Estimating Genetic Diversity and Population Structure of Founder Lines in a Pennycress Breeding Program

Katherine Frels[1], Ratan Chopra[1], Donald Wyse[1], M. David Marks[2], James Anderson[1]
[1]Dept. of Agronomy and Plant Genetics, University of Minnesota, Saint Paul, MN
[2]Dept. of Plant and Microbial Biology, University of Minnesota, Saint Paul, MN

Introduction

Pennycress (*Thlaspi arvense* L.) is a winter annual oilseed species native to Eurasia. As part of the Forever Green Initiative at the University of Minnesota, breeders and geneticists are domesticating pennycress for use as a cash cover crop (Figure 1).

Pennycress has no breeding history and has many undesirable weedy traits such as uneven germination, small seed, and seedpod shatter prior to harvest. However, despite these negative traits, pennycress has many useful traits that make it a species amenable to rapid domestication including:

- High seed yields
- High oil content
- Self-pollinated
- Diploid (2n=14)
- Small genome size (539 Mbp)
- Draft genome published by Dorn et al. 2015[1]
- Close Arabidopsis relative

A pennycress breeding program was initiated at the University of Minnesota in 2013 with pennycress accessions collected by UMN researchers and from GRIN. These accessions have some phenotypic evaluation, but no genotypic characterization has been completed until now.

Figure 1. Pennycress life cycle as a cash cover crop. Pennycress is planted in the fall after silage corn or other crops, overwinters in rosette form, flowers in early spring and is harvested in June. Soybean can be planted in the pennycress stand shortly after flowering. Once the pennycress is harvested, the soybeans can grow as normal.

Objectives

- Evaluate the genetic variation present in the foundation breeding population.
  - Are all accessions genetically distinct?
  - Are sub-populations present?
- Identify parent lines that could be useful to the breeding program.

Materials

- 122 pennycress wild accessions collected from North America, Europe, and West Asia.
  - 40 collections made by UMN researchers
  - 7 collections made by IL researchers
  - 75 collections from GRIN
  - 39 spring annual, 74 winter annual, 5 unconfirmed accessions

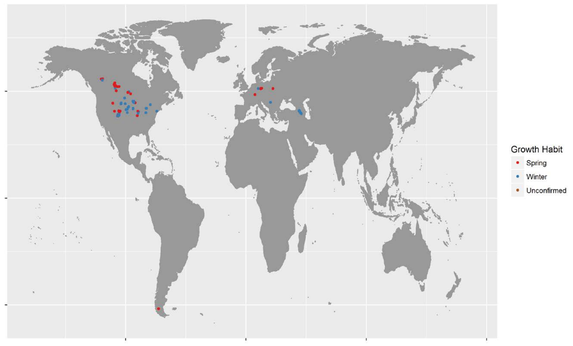


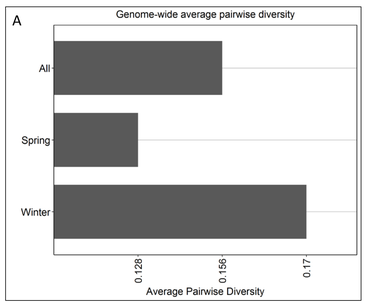
Figure 2. Locations of original collection sites.

Methods

- Sequencing
  - Genotyping-by-sequencing performed by BGI Inc.
    - DNA extracts digested with ApeKI
    - Libraries were multiplex sequenced on Illumina Hi-Seq 2000
    - Alignment to genome with SOAP[2]
    - 213,760 SNPs initially identified
- Filtering
  - SNPs further filter based on read depth: 56,248 SNPs
  - After MAF <4% removed= 24,020 SNPs
- Genome-wide pairwise genetic diversity
  - APE package in R[3]
    - Heterozygous or ambiguous SNPs treated as missing data
- Principal Components Analysis
  - Dataset further filtered to reduce redundant SNPs prior to PCA using PLINK v1.9[4]= 11,747 SNPs
  - PCA completed in TASSEL 5[5]

Results

Genome-wide pairwise genetic diversity



B

Line1	Origin1	Line2	Origin2	GenDist
TAMN107	CoatesMN	TAMN104	RosemountMN	0.021
TAMN108	SherburneMN	TAMN105	CoatesMN	0.021
TAMN111	NewYorkNY	TAMN104	RosemountMN	0.022
TAMN111	NewYorkNY	TAMN107	CoatesMN	0.022
TAMN129	Montana	TAMN109	RosemountMN	0.022
TAMN136	Montana	TAMN134	Montana	0.022
TAMN130	Montana	Ames_31016	CO	0.023
TAMN136	Montana	TAMN135	Montana	0.025
TAMN112	Illinois	Ames_30983	IllinoisGlennBrosFarm	0.026
Ames_31011	Briggsdale, CO	Ames_31010	Weld County, CO	0.027
TAMN124	RoseauMN	TAMN110	RosemountMN	0.027
TAMN128	Montana	Ames_31487	Harrow, Ontario, Can	0.028
TAMN117	SpringfieldIL	TAMN116	SpringfieldIL	0.028
TAMN132	Montana	TAMN131	Montana	0.028
TAMN135	Montana	TAMN134	Montana	0.028
TAMN121	RoseauMN	Ames_31495	Saskatchewan, Canad	0.029
TAMN136	Montana	Buf_1.A		0.029
Ames_30982	Iowa- Saylorville	Ames_29513	Iowa- PI Station	0.03
TAMN119	RoseauMN	Ames_30997	Ft. Collins, Colorado	0.03
Ames_32872	Armenia	Ames_32871	Armenia	0.03
Ames_32873	Armenia	Ames_32871	Armenia	0.03
TAMN134	Montana	Buf_1.A		0.03
Ames_32873	Armenia	Ames_32872	Armenia	0.032
Ames_32867	Armenia	Ames_29509	Ohio	0.033
TAMN135	Montana	Buf_1.A		0.034
TAMN119	RoseauMN	TAMN110	RosemountMN	0.034
Ames_32757	Loveland, Colora	Ames_31025	Sargents, Colorado	0.035
TAMN124	RoseauMN	TAMN119	RoseauMN	0.036
Ames_31001	Ft. Collins, Colora	Ames_31000	Ft. Collins, Colorado	0.037
TAMN123	RoseauMN	TAMN110	RosemountMN	0.038
TAMN119	RoseauMN	TAMN103	RosemountMN	0.039
TAMN124	RoseauMN	TAMN123	RoseauMN	0.039

Figure 3. Genome-wide pairwise diversity estimates reveal overall diversity in the foundation pennycress population as well as identifying accessions that are likely from selfed populations. A) Genome-wide pairwise diversity averaged across all accessions, spring, and winter accessions. B) Accessions with very low pairwise diversity suggesting these lines are from the same populations.

Results cont.

Principal components analysis

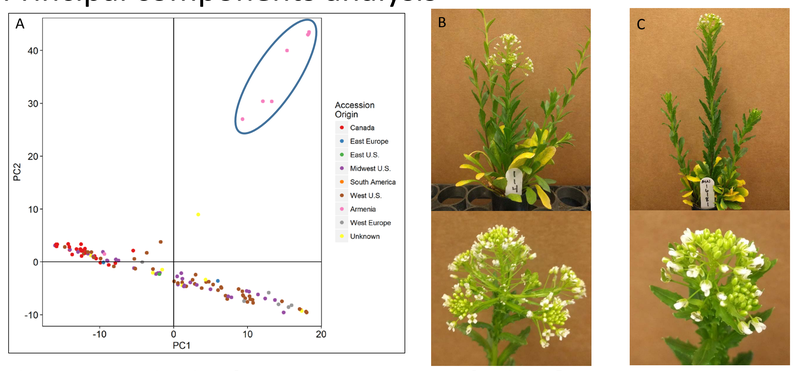


Figure 4. Principal components analysis of 122 pennycress collections. A) Seven pennycress accessions collected in Armenia are divergent from all others. Principal components 1 and 2 explain 25% of the cumulative variation.
B-C) Phenotypically the Armenian accessions (B) appear to be the same as other accessions (C). More work will be needed to determine if these lines are indeed *T. arvense*, a related species, or polyploid. If these accessions are *T. arvense*, they may be valuable resources for introducing genetic variation into the breeding program.

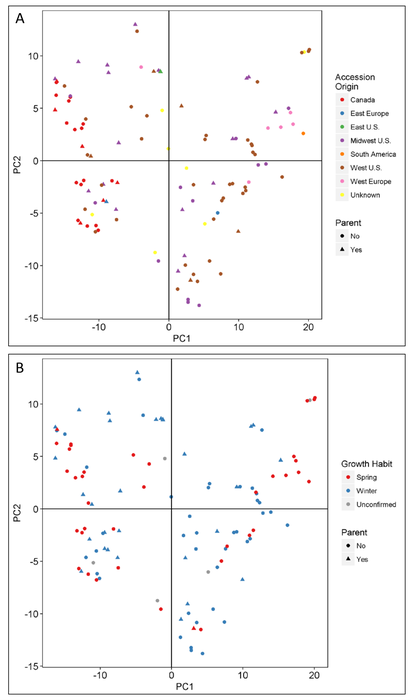


Figure 5. Principal components analysis of non-Armenian accessions. A) Without Armenian lines, PCs 1 and 2 explain 21% of the cumulative variation. Collections from similar geographic regions do not cluster together with the exception of Canadian and Western European accessions. Accessions used as parents in the breeding program represent most areas of the chart.
B) No obvious differentiation between spring and winter annual accessions.

Conclusions and Future Work

Conclusions

- Genome-wide diversity is low, but slightly higher in winter annual accessions.
- No evidence for spring vs winter sub populations beyond the difference in growth habit.
- Some evidence for geographic population differences particularly for Canadian and Armenian accessions.

Future work

- Determine the reason for genetic differences between Armenian accessions and all other accessions.
  - Tetraploid?
  - Different species? Possibly *Thlaspi caerulescens*?
- Identify geographic locations where more wild collections would be valuable.

Acknowledgements: Project funding was provided by Clean Water Funds through the Minnesota Department of Agriculture, and the Forever Green Initiative.

References: [1]Dorn et al. 2015. DNA Research 22.2: 121-131.; [2]Li et al. 2008. Bioinformatics 24.5: 713-714.; [3]Paradis et al. 2004. Bioinformatics 20.2: 289-290.; [4]Purcell et al. 2007. The American Journal of Human Genetics 81.3: 559-575.; [5]Glaubitz et al. 2014. PloS one 9.2: e90346.