	0.0	0.1	0.0
087	stB1.1a	Transporter	22.9	16.7	70.8	66.1
088	stD1.1a	Transporter	18.7	1.0	57.9	6.1
089	HvHKT1	Transporter	2.3	2.3	3.8	6.3
090	ZIP7	Transporter	0.0	0.1	0.0	0.0
091	mtp1	Transporter	94.6	44.4	146.4	164.0
092	ST3.5	Transporter	71.7	43.4	101.8	137.0
093	NRT2.3	Transporter	30.8	21.0	47.1	64.0
094	HvPT2	Transporter	14.5	10.7	17.6	17.9
095	HvPT5	Transporter	11.7	6.5	19.7	25.8
096	HvPT7	Transporter	11.4	5.4	15.7	16.6
097	HvPT8	Transporter	35.1	19.8	58.7	68.0
098	HvPT4	Transporter	19.3	18.7	18.9	22.1
099	HvST1	Transporter	14.4	9.3	18.5	17.2
100	PT2-1	Transporter	75.8	38.5	125.3	145.4
101	AMT1	Transporter	17.9	9.9	30.6	41.2
102	Amt2.1	Transporter	33.7	19.8	52.5	56.0
103	Amt1;1	Transporter	9.2	7.6	19.6	27.4
104	Amt1;2	Transporter	7.9	2.1	17.8	27.4
105	NRT1.2	Transporter	74.1	30.7	117.9	135.0
106	NRT1.1	Transporter	88.1	38.9	203.4	125.2
107	ABC	Transporter	112.4	63.1	169.3	196.2
108	MOT	Transporter	5.1	6.7	10.1	13.9
109	TaHKT1	Transporter	32.9	18.2	59.1	96.4
110	ALMT1	Transporter	15.0	4.8	20.1	6.2
111	ALMT1	Transporter	2.2	1.6	2.3	2.0
112	ABC	Transporter	104.3	60.5	134.1	131.4

ID	Gene name	Gene category	*T.* cf. *aestivum* ACAD16052	*T.* cf. *timopheevii* ACAD16053	*T.* cf. *turgidum* ACAD16054	*T.* cf. *turgidum* ACAD16055
113	ZIP3	Transporter	11.6	4.2	19.2	28.6
114	ZIP5	Transporter	28.6	9.6	33.7	31.1
115	ZIP1	Transporter	31.8	18.0	50.7	63.2
116	ZIP7	Transporter	31.2	13.2	49.5	48.7
117	TmHKT8	Transporter	23.5	1.7	41.0	36.9
118	HKT1;5	Transporter	4.7	2.7	4.1	4.0
119	HKT7-A2	Transporter	16.9	14.0	32.2	50.9
120	HKT7-A1	Transporter	35.5	13.5	51.2	53.2
121	ALMT2	Transporter	41.8	18.4	66.6	60.6
122	ALMT3	Transporter	16.8	10.1	33.9	34.5
123	TaSTRG	Transporter	57.6	18.0	109.7	111.7
124	Bot1.b/Bot1(Dp).b	Transporter	14.1	7.4	21.8	22.7
125	CPK1	Transporter	86.3	38.1	115.9	145.7
126	CPK4	Transporter	58.1	24.9	93.0	95.4
127	CPK5	Transporter	93.7	48.5	146.7	172.5
128	CPK7	Transporter	128.3	78.3	162.3	183.5
129	CPK8	Transporter	83.6	42.1	124.6	147.5
130	TaBOR2	Transporter	65.3	36.6	100.1	109.9
131	HvBOR2	Transporter	0.0	0.1	0.0	0.0
132	ZIP3	Transporter	0.1	0.0	0.1	0.0
133	ZIP5	Transporter	28.3	25.1	42.6	48.5
134	ZIP8	Transporter	24.5	9.0	36.5	40.1
135	htk1	Transporter	28.3	0.0	47.2	60.2
136	htk1	Transporter	10.2	0.0	21.1	0.0
137	ST3.1	Transporter	22.9	13.7	31.4	38.9
138	st3.2	Transporter	44.6	22.0	66.9	74.0
139	ST5.1	Transporter	8.3	4.7	18.1	13.0
140	ST5.2	Transporter	31.8	15.4	47.6	58.6
141	hma1	Transporter	69.6	37.2	93.7	102.5
142	HvALMT	Transporter	90.8	50.1	127.2	136.7
143	HvALMT	Transporter	34.1	14.9	57.1	65.5
144	BT1	Transporter	18.6	13.9	35.1	63.4
145	BT1	Transporter	14.3	2.3	25.0	8.5
146	hkt1	Transporter	15.7	0.0	31.3	59.5
147	ZIP1	Transporter	15.9	2.3	20.7	37.6
148	NPF6.7	Transporter	27.2	15.0	45.7	56.6
149	si1	Transporter	8.5	3.7	12.5	6.8
150	HKT4	Transporter	26.3	11.5	40.4	33.7

ID	Gene name	Gene category	*T.* cf. *aestivum* ACAD16052	*T.* cf. *timopheevii* ACAD16053	*T.* cf. *turgidum* ACAD16054	*T.* cf. *turgidum* ACAD16055
151	HKT1	Transporter	43.0	20.6	51.7	68.6
152	MAs	Transporter	42.4	35.9	52.0	59.8
153	MAs	Transporter	94.3	37.0	124.2	114.6
154	TMPIT1	Transporter	128.1	79.8	115.8	102.6
155	HMA2	Transporter	47.8	36.0	53.6	38.2
156	HMA2	Transporter	35.2	22.5	90.3	106.5
157	HMA2	Transporter	19.6	15.5	38.9	39.0
158	HKT1	Transporter	53.7	17.9	85.3	84.2
159	MTP8.1	Transporter	67.5	37.5	95.7	106.7
160	MTP8.2	Transporter	23.3	14.0	39.1	40.6
161	CDPK12	Transporter	2.0	2.6	3.9	6.2
162	Sultr1.1	Transporter	2.0	1.5	3.3	5.9
163	Sultr1.3	Transporter	98.7	61.7	112.6	157.8
164	CT1-5A	Transporter	11.9	10.0	20.7	39.7
165	CT1-5B	Transporter	18.1	5.7	29.5	44.4
166	CT1-5D	Transporter	8.1	0.4	12.1	1.3
167	Bot-B5a	Transporter	25.7	18.0	76.5	7.7
168	HKT1;4-1	Transporter	6.0	0.9	8.0	16.4
169	HKT1;4-2	Transporter	9.3	0.2	18.3	30.7
170	PHT1.2-B1	Transporter	12.1	24.4	26.1	42.5
171	PHT1.10-D1	Transporter	43.5	26.4	93.7	108.6
172	HKT1_5	Transporter	15.3	0.0	35.3	19.3
173	TaHKT2;2	Transporter	18.0	12.0	35.8	60.0
174	TaHKT2;2	Transporter	29.2	5.2	49.7	85.9
175	TaHKT2;2	Transporter	15.4	1.2	33.1	2.9
176	TaHKT2;1	Transporter	48.8	26.2	82.5	105.8
177	HKT8D	Transporter	19.9	1.6	30.6	15.8
178	HvALMT	Transporter	78.7	35.7	118.3	138.0
179	HvALMT	Transporter	20.8	12.9	33.1	41.5
180	HvALMT	Transporter	24.7	12.6	36.1	41.5
181	HvALMT	Transporter	16.1	8.9	24.7	34.9
182	HvALMT	Transporter	0.1	0.0	0.1	0.1
183	HvALMT2	Transporter	10.0	5.8	21.5	16.7
184	pBCH1	Transporter	7.6	6.4	8.7	5.9
185	pBCH2	Transporter	41.6	21.2	63.8	60.4

Supplementary Table 2. Number of single nucleotide variants (SNVs) and SNVs per kilobase (SNVs/kb) detected in 81 nuclear genes sequenced from three historical Georgian polyploid wheat samples. Additional information about gene accession numbers and is given in Table 2.

ID	Gene category	Gene name	Length (bp)	*T. timopheevii* ACAD16053		*T. turgidum* (ACAD16055)		*T. aestivum* (ACAD16052)	
				SNVs	SNVs/Kb	SNVs	SNVs/Kb	SNVs	SNVs/Kb
001	Disease resistance	HvGerA	1243	74	59.5	60	48.3	73	58.7
005	Disease resistance	HvGLP8	2741	50	18.2	67	24.4	62	22.6
008	Disease resistance	PM3D	8816	153	17.4	195	22.1	158	17.9
011	Disease resistance	MRP1	3024	128	42.3	104	34.4	53	17.5
012	Disease resistance	HLRG	2396	58	24.2	56	23.4	61	25.5
013	Disease resistance	Rpg5	5280	254	48.1	247	46.8	232	43.9
015	Disease resistance	Bln2	605	30	49.6	31	51.2	33	54.5
018	Disease resistance	TaPrx	1374	26	18.9	26	18.9	53	38.6
019	Disease resistance	HvPPT1	4345	192	44.2	219	50.4	187	43.0
020	Disease resistance	Mla1	3928	171	43.5	159	40.5	189	48.1
027	Disease resistance	Pm8	7113	210	29.5	222	31.2	201	28.3
028	Disease resistance	Pm3-1B	4527	180	39.8	66	14.6	86	19.0
029	Disease resistance	Germin-9f-2.8	2822	122	43.2	124	43.9	72	25.5
030	Disease resistance	Germin-9f-3.8	3761	146	38.8	136	36.2	97	25.8
031	Domestication	ACT-1	1163	16	13.8	6	5.2	16	13.8
040	Domestication	CesA4	5700	178	31.2	228	40.0	198	34.7
041	Domestication	HvAP2	1015	17	16.7	29	28.6	29	28.6
042	Domestication	HvCO1	3908	184	47.1	192	49.1	190	48.6
043	Domestication	HvCO2	2097	94	44.8	98	46.7	78	37.2
045	Domestication	HvFT2	4008	80	20.0	124	30.9	110	27.4
046	Domestication	HvFT3	1853	76	41.0	67	36.2	61	32.9
047	Domestication	HvFT4	2419	52	21.5	82	33.9	62	25.6
051	Domestication	HvGI	7393	286	38.7	268	36.3	238	32.2
052	Domestication	HvPhyA	1531	65	42.5	84	54.9	72	47.0
053	Domestication	HvPhyB	1336	36	26.9	32	24.0	29	21.7
054	Domestication	HvPhyC	3842	151	39.3	182	47.4	153	39.8
055	Domestication	HvSOC1	1050	47	44.8	75	71.4	62	59.0
056	Domestication	HvTFL1	1410	58	41.1	63	44.7	55	39.0
057	Domestication	HvVRN1	17049	714	41.9	754	44.2	900	52.8
062	Domestication	Glu-D1-2	1980	76	38.4	80	40.4	20	10.1
065	Transporter	HvLsi2	1969	40	20.3	72	36.6	59	30.0
066	Transporter	HvLsi6	1588	58	36.5	69	43.5	55	34.6
072	Transporter	TaNRT2	1796	23	12.8	34	18.9	6	3.3

073	Transporter	TaNRT2.1	1750	25	14.3	35	20.0	54	30.9
077	Transporter	HvHak1	1205	49	40.7	66	54.8	66	54.8
078	Transporter	HvHak4	1183	64	54.1	49	41.4	55	46.5
079	Transporter	ABC	966	53	54.9	47	48.7	45	46.6
080	Transporter	TaSUT1D	6582	284	43.1	278	42.2	134	20.4
081	Transporter	HvPPA1	808	49	60.6	44	54.5	45	55.7
087	Transporter	stB1.1a	4071	17	4.2	20	4.9	10	2.5
091	Transporter	mtp1	1404	45	32.1	37	26.4	32	22.8
092	Transporter	ST3.5	894	19	21.3	11	12.3	16	17.9
093	Transporter	NRT2.3	1744	21	12.0	31	17.8	24	13.8
097	Transporter	HvPT8	2276	89	39.1	51	22.4	97	42.6
098	Transporter	HvPT4	6561	51	7.8	65	9.9	48	7.3
100	Transporter	PT2-1	2094	53	25.3	45	21.5	26	12.4
102	Transporter	Amt2.1	1521	19	12.5	33	21.7	30	19.7
105	Transporter	NRT1.2	1783	46	25.8	45	25.2	25	14.0
106	Transporter	NRT1.1	1811	42	23.2	45	24.8	39	21.5
107	Transporter	ABC	2584	72	27.9	55	21.3	63	24.4
109	Transporter	TaHKT1	1981	17	8.6	4	2.0	5	2.5
112	Transporter	ABC	2998	110	36.7	110	36.7	77	25.7
115	Transporter	ZIP1	1068	36	33.7	2	1.9	13	12.2
121	Transporter	ALMT2	1368	13	9.5	28	20.5	12	8.8
123	Transporter	TaSTRG	879	17	19.3	9	10.2	30	34.1
125	Transporter	CPK1	4118	98	23.8	47	11.4	79	19.2
126	Transporter	CPK4	3757	82	21.8	87	23.2	42	11.2
127	Transporter	CPK5	3552	75	21.1	65	18.3	40	11.3
128	Transporter	CPK7	3175	114	35.9	95	29.9	88	27.7
129	Transporter	CPK8	3492	75	21.5	66	18.9	21	6.0
130	Transporter	TaBOR2	2338	56	24.0	55	23.5	42	18.0
133	Transporter	ZIP5	1053	46	43.7	48	45.6	39	37.0
138	Transporter	st3.2	959	27	28.2	26	27.1	13	13.6
140	Transporter	ST5.2	999	14	14.0	15	15.0	16	16.0
141	Transporter	hma1	9488	366	38.6	394	41.5	358	37.7
142	Transporter	HvALMT	1767	53	30.0	55	31.1	64	36.2
148	Transporter	NPF6.7	834	13	15.6	19	22.8	7	8.4
151	Transporter	HKT1	2319	88	37.9	109	47.0	105	45.3
152	Transporter	MAs	2109	25	11.9	36	17.1	26	12.3
153	Transporter	MAs	2037	78	38.3	83	40.7	71	34.9
154	Transporter	TMPIT1	660	13	19.7	6	9.1	16	24.2
155	Transporter	HMA2	2221	24	10.8	27	12.2	18	8.1
156	Transporter	HMA2	2369	36	15.2	35	14.8	29	12.2
157	Transporter	HMA2	1519	26	17.1	24	15.8	18	11.8

158	Transporter	HKT1	2505	82	32.7	152	60.7	129	51.5
159	Transporter	MTP8.1	1203	25	20.8	24	20.0	23	19.1
163	Transporter	Sultr1.3	2114	43	20.3	33	15.6	43	20.3
171	Transporter	PHT1.10-D1	1578	15	9.5	1	0.6	5	3.2
176	Transporter	TaHKT2;1	2144	15	7.0	15	7.0	20	9.3
178	Transporter	HvALMT	1587	68	42.8	61	38.4	54	34.0
185	Transporter	pBCH2	1741	66	37.9	100	57.4	84	48.2
Total			**224251**	**6859**	**30.6**	**7039**	**31.4**	**6346**	**28.3**

Chapter 5

Ancient chloroplast and nuclear genomes provide insights into the evolutionary history of quinoa (*Chenopodium quinoa* Willd.)

Manuscript prepared for publication

Statement of authorship

Title of Paper	Ancient chloroplast and nuclear genomes provide insights into the evolutionary history of quinoa (*Chenopodium quinoa* Willd.)
Publication Status	☐ Published ☐ Accepted for Publication ☐ Submitted for Publication ☑ Unpublished and Unsubmitted w ork w ritten in manuscript style
Publication Details	Manuscript prepared for submission to the journal *Frontiers in Plant Science*

Principal Author

Name of Principal Author (Candidate)	Oscar Estrada		
Contribution to the Paper	Contributed to the conception and experimental design, collected samples, performed DNA extractions, constructed Illumina DNA sequencing libraries, performed bioinformatics data analyses, performed phylogenetic analyses, interpreted results, wrote and edited the manuscript.		
Overall percentage (%)	85		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	08/03/2019

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	James Breen		
Contribution to the Paper	Contributed to the conception and experimental design, advised the data analyses, reviewed and edited the manuscript.		
Signature		Date	08/03/2019

Name of Co-Author	Stephen M Richards
Contribution to the Paper	Contributed to the conception and design of the manuscript, reviewed and edited the manuscript.
Signature	Date 09/03/2014

Please cut and paste additional co-author panels here as required.

Name of Co-Author	Alan Cooper
Contribution to the Paper	Contributed to the conception and design of the manuscript, reviewed and edited the manuscript.
Signature	Date 8/3/19

Ancient chloroplast and nuclear genomes provide insights into the evolutionary history of quinoa (*Chenopodium quinoa* Willd.)

Oscar Estrada[1], James Breen[2], Stephen M Richards[1], Alan Cooper[1]

[1]Australian Centre for Ancient DNA, University of Adelaide, North Terrace, South Australia 5005, Australia.
[2]Robinson Research Institute, University of Adelaide, SA, 5005, Australia.

Abstract

Quinoa (*Chenopodium quinoa*), an important crop for the global food security, was domesticated in the Andean highlands of South America about 7,000 years ago, where wild relatives still grow today. Archaeological studies have reported quinoa remains dating back to 4,000 years before present, but the ancient genetic diversity remains uncharacterised. Recent sequencing studies of the quinoa nuclear genome suggested domestication took place independently in highland and coastal environments. However, the origin and exact timing of domestication, and the genomic relationships among the tetraploid relatives (*C. hircinum* and *C. berlandieri*) of quinoa are not fully understood. Here, we report the nuclear and chloroplast genome sequences of four ~1,400 years old archaeological samples of *C. quinoa* from the highlands of northern Argentina. Phylogenetic analyses show that wild, weedy and cultivated accessions from different species cluster in a monophyletic clade, indicative of historical hybridisation of closely related tetraploid species of *Chenopodium*. Two of the identified chloroplast haplogroups in the ancient samples appear to be absent from modern quinoa accessions, indicating a genetic bottleneck in the recent past. Further, sequence alignment to the nuclear and chloroplast reference genomes of *C. quinoa* identified a large number of nucleotide polymorphisms in the ancient specimens, providing resources for further analysis of domestication history as well as genetic changes in loci of breeding value.

Introduction

Quinoa (*Chenopodium quinoa* Willd.) is an Andean pseudocereal that is increasingly recognised as a promising crop to sustain food security. The seeds of quinoa have a higher nutritional value than cereal grains, including high protein content (12-22%), lack of gluten, and an excellent balance of amino acids, carbohydrates, vitamins, and minerals essential to the human diet (Vega-Gálvez *et al.*, 2010; Gordillo-Bastidas & Díaz-Rizzolo, 2016; Nowak *et al.*, 2016). Quinoa is native to all countries of the Andean region, from Colombia to the south of Chile (Bazile *et al.*, 2015). Across its natural distribution, quinoa has adapted to a varied range of environments, from sea level in Chile to the highlands in Bolivia at 3,800 meters above sea level (Jacobsen *et al.*, 2003; Bazile *et al.*, 2016). Also, the quinoa crop is highly resilient to different temperatures (from −8°C to 38°C), annual precipitation (80-2000 mm/year), and is tolerant to frost, drought, and salinity (Jacobsen *et al.*, 2003; Hariadi *et al.*, 2011; Bazile *et al.*, 2016). Given the high nutritional value and adaptability of quinoa, the Food and Agriculture Organization of the United Nations (FAO) declared 2013 as the 'International Year of Quinoa' (Bazile *et al.*, 2015).

Archaeological evidence suggests that the domestication of quinoa started approximately 7,000 years ago in the highlands of Argentina, Chile and Peru, when humans started the transition from hunters-gatherers to an agropastoralism-based society (Uhle, 1919; Weber, 1978; Bazile *et al.*, 2015; Hocsman & Babot, 2018). Quinoa, together with potato and maize, was widely cultivated by early South American people and constituted the principal staple food of the Pre-Columbian cultures. (Maughan *et al.*, 2007; Bazile *et al.*, 2015). However, social and environmental changes following the Spanish arrival in the 15th century led to a decline in the cultivation of quinoa (Binford *et al.*, 1997; Chepstow-Lusty *et al.*, 2009; Winkel *et al.*, 2018). It is estimated that in 1961 quinoa was harvested from around 32,435 hectares with a total production of 52,555 tonnes (Figure 1) (FAOSTAT, 2019). It was not until the mid-1970s, when researchers from South American countries discovered the nutritional factors and the ability of quinoa to grow in harsh environments, that it started to be promoted as an important crop (Weber, 1978; Maughan *et al.*, 2007). In 2017, the area under harvest reached 173,242 hectares with a total production of 146,735 tonnes (Figure 1) (FAOSTAT, 2019). Despite the increases in production, breeding efforts are clearly needed to increase quinoa production levels to meet the current and near-future food demands in the world.

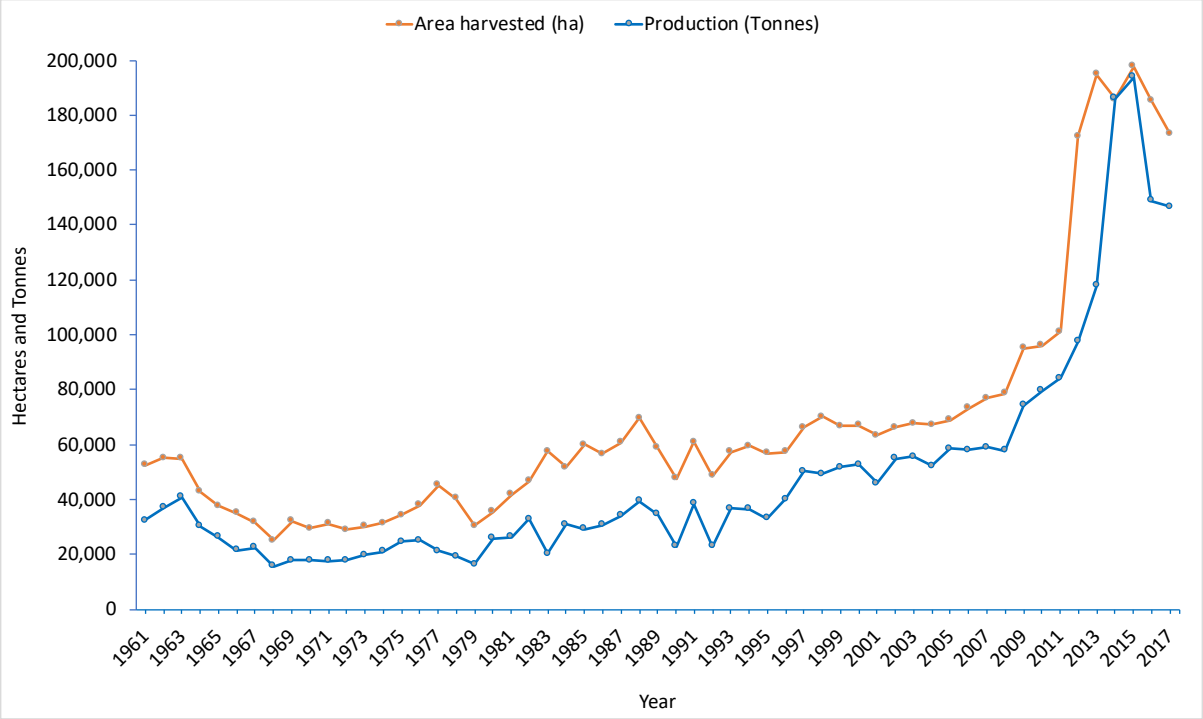


Figure 1. Area harvested (hectares) and production (tonnes) of quinoa worldwide between 1961 and 2017. Data compiled from FAOSTAT (FAOSTAT, 2019).

Chenopodium quinoa is a member of the Amaranthaceae family and botanically related to other important crops such as amaranths (*Amaranthus hypochondriacus* L.), spinach (*Spinacia oleracea* L.), and sugar beet (*Beta vulgaris* L.) (Kadereit *et al.*, 2003). Quinoa is part of a complex of interfertile wild, weedy, and domesticated ecotypes of three separate allotetraploid taxa (Jellen *et al.*, 2011). This interfertile complex includes the weedy South American *C. hircinum* Schrad., the weedy ecotypes of North American *C. berlandieri* Moq., and the extinct or surviving domesticates of *C. berlandieri* (Jellen *et al.*, 2011).

Molecular and cytological data show that *C. quinoa* is an allotetraploid species (2n=4x=36, AABB) that arose from the hybridisation of two unknown diploid species from the A and B genomes around 3.3-6.3 million years ago (Kolano *et al.*, 2016; Jarvis *et al.*, 2017). Molecular markers show that quinoa clusters in two main ecotypes: the coastal ecotype from the lowlands of Chile, and the highland ecotype adapted to the high-altitude environments of Colombia, Peru, Bolivia, Ecuador, Argentina, and extreme northeastern Chile (Wilson, 1988; Christensen *et al.*, 2007). It was traditionally thought that quinoa was domesticated from *C. hircinum* about 7,000 years ago in the Lake Titicaca Basin, with subsequent dispersal north to Colombia and south to the Chilean coast (Wilson, 1988, 1990). The recent sequencing of the quinoa genome has provided evidence that suggests highland and coastal quinoas might have been

independently domesticated (Jarvis *et al.*, 2017; Maughan *et al.*, 2019). However, the evolutionary history, origin, and exact timing of the domestication, and the genomic relationships among diploid and tetraploid relatives of quinoa remain unclear. Furthermore, the genetics of quinoa adaptation to new environmental and agricultural conditions remain largely uncharacterised.

Genetic analysis of archaeobotanical remains can provide a wide range of information about the origin and spread of agricultural species, as demonstrated by ancient DNA (aDNA) studies of important crops such as maize (Fordyce *et al.*, 2013; da Fonseca *et al.*, 2015; Vallebueno-Estrada *et al.*, 2016; Swarts *et al.*, 2017), barley (Mascher *et al.*, 2016) and cotton (Palmer *et al.*, 2012). Desiccated and well-preserved archaeological remains of quinoa (seeds, spikelets, roots, and stems) have been recovered from the highlands of northern Argentina spanning over 1,800 years (Babot & Hocsman, 2015). The genetic diversity of archaeological seeds throughout this record have been analysed using microsatellite markers amplified by PCR (Winkel *et al.*, 2018), although appropriate aDNA authentication criteria (Cooper & Poinar, 2000; Paabo *et al.*, 2004; Llamas *et al.*, 2017) were not used.

To investigate the evolutionary and domestication history of quinoa, we report chloroplast and nuclear sequences of 4 quinoa specimens from northern Argentina radiocarbon-dated at 1,364 years BP (Babot & Hocsman, 2015; Winkel et al., 2018). We investigated the preservation and recovery of aDNA molecules from two types of archaeological specimens: seeds and stems. Additionally, we compared the chloroplast genomes to sequences from several extant *Chenopodium* species and reconstructed the phylogenetic relationships. Furthermore, we characterised the chloroplast, mitochondrial, and nuclear genetic variability of the ancient quinoa.

Materials and Methods

Archaeological material

Ancient plant material was provided by the Institute of Archaeology and Museum of the National University of Tucuman, Argentina. Desiccated quinoa remains including seeds, stems and panicles were excavated from the Alero 1 in the Punta de la Peña 9.I archaeological site in Northern Argentina (Babot & Hocsman, 2015). The archaeological site is located in the dry and cold Andean highlands between 3,600 and 3,700 meters above sea level and corresponds

to agro-pastoral occupations between 1,500 and 1,100 years before present (Babot & Hocsman, 2015). Abundant archaeobotanical materials were found in a sandy matrix together with faunal remains and lithic and ceramic materials (Babot & Hocsman, 2015; Winkel *et al.*, 2018). Quinoa seeds present a variety of colours from white to red or black, depending on the cultivar. In the present study, four quinoa samples were selected for aDNA analysis, including naturally dark-coloured seeds, white seeds, and stems (Figure 2). Quinoa seeds were previously carbon-dated 1364 ± 20 BP (AA-107154, cal. 2σ, 95.4%: 655-766 CE) (Babot & Hocsman, 2015; Winkel *et al.*, 2018).

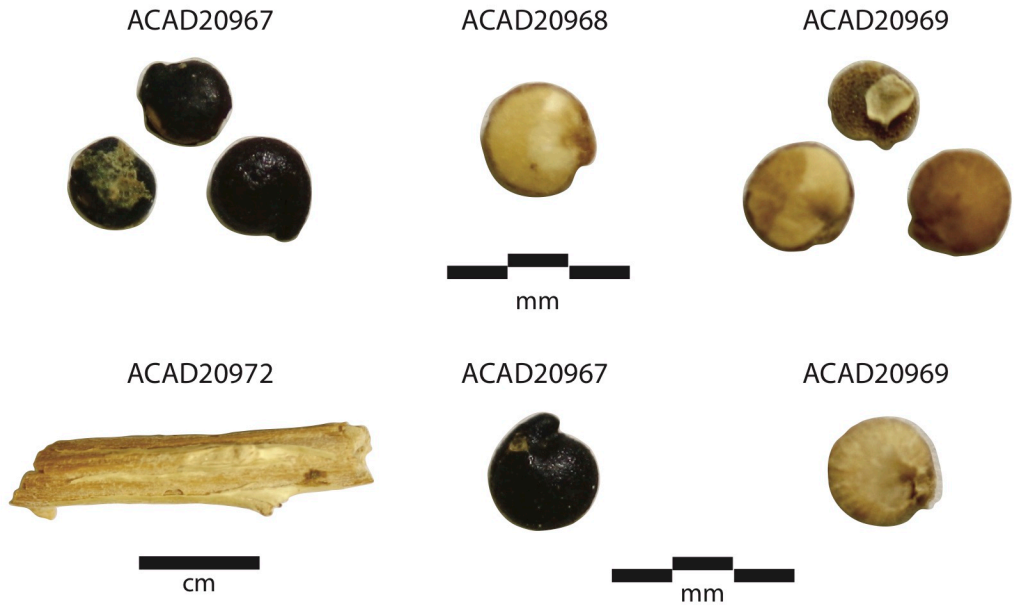


Figure 2. Archaeological remains of quinoa (*C. quinoa*) excavated from the Alero 1 in the Punta de la Peña 9.I archaeological site (Argentina) and analysed in this study.

Ancient DNA extraction

DNA extraction and library preparation steps were performed at the Australian Centre for Ancient DNA at the University of Adelaide using standard approaches for aDNA research (Cooper & Poinar, 2000; Paabo *et al.*, 2004; Llamas *et al.*, 2017) such as complete physical isolation, HEPA filtered positive air flow systems, and regularly cleaning with bleach and UV light to minimize potential DNA contamination. Standard ancient DNA practices were followed including negative controls for all extractions and amplifications.

The archaeological seeds were <2mm in size and weighed ~1.5 mg, making DNA isolation challenging due to the amount of starting tissue. Therefore, to assess the preservation and yield of aDNA molecules in archaeological specimens of quinoa, an initial DNA extraction was conducted on three types of specimen: a quinoa stem of 67.8 mg, a single white seed of 1.6 mg, a bulk sample of three naturally dark-coloured seeds (4.5 mg), and a bulk sample of three white seeds (5.1 mg) (Table 1). After the initial assessment of aDNA content in the different specimens showed positive results, a second round of DNA extractions was conducted from single seeds of the accessions that were previously extracted in the bulk samples. These samples were a single dark seed (1.4 mg), and a single white seed (1.5 mg) (Table 1).

Table 1. Details of the archaeological remains of quinoa (*C. quinoa*) and amount of tissue used for DNA extraction. Specimens marked with (*) were used for a deeper sequencing effort (aiming for a high number of unique reads covering each region of a sequence). The accession numbers of the Australian Centre for Ancient DNA (ACAD) biological collection are provided.

Accession number	Species	Specimen	Weight (mg)
ACAD20967	*C. quinoa*	Bulk of three dark seeds	4.5
ACAD20968	*C. quinoa*	Single white seed*	1.6
ACAD20969	*C. quinoa*	Bulk of three white seeds	5.1
ACAD20972	*C. quinoa*	Stem*	67.8
ACAD20967	*C. quinoa*	Single dark seed*	1.4
ACAD20969	*C. quinoa*	Single white seed*	1.5

The quinoa stem was cut into small pieces with a new sterile scalpel blade and all specimens were washed in a RMS6 rotor (Ratek Instruments, Australia) for two minutes with 1 ml absolute ethanol and rinsed twice with 2 ml of ultrapure water prior to DNA extraction. After rinsing, genomic DNA was isolated from each specimen using the DNeasy PowerPlant Pro Kit (Qiagen, USA) following the manufacturer's directions. In brief, quinoa samples were transferred to a 2 ml bead tube containing the lysis buffer and 2.38 mm metal beads, and ground to powder in a FastPrep 120 instrument (Thermo Savant, USA), for 1 min at 4000 rpm. The lysis buffer was composed of 410 µL of bead solution, 40 µl of the phenolic separation solution, and 50 µL of solution SL. The samples were incubated in the lysis buffer for 24 hours at room temperature in a shaker with constant agitation. DNA was purified according to the instructions of the manufacturer and eluted twice in 50 µl.

DNA library preparation and sequencing

Ancient DNA sequencing libraries were constructed from each sample following the protocol outlined by Meyer & Kircher, (2010), with some modifications. DNA extracts were treated with *E. coli* uracil-DNA-glycosylase (UDG) and endonuclease VIII to minimise misincorporations during sequencing due to deaminated cytosines, which are characteristic of ancient DNA (Gansauge & Meyer, 2014; Rohland *et al.*, 2015). Genomic DNA was enzymatically repaired and blunt-ended using T4 Polynucleotide Kinase PNK (NEB, USA) and T4 DNA polymerase (NEB, USA). Custom Illumina P5 and P7 adapters with seven bp internal barcodes were ligated to the DNA using T4 ligase (ThermoFisher Scientific, Germany). The Illumina DNA libraries were completed by PCR amplification using GAII-index and sequencing primers. Each library was split into eight PCR replicates as a means to reduce clonality. Individual PCR reactions of 25 µl consisted of 5 µl of library DNA, 1X High Fidelity PCR buffer, 2 mM $MgSO_4$, 0.2 mM dNTPs, 0.2 mM of each primer, and 0.1 µl Platinum Taq DNA Polymerase High Fidelity (5 U/µL). The following thermocycling conditions were utilised: 94 ˚C for 2 min, 13 cycles of 94 ˚C for 10 s, 58 ˚C for 30 s, and 68 ˚C for 40 s. Indexed libraries were purified using Ampure XP magnetic beads (Beckman Coulter, USA) and quantified using the Agilent 2200 TapeStation (Agilent Technologies, USA).

Four DNA libraries generated for initial screening were pooled at equimolar concentrations and sequenced on the Illumina NextSeq platform (2 x 75 bp, paired-end) (Illumina, USA) at the Australian Cancer Research Foundation (ACRF) Cancer Genomics Facility in South Australia. Following the initial screening, four libraries were selected for deeper sequencing of the DNA extracts of single seeds and stem, aiming for a high number of unique reads covering each region of a sequence. These libraries were pooled at 1.5 nM and run on an Illumina HiSeq X Ten using 300 cycles (2 x 150 bp, paired-end) sequencing chemistry at the Garvan Institute of Medical Research, Sydney, Australia.

Data analysis

Sequencing reads were demultiplexed according to the unique P5 and P7 barcoded adapters using S*abre* version 1.0 (https://github.com/najoshi/sabre) with default parameters and no mismatches permitted. Demultiplexed reads were trimmed of adapters and merged using *BBmerge* (*BBtools* version 36.64 https://sourceforge.net/projects/bbmap), keeping overlapped reads with base quality higher than four and length higher than 25 bp. The qualities of the reads

were checked using *fastQC* version 0.11.5 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc) before and after adapter trimming.

Merged sequencing reads were aligned to the nuclear (NCBI BioSample accession: SAMN04338310) (Jarvis *et al.*, 2017), chloroplast (NCBI accession: NC_034949.1) (Hong *et al.*, 2017) and mitochondria (NCBI accession: MK182703) reference genomes of *C. quinoa* using the *BWA aln* and *samse* algorithms version 0.7.15 (Li & Durbin, 2009; Li, 2013) with recommended mapping parameters for aDNA (-l 1024 -n 0.01 -o 2) (Schubert *et al.*, 2012). Binary sequence alignment files were created using *SAMTools* version 1.3.1 (Li *et al.*, 2009) keeping reads with a minimum mapping quality of 30. Alignment files were coordinate sorted and PCR duplicates removed using *SortSam* and *MarkDuplicates* from the *Picard* package version 2.1.1 (https://broadinstitute.github.io/picard). Aligned nuclear sequences were analysed for patterns of DNA damage such as fragment size distribution and C to T and G to A substitutions using *mapDamage* version 2.0.6 (Jónsson *et al.*, 2013).

Summary statistics for each sequenced library such as number of mapped reads, number of unique reads, endogenous content, library complexity, clonality, coverage, GC content, and fragment lengths were estimated using *SAMTools* version 1.3.1 (Li *et al.*, 2009) and *Qualimap* version 2.2.1 (Okonechnikov *et al.*, 2016). Plots and figures were generated using R 3.5.1 base graphics (R CoreTeam, 2018) and the ggplot2 package (Wickham, 2016).

Evaluation of ancient DNA preservation in archaeological quinoa

The shotgun sequencing data was used to calculate initial aDNA yield from the different types of specimens (stem, single seed, and bulk samples of three seeds). To assess whether the DNA extraction from single seeds or bulk seeds affected the number of genotypes recovered in a DNA library, we searched for heterozygous variants in chloroplast sequences using *freebayes* version 1.2.0 (Garrison & Marth, 2012), with a minimum coverage of 5x to call a variant.

Characterisation of the nuclear genome

We surveyed the variant profiles of the nuclear genes of each ancient specimen using the quinoa nuclear genome as a reference (NCBI BioSample accession: SAMN04338310) (Jarvis *et al.*, 2017). Gene information from the quinoa reference genome was downloaded from the *Phytozome* database (http://www.phytozome.net) (Goodstein *et al.*, 2012) and variant profiles

were computed for the overlapped regions using *freebayes* version 1.2.0 (Garrison & Marth, 2012) with a minimum coverage of 5x to call a variant. Quinoa is an allotetraploid species, and therefore multiple distinct copies of each nuclear gene derived from a different ancestor (homoeologous locus) are expected. We estimated the variant profiles using *freebayes,* by analysing sequences of polyploid organisms to detect multi-allelic haplogroups based on the most likely combination of genotypes at each position in the reference sequence (Garrison & Marth, 2012).

Chloroplast phylogenetic analysis

Merged reads from each ancient library were mapped as before to the chloroplast genome of *C. quinoa* (NCBI accession: NC_034949.1) with one of the two inverted repeats. Consensus sequences and variant profiles were generated from the unique mapped alignments using *Geneious* 10.2.6 (https://www.geneious.com). The *'Highest Quality Threshold'* setting in *Geneious* was selected to produce the consensus sequences and a minimum of 5X coverage and frequency of one (100% of the alignments) to call a variant. The consensus sequences from all four specimens were aligned with 26 previously published chloroplast genomes of species from the *Chenopodium* genus and one outgroup (Table 1) using *MAFFT* (Katoh & Standley, 2013) in *Geneious* 10.2.6 (https://www.geneious.com). Following the multiple sequencing analysis, a maximum likelihood (ML) phylogenetic tree was built using *RAxML GTRGAMMA* model (Stamatakis, 2014) in *Geneious* 10.2.6 (https://www.geneious.com) with 1,000 bootstrapping replicates. The chloroplast genomes of the hexaploid of *C. album* (NCBI accession: NC_034950.1) was used as outgroup (Table 1). *RAxML* version 8.2.0 was also used to perform a *Shimodaira–Hasegawa* test (Shimodaira, 1998; Shimodaira & Hasegawa, 1999) to compare our best tree topology to three alternative topologies (using the "-f h" command): (1) a tree where all *C. quinoa* and *C. hircinum* samples form a monophyletic clade; (2) a tree where all *C. berlandieri* samples cluster in a monophyletic clade; and (3) a tree where *C. berlandieri* constitutes a monophyletic group and is a sister-taxon to a monophyletic clade of all *C. quinoa* and *C. hircinum* samples. The phylogenetic tree was visualised and edited using *Figtree* version 1.4.4 (http://tree.bio.ed.ac.uk/software/figtree).

Table 2. Details of the chloroplast genomes of 26 species of *Chenopodium* used in this study for phylogenetic analyses. The chloroplast genome of *C. album* was used as outgroup. Samples IDs for most of the species were taken from (Maughan *et al.*, 2019). (*) denotes NCBI accession numbers.

Species	ID	Origin	Type	Ploidy
C. album	NC_034950.1*	Korea	Wild/weedy	2n=6x=54
C. berlandieri subsp. *nuttalliae*	PI 568156	Mexico	Cultivated (huauzontle)	2n=4x=36
C. berlandieri var. *boscianum*	BYU 937	Texas, US	Wild/weedy	2n=4x=36
C. berlandieri var. *macrocalycium*	PI 666279	Maine, US	Wild/weedy	2n=4x=36
C. berlandieri var. *sinuatum*	Ames 33103	Arizona, US	Wild/weedy	2n=4x=36
C. berlandieri var. *zschackei*	BYU 1314	Utah, US	Wild/weedy	2n=4x=36
C. hircinum	BYU 1101	Argentina	Weedy (pampas)	2n=4x=36
C. hircinum	BYU 566	Chile	Weedy (desert valley)	2n=4x=36
C. pallidicaule	Ames 13221	Bolivia	Cultivated (kaniwa)	2*n*=2*x*=18
C. quinoa	NC_034949.1*	Peru	Highland	2n=4x=34
C. quinoa	MK159176.1*	Chile	Coastal	2n=4x=35
C. quinoa	0654	Peru	Highland	2n=4x=36
C. quinoa	Cherry Vanilla	Oregon, US	Coastal	2n=4x=36
C. quinoa	Chucapaca	Bolivia	Highland	2n=4x=36
C. quinoa	CICA-17	Peru	Highland	2n=4x=36
C. quinoa	G-205-95DK	Denmark	Coastal	2n=4x=36
C. quinoa	Ku-2	Chile	Coastal	2n=4x=36
C. quinoa	Kurmi	Bolivia	Highland	2n=4x=36
C. quinoa	Ollague	Chile	Highland	2n=4x=36
C. quinoa	Pasankalla	Peru	Highland	2n=4x=36
C. quinoa	PI 614886	Chile	Coastal	2n=4x=36
C. quinoa	PI 634921	Chile	Coastal	2n=4x=36
C. quinoa	Real	Bolivia	Highland	2n=4x=36
C. quinoa	Regalona	Chile	Coastal	2n=4x=36
C. quinoa	Salcedo INIA	Peru	Highland	2n=4x=36
C. suecicum	328 / 6	Czech Republic		2*n*=2*x*=18

Results

Ancient DNA preservation in archaeological quinoa remains

Genome-wide shotgun sequencing of the four ancient DNA libraries yielded between 7.2 - 16.1 million paired-end reads (Table 3). The percentage of endogenous content was calculated as the fraction of unique reads mapping to the reference genomes, and revealed most samples possessed around 42% endogenous nuclear content, except for the stem specimen which had ~30% (Table 3 and Figure 3). Sequence clonality, estimated as the proportion of the aligned reads that are PCR duplicates, ranged between 2.27% and 7.68%. The least diverse library was the white single seed (7.68% PCR duplicates), versus the bulk sample of white seeds (3.68%), the stem (3.28%), or the bulk sample of dark seeds (2.27%) (Figure 3). The relatively low clonality indicates a high level of complexity in the DNA libraries, suggesting they are good candidates for deeper sequencing efforts.

Table 3. Summary of the shotgun sequencing results, endogenous content, clonality, and average fragment read length of four archaeological specimens of quinoa *C. quinoa*. The accession numbers of the Australian Centre for Ancient DNA (ACAD) biological collection are provided.

Accession	Specimen	Total number of reads	Endogenous nuclear reads	Clonality	Average read length (bp)
ACAD20967	Bulk sample 3 dark seeds	16,108,768	43.88%	2.27%	43.88
ACAD20968	Single white seed	13,191,794	41.80%	7.68%	41.80
ACAD20969	Bulk sample 3 white seeds	11,872,795	41.91%	3.68%	41.91
ACAD20972	Stem	7,208,930	29.70%	3.28%	29.70

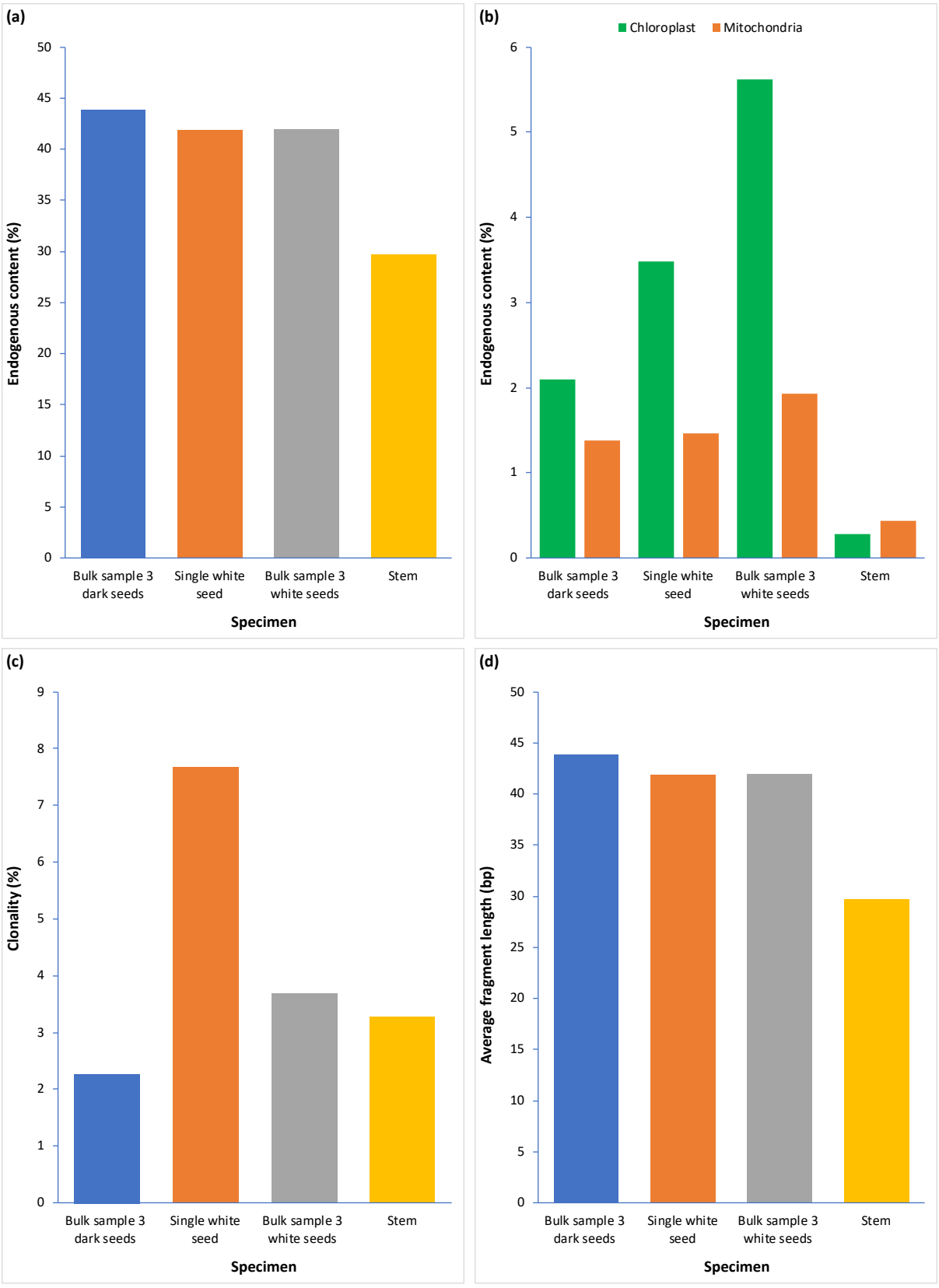


Figure 3. Plot of the shotgun sequencing results of four archaeological specimens of quinoa *C. quinoa*. (a) Nuclear endogenous content. (b) Chloroplast and mitochondria endogenous content. (c) Clonality. (d) Average fragment length.

For all specimens, the percentage of endogenous reads mapping to the mitochondrial and chloroplast genomes were less than 2% and 6%, respectively (Figure 3). The quinoa stem yielded a lower percentage of unique reads mapping to the chloroplast (0.28%) and mitochondria (0.28%) than those constructed from seeds (mean chloroplast= 3.73%, mean mitochondria= 1.59%). Libraries generated from white seeds (single and bulk seeds) produced more endogenous chloroplast and mitochondrial reads (mean= 4.55% and 1.70%) than those retrieved from dark seeds (2.09% and 1.38%). The bulk sample of white seeds generated almost double endogenous chloroplast reads (5.62%) than single white seeds (3.48%), but only a slightly higher proportion of mitochondria content (single white seed= 1.46% and bulk sample of white seeds= 1.93%) (Figure 3). Although the use of bulk samples increases the yield of unique chloroplast sequences, this does not have a significant impact in the number of nuclear and mitochondrial sequences recovered. However, the increase in chloroplast sequences in the bulk sample may be due to the presence of different genotypes in the DNA library.

To further examine the number of genotypes recovered from single or multiple seed samples, we used *freebayes* version 1.2.0 (Garrison & Marth, 2012) to identify heterozygous variants in the chloroplast alignment files. The bulk seed samples contained about six times more heterozygous variants than those constructed from the single seeds or the stem (Figure 4). This increased in the number of heterozygous variants in the bulk seed samples is likely to represent alleles from different plastid haplogroups, which can come from either heteroplasmies (presence of more than one type of organellar genome within a cell or individual) or haplogroups belonging to the different individual seeds. The use of multiple seeds did increase the percentage of chloroplast and mitochondrial reads, somewhat strangely, but not the nuclear reads – as expected (Figure 3). These results, however, are dramatically impacted by the preservation of the samples.

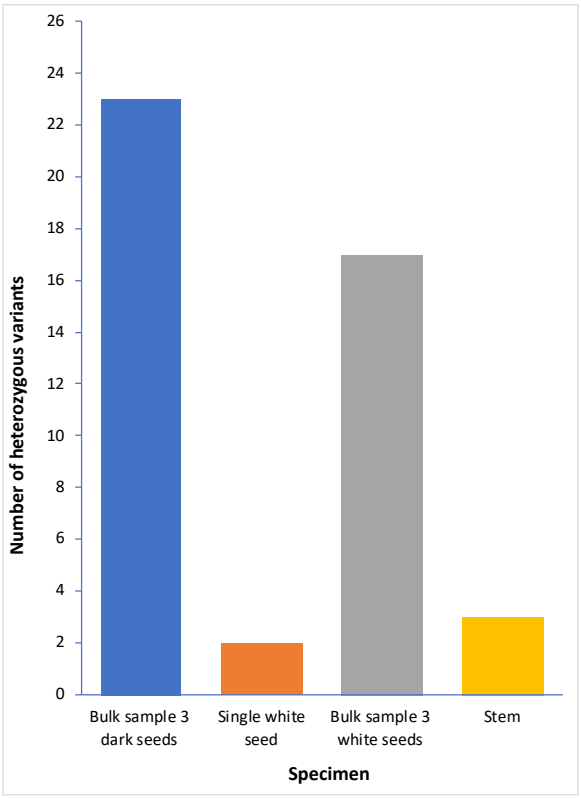


Figure 4. Number of heterozygous variants identified in the chloroplast sequence alignments of four archaeological specimens of quinoa *C. quinoa*.

A powerful quality control analysis in ancient DNA studies is to confirm whether the sequences have characteristic profiles of degraded DNA samples, such as fragmented molecules and signs of DNA damage at the end of fragments (Knapp & Hofreiter, 2010; Orlando *et al.*, 2015). We used *mapDamage* version 2.0.6 (Jónsson *et al.*, 2013) to assess patterns of DNA damage in all the ancient quinoa samples. The mean read length of endogenous nuclear DNA for the archaeological specimens ranged from 57.4 to 78.1 bp, with an overall mean of 71.1 bp (Figure 3). All ancient quinoa specimens showed a high proportion of C to T and G to A misincorporations at the first and last positions of DNA fragments (Supplementary Figures 1-4). This pattern of misincorporations together with the small fragment length distribution is consistent with the characteristics of degraded ancient DNA (Briggs *et al.*, 2007; Dabney *et al.*, 2013; Jónsson *et al.*, 2013).

Deep sequencing results

The genetic diversity within the four samples identified for deep sequencing (Table 1) was consistent with the preliminary screening. Between 105 and 225 million reads were produced and revealed ~42-45% nuclear endogenous content for the seeds, and ~30% for the stem specimen (Table 4). Chloroplast reads varied from 0.27% (stem) to 3.36% (white seed) and were generally slightly higher than mitochondrial reads (0.42-1.83%) although the stem was proportionally elevated at 0.42% (Table 4).

Table 4. Summary of the deep sequencing results and endogenous content of four archaeological specimens of quinoa *C. quinoa*. The accession numbers of the Australian Centre for Ancient DNA (ACAD) biological collection are provided.

Accession	Specimen	Total number of reads	Endogenous chloroplast	Endogenous mitochondria	Endogenous nuclear
ACAD20967	Dark seed	225,658,175	1.96%	1.29%	47.09%
ACAD20968	White seed	176,311,686	3.28%	1.36%	44.59%
ACAD20969	White seed	159,257,423	3.36%	1.83%	44.90%
ACAD20972	Stem	105,392,444	0.27%	0.42%	29.71%

Characterisation of the ancient nuclear genomes of quinoa

We generated the genome sequences of four ~1,400 years BP archaeological samples of *C. quinoa*. Encouragingly, the whole-genome shotgun sequencing of the ancient quinoa specimens recovered between 25.45% and 57% of the 1.39 gigabases (Gb) of the nuclear reference genome of *C. quinoa* (NCBI BioSample accession: SAMN04338310) (Jarvis *et al.*, 2017) (Figure 5). The sequence alignments of each sample appeared to be are evenly distributed across the nuclear genome (Supplementary Figures 6-9) with a mean coverage of 5.2X for the dark seed (ACAD20967), 2.6-3.2X for the white seeds (ACAD20968 and ACAD20969) and 0.9X for the stem (ACAD20972) (Figure 5).

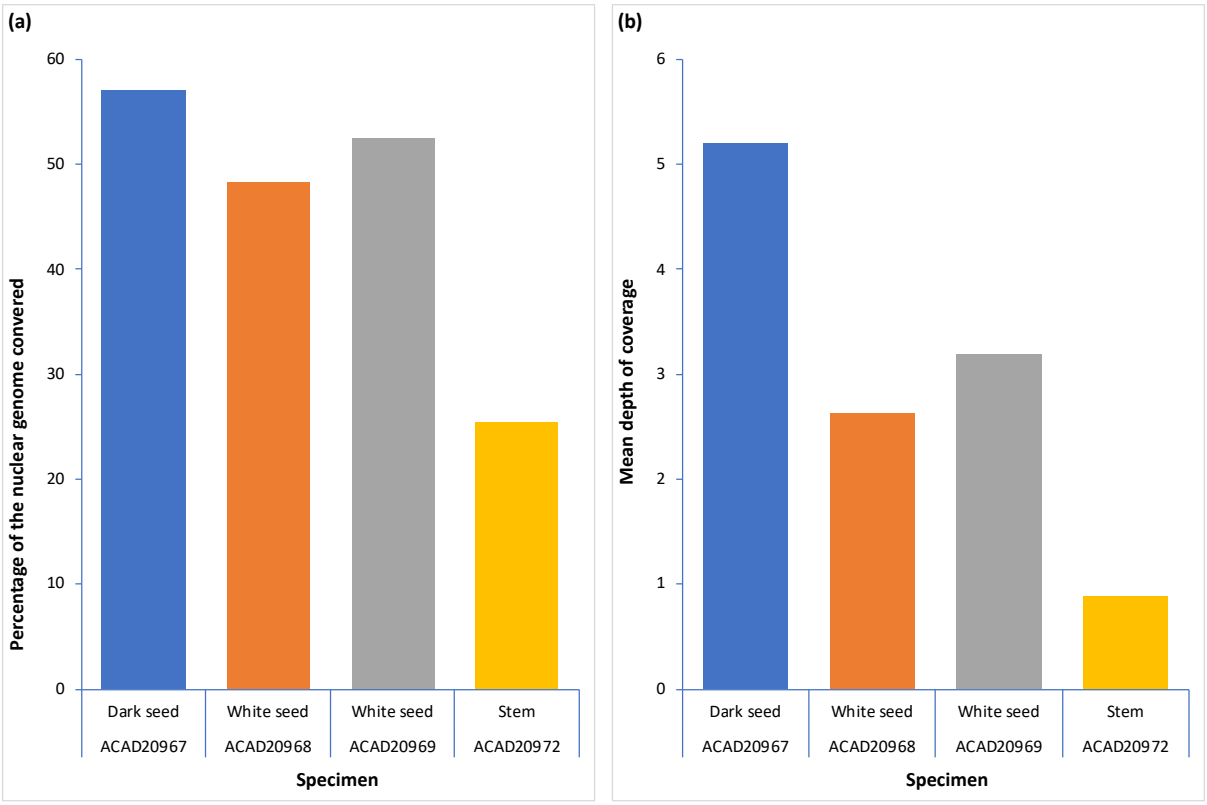


Figure 5. Percentage (a) and mean depth of coverage (b) of the nuclear genome of four archaeological specimens of quinoa *C. quinoa*.

Using the coordinates of the nuclear genes in the quinoa genome (44,776 genes ~57 Megabase-pairs), we characterised the variant profiles of the ancient quinoa specimens to identify an estimated level of diversity between the ancient and modern samples. A total of 253,843 variants, including 247,468 single nucleotide variants (SNVs) and 6,375 indels (insertions/deletions), were identified in the ancient quinoa samples (Figure 6). The dark seed specimen (ACAD20967) held 43% of the total nuclear gene variants (106,361 SNVs and 2,903 indels) while the two white seeds presented much less, 21.2% (ACAD20968: 52,605 SNVs and 1,259 indels) and 26.9% (ACAD20969: 66,797 SNVs and 1,613 indels). The stem specimen exhibited the lowest number of nuclear gene variants with 8.8% (21,705 SNVs and 600 indels) (Figure 6). As expected, these results are influenced by the low average coverage of the target regions for the stem (ACAD20972 1.38x) and white seeds specimens (ACAD201968: 4x and ACAD20969: 5x) compare to the dark seeds (ACAD20967 9x), which limits the number of variants that can be called at a minimum coverage of 5X.

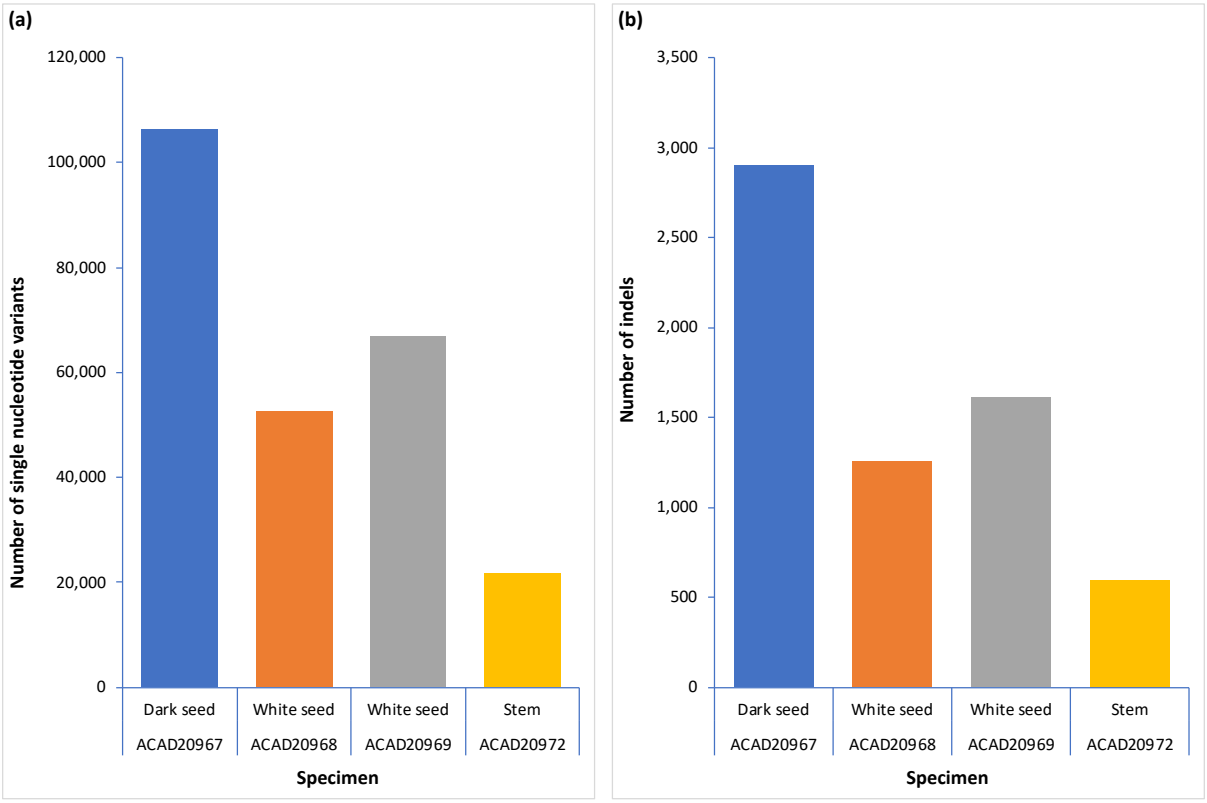


Figure 6. Number polymorphisms detected in 44,776 genes of four archaeological specimens of quinoa *C. quinoa*. (a) Number of single nucleotide variants (SNVs). (b) Number of insertions and deletions (indels).

Phylogenetic characterisation of ancient quinoa chloroplast genomes

To assess the variability of the chloroplast genomes of the ancient samples of quinoa, sequencing reads were mapped to the reference chloroplast genome of *C. quinoa* without one of the two inverted repeats. The inverted repeat regions within the chloroplast are almost identical to each other and combined with the small fragment size of ancient DNA are almost impossible to accurate map. We recovered the chloroplast genome from the four ancient quinoa specimens with a mean coverage of 139.3X for the dark seed (ACAD20967), 122.9X and 169.2X for the white seeds ACAD20968 and ACAD20969, and 29.3X for the stem (ACAD20972). A total of 1,680 polymorphisms, including 1,616 single nucleotide variants (SNVs) and 64 indels (insertions/deletions), were identified in the four ancient quinoa samples. As for the nuclear variants, the dark seed (ACAD20967) held most (43%) of the chloroplast diversity (697 SNVs and 24 indels) while the white seeds had 9-24% (ACAD20969: 128 SNVs and 18 indels; and ACAD20968: 394 SNVs and 16 indels) (Figure 7). Unlike the nuclear genomic diversity, the stem specimen held a similar or greater amount of chloroplast genetic diversity than the white seeds, around 24% of the variants (397 SNVs and six indels) (Figure 7).

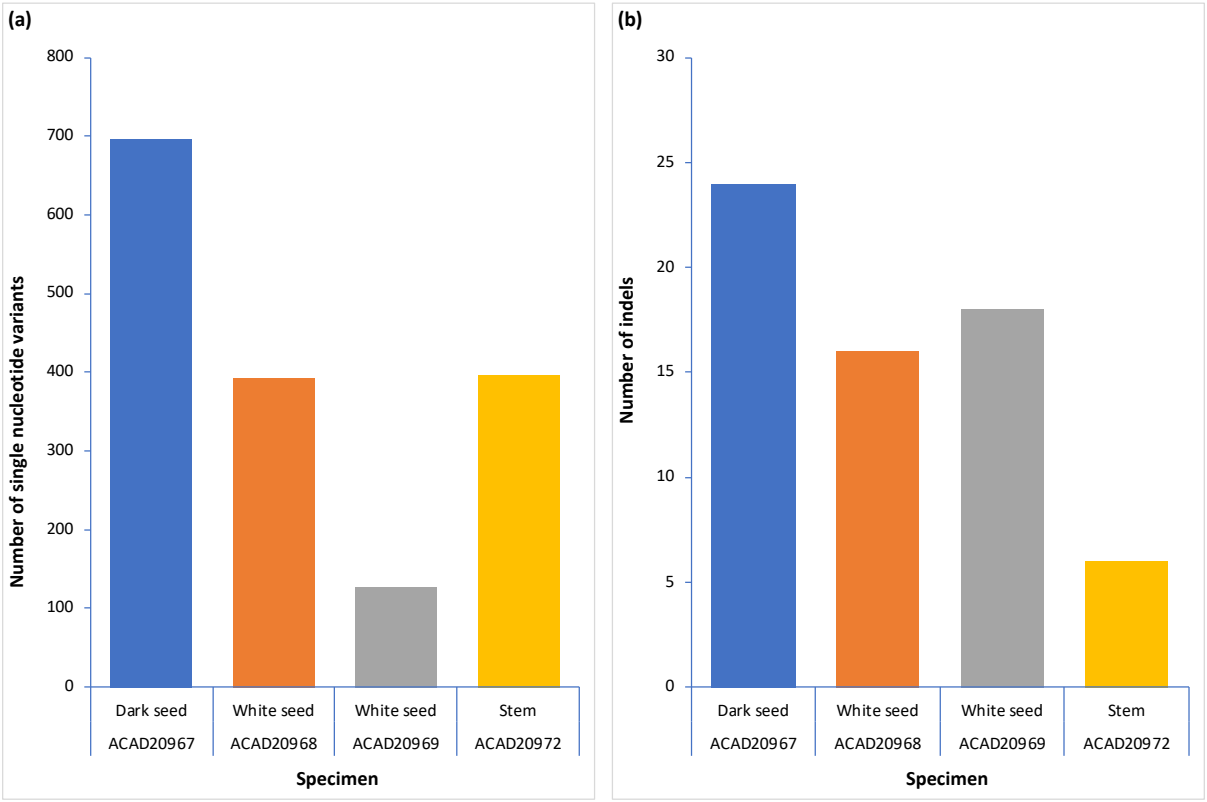


Figure 7. Number polymorphisms detected in the chloroplast genomes of four archaeological specimens of quinoa *C. quinoa*. (a) Number of single nucleotide variants (SNVs). (b) Number of insertions and deletions (indels).

In order to reconstruct the phylogenetic history of the archaeological samples of quinoa, we aligned the ancient chloroplast sequences together with the chloroplast genomes of 26 species of the *Chenopodium* genus (Table 2). Phylogenetic analyses revealed strong support for a monophyletic clade including all the tetraploid species to the exclusion of the diploid species, but the monophyly of the individual tetraploid species was not supported (Figure 8 and Supplementary Figure 5). The diploid *C. pallidicaule* was the sister-taxon to the tetraploid clade with 100% bootstrap support (Figure 8 and Supplementary Figure 5), concordant with the hypothesis that an ancestral species of the *pallidicaule* lineage was the maternal donor of the plasmons to the tetraploid species of *Chenopodium* (Maughan *et al.*, 2019).

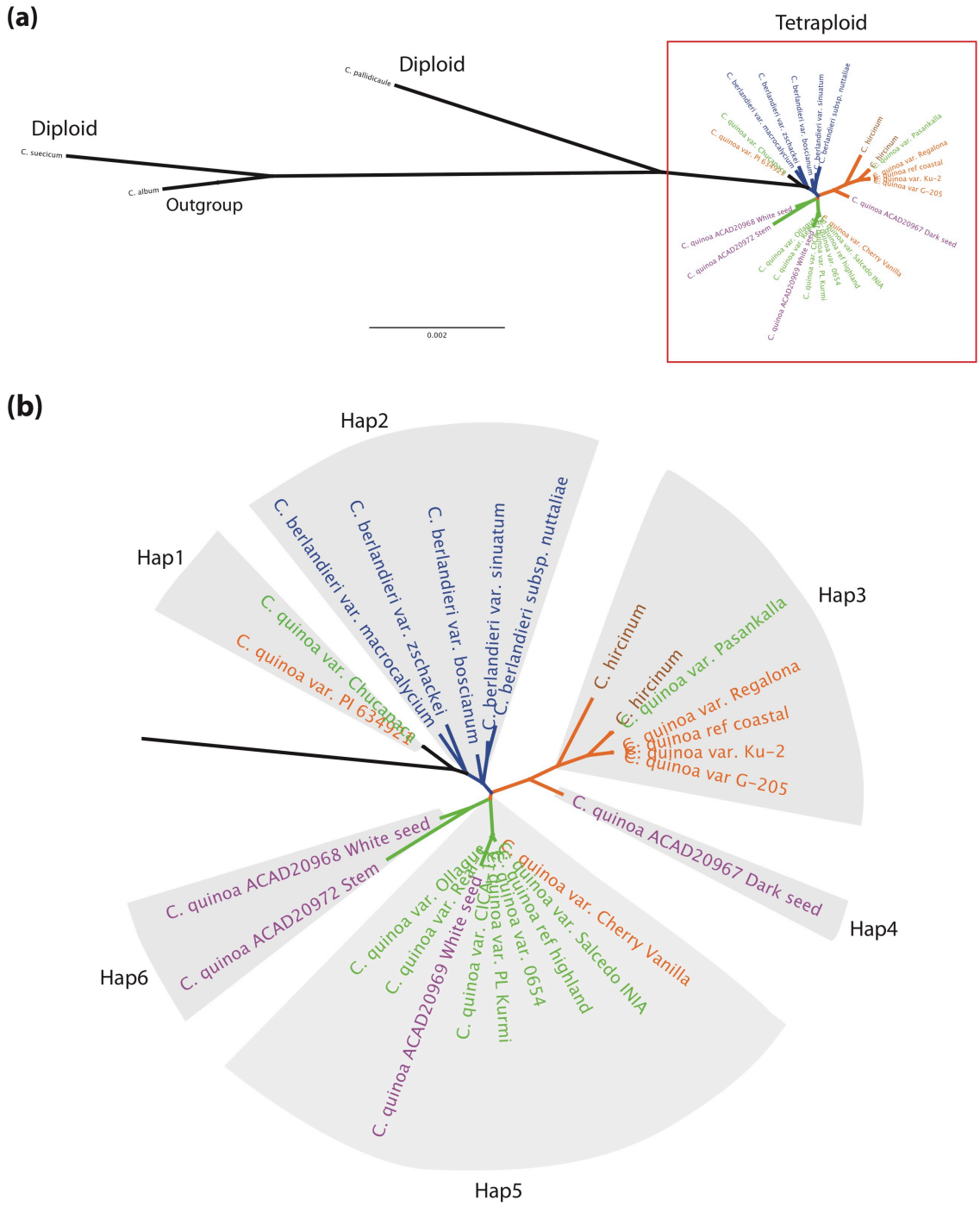


Figure 8. Chloroplast phylogeny of four archaeological specimens of quinoa *C. quinoa* (purple text) estimated following a multiple sequence alignment with 26 previously published chloroplast genomes of species from the *Chenopodium* genus (Table 2). (a) The best maximum likelihood tree built using *RAxML* with 1,000 bootstrapping replicates and *C. album* as outgroup. (b) Enlarged section of the three highlighted with the red box. Shaded groups and text boxes illustrate the main haplogroups found in the phylogeny. Ancient samples of *C. quinoa* in purple text. Accessions of *C. album*, *C. suecicum* and *C. pallidicaule* in black text. Accessions of *C. berlandieri* in blue text. Accessions of *C. hircinum* in brown text. Accessions of highland ecotypes of *C. quinoa* in green text. Accessions of coastal ecotypes of *C. quinoa* in orange text. The phylogenetic tree with bootstrap support values is presented in the Supplementary Figure 5.

The best scoring tree from our maximum likelihood analysis showed that *C. quinoa* is a paraphyletic group with respect to *C. berlandieri* and *C. hircinum* (Figure 8 and Supplementary Figure 5). In general, all samples of *C. quinoa*, *C. berlandieri* and *C. hircinum* could be divided into 6 haplogroups (Figure 8). The first haplogroup (Hap1) appears to be constituted of two samples of C. *quinoa*, the highland variety *C. quinoa* var. *chucapaca* and the coastal accession PI 634921, which cluster basally (100% bootstrap support) to the samples *C. berlandieri*, *C. hircinum* and *C. quinoa*. The second haplogroup (Hap2) is formed by the wild and weedy North American specimens of *C. berlandieri*, which contains multiple lineages and is separated from the South American specimens of *C. quinoa* with low bootstrap support (68% bootstrap value). The third haplogroup (Hap3) encompasses the specimens of the weedy species *C. hircinum*, most of the coastal ecotypes of *C. quinoa* and the highland quinoa variety *C. quinoa* var. *pasankalla*. The ancient dark quinoa seed (ACAD20967) has a unique haplogroup (Hap4) and is a sister-taxa to the coastal ecotypes (92% bootstrap support). The fifth haplogroup (Hap5) includes most of the highland ecotypes and the coastal ecotype *C. quinoa* var. *cherry vanilla*, which clusters with the ancient white quinoa seed (ACAD20969) (100% bootstrap support). Lastly, the ancient white quinoa seed (ACAD20968) together with the ancient quinoa stem (ACAD20972) composes a second unique haplogroup (Hap6) and is the sister-taxon a clade comprising most of the highland ecotypes (Figure 8 and Supplementary Figure 5).

Notably, haplogroups three to six (Hap3-Hap6), which include the specimens of *C. hircinum* and most of the South American samples of *C. quinoa* form a monophyletic clade with strong support (100% bootstrap support) (Figure 8 and Supplementary Figure 5). However, two specimens of quinoa (haplogroup Hap1) are outside of the core group of *C. quinoa* and positioned closer to the root of the tree. Also, the haplogroup Hap4 (*C. berlandieri* samples) has branches with poor bootstrap support (between 68% and 89%) and is likely to include several lineages. Therefore, we tested various hypothetical alternative topologies to confirm whether the samples of *C. quinoa* and *C. berlandieri* constitute monophyletic groups (Figure 9). Alternative phylogenetic positions could be rejected at $p < 0.01$ by a *Shimodaira-Hasegawa* test, including a tree in which *C. quinoa* and *C. hircinum* samples form monophyletic clade (H_1), a tree where all samples of *C. berlandieri* cluster in a monophyletic clade (H_2), and a tree where *C. berlandieri* constitutes a monophyletic group and is a sister-taxon to a monophyletic clade of all *C. quinoa* and *C. hircinum* samples (H_3) (Figure 9).

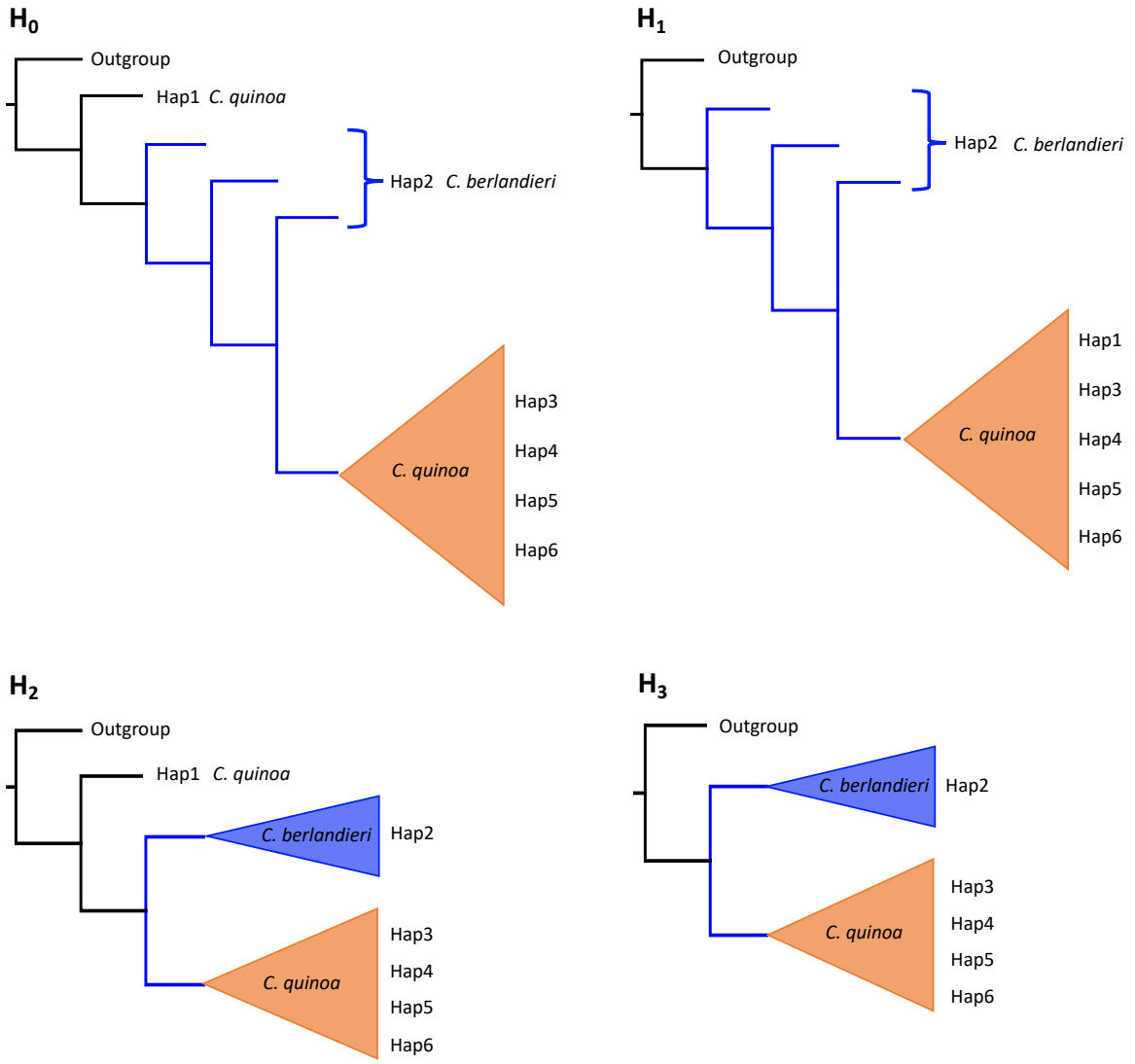


Figure 9. Illustration of the possible phylogenetic hypotheses of the monophyly of *C. quinoa* and *C. berlandieri*. **H_0)** maximum likelihood phylogeny from *RAxML,* all *C. quinoa* and *C. hircinum* samples (Hap1, Hap3, Hap4, Hap5, and Hap6) form a paraphyletic group with respect to *C. berlandieri* (Hap2). **H_1)** a tree where all *C. quinoa* and *C. hircinum* samples (Hap1, Hap3, Hap4, Hap5, and Hap6) form a monophyletic clade. **H_2)** a tree where all *C. berlandieri* samples (Hap2) cluster in a monophyletic clade. **H_3)** a tree where *C. berlandieri* (hap2) constitutes a monophyletic group and is a sister-taxon to a monophyletic clade of all *C. quinoa* and *C. hircinum* samples (Hap1, Hap3, Hap4, Hap5, and Hap6). We were able to reject all alternative hypotheses at $p < 0.01$ using the *Shimodaira-Hasegawa* test.

Discussion

Our current knowledge on quinoa domestication is primarily derived from morphological analysis of archaeobotanical remains (Bruno & Whitehead, 2003), analysis of microfossils (Babot, 2009), and genetic analysis of modern samples (Jarvis *et al.*, 2017). Although domesticated quinoa appears in the archaeological record by 7,000 years ago (Uhle, 1919), the ancient genomic diversity remains poorly studied. To reduce this knowledge gap, we characterised the nuclear and chloroplast genome sequences of four ~1,400 years BP archaeological samples of *C. quinoa* from the highlands of northern Argentina. To our knowledge, these are the first known ancient plant genome sequences generated from a crop domesticated in South American.

Expanding on previous studies that reported the PCR amplification of the chloroplast *rbcL* gene from rodent middens (Kuch *et al.*, 2002) and microsatellites markers from ancient quinoa seeds (Winkel *et al.*, 2018), we have confirmed it is possible to recover authentic aDNA molecules from desiccated plant remains including stems, as well as seeds, preserved under the dry and arid conditions of the high Andes. Our results show sufficient amounts of endogenous DNA can be obtained from very limited amounts of plant tissue (less than 1.7 mg) to examine genome-wide variability. Notably, our retrieval of aDNA from ancient quinoa stems, rather than seeds, provides novel evidence about the active cultivation and harvesting of quinoa in the highlands of Argentina at least some 1,400 years ago. This is because the presence of seeds in a given archaeological site does not necessarily indicate that the species was cultivated within the area, as seeds could quite easily represent the trade or exchange of plant products (Day, 2013). Therefore, our findings demonstrate the value of different sources of plant ancient DNA to support archaeobotanical studies on the origin, domestication and early use of crop plants.

Phylogenetic analyses based on nuclear data from several *Chenopodium* taxa have shown that *C. quinoa* and *C. hircinum* form a monophyletic group with *C. berlandieri* as sister-taxa (Jarvis *et al.*, 2017). In contrast, chloroplast and mitochondrial phylogenies on the same dataset indicated that *C. quinoa*, *C. hircinum* and *C. berlandieri* cluster in a monophyletic clade and that *C. quinoa* accessions form a paraphyletic group with respect to *C. berlandieri* and *C. hircinum* (Maughan *et al.*, 2019). We generated a maximum likelihood phylogeny using the dataset published by Maughan *et al.*, (2019) together with our newly generated ancient quinoa chloroplast genomes. Although our maximum likelihood tree strongly supports the monophyly

of a clade comprising all the tetraploid species (100% bootstrap support), the reciprocal monophyly of the individual tetraploid species could not be confirmed. In fact, we explicitly rejected (at $p < 0.01$) alternative chloroplast phylogenies testing the reciprocal monophyly of *C. quinoa* and *C. berlandieri*.

The structure of the chloroplast phylogeny, in which wild, weedy and cultivated accessions from different species are clustering in a monophyletic clade, is likely to indicate reticulate evolution (i.e. hybridisation of closely related species) and incomplete lineage sorting in the tetraploid species of *Chenopodium*. This finding is consistent with the traditional practices of quinoa cultivation in South America, where crops are characterised by a broad range of seed diversity produced through the hybridisation between cultivated, weedy and wild relatives and trade between indigenous peoples (Fuentes *et al.*, 2009; Bazile *et al.*, 2015). Interspecific genetic exchange is an important feature in plant evolution and incomplete lineage sorting has been documented in several domesticated plants such as wheat (Middleton *et al.*, 2014), cotton (Cronn & Wendel, 2003) and sunflower (Sambatti *et al.*, 2012).

We also provide evidence of a major bottleneck in the chloroplast diversity of quinoa occurring at some point in the last 1,400 years. The best scoring tree from our maximum likelihood analysis showed that the ancient specimens of quinoa fall in three distinct chloroplast haplogroups, two of which are not present in the modern accessions. Major social and environmental changes occurred after the Spanish arrival to South America in the 15th century, and together with a massive human migration from rural zones of the high Andes to urban centres in the 20th century, led to a decline in the cultivation of quinoa (Binford *et al.*, 1997; Chepstow-Lusty *et al.*, 2009; Jellen *et al.*, 2011; Winkel *et al.*, 2018). These major demographic shifts are likely to have also led to a reduction in the genetic diversity of quinoa.

Furthermore, sequence alignment to the nuclear and chloroplast reference genomes of *C. quinoa* identified 253,843 and 1,680 nucleotide polymorphisms, respectively. These newly identified variants highlight the potential information that could be accessed in the ancient quinoa specimens to answer fundamental questions such as wild progenitors, domestication traits, timing and regions of domestication, and the genetic variability associated with the wide range of environmental adaptation. Ancient genetic information generated in this study could be used in the future to investigate the genetic changes that occurred in loci of breeding value such as those related to high nutrient content, drought tolerance, sodium tolerance (Zou *et al.*, 2017), and saponin production (Jarvis *et al.*, 2017). Selective breeding strategies, mutant

screening or artificial genetic screens with technologies such as CRISPR/Cas9 system could be applied to incorporate ancient polymorphisms into modern quinoa germplasm to both confirm functionality and assess the potential value of reincorporating ancient diversity into modern cultivars.

References

Babot M del P. **2009**. La cocina, el taller y el ritual: Explorando las trayectorias del procesamiento vegetal en el noroeste Argentino. *Darwiniana* **47**: 7–30.

Babot M del P, Hocsman S. **2015**. Quinoa. A Millenary Grain in Northern Argentina. In: Selin H, ed. Encyclopaedia of the History of Science, Technology, and Medicine in Non-Western Cultures. Dordrecht: Springer Netherlands, 1–16.

Bazile D, Bertero HD, Nieto C. **2015**. State of the art report on quinoa around the world in 2013.

Bazile D, Jacobsen S-E, Verniau A. **2016**. The Global Expansion of Quinoa: Trends and Limits. *Frontiers in Plant Science* **7**: 1–6.

Binford MW, Kolata AL, Brenner M, Janusek JW, Seddon MT, Abbott M, Curtis JH. **1997**. Climate Variation and the Rise and Fall of an Andean Civilization. *Quaternary Research* **47**: 235–248.

Briggs AW, Stenzel U, Johnson PLF, Green RE, Kelso J, Prufer K, Meyer M, Krause J, Ronan MT, Lachmann M, *et al.* 2007. Patterns of damage in genomic DNA sequences from a Neandertal. *Proceedings of the National Academy of Sciences* **104**: 14616–14621.

Bruno MC, Whitehead WT. **2003**. Chenopodium Cultivation and Formative Period Agriculture at Chiripa, Bolivia. *Latin American Antiquity* **14**: 339–355.

Chepstow-Lusty AJ, Frogley MR, Bauer BS, Leng MJ, Boessenkool KP, Carcaillet C, Ali AA, Gioda A. **2009**. Putting the rise of the Inca Empire within a climatic and land management context. *Climate of the Past* **5**: 375–388.

Christensen SA, Pratt DB, Pratt C, Nelson PT, Stevens MR, Jellen EN, Coleman CE, Fairbanks DJ, Bonifacio A, Maughan PJ. **2007**. Assessment of genetic diversity in the USDA and CIP-FAO international nursery collections of quinoa (Chenopodium quinoa Willd.) using microsatellite markers. *Plant Genetic Resources: Characterization and Utilization* **5**: 82–95.

Cooper A, Poinar H. **2000**. Ancient DNA: Do It Right or Not at All. *Science* **289**: 1139b–1139.

Cronn R, Wendel JF. **2003**. Cryptic trysts, genomic mergers, and plant speciation. *New Phytologist* **161**: 133–142.

Dabney J, Meyer M, Pa S. **2013**. Ancient DNA Damage. *Cold Spring Harb Perspect Biol* **5**: 1–7.

Day J. **2013**. Botany meets archaeology: People and plants in the past. *Journal of Experimental Botany* **64**: 5805–5816.

FAOSTAT D. **2019**. Food and agriculture organization of the United Nations. Statistical database.

da Fonseca RR, Smith BD, Wales N, Cappellini E, Skoglund P, Fumagalli M, Samaniego JA, Carøe C, Ávila-Arcos MC, Hufnagel DE, *et al.* 2015. The origin and evolution of maize in the Southwestern United States. *Nature Plants* **1**: 14003.

Fordyce SL, Ávila-Arcos MC, Rasmussen M, Cappellini E, Romero-Navarro JA, Wales N, Alquezar-Planas DE, Penfield S, Brown T a., Vielle-Calzada JP, *et al.* 2013. Deep Sequencing of RNA from Ancient Maize Kernels. *PLoS ONE* **8**: 1–9.

Fuentes FF, Martinez EA, Hinrichsen P V., Jellen EN, Maughan PJ. **2009**. Assessment of genetic diversity patterns in Chilean quinoa (Chenopodium quinoa Willd.) germplasm using multiplex fluorescent microsatellite markers. *Conservation Genetics* **10**: 369–377.

Gansauge M-T, Meyer M. **2014**. Selective enrichment of damaged DNA molecules for ancient genome sequencing. *Genome research* **24**: 1543–9.

Garrison E, Marth G. **2012**. Haplotype-based variant detection from short-read sequencing. : 1–9.

Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N, *et al.* 2012. Phytozome: A comparative platform for green plant genomics. *Nucleic Acids Research* **40**: 1178–1186.

Gordillo-Bastidas E, Díaz-Rizzolo D. **2016**. Quinoa (Chenopodium quinoa Willd), from Nutritional Value to Potential Health Benefits: An Integrative Review. *Journal of Nutrition & Food Sciences* **06**.

Hariadi Y, Marandon K, Tian Y, Jacobsen S-E, Shabala S. **2011**. Ionic and osmotic relations in quinoa (Chenopodium quinoa Willd.) plants grown at various salinity levels. *Journal of Experimental Botany* **62**: 185–193.

Hocsman S, Babot M del P. **2018**. La transición de cazadores-recolectores a sociedades agropastoriles en Antofagasta de la Sierra (Puna de Catamarca, Argentina): Perspectivas desde la agencia y las prácticas. *Chungará (Arica)* **254**: 0–0.

Hong S-Y, Cheon K-S, Yoo K-O, Lee H-O, Cho K-S, Suh J-T, Kim S-J, Nam J-H, Sohn H-B, Kim Y-H. **2017**. Complete Chloroplast Genome Sequences and Comparative Analysis

of Chenopodium quinoa and C. album. *Frontiers in Plant Science* **8**: 1696.

Jacobsen S-E, Mujica A, Jensen CR. **2003**. The Resistance of Quinoa (Chenopodium quinoa Willd.) to Adverse Abiotic Factors. *Food Reviews International* **19**: 99–109.

Jarvis DE, Ho YS, Lightfoot DJ, Schmöckel SM, Li B, Borm TJA, Ohyanagi H, Mineta K, Michell CT, Saber N, *et al.* 2017. The genome of Chenopodium quinoa. *Nature* **542**: 307–312.

Jellen EN, Kolano BA, Sederberg MC, Bonifacio A, Maughan PJ. **2011**. Chenopodium. In: Kole C, ed. Wild Crop Relatives: Genomic and Breeding Resources. Berlin, Heidelberg: Springer Berlin Heidelberg, 35–61.

Jónsson H, Ginolhac A, Schubert M, Johnson PLF, Orlando L. **2013**. MapDamage2.0: Fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* **29**: 1682–1684.

Kadereit G, Borsch T, Weising K, Freitag H. **2003**. Phylogeny of Amaranthaceae and Chenopodiaceae and the Evolution of C 4 Photosynthesis. *International Journal of Plant Sciences* **164**: 959–986.

Katoh K, Standley DM. **2013**. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution* **30**: 772–780.

Knapp M, Hofreiter M. **2010**. Next generation sequencing of ancient DNA: Requirements, strategies and perspectives. *Genes* **1**: 227–243.

Kolano B, McCann J, Orzechowska M, Siwinska D, Temsch E, Weiss-Schneeweiss H. **2016**. Molecular and cytogenetic evidence for an allotetraploid origin of Chenopodium quinoa and C. berlandieri (Amaranthaceae). *Molecular Phylogenetics and Evolution* **100**: 109–123.

Li H. **2013**. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint arXiv:1303.3997*.

Kuch M, Rohland N, Betancourt JL, Latorre C, Steppan S, Poinar HN. **2002**. Molecular analysis of an 11,700-year-old rodent midden from the Atacama Desert, Chile. *Molecular ecology* **11**: 913–924.

Li H, Durbin R. **2009**. Fast and accurate short read alignment with Burrows – Wheeler transform. **25**: 1754–1760.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. **2009**. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078–2079.

Llamas B, Valverde G, Fehren-Schmitz L, Weyrich LS, Cooper A, Haak W. **2017**. From the field to the laboratory: Controlling DNA contamination in ancient DNA research. *STAR-Science and Technology in Archaeological Research* **??** ??-??

Mascher M, Schuenemann VJ, Davidovich U, Marom N, Himmelbach A, Hübner S, Korol A, David M, Reiter E, Riehl S, *et al.* 2016. Genomic analysis of 6,000-year-old cultivated grain illuminates the domestication history of barley. *Nature Genetics*.

Maughan PJ, Bonifacio A, Coleman CE, Jellen EN, Stevens MR, Fairbanks DJ. **2007**. Quinoa (Chenopodium quinoa). In: Pulses, Sugar and Tuber Crops. Berlin, Heidelberg: Springer Berlin Heidelberg, 147–158.

Maughan PJ, Chaney L, Lightfoot DJ, Cox BJ, Tester M, Jellen EN, Jarvis DE. **2019**. Mitochondrial and chloroplast genomes provide insights into the evolutionary origins of quinoa (Chenopodium quinoa Willd.). *Scientific Reports* **9**: 185.

Meyer M, Kircher M. **2010**. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor Protocols* **5**: 1–11.

Middleton CP, Senerchia N, Stein N, Akhunov ED, Keller B, Wicker T, Kilian B. **2014**. Sequencing of Chloroplast Genomes from Wheat, Barley, Rye and Their Relatives Provides a Detailed Insight into the Evolution of the Triticeae Tribe. *PLoS ONE* **9**: e85761.

Nowak V, Du J, Charrondière UR. **2016**. Assessment of the nutritional composition of quinoa (Chenopodium quinoa Willd.). *Food Chemistry* **193**: 47–54.

Okonechnikov K, Conesa A, Garcı F. **2016**. Genome analysis Qualimap 2 : advanced multi-sample quality control for high-throughput sequencing data. **32**: 292–294.

Orlando L, Gilbert MTP, Willerslev E. **2015**. Reconstructing ancient genomes and epigenomes. *Nature Reviews Genetics* **16**: 395–408.

Paabo S, Poinar H, Serre D, Jaenicke-Despres V, Hebler J, Rohland N, Kuch M, Krause J, Vigilant L, Hofreiter M. **2004**. Genetic analyses from ancient DNA. *Annual Review of Genetics* **38**: 645–679.

Palmer SA, Clapham AJ, Rose P, Freitas FO, Owen BD, Beresford-Jones D, Moore JD, Kitchen JL, Allaby RG. **2012**. Archaeogenomic evidence of punctuated genome evolution in gossypium. *Molecular Biology and Evolution* **29**: 2031–2038.

R CoreTeam. **2018**. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2016.

Rohland N, Harney E, Mallick S, Nordenfelt S, Reich D. **2015**. Partial uracil-DNA-glycosylase treatment for screening of ancient DNA. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **370**: 20130624.

Sambatti JBM, Strasburg JL, Ortiz-Barrientos D, Baack EJ, Rieseberg LH. **2012**. RECONCILING EXTREMELY STRONG BARRIERS WITH HIGH LEVELS OF GENE EXCHANGE IN ANNUAL SUNFLOWERS. *Evolution* **66**: 1459–1473.

Schubert M, Ginolhac A, Lindgreen S, Thompson JF, Al-rasheid KAS, Willerslev E,

Krogh A, Orlando L. **2012**. Improving ancient DNA read mapping against modern reference genomes.

Shimodaira H. **1998**. An Application of Multiple Comparison Techniques to Model Selection. *Annals of the Institute of Statistical Mathematics* **50**: 1–13.

Shimodaira H, Hasegawa M. **1999**. Multiple Comparisons of Log-Likelihoods with Applications to Phylogenetic Inference Letter to the Editor Multiple Comparisons of Log-Likelihoods with Applications to Phylogenetic Inference. *Mol. Biol. Evol* **16**: 1114–1116.

Stamatakis A. **2014**. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**: 1312–1313.

Swarts K, Gutaker RM, Benz B, Blake M, Bukowski R, Holland J, Kruse-Peeples M, Lepak N, Prim L, Cinta Romay M, *et al.* 2017. Genomic estimation of complex traits reveals ancient maize adaptation to temperate North America. *Science* **357**: 512–515.

Uhle M. **1919**. La arqueología de Arica y Tacna. *Boletín de la Sociedad Ecuatoriana de Estudios Históricos Americanos* **3**: 7.

Vallebueno-Estrada M, Rodríguez-Arévalo I, Rougon-Cardoso A, Martínez González J, García Cook A, Montiel R, Vielle-Calzada J-P. **2016**. The earliest maize from San Marcos Tehuacán is a partial domesticate with genomic evidence of inbreeding. *Proceedings of the National Academy of Sciences* **113**: 14151–14156.

Vega-Gálvez A, Miranda M, Vergara J, Uribe E, Puente L, Martínez EA. **2010**. Nutrition facts and functional potential of quinoa (Chenopodium quinoa willd.), an ancient Andean grain: a review. *Journal of the Science of Food and Agriculture* **90**: 2541–2547.

Weber EJ. **1978**. The Inca's ancient answer to food shortage. *Nature* **272**: 486–486.

Wickham H. **2016**. *ggplot2: elegant graphics for data analysis*. Springer.

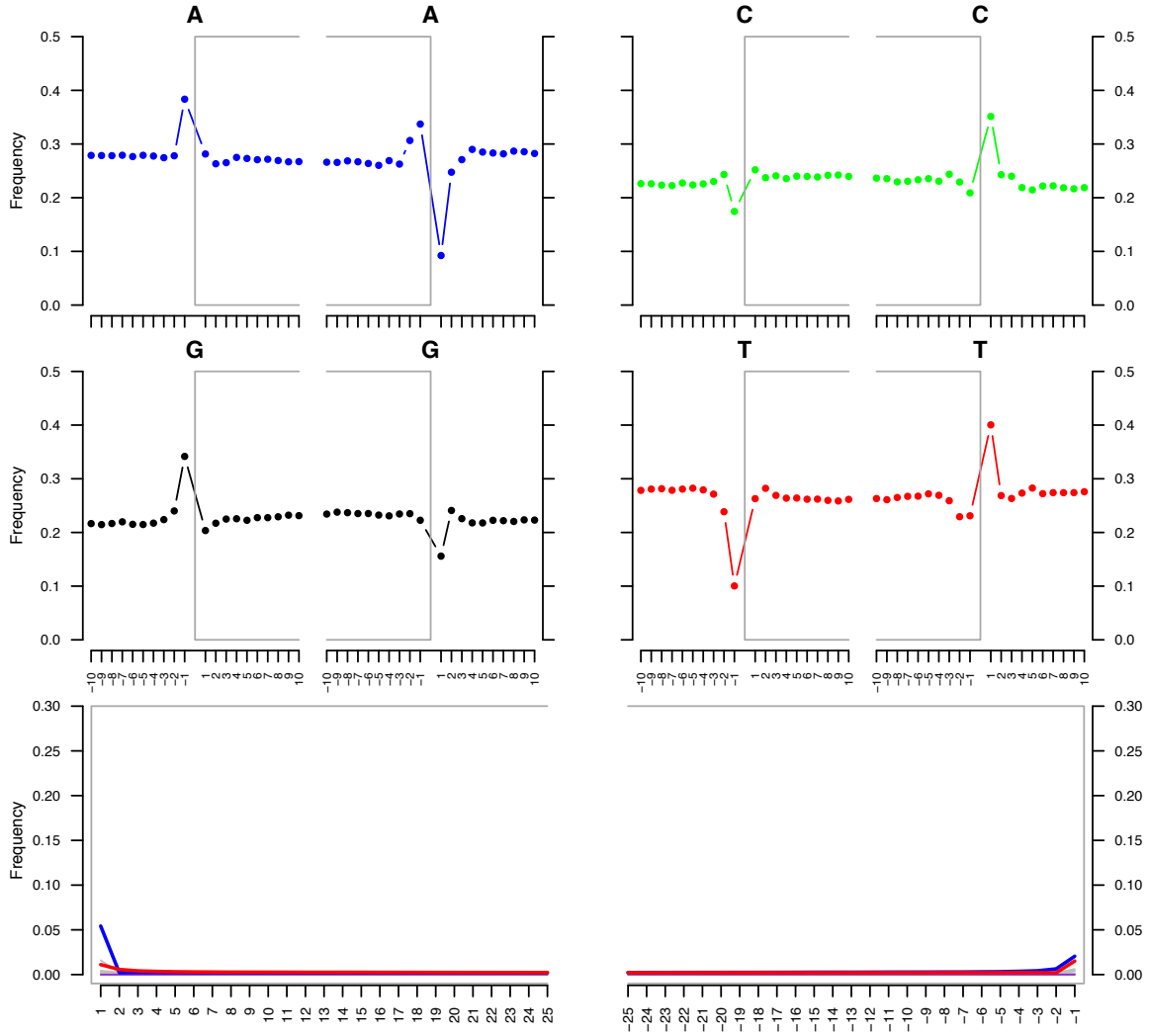
Wilson HD. **1988**. Quinua biosystematics II: Free-living populations. *Economic Botany* **42**: 478–494.

Wilson HD. **1990**. Quinua and Relatives (Chenopodium sect. Chenopodium subsect. Celluloid). *Economic Botany* **44**: 92–110.

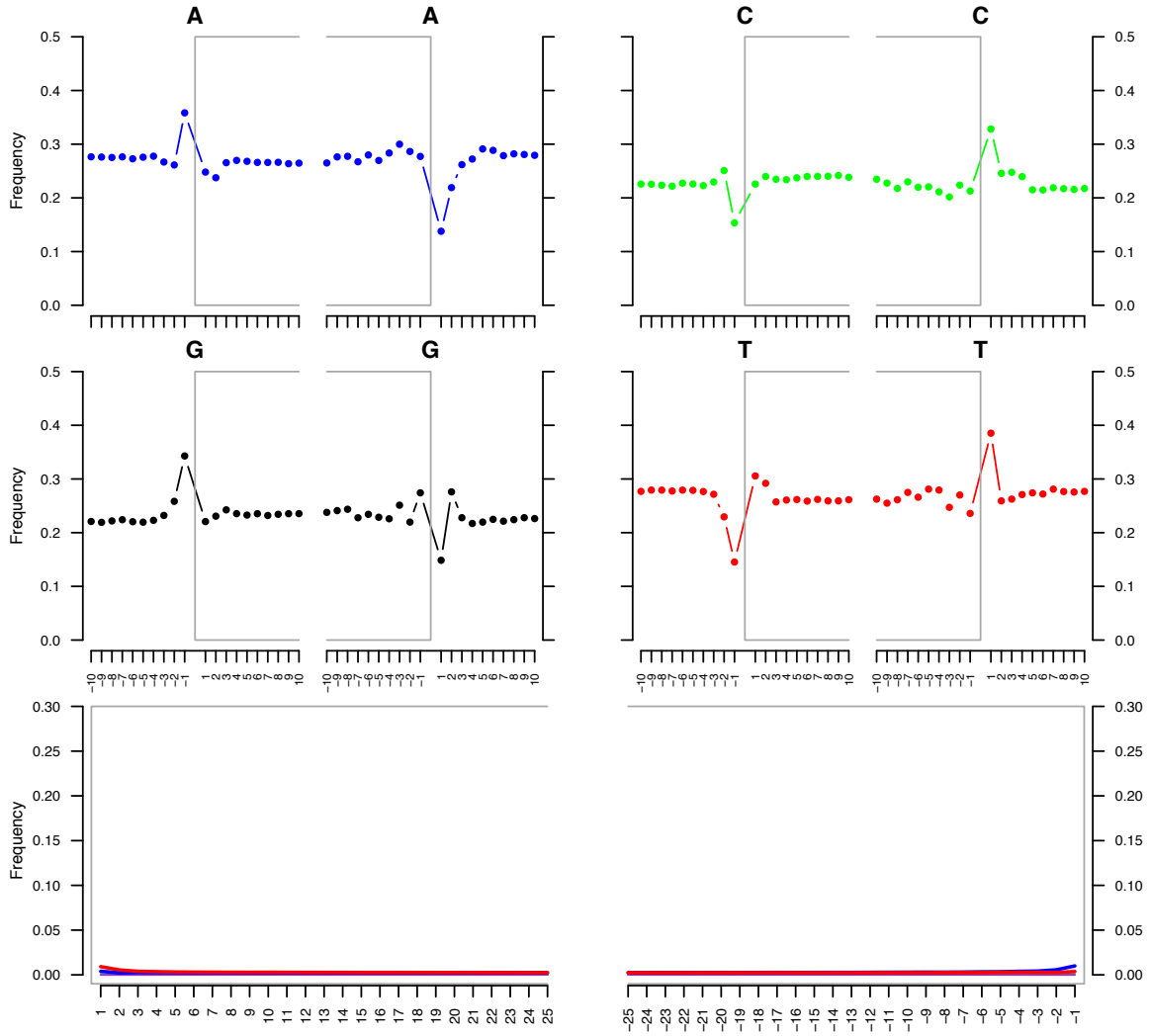
Winkel T, Aguirre MG, Arizio CM, Aschero CA, Babot M del P, Benoit L, Burgarella C, Costa-Tártara S, Dubois M-P, Gay L, *et al.* 2018. Discontinuities in quinoa biodiversity in the dry Andes: An 18-century perspective based on allelic genotyping (E Jellen, Ed.). *PLOS ONE* **13**: e0207519.

Zou C, Chen A, Xiao L, Muller HM, Ache P, Haberer G, Zhang M, Jia W, Deng P, Huang R, *et al.* 2017. A high-quality genome assembly of quinoa provides insights into the molecular basis of salt bladder-based salinity tolerance and the exceptional nutritional value. *Cell Research* **27**: 1327–1340.

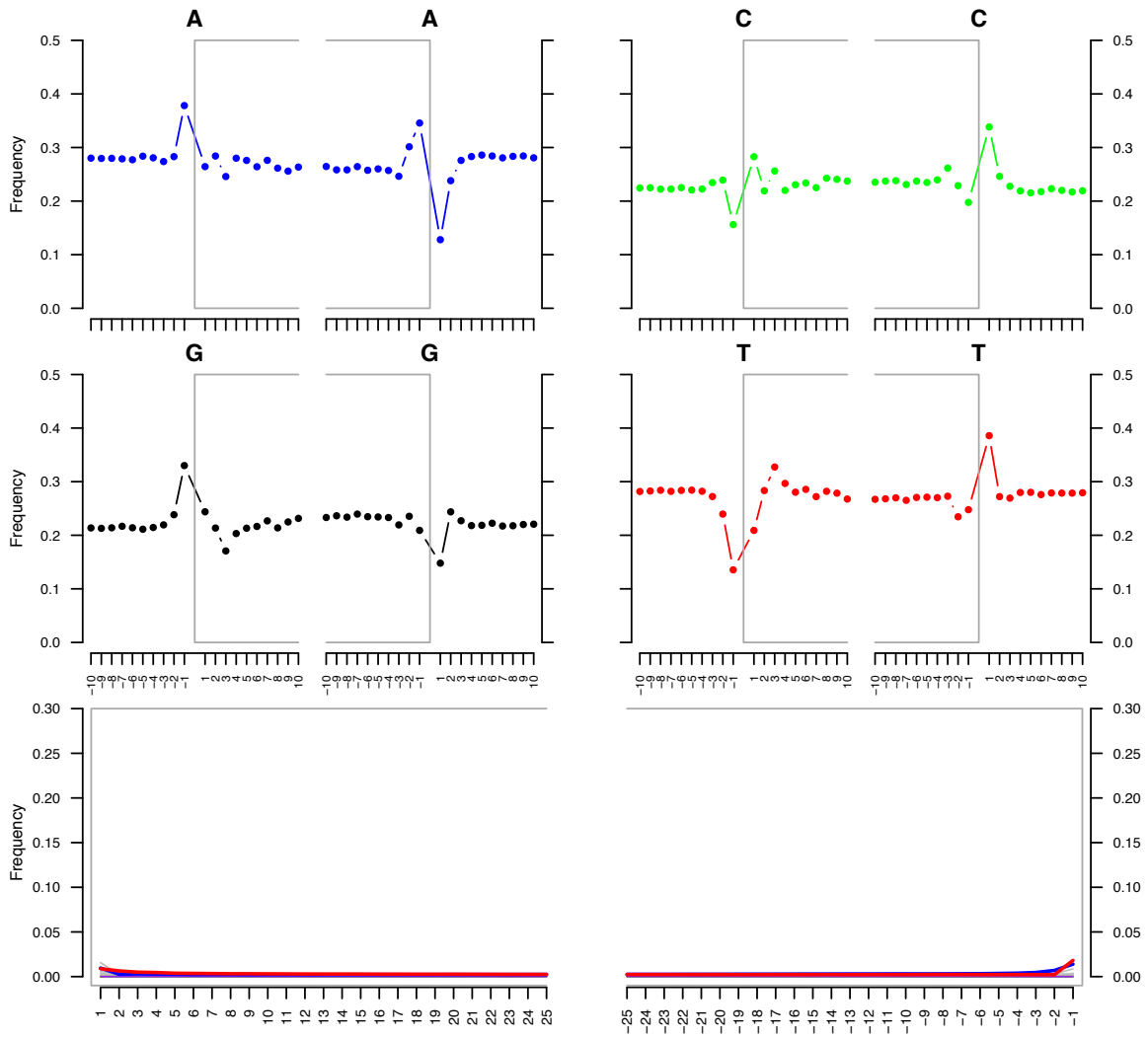
Supplementary information



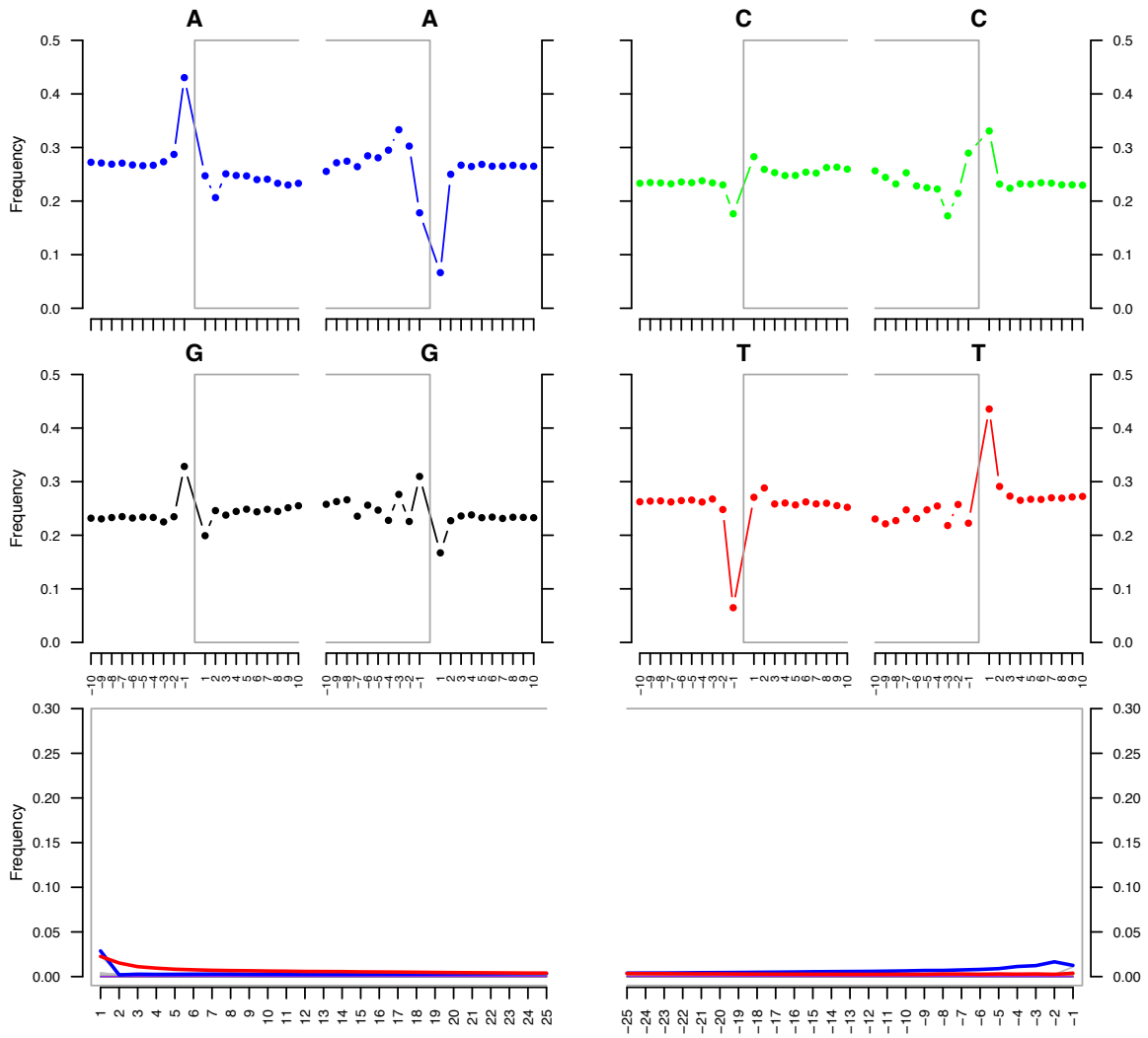
Supplementary Figure 1. MapDamage report for the shotgun sequenced library of the aDNA isolated from a bulk of three archaeological dark seeds of quinoa (Accession number: ACAD29067). The two top panels show the frequency of bases outside and in the reads. The two lower panels show the specific positions in which substitutions occurred. Note the high proportion of 5' C to T and 3' G to A substitutions accumulated at the start of the read characteristic of ancient DNA, shown in red and blue, respectively. All other substitutions are shown in grey. Deletions and insertions are shown in green and blue. Soft-clipped bases are shown in orange.



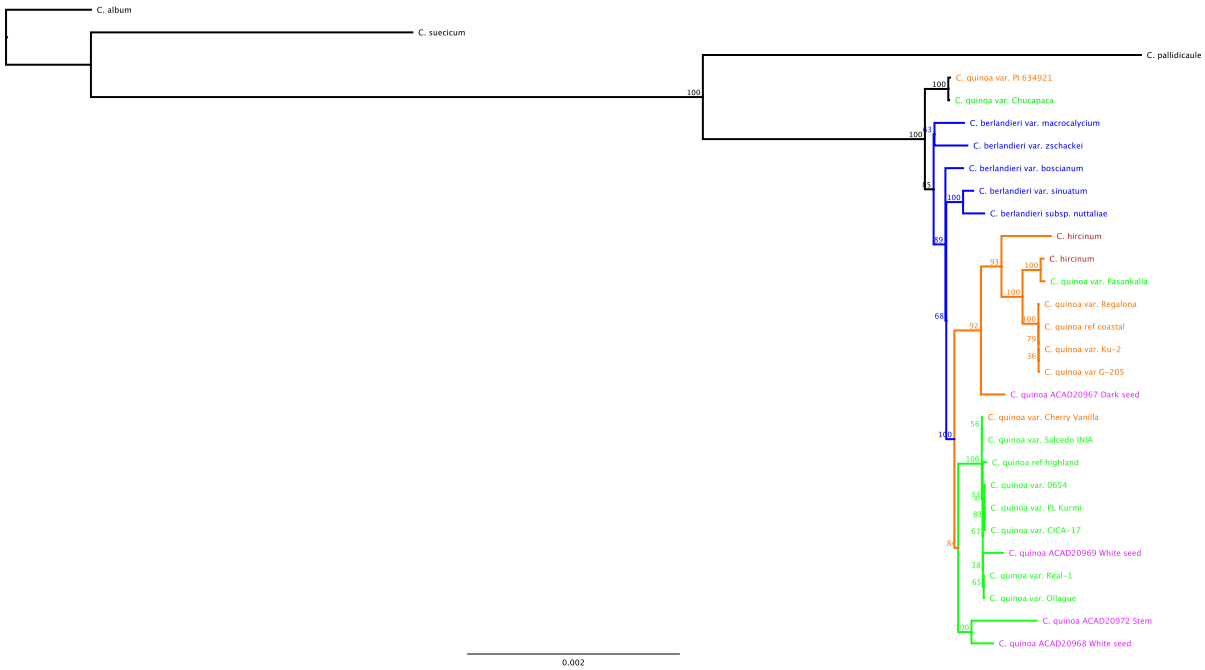
Supplementary Figure 2. MapDamage report for the shotgun sequenced library of the aDNA isolated from a single archaeological white seed of quinoa (Accession number: ACAD29068). The two top panels show the frequency of bases outside and in the reads. The two lower panels show the specific positions in which substitutions occurred. Note the high proportion of 5' C to T and 3' G to A substitutions accumulated at the start of the read characteristic of ancient DNA, shown in red and blue, respectively. All other substitutions are shown in grey. Deletions and insertions are shown in green and blue. Soft-clipped bases are shown in orange.



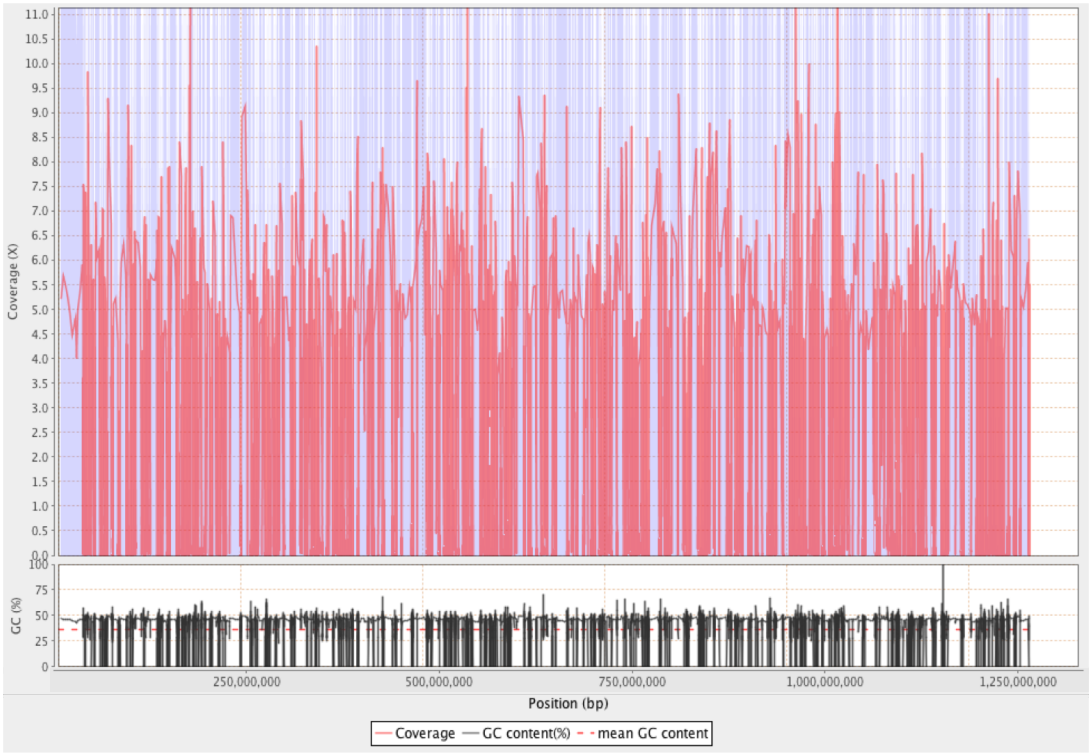
Supplementary Figure 3. MapDamage report for the shotgun sequenced library of the aDNA isolated from a bulk of three archaeological white seeds of quinoa (Accession number: ACAD29068). The two top panels show the frequency of bases outside and in the reads. The two lower panels show the specific positions in which substitutions occurred. Note the high proportion of 5' C to T and 3' G to A substitutions accumulated at the start of the read characteristic of ancient DNA, shown in red and blue, respectively. All other substitutions are shown in grey. Deletions and insertions are shown in green and blue. Soft-clipped bases are shown in orange.



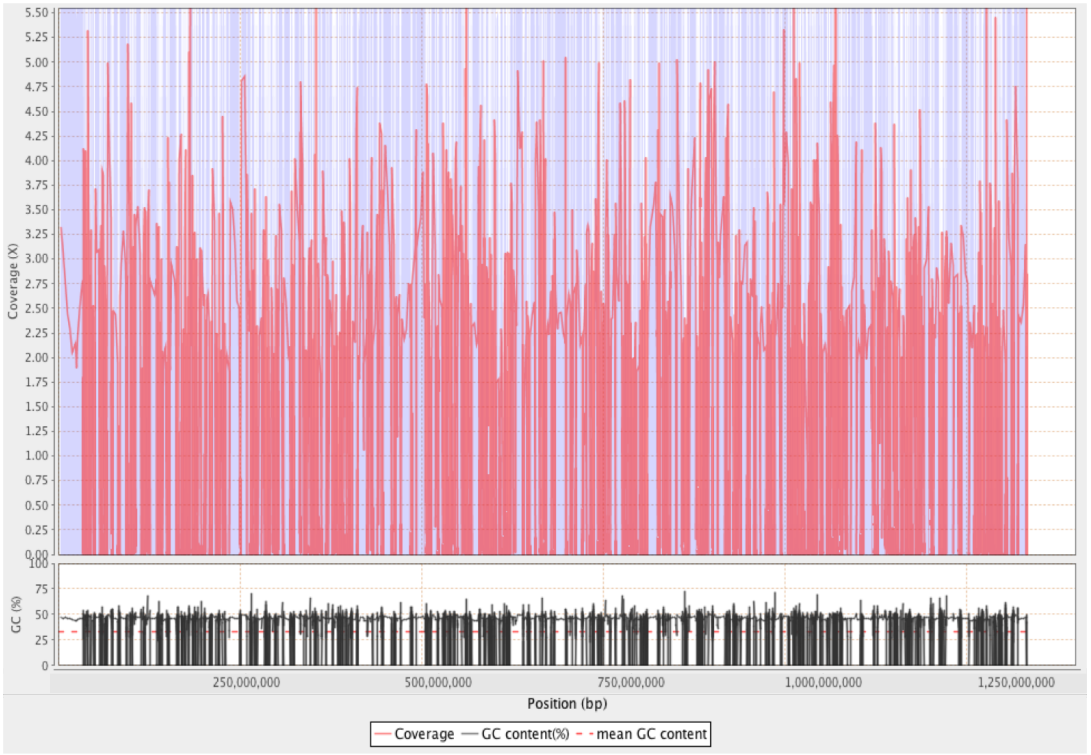
Supplementary Figure 4. MapDamage report for the shotgun sequenced library of the aDNA isolated from an archaeological stem of quinoa (Accession number: ACAD29072). The two top panels show the frequency of bases outside and in the reads. The two lower panels show the specific positions in which substitutions occurred. Note the high proportion of 5' C to T and 3' G to A substitutions accumulated at the start of the read characteristic of ancient DNA, shown in red and blue, respectively. All other substitutions are shown in grey. Deletions and insertions are shown in green and blue. Soft-clipped bases are shown in orange.



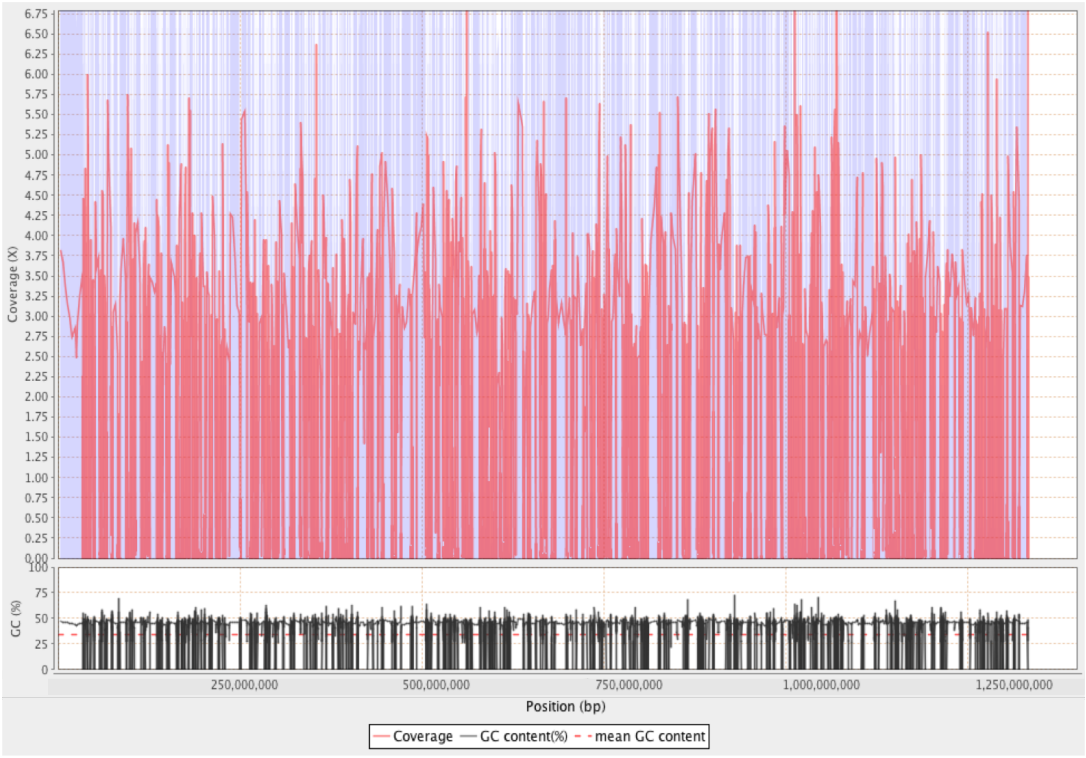
Supplementary Figure 5. Maximum likelihood phylogeny from *RAxML* analysis of the chloroplast sequences of four archaeological specimens of quinoa *C. quinoa* estimated following a multiple sequence alignment with 26 previously published chloroplast genomes of species from the *Chenopodium* genus (Table 2). The best maximum likelihood tree built using *RAxML* with 1,000 bootstrapping replicates and *C. album* as outgroup. Ancient samples of *C. quinoa* in purple text. Accessions of *C. album*, *C. suecicum* and *C. pallidicaule* in black text. Accessions of *C. berlandieri* in blue text. Accessions of *C. hircinum* in brown text. Accessions of highland ecotypes of *C. quinoa* in green text. Accessions of coastal ecotypes of *C. quinoa* in orange text. Branch values correspond to maximum likelihood bootstrap (%).



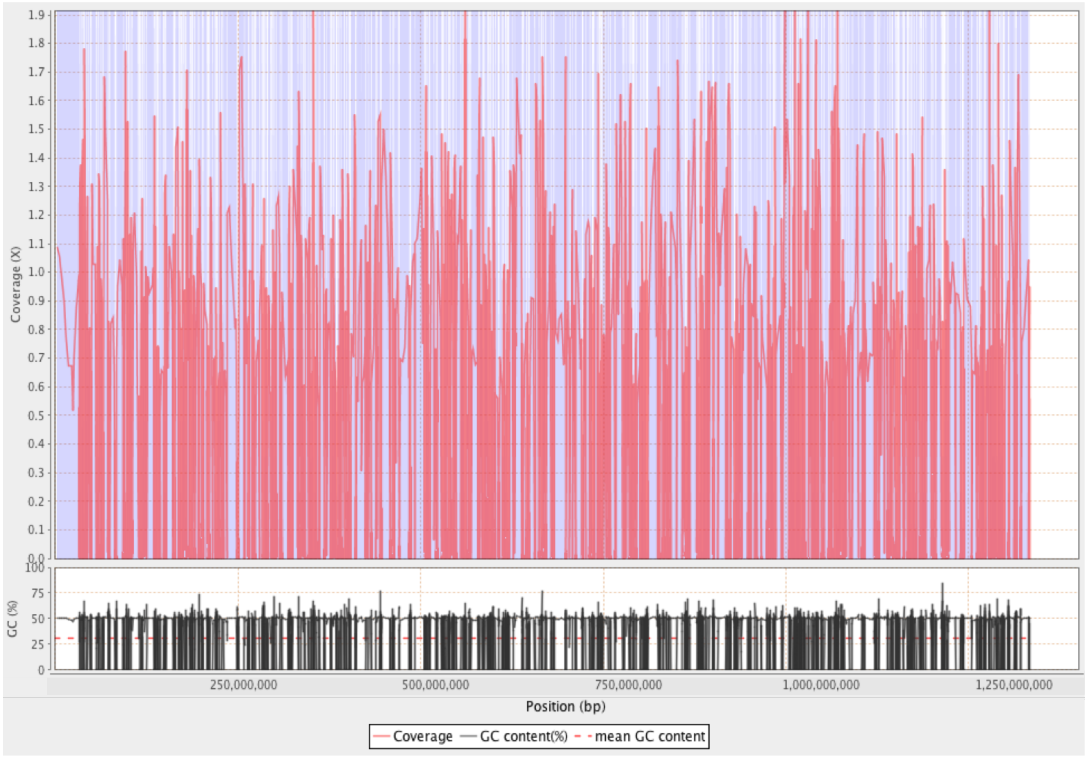
Supplementary Figure 6. Coverage across the genome and GC content of the shotgun sequenced aDNA library constructed from an archaeological dark seed of *C. quinoa* (Accession number: ACAD29067).



Supplementary Figure 7. Coverage across the genome and GC content of the shotgun sequenced aDNA library constructed from an archaeological white seed of *C. quinoa* (Accession number: ACAD29068).



Supplementary Figure 8. Coverage across the genome and GC content of the shotgun sequenced aDNA library constructed from an archaeological white seed of *C. quinoa* (Accession number: ACAD29068).



Supplementary Figure 9. Coverage across the genome and GC content of the shotgun sequenced aDNA library constructed from an archaeological stem of *C. quinoa* (Accession number: ACAD29072).

Chapter 6

Conclusions

Conclusions

Ancient plant specimens remain an understudied area of ancient DNA research, which has historically focused on investigating the domestication and population movements of animals such as chickens, pigs, cattle, bison and of course humans. These studies have been important in recovering valuable new perspectives on the history of human evolution and how climate change has impacted population structures over the past 100,000 years. However, while the anthropological insight extracted from these studies are important in a historical context, ancient plant DNA holds significant promise as an important tool for crop improvement for modern crops with a limited amount of potential genetic diversity. Here, I discuss three main research themes that are explored in this thesis:

1. What issues impact the recovery of ancient plant DNA and can current molecular techniques be optimised to extract additional plant information from ancient DNA samples?
2. How can we utilise modern agricultural genomic information to investigate ancient and historical crop samples?
3. How much genetic diversity or "allelic potential" is contained within historical crops (from a pre-'high-density agriculture' era) and how can this be utilised in modern agricultural systems?

Ancient plant DNA recovery

The recovery of genomic information from subfossilised specimens of animals, hominids, plants, and microorganisms has revolutionised our understanding of evolution (Orlando & Cooper, 2014; Weyrich *et al.*, 2015; Orlando *et al.*, 2015; MacHugh *et al.*, 2016). By applying new genomic technologies such as high-throughput sequencing (HTS), studies in ancient DNA have provided meaningful insights into human evolution, phylogenetics, speciation, paleoenvironments, and domestication (Orlando & Cooper, 2014; Weyrich *et al.*, 2015; MacHugh *et al.*, 2016; Estrada *et al.*, 2018). For instance, recent studies using HTS have recovered ancient genomes of important crop species such as barley and maize (Mascher *et al.*, 2016; Ramos-Madrigal *et al.*, 2016). However, despite the progress in these specific species, there is still a limited number of ancient DNA studies in plants compared to those in

humans and animals (Gugerli *et al.*, 2005; Schlumbaum *et al.*, 2008; Palmer *et al.*, 2012; Parducci *et al.*, 2017). As described in chapter 2, this is largely due to the lack of well-preserved samples, the large and complex plant genomes, and the absence of optimised plant-specific methods (Samarakoon *et al.*, 2013; Nistelberger *et al.*, 2016; Jiao & Schneeberger, 2017).

One of the most critical aspects hindering ancient plant DNA research is the restricted number of well-preserved archaeological samples suitable for aDNA isolation (Gugerli *et al.*, 2005; Nistelberger *et al.*, 2016; Estrada *et al.*, 2018). The preservation of DNA in plant subfossils is determined by several environmental conditions and taphonomic processes (described in chapter 2). However, the process of DNA degradation and fossilisation of plant tissues remains poorly understood (Weiß *et al.*, 2016). Authentic plant aDNA sequences have been retrieved from specimens with exceptional preservation found under desiccated and waterlogged conditions (Kistler *et al.*, 2014; Wales *et al.*, 2014; Mascher *et al.*, 2016) but specimens with such preservation are uncommon. Many archaeobotanical remains are found in a charred state, and less likely to yield endogenous DNA molecules (Nistelberger *et al.*, 2016). Also, plant domestication generally took place in areas of low latitude and relatively warm temperatures (Gepts, 2014), as opposed to cold or permafrost environments, limiting the time depth from which aDNA can be recovered (Palmer *et al.*, 2009; Allaby *et al.*, 2015).

Standard methods for the recovery and analysis of aDNA have been optimised for animal tissues and do not consider underlying characteristics of plants (Wales *et al.*, 2014). Plants are rich in secondary compounds that hinder DNA retrieval and may act as inhibitors in chemical reactions applied in aDNA methods (described in detail in chapter 2) (Bessetti, 2007; Samarakoon *et al.*, 2013). Therefore, modifications to standard aDNA procedures are needed to optimise recovery from a wide range of ancient plant remains preserved in different conditions. The examination of four DNA extraction methods showed that the combination of mechanical and chemical cell lysis, together with the PCR inhibitor removal technology applied in the DNeasy PowerPlant Pro protocol (Qiagen, USA), favoured the recovery of endogenous DNA from historical (60-100-years-old) and ancient wheat seeds (~1200 BP). This protocol was subsequently used to isolate DNA from ancient seeds and stems of archaeological quinoa (~1,400 BP), demonstrating the utility of this method for the retrieval of authentic aDNA molecules from other plant tissues even from low amounts of starting material (< 1.7 mg).

These findings may help the rapid advancement of several areas of plant research such as ancient plant DNA, phylogenetics, and plant breeding, where the isolation of DNA from limited amount of tissues could save time, effort, and provide access to genetic information that otherwise would not be accessible. Future development and standardisation of DNA isolation and library preparation methods will greatly enhance our ability to retrieve genomic information from a broader range of plant tissues preserved under different environmental conditions. Recent innovations in aDNA extraction and library preparation methods, such as single stranded library approaches, have permitted the retrieval of authentic ancient molecules from highly degraded animal bones and teeth and sediments (Rohland *et al.*, 2018). The application of these methodologies together with target enrichment of mitochondrial DNA even allowed the recovery of Neandertal and Denisovan DNA simply from Pleistocene cave sediments (Slon *et al.*, 2017). Similarly, the application of the DNeasy PowerPlant Pro protocol (Qiagen, USA) together with target enrichment arrays could open access to new sources of plant DNA such as sediments, plant traces on lithic artefacts, or even partially charred plant specimens.

Using modern agricultural information to study ancient plant material

Another challenge for ancient DNA research is the recovery of meaningful endogenous DNA from the complex metagenomic mixture in DNA extracts from degraded specimens (Paabo *et al.*, 2004; Willerslev & Cooper, 2005). Endogenous DNA of degraded samples is often embedded in a complex matrix containing exogenous DNA from organisms that were in contact before and after death (Willerslev & Cooper, 2005; Orlando *et al.*, 2015). Molecular procedures employed during DNA library preparation for high-throughput sequencing are highly sensitivity to any source of DNA, making the amplification and sequencing of either abundant or well-preserved contaminating DNA more likely than the low content of fragmented endogenous aDNA (Orlando *et al.*, 2015; Der Sarkissian *et al.*, 2015; Llamas *et al.*, 2017). Therefore, several innovations have been applied in ancient DNA research to increase the amount of endogenous aDNA content (Orlando *et al.*, 2015). One of the most commonly used strategies which I used in *Chapter 4* is the selective enrichment of endogenous molecules by hybridisation capture through the annealing of DNA fragments to pre-designed sets of RNA probes (Hodges *et al.*, 2007; Gnirke *et al.*, 2009).

In addition to the complex metagenomic mixture of ancient DNA extracts, plant genomes are also a mixture large gene families and abundant pseudogenes among a background of high copy repeats (usually more than 80%), which can hinder sequencing and assembly processes (Egan *et al.*, 2012; Jiao & Schneeberger, 2017). Again, hybridisation capture has proven to be an innovative and efficient tool for targeting and enriching specific genomic regions in complex plant genomes (Cronn *et al.*, 2012; Bevan *et al.*, 2017). The flexibility of this technology can also permit complexity reduction of ancient DNA libraries in a two-step or subtractive hybridisation capture protocol, where the first set of probes removes repetitive elements (*e.g.* transposable elements) from the library and the second set enriches for target loci (Fu *et al.*, 2010). Other methodologies such as C0t and methylation-filtration (explained in detail in *chapter 2*) that have proved to be useful to reduce the proportion of repetitive elements in complex plant genomes are yet to be applied to ancient samples. The inclusion of such strategies in future methodological improvements promises to further improve the recovery of target sequences.

In this thesis, bioinformatic data mining and analysis were used to design two hybridisation capture arrays as a strategy to characterise the plastid and nuclear diversity of a set of historical samples (60-100-years-old) of polyploid wheat species (*T. aestivum*, *T. timopheevii*, and *T. turgidum*) from Georgia. Analysis of HTS data generated from the enriched libraries demonstrated that both capture arrays efficiently recover the target sequences and reduce the complexity of the historical wheat DNA libraries, whose genomes feature different levels of ploidy (2x, 4x, 6x), large sizes (4-16Gbp), and a major composition of repetitive elements (80-90%) (Mayer *et al.*, 2014; Avni *et al.*, 2017; Appels *et al.*, 2018; Ling *et al.*, 2018). The high-throughput plastid and nuclear arrays developed in this study are a powerful resource that will facilitate future analyses of genetic resources and evolutionary processes in wheat as well as other important cereal species such as barley (*Hordeum vulgare*), oat (*Avena sativa* L.), rye (*Secale cereale* L.), and millet (*Cenchrus americanus* L., *Setaria* sp L. and *Panicum* sp. L.).

Allele-mining, and utilising ancient genomic information to improve crop yields

Currently, modern society relies on the cultivation of a small group of highly productive crop species to serve as staples for human and animal nutrition (Khoury *et al.*, 2014). Most of these crop species were first domesticated from wild relatives in the Holocene in quite different regions of the world (Larson *et al.*, 2014; Khoury *et al.*, 2016). The domestication process involved a strong selective pressure exerted by humans on the original wild genetic pool, which resulted in the generation of new plant species with a high frequency of desired traits but a much narrowed genetic pool (Meyer & Purugganan, 2013; Gepts, 2014). The diversity of domesticated species has been further winnowed by socio-political changes, environmental and climate events, and more importantly, intensive breeding (Doebley *et al.*, 2006; Chepstow-Lusty *et al.*, 2009; Bevan *et al.*, 2017). Maintaining genetic diversity is fundamental to crop improvement, yet the breeding process is generally counterproductive to this and threatens the genetic base on which improvement depends and in turn the food security of humankind (Tanksley & McCouch, 1997; Reif *et al.*, 2005; Doebley *et al.*, 2006).

Hybridisation capture and HTS, combined with bioinformatic data mining and analysis, were used to characterise the phylogenetic history and the genetic diversity of modern and historical (60-100-year-old) specimens of wheat (*Triticum* spp.), and ancient (~1,400 BP) specimens of quinoa (*Chenopodium quinoa*). The data analyses of the historical wheat species unearthed large amounts of genetic variability occurring 100 years ago in Georgia, a region with a significant number of native varieties. Notably, these historical wheat samples existed before the Green Revolution; one of the most intensive breeding processes in the last 60 years (Khush, 2001; Hedden, 2003), and thus represent part of the genetic pool that could have been lost during this massive genetic bottleneck. Similarly, the ancient quinoa specimens from the highlands of northern Argentina contained three distinct haplogroups, two of which are absent from present-day accessions. These changes in the genetic structure are likely to be the result of a reduction in the cultivation of quinoa as a consequence of the socio-political and environmental changes that led to the fall of the Inca civilisation after the Spanish arrival in the 15th century (Binford *et al.*, 1997; Chepstow-Lusty *et al.*, 2009; Fehren-Schmitz *et al.*, 2014). Consequently, the historical wheat samples from Georgia and the ancient quinoa specimens from Argentina both represent a promising reservoir to mine missing genes lost during domestication and breeding, as well as to discover critical genetic changes in the evolutionary history of these crops.

This study discovered a large number of nucleotide polymorphisms in the plastid and nuclear genomes of the historical wheat samples and the ancient quinoa specimens. These newly identified polymorphisms provide evidence of past selective pressures that potentially led to a reduction in the gene pool of present-day varieties. Further research efforts based on the historical and ancient diversity identified throughout this thesis promise to shed light on the nature of the genetic changes and evaluate the potential reintroduction of this ancestral diversity to improve present-day crop varieties. Bioinformatic data mining and analysis could be used to identify allelic variants that have changed through time in target loci of breeding value such as nutritional content, tolerance to abiotic stress, and disease resistance or susceptibility.

Breeding research in modern plant varieties has already applied technologies such as mutant screening, genome-wide association, genetic engineering, and targeted genome editing (*e.g.* CRISPR–Cas9 system) to mine genetic diversity and introduce alleles of agronomic importance (Kim *et al.*, 2015; Hilscher *et al.*, 2017; Arora *et al.*, 2019). Such technologies could also be applied in the historical and ancient polymorphisms detected in this study to confirm functionality and assess the potential reintroduction into modern cultivars. For instance, target enrichment of ancient variants and HTS can be used to screen seed banks, biological collections or chemically-induced mutant populations and identify accessions carrying the ancestral variants. Following the identification of plants carrying the ancestral polymorphisms, phenotypic characterisation and gene expression assays would allow identification of variants associated with targeted traits. Also, a more straightforward strategy could be the use of genetic engineering or genome editing (*e.g CRISPR-Cas9*) to introduce ancient variants directly into the target genomes, with subsequent phenotype-genotype association studies. In cases in which domesticated species have complex genomes or limited genomic information, plant species such as *Brachypodium distachyon* or *Arabidopsis thaliana* could be used as a model for the functional genomic analyses. The identified candidate loci could then be used for the genetic improvement of commercially grown cultivars.

In summary, although ancient plant DNA studies are more challenging and have thus lagged behind those in vertebrates and microorganisms, this thesis contributes critical new knowledge for the retrieval and analysis of degraded plant materials that can facilitate future research in ancient plant DNA. Also, by applying modern genomic technologies such as hybridisation capture and HTS, this research provides new insights into the evolutionary history of wheat and quinoa. Despite the agronomic potential and nutritional value of quinoa, little is known about the evolutionary and domestication history and the first ancient genomes generated in this thesis greatly enhance our understanding and provide a novel resource for future allele mining. There is also a further opportunity to expand on aDNA studies of mammals to infer phenotypes (Weyrich *et al.*, 2017; Posth *et al.*, 2018), and utilise current technologies to determine ancient plant phenotypes and the subsequent reconstruction of past environments and selective pressures. These features give ancient plant DNA incredible power to generate research information with direct application to the agricultural industry, contributing valuable information to human food security in the face of climate change and the projected population growth.

References

Allaby RG, Gutaker R, Clarke AC, Pearson N, Ware R, Palmer SA, Kitchen JL, Smith O. **2015**. Using archaeogenomic and computational approaches to unravel the history of local adaptation in crops. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **370**: 20130377.

Appels R, Eversole K, Feuillet C, Keller B, Rogers J, Stein N, Pozniak CJ, Choulet F, Distelfeld A, Poland J, *et al.* 2018. Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* **361**.

Arora S, Steuernagel B, Gaurav K, Chandramohan S, Long Y, Matny O, Johnson R, Enk J, Periyannan S, Singh N, *et al.* 2019. Resistance gene cloning from a wild crop relative by sequence capture and association genetics. *Nature Biotechnology* **37**: 139–143.

Avni R, Nave M, Barad O, Baruch K, Twardziok SO, Gundlach H, Hale I, Mascher M, Spannagl M, Wiebe K, *et al.* 2017. Wild emmer genome architecture and diversity elucidate wheat evolution and domestication. *Science* **357**: 93–97.

Bessetti J. **2007**. An Introduction to PCR Inhibitors. *Promega Notes*: 9–10.

Bevan MW, Uauy C, Wulff BBH, Zhou J, Krasileva K, Clark MD. **2017**. Genomic innovation for crop improvement. *Nature* **543**: 346–354.

Binford MW, Kolata AL, Brenner M, Janusek JW, Seddon MT, Abbott M, Curtis JH. **1997**. Climate Variation and the Rise and Fall of an Andean Civilization. *Quaternary Research* **47**: 235–248.

Chepstow-Lusty AJ, Frogley MR, Bauer BS, Leng MJ, Boessenkool KP, Carcaillet C, Ali AA, Gioda A. **2009**. Putting the rise of the Inca Empire within a climatic and land management context. *Climate of the Past* **5**: 375–388.

Cronn R, Knaus BJ, Liston A, Maughan PJ, Parks M, Syring J V., Udall J. **2012**. Targeted enrichment strategies for next-generation plant biology. *American Journal of Botany* **99**: 291–311.

Doebley JF, Gaut BS, Smith BD. **2006**. The molecular genetics of crop domestication. *Cell* **127**: 1309–21.

Egan AN, Schlueter J, Spooner DM. **2012**. Applications of next-generation sequencing in plant biology. *American Journal of Botany* **99**: 175–185.

Estrada O, Breen J, Richards SM, Cooper A. **2018**. Ancient plant DNA in the genomic era . *Nature Plants*.

Fehren-Schmitz L, Haak W, Mächtle B, Masch F, Llamas B, Tomasto Cagigao E, Sossna V, Schittek K, Isla Cuadrado J, Eitel B, *et al.* 2014. Climate change underlies global demographic, genetic, and cultural transitions in pre-Columbian southern Peru. *Proceedings of the National Academy of Sciences* **111**: 9443–9448.

Gepts P. **2014**. Domestication of Plants. In: Encyclopedia of Agriculture and Food Systems. Elsevier, 474–486.

Gnirke A, Melnikov A, Maguire J, Rogov P, LeProust EM, Brockman W, Fennell T, Giannoukos G, Fisher S, Russ C, *et al.* 2009. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nature biotechnology* **27**: 182–9.

Gugerli F, Parducci L, Petit RJ. **2005**. Ancient plant DNA: review and prospects. *New Phytologist* **166**: 409–418.

Hedden P. **2003**. The genes of the Green Revolution. *Trends in Genetics* **19**: 5–9.

Hilscher J, B??rstmayr H, Stoger E. **2017**. Targeted modification of plant genomes for precision crop breeding. *Biotechnology Journal* **12**.

Hodges E, Xuan Z, Balija V, Kramer M, Molla MN, Smith SW, Middle CM, Rodesch MJ, Albert TJ, Hannon GJ, *et al.* 2007. Genome-wide in situ exon capture for selective resequencing. *Nature genetics* **39**: 1522–7.

Jiao WB, Schneeberger K. **2017**. The impact of third generation genomic technologies on plant genome assembly. *Current Opinion in Plant Biology* **36**: 64–70.

Khoury CK, Achicanoy HA, Bjorkman AD, Navarro-Racines C, Guarino L, Flores-

Palacios X, Engels JMM, Wiersema JH, Dempewolf H, Sotelo S, *et al.* 2016. Origins of food crops connect countries worldwide. *Proceedings of the Royal Society B: Biological Sciences* **283**: 20160792.

Khoury CK, Bjorkman AD, Dempewolf H, Ramirez-Villegas J, Guarino L, Jarvis A, Rieseberg LH, Struik PC. 2014. Increasing homogeneity in global food supplies and the implications for food security. *Proceedings of the National Academy of Sciences* **111**: 4001–4006.

Khush GS. 2001. Green revolution: the way forward. *Nature Reviews Genetics* **2**: 815–822.

Kim H, Kim S-T, Kim S-G, Kim J-S. 2015. Targeted Genome Editing for Crop Improvement. *Plant Breeding and Biotechnology* **3**: 283–290.

Kistler L, Montenegro A, Smith BD, Gifford J a, Green RE, Newsom L a, Shapiro B. 2014. Transoceanic drift and the domestication of African bottle gourds in the Americas. *Proceedings of the National Academy of Sciences of the United States of America* **111**: 1–5.

Larson G, Piperno DR, Allaby RG, Purugganan MD, Andersson L, Arroyo-Kalin M, Barton L, Climer Vigueira C, Denham T, Dobney K, *et al.* 2014. Current perspectives and the future of domestication studies. *Proceedings of the National Academy of Sciences* **111**: 6139–6146.

Ling H-Q, Ma B, Shi X, Liu H, Dong L, Sun H, Cao Y, Gao Q, Zheng S, Li Y, *et al.* 2018. Genome sequence of the progenitor of wheat A subgenome Triticum urartu. *Nature* **557**: 424–428.

Llamas B, Valverde G, Fehren-Schmitz L, Weyrich LS, Cooper A, Haak W. 2017. From the field to the laboratory: Controlling DNA contamination in ancient DNA research. *STAR-Science and Technology in Archaeological Research* **??** ??-??

MacHugh DE, Larson G, Orlando L. 2016. Taming the Past: Ancient DNA and the Study of Animal Domestication. *Annual Review of Animal Biosciences* **5**: 329–351.

Mascher M, Schuenemann VJ, Davidovich U, Marom N, Himmelbach A, Hübner S, Korol A, David M, Reiter E, Riehl S, *et al.* 2016. Genomic analysis of 6,000-year-old cultivated grain illuminates the domestication history of barley. *Nature Genetics*.

Mayer KFX, Rogers J, Dole el J, Pozniak C, Eversole K, Feuillet C, Gill B, Friebe B, Lukaszewski AJ, Sourdille P, *et al.* 2014. A chromosome-based draft sequence of the hexaploid bread wheat (Triticum aestivum) genome. *Science* **345**: 1251788–1251788.

Meyer RS, Purugganan MD. 2013. Evolution of crop species: Genetics of domestication and diversification. *Nature Reviews Genetics* **14**: 840–852.

Nistelberger HM, Smith O, Wales N, Star B, Boessenkool S. 2016. The efficacy of high-throughput sequencing and target enrichment on charred archaeobotanical remains. *Scientific*

Reports **6**: 37347.

Orlando L, Cooper A. **2014**. Using Ancient DNA to Understand Evolutionary and Ecological Processes. *Annual Review of Ecology, Evolution, and Systematics* **45**: 573–598.

Orlando L, Gilbert MTP, Willerslev E. **2015**. Reconstructing ancient genomes and epigenomes. *Nature Reviews Genetics* **16**: 395–408.

Paabo S, Poinar H, Serre D, Jaenicke-Despres V, Hebler J, Rohland N, Kuch M, Krause J, Vigilant L, Hofreiter M. **2004**. Genetic analyses from ancient DNA. *Annual Review of Genetics* **38**: 645–679.

Palmer SA, Moore JD, Clapham AJ, Rose P, Allaby RG. **2009**. Archaeogenetic evidence of ancient nubian barley evolution from six to two-row indicates local adaptation. *PLoS ONE* **4**: 2–8.

Palmer SA, Smith O, Allaby RG. **2012**. The blossoming of plant archaeogenetics. *Annals of Anatomy* **194**: 146–156.

Parducci L, Bennett KD, Ficetola GF, Alsos IG, Suyama Y, Wood JR, Pedersen MW. **2017**. Ancient plant DNA in lake sediments. *New Phytologist* **214**: 924–942.

Posth C, Begg T, Nieselt K, Krause J, Mittnik A, Peltzer A, Wang C. **2018**. Inferring genetic origins and phenotypic traits of George Bähr, the architect of the Dresden Frauenkirche. *Scientific Reports* **8**: 2115.

Ramos-Madrigal J, Smith BD, Moreno-Mayar JV, Gopalakrishnan S, Ross-Ibarra J, Gilbert MTP, Wales N. **2016**. Genome Sequence of a 5,310-Year-Old Maize Cob Provides Insights into the Early Stages of Maize Domestication. *Current Biology* **26**: 3195–3201.

Reif JC, Zhang P, Dreisigacker S, Warburton ML, van Ginkel M, Hoisington D, Bohn M, Melchinger AE. **2005**. Wheat genetic diversity trends during domestication and breeding. *Theoretical and Applied Genetics* **110**: 859–864.

Samarakoon T, Wang SY, Alford MH. **2013**. Enhancing PCR Amplification of DNA from Recalcitrant Plant Specimens Using a Trehalose-Based Additive. *Applications in Plant Sciences* **1**: 1200236.

Der Sarkissian C, Allentoft ME, Avila-Arcos MC, Barnett R, Campos PF, Cappellini E, Ermini L, Fernandez R, da Fonseca R, Ginolhac A, *et al.* 2015. Ancient genomics. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **370**: 20130387–20130387.

Schlumbaum A, Tensen M, Jaenicke-Després V. **2008**. Ancient plant DNA in archaeobotany. *Vegetation History and Archaeobotany* **17**: 233–244.

Shapiro B, Hofreiter M. **2012**. *Ancient DNA: Methods and Protocols* (B Shapiro and M Hofreiter, Eds.). Totowa, NJ: Humana Press.

Tanksley SD, McCouch SR. **1997**. Seed Banks and Molecular Maps: Unlocking Genetic Potential from the Wild. *Science* **277**: 1063–1066.

Wales N, Andersen K, Cappellini E, ??vila-Arcos MC, Gilbert MTP. **2014**. Optimization of DNA recovery and amplification from non-carbonized archaeobotanical remains. *PLoS ONE* **9**.

Weiß CL, Schuenemann VJ, Devos J, Shirsekar G, Reiter E, Gould BA, Stinchcombe JR, Krause J, Burbano HA. **2016**. Temporal patterns of damage and decay kinetics of DNA retrieved from plant herbarium specimens. *Royal Society Open Science* **3**: 160239.

Weyrich LS, Dobney K, Cooper A. **2015**. Ancient DNA analysis of dental calculus. *Journal of Human Evolution* **79**: 119–124.

Weyrich LS, Duchene S, Soubrier J, Arriola L, Llamas B, Breen J, Morris AG, Alt KW, Caramelli D, Dresely V, *et al.* 2017. Neanderthal behaviour, diet, and disease inferred from ancient DNA in dental calculus. *Nature* **544**: 357–361.

Willerslev E, Cooper A. **2005**. Review Paper. Ancient DNA. *Proceedings of the Royal Society B: Biological Sciences* **272**: 3–16.

Appendix 1

Perspectives: Ancient plant DNA in the genomic era

Manuscript published in *Nature Plants*

Estrada, O., Breen, J., Richards, S. M., & Cooper, A. (2018). Ancient plant DNA in the genomic era . *Nature Plants*. https://doi.org/10.1038/s41477-018-0187-9

Statement of authorship

Title of Paper	Ancient plant DNA in the genomic era
Publication Status	☑ Published ☐ Accepted for Publication ☐ Submitted for Publication ☐ Unpublished and Unsubmitted w ork w ritten in manuscript style
Publication Details	Estrada, O., Breen, J., Richards, S. M., & Cooper, A. (2018). Ancient plant DNA in the genomic era. Nature Plants. https://doi.org/10.1038/s41477-018-0187-9

Principal Author

Name of Principal Author (Candidate)	Oscar Estrada		
Contribution to the Paper	Contributed to the conception and design of the manuscript, reviewed literature, wrote and edited the manuscript.		
Overall percentage (%)	70		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	08/03/2019

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	James Breen		
Contribution to the Paper	Contributed to the conception and design of the manuscript, reviewed and edited the manuscript.		
Signature		Date	08/03/2019

Name of Co-Author	Stephen M Richards
Contribution to the Paper	Contributed to the conception and design of the manuscript, reviewed and edited the manuscript.
Signature	Date 08/03/2019

Please cut and paste additional co-author panels here as required.

Name of Co-Author	Alan Cooper
Contribution to the Paper	Contributed to the conception and design of the manuscript, reviewed and edited the manuscript.
Signature	Date 8/3/19

ARCHAEOBOTANY

Ancient plant DNA in the genomic era

Next-generation sequencing technologies have significantly changed the scope of ancient plant DNA research, moving from analysis of a few loci to generation of ancient genomes. Future research could refine our understanding of plant evolution and adaptation, and provide information for conservation, crop breeding and food security.

Oscar Estrada, James Breen, Stephen M. Richards and Alan Cooper

Under favourable conditions, DNA fragments can be preserved in biological remains for long periods of time, as demonstrated by its recovery from ancient animal, plant and invertebrate specimens dating back 450–800 thousand years[1]. Accessing the genetic information of such material provides a unique means to study how species have changed and adapted through time, enabling the reconstruction of past genetic diversity and interaction with past environments[2]. Plant ancient DNA (aDNA) can be recovered from a variety of sub-fossilised macro-remains such as fruits, seeds, leaves and wood, and can even be recovered from animal and human faeces as well as secondary components in sediments[3] (Fig. 1).

Studies in ancient plant DNA have lagged behind those of humans and vertebrate taxa[3], although this is partly due to the lack of well-preserved samples suitable for aDNA research. Moreover, standard methods used in aDNA studies are not explicitly designed for botanical remains, and thus do not consider the intrinsic characteristics of plants. Lastly, studies in agricultural plant species, such as bread wheat, barley and maize, are hindered by factors including large genome sizes (2–16 Gbp), high composition of repetitive elements (60–90%) and diverse ploidy levels[4].

Many archaeobotanical samples (most notably seeds or grain material) are found in a charred state, with samples having been burned through the deliberate use of fire by humans in activities such as food processing (cooking, drying or disposal of refuse), heating, pottery, crop management and razing[5]. Such exposure to high temperatures transforms starches and proteins into high-molecular-weight melanoidins, as well as other organic compounds into charcoal[5]. Since decomposing organisms do not affect melanoidins and charcoal, charred remains can survive for long periods of time in different environments[5]. While charred specimens can retain most of their morphological features, extensive damage at the molecular level makes the preservation and recovery of nucleic acids unlikely[6].

Plant materials are extremely rich in polyphenols, proteins and polysaccharides, while ancient plant remains can also contain humic acids and salts produced by organic decomposition or associated sediments[7]. These macromolecules represent two major obstacles in the ability to extract DNA from ancient plants. First, crosslinks between polyphenols, proteins and nucleic acids are formed through the decay of plant tissues and such cross-linked nucleic acids are lost during extraction steps, decreasing the number of endogenous DNA molecules recovered. Second, crosslink formations, polyphenols, polysaccharides and humic acids can be co-extracted with nucleic acids and affect polymerase activity, interfere with absorbance-based quantification methods and inhibit fluorescence in real-time PCR (ref. [8]). The presence of inhibitor compounds in aDNA extracts is also likely to interfere with the activity of ligases and endonucleases used in the preparation of DNA sequencing libraries.

The most successful plant aDNA studies have made use of desiccated specimens from extremely dry and arid environments such as caves or desert areas[9–11]. In addition, the anoxic conditions in peat bogs, deep sediments of lakes and seas, or even in wells and latrines, allow good preservation of waterlogged plant samples, which have also provided aDNA sequences[7,12,13]. Seeds are the most common plant material used in ancient DNA studies. As a reproductive body, seeds are adapted to protect the embryo and remain in dormancy for long periods of time before germination, in some cases germinating after millennia[14]. This feature makes seeds an important vessel to preserve DNA under favourable conditions. Other ancient plant remains that have been reported to yield significant amounts of aDNA include maize cobs, bottle gourd rinds, foliar tissues and wood[10–13,15].

Agricultural species have received the most research interest with a key area being the evolution and domestication of plants. Traditionally, standard PCR-based assays have been used to address questions such as taxonomy of archaeobotanical specimens, and the origin and spread of agricultural species such as wheat, maize and barley[3]. However, conventional PCR and chain-termination sequencing constrained plant aDNA research to a range of loci previously annotated in modern species. The adoption of next-generation sequencing (NGS) technologies significantly expanded the range of studies possible, enabling the analysis of full chloroplast[12,15], mitochondrial[15] and nuclear genomes[9,10] from archaeological samples of several cultivated species dating back to the early-mid Holocene epoch (~10,000 years ago).

NGS has also been used to obtain RNA and pathogen genome sequences from ancient plant remains. Mature seeds act as reservoirs of a large number of RNA molecules generated during late embryogenesis and following preparation for germination[16]. This characteristic, coupled with the ability of seeds to remain dormant over time, has permitted the preservation and subsequent recovery of ancient RNA from historic maize[17] (750 years old) and barley[18] (1,100–1,400 years old) samples. In addition, significant plant pathogens have also been recovered and sequenced, including the potato blight fungus *Phytophthora infestans* — the cause of the Irish potato famine — which was obtained from herbarium specimens from the nineteenth century[19,20], and the Barley Stripe Mosaic Virus isolated from a 750 year old barley grain found in North Africa[21].

Recent advances in aDNA methods have also allowed the characterisation of ancient epigenomes. Plants have historically been a key element in the investigation of epigenetic mechanisms of environmental adaptation[22], due to their ability to respond to abiotic and biotic stresses[23] as well as engage somatic and transgenerational epigenetic memory[24]. In ancient specimens, epigenetic signals, such as the methylation of cytosines, can be detected directly

Fig. 1 | Sub-fossilised plant macro-remains sources of ancient plant DNA. a, Desiccated Eucalyptus leaves from Robertson Cave (Naracoorte, Australia). **b**, Desiccated Casuarina leaves from Robertson Cave (Naracoorte, Australia). **c**, Charred wheat seeds from Areni-1 cave (Armenia). **d**, Desiccated wheat seeds from Areni-1 cave (Armenia).

by bisulphite sequencing[25] or indirectly from the distribution of DNA damage patterns in NGS data[26]. These methods have been applied successfully to a panel of archaeological barley seeds spanning ~2,600 years, allowing the association of high levels of methylated sites with the response to a viral infection and the identification of a tendency towards decreasing methylated cytosines as a diagenetic process[27]. The success of these methods, however, is impacted by the environmental conditions in which the samples were preserved.

Future research on ancient plant DNA has the potential to reconstruct many of the key evolutionary processes that have shaped wild and domesticated plants. It is estimated that approximately 2,500 species of plants have been subject to domestication since the last glacial period, over 12,000 years ago[28]. However, the domestication history of most crop species remains unclear, with several fundamental questions regarding wild progenitor species, timing of domestication and geographic regions of domestication, yet to be answered. Moreover, domestication, as an extensive selection process, can also lead to a reduction in the genetic diversity of cultivated species. These genetic bottlenecks present a major limitation for the continual improvement of commercially grown plants[29]. Hence, investigating new sources of genetic variation is critical for crop improvement, not only to enhance crop performance but also to increase the efficiency of plant breeding[29]. Ancient DNA opens a new path to directly access the genetic information that has changed at various stages of plant domestication, enabling the identification of lost genetic diversity and the potential to reintroduce extinct loci related to desired traits. Studies of modern plant DNA have already detected several genetic loci associated with tolerance to abiotic and biotic factors such as water stress, ion toxicity or deficiency, extreme temperatures, and disease resistance or susceptibility[30,31]. DNA capture-enrichment methods and NGS can be used to investigate such genes, exomes or whole genomes of ancient plants, and identify allelic variants that have changed through time. Mutant screening or genome editing techniques (for example, the CRISPR–Cas9 system) could then be applied on ancient polymorphisms across seed banks and biological collections to confirm functionality and assess the potential reincorporation of ancient genomic variation into modern cultivars.

The investigation of plant domestication is central to accelerating efforts to domesticate new species with high environmental adaptability and better yield. Industrialised production and global trade are homogenising supplies and, as a consequence, a reduced group of species is providing most of the plant commodities for the entire human population, threatening sustainability[32]. One strategy to overcome this limitation is to diversify plant sources by domesticating new wild plants, completing the domestication of partial domesticates and improving underutilised crops[29]. By studying the evolutionary history of crop species, ancient plant DNA studies can identify past plant uses, lost crop progenitors and partial domesticates while investigating their genetic composition. For example, recent domestication and cultivation efforts of several ancient and underutilised crops, such as quinoa, amaranth, chenopod and teff, have demonstrated the potential of adopting new crops that can thrive in harsh environments, while contributing to a more varied and nutritious human diet[33,34].

Despite the limited number of studies in ancient plant DNA compared to those in other organisms, plant aDNA research has progressed from the analysis of a few loci to the retrieval of whole ancient genomes, transcriptomes and epigenomes. Looking forward to the challenges that natural populations are currently facing in climate change and habitat modification, aDNA research will play a significant role in studying how plant species have evolved and adapted to environmental changes through time. Due to their stationary existence and ubiquitous distribution across the Earth's surface since the Paleozoic era (541–252 million years ago), the genomes and epigenomes of plant species are incredibly valuable markers of large-scale climate upheaval. The ability of plants to respond to climate change through natural selection, and absorb fine-scale climate shifts through epigenetic mechanisms, have been key drivers of their distribution across the planet — in-turn enabling modern, human community development and cultural evolution. By using aDNA to track variations in allele frequency at critical genetic loci and tracing epigenetic changes across gene regulatory networks, researchers will be able to elucidate exact timings of historical climate events and directly observe evolutionary changes. ❐

Oscar Estrada[1], James Breen[2,3]*, Stephen M. Richards[1] and Alan Cooper[1]
[1]*Australian Centre for Ancient DNA, University of Adelaide, South Australia, Australia.* [2]*Robinson Research Institute, University of Adelaide, South Australia, Australia.* [3]*University of Adelaide Bioinformatics Hub, School of Biological Sciences, University of Adelaide, South Australia, Australia.*
**e-mail: jimmy.breen@adelaide.edu.au*

Published online: 18 June 2018
https://doi.org/10.1038/s41477-018-0187-9

References

1. Willerslev, E. et al. *Science* **317**, 111–114 (2007).
2. Orlando, L. & Cooper, A. *Annu. Rev. Ecol. Evol. Syst.* **45**, 573–598 (2014).
3. Palmer, S. A., Smith, O. & Allaby, R. G. *Ann. Anat.* **194**, 146–156 (2012).
4. Jiao, W. B. & Schneeberger, K. *Curr. Opin. Plant Biol.* **36**, 64–70 (2017).
5. Styring, A. K. et al. *J. Archaeol. Sci.* **40**, 4767–4779 (2013).
6. Nistelberger, H. M., Smith, O., Wales, N., Star, B. & Boessenkool, S. *Sci. Rep.* **6**, 37347 (2016).
7. Wales, N., Andersen, K., Cappellini, E., Avila-Arcos, M. C. & Gilbert, M. T. P. *PLoS ONE* **9**, e86827 (2014).
8. Schrader, C., Schielke, A., Ellerbroek, L. & Johne, R. *J. Appl. Microbiol.* **113**, 1014–1026 (2012).
9. Mascher, M. et al. *Nat. Genet.* **48**, 1089–1093 (2016).
10. Ramos-Madrigal, J. et al. *Curr. Biol.* **26**, 3195–3201 (2016).
11. da Fonseca, R. R. et al. *Nat. Plants* **1**, 14003 (2015).
12. Kistler, L. et al. *Proc. Natl Acad. Sci. USA* **111**, 1–5 (2014).
13. Wagner, S. et al. *Mol. Ecol.* **27**, 1138–1154 (2018).
14. Sallon, S. et al. *Science* **320**, 1464 (2008).
15. Pérez-Zamorano, B. et al. *Genome Biol. Evol.* **9**, 904–915 (2017).
16. Nakabayashi, K., Okamoto, M., Koshiba, T., Kamiya, Y. & Nambara, E. *Plant J.* **41**, 697–709 (2005).
17. Fordyce, S. L. et al. *PLoS ONE* **8**, 1–9 (2013).
18. Smith, O. et al. *Mol. Biol. Evol.* **37**, 2555–2562 (2017).
19. Yoshida, K. et al. *eLife* **2**, 1–25 (2013).
20. Martin, M. D. et al. *Nat. Commun.* **4**, 2172 (2013).
21. Smith, O. et al. *Sci. Rep.* **4**, 4003 (2014).
22. Cubas, P., Vincent, C. & Coen, E. *Nature* **401**, 157–161 (1999).
23. Springer, N. M. & Schmitz, R. J. *Nat. Rev. Genet.* **18**, 563–575 (2017).
24. Lämke, J. & Bäurle, I. *Genome Biol.* **18**, 1–11 (2017).
25. Llamas, B. et al. *PLoS One* **7**, e30226 (2012).
26. Gokhman, D. et al. *Science* **344**, 523–527 (2014).
27. Smith, O. et al. *Sci. Rep.* **4**, 5559 (2015).
28. Meyer, R. S., DuVal, A. E. & Jensen, H. R. *New Phytol.* **196**, 29–48 (2012).
29. Bevan, M. W. et al. *Nature* **543**, 346–354 (2017).
30. Gururani, M. A. et al. *Physiol. Mol. Plant Pathol.* **78**, 51–65 (2012).
31. Mickelbart, M. V., Hasegawa, P. M. & Bailey-Serres, J. *Nat. Rev. Genet.* **16**, 237–251 (2015).
32. Khoury, C. K. et al. *Proc. Natl Acad. Sci. USA* **111**, 4001–4006 (2014).