Studies of genetic diversity in wild and cultivated barley and wheat

A thesis submitted to the University of Manchester for the degree of Doctor of Philosophy (PhD by Published Work) in the Faculty of Science and Engineering

2019

Beata Izabela Czajkowska

Contents list Page

Title page	**1**
Contents	**2**
Abstract	**4**
Declaration	**5**
Declaration Part i	5
Declaration Part ii	6
Declaration Part iii	6
Copyright statement	**7**
Copyright statement Part i	7
Copyright statement Part ii	7
Copyright statement Part iii	7
Copyright statement Part iv	7
Statement	**8**
Statement Part i	8
Statement Part ii	13
Statement Part iii	13
Journal standard	13
Current state of knowledge	14
Contribution to field	22
Future work	34
References	36
Acknowledgments	**47**

Copies of each of the publications 48

1. Czajkowska BI, Finlay CM, Jones G, Brown TA (2019) *Diversity of a cytokinin dehydrogenase gene in wild and cultivated barley*. PLoS ONE 14(12): e0225899. https://doi.org/10.1371/journal.pone.0225899. 48

2. Czajkowska BI, Jones G, Brown TA (2019) *Diversity of a wall-associated kinase gene in wild and cultivated barley*. PLoS ONE 14(6):e0218526. https://doi.org/10.1371/journal. pone.0218526. 49

3. Czajkowska BI, Oliveira HR, Brown TA (2019) *A discriminatory test for the wheat B and G genomes reveals misclassified accessions* of *Triticum timopheevii and Triticum turgidum.* PLoS ONE 14(4):e0215175. https://doi.org/10.1371/journal.pone.0215175. 50

Supplementary material

An electronic version of supporting information (SI),
Diploma Certificate and Diploma Supplement
in the form of a compact disc attached to the back of the thesis

Abstract

The University of Manchester
Mgr Beata Izabela Czajkowska, MIScT, RSci
Doctor of Philosophy
Studies of genetic diversity in wild and cultivated barley and wheat
2019

I am presenting three first author research publications in my PhD thesis. Copies of my manuscripts in their published form are preceded by a Statement (part iii) that explores the scientific information at a more detailed and advanced level, summarizes the aims and achievements of my work in the light of current knowledge, incorporates the reviewers comments and discusses my contribution to the field.

I studied advanced genetic implications of the process of domestication: the nature and consequences of selective pressures acting on the barley crop in prehistoric agriculture as well as molecular differences within highly similar wheat species to enable reassessment of their identity, phylogenetic affinity and the composition of the Neolithic crop package.

The first two papers investigated the molecular patterns of traits assumed to have been targeted during selection for domestication in barley. They report on two phenotypic traits, water utilization and growth vigour, as potentially novel traits targetted during the selection for domestication. Molecular analysis of the sequence diversity of a cytokinin dehydrogenase gene *CKX2.1*, associated with water utilization, and a wall-associated kinase gene *WAK1*, involved in root proliferation, in an extensive set of wild barley accessions and historical barley domesticates, highlights the possibility that these genes evolved under selection for domestication and are candidates for genes controlling these phenotypic traits. The observed extreme changes occurred within the coding region of the *CKX2.1* gene and within *WAK1 cis*-regulatory sequences. The absence of *CKX2.1* haplotype 3 in landraces compared to wild accessions raised the possibility that selection for domestication was accompanied by water utilization changes imposed by husbandry techniques and that this physiological adaptation to cultivation practices is likely to be a component of the domestication syndrome. Fixation of SNPs within the *WAK1* promoter in landraces supports the proposed scenario that our ancestors selected for the overall growth vigour of the plant rather than directly for seed size and proposes plant vigour as a conceivable component of the domestication suite.

The third paper is on the molecular identification of hulled, tetraploid wheat species. It describes first, a straightforward and unambiguous PCR test directed at the wheat B and G genome-specific *Ppd-1* gene that I have developed in order to distinguish between *dicoccoides-dicoccum* group, (*Turgidum*/emmer wheats, A_uA_uBB) and *araraticum-timopheevii* group (*Timopheevii* wheats, A_tA_tGG), for which morphological misidentification occurs. The test revealed the morphologically misclassified wheat accessions and proved consistent with the results of whole-genome analysis. The discriminatory test for the B and G genomes not only allows fast and correct classification of extant groups, which is a prerequisite for inferring accurate phylogenetic relationships, but it is also suitable for ancient DNA (aDNA) to establish which tetraploid wheat group the archaeological, extinct today, so-called New Glume Wheat, NGW, belongs to.

Declaration

i. Nature and extent of the candidate's contribution and the co-authors

1. Czajkowska BI, Finlay CM, Jones G, Brown TA (2019) *Diversity of a cytokinin dehydrogenase gene in wild and cultivated barley*. PLoS ONE 14(12): e0225899. https://doi.org/10.1371/journal.pone.0225899.

    Beata I. Czajkowska

    Project design, Methodology – wet lab work, bioinformatics and data analysis, Writing – original draft, review & editing

    Glynis E. Jones

    Conceptualization, Funding acquisition, Writing – review & editing

    Terence A. Brown

    Conceptualization, Data analysis, Funding acquisition, Project administration, Supervision, Writing – review & editing

2. Czajkowska BI, Jones G, Brown TA (2019) *Diversity of a wall-associated kinase gene in wild and cultivated barley*. PLoS ONE 14(6):e0218526. https://doi.org/10.1371/journal. pone.0218526.

    Beata I. Czajkowska

    Project design, Methodology – wet lab work, bioinformatics and data analysis, Writing – original draft, review & editing

    Glynis E. Jones

    Conceptualization, Funding acquisition, Writing – review & editing

    Terence A. Brown

    Conceptualization, Data analysis, Funding acquisition, Project administration, Supervision, Writing – review & editing

3. Czajkowska BI, Oliveira HR, Brown TA (2019) *A discriminatory test for the wheat B and G genomes reveals misclassified accessions of Triticum timopheevii and Triticum turgidum*. PLoS ONE 14(4):e0215175. https://doi.org/10.1371/journal.pone.0215175.

    Beata I. Czajkowska

        Conceptualization, Project design, Methodology – wet lab work, bioinformatics and data analysis, Writing – original draft, review & editing

    Hugo R. Oliveira

        Methodology, Writing – review & editing

    Terence A. Brown

        Conceptualization, Funding acquisition, Project administration, Supervision, Writing – review & editing

ii. All of the presented Work has been completed while the author was a member of staff of the University of Manchester. Work has been carried out with Professor Terry Brown since 21 January 2013 in the field of Evolutionary Genetics.

iii. None of the work presented has been submitted in support of a successful or pending application for any other degree or qualification of this or any other University or of any professional or learned body.

Copyright statement

i. The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the "Copyright") and s/he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

ii. Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made only in accordance with the Copyright Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

iii. The ownership of certain Copyright, patents, designs, trade marks and other intellectual property (the "Intellectual Property") and any reproductions of copyright works in the thesis, for example graphs and tables ("Reproductions"), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

iv. Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy
(see http://documents.manchester.ac.uk/DocuInfo.aspx?DocID=487),
in any relevant Thesis restriction declarations deposited in the University Library, The University Library's regulations
(see http://www.manchester.ac.uk/library/aboutus/regulations)
and in The University's policy on Presentation of Theses.

Statement

i. Particulars of the candidate

Degree

Awarded Magister (Mgr) in Faculty of Biology and Environment Protection, University of Silesia, Katowice, Poland.

Magister Research Project: *Analysis of mutagens induced mutation in Arabidopsis thaliana (L.) Heynh, by application of DNA markers*. Biotechnology of Plants and Microorganisms, Department of Genetics.

- Achieved the highest professional title Magister awarded to graduates of institutions of higher education following the completion of uniform 5-year Magister level course in field of Biology.
- Final Overall Grade: Very Good
  Arithmetic mean of all obtained grades during the whole study, grade obtained as a result of the Magister Thesis evaluation and grade from the Magister Exam in the proportion of ½: ¼ : ¼, respectively.
- 5 years of studies included 3 years of specialization in Biotechnology of Plants and Microorganisms with Specialized Subjects among others: Molecular Biology, Genetic Engineering, Genetics, Cytogenetics, Biotechnology of Plants, Instrumental Analysis, Molecular Markers.
- Magister program included 927 hours of lectures and 2715 hours of laboratories. Additionally 3 years Magister project evaluation in Department of Genetics.

Research Experience at University of Manchester

- 1 October 2019 – present. Senior Research Technician. Plant Stress group, Giles Johnson's lab.
  Research project: Plastid Terminal Oxidase – a target for improving food security.
- 1 May 2019 – present. Senior Research Technician. Geomicrobiology group, Jonathan Lloyd's lab.
- 1 August 2016 – present. Permanent work in the role of Senior Research Technician. University of Manchester, Faculty of Science and Engineering.

- 12 May 2008 – 31 July 2016. Permanent work in the role of Senior Research Technician (from 30 April 2012) and Research Technician. University of Manchester, Faculty of Life Sciences.
- 21 January 2013 – 30 April 2019. Senior Research Technician. Biomolecular Archaeology group, Terry Brown's lab.
  Research projects: The Evolutionary Origins of Agriculture: Genetic implications of a protracted crop domestication process for the origins of agriculture. The identity of the mysterious 'new glume wheat' of early European agriculture.
- 12 November 2012 – 18 January 2013. Plant Science group, Simon Turner's lab. *Temporary affiliation, research support.*
- 7 November 2012 – 27 August 2013. Synthetic Biology group, Eriko Takano's lab. *Lab establishment and management.*
- 18 October 2012 – 9 November 2012. Manchester Collaborative Centre for Inflammation Research (MCCIR). *Centre starting and establishment.*
- 16 February 2009 – 9 November 2012. Immunology group. Mark Travis's lab. *Laboratory Floor Coordinator/Representative 1 November 2009 – 9 November 2012.*
  Research project: How does integrin alphav beta8 signal to control immune responses?
- 21 July 2008 – 13 February 2009. Microbiology group. Dennis Linton's lab.
  Research project: Exploitation of Camplylobacter glycosylation system – a new era for glycoengineering and vaccine development.
- 12 May 2008 – 18 July 200. Research Technician Plant Science group. Minsung Kim's lab.
  Research project: Plant hormones and flower development in Asteraceae.

Publications

http://orcid.org/0000-0002-8596-9288

1. **Czajkowska BI**, Brown TA *Efficient sequencing of long-range amplicons generated by real-time PCR.* In preparation.
2. Jones G, Cunniff J, Kluyver T, Preece C, Martin G, Forster EE, Wallace M, Frenck G, Bennett C, Burnet K, **Czajkowska BI**, Brown TA, Charles M,

Rees M, Osborne C *The Origins of Agriculture: Intentions and Consequences*. In preparation.

3. **Czajkowska BI**, Jones G, Brown TA *Ancient DNA typing indicates that the 'new glume wheat' of early Eurasian agriculture is related to Triticum timopheevii.* In preparation.
4. Oliveira HR, Jacocks L, **Czajkowska BI**, Kennedy SL, Brown TA *Multiregional origins of the domesticated tetraploid wheats*. PLoS ONE PONE-D-19-23497. Accepted/ in press.
5. **Czajkowska BI**, Finlay CM, Jones G, Brown TA (2019) *Diversity of a cytokinin dehydrogenase gene in wild and cultivated barley*. PLoS ONE 14(12): e0225899.
6. **Czajkowska BI**, Jones G, Brown TA (2019) *Diversity of a wall-associated kinase gene in wild and cultivated barley*. PLoS ONE 14(6):e0218526.
7. **Czajkowska BI**, Oliveira HR, Brown TA (2019*) A discriminatory test for the wheat B and G genomes reveals misclassified accessions of Triticum timopheevii and Triticum turgidum*. PLoS ONE 14(4):e0215175.
8. Steel N, Faniyi AA, Rahman S, Swietlik S, **Czajkowska BI**, Chan BT, Hardgrave A, Steel A, Sparwasser TD, Assas MB, Grencis RK, Travis MA, Worthington JJ (2019) *TGFβ-activation by dendritic cells drives Th17 induction and intestinal contractility and augments the expulsion of the parasite Trichinella spiralis in mice.* PLoS Pathog 15(4):e1007657. Editorial feature.
9. Palasz A, Lapray D, Peyron C, Rojczyk-Golebiowska E, Skowronek R, Markowski G, **Czajkowska B**, Krzystanek M, Wiaderkiewicz R (2014) *Dual orexin receptor antagonists – promising agents in the treatment of sleep disorders.* Int J Neuropsychopharmacol 17(1):157-168.
10. Worthington JJ, Klementowicz JE, Rahman S, **Czajkowska BI**, Smedley C, Waldmann H, Sparwasser T, Grencis RK, Travis MA (2013) *Loss of the TGFβ-activating integrin αvβ8 on dendritic cells protects mice from chronic intestinal parasitic infection via control of type 2 immunity*. PLoS Pathog 9(10):e1003675.
11. Worthington JJ, Fenton TM, **Czajkowska BI**, Klementowicz JE, Travis MA (2012) *Regulation of TGFβ in the immune system: an emerging role for integrins and dendritic cells*. Immunobiology 217(12):1259-1265.

12. Palasz A, Krzystanek M, Worthington J, **Czajkowska B**, Kostro K, Wiaderkiewicz R, Bajor G (2012) *Nesfatin-1, a unique regulatory neuropeptide of the brain.* Neuropeptides 46(3):105-112.
13. Pałasz A, Wiaderkiewicz A, Wiaderkiewicz R, Czekaj P, **Czajkowska B**, Lebda-Wyborny T, Piwowarczyk A, Bryzek A (2012) *Age-related changes in the mRNA levels of CYP1A1, CYP2B1/2 and CYP3A1 isoforms in rat small intestine.* Genes Nutr 7(2):197-207.
14. Worthington JJ, **Czajkowska BI**, Melton AC, Travis MA (2011) *Intestinal dendritic cells specialize to activate transforming growth factor-β and induce Foxp3+ regulatory T-cells via integrin αvβ8.* Gastroenterology 141(5):1802-1812. Editorial feature.
15. Worthington JJ**, Czajkowska B**, Travis MA (2010) *Enhanced induction of Foxp3+regulatory T-cells by specialised gut dendritic cells is dependent on TGF-beta activation by the integrin alphaVbeta8.* Immunology 131:35-35.
16. Wiaderkiewicz R, Zasadowski A, Czekaj P, Wiaderkiewicz A, **Czajkowska B**, Barski D, Pałasz A (2006) *Effect of pyrantel and dimethoate administration on rat liver cytochrome P450 system.* Bull Vet Inst Pulawy 50(2):253-257.

Presentations

1. 3-7 July 2017 – Society for Experimental Biology Conference (SEB Gothenburg 2017). Czajkowska BI, Brown TA. *Comparative molecular analysis of genes underlying domestication traits in barley.* Crop Molecular Genetics. Results for papers 1 and 2.
2. 5-8 April 2017 – the UK Archaeological Science Conference (UKAS London 2017). Czajkowska BI, Brown TA. *The identity of the mysterious 'new glume wheat' of early European agriculture.* B and G genome specific nuclear gene variants. Data and sources for paper 3.
3. 14-16 September 2016 – 7th International Symposium on Biomolecular Archaeology (ISBA7 Oxford 2016). Czajkowska BI, Brown TA. *The identity of the mysterious 'new glume wheat' of early European agriculture.* B and G genome specific nuclear gene variants. Data and sources for paper 3.
4. 4-9 July 2016 – 17th conference of the International Work Group for Palaeoethnobotany (IWGP17 Paris 2016). Czajkowska BI, Brown TA. *The*

identity of the mysterious ‘new glume wheat’ of early European agriculture. Possible diagnostic DNA loci of wheat plastome and nuclear genome. Data and sources for manuscript in preparation.

Progress reports and journal club talks during the Brown lab meetings

1. 23/03/16 – Progress Report. Results for the manuscripts in preparation.
2. 17/06/15 – Journal Club. *RNA from ancient maize kernels*.
3. 24/11/14 – Progress report. Results for paper 1, 2 and manuscript in preparation.
4. 01/07/14 – Journal Club. *The evolutionary origin of cancer*.
5. 14/03/14 – Progress report. Results for paper 1 and manuscript in preparation.

Magister talks

1. Tenth semester talk. *Methods of mutations detection.*
2. Ninth semester talk. *DNA markers in research on stability and variability of the plant genomes*.
3. Eighth semester talk. *The problems associated with the expression of introduced genes (‘gene silencing’).*
4. Seventh semester talk. *The application of molecular markers in genetic and phylogenetic research.*
5. Sixth semester talk. *Prions.*

Highlighted Awards

- 8 October 2013. Identified as a scientific professional and offered a consultant role by The Science Advisory Board, an exclusive global community of life science and medical professionals.
- 6 November 2013. Awarded Registered Scientist (RSci) status, professional registration with the Science Council awarded to those working in scientific and higher technical roles.

- 1 June 2013. Awarded Full Corporate grade of Member (MIScT) of the Institute of Science and Technology (IST).
- 1 May 2012. Role approved by HERA and HR team for a change to Senior Research Technician.
- Achieved the highest professional title, Magister at University of Silesia, awarded to graduates of institutions of higher education following the completion of uniform 5-year Magister level course in field of Biology.
- Awarded Teaching Certificate – achieved detailed qualification required from teacher (5 years teaching course in parallel with Magister course).
- Given Annual Award of Excellence for the best fourth year biology student of University of Silesia.

ii. List of publications submitted

1. **Czajkowska BI**, Finlay CM, Jones G, Brown TA (2019) *Diversity of a cytokinin dehydrogenase gene in wild and cultivated barley*. PLoS ONE 14(12): e0225899. https://doi.org/10.1371/journal.pone.0225899.
2. **Czajkowska BI**, Jones G, Brown TA (2019) *Diversity of a wall-associated kinase gene in wild and cultivated barley*. PLoS ONE 14(6):e0218526. https://doi.org/10.1371/journal. pone.0218526.
3. **Czajkowska BI**, Oliveira HR, Brown TA (2019*) A discriminatory test for the wheat B and G genomes reveals misclassified accessions of Triticum timopheevii and Triticum turgidum*. PLoS ONE 14(4):e0215175. https://doi.org/10.1371/journal.pone.0215175.

iii. An overall summary of the aims and achievement of the work

Journal standard

All my work has been published in PLoS ONE which is a multidisciplinary and also interdisciplinary Open Access publisher. The Public Library of Science (PLoS) describes itself as an innovative organization with a mission to advance progress in science and medicine. PLoS ONE is fully peer reviewed journal with a rigorous multi stage editorial assessment process. In order to publish in PLoS ONE, the following criteria must be met:

1. The study presents the results of original research.
2. Results reported have not been published elsewhere.
3. Experiments, statistics, and other analyses are performed to a high technical standard and are described in sufficient detail.
4. Conclusions are presented in an appropriate fashion and are supported by the data.
5. The article is presented in an intelligible fashion and is written in standard English.
6. The research meets all applicable standards for the ethics of experimentation and research integrity.
7. The article adheres to appropriate reporting guidelines and community standards for data availability

PLoS ONE considers original research in all fields and has a 2018 impact factor of 2.776. It provides a comprehensive assessment of research impact with Article-Level Metrics (ALMs) which follow the impact of a paper in real time, by considering viewership, download rates, social sharing and citations.
(https://journals.plos.org/plosone/s/journal-information).

Due to very recent publishing record, it is too early to expect any citation of my research. However, available ALM elements are incorporated in the discussion below.

Current state of knowledge

The relatively recent human transition from a mobile hunter-gatherer lifestyle to agricultural settings around 10,000 years ago is indisputably the most important switch in 250,000 years of human history (Brown et al, 2009; Zohary et al, 2012). The sedentary farming existence stimulated the development of advanced political, economic and technological systems leading to the complex civilization of human populations. In order to manage limited food resources, mobile foraging groups developed a mutualistic relationship with the natural environment by growing crops and herding animals (Diamond, 2002; Barker, 2006; Morrell and Clegg, 2007; Ruddiman, 2007; Kaplan et al, 2010; Kutzbach et al, 2010; Lemmen et al, 2011). This shift transformed global history by inducing fundamental ecological and sociological changes that led to humans taking control of the natural environment (Brown, 1999; Brown et al, 2009) and becoming less dependant on seasonal timing

(Fuller, 2007). This switch in the method of sustainability is significantly marked by the evolution of domesticated crops, upon which agriculture is founded, and therefore the mechanism of the transition in human subsistence behaviour is often reflected in changes to crop plants that occurred during their domestication.

The earliest evidence of domesticated crops occurs in southwest Asia at the beginning of the 10th millennium before present (BP) in the western Fertile Crescent of southeast Turkey, northern Syria and southern Levant (Zohary et al, 2012) for three cereal crops: einkorn wheat (*Triticum monococcum* L.), emmer wheat (*T. turgidum* L. subsp. *dicoccum* [Schrank ex Schübl.] Thell.) and barley (*Hordeum vulgare* L.). The speed and location of domestication has been subject to challenging dispute, with one view supporting a rapid revolution which occurred in an isolated geographical area (e.g. Zohary, 1999) and the other a gradual widespread process occurring throughout western Asia (e.g. Willcox et al, 2008) (Table 1).

Protracted model	Rapid model
Polycentric: recurrent event, gradual and widespread process, multiple events	Monocentric: singular timing, domestication episode
Polyphyletic evolution	Monophyletic evolution
Relatively large portion of genetic diversity at multiple loci in modern crops and profiles of diversity across geographic space	Extensive reduction in genetic diversity in modern crops throughout the genome: “domestication bottleneck”
Slow fixation of domesticated traits	Rapid and simultaneous fixation of domesticated traits
Even balance between artificial and natural selective pressure mediated by a gen flow	Artificial selective pressure much stronger than natural selection pressure
Post-Neolithic extinction of potential founder crops	Classic Neolithic eight founder crops package
Knowledge transfer or independent innovation	Outward rapid radiation of domesticated species
Proponents: Gebel, 2004; Willcox, 2004; Molina-Cano et al, 2005; Willcox, 2005; Fuller, 2007; Morrell and Clegg, 2007; Saisho and Purugganan, 2007; Allaby et al, 2008; Jones et al, 2008; Willcox et al, 2008; Brown et al, 2009; Jones et al, 2011; Morrell et al, 2014; Brown, 2018; Allaby et al, 2019; Brown, 2019.	Proponents: Heun et al, 1997; Zohary, 1999; Badr et al, 2000; Özkan et al, 2002; Doebley et al, 2006; Abbo et al, 2013; Abbo et al, 2014.

Table 1. Comparison of the protracted and rapid transition models of domestication.

The rapid domestication model was developed in the early 1990s and was supported by primary genetic data that suggested that domestication was a localized event that occurred only once, or at least very rarely (Heun et al, 1997; Badr et al, 2000; Özkan et al, 2002) and showed that the process could have been relatively fast (Gepts, 2004). Domestication under a fast radiation model would have proceeded in three principal steps strictly following the climatic changes between the Pleistocene and Holocene:

1. True hunting and gathering
2. A small amount of pre-domestication cultivation
3. A rapid rise of domesticated crops and an explosive expansion of agriculturists out of the centre(s) of origin (Blumler, 1992; Diamond, 1997; Allaby et al, 2008).

According to this determinist model for the origins of agriculture, the climatic deterioration of the Younger Dryas in the late Pleistocene led to a reduction in the distribution and availability of previously collected wild cereals that prompted hunter-gatherers to start their intense cultivation (Willcox, 2005). The Younger Dryas was a period of rapid cooling during the Early Upper Palaeolithic, approximately 12,800 years ago, followed by sharp warming that brought the last Ice Age to a sudden end 11,500 years ago during the Late Upper Palaeolithic (Grosman and Belfer-Cohen, 2002). However, the increase in archaeological evidence and the development of sophisticated molecular techniques has brought to an end the illusion that the process was rapid, and has shifted the view toward a protracted, widespread, and quantitative process of change with prolonged predomestication cultivation (Gepts, 2004; Willcox, 2005; Allaby et al, 2008; Brown et al, 2009; Allaby, 2010). The notion of a “Neolithic Revolution” has also been challenged by computational simulations, which have proved that previous genetic studies were incorrectly performed and misanalysed (Allaby et al, 2008; Brown et al, 2009). Further revisions revealed that the previous genetic studies were affected by genetic events occurring during the recent Green Revolution and not just the origin of agriculture. Formal plant breeding often replaced landraces, historical domesticates, with modern cultivars, which have no association with traditional farming systems. Only landraces have historical origins, high local genetic adaptation, recognizable identities, genetic diversity and a lack of formal genetic crop improvement (Jones et al, 2008; Morrel et al, 2014). A significant amount of the undeniable evidence obtained from numerous disciplines suggests that the origin of agriculture was not an abrupt, highly localised and

episodic event as originally thought, but the result of a gradual cultural evolution with a prolonged multistage transition (Fig 1). This protracted transition model is also advocated by extensive evidence of 20 millennia of significant exploitation of wild plant resources, including the wild progenitors of domesticated crops (Willcox et al, 2008). Archaeological data for the reduction and elimination of the natural seed dispersal process, resulting in the emergence of crops displaying indehiscence, which is a key trait in the domestication of cereals, also shows that process was slow (Tanno and Willcox, 2006). Now that it has been widely accepted that domestication is a multistage process, a classic model of Harris (1989) can be used to distinguish four general stages based on their archaeological visibility:

1. True hunting and gathering (wild plant food procurement)
2. The very beginnings of cultivation (wild plant food production)
3. Systematic cultivation (of morphologically wild plants)
4. Agriculture based on domesticated plants

The above is a simplified description to represent the mode of prolonged transition in the context of the demographical model of rapid change. It is has been suggested that the difference between wild plant procurement and wild plant production is very subtle: for example manipulation of wild stands resulting in improved productivity is still a form of gathering but also can be looked on as a type of cultivation. Archaeological examination show that our ancestors began to manipulate wild stands of cereals by soil improvement and weeding around 2000 years before fully domesticated crops first appear in the archaeobotanical record (Brown, 2018).

Selection is a key mechanism of evolution that acts on any heritable phenotypic trait. Plant domestication results from selective processes and can be defined as the increased adaptation of plants to cultivation in order to sustain humans (Gepts, 2004). Selection for domestication is a process of preferential elimination or enhancement of particular traits by environmental and agricultural means. Natural pressures act towards the species' ability to reproduce and survive, while artificial controlling factors imposed by human action may enhance or repress an organism's genetic traits through selective breeding which often means the species being subjected to an opposing evolutionary path in order to meet human needs (Brown et al, 2009). Domestication is a complex evolutionary process driven by cultural and ecological selection and resulting in genetic modification of a wild species to create

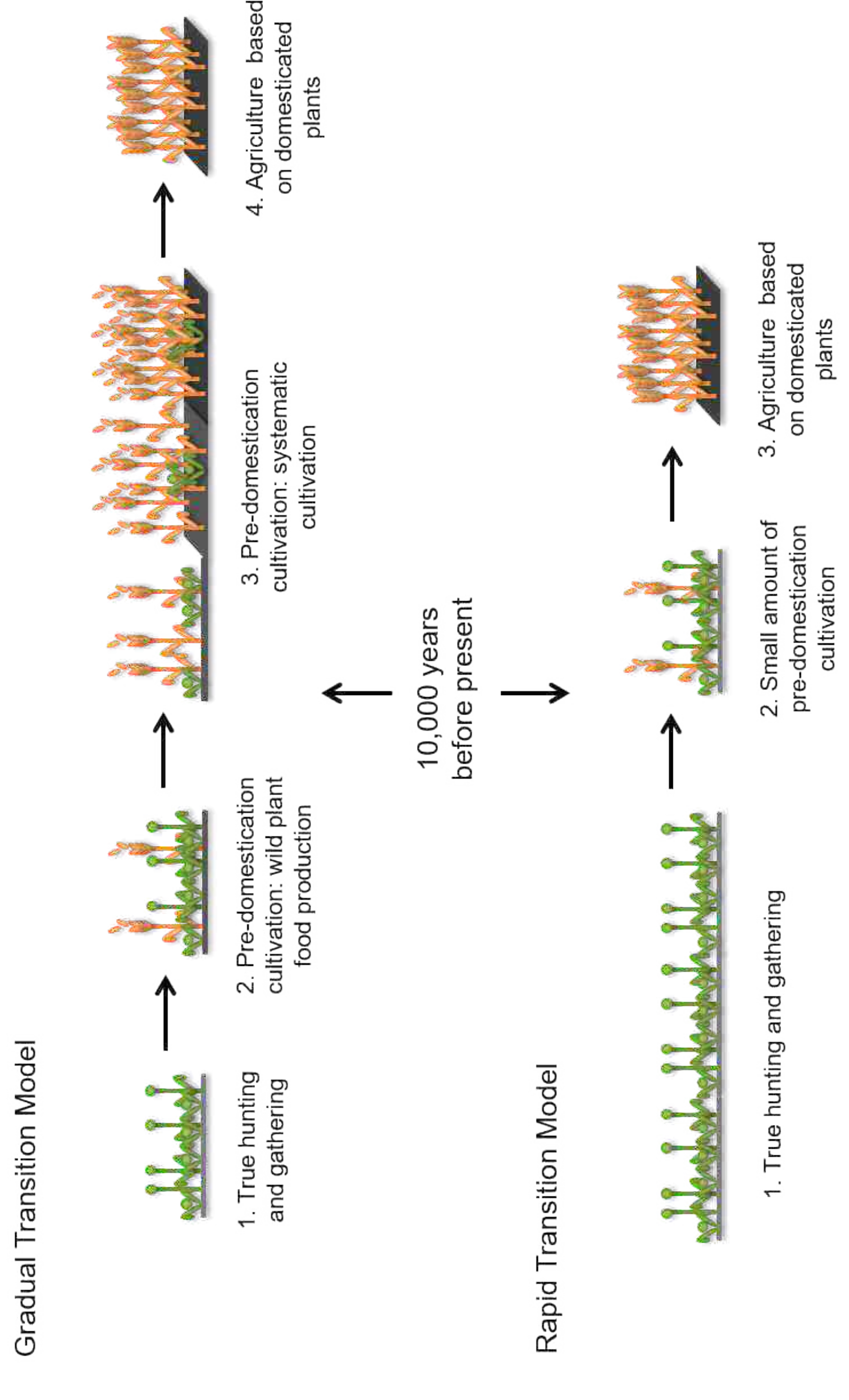


Fig 1. Stages of the gradual and determinist models of domestication (adapted from Brown, 2018).

a new form altered by and for humans, with those genes controlling domestication traits subjected to the competing effects of natural selection as well as the artificial selective forces imposed by cultivation. A combination of indirect and direct selection may have operated in successful domestication, with the former operating in the early stages.

A compendium of morphological and physiological features of cultivated species called the “domestication syndrome” distinguishes crop plants from their wild ancestors. Studies of the domestication syndrome have shown that similar traits have been selected during the domestication of different crops, these including, for cereals:

1. Elimination/reduction of natural seed dispersal (indehiscence)
2. Reduction in seed dispersal aids
3. Trends towards increasing seed/fruit size
4. Loss of germination inhibition (lack of or reduced dormancy)
5. Synchronous tillering and ripening
6. More compact growth habit (Hammer, 1984; Harlan, 1992)

Selection can be identified by studying genetic diversity and its effect on a trait. Studying sequence variability of the genes underling the domestication syndrome has proven successful and complementary to archaeological research (Jones et al, 2008; Pourkheirandish et al, 2015). Most domestication traits have zero or insignificant adaptive advantage in the wild but high adaptive advantage in the crop, and so are assumed to have been selected as a result of cultivation practices (Zohary, 1996). Apart from traditional approaches such as candidate gene resequencing and production of classic mapping populations, significant advances have been made in gene identification methods based on association studies with more diverse mapping populations, such as nested association mapping (NAM) (Buckler et al, 2009) and a genome-wide association study (GWA study, or GWAS) (Ramsay et al, 2011). NAM is a method for the whole genome linkage analysis of a complex trait in a single mapping population, while GWAS is a “phenotype first” approach that relies on observational study of a genome-wide set of traits. However, all mapping methods require prior information on the traits that are the subject of investigation (Olsen and Wendel, 2013), therefore resequencing candidate genes remains a very practical and competent approach, especially valuable in identification of new genetic targets underlying novel domesticated phenotypes.

Agriculture sits in the nexus for the cultural changes occurring during the Neolithic, and ascribing domesticated plants to the classic “Neolithic package” is currently of special interest for archaeologists. The conventional Neolithic package, as described by Vere Gordon Childe (1928), comprises herding, polished stone axes, timber long houses, pottery, and animals and plants that were originally domesticated in the Near East and whose domesticated forms spread into Europe. The Neolithic package is marked by a classic set of eight founder crops, which form the basis to the hypothesis that agriculture emerged in a “core area” in the Near East. These founder crops are:

- diploid einkorn wheat (*Triticum monococcum* L.)
- tetraploid emmer wheat (*Triticum turgidum* L.)
- barley (*Hordeum vulgare* L.)
- lentil (*Lens culinaris* Medikus)
- pea (*Pisum sativum* L.)
- chickpea (*Cicer arietinum* L.)
- bitter vetch [*Vicia ervilia* (L.) Willd.]
- flax (*Linum usitatissimum* L.)

However, archaeological data upon which the diffused transition model is founded indicates that a larger number of domesticated species should be addressed, as additional taxa could have existed at certain points in time and in certain locations, some as genuine crops that were later abandoned (Abbo et al, 2013). The so-called "new glume wheat" (NGW) is an example, being an unidentified type of domesticated wheat, which is extinct today, formerly classified as "aberrant" or "slender emmer". In prehistoric times the NGW could have been a widespread crop with archaeobotanical records ranging from Anatolia to Western Germany, from various cultures from the Early Neolithic to the Iron Age (Jones et al, 2000; Kohler-Schneider, 2003; Kenéz et al, 2014). The unique morphological characteristics of the NGW, although indicative of a genetically distinct type, do not provide an accurate taxonomic identification. Wheat belongs to the genus *Triticum*, an allopolyploid complex that includes among others the tetraploid *araraticum-timopheevii* (*Timopheevii* wheats) and *dicoccoides-dicoccum* (emmer wheats) groups. These two groups have very similar morphology and therefore misidentification of their accessions occasionally occurs. Although the morphological evidence supports a relationship of the NGW with the *araraticum-timopheevii* group, it is also conceivable that the NGW is descended from *T.*

dicoccoides, for example by a separate domestication to that giving rise to *T. dicoccum* or by mutation of an existing *T. dicoccum* population. The question might be resolved because of unambiguous differences at the molecular level: *Timopheevii* wheat possesses the G and A_u genomes; whereas emmer's genomic composition is B and A_u (Fig 2). Domesticated *T. timopheevii* is not a common species, being restricted to localized regions of Georgia, it has been given a "secondary role" in agriculture due to its uncommon existence and limited exploitation. The widespread presence in Europe of domesticated wheat related to *Timopheevii* would force a debate on the composition of the conventional Neolithic package, as well as the geographical origins of agriculture in the Fertile Crescent. Archaeobotanical studies, on their own, cannot identify the NGW. One possibility would be genetic analysis of ancient DNA (aDNA) from NGW samples, but a major challenge lies with the nature of the archaeobotanical specimens that are available. Like most plant remains from Eurasian archaeological sites, virtually all available samples of the NGW have been preserved by charring from which it is difficult to obtain aDNA, and when genuine aDNA is obtained, it is highly fragmented and often contaminated by environmental DNA.

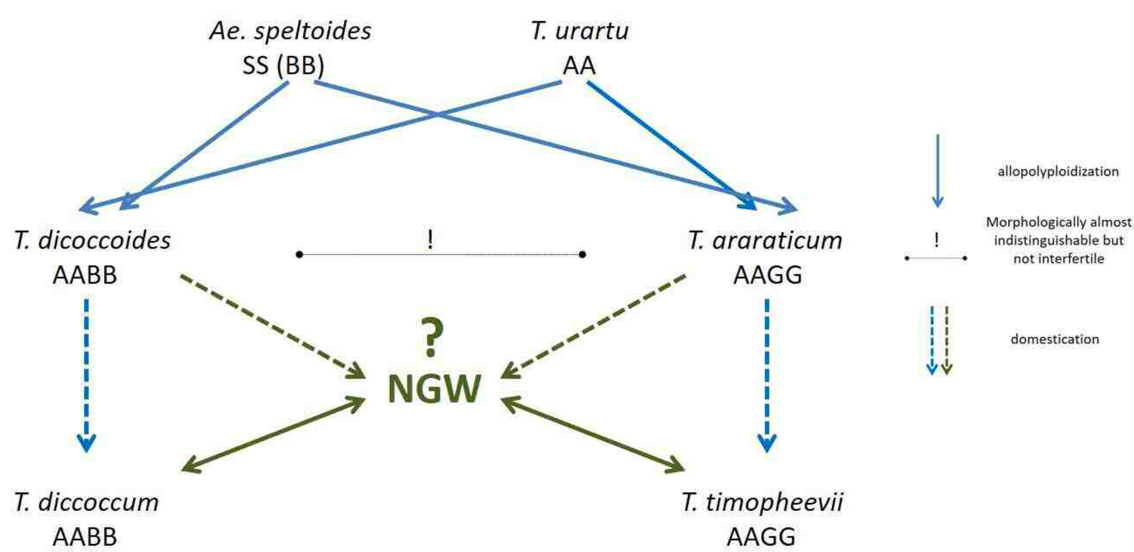


Fig 2. Schematic diagram of *Timopheevii* and *Turgidum* wheat origins and their ancestors with the possible position of the unidentified NGW indicated (adapted from Kilian et al, 2007).

Contribution to field

Work on the origin and spread of domesticated plants is an interactive aspiration of modern molecular genetics, archaeology, geography and history; with the evidence gathered from living plants and archaeological remains being the principal sources. My research has studied two aspects of the origins and development of prehistoric agriculture. The study of extant plants and in particular, the investigation of their main traits that evolved under domestication and resulting genetic differences between the crops and their wild relatives, delivers the principal contribution to our current understanding of domestication. Enabling molecular examination of plant remains retrieved from archaeological excavations and in particular, remains of species that are extinct today, adds an invaluable dimension to the modern synthesis of research on domestication.

The first aim of my research has been to investigate the selective pressures acting on the crop at the beginning of agriculture through the molecular analysis of traits assumed to have been targeted during domestication of barley. This includes identification of the potential genetic targets of selection and functional analysis of genotype-phenotype relationships. My results are presented in the first and the second publications.

Cultivated barley *Hordeum vulgare* L. was domesticated from its wild progenitor, *Hordeum spontaneum* C. Koch and became one of the principal crops that assisted the spread of agriculture into Europe in 8th and 7th millennia BP (Jones et al, 2011). Because of its abiotic stress-tolerance – enabling it to withstand arid conditions and poor soil – as well as some saline tolerance, it has an extensive natural distribution and is widely sampled for seeds. Being one of the main cereals of Mediterranean agriculture and a “founder crop” of Old World Neolithic food production, it is currently the fourth most important cereal crop worldwide (Mascher et al, 2013). Being a true diploid (2n=2x=14 chromosomes, no need to determine which genome is being observed), with a relatively high level of nucleotide sequence diversity (Morell and Clegg, 2007) and predominantly autogamous (like most domesticated seed crops, selfing rate of ~98% but rate can vary across populations, Morell et al, 2005), barley has been proposed as a model for genomic research in the *Triticeae* tribe (Schulte et al, 2009).

Molecular evidence gathered about the domestication traits is used to understand the way these traits are genetically controlled. These traits are regulated by both genetic and epigenetic factors and thus present the underlying complexity of inheritance. The genetic control of traits can be monogenic (e.g. tough rachis), digenic or polygenic (e.g. seed size). A trait presenting a clear dichotomy between the wild and domesticated forms (e.g., brittle versus non-brittle rachis in cereals, hard seeded versus free germinating in legumes) corresponds with mono- or di-genic inheritance. Some of these genes segregate as Mendelian loci whereas some have been discovered as quantitative trait loci, QTLs (Gepts, 2004; Allaby et al, 2008; Abbo et al, 2014). The rise of the initial domestication syndrome was likely to have been primarily driven by the regulatory genes followed by the structural genes that subsequently triggered selection for trait distinctiveness (Brown et al, 2009). The number of identified genes contributing to the domestication syndrome continues to grow, from 7 candidates listed in 2006 (Doebley, 2006) to 52 genes until 2013 (Olsen and Wendel, 2013) for which molecular structure and causative changes have been characterised, assuming these alterations are the product of selection for domestication and not recent crop improvement, discrimination between the two sometimes being subtle and difficult. New phenotypic end products can be produced through changes in the proteome or transcriptome, in coding regions resulting in a protein of altered activity or via changes in *cis*-regulatory sequences resulting in different expression profiling, respectively (Hufford et al, 2012).

Studying the pace and mechanism of the fixation of domestication traits delivers information on how rapidly domestication occurred. There is solid evidence that different elements of the domestication syndrome arose during different stages in the transition from gathering to farming rather than as a "package" in which all the traits arose simultaneously (Fuller, 2007) supporting a protracted, diffused process of transition. The major corollary of the rapid-transition model lies within its biological significance that the artificial selective pressure provided by cultivation practices quickly dominated the natural selective pressures to which the wild crops were subjected. In contrast, a prolonged-transition does not require human agency, as this transition could come about via co-evolutionary events (Fig 3), that is by an even balance between artificial and natural selection pressure mediated by gene flow (Allaby et al, 2008). This selection could have been attained directly by humans or unintentionally as an inevitable secondary result of their plant cultivation. An insight

into the human events that resulted in this transition is obtained by examining the selective pressures that led to fixation of the domestication traits in the crop plants that became domesticated.

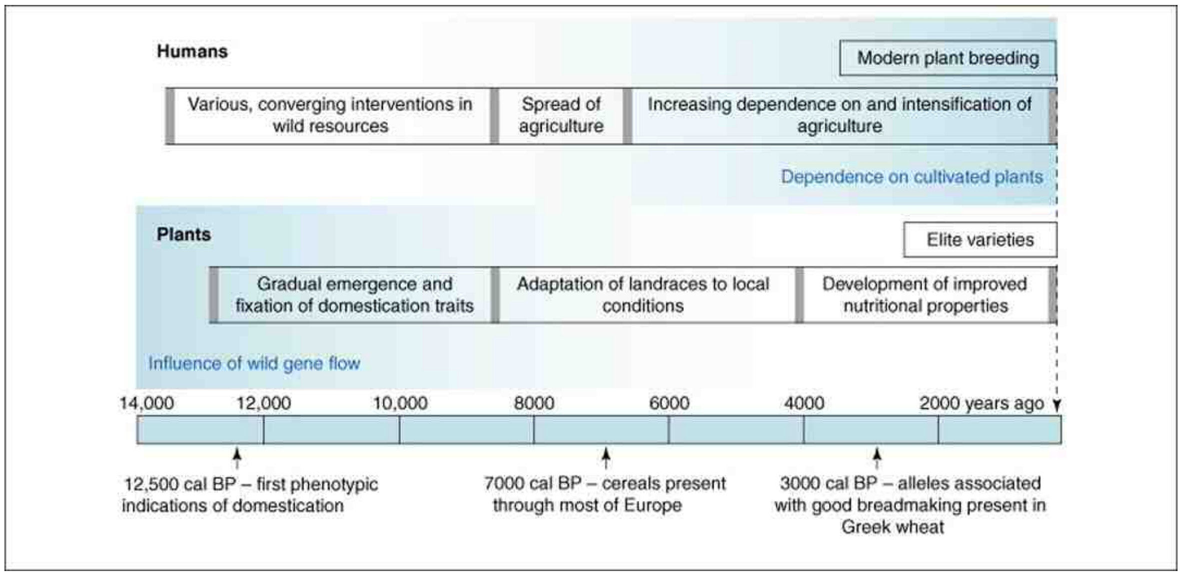


Fig 3. Human-plant interactions in the context of prolonged model of domestication (adopted from Brown et al, 2009).

The classic "domesticated phenotype" includes a set of observable, morphological and physiological, characteristics that distinguish domesticates from their wild progenitors. It has been emphasised, however, that the domestication suite is a vast directional phenotype of developmental, physiological, biochemical and metabolic adaptations among others to the artificial environment produced by our ancestors. Cultivation likely imposed selection for enhanced culinary chemistry, such as improved bread-making quality of wheat and changes to the sugar–starch balance in maize (Brown, 1999; Jaenicke-Després et al, 2003) and for a spectrum of physiological changes (Cunniff et al, 2014; Preece et al, 2017) driven by husbandry practices with water management as one example.

Water availability as the principal factor influencing plant vitality and yield could have prompted early farmers to develop artificial watering and irrigation systems in order to supplement insufficient natural rainfall. It is therefore conceivable that selection for domestication was mediated by changes to the water utilization properties of cultivated plants. The barley cytokinin oxidase/ dehydrogenase 2.1 gene *CKX2.1* is associated with water stress and is controlled epigenetically by the mechanism called gene silencing. Posttranscriptional gene silencing (PTGS) is

mediated by microRNAs (miRNAs) that promote *CKX2.1* transcript degradation. Transcriptional gene silencing (TGS) is driven by the most abundant drought-responsive heterochromatic small interfering RNAs (hc-siRNAs) detected in barley caryopses and is achieved by reduced gene expression in drought stressed plants due to methylation of the *CKX2.1* promoter (Surdonja et al, 2017). To study the nature of the selective pressures leading to an increased adaptation to drought during the domestication of barley, I investigated the diversity of *CKX2.1* in an extensive geo-referenced assemblage of wild barley accessions (*Hordeum vulgare* ssp. *spontaneum*) and historical barley domesticates (*Hordeum vulgare* ssp. *vulgare)*. I found that haplotype 3, being the major haplotype in the wild, is absent from landraces. Amino acid substitutions specific to haplotype 3 are likely to affect the functional properties of the CKX2.1 protein. Furthermore, there is no correlation between haplotype 3 and the geographic distribution of accessions in our wild collection, therefore we believe that domestication was accompanied by the changes in *CKX2.1* locus that might be associated with the artificial irrigation and watering regimes used by early farmers. We suggest that elimination of haplotype 3 was advantageous in a human irrigation environment and therefore water utilization properties should be considered as one of physiological adaptations accompanying domestication. We propose the novel domestication trait, potential underlying candidate gene and its structural changes accompanying selection for domestication.

Reviewer comments:

"The results of the current study agree with large number of studies showing that wild barley accessions (as well as other crop wild relatives), collected in natural populations display non-random genetic diversity which is partly correlated with eco-geographic parameters. For example, a diversity study was described for the dehydrin gene family in wild barley from wide geographical distribution. This gene family is specifically associated with plant dehydration, while the cytokinin dehydrogenase gene is associated with water stress but also to other functions in the plant growth. The molecular methods used this study are well performed and described, this include sequencing of the gene in large number of accessions, sequence analysis and bioinformatics, comparative genomics among other cereals, and protein analysis. The authors assembled large collection of wild barley and landraces from wide geographical rage, and tested some phenotypic parameters."

"I would recommend this group of scientists on their effort to find the genetic evidence for early selection for water utility during barley domestication. The research was well thought and executed, and the findings are original and useful for understanding barley domestication. This is the first time over the last 30 years of my scientific career that I recommend a manuscript accepted as it is after a review".
ALMs: 5 December 2019 (publication date) through 13 December 2019
Total article views: 166
HTML page views (PLoS and PMC): 97
Downloads (PDF and XML): 69
63.92 % of article views led to PDF downloads

It is worth mentioning the only comprehensive paper on *CKX2.1* gene function in barley is Surdonja et al (2017). Most of the proteomic literature uses an incorrect annotation of *CKX2* after Galuszka et al (2004) (e.g. Zalewski et al, 2012) while in fact the subject of their research was *CKX9*. This was clarified by Mameaux et al (2012) in their phylogenetic study.

Seed size together with the loss of natural seed dispersal are the easiest traits to study from archaeological remains (Fuller, 2007). Seed size positively correlates with yield which is the ultimate agronomic trait (Gepts, 2004), therefore selection for increased seed size is an early adaptive response to human cultivation. Selection for seed enlargement has been a matter of debate, whether it is a result of direct human selection for larger seeds or an indirect consequence of selection for traits that are correlated with seed size, for example plant vigour.

WAK1 (wall associated kinase gene) is a member of a gene family specifying wall-associated receptor-like kinases that has specific impact on root growth (Kaur et al, 2013). Root proliferation has been positively correlated with plant biomass and vigorous growth. The root system responds to both environmental and genetic stimuli and can be very complex owing to its principal role in sustaining the plant needs. Root structure is regulated by genetic factors associated with plant anchorage and the features of the species as well the need to perceive environmental cues, such as soil composition; nutrients, water and air availability; light, gravity and other physical factors. Root morphology therefore significantly influences the overall plant growth vigour. *WAK*s play essential roles in plant growth and development and are also involved in various biotic and abiotic stresses and

regulate both cell expansion and response to pathogens (Kohorn, 2015). Comparison between the diversity patterns for the *WAK1* locus in wild and cultivated barley might indicate whether alterations in seed mediated by human agency on the crop at the beginning of agriculture were due to direct selection for seed augmentation or rather indirectly, for more vigorous growth habit. I therefore investigated the diversity of *WAK1* in the same extensive collection of geo-referenced wild barley accessions and historical barley domesticates as used in the *CKX2.1* study. Comparison of the sequence diversity of *WAK1* in wild barley accessions and landraces suggests that domestication resulted in enhancement for particular variants of the *cis*-regulatory region within domesticates, potentially reflecting directional selection for a reduction in *cis*-regulatory variation thus reducing variability in expression profiles. In their review, Olsen and Wendel (2013) report 7 traits being transformed by changes in *cis*-regulatory sequences. Selection for a particular *cis* variant that would increase gene expression level is likely to be advantageous by promoting greater growth vigour. Our results support the hypothesis that increase in seed size is most likely the indirect consequence of selection for more vigorous growth habit; we propose growth vigour as a novel trait, and describe potential functional changes and the underlying genetic basis of the candidate gene.

Reviewer comments:

“The sequencing and analysis experiments are comprehensive. I think the data presented in this paper has been produced in a competent fashion. The analysis does support the conclusion that cis-regulatory elements were selected for during barley domestication, provided the results of several predictive algorithms with respect to protein structure, TATA box location, the putative presence of the HORPIA element and TSSPlant likelihood scores are correct”.

ALMs: 27 June 2019 (publication date) through 13 December 2019
Total article views: 337
HTML page views (PLoS and PMC): 211
Downloads (PDF and XML): 126
54.98 % of article views led to PDF downloads

During the course of the work for the two papers above, I developed a pipeline for the rapid and efficient amplification and sequencing of the *CKX2.1* and *WAK1* genes to complement the laborious and time-consuming but highly accurate Sanger

sequencing. This pipeline enables multiple accessions (444 for *CKX2.1* and 420 for *WAK1*) to be screened in minimum time before being sequenced with a higher degree of base call accuracy and unbiased consensus assemblies accuracy than is possible by any next generation sequencing (NGS; barcoded Illumina sequencing or PacBio). The methodology is the subject of a paper being written (Czajkowska at al, in preparation).

The second aim of my study was to investigate the genetic differences between the *araraticum-timopheevii* (*Timopheevii* wheats) and *dicoccoides-dicoccum* (emmer wheats) groups in order to design a species diagnostic test that allows accurate and fast identification of living plants, inference of the correct evolutionary affinities and which can be also applied to ancient DNA (aDNA) to identify which group the new glume wheat (NGW) belongs to.

Wheat is the one of the most important staple crops and is cultivated all over the world. Today, it accounts for more than 20% of total human food calories (Takenaka and Kawahara, 2013). The genus *Triticum* is an allopolyploid complex and comprises six wheat species at the diploid (AA genomes), tetraploid (AABB and AAGG) and hexaploid (AABBDD and AAAAGG) levels. Hulled AABB tetraploid *Triticum turgidum* L. includes wild, subsp. *dicoccoides* [Korn.ex Asch. & Graebn.] Thell. and cultivated emmer, subsp. *dicoccum* [Schrank ex Schübl.] Thell. The AAGG tetraploid *T. timopheevii* (Zhuk.) Zhuk. is exclusively hulled and made of its wild, subsp. *araraticum* [Jakubz.] Slageren) and also domesticated version, subsp. *timopheevii*. *Triticum* species originated from hybridization events involving the genus *Aegilops*. The diploid wheat *T. urartu* (AA), as the male parent, contributed the A genome to the emmer and *Timopheevii* lineages (Dvořák et al, 1988; Kerby and Kuspira, 1988). *Aegilops speltoides* (SS), as the female parent, contributed cytoplasm (Wang et al, 1997) and the G genome (Kimber, 1974; Dvořák and Zhang, 1992) to *T. timopheevii*. The B nucleus and cytoplasm of the emmer lineage could have been derived from a different genotype of *A. speltoides* (Wang et al, 1997, Kilian et al, 2007; Gornicki et al, 2014).

Timopheevii and emmer wheats have very similar morphology, and the key phenotypic difference is the hairiness and coarseness of the leaf surfaces (Tanaka and Kawahara, 1976; Gornicki et al, 2014). Cytogenetic methods of differentiation based on heterochromatic structure and species-specific cyclic translocations (Gill and Chen, 1987; Jiang and Gill, 1994; Brown-Guedira et al, 1996; Feldman, 1996)

are tremendously costly in terms of time and thus not routinely applied, and are inapplicable to seed remains from archaeological sites preserved by charring (Tanno et al, 2018). Owing to the challenges related to the potential misidentification of the accessions in germplasm collections due to high morphological similarity, robust DNA typing would be of great advantage. The objective was to identify DNA loci that enable the modern species to be distinguished unambiguously.

The most significant amount of resequencing data is available for chloroplast (plastome) and mitochondrial genome (chondriome) but these tests may not be accurate due to the number of nuclear-plasmon hybridization events that occurred in wheat history. Another research group had previously published a paper suggesting that plasmon variation (mitochondrial and chloroplast DNA) might be used to distinguish the emmer and *Timopheevii* wheat groups (Gornicki et al, 2014; Tanno et al, 2018). My additional study of plasmon variation led me to discover that it may only be possible to establish the prevalence of major plasmon haplotypes, leaving *Triticum* species in question identification unresolved (Czajkowska et al, in preparation). Limited ability of plasmon based typing is further indicated by:

- Nuclear and cytoplasmically inherited markers yielded contrasting results (Wang et al, 1997; Kilian et al, 2007)
- The chloroplast NADH dehydrogenase F (*ndhF)* data assign *A. speltoides* cytoplasm to both *T. dicoccoides* and *T. araraticum* (Kilian et al, 2007)
- Full-length ribulose-1,5 bisphosphate-carboxylase large subunit (*rbcL*) and full-length hypothetical chloroplast open reading frame 1 (*ycf1*) sequence data failed to discriminate between *Timopheevii* wheats and *T. dicoccoides* (Awad et al, 2017).

An alternative and more robust approach would account for nuclear gene differences. There is already existing information and a screening method based on the Hardness locus (*Ha*) that determines grain texture in wheat. It consists of the genes Puroindoline a (*Pina*) and Puroindoline b (*Pinb*). Functional deletion has been reported; *Pina* and *Pinb* were deleted from the A and B genomes of *T. turgidum* and also eliminated from the G genome of *T. timopheevii* but maintained in the A genome of *T. timopheevii* (Morris, 2002; Li et al, 2008). A *Pinb-A* assay, therefore, gives a null signal for *Turgidum* wheat and positive for *Timopheevii* with an amplicon size of 484bp. Such a design has the weakness that a null result cannot be distinguished from PCR failure, and therefore limits the robustness of *Turgidum* screening; such a test is unsuitable for identification of archaeological remains where PCR failures

are common due to the poor aDNA preservation and the amplicon size being beyond the expectation for aDNA. Therefore, I focussed my attention on B and G genome-specific nuclear gene variants, for which sequence data for *Turgidum* and for *Timopheevii* wheat exist (Table 2). The first locus I examined on was the High Molecular Weight (HMW) glutenin promoter. Initial tests proved the reported changes indeed display species-specific variability (Allaby et al, 1999). However, this test would need to use amplified DNA fragments, subsequently digested with a restriction enzyme and gel electrophoresed, an assay called Cleaved Amplified Polymorphic Sequences (CAPS). An optimal genotyping method should not only be accurate but also fast, inexpensive and of little effort.

Gene	Symbol	Chromosome	Function	References
Photoperiod response gene 1/ pseudo-response regulator 1 (*PRP1*)	*Ppd-1*	2	Photoperiod response	Takenaka and Kawahara, 2012; Takenaka and Kawahara, 2013
No apical meristem (*NAC* transcription factor gene)	*NAM*	6	Grain protein content (GPC)	Hu et al, 2013
Acetyl-CoA carboxylase 1	*ACC1*	3	Fatty acid biosynthesis	Faris et al, 2001; Kilian et al, 2007
Acetyl-CoA carboxylase 2	*ACC2*	5	Fatty acid biosynthesis	Faris et al, 2001
Dihydroflavonol-4-reductase	*DFR*	3	Flavonoid metabolism	Sequence data are available from GenBank
High-molecular-weight glutenin promoter	*GLUT*	1	Seed storage proteins	Allaby et al, 1999
5´ external transcribed spacers of the 18S ribosomal DNA genes	*18S rDNA ETS*	Multi-copy	Ribosome synthesis	Sallares and Brown, 2004
Glucose-6-phosphate dehydrogenase	*G6PDH*	Not mapped	PPC-pathway	Kilian et al, 2007
Glucose-6-phosphate/phosphate translocator	*GPT*	Not mapped	Starch biosynthesis	Kilian et al, 2007
3-phosphoglycerate kinase	*PGK1*	Not mapped	Calvin cycle	Kilian et al, 2007
Q-Locus	*Q*	5	Free threshing	Kilian et al, 2007
TmAP1-gene	*VRN1*	5	Vernalisation	Kilian et al, 2007

Table 2. B and G genome-specific nuclear gene variants.

The photoperiod response gene 1 (*Ppd-1*), for which the greatest amount of resequencing data exists (Takenaka and Kawahara, 2012; Takenaka and Kawahara, 2013), controls the time of flowering, which is one of the most important factors of crop adaptation. In emmer and tetraploid *Timopheevii* lineages, the major *Ppd-1* gene is present as two homeologous copies located on the short arm of chromosome 2, the *Ppd-A1* and *Ppd-B1* genes in emmer and *Ppd-A1* and *Ppd-G1* genes in *Timopheevii*. From examination of published sequences I discovered that

there are unambiguous differences within the *Ppd-B1* and *Ppd-G1* sequences (Fig 4) that can be used to distinguish between emmer and *Timopheevii* wheats. I designed a series of PCRs, each giving length polymorphisms for the B and G genomes and tested these with a panel of emmer and *Timopheevii* wheats. This revealed that several accessions had been misidentified by the germplasm collections.

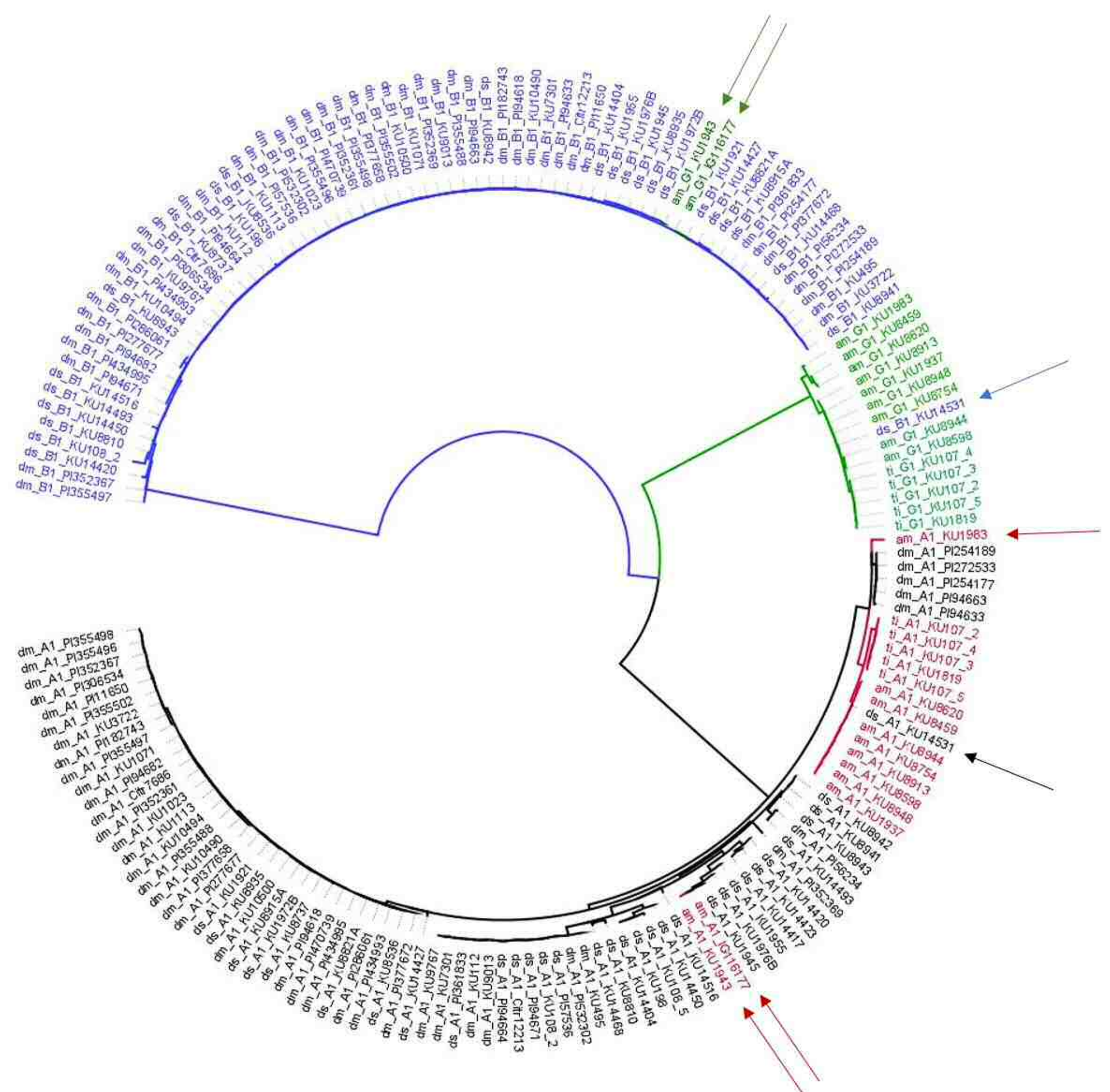


Fig 4. Maximum likelihood (ML) tree inferred from the sequence databases for a 3522 bp region of *Ppd-1* (3076-3463bp), displaying genome specific variability between the A, B and G genomes. Two *T. araraticum* accessions, KU1943 and IG116177, and one *T. dicoccoides,* accession, KU14531, are shown to be misidentified. *Ppd-B1* of *T. dicoccum* (dm) and *T. dicoccides* (ds) – blue, *Ppd-G1* of *T. timopheevii* (ti) and *T. araraticum* (am) – green, *Ppd-A1* of *T. dicoccum* (dm) and *T. dicoccides* (ds) – black, *Ppd-A1* of *T. timopheevii* (ti) and *T. araraticum* (am) – red. Smart Model Selection (SMS) was used to determine the evolutionary model

that best fit the input sequences, GTR+G. PhyML online execution (v.3.0; Guindon at al, 2010) was explored to search the tree space deterministically using a Subtree Pruning and Regrafting (SPR) algorithm. SPR topological moves started from 10 random trees and BioNJ tree. Support for inferred relationships was evaluated by nonparametric branch support measure, nonparametric version of Approximate Likelihood-Ratio (aLRT), Shimodaira-Hasegawa-like (SHlike) procedure, (aLRTSHlike). Computed inference was viewed and analysed in FigTree (v. 1.4.3) (http://tree.bio.ed.ac.uk/software/figtree/).

Reviews considered it a useful technique for identification/classification of accessions of *T. timopheevii* and *T. turgidum* supported by genotyping by sequencing (GBS) analysis.

ALMs: 10 April 2019 (publication date) through 13 December 2019
Total article views: 549
Page views (PLoS and PMC): 292
Downloads (PDF and XML): 257
53.42 % of article views led to PDF downloads

I also used the *Ppd-1* method to attempt to identify the extinct today, so called New Glume Wheat from charred specimens (Czajkowska et al, in preparation). This is the first attempt at molecular classification of the NGW, the previously reported analysis of the *rbcL* gene yielded positive results from aDNA from NGW samples from Stillfried but did not allow for differentiation of *Triticum* species (Blatter et al, 2002). Work with ancient samples is not commonly undertaken due to the seemingly high risk of obtaining no results.

Analysis of DNA retrieved from charred archaeological samples is the most challenging procedure (Giles and Brown, 2008), as the samples contain no or little DNA, which is very fragmented, degraded and modified, and often contains absorbed environmental DNA. Therefore, an effort must be made towards designing a dedicated extraction method that would allow extremely short DNA to be obtained and if possible, an amplification must make use of very specific primers. Nucleic acids (NAs) were extracted with NucleoSpin totalRNA FFPE (Macherey-Nagel, 740982) according to an optimised protocol that I designed. The method is based on the silica-binding procedure because this allows more effective washing than is

possible with a DNA pellet, and the loss of NAs is much lower. This procedure enables recovery of all NAs, both double and single stranded, down to 18 nucleotides with high efficiency. Most of the commonly used DNA extraction kits recover only double stranded DNA (dsDNA) with a lower cut off of 100bp. Such a design works well for modern DNA as even though slightly less DNA is isolated, contaminants that would be recovered with a lower cut off are removed. This is not applicable for aDNA from charred material since it is only obtained as fragments of 100bp maximum. PCR purification kits feature lower cut off (currently down to 50bp) but these kits are very basic and designed to bind only DNA which is already released in the solution. The lysis step that is included in extraction kits is necessary to bring all existing NA molecules into the solution. My protocol greatly enhances the likelihood of retrieving all the DNA molecules in a sample unless they are unrecoverable because they have become irreversibly and covalently modified by the Maillard reaction or by caramelization during the course of charring. The procedure is quick and consists of a minimal number of steps to reduce the likelihood of contamination.

Sample	Location	Sample ID	No of grains	No of chaff fragments
1	Feudvar, Serbia	W3063	100	
2	Çatalhöyük, southern Anatolia, Turkey	FI# 10830, unit 20703	30	
3	Çatalhöyük, southern Anatolia, Turkey	FI# 11995, unit 22656	30	
4	Çatalhöyük, southern Anatolia, Turkey	FI# 12015, unit 22637	25	
5	Assiros, Greece	4355	30	
6	Stillfried, eastern Austria	ST 11392	30	
7	Assiros, Greece	4355		25
8	Stillfried, eastern Austria	ST 11392		25
9	Miechowice, Poland	M4, pit 30		25

Table 3. NGW samples used in this study. Five grains or twenty five chaff fragments were used per single extraction. For the grain samples, six extractions (and also seven for the Feudvar specimen) from the same location were further pooled together prior to amplification.

Having devised an efficient aDNA extraction method, I applied the *Ppd-B1* and *Ppd-G1* based PCR tests to samples of the NGW (Table 3), using extracts from charred grains as well charred chaff fragments (single or paired glume bases). All

amplicons that were obtained were cloned and sequenced by the Sanger method. In total, I screened 32 *Ppd-G1* clones and 36 *Ppd-B1* clones. PCR 1 (for *Ppd-B1*) gave no products during amplification. I obtained ten *Ppd-G1* reads, nine with PCR 3 and one with PCR 4. Five of the PCR 3 reads were retrieved for the charred seed sample from Çatalhöyük, southern Anatolia, Turkey (FI# 12015, unit 22637) and four PCR 3 reads were obtained from the Neolithic charred chaff sample from Miechowice, Poland. One PCR 4 read was obtained from the Miechowice sample and also for this sample two PCR 2 (for *Ppd-B1*) reads were identified. Therefore, the Miechowice sample delivered both *Ppd-G1* and *Ppd-B1* reads while the Çatalhöyük sample returned only *Ppd-G1* reads. It is interesting to mention that only ssDNA was detectable for the Miechowice sample and it is also the very first successful attempt to retrieve DNA from chaff. I have not identified any sample containing solely *Ppd-B1* reads. The weight of evidence therefore suggests that the NGW contains a G genome and hence is a member of the *araraticum-timopheevii* group.

Demonstration that the NGW is a member of the *araraticum-timopheevii* group, would have a huge significance, not just because it would force a reappraisal of the composition of the package of crops that drove the establishment of prehistoric agriculture in western Asia and Europe. The discovery would also have more general implications for our understanding of the origin of agriculture in southwest Asia. *Triticum araraticum* has never been considered an important crop progenitor, because *T. timopheevii*, the only domesticate at present known to be derived from *T. araraticum*, is confined to a small part of eastern Georgia, and it is looked on as a “local episode in wheat-crop evolution” (Zohary et al, 2012). If the NGW, which was widely distributed in western Asia and Europe, is shown to be a third member of the *araraticum-timopheevii* group, then the role of *T. araraticum* in the origins of agriculture has to be re-evaluated.

Future Work

I have reported two phenotypic features, water utilisation and growth vigour, in the context of selection during domestication, and propose that both are important components of the domestication suite. Extreme changes in diversity in landraces compared to wild accessions – the absence of haplotype 3 in *CKX2.1* and the fixation of the promoter SNPs in *WAK1* – highlight the possibility that these genes

evolved under selection for domestication and are the candidates for genes involved in controlling these phenotypic traits. I presented potential causes and possible effects of these traits, however much more work needs to be done in order to confirm that the reported molecular signatures result in the proposed effect on the traits and so are the result of human-driven selection for domestication.

The observed reduction of wild diversity within the coding regions of the *CKX2.1* gene and the extensive reduction of wild diversity within the promoter of the *WAK1* gene could possibly be ascribed to, although unpragmatically, the result of a domestication bottleneck (Meyer and Purugganan, 2013). A genetic bottleneck is defined as the outcome of strong and rapid selective pressure imposed on a desired trait; in the case of the domestication bottleneck this pressure is exclusively a human driven action. The domestication bottleneck is expected to operate at the multilocus level and to result in a drastic reduction of genetic diversity in modern crops throughout the genome. During the bottleneck, direct selection for the desired, adaptive trait, will also indirectly select for other traits that are present in the original genepool taken into cultivation. As such, the population becomes enriched in selected adaptive traits but also is marked by an abundance of non-adaptive by-products. Rather than performing whole genome analysis on our barley assemblage, one possibility would be to re-sequence one or more neutral genes in our panel of accessions to confirm that the changes reported in the *CKX2.1* and *WAK1* genes are not the outcomes of a domestication bottleneck, and are in fact more likely to be the outcome of selection imposed by our ancestors. However, virtually any gene could be involved in the mechanism of domestication, hence choosing an appropriate candidate is almost an impossible task. Insofar, there is only a one report of a gene shown to behave neutrally, *G3PDH* that encodes glyceraldehyde-3-phosphate dehydrogenase, a relatively stable enzyme involved in the second phase of glycolysis (Caldwell et al, 2006). Noticeably, these authors used only 23 landraces and 34 wild barley accessions; as such, re-sequencing more individuals may prove the gene is not neutral and indeed was involved in the domestication process. The notion of a genetic bottleneck during domestication is in fact being questioned, and is being replaced by the concept of a more gradual loss of diversity that occurred after domestication. This is because the claims for occurrence of a bottleneck are not supported by specially designed studies on ancient genomes, nor by computer simulations that show no evidence for the

genetic bottleneck and advocate a steady decline of diversity as a function of time (Allaby et al, 2019; Brown, 2019; Smith et al, 2019).

To demonstrate the adaptive nature of the features we report, detailed transcriptome profiling analysis would be necessary in order to investigate the correlation between particular *cis* variants of the *WAK1* gene and their effect on expression profiles and subsequently, a direct effect on growth vigour. Unfortunately, no study has been reported that enables the gene expression profiles in wild accessions and landraces to be compared. A proteomic approach would be needed to study the activity of different *CKX2.1* products. It would also be necessary to perform a number of complex physiological tests to investigate the strength of association between particular coding variants of *CKX2.1* and their effect on water utilization as well as particular *cis* variants of *WAK1* and their effect on plant vigour.

I have also studied genetic differences between two morphologically similar group of wheat species. I provided the first discriminatory test for the B and G genomes and I hope it will be used by other researchers and germplasm collections in order to subsequently produce accurate evolutionary trajectories. I also report the first attempts at the molecular identification of the extinct and mysterious NGW. Confirmation of this result with additional samples of the NGW would stimulate new research on the nature of the Neolithic crop package and the pattern of domestication in the Fertile Crescent.

References

- Abbo S, Lev-Yadun S, Heun M, Gopher A (2013) *On the 'lost' crops of the neolithic Near East*. J Exp Bot 64:815-822.

- Abbo S, Pinhasi van-Oss R, Gopher A, Saranga Y, Ofner R, Peleg Z (2014) *Plant domestication versus crop evolution: a conceptual framework for cereals and grain legumes*. Trends Plant Sci 19:351-360.

- Allaby RG, Banerjee M, Brown TA (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, Fuller DQ, Brown TA (2008) *The genetic expectations of a protracted model for the origins of domesticated crops*. Proc Natl Acad Sci USA 105:13982-13986.

- Allaby R (2010) *Integrating the processes in the evolutionary system of domestication.* J Exp Bot 61:935-944.

- Allaby RG, Ware RL, Kistler L (2019) *A re-evaluation of the domestication bottleneck from archaeogenomic evidence*. Evol Appl 12:29-37.

- Awad M, Fahmy RM, Mosa KA, Helmy M, El-Feky FA (2017) *Identification of effective DNA barcodes for Triticum plants through chloroplast genome-wide analysis*. Comput Biol Chem 71:20-31.

- Badr A, Müller K, Schäfer-Pregl R, El Rabey H, Effgen S, Ibrahim HH, Pozzi C, Rohde W, Salamini F (2000) *On the origin and domestication history of barley*. Mol Biol Evol 17:499-510.

- Barker G (2006) *The agricultural revolution in prehistory. Why did Foragers Become Farmers*? Oxford: Oxford University Press.

- Blatter RHE, Jacomet S, Schlumbaum A (2002) *Little evidence for the preservation of a single-copy gene in charred archaeological wheat*. Anc Biomol 4:65-77.

- Blumler MA (1992) *Independent inventionism and recent genetic evidence on plant domestication*. Econ Bot 46:98-111.

- Brown TA (1999) *How ancient DNA may help in understanding the origin and spread of agriculture*. Philos Trans R Soc Lond B Biol Sci 354:89-98.

- Brown TA, Jones MK, Powell W, Allaby RG (2009) *The complex origins of domesticated crops in the Fertile Crescent*. Trends Ecol Evol 24:103-109.

- Brown TA (2018) *The role of humans in a protracted transition from hunting-gathering to plant domestication in the Fertile Crescent*. Front Plant Sci 9:1287.

- Brown TA (2019) *Is the domestication bottleneck a myth?* Nat Plants 5:337-338.

- Brown-Guedira GL, Badaeva ED, Gill BS, Cox TS (1996) *Chromosome substitutions of Triticum timopheevii in common wheat and some observations on the evolution of polyploid wheat species*. Theor Appl Genet 93:1291-1298.

- Buckler ES, Holland JB, Bradbury PJ, Acharya CB, Brown PJ, Browne C, Ersoz E, Flint-Garcia S, Garcia A, Glaubitz JC, Goodman MM, et al (2009) *The genetic architecture of maize flowering time*. Science 325:714-718.

- Caldwell KS, Russell J, Langridge P, Powell W (2006) *Extreme Population-Dependent Linkage Disequilibrium Detected in an Inbreeding Plant Species, Hordeum vulgare*. Genetics 172:557-567.

- Childe VG (1928) *The most ancient East: the oriental prelude to European prehistory.* London: Kegan Paul.

- Cunniff J, Wilkinson S, Charles M, Jones G, Rees M, Osborne CP (2014) *Functional traits differ between cereal crop progenitors and other wild grasses gathered during the Neolithic in southwest Asia*. PLoS One 9(1):e87586.

- Diamond J (1997) *Location, location, location: the first farmers*. Science 278:1243-1244.

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

- Dvořák J, McGuire PE, Cassidy B (1988) *Apparent sources of the A genomes of wheats inferred from polymorphism in abundance and restriction fragment length of repeated nucleotide sequences*. Genome 30:680-689.

- Dvořák J, Zhang H-B (1992) *Reconstruction of the phylogeny of the genus Triticum from variation in repeated nucleotide sequences*. Theor Appl Genet 84:419-429.

- Doebley JF, Gaut BS, Smith BD (2006) *The molecular genetics of crop domestication*. Cell 127:1309-1321.

- Faris J, Sirikhachornkit A, Haselkorn R, Gill B, Gornicki P (2001) *Chromosome mapping and phylogenetic analysis of the cytosolic acetyl-CoA carboxylase loci in wheat*. Mol Biol Evol 18:1720-1733.

- Feldman M (1966) *The mechanism regulating pairing in Triticum timopheevii*. Wheat Inf Serv 2:1-2.

- FigTree software page [last accessed 29 July 2019].
  Available from: http://tree.bio.ed.ac.uk/software/figtree/.

- Fuller DQ (2007) *Contrasting patterns in crop domestication and domestication rates: Recent archaeological insights from the Old World*. Ann Bot 100:903-924.

- Galuszka P, Frébortová J, Werner T, Yamada M, Strnad M, Schmülling T, Frébort I (2004) *Cytokinin oxidase/dehydrogenase genes in barley and wheat cloning and heterologous expression.* Eur J Biochem 271:3990-4002.

- Gebel HGK (2004) *There was no centre: the polycentric evolution of the Near Eastern Neolithic*. Neo-Lithics 1:28-32.

- Gepts P (2004) *Crop domestication as a long-term selection experiment*. Plant Breed Rev 24:1-44.

- Giles RJ, Brown TA (2008) *Improved methodology for extraction and amplification of DNA from single grains of charred wheat*. J Archaeol Sci 35:2585-2588.

- Gill BS, Chen PD (1987) *Role of cytoplasmic specific introgression in the evolution of the polyploid wheats*. Proc Natl Acad Sci USA 84:6800-6804.

- Grosman L, Belfer-Cohen A (2002) *Zooming on to the Younger Dryas*. In: Cappers RTJ, Bottema S, editors. *The dawn of farming in the Near East*. *Studies in early Near Eastern production, subsistence, and environment*. Berlin: Ex oriente pp 49-54.

- 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 Phytol. 204:704-714.

- Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O (2010) *New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0*. Syst Biol 59:307-321.

- Hammer K (1984) *Das Domestikationssyndrom.* Kulturpflanze 32:11-34.

- Harris DR (1989) *An evolutionary continuum of people-plant interaction*. In: Harris DR, Hillman GC, editors. *Foraging and farming: the evolution of plant exploitation*. Abingdon: Routledge Library Edition: Archaeology pp 11-26.

- Harlan JR (1992) *Crops and Man.* 2nd ed. Madison: American Society of Agronomy and Crop Science Society of America.

- Hu X-G, Wu B-H, Liu D-C, Wei Y-M, Gao S-B, Zheng Y-L (2013) *Variation and their relationship of* NAM-G1 *gene and grain protein content in Triticum timopheevii Zhuk*. J Plant Physiol 170:330-337.

- Heun M, Schäfer-Pregl R, Klawan D, Castagna R, Accerbi M, Borghi B, Salamini F (1997) *Site of einkorn wheat domestication identified by DNA fingerprinting*. *Science* 278:1312-1314.

- Hufford MB, Xu X, van Heerwaarden J, Pyhäjärvi T, Chia JM, Cartwright RA, Elshire RJ, Glaubitz JC, Guill KE, Kaeppler SM, Lai J, et al (2012) *Comparative*

population genomics of maize domestication and improvement. Nat Genet 44:808-811.

- 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 302:1206-1208.

- Jiang J, Gill BS (1994) *Different species-specific translocations in T. timopheevii and T. turgidum support the diphyletic origin of polyploid wheats*. Chromosome Res 2:59–64.

- Jones G, Valamoti S, Charles M (2000) *Early crop diversity: a "new" glume wheat from northern Greece*. Veg Hist Archaeobot 9:133-146.

- Jones H, Leigh FJ, Mackay I, Bower MA, Smith LMJ, Charles MP, Jones G, Jones MK, Brown TA, Powell W (2008) *Population-based resequencing reveals that the flowering time adaptation of cultivated barley originated east of the Fertile Crescent*. Mol Biol Evol 25:2211-2219.

- Jones H, Civáň P, Cockram J, Leigh FJ, Smith LMJ, Jones MK, Charles MP, Molina-Cano JL, Powell W, Brown TA (2011) *Evolutionary history of barley cultivation in Europe revealed by genetic analysis of extant landraces*. BMC Evol Biol 11:320.

- Kaplan JO, Krumhardt KM, Ellis EC, Ruddiman WF, Lemmen C, Goldewijk KK (2010) *Holocene carbon emissions as a result of anthropogenic land cover change*. Holocene 21:775-791.

- Kaur R, Singh K, Singh J (2013) *A root-specific wall-associated kinase gene, HvWAK1, regulates root growth and is highly divergent in barley and other cereals*. Funct Integr Genomics 13:167-177.

- Kenéz A, Pető A, Gyulai F (2014) *Evidence of 'new glume wheat' from the Late Neolithic (Copper Age) of south-eastern Hungary (4th millennium cal. B.C.)* Veg Hist Archaeobot 23:551-566.

- Kerby K, Kuspira J (1988) *Cytological evidence bearing on the origin of the B genome in polyploid wheats*. Genome 30:36-43.

- Kilian B, Özkan H, Deusch O, Effgen S, Brandolini A, Kohl J, Martin W, Salamini F (2007) *Independent Wheat B and G Genome Origins in Outcrossing Aegilops Progenitor Haplotypes.* Mol Biol Evol 24:217-227.

- Kimber G (1974) *A reassessment of the origin of the polyploid wheats*. Genetics 78:487-492.

- Kohler-Schneider M (2003) *Contents of a storage pit from late Bronze Age Stillfried, Austria: another record of the "new"glume wheat*. Veg Hist Archaeobot 12:105-111.

- Kohorn BD (2015) *Cell wall-associated kinases and pectin perception.* J Exp Bot 67:489-494.

- Kutzbach JE, Ruddiman WF, Vavrus SJ, Philippon G (2010) *Climate model simulation of anthropogenic influence on greenhouse-induced climate change (early agriculture to modern): the role of ocean feedbacks*. Clim Change 99:351-381.

- Lemmen C, Gronenborn D, Wirtz KW (2011) *A simulation of the Neolithic transition in Western Eurasia.* J Archaeol Sci 38:3459-3470.

- Li W, Huang L, Gill BS (2008) *Recurrent deletions of puroindoline genes at the grain hardness locus in four independent lineages of polyploid wheat*. Plant Physiol 146:200-212.

- Liu B, Segal G, Rong JK, Feldman M (2003) *A chromosome-specific sequence common to the B genome of polyploid wheat and Aegilops searsii*. Plant Syst Evol 241:55-66.

- Mameaux S, Cocram J, Thiel T, Steuernagel B, Stein N, Taudien S, Jack P, Werner P, Gray JC, Greenland AJ, Powell W (2012) *Molecular, phylogenetic and comparative genomic analysis of the cytokinin oxidase⁄dehydrogenase gene family in the Poaceae*. Plant Biotechnol J 10:67-82.

- Mascher M, Richmond TA, Gerhardt DJ, Himmelbach A, Clissold L, Sampath D, Ayling S, Steuernagel B, Pfeifer M, D'Ascenzo M, Akhunov ED, et al (2013) *Barley whole exome capture: a tool for genomic research in the genus Hordeum and beyond*. Plant J 76:494–505.

- Meyer RS, Purugganan MD (2013) *Evolution of crop species: genetics of domestication and diversification*. Nat Rev Genet 14:840-852.

- Molina-Cano JL, Russell JR, Moralejo MA, Escacena JL, Arias G, Powell W (2005) *Chloroplast DNA microsatellite analysis supports a polyphyletic origin for barley.* Theor Appl Genet 110:613-619.

- Morrell PL, Toleno DM, Lundy KE, Clegg MT (2005) *Low levels of linkage disequilibrium in wild barley (Hordeum vulgare ssp. spontaneum) despite high rates of self fertilization*. Proc Natl Acad Sci USA 102:2442-2447.

- Morrell PL, Clegg MT (2007) *Genetic evidence for a second domestication of barley (Hordeum vulgare) east of the Fertile Crescent*. Proc Natl Acad Sci USA 104:3289-3294.

- Morrell PL, Gonzales AM, Meyer KKT, Clegg MT (2014) *Resequencing data indicate a modest effect of domestication on diversity in barley: a cultigen with multiple origins.* J Hered 105:253-264.

- Morris CF (2002) *Puroindolines: the molecular genetic basis of wheat grain hardness*. Plant Mol Biol 48:633-647.

- Olsen KM, Wendel JF (2013) *Crop plants as models for understanding plant adaptation and diversification*. Front. Plant Sci 4:290.

- Özkan H, Brandolini A, Schäffer-Pregl R, Salamini F (2002) *AFLP analysis of a collection of tetraploid wheat indicates the origin of emmer and hard wheat in Southeast Turkey*. Mol Biol Evol 19:1797-1801.

- PLoS ONE journal page [last accessed 29 July 2019]. Available from: https://journals.plos.org/plosone/s/journal-information.

- Pourkheirandish M, Hensel G, Kilian B, Senthil N, Chen G, Sameri M, Azhaguvel P, Sakuma S, Dhanagond S, Sharma R, Mascher M, et al (2015) *Evolution of the grain dispersal system of barley*. Cell 162:527-539.

- Preece C, Livarda A, Christin P-A, Wallace M, Martin G, Charles M, Jones G, Rees M, Osborne CP (2017) *How did domestication of Fertile Crescent grain crops increase their yields?* Funct Ecol 31:387-397.

- Ramsay L, Comadran J, Druka A, Marshall DF, Thomas WTB, Macaulay M, MacKenzie K, Simpson C, Fuller J, Bonar N, Hayes PM, et al (2011) *INTERMEDIUM-C, a modifier of lateral spikelet fertility in barley, is an ortholog of the maize domestication gene TEOSINTE BRANCHED 1*. Nat.Genet 43:169-172.

- Ruddiman WF (2007) *The early anthropogenic hypothesis: challenges and responses.* Rev of Geophys 45:RG4001:1-37.

- Saisho D, Purugganan MD (2007) *Molecular Phylogeography of Domesticated Barley Traces Expansion of Agriculture in the Old World*. Genetics 177:1765-1776.

- Sallares R, Brown TA (2004) *Phylogenetic analysis of complete 5´ external transcribed spacers of the 18S ribosomal RNA genes of diploid Aegilops and related species (Triticeae, Poaceae)*. Genet Resour Crop Evol 51: 701-712.

- Schulte D, Close TJ, Graner A, Langridge P, Matsumoto T, Muehlbauer G, Sato K, Schulman AH, Waugh R, Wise RP, Stein N (2009) *The international barley*

sequencing consortium – at the threshold of efficient access to the barley genome. Plant Physiol 149:142-147.

- Smith O, Nicholson WV, Kistler L, Mace E, Clapham A, Rose P, Stevens C, Ware R, Samavedam S, Barker G, David Jordan D, et al (2019) *A domestication history of dynamic adaptation and genomic deterioration in Sorghum*. Nat Plants 5:369-379.

- Surdonja K, Eggert K, Hajirezaei M-R, Harshavardhan VT, Seiler C, von Wirén N, Sreenivasulu N, Kuhlmann M (2017) *Increase of DNA methylation at the HvCKX2.1 promoter by terminal drought stress in barley*. Epigenomes 1:9.

- Takenaka S, Kawahara T (2012) *Evolution and dispersal of emmer wheat (Triticum sp.) from novel haplotypes of Ppd-1 (photoperiod response) genes and their surrounding DNA sequences*. Theor Appl Genet 125:999-1104.

- Takenaka S, Kawahara T (2013) *Evolution of tetraploid wheat based on variations in 5´ UTR regions of Ppd-A1: evidence of gene flow between emmer and Timopheevi wheat*. Genet Resour Crop Evol 60:2143-2155.

- Tanaka M, Kawahara T (1976) *Wild tetraploid wheats from northern Iraq cytogenetically closely related to each other*. Wheat Inf Serv 43:3-4.

- Tanno K-I, Willcox G (2006) *How fast was wild wheat domesticated?* Science 311:1886.

- Tanno K-I, Takeuchi A, Akahori E, Kobayashi K, Kawahara T, Yamane K (2018) *Multiplex PCR effectively identifies tetraploid Triticum AABB–or AAGG–genome species*. Plant Genet Resour 16:279-283.

- Wang G-Z, Miyashita NT, Tsunewaki K (1997) *Plasmon analyses of Triticum (wheat) and Aegilops: PCR-single-strand conformational polymorphism (PCR-SSCP) analyses of organellar DNAs*. Proc Natl Acad Sci USA 94:14570-14577.

- Willcox G (2004) *Measuring grain size and identifying Near Eastern cereal domestication: evidence from the Euphrates valley*. J Archaeol Sci 31:145-150.

- Willcox G (2005) *The distribution, natural habitats and availability of wild cereals in relation to their domestication in the Near East: multiple events, multiple centres*. Veg Hist Archaeobot 14:534–541.

- Willcox G, Fornite S, Herveux L (2008) *Early Holocene cultivation before domestication in northern Syria*. Veget Hist Archaeobot 17:313-325.

- Zalewski W, Orczyk W, Gasparis S, Nadolska-Orczyk A (2012) *HvCKX2 gene silencing by biolistic or Agrobacterium-mediated transformation in barley leads to different phenotypes*. BMC Plant Biol 12:206.

- Zohary D (1996) *The mode of domestication of the founder crops of Southwest Asian agriculture*. In: Harris DR, editor. *The origins and spread of agriculture and pastoralism in Eurasia*. London: UCL Press pp 142-158.

- Zohary D (1999) *Monophyletic vs. polyphyletic origin of crops on which agriculture was founded in the Near East*. Genet Resour Crop Evol 46:133-142.

- Zohary D, Hopf M, Weiss E (2012) *Domestication of plants in the Old World*. 4th ed. Oxford: Oxford University Press.

Acknowledgments

I would like to sincerely thank my supervisor Professor Terry Brown for not only being an exceptional scientific supervisor, but also an extraordinary mentor who gave me all the encouragement, motivation and appreciation over the past six years. This has made my PhD journey both possible and enjoyable experience. I am aware how extremely fortunate I am, and I will be forever grateful for the wonderful opportunity he gave me.

I express my thanks to my co-authors, especially Professor Glynis Jones, for amazing contribution to my research papers.

Chciałabym również bardzo podziękować Mojej Rodzinie za cierpliwość, zrozumienie i wiarę we mnie.
(I would also like to say thanks to My Family for their patience, understanding and believing in me).

Copies of each of the publications

1.

Diversity of a cytokinin dehydrogenase gene in wild and cultivated barley.

Czajkowska BI, Finlay CM, Jones G, Brown TA

Published on 5 December 2019

in

PLoS ONE

14(12): e0225899

https://doi.org/10.1371/journal.pone.0225899

RESEARCH ARTICLE

Diversity of a cytokinin dehydrogenase gene in wild and cultivated barley

Beata I. Czajkowska[1], **Conor M. Finlay**[2], **Glynis Jones**[3], **Terence A. Brown**[1]*

1 Department of Earth and Environmental Sciences, Manchester Institute of Biotechnology, University of Manchester, Manchester, England, United Kingdom, **2** Lydia Becker Institute of Immunology and Inflammation, School of Biological Sciences, University of Manchester, Manchester, England, United Kingdom, **3** Department of Archaeology, University of Sheffield, Northgate House, West Street, Sheffield, England, United Kingdom

* terry.brown@manchester.ac.uk

OPEN ACCESS

Citation: Czajkowska BI, Finlay CM, Jones G, Brown TA (2019) Diversity of a cytokinin dehydrogenase gene in wild and cultivated barley. PLoS ONE 14(12): e0225899. https://doi.org/10.1371/journal.pone.0225899

Editor: Chengdao Li, Murdoch University, AUSTRALIA

Received: April 14, 2019

Accepted: November 14, 2019

Published: December 5, 2019

Copyright: © 2019 Czajkowska et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the manuscript and its Supporting Information files.

Funding: GJ received 269830 from the European Research Council, https://erc.europa.eu/. TAB received 339941 from the European Research Council, https://erc.europa.eu/. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Abstract

The cytokinin dehydrogenase gene *HvCKX2.1* is the regulatory target for the most abundant heterochromatic small RNAs in drought-stressed barley caryopses. We investigated the diversity of *HvCKX2.1* in 228 barley landraces and 216 wild accessions and identified 14 haplotypes, five of these with ten or more members, coding for four different protein variants. The third largest haplotype was abundant in wild accessions (51 members), but absent from the landrace collection. Protein structure predictions indicated that the amino acid substitution specific to haplotype 3 could result in a change in the functional properties of the HvCKX2.1 protein. Haplotypes 1–3 have overlapping geographical distributions in the wild population, but the average rainfall amounts at the collection sites for haplotype 3 plants are significantly higher during November to February compared to the equivalent data for plants of haplotypes 1 and 2. We argue that the likelihood that haplotype 3 plants were excluded from landraces by sampling bias that occurred when the first wild barley plants were taken into cultivation is low, and that it is reasonable to suggest that plants with haplotype 3 are absent from the crop because these plants were less suited to the artificial conditions associated with cultivation. Although the cytokinin signalling pathway influences many aspects of plant development, the identified role of *HvCKX2.1* in the drought response raises the possibility that the particular aspect of cultivation that mitigated against haplotype 3 relates in some way to water utilization. Our results therefore highlight the possibility that water utilization properties should be looked on as a possible component of the suite of physiological adaptations accompanying the domestication and subsequent evolution of cultivated barley.

Introduction

The transition from hunting-gathering to agriculture is arguably the most fundamental change in human history [1–6] and the factors responsible for and influencing the domestication of crop plants remain a subject of intense debate. Agriculture began independently in several different parts of the world, one of these locations being the Fertile Crescent of southwest Asia,

where the earliest evidence for the appearance of domesticated grain crops–einkorn wheat (*Triticum monococcum* L.), emmer wheat (*T*. *dicoccum* (Schrank) Schübl.) and barley (*Hordeum vulgare* L.)–occurs during the 8th millennium bc [7]. The domesticated version of a plant is distinguished from its wild ancestor by a set of phenotypic features collectively referred to as the domestication syndrome [8,9]. A comparison of cultivated species has shown that a similar set of morphological and physiological traits has been selected during domestication of different crops, these including, for cereals, loss of the natural seed dispersal mechanisms and aids, insensitivity to environmental cues that inhibit germination, and an increase in seed size [9,10]. Most domestication traits are looked on as having zero or low adaptive advantage in the wild but high adaptive advantage in the crop [4], and so are assumed to have been selected as a result of cultivation practices.

Although not conventionally looked on as a domestication trait, it is possible that cultivation also resulted in changes to the water utilization properties of crop plants. Water availability is often the main factor limiting the yield of wild and cultivated grain plants such as wheat and barley in the arid to sub-arid environments of southwest Asia [11,12]. Archaeologically recognisable irrigation systems do not appear until the Early Bronze Age, some 4000–5000 years after the beginning of agriculture, and the extent to which the earliest cultivators compensated for aridity through artificial water management is unknown [13]. If practised, such early watering is likely to have been variable and of low intensity, perhaps involving small-scale earthen channels or watering by hand [14]. Because of their transient nature, direct evidence regarding these early water management practices is difficult to obtain, but water availability during plant growth can be inferred from the stable carbon isotope ratios within tissues [15], including archaeologically charred grain [12, 14, 16–20], and by examination of the weed seeds accompanying crops in archaeobotanical assemblages [21–26]. There is, for example, isotopic evidence suggesting water management or low intensity watering of cereal crops at some Neolithic sites in Western Asia prior to the emergence of fully developed irrigation, though any artificial watering used by these early cultivators may have only partially alleviated the effects of the arid and sub-arid environments within which crops were grown [12,14].

The physiological and genetic response of plants to drought conditions is complex [27–29] and is controlled by a variety of phytohormones including abscisic acid (ABA), cytokinins and ethylene [30]. The initial response to drought stress appears to be mediated by ABA, which is synthesized in roots and transported to other parts of the plant [31,32], resulting in changes in gene expression that adapt the plant to the stress conditions [33]. One way in which phytohormones influence gene expression is via small RNAs including microRNAs (miRNAs), which suppress gene activity by increasing transcript degradation and inhibiting translation [34], and heterochromatic small RNAs (hc-siRNAs), which remodel DNA methylation patterns within the promoters of target genes [35]. Both miRNAs [36] and hc-siRNAs [37] have been implicated in the drought response of cereals. In particular, 24-nucleotide hc-siRNAs have been identified that are present in barley caryopses subject to terminal drought stress but absent in control caryopses grown under normal conditions [37]. The most abundant of these hc-siRNAs was homologous to a region within the promoter of the barley *HvCKX2.1* gene, this promoter also containing a binding site for a second, less abundant member of the drought-specific hc-siRNA set. *HvCKX2.1* codes for a cytokinin dehydrogenase, a type of enzyme that regulates cytokinin activity by carrying out an oxidoreduction that degrades the target hormone molecules [38]. In seedlings derived from drought stressed barley caryopses, the *HvCKX2.1* promoter displays increased methylation, the *HvCKX2.1* mRNA content is reduced, and isopentenyladenine and trans-zeatine, which are the target cytokinins for the HvCKX2.1 protein, accumulate [37].

In this paper, we report the diversity of *HvCKX2.1* in an extensive range of georeferenced wild barley accessions and cultivated barley landraces and, from the data, suggest that water

utilization properties should be looked on as an possible component of the suite of physiological adaptations accompanying the domestication and subsequent evolution of cultivated barley.

Materials and methods

Barley accessions

Seeds of 228 barley landraces and 216 wild barley accessions (S1 Table, S2 Table) were obtained from the United States Department of Agriculture–Agricultural Research Service (USDA-ARS) Small Grains Collection (NSGC). Seeds were germinated and seedlings grown in Petri dishes in hydroponic conditions at room temperature (c.22˚C). Once the coleoptiles emerged, the seeds were placed on moist filter paper. Fresh leaf material was collected when the seedlings were 21 days old and DNA extracted using the ISOLATE II Plant DNA kit (Bioline).

DNA sequencing

A 1321 bp segment of the *HvCKX2.1* gene, beginning upstream of the start codon and spanning the first and second exons and the intron between these exons, was amplified as two overlapping fragments (amp1 primers: forward 5´-TACCTATACACAAGGTGCCC-3´, reverse 5´-CCCGAGCCCTACATATCAG-3´, 877 bp product, annealing temperature 65˚C; amp2 primers: forward 5´-TGGACATGATGTCGCTCGGG-3´, reverse 5´-GATCGACGTCAGACTCACCG-3´, 791 bp, 73˚C) and as a single intact product (amp1+2 primers: forward 5´-GAGGGAGTACAGTGTATGCGTATT-3´, reverse 5´-TGATCGACGTCAGACTCACC-3´, 1321 bp, 65˚C). PCRs were carried out in a LightCycler480 (Roche) in 20 μl reaction volumes comprising 100 ng DNA extract, 1× SensiFAST SYBR No-ROX PCR master mix (Bioline), 100 nM forward primer, 100 nM reverse primer and PCR grade water. Cycling parameters were: 95˚C for 5 min; followed by 35 cycles of 30 s at 95˚C, 30 s at the annealing temperature, 60 s at 72˚C. Product formation was assayed using the SYBR Green I/HRM Dye detection format (465 nm excitation, 510 nm emission), and melting data were obtained by first cooling the product to 55˚C for 30 s and then heating to 99˚C with five data acquisitions/˚C. PCR products were purified with the High Pure PCR Product Purification Kit (Roche) and sequenced using the BigDye Terminator v3.1 kit chemistry (Applied Biosystems). Standard sequencing reactions comprised 40 ng PCR product, 1× BigDye sequencing buffer, 0.125× BigDye reaction mix, 4 pmoles primer and UltraPure DNase/RNase-free distilled water to give a final volume of 20 μl. Modified reactions comprised 40 ng PCR product, 1× BigDye sequencing buffer, 0.125× BigDye v3.1 reaction mix, 0.0625× dGTP BigDye v3.0 reaction mix, 4 pmoles primer, 0.95 M betaine (Sigma), 5% (v/v) dimethyl sulfoxide (Sigma) and UltraPure DNase/RNase-free distilled water to give a final volume of 20.05 μl. The modified reactions were carried out to avoid early signal loss when sequencing difficult regions such as those with high GC/GT/G content and/or containing small hairpins or other secondary structures. Cycling parameters were: 2 min at 96˚C; 35 cycles of 40 s at 96˚C, 15 s at 50˚C, 4 min at 60˚C; with products held at 4˚C before purification (Agencourt CleanSEQ; Beckman Coulter) and reading of paired-end sequences by capillary electrophoresis in a 3730 DNA Analyser (Applied Biosystems).

Data analysis

HvCKX2.1 sequences for individual barley accessions were assembled using Geneious version R10 (https://www.geneious.com, [39]) and multiple alignments of assembled sequences from different accessions were generated by the ClustalW, Muscle and Mafft programs. The

consensus sequence of the multiple alignment was identical to the corresponding part of Genbank entry JF495488.1 (*Hordeum vulg*are subsp. *vulgare* cultivar Morex cytokinin oxidase/dehydrogenase [CKX2.1] gene, complete cds). Single nucleotide polymorphisms were identified using the prediction software in Geneious at various settings for maximum variant *p*-value and minimum sequence coverage. Median joining haplotype networks were generated using Network 4 [40] and PopART [41]. Multiple alignment of cytokinin dehydrogenase DNA and protein sequences from different species was carried out online with Clustal Omega [42] at EMBL-EBI. Protein secondary structures were predicted using the garnier tool of EMBOSS [43], operated as a Geneious plug-in. Geographical distribution maps were plotted using ArcMap 10.2.1 of ArcGIS (ESRI. ArcGIS Desktop: Release 10. Redlands, CA: Environmental Systems Research Institute 2011) and correlations between haplotype distribution and modern precipitation data (WorldClim version 2, [44]) were assessed by principal components analysis (PCA) performed with PAST 3.19 [45], and by t-distributed stochastic neighbour embedding (tSNE) and uniform manifold approximation and projection (UMAP) using the Rtsne and umap packages, respectively, of R [46]. A χ^2 test was performed using GraphPad Prism 8.

Results

Diversity of the barley *HvCKX2.1* gene and predicted translation product

We sequenced *HvCKX2.1* in 228 barley landraces and 216 wild barley accessions (S1 Table, S2 Table). Alignment of the sequences revealed multiple variable positions, of which six were identified as high confidence single nucleotide polymorphisms (SNPs) at a maximum variant *p*-value of 10^{-9} and minimum coverage of 393, and a further three were identified as medium confidence SNPs at lower stringency settings (Fig 1, S3 Table). Three of the SNPs, at positions 432, 1112 and 1220 of the amplified region, lie at the third positions within their codons, and do not affect the sequence of the translation product (Table 1). Four other SNPs, at positions 263, 277, 572 and 707, affect the first or second nucleotide of a codon, resulting in the following substitutions: alanine/valine at position 46 of the predicted translation product, histidine/aspartic acid at position 51, isoleucine/threonine at position 149, and glycine/alanine at position 194. The two remaining SNPs, at positions 110 and 113, lie upstream of the initiation codon as listed in the Genbank entry for *HvCKX2.1* (accession number JF495488.1), but lie within the coding region of the entry for *HvCKX2.1* given in the morexGenes database (sequence ID MLOC_53923.1). The discrepancy is because the morexGenes entry uses an upstream `ATG` as the initiation codon, increasing the *N*-terminal region of the predicted translation product by 25 amino acids. According to this translation, SNPs 110 and 113 both affect the second position of a codon resulting in leucine/proline and lysine/arginine substitutions, respectively. However, this upstream `ATG` is not present in the cytokinin dehydrogenase 2 gene of the related grass *Brachypodium distachyon* (S1 Fig), which suggests that for the barley gene the initiation codon used in the Genbank entry is the correct one and that SNPs 110 and 113 do not result in amino acid substitutions.

Complete data for each of the nine SNP positions were available for 372 accessions. These accessions fall into 14 haplotypes, five of which can be looked on as major haplotypes, comprising 232, 54, 51, 11 and 10 accessions, with the remaining nine haplotypes having four or fewer members each (Table 2, S4 Table). Each of the five major haplotypes is present in wild accessions, and haplotypes 1, 2, 4 and 5 are also represented in the landrace collection. In contrast haplotype 3 is absent in the landraces that we studied. Two other minor haplotypes have multiple members: haplotype 6 with four members, each of these landraces, and haplotype 7 comprising two landraces and one wild accession. The other seven haplotypes have one member each, wild accessions for haplotypes 8 and 11–14, and landraces for haplotypes 9 and 10.

```
gaggggagtacagtgtatgcgtattctacctatacacaaggtgcccttatttatgcccatctctccgcttgaaaggagcgtcaccaaacac   90
                                                             m  p  I  c  p  l  e  r  s  v  t  k  h
agagagatcgatcgatcactcaagctaggactacagATGGCCGTGAGGAACAAAGCGCGAGAGATGAGACTGATAGCCGCTCTCTTCGTG  180
 r  e  I  d  r  s  l  k  l  g  l  q  M  A  V  R  N  K  A  R  E  M  R  L  I  A  A  L  F  V   18
CTCAACTGCCTCCTCAAGACAACGACGACGAGCCCCGCCACCTCGGCGTACTCACGTGGTGATGCGCTCGCGTGGAGGCCAGCCACCTCC  270
 L  N  C  L  L  K  T  T  T  T  S  P  A  T  S  A  Y  S  R  G  D  A  L  A  W  R  P  A  T  S   48
TTACTCCACCAGCTCCGCCACCTCGGGGTTCGCGCGCTGATCCGTGACGATGCCGAGGCCACCGCGCTGGCGTCCGCCGACTTCGGCAAC  360
 L  L  H  Q  L  R  H  L  G  V  R  A  L  I  R  D  D  A  E  A  T  A  L  A  S  A  D  F  G  N   78
GTGTCGGACGCCCAGCCGCCGCCGGCAGCCGTGCTCTACCCGTCGCTCCCCGAGGACATCGCCAAGCTGCTGCGCGCCTCGTGCACGCAC  450
 V  S  D  A  Q  P  P  P  A  A  V  L  Y  P  S  L  P  E  D  I  A  K  L  L  R  A  S  C  T  H  108
CCCTCCCCGTTCCCCGTGTCTGCCCGGGGCTGCGGCCACTCGACCCGAGGCCAGGCGTCCGCACCCCGCGGCGTAGTCGTGGACATGATG  540
 P  S  P  F  P  V  S  A  R  G  C  G  H  S  T  R  G  Q  A  S  A  P  R  G  V  V  V  D  M  M  138
TCGCTCGGGTGCCACGCCGGCGGCTCAAGCATCCGCCTCTCCGTCTCGGTCGACGGCCGCTACGTCGACGCCGGCGGCGAGCAGCTGTGG  630
 S  L  G  C  H  A  G  G  S  S  I  R  L  S  V  S  V  D  G  R  Y  V  D  A  G  G  E  Q  L  W  168
GTGGACGTGCTGCGTGCAGCCCTGGCGCACGGCCTCACCCCGTGGTCGTGGACCGACTACCTGCACCTCACCGTCGGCGGCACCCTCTCC  720
 V  D  V  L  R  A  A  L  A  H  G  L  T  P  W  S  W  T  D  Y  L  H  L  T  V  G  G  T  L  S  198
AACGCGGGAATCAGCGGCCAGGCCTTCCGCCACGGCCCCCAAATATCCAGCGTCCAAGAACTCGACGTGATCAGCGgtacgtgtgtggac  810
 N  A  G  I  S  G  Q  A  F  R  H  G  P  Q  I  S  S  V  Q  E  L  D  V  I  S                  223
gacggcacaagatgaacccgtttggttcttcatgtctcgttcacgtactagtagtactaaagattgtgaattactgatatgtagGGCTCG  900
                                                                                      G  L  225
GGGAGATGGTGACATGCTCAAAGGATAAGCACGCGGACCTGTTCGACGCCGTGCTGGGCGGGCTGGGGCAGTTCGGTGTCATAACGCGGG  990
G  E  M  V  T  C  S  K  D  K  H  A  D  L  F  D  A  V  L  G  G  L  G  Q  F  G  V  I  T  R   255
CGCGCATCCCGCTGATGCCGGCGCCGGCGAGGGCGCGCTGGGTGCGACTCTTCTACACGAGTGCGGCGGCGCTCACGGCCGATCAGGAGG 1080
A  R  I  P  L  M  P  A  P  A  R  A  R  W  V  R  L  F  Y  T  S  A  A  A  L  T  A  D  Q  E   285
GGCTCATTGGCATCAACCTCGGCACCGCAGTCTCCGGACTCATGGACTACGTCGAGGGCTCGGTGGTGCTCGCAGACCAGGGCCGAGCAG 1170
G  L  I  G  I  N  L  G  T  A  V  S  G  L  M  D  Y  V  E  G  S  V  V  L  A  D  Q  G  R  A   315
GGAGCTGGCGCTCGTCCTTCTTCTCGGACGCCGACGCGGCGCGCATCGCCGCGCTCGCGGAGGAGGCGGGCGGCATCCTTTACTGCCTCG 1260
G  S  W  R  S  S  F  F  S  D  A  D  A  A  R  I  A  A  L  A  E  E  A  G  G  I  L  Y  C  L   345
AGGGAGCGCTATACTACGGCGGCGCTGCCGCTGCTCGCGGCGGTGAGTCTGACGTCGATCA                              1321
E  G  A  L  Y  Y  G  G  A  A  A  A  R  G  G  E  S  D  V  D  Q                              366
```

Fig 1. The barley cytokinin dehydrogenase gene *HvCKX2.1*. The top panel shows the structure of the gene and the location of the sequenced region. The lower panel shows the sequence of the gene with lower case letters used for the leader sequence and intron, and upper case letters used for the coding region. Primer annealing positions are indicated: amp1 primers, red lettering; amp2 primers, blue lettering; amp1+2 primers, yellow highlight. The possible ATG initiation codons are shown with green highlight. High confidence SNPs and their amino acid substitutions are highlighted in turquoise and medium confidence SNPs and substitutions in grey.

https://doi.org/10.1371/journal.pone.0225899.g001

Table 1. SNPs identified at the *HvCKX2.1* locus.

Position	Variants	Location in gene	Position in codon	Codon variants	Amino acid variants	Amino acid position in protein
110	T, C	upstream	–	–	–	–
113	A, G	upstream	–	–	–	–
263	C, T	exon 1	second	GCC, GTC	ala, val	46
277	C, G	exon 1	first	CAC, GAC	his, asp	51
432	G, A	exon 1	third	CTG, CTA	none	102
572	T, C	exon 1	second	ATC, ACC	ile, thr	149
707	G, C	exon 1	second	GGC, GCC	gly, ala	194
1112	C, G	exon 2	third	GTC, GTG	none	296
1220	C, T	exon 2	third	GCC, GCT	none	332

https://doi.org/10.1371/journal.pone.0225899.t001

Table 2. *HvCKX2.1* haplotypes.

Haplotype	SNP positions									Number of accessions		
	110	113	263	277	432	572	707	1112	1220	Wild	Landraces	Total
1	T	A	C	C	G	T	G	C	C	67	165	232
2	T	A	C	C	G	T	G	G	T	43	11	54
3	C	A	T	C	A	C	G	G	T	51	0	51
4	T	G	C	C	G	T	G	C	C	2	9	11
5	T	A	C	G	G	T	C	C	C	3	7	10
6	T	A	C	C	G	T	G	C	T	0	4	4
7	T	A	T	C	G	T	G	C	C	1	2	3
8	T	A	C	C	G	T	G	G	C	1	0	1
9	T	G	C	G	G	T	C	C	C	0	1	1
10	T	G	C	C	G	T	G	G	T	0	1	1
11	C	A	C	C	G	T	G	G	C	1	0	1
12	C	A	C	C	G	T	G	G	T	1	0	1
13	C	A	T	C	G	C	G	G	C	1	0	1
14	C	A	T	C	G	T	G	G	T	1	0	1

https://doi.org/10.1371/journal.pone.0225899.t002

Those landrace haplotypes with more than two members each include accessions with different growth habits, ear row number and caryopsis structure, and the wild haplotypes are similarly variable for growth habit (S5 Table). Comparing the growth habit phenotype of haplotype 3 to that of all other haplotypes by a χ^2 test yielded a p value of 0.4186, demonstrating that haplotype 3 did not significantly associate with a particular growth habit in the wild population. For 72 accessions, missing data prevented identification of the complete haplotype (S4 Table). For 69 of these accessions, replacement of the unidentified nucleotides could give one of the identified haplotypes, with 17 of these, all wild accessions, being possible additional members of haplotype 3. The remaining three of the 72 accessions have partial haplotypes that cannot be extended into either of the 14 identified haplotypes and which therefore represent additional diversity within the *HvCKX2.1* gene.

When the amino acid substitutions at positions 46, 51, 149 and 194 of the predicted HvCKX2.1 translation product are considered, there are four protein variants (Table 3). All four variants are present in the wild population but variant B is absent from landraces. Variant B is the only type with a threonine at amino acid position 149, which means that all 200 landraces with complete SNP haplotypes have an isoleucine at this position, whereas this position is threonine for 52 of the 172 wild accessions.

Network analysis (Fig 2) placed major haplotype 1 at a principal position, connected by a maximum of three SNP differences to each of the other haplotypes (haplotypes 2, 4, 6, 8, 10–12) specifying protein variant A (ala-his-ile-gly). Haplotypes 5 and 9, giving protein variant C

Table 3. HvCKX2.1 protein variants.

Variant	Amino acid sequence[a]	DNA haplotype(s)	Number of accessions		
			Wild	Landraces	Total
A	ala-his-ile-gly	1, 2, 4, 6, 8, 10, 11, 12	115	190	305
B	val-his-thr-gly	3, 13	52	0	52
C	ala-asp-ile-ala	5, 9	3	8	11
D	val-his-ile-gly	7, 14	2	2	4

[a]The amino acids at positions 46, 51, 149 and 194 are given in order.

https://doi.org/10.1371/journal.pone.0225899.t003

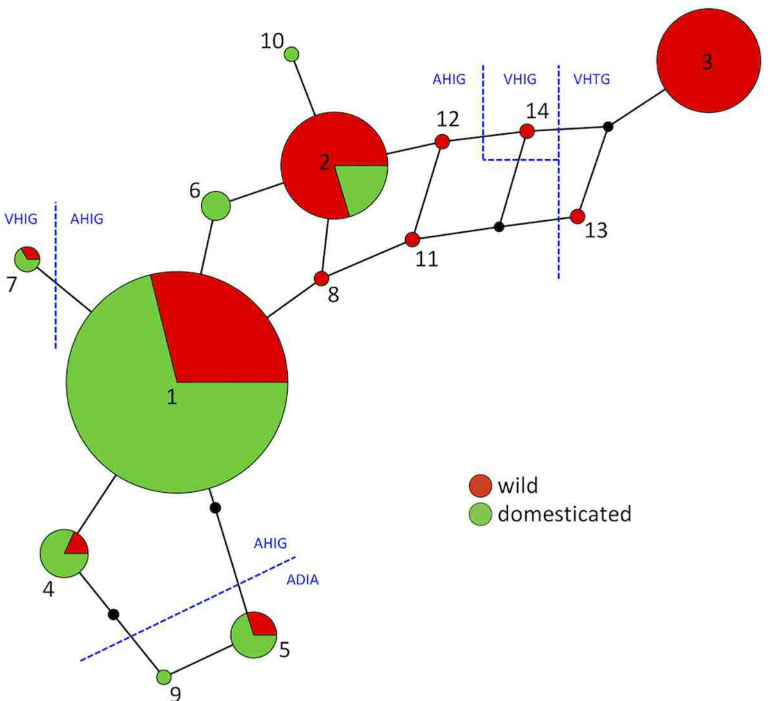


Fig 2. Network displaying the relationships between the fourteen DNA haplotypes of *HvCKX2.1*. Node sizes are proportional to numbers of accessions and empty nodes are shown as black dots. The proportion of wild and domesticated accessions for each haplotype are shown in red and green, respectively. The locations within the network of the four protein variants are shown with the amino acid sequences given in the IUPAC single-letter code.

https://doi.org/10.1371/journal.pone.0225899.g002

(ala-asp-ile-ala), form a pair of linked nodes attached to haplotype 1. Protein variant D (val-his-ile-gly) is specified by haplotypes 7 and 14, which occupy different parts of the network, reflecting their dissimilarity at the DNA level (three out of nine SNP differences). Protein variant B (val-his-thr-gly) is coded by the exclusively wild haplotypes 3 and 13, which occupy a distal part of the network.

Potential effect of *HvCKX2.1* gene diversity on the structure of the HvCKX2.1 protein

The potential impact of the four amino acid substitutions on the structure of the barley HvCKX2.1 translation product was assessed by aligning the barley sequence, with and without the substitutions, with the sequences of cytokinin dehydrogenase proteins from related grasses, and then comparing the predicted secondary structures for each of these proteins (Fig 3). The alanine/valine and histidine/aspartic acid substitutions at positions 46 and 51 of the barley protein, respectively, lie within a relatively non-conserved part of the amino acid sequence alignment, although the two positions are alanine and histidine in the most similar wheat protein, and position 46 is alanine in a cytokinin dehydrogenase 2 protein of *Aegilops tauschii*. The two substitutions are predicted to have minor impact on the secondary structure of the barley protein. The alanine/valine substitution affects the length of a short turn in the predicted barley protein, but this turn is not predicted at the equivalent positions of the rice, *B*. *distachyon* and *Ae*. *tauschii* sequences. The histidine/aspartic acid substitution at position 51 affects the length of helical region, which is absent in sorghum and rice and of variable lengths in the other grass proteins. In contrast, the substitutions at positions 149 and 194 of the barley protein lie in

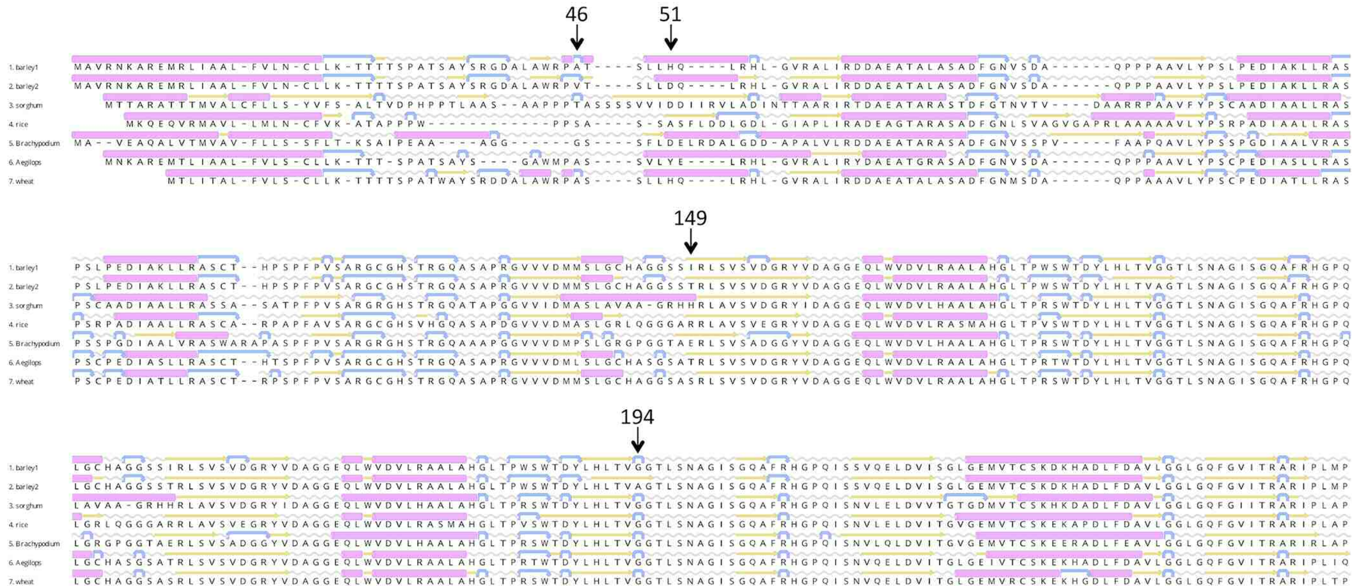


Fig 3. Secondary structure predictions for the barley HvCKX2.1 protein and various other grass cytokinin dehydrogenases. The parts of the protein alignments containing the four substitutions (positions 46, 51, 149 and 194) in HvCKX2.1 are shown. Structural codes: pink barrel, α-helix; yellow arrow, β-strand; blue hooked arrow, turn; grey wavy line, coil. The barley 1 and barley 2 sequences are HvCKX2.1 incorporating the alternative versions of each of the four substitutions. The other sequences are taken from Genbank: sorghum, *Sorghum bicolor* cytokinin dehydrogenase 2, XP_002455003.1; rice, *Oryza sativa japonica* group cytokinin dehydrogenase 2, XP_015629416; Brachypodium, *Brachypodium distachyon* cytokinin dehydrogenase 2-like protein, XP_003564990.3; Aegilops, Aegilops *tauschii* subsp. *tauschii* cytokinin dehydrogenase 2-like protein, XP_020183514.1; wheat, *Triticum aestivum* cytokinin oxidase 2, ADG57787.1.

https://doi.org/10.1371/journal.pone.0225899.g003

regions that display both primary and secondary structural conservation in the grass proteins as a whole. Position 149 is not itself conserved but lies at the *N*-terminus of a predicted β-strand whose length and position is very similar in each sequence. Presence of a threonine at position 149 (the variant absent in landraces) is predicted to stabilise this strand by removing a short turn that is located in the middle of the strand when the isoleucine is present. Position 194 is glycine in each of the other grass proteins, and is located within a 30-amino-acid region that is identical in each of these sequences. The alanine substitution is predicted to move the conserved *C*-terminal position of a β-strand and result in loss of a short conserved turn structure.

To obtain additional insights into the potential impact of the amino acid substitutions on the barley protein, the alignment was extended to include a maize cytokinin dehydrogenase whose X-ray crystallographic structure is known [46]. The maize protein consists of a cytokinin-binding domain and a bipartite binding domain for a flavin adenine dinucleotide (FAD) cofactor. The first part of the FAD binding domain is specified by amino acids 40–244 of the maize protein, which correspond to amino acids 54–262 of the barley version (S2 Fig). The alanine/valine and histidine/aspartic acid substitutions at positions 46 and 51 of the barley protein, in a region that displays poor primary and secondary structure conservation in the grass proteins, are therefore immediately upstream of the FAD binding domain. The isoleucine/threonine and glycine/alanine substitutions at positions 149 and 194 both lie within the FAD binding domain, corresponding to positions 131 and 176 of the maize protein. The first of these positions lies within a part of the maize polypeptide that is located at the protein surface, where a β-strand–turn–β-strand motif forms a finger that protrudes slightly away from the main body of the protein (Fig 4). The β-strand–turn–β-strand structure is predicted for the variant of the barley protein with an isoleucine at position 149, but the turn is not predicted

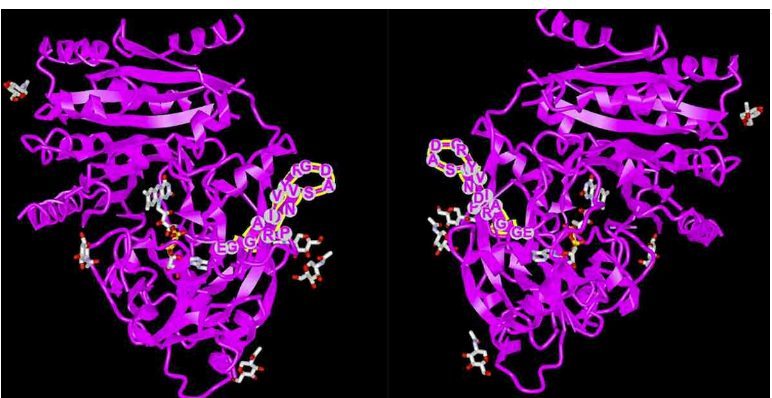


Fig 4. Two views of the X-ray crystallographic structure of a maize cytokinin dehydrogenase protein. The β-strand–turn–β-strand motif that forms a finger on the surface of the protein is highlighted. The structure is PDB ID IW10 and is viewed using the iCnS3D web-base structure viewer at https://www.ncbi.nlm.nih.gov/Structure/icn3d/full.html?complexity=3&buidx=1&showseq=1&mmdbid=29226.

https://doi.org/10.1371/journal.pone.0225899.g004

when the isoleucine is replaced by threonine (see Fig 3). To test whether this comparison between the actual structure of the maize protein and the predicted structure of the barley protein is valid, we also predicted the secondary structure of the FAD domain of the maize protein from its amino acid sequence. There was good agreement between the predicted and actual structures of the maize protein in the region surrounding position 131, with the prediction identifying the β-strand–turn–β-strand motif at the correct position with only minor conformational differences compared with the actual motif (Fig 5). The accurate prediction of this motif from the maize amino acid sequence, and the agreement between the maize and barley predictions in this region, suggests that the β-strand–turn–β-strand structure is also likely to be a genuine feature of the barley protein when isoleucine is present at position 149, and that this motif might be disrupted by replacement of the isoleucine by threonine in the *HvCKX2.1* haplotype that is absent in landraces. The maize structure also includes an asparagine (position 134 of the maize protein) that is the binding site for an *N*-linked *N*-acetylglucosamine sugar residue [47]. The barley protein does not have a potential *N*-linked binding site in this region, but the threonine substitution at barley position 149 would create a potential *O*-linked site. Finally, position 176 of the maize protein (corresponding to the glycine/alanine variation at position 194 of the barley protein) is a glycine located in the same conserved 30-amino-acid region noted above for the other grass proteins, this conserved region including an aspartic acid (position 169 in the maize protein) which is thought to play a critical role as a hydrogen bond acceptor during cytokinin binding [47,48]. In this region, there is poor agreement between the predicted and actual secondary structures of the maize protein, invalidating any further comparisons with the predicted secondary structure of the barley protein.

Geographical distributions of the *HvCKX2.1* haplotypes

We examined the geographical distributions of wild plants of different haplotypes, to assess if the absence of haplotype 3 in landraces could be due to the geographical location(s) of the earliest farming sites in the Fertile Crescent being such that haplotype 3 was not sampled when wild plants were first taken into cultivation. Haplotypes 1–3, which comprise 67, 43 and 51 wild accessions, respectively, have overlapping geographical distributions in the wild population (Fig 6). Wild accessions with haplotype 1 are distributed throughout the Fertile Crescent and are also present in central Asia, including the Balkan region of western Turkmenistan.

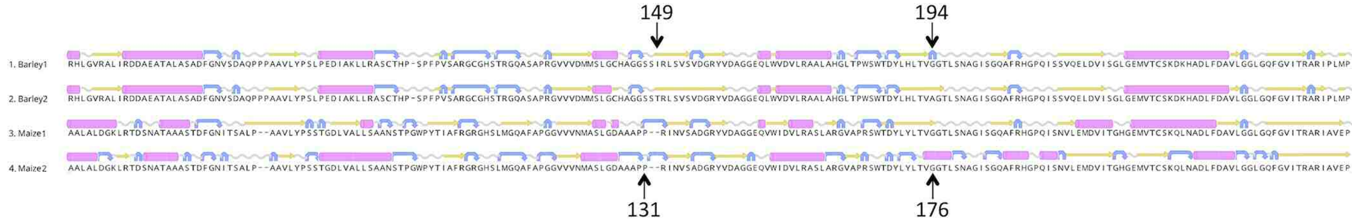
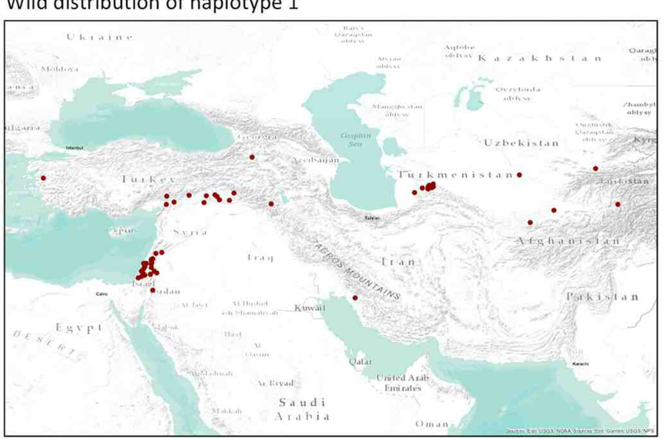
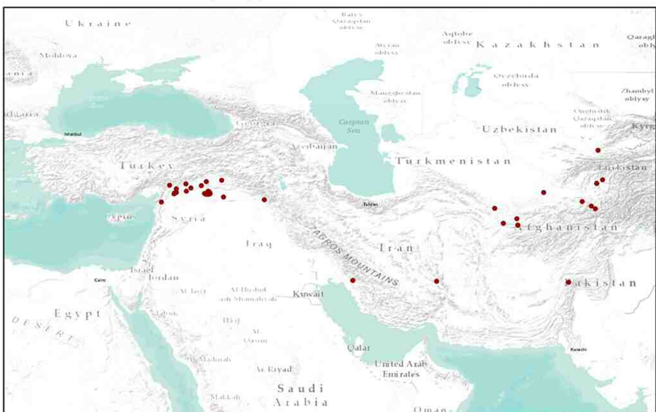


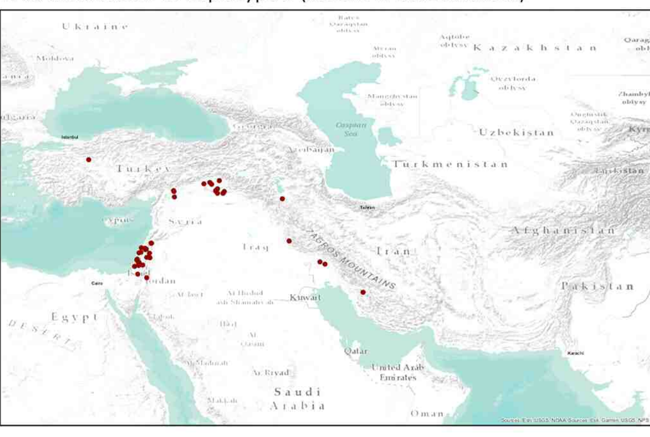
Fig 5. Secondary structure of the FAD binding domains of the barley HvCKX2.1 protein and a maize cytokinin dehydrogenase protein. The Barley 1 and Barley 2 sequences are HvCKX2.1 incorporating the alternative versions of the substitutions at positions 149 and 194. Maize 1 is the secondary structure of a maize cytokinin oxidase/dehydrogenase as predicted from its amino acid sequence, and Maize 2 is the actual secondary structure of this maize cytokinin oxidase/dehydrogenase according to the DSSP analysis of the X-ray crystallographic data of the protein complexed with N6-(3-methoxy-phenyl)adenine (PDB 3DQ0 available at https://www.rcsb.org/structure/3DQ0). Structural codes: pink barrel, α-helix; yellow arrow, β-strand; blue hooked arrow, turn; grey wavy line, coil. Note that the alignment between and the barley and maize amino acid sequences around barley position 149 is slightly different to that shown in S2 Fig.

https://doi.org/10.1371/journal.pone.0225899.g005

Haplotype 2 is present in the northern Fertile Crescent and central Asia, but absent from the wild barley population in the southern Levant. Wild accessions with haplotype 3, the haplotype absent in domesticated barley, are distributed throughout the Fertile Crescent. The current







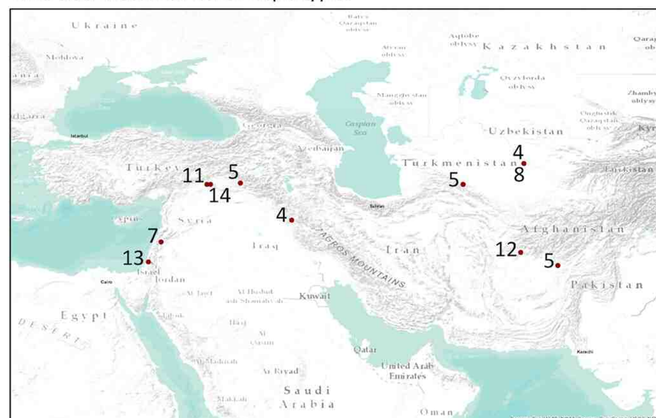


Fig 6. Distributions of wild accessions belonging to different haplotypes. Maps were plotted using ArcMap 10.2.1 of ArcGIS (ESRI. ArcGIS Desktop: Release 10. Redlands, CA: Environmental Systems Research Institute 2011).

https://doi.org/10.1371/journal.pone.0225899.g006

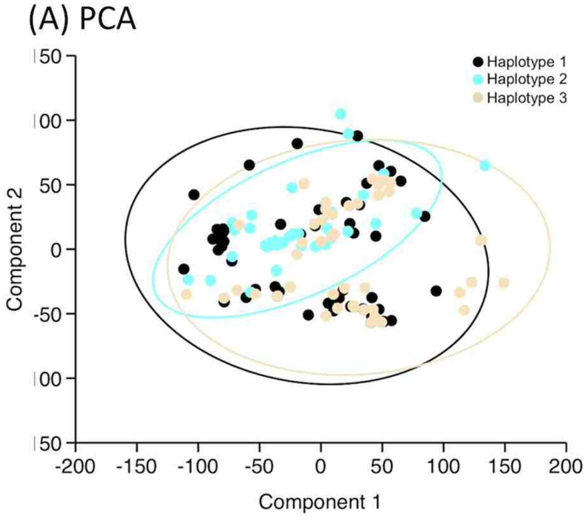
distributions within the Fertile Crescent of haplotypes 1 and 3 are therefore very similar, and in the northern part of the arc both distributions include the full range of haplotype 2.

To explore whether wild accessions belonging to haplotype 3 might respond differently to precipitation, we carried out a PCA using as input data the combined monthly precipitation amounts for the collection sites of each wild accession. When the rainfall data for all months are combined, the resulting plot (Fig 7) shows extensive overlap between the precipitation envelopes for each of the three haplotypes, although wild plants belonging to haplotype 2 occupy a smaller precipitation envelope than either haplotypes 1 or 3, consistent with the less broad geographical distribution of haplotype 2. The envelope for haplotype 3 extends slightly outside of the range of the other two haplotypes in PC1, and excludes an area of the plot occupied by outliers of haplotypes 1 and 2; otherwise, the envelope for haplotype 3 shows no significant difference compared to the combined envelopes for haplotypes 1 and 2. The small differences described above were not apparent when the annual rainfall data for the three haplotypes were analysed by tSNE and UMAP (Fig 7B and 7C).

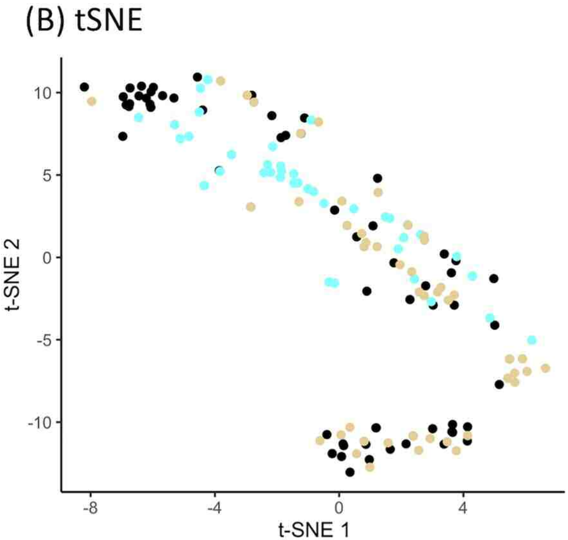
To assess if there was any correlation between haplotype and seasonal rainfall patterns, PCA and tSNE were also performed with the rainfall data for different bimonthly periods (S3 Fig, S4 Fig). Again the PCA, but not tSNE, suggested small differences in the rainfall envelope of haplotype 3 compared to the envelopes for haplotypes 1 and 2, at least for the bimonthly periods October/November to March/April. To investigate further, graphs were drawn plotting the average rainfall per month for the collection sites of wild accessions belonging to haplotypes 1, 2 and 3 (Fig 8). The graphs revealed a significant difference in the rainfall data for haplotype 3, these accessions coming from areas with higher rainfall during November to February. This feature was apparent when all the accessions were considered together (Fig 8A) and when the winter barleys were considered on their own (Fig 8B). However, when the springs barleys were examined, there was no significant differences between the plots for haplotypes 1 and 3 (Fig 8C).

Discussion

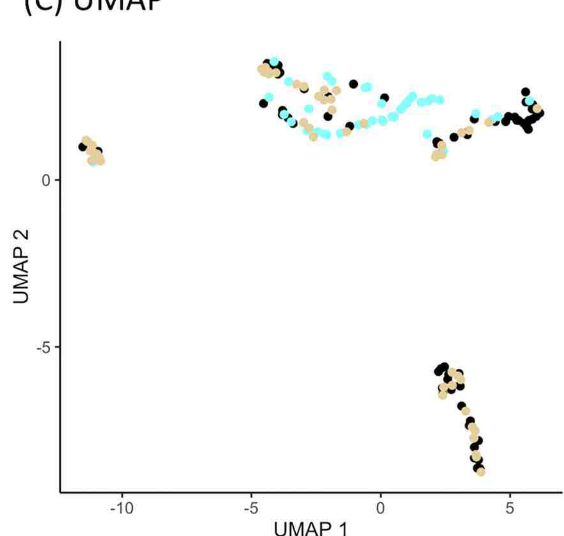
We studied the diversity of the barley cytokinin dehydrogenase gene *HvCKX2.1* in an extensive range of georeferenced wild barley accessions and cultivated barley landraces. The role of *HvCKX2.1* as the target of the most abundant drought-responsive hc-siRNAs in barley caryopses, and the reduced *HvCKX2.1* expression that occurs in seedlings derived from drought stressed plants, indicates that this gene contributes to the water utilization properties of barley plants. Our results show that cultivated barley landraces lack one of the five major haplotypes of the *HvCKX2.1* gene present in the wild population, resulting in the absence in landraces of a version of the cytokinin dehydrogenase protein with a threonine rather than isoleucine at position 149 of the predicted translation product. This position lies within the FAD binding domain of the protein, comparison with the X-ray structure of a maize cytokinin dehydrogenase suggesting that the isoleucine/threonine substitution affects the conformation of a finger motif that projects from the surface of the protein. According to secondary structure prediction, this finger motif is conserved in the isoleucine version of the barley protein, but is disrupted by the threonine substitution, resulting in loss of the turn linking the two β-strands of the finger. Additionally, the maize motif is linked to an *N*-acetylglucosamine sugar residue which would be absent from the isoleucine version of the barley protein because this sequence lacks a potential glycosylation site, though such a site would be created by the threonine substitution. Although there are no published data reporting a role for the finger motif in the function of the maize protein, the predicted structural changes that we describe make it possible



(A) PCA
Haplotype 1
Haplotype 2
Haplotype 3
Component 2
Component 1



(B) tSNE
t-SNE 2
t-SNE 1



(C) UMAP
UMAP 2
UMAP 1

Fig 7. Dimensionality reduction analyses using as input data the combined monthly precipitation amounts (WorldClim version 2) for the collection sites of each wild accession. (A) PCA. Component 1 accounted for 67.8% of the total variability and component 2 accounted for 25.3%. The ellipses indicate the regions within which 95% of the data points for each haplotype are expected to fall. (B) tSNE, run with perplexity = 30, iterations = 2000 and theta = 0.5. (C) UMAP. Black dots and ellipses, haplotype 1; cyan, haplotype 2; orange, haplotype 3.

https://doi.org/10.1371/journal.pone.0225899.g007

that the presence of threonine rather than isoleucine at position 149 results in a change in the properties of the HvCKX2.1 protein.

There are several possible explanations for the apparent absence of haplotype 3 in landraces. This first is that this is simply due to sampling bias that occurred when we assembled our landrace collection. However, the likelihood of haplotype 3 being excluded by sampling bias is low: if, in reality, haplotype 3 is present in landraces at a frequency of 29.65% (the frequency of this

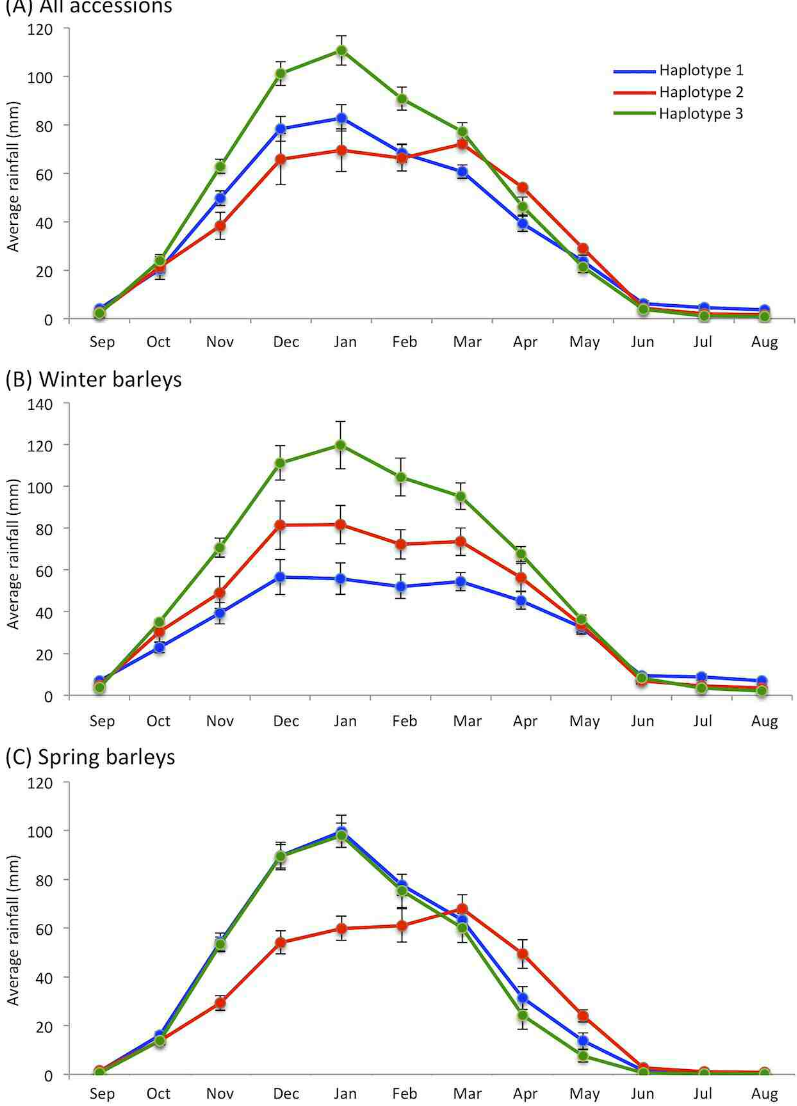


Fig 8. Average monthly precipitation amounts (WorldClim version 2) at the collections sites for wild accessions of haplotypes 1, 2 and 3. (A) All accessions; (B) winter barleys; (C) spring barleys. Blue, haplotype 1; red, haplotype 2; green, haplotype 3. Bars indicate standard error.

https://doi.org/10.1371/journal.pone.0225899.g008

haplotype in the wild accessions) then the probability of haplotype 3 being absent in a random collection of 200 landraces is 2.84×10^{-31}. Our landrace collection had a broad geographical distribution (see S1 Table) and hence was unlikely to be so non-random as to bias this probability to the extent that haplotype 3 was missed due to sampling effects.

A second possibility is that the absence of haplotype 3 in landraces is due to sampling bias that occurred when the first wild barley plants were taken into cultivation. If barley was initially domesticated from a wild population that contained relatively few genotypes then it is possible that haplotype 3 was missed purely by chance, and hence never made its way into the crop. In our view, two factors reduce the likelihood of this scenario. First, in the modern wild population, the geographical distributions of haplotypes 1 and 3 are very similar and both encompass the full range of haplotype 2. If the modern phylogeography reflects the haplotype frequency and distribution when barley was first taken into cultivation, then this would appear to mitigate against the possibility that the early farmers, purely by chance, only domesticated wild plants belonging to haplotypes 1 and 2, to which the majority of the landraces belong, when haplotype 3 plants were growing in similar areas. The second argument which in our view makes its unlikely that haplotype 3 was excluded by chance from the crop is the evidence from genome-wide studies that cultivated barley emerged as a genetic mosaic of wild source populations, with the diversity of the crop established in part by hybridization between early cultivated forms and various wild populations [49,50]. Exclusion of haplotype 3 purely by chance therefore requires not only the absence of this haplotype among the initial set(s) of plants taken into cultivation, but also the absence of haplotype 3 in any of the wild populations with which the early crop subsequently hybridized.

From the above considerations it seems plausible that the absence of haplotype 3 in landraces is due to these plants being less suited to the artificial conditions associated with cultivation. Cytokinin dehydrogenases are one of a number of enzyme families that participate in the cytokinin signalling pathway of plants [51], this pathway regulating diverse physiological processes involved in plant growth, development and the response to stresses such as drought and heat. Although *HvCKX2.1* has been highlighted as the regulatory target for the most abundant hc-siRNAs in barley caryopses subject to terminal drought stress [37], this does not preclude the possibility that the HvCKX2.1 protein has other, as yet undetected functions in those parts of the cytokinin signalling pathway that operate in developing caryopses and/or in seedlings up to 12–24 hours after imbibition, these being the growth stages when *HvCKX2.1* RNA is present in plant tissue [37]. If the protein has other such roles, then any one of these, or a combination, could underlie the absence of haplotype 3 in landraces. However, we believe that it is reasonable based on what is known about the role of *HvCKX2.1* in the drought response to propose as a working hypothesis that the particular aspect of cultivation that mitigates against haplotype 3 relates in some way to water utilization. The rainfall analysis would appear to support this hypothesis, by suggesting that there are differences in the preferred precipitation patterns for plants of different haplotypes in the natural environment, manifested most clearly by the significantly higher rainfall at the collection sites of haplotype 3 plants during November to March (Fig 8), which includes the period when grain from plants with a winter growth habit is undergoing germination and early seedling growth.

The hypothesis that haplotype 3 plants are less suited to the artificial hydrological conditions associated with cultivation is prompted by the genetic data that we report in this paper, but comparison between the genetic data and environmental factors can only ever provide indirect support for such a hypothesis. Confirmation of the hypothesis would require detailed functional studies aimed at discerning some difference between the physiological properties of plants belonging to haplotype 3 (or more specifically to plants whose HvCKX2.1 protein contains a threonine rather than isoleucine at position 149) and plants carrying other versions of

HvCKX2.1. Transgenic experiments or gene editing could be used to ensure that the properties of different *HvCKX2.1* variants are examined in a uniform genetic background. Such studies would be complex, as the precise nature of any phenotypic change cannot be predicted and could be subtle, but the presence of *HvCKX2.1* mRNA in developing caryopses and in seedlings up to 12–24 hours after imbibition [37] suggests that the altered phenotype is likely to be expressed during grain filling and/or germination. Sequencing of *HvCKX2.1* transcripts might also be carried out to check if there are any differences in splice site selection and the usage of transcription start sites and polyadenylation sites in wild and domesticated plants of different haplotypes.

Conclusion

The traditionally recognised traits characterizing the domestication syndrome for grain crops such as barley include loss of the natural seed dispersal mechanisms, increase in seed size, and insensitivity to environmental cues that inhibit germination [8–10]. It has been suggested, however, that domestication of wild grasses was also accompanied by selection for physiological changes driven by early cultivation practices [52,53]. Our results highlight the possibility that one of these practices was water management, and that water utilization properties should be looked on as a possible component of the suite of physiological adaptations accompanying the domestication of barley and, by implication, other grain crops that were domesticated in arid or semi-arid environments. By raising the possibility that genetic adaptation occurred in response to the artificial hydrological conditions associated with cultivation, our results also emphasise the important role that water availability played during the emergence of agriculture in the Fertile Crescent, and indicate that the development of crop husbandry techniques able to mitigate against water stress could have been a major factor in ensuring the sustainability of early cultivation in the region.

Supporting information

S1 Table. Barley landraces used in this study.
(XLSX)

S2 Table. Wild barley accessions used in this study.
(XLSX)

S3 Table. Positions of SNPs identified at difference confidence settings.
(XLSX)

S4 Table. Haplotype identities for the landraces and wild barley accessions.
(XLSX)

S5 Table. Phenotypes of accessions belonging to different HvCKX2.1 haplotypes.
(XLSX)

S1 Fig. Alignment between the upstream regions of the *Brachypodium distachyon* cytokinin dehydrogenase 2-like gene and the barley *HvCKX2.1* gene. Nucleotides in the upstream regions are in lower case and those in the coding regions in uppercase, with nucleotide identities indicated by asterisks. Potential initiation codons are highlighted in green. The *Brachypodium* gene is Genbank accession number XM_003564942.3 and the barley gene is accession number JF495488.1.
(TIFF)

S2 Fig. Alignment between different cytokinin dehydrogenase genes. The location of the FAD binding domain is shown in red, based on the sequence of the maize protein. Amino acids highlighted in green in the barley sequence are those substituted in the different HvCKX2.1 protein variants. Asterisks indicate positions where the amino acid is identical in each sequence; colons indicate positions occupied by amino acids with strongly similar properties (>0.5 in the Gonnet point accepted mutation [PAM] 250 matrix); periods indicate positions occupied by amino acids with weakly similar properties (<0.5 and >0 in the Gonnet PAM 250 matrix). The barley sequence is the HvCKX2.1 consensus sequence as determined in this study (identical with Genbank JF495488.1). The other sequences are taken from Genbank: maize, *Zea mays* cytokinin oxidase 1, accession number ONM29023.1; sorghum, *Sorghum bicolor* cytokinin dehydrogenase 2, XP_002455003.1; rice, *Oryza sativa japonica* group cytokinin dehydrogenase 2, XP_015629416; Brachypodium, *Brachypodium distachyon* cytokinin dehydrogenase 2-like protein, XP_003564990.3; Aegilops, *Aegilops tauschii* subsp. *tauschii* cytokinin dehydrogenase 2-like protein, XP_020183514.1; wheat, *Triticum aestivum* cytokinin oxidase 2, ADG57787.1.
(TIFF)

S3 Fig. PCAs using as input data the bimonthly precipitation amounts (WorldClim version 2) for the collection sites of each wild accession. The ellipses indicate the regions within which 95% of the data points for each haplotype are expected to fall. Black dots and ellipses, haplotype 1; cyan, haplotype 2; orange, haplotype 3.
(TIFF)

S4 Fig. tSNE using as input data the bimonthly precipitation amounts (WorldClim version 2) for the collection sites of each wild accession. tSNE was run with perplexity = 30, iterations = 2000 and theta = 0.5. Black dots and ellipses, haplotype 1; cyan, haplotype 2; orange, haplotype 3.
(TIFF)

Author Contributions

Conceptualization: Beata I. Czajkowska, Glynis Jones, Terence A. Brown.

Data curation: Terence A. Brown.

Formal analysis: Beata I. Czajkowska, Conor M. Finlay, Terence A. Brown.

Funding acquisition: Glynis Jones, Terence A. Brown.

Investigation: Beata I. Czajkowska.

Methodology: Beata I. Czajkowska.

Project administration: Terence A. Brown.

Supervision: Terence A. Brown.

Validation: Beata I. Czajkowska, Terence A. Brown.

Writing – original draft: Beata I. Czajkowska, Terence A. Brown.

Writing – review & editing: Beata I. Czajkowska, Glynis Jones, Terence A. Brown.

References

1. Abbo S, Gopher A, Rubin B, Lev-Yadun S. On the origin of Near Eastern founder crops and the 'dump-heap hypothesis'. Genet Resour Crop Evol. 2005; 52: 491–495.

2. Zeder MA. Central questions in the domestication of plants and animals. Evol Anthropol. 2006; 15: 105–117.
3. Zeder MA. Domestication and early agriculture in the Mediterranean Basin: origins, diffusion, and impact. Proc Natl Acad Sci USA. 2008; 105: 11597–11604. https://doi.org/10.1073/pnas.0801317105 PMID: 18697943
4. Brown TA, Jones MK, Powell W, Allaby RG. The complex origins of domesticated crops in the Fertile Crescent. Trends Ecol Evol. 2009; 24: 103–109. https://doi.org/10.1016/j.tree.2008.09.008 PMID: 19100651
5. Fuller DQ. An emerging paradigm shift in the origins of agriculture. Gen Anthropol. 2010; 17(1): 8–12.
6. Abbo S, Gopher A. Near Eastern plant domestication: a history of thought. Trends Plant Sci. 2017; 22: 491–511. https://doi.org/10.1016/j.tplants.2017.03.010 PMID: 28434795
7. Zohary D, Hopf M, Weiss E. Domestication of Plants in the Old World, 4th edn. Oxford: Oxford University Press; 2012.
8. Hammer K. Das Domestikationssyndrom. Kulturpflanze. 1984; 32: 11–34.
9. Fuller DQ. Contrasting patterns in crop domestication and domestication rates: recent archaeobotanical insights from the Old World. Annals Bot. 2007; 100: 903–924.
10. Gepts P. Crop domestication as a long-term selection experiment. Plant Breed Rev. 2004; 24: 1–44.
11. Abbo S, Lev-Yadun S, Gopher A. Agricultural origins: centers and non-centers; a Near Eastern reappraisal. Crit Rev Plant Sci. 2010; 29: 317–328.
12. Riehl S, Pustovoytov KE, Weippert H, Klett S, Hole F. Drought stress variability in ancient Near Eastern agricultural systems evidenced by $\delta^{13}C$ in barley grain. Proc Natl Acad Sci USA. 2014; 111: 12348–12353. https://doi.org/10.1073/pnas.1409516111 PMID: 25114225
13. Mithen S. The domestication of water: water management in the ancient world and its prehistoric origins in the Jordan Valley. Phil Trans R Soc A. 2010; 368: 5249–5274. https://doi.org/10.1098/rsta.2010.0191 PMID: 20956370
14. Wallace MP, Jones G, Charles M, Fraser R, Heaton THE, Bogaard A. Stable carbon isotope evidence for Neolithic and Bronze Age crop water management in the Eastern Mediterranean and Southwest Asia. PLoS ONE. 2015; 10(6): e0127085. https://doi.org/10.1371/journal.pone.0127085 PMID: 26061494
15. Farquhar G, Richards R. Isotopic composition of plant carbon correlates with water-use efficiency of wheat genotypes. Aus J Plant Physiol. 1984. 11: 539–552.
16. Araus J, Buxó R. Changes in carbon isotope discrimination in grain cereals from the north-western Mediterranean Basin during the past seven millennia. Aus J Plant Physiol. 1993; 20: 117–128.
17. Araus J, Febrero A, Buxó R, Rodríguez-Ariza M, Molina F, Camalich MD, et al. Identification of ancient irrigation practices based on the carbon isotope discrimination of plant seeds: a case study from the South-East Iberian Peninsula. J Archaeol Sci. 1997; 24: 729–740.
18. Araus J, Febrero A, Catala M, Molist M, Voltas J, Romagosa I, et al. Crop water availability in early agriculture: evidence from carbon isotope discrimination of seeds from a tenth millennium BP site on the Euphrates. Glob Change Biol. 1999; 5: 201–212.
19. Ferrio JP, Araus JL, Buxó R, Voltas J, Bort J. Water management practices and climate in ancient agriculture: inferences from the stable isotope composition of archaeobotanical remains. Veget Hist Archaeobot. 2005; 14: 510–517.
20. Riehl S, Bryson R, Pustovoytov K. Changing growing conditions for crops during the Near Eastern Bronze Age (3000–1200 BC): the stable carbon isotope evidence. J Archaeol Sci. 2008; 35: 1011–1022.
21. Jones G, Charles M, Colledge S, Halstead, P. Towards the archaeobotanical recognition of winter cereal irrigation: an investigation of modern weed ecology in northern Spain. In Kroll H, Pasternak R, editors. Res Archaeobotanicae–International Workgroup for Palaeoethnobotany (Proceedings of the 9th Symposium, Kiel 1992). Kiel: Oetker-Voges-Verlag; 1995. pp. 49–68.
22. Jones G, Charles M, Bogaard A, Hodgson J, Palmer C. The functional ecology of present-day arable weed floras and its applicability for the identification of past crop husbandry. Veget Hist Archaeobot. 2005; 14: 493–504.
23. Jones G, Charles M, Bogaard A, Hodgson J. Crops and weeds: the role of weed functional ecology in the identification of crop husbandry methods. J Archaeol Sci. 2010; 37: 70–77.
24. Charles M, Jones G. FIBS in archaeobotany: functional interpretation of weed floras in relation to husbandry practices. J Archaeol Sci. 1997; 24: 1151–1161.
25. Charles M, Hoppé C. The effects of irrigation on the weed floras of winter cereal crops in Wadi Ibn Hamad (Southern Jordan). Levant. 2003; 35: 213–230.

26. Charles M, Hoppé C, Jones G, Bogaard A, Hodgson J. Using weed functional attributes for the identification of irrigation regimes in Jordan. J Archaeol Sci. 2003; 30: 1429–1441.
27. Honsdorf N, March TJ, Berger B, Tester M, Pillen K. High-throughput phenotyping to detect drought tolerance QTL in wild barley introgression lines. PLoS ONE. 2014; 9(5): e97047. https://doi.org/10.1371/journal.pone.0097047 PMID: 24823485
28. Hu H, Xiong L. Genetic engineering and breeding of drought-resistance crops. Ann Rev Plant Biol. 2014; 65: 715–741.
29. Nuccio ML, Paul M, Bate NJ, Cohn J, Cutler SR. Where are the drought tolerant crops? An assessment of more than two decades of plant biotechnology effort in crop improvement. Plant Sci. 2018; 273: 110–119. https://doi.org/10.1016/j.plantsci.2018.01.020 PMID: 29907303
30. Wilkinson S, Kudoyarova GR, Veselov DS, Arkhipova TN, Davies WJ. Plant hormone interactions: innovative targets for crop breeding and management. J Exp Bot. 2012; 63: 3499–3509. https://doi.org/10.1093/jxb/ers148 PMID: 22641615
31. Wilkinson S, Davies WJ. Drought, ozone, ABA and ethylene: new insights from cell to plant to community. Plant Cell Environ. 2010; 33: 510–525. https://doi.org/10.1111/j.1365-3040.2009.02052.x PMID: 19843256
32. Basu S, Ramegowda V, Kumar A, Pereira A. Plant adaptation to drought stress. F1000 Res. 2016; 5: 1554.
33. Cutler SR, Rodriguez PL, Finkelstein RR, Abrams SR. Abscisic acid: emergence of a core signaling network. Ann Rev Plant Biol. 2010; 61: 651–679.
34. Budak H, Akpinar BA. Plant miRNAs: biogenesis, organization and origins. Funct Integr Genet. 2015; 15: 523–531.
35. Matzke MA, Kanno T, Matzke AJ. RNA-directed DNA methylation: the evolution of a complex epigenetic pathway in flowering plants. Ann Rev Plant Biol. 2015; 66: 243–267.
36. Ferdous J, Hussain SS, Shi BJ. Role of microRNAs in plant drought tolerance. Plant Biotechnol J. 2015; 13: 293–305. https://doi.org/10.1111/pbi.12318 PMID: 25583362
37. Surdonja K, Eggert K, Hajirezaei M-R, Harshavardhan VT, Seiler C, von Wirén N, et al. Increase of DNA methylation at the *HvCKX2.1* promoter by terminal drought stress in barley. Epigenomes. 2017; 1: 9.
38. Galuszka P, Frébort I, Šebela M, Sauer P, Jacobsen S, Peč P. Cytokinin oxidase or dehydrogenase? Mechanism of cytokinin degradation in cereals. Eur J Biochem. 2001; 268: 450–461. https://doi.org/10.1046/j.1432-1033.2001.01910.x PMID: 11168382
39. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 2012; 28: 1647–1649. https://doi.org/10.1093/bioinformatics/bts199 PMID: 22543367
40. Bandelt H-J, Forster P, Röhl A. Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol. 1999; 16: 37–48. https://doi.org/10.1093/oxfordjournals.molbev.a026036 PMID: 10331250
41. Leigh JW, Bryant D. PopART: Full-feature software for haplotype network construction. Methods Ecol Evol. 2015; 6: 1110–1116.
42. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol. 2011; 7: 539. https://doi.org/10.1038/msb.2011.75 PMID: 21988835
43. Rice P, Longden I, Bleasby A. EMBOSS: The European Molecular Biology Open Software Suite. Trends Genet. 2000; 16: 276–277. https://doi.org/10.1016/s0168-9525(00)02024-2 PMID: 10827456
44. Fick SE, Hijmans RJ. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. Int J Climatol. 2017; 37: 4302–4315.
45. Hammer Ø, Harper DAT, Ryan PD. PAST: Paleontological statistics software package for education and data analysis. Palaeontol Electron. 2001; 4: 9–18.
46. R Core Team. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2017. Available from: https://www.R-project.org/
47. Malito E, Coda A, Bilyeu KD, Fraaije MW, Mattevi A. Structures of Michaelis and product complexes of plant cytokinin dehydrogenase: implications for flavoenzyme catalysis. J Mol Biol. 2004; 341: 1237–1249. https://doi.org/10.1016/j.jmb.2004.06.083 PMID: 15321719
48. Kopečný D, Končitíková R, Popelka H, Briozzo P, Vigouroux A, Kopečná M, et al. Kinetic and structural investigation of the cytokinin oxidase/dehydrogenase active site. FEBS J. 2016; 283: 361–377. https://doi.org/10.1111/febs.13581 PMID: 26519657

49. Poets AM, Fang Z, Clegg MT, Morrell PL. Barley landraces are characterized by geographically heterogeneous genomic origins. Genome Biol. 2015; 16: 173. https://doi.org/10.1186/s13059-015-0712-3 PMID: 26293830

50. Pankin A, Altmüller J, Becker C., von Korff M. Targeted resequencing reveals genomic signatures of barley domestication. New Phytol. 2018; 218: 1247–1259. https://doi.org/10.1111/nph.15077 PMID: 29528492

51. Keiber JJ, Schaller GE. Cytokinin signalling in plant development. Development 2018; 145: dev149344. https://doi.org/10.1242/dev.149344 PMID: 29487105

52. Cunniff J, Wilkinson S, Charles M, Jones G, Rees M, Osborne CP. Functional traits differ between cereal crop progenitors and other wild grasses gathered during the Neolithic in southwest Asia. PLoS ONE. 2014; 9(1): e87586. https://doi.org/10.1371/journal.pone.0087586 PMID: 24489941

53. Preece C, Livarda A, Christin P-A, Wallace M, Martin G, Charles M, et al. How did domestication of Fertile Crescent grain crops increase their yields? Funct Ecol. 2017; 31: 387–397. https://doi.org/10.1111/1365-2435.12760 PMID: 28286354

2.

Diversity of a wall-associated kinase gene in wild and cultivated barley.

Czajkowska BI, Jones G, Brown TA

Published on 27 June 2019

in

PLoS ONE

14(6):e0218526

https://doi.org/10.1371/journal. pone.0218526

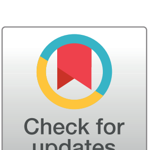
RESEARCH ARTICLE

Diversity of a wall-associated kinase gene in wild and cultivated barley

Beata I. Czajkowska[1], **Glynis Jones**[2], **Terence A. Brown**[1]*

1 School of Earth and Environmental Sciences, Manchester Institute of Biotechnology, University of Manchester, Manchester, United Kingdom, 2 Department of Archaeology, University of Sheffield, Northgate House, Sheffield, United Kingdom

* terry.brown@manchester.ac.uk



OPEN ACCESS

Citation: Czajkowska BI, Jones G, Brown TA (2019) Diversity of a wall-associated kinase gene in wild and cultivated barley. PLoS ONE 14(6): e0218526. https://doi.org/10.1371/journal.pone.0218526

Editor: Chengdao Li, Murdoch University, AUSTRALIA

Received: December 18, 2018

Accepted: June 4, 2019

Published: June 27, 2019

Copyright: © 2019 Czajkowska et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the manuscript and its Supporting Information files.

Funding: Funding was provided by the European Research Council, grants 339941 to TAB and 269830 to GJ, https://erc.europa.eu/. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Abstract

Domestication of barley and other cereals was accompanied by an increase in seed size which has been ascribed to human selection, large seeds being preferred by early farmers or favoured by cultivation practices such as deep sowing. An alternative suggestion is that the increase in seed size was an indirect consequence of selection for plants with more vigorous growth. To begin to address the latter hypothesis we studied the diversity of *HvWAK1*, a wall-associated kinase gene involved in root proliferation, in 220 wild barley accessions and 200 domesticated landraces. A 3655-bp sequence comprising the gene and upstream region contained 69 single nucleotide polymorphisms (SNPs), one indel and four short tandem repeats. A network of 50 haplotypes revealed a complex evolutionary relationship, but with landraces largely restricted to two parts of the topology. SNPs in the *HvWAK1* coding region resulted in nonsynonymous substitutions at nine positions in the translation product, but none of these changes were predicted to have a significant effect on the protein structure. In contrast, the region upstream of the coding sequence contained five SNPs that were invariant in the domesticated population, fixation of these SNPs decreasing the likelihood that the upstream of a pair of TATA boxes and transcription start sites would be used to promote transcription of *HvWAK1*. The sequence diversity therefore suggests that the *cis*-regulatory region of *HvWAK1* might have been subject to selection during barley domestication. The extent of root proliferation has been linked with traits such as above-ground biomass, so selection for particular *cis*-regulatory variants of *HvWAK1* would be consistent with the hypothesis that seed size increases during domestication were the indirect consequence of selection for plants with increased growth vigour.

Introduction

The transition from hunting-gathering to agriculture in the Fertile Crescent of southwest Asia was to a large extent based on the domestication of three cereal crops, einkorn wheat (*Triticum monococcum* L.), emmer wheat (*T. turgidum* L. subsp. *dicoccum* [Schrank ex Schübl.] Thell.) and barley (*Hordeum vulgare* L.) [1]. Although the pace and biogeography of cereal domestication is still disputed [2–4], the process resulted in crop plants that display phenotypic

differences compared with their wild ancestors. These features are collectively referred to as the domestication syndrome [5,6], and include traits such as loss of the natural seed dispersal mechanisms and aids, insensitivity to environmental cues that inhibit germination and/or flowering, and an increase in size of individual seeds and/or the overall yield of grain per plant [6,7].

The domestication traits are assumed to have been selected by early cultivation practices, countering their low or zero adaptive advantage in wild populations [8], but the nature of the selective pressures that resulted in these evolutionary changes, and the role of conscious human agency in driving those changes, is the subject of intense debate [4, 9–13]. The increase in seed size that occurs during domestication provides an example. The size increase could have resulted from selection for large seeds, intended to provide greater food yield [14,15], or possibly because these were easier to handle [16]. Alternatively, cultivation practices might have involved burial of seeds to lower depths than encountered in the natural ecosystem, resulting in selection of seeds that had sufficient stored resource to support growth of the seedling during the increased period that would elapse before its emergence from the soil [14,17]. However, the first of these scenarios appears inconsistent with the relatively slow rate at which seed size increases over time, as indicated by the archaeobotanical record for at least some early farming sites [18], and the second is contradicted by experimental studies which have shown that, at least with the majority of legume species, the increased size of domesticated seeds compared with wild ones has no significant effect on the ability of seedlings to emerge after sowing at different depths [19]. An alternative explanation is that seed size increased as a pleiotropic effect of domestication, perhaps linked to the selection of plants with more vigorous growth. This hypothesis is supported by the observation that compared to their wild species, many domesticates, including southwest Asian cereals, have greater aboveground biomass with larger leaves and hence enhanced photosynthetic capabilities, as well as tall canopies and high leaf nitrogen contents [20–22]. Selection for growth vigour could also explain why the seeds of vegetables such as beet and carrot increase as a result of domestication, despite these crops being harvested for their roots or leaves rather than their seeds [23].

The preserved plant material at early agricultural sites mainly comprises seeds and, for cereals, intact spikelets and spikelet fragments such as glume bases [1,24], none of which provide information on the overall architecture or growth rate of the plants from which they were derived. The archaeobotanical record therefore does not allow us to address the hypothesis that early farmers selected for cereals with increased growth vigour. An alternative approach is to examine the sequence variability of genes influencing growth properties in living examples of the wild and domesticated versions of a crop, and then using the pattern of variability to assess whether particular haplotypes of those genes might have been selected during the domestication process. In barley, this approach has been successful in understanding selection for phenotypes such as flowering time [25] and seed shattering [26]. Using gene diversity to study selection for growth properties is more challenging because the genetic basis to growth vigour is a complex trait, but some individual genes that affect relevant phenotypes are known. An example in barley is *HvWAK1*, which codes for a member of the cell wall-associated kinase (WAK) group of receptor-like proteins. WAK proteins are located in plant cell walls and regulate disparate events including cell expansion and response to pathogens [27]. *HvWAK1* is expressed in germinating embryos and barley roots up to at least 28 days after planting, with no detectable expression in other tissues such as shoots, leaves, inflorescences and developing grain [28,29], and inactivation of *HvWAK1* results in a significant decrease in the rate of root elongation [28]. These expression and functional studies therefore indicate that the HvWAK1 protein plays a role in the control of root development, but is not directly involved in seed development. The extent of root proliferation has been linked with traits such as shoot growth

rate, plant height and biomass production in rice and wheat [30,31], and hence can be used as a proxy for growth vigour. To begin to address the hypothesis that the increase in seed size that accompanied barley domestication was the indirect effect of selection by early farmers for plants with more vigorous growth, we examined the diversity of *HvWAK1* in an extensive range of georeferenced wild barley accessions and domesticated barley landraces.

Materials and methods

Barley accessions

The study material was 200 barley landraces (*H. vulgare* L.) and 220 wild barley accessions (*H. spontaneum* [K. Koch] Thell.) (S1 Table) obtained from the United States Department of Agriculture–Agricultural Research Service (USDA-ARS) Small Grains Collection (NSGC). Seeds were germinated at room temperature (c.22˚C) in Petri dishes in hydroponic conditions. When coleoptiles emerged, the seeds were transferred to moist filter paper and seedlings grown until 21 days old. Fresh leaf material was then collected and DNA extracted using the ISOLATE II Plant DNA kit (Bioline).

DNA sequencing

A 3655 bp sequence comprising the entire *HvWAK1* gene (barleyGenes sequence MLOC_ 68187.1, EnsemblPlants locus HORVU5Hr1G087560|chr5H:578907334–57891028) along with the immediate upstream and downstream regions (Fig 1) was amplified as four overlapping fragments (amp1 primers: forward 5´-GGTGGCATTGTCTTCATGC-3´, reverse 5´-GATC CGGGAATCGGTCAG-3´, 1302 bp product, annealing temperature 68˚C; amp2 primers: forward 5´-AGGCATGAGTACGTCCAGCTA-3´, reverse 5´-AATGTATGGGTTGCCATCGT-3´, 997 bp, 68˚C; amp3 primers: forward 5´-GCGTGAGCTACAAGCACAAC-3´, reverse 5´-GCA TCAACTTCAAGGCAACA-3´, 1238 bp, 68˚C; amp4 primers: forward 5´-TGGTATTCTCCTC ATGGTGATTC-3´, reverse 5´-GATGCAGCGTACAAGCATTC-3´, 1105 bp, 67˚C). Polymerase chain reactions (PCRs) were carried out in a LightCycler480 (Roche) in 20 μl reaction volumes comprising 100 ng DNA extract, 1x SensiFAST SYBR No-ROX PCR master mix (Bioline), 100 nM forward primer, 100 nM reverse primer and PCR grade water. Cycling parameters were: 95˚C for 5 min; followed by 35 cycles of 30 s at 95˚C, 30 s at the annealing temperature, 60 s at 72˚C. Product formation was assayed using the SYBR Green I/HRM Dye detection format (465 nm excitation, 510 nm emission), and melting data were obtained by first cooling the product to 55˚C for 30 s and then heating to 99˚C with five data acquisitions/˚C. Melting peaks were obtained by plotting–(δF/δT) against temperature. PCR products were purified with the High Pure PCR Product Purification Kit (Roche) and sequenced using the BigDye Terminator v3.1 kit chemistry (Applied Biosystems). Standard sequencing reactions of 20 μl comprised 40 ng PCR product, 1x BigDye sequencing buffer, 0.125x BigDye reaction mix, 4 pmoles primer and Ultra-Pure DNase/RNase-free distilled water. To avoid early signal loss when sequencing difficult regions (high GC/GT/G content, small hairpins or other secondary structures) a modified reaction of 20.05 μl was used, comprising 40 ng PCR product, 1x BigDye sequencing buffer, 0.125x BigDye v3.1 reaction mix, 0.0625x dGTP BigDye v3.0 reaction mix, 4 pmoles primer, 0.95 M betaine (Sigma), 5% (v/v) dimethyl sulfoxide (Sigma), UltraPure DNase/RNase-free distilled water. Cycling parameters were: 2 min at 96˚C; 35 cycles of 40 s at 96˚C, 15 s at 50˚C, 4 min at 60˚C; with products held at 4˚C before purification (Agencourt CleanSEQ; Beckman Coulter) and reading of paired-end sequences by capillary electrophoresis in a 3730 DNA Analyser (Applied Biosystems).

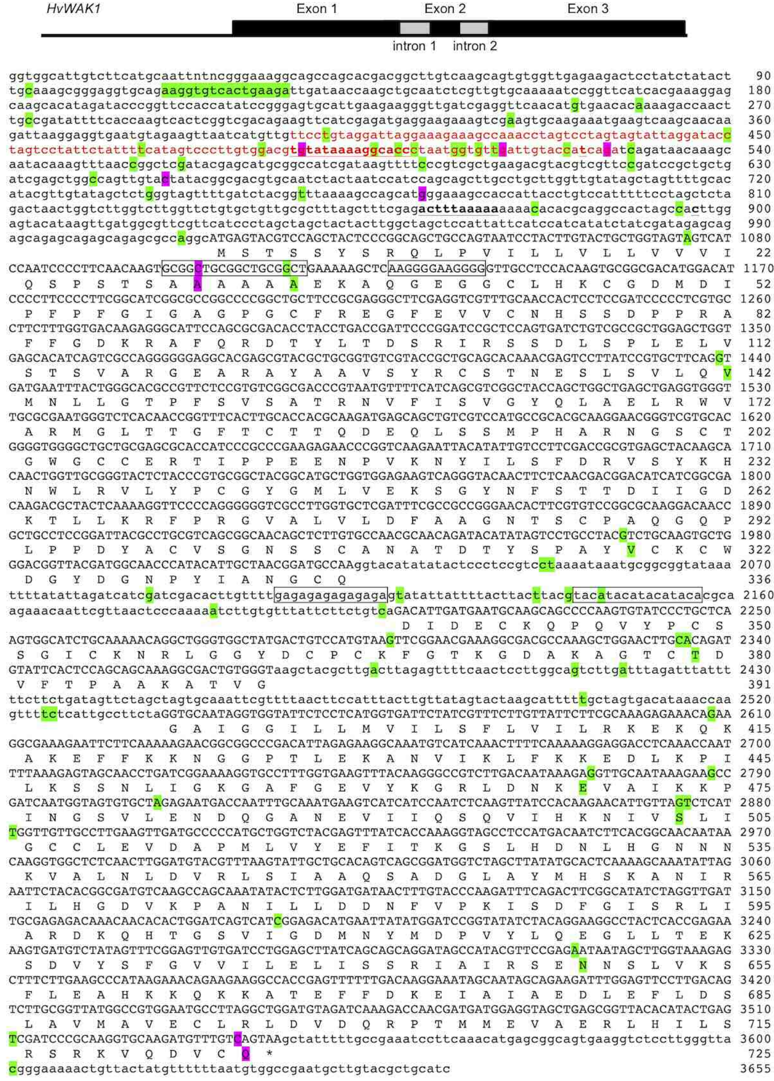


Fig 1. Consensus sequence of the *HvWAK1* gene and adjacent regions from 200 barley landraces and 220 wild accessions. The coding sequence is shown in upper case, and the sequence upstream of the initiation codon, the introns, and the sequence downstream of the termination codon in lower case. SNPs and the indel are highlighted in green, or purple for those seven SNPs that are invariant among landraces. The four STRs are boxed. In the upstream sequence, the region with closest similarity with the HORPIA–2 retrotransposon (see Results) is shown in red typeface, and the two predicted TATA boxes and their associated transcription start sites are shown in bold underlined.

https://doi.org/10.1371/journal.pone.0218526.g001

Data analysis

HvWAK1 sequences for individual barley accessions were assembled from overlapping reads using Geneious version R10 (https://www.geneious.com, [32]). Multiple alignments were generated by the ClustalW, Muscle and Mafft programs and single nucleotide polymorphisms (SNPs) identified using the prediction software in Geneious at a maximum variant *p*-value of 10^{-6} and minimum sequence coverage of 250. Alignments of the *HvWAK1* consensus with other sequences was carried out with Clustal Omega [33] online at EMBL-EBI. Median joining haplotype networks were generated using Network 4 [34] and PopART [35]. Protein secondary structures were predicted using the garnier tool of EMBOSS [36], operated as a Geneious plug-in, and sequence features within the *cis*-regulatory region were identified using the

transcription factor binding site function of PlantPAN 2.0 [37] and the TATA-TSS search function of TSSPlant [38], the latter online at Softberry (http://www.softberry.com/). Geographical distribution maps were plotted using ArcMap 10.2.1 of ArcGIS (ESRI. ArcGIS Desktop: Release 10. Redlands, CA: Environmental Systems Research Institute 2011).

Results

Based on the published sequence of *HvWAK1* [28], we designed primers to amplify and sequence the gene in 200 landraces and 220 wild accessions of barley. The resulting consensus sequence, which differs in places from the published variant is 3655 bp, comprising 1016 bp upstream of the initiation codon, 2528 bp of coding sequence and introns, and 111 bp downsteam of the termination codon (Fig 1). The individual sequences displayed extensive variability (S2 Table) with 69 SNPs and one indel identified using the stringency parameters described in Materials and Methods. In addition, there are four short tandem repeats (STRs), two in the first exon of the coding sequence and two in the first intron.

For 137 accessions (82 landraces and 55 wild barleys) complete sequences were obtained for all 69 SNPs, the indel and the four STRs, with no missing data. When variants at the SNPs and indel were taken into account, these sequences fell into 50 haplotypes (Table 1, S3 and S4 Tables). A single major haplotype included 48 landraces and 2 wild accessions, eleven other haplotypes had 2–12 members, and the remaining 38 were singletons comprising just one accession each. Inclusion of the STRs in the haplotype analysis did not result in significant changes: each of the 50 haplotypes was monoallelic for STRs 1, 2 and 4, and 47 haplotypes were also monoallelic for STR3. The exceptions were haplotypes 1, 2 and 5, each of which included two variants of the $(GA)_n$ STR3 located within intron 1. Network analysis (Fig 2) revealed a complex evolutionary relationship between the 50 haplotypes, but with landraces largely restricted to two areas of the network, the first of these regions comprising major haplotype 1 and the singleton landrace haplotypes 13, 30 and 43, and the second made up of a starburst topology centered on haplotype 4 with haplotypes 8, 18, 19, 25, 26, 28, 32, 37, and 41 as satellites.

Nineteen of the SNPs are located within the coding sequence of the gene, nine of these SNPs resulting in amino acid changes in the predicted translation product, another nine giving

Table 1. *HvWAK1* haplotypes.

Haplotype	Number of accessions		
	Wild	Landraces	Total
1	2	48	50
2	8	4	12
3	7	0	7
4	0	7	7
5	5	0	5
6	4	0	4
7	3	0	3
8	0	3	3
9	2	0	2
10	0	2	2
11	2	0	2
12	0	2	2
13–50[a]	22	16	38

[a]Singleton haplotypes with one member each.

https://doi.org/10.1371/journal.pone.0218526.t001

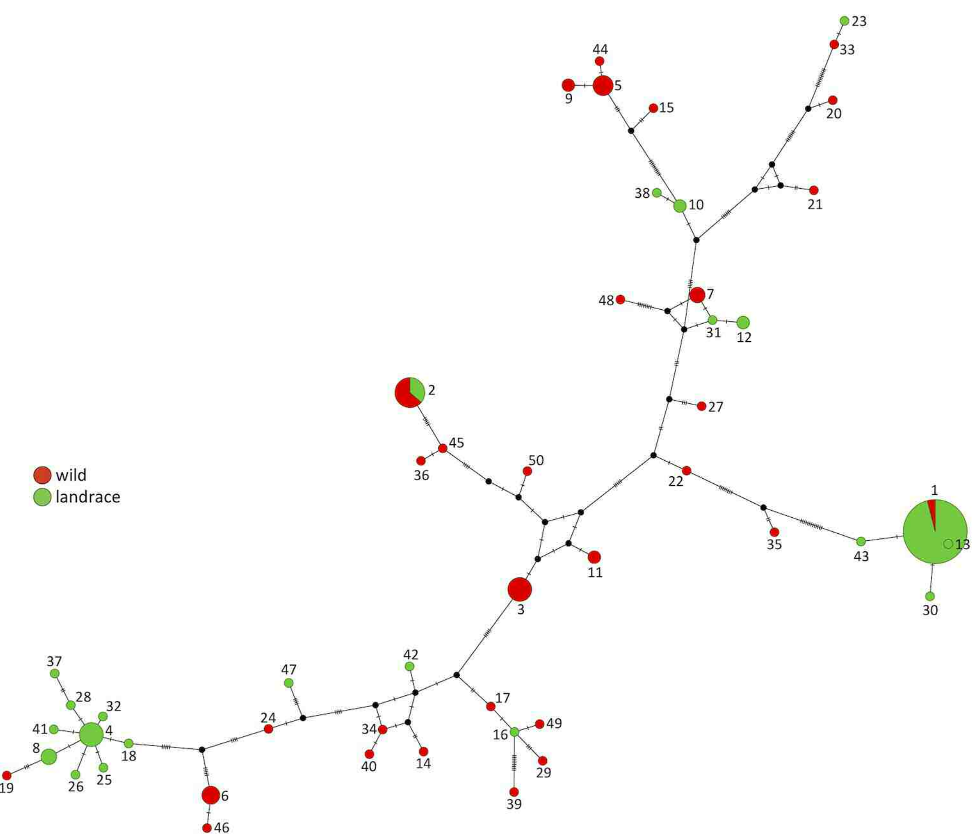


Fig 2. Network displaying the relationships between the 50 haplotypes of *HvWAK1*. Node sizes are proportional to numbers of accessions and empty nodes are shown as black dots. Short dashes on the edges indicate the number of point changes between pairs of nodes. The proportion of wild accessions and landraces for each haplotype are shown in red and green, respectively.

https://doi.org/10.1371/journal.pone.0218526.g002

synonymous codon changes, and one introducing a premature termination codon that shortens the predicted translation product by one amino acid (Table 2). The two STRs in the first exon of the coding sequence also result in amino acid sequence changes (S5 Table). The first of these STRs is $(\mathtt{GCGGCT})_{2-4}$, specifying a series of 4–8 alanines, although the second of these alanines is valine in some accessions due to the presence of SNP35, and the fifth of can be proline due to SNP36. The second STR is $(\mathtt{AAGGGG})_{2-4}$, which is translated into glutamine-glycine followed by 1–3 repeats of glutamic acid-glycine, followed by a single glycine.

Complete data for the nine nonsynonymous SNPs were available for 298 accessions (154 wild accessions and 144 landraces). The amino acid substitutions resulting from these SNPs combine to produce 28 protein variants (Table 3 and S6 Table), of which 16 are unique to the wild population, two are unique to landraces (these two variants comprising only three accessions) and the remaining ten are shared by wild accessions and landraces.

The nonsynonymous amino acid changes include three that result in significant changes in chemical properties [39]. These are: the serine to arginine substitution at amino acid position 503 of the predicted translation product, caused by SNP63; the conversion of the same serine to lysine due to the presence of SNPs 62 and 63 in the codon specifying amino acid 503; and the asparagine to histidine change at position 649, caused by SNP66 (see Table 2). However, secondary structure predictions do not suggest that these or any other of the substitutions occurring in the coding sequence would have a significant impact on the structure and, by

Table 2. SNPs in the coding region of the *HvWAK1* gene.

SNP number	Nucleotide position	Amino acid position	Codon sequence		Amino acid identity	
			Consensus	Variant	Consensus	Variant
SNP34	1075	20	GTA	GTC	val	val
SNP35	1104	30	GCT	GTT	ala	val
SNP36	1115	34	GCT	CCT	ala	pro
SNP37	1439	142	GTG	CTG	val	leu
SNP38	1967	318	GTC	ATC	val	ile
SNP49	2298	366	AAG	AAA	lys	lys
SNP50	2334	378	TGC	TGT	cys	cys
SNP51	2335	379	ACA	TCA	thr	ser
SNP58	2608	414	CAG	CAA	gln	gln
SNP59	2773	469	GAG	GAT	glu	asp
SNP60	2788	474	AAG	AAA	lys	lys
SNP61	2809	481	CTA	CTC	leu	leu
SNP62	2874	503[a]	AGT	AAT	ser	asn
SNP63	2875	503[a]	AGT	AGG	ser	arg
SNP64	2881	505	ATT	ATC	ile	ile
SNP65	3184	606	ATC	ATT	ile	ile
SNP66	3311	649	AAT	CAT	asn	his
SNP67	3511	715	AGT	AGC	ser	ser
SNP68	3539	725	CAG	TAG	gln	stop

[a]When variants are present at both SNP62 and SNP63, amino acid position 503 is converted to AAG = lys.

https://doi.org/10.1371/journal.pone.0218526.t002

inference, function of the *HvWAK1* protein (S1 Fig). The three possible substitutions at amino acid position 503 (serine to asparagine, arginine or lysine) are predicted to affect the precise conformation of a short series of turns, but substitutions elsewhere in the protein either have no effect on the predicted secondary structure, or change the lengths of the predicted structural elements in minor ways.

Each of the 69 SNPs, as well as the indel and each of the four STRs, displays variation in the wild population (S7 Table). In contrast, seven of the SNPs are invariant in the landraces that were sequenced, each displaying the major allele present in wild plants. Two of the invariant SNPs are located in the coding sequence, with the consequence that all of the landraces have alanine at amino acid position 30, and none of the landraces have the premature termination codon. The other five invariant SNPs are numbers 10, 18, 20, 27 and 30, located upstream of the coding region at positions 487, 512, 525, 650 and 772, respectively. Two potential recognition sites for the TATA-binding protein were identified in this region by the promoter analysis function of PlantPAN2.0, and both of these sites were independently identified as TATA boxes by TSSPlant (see Fig 1). The first of these TATA boxes lies at positions 486–500, with a predicted transcription start site (TSS) at position 522, giving an mRNA with a 5´-untranslated region (5´-UTR) of 494 nucleotides. This TATA box and its TSS both lie within a region of the *HvWAK1* sequence that shows similarity with part of the 3´ long terminal repeat (3´-LTR) of the barley *copia*-like retrotransposon HORPIA–2 [40] (S2 Fig). The second TATA box is predicted at positions 862–871, with the TSS at position 896 and a 5´-UTR of 120 nucleotides. When the sequence containing the minor variants of SNPs 10, 18, 20, 27 and 30 (i.e. the sequence not found in landraces) is used as the query, the TSSPlant prediction tool assigns a higher score to the upstream TATA box (Table 4). Replacing the SNPs with their major variants (i.e. the sequence found in landraces), has no effect on the likelihood scores for the

Table 3. Variants of the *HvWAK1* protein sequence.

Variant number	Amino acid sequence[a]	Number of accessions		
		Wild	Landraces	Total
1	AAVVTESNQ	50	26	76
2	A-VVTERHQ	4	69	73
3	AAVITDSNQ	4	27	31
4	AAVVSESNQ	17	–	17
5	A-VVTESHQ	12	4	16
6	AAVVTEKNQ	13	–	13
7	APVVTESNQ	9	1	10
8	AALVTEKN*	9	–	9
9	V-VVTENNQ	3	4	7
10	A-VVTESNQ	7	–	7
11	V-VITDSNQ	1	5	6
12	AALVTESNQ	3	3	6
13	AAVITESNQ	5	–	5
14	AAVVTDSNQ	3	–	3
15	A-VVTEKNQ	1	1	2
16	AAVVTESN*	2	–	2
17	AAVVTESNQ	–	2	2
18	A-VITDSNQ	1	1	2
19	V-VVTESNQ	2	–	2
20	A-LVTDSNQ	1	–	1
21	A-VVTDSNQ	1	–	1
22	AALITDSNQ	1	–	1
23	AALVTEKNQ	1	–	1
24	AAVITDNNQ	1	–	1
25	AAVITENNQ	1	–	1
26	APVVTEKNQ	1	–	1
27	APVVTERNQ	–	1	1
28	V-VITESNQ	1	–	1

[a]Amino acid identities at positions 30, 34, 142, 318, 379, 469, 503, 649 and 725, given in the IUPAC one-letter code.

* An asterisk indicates a stop codon at position 725.

https://doi.org/10.1371/journal.pone.0218526.t003

downstream TATA-TSS combination, but decreases the likelihood scores for the TATA box at positions 486–500 and the TSS at position 522. Examination of each SNP in turn reveals that SNPs 10 and 20 have the greatest impact: replacement of the major variant of SNP10 increases the score for the upstream TATA box to the same figure as obtained with minor versions of all five SNPs, and replacement of the major variant of SNP20 has a similar effect on the TSS. Altogether, 112 wild accessions have the minor variant at one or more of SNPs 10, 18, 20, 27 and 30 (S8 Table), these accessions distributed throughout the entire geographical range of wild barley (S3A Fig). These 111 accessions include 25 with the minor variant of SNP10, distributed in the southern Levant and southeast Turkey, with an outlier to the east (S3B Fig), and 50 with one of the minor variants of SNP20, which have a similar distribution (S3C Fig).

Discussion

We sequenced the *HvWAK1* gene–coding for a wall-associated kinase that influences root proliferation–in a large set of barley landraces and wild accessions. The gene displayed extensive

Table 4. Results of promoter analysis of the *HvWAK1* upstream region.

	Sequence					TSSPlant likelihood scores			
Version	SNP10	SNP18	SNP20	SNP27	SNP30	TATA-box 486–500	TSS 522	TATA-box 862–871	TSS 896
Minor variants[a]	A	A	A	T	A	7.5504	1.9686	5.2759	1.9787
Minor variants[a]	A	A	C	T	A	7.5504	1.9712	5.2759	1.9787
Major variants	G	G	G	C	G	6.6702	1.8783	5.2759	1.9776
Minor SNP10	A	G	G	C	G	7.5504	1.8930	5.2759	1.9776
Minor SNP18	G	A	G	C	G	6.6702	1.8927	5.2759	1.9776
Minor SNP20[a]	G	G	A	C	G	6.6702	1.9588	5.2759	1.9776
Minor SNP20[a]	G	G	C	C	G	6.6702	1.9660	5.2759	1.9776
Minor SNP27	G	G	G	T	G	6.6702	1.8783	5.2759	1.9776
Minor SNP30	G	G	G	C	A	6.6702	1.8783	5.2759	1.9787

[a]SNP20 has two variant forms, A present in 44 wild accessions, and C in six accessions.

https://doi.org/10.1371/journal.pone.0218526.t004

diversity with nineteen SNPs and two STRs in the coding sequence and an additional 50 SNPs, two STRs and an indel in the two introns and the regions immediately upstream and downstream of the coding sequence. Despite this complexity, the diversity was structured into 50 haplotypes that could be displayed as a median joining network with minimal character conflict (see Fig 2). Landraces were distributed non-randomly within this network: of the 82 landraces with complete data for every variable site, 51 were located at a peripheral position comprising haplotype 1 and the singleton haplotypes 13, 30 and 43, and another 17 landraces were located at a second peripheral position, in a starburst comprising haplotype 4 as the principal node and eight other landrace haplotypes as satellites. The events occurring during and/or after domestication have therefore resulted in some parts of the wild diversity *of HvWAK1* being lost and segments of the remaining diversity proliferating in the landrace population, consistent with the expectations of a domestication bottleneck [41] or the more gradual loss of diversity following domestication suggested by examination of time-series of ancient crop genomes [42,43].

The 19 SNPs in the coding region include nine that result in nonsynonymous codon changes (two of these changes in the same codon, meaning that there are ten nonsynonymous amino acid substitutions in total), and one that introduces a premature stop codon, one position upstream of the stop codon used in the majority of accessions. A total of 26 protein variants are present in the wild population, but the reduction in diversity that accompanied domestication means that only ten of these variants are seen in landraces. The majority (122/144) of the landraces belong to variants 1–3, which have dissimilar amino acid combinations at the nine nonsynonymous sites, with only four out of eight amino acid identities (alanine at position 30, valine at 142, threonine at 379 and glutamine at 725 –see Table 3). Those three amino acids are also present in fifteen other protein variants, nine of which are not seen in landraces, suggesting that the reduced landrace diversity is not due to selection for particular protein variants. Additional evidence against selection for protein variants is provided by predictions of the secondary structures that would form in the protein, this analysis (S1 Fig) indicating that the amino acid substitutions resulting from the nonsynonymous SNPs would not have a significant effect on the structures of the protein variants. Secondary structure prediction is of limited utility as it has only c.65% accuracy, and cannot reveal changes in longer range interactions within a protein, so the possibility remains that one or more of the amino acid substitutions has a significant effect on protein function, but the available information provides no evidence for this.

To gain further insight into possible functional differences that might arise as a result of the diversity of the *HvWAK1* sequence, we compared the frequencies of the major and minor variants of every SNP in the wild accessions and landraces (S7 Table). This analysis revealed that for seven SNPs the major variant is fixed in the landrace population. Five of these 'invariant' SNPs lie upstream of the *HvWAK1* initiation codon, in the *cis*-regulatory region predicted to contain two TATA boxes and their associated transcription start sites. The upstream of the two TATA-TSS combinations lies within a sequence segment that is highly similar to the 3 ′-LTR of the barley retrotransposon HORPIA-2, and may have been captured by the *HvWAK1* gene from an adjacent retrotransposon at some time in the past. Analysis of the region containing the TATA boxes and TSSs with the TSSPlant predictive software [37] suggests that fixation of SNPs 10 and 20 (at positions 487 and 525) decreases the likelihood of the upstream TATA box and TSS being used to promote transcription of *HvWAK1*. As well as the five invariant SNPs, several other SNPs in the *cis*-regulatory region have significantly different major/minor variant frequencies in landraces compared with the wild accessions. In particular, the minor variant of SNP11, located within the upstream TATA box, is present in only one landrace, and the minor variants of SNPs 13 and 17, which are between this TATA box and its TSS, are much more frequent in landraces compared with wild accessions. The latter is also the case for two additional SNPs, numbers 25 and 28, which are located between the two predicted TATA boxes.

Comparison of the sequence diversity of *HvWAK1* in wild barley accessions and landraces therefore suggests that the events during and/or after domestication might have resulted in selection of particular variants of the *cis*-regulatory region. The alternative, that fixation of the 'invariant' SNPs was due to stochastic effects resulting from the geographical distribution of the SNP variants in the wild population, is less likely, as wild accessions with the minor versions of the invariant SNPs are found throughout the Fertile Crescent, with the minor variants of SNPs 10 and 20, which according to the predictive software have the greatest impact on TATA-TSS usage, located in both the southern Levant and the Syria/southeast Turkey border, areas proposed as the sites of barley domestication [26]. It therefore seems less likely that the invariant SNPs became fixed as a result of chance sampling only of wild plants lacking the minor versions of these SNPs during the initial cultivation of barley.

The suggestion that the phenotypic features that characterize the domesticated version of a plant result not only from changes in protein structures but also changes in the expression patterns of key genes [44], has been confirmed by transcriptome studies of crops as diverse as maize [45–47], tomato [48,49], cotton [50,51], soybean [52] and carrot [53]. It now accepted that human selection during plant domestication acted on *cis*-regulatory regions as well as the coding sequences of genes [54]. The evidence that we present in this paper indicates that the *cis*-regulatory region of *HvWAK1* might have been subject to selection during the domestication of barley. Confirmation of this hypothesis would require functional studies that tested whether fixation of one or more of SNPs 10, 18, 20, 27 and 30 results in changes in the expression pattern of *HvWAK1* and/or can be associated with phenotypic changes such an alteration in the dynamics or extent of root proliferation. Such studies might be challenging due to the need to address the possibility that the functional effects of fixation of these SNPs might only be apparent under certain cultivation conditions.

As *HvWAK1* plays a role in root proliferation [28], selection for particular *cis*-regulatory variants of this gene would be consistent with the hypothesis that plants with increased growth vigour had a selective advantage under cultivation, possibly because they were able to outcompete plants with lower biomass and photosynthetic capability, and/or because farmers displayed a preference for larger and more vigorous plants. Our results therefore provide preliminary support for the proposal [19,23] that the increase in seed size that occurred during barley

domestication was the indirect consequence of selection for increased growth vigour, rather than a direct consequence of the selection for large seeds.

Supporting information

S1 Fig. Secondary structure predictions for the barley *HvWAK1* protein. (A) The consensus amino acid sequence. (B) The amino acid sequence containing all variants, with asparagine at position 503. (C) Comparison of the predicted structures for the region surrounding each of the position 503 variants. Structural codes: pink barrel, α-helix; yellow arrow, β-strand; blue hooked arrow, turn; grey wavy line, coil.
(TIFF)

S2 Fig. Alignment between the *HvWAK1* upstream sequence and the 3′-LTR of the barley retrotransposon HORPIA-2. Nucleotide identities are indicated by asterisks and the two predicted TATA boxes and their associated transcription start sites are highlighted in yellow. The *HvWAK1* sequence is numbered as in Fig 1. The HORPIA–2 sequence is taken from positions 35775–36099 of Genbank entry AH014393.2. Upstream and downstream of these positions in AH014393.2 there is no significant similarity with the *HvWAK1* sequence.
(TIFF)

S3 Fig. Locations of the collection sites for wild accessions. (A) All wild accessions with the minor variant at one or more of SNPs 10, 18, 20, 27 and 30; (B) accessions with the minor variant at SNP10; (C) accessions with the minor variant at SNP20.
(TIFF)

S1 Table. Barley landraces and wild accessions used in this study.
(XLSX)

S2 Table. SNPs and other variants in the *HvWAK1* sequence.
(XLSX)

S3 Table. Variant identities for all haplotypes.
(XLSX)

S4 Table. Haplotype memberships.
(XLSX)

S5 Table. STR variants in the *HvWAK1* coding sequence.
(XLSX)

S6 Table. Protein variant memberships.
(XLSX)

S7 Table. Allele frequencies at the variable positions in the *HvWAK1* sequence.
(XLSX)

S8 Table. Wild accessions with minor variants at SNPs 10, 18, 20, 27 or 30.
(XLSX)

Acknowledgments

We thank Harold Bockelman (NSGC) for provision of barley accessions, Paul Fullwood, Fraser Combe and Graeme Fox (University of Manchester) for DNA sequencing, and Sandra Kennedy (University of Manchester) for assistance with data handling.

Author Contributions

Conceptualization: Glynis Jones, Terence A. Brown.

Formal analysis: Terence A. Brown.

Funding acquisition: Glynis Jones, Terence A. Brown.

Investigation: Beata I. Czajkowska.

Methodology: Beata I. Czajkowska.

Project administration: Terence A. Brown.

Supervision: Terence A. Brown.

Writing – original draft: Beata I. Czajkowska.

Writing – review & editing: Glynis Jones, Terence A. Brown.

References

1. Zohary D, Hopf M, Weiss E. Domestication of plants in the Old World, 4th ed. Oxford: Oxford University Press; 2012.
2. Tanno K, Willcox G. How fast was wild wheat domesticated? Science. 2006; 311: 1886. https://doi.org/10.1126/science.1124635 PMID: 16574859
3. Brown TA, Jones MK, Powell W, Allaby RG. The complex origins of domesticated crops in the Fertile Crescent. Trends Ecol Evol. 2009; 24; 103–109. https://doi.org/10.1016/j.tree.2008.09.008 PMID: 19100651
4. Abbo S, Gopher A. Near Eastern plant domestication: a history of thought. Trends Plant Sci. 2017; 22: 491–511. https://doi.org/10.1016/j.tplants.2017.03.010 PMID: 28434795
5. Hammer K. Das Domestikationssyndrom. Kulturpflanze. 1984; 32: 11–34.
6. Fuller DQ. Contrasting patterns in crop domestication and domestication rates: recent archaeobotanical insights from the Old World. Ann Bot. 2007; 100: 903–924. https://doi.org/10.1093/aob/mcm048 PMID: 17495986
7. Gepts P. Crop domestication as a long-term selection experiment. Plant Breed Rev. 2004; 24: 1–44.
8. Zohary D. The mode of domestication of the founder crops of Southwest Asian agriculture. In: Harris DR, editor. The origins and spread of Agriculture and pastoralism in Eurasia. London: UCL Press; 1996. pp. 142–158.
9. Abbo S, Lev-Yadun S, Gopher A. Agricultural origins: centers and non-centers; a Near Eastern reappraisal. Crit Rev Plant Sci. 2010; 29: 317–328.
10. Abbo S, Lev-Yadun S, Gopher A. The 'human mind' as a common denominator in plant domestication. J Exp Bot. 2014; 65: 1917–1920. https://doi.org/10.1093/jxb/eru068 PMID: 24638899
11. Allaby R. Integrating the processes in the evolutionary system of domestication. J Exp Bot. 2010; 61: 935–944. https://doi.org/10.1093/jxb/erp382 PMID: 20067978
12. Fuller DQ, Allaby RG, Stevens C. Domestication as innovation: the entanglement of techniques, technology and chance in the domestication of cereal crops. World Archaeol. 2010; 42: 13–28.
13. Brown TA. The role of humans in a protracted transition from hunting-gathering to plant domestication in the Fertile Crescent. Front Plant Sci. 2018; 9: 1287. https://doi.org/10.3389/fpls.2018.01287 PMID: 30245699
14. Harlan JR, Wet JMJD, Price EG. Comparative evolution of cereals. Evolution. 1973; 27: 311–325. https://doi.org/10.1111/j.1558-5646.1973.tb00676.x PMID: 28564784
15. Harlan JR. Crops and Man, 2nd ed. Madison: American Society of Agronomy and Crop Science Society of America; 1992.
16. Abbo S, Van-Oss RP, Gopher A, Saranga Y, Ofner R, Peleg Z. Plant domestication versus crop evolution: a conceptual framework for cereals and grain legumes. Trends Plant Sci. 2014; 19: 351–360. https://doi.org/10.1016/j.tplants.2013.12.002 PMID: 24398119
17. Zohary D. Unconscious selection and the evolution of domesticated plants. Econ Bot. 2004; 58: 5–10.

18. Purugganan MD, Fuller DQ. Archaeological data reveal slow rates of evolution during plant domestication. Evolution. 2011; 65: 171–183. https://doi.org/10.1111/j.1558-5646.2010.01093.x PMID: 20666839

19. Kluyver TA, Charles M, Jones G, Rees M, Osborne CP. Did greater burial depth increase the seed size of domesticated legumes? J. Exp. Bot. 2013; 64: 4101–4108. https://doi.org/10.1093/jxb/ert304 PMID: 24058143

20. Milla R, Matesanz S. Growing larger with domestication: a matter of physiology, morphology or allocation? Plant Biol. 2017; 19: 475–483. https://doi.org/10.1111/plb.12545 PMID: 28075047

21. Preece C, Livarda A, Christin P-A, Wallace M, Martin G, Charles M, et al. How did the domestication of Fertile Crescent grain crops increase their yields? Funct Ecol. 2017; 31: 387–397. https://doi.org/10.1111/1365-2435.12760 PMID: 28286354

22. Milla R, Bastida JM, Turcotte MM, Jones G, Violle C, Osborne CP, et al. Phylogenetic patterns and phenotypic profiles of the species of plants and mammals farmed for food. Nature Ecol Evol. 2018; 2: 1808–1817.

23. Kluyver TA, Jones G, Pujol B, Bennett C, Mockford EJ, Charles M, et al. Unconscious selection drove seed enlargement in vegetable crops. Evol Lett. 2017; 1–2: 64–72.

24. Tanno K, Willcox G. Distinguishing wild and domestic wheat and barley spikelets from early Holocene sites in the Near East. Veg Hist Archaeobot. 2012; 21: 107–115.

25. Jones H, Leigh FJ, Mackay I, Bower MA, Smith LM, Charles MP, et al. Population-based resequencing reveals that the flowering time adaptation of cultivated barley originated east of the Fertile Crescent. Mol Biol Evol. 2008; 25: 2211–2219. https://doi.org/10.1093/molbev/msn167 PMID: 18669581

26. Pourkheirandish M, Hensel G, Kilian B, Senthil N, Chen G, Sameri M, et al. Evolution of the grain dispersal system of barley. Cell. 2015; 162: 527–539. https://doi.org/10.1016/j.cell.2015.07.002 PMID: 26232223

27. Kohorn BD. Cell wall-associated kinases and pectin perception. J. Exp Bot. 2015; 67: 489–494. https://doi.org/10.1093/jxb/erv467 PMID: 26507892

28. Kaur R, Singh K, Singh J. A root-specific wall-associated kinase gene, *HvWAK1*, regulates root growth and is highly divergent in barley and other cereals. Funct Integr Genomics. 2013; 13: 167–177. https://doi.org/10.1007/s10142-013-0310-y PMID: 23443578

29. BARLEX Barley Genome Browser [accessed 2019 May 10]. Available from: https://apex.ipk-gatersleben.de/apex/f?p=284:45:::NO::P45_GENE_NAME:HORVU5Hr1G087560.1.

30. Huang M, Chen J, Cao F, Jiang L, Zou Y. Root morphology was Improved in a late-stage vigor super rice cultivar. PLoS ONE. 2015; 10(11): e0142977. https://doi.org/10.1371/journal.pone.0142977 PMID: 26566229

31. Colombi T, Walter A. Genetic diversity under soil compaction in wheat: root number as a promising trait for early plant vigor. Front Plant Sci. 2017; 8: 420. https://doi.org/10.3389/fpls.2017.00420 PMID: 28400783

32. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 2012; 28: 1647–1649. https://doi.org/10.1093/bioinformatics/bts199 PMID: 22543367

33. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol. 2011; 7: 539. https://doi.org/10.1038/msb.2011.75 PMID: 21988835

34. Bandelt H-J, Forster P, Röhl A. Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol. 1999; 16: 37–48. https://doi.org/10.1093/oxfordjournals.molbev.a026036 PMID: 10331250

35. Leigh JW, Bryant D. PopART: full-feature software for haplotype network construction. Methods Ecol Evol. 2015; 6: 1110–1116.

36. Rice P, Longden I, Bleasby A. EMBOSS: The European Molecular Biology Open Software Suite. Trends Genet. 2000; 16: 276–277. PMID: 10827456

37. Chow CN, Zheng HQ, Wu NY, Chien CH, Huang HD, Lee TY, et al. PlantPAN 2.0: an update of plant promoter analysis navigator for reconstructing transcriptional regulatory networks in plants. Nucl Acids Res. 2016; 44(D1): D1154–D1160. https://doi.org/10.1093/nar/gkv1035 PMID: 26476450

38. Shahmuradov IA, Umarov RK, Solovyev VV. TSSPlant: a new tool for prediction of plant Pol II promoters. Nucl Acids Res. 2017; 45: e65. https://doi.org/10.1093/nar/gkw1353 PMID: 28082394

39. Sneath PH. Relations between chemical structure and biological activity in peptides. J Theor Biol. 1966; 12: 157–195. PMID: 4291386

40. Schulman AH, Kalendar P. A movable feast: diverse retrotransposons and their contribution to barley genome dynamics. Cytogenet Genome Res. 2017; 110: 598–605.

41. Meyer RS, Purugganan MD. Evolution of crop species: Genetics of domestication and diversification. Nat Rev Genet. 2013; 14: 840–852. https://doi.org/10.1038/nrg3605 PMID: 24240513
42. Allaby RG, Ware RL, Kistler L. A re-evaluation of the domestication bottleneck from archaeogenomic evidence. Evol Appl. 2019; 12: 29–37. https://doi.org/10.1111/eva.12680 PMID: 30622633
43. Brown TA. Is the domestication bottleneck a myth? Nat Plants. 2019; 5: 337–338. https://doi.org/10.1038/s41477-019-0404-1 PMID: 30962526
44. Doebley JF, Gaut BS, Smith BD. The molecular genetics of crop domestication. Cell. 2006; 127: 1309–1321. https://doi.org/10.1016/j.cell.2006.12.006 PMID: 17190597
45. Hufford MB, Xu X, van Heerwaarden J, Pyhäjärvi T, Chia JM, Cartwright RA, et al. Comparative population genomics of maize domestication and improvement. Nature Genet. 2012; 44: 808–811. https://doi.org/10.1038/ng.2309 PMID: 22660546
46. Swanson-Wagner R, Briskine R, Schaefer R, Hufford MB, Ross-Ibarra J, Myers CL, et al. Reshaping of the maize transcriptome by domestication. Proc Natl Acad Sci USA. 2012; 109: 11878–11883. https://doi.org/10.1073/pnas.1201961109 PMID: 22753482
47. Lemmon ZH, Bukowski R, Sun Q, Doebley JF. The role of *cis* regulatory evolution in maize domestication. PLoS Genet. 2014; 10: e1004745. https://doi.org/10.1371/journal.pgen.1004745 PMID: 25375861
48. Koenig D, Jiménez-Gómez JM, Kimura S, Fulop D, Chitwood DH, Headland LR, et al. Comparative transcriptomics reveals patterns of selection in domesticated and wild tomato. Proc Natl Acad Sci USA. 2013; 110: E2655–E2662. https://doi.org/10.1073/pnas.1309606110 PMID: 23803858
49. Sauvage C, Rau A, Aichholz C, Chadoeuf J, Sarah G, Ruiz M, et al.Domestication rewired gene expression and nucleotide diversity patterns in tomato. Plant J. 2017; 91: 631–645. https://doi.org/10.1111/tpj.13592 PMID: 28488328
50. Yoo M-J, Wendel JF. Comparative evolutionary and developmental dynamics of the cotton (*Gossypium hirsutum*) fiber transcriptome. PLoS Genet. 2014; 10: e1004073. https://doi.org/10.1371/journal.pgen.1004073 PMID: 24391525
51. Wang M, Tu L, Lin M, Lin Z, Wang P, Yang Q, et al. Asymmetric subgenome selection and *cis*-regulatory divergence during cotton domestication. Nature Genet. 2017; 49: 579–587. https://doi.org/10.1038/ng.3807 PMID: 28263319
52. Yu C, Qu Z, Zhang Y, Zhang X, Lan T, Adelson DL, et al. Seed weight differences between wild and domesticated soybeans are associated with specific changes in gene expression. Plant Cell Rep. 2017; 36: 1417–1426. https://doi.org/10.1007/s00299-017-2165-5 PMID: 28653111
53. Rong J, Lammers Y, Strasburg JL, Schidlo NS, Ariyurek Y, de Jong TJ, et al. (2014) New insights into domestication of carrot from root transcriptome analyses. BMC Genomics. 2014; 15: 895. https://doi.org/10.1186/1471-2164-15-895 PMID: 25311557
54. Swinnen G, Goosens A, Pauwels L. Lessons from domestication: targeting cis-regulatory elements for crop improvement. Trends Plant Sci. 2016. 21; 506–515. https://doi.org/10.1016/j.tplants.2016.01.014 PMID: 26876195

3.

A discriminatory test for the wheat B and G genomes reveals misclassified accessions of Triticum timopheevii and Triticum turgidum.

Czajkowska BI, Oliveira HR, Brown TA

Published on 10 April 2019
in
PLoS ONE
14(4):e0215175
https://doi.org/10.1371/journal.pone.0215175

RESEARCH ARTICLE

A discriminatory test for the wheat B and G genomes reveals misclassified accessions of *Triticum timopheevii* and *Triticum turgidum*

Beata I. Czajkowska, Hugo R. Oliveira¤, Terence A. Brown*

School of Earth and Environmental Sciences, Manchester Institute of Biotechnology, University of Manchester, Manchester, United Kingdom

¤ Current address: Interdisciplinary Center for Archaeology and Evolution of Human Behaviour, Universidade do Algarve, Faro, Portugal
* terry.brown@manchester.ac.uk

OPEN ACCESS

Citation: Czajkowska BI, Oliveira HR, Brown TA (2019) A discriminatory test for the wheat B and G genomes reveals misclassified accessions of *Triticum timopheevii* and *Triticum turgidum*. PLoS ONE 14(4): e0215175. https://doi.org/10.1371/journal.pone.0215175

Editor: Aimin Zhang, Institute of Genetics and Developmental Biology Chinese Academy of Sciences, CHINA

Received: December 8, 2018

Accepted: March 27, 2019

Published: April 10, 2019

Copyright: © 2019 Czajkowska et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the manuscript and its Supporting Information files.

Funding: This work was supported by TAB 339941 European Research Council, https://erc.europa.eu, no involvement, TAB RPG-2015-043 Leverhulme Trust, https://www.leverhulme.ac.uk, no involvement.

Competing interests: The authors have declared that no competing interests exist.

Abstract

The tetraploid wheat species *Triticum turgidum* and *Triticum timopheevii* are morphologically similar, and misidentification of material collected from the wild is possible. We compared published sequences for the *Ppd-A1*, *Ppd-B1* and *Ppd-G1* genes from multiple accessions of *T. turgidum* and *T. timopheevii* and devised a set of four polymerase chain reactions (PCRs), two specific for *Ppd-B1* and two for *Ppd-G1*. We used these PCRs with 51 accessions of *T. timopheevii* and 20 of *T. turgidum*. Sixty of these accessions gave PCR products consistent with their taxon identifications, but the other eleven accessions gave anomalous results: ten accessions that were classified as *T. turgidum* were identified as *T. timopheevii* by the PCRs, and one *T. timopheevii* accession was typed as *T. turgidum*. We believe that these anomalies are not due to errors in the PCR tests because the results agree with a more comprehensive analysis of genome-wide single nucleotide polymorphisms, which similarly suggest that these eleven accessions have been misclassified. Our results therefore show that the accepted morphological tests for discrimination between *T. turgidum* and *T. timopheevii* might not be entirely robust, but that species identification can be made cheaply and quickly by PCRs directed at the *Ppd-1* gene.

Introduction

Wild and cultivated wheats comprise an allopolyploid complex of diploid (AA genomes), tetraploid (AABB and AAGG) and hexaploid forms (AABBDD and AAAAGG). The AABB species is called *Triticum turgidum* L. and includes wild and cultivated emmer (subsp. *dicoccoides* [Korn. ex Asch. & Graebn.] Thell. and subsp. *dicoccum* [Schrank ex Schübl.] Thell., respectively), both of which are hulled, meaning that the kernels are enclosed in toughened husks called glumes. Additionally, there is a series of cultivated emmer derivatives such as durum wheat (subsp. *durum* [Desf.] Husn.) and rivet wheat (subsp. *turgidum* (Desf.] Husn.), which are called naked or free-threshing wheats because they have thinner glumes that enclose the kernels less tightly. The AAGG tetraploid, *T. timopheevii* (Zhuk.) Zhuk., also has wild (subsp. *armeniacum*

[Jakubz.] Slageren) and domesticated forms (subsp. *timopheevii*), both of which are hulled. *T. turgidum* and *T. timopheevii* can be crossed to produce F_1 progeny (e.g. [1]), but these plants are sterile and the two species are thought to be non-interfertile due to failures in chromosome pairing [2].

The wild versions of *T. turgidum* and *T. timopheevii* have restricted geographical ranges, overlapping in southeast Turkey, northwest Syria and in the mountainous regions of eastern Iraq/western Iran, with *T. turgidum* additionally present in the upper Jordan valley and *T. timopheevii* in the Caucusus [3,4]. Although both species were domesticated by early farmers, only cultivated *T. turgidum* is considered to be a major crop, being grown extensively at Neolithic sites throughout the Fertile Crescent [3,5,6], and forming part of the package of crops whose cultivation spread into Europe, Asia and North Africa [3]. In contrast, *T. timopheevii* is looked on as a secondary crop, being found today only in western Georgia [3], although it has been suggested that the 'new glume wheat', which was grown by prehistoric farmers throughout western Asia and eastern Europe but is extinct today, might have been a form of *T. timopheevii* [7].

The hulled subspecies of *T. turgidum* and *T. timopheevii* have very similar morphologies and taxonomic identification is based mainly on the greater degree of hairiness of the culm internodes and leaf sheaths of *T. timopheevii* [8]. Misclassification is therefore possible, and DNA typing methods that can make unambiguous and correct identifications of the two species have been sought. However, identification of diagnostic DNA markers is complicated by the divergence time of the B and G genomes, which at 2.5–3.5 million years ago [9] is very recent in evolutionary terms, meaning that the two genomes share extensive DNA sequence identity. Additionally, in order to discriminate between *T. turgidum* and *T. timopheevii*, a marker must also give a null or diagnostic signal for the A genome, which diverged from the ancestor of the B and G genomes approximately 7 million years ago [9,10] and so also has extensive sequence similarity. Early studies indicated that the multicopy ribosomal DNA (rDNA) transcription units have features that enable the three genomes to be distinguished [11,12], and two polymerase chain reactions (PCRs) intended to be specific for the internal transcribed spacer of the G genome rDNA units were designed for identification of archaeological specimens [13]. However, one of these PCRs gave nonspecific amplification products with modern *T. turgidum* accessions and neither were successful with the ancient material. More recently, PCRs targeting chloroplast and mitochondrial DNA markers have been used [14,15], but these tests assume that the cytotype is an accurate proxy for the nuclear genome, which may not always be the case [14].

In order to identify nuclear markers for discrimination between *T. turgidum* and *T. timopheevii*, gene resequencing data (i.e. the sequences of orthologous genes from multiple accessions of the two species) are required so that species-specific sequence variations can be identified. The wheat gene for which the greatest amount of resequencing data is available is *Ppd-1*, coding for the major photoperiod response protein, with complete sequences in Genbank for 74 copies of *Ppd-B1*, 16 *Ppd-G1*, and 93 *Ppd-A1* (77 from *T. turgidum* and 16 from *T. timopheevii*) [16,17]. From this information we designed two PCRs that are specific for *Ppd-B1* and another two specific for *Ppd-G1*. Through use of these PCRs, we identify germplasm accessions of *T. turgidum* that have been misclassified as *T. timopheevii*, and vice versa.

Materials and methods

Accessions of *T. turgidum* L. subsp. *dicoccoides* (Korn. ex Asch. & Graebn.) Thell., *T. turgidum* L. subsp. *dicoccum* (Schrank ex Schübl.) Thell., *T. timopheevii* (Zhuk.) Zhuk. subsp. *armeniacum* (Jakubz.) Slageren and *T. timopheevii* (Zhuk.) Zhuk. subsp. *timopheevii* (S1 Table) were

obtained from: the Centre for Genetic Resources (CGN), Wageningen, Netherlands; the International Center for Agricultural Resources in the Dry Areas (ICARDA), Beirut, Lebanon; the Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Gatersleben, Germany; and the National Small Grains Collection (NSGC), Aberdeen, Idaho, USA. Seeds were germinated at room temperature (c.22˚C) in Petri dishes in hydroponic conditions until coleoptiles emerged. Seeds were then transferred to moist filter paper and seedlings grown until 21 days old. Fresh leaf material was collected and DNA extracted using the ISOLATE II Plant DNA kit (Bioline).

DNA sequences were downloaded from Genbank for *Ppd-B1* from 24 accessions of *T. turgidum* subsp. *dicoccoides* and 50 *T. turgidum* subsp. *dicoccum*, *Ppd-G1* from 11 *T. timopheevii* subsp. *armeniacum* and 5 *T. timopheevii* subsp. *timopheevii*, and *Ppd-A1* from 32 *T. turgidum* subsp. *dicoccoides*, 45 *T. turgidum* subsp. *dicoccum*, 11 *T. timopheevii* subsp. *armeniacum* and 5 *T. timopheevii* subsp. *timopheevii* (S2 Table). Sequences were aligned using the ClustalW, Muscle and Mafft programs in Geneious version R10 (https://www.geneious.com, [18]) and single nucleotide polymorphisms (SNPs) that are specific to the different genomes identified. Primer pairs were identified for four PCRs (Table 1), two specific for *Ppd-B1* and two for *Ppd-G1*. PCRs were carried out in a LightCycler480 (Roche) in 20 μl reaction volumes comprising 100 ng DNA extract, 1x SensiFAST SYBR No-ROX PCR master mix (Bioline), 100 nM forward primer, 100 nM reverse primer and PCR grade water. Cycling parameters were: 95˚C for 5 min; followed by 35 cycles of 20 s at 95˚C, 20 s at the annealing temperature, 20 s at 72˚C; followed by a final extension at 72˚C for 10 min. Product formation was assayed using the SYBR Green I/HRM Dye detection format (465 nm excitation, 510 nm emission) by melt curve analysis. Melting data were obtained by heating the products to 95˚C for 5 s, cooling to 55˚C for 30 s and then heating to 99˚C with five data acquisitions/˚C. Melting peaks were obtained by plotting–(δF/δT) against temperature. PCR products were additionally visualized by electrophoresis in 3% agarose gels to confirm they were the correct length.

Prior to sequencing, PCR products were cloned (Invitrogen TOPO TA Cloning Kit for Subcloning, with One Shot TOP10 chemically competent *E. coli* cells) and reamplified, using the conditions described above except for the final extension at 72˚C, with forward and reverse M13 primers (annealing temperature 55˚C) and recombinant colonies added directly to the PCR mixture. PCR products were purified with the NucleoSpin Gel and PCR Clean-up kit (Macherey-Nagel) and sequenced using the BigDye Terminator v3.1 kit chemistry (Applied Biosystems). Standard sequencing reactions of 20 μl comprised 20 ng PCR product, 1x BigDye sequencing buffer, 0.125x BigDye reaction mix, 4 pmoles M13 primer and UltraPure DNase/RNase-free distilled water. Cycling parameters were: 2 min at 96˚C; 35 cycles of 40 s at 96˚C, 15 s at 50˚C, 4 min at 60˚C; with products held at 4˚C before purification (Beckman Coulter

Table 1. Details of PCRs.

PCR	Primers	Annealing temperature (˚C)	Product size (bp)	Specific for
1	Forward: 5´-TGAAGCACAGAGCAAACACC-3´	67	84	*Ppd-B1*
	Reverse: 5´-TTGATCACGTTGGACTGAGC-3´			
2	Forward: 5´-TCTGAAAGCCGATTTCGTTT-3´	66	100	*Ppd-B1*
	Reverse: 5´-GCACCTGCAAAAGGAATGAT-3´			
3	Forward: 5´-TGAACACAGACGGTCAGTCC-3´	64	61	*Ppd-G1*
	Reverse: 5´-CGTCCATTATCGGTTGGTTT-3´			
4	Forward: 5´-GGGAAGGAGCTGGAGATAGG-3´	67	69	*Ppd-G1*
	Reverse: 5´-ACTCTCATTCGGGGAGGACT-3´			

https://doi.org/10.1371/journal.pone.0215175.t001

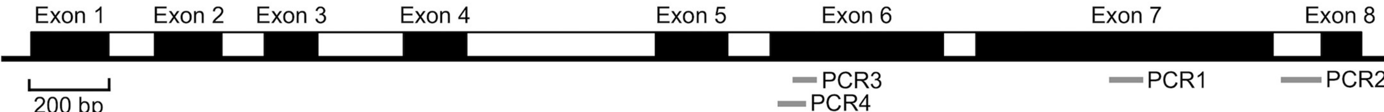


Fig 1. Schematic of the wheat *Ppd-1* gene. Exons are shown as closed boxes and introns as open boxes. The positions of the four PCRs are indicated.

https://doi.org/10.1371/journal.pone.0215175.g001

Agencourt CleanSEQ kit) and reading of paired-end sequences by capillary electrophoresis in a 3730 DNA Analyser (Applied Biosystems).

Genotyping-by-sequencing (GBS) was carried out (Genomic Diversity Facility, Cornell University) with a panel of 138 tetraploid wheats comprising 76 *T. turgidum* subsp. *dicoccoides*, 43 *T. turgidum* subsp. *dicoccum*, 11 *T. timopheevii* subsp. *armeniacum* and 8 *T. timopheevii* subsp. *timopheevii*, using a standard method [19]. Unique sequence tags were aligned to release 31 of the genome of *Triticum aestivum* L. [20] using BWA v.0.7.8-r455 [21] and SNPs identified with the TASSEL-GBS pipeline [22]. Principal components analysis (PCA) was performed with TASSEL [23].

Results

The consensus sequence resulting from multiple alignment of the 173 *Ppd-1* Genbank entries had a total length of 7302 bp with the first nucleotide of the initiation codon at position 3604 and the last nucleotide of the termination codon at position 6819. The alignment was used to design two PCRs specific for *Ppd-B1*, one of these located within exon 7 of the gene and the second mainly in intron 7 but with its 3´–terminus extending a short distance into exon 8, and a further two PCRs specific for *Ppd-G1*, both of these targeting sequences within exon 6 (Fig 1). The PCRs were designed so that each primer pair had a 100% match with their annealing sites on the target genome, but at least two mismatches with the equivalent sites on the non-target genomes (Table 2). Each primer pair gave a single product of the expected size

Table 2. Differences between the primer sequences and the equivalent sequences on the non-target genomes.

PCR1	**Forward primer**	**Reverse primer**
B genome (target sequence)	TGAAGCACAGAGCAAACACC	GCTCAGTCCAACGTGATCAA
G genome	TGAAGCACAGAGCAAACA**T**C	GCTCAGTCCA**GTT**TG**G**TCAA
A genome	TGAAGCACAGAGCAAACACC	GCTCAGTCCA**GTT**TG**G**TCAA
PCR2	**Forward primer**	**Reverse primer**
B genome (target sequence)	TCTGAAAGCCGATTTCGTTT	ATCATTCCTTTTGCAGGTGC
G genome	**–**C**CT**AAAGCCG**TC**TT**G**GT**C**T	**G**TCATT**GA**TTTT**T**CAGGTGC
A genome	**–**C**C**GAAAGCCGATT**C**CGT**C**T	**G**T**A**A**C**TC**A**TTTTGCAGGTGC
PCR3	**Forward primer**	**Reverse primer**
G genome (target sequence)	TGAACACAGACGGTCAGTCC	AAACCAACCGATAATGGACG
B genome	TGAACACAGA**T**G**A**TCA**A**TCC	AAACCAAC**A**GATAATGGACG
A genome	TGAACACAGA**T**G**A**TCAGTCC[a]	AAACCAAC**T**GAT**–––**GGACG[a]
PCR4	**Forward primer**	**Reverse primer**
G genome (target sequence)	GGGAAGGAGCTGGAGATAGG	AGTCCTCCCCGAATGAGAGT
B genome	GGGAAGGAGCTGGAGATAGG	A**A**TCCTCCCCGAA**C**GAGAGT
A genome	GGGAAGGAG**T**TGGAGATAGG[a]	AGTCCTCCCCGAA**C**GAGAGT[a]

Differences between the sequences of the primers and the non-target genomes are shown in bold.

[a] In some accessions of *T. turgidum* subsp. *dicoccum* the target sequence is absent due to a larger deletion in the *Ppd-A1* gene.

https://doi.org/10.1371/journal.pone.0215175.t002

Table 3. Accessions giving anomalous results after *Ppd-1* typing.

Accession number	Original classification	Species according to *Ppd-1* typing	Collection site		
			Country	Latitude	Longitude
PI 560697	*T. turgidum* subsp. *dicoccoides*	*T. timopheevii*	Turkey	37.58333	42.38333
PI 560873	*T. turgidum* subsp. *dicoccoides*	*T. timopheevii*	Turkey	37.47	42.03
PI 560877	*T. turgidum* subsp. *dicoccoides*	*T. timopheevii*	Turkey	38.13	41.26
PI 656869	*T. turgidum* subsp. *dicoccoides*	*T. timopheevii*	Turkey	37.2214	37.3303
PI 656872	*T. turgidum* subsp. *dicoccoides*	*T. timopheevii*	Turkey	37.2026	37.0925
PI 656873	*T. turgidum* subsp. *dicoccoides*	*T. timopheevii*	Turkey	37.1939	37.0944
CGN 16098	*T. turgidum* subsp. *dicoccoides*	*T. timopheevii*	Iran	37.28083	49.58306
CGN 16102	*T. turgidum* subsp. *dicoccoides*	*T. timopheevii*	Iraq	33.138	44.43333
CGN 13161	*T. turgidum* subsp. *dicoccoides*	*T. timopheevii*	Iraq	33.639	44.43333
CGN 24296	*T. turgidum* subsp. *dicoccoides*	*T. timopheevii*	Iraq	33.334	44.43333
PI 427998	*T. timopheevii* subsp. *armeniacum*	*T. turgidum*	Lebanon	33.51667	35.86667

https://doi.org/10.1371/journal.pone.0215175.t003

when used with DNA from its target species, and no product with the non-target species (S1 and S2 Figs), confirming the specificities of the PCRs.

The PCRs were used with 51 accessions of *T. timopheevii* and 20 of *T. turgidum* (S3 Table). Sixty accessions gave PCR products consistent with their taxon identifications. The other eleven accessions gave anomalous results (Table 3). These accessions comprised ten that were classified as *T. turgidum* subsp. *dicoccoides* but which gave positive results with the *Ppd-G1* but not the *Ppd-B1* PCRs, and which were therefore typed as *T. timopheevii*, and one *T. timopheevii* subsp. *armeniacum* accession which gave positive results for *Ppd-B1* but not *Ppd-G1*, and so was identified as *T. turgidum* (Fig 2). For each of these eleven anomalous accessions, the PCR products that were obtained were sequenced and their authenticity as *Ppd-B1* or *Ppd-G1* products confirmed from the presence of specific variations within the internal part of the amplicon (Fig 3).

GBS was carried out with 138 tetraploid wheats including each of the eleven accessions that gave anomalous results by *Ppd-1* typing. The resulting dataset of 1,172,469 SNPs was examined

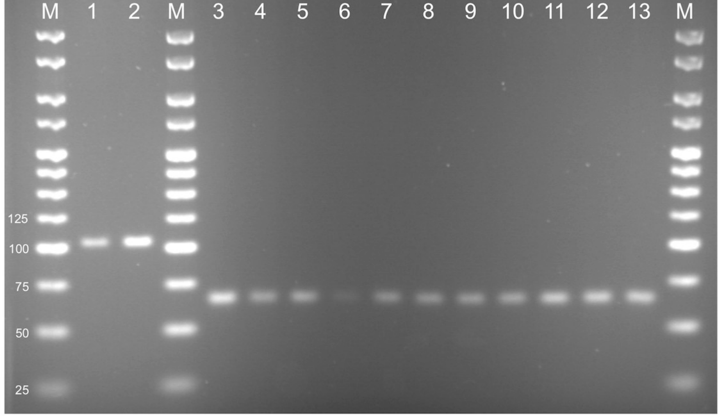


Fig 2. PCR products obtained from eleven anomalous accessions. Lanes 1 and 2: results of PCR2, specific for *Ppd-B1*, with PI 286061 (lane 1, authentic *T. turgidum* subsp. *dicoccum*) and PI 427998 (lane 2, classified as *T. timopheevii* subsp. *armeniacum*). Lanes 3–13: results with PCR3, specific for *Ppd-G1*, with PI 341802 (lane 3, authentic *T. timopheevii* subsp. *timopheevii*), PI 560697 (lane 4), PI 560873 (lane 5), PI 560877 (lane 6), PI 656869 (lane 7), PI 656872 (lane 8), PI 656873 (lane 9), CGN 16098 (lane 10), CGN 16102 (lane 11), CGN 13161 (lane 12) and CGN 24296 (lane 13) (all classified as *T. turgidum* subsp. *dicoccoides*). M, size markers (bp).

https://doi.org/10.1371/journal.pone.0215175.g002

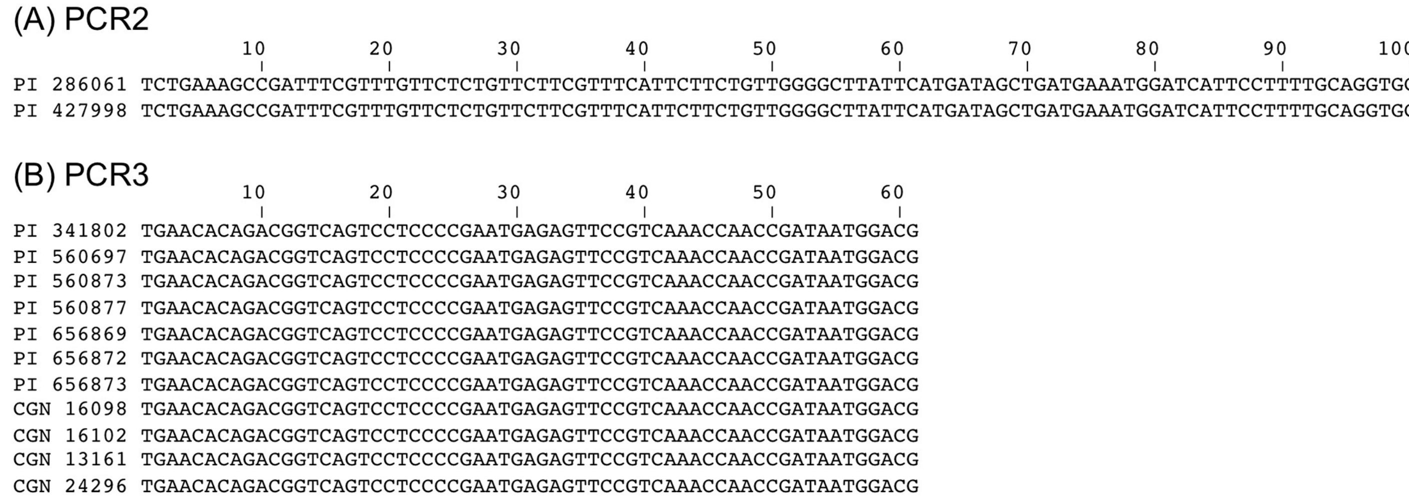


Fig 3. Sequences of PCR products obtained from eleven anomalous accessions. (A) PCR2, specific for *Ppd-B1*, with PI 286061 (authentic *T*. *turgidum* subsp. *dicoccum*) and PI 427998 (classified as *T*. *timopheevii* subsp. *armeniacum*). (B) PCR3, specific for *Ppd-G1*, with PI 341802 (authentic *T*. *timopheevii* subsp. *timopheevii*), PI 560697, PI 560873, PI 560877, PI 656869, PI 656872, PI 656873, CGN 16098, CGN 16102, CGN 13161 and CGN 24296 (all classified as *T*. *turgidum* subsp. *dicoccoides*).

https://doi.org/10.1371/journal.pone.0215175.g003

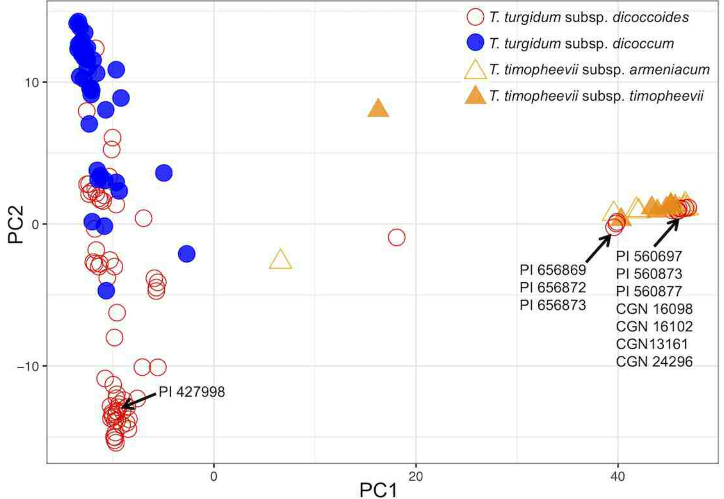


Fig 4. PCA of 138 tetraploid wheat accessions based on 1,172,469 SNPs obtained by GBS. The positions of the eleven anomalous accessions are indicated. PC1 accounts for 5.59% of the variance and PC2 accounts for 2.26%.

https://doi.org/10.1371/journal.pone.0215175.g004

by PCA. The first principal component (PC1) separated the *T*. *turgidum* and *T*. *timopheevii* accessions into distinct clusters (Fig 4). Each of the ten accessions classified as *T*. *turgidum* subsp. *dicoccoides* but identified as *T*. *timopheevii* by *Ppd-1* typing were positioned within the *T*. *timopheevii* cluster, and the single accession classified as *T*. *timopheevii* subsp. *armeniacum* but identified as *T*. *turgidum* by *Ppd-1* typing was located within the *T*. *turgidum* cluster.

Discussion

We designed two PCRs specific for the *Ppd-B1* gene and two for *Ppd-G1* and tested these with 71 *T*. *timopheevii* and *T*. *turgidum* accessions. For 60 accessions, the results of the PCRs were

consistent with the species identification, giving positive results for *Ppd-B1* and negative for *Ppd-G1*, or vice versa, indicating that the PCRs were specific for their target sequences and that neither of the PCRs gave products with the *Ppd-A1* gene on the A genome.

There were, however, eleven anomalous accessions, ten which gave positive results for *Ppd-G1* despite being classified as *T*. *turgidum*, and one classified as *T*. *timopheevii* that was typed positive for *Ppd-B1*. Previous contradictions between the outcomes of PCR typing and the morphological identification of a wheat as *T*. *timopheevii* or *T*. *turgidum* have been dismissed as errors in the DNA method [17]. However, we believe that with the anomalies we report our DNA typing results are correct and the accessions have previously been misclassified. This is because each of these eleven accessions were included in a larger group of 138 *T*. *timopheevii* and *T*. *turgidum* wheats for which we obtained GBS data. PCA of the resulting SNPs separated the 138 accessions into two clusters, one cluster comprising *T*. *timopheevii* wheats plus the ten accessions that were classified as *T*. *turgidum* but which gave a positive result for *Ppd-G1*, and the second cluster made up of *T*. *turgidum* plus the one accession that was classified as *T*. *timopheevii* but which gave a *Ppd-B1* result. As the SNPs used in the PCA mapped to all 14 tetraploid wheat chromosomes, with $>$59,000 markers per chromosome, we can be confident that the clustering reflects genome-wide differences between the groups of accessions, and therefore is giving an accurate identification of whether each wheat has an AABB or AAGG genome set. The agreement between the PCAs and the *Ppd-1* typing therefore confirms that these eleven accessions have been misclassified, and that *Ppd-1* typing (which is much less time-consuming and costly than GBS analysis) is an accurate means of distinguishing between *T*. *timopheevii* and *T*. *turgidum*.

The entries for the eleven misclassified accessions in the Germplasm Resources Information Network (GRIN) and the European Wheat Database (EWDB) give no indications that the original material that was collected might have been misidentified. However, the ten accessions misclassified as *T*. *turgidum* were collected from Turkey, Iran and Iraq, which are within the distribution range for wild *T*. *timopheevii*, and the one misidentified as *T*. *timopheevii* was collected in the Lebanon, which is outside of the area normally associated with *T*. *timopheevii* [3]. Three of the accessions misidentified as *T*. *turgidum* (PI 560697, PI 560873 and PI 560877) were previously reclassified by us as *T*. *timopheevii* based on the pattern of retrotransposon insertions in the 5S rDNA arrays [24], and two (PI 560697 and PI 560877) were similarly classified as *T*. *timopheevii* in a study of the grain *Hardness* locus [25]. In contrast, PI 560697 was included in a panel of 113 wild *T*. *turgidum* accessions used in a survey of allelic diversity at the ear-shattering loci, *TtBtr1-A* and *TtBtr1-B* [26], although PI 560697 gave an unusual result, being one of only two accessions that possessed the domesticate allele at *TtBtr1-A*. None of the other seven accessions that we reclassify as *T*. *timopheevii* (PI 656869, PI 656872, PI 656873, CGN 16098, CGN 16102, CGN 13161, CGN 24296) appear to have been extensively studied in the past. The single accession that we reclassify from *T*. *timopheevii* to *T*. *turgidum* (PI 427998) was listed as *Triticum boeoticum*, a wild diploid wheat, now called *Triticum monococcum* L. subsp. *aegilopoides* (Link) Thell., in a study of molecular diversity at 18 genetic loci [27], but was subsequently looked on as *T*. *turgidum* in the retrotransposon and *Hardness* projects mentioned above [24,25].

Conclusion

We show that the *Ppd-1* gene of wheat displays species-specific variations that enable the B and G genomes to be distinguished via simple PCR tests, the outcomes of these tests agreeing with identifications made by more comprehensive, but more time-consuming and expensive, analysis of genome-wide SNPs. The use of *Ppd-1* typing reveals a significant number of

misclassified accessions, in particular wheats initially identified as *T*. *turgidum* but which we show to be *T*. *timopheevii*, suggesting that the accepted morphological tests for discrimination between the two species might not be entirely robust. The short lengths of the amplicons (61–100 bp) means that the tests we report would be particularly suitable for typing ancient DNA, which is typically obtained as fragments <100 bp [28]. Among other archaeological applications, these tests might therefore make it possible to establish if the new glume wheat [7] is a type of *T*. *turgidum* or *T*. *timopheevii*.

Supporting information

S1 Fig. Agarose gel showing products of B- and G-specific PCRs. Within each set of four lanes the PCR has been carried out with (left to right) *T*. *timopheevii* subsp. *timopheevii* PI 341802, *T*. *timopheevii* subsp. *armeniacum* Cltr 17678, *T*. *turgidum* subsp. *dicoccum* PI 286061, *T*. *turgidum* subsp. *dicoccoides* PI 428143. Lanes 1, 10 and 19 are DNA size markers.
(TIFF)

S2 Fig. Melting peaks of products of B- and G-specific PCRs. (A) PCR1 with *T*. *turgidum* subsp. *dicoccum* PI 286061 and *T*. *turgidum* subsp. *dicoccoides* PI 428143; (B) PCR2 with *T*. *turgidum* subsp. *dicoccum* PI 286061 and *T*. *turgidum* subsp. *dicoccoides* PI 428143; (C) PCR3 with *T*. *timopheevii* subsp. *timopheevii* PI 341802 and *T*. *timopheevii* subsp. *armeniacum* Cltr 17678; (D) PCR4 with *T*. *timopheevii* subsp. *timopheevii* PI 341802 and *T*. *timopheevii* subsp. *armeniacum* Cltr 17678. The blue lines are no-template controls. Melting peak analysis enables PCR specificity to be confirmed because products with different sequences melt at different temperatures. A single peak therefore indicates that a single PCR product has been formed.
(TIFF)

S1 Table. Wheat accessions used in this study.
(XLSX)

S2 Table. Genbank entries for *Ppd-1* used in design of PCRs.
(XLSX)

S3 Table. Results of PCRs.
(XLSX)

Author Contributions

Conceptualization: Beata I. Czajkowska, Terence A. Brown.

Formal analysis: Beata I. Czajkowska, Terence A. Brown.

Funding acquisition: Terence A. Brown.

Investigation: Beata I. Czajkowska, Terence A. Brown.

Methodology: Beata I. Czajkowska, Hugo R. Oliveira, Terence A. Brown.

Project administration: Terence A. Brown.

Supervision: Terence A. Brown.

Writing – original draft: Beata I. Czajkowska, Terence A. Brown.

Writing – review & editing: Beata I. Czajkowska, Hugo R. Oliveira, Terence A. Brown.

References

1. Bhagyalakshmi K, Vinod KK, Kumar M, Arumugachamy S, Prabhakaran A, Raveendran TS. Interspecific hybrids from wild x cultivated *Triticum* crosses–a study on the cytological behaviour and molecular relations. J Crop Sci Biotechnol. 2008; 11: 257–262.
2. Wagenaar EB. Studies on the genome constitution of *Triticum timopheevi* ZHUK. II. The *T. timopheevi* complex and its origin. Evolution. 1966; 20: 150–164. https://doi.org/10.1111/j.1558-5646.1966.tb03351.x PMID: 28563623
3. Zohary D, Hopf M, Weiss E. Domestication of plants in the Old World, 4th ed. Oxford: Oxford University Press; 2012.
4. Nevo E. Evolution of wild emmer wheat and crop improvement. J Syst Evol. 2014; 52: 673–696.
5. Arranz-Otaegui A, Colledge S, Zapata L, Teira-Mayolini LC, Ibáñez JJ. Regional diversity on the timing for the initial appearance of cereal cultivation and domestication in southwest Asia. Proc Natl Acad Sci USA. 2016; 113: 14001–14006. https://doi.org/10.1073/pnas.1612797113 PMID: 27930348
6. Weide A, Riehl S, Zeidi M, Conard NJ. A systematic review of wild grass exploitation in relation to emerging cereal cultivation throughout the Epipalaeolithic and aceramic Neolithic of the Fertile Crescent. PLoS ONE. 2018; 13: e0189811. https://doi.org/10.1371/journal.pone.0189811 PMID: 29293519
7. Jones G, Valamoti S, Charles M. Early crop diversity: a "new" glume wheat from northern Greece. Veg Hist Archaeobot. 2000; 9: 133–146.
8. Clayton WD, Vorontsova MS, Harman KT, Williamson H. GrassBase—the online world grass flora. http://www.kew.org/data/grasses-db.html; 2006 onwards.
9. Allaby RG, Banerjee M, Brown TA. Evolution of the high-molecular-weight glutenin loci of the A, B, D and G genomes of wheat. Genome. 1999; 42: 296–307. PMID: 10231962
10. Marcussen T, Sandve SR, Heier L, Spannagl M, Pfeifer M, The International Wheat Genome Sequencing Consortium, et al. Ancient hybridizations among the ancestral genomes of bread wheat. Science. 2014; 345: 1250092. https://doi.org/10.1126/science.1250092 PMID: 25035499
11. Sallares R, Brown TA. PCR-based analysis of the intergenic spacers of the *Nor*-loci of the A genomes of *Triticum* diploids and polyploids. Genome. 1999; 42: 116–128. PMID: 10208005
12. Sallares R, Brown TA. Phylogenetic analysis of complete 5´ external transcribed spacers of the 18S ribosomal RNA genes of diploid *Aegilops* and related species (Triticeae, Poaceae). Genet Resour Crop Evol. 2004; 51: 701–712.
13. Boscato P, Carioni C, Brandolini A, Sadori L, Rottoli A. Molecular markers for the discrimination of *Triticum turgidum* L. subsp. *dicoccum* (Schrank ex Schübl.) Thell. and *Triticum timopheevii* (Zhuk.) Zhuk. subsp. *timopheevii*. J Archaeol Sci. 2008; 35: 239–246.
14. Gornicki P, Zhu H, Wang J, Challa GS, Zhang Z, Gill BS, et al. The chloroplast view of the evolution of polyploid wheat. New Phytol. 2014; 204: 704–714. https://doi.org/10.1111/nph.12931 PMID: 25059383
15. Tanno K-I, Takeuchi A, Akahori E, Kobayashi K, Kawahara T, Yamane K. Multiplex PCR effectively identifies tetraploid *Triticum* AABB–or AAGG–genome species. Plant Genet Resour. 2018; 16: 279–283.
16. Takenaka S, Kawahara T. Evolution and dispersal of emmer wheat (*Triticum* sp.) from novel haplotypes of *Ppd-1* (photoperiod response) genes and their surrounding DNA sequences. Theor Appl Genet. 2012; 125: 999–1104. https://doi.org/10.1007/s00122-012-1890-y PMID: 22639190
17. Takenaka S, Kawahara T. Evolution of tetraploid wheat based on variations in 5´ UTR regions of *Ppd-A1*: evidence of gene flow between emmer and timopheevi wheat. Genet Resour Crop Evol. 2013; 60: 2143–2155.
18. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 2012; 28: 1647–1649. https://doi.org/10.1093/bioinformatics/bts199 PMID: 22543367
19. Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, et al. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS ONE. 2011; 6: e19379. https://doi.org/10.1371/journal.pone.0019379 PMID: 21573248
20. Ensembl Plants Release 31. http://mar2016-plants.ensembl.org/index.html.
21. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics. 2009; 25: 1754–1760. https://doi.org/10.1093/bioinformatics/btp324 PMID: 19451168
22. Glaubitz JC, Casstevens TM, Lu F, Harriman J, Elshire RJ, Sun Q, et al. TASSEL-GBS: a high capacity genotyping by sequencing analysis pipeline. PLoS ONE. 2014; 9: e90346. https://doi.org/10.1371/journal.pone.0090346 PMID: 24587335

23. Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics. 2007; 23: 2633–2635. https://doi.org/10.1093/bioinformatics/btm308 PMID: 17586829
24. Civáň P, Ivaničová Z, Brown TA. Reticulated origin of domesticated emmer wheat supports a dynamic model for the emergence of agriculture in the Fertile Crescent. PLoS ONE. 2013; 8: e81955. https://doi.org/10.1371/journal.pone.0081955 PMID: 24312385
25. Shaaf S, Sharma R, Baloch FS, Badaeva ED, Knüpffer H, Kilian B, et al. The grain *Hardness* locus characterized in a diverse wheat panel (*Triticum aestivum* L.) adapted to the central part of the Fertile Crescent: genetic diversity, haplotype structure, and phylogeny. Mol Genet Genomics. 2016; 291: 1259–1275. https://doi.org/10.1007/s00438-016-1180-5 PMID: 26898967
26. Avni R, Nave M, Barad O, Baruch K, Twardziok SO, Gundlach H, et al. Wild emmer genome architecture and diversity elucidate wheat evolution and domestication. Science. 2017; 357: 93–97. https://doi.org/10.1126/science.aan0032 PMID: 28684525
27. Kilian B, Özkan H, Walther A, Kohl J, Dagan T, Salamini F, et al. Molecular diversity at 18 loci in 321 wild and 92 domesticate lines reveal no reduction of nucleotide diversity during *Triticum monococcum* (einkorn) domestication: implications for the origin of agriculture. Mol Biol Evol. 2007; 24: 2657–2668. https://doi.org/10.1093/molbev/msm192 PMID: 17898361
28. Brown TA, Cappellini E, Kistler L, Lister DL, Oliveira HR, Wales N, et al. Recent advances in ancient DNA research and their implications for archaeobotany. Veg Hist Archaeobot. 2015; 24: 207–214.