Small-scale spatial genetic structure of an allotetraploid weed, *Capsella bursa-pastoris*

Byungwook Choi[1,P], Eunsuk Kim[1], and Sang-Tae Kim[2,C]

[1,P] School of Earth Sciences and Environmental Engineering, Gwangju Institute of Science and Technology, South Korea
[2,C] Department of Medical and Biological Sciences, The Catholic University of Korea, South Korea

[P] Presenter; wook1243@gm.gist.ac.kr | [C] Corresponding author; stkim@catholic.ac.kr

Introduction

***Capsella bursa-pastoris* (Shepherd's purse)**

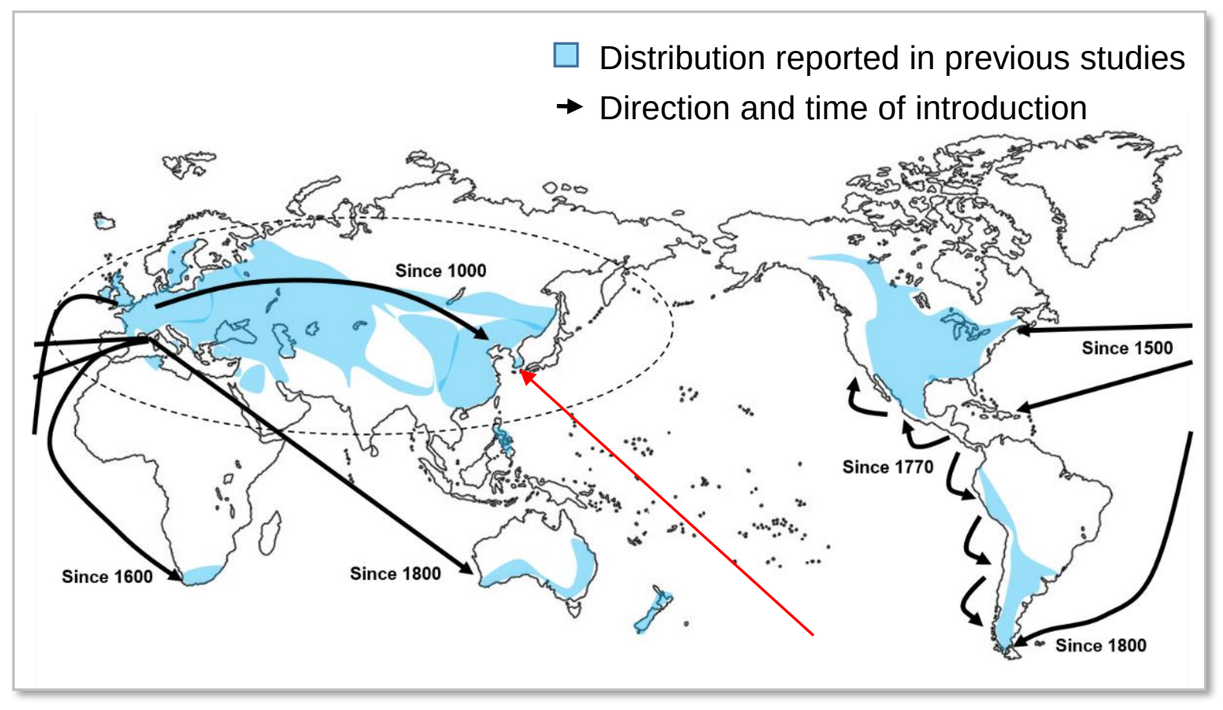
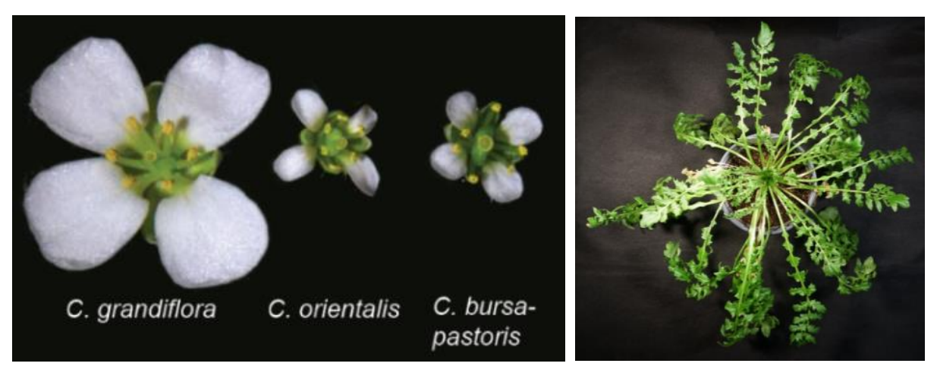


Fig 1. Distribution, introduction pathway and its year of *C. bursa-pastoris*



(adapted from Slotte et al. 2009)

- Globally distributed species
- Evolution by hybridization of two diploid species (Douglas *et al.* 2015)
- Evidence of colonization based on genetics and GBS (Slotte *et al.* 2008, Cornille *et al.* 2016, Kryvokhyzha *et al.* 2019)
- Geographical pattern of genetic (allozymic and isozymic) diversity (Wesse *et al.* 2020)
- Population genetics patterns at varying geographic scales (Bomblies *et al.* 2010)
- Overall patterns of IBD (Gene flows) and IBE (Local adaptation)

>> To unveil population genetics in local-scale geography

- RAD-seq; one of powerful tools to reveal genetic structure of populations

- Polyploidy species; a challenge with high heterozygosity
- Species without well-established reference genomes
- In *C. bursa-pastoris*; previous studies with partial genomes; *C. rubella* and *C. grandiflora* (Cornille *et al.* 2016)

- PolyRAD; a baysian genotyping approach (Clark *et al.* 2019), employed population genetics of autotetraploidy species *Alkanna tinctoria* (Ahmad *et al.* 2021)

>> Comparative analyses with and without reference genomes using a genotyper for polyploidy species

Research question and Purposes

"Can we demonstrate the genetic differentiation and structure of a globally distributed species within a small-scale landscape by identifying SNPs through sequencing and pipelines, especially for an allotetraploid species?"

1. To depict a globally distributed species' genetic differentiation and structure in a small-scale landscape.
2. To identify the factors contributing to genetic differentiation, such as geography and environment

Materials and Methods

Seed collections

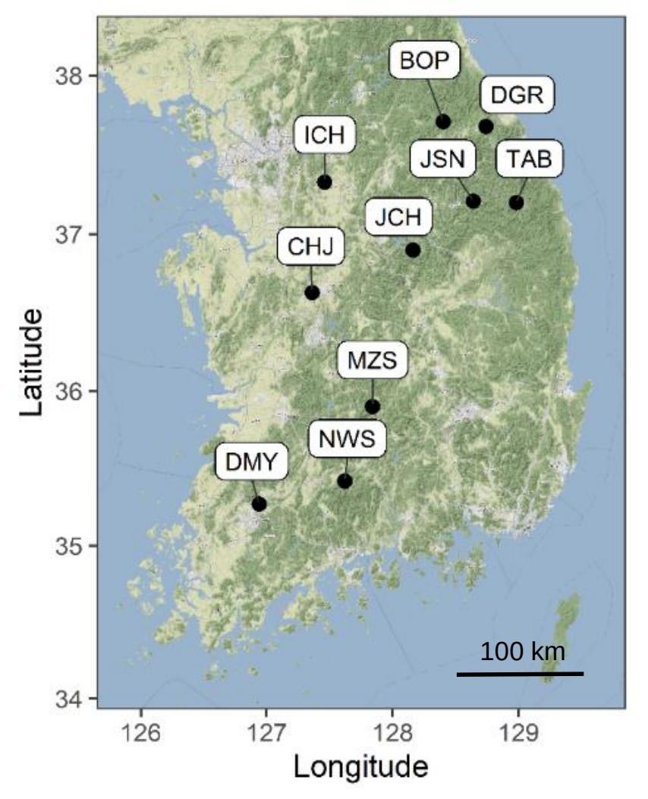
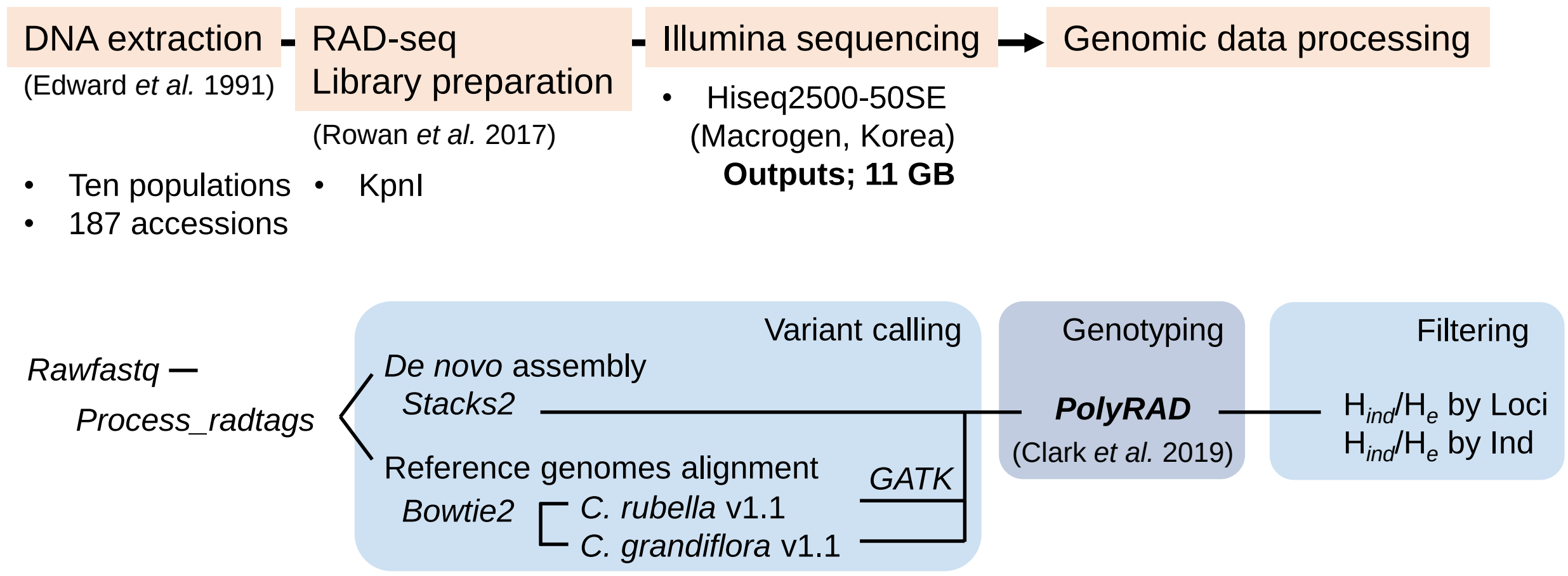


Fig 2. Map of seed collections (2018-2019)

Workflows



DNA extraction (Edward *et al.* 1991) – RAD-seq Library preparation (Rowan *et al.* 2017) – Illumina sequencing → Genomic data processing

- Ten populations
- 187 accessions

- KpnI

- Hiseq2500-50SE (Macrogen, Korea)

Outputs; 11 GB

Rawfastq – *Process_radtags*

Variant calling: *De novo* assembly – *Stacks2*; Reference genomes alignment – *Bowtie2* – *C. rubella* v1.1, *C. grandiflora* v1.1 – *GATK*

Genotyping: ***PolyRAD*** (Clark *et al.* 2019)

Filtering: H_{ind}/H_e by Loci, H_{ind}/H_e by Ind

Environmental variables

- Geographic information

Altitudes, Latitudes, and Elevation (GPS)

- Climatic variables

Public database from Automatic Weather Station (AWS) by Korea Meteoroloigcal Administration (KMA) was to set 19 Environmental Variables of WorldClim. Of them, six variables were employed to calculate environmental distances.

Statistical analyses

All statistical analyses were performed in R (functions and packages).

- Loading vcf dataset (*read.vcfR* in *vcfR*)
- Genetic diversity (Heterozygosity values) (*basic.stats* in *hierfstat*)
- Population differentiation (Fst (WC84) and DAPC for ordination by genotypes) (*mat_gen_dist* in *graph4lg* and *dapc* in *adegenet*)
- Analysis of Molecular Variance (AMOVA) (*poppr.amova* in *poppr* and *randtest* in *ade4*)
- Mantel's test for IBD (Isolation by Distance) and IBE (Isolation by Environment) (*mantel* in *vegan*)

>> Geographic distances by "Euclidean" and Climatic distances by "Gower"

- Multiple Matrix Regression with Randomization (MMRR or MDMR) (*lgrMMRR* in *PopGenReport*)
- Commonality analysis (CA) (*commonalityCoefficients* in *yhat*)

Results & Discussions

- The number of SNPs and heterozygosity indices were contingent upon the presence of reference genomes and the specific species used as the reference.

The number of SNPs

De novo assembly	*Raw*	*Genotyped*	*Filtered*
Ref.Gen_*C. rubella*	27,947	11,862	**1,754**
Ref.Gen_*C. grandiflora*	18,889	3,986	**142**
	17,600	3,802	**126**

Tab 1. Basic information after PolyRAD and its genetic heterozygosity indices

	A) *De novo*	B) *C. rubella*	C) *C. grandiflora*
Individuals	180	181	181
SNPs	1,754	142	126
Heterozygosity indices			
H_o	0.3467	0.3969	0.4147
H_e	0.2719	0.3362	0.3523
F_{IS}	-0.2750	-0.1808	-0.1776

- Despite minor variations in genetic differentiation among the tested populations arising from different methodologies, the overall level of the genetic differentiation remained low. Notably, only the NWS population exhibited relatively marginal genetic differentiation from the other populations ($F_{ST} < 0.15$).

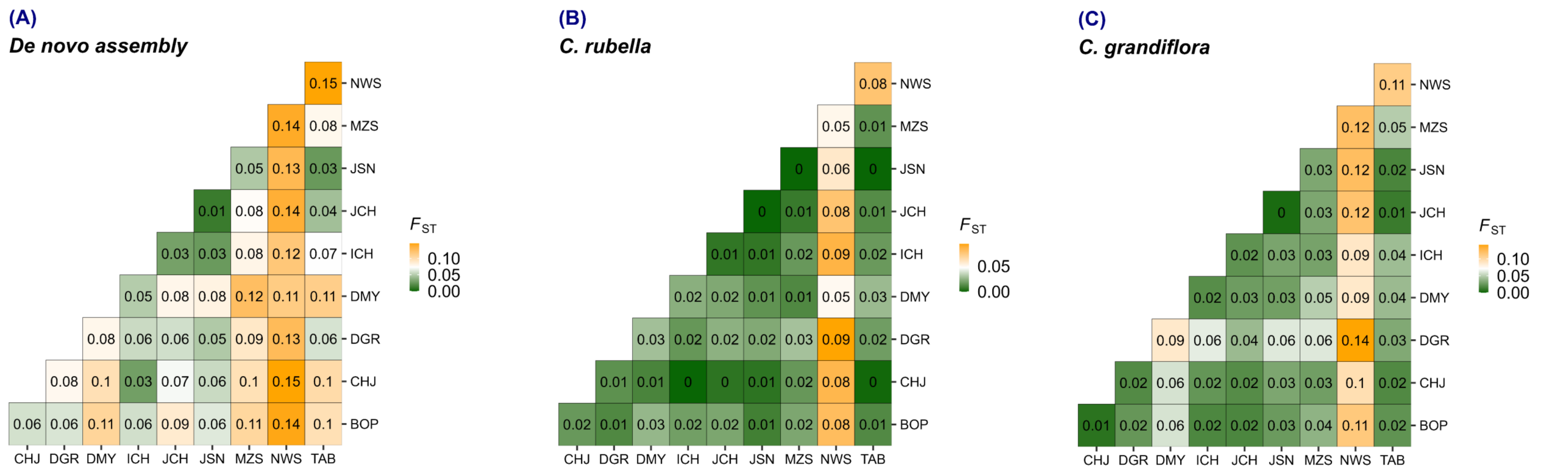


Fig 3. Heat-maps of pairwise FST between all testing populations

- The low genetic differentiation observed in these populations can be attributed to the close genetic similarity within the majority of genotypes (individuals within each population). This was statically demonstrated by the results of AMOVA, revealing a covariance component for differentiation that elucidates variances between populations (Between Samples) and within populations (within Samples).

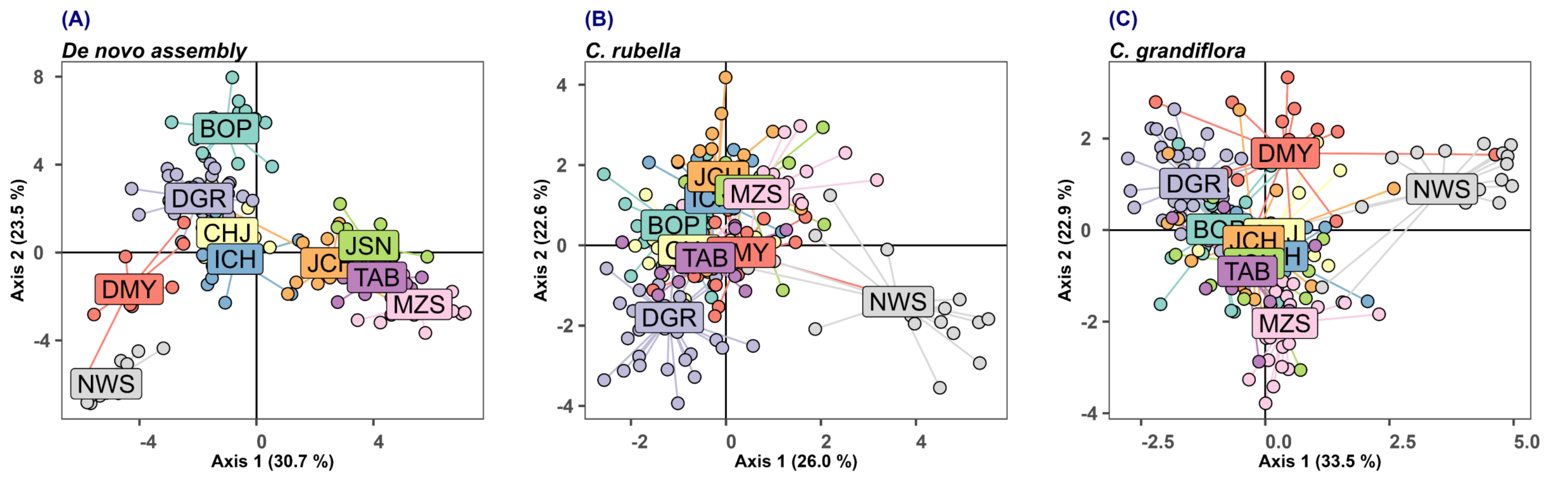


Fig 4. Discriminant Analysis of Principal Components (DAPC) scatter plots

- Model; Fst ~ Populations

Tab 2. Analysis of the Molecular Variance (AMOVA) table statistics

A	Df	Sigma	Percent
Between Samples	9	145.5321	30.91
Within Samples	170	325.3682	69.09
Total Variations	179	470.9004	100.00
			P < 0.001

B	Df	Sigma	Percent
Between Samples	9	4.1497	9.66
Within Samples	170	38.8106	90.34
Total Variations	179	42.9604	100.00
			P < 0.001

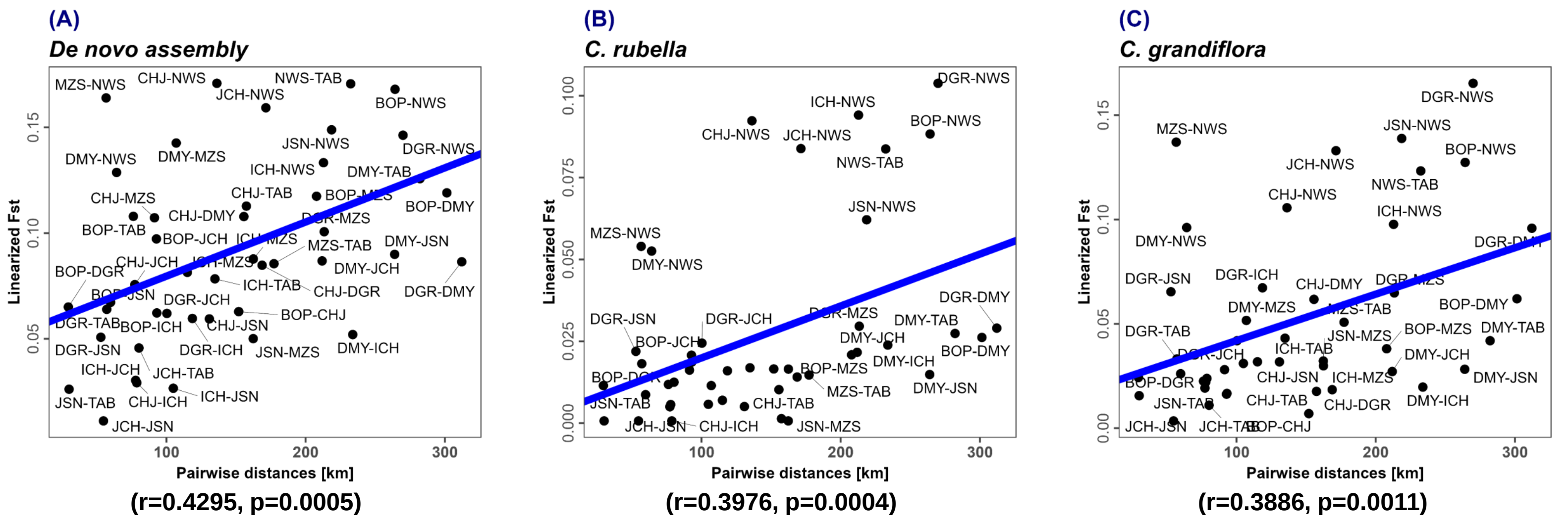
C	Df	Sigma	Percent
Between Samples	9	8.2914	18.07
Within Samples	170	37.5925	81.93
Total Variations	179	45.8839	100.00
			P < 0.001

- Although the outcomes of STRUCTURE analysis (using LEA packages for polyploidy species) on all individuals within each population were not presented here, the results indicated a substantial admixture composition and a high value for the ancestral population (K=6; De novo and K=10; Reference genomes)

- Selected environmental variables to calculate environmental distances (exclude high correlation coefficients (> 0.8))

BIO1; Annual Mean Temperature
BIO3; Isothermality (BIO2/BIO7 * 100)
 # BIO2; Mean Diurnal Range (Mean of monthly Temp.Max - Temp.Min)
 # BIO7; Temperature Annual Range (BIO5-BIO6)
BIO4; Temperature Seasonality (Standard deviation * 100)
BIO12; Annual precipitation
BIO14; Precipitation of Driest Month
BIO18; Precipitation of Warmest Quarter

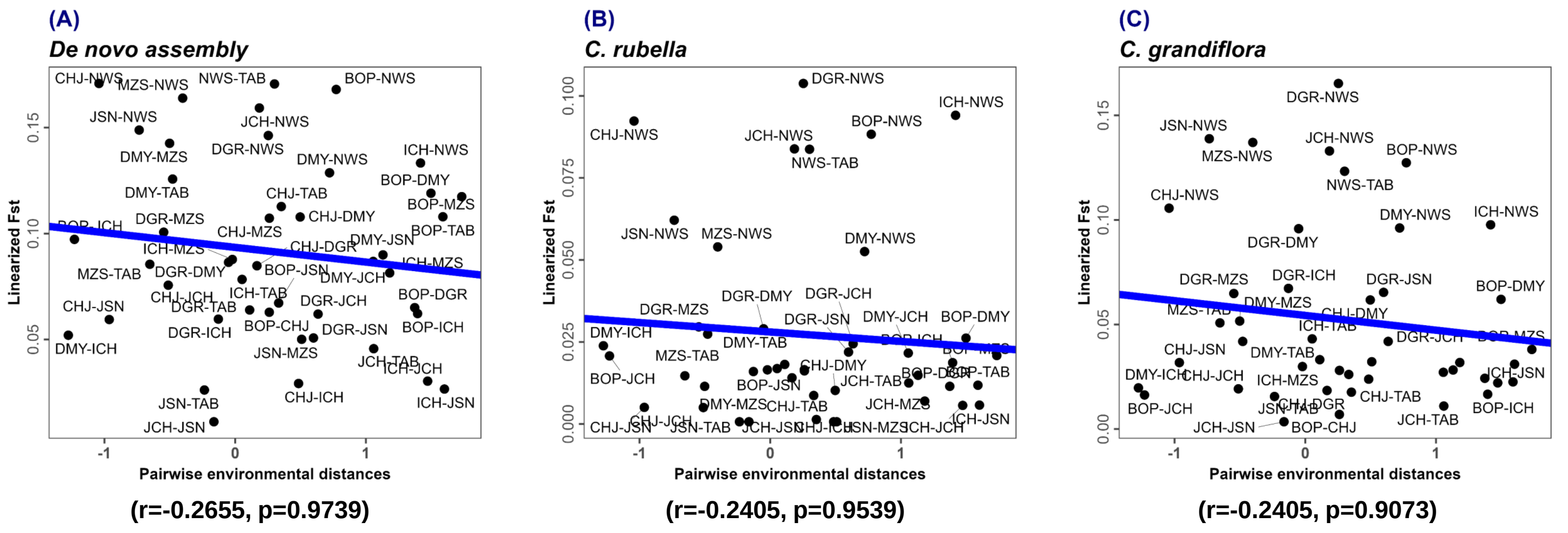
- With increases in geographic distances, genetic distances among populations also increased. This suggests that gene flows occur between geographically closer populations, contributing to the reduction of genetic differentiation.



(r=0.4295, p=0.0005) **(r=0.3976, p=0.0004)** **(r=0.3886, p=0.0011)**

Fig 5. Isolation-by-distance (IBD) plots with simple regression (blue) and partial Mantel's test results

- Nevertheless, there was no correlation between environmental distances and genetic distances, suggesting that the environment might not play a role in genetic differentiation among populations.



(r=-0.2655, p=0.9739) **(r=-0.2405, p=0.9539)** **(r=-0.2405, p=0.9073)**

Fig 6. Isolation-by-environment (IBE) plots with simple regression (blue) and partial Mantel's test results

- Model; Fst ~ Mat.GEO + Mat.Env

Tab 3. Results of Multiple Matrix Regression (MMRR) and Commonality analysis (CA)

A	layer	coefficient	t_{value}	$t_{p\text{-}value}$	F_{value}	$F_{p\text{-}value}$	r^2
	geography	**0.0183**	**3.058**	**0.0007**	**6.624324**	**3.00E-04**	**0.2398**
	Intercept	0.0801	16.225	0.0016			
	environment	**-0.0119**	**-1.748**	**0.0446**			

	Coefficient %	Total
Unique to GEO	**0.0682**	**80.34**
Unique to ENV	0.014	16.53
Common to GEO and ENV	0.0027	3.13
Total	0.0849	100

B	layer	coefficient	t_{value}	$t_{p\text{-}value}$	F_{value}	$F_{p\text{-}value}$	r^2
	geography	**0.0123**	**2.768**	**0.0002**	**5.35676**	**2.00E-04**	**0.20324**
	environment	-0.0078	-1.542	0.0735			
	Intercept	0.0242	6.585	0.168			

	Coefficient %	Total
Unique to GEO	**0.0527**	**88.07**
Unique to ENV	0.0056	9.42
Common to GEO and ENV	0.0015	2.51
Total	0.0599	100

C	layer	coefficient	t_{value}	$t_{p\text{-}value}$	F_{value}	$F_{p\text{-}value}$	r^2
	geography	**0.0169**	**2.686**	**0.0016**	**4.68199**	**0.001**	**0.18231**
	Intercept	0.0459	8.829	0.1122			
	environment	-0.0091	-1.268	0.1582			

	Coefficient %	Total
Unique to GEO	**0.0227**	**55.3**
Unique to ENV	0.0167	40.68
Common to GEO and ENV	0.0017	4.03
Total	0.041	100

References

- Bomblies, K., Yant, L., Laitinen, R. A., Kim, S. T., Hollister, J. D., Warthmann, N., ... & Weigel, D. (2010). PLoS Genetics, 6(3), e1000890.
- Clark, L. V., Lipka, A. E., & Sacks, E. J. (2019). *G3: Genes, Genomes, Genetics*, 9(3), 663-673.
- Cornille, A., Salcedo, A., Kryvokhyzha, D., Glémin, S., Holm, K., Wright, S. I., & Lascoux, M. (2016). *Molecular ecology*, *25*(2), 616-629.
- Douglas, G. M., Gos, G., Steige, K. A., Salcedo, A., Holm, K., Josephs, E. B., ... & Wright, S. I. (2015). *Proceedings of the National Academy of Sciences*, *112*(9), 2806-2811.
- Edwards, K., Johnstone, C., & Thompson, C. (1991). Nucleic acids research, 19(6), 1349.
- Kryvokhyzha, D., Salcedo, A., Eriksson, M. C., Duan, T., Tawari, N., Chen, J., ... & Lascoux, M. (2019). *PLoS genetics*, *15*(2), e1007949.
- Rowan, B. A., Seymour, D. K., Chae, E., Lundberg, D. S., & Weigel, D. (2017). *Genotyping: Methods and Protocols*, 221-242.
- Slotte, T., Huang, H., Lascoux, M., & Ceplitis, A. (2008). *Molecular Biology and Evolution*, *25*(7), 1472-1481.

Acknowledgments

This work was supported by AI-based GIST Research Scientist Project grant funded by the GIST (Institute of Artificial Intelligence) in 2023.

Summary

- Low genetic differentiation between populations in the small-scale landscape

SNPs or heterozygosity values varied depending on the methods. However, the differentiation between populations was determined by RADseq – AMOVA / STRUCTURE.
The low genetic differentiation stems from the genetic structure of individuals between populations.

- On the small-scale, geographic distances are a vital factor in determining genetic differentiation, like on a global scale, but environmental distances are a minor factor in affecting genetic variation

Isolation by distances; significant correlation and matrix regression can explain their genetic differentiation in a small-scale Isolation by environments; no effects on the differentiation Since environmental gradients on a small-scale
c.f.) significant IBE in global range study (Wesse et al. 2021)

- Though it is a worldwide species, they have a low genetic differentiation within a small-scale landscape. It also suggests that local population genetic studies are critical to understanding the natural populations in certain regions.
- We still wonder whether diverse possible scenarios or various factors to influence genetic differentiation and sequential events can happen.