Supplementary Materials

Genotyping-by-sequencing reasserts the close relationship between tef and its putative wild *Eragrostis* progenitors

Dejene Girma[1,2], Gina Cannarozzi[1], Annett Weichert[1], Zerihun Tadele[1,3,4,*]

[1] Institute of Plant Sciences, University of Bern, Altenbergrain 21, 3013 Bern, Switzerland
[2] Ethiopian Institute of Agricultural Research, National Agricultural Biotechnology Research Center P.O. Box 249 Holetta, Ethiopia
[3] Addis Ababa University, Institute of Biotechnology, P.O. Box 32853 Addis Ababa, Ethiopia
[4] Center for Development and Environment (CDE), University of Bern, Bern 3012, Switzerland
* Correspondence: zerihun.tadele@ips.unibe.ch; Tel: +41 31 631 49 56

Table S1. Germplasm class information and origin of materials used in the experiment. IPS: Institute of Plant Sciences; m a.s.l.: meter above sea level; MoA: Ministry of Agriculture; USDA: United States Department of Agriculture.

Common/scientific name	**Germplasm origin**			**Reference**
	Location	**Coordinates**	**Altitude (m a.s.l)**	
1. Tef cultivars				
Ada	Debre Birhan, North Shewa	9°41´N, 39°32´E	2750	(Ebba 1975)
Addissie	Debre Zeit, East Shewa	8°42´N, 39°5´E	2000	(Ebba 1975)
Adoensis	Nekemte, Wollega	9°1´N, 36°30´E	2400	(Ebba 1975)
Balami	Debre Zeit, East Shewa	8°42´N, 39°5´E	2000	(Ebba 1975)
Beten	Debre Zeit, East Shewa	8°42´N, 39°5´E	2000	(Ebba 1975)
Bunninye	Nejo, Wellega	9°30´N, 35°30´E	2000	(Ebba 1975)
Burssa	Harar, East Hararghe	9°21´N, 42°8´E	1856	(Ebba 1975)
Curati	Damota, Gojam	10°33´N, 36°35´E	2400	(Ebba 1975)
Dabbi	Bahir Dar, Gojam	10°33´N, 38°27´E	1840	(Ebba 1975)
Denkeye	Harar, East Hararghe	9°21´N, 42°8´E	1856	(Ebba 1975)
Dschanger	Unknown			(Ebba 1975)
Enatite	Debre Zeit, East Shewa	8°42´N, 39°5´E	2000	(Ebba 1975)
Fesho	Wolisso, West Shewa	8°35´N, 38°0´E	2200	(Ebba 1975)
Gofarie	Jimma, Keffa	7°35´N, 36°45´E	1750	(Ebba 1975)
Gommadie	Agaro, Keffa	7°47´N, 36°36´E	1870	(Ebba 1975)
Gorradie	Haramaya, East Hararghe	9°24´N, 42°26´E	2100	(Ebba 1975)
Hamrawe-murri	Ameressa, East Hararghe	9°23´N, 42°7´E	1860	(Ebba 1975)
Hatalla	Harar, East Hararghe	9°21´N, 42°8´E	1856	(Ebba 1975)
Janno	Damota, Gojam	10°43´N, 37°2´E	2310	(Ebba 1975)
Karadebie	Ghimbi, Wollega	9°9´N, 35°37´E	2100	(Ebba 1975)
Kaye-murri	Hidabu-Gensay, Wollega	9°31´N, 35°28´E	2000	(Ebba 1975)
Keyagachem	Debre Zeit, East Shewa	8°42´N, 39°5´E	2000	(Ebba 1975)
Manya	Debre Zeit, East Shewa	8°44´N, 39°0´E	2100	(Ebba 1975)
Purpurea	Unknown			(Ebba 1975)
Rosea	Debre Zeit, East Shewa	8°42´N, 39°5´E	2000	(Ebba 1975)
Rubicunda	Unknown			(Ebba 1975)
Shewa-gemerra	Shewa-Gemerra, Keffa	7°1´N, 35°50´E	2743	(Ebba 1975)
Trotteriana	Damota, Gojam	10°44´N, 35°45´E	2480	(Ebba 1975)

Tullu-Nasy	Nasy, Nejo, Wollega	9°40′N, 35°25′E	2400	(Ebba 1975)
Viridis	Unknown			(Ebba 1975)
Zuccagniana	Damota, Gojam	10°44′N, 35°45′E	2480	(Ebba 1975)
2. *Improved varieties*				
DZ-01-1285 (*Koye*)	Debre Zeit Res. Center	8°45′N, 38°59′E	1900	(MoA 2014)
DZ-01-196 (*Manya*)	Debre Zeit Res. Center	8°45′N, 38°59′E	1900	(MoA 2014)
DZ-01-2675 (*Dega Tef*)	Debre Zeit Res. Center	8°45′N, 38°59′E	1900	(MoA 2014)
DZ-01-354 (*Enatite*)	Debre Zeit Res. Center	8°45′N, 38°59′E	1900	(MoA 2014)
DZ-01-787 (*Wellonkomi*)	Debre Zeit Res. Center	8°45′N, 38°59′E	1900	(MoA 2014)
DZ-01-899 (*Gimbichu*)	Debre Zeit Res. Center	8°45′N, 38°59′E	1900	(MoA 2014)
DZ-01-99 (*Asgori*)	Debre Zeit Res. Center	8°45′N, 38°59′E	1900	(MoA 2014)
DZ-Cr-387 (*Quncho)*	Debre Zeit Res. Center	8°45′N, 38°59′E	1900	(MoA 2014)
DZ-Cr-438 (*Kora)*	Debre Zeit Res. Center	8°45′N, 38°59′E	1900	(MoA 2014)
DZ-Cr-44 (*Menagesha*)	Debre Zeit Res. Center	8°45′N, 38°59′E	1900	(MoA 2014)
3. Mutant line				
GA10	IPS, University of Bern			(Tadele 2013)
4. Wild *Eragrostis* species				
E. acutiflora	Brazil	Not available		(USDA 2015)
E. acutiglumis	Brazil			(USDA 2015)
E. aethiopica	South Africa			(USDA 2015)
E. airoides	Brazil			(USDA 2015)
E. bahiensis	Brazil			(USDA 2015)
E. barbonodis	South Africa			(USDA 2015)
E. bicolor	South Africa			(USDA 2015)
E. capensis	Zimbabwe			(USDA 2015)
E. cilianesis	USA			(USDA 2015)
E. cilindriflora	South Africa			(USDA 2015)
E. congesta	Zimbabwe			(USDA 2015)
E. curvula	South Africa			(USDA 2015)
E. dielsii	Australia			(USDA 2015)
E. echinochloidea	South Africa			(USDA 2015)
E. eriopoda	Australia			(USDA 2015)
E. ferruginea	Papua New Guinea			(USDA 2015)
E. heteromera	South Africa			(USDA 2015)
E. humidicola	Kenya			(USDA 2015)
E. intermedia	South Africa			(USDA 2015)
E. lehmanniana	South Africa			(USDA 2015)
E. lugens	Uruguay			(USDA 2015)
E. mexicana	Uruguay			(USDA 2015)
E. nigra	India			(USDA 2015)
E. nindensis	South Africa			(USDA 2015)
E. nutans	India			(USDA 2015)
E. obtusa	South Africa			(USDA 2015)
E. pallens	South Africa			(USDA 2015)
E. papposa	Tunisia			(USDA 2015)
E. patentipilosa	South Africa			(USDA 2015)
E. pectinacea	Canada			(USDA 2015)
E. pilosa	France			(USDA 2015)

E. plana	Lesotho	(USDA 2015)
E. rigidior	South Africa	(USDA 2015)
E. samentosa	South Africa	(USDA 2015)
E. superba	South Africa	(USDA 2015)
E. tenella	Sierra Leone	(USDA 2015)
E. tremula	Nigeria	(USDA 2015)
E. trichodes	USA	(USDA 2015)
E. trichopora	South Africa	(USDA 2015)
E. viscosa	Zimbabwe	(USDA 2015)

Table S2. Summary of the ApeKI *Eragrostis* GBS.

GBS statistics	**Number**
• Samples sequenced (total)	95 + BLANK
• Reads in the plate (total)	237,134,465
• Mean read count	2,491,878
• Good barcoded reads	144,938,479
• Low-quality reads	92,198,148
• Failed samples	3

Table S3. Number of read counts and mapping rate to the tef genome.

Common/scientific name	**Read counts (in millions of reads)**	**Mapping rate (%)**
1. Tef species		
Ada	3.15	94.2
Addissie	3.20	93.6
Adoensis	2.92	94.0
Balami	4.05	93.9
Beten	3.07	93.7
Bunninye	2.53	94.1
Burssa	2.38	93.7
Curati	2.09	93.5
Dabbi	3.14	94.0
Denkeye	3.35	94.3
Dschanger	2.33	94.5
Enatite	2.83	94.3
Fesho	3.02	93.9
Gofarie	3.55	93.9
Gommadie	3.34	93.5
Gorradie	2.57	94.2
Hatalla	2.41	93.4
Karadebie	2.41	94.2
Kaye-murri	1.74	94.3
Keyagachem	2.39	94.2
Manya	2.56	93.9
Purpurea	2.35	93.4
Rosea	2.58	93.9
Rubicunda	2.61	93.7
Shewa-gemerra	2.89	93.8
Trotteriana	2.94	93.8
Tullu-Nasy	2.05	93.9
Viridis	2.12	94.3
Zuccagniana	2.52	93.3
DZ-01-1285 (*Koye*)	2.49	94.0
DZ-01-196 (*Manya*)	2.33	93.8
DZ-01-2675 (*Dega Tef*)	2.86	93.9
DZ-01-354 (*Enatite*)	3.29	93.9
DZ-01-787 (*Wellonkomi*)	2.52	94.4
DZ-01-899 (*Gimbichu*)	2.73	94.0
DZ-01-99 (*Asgori*)	2.99	94.6
DZ-Cr-387 *(Quncho)*	2.87	93.9
DZ-Cr-438 *(Kora)*	2.42	93.7
DZ-Cr-44 (*Menagesha*)	2.17	93.9
GA10 (mutant line)	3.14	94.0

2. Wild *Eragrostis* species		
E. acutiflora	1.62	63.6
E. acutiglumis	1.60	56.0
E. aethiopica	2.31	93.4
E. airoides	2.81	73.5
E. bahiensis	2.13	77.8
E. barbonodis	2.18	80.5
E. bicolar	2.53	77.6
E. capensis	2.08	75.2
E. cilianensis	1.80	69.7
E. cilindriflora	2.03	68.4
E. congesta	2.69	79.8
E. curvula	3.09	79.0
E. dielsii	2.76	69.3
E. echinochloidea	1.80	77.0
E. eriopoda	1.94	89.3
E. ferruginea	2.81	93.4
E. heteromera	2.16	67.8
E. himidicola	2.30	57.2
E. intermedia	2.05	80.1
E. lehmanniana	2.66	91.5
E. lugens	2.09	92.7
E. mexicana	2.41	83.1
E. nigra	2.96	73.3
E. nindensis	3.06	82.1
E. nutans	2.23	81.7
E. obstua	2.16	91.3
E. pallens	2.35	77.5
E. papposa	2.16	57.2
E. patentipilosa	1.94	82.3
E. pectinacea	2.80	75.5
E. pilosa	3.10	94.5
E. plana	1.93	82.7
E. rigidior	2.23	76.6
E. samentosa	2.50	78.6
E. superba	2.85	78.8
E. tenella	2.55	78.7
E. tremula	3.08	73.5
E. trichopora	3.13	68.5
E. viscosa	2.60	80.0

Table S4. Number of SNPs detected on individual pseudomolecules and SNP density. *The number of SNPs listed here is obtained after filtering the raw SNPs detected per pseudomolecule by allelic state (biallelic), coverage (60%), and minor allele frequency (MAF ≥ 0.05).

Pseudo-chromosome	Sequence length (bases)	No. of SNPs*	SNP Mb^{-1}
Chr1	73,017,089	17,226	236
Chr2	55,089,357	12,951	235
Chr3	58,718,112	14,563	248
Chr4	43,964,424	11,540	262
Chr5	19,121,838	3,234	169
Chr6	34,025,776	8,777	258
Chr7	26,054,076	5,639	216
Chr8	20,774,635	3,748	180

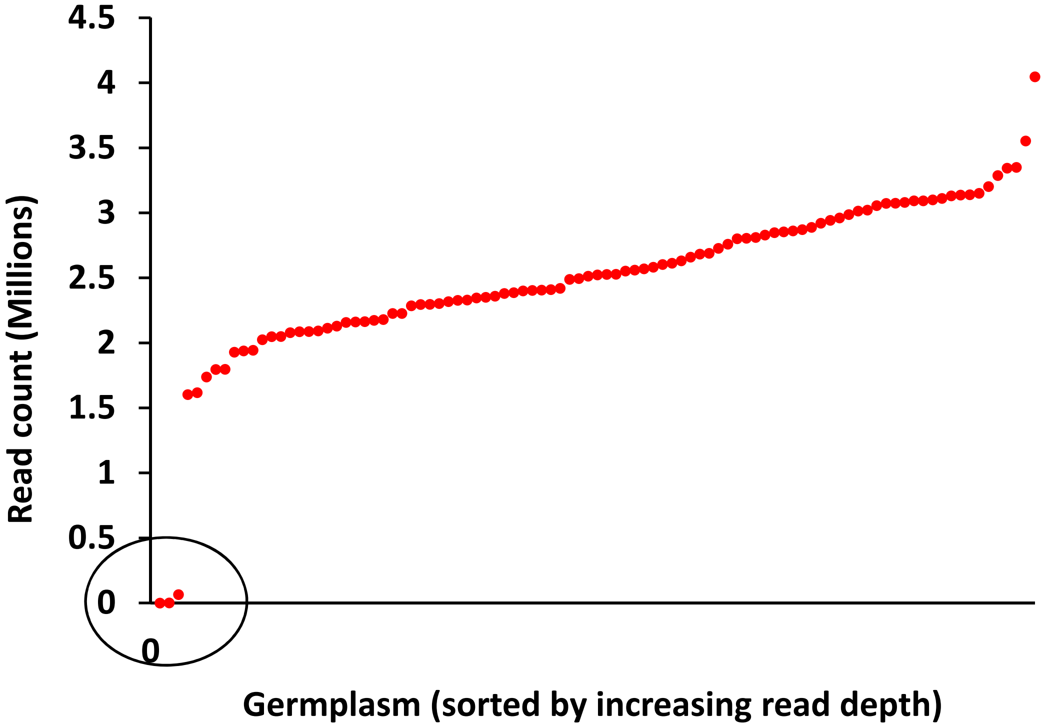
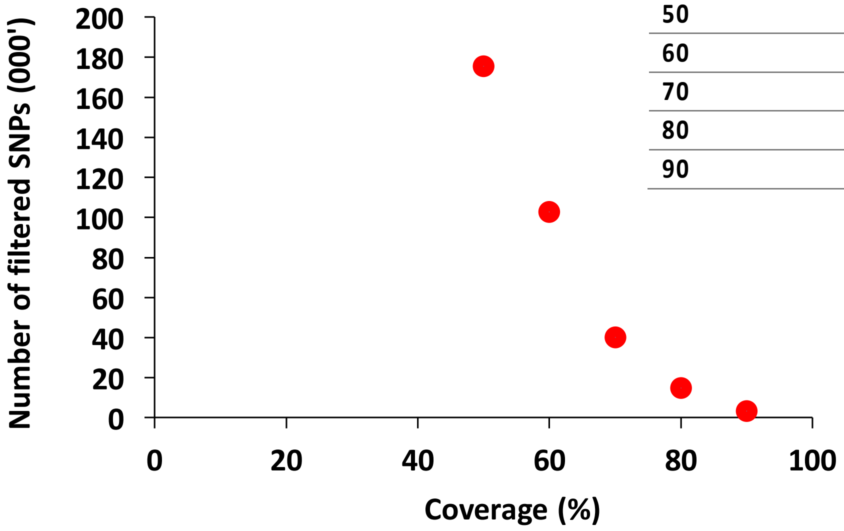


Figure S1. Mean sequencing depth (expressed as read count per germplasm) of the ApeKI Eragrostis library generated by the GBS protocol. The number of read counts per individual germplasm is shown in order of increasing number of reads generated. Three *(E. trichodes, Hamrawi-murri* and *Jano)* of the 95 samples were excluded from further analysis due to very low read counts and are indicated by red points inside the circle.



Coverage (%)	**Number of SNPs**
50	175,498
60	103,037
70	40,155
80	14,984
90	3,563

Figure S2. SNP coverage statistics. We filtered the raw SNP dataset using ranges of coverages as a cutoff. By coverage we mean the number of sites (SNPs) based on the proportion of missing data defined to be 50, 60, 70, 80, and 90. The number of SNPs retained is plotted against the corresponding coverage. The number of filtered SNPs dramatically decreases as the coverage increases (in other words as the amount of missing data is reduced).

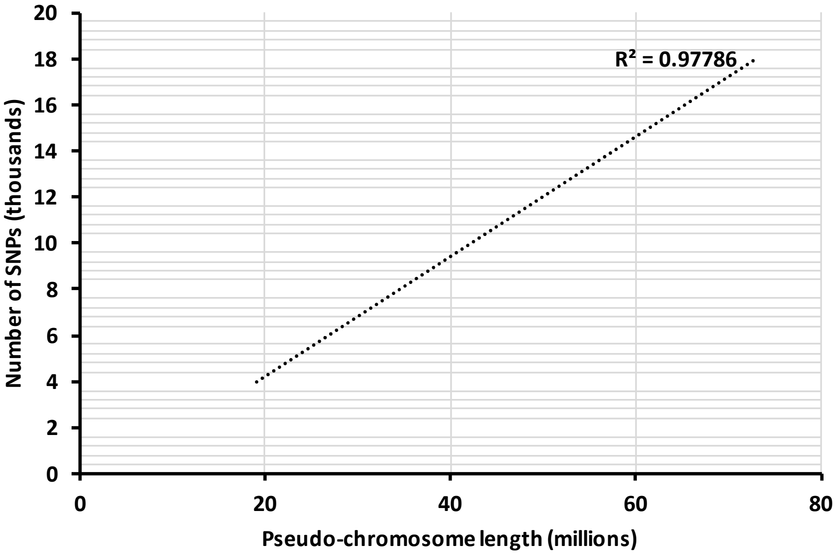


Figure S3. SNP number vs. pseudomolecule length. The number of SNPs positively and linearly increases with pseudo-chromosome length. The strength of correlation is measured by the correlation coefficient (R^2) and is 0.978.

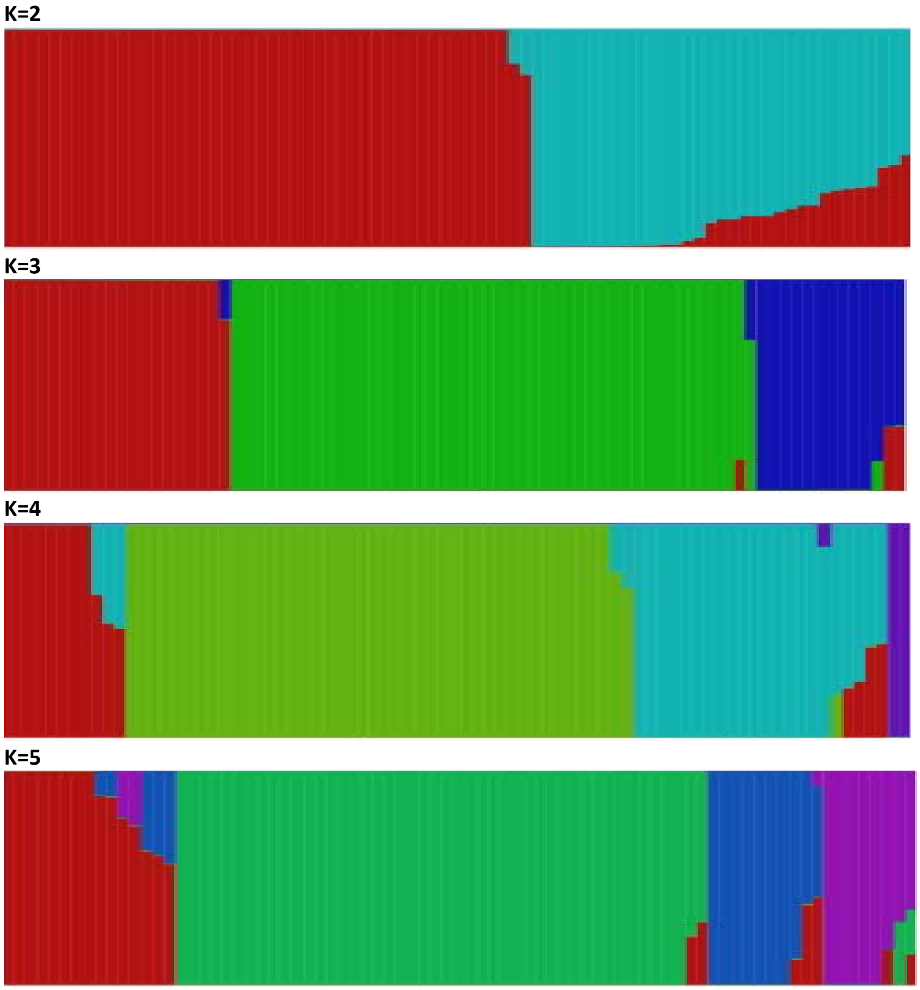


Figure S4. Population structure at different (K) clustering levels. We used fastStructure, an algorithm for inferring population structure from large SNP genotype data. Estimation of the most probable number of clusters (K), was based on four independent runs and K ranging from 2 to 5.

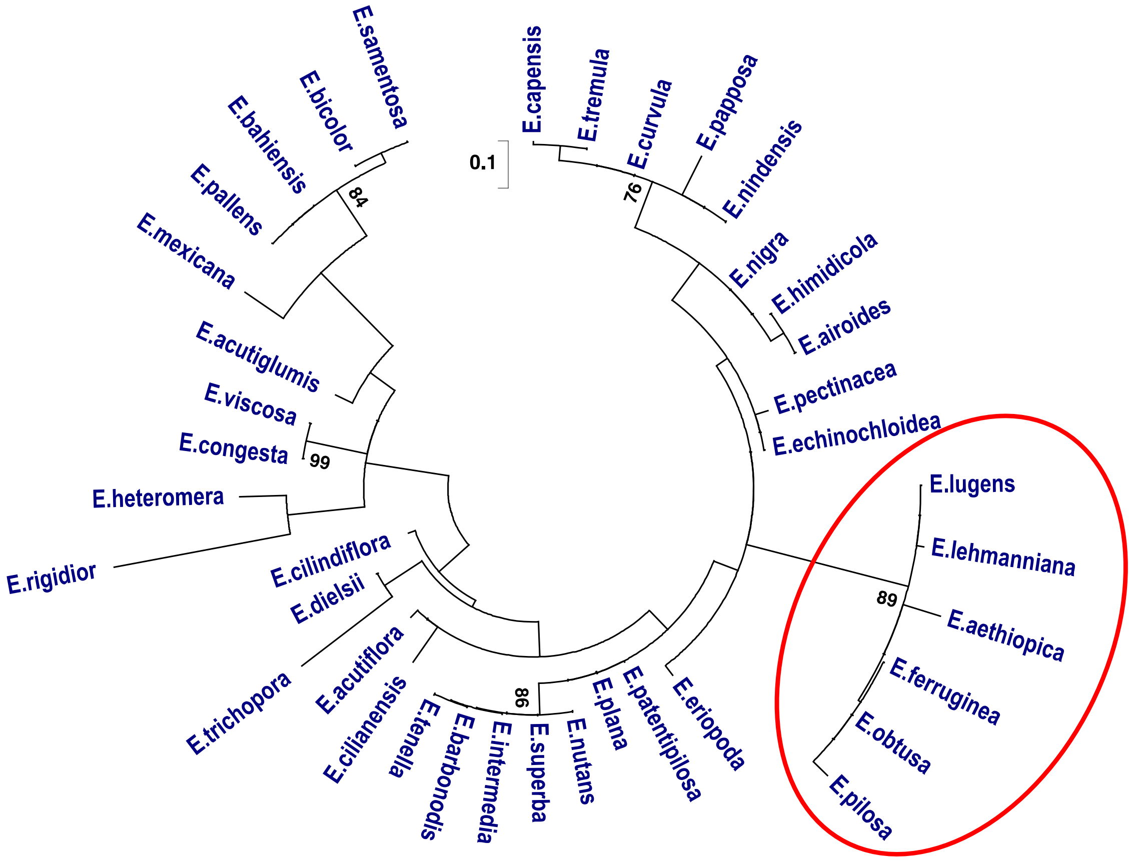


Figure S5. Phylogenetic tree of the wild *Eragrostis* species. Very little is known about the intraspecific relationships among the wild *Eragrostis* species, thus it was difficult to make sense of their grouping pattern displayed on the phylogenetic tree shown above. We subsampled these species and inferred a separate tree to visualize their relationships. The species that showed grouping with the tef species are clustered into one clade (red circle).

References

Ebba T (1975) Tef cultivars: morphology and classification. Addis Ababa University, College of Agriculture, Dire Dawa, Ethiopia

MoA (2014) Crop Variety Register Issue No. 15 Ministry of Agriculture, Animal and Plant Health Regulatory Directorate, Addis Ababa, Ethiopia

Tadele Z (2013) Tef Improvement Project: harnessing genetic and genomic tools to boost productivity. In: Assefa K, Chanyalew S, Tadele Z (eds) Achievements and Prospects of Tef Improvement. EIAR-University of Bern, Bern, Switzerland, pp 333-342

USDA (2015) Germplasm Resources Information Network. https://www.ars-grin.gov/. Accessed March 1, 2018