Population genetics of foxtail millet and its wild ancestor

Chunfang Wang1,2,4†, Jinfeng Chen3†, Hui Zhi1,2, Lu Yang3, Wei Li2, Yongfang Wang2, Haiquan Li2, Baohua Zhao4, Mingsheng Chen3*, Xianmin Diao1,2*

Abstract

Background: Foxtail millet (Setaria italica (L.) P. Beauv.), one of the most ancient domesticated crops, is becoming a model system for studying biofuel crops and comparative genomics in the grasses. However, knowledge on the level of genetic diversity and linkage disequilibrium (LD) is very limited in this crop and its wild ancestor, green foxtail (Setaria viridis (L.) P. Beauv.). Such information would help us to understand the domestication process of cultivated species and will allow further research in these species, including association mapping and identification of agricultural significant genes involved in domestication.

Results: In this study, we surveyed DNA sequence for nine loci across 50 accessions of cultivated foxtail millet and 34 of its wild progenitor. We found a low level of genetic diversity in wild green foxtail (θ = 0.0059), Watterson’s estimator of θ. Despite of a 55% loss of its wild diversity, foxtail millet still harbored a considerable level of diversity (θ = 0.0027) when compared to rice and sorghum (θ = 0.0024 and 0.0034, respectively). The level of LD in the domesticated foxtail millet extends to 1 kb, while it decayed rapidly to a negligible level within 150 bp in wild green foxtail. Using coalescent simulation, we estimated the bottleneck severity at k = 0.6095 when r/θ = 1. These results indicated that the domestication bottleneck of foxtail millet was more severe than that of maize but slightly less pronounced than that of rice.

Conclusions: The results in this study establish a general framework for the domestication history of foxtail millet. The low level of genetic diversity and the increased level of LD in foxtail millet are mainly caused by a population bottleneck, although gene flow from foxtail millet to green foxtail is another factor that may have shaped the pattern of genetic diversity of these two related gene pools. The knowledge provided in this study will benefit future population based studies in foxtail millet.
domestication bottleneck and/or reduction of effective recombination rate due to self-fertilization [5]. Additionally, selection can have a similar effect on LD in the target region [13]. Studies that focus on the level and organization of genetic variation in these major crops are essential for our understanding of the process of domestication and are instructive for crop improvement or other research that is based on these population parameters, such as association mapping [2,12,14].

To study the patterns of genetic diversity within and between populations, as well as to trace the demographic history of crops and their wild relatives, multilocus surveys of population sequence data have been widely used in recent years [3,6,12,15]. Since selection acts on some but not all genes in genome, a multilocus sample increases the probability that both loci that are under selection and those that are neutrally-evolving will be sampled. Multi-locus sampling is also necessary to understand the demographic history of populations, and enables targets of natural or artificial selection to be more reliably identified. Many studies have been carried out to investigate nucleotide diversity in plants, yet only a few have been focused on their demographic history. Understanding the demographic history of populations will also help in interpretation of population genetic neutrality tests, which are based on the comparison of observed and expected polymorphism patterns under the neutral equilibrium model (NE) [16,17]. This model assumes random outcrossing and a large stable population size, assumptions which may not be valid in many domesticated crops [16]. Without a reliable knowledge of the demographic history, it is hard to interpret the statistical results of neutrality tests.

Foxtail millet (Setaria italica (L.) P. Beauv) has been a very important cereal since ancient times in Eurasia and has contributed greatly to human civilization both in Asia and Europe [18]. The most recent archaeological evidence demonstrates that foxtail millet is one of the most ancient crops as its domestication in China dates back to 8,700 years ago [19]. With the rapid development of maize and other modern crops, foxtail millet has gradually become a minor crop in the last 80 years, but is nonetheless still widely cultivated in Asia, Europe, North America, Australia and North Africa as grain food or forage [20]. Green foxtail (Setaria viridis (L.) Beauv.), a weed distributed worldwide, is the presumed wild progenitor of domesticated foxtail millet, based on cytological evidence and RAPD, AFLP, and other markers [21,22]. But the number of domestication centers and the age of domestication of foxtail millet remains controversial [23]. As revealed by earlier studies, genetic diversity is low in foxtail millet [24], yet these analyses were based on genetic markers that only represent a subset of the possible information that can be gained regarding genomic polymorphism [16]. However, research at the sequence level is very limited in foxtail millet and its close relatives, although sequencing of the foxtail millet genome is now nearly completed [25]. This will provide another cereal model system for comparative and functional genomics and model for studying other biofuel crops such as switchgrass (Panicum virgatum), and napiergrass (Pennisetum purpureum). Studies on domestication or other agricultural related characteristics are ongoing [26-28]. In this study, we conduct a multilocus analysis of nucleotide variation in foxtail millet and green foxtail to reveal the pattern of genetic diversity within and between these two species and to establish a population genetic framework for further analysis of the effects of domestication in foxtail millet.

Results
Nucleotide diversity and neutrality test
We collected 50 cultivated foxtail millets and 34 wild green foxtails to represent the broad diversity of these two species (Table 1). Nine loci randomly selected from the genome were used to survey genetic diversity from both species (Table 2 & Additional file 1, Fig S1). The alignment length of each locus after excluding gaps and missing data varied from 431 bp to 996 bp. All nine loci contained both coding and noncoding sequences. However, ninety percent of the final alignments were non-coding sequence because the primers were designed to amplify a large proportion of intron fragments. A total of 160 SNPs were found in the nine loci across the 84 accessions, with an average density of 52 bp/SNP. The wild progenitor had more SNPs (147 SNPs with a density of 43 bp/SNP) compared to the cultivars (75 SNPs with a density of 87 bp/SNP). Indel polymorphisms were infrequent across loci and most had a small size of 1 to 3 basepairs. Three large indels were present in ADTY (143 bp), UPL (112 bp) and TIFIIE (43) bp. These indel polymorphisms were excluded from the subsequent analysis.

The nucleotide diversity of the nine loci for each species is summarized in Table 3. For both $\theta$ and $\pi$, the values for each locus were slightly lower than values for silent sites, presumably due to strong functional constraint in coding regions. Considering individual loci, the most variable gene was MDEH with a mean of 0.01 for $\theta_{sil}$ across all accessions. Compared to MDEH, PP2C was the least variable site with a value of only 0.0016 for $\theta_{sil}$. At the taxon level, the diversity of wild green foxtail (mean $\theta_{sil}$ 0.0059) was much higher than the domesticated foxtail millet (mean $\theta_{sil}$ 0.0027). On average, the cultivars lost 55% of the diversity harbored by the wild progenitor during the domestication process. Some loci in cultivars, such as ADTY and PP2C, lost more than 86% and 75% of their diversity in the wild progenitor.
| Taxon | Accessions* | Origin | Code | Variety Name |
|-------|-------------|--------|------|--------------|
| Setaria italica ssp. italica |           |        |      |              |
| 00014654 | Gongyang, Yunnan, China | g137 | Yunnan NoMiXiaoChuiNiu |
| 00014528 | Anyang, Henan, China | g141 | XiaoHongGu |
| 00012913 | Qihe, Shandong, China | g133 | HongMaoGu |
| 00014617 | Ninghai, Zhejiang, China | g136 | NingHaiGuZi |
| 00014570 | Shaanxi, China | g135 | LaoLaiBian |
| 00014625 | Yuanling, Hunan, China | n793 | HuangBangTou |
| 00022285 | Leshan, Sichuan, China | n1549 | LeShanBaiNuo |
| 00005252 | Fanshi, Shenxi, China | n286 | LiuLengGu |
| 00001532 | Kangping, Liaoning, China | n107 | HeiNianGu |
| 00000077 | Nenjiang, Heilongjiang, China | n6 | HeiLongjiangCaoPiYiDaoBaQi |
| 00000158 | Baiquan, Heilongjiang, China | n9 | HeiLongjiangFuoDingZhu |
| 00021705 | Gansu, China | n1473 | DaLiangZhouGu |
| 00018757 | Hualong, Qinghai, China | n1187 | XiaoHongGu |
| 00018782 | Minhe, Qinghai, China | n1199 | DailiangGu |
| 00018783 | Minhe, Qinghai, China | n1200 | DailiangGu |
| 00014609 | Xinjiang, China | n779 | XinjiangEminGuzi |
| 00014612 | Xinjiang, China | n782 | XinjiangShawangzi |
| 00018751 | Guyuan, Ningxia, China | n1185 | XiaoMiaoGu |
| PI 433458 | Taiwan, China | g45 | Megalaoud |
| PI 433396 | Taiwan, China | g50 | Pagarugareano |
| PI 433465 | Taiwan, China | g54 | Balahigh |
| PI 433481 | Taiwan, China | g121 | Ishsumsum |
| 00014695 | Japan | n822 | ZhaoheliNuo |
| PI 464241 | Bihar, India | g27 | I.S. 83 |
| PI 454359 | Maharashtra, India | g32 | I.S. 210-B |
| PI 464157 | Andhra Pradesh, India | g33 | I.S. 1 |
| PI 464287 | Kerala, India | g36 | I.S. 141 |
| PI 464457 | Punjab, India | g39 | I.S. 304 |
| PI 472256 | Nepal | g16 | Nepal |
| PI 269972 | Pakistan | g21 | Pakistan |
| PI 251395 | Iran | g22 | BAJRA |
| PI 220634 | Afghanistan | g18 | Afghanistan |
| PI 473601 | Lebanon | g6 | Lebanon |
| Ames 21521 | Kazakhstan | g14 | VYSOKOROSKY |
| PI 177543 | Turkey | g5 | KUMDARI |
| 00014986 | Netherland | n836 | Kraftborn |
| 00015029 | German | n840 | Set64/82 |
| PI 442551 | Belgium | g51 | Belgium |
| PI 283988 | Spain | g15 | Spain |
| PI 464567 | Switzerland | g68 | I.S. 663 |
| PI 290459 | Hungary | g3 | Hungary |
| 00014980 | Romania | g139 | Romania |
| 00014968 | Russian | g132 | Russian |
| PI 464544 | Ethiopia | g28 | Ethiopia |
| PI 209909 | South Africa | g59 | SouthAfrica |
| PI 517051 | Morocco | g7 | Morocco |
| PI 015040 | USA | n842 | USA |
| PI 015042 | USA | n843 | USA |
| 00015044 | USA | n844 | USA |
Under the assumption of a neutral equilibrium model, the population silent diversity ($\theta_{sil}$) and population divergence (Ks) should be correlated with each other across the loci. We calculated the Pearson correlation between them to test if both species were under neutral evolution. The correlation for wild species is high and significant (pearson cor = 0.93, p value = 0.0001), whereas it is not significant for the cultivar (pearson cor = 0.32, p = 0.19). When excluding the most diverged locus ADTY, significant positive correlations were found in both wild (pearson cor = 0.89, p = 0.002) and cultivars species (pearson cor = 0.69, p = 0.03). The significant correlations suggested that most of loci were under neutral evolution in both species, except for ADTY in cultivars. This locus may be under directional selection or influenced by a selective sweep on neighboring loci.

We used a series of neutrality tests to determine the fit of our data to a neutral equilibrium model. Tajima’s D and Fu and Li’s D* and F* were used to examine the allele frequency spectrum in polymorphism data for each locus. In the wild population, all the loci showed a negative value for D, D* and F*. Two loci, SIGT and MDEH, had a significant negative value for both tests. However, loci in the cultivars exhibited positive values or less negative values than the wild species for both tests except for DACP and ADTY. The results were unsurprising since we expected D to be higher when the species had experienced a recent population bottleneck[29]. The loss of low

Table 1: Plant materials used in this study (Continued)

| Setaria italica ssp. viridis | 8199W64 | England | q140England |
|-------------------------------|---------|---------|-------------|
| 8200W65                      | France  | q131France |
| 8201W66                      | German  | q94German |
| 8012Q10                      | Bashang, Hebei, China | q90Hebei |
| 8019Q17                      | Baoding, Hebei, China | q91Hebei |
| 8003Q03                      | Qinghai, China | q92Qinghai |
| 8044Q35                      | Chifeng, Neimeng, China | q88Neimeng |
| 8045Q36                      | Heilongjiang, China | q95Heilongjiang |
| 8049Q39                      | Liaoning, China | q96Liaoning |
| 8004Q24                      | Lanzhou, Gansu, China | q93Gansu |
| 8058Q46                      | Changzhi, Shanxi, China | q143Shanxi |
| 8005Q05                      | Xinjiang, China | q89Xinjiang |
| 8066Q54                      | Shandong, China | q87Shandong |
| 8009Q7-1                     | Kunming, Yunnan, China | q142Yunnan |
| GZ001                        | Guangzhou, China | qGuangzhou |
| UC9001                       | Uzbekistan | qUzbekistan |
| UC9012                       | Japan | qJapan |
| ZH001                        | Zhejiang, China | qZhejiang1 |
| Ames 21519                   | Russian | q77Russian |
| Ames 21520                   | Russian | q134Russian |
| PI 202407                    | Chile | q80Chile |
| PI 204624                    | Turkey | q75Turkey |
| PI 204625                    | Turkey | q76Turkey |
| PI 204628                    | Turkey | q81Turkey |
| PI 204727                    | Turkey | q83Turkey |
| PI 204730                    | Turkey | q86Turkey |
| PI 212625                    | Afghanistan | q73Afghanistan |
| PI 221960                    | Afghanistan | q74Afghanistan |
| PI 223677                    | Iran | q78iran |
| PI 230134                    | Iran | q79iran |
| PI 230135                    | Iran | q85iran |
| PI 408810                    | Changchun, Jilin, China | q138Jilin |
| PI 408811                    | Shaanxi, China | q84Shaanxi |
| PI 442553                    | Belgium | q148Belgium |

* Accession numbers started with PI or Ames were from the National Germplasm Resources Laboratory, Beltsville, Maryland, USA; Others samples were from Chinese National Germplasm Bank in Institute of Crop Sciences, Chinese Academy of Agricultural Sciences.
frequency alleles during the bottleneck process will increase the D for cultivars. On the other hand, a significant negative D value means the locus has an excess of low frequency variants due to population size expansion and/or purifying or directional selection. We also used a multilocus HKA test to determine whether the level of polymorphism and divergence were correlated across the loci. A significant result was found when we used all nine loci ($X^2 = 12.26, p < 0.007$). As shown in Figure 1, locus ADTY contributed a large proportion to the overall

### Table 2 Summary of the genes surveyed and the primer sequences used in the study

| Gene Name | Rice/Millet | Putative function | Primers |
|-----------|-------------|-------------------|---------|
| DACP      | LOC_Os01g21160.2/ECG12491 | Dihydropolyamines-residue acetyltransferase | F:5′-3′-CACTTCTCCTCCGTTCCCTCAT |
| Sigt      | LOC_Os01g09120.2/ECG13421 | Signal transducer/two-component sensor molecule | F:5′-3′-ATCCACGACTCAGTTCCCTCAT |
| Adty      | LOC_Os02g56550.1/ECG12081 | ATP-dependent transporter YFL028C | F:5′-3′-TCCATACAGGGCGATTCTT |
| Pp2c      | LOC_Os03g60505.1/ECG12551 | Catalytic/protein phosphatase type 2C | F:5′-3′-TGTGACCGCTCGCTTCAG |
| Sps1      | LOC_Os08g20660.2/ECG12114 | Sucrose-phosphate synthase 1 | F:5′-3′-GACGCCCTTCTGGATACTCT |
| Upl       | LOC_Os10g41360.1/ECG11973 | Ubiquitin-protein ligase | F:5′-3′-GACGCCCTTCTGGATACTCT |
| Tififf    | LOC_Os10g10920.3/ECG13446 | Transcription initiation factor IIF, alpha subunit | F:5′-3′-GACGCCCTTCTGGATACTCT |
| Tran      | LOC_Os11g37980.1/ECG12732 | Transferase, transferring glycosyl groups | F:5′-3′-GACGCCCTTCTGGATACTCT |
| Mdeh      | LOC_Os12g43630.1/ECG13245 | Malate dehydrogenase, glyoxysomal precursor | F:5′-3′-GACGCCCTTCTGGATACTCT |

### Table 3 Summary of nucleotide diversity and neutrality tests

| Loci          | Length* | Silent | Accessions* | S | π | nsilent | θ | θsilent | D | D* | F* | p | Rm | Hap | Hdiv |
|---------------|---------|-------|-------------|---|---|--------|---|---------|---|----|----|---|----|-----|-----|
| Setaria italica ssp. viridis |         |       |             |   |   |        |   |         |   |    |    |   |    |     |     |
| DACP          | 1014    | 944.833 | 29          | 27 | 0.0062 | 0.0067 | 0.0068 | 0.0073 | -0.3118 | -0.4487 | -0.4763 | 0.0002 | 0 | 13 | 0.877 |
| Sigt          | 732     | 681.833 | 29          | 10 | 0.0141 | 0.0151 | 0.0035 | 0.0037 | -1.9503 | -2.8589 | -3.0169 | 0.0052 | 0 | 8  | 0.643 |
| Adty          | 997     | 892.011 | 31          | 41 | 0.0111 | 0.0115 | 0.0112 | -0.1808 | -0.5215 | -0.5215 | 0.0076 | 0 | 17 | 0.931 |
| Pp2c          | 815     | 811.333 | 31          | 8  | 0.0221 | 0.0224 | 0.0224 | -0.3792 | -0.6279 | -0.6449 | 0.1034 | 1 | 8  | 0.799 |
| Sps1          | 716     | 716.333 | 31          | 10 | 0.0181 | 0.0181 | 0.0035 | 0.0035 | -1.5505 | -1.6212 | 0  | 0  | 7  | 0.452 |
| Upl           | 656     | 636.667 | 32          | 9  | 0.0224 | 0.0224 | 0.0224 | -0.8911 | -1.0025 | -1.1326 | 0.0027 | 0 | 6  | 0.609 |
| Tififf        | 569     | 521.222 | 29          | 9  | 0.0235 | 0.0237 | 0.0094 | 0.0044 | -1.1863 | -0.9376 | -1.1834 | 0.0253 | 0 | 9  | 0.717 |
| Tran          | 440     | 437.333 | 32          | 14 | 0.0462 | 0.0463 | 0.0085 | 0.0085 | -0.8844 | -0.6469 | -0.7027 | 1.1713 | 0 | 12 | 0.891 |
| Mdeh          | 525     | 525.333 | 31          | 19 | 0.0242 | 0.0242 | 0.0091 | 0.0091 | -1.8459 | -2.8617 | -2.8235 | 0.0315 | 0 | 14 | 0.819 |
| Average       | 718     | 658.614 | 30.8        | 16.3 | 0.004 | 0.004 | 0.057 | 0.059 | -1.02 | -1.2039 | -1.347 | 0.1497 | 0.1 | 10.4 | 0.748 |

Note: Length, alignment length; Silent, alignment length on silent site; Accessions, number of accessions sequenced for each locus; S, the number of segregating sites; π, average number of nucleotide differences between 2 sequences; nsilent, π on silent sites; θ, the Watterson estimator of population mutation rate; D, Tajima’s D; D* and F*, D* and F* of Fu and Li; p, the population recombination rate; Rm, the minimum number of recombination events; Hap, number of haplotype; Hdiv, haplotype diversity.
deviation. Further analysis by removing ADTY showed that no significant result was found across the loci (X^2 = 6.92, p < 0.21).

Population divergence

The level of population differentiation of the two related species was examined by the values of Fst and shared, fixed and unique polymorphisms in the two species, as well as the phylogenetic relationship of these accessions. Fst varied from locus to locus with a mean of 0.1536 (Table 4). Although some loci, such as ADTY and UPL, had Fst values over 0.40, the remaining loci were differentiated at a very low level with Fst values between 0.0254-0.1546. No fixed differences were detected between the cultivars and their wild relatives. This was consistent with a low divergence level due to the short history of domestication. However, a high proportion of shared polymorphisms were observed, particularly for the cultivars. The domesticated foxtail millet shared almost 75% of its polymorphism with green foxtail on average, whereas the proportion for green foxtail was 36%. Unique polymorphisms were present in both species, but the number was much less in cultivars compared to the wild species. Due to the nature of the domestication process, we expected that the cultivars had lost much of their diversity compared to its wild progenitor (55%). The unique polymorphisms in the cultivars suggested that new mutations occurred after domestication or that the wild gene pool was insufficiently sampled.

Linkage disequilibrium

The level of linkage disequilibrium measured as squared allele-frequency correlations were plotted against the distance between pairwise SNPs (Figure 2). The regression curves show that the decay of linkage disequilibrium along the distance for the cultivars was much slower than for the wild. The expected value of for wild green foxtail dropped rapidly to 0.1 within 150 bp, whereas a much higher level (> 0.1) extended to 1000 bp in foxtail millet. Comparison of the population recombination rate, \(r\), in the two species implied that recombination was more frequent in the wild species (mean = 0.1497) than that in the domesticated species (mean = 0.0329). Together with the data of haplotype number and haplotype diversity (Table 3), the cultivar foxtail millet showed increased linkage disequilibrium level compared to the wild progenitor.

Bayesian estimate of population mutation rate \(\theta\) and recombination rate/mutation rate \(r/\theta\)

Under the standard neutral model, we drew the prior distribution of \(\theta\) and \(r\) from uniform distribution within

| Loci  | S\text{Shared} | S\text{Fixed} | S\text{wild specific} | S\text{cultivar specific} | Fst     |
|-------|----------------|---------------|-----------------------|--------------------------|---------|
| DACP  | 18             | 0             | 9                     | 1                        | 0.0254  |
| SIGT  | 5              | 0             | 4                     | 1                        | 0.0289  |
| ADTY  | 4              | 0             | 37                    | 1                        | 0.4085  |
| PP2C  | 2              | 0             | 4                     | 0                        | 0.0933  |
| SPS1  | 2              | 0             | 8                     | 1                        | 0.1546  |
| UPL   | 1              | 0             | 8                     | 4                        | 0.4082  |
| TIFIF | 3              | 0             | 6                     | 1                        | 0.1002  |
| TRAN  | 7              | 0             | 8                     | 3                        | 0.1075  |
| MDEH  | 10             | 0             | 9                     | 6                        | 0.0556  |
| Average| 5.8           | 0             | 10.3                  | 2                        | 0.1536  |
intervals 0-0.03 and 0-0.15. Using rejection algorithm, we obtained 1000 samples from simulations. The posterior distribution of $\theta$, $\rho$ and $r/\theta$ are shown in Figure 3. Compared with the mean values calculated from the sequence data, values estimated by simulations were quite similar for $\theta$ in both domesticated (0.0022) and wild species (0.0053). However, the maximum posterior estimates of $\rho$ were not as large as observed data ($\rho = 0.001$, $r/\theta = 0.43$ for the domesticated and $\rho = 0.005$ and $r/\theta = 0.836$ for the wild). Since the sequence used in this study were short, recombination that could be detected in such a short distance would be rare. The method used for calculating recombination rates here may not give correct values. We also found variations among loci for the values of $\rho$ (0-1.1713 for the cultivated and 0-0.2204 for the wild). This method may over-estimate $\rho$ for some loci. We used a grid of values 0, 1, 2, 5, 10 for $r/\theta$ in the later analysis, since previous studies showed that the level of recombination had influence on posterior estimation of parameters [30].

Bayesian inference of bottleneck process

Domestication process can be modeled by coalescent simulation using a simple bottleneck model. The model has been described in Maize, Rice, Wheat and other crops, although for species such as Asian rice the domestication scenario may be more complex[2,3,6,15]. In the bottleneck model, it was assumed that the domesticated species experienced a reduction in population size in the initial of domestication state and then the population size increased after the domesticated species were wildly distributed. We can use this model to infer the domestication process based on population structure and other known information. Upon the assumption of the domestication model, a series of coalescent simulations were done to infer the parameters of the bottleneck model. The parameters that need to be estimated were drawn from a uniform distribution with a specific interval ($t_2 = 5000-15000$, $d = 100-3000$, $k = 0.1-10$, $m_{12} = 0-100$, $m_{21} = 0-100$) (table 3). We did simulations for $r/\theta = 0$, 1, 2, 5, 10, respectively. As the acceptance rates for simulations of $r/\theta > 2$ were very low, we used a larger interval (60%) for the rejection algorithm in these simulations. For each simulation, we collected 10,000 samples using the rejection algorithm with summary statistics of the wild species. Then the samples were used to fit with summary statistics of the cultivated species. The number of fitted samples used for posterior prediction varied from 702 for $r/\theta = 1$ to 2571 for $r/\theta = 2$. The posterior distribution of $t_2$, $d$, $M_{12}$ and $M_{21}$ showed no large peak. However, the posterior distribution of the bottleneck intensity $k$ did show a clear peak and depended on the ratio of $r/\theta$ (Figure 4). The domestication bottleneck was more severe when $r/\theta$ changed to large values. Since the ratio of $r/\theta$ estimated by Bayesian estimation in this study was 0.836, the severity of the bottleneck for foxtail millet should be similar to that of $r/\theta = 1$ (0.6095). If the ratio of $r/\theta$ was underestimated, the domestication process would be much more intense. To further estimate the rate of migration between the two species, we employed a Markov Chain Monte Carlo method as implemented in MIMAR [31]. We ran 1.1e7 steps with 1e6 burnin steps and considered convergence was reached when the posterior distribution of two independent chains were similar (Figure 5). The maximum estimated migration from cultivar to wild is 0.3174, whereas the reverse process was 0.1712.

Discussion

Sequence diversity

Previous studies based on isozymes and DNA markers showed a high level of genetic diversity in foxtail millet [32,33]. However, diversity at the DNA sequence level of

![Figure 3 Posterior distributions of $\theta$, $\rho$ and $r/\theta$ as estimated using approximate Bayesian approach in foxtail millet (solid line) and green foxtail(dashed line).](image.png)
Foxtail millet was not well documented and comparisons with other crops, especially cereals, was absent. In this study, nine loci were surveyed to investigate the genetic diversity in foxtail millet and its wild progenitor green foxtail at the DNA sequence level. We found a low level of nucleotide variation in both foxtail millet and green foxtail as compared with other domesticated crops and their wild relatives, such as sunflower, barley, maize and rice [3,6,12,34]. The silent nucleotide variation for green foxtail estimated here was 0.0059, which was lower than most of these wild crop relatives but higher than wild soybean. Compared to other species, nucleotide variation indicated by $\theta_{\text{sil}}$ was higher in wild maize *Zea mays* ssp. *parviglumis* (0.0247), wild sunflower *Helianthus annuus* (0.0234), etc [3,6,12,34]. An exception was found in wild soybean *Glycine soja*, which had a $\theta_{\text{sil}}$ as low as 0.00235 [35]. Further investigation of these wild relatives showed that most wild species with a higher nucleotide variation have an outcrossing mating system, while nucleotide variation was lower for inbreeding species, such as wild barley and wild soybean [34,36,37]. Green foxtail, as well as foxtail millet, are self-pollinated plant that have a 0.3% to 4% outcrossing rate in natural conditions [38,39]. Lower levels of nucleotide variation in these wild species is expected by their mating system, although the samples collected in this study may also influence the estimation. This is because some wild accessions may not represent a local original wild species but a weedy form that derived from the gene flow between the cultivars and their wild relatives. However, the underestimation influenced by sampling may have been very limited because several accessions were collected in the areas where a weedy form was recognized. The low level of genetic diversity of wild green foxtail might be caused mainly by its high inbreeding rate and/or other demographic factors, particularly changes in population size. Consistent with its wild relatives, foxtail millet also showed a lower level of silent nucleotide variation (0.0027). It was much lower than that of maize (0.0149) or sunflower (0.0072), but similar to that of rice (0.0024) and sorghum (0.0034) [3,6,12,40]. The domestication of maize, barley, soybean, and foxtail millet did not involve changes in mating systems. However, the mating system was changed from outcrossing to inbreeding in rice and sunflower [3,12]. Based on this information, we can conclude that the level of genetic diversity of foxtail millet, which maintained 45% of its wild diversity, was mainly a result of its change in population size during domestication process, followed by mutation accumulated after divergence. To further explore the impact of domestication on genetic diversity, we used Tajima’s D test to detect the change of SNP frequency in the species after domestication. As indicated by Tajima’ test, D values of most loci in domesticated species were higher than those of the wild species, but only two of them had a positive D value. The increase of D values in the domesticated species is likely because low frequency alleles were preferentially lost during the domestication bottleneck. Detailed site frequency spectra are shown in Figure 6. We detected an excess in both low and high frequency alleles in the
domesticated species compared to the wild one. The excess of high frequency variants in the domesticated species was also found in domesticated Asian rice [41]. Taking into account the higher level of LD in the domesticated species, this may mean that artificial selection active during the domestication process may have extended over a long distance. This phenomenon has been observed for tb1 in maize and waxy in rice, that the influence of directional selection on certain loci had a large effect on regions around them [8,42].

Linkage disequilibrium
Several studies that focus on the domestication of crops indicated that there was an increased level of linkage disequilibrium in domesticated species as compared to their wild relatives [3,12,43]. With the bottleneck effect of domestication, the genome-wide level of nonrandom association tends to increase. In some loci that were targeted by artificial selection, LD could extend to a long distance up to 100 kb, such as tb1 in maize and waxy in rice [8,42]. In addition to the influence of population size and direction selection, mating system was also shown to affect LD level [16]. In selfing Arabidopsis thaliana, LD can persist for 250 kb. However, in outcrossing maize, LD declined to a very low level of less than 1 kb [43,44]. Using the same set of sequences, it was shown that the LD level in O. nivara was slightly higher than that in O. rufipogon, which displayed a higher rate of outcrossing compared to O. nivara [3]. In this study, the higher level of LD observed in foxtail millet compared with green foxtail was similar to what has been observed in rice and sunflower [3,12]. In wild species, which were shown to have a higher outcrossing rate, the LD level declined rapidly within 100-200 bp. In contrast, LD in domesticated foxtail millet extended to 1000 bp. Based on the fact that the mating system did not change after domestication of foxtail millet, and that the influences of directional selection was likely focused on a small proportion of loci, it is suggested that the increased level of LD in the cultivated foxtail millet was mainly due to the change of population size during the domestication process.

Gene flow between the domesticated and the wild
As indicated previously, foxtail millet maintained 45% of its wild diversity. The proportion was similar with that of sorghum (60-70%) and sunflower (40-59%) [12,40]. However, diversity retention was higher in maize (80%) and very low in rice (10-20%) [3,16]. The retention of wild diversity is a reflection of the bottleneck intensity (such as in rice) or the mating system (such as in maize). We used coalescent simulation to infer the bottleneck severity during the domestication process. The Bayesian estimate of bottleneck severity k was 0.6095 if we set $r/\theta = 1$, which is near the maximum estimate of $r/\theta$. Compared with the bottleneck severity that modeled in maize (2.45) and in rice (0.2 for japonica and 0.5 for indica) [3,45], the severity estimated here was compatible with the loss of diversity from its wild relatives. We also incorporated gene flow into a simulation model. The results suggested that there were low levels of gene flow from the cultivated species to wild species, but the reverse process was even smaller. As foxtail millet and green foxtail can have low levels of cross-pollination and the hybridization between them are compatible [46], we expected that gene flow between the two species would be frequent. By intensive human selection, alleles from wild relatives to cultivated foxtail millet were mostly eliminated. But genes introgressed into the wild species might be retained at a higher level, as evidenced by many weedy types morphologically similar to foxtail millet in and around millet fields. This has become a serious problem for field management. Taken together, the population bottleneck and gene flow both contributed to the present pattern of nucleotide diversity of these two species.

Loci under selection in the domestication
Another factor in the domestication process is human selection [16]. We intended to select nearly neutral loci in order to make inferences about the domestication process in this study. However, ADTY had a significant negative D value in the domesticated species. This observation, taken together with an 86% loss of diversity compared to the wild type and a large deviation in the HKA test, suggested that this locus was likely under directional selection during domestication. We used the estimated parameter of bottleneck severity to infer
whether this locus was under selection. Two loci were detected to be under selection in this model, ADTY and PP2C (p-value = 0.014 and 0.008). Except for a loss of 75% of its wild diversity, other statistical tests did not show any signal of selection for PP2C, indicating that the simulation may have given a false positive result for this locus due to low diversity in both cultivated and wild species. Furthermore, the model used to detect selection may not be robust enough to give a result with only a few false positive genes. The locus ADTY may be a target gene of human selection or located in a selected region. However, excluding this locus did not affect the calculated parameters and the results of the simulation. The whole genome sequence of foxtail millet will be released very soon; and further work that focuses on whole genome analysis of genes involved in domestication will likely indicate to what extent human selection has acted in the domestication process.

Conclusions
We found a 55% loss of diversity in foxtail millet and an increased level of LD that influenced the genetic diversity of these two species. The whole genome sequence of foxtail millet to green foxtail may have been another factor that influenced the genetic diversity of these two species.

Methods
Plant materials and DNA sequencing
We collected 84 accessions to survey DNA sequence variation in this study, including 50 cultivated foxtail millets and 34 wild green foxtails (Table 1). Accession numbers started with PI or Ames were obtained from the National Germplasm Resources Laboratory, Beltsville, Maryland, USA; others samples were from the Chinese National Germplasm Bank in the Institute of Crop Sciences, Chinese Academy of Agricultural Sciences. The cultivated materials were sampled to represent a broad diversity of foxtail millet, in which all Chinese accessions are landraces. The wild samples were collected throughout the Eurasian continent to cover the distribution range of green foxtail where foxtail millet was claimed to have been domesticated [23,24]. All the sample seeds were planted in autoclaved soil and fresh leaves were collected to extract genomic DNA using a modified CTAB protocol.

Genomic fragments between 500 and 1500 bp were amplified and sequenced from nine unlinked loci (Table 2 and Additional file 1: figure S1). Based on the high colinearity between the genetic map of foxtail millet and rice [47], we used the rice genome sequence as a reference when selecting the loci so as to have wide coverage of the genome. All EST sequences of Setaria italica were downloaded from the NCBI ftp site http://ftp.ncbi.nlm.nih.gov. The EST set was used to search rice gene models and only those with a single hit were retained for further studies http://rice.plantbiology.msu.edu/. We checked if the homologous region covers an intron in the rice genome and designed primers to amplify the intron.

The nine loci were amplified in the two species using a modified PCR reaction system: 50 ng genomic DNA, 0.2 μmol/l of each primer, 0.2 mM dNTP, 1 U ExTaq DNA polymerase (TaKaRa), 2 μl PCR mix buffer and dH2O to a final volume of 20 μl. After amplification, the products were separated by electrophoresis on a 2% agarose gel. DNA bands were excised, purified, and directly sequenced on an Applied Biosystems 3730 DNA Sequencer. For individual with heterozygous alleles, it is impossible to choose a true allele if it contains more than one heterozygous polymorphism. By directly sequencing the PCR product, the haplotype that was chosen randomly may not be a true haplotype. However both foxtail millet and green foxtail have a low outcross rate, we expected that the influence of heterozygous polymorphism on haplotype inference is very limited. Single base pair changes were further confirmed by PCR and DNA sequencing.

Sequence analysis
The raw sequence trace files were collected and assembled by Phred/Phrap [48,49]. Alleles of each locus were aligned by ClustalW 1.81 with further manual check [50]. All the alleles containing singletons were subject to a check process, in which we amplified and sequenced the product again to confirm the sequence quality and update the alignment. The alignment files were imported to DnaSP 4.5 with coding regions assigned according to the rice gene annotation [51]. For each locus and species, we calculated the number of segregating sites (S), the population recombination rate (ρ), minimum number of recombination events (Rm), number of haplotypes (Hap), haplotype diversity (Hdiv), average number of nucleotide difference per site between two sequences (π), and the watterson estimator of population mutation rate (θ). To test for neutrality, we calculated Tajima’s D [52], and D* and F* of Fu and Li [53] test without outgroup. To access the level of species divergence, we calculated shared, fixed, species-specific S and Fst for the two species. Multilocus HKA test was done by HKA http://lifesci.rutgers.edu/~heylab/ for the nine loci and the results was parsed to R for further analysis [54]. The decay of linkage disequilibrium (LD) with physical distance was described using a nonlinear regression analysis. The expected value of squared allele-frequency
correlations ($r^2$) at drift-recombination equilibrium is, 

$$E(r^2) = \frac{10 + \rho}{(2 + \rho)(11 + \rho)} + \frac{(3 + \rho)(12 + 12\rho + \rho^2)}{\pi(2 + \rho)(11 + \rho)}$$

where n is the sample size of sequences [55]. To introduce the distance between pairs of SNP sites (d) into the formula, we use rho’d to replace $\rho$, in which rho is the recombination rate per basepair. $r^2$ between pairs of polymorphism was estimated using DnaSP 4.5 for each locus and species. $r^2$ and d were pooled across the loci for each species. The nonlinear regression analysis was performed with the NLS function in the R statistical package http://www.r-project.org.

**Coalescent simulation**

Coalescent simulation was used to model the process of domestication, as well as to estimate the population mutation rate $\theta$ and population recombination rate $\rho$ for each species. The simulations were done using Hudson’s ms [56]. In each simulation, we used an rejection-based approximate Bayesian computation approach to obtain a posterior distribution for parameters of interest [57,58]. Briefly, the initial values of parameters were drawn from a user-specified prior distribution, and, starting with these parameters, the simulation was run under a defined model; For each simulated datasets, several summary statistics were calculated and compared with the observed values. The data was accepted if it was within a defined interval of observed data; and then the parameter set that generated acceptable data were used to obtain a posterior distribution for each parameter. In this study, we used a multilocus approach to assess the acceptability of the data. Summary statistics were calculated for each locus and summarized by mean and/or variance across the loci. The values of mean and/or variance of each summary statistics were compared between simulated and observed data using. The accepted datasets were used for Bayesian inference of selected parameters. Doing this incorporates the variation among the loci into the simulations.

**Approximate Bayesian estimate of $\theta$ and $\rho/\theta$**

To estimate $\theta$ and $\rho$, the simulations were run under the standard neutral model for the wild and domesticated species separately. The means for each four summary statistics ($S$, $\pi$, Hap, Hdv) were used to access the acceptable of the simulated data. The data was accepted if three of them were within 20% of the observed data [59].

**Approximate Bayesian inference of bottleneck parameters**

The model used to investigate the bottleneck process was similar as described for maize and rice [3,6,45]: $N_a$ is the effective population size of the ancestor of the two species, $\mu$ is the mutation rate for the ancestor. The values of $\mu$ were based on the synonymous substitution rate or calculated by $\theta = 4N\mu$. The recombination rate $\rho$ together with $N_a$ and $\mu$ defined the ancestor population. At time $t_2$ generations ago, a new population was derived from the ancestor with a population size of $N_b$ and expanded to a population with a size of $N_p$ at $t_1$ generations ago. Then, the ancestor population and the newly derived population evolved with constant population size until present. To incorporate gene flow in this model, we defined $M_{12}$ to be migration rate from wild to cultivar and $M_{21}$ for the rate for the reverse rate. In this model, the parameters for the wild and domesticated species were calculated by the sequence or estimated using approximate Bayesian estimate. We defined $k$ to be the ratio of $N_b$ and $d$, $d$ being the duration of the bottleneck in generations ($d = t_2 - t_1$). Previous studies suggested that $N_b$ is positively correlated with $d$; $k$ is a good indicator for the stringency of the bottleneck [3,6].

Under this model, we used the mean and variance for each four summary statistics ($S$, $\pi$, Hap, Hdv) to assess the acceptability of simulated wild species data. The simulation was accepted if both the mean and variance fell within 30% of the observed data. To fit the cultivated data, we use the mean of $S$, $\rho$, Hap, and Hdv to assess the fitness of the simulated data, but we accepted if three of the four fell within 30% of the observed data.

**Additional material**

Additional file 1: Schematic diagrams of nine loci and sequenced regions in this study. Exons, introns and UTRs are indicated by blue boxes, lines and open boxes. The primers that were used to PCR and sequencing are marked with black arrowhead, where F and R stand for forward primer and reverse primer respectively.

**Acknowledgements**

This paper was supported by National Sciences Foundation of China (30630045, 30471117), the earmarked fund for Modern Agro-industry Technology Research System of China, National Technology Supporting Program (2006BAD09B20), National 863 program (2006AA102157, 2008AA102158) and Hebei Natural Sciences Foundation (C2004000697, C2006000725, C2007000972). We thank Dr. Song Ge from the Institute of Botany, Chinese Academy of Sciences, Dr. Andrew Doust from Oklahoma State University, USA, for their critical reading and English modification of the manuscript.
Author details
1 The National Key Facility for Crop Gene Resources and Genetic Improvement (NFCRI), Institute of Crop Science, the Chinese Academy of Agricultural Sciences (CAAS), Beijing 100081, China. 2 Institute of Millet Crops, Hebei Academy of Agricultural and Forestry Science, Shijiazhuang 050031, China. 3 State Key Laboratory of Plant Genomics, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, China. 4 College of Life Sciences, Hebei Normal University, Shijiazhuang 050012, China.

Authors’ contributions
XD and MC designed the study, supervised the experiment. HZ and WL collected and planted the samples used and made help in laboratory work. JC, BZ, CW and XD performed the statistical analysis. JC prepared the first draft of the manuscript and XD made optimization of the manuscript. All authors discussed the result and conclusion and read and approved the final manuscript.

Received: 14 May 2010 Accepted: 11 October 2010
Published: 11 October 2010

References
1. Diamond J. Evolution, consequences and future of plant and animal domestication. Nature 2002, 418(6908):700-707.
2. Eyre-Walker A, Gout RL, Hilton H, Feldman DL, Gaut BS. Investigation of the bottleneck leading to the domestication of maize. Proc Natl Acad Sci USA 1998, 95(8):4441-4446.
3. Zhu Q, Zheng X, Luo J, Gaut BS, Ge S. Multilocus analysis of nucleotide variation of Oryza sativa and its wild relatives: severe bottleneck during domestication. Mol Biol Evol 2007, 24(3):875-888.
4. Pollak E. On the theory of partially inbreeding finite populations. I. Partial selfing. Genetics 1987, 117(2):353-360.
5. Nordborg M: Linkage disequilibrium, gene trees and selfing: an ancestral recombination graph with partial self-fertilization. Genetics 2000, 154(2):923-929.
6. Tenenon ML, Lefèvre J, Tenenon Q, Gaut BS: Selection versus demography: a multilocus investigation of the domestication process in maize. Mol Biol Evol 2004, 21(7):1214-1225.
7. Hanson MA, Gaut BS, Stec AO, Fuerstenberg SI, Goodman MM, Coe EH, Wang et al. Genetic diversity and its wild ancestor (Zea mays ssp. parviglumis).
8. Wang RL, Wendel JF, Doebley JF: Molecular signatures of natural selection. BMC Genetics 2010, 11(40).
9. Wright SI, Gaut BS: Molecular signatures of natural selection. Annu Rev Genet 2005, 39:197-218.
10. Li Y, Wu S: Traditional maintenance and multiplication of foxtail millet (Setaria italica (L.) Beauv.) landraces in China. Euphytica 1996, 87(1):33-38.
11. Lu H, Zhang J, Liu KB, Wu N, Li Y, Zhou K, Ye M, Zhang T, Zhang H, Yang X, et al. Earliest domestication of common millet (Panicum miliaceum) in East Asia extended to 10,000 years ago. Proc Natl Acad Sci USA 2009, 106(18):7367-7372.
12. Austin DF: Fox-tail Millets (Setaria: Poaceae)—Abandoned Food in Two Hemispheres. Economic Botany 2006, 60(2):143-158.
13. Fukunaga K, Kato K: Mitochondrial DNA variation in foxtail millet, Setaria italica (L.) Beauv. Euphytica 2003, 129(1):7-13.
14. Fukunaga K, Domon E, Kawase M: Ribosomal DNA variation in foxtail millet, Setaria italica (L.) Beauv., and a survey of variation from Europe and Asia. Theoretical and Applied Genetics 1997, 95(6):751-756.
15. Fukunaga K, Ichitani K, Kawase M: Phylogenetic analysis of the rDNA intergenic spacer subrepeats and its implication for the domestication history of foxtail millet, Setaria italica. Theor Appl Genet 2008, 113(2):261-269.
16. Li Y, Jia J, Wang Y, S W: Intraspecific and interspecific variation in Setaria revealed by RAPD analysis. Genetic Resources and Crop Evolution 1998, 45(3):279-285.
17. Doust AN, Kellogg EA, Devos KM, Bennetzen JL: Foxtail millet: a sequence-driven grass model system. Plant Physiol 2009, 149(1):137-141.
18. Doust AN, Devos KM, Gadberry MD, Gale MD, Kellogg EA: Genetic control of branching in foxtail millet. Proc Natl Acad Sci USA 2004, 101(24):9045-9050.
19. Doust AN, Devos KM, Gadberry MD, Gale MD, Kellogg EA: The genetic basis for inbreeding variation between foxtail and green millet (poaceae). Genetics 2005, 169(3):1659-1672.
20. Doust AN, Kellogg EA: Effect of genotype and environment on branching in weedy green millet (Setaria viridis) and domesticated foxtail millet (Setaria italica) (Poaceae). Mol Ecol 2006, 15(5):1335-1349.
21. Tenellon ML, Sawkins MC, Long AD, Gaut RL, Doebley JF, Gaut BS: Patterns of DNA sequence polymorphism along chromosome 1 of maize (Zea mays ssp. mays L.). Proc Natl Acad Sci USA 2001, 98(16):9161-9166.
22. Thornton K, Andolfatto P: Approximate Bayesian inference reveals evidence for a recent, severe bottleneck in a Nethersol population of Drosophila melanogaster. Genetics 2006, 172(3):1607-1619.
23. Becquet C, Przeworski M: A new approach to estimate parameters of speciation models with application to apes. Genome Res 2007, 17(10):1505-1519.
24. Wang RL, Wendel JF, Dekker JH: Weedy Adaptation in Setaria spp. I. Isozyme Analysis of Genetic Diversity and Population Genetic Structure in Setaria viridis. American Journal of Botany 1995, 82:308-317.
25. Le Thierry d’Ennequin M, Panaud O, Toupance B, Sarr A: Assessment of genetic relationships between Setaria italica and its wild relative S. viridis using AFLP markers. TAG Theoretical and Applied Genetics 2000, 100(7):1061-1066.
26. Remor PL, Tokano DM, Lundy KE, Clegg MT: Low levels of linkage disequilibrium in wild barley (Hordeum vulgare ssp. spontaneum) despite high rates of self-fertilization. Proc Natl Acad Sci USA 2005, 102(7):2442-2447.
27. Hyten DL, Song Q, Zhu Y, Choi IY, Nelson RL, Cortez JM, Specht JE, Shoemaker RC, Cregan PB: Structure of Wild Soybean (Glycine soja) in Japan. Crop Science 2008, 48(3):1071-1079.
28. Fujita R, Ohara M, Okazaki K, Shimamoto Y: The Extent of Natural Cross-Pollination in Wild Soybean (Glycine soja). Journal of Heredity 1997, 88(2):124-128.
29. Li H, Li C, Piao W: Cytological and genetical studies of the interspecific cross of the cultivated foxtail millet, Setaria-italica (L) Beauv. and the green foxtail millet, S-Vindills L. J Am Soc Agron 1945, 39:25-34.
30. Till-Bottraud I, Reboud X, Brabant P, Le Franc M, Rhesa B, Veldt F: Damency H: Outcrossing and hybridization in wild and cultivated foxtail millets: consequences for the release of transgenic crops. TAG Theoretical and Applied Genetics 1992, 83(8):940-946.
40. Hamblin MT, Casa AM, Sun H, Murray SC, Paterson AH, Aquadro CF, Kresovich S: Challenges of detecting directional selection after a bottleneck: lessons from Sorghum bicolor. Genetics 2006, 173(2):953-964.
41. Caicedo AL, Williamson SH, Hernandez RD, Boyko A, Fiedel-Alan A, York TL, Polato NR, Olsen KM, Nielsen R, McCouch SR, et al: Genome-wide patterns of nucleotide polymorphism in domesticated rice. PLoS Genet 2007, 3(9):1745-1756.
42. Olsen KM, Caicedo AL, Polato N, McClung A, McCouch S, Purugganan MD: Selection Under Domestication: Evidence for a Sweep in the Rice Waxy Genomic Region. Genetics 2006, 173(2):975-983.
43. Remington DL, Thornsberry JM, Matsuoka Y, Wilson LM, Whitt SR, Doebley J, Kresovich S, Goodman MM, Buckler ESt: Structure of linkage disequilibrium and phenotypic associations in the maize genome. Proc Natl Acad Sci USA 2001, 98(20):11479-11484.
44. Nordborg M, Innan H: Molecular population genetics. Current Opinion in Plant Biology 2002(2):69-73.
45. Wright SI, Bi IV, Schroeder SG, Yamasaki M, Doebley JF, McMullen MD, Gaut BS: The effects of artificial selection on the maize genome. Science 2005, 308(5726):1310-1314.
46. Darmency H, Zangre GR, Pernes J: The wild-weed-crop complex in Setaria: a hybridization study. Genetica 1987, 75(2):103-107.
47. Devos KM, Wang ZM, Beales J, Sasaki T, Gale MD: Comparative genetic maps of foxtail millet (Setaria italica) and rice (Oryza sativa). Theoretical and Applied Genetics 1998, 96(1):63-68.
48. Ewing B, Hillier L, Wendl MC, Green P: Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res 1998, 8(3):175-185.
49. Ewing B, Green P: Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res 1998, 8(3):186-194.
50. Thompson JD, Higgins DG, Gibson TJ: CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. Nucleic Acids Research 1994, 22:4673-4680.
51. Rozas J, Sanchez-DelBarrio JC, Mesquida X, Rozas R: DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 2003, 19(18):2496-2497.
52. Tajima F: Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 1989, 123(3):585-595.
53. Fu YX, Li WH: Statistical tests of neutrality of mutations. Genetics 1993, 133(3):693-709.
54. Hudson RR, Kreitman M, Aguade M: A test of neutral molecular evolution based on nucleotide data. Genetics 1987, 116(1):153-159.
55. Hill WG, Weir BS: Variances and covariances of squared linkage disequilibria in finite populations. Theor Popul Biol 1988, 33(4-78).
56. Hudson RR: Generating samples under a Wright-Fisher neutral model of genetic variation. Bioinformatics 2002, 18(2):337-338.
57. Beaumont MA, Zhang W, Balding DJ: Approximate Bayesian computation in population genetics. Genetics 2002, 162(4):2025-2035.
58. Marjoram P, Tavare S: Modern computational approaches for analysing molecular genetic variation data. Nat Rev Genet 2006, 7(10):759-770.
59. Weiss G, von Haeseler A: Inference of population history using a likelihood approach. Genetics 1998, 149(3):1539-1546.

doi:10.1186/1471-2156-11-90
Cite this article as: Wang et al.: Population genetics of foxtail millet and its wild ancestor. BMC Genetics 2010 11:90.