

# Supplementary materials

Here, you will find the description of the different files used in the analysis for each step of the process. This file was retrieved from [Figshare](#).

## Contents

Supplementary 1 .....	2
Supplementary 2 .....	3
Supplementary 3 .....	4
Supplementary 4 .....	5
Supplementary 5 .....	6
Supplementary 6 .....	7
References .....	8

# Supplementary 1

**S1:** This directory contains all the wrangled data from formatted original files to subsets of the formatted files.

**S1.1:** This directory contains CSV files of the different wrangled data frames of the Original downloaded files from Daniëls et al. (2013). In the thesis, these files are referred to as the “Arctic plant Inventory List”.

**S1.1.1:** The formatted version of the original Daniëls et al. (2013). The headers 1 to 6 are the full taxonomical names, 7 to 27 are the floristic province codes, 28 to 32 are the Arctic subzones A to E, while header 33, “N”, are distributions outside of the Arctic in the neighboring boreal or boreal-alpine zone. Headers 34 to 41 are descriptions of a species' status. Header 42, “paf”, refers to the Pan-Arctic Flora code, and “genusCount” and “familyCount” in headers 43 and 44 refer to the number of genera and families.

**S1.1.2:** List of 136 species absent from the Arctic based on S1.1.1. This list was created by subsetting species with the codes: “-”, “?”, “b”, “\*\*\*”. This subset borderline species, absent species, uncertain species, and casual species (Daniëls et al., 2013).

**S1.1.3:** List of 2,082 species present in the Arctic based on S1.1.1. This list was created by subsetting species with the codes: “r/R”, “s/S”, “f/F”, “\*”. This subset species with the status rare, scattered, frequent, and stabilized introductions (Daniëls et al., 2013).

**S1.2:** This directory contains CSV files for the different wrangled data frames of the Original downloaded files from Wasowicz et al. (2020). In the thesis, these files are referred to as the “Arctic Alien Plant List”.

**S1.2.1:** The formatted version of the original Wasowicz et al. (2020). Headers 2 to 24 are codes for the floristic provinces in the Arctic, as used in the original study.

**S1.2.2:** List of 166 species absent from the Arctic based on S1.2.1. This list was created by subsetting species with the codes: “o”, “?”, “x”. This subset casual aliens, uncertain species, and excluded although present on previous lists (Wasowicz et al., 2020).

**S1.2.3:** List of 169 species absent from the Arctic based on S1.2.1. This list was created by subsetting species with the codes: “•”, “IT”, “IR”. This subset naturalized aliens, ‘Invasive transformer, and Invasive Rapidly spreading species (Wasowicz et al., 2020).

**S1.3:** The directory contains the CSV file of the output from the wrangled original GloNAF database from Kleunen et al. (2019).

**S1.3.1:** List of 13,939 species with unknown Arctic presence, subset from the original GloNAF database (Kleunen et al., 2019), with the codes “alien” and “naturalized”.

## Supplementary 2

**S2:** This directory contains CSV files of the different outputs from the `WFO.match` and `WFO.one` functions from the `WFO` package in R (Kindt, 2020). `WFO.match` outputs species names and corresponding synonyms, while `WFO.one` chooses the best match among all of these.

**S2.1:** This directory contains CSV files for the Arctic Plant Inventory List.

**S2.1.1:** All columns are returns of the `WFO.match` function for the Inventory Arctic Absent List except columns 1 and 2, which are the species from the wrangled data (S1.1.2). The function returned 161 scientific names.

**S2.1.2:** All columns are returns of the `WFO.match` function for the Inventory Arctic Present List except columns 1 and 2, which are the species from the wrangled data (S1.1.3). The function returned 2,491 scientific names.

**S2.1.3:** All columns are returns of the `WFO.one` function for the output from `WFO.match` in S2.1.1, which corresponds to the Inventory Arctic Absent List. The function returned 135 scientific names.

**S2.1.4:** All columns are returns of the `WFO.one` function for the output from `WFO.match` in S2.1.2, which corresponds to the Inventory Arctic Present List. The function returned 1,936 scientific names.

**S2.2:** This directory contains CSV files for the Arctic Alien Plant List.

**S2.2.1:** All columns are returns of the `WFO.match` function for the Absent Arctic Aliens except columns 1 and 2, which are the species from the wrangled data (S1.2.2). The function returned 269 scientific names.

**S2.2.2:** All columns are returns of the `WFO.match` function for the Present Arctic Aliens except columns 1 and 2, which are the species from the wrangled data (S1.2.3). The function returned 244 scientific names.

**S2.2.3:** All columns are returns of the `WFO.one` function for the output from `WFO.match` in S2.2.1, which corresponds to the Arctic Absent Aliens. The function returned 164 scientific names.

**S2.2.4:** All columns are returns of the `WFO.one` function for the output from `WFO.match` in S2.2.2, which corresponds to the Present Arctic Aliens. The function returned 164 scientific names.

**S2.3:** This directory contains CSV files for the GloNAF species list.

**S2.3.1:** All columns are returns of the `WFO.match` function for the GloNAF list of species except columns 1 and 2, which are the species from the wrangled data (S1.3.1). The function returned 17,381 scientific names.

**S2.3.2:** All columns are returns of the `WFO.one` function for the output from `WFO.match` in S2.3.1, corresponding to the GloNAF species list. The function returned 13,433 scientific names.

## Supplementary 3

**S3:** This directory contains CSV files of the filtered lists of standardized scientific names, in addition to species usage keys used in the occurrence download from the Global Biodiversity Information Facility (Derived dataset GBIF.org, 2024). Moreover, this includes the CSV files of the species excluded before the climatic analysis because of occurrences less or equal to 55.

**S3.1:** This directory contains CSV files for the Filtered Arctic Present and Filtered Arctic Absent species lists.

**S3.1.1:** This file contains a list of 1,988 scientific names from the Filtered Arctic Present, which is a combination of unique scientific names from Inventory Arctic Present Standardized and Present Arctic Aliens Standardized.

**S3.1.2:** This file contains a list of 229 scientific names from the Filtered Arctic Absent, which is a combination of unique scientific names from Inventory Arctic Absent Standardized and Absent Arctic Aliens Standardized.

**S3.2:** This directory contains CSV files for the Filtered GloNAF Arctic Present and Filtered GloNAF Arctic Absent species lists.

**S3.2.1:** This file contains a list of 475 scientific names of species that appeared in both the Filtered Arctic Present list and the GloNAF Standardized list.

**S3.2.2:** This file contains a list of 12,838 scientific names from the Filtered GloNAF Arctic Absent list, which was made by removing species from the Filtered Arctic Present and Filtered Arctic Absent lists.

**S3.2.3:** This file contains a list of 12,826 species with corresponding usage keys collected with the `name_backbone` function in `rgbif` (Chamberlain et al., 2024). Twelve species returned without any usage key and were excluded from the climate analysis. See List A1 in the thesis for the full list of the excluded species.

**S3.3:** This directory contains the CSV file of the excluded species with occurrences less than or equal to 55.

**S3.3.1:** This file contains 4,050 species that were excluded before the climate analysis. It includes information on the number of observations or occurrences each species included, the logarithm, and the number of dimensions. Absolute median latitude is also included for each species.

## **Supplementary 4**

**S4:** This directory contains the CSV file of the correlation matrix for the 19 bioclimatic variables in the Arctic.

**S4.1:** This CSV file contains the Pearson correlation coefficient for the Arctic's 19 different bioclimatic variables (bio\_n). The study utilized  $|r_p| < 0.5$  and selected bio\_18 (precipitation of the warmest quarter), bio\_10 (mean temperature of the warmest quarter), bio\_3 (isothermality), and bio\_4 (and temperature seasonality). For a visual representation, see Figure A1 in the thesis.

## Supplementary 5

**S5:** This directory contains the CSV files of the outputs from climatic niche space analysis. It includes a cleaned scientific name labeled “clean name,” the iteration number from the climate analysis, the number of observations, and dimensions. It includes information on the samples per point, and random points calculated by the hypervolume function in R, in addition to the result of the hypervolume\_overlap\_statistics results (Blonder et al., 2023). This returned the “jaccard”, “sorensen”, “fracVolumeSpecies”, and “fracVolumeRegion”. “fracVolume-” refers to the unique fraction of the species and the Arctic region and describes how much of the volume is unique to each.

The “realizedNiche” describes the portion of how much of the species’ niche space is that is found within the Arctic, and the “overlapRegion” describes the extent to which the species’ climate niche overlaps with the climatic conditions in the Arctic. “IncludedOverlap” describes the simplest and fastest overlap statistic used as a first exclusion step. Taxonomic names are included in the CSV files as well as country of origin and country codes. ‘meanLong’ and ‘meanLat’ are provided to understand the geographical distribution of the species, referring to the mean longitude and latitude for each species in each country. This implies that the mean was calculated from the occurrence data by each country's mean longitude and latitude. Exclusion status is also included in the CSV files.

**S5.1:** This CSV file contains the data explained in S4 for the 2,318 species included for post-analysis processing.

**S5.2:** This CSV file contains the data explained in S4 for the 5,541 species excluded during the climate analysis process.

## Supplementary 6

**S6:** This directory contains the CSV files, each of which holds crucial data from the post-analysis process. These files are particularly significant as they provide insights into the species included after the climate niche space analysis.

**S6.1:** This CSV file explains the potential area of occupancy of the included species. Here, “TPAoO” refers to Total Potential Area of Occupancy, “PAoO” refers to the Potential Area of Occupancy in each floristic province, and “prop-“ refers to the proportion of the respective unit for each species. Thus “propPAoO” refers to the proportion of the potential area of occupancy in each floristic province and “propTPAoO” refers to the proportion of the Arctic. This was used to create Figure 6 in the thesis.

**S6.2:** This CSV file explains the potential probability density of the included species. It is separated into total (the entire Arctic) and each region (floristic province), calculated using different methods of min, mean, median, and max probability densities. This was used to create Figure 7 in the thesis.

**S6.3:** This CSV file explains the potential relative order and group richness of the included species. Here, relative richness refers to the relative richness of each order in each floristic province, while relative group richness refers to the relative richness of each group in each floristic province. This was used to create Figure 8 in the thesis.

**S6.4:** This CSV file explains the potential connections for each included species. Here, connections refer to the count of a unique combination of species, floristic province, and origin country. This was used to create Figure 9 in the thesis.

**S6.5:** This CSV file provides the potential climatic overlap for each included species and its respective latitude native range. This was used to create the thesis's linear regression model and Figure 10.

## References

- Blonder, B., Morrow, with contributions from C. B., Brown, S., Butruille, G., Chen, D., Laini, A., & Harris, D. J. (2023). *hypervolume: High Dimensional Geometry, Set Operations, Projection, and Inference Using Kernel Density Estimation, Support Vector Machines, and Convex Hulls*. <https://CRAN.R-project.org/package=hypervolume>
- Chamberlain, S., Barve, V., Mcglinn, D., Oldoni, D., Desmet, P., Geffert, L., & Ram, K. (2024). *rgbif: Interface to the Global Biodiversity Information Facility API*. <https://CRAN.R-project.org/package=rgbif>
- Daniëls, F. J. A., Gillespie, L. J., & Poulin, M. (Eds.). (2013). Appendix 9.1. In *Arctic Biodiversity Assessment: Status and trends in Arctic biodiversity* (p. 35). The Conservation of Arctic Flora and Fauna.
- Derived dataset GBIF.org. (2024). *Filtered export of GBIF occurrence data* [dataset]. <https://doi.org/10.15468/dd.3tkkeb>
- Kindt, R. (2020). WorldFlora: An R package for exact and fuzzy matching of plant names against the World Flora Online taxonomic backbone data. *Applications in Plant Sciences*, 8(9). <https://doi.org/10.1002/aps3.11388>
- Kleunen, M., Pyšek, P., Dawson, W., Essl, F., Kreft, H., Pergl, J., Weigelt, P., Stein, A., Dullinger, S., König, C., Lenzner, B., Maurel, N., Moser, D., Seebens, H., Kartesz, J., Nishino, M., Aleksanyan, A., Ansong, M., Antonova, L. A., ... Winter, M. (2019). The Global Naturalized Alien Flora (GloNAF) database. *Ecology*, 100(1). <https://doi.org/10.1002/ecy.2542>
- Wasowicz, P., Sennikov, A. N., Westergaard, K. B., Spellman, K., Carlson, M., Gillespie, L. J., Saarela, J. M., Seefeldt, S. S., Bennett, B., Bay, C., Ickert-Bond, S., & Väre, H. (2020). Non-native vascular flora of the Arctic: Taxonomic richness, distribution and pathways. *Ambio*, 49(3), 693–703. <https://doi.org/10.1007/s13280-019-01296-6>



