EXTENDED PDF FORMAT
SPONSORED BY
Travel Grants Available
Neuroscience 2014
Application Deadline: July 31st
>Apply Now
R&D systems
a bio-techne brand
www.rndsystems.com

Science
AAAS

Ancient hybridizations among the ancestral genomes of bread wheat

Thomas Marcussen *et al.*

Science **345**, (2014);
DOI: 10.1126/science.1250092

This copy is for your personal, non-commercial use only.

If you wish to distribute this article to others, you can order high-quality copies for your colleagues, clients, or customers by clicking here.

Permission to republish or repurpose articles or portions of articles can be obtained by following the guidelines here.

The following resources related to this article are available online at www.sciencemag.org (this information is current as of July 17, 2014):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:
http://www.sciencemag.org/content/345/6194/1250092.full.html

Supporting Online Material can be found at:
http://www.sciencemag.org/content/suppl/2014/07/16/345.6194.1250092.DC1.html

A list of selected additional articles on the Science Web sites **related to this article** can be found at:
http://www.sciencemag.org/content/345/6194/1250092.full.html#related

This article **cites 43 articles**, 22 of which can be accessed free:
http://www.sciencemag.org/content/345/6194/1250092.full.html#ref-list-1

This article has been **cited by** 1 articles hosted by HighWire Press; see:
http://www.sciencemag.org/content/345/6194/1250092.full.html#related-urls

This article appears in the following **subject collections:**
Evolution
http://www.sciencemag.org/cgi/collection/evolution
Genetics
http://www.sciencemag.org/cgi/collection/genetics

Downloaded from www.sciencemag.org on July 17, 2014

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published weekly, except the last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. Copyright 2014 by the American Association for the Advancement of Science; all rights reserved. The title *Science* is a registered trademark of AAAS.

A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome

The International Wheat Genome Sequencing Consortium (IWGSC)

An ordered draft sequence of the 17-gigabase hexaploid bread wheat (*Triticum aestivum*) genome has been produced by sequencing isolated chromosome arms. We have annotated 124,201 gene loci distributed nearly evenly across the homeologous chromosomes and subgenomes. Comparative gene analysis of wheat subgenomes and extant diploid and tetraploid wheat relatives showed that high sequence similarity and structural conservation are retained, with limited gene loss, after polyploidization. However, across the genomes there was evidence of dynamic gene gain, loss, and duplication since the divergence of the wheat lineages. A high degree of transcriptional autonomy and no global dominance was found for the subgenomes. These insights into the genome biology of a polyploid crop provide a springboard for faster gene isolation, rapid genetic marker development, and precise breeding to meet the needs of increasing food demand worldwide.

Lists of authors and affiliations are available in the full article online.
Corresponding author: K. X. Mayer, e-mail: k.mayer@helmholtz-muenchen.de

Read the full article at http://dx.doi.org/10.1126/science.1251788

Ancient hybridizations among the ancestral genomes of bread wheat

Thomas Marcussen, Simen R. Sandve,* Lise Heier, Manuel Spannagl, Matthias Pfeifer, The International Wheat Genome Sequencing Consortium,† Kjetill S. Jakobsen, Brande B. H Wulff, Burkhard Steuernagel, Klaus F. X. Mayer, Odd-Arne Olsen

The allohexaploid bread wheat genome consists of three closely related subgenomes (A, B, and D), but a clear understanding of their phylogenetic history has been lacking. We used genome assemblies of bread wheat and five diploid relatives to analyze genome-wide samples of gene trees, as well as to estimate evolutionary relatedness and divergence times. We show that the A and B genomes diverged from a common ancestor ~7 million years ago and that these genomes gave rise to the D genome through homoploid hybrid speciation 1 to 2 million years later. Our findings imply that the present-day bread wheat genome is a product of multiple rounds of hybrid speciation (homoploid and polyploid) and lay the foundation for a new framework for understanding the wheat genome as a multilevel phylogenetic mosaic.

The list of author affiliations is available in the full article online.***Corresponding author. E-mail: simen.sandve@nmbu.no** †The International Wheat Genome Sequencing Consortium (IWGSC) authors and affiliations are listed in the supplementary materials.

Read the full article at http://dx.doi.org/10.1126/science.1250092

Ancestral wheat

Wheat varieties and species (shown) believed to be the closest living relatives of modern bread wheat (*T. aestivum*). Multiple ancestral hybridizations occurred among most of these species, many of which are cultivated, and along with *T. aestivum* represent a dominant source of global nutrition.

Published by AAAS

PHOTOS: SUSANNE STAMP, ERNST MERZ/ETH ZURICH

Genome interplay in the grain transcriptome of hexaploid bread wheat

Matthias Pfeifer, Karl G. Kugler, Simen R. Sandve, Bujie Zhan, Heidi Rudi, Torgeir R. Hvidsten, International Wheat Genome Sequencing Consortium,* Klaus F. X. Mayer, Odd-Arne Olsen†

Allohexaploid bread wheat (*Triticum aestivum* L.) provides approximately 20% of calories consumed by humans. Lack of genome sequence for the three homeologous and highly similar bread wheat genomes (A, B, and D) has impeded expression analysis of the grain transcriptome. We used previously unknown genome information to analyze the cell type–specific expression of homeologous genes in the developing wheat grain and identified distinct co-expression clusters reflecting the spatiotemporal progression during endosperm development. We observed no global but cell type– and stage-dependent genome dominance, organization of the wheat genome into transcriptionally active chromosomal regions, and asymmetric expression in gene families related to baking quality. Our findings give insight into the transcriptional dynamics and genome interplay among individual grain cell types in a polyploid cereal genome.

The list of author affiliations is available in the full article online. *The International Wheat Genome Sequencing Consortium (IWGSC) authors and affiliations are listed in the supplementary materials. †**Corresponding author. E-mail: odd-arne.olsen@nmbu.no**

Read the full article at http://dx.doi.org/10.1126/science.1250091

Structural and functional partitioning of bread wheat chromosome 3B

Frédéric Choulet,* Adriana Alberti, Sébastien Theil, Natasha Glover, Valérie Barbe, Josquin Daron, Lise Pingault, Pierre Sourdille, Arnaud Couloux, Etienne Paux, Philippe Leroy, Sophie Mangenot, Nicolas Guilhot, Jacques Le Gouis, Francois Balfourier, Michael Alaux, Véronique Jamilloux, Julie Poulain, Céline Durand, Arnaud Bellec, Christine Gaspin, Jan Safar, Jaroslav Dolezel, Jane Rogers, Klaas Vandepoele, Jean-Marc Aury, Klaus Mayer, Hélène Berges, Hadi Quesneville, Patrick Wincker, Catherine Feuillet

We produced a reference sequence of the 1-gigabase chromosome 3B of hexaploid bread wheat. By sequencing 8452 bacterial artificial chromosomes in pools, we assembled a sequence of 774 megabases carrying 5326 protein-coding genes, 1938 pseudogenes, and 85% of transposable elements. The distribution of structural and functional features along the chromosome revealed partitioning correlated with meiotic recombination. Comparative analyses indicated high wheat-specific inter- and intrachromosomal gene duplication activities that are potential sources of variability for adaption. In addition to providing a better understanding of the organization, function, and evolution of a large and polyploid genome, the availability of a high-quality sequence anchored to genetic maps will accelerate the identification of genes underlying important agronomic traits.

The list of author affiliations is available in the full article online.
***Corresponding author. E-mail: frederic.choulet@clermont.inra.fr**

Read the full article at http://dx.doi.org/10.1126/science.1249721

Published by AAAS

WHEAT GENOME

Ancient hybridizations among the ancestral genomes of bread wheat

Thomas Marcussen,[1*] Simen R. Sandve,[1*†] Lise Heier,[2] Manuel Spannagl,[3] Matthias Pfeifer,[3] The International Wheat Genome Sequencing Consortium,[‡] Kjetill S. Jakobsen,[4] Brande B. H. Wulff,[5] Burkhard Steuernagel,[5] Klaus F. X. Mayer,[3] Odd-Arne Olsen[1]

The allohexaploid bread wheat genome consists of three closely related subgenomes (A, B, and D), but a clear understanding of their phylogenetic history has been lacking. We used genome assemblies of bread wheat and five diploid relatives to analyze genome-wide samples of gene trees, as well as to estimate evolutionary relatedness and divergence times. We show that the A and B genomes diverged from a common ancestor ~7 million years ago and that these genomes gave rise to the D genome through homoploid hybrid speciation 1 to 2 million years later. Our findings imply that the present-day bread wheat genome is a product of multiple rounds of hybrid speciation (homoploid and polyploid) and lay the foundation for a new framework for understanding the wheat genome as a multilevel phylogenetic mosaic.

The rise of modern agriculture and wheat domestication in the Fertile Crescent ~10,000 years ago (*1*–*4*) was pivotal in shaping modern human history. Early farming practices made use of wild diploid wheat species (i.e., *Aegilops* and *Triticum* species), but as agriculture evolved, wild crops were gradually substituted with domesticated diploid and polyploid wheat varieties (*3*, *4*). Presently, the allohexaploid bread wheat (*Triticum aestivum*, $2n = 6x = 42$ chromosomes; genomic code AABBDD) dominates global wheat production. Because of its economic value and the desire for its genetic improvement, questions concerning the evolution and domestication of wheat have been under intense scientific scrutiny (*5*, *6*).

The bread wheat subgenomes A, B, and D were originally derived from three diploid ($2x$; $2n = 14$) species within tribe Triticeae [see figure 1 in (*7*)]: *Triticum urartu* (AA), an unknown close relative of *Aegilops speltoides* (BB), and *Ae. tauschii* (DD) (*4*, *8*). The initial allopolyploidization event is hypothesized to have involved the A and B genome donors, resulting in the extant tetraploid emmer wheat (*T. turgidum*; AABB). This species subsequently hybridized with the D genome donor to form modern hexaploid bread wheat (AABBDD) (*4*, *8*).

Tetraploid emmer wheat is believed to have originated within the past few hundred thousand years (*9*), whereas hexaploid bread wheat is thought to have originated with modern agriculture ~10,000 years ago (*4*). The time of origin for hexaploid bread wheat is currently supported solely by archeological evidence (*2*, *3*) and the apparent absence of hexaploid wheats in wild populations (*4*). Although the relatedness between the bread wheat subgenomes and diploid wheat species has been well documented (*8*, *10*), a clear understanding of the phylogenetic history and divergence times among the three A, B, and D genome lineages is still lacking (*9*, *11*–*13*). This knowledge gap is mainly a consequence of the paucity of Triticeae fossils (*14*), which has prevented investigations of diversification through time; extensive topological discordance between wheat gene trees (*15*); and, most importantly, the lack of genome sequences of the hexaploid bread wheat and its close diploid relatives. Improved understanding of the phylogenetic relationships among the diploid species of wheat and the bread wheat subgenomes is important for understanding genome function and for future agricultural crop improvement in light of a changing global climate (*16*).

Gene tree topology analyses

We used the genome sequences of hexaploid bread wheat subgenomes (denoted TaA, TaB, and TaD) and five diploid relatives (*T. monococcum*, *T. urartu*, *Ae. sharonensis*, *Ae. speltoides*, and *Ae. tauschii*) (*7*, *17*, *18*) to generate a genome-wide sample of 275 gene trees and to estimate the phylogenetic history of the A, B, and D genome lineages. Barley (*Hordeum vulgare*), *Brachypodium distachyon*, and rice (*Oryza sativa*) were used as outgroup species. To generate multiple alignments of ortholog genes, we employed a phylogeny-aware strategy (*19*), which simultaneously filters alignments for unreliably aligned codon sites and putative erroneously predicted ortholog sequences (fig. S1 and supplementary materials and methods). Finally, we used BEAST (*20*) to calculate gene trees topologies.

We found that the basal relatedness among the three lineages A, B, and D varied substantially among the 275 gene trees, with the lineage topologies A(B,D) and B(A,D) each being about twice as common as D(A,B) (Fig. 1A and Table 1). Stochastic population genetic processes typically cause incomplete lineage sorting (ILS), which results in topological discordance (i.e., variation in topology) among individual gene trees. For three taxa under ILS alone, the gene tree topology that equals the species tree topology is expected to be more common than the other

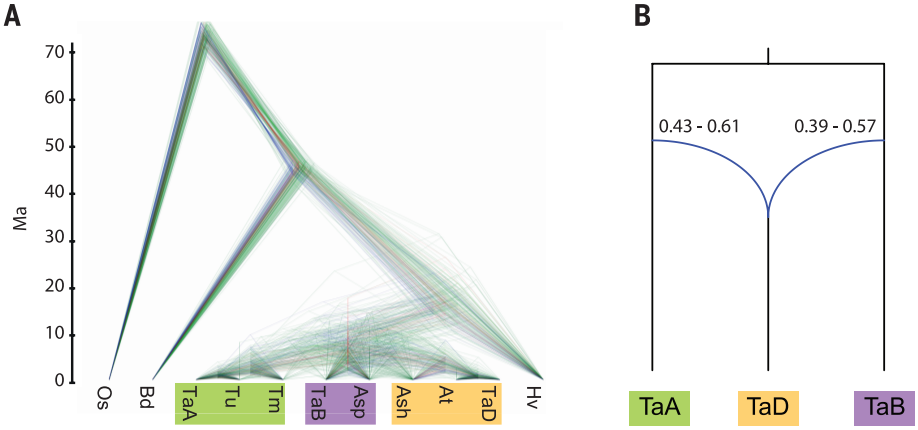


Fig. 1. Analyses of gene tree topologies. (**A**) Superimposed ultrametric gene trees in a consensus DensiTree plot. The branch color changes for every 100 trees plotted. (**B**) Topology-based species phylogeny, assuming incomplete lineage sorting using a data set of 2269 gene trees inferred by PhyloNet. The results presented represent analyses of all gene trees (2269). The numbers on the branches represent estimates of parental contributions to the hybrid. Range estimates of parental contribution are extrapolated from results reported in Table 1. Species names are abbreviated as follows: Ash, *Aegilops sharonensis*; Asp, *Ae. speltoides*; At, *Ae. tauschii*; Tm, *Triticum monococcum*; Tu, *T. urartu*; Hv, *Hordeum vulgare*; TaA, *T. aestivum* A subgenome; TaB, *T. aestivum* B subgenome; TaD, *T. aestivum* D subgenome; Bd, *Brachypodium distachyon*; Os, *Oryza sativa*.

[1]Department of Plant Sciences, Norwegian University of Life Sciences, 1432 Ås, Norway. [2]Strømsveien 78 B, 0663 Oslo, Norway. [3]Plant Genome and Systems Biology, Helmholtz Center Munich, Ingolstädter Landstrasse 1, 85764 Neuherberg, Germany. [4]Centre for Ecological and Evolutionary Synthesis, Department of Biosciences, University of Oslo, 0316 Oslo, Norway. [5]The Sainsbury Laboratory, Norwich Research Park, Norwich NR4 7UH, UK.
*These authors contributed equally to this work. **†Corresponding author. E-mail: simen.sandve@nmbu.no** ‡The International Wheat Genome Sequencing Consortium (IWGSC) authors and affiliations are listed in the supplementary materials.

Table 1. Distribution of ABD lineage topologies in gene trees. Analyses were made on different data subsets: diploid genomes only (2*x*), hexaploid genomes only (6*x*), and either whole genomes or gene trees from individual chromosomes (Chr.). Interrelationships within the A, D, and B clades are not considered in the data set that includes diploids. Topologies including diploids were estimated with Bayesian MCMC sampling using the HKY+G nucleotide substitution model, whereas topologies excluding diploids were taken from IWGSC (7) and represent maximum likelihood topologies under the GTR+I+G model. Bold numbers represent the largest topology group. The likelihood ratio test was used to test the probability (*P*) of observing the data under the model of multispecies coalescent and the (conservative) assumption that the most common observed tree topology equaled the species tree topology.

Genomes	Genes	Sample	Observed topologies			Proportion of genes with deep coalescence	Parental contributions in D		Likelihood ratio test *P*
			A,(B,D)	B,(A,D)	D,(A,B)		A	B	
2*x*	275	Whole genome*	**112**	100	63	0.69	0.43	0.57	0.0036‡
	275	Whole genome†	**109**	101	65	0.71	0.45	0.55	0.0050‡
6*x*	275	Whole genome	**107**	105	63	0.69	0.49	0.51	0.0058‡
	2269	Whole genome	786	**909**	574	0.76	0.61	0.39	8.4×10^{-9}‡
	324	Chr. 1	109	**137**	78	0.72	0.66	0.34	0.023§
	428	Chr. 2	131	**191**	106	0.74	0.77	0.23	0.10
	321	Chr. 3	111	**141**	69	0.65	0.63	0.37	0.0016‡
	305	Chr. 4	**121**	102	82	0.81	0.34	0.67	0.14
	347	Chr. 5	127	**129**	91	0.79	0.52	0.49	0.015§
	290	Chr. 6	99	**117**	74	0.77	0.63	0.37	0.057
	254	Chr. 7	88	**92**	74	0.87	0.56	0.44	0.27

*A lineage represented by *T. monococcum*, excluding *T. urartu*. †A lineage represented by *T. urartu*, excluding *T. monococcum*. ‡Significant at $P < 0.01$. §Significant at $P < 0.05$.

two. However, for our data the observed lineage topology ratios differed significantly [$P < 0.01$; likelihood ratio test; df = 1 (Table 1)] from this expectation. This suggests the presence of phylogenetic signals additional to ILS in the data. Except in rare instances of deep coalescence, bread wheat homeologs consistently formed monophyletic clades with orthologs of their close diploid relatives (Fig. 1A), and never with each other. This rules out nonhomologous gene conversion and nonhomologous recombination as explanations for the observed topological discordance. To enable analyses of individual chromosomes, we also made use of a considerably larger data set of 2269 maximum likelihood gene tree topologies taken from (7), which did not include the diploid *Triticum* and *Aegilops* species. These gene trees support genome-wide signals of significantly skewed topology frequencies, with B(A,D) and A(D,B) topologies being most common (Table 1). Taken together, these results suggest that lineages A and B are more closely related to D individually than to each other, which agrees with a model of hybrid origin of the D lineage (Fig. 1B).

Under the assumption of ILS and a hybrid origin of the D lineage from the A and B lineages, analysis of tree topology frequencies revealed roughly equal contributions of each parental lineage to the D lineage (Table 1). There was considerable stochasticity among the different data subsets, and between 65 and 87% of the genes displayed deep coalescence for the target nodes. Similar results were obtained with a topology-based parsimony approach to estimate phylogenetic genome networks under the assumption of a single homoploid hybridization at the genome-wide level (Fig. 1B) and for the majority of the chromosome-specific analyses (table S4). Finally, ~80% of the genes could be anchored to a chromosome position in the hexaploid genome using the in silico gene order predictions from the bread wheat genome sequence (7). Such positional information can be used to investigate whether different regions of the genome have distinct phylogenetic signals—that is, conserved chromosome blocks from the parental genomes. Homeologs within gene trees showed highly conserved syntenic relationships (fig. S3); however, anchoring of gene tree topologies to chromosome positions in bread wheat did not support the presence of larger chromosome blocks with a single parental origin (7), indicating a relatively homogeneous hybrid signal throughout the D subgenome.

Genome divergence times

Because topology-based phylogenetic analyses do not consider the temporal scale, we used a Bayesian hierarchical model to further estimate the genome divergence times under the multispecies coalescent (21) from pairwise ortholog coalesent distributions. This approach does not assume a treelike species phylogeny and can handle very large data sets.

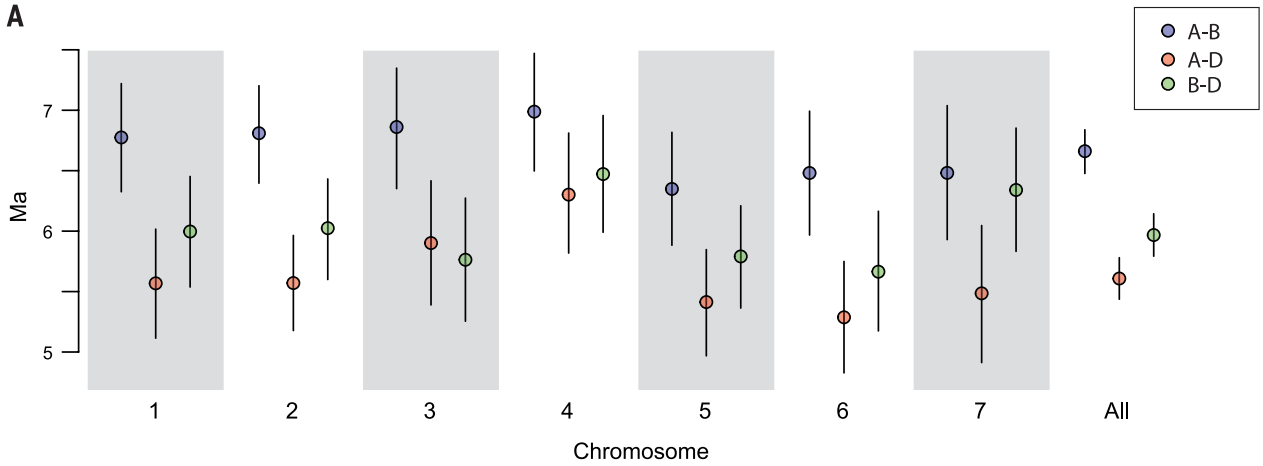
Concurrent with the topology-based analyses (Fig. 2, A and B, and fig. S4), genome divergence estimates showed no signs of being affected by nonhomologous gene conversion or recombination within the hexaploid genome, which would have resulted in shallow coalescence of bread wheat subgenomes (figs. S4 and S5). The basal divergence in *Aegilops*/*Triticum* was estimated to have occurred between the A and the B genome lineages ~7 million years ago (Ma) (Table 2). Both A-D and B-D divergence times overlap and are estimated to be 1 to 2 million years younger than the A-B divergence (Fig. 2, A and B). This contradicts a treelike phylogeny for the three subgenome lineages (Table 2 and Fig. 2A) and favors a model of hybridization between the A and B genomes, giving rise to the D genome. Genome divergence times did not support the more complex models of hybridization patterns, as suggested by the topology analyses assuming two hybridization events (table S4). Furthermore, the majority of the analyses produced slightly younger divergence of A and D lineages compared with B and D lineages (Fig. 2A and Table 2), indicating that gene flow from A to D may have persisted after gene flow from B to D had ceased.

The identification of hybridization events in phylogenies strongly depends on taxon sampling. Nevertheless, given that the hybridization event happened basally in the Triticum/Aegilops clade and that the 15 extant diploid species all seem to fall within one of the three lineages A, B, and D (fig. S6), the hybridization pattern is likely to remain unaltered, even with a denser sampling of wheat species. Beyond the phylogenetic evidence presented here, support for a homoploid hybrid origin of the D genome is found in independent analyses using the genome sequence of bread wheat. Both at the base-pair level (7, 22) as well as in gene content (7), the A and B lineages are more similar to the D genome lineage than they are to each other.

Although the existence of homoploid hybrid speciation has been acknowledged for well more than a century (23), it has only recently been recognized as a relatively common phenomenon (24–26). Our data support a homoploid hybrid origin of the bread wheat D genome lineage ancestor more than 5 Ma, which is among the oldest cases of homoploid hybrid speciation reported to date.

Divergence of polyploid genomes from diploid relatives

Genome divergence estimates showed that *T. monococcum* and *T. urartu* are successive



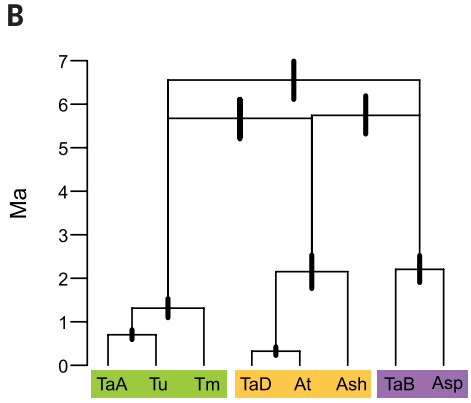


Fig. 2. Coalescent-based genome divergence analyses. Coalescence times were estimated as the median of Bayesian MCMC sampling in BEAST. Genome divergence estimates were inferred with a Bayesian hierarchical model through WinBUGS and the R2OpenBUGS R package (*35*). (**A**) Divergence times (mean, 95% credibility interval) for the genome lineages A, B, and D for 2269 gene trees, excluding diploid species. A-B, blue; A-D, red; B-D, green. (**B**) Genome divergence network including diploid and hexaploid wheat genomes. Node age is given as mean genome divergence time, estimated independently for each pair of species representing that node. For nodes with more than two decendant tips, age is given as the mean for all relevant pairwise species comparisons, and bars span from the lowest minimal to the highest maximal 95% bound for their credibility intervals. Due to evidence of recent interlineage hybridizations (both in topology and coalescence analyses) in the *Ae. sharonensis* and *Ae. speltoides* genomes, these species are not considered in the estimation of the ancestral A, B, and D lineage divergence.

Table 2. Estimated genome divergence times. All age estimates are given in units of million years ago as 95% credibility intervals (CIs). The CI of the Tm-TaA divergence and the CI of the At-TaD divergence represent the summarized CI ranges of two hierarchical Bayesian models using median plus median. The Tm-TaA divergence and the At-TaD divergence are expected to be overestimates of the actual polyploidization times due to the fact that the true ancestral populations to the A and D subgenomes in bread wheat were not sampled. Species names are abbreviated as follows: At, *Aegilops tauschii*; Tm, *Triticum monococcum*; TaA, *T. aestivum* A subgenome; TaD, *T. aestivum* D subgenome. Dashes indicate no data.

Data set	TMRCA				
	A-B	A-D	B-D	Tm-TaA	At-TaD
275 gene trees	6.11–6.99	5.20–6.11	5.31–6.19	0.59–0.82	0.23–0.43
2269 gene trees	6.47–6.83	5.43–5.78	5.79–6.14	–	–

sisters to TaA, *Ae. speltoides* is sister to TaB, and *Ae. sharonensis* and *Ae. tauschii* are successive sisters to TaD (Fig. 2B and table S5). To elucidate the timing of the polyploid speciations giving rise to emmer and bread wheat, we analyzed the pairwise coalescence distributions of the TaA and TaD genomes and the closest related diploid species, *T. urartu* and *Ae. tauschii*, respectively (Fig. 2B). We estimated the *T. urartu*-TaA divergence to 0.58 to 0.82 Ma and the *Ae. tauschii*–TaD divergence to 0.23 to 0.43 Ma (Fig. 2B and Table 1), suggesting that the age of these two polyploidization events could be older than previously suggested (*4*, *9*). It is however important to note that this study includes only single-genome samples from *T. urartu* and *Ae. tauschii* and that these samples are not likely to represent plants from the actual ancestor populations that gave rise to the A and D genome constituents. Hence, our genome divergence estimates between diploid and hexaploid genomes are likely to reflect population divergence events before the actual polyploidization times, and additional analyses of broad samples of both *T. urartu* and *Ae. tauschii* populations are needed to further improve the accuracy of the polyploidization time estimates (*27*).

Conclusions

Our study exemplifies how analyses of whole-genome data sets can aid in resolving convoluted patterns of genome evolution caused by ancient hybridization events. We elucidate genome-wide signatures of hybrid ancestry of the wheat D lineage from A and B lineage ancestors (Fig. 3). Not only is bread wheat a product of hybridization and allopolyploidization involving the A, B, and D genomes, but also the ancestral lineages of these three genomes are the result of ancestral hybridization events among themselves. Our findings could have broad implications for understanding genome function and, thus, cultivar improvement in bread wheat.

Methods

Sequence data for *T. urartu* and *Ae. tauschii* were downloaded from GenBank (accession numbers AOTI00000000 and AOCO000000000.1, respectively). All other sequences were downloaded from the International Wheat Genome Sequencing Consortium (IWGSC) sequence repository at Institut National de la Recherche Agronomique (http://wheat-urgi.versailles.inra.fr/Seq-Repository/Genes-annotations). Preliminary ortholog predictions were carried out with OrthoMCL (*28*). Only orthologous gene sets containing single-gene copies from all species were used. A phylogeny-aware sequence alignment pipeline modified from (*19*) was implemented in R language using MAFFT (*29*) to construct multiple sequence alignments, GUIDANCE (*30*) to assess sequence alignment quality, and phangorn (*31*) to construct maximum likelihood topologies, which were used to evaluate ortholog relationships within gene trees. Ultrametric gene trees were estimated in BEAST (*20*), based on a secondary calibration on the

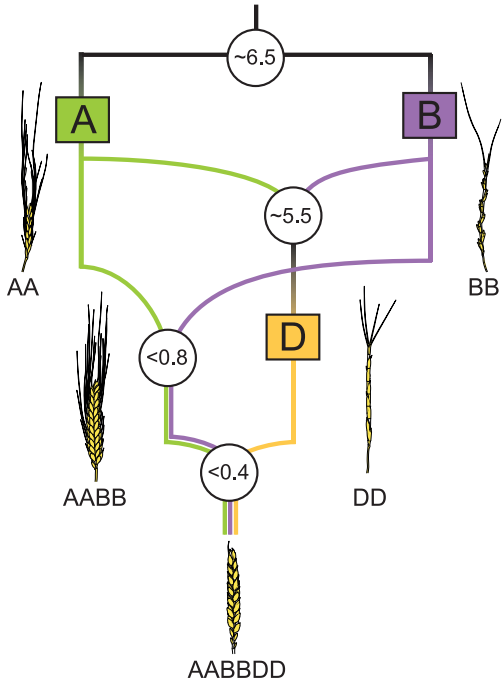


Fig. 3. Model of the phylogenetic history of bread wheat (*Triticum aestivum*; AABBDD). Approximate dates for divergence and the three hybridization events are given in white circles in units of million years ago. Differentiation of the wheat lineage (*Triticum* and *Aegilops*) from a common ancestor into the A and B genome lineages began ~6.5 Ma. The first hybridization occurred ~5.5 Ma between the A and B genome lineages and led to the origin of the D genome lineage by homoploid hybrid speciation. The second hybridization, between a close relative (BB) of *Ae. speltoides* and *T. urartu* (AA), gave rise to the allotetraploid emmer wheat (*T. turgidum*; AABB) by polyploidization. Bread wheat originated by allopolyploidization from a third hybridization, between emmer wheat and *Ae. tauschii* (DD). The three diploid lineages are indicated with color and labels. Inflorescences (spikes) illustrate extant species closely related to those involved in the polyploidizations.

Brachypodium crown node (fig. S2). All analyses of topologies were carried out with the APE R package (*32*). Estimations of parental genome contributions and species networks from gene tree topologies were calculated using parsimonious inference of hybridization in the presence of ILS (*33*) implemented in PhyloNet (*34*) and with a coalescent-based method (described in the supplementary materials). Topology-independent genome divergence times were estimated under the multispecies coalescent model with hierarchical modeling using BUGS (Bayesian inference Using Gibbs Sampling) (www.openbugs.net/w/FrontPage) through the R2OpenBUGS R package (*35*).

REFERENCES AND NOTES

1. M. Heun *et al*., Site of einkorn wheat domestication identified by DNA fingerprinting. *Science* **278**, 1312–1314 (1997). doi: 10.1126/science.278.5341.1312
2. S. Lev-Yadun, A. Gopher, S. Abbo, The cradle of agriculture. *Science* **288**, 1602–1603 (2000). doi: 10.1126/science.288.5471.1602; pmid: 10858140
3. S. Riehl, M. Zeidi, N. J. Conard, Emergence of agriculture in the foothills of the Zagros Mountains of Iran. *Science* **341**, 65–67 (2013). doi: 10.1126/science.1236743; pmid: 23828939
4. F. Salamini, H. Özkan, A. Brandolini, R. Schäfer-Pregl, W. Martin, Genetics and geography of wild cereal domestication in the near east. *Nat. Rev. Genet.* **3**, 429–441 (2002). pmid: 12042770
5. E. S. McFadden, E. R. Sears, The origin of *Triticum spelta* and its free-threshing hexaploid relatives. *J. Hered.* **37**, 81–107, 107 (1946). pmid: 20985728
6. J. Dubcovsky, J. Dvorak, Genome plasticity a key factor in the success of polyploid wheat under domestication. *Science* **316**, 1862–1866 (2007). doi: 10.1126/science.1143986; pmid: 17600208
7. International Wheat Genome Sequencing Consortium, A chromosome-based draft sequence of the hexaploid bread wheat genome. *Science* **345**, 1251788 (2014).
8. G. Petersen, O. Seberg, M. Yde, K. Berthelsen, Phylogenetic relationships of *Triticum* and *Aegilops* and evidence for the origin of the A, B, and D genomes of common wheat (*Triticum aestivum*). *Mol. Phylogenet. Evol.* **39**, 70–82 (2006) and references therein. doi: 10.1016/j.ympev.2006.01.023; pmid: 16504543
9. S. Huang *et al*., Genes encoding plastid acetyl-CoA carboxylase and 3-phosphoglycerate kinase of the *Triticum/Aegilops* complex and the evolutionary history of polyploid wheat. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 8133–8138 (2002). doi: 10.1073/pnas.072223799; pmid: 12060759
10. E. D. Akhunov, A. R. Akhunova, J. Dvorák, BAC libraries of *Triticum urartu*, *Aegilops speltoides* and *Ae. tauschii*, the diploid ancestors of polyploid wheat. *Theor. Appl. Genet.* **111**, 1617–1622 (2005) and references therein. doi: 10.1007/s00122-005-0093-1; pmid: 16177898
11. D. Chalupska *et al*., *Acc* homoeoloci and the evolution of wheat genomes. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 9691–9696 (2008). doi: 10.1073/pnas.0803981105; pmid: 18599450
12. J. Dvorak, E. D. Akhunov, Tempos of gene locus deletions and duplications and their relationship to recombination rate during diploid and polyploid evolution in the Aegilops-Triticum alliance. *Genetics* **171**, 323–332 (2005). doi: 10.1534/genetics.105.041632; pmid: 15996988
13. X. Fan *et al*., Phylogenetic reconstruction and diversification of the Triticeae (Poaceae) based on single-copy nuclear *Acc1* and *Pgk1* gene data. *Biochem. Syst. Ecol.* **50**, 346–360 (2013). doi: 10.1016/j.bse.2013.05.010
14. C. A. E. Strömberg, Evolution of grasses and grassland ecosystems. *Annu. Rev. Earth Planet. Sci.* **39**, 517–544 (2011). doi: 10.1146/annurev-earth-040809-152402
15. J. S. Escobar *et al*., Multigenic phylogeny and analysis of tree incongruences in Triticeae (Poaceae). *BMC Evol. Biol.* **11**, 181–198 (2011). doi: 10.1186/1471-2148-11-181; pmid: 21702931
16. D. B. Lobell, W. Schlenker, J. Costa-Roberts, Climate trends and global crop production since 1980. *Science* **333**, 616–620 (2011). doi: 10.1126/science.1204531; pmid: 21551030
17. H. Q. Ling *et al*., Draft genome of the wheat A-genome progenitor *Triticum urartu*. *Nature* **496**, 87–90 (2013). doi: 10.1038/nature11997; pmid: 23535596
18. J. Jia *et al*., *Aegilops tauschii* draft genome sequence reveals a gene repertoire for wheat adaptation. *Nature* **496**, 91–95 (2013). doi: 10.1038/nature12028; pmid: 23535592
19. M. D. Vigeland *et al*., Evidence for adaptive evolution of low-temperature stress response genes in a Pooideae grass ancestor. *New Phytol.* **199**, 1060–1068 (2013). doi: 10.1111/nph.12337; pmid: 23701123
20. A. J. Drummond, M. A. Suchard, D. Xie, A. Rambaut, Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* **29**, 1969–1973 (2012). doi: 10.1093/molbev/mss075; pmid: 22367748
21. J. H. Degnan, N. A. Rosenberg, Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends Ecol. Evol.* **24**, 332–340 (2009). doi: 10.1016/j.tree.2009.01.009; pmid: 19307040
22. M. Pfeifer *et al*., Genome interplay in the grain transcriptome of hexploid bread wheat. *Science* **345**, 1250091 (2014).
23. J. P. Lotsy, *Evolution by Means of Hybridization* (Martinus Nijhoff, The Hague, Netherlands, 1916).
24. R. Abbott *et al*., Hybridization and speciation. *J. Evol. Biol.* **26**, 229–246 (2013). doi: 10.1111/j.1420-9101.2012.02599.x; pmid: 23323997
25. F. Eroukhmanoff, R. I. Bailey, G.-P. Sætre, Hybridization and genome evolution I: The role of contingency during hybrid speciation. *Curr. Zool.* **59**, 667–674 (2013).
26. J. Mallet, Hybrid speciation. *Nature* **446**, 279–283 (2007). doi: 10.1038/nature05706; pmid: 17361174
27. J. J. Doyle, A. N. Egan, Dating the origins of polyploidy events. *New Phytol.* **186**, 73–85 (2010). doi: 10.1111/j.1469-8137.2009.03118.x; pmid: 20028472
28. L. Li, C. J. Stoeckert Jr., D. S. Roos, OrthoMCL: Identification of ortholog groups for eukaryotic genomes. *Genome Res.* **13**, 2178–2189 (2003). doi: 10.1101/gr.1224503; pmid: 12952885
29. K. Katoh, G. Asimenos, H. Toh, Multiple alignment of DNA sequences with MAFFT. *Methods Mol. Biol.* **537**, 39–64 (2009). doi: 10.1007/978-1-59745-251-9_3; pmid: 19378139
30. O. Penn, E. Privman, G. Landan, D. Graur, T. Pupko, An alignment confidence score capturing robustness to guide tree uncertainty. *Mol. Biol. Evol.* **27**, 1759–1767 (2010). doi: 10.1093/molbev/msq066; pmid: 20207713
31. K. P. Schliep, phangorn: Phylogenetic analysis in R. *Bioinformatics* **27**, 592–593 (2011). doi: 10.1093/bioinformatics/btq706; pmid: 21169378
32. E. Paradis, J. Claude, K. Strimmer, APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics* **20**, 289–290 (2004). doi: 10.1093/bioinformatics/btg412; pmid: 14734327
33. Y. Yu, R. M. Barnett, L. Nakhleh, Parsimonious inference of hybridization in the presence of incomplete lineage sorting. *Syst. Biol.* **62**, 738–751 (2013). doi: 10.1093/sysbio/syt037; pmid: 23736104
34. C. Than, D. Ruths, L. Nakhleh, PhyloNet: A software package for analyzing and reconstructing reticulate evolutionary relationships. *BMC Bioinformatics* **9**, 322 (2008). doi: 10.1186/1471-2105-9-322; pmid: 18662388
35. S. Sturtz, U. Ligges, A. Gelman, R2WinBUGS: A package for running WinBUGS from R. *J. Stat. Softw.* **12**, 1–16 (2005).

ACKNOWLEDGMENTS

This work was financed by Norwegian Research Council grant 199387 to Norwegian University of Life Sciences and Graminor AS to O.-A.O. The phylogenetic analyses were run, in part, on the Bioportal supercomputing facility at the University of Oslo. We thank C.-P. Antoine and two anonymous reviewers for valuable comments on the manuscript and S. Sæbø for help on implementation of the OpenBUGS analyses in R. Ortholog alignments, gene trees, and the OpenBUGS model used for multispecies coalescent modeling can be downloaded from Dryad (data available from the Dryad Digital Repository: http://doi.org/10.5061/dryad.f6c34).

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/345/6194/1250092/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S6
Tables S1 to S5
References (*36*–*45*)
IWGSC Author List

23 December 2013; accepted 20 May 2014
10.1126/science.1250092