nature genetics

A distant upstream enhancer at the maize domestication gene *tb1* has pleiotropic effects on plant and inflorescent architecture

Richard M Clark[1,2], Tina Nussbaum Wagler[1], Pablo Quijada[1] & John Doebley[1]

Although quantitative trait locus (QTL) mapping has been successful in describing the genetic architecture of complex traits[1–4], the molecular basis of quantitative variation is less well understood, especially in plants such as maize that have large genome sizes. Regulatory changes at the *teosinte branched1* (*tb1*) gene have been proposed to underlie QTLs of large effect for morphological differences that distinguish maize (*Zea mays* ssp. *mays*) from its wild ancestors, the teosintes (*Z. mays* ssp. *parviglumis* and *mexicana*)[1,5–7]. We used a fine mapping approach to show that intergenic sequences ~58–69 kb 5′ to the *tb1* cDNA confer pleiotropic effects on *Z. mays* morphology. Moreover, using an allele-specific expression assay, we found that sequences >41 kb upstream of *tb1* act in *cis* to alter *tb1* transcription. Our findings show that the large stretches of noncoding DNA that comprise the majority of many plant genomes can be a source of variation affecting gene expression and quantitative phenotypes.

Although a large number of QTLs have been mapped in various organisms, comparatively few have been subsequently resolved to single genes. In plants, this number now stands at only ~12 (refs. 8,9), and the relationship between sequence variation and phenotypic effects is either completely unknown or inferred but not proven for many of these. In six cases, QTLs have been resolved to either single amino acid substitutions or premature stop codons, and in two cases regulatory changes have been inferred although the exact regulatory sequence has not been identified[6,10]. Our understanding of quantitative inheritance in plants is further limited in that most cloned QTLs are known from either *Arabidopsis thaliana* or rice[8,11], which have relatively small genomes. In contrast, plants such as maize have large genomes with many repetitive elements[12]. Virtually nothing is known about whether and how sequence variation in the 'junk' DNA in such genomes contributes to phenotypic variation[13].

In *Z. mays*, QTLs for morphological traits distinguishing maize from teosinte have been mapped to the *tb1* genomic region[14–16], and genetic complementation indicates that *tb1*, which encodes a transcriptional regulator[6], underlies at least some of these QTLs (ref. 5). Recently, we showed that a large intergenic region of 161 kb separates *tb1* from the nearest upstream gene (PG3, a predicted gene of unknown function; **Fig. 1**) and that sequences extending as far as 90 kb upstream of *tb1* were the target of a selective sweep during maize domestication[17]. Although it is expected that at least one causal site controlling domestication traits lies within the region of the sweep, the exact location and molecular action of such sites remains unknown.

To map the factors controlling phenotypes at *tb1*, we backcrossed a chromosome segment containing *tb1* from a single teosinte plant from Panindicuaro, Mexico, into the maize inbred line W22 for six generations to generate near isogenic lines 1 and 2 (denoted I1 and I2, **Fig. 1**). The I1 and I2 segments harbor Panindicuaro sequence for the region from *tb1* to PG3 (although they differ at the flanking marker *UMC140*). We then screened 20,878 F2 individuals (41,756 meioses) derived from crosses of I1, I2 or derivative lines to W22 for crossovers in and between *tb1* and PG3 (**Supplementary Methods** online). From this work, we recovered 14 recombinant chromosomes with breakpoints in the *tb1* or PG3 ORFs or in the intervening region that we used to characterize quantitative variation at *tb1* (segments I6–I19; **Fig. 1**, and **Supplementary Table 1** and **Supplementary Fig. 1** online). Lines I3–I5 were recovered from segregating populations as nonrecombinant controls.

To test for phenotypic effects of the introgressed Panindicuaro chromosome segments, we crossed the introgression lines to W22 and grew F2 progeny from 24 families in fully randomized field trials over 2 years (replicate families for five lines were grown between or within years: for example, I6-1 and I6-2, **Fig. 1**). A total of 2,146 F2 individuals of known genotype were then scored for eight traits previously associated with *tb1* (**Table 1**). Within each family, we determined trait values of individuals for the Panindicuaro introgression class (I/I), the W22 maize class (W/W) and the heterozygous class (I/W) (**Supplementary Fig. 2** online). For each trait, means for the W/W control class varied significantly among families ($P < 0.001$), suggesting that family effects due to either genetic background (lines are not completely isogenic), maternal influences (for example, maternal ear quality) or environmental influences (for example, different years) affected variation in absolute trait values. We

[1]Laboratory of Genetics, University of Wisconsin, Madison, Wisconsin 53706, USA. [2]Present address: Max Planck Institute for Developmental Biology, D-72076 Tübingen, Germany. Correspondence should be addressed to J.D. (jdoebley@wisc.edu).

Received 4 January; accepted 16 March; published online 23 April 2006; doi:10.1038/ng1784

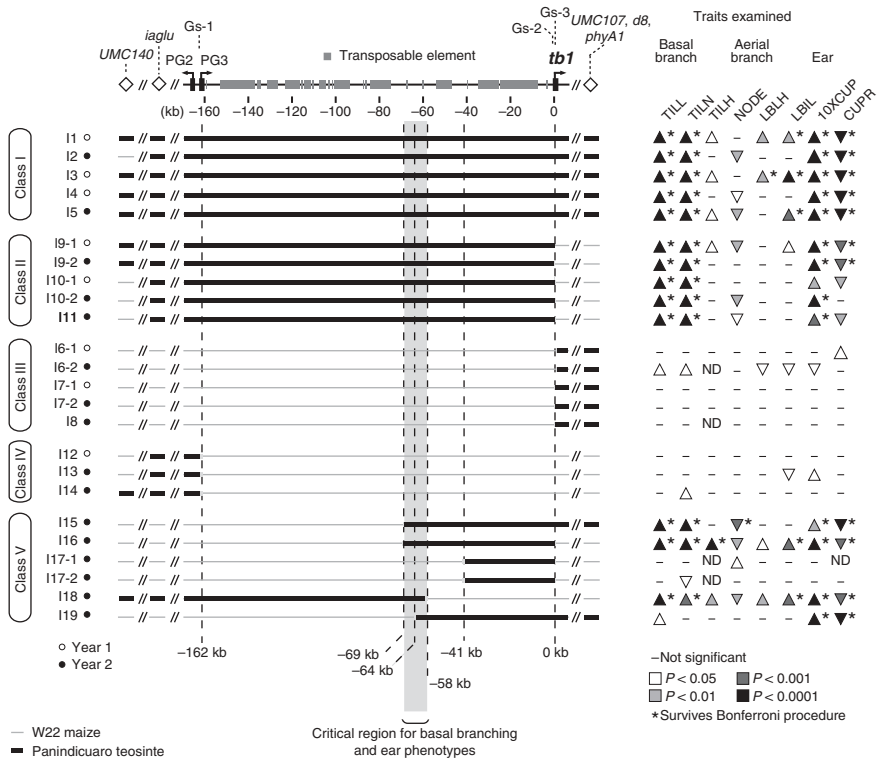

Figure 1 Introgression lines and effects on traits. Sequence features and flanking genetic markers for the *tb1* region are indicated at top; breakpoints for introgression chromosomes are indicated below. Statistical effects of introgression lines are indicated at right. Where significant differences were observed, up or down triangles indicate the direction of the effect of the homozygous introgression class relative to the within-family homozygous W22 maize class. Degree of significance is indicated by shading (see key), and *P*-values were not determined (ND) when mean homozygous class sizes were <10. Replicates are indicated by dashes followed by replicate number (for example, I6-1 and I6-2 are replicates of the I6 introgression). Markers *UMC140* and *UMC107* have previously been mapped 3.9 cM 5′ to *tb1* and 1.3 cM 3′ to *tb1*, respectively[5]. In our crosses, the distances between the genotyping markers Gs-1 and Gs-2 (162.7 kb) and Gs-2 and Gs-3 (1.4 kb) were 0.63 cM and 0.16 cM, respectively (data not shown). The critical region for basal branching and ear phenotypes (traits TILL, TILN, 10XCUP and CUPR) is indicated by the gray column (middle).

therefore tested for significant differences in phenotype among the genotypic classes I/I, I/W and W/W within families (**Fig. 1** and **Supplementary Table 2** online).

For 13 of 19 introgressed segments, we observed highly significant effects ($P < 0.001$) on the production of basal branches (TILL and TILN traits) that remained significant after correction for multiple tests, and these effects were consistent when replicates were performed (**Fig. 1**). A nearly identical pattern of effects was observed for ear traits (10XCUP and CUPR), and the direction of effects of Panindicuaro sequences was as expected from earlier studies (**Table 1**). Whereas introgressed segments with Panindicuaro sequence encompassing the entire *tb1* to PG3 region (class I introgressions) or the region 5′ to the *tb1* ORF (class II introgressions) affected basal branch and ear phenotypes, segments with Panindicuaro sequence 5′ to PG3 (class IV introgressions) did not have consistent or highly significant effects on phenotypes (no significant tests at $P < 0.05$ survived correction for multiple testing). Similarly, introgressions containing Panindicuaro sequence extending 3′ to *tb1* (class III introgressions) that include parts or all of the *tb1* coding sequence (**Supplementary Fig. 1**) and the linked regulatory genes *d8* (ref. 18) and *phyA1* (ref. 19) showed no consistent effect on phenotypes. These findings localize sequences affecting phenotypes to the segment between *tb1* and PG3 and suggest that coding sequence polymorphisms distinguishing the Panindicuaro and W22 haplotypes do not contribute to variation in basal branch and ear architecture. An examination of class V recombinants, which have breakpoints within the PG3-*tb1* intergenic region, further localized the causal region for both ear and basal branch phenotypes to ~58–69 kb upstream of *tb1* (compare I15–I19).

Further inspection of the effects associated with individual introgression lines provides additional insights. First, comparison of I19 to I18 suggests that functional polymorphisms for basal branching and ear phenotypes are located ~64–69 kb and ~58–64 kb upstream of *tb1*, respectively. Although this finding requires additional validation, it hints that *tb1* has a modular control region with multiple tissue-specific elements. Second, although aerial branch traits (NODE, LBLH and LBIL) have weaker effects than basal branch and ear traits, their effects show the same trend, suggesting that the causative site(s) for these traits also lie(s) in the ~58–69 kb upstream region. Finally, the minimal introgression to have strong effects on basal branching and ear

Table 1 Traits analyzed and phenotypic effects

Trait	Description	Effect of substituting teosinte for maize haplotypes on traits	
		tb1 chromosome region (earlier studies[a])	*tb1* locus (this study[b])
TILL	Length of all tillers[c]/height of main culm	↑	↑
TILN	Number of tillers per plant	↑	↑
TILH	Mean tiller height/height of main culm	ND	(↑)
NODE	Number of nodes in uppermost lateral branch	↓[d]	(↓)
LBLH	Length of uppermost lateral branch	↑[d]	(↑)
LBIL	Mean internode length of uppermost lateral branch	↑	(↑)
10XCUP	Length between 10 cupules at center of ear	↑[e]	↑
CUPR	Cupules (or kernels of grain) per row in the ear	↓	↓

[a]Where traits were not examined previously[5,16,29], the direction of effect is indicated as not determined (ND). [b]Where effects in the current study were not highly significant or consistent across introgressed segments, direction of effects are indicated in parentheses. [c]We refer to a tiller as a branch at ground level (also called a basal branch). [d]J.D., unpublished data, available upon request. [e]Direction of effect inferred from a nearly identical trait measure (CUPL) described previously[5].

Table 2 Effects of I16 introgression on phenotypes

Trait	Trait mean ± 2 s.e.m. (sample size) I16/I16	I16/W22	W22/W22	Phenotypic effect[a]	*P* value
TILL	1.71 ± 0.25 (27)	1.01 ± 0.17 (30)	0.56 ± 0.18 (32)	1.15	7.6×10^{-9}
TILN	2.41 ± 0.27 (27)	1.57 ± 0.23 (30)	1.28 ± 0.31 (32)	1.13	3.4×10^{-6}
TILH	0.70 ± 0.06 (27)	0.65 ± 0.06 (29)	0.42 ± 0.08 (25)	0.28	4.0×10^{-6}
NODE	7.09 ± 0.55 (23)	8.00 ± 0.44 (30)	8.19 ± 0.34 (27)	–1.10	0.0024
LBLH (cm)	22.71 ± 2.71 (23)	25.62 ± 1.54 (30)	21.57 ± 1.84 (27)	1.14	0.0119
LBIL (cm)	3.21 ± 0.29 (23)	3.24 ± 0.21 (30)	2.65 ± 0.23 (27)	0.56	8.3×10^{-4}
10XCUP (mm)	44.68 ± 0.74 (22)	42.33 ± 0.78 (27)	41.00 ± 0.93 (25)	3.68	2.4×10^{-7}
CUPR[b]	30.00 ± 0.89 (16)	32.95 ± 1.19 (20)	33.33 ± 1.04 (21)	–3.33	1.4×10^{-4}

[a]The effect of substituting two copies of the I16 introgression for two copies of the W22 chromosomal segment. [b]This subtracts 3.33 cupules from each rank. W22 has ~7 ranks of cupules and each cupule contains two kernels. Thus, I16 reduces the number of kernels per ear by about 46.

phenotypes (I16) affected all traits (**Table 2**). As compared to pure W22, this line adds one tiller as long as the main stalk and subtracts about 46 kernels from the ear.

Taken together, these findings show that the QTLs that map to the *tb1* region do not result from a cluster of linked genes. Rather, a single noncoding segment ~58–69 kb upstream of *tb1* controls multiple traits pleiotropically, and we found no evidence that tightly linked loci, such as *d8* and *phyA1*, contribute to QTL previously associated with *tb1* (compare I6–I8). Our findings thus confirm that loci with large and diverse pleiotropic effects can and have contributed to crop domestication (see also refs. 20,21). An unanticipated result was that the Panindicuaro introgression did not cause a partial conversion of the ear into tassel (data not shown) as seen with another teosinte *tb1* introgression[5]. This may indicate that Panindicuaro has a weak allele for this trait that is dependent on a teosinte genetic background to affect inflorescence sex[16].

tb1 RNA levels are higher in ear primordia with a W22 maize allele than in those with a teosinte allele, and the difference in expression level seems to underlie differences in phenotypes[6]. Our finding that at least one factor controlling the differences in phenotypes is located as far as ~58–69 kb upstream of the *tb1* transcribed sequence is consistent with a *cis*-regulatory basis for QTLs that map to *tb1*. However, in plants, virtually all characterized *cis*-regulatory sequences are located very close to genes, and in only one case, at the *b1* locus in maize, has a regulatory element been proposed to act over a longer physical distance (~8–45 kb)[22]. We therefore also considered the alternative possibility that an unrecognized factor (either a protein or a functional RNA species) derived from the upstream *tb1* region affects *tb1* RNA levels and phenotypes in *trans*.

To distinguish between these possibilities, we performed an allele-specific expression assay on RNA samples obtained from immature ears from three heterozygous introgression stocks that differed in the extent of Panindicuaro sequence for the *tb1* intergenic region (**Fig. 2** and **Supplementary Table 3** online). The assay took advantage of a 6-bp insertion or deletion (indel) in the *tb1* ORF that distinguishes the W22 and Panindicuaro alleles (**Supplementary Methods**). We compared relative *tb1* expression levels among three introgressions after normalization against the full Panindicuaro segment (I1) that was common to each heterozygous stock. Using this design, differences in *tb1* RNA levels produced from introgressions in contrasting stocks are expected if upstream factor(s) control(s) *tb1* transcription in *cis*, but not in *trans*. Consistent with a *cis* effect, relative message levels differed between the three stocks ($P < 10^{-10}$). We did not observe expression differences between W22 and I17 ($P = 0.90$), but contrasts between

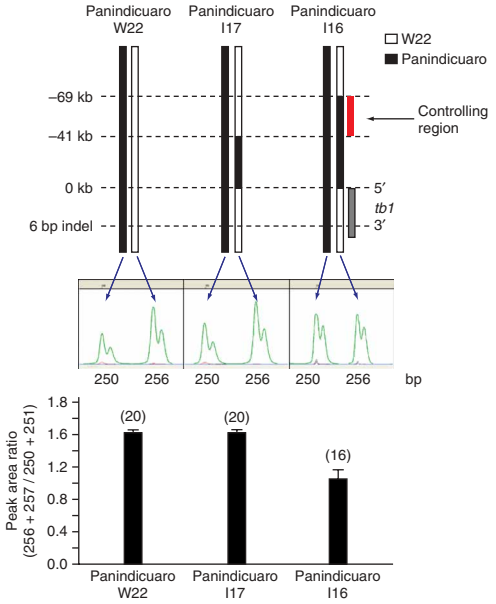

Figure 2 Allele-specific *tb1* expression assay. *tb1* allele–specific expression was assayed for three heterozygous stocks (top): Panindicuaro/W22, Panindicuaro/I17 and Panindicuaro/I16 (transcription from the W22, I17 and I16 segments can thereby be compared to the large Panindicuaro segment, I1, that is shared among stocks). I16 and I17 were analyzed because they represent the minimal Panindicuaro introgressions that either affect or do not affect traits. Expression was assayed using immature ear RNA collected at the time that ear phenotypes are being established (cupule formation, 10XCUP and CUPR traits). RT-PCR with total RNA and a fluorescently tagged primer was used to measure the relative abundance of *tb1* mRNA carrying the 250-bp versus 256-bp fragments that differ for a 6-bp indel in the ORF. *Taq* polymerase +A activity produced additional 251-bp and 257-bp peaks on an ABI 3700 fragment analyzer (middle, and **Supplementary Methods**). For the Panindicuaro/W22 and Panindicuaro/I17 stocks, the ratio of the area of the 256/257-bp peaks to that of the 250/251-bp peaks was ~1.6, indicating that mRNA from the W22 and I17 chromosomes was more abundant than that from the Panindicuaro chromosome (bottom). For the Panindicuaro/I16 stock, the 250/251 and 256/257 peaks had equivalent areas, indicating equivalent mRNA levels for the I16 and Panindicuaro chromosomes. Sample sizes (*n*) are indicated in parentheses, and error bars represent ± 2 s.e.m. (bottom panel).

I16 and both W22 and I17 were highly significant with reduced expression from the I16 introgression ($P < 10^{-10}$ for each test). I16 and I17 harbor identical breakpoints in the *tb1* ORF (**Supplementary Fig. 1**) and differ only in that I16 has Panindicuaro sequence for the region ~41–69 kb upstream of *tb1* (**Fig. 1**). These findings localize *cis* factor(s) controlling *tb1* message levels between –41 kb and –69 kb upstream of *tb1* to an interval that contains the region that, according to phenotypic analysis, controls *tb1*-related phenotypes.

Even among close relatives, genome size can vary markedly in plants[23,24]. In large part, these differences result from copy number variation in mobile repetitive elements; in cereals, moderately to highly repetitive tracts of long nongenic sequences typify species with large genomes such as maize, barley and wheat[12,25–27]. Although differences in DNA content have been correlated with life history traits[13], the contribution of such intergenic sequences to quantitative variation has been unclear. We have shown that such intergenic sequences, often considered junk DNA, can harbor QTL and *cis*-regulatory sequences that act over long physical distances. The causative region for phenotypes and expression differences in the *tb1* upstream intergenic region harbors a mixture of repetitive and unique sequences that is representative of much of the maize genome. Establishing whether functional variation in these sequences is common or rare will be important for understanding the range and types of sequence variation that contribute to phenotypic variation. These results also indicate that linkage disequilibrium mapping in maize should look beyond coding sequences and their nearby regulatory regions.

METHODS

See **Supplementary Methods** for biological material, backcrossing of Panindicuaro chromosome segments into W22 maize, identification and genealogies of introgressed segments (see also **Supplementary Table 1**), characterization of recombination breakpoints in the *tb1* region, detection of allele-specific *tb1* transcription, primers used for genotyping (**Supplementary Table 4** online), genotyping assays (**Supplementary Table 5** online) and genomic annotation.

Phenotypic characterization of introgressed chromosome segments. F2 progeny from introgressed Panindicuaro chromosome segments crossed with W22 maize were grown in a field in Madison, Wisconsin, in the summers of 2003 and 2004 for phenotypic characterization. Plants were grown with 0.9 m spacing on each side using a fully randomized design. Between 120 and 160 seeds were planted for each family, depending on seed availability and quality. Homozygous introgression (I/I) and homozygous W22 individuals (W/W), as well ~30–40 heterozygous individuals (I/W) chosen at random within each family, were phenotyped for the traits given in **Table 1**. In cases where plants lacked basal branches (tillers), the TILH trait was not calculated, and sample sizes for late traits (for example, ear traits) were reduced as a result of pest damage.

Statistical analysis. Statistical analyses were performed using R (ref. 28) or with JMP-IN software (SAS Institute). The significance of differences in trait means between genotypic classes within families segregating for introgressed Panindicuaro chromosome segments was assessed using one-way analysis of variance (ANOVA) for traits with approximately normal distributions (NODE, LBLH, LBIL, 10XCUP and CUPR) and the Kruskal-Wallis test where deviations from normality were observed (TILL, TILN and TILH). To correct for multiple tests among families for each trait, the conservative Bonferroni procedure was performed. One-way ANOVA was used to test for differences in *tb1* RNA accumulation between test stocks.

URLs. R software: http://cran.r-project.org.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

We thank L. Rentmeester, B. Oldham, A.-C. Thuillet, K. Bomblies, Y. Vigouroux and other laboratory members for help with fieldwork and for suggestions that improved the experimental design, and K. Bomblies and S. Balasubramanian for comments on the manuscript. This work was supported by US National Institutes of Health grant GM-58816 (to J.D.) and National Institutes of Health award F32 GM-65008 (to R.M.C.).

AUTHOR CONTRIBUTIONS

R.M.C. and J.D. conceived the experiments; R.M.C., T.N.W., P.Q., and J.D performed the experiments; and R.M.C. and J.D. analyzed the data and wrote the manuscript.

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Published online at http://www.nature.com/naturegenetics
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/

1. Doebley, J. The genetics of maize evolution. *Annu. Rev. Genet.* **38**, 37–59 (2004).
2. Mackay, T.F. The genetic architecture of quantitative traits. *Annu. Rev. Genet.* **35**, 303–339 (2001).
3. Mackay, T.F. The genetic architecture of quantitative traits: lessons from *Drosophila*. *Curr. Opin. Genet. Dev.* **14**, 253–257 (2004).
4. Tanksley, S.D. The genetic, developmental, and molecular bases of fruit size and shape variation in tomato. *Plant Cell* **16** (Suppl.), S181–S189 (2004).
5. Doebley, J., Stec, A. & Gustus, C. *teosinte branched1* and the origin of maize: evidence for epistasis and the evolution of dominance. *Genetics* **141**, 333–346 (1995).
6. Doebley, J., Stec, A. & Hubbard, L. The evolution of apical dominance in maize. *Nature* **386**, 485–488 (1997).
7. Wang, R.L., Stec, A., Hey, J., Lukens, L. & Doebley, J. The limits of selection during maize domestication. *Nature* **398**, 236–239 (1999).
8. Alonso-Blanco, C., Mendez-Vigo, B. & Koornneef, M. From phenotypic to molecular polymorphisms involved in naturally occurring variation of plant development. *Int. J. Dev. Biol.* **49**, 717–732 (2005).
9. Salvi, S. & Tuberosa, R. To clone or not to clone plant QTLs: present and future challenges. *Trends Plant Sci.* **10**, 297–304 (2005).
10. Frary, A. *et al. fw2.2*: a quantitative trait locus key to the evolution of tomato fruit size. *Science* **289**, 85–88 (2000).
11. Koornneef, M., Alonso-Blanco, C. & Vreugdenhil, D. Naturally occurring genetic variation in *Arabidopsis thaliana*. *Annu. Rev. Plant Biol.* **55**, 141–172 (2004).
12. SanMiguel, P. *et al.* Nested retrotransposons in the intergenic regions of the maize genome. *Science* **274**, 765–768 (1996).
13. Meagher, T.R. & Vassiliadis, C. Phenotypic impacts of repetitive DNA in flowering plants. *New Phytol.* **168**, 71–80 (2005).
14. Doebley, J. & Stec, A. Genetic analysis of the morphological differences between maize and teosinte. *Genetics* **129**, 285–295 (1991).
15. Doebley, J., Stec, A., Wendel, J. & Edwards, M. Genetic and morphological analysis of a maize-teosinte F2 population: implications for the origin of maize. *Proc. Natl. Acad. Sci. USA* **87**, 9888–9892 (1990).
16. Lukens, L.N. & Doebley, J. Epistatic and environmental interactions for quantitative trait loci involved in maize evolution. *Genet. Res.* **74**, 291–302 (1999).
17. Clark, R.M., Linton, E., Messing, J. & Doebley, J.F. Pattern of diversity in the genomic region near the maize domestication gene *tb1*. *Proc. Natl. Acad. Sci. USA* **101**, 700–707 (2004).
18. Thornsberry, J.M. *et al. Dwarf8* polymorphisms associate with variation in flowering time. *Nat. Genet.* **28**, 286–289 (2001).
19. Sheehan, M.J., Farmer, P.R. & Brutnell, T.P. Structure and expression of maize phytochrome family homeologs. *Genetics* **167**, 1395–1405 (2004).
20. Simons, K.J. *et al.* Molecular characterization of the major wheat domestication gene. *Q. Genetics* **172**, 547–555 (2006).
21. Wang, H. *et al.* The origin of the naked grains of maize. *Nature* **436**, 714–719 (2005).
22. Stam, M. *et al.* The regulatory regions required for *B′* paramutation and expression are located far upstream of the maize *b1* transcribed sequences. *Genetics* **162**, 917–930 (2002).
23. Wendel, J.F., Cronn, R.C., Johnston, J.S. & Price, H.J. Feast and famine in plant genomes. *Genetica* **115**, 37–47 (2002).
24. Bennetzen, J.L., Ma, J. & Devos, K.M. Mechanisms of recent genome size variation in flowering plants. *Ann. Bot. (Lond.)* **95**, 127–132 (2005).
25. Rostoks, N. *et al.* Genomic sequencing reveals gene content, genomic organization, and recombination relationships in barley. *Funct. Integr. Genomics* **2**, 51–59 (2002).
26. Wicker, T. *et al.* Analysis of a contiguous 211 kb sequence in diploid wheat (*Triticum monococcum* L.) reveals multiple mechanisms of genome evolution. *Plant J.* **26**, 307–316 (2001).
27. Wicker, T. *et al.* A detailed look at 7 million years of genome evolution in a 439 kb contiguous sequence at the barley Hv-eIF4E locus: recombination, rearrangements and repeats. *Plant J.* **41**, 184–194 (2005).
28. Ihaka, R. & Gentleman, R.R. A language for data analysis and graphics. *J. Comput. Graph. Stat.* **5**, 299–314 (1996).
29. Doebley, J. & Stec, A. Inheritance of the morphological differences between maize and teosinte: comparison of results for two F2 populations. *Genetics* **134**, 559–570 (1993).