Supplementary material, S1
Descriptor list for the stem rust set

Predictive association between biotic stress traits and ecogeographic data for wheat and barley landraces

Dag Terje Filip Endresen, Kenneth Street, Michael Mackay, Abdallah Bari, Eddy De Pauw

Introduction

This is the descriptor list for the stem rust dataset (supplementary material, S1). The passport data for the genebank accessions (wheat landraces) is presented together with the trait measurements from the field trials, and the ecoclimatic data. This dataset includes a total of 6890 records (trait observations) for 4933 genebank accessions collected from a total of 2014 different collecting sites. Notice that the accession with catalog number PI 212925 is included in this dataset, but that this accession was removed for the data analysis experiments reported in the manuscript (identified as a very atypical outlier).

The passport data and the trait experiment data are available online from the USDA NPGS GRIN database (http://www.ars-grin.gov/cgi-bin/npgs/html/desc.pl?65049, verified 17 Dec 2010). Bonman et al. (2007) described the experimental design for the field trials. Susceptibility to stem rust (*Puccinia graminis* Pers. f.sp. *tritici*) was measured for six different years (during 1988-1994) at the USDA ARS (Agricultural Research Station) St Paul (44º59'17" N, 93º10'48 W) and Rosemount (44º43'01" N, 93º05'56" W) located in Minnesota in the northern USA. Dr Don V. McVey made all of the trait observations for both locations.

The ICARDA ecoclimatic layers were extracted from the ICARDA ecoclimatic information system (De Pauw, 2008). The layers are designed as 'climate surfaces' with a 30 arc-second resolution grid (approximately 1 km^2). The climate data for this study was extracted for each accession using the longitude and latitude coordinates of the original collecting site.

The WorldClim ecoclimatic layers were extracted from the online WorldClim dataset (http://www.worldclim.org; Hijmans et al., 2005). The WorldClim dataset is available in different spatial resolutions: 30 arc-seconds (approximately 1 km), 2.5 minutes (approximately 4.5 km), 5 minutes (approximately 9.3 km), and 10 minutes (approximately 18.5 km). Included to this dataset is the resolution of 2.5 minutes. The BIOCLIM layers were also extracted from the online WorldClim dataset (and with the 2.5 minutes resolution included in the dataset presented here).

General format rules
• The period (dot) "." is used as decimal mark (separator).
• A field for which no value is available is left empty (missing values).

References

Bonman, J.M., H.E. Bockelman, Y. Jin, R.J. Hijmans, and A.I.N. Gironella. 2007. Geographic Distribution of Stem Rust Resistance in Wheat Landraces. Crop Science 47:1955-1963. DOI: 10.2135/cropsci2007.01.0028

Hijmans, R.J., S.E. Cameron, J.L. Parra, P.G. Jones, and A. Jarvis. 2005. Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology 25:1965-1978. DOI: 10.1002/joc.1276.

De Pauw, E. 2008. Climatic and Soil Datasets for the ICARDA Wheat Genetic Resource Collections of the Eurasia Region. Explanatory Notes. ICARDA GIS Unit, Aleppo, Syria. 68 pages. Available online at http://geonet.icarda.cgiar.org/geonetwork/data/regional/GRU_NetBlotch/Doc/Report_NetBlotch.pdf (6.6 MB, verified 4 Dec 2010).

Descriptors (stem rust set)

Object ID number **(OID)**

The object ID number is a record number and is unique across the entire stem rust dataset for this experiment.

Example: `1`

Accession number **(ACCENUMB)**

This number serves as a unique identifier for accessions within a genebank collection (here the USDA GRIN NPGS). The accession number is a combination of the accession ID prefix and the accession ID number.

Example: `PI 430058`

Accession ID prefix **(ACP)**

Prefix for the accession catalog number used in the USDA NPGS GRIN system.

Example: `PI`

Accession ID number **(ACNO)**

This number serves as a unique identifier for accessions within a genebank collection (here the USDA GRIN NPGS).

Example: `430058`

Full scientific name **(TAXON)**

Full scientific name, binominal nomenclature name following the International Code of Botanical Nomenclature (CBN).

Example: `Triticum aestivum subsp. aestivum`

Longitude of collecting site (decimal) **(LONGITUDE)**

Longitude in decimal format for the collecting site of the source location where the landrace originates. Positive longitude values for the western hemisphere and negative values for the western hemisphere.

Example: `75.95`

Latitude of collecting site (decimal) **(LATITUDE)**

Latitude in decimal format for the collecting site of the source location where the landrace originates. Positive latitude values for the northern hemisphere and negative values for the southern hemisphere.

Example: `32.51667`

Elevation of collecting site **(ELEVATION)**

Elevation of collecting site expressed in meters above sea level. Negative values are allowed.

Example: `1362`

Collecting site code **(SITE_CODE)**

A code to identify the collecting site where the sample was originally collected.

Example: `NS_1162`

Country of origin **(COUNTRY)**

Country in which the sample was originally collected.

Example: `India`

State of collecting site **(STATE)**

State or administrative area in which the sample was originally collected.

Example: `Himachal Pradesh`

Location of collecting site **(LOCALITY)**

Location information below the country level that describes where the accession was collected. This might include the distance in kilometres or miles and direction from the nearest town, village or map grid reference point

Example: `Megzin, near Kunan`

Stem rust trait measurements **(STEM_RUST)**

The original stem rust trait score from the USDA NPGS GRIN dataset. The stem-rust trait ratings 0-3 (1915 landraces, 28% of the total) were considered as resistant to stem rust, ratings 4-6 (2729 landraces, 40%) as medium, and ratings 7-9 (2246 landraces, 32%) as susceptible. (Note that PI 212925 is included in these figures).

Values: `0, 1, 2, 3, 4, 5, 6, 7, 8, 9`

Stem rust trait measurements (reclassified to 9 categories) **(S9)**

Disease scores reclassified to 9 classes (0-1), 2, 3, 4, 5, 6, 7, 8, and 9. Stem rust ratings 0 and 1 on the

original measurement scale were combined as the reclassified class 1.

Values: `1, 2, 3, 4, 5, 6, 7, 8, 9`

Stem rust trait measurements (reclassified to 3 categories)	(S3)

Disease scores reclassified to 3 classes (0-3), (4-6), and (7-9). Stem rust ratings between 0 and 3 on the original measurement scale were combined as the reclassified class 1, ratings between 4 and 6 as class 2, and ratings between 7 and 9 as class 3.

Values: `1, 2, 3`

Stem rust trait measurements (reclassified to 2 categories)	(S2)

Disease scores reclassified to 2 classes (0-3), and (4-9). Stem rust ratings between 0 and 3 on the original measurement scale were combined as the reclassified class 1, ratings between 4 and 9 as class 0.

Values: `0, 1`

Trial experiment location	(E_SITE)

Location for the trait experiment (trial). This is the Agricultural Research Station (USDA ARS) were the trait measurement was made.

Values: `Rosemount, StPaul`

Trial experiment year	(E_YEAR)

Year of the trait experiment (trial). This is the year of the growth season when the trait measurement was made.

Values: `1988, 1989, 1991, 1992, 1993, 1994`

Separator column for the ecoclimatic layers	(ECO)

This is used as a single value alphanumeric column to separate the ecoclimatic layers from the passport and trait dataset.

Value: `eco`

Mean monthly precipitation (January to December)	(PREC01 - PREC12)

Ecoclimatic layers for the mean monthly precipitation from the ICARDA ecoclimatic information system. One layers for each month where PREC01 is for January, PREC02 is for February, PREC03 is for March, etc.

Example: `167.26`

Mean monthly minimum temperature (January to December)	(TMIN01 - TMIN12)

Ecoclimatic layers for the mean monthly minimum temperature from the ICARDA ecoclimatic information system. One layers for each month where TMIN01 is for January, TMIN02 is for February, TMIN03 is for March, etc.

Example: `1.38`

Mean monthly maximum temperature (January to December)	(TMAX01 - TMAX12)

Ecoclimatic layers for the mean monthly maximum temperature from the ICARDA ecoclimatic information system. One layers for each month where TMAX01 is for January, TMAX02 is for February, TMAX03 is for March, etc.

Example: `10.85`

Mean monthly potential evapotranspiration (January to December)	(PET01 - PET12)

Ecoclimatic layers for the mean monthly potential evapotranspiration from the ICARDA ecoclimatic information system. One layers for each month where PET01 is for January, PET02 is for February, PET03 is for March, etc.

Example: `37.95`

Mean monthly precipitation (January to December)	(PREC01_WC - PREC12_WC)

Ecoclimatic layers for the mean monthly precipitation from the online WorldClim ecoclimatic dataset. One layers for each month where PREC01_WC is for January, PREC02_WC is for February, PREC03_WC is for March, etc.

Example: `155`

Mean monthly minimum temperature (January to December)	(TMIN01_WC - TMIN12_WC)

Ecoclimatic layers for the mean monthly minimum temperature from the online WorldClim ecoclimatic dataset. One layers for each month where TMIN01_WC is for January, TMIN02_WC is for February, TMIN03_WC is for March, etc.

Example: `4.4`

Mean monthly maximum temperature (January to December)   (TMAX01_WC - TMAX12_WC)

Ecoclimatic layers for the mean monthly maximum temperature from the online WorldClim ecoclimatic dataset. One layers for each month where TMAX01_WC is for January, TMAX02_WC is for February, TMAX03_WC is for March, etc.

Example: `13.2`

BIOCLIM ecoclimatic layers (1 - 19)   (BIO1 - BIO19)

The BIOCLIM ecoclimatic layers available from the online WorldClim and also derived from the temparature and precipitation layers of this dataset.

1) BIO1 = Annual Mean Temperature
2) BIO2 = Mean Diurnal Range (Mean of monthly (maximum temperature - minimum temperature))
3) BIO3 = Isothermality (P2/P7) (* 100)
4) BIO4 = Temperature Seasonality (standard deviation *100)
5) BIO5 = Max Temperature of Warmest Month
6) BIO6 = Min Temperature of Coldest Month
7) BIO7 = Temperature Annual Range (P5-P6)
8) BIO8 = Mean Temperature of Wettest Quarter
9) BIO9 = Mean Temperature of Driest Quarter
10) BIO10 = Mean Temperature of Warmest Quarter
11) BIO11 = Mean Temperature of Coldest Quarter
12) BIO12 = Annual Precipitation
13) BIO13 = Precipitation of Wettest Month
14) BIO14 = Precipitation of Driest Month
15) BIO15 = Precipitation Seasonality (Coefficient of Variation)
16) BIO16 = Precipitation of Wettest Quarter
17) BIO17 = Precipitation of Driest Quarter
18) BIO18 = Precipitation of Warmest Quarter
19) BIO19 = Precipitation of Coldest Quarter

Source: http://www.worldclim.org/bioclim (verified 17 Dec 2010)